Coronavirus Nonstructural Protein Interactions with
Endoplasmic Reticulum Proteostasis Factors Mediate Infection

By

Jonathan Patrick Davies

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

May 10, 2024

Nashville, Tennessee

Approved:

Todd Graham, Ph.D.

Jared Nordman, Ph.D.

Kristen Ogden, Ph.D.

Manuel Ascano, Jr., Ph.D.

Lars Plate, Ph.D.

To my loving partner, Esther

ACKNOWLEDGMENTS

It takes a village to raise a graduate student and I could not have completed this Ph.D. without the help of many, many people.

I would like to thank my thesis committee members, Drs. Todd Graham (chair), Jared Nordman, Kristen Ogden, and Manuel Ascano, Jr. for their invaluable guidance, feedback, and advice in matters related both to my thesis work as well as in professional development. Their perceptive questions and constructive criticism pushed me to think critically about my work and push forward with intention.

An enormous thank you to Dr. Mark Denison and Xiaotao Lu for their generosity in providing reagents, protocols, and expert advice formed over decades of coronavirus research which enabled a large portion of this thesis.

I am fortunate to have many past mentors who have helped me develop my passion for science. Thank you to the instructors and staff at the School for Science and Math at Vanderbilt, including Mrs. Amanda Dixon, Drs. Angela Eeds, Chris Vanags, and Jonathan Creamer, who have invested incredible amounts of energy into nourishing the scientific interests of Nashville high school students. Thank you to my undergraduate research advisor, Dr. Jason Sello, and summer research advisors Dr. Borden Lacy and Dr. Heather Kroh, whose mentorship was instrumental in fostering my desire to go to graduate school and have written many a letter of recommendation on my behalf.

My interests in global health and virology are thanks in large part to the international partners who have taken the time to host and work with me. Thank you to Dr. Marisa Romero and the Food Science team at the Philippines Rice Research Institute, who hosted me in the year prior to graduate school. Thank you to Dr. Claudio Lanata, Dr. Luis Franchi, Ana Gil, Mayra Ochoa, Bia Peña, Alejandra Ingunza, and the rest of the Bioequipo team at the Instituto de Investigación Nutricional in Lima, Peru, for allowing me to learn from them for two months and for our continued collaboration. That internship would not have been possible without the initiative and generous support of the Vanderbilt Institute of Infection, Immunology and Inflammation (VI4) and the Vanderbilt Institute of Global Health (VIGH), and in particular Drs. Eric Skaar, Muktar Aliyu, Carlos Grijalva, and Leigh Howard.

I have spent the majority of my waking hours in this PhD with my fellow lab mates in the Plate Lab. They have made lab a fun and inspiring environment with their conversations and comradery. They are both interesting people and sharp scientists who tolerate my dad jokes, and I could not have asked for better colleagues. In particular, a huge thank you to Katherine Almasy, who was my rotation and virology mentor and who worked jointly with me on much of the research herein. Also thank you to my undergraduate/high school mentees, Joyce Kariuki, Athira Sivadas, Lily Kim, and Andrew Spencer, whose curiosity and enthusiasm for science continually reinvigorated my own.

A PhD advisor can make or break the graduate school experience, and I have been fortunate enough to have Lars Plate as mine. I am immensely grateful for the mentorship, guidance, and insight he has provided over the last five years. From teaching hands-on mass spectrometry techniques to discussing project strategy to supporting professional development and allowing me to take ownership of my PhD path, he has helped me grow into a better scientist and thinker.

Research can be rough sometimes, but having a quality community to support you makes all the difference. Thank you to my friends, both in and outside of Vanderbilt, who have humored my obsession with viruses and reminded me of other things in life when research got tough.

My family have been pillars of support throughout graduate school and I am so, so lucky to have them nearby. From Sunday dinners to hikes together, they have provided me with so much love and care. Thank you to my mom, Anne Marie Davies, for always dropping off home-cooked meals to me in lab and making sure I am doing okay. Thank you to my dad, Dr. Sean Davies, whose own passion for science has always inspired mine. Our chats about experiments and grants around the dinner table, while perhaps less enthralling for the rest of the family, have always been a highlight of my day.

Last, but most certainly not least, I would like to thank my loving partner, Esther Kang, who has lived this PhD every step of the way with me. I am in awe of her unending patience with late-night checks on the mass spec, Western blot lab dates, and the general chaos of grant/paper/qualifying exam/dissertation writing. From bringing dinner to lab to listening to me vent about yet another failed experiment, she has provided the love and support I needed to keep going. She has celebrated every success and commiserated every disappointment the PhD has dealt, and I truly could not have done this without her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1   Coronavirus Biology

1.1.1   Emergence of SARS-CoV-2 and the COVID-19 pandemic

In late 2019, a novel coronavirus (CoV) emerged in Wuhan, China, causing severe acute respiratory illness and death in some cases (1,2). By early 2020, the virus had made its way across the globe and the World Health Organization (WHO) declared the outbreak a pandemic on March 11th, 2020 (3). The disease was named Coronavirus Disease 2019 (COVID-19, (4)) and the virus, after an initial title of 2019 novel CoV (2019-nCoV), was named SARS-CoV-2 (also referred to as SARS-2 within this text) (5). Symptoms included fever, chills, difficulty breathing, and in severe cases, pneumonia, and respiratory failure (6). As of January 2024, the current human toll stands at over 773 million reported cases and close to 7 million deaths (7). In the ensuing weeks, months, and years after the emergence of SARS-CoV-2, the medical and scientific community has conducted swift and concentrated research on the virus and COVID-19, pursuing therapeutic strategies and a better understanding of how this new virus works. This dissertation is a small piece of these efforts, initiated in late February of 2020 after conservations between myself, my advisor, Lars Plate, and a fellow graduate student, Katherine Almasy, on how we might contribute to the larger CoV research endeavor. We began our work just as the university and the city implemented shut-down measures to limit the outbreak spread, setting out to study the mechanisms by which SARS-CoV-2 and other CoVs hijack a host cell for infection.

SARS-CoV-2 is the third highly pathogenic beta-CoV to emerge in the modern era. In late 2002, an outbreak of SARS-CoV (also referred to as SARS-CoV-1 or SARS-1 within this text) occurred in Guangdong Province, China and spread to Hong Kong, Singapore, Canada, and Vietnam (8,9). The outbreak ended in 2003 with 8,098 total cases and 774 deaths (10). A decade later, Middle Eastern Respiratory Syndrome- CoV (MERS-CoV; also referred to as MERS within this text) broke out in Saudi Arabia in 2012 (11). To date, there have been 2,617 confirmed cases and 947 deaths (12). MERS-CoV spilled over from bats to camels and eventually to humans (13,14). SARS-CoV also emerged from bats (15,16) and there is evidence SARS-CoV-2 is likely to have bat origins as well (17–20). Further outbreaks of highly pathogenic CoVs are likely, due to the large circulating pool of animal CoVs with spill-over potential, viral recombination events, and increasing contact between wildlife and human populations (16,21,22). As such, continued study of this virus family is imperative for preparing for future outbreaks.

Besides the three severely pathogenic CoVs, four human CoVs (hCoVs) associated with the common cold have been characterized, including the alpha-CoVs hCoV-229E (23) and hCoV-NL63 (24), and the beta-CoVs hCoV-OC43 (25), and hCoV-HKU1 (26). The range of pathogenicity of human CoVs is notable and continued research is needed to understand the molecular mechanisms governing these differential outcomes between viruses.

Additionally, there are a wide array of well-studied animal CoVs that have proven useful models for elucidating fundamental principles of CoV biology. These include avian infectious bronchitis virus (IBV) (27) and murine hepatitis

virus (MHV) (28). MHV in particular has served as a workhorse infection model, due to its ease of handling and permissibility for cell culture (29,30).

1.1.2 Role of nonstructural proteins (nsps) in CoV infection

1.1.2.1 CoV replication cycle

CoVs subvert host cell functions to replicate and produce progeny virus. The events and molecular mechanisms mediating infection have been studied extensively and are comprehensively reviewed by Perlman and Netland (31) and V'kovski et al. (32), among many others. Herein, I will briefly review aspects of the CoV replication cycle with a focus on the role of nonstructural proteins (nsps) and in particular the subjects of this dissertation: nsp2, nsp3, and nsp4.
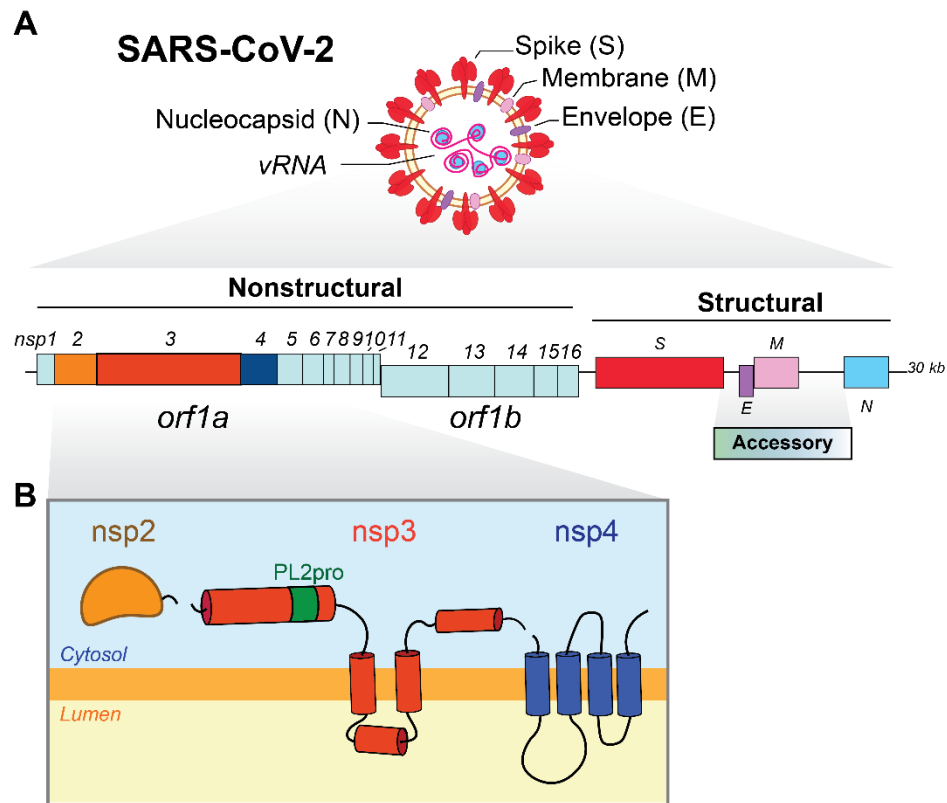
Figure 1-1. SARS-CoV-2 virion and genome composition.
  A. SARS-CoV-2 virion composition, including labeled structural proteins, which deliver the viral RNA genome (vRNA). The vRNA is 30,000 bases long and encodes for *Orf1a* and *Orf1b* from which 16 nonstructural proteins are produced. Structural and accessory proteins are encoded in the last two-thirds of the genome.
  B. This dissertation focuses on the functions and host interactions of nsp2, nsp3 (containing a protease domain, PL2pro), and nsp4.

CoV virions are enveloped in host-derived membranes, inside which is the viral positive-sense, single-stranded RNA genome (gRNA) wrapped around nucleocapsid (N) protein (**Figure 1-1A**). The viral membrane is studded with integral membrane structural proteins, including membrane (M), envelope (E), and spike (S), as well as hemagglutinin

esterase (HE) in the case of MHV. The spike protein mediates cell attachment and entry, binding to host protein receptors on the plasma membrane such as ACE2 (SARS-CoV-1, -2, hCoV-NL63) (33–35), DPP4 (MERS-CoV)(36), and CEACAM1 (MHV) (37) (**Figure 1-2**). Virions release the gRNA into the cytoplasm after fusion with either endosomes or the plasma membrane and uncoating of nucleocapsid proteins. CoV gRNA is roughly 30,000 bases in length, one of the largest known RNA viral genomes, with the first ~20,000 bases encoding two open reading frames (Orf1a, Orf1b) containing sixteen nonstructural proteins (nsp1-16) (**Figure 1-1A**). The last ~10,000 bases encode for the structural proteins (S, E, M, N) with accessory protein genes located between (31).
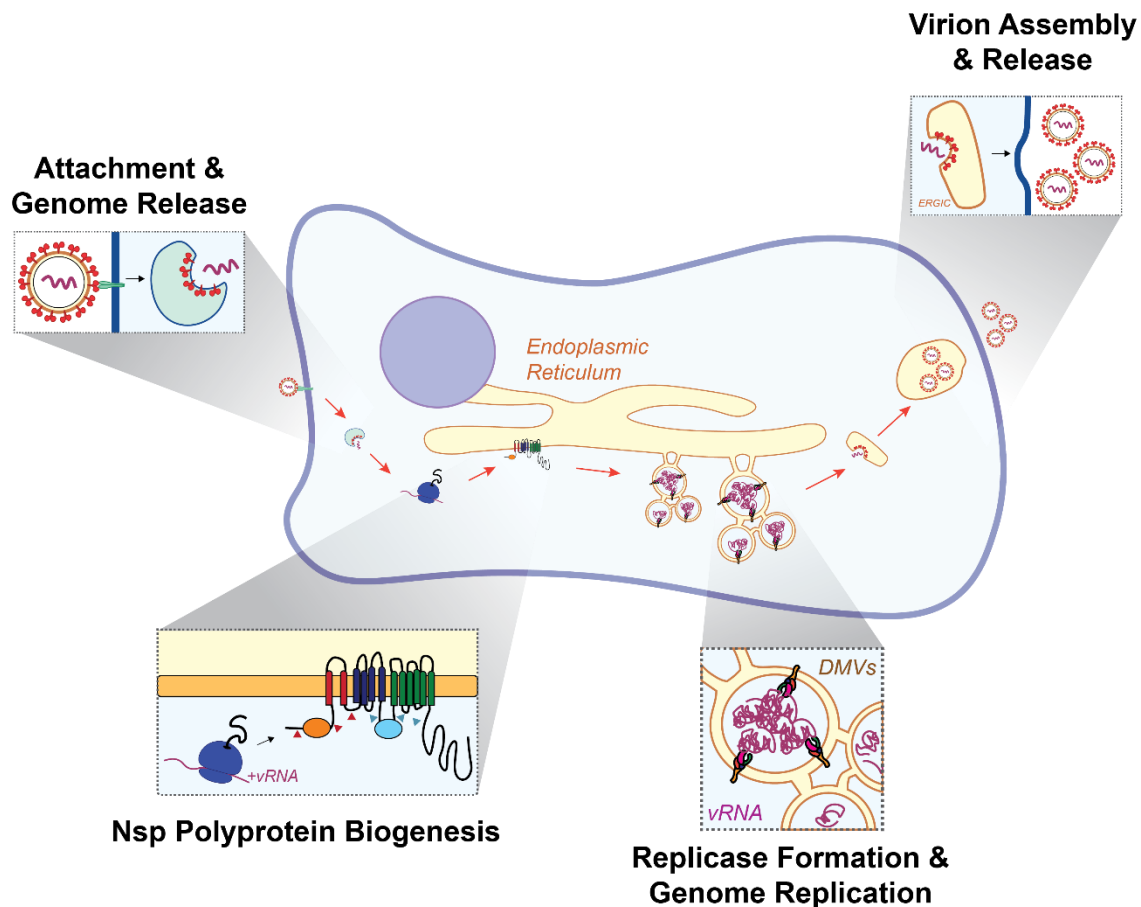
Figure 1-2. The CoV replication cycle.
Prominent aspects of the replication cycle are highlighted, including attachment, genome release into the cytosol, production of nsp polyprotein, replicase formation, genome replication, virion assembly, and release.

The Orf1a/b portion of viral gRNA is translated first by host ribosomes, resulting in two polyproteins, pp1a and pp1ab, through the use of a -1 ribosomal frame shift at nsp11 (**Figure 1-2**). These polyproteins are inserted into host ER membranes and self-cleaved by viral proteases (nsp3 and nsp5) to release the nsps (38). Nsps form double membrane vesicles (DMVs) from ER-derived membranes and assemble into viral replication/transcription complexes (RTCs) within DMVs (31). RTCs mediate viral gRNA replication and processing, in which negative-sense gRNA ((-)gRNA) is synthesized and used as a template, both for gRNA replication and production of subgenomic RNAs (sgRNAs) that encode for structural and accessory proteins. New gRNA is transferred to N protein in the cytoplasm and packaged with structural proteins into

new virions in the ER-Golgi-Intermediate Compartment (ERGIC). Virions are canonically thought to exit the cell through the secretory pathway, though recent evidence suggests this may be mediated by the lysosomal trafficking for MHV and SARS-CoV-2 (39).

Nsps mediate most CoV replication events between entry and egress, in large part through virus-virus and virus-host protein interactions (PPIs) (40,41). Likely due to this essential nature, nsp1-16 are found in almost all CoVs and in general have high sequence conservation amongst homologs (with the exception of nsp2), which is in contrast to the tremendous sequence variation for the spike protein. Nsp3 and nsp6, in combination with nsp4, are the only multi-pass transmembrane nsps and rearrange host ER membranes into DMVs (42–45). Recent work has also shown that nsp6 acts as a gatekeeper between the ER and DMVs, presumably to prevent host proteins from gaining access to the DMV interior (46). Emergence of sequence variants in nsp3, nsp4, and nsp6 have been observed over the course of the COVID-19 pandemic and some have been associated with alterations in SARS-CoV-2 pathogenesis (47–49).

Nsp12-16 complex together to form the core viral genome replication complex of the RTC with the assistance of nsp7-10. Nsp12 contains the RNA-dependent RNA polymerase which synthesizes new viral RNA with the aid of nsp7 and nsp8 (50). Viral gRNA is capped at the 5' end by nsp9 and nsp12 and the cap is methylated by nsp14 and nsp16 (51). New gRNA is also scanned by nsp14, which is a 3' to 5' exoribonuclease with proofreading and recombinase abilities, an unusual feature for an RNA virus (52,53).

1.1.2.2   Nsp2 overview

Nsp2 is a soluble protein with an ambiguous role in replication (**Figure 1-1B**). It is dispensable for infection in SARS-CoV-1 and MHV cell culture models, albeit with significant stunting in viral titers ($\sim$two $\log_{10}$ plaque forming units (PFU)/mL) (54); however, its role may be more necessary during infection in animals and humans, perhaps in contending with host immunity. There is a high degree of amino acid sequence variation between nsp2 homologs (32), and early sequence analysis of SARS-CoV-2 identified regions of positive selection pressure in nsp2 (55). Nsp2 has also been shown to associate with other nsps amid the viral RTC during infection (56–58), indicating likely involvement in supporting either host modification or viral replication.

Structural characterization of SARS-CoV-2 nsp2 has been carried out through several different methods, such as crosslinking mass spectrometry (59), X-ray crystallography of the N-terminus (60), and Cryo-electron microscopy (Cryo-EM) aided by AlphaFold2 (61). These studies have identified well-conserved Zinc-binding sites in the N-terminus and suggest potential RNA-binding activity.

1.1.2.3   Nsp3 overview

Nsp3 is a glycoprotein and the largest nsp, with a large N-terminal cytosolic portion, followed by two transmembrane domains with a small luminal Ecto domain and a membrane-associated alpha helix domain on the C-terminus (43) (**Figure 1-1B**). There are many domains within nsp3 with some variation between different CoVs (see review (62)). In the N-terminal region, the ubiquitin-like domain (UBL) 1, UBL2, and macrodomain 1 (Mac1) domains are shared

across homologs, while SARS-CoV and SARS-CoV-2 contain a SARS-unique domain (SUD). hCoV-OC43, hCoV-229E, and MHV contain two papain-like protease domains (PL1pro and PL2pro), whereas SARS-1, SARS-2 and MERS only have the one PL2pro domain (also annotated as PLpro). The PL2pro domain has been the subject of much research, as besides cleavage of the nsp1/2/3/4 polyprotein sites, this domain also possesses deubiquitinase and deISGylation activity against host proteins (63,64). The PL2pro domain functions as a Cysteine protease and inhibitors have been developed to suppress viral infection, although none have been approved for treatment (65).

As noted above, nsp3 is involved in DMV formation. Recent work has found expression of SARS-CoV-2 nsp3 and nsp4 are sufficient to form DMVs similar to those observed in infection (45). Cryo-electron tomography (Cryo-ET) studies have revealed molecular pores doting DMVs during MHV infection, which appear to be produced by a crown-like oligomerization of nsp3 and nsp4 (66). Further cryo-ET work with SARS-2 nsp3-nsp4 expression show that these proteins are sufficient to form DMV pores (67). These pores are postulated to act as gateways for viral RNA to be transported to the outside of DMVs for virion encapsulation.

1.1.2.4   Nsp4 overview

Nsp4 contains four transmembrane domains with two luminal loops, one cytosolic loop, and both N- and C-termini in the cytosol (68,69) (**Figure 1-1B**). It retains high sequence conservation between homologs and, like most nsps, is indispensable for infection (70). It primarily acts to remodel host membranes for DMVs (42,44), forming homodimers with itself and heterodimers with nsp3 and nsp6 to pair host membranes together (71). The larger luminal loop of nsp4 is *N*-glycosylated at two sites in MHV (N176, N237) (72) and one site in SARS-CoV and SARS-CoV-2 (N131) (43). This glycosylation has been shown to be crucial for viral infection and DMV morphology, as mutations at these residues hinder viral fitness (73).

1.1.3   ER protein homeostasis (proteostasis) network

1.1.3.1   General purpose of the proteostasis network

Human cells genomically encode approximately 20,000 different proteins (74), all with varying structural features, isoforms, half-lives, spatial organization, and functions. To manage this vast proteome and maintain protein homeostasis (proteostasis), cells use a collection of interconnected quality control systems which are collectively called the proteostasis network. These systems can be broadly binned into three categories: 1) protein production, 2) protein degradation, and 3) surveillance and regulation. They cover a range of activities, from protein folding and addition of post-translational modifications to clearing misfolded proteins to responding to heat shock and other stress conditions (75). Key to these networks is the coordination of PPIs between proteostasis factors and client proteins (76). The host proteostasis network is particularly prominent in the endoplasmic reticulum (ER), where roughly a third of the cellular proteome is processed and some are modified for secretion. The ER is an oxidizing environment akin to the outside of the cell and contains a high

concentration of calcium and chaperones dedicated to folding new proteins. Given the centrality of the ER to proteostasis, I will primarily focus on reviewing the relevant ER proteostasis systems, with an emphasis on processes known to be hijacked by CoV replication.

1.1.3.2 *Protein production and biogenesis systems*

Proteins are produced by ribosomes, which translate genetic information encoded in mRNA transcripts into nascent polypeptides. The essential nature of this process means every sequential event is heavily regulated. Translation initiation requires the temporal assembly of various factors to assemble ribosomes and load mRNAs (77). Ribosomes can sometimes stall on transcripts or transcripts may need to be temporally blocked from being translated, so components of the Ribosomal Quality Control (RQC) complex act to prevent and remove problematic ribosome collisions (78). As polypeptides emerge from the ribosome, they are scanned by the Signal Recognition Particle (SRP) for signal peptides that indicate ER membrane localization and upon recognition, the ribosome is brought to the rough ER (79,80). Here proteins are threaded co-translationally into the ER membrane by the Sec61 translocon, with the help of the ER membrane protein complex (EMC) to insert transmembrane helices (81,82).

Within the ER, a whole slew of factors engage with client proteins for further processing. There is a large concentration of molecular chaperones and co-chaperones, such as BiP (HSPA5) and GRP94, which help a variety of client proteins fold into functional three-dimensional shapes (83). Post-translational modifications (PTMs) are added to modify function or structure, such as disulfide bonds between two Cystine residues formed by protein disulfide isomerases (PDIs) (84)
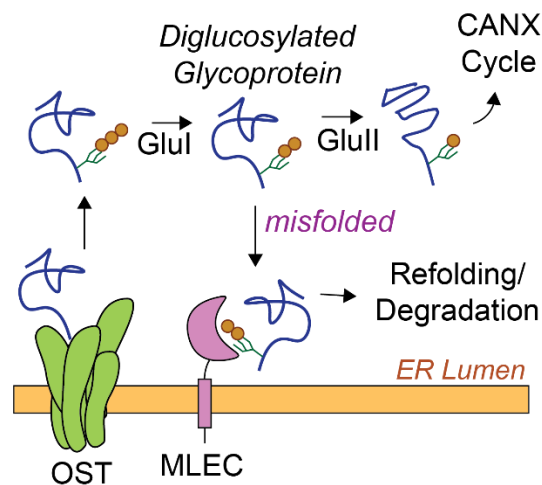
Figure 1-3. Role of Malectin (MLEC) in glycoprotein biogenesis.

An important PTM is *N*-linked glycosylation, where an oligosaccharide is placed on an Asparagine residue in a client protein (termed glycoprotein). Lipid-linked oligosaccharides (LLOs) are synthesized on the cytoplasmic face of the ER before being brought into the lumen and placed on target proteins either co- or post-translationally by the

Oligosaccharyltransferase (OST) Complex at Asn-X-Ser/Thr sequons (85). These sugars initially contain terminal glucosyl groups which are used as recognition sites to monitor the status of glycoprotein biogenesis by the glycoprotein quality control system (86) (**Figure 1-3**). Initially after OST transfer, the oligosaccharide has three glucosyl moieties and is trimmed to diglucosyl by glucosidase I (GluI) and then subsequently into a monoglucosyl group by glucosidase II (GluII) (87–89). From there, client glycoproteins enter the calnexin (CANX) folding cycle and the final glucosyl group is removed by GluII if proper folding is achieved prior to further secretion. If the glycoprotein fails to reach the folded state, it is re-glucosylated by UGGT1 and re-enters the CANX folding cycle. If several iterations of this fail, the glycoprotein is shuttled for degradation (86,90–92)

The trimming sequence from tri- to di- to monoglucosyl moieties was initially thought to be rapid. However, an ER-resident lectin protein, malectin (MLEC), was discovered to have strong and specific affinity for diglucosyl moieties, indicating a potential subset of glycoproteins that contain the diglucosyl sugars (93,94) (**Figure 1-3**). MLEC expression was shown to be upregulated during ER stress and found to retain diglucosylated glycoproteins in the ER lumen (95). The improper trimming is postulated to result from misfolding of the glycoprotein, preventing GluII full access to the glycan, and so MLEC retains the client until proper folding and trimming can be achieved or the protein degraded. Given its role in preventing secretion of misfolded glycoproteins, MLEC has been added to the list of glycoprotein quality control factors (95,96). MLEC has also been shown to associate with ribophorin I (RPN1)(97–99), a component of the OST complex (100), likely indicating a tight and early linkage between the initial addition of the triglucosylated oligosaccharide by the OST complex and the surveillance function of MLEC. The role of MLEC in host cell proteostasis still requires further research and this is further explored in the context of CoV replication in Chapter 4.

1.1.3.3   *Degradation systems*

Whether through the natural course of their lifespan or because of some external insult, proteins must be regularly degraded to make room for new proteins and avoid toxic protein aggregation. Cells use several systems to clear proteins, ranging from targeting individual proteins to whole organelles for destruction.

The canonical degradation apparatus for individual proteins is the ubiquitin-proteosome system (UPS). Target proteins are tagged with polymers of the small protein ubiquitin by E3-ubiquitin ligases (101). The ubiquitin tags are recognized by the proteosome, a large barrel-like complex with catalytic proteases in the interior, for subsequent proteolysis and recycling of amino acids (102,103). Proteosomes are not located in the lumen of the ER and so there is a physical barrier to destroying misfolded proteins in the ER. This is overcome by the ER-associated degradation (ERAD) pathway. In brief, components of ERAD recognize misfolded proteins and retrotranslocate the clients across the ER lumen into the cytosol, where the unfolded protein is tagged with ubiquitin and shunted to the aforementioned proteasome for degradation (104).

There are also systems in place for bulk-degradation, namely autophagy, in which large groups of protein aggregates are encompassed by double membranes into vesicles (autophagosomes) that then fuse with the lysosome and general degradation (105). The specific form of bulk degradation of whole portions of the ER is termed ER-phagy and is discussed comprehensively in a recent review (106). New work has begun to decipher the molecular mechanisms which regulate ER-phagy, including the identification of ER-phagy receptors, including RTN3, FAM134B, CCPG1, TEX264, and SEC62 (107–

112). The specificity and operational principles of how regions and clients are selected is actively being researched and may also have implications for protein secretion (76).

1.1.3.4 Regulation by the Unfolded Protein Response (UPR)

Cells encode a variety of stress response systems to sense and adapt to various insults to the proteome which might jeopardize cell viability. One of these systems is the Unfolded Protein Response (UPR), which responds to accumulation of misfolded proteins in the ER. The UPR temporarily shuts off new protein translation to decrease incoming client proteins, upregulates ER chaperones to fold existing misfolded proteins, and expands ER volume (113), all in an attempt to mitigate the misfolded protein burden. If unsuccessful and left chronically activated, the UPR can trigger apoptosis (114). The UPR consists of three branches that signal downstream of their respective ER-membrane localized stress sensors to carry-out the aforementioned functions: 1) protein kinase R-like ER kinase (PERK), 2) Inositol requiring enzyme 1α (IRE1α), and 3) activating transcription factor 6 (ATF6) pathways (115,116).

1.2   Nsps and the Proteostasis Network

CoV infection necessitates the rapid and large-scale production of viral proteins to produce new viral genomes and assemble progeny virions against the impending pressure of awakening host defense systems. Furthermore, viruses must remodel host proteins and networks into a permissive cellular environment for replication. Because the host proteostasis network is foundational to all aspects of these goals, CoVs have evolved diverse strategies to commandeer this essential pathway.

1.2.1   Proteomics as a tool to measure virus-host interactions

Protein-protein interactions (PPIs) are one of the key methods used by viruses to hijack the host cell during infection (**Figure 1-4A**). Defining the interactions between nsps and the host proteostasis network thus offers a foothold to characterize mechanisms of viral infection. The application of mass spectrometry to measure proteins (proteomics) has been leveraged extensively to explore the interface between viruses and the host cell, including measuring changes in host or virus proteome levels (global proteomics), characterizing post-translational modifications (e.g. phosphoproteomics, glycoproteomics), defining the proximal protein environment (proximity-labeling), and identifying virus-host protein-protein interactions (PPIs) (interactomics/affinity-purification mass spectrometry (AP-MS)). Comparative AP-MS has proven to be a valuable asset in studying viral infections of pathogens with high human health relevance (117).

Proteomics was again successfully deployed in response to the COVID-19 pandemic to characterize the molecular underpinnings of SARS-CoV-2 as well as to advance therapeutic development. Several landmark studies include a SARS-CoV-2 protein interactome map (118) and comparative interactomics profiling of SARS-CoV, SARS-CoV-2, and MERS-CoV homologs (119) from a large consortium led by researchers at the University of California, San Francisco, as well as

global proteomics profiling of SARS-CoV-2 infection (120) and a comparative multi-omics study on SARS-CoV, SARS-CoV-2, hCoV-NL63, and hCoV-229E (121). These studies were carried out and published in a parallel time frame with the comparative interactomics work described in this dissertation (Chapters 2 & 3) and highlight the importance of researching CoV-host interactions.
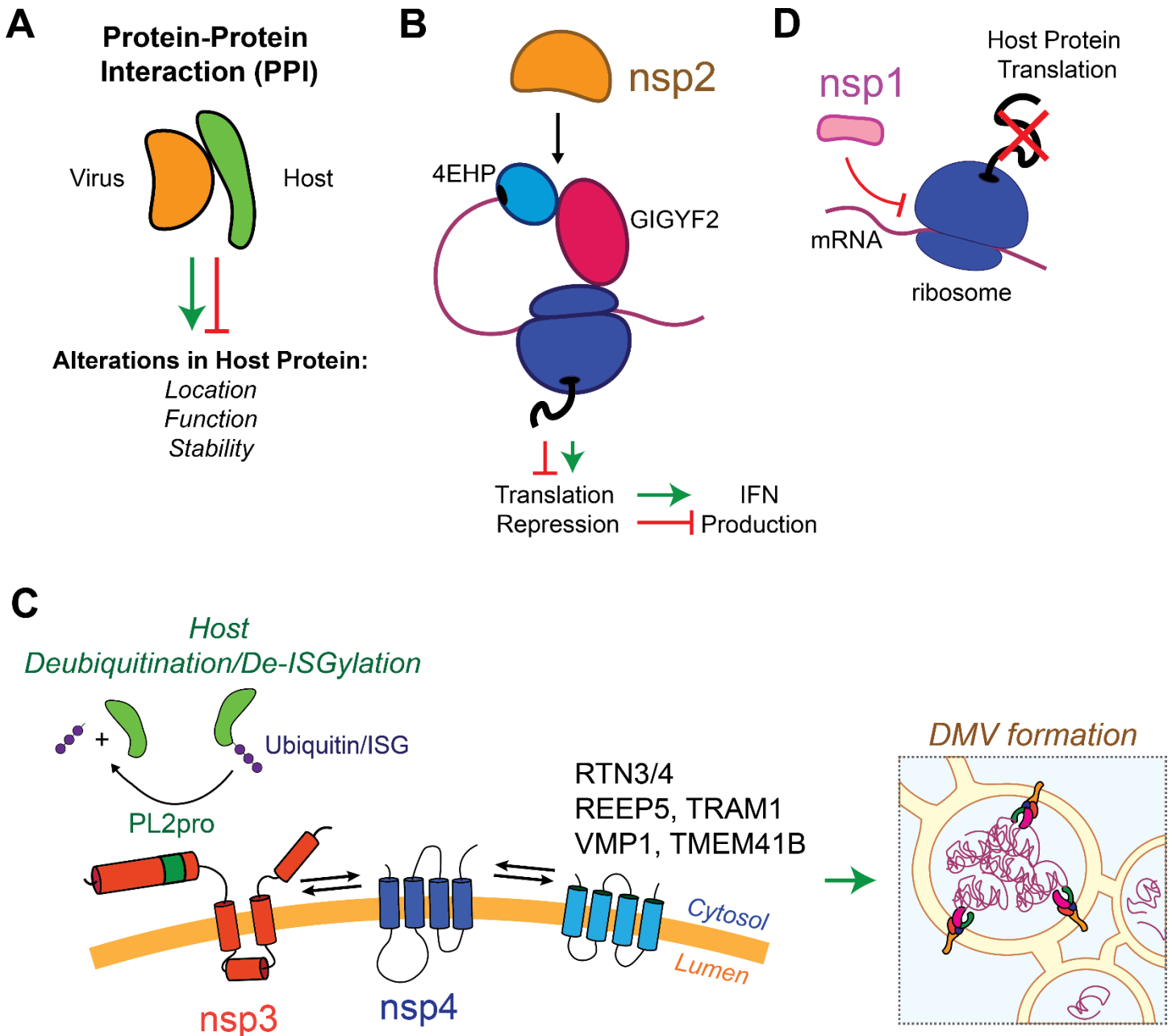
Figure 1-4. CoV nsp-host protein-protein interactions (PPI) mediate infection.
- A. Virus-host PPIs can alter host protein localization, function, and stability.
- B. Nsp2 interacts with 4EHP/GIGYF2 to modify translation repression and interferon (IFN) production. There are conflicting studies on whether this interaction increases or suppresses translation repression.
- C. Nsp3 contains a PL2pro protease domain, which in addition to cleaving the viral polyprotein, can cleave ubiquitin or ISG-ylation post-translational modifications on host proteins, leading to modifications in host protein function. Nsp3 and nsp4 interact with a variety of host proteins (RTN3/4, REEP5, TRAM1, VMP1, TMEM41B) to mediate DMV formation.
- D. Nsp1 inhibits translation by binding the mRNA intake site on ribosomes.

9

Herein, I will review recent advances in our understanding of the interactions between CoV nsps and the host proteostasis network, with a particular emphasis on the application of interactomics to study this interface and the nsps which this dissertation focuses on (nsp2, nsp3, nsp4).

1.2.2 Nsp2 & host translation regulators

Nsp2 has been studied in several different interactomics studies and found to interface with several aspects of the proteostasis network. Early work expressing tagged SARS-CoV nsp2 in cells during SARS-CoV infection by Cornillez-Ty et al. identified interactions between nsp2 and GIGYF2 and EIF4E2 (4EHP) (122), which act as translation repressors in conjunction with the Ribosome Quality Control (RQC) system (123–125) (**Figure 1-4B**). This interaction was later reported to be conserved with SARS-CoV-2 nsp2 (118,119), though work from our group would suggest it is a SARS-1 specific interaction (126).

Characterization of the structural determinants of these interactions has led to varied models on the binding interfaces. One study reported that SARS-CoV-2 nsp2 binds directly to GIGYF2 (127), narrowing down the binding site on GIGYF2 to amino acids 860-919 using IP and proximity ligation assays (PLA). In contrast, another study suggested multiple binding interfaces between SARS-CoV-2 nsp2, 4EHP, and GIGYF2 using a series of truncation IP assays (128). The authors additionally narrowed down the binding region of nsp2 to the N-terminal region (1-350 aa), consistent with findings by AP-MS that naturally occurring variants with G262V and G265V substitutions in SARS-CoV-2 nsp2 reduce 4EHP-GIGYF2 binding (61). There is further conflicting evidence on the functional output of this interaction, with nsp2 either interfering with the translation repression activity of the 4EHP/GIGYF2 complex (128) or enhancing translation repression, particularly against IFN-β production (127) (**Figure 1-4B**). These discrepancies may result from differences in interaction and functional assays used. Regardless, 4EHP and GIGYF2 have been shown to be important for SARS-CoV-2 infection, with genetic depletion of these factors leading to reduced viral replication (127,129). Whether these functions apply to other nsp2 homologs, such as SARS-CoV and MERS-CoV, remains to be tested.

Other interactions between nsp2 and host translation regulation have been reported. The MERS-CoV nsp2 homolog was shown to interact with several components of the RQC (119), including the ASC-1 complex, which disassembles collided ribosomes (130), and TCF25, an RQC complex subunit (131). This study also identified an interaction with GALNT2, which catalyzes a portion of the *O*-linked oligosaccharide biosynthesis pathway (132). The implications of nsp2 interactions with glycoprotein biogenesis pathways require further investigation (see Chapter 4). Beyond PPIs, proximity-labeling proteomics has also shown that MHV and MERS-CoV nsp2 are in close proximity to host translation initiation factors and autophagy factors at Replication-Transcription Complexes (RTCs) during infection (133).

Clearly, there is a consistent connection between nsp2 and the protein biogenesis pathway, which is the subject of further investigation in Chapter 4.

1.2.3 Nsp3 PLpro & host protein modifications

Nsp3 was largely absent from large-scale AP-MS studies published at the advent of the COVID-19 pandemic, due in part to the protein's large size and poor expression. Chapter 3 details our lab's work to fill this gap using comparative interactomics. Early work has primarily focused on domain-specific interactions, though improved protein expression models have enabled studies of full-length nsp3 (45,134).

Much of the interface between nsp3 and the host proteostasis system is centered around its function as a viral protease and manipulation of ER membranes to form DMVs. The papain-like protease domain (PLpro or PL2pro) cleaves the orf1a polyprotein between nsp1/nsp2/nsp3/nsp4 (135) and has been shown to remove post-translational modifications (PTMs) on host proteins, namely through deubiquitination and deISGylation (62) (**Figure 1-4C**). Ubiquitination and ISGylation are pivotal modifications that influence the stability and trafficking of the host proteome, particularly in organizing host anti-viral defense (136), making these PTMs important targets for viral interference.

The deubiquitinating activity of PLpro is conserved across CoV nsp3 homologs. The SARS-CoV PLpro domain was shown to have deubiquitinating activity by Linder et al. (137) and later structural work revealed the binding interface for MERS-CoV PLpro with ubiquitin (138). SARS-CoV-2 PLpro was shown to diverge in client specificity from SARS-1, with SARS-CoV-2 preferring to cleave ISG-ylated substrates in contrast to SARS-CoV preferences for cleaving ubiquitinated substrates. This difference in substrate preference was shown to lead to a reduction in the interferon response to SARS-CoV-2 infection (139,140). Additional proteomics work has shown that many of the interactions of the N-terminal cytoplasmic arm of nsp3 (which contains PLpro) are largely oriented around the PLpro domain, highlighting the importance of this domain in host interactions (141).

1.2.4   Nsp3, nsp4, & ER membrane proteins

Both nsp3 and nsp4 are embedded in ER-derived membranes, placing these proteins in a prime location to modify host membranes and the tightly linked proteostasis processes linked to the ER environment. SARS-CoV-2 nsp3 and nsp4 can be expressed as an auto-cleaving polyprotein and are sufficient to form DMVs similar to those observed during infection (45). As such, much recent work has focused on the dual interface between nsp3/nsp4 and the host factors needed to reshape ER-derived membranes for DMVs (**Figure 1-4C**).

One of the primary proteostasis networks that has been implicated in DMV formation are vesicle-based protein degradation pathways, such as autophagy (58,142) and ERAD-associated vesicles (134,143). In a typical subversive fashion, CoVs repurpose cellular waste-disposal mechanisms to shield viral propagation.

Recent work has highlighted how this subversion is mediated by nsp3/4 PPIs with host regulators of autophagy. Williams, et al. reported PPIs between ER reticulon proteins RTN3/4 with SARS-CoV-2 nsp3 and nsp4 (144). RTN3 was recently shown to serve as an ER autophagy (ER-phagy) receptor (107). The authors further show that these interactions mediate DMV formation and are pro-viral for SARS-CoV-2 infection.

Concurrently, the ER-resident autophagy regulators VMP1 and TMEM41B (145,146) have been shown to be key host dependency factors in SARS-CoV-2 and common cold CoV strains (147,148). Further biochemical study of this dependency found that VMP1 directly interacts with both SARS-2 nsp3 and nsp4 through its ER luminal loop, while TMEM41B binds just nsp4 to mediate DMV formation (148). Further, TMEM41B regulates nsp3 and nsp4 intraviral PPIs,

with TMEM41B knockout weakening interactions between the two. TMEM41B has also been shown to play a broader role in RNA virus modulation of host ER membranes. It interacts with DENV NSP4 to facilitate replication sites around the ER membrane (129), which points towards pan-RNA virus therapeutics if TMEM41B-specific inhibitors can be developed. Nsp3/4-interactions with ER-phagy/autophagy factors may also influence host protein degradation regulation and should be studied further.

A common caveat of AP-MS studies using individual viral protein expression is the loss of infection context which may drastically reshape the nsp interactome. Pahmeier et al. partially circumvent this using the self-cleaving SARS-CoV-2 tagged nsp3/nsp4 polyprotein construct for AP-MS (149). Using full-length nsp3 increases the relevance of identified interactors, overcoming limitations of protein truncation. As nsp3 and nsp4 are sufficient for SARS-CoV-2 DMV formation, the authors were able to identify host interactors with the nsp3/nsp4 polyprotein and use individual nsp3 or nsp4 IPs to subtract out interactors likely not associated with DMV formation. They identified several ER proteostasis interactors, including TMEM33 (a regulator of UPR activator (150)) and OST components (OSTC, MAGT1), however knockdown of these interactors did not lead to significant loss in DMV formation or SARS-CoV-2 infection. Intriguingly, the host factors FAM149B1, CCAR2, and ZC3HAV1 were found to promote DMV formation and SARS-CoV-2, despite these factors having cytosolic localization.

Continuing to bridge the gap between single protein and virus infection AP-MS, a recent study profiled the interactome of SARS-CoV-2 nsp3, nsp4, and nsp6 expressed individually or in pairs. They identified a wide swath of ER-resident interactors, including the nsp3-interacting REEP5 and TRAM1 (151). REEP5 induces ER membrane curvature (152) and TRAM1 works in complex with Sec61 to translocate secretory proteins into the ER membrane (153). The authors demonstrate that this interaction is maintained during SARS-CoV-2 infection and that knockout of REEP5 severely attenuates replication.

Additional AP-MS work performed in the early stage of the pandemic found nsp4 interacts with a wide range of ER membrane, protein biogenesis, and glycosylation factors (119). MERS nsp4 has been shown to interact with TOR1AIP2 (119), a factor that can induce double membrane formation in the ER (154). A conserved interaction amongst SARS-1 and SARS-2 nsp4 homologs is with ALG11 (118,119), the final enzyme in oligosaccharide synthesis process on the cytoplasmic face of the ER prior to transfer by the OST complex (155). Additionally, the MERS homolog interacts with the mannose lectin LMAN1, which acts as a cargo receptor to sort proteins for ER-to-Golgi trafficking (156), and the prolyl 4-hydroxylase subunit P4HA1, which catalyzes the formation of 4-hydroxyprolines in collagen (119,157). As nsp4 is also a glycoprotein, it remains to be determined whether interactions with PTM factors occur as a client protein or as a modifier of host machinery.

1.2.5   Other nsps at the host proteostasis interface

While this dissertation focuses on nsp2, nsp3, and nsp4, several other CoV nsps have been documented to interact with the proteostasis network, primarily nsp1, nsp6, and nsp8.

1.2.5.1 Nsp1

SARS-CoV nsp1 was the subject of a yeast-two hybrid screen which identified a large cluster of peptidyl-prolyl isomerases (immunophilins: PPIA, PPIG, PPIH, FKBP1A/B) interacting with the N-terminus of nsp1 (158). This interaction was shown to induce upregulation of cytokine production. Inhibition of the interaction suppressed both SARS-CoV and hCoV-229E replication.

Nsp1 has also been shown to interfere with host translation. SARS-CoV nsp1 was shown to suppress host translation through binding with the 40S ribosomal subunit and inhibiting ribosomal translation, leading to suppression of interferon (IFN) production and host anti-viral activity (159,160). This function is conserved with SARS-CoV-2 nsp1, which was further found to bind the 18S ribosomal RNA within mRNA entry channel and block translation initiation and protein production (161) (**Figure 1-4D**). This interaction is disrupted by the 5' leader sequence on viral subgenomic RNAs (sgRNAs), allowing for selective viral translation while repressing host IFN translation. Using ribosomal profiling, a recent preprint has also reported that SARS-CoV-2 nsp1 enhances viral genomic RNA (gRNA) translation efficiency through interactions with the host translation initiation factor EIF1A (162). These studies showcase how nsp1 modifies host translation mechanisms to support viral protein production.

1.2.5.2 Nsp6

Nsp6 has been shown to interact with several chaperones. AP-MS identified an interaction between MERS nsp6 and the co-chaperones DNAJB1/4 (119) and between SARS-CoV-2 nsp6 and the Sigma-1 receptor (S1R) (118). S1R is an ER-membrane-resident chaperone that regulates ER-to-mitochondria calcium signaling through interactions with its client protein inositol 1,4,5-triphosphate receptor (IP3R) (163) as well as ER stress signaling (164). S1R is a common target for several antipsychotics, antihistamines, and antimalarials and a wide array of licensed S1R ligands have been tested against SARS-CoV-2 infection (118). Several of these compounds had antiviral activity (PB28, haloperidol, clemastine, cloperastine), except for one (dextromethorphan) which increased infection. While it remains to be tested whether these compounds disrupt the nsp6-S1R interaction, efforts have been made to target this interface through *in-silico* drug testing (165) and it is clear the S1R plays an important role in viral replication, as knockout of S1R impedes both SARS-CoV and SARS-CoV-2 infection (119,166).

1.2.5.3 Nsp8 & nsp9

Nsp8 and nsp9 have been shown to interact with host pathways involved in membrane protein biogenesis. Most notably, AP-MS studies identified a conserved interaction between SARS-CoV-2, SARS-CoV, and MERS-CoV nsp8 homologs with components of the signal recognition particle (SRP; SRP19/54/72) (118,119). The SRP complex is composed of proteins scaffolded on a long noncoding RNA (lncRNA; 7SL) and co-translationally directs nascent proteins for ER membrane insertion (79,80). Further work by Banerjee, et al. found that nsp8 and nsp9 bind to the 7SL scaffold at sites where SRP54 and SRP19 normally bind, displacing the host proteins and preventing ER-membrane insertion of client

proteins and leading to IFN suppression (161). It remains to be seen whether this mechanism can differentiate between host versus viral proteins, as many CoV proteins require ER-membrane insertion.

1.2.5.4   Nsps and the Unfolded Protein Response

Coronavirus replication requires extensive production, folding, and modification of viral proteins (167), as well as alteration of ER-derived membranes to form DMVs for replication sites. These processes can trigger ER stress and activate the Unfolded Protein Response (UPR) (168). SARS-CoV-2 has been shown to activate all three branches during infection (169) and MHV triggers activation of the XBP1 and PERK pathways while also hindering production of certain UPR-responsive genes such as CHOP (170). The contribution of individual CoV proteins has also been demonstrated, such as SARS-CoV-2 Spike and orf8, which are sufficient to activate the UPR (169), and the SARS-CoV accessory proteins orf3a and orf8ab, which activate the PERK and ATF6 pathways respectively (171,172). This previous work shows that CoVs can modulate stress responses from multiple angles. Interestingly, relatively little is known about the impact of CoV nsps on the UPR, despite the central role nsps play in modifying host membranes for infection. This gap is further explored in this dissertation Chapter 5.

1.3   Dissertation Outline

1.3.1   Pressing questions

With the onset of the COVID-19 pandemic, several pressing scientific questions arose surrounding the virus of the moment, SARS-CoV-2. What parts of a host cell does SARS-CoV-2 use for infection? How is SARS-CoV-2 different or similar from other highly pathogenic CoVs that we have encountered before? What infection mechanisms differentiate a highly pathogenic CoV from a common cold CoV? Which host cell dependencies might current and future CoVs need for infection? And can we exploit these dependencies for anti-viral therapies against SARS-CoV-2 and perhaps a broader range of CoVs?

In parallel, the outbreak brought into focus broader questions about pandemic preparedness. How can we quickly assess emerging pathogens for their molecular mechanisms of infection? Can we identify host dependencies that are shared among a family of viruses and develop anti-viral therapies to have on hand prior to new outbreaks? And are there pan-viral targets that can be exploited for broad spectrum anti-virals against multiple families of viruses?

1.3.2   Aims and dissertation layout

In this dissertation, I set out to answer some of the pressing questions listed above by exploring the interactions between CoV nsps and the host cell. In particular, I chose to focus on how nsps hijack host proteostasis systems and whether these mechanisms are the same or different between SARS-CoV-2 and other CoVs. The goals of this work roughly fit into two

phases, beginning with a discovery phase of nsp-host interactions (Aim 1 & 2) and transitioning to focus on functional characterization of these interactions (Aims 3 & 4) and are as follows:

Aim 1. Use comparative proteomics to rapidly profile unique and conserved host interactors of CoV nsp homologs.

Aim 2. Apply functional genomics screening to determine which interactors are important for infection (pro-viral) versus host defense (anti-viral), with a particular focus on host proteostasis interactors.

Aim 3. Investigate how *individual* nsp-host proteostasis interactions mediate CoV infection.

Aim 4. Investigate how nsps modulate host proteostasis *systems* like the Unfolded Protein Response through PPIs to modify the host cell environment.

Aim 1 is encompassed in Chapters 2 & 3, which were a joint, collaborative effort between Dr. Katherine M. Almasy, a fellow graduate student in the Plate Lab, and me. **Chapter 2** showcases the rapid deployment of a comparative interactomics workflow to profile the interactomes of SARS-CoV-2 and other CoV nsp2 and nsp4 homologs (based on published work (126)). **Chapter 3** details an expanded application of this workflow to nsp3 homologs and includes a foray into the goals of Aim 4, looking at the interface of nsp3 and the Unfolded Protein Response (UPR) (based on published work (173)). **Chapter 4** encompasses Aims 2 & 3, first applying functional genomics screening to identify relevant interactors from Chapter 2 and then focusing on one of the most impactful hits, the glycoprotein quality control factor Malectin (MLEC), and characterizing its function in infection (unpublished work in preparation for submission). **Chapter 5** then builds off work from Chapter 3 and investigates how nsp3 and nsp4 tune the UPR using a systems biology approach (based on published work (174)). The work in this chapter was a joint project with an undergraduate mentee, Athira Sivadas, and in collaboration with others. Lastly, in **Chapter 6**, I discuss conclusions, impact, and limitations of the work, and outline future directions.

1.4   Potential Impact

The interactions between viral nsps and host proteins are the foundation upon which CoV infection is built. Thus, studying these interactions advances both our knowledge of the basic biology of infection and lends crucial insight into strategies we might take to stop infection and preserve human health. SARS-CoV-2 is a novel human pathogen. To understand the molecular mechanisms which govern infection, we must both draw from our knowledge of related CoVs and use discovery-oriented methods. This work can have an immediate impact in expanding our knowledge of how SARS-CoV-2 infects cells. A comparative approach can also widen our understanding of how CoVs operate as a family and what differentiates these viruses from one another. Additionally, viruses are masters of manipulating host processes, so examining viral mechanisms of replication invariably sheds light on the fundamental biology of human cells. This work will help reveal guiding principles for host proteostasis networks, such as the understudied glycoprotein quality control factor, Mlec, and systems regulation of proteostasis through the Unfolded Protein Response.

Furthermore, this work can enhance approaches to pandemic preparedness. Developing a comparative interactomics workflow that can rapidly assess the workings of emerging pathogens will provide valuable information on a novel virus like SARS-CoV-2. This can identify potential host dependencies that can be targeted with existing drugs for accelerated

therapeutic testing. Outside of an outbreak setting, this work also provides a framework for finding host dependencies shared across multiple viruses within families that have a known propensity for causing severe disease in humans, like coronaviruses. Targeting host dependencies can also decrease the likelihood of drug-resistance development. Given that host proteostasis systems are heavily used by other RNA viruses for infection, detailed research into CoV-proteostasis interactions may also be widely applicable for developing broad spectrum antivirals that target these dependencies.

1.5   References

1. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579,** 265–269 (2020).

2. Wu, A. *et al.* Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **27,** 325–328 (2020).

3. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. at <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>;

4. Novel Coronavirus(2019-nCoV) Situation Report – 22. at <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf>;

5. Viruses, C. S. G. of the I. C. on T. of *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5,** 536–544 (2020).

6. Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J. & Prescott, H. C. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19). *JAMA* **324,** 782–793 (2020).

7. COVID-19 deaths | WHO COVID-19 dashboard. at <https://data.who.int/dashboards/covid19/deaths >;

8. G., K. T. *et al.* A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **348,** 1953–1966 (2003).

9. Cherry, J. D. & Krogstad, P. SARS: The First Pandemic of the 21st Century. *Pediatr. Res.* **56,** 1–5 (2004).

10. SARS | Basics Factsheet | CDC. at <https://www.cdc.gov/sars/about/fs-sars.html>;

11. M., Z. A., Sander, van B., M., B. T., D.M.E., O. A. & A.M., F. R. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367,** 1814–1820 (2012).

12. MERS-CoV worldwide overview. at <https://www.ecdc.europa.eu/en/middle-east-respiratory-syndrome-coronavirus-mers-cov-situation-update>

13. Memish, Z. A. *et al.* Middle East Respiratory Syndrome Coronavirus in Bats, Saudi Arabia - Volume 19, Number 11—November 2013 - Emerging Infectious Diseases journal - CDC. *Emerg. Infect. Dis.* **19,** 1819–1823 (2013).

14. Haagmans, B. L. *et al.* Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *Lancet Infect. Dis.* **14,** 140–145 (2014).

15. Lau, S. K. P. *et al.* Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci.* **102,** 14040–14045 (2005).

16. Li, W. *et al.* Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* **310,** 676–679 (2005).

17. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579,** 270–273 (2020).

18. Lau, S. K. P. *et al.* Early Release - Possible Bat Origin of Severe Acute Respiratory Syndrome Coronavirus 2 - Volume 26, Number 7—July 2020 - Emerging Infectious Diseases journal - CDC. *Emerg. Infect. Dis.* **26,** 1542–1547 (2020).

19. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26,** 450–452 (2020).

20. Alwine, J. C., Casadevall, A., Enquist, L. W., Goodrum, F. D. & Imperiale, M. J. A Critical Analysis of the Evidence for the SARS-CoV-2 Origin Hypotheses. *mBio* **14,** e00583-23 (2023).

21. Menachery, V. D., Graham, R. L. & Baric, R. S. Jumping species—a mechanism for coronavirus persistence and survival. *Curr Opin Virol* **23,** 1–7 (2017).

22. Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nat. Rev. Microbiol.* **20,** 299–314 (2022).

23. Hamre, D. & Procknow, J. J. A New Virus Isolated from the Human Respiratory Tract.∗. *Proc. Soc. Exp. Biol. Med.* **121,** 190–193 (1966).

24. Hoek, L. van der *et al.* Identification of a new human coronavirus. *Nat. Med.* **10,** 368–373 (2004).

25. McIntosh, K., Dees, J. H., Becker, W. B., Kapikian, A. Z. & Chanock, R. M. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci.* **57,** 933–940 (1967).

26. Woo, P. C. Y. *et al.* Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J. Virol.* **79,** 884–895 (2005).

27. Boursnell, M. E. G. *et al.* Completion of the Sequence of the Genome of the Coronavirus Avian Infectious Bronchitis Virus. *J. Gen. Virol.* **68,** 57–77 (1987).

28. Cheever, F. S., Daniels, J. B., Pappenheimer, A. M. & Bailey, O. T. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin. *J. Exp. Med.* **90,** 181–194 (1949).

29. Weiss, S. R. & Leibowitz, J. L. Chapter 4 Coronavirus Pathogenesis. *Adv Virus Res* **81,** 85–164 (2011).

30. Körner, R. W., Majjouti, M., Alcazar, M. A. A. & Mahabir, E. Of Mice and Men: The Coronavirus MHV and Mouse Models as a Translational Approach to Understand SARS-CoV-2. *Viruses* **12,** 880 (2020).

31. Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* **7,** 439–450 (2009).

32. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19,** 155–170 (2021).

33. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426,** 450–454 (2003).

34. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181,** 271-280.e8 (2020).

35. Hofmann, H. *et al.* Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci.* **102,** 7988–7993 (2005).

36. Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495,** 251–254 (2013).

37. Williams, R. K., Jiang, G. S. & Holmes, K. V. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc. Natl. Acad. Sci.* **88,** 5533–5536 (1991).

38. Denison, M. R. *et al.* Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. *Virology* **189,** 274–284 (1992).

39. Ghosh, S. *et al.* β-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell* **183,** 1520-1535.e14 (2020).

40. Xu, W. *et al.* Compartmentalization-aided interaction screening reveals extensive high-order complexes within the SARS-CoV-2 proteome. *Cell Reports* **36,** 109482 (2021).

41. Brunn, A. von *et al.* Analysis of Intraviral Protein-Protein Interactions of the SARS Coronavirus ORFeome. *Plos One* **2,** e459 (2007).

42. Angelini, M. M., Akhlaghpour, M., Neuman, B. W. & Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. *Mbio* **4,** e00524-13 (2013).

43. Oostra, M. *et al.* Topology and Membrane Anchoring of the Coronavirus Replication Complex: Not All Hydrophobic Domains of nsp3 and nsp6 Are Membrane Spanning ▽ †. *J Virol* **82,** 12392–12405 (2008).

44. Oudshoorn, D. *et al.* Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication. *Mbio* **8,** e01658-17 (2017).

45. Tabata, K. *et al.* Convergent use of phosphatidic acid for hepatitis C virus and SARS-CoV-2 replication organelle formation. *Nat Commun* **12,** 7276 (2021).

46. Ricciardi, S. *et al.* The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature* **606,** 761–768 (2022).

47. Taha, T. Y. *et al.* A single inactivating amino acid change in the SARS-CoV-2 NSP3 Mac1 domain attenuates viral replication and pathogenesis in vivo. (2023). doi:10.1101/2023.04.18.537104

48. Taha, T. Y. *et al.* Rapid assembly of SARS-CoV-2 genomes reveals attenuation of the Omicron BA.1 variant through NSP6. *Nat Commun* **14,** 2308 (2023).

49. Lin, X. *et al.* The NSP4 T492I mutation increases SARS-CoV-2 infectivity by altering non-structural protein cleavage. *Cell Host Microbe* **31,** 1170-1184.e7 (2023).

50. Snijder, E. J., Decroly, E. & Ziebuhr, J. Chapter Three The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv. Virus Res.* **96,** 59–126 (2016).

51. Park, G. J. *et al.* The mechanism of RNA capping by SARS-CoV-2. *Nature* **609,** 793–800 (2022).

52. Moeller, N. H. *et al.* Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease ExoN. *Proc. Natl. Acad. Sci.* **119,** e2106379119 (2022).

53. Gribble, J. *et al.* The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* **17,** e1009226 (2021).

54. Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S. & Denison, M. R. The nsp2 Replicase Proteins of Murine Hepatitis Virus and Severe Acute Respiratory Syndrome Coronavirus Are Dispensable for Viral Replication. *Journal of Virology* 13399–13411 (2005). doi:10.1128/jvi.79.21.13399-13411.2005

55. Angeletti, S. *et al.* COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol* **92,** 584–588 (2020).

56. Freeman, M. C., Graham, R. L., Lu, X., Peek, C. T. & Denison, M. R. Coronavirus Replicase-Reporter Fusions Provide Quantitative Analysis of Replication and Replication Complex Formation. *J Virol* **88,** 5319–5327 (2014).

57. Hagemeijer, M. C. *et al.* Dynamics of Coronavirus Replication-Transcription Complexes. *J Virol* **84,** 2134–2149 (2009).

58. Prentice, E., McAuliffe, J., Lu, X., Subbarao, K. & Denison, M. R. Identification and Characterization of Severe Acute Respiratory Syndrome Coronavirus Replicase Proteins. *J. Virol.* **78,** 9977–9986 (2004).

59. Slavin, M. *et al.* Targeted in situ cross-linking mass spectrometry and integrative modeling reveal the architectures of three proteins from SARS-CoV-2. *Proc. Natl. Acad. Sci.* **118,** e2103554118 (2021).

60. Ma, J., Chen, Y., Wu, W. & Chen, Z. Structure and Function of N-Terminal Zinc Finger Domain of SARS-CoV-2 NSP2. *Virol. Sin.* **36,** 1104–1112 (2021).

61. Gupta, M. *et al.* CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. *bioRxiv* 2021.05.10.443524 (2021). doi:10.1101/2021.05.10.443524

62. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antivir Res* **149,** 58–74 (2018).

63. Báez-Santos, Y. M., John, S. E. St. & Mesecar, A. D. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antivir Res* **115,** 21–38 (2015).

64. Clementz, M. A. *et al.* Deubiquitinating and Interferon Antagonism Activities of Coronavirus Papain-Like Proteases. *J. Virol.* **84,** 4619–4629 (2010).

65. Jiang, H., Yang, P. & Zhang, J. Potential Inhibitors Targeting Papain-Like Protease of SARS-CoV-2: Two Birds With One Stone. *Front. Chem.* **10,** 822785 (2022).

66. Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369,** 1395–1398 (2020).

67. Zimmermann, L. *et al.* SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a pore spanning replication organelle. *Nat. Commun.* **14,** 7894 (2023).

68. Gadlage, M. J. *et al.* Murine Hepatitis Virus Nonstructural Protein 4 Regulates Virus-Induced Membrane Modifications and Replication Complex Function ▿ . *J Virol* **84,** 280–290 (2010).

69. Oostra, M. *et al.* Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication ▿ . *J Virol* **81,** 12323–12336 (2007).

70. Sparks, J. S., Lu, X. & Denison, M. R. Genetic Analysis of Murine Hepatitis Virus nsp4 in Virus Replication. *J Virol* **81,** 12554–12563 (2007).

71. Hagemeijer, M. C. *et al.* Mobility and Interactions of Coronavirus Nonstructural Protein 4. *J. Virol.* **85,** 4572–4577 (2011).

72. Clementz, M. A., Kanjanahaluethai, A., O'Brien, T. E. & Baker, S. C. Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles. *Virology* **375,** 118–129 (2008).

73. Beachboard, D. C., Anderson-Daniels, J. M. & Denison, M. R. Mutations across Murine Hepatitis Virus nsp4 Alter Virus Fitness and Membrane Modifications. *J Virol* **89,** 2080–2089 (2015).

74. Adhikari, S. *et al.* A high-stringency blueprint of the human proteome. *Nat. Commun.* **11,** 5301 (2020).

75. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting Proteostasis for Disease Intervention. *Science* **319,** 916–919 (2008).

76. Wright, M. T., Timalsina, B., Lopez, V. G. & Plate, L. Time-Resolved Interactome Profiling Deconvolutes Secretory Protein Quality Control Dynamics. *bioRxiv* 2022.09.04.506558 (2022). doi:10.1101/2022.09.04.506558

77. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11,** 113–127 (2010).

78. Joazeiro, C. A. P. Mechanisms and functions of ribosome-associated protein quality control. *Nat. Rev. Mol. Cell Biol.* **20,** 368–383 (2019).

79. Walter, P., Gilmore, R. & Blobel, G. Protein translocation across the endoplasmic reticulum. *Cell* **38,** 5–8 (1984).

80. Walter, P. & Blobel, G. Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. *Proc. Natl. Acad. Sci.* **77,** 7112–7116 (1980).

81. Guna, A., Volkmar, N., Christianson, J. C. & Hegde, R. S. The ER membrane protein complex is a transmembrane domain insertase. *Science* **359,** 470–473 (2018).

82. Shurtleff, M. J. *et al.* The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7,** e37018 (2018).

83. Kleizen, B. & Braakman, I. Protein folding and quality control in the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **16,** 343–349 (2004).

84. Khan, H. A. & Mutus, B. Protein disulfide isomerase a multifunctional protein with multiple physiological roles. *Front Chem* **2,** 70 (2014).

85. Breitling, J. & Aebi, M. N-Linked Protein Glycosylation in the Endoplasmic Reticulum. *Cold Spring Harb. Perspect. Biol.* **5,** a013359 (2013).

86. Hammond, C., Braakman, I. & Helenius, A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl. Acad. Sci.* **91,** 913–917 (1994).

87. Totani, K., Ihara, Y., Matsuo, I. & Ito, Y. Substrate Specificity Analysis of Endoplasmic Reticulum Glucosidase II Using Synthetic High Mannose-type Glycans*. *J. Biol. Chem.* **281,** 31502–31508 (2006).

88. Deprez, P., Gautschi, M. & Helenius, A. More Than One Glycan Is Needed for ER Glucosidase II to Allow Entry of Glycoproteins into the Calnexin/Calreticulin Cycle. *Mol. Cell* **19,** 183–195 (2005).

89. D'Alessio, C., Caramelo, J. J. & Parodi, A. J. UDP-GlC:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin. Cell Dev. Biol.* **21,** 491–499 (2010).

90. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4,** 181–191 (2003).

91. Adams, B. M., Canniff, N. P., Guay, K. P., Larsen, I. S. B. & Hebert, D. N. Quantitative glycoproteomics reveals cellular substrate selectivity of the ER protein quality control sensors UGGT1 and UGGT2. *eLife* **9,** e63997 (2020).

92. Guay, K. P. *et al.* ER chaperones use a protein folding and quality control glyco-code. *Mol. Cell* (2023). doi:10.1016/j.molcel.2023.11.006

93. Schallus, T., Fehér, K., Sternberg, U., Rybin, V. & Muhle-Goll, C. Analysis of the specific interactions between the lectin domain of malectin and diglucosides. *Glycobiology* **20,** 1010–1020 (2010).

94. Schallus, T. *et al.* Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N-Glycosylation. *Mol Biol Cell* **19,** 3404–3414 (2008).

95. Galli, C., Bernasconi, R., Soldà, T., Calanca, V. & Molinari, M. Malectin Participates in a Backup Glycoprotein Quality Control Pathway in the Mammalian ER. *Plos One* **6,** e16304 (2011).

96. Chen, Y. *et al.* Role of malectin in Glc2Man9GlcNAc2-dependent quality control of α1-antitrypsin. *Mol Biol Cell* **22,** 3559–3570 (2011).

97. Takeda, K., Qin, S.-Y., Matsumoto, N. & Yamamoto, K. Association of malectin with ribophorin I is crucial for attenuation of misfolded glycoprotein secretion. *Biochem Bioph Res Co* **454,** 436–440 (2014).

98. Qin, S.-Y. *et al.* Malectin Forms a Complex with Ribophorin I for Enhanced Association with Misfolded Glycoproteins*. *J Biol Chem* **287,** 38080–38089 (2012).

99. Ramírez, A. S., Kowal, J. & Locher, K. P. Cryo–electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science* **366,** 1372–1375 (2019).

100. Wilson, C. M., Roebuck, Q. & High, S. Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proc National Acad Sci* **105,** 9534–9539 (2008).

101. Fenech, E. J. *et al.* Interaction mapping of endoplasmic reticulum ubiquitin ligases identifies modulators of innate immune signalling. *Elife* **9,** e57306 (2020).

102. Ciechanover, A. & Schwartz, A. L. The ubiquitin-proteasome pathway: The complexity and myriad functions of proteins death. *Proc. Natl. Acad. Sci.* **95,** 2727–2730 (1998).

103. Pla-Prats, C. & Thomä, N. H. Quality control of protein complex assembly by the ubiquitin–proteasome system. *Trends Cell Biol.* **32,** 696–706 (2022).

104. Thibault, G. & Ng, D. T. W. The Endoplasmic Reticulum-Associated Degradation Pathways of Budding Yeast. *Csh Perspect Biol* **4,** a013193 (2012).

105. Takeshige, K., Baba, M., Tsuboi, S., Noda, T. & Ohsumi, Y. Autophagy in yeast demonstrated with proteinase-deficient mutants and conditions for its induction. *J. cell Biol.* **119,** 301–311 (1992).

106. Molinari, M. ER-phagy responses in yeast, plants, and mammalian cells and their crosstalk with UPR and ERAD. *Dev Cell* **56,** 949–966 (2021).

107. Grumati, P. *et al.* Full length RTN3 regulates turnover of tubular endoplasmic reticulum via selective autophagy. *eLife* **6,** e25555 (2017).

108. Kohno, S., Shiozaki, Y., Keenan, A. L., Miyazaki-Anzai, S. & Miyazaki, M. An N-terminal–truncated isoform of FAM134B (FAM134B-2) regulates starvation-induced hepatic selective ER-phagy. *Life Sci. Alliance* **2,** e201900340 (2019).

109. Smith, M. D. *et al.* CCPG1 Is a Non-canonical Autophagy Cargo Receptor Essential for ER-Phagy and Pancreatic ER Proteostasis. *Dev Cell* **44,** 217-232.e11 (2018).

110. An, H. *et al.* TEX264 Is an Endoplasmic Reticulum-Resident ATG8-Interacting Protein Critical for ER Remodeling during Nutrient Stress. *Mol. Cell* **74,** 891-908.e10 (2019).

111. Chino, H., Hatta, T., Natsume, T. & Mizushima, N. Intrinsically Disordered Protein TEX264 Mediates ER-phagy. *Mol Cell* **74,** 909-921.e6 (2019).

112. Fumagalli, F. *et al.* Translocon component Sec62 acts in endoplasmic reticulum turnover during stress recovery. *Nat. Cell Biol.* **18,** 1173–1184 (2016).

113. Hetz, C., Zhang, K. & Kaufman, R. J. Mechanisms, regulation and functions of the unfolded protein response. *Nat Rev Mol Cell Bio* **21,** 421–438 (2020).

114. Szegezdi, E., Logue, S. E., Gorman, A. M. & Samali, A. Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Rep.* **7,** 880–885 (2006).

115. Shoulders, M. D. *et al.* Stress-Independent Activation of XBP1s and/or ATF6 Reveals Three Functionally Diverse ER Proteostasis Environments. *Cell Reports* **3,** 1279–1292 (2013).

116. Walter, P. & Ron, D. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science* **334,** 1081–1086 (2011).

117. Shah, P. S. *et al.* Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **175,** 1931-1945.e18 (2018).

118. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583,** 459–468 (2020).

119. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370,** eabe9403 (2020).

120. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583,** 469–472 (2020).

121. Stukalov, A. *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594,** 246–252 (2021).

122. Cornillez-Ty, C. T., Liao, L., Yates, J. R., Kuhn, P. & Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling ▽ . *J Virol* **83,** 10314–10318 (2009).

123. Christie, M. & Igreja, C. eIF4E-homologous protein (4EHP): a multifarious cap-binding protein. *FEBS J.* **290,** 266–285 (2021).

124. Chapat, C. *et al.* Cap-binding protein 4EHP effects translation silencing by microRNAs. *Proc. Natl. Acad. Sci.* **114,** 5425–5430 (2017).

125. Morita, M. *et al.* A Novel 4EHP-GIGYF2 Translational Repressor Complex Is Essential for Mammalian Development. *Mol. Cell. Biol.* **32,** 3585–3593 (2012).

126. Davies, J. P., Almasy, K. M., McDonald, E. F. & Plate, L. Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *Acs Infect Dis* **6,** 3174–3189 (2020).

127. Xu, Z. *et al.* SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc National Acad Sci* **119,** e2204539119 (2022).

128. Zou, L., Moch, C., Graille, M. & Chapat, C. The SARS-CoV-2 protein NSP2 impairs the silencing capacity of the human 4EHP-GIGYF2 complex. *Iscience* **25,** 104646 (2022).

129. Hoffmann, H.-H. *et al.* Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors. *Cell Host Microbe* **29,** 267-280.e5 (2021).

130. Juszkiewicz, S., Speldewinde, S. H., Wan, L., Svejstrup, J. Q. & Hegde, R. S. The ASC-1 Complex Disassembles Collided Ribosomes. *Mol. Cell* **79,** 603-614.e8 (2020).

131. Kuroha, K., Zinoviev, A., Hellen, C. U. T. & Pestova, T. V. Release of Ubiquitinated and Non-ubiquitinated Nascent Chains from Stalled Mammalian Ribosomal Complexes by ANKZF1 and Ptrh1. *Mol. Cell* **72,** 286-302.e8 (2018).

132. Zilmer, M. *et al.* Novel congenital disorder of O -linked glycosylation caused by GALNT2 loss of function. *Brain* **143,** 1114–1126 (2020).

133. V'kovski, P. *et al.* Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8,** e42037 (2019).

134. Twu, W.-I. *et al.* Contribution of autophagy machinery factors to HCV and SARS-CoV-2 replication organelle formation. *Cell Reports* **37,** 110049 (2021).

135. Graham, R. L. & Denison, M. R. Replication of Murine Hepatitis Virus Is Regulated by Papain-Like Proteinase 1 Processing of Nonstructural Proteins 1, 2, and 3. *J Virol* **80,** 11610–11620 (2006).

136. Zhang, M. *et al.* ISGylation in Innate Antiviral Immunity and Pathogen Defense Responses: A Review. *Front. Cell Dev. Biol.* **9,** 788410 (2021).

137. Lindner, H. A. *et al.* The Papain-Like Protease from the Severe Acute Respiratory Syndrome Coronavirus Is a Deubiquitinating Enzyme. *J. Virol.* **79,** 15199–15208 (2005).

138. Bailey-Elkin, B. A. *et al.* Crystal structure of the Middle East respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of deubiquitinating activity to demonstrate its role in innate immune suppression. *J. Biol. Chem.* **289,** 34667–82 (2014).

139. Shin, D. *et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587,** 657–662 (2020).

140. Lei, X. *et al.* Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun* **11,** 3810 (2020).

141. Armstrong, L. A. *et al.* Biochemical characterization of protease activity of Nsp3 from SARS-CoV-2 and its inhibition by nanobodies. *PLoS ONE* **16,** e0253364 (2021).

142. Prentice, E., Jerome, W. G., Yoshimori, T., Mizushima, N. & Denison, M. R. Coronavirus Replication Complex Formation Utilizes Components of Cellular Autophagy*. *J Biol Chem* **279,** 10136–10141 (2004).

143. Reggiori, F. *et al.* Coronaviruses Hijack the LC3-I-Positive EDEMosomes, ER-Derived Vesicles Exporting Short-Lived ERAD Regulators, for Replication. *Cell Host Microbe* **7,** 500–508 (2010).

144. Williams, J. M., Chen, Y.-J., Cho, W. J., Tai, A. W. & Tsai, B. Reticulons promote formation of ER-derived double-membrane vesicles that facilitate SARS-CoV-2 replication. *J Cell Biol* **222,** e202203060 (2023).

145. Morita, K. *et al.* Genome-wide CRISPR screen identifies TMEM41B as a gene required for autophagosome formation. *J. Cell Biol.* **217,** 3817–3828 (2018).

146. Shoemaker, C. J. *et al.* CRISPR screening using an expanded toolkit of autophagy reporters identifies TMEM41B as a novel autophagy factor. *PLoS Biol.* **17,** e2007044 (2019).

147. Schneider, W. M. *et al.* Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks. *Cell* **184,** 120-132.e14 (2021).

148. Ji, M. *et al.* VMP1 and TMEM41B are essential for DMV formation during β-coronavirus infection. *J Cell Biol* **221,** e202112081 (2022).

149. Pahmeier, F. *et al.* Identification of host dependency factors involved in SARS-CoV-2 replication organelle formation through proteomics and ultrastructural analysis. *J. Virol.* **97,** e00878-23 (2023).

150. Sakabe, I., Hu, R., Jin, L., Clarke, R. & Kasid, U. N. TMEM33: a new stress-inducible endoplasmic reticulum transmembrane protein and modulator of the unfolded protein response signaling. *Breast Cancer Res Tr* **153,** 285–297 (2015).

151. Li, J. *et al.* The REEP5/TRAM1 complex binds SARS-CoV-2 NSP3 and promotes virus replication. *J. Virol.* **97,** e00507-23 (2023).

152. Hu, J. *et al.* Membrane Proteins of the Endoplasmic Reticulum Induce High-Curvature Tubules. *Science* **319,** 1247–1250 (2008).

153. Görlich, D., Hartmann, E., Prehn, S. & Rapoport, T. A. A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* **357,** 47–52 (1992).

154. Rose, A. E., Zhao, C., Turner, E. M., Steyer, A. M. & Schlieker, C. Arresting a Torsin ATPase Reshapes the Endoplasmic Reticulum*. *J. Biol. Chem.* **289,** 552–564 (2014).

155. Huffaker, T. C. & Robbins, P. W. Temperature-sensitive yeast mutants deficient in asparagine-linked glycosylation. *J. Biol. Chem.* **257,** 3203–10 (1982).

156. Nufer, O., Kappeler, F., Guldbrandsen, S. & Hauri, H.-P. ER export of ERGIC-53 is controlled by cooperation of targeting determinants in all three of its domains. *J. Cell Sci.* **116,** 4429–4440 (2003).

157. Annunen, P. *et al.* Cloning of the Human Prolyl 4-Hydroxylase α Subunit Isoform α(II) and Characterization of the Type II Enzyme Tetramer THE α(I) AND α(II) SUBUNITS DO NOT FORM A MIXED α(I)α(II)β2 TETRAMER*. *J. Biol. Chem.* **272,** 17342–17348 (1997).

158. Pfefferle, S. *et al.* The SARS-Coronavirus-Host Interactome: Identification of Cyclophilins as Target for Pan-Coronavirus Inhibitors. *Plos Pathog* **7,** e1002331 (2011).

159. Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K. G. & Makino, S. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat. Struct. Mol. Biol.* **16,** 1134–1140 (2009).

160. Tanaka, T., Kamitani, W., DeDiego, M. L., Enjuanes, L. & Matsuura, Y. Severe Acute Respiratory Syndrome Coronavirus nsp1 Facilitates Efficient Propagation in Cells through a Specific Translational Shutoff of Host mRNA. *J. Virol.* **86,** 11128–11137 (2012).

161. Banerjee, A. K. *et al.* SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* **183,** 1325-1339.e21 (2020).

162. Aviner, R. *et al.* SARS-CoV-2 Nsp1 regulates translation start site fidelity to promote infection. *bioRxiv* 2023.07.05.547902 (2023). doi:10.1101/2023.07.05.547902

163. Hayashi, T. & Su, T.-P. Sigma-1 Receptor Chaperones at the ER- Mitochondrion Interface Regulate Ca2+ Signaling and Cell Survival. *Cell* **131,** 596–610 (2007).

164. Mori, T., Hayashi, T., Hayashi, E. & Su, T.-P. Sigma-1 Receptor Chaperone at the ER-Mitochondrion Interface Mediates the Mitochondrion-ER-Nucleus Signaling for Cellular Survival. *PLoS ONE* **8,** e76941 (2013).

165. Abdelkader, A. *et al.* In-Silico targeting of SARS-CoV-2 NSP6 for drug and natural products repurposing. *Virology* **573,** 96–110 (2022).

166. Daniloski, Z. *et al.* Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells. *Cell* **184,** 92-105.e16 (2021).

167. Aviner, R. & Frydman, J. Proteostasis in Viral Infection: Unfolding the Complex Virus–Chaperone Interplay. *Csh Perspect Biol* **12,** a034090 (2019).

168. Fung, T. S. & Liu, D. X. Coronavirus infection, ER stress, apoptosis and innate immunity. *Front Microbiol* **5,** 296 (2014).

169. Echavarría-Consuegra, L. *et al.* Manipulation of the unfolded protein response: A pharmacological strategy against coronavirus infection. *Plos Pathog* **17,** e1009644 (2021).

170. Bechill, J., Chen, Z., Brewer, J. W. & Baker, S. C. Coronavirus Infection Modulates the Unfolded Protein Response and Mediates Sustained Translational Repression ▽. *J Virol* **82,** 4492–4501 (2008).

171. Minakshi, R. *et al.* The SARS Coronavirus 3a Protein Causes Endoplasmic Reticulum Stress and Induces Ligand-Independent Downregulation of the Type 1 Interferon Receptor. *PLoS ONE* **4,** e8342 (2009).

172. Sung, S.-C., Chao, C.-Y., Jeng, K.-S., Yang, J.-Y. & Lai, M. M. C. The 8ab protein of SARS-CoV is a luminal ER membrane-associated protein and induces the activation of ATF6. *Virology* **387,** 402–413 (2009).

173. Almasy, K. M., Davies, J. P. & Plate, L. Comparative Host Interactomes of the SARS-CoV-2 Nonstructural Protein 3 and Human Coronavirus Homologs. *Mol Cell Proteomics* **20,** 100120 (2021).

174. Davies, J. P. *et al.* Expression of SARS-CoV-2 Nonstructural Proteins 3 and 4 Can Tune the Unfolded Protein Response in Cell Culture. *J. Proteome Res.* (2023). doi:10.1021/acs.jproteome.3c00600

CHAPTER 2

COMPARATIVE MULTIPLEXED INTERACTOMICS OF SARS-COV-2 AND HOMOLOGOUS CORONAVIRUS NONSTRUCTURAL PROTEINS 2 AND 4

This chapter is adapted from "Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies" published in *ACS Infectious Disease* and has been reproduced with the permission of the publisher and my co-authors: Katherine M. Almasy, Eli F. McDonald, Lars Plate.

2.1   Summary
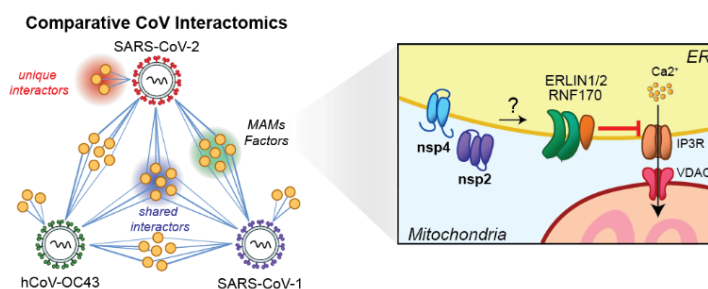
Human coronaviruses (hCoV) have become a threat to global health and society, as evident from the SARS outbreak in 2002 caused by SARS-CoV-1 and the most recent COVID-19 pandemic caused by SARS-CoV-2. Despite high sequence similarity between SARS-CoV-1 and -2, each strain has distinctive virulence. A better understanding of the basic molecular mechanisms mediating changes in virulence is needed. Here, we profile the virus-host protein-protein interactions of two hCoV non-structural proteins (nsps) that are critical for virus replication. We use tandem mass tag-multiplexed quantitative proteomics to sensitively compare and contrast the interactomes of nsp2 and nsp4 from three betacoronavirus strains: SARS-CoV-1, SARS-CoV-2, and hCoV-OC43 – an endemic strain associated with the common cold. This approach enables the identification of both unique and shared host cell protein binding partners and the ability to further compare the enrichment of common interactions across homologs from related strains. We identify common nsp2 interactors involved in endoplasmic reticulum (ER) Ca^{2+} signaling and mitochondria biogenesis. We also identify nsp4 interactors unique to each strain, such as E3 ubiquitin ligase complexes for SARS-CoV-1 and ER homeostasis factors for SARS-CoV-2. Common

nsp4 interactors include *N*-linked glycosylation machinery, unfolded protein response (UPR) associated proteins, and anti-viral innate immune signaling factors. Both nsp2 and nsp4 interactors are strongly enriched in proteins localized at mitochondrial-associated ER membranes suggesting a new functional role for modulating host processes, such as calcium homeostasis, at these organelle contact sites. Our results shed light on the role these hCoV proteins play in the infection cycle, as well as host factors that may mediate the divergent pathogenesis of OC43 from SARS strains. Our mass spectrometry workflow enables rapid and robust comparisons of multiple bait proteins, which can be applied to additional viral proteins. Furthermore, the identified common interactions may present new targets for exploration by host-directed anti-viral therapeutics.

All supplemental materials can be found in Appendix 1 or accompanying the online publication.

2.2   Introduction

Coronaviruses (CoVs) are positive single-stranded RNA viruses capable of causing human disease with a range of severity. While some strains, such as endemic hCoV-OC43, cause milder common-cold like symptoms, other strains are associated with more severe pathogenesis and higher lethality, including SARS-CoV-1 (emerged in 2002), MERS-CoV (in 2012), and most recently SARS-CoV-2, the causative agent of COVID-19(1,2). Despite the relevance of CoVs for human health, our understanding of the factors governing their divergent pathogenicity remains incomplete. Pathogenicity may be mediated by a variety of factors, including different specificities and affinity for different cell surface receptors such as angiotensin-converting enzyme 2 (ACE2) for SARS-CoV-1 and SARS-CoV-2(1,3) or 9-*O*-acetylated sialic acid for hCoV-OC43(4). CoV strains also engage a variety of host immune processes in infected cells. Pathogenic strains more strongly interfere with interferon I signaling (4,5) and induce apoptosis and pyroptosis (6–9). Ensembles of virus-host protein-protein interactions (PPIs) orchestrate the reprogramming of these processes during infection.

Coronaviruses possess the largest known RNA viral genomes, approximately 30kbp in length. The 5' 20 kb region of the genome encodes for two open reading frames (orf1a/1ab) that produce 16 non-structural proteins (nsp1-nsp16) needed to form the viral replication complex, while the 3' proximal region encodes for the structural proteins and several accessory factors with varying roles (**Figure 2-1A**). Previous protein-protein interaction studies of individual CoV proteins have shed light on their functions in the infected host cells and putative roles during pathogenesis. Yeast-two hybrid studies of coronavirus proteins have identified intraviral interactions(10) and interactions between nsp1 and immunophilins(11), and a proximity-labeling approach was used to determine host proteins concentrated in sites of replication(12).

Affinity purification-mass spectrometry (AP-MS) is a powerful tool to study virus-host interactions and has been used extensively to examine how viruses reorganize host cells(13–16). A prior AP-MS study of SARS-CoV-1 nsp2 identified multiple host interactors including prohibitin 1/2 (PHB1/2)(17). Most notably, Gordon et al. recently profiled host interactors for 26 SARS-CoV-2 proteins(18). While these studies enabled important insight on individual viral protein functions, they focused on single CoV strains, limiting direct cross-strain comparisons.

Here, we sought to profile and compare the host interaction profiles of nsps from multiple hCoVs, namely hCoV-OC43, SARS-CoV-1, and SARS-CoV-2. Through comparative interactomics, we identify both conserved and unique

interactors across various strains. Notably, quantitative analysis of interaction enrichment enables nuanced differentiation between shared interactions for each coronavirus protein. Through this approach we discovered both conserved and novel functions of viral proteins and the pathways by which they manipulate cellular processes. Comparisons across strains may also provide clues into the evolutionary arms race between virus and host proteins to hijack or protect protein-protein interfaces(19). Additionally, identified host dependencies can potentially be exploited as targets for host-directed antiviral therapeutics.

In particular, we focus on the host interactors of nsp2 and nsp4. Nsp2 has been suggested to play a role in modifying the host cell environment, although its precise function remains unknown(17). Nsp2 is dispensable for infection in SARS-CoV-1 (20) and has pronounced amino acid sequence differences across coronavirus strains (**Figure 2-1B, Figure 7-1**). Additionally, early sequence analysis of SARS-CoV-2 identified regions of positive selection pressure in nsp2(21). Given the variability of sequence across strains and the ambiguous function, a comparison of interaction profiles across strains can yield insights into the role of nsp2. In contrast, the role of nsp4, a transmembrane glycoprotein, is better defined, most notably in formation of the double-membrane vesicles associated with replication complexes(22,23). Unlike nsp2, nsp4 has a high degree of sequence similarity across human coronavirus strains (**Figure 2-1B, Figure 7-1**).

In this study, we use affinity purification-proteomics to identify interactors of nsp2 from two human coronaviruses (SARS-CoV-1 and SARS-CoV-2) and interactors of nsp4 from three strains (OC43, SARS-CoV-1, and SARS-CoV-2). Quantitative comparative analysis of nsp2 interactors identifies common protein binding partners, including the ERLIN1/2 complex and prohibitin complex involved in regulation of mitochondrial function and calcium flux at ER-mitochondrial contact sites. We also identify overlapping nsp4 interactors, including *N*-linked glycosylation machinery, UPR associated factors, and anti-viral innate immune signaling proteins. Unique interactors of different nsp4 homologs include E3 ubiquitin ligase complexes for SARS-CoV-1, and ER homeostasis factors for SARS-CoV-2. In particular, we found nsp2 and nsp4 interactors are strongly enriched for mitochondria-associated ER membranes (MAM) factors, suggesting a potential mechanism to affect calcium homeostasis and other host processes at these organelle contact sites.

2.3   Results

2.3.1   Design and validation of expression of CoV nsp2 and nsp4 constructs for affinity purification

The two main open reading frames of the CoV viral genome, orf1a and orf1ab, encode for 16 non-structural proteins which perform a variety of tasks during the infection cycle (**Figure 2-1A**). We focus our analysis on two of these proteins, nsp2 and nsp4. Nsp2 is a less functionally-well understood protein with less than 70% amino acid sequence identity between the SARS-CoV-1 and SARS-CoV-2 homologs (**Figure 2-1B, Figure 7-1**). Nsp4 is a component of the CoV replication complex that is 80 % identical between SARS strains, but only 42% between SARS and OC43 strains, a less clinically severe human CoV (**Figure 2-1B, Figure 7-1**).

To compare the virus-host protein-protein interactions of nsp2 and nsp4 across multiple CoV strains, we designed FLAG-tagged expression constructs for affinity purification (**Figure 2-1C**). SARS-CoV-1 and SARS-CoV-2 nsp2 constructs contain an N-terminal FLAG-tag, while the SARS-CoV-1, SARS-CoV-2, and OC43 nsp4 constructs contain a C-terminal

FLAG-tag. In addition, nsp4 constructs contain a 19 amino acid leader sequence corresponding to the C-terminus of nsp3, which includes the nsp3-PL2^{pro} cleavage site necessary for proper nsp4 translocation into the ER membrane as has been shown previously(24,25). Improper membrane insertion would likely alter the observed interactome as compared to the native state.
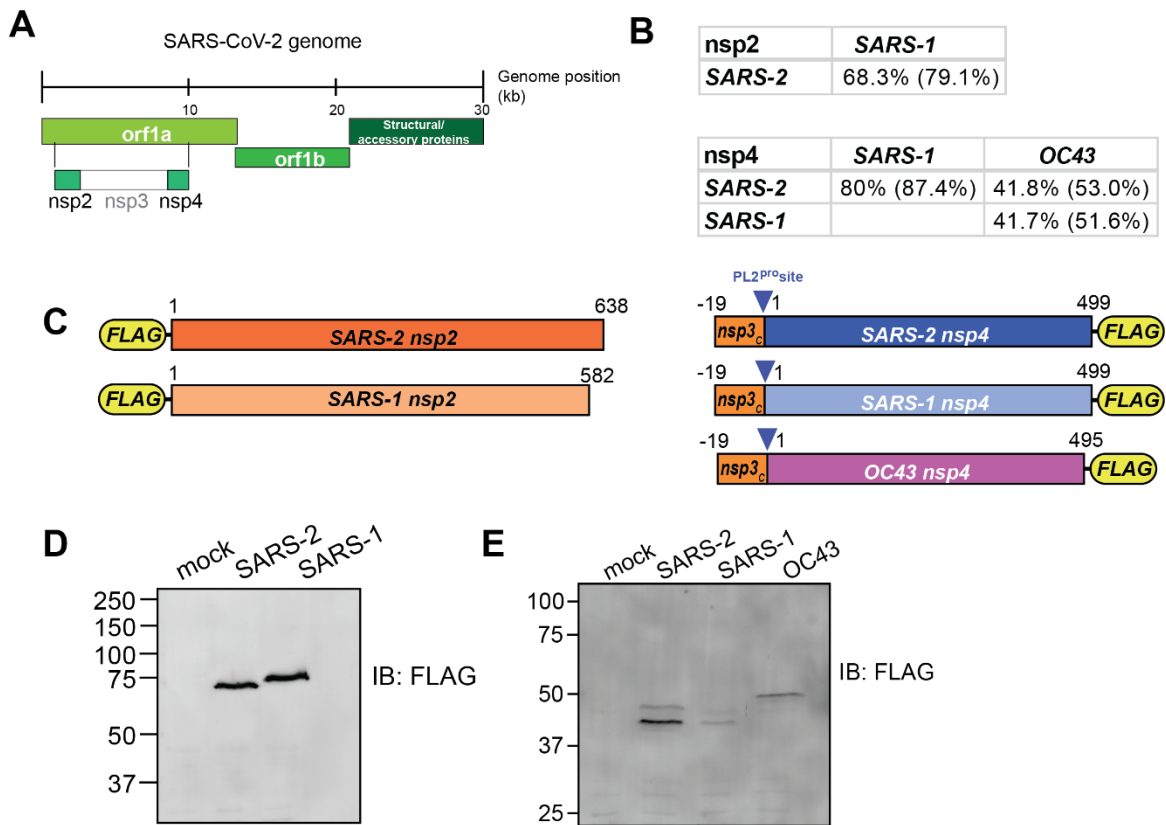
Figure 2-1. Design and validation of expression of CoV nsp2 and nsp4 constructs for affinity purification.
- A. Schematic of SARS-CoV-2 genome organization.
- B. Amino acid sequence identity and similarity (in parentheses) for comparisons of nsp2 and nsp4 homologs. Sequence alignments are shown in **Figure 7-1A-B**.
- C. Nsp2 and nsp4 FLAG-tagged construct designs. Nsp2 constructs contain an N-terminal FLAG-tag. Nsp4 constructs contain a 19 amino acid leader sequence from nsp3 at the N-terminus, including the PL2^{pro} cleavage site, along with a C-terminal FLAG-tag.
- D. Western blot of nsp2 homologs expressed in HEK293T cells. Cells were transiently transfected with FLAG-nsp2. Proteins were detected using an anti-FLAG antibody.
- E. Same as in (D), of nsp4 homologs. Cells were transiently transfected with nsp4-FLAG.

Protein constructs were transiently transfected into HEK293T cells and proteins were detected by immunoblotting for the FLAG-tag. While HEK293T cells are not representative of the primary physiological target tissue, these cells are permissive to infection, and were able to recapitulate strong interactors expected in lung tissue in a prior SARS-CoV-2 interactome study(18). The nsp2 constructs were detectable as a single protein band at the expected molecular weight (**Figure 2-1D**), while nsp4 constructs displayed two distinct bands at a lower size than its expected molecular weight (**Figure 2-1E**). This lower apparent molecular weight was previously reported and the different bands likely correspond to different glycosylation states(18). To ensure nsp4 is expressed fully and the detected products do not correspond to a truncated

protein, we immunopurified the protein using FLAG-agarose beads and analyzed the purified protein by LC-MS. We detected peptide fragments spanning the N- and C-termini with overall sequence coverage of up to 62% (**Figure 7-1**) confirming expression of the full proteins.

2.3.2 Affinity purification-mass spectrometry identifies nsp2 interactors

To identify host cell interaction partners of the distinct CoV non-structural proteins, we employed an affinity purification-mass spectrometry workflow (**Figure 2-2A**). The protein constructs were expressed in HEK293T cells, gently lysed in mild detergent buffer, and co-immunopurified from whole cell lysates using anti-FLAG agarose beads. Virus-host protein complexes were then reduced, alkylated, and trypsin digested. Importantly, we used tandem mass tag (TMT)-based multiplexing using TMTpro-16plex or TMT-11plex for relative quantification of protein abundances. For this purpose, 4 – 6 co-immunoprecipitation (Co-IP) replicates for respective nsp2 homologs were pooled into a single MS run. Co-IPs from mock GFP transfected cells were included to differentiate non-specific background proteins (**Figure 2-2B**). Overall, the data set included three individual MS runs containing 34 Co-IP (SARS-CoV-2 n = 13; SARS-CoV-1 n = 9; GFP (mock) n = 12) (**Figure 7-2A**).

We first determined interactors of the individual nsp2 homologs by comparing the log-transformed TMT intensity differences for prey proteins between bait and GFP samples (**Figure 2-2C-D**). We optimized variable cutoffs for high- and medium-confidence interactors based on their magnitude of enrichment compared to the GFP samples and confidence as defined by adjusted p-values (**Figure 2-2C-D, Figure 7-2B-C**). Using the most stringent cutoff, we identified 6 and 11 high-confidence interactors for SARS-CoV-2 and SARS-CoV-1 nsp2 respectively (**Figure 2-2C-D**). Including medium-confidence interactors, we identified 20 nsp2 interactors for each homolog, including four overlapping proteins, ERLIN1, ERLIN2, RNF170, and TMEM199 (**Figure 2-2E**).

Gene enrichment analysis shows nsp2 interactors are involved in a number of host cell processes, including metabolic processing and transport (**Figure 7-3A**). A number of these interactors are membrane-associated proteins in the ER and nucleus (**Figure 7-3B**). Detailed comparisons of gene set enrichments for individual nsp2 homologs revealed several pathways preferentially enriched for SARS-CoV-1, such as mitochondrial calcium ion transport, protein deacetylation, and negative regulation of gene expression (**Figure 7-3C**). We confirmed by immunofluorescence that SARS-CoV-1 and SARS-CoV-2 nsp2 are largely localized perinuclear and co-localize partially with the ER marker PDIA4 (**Figure 7-3D**). Nsp2 expression appears to be limited to a subset of cells as seen by immunofluorescence staining, indicating low transfection efficiency. This indicates that the identified interactions are occurring in the subset of transfected cells processed in this study. Nonetheless, the transfection efficiency was sufficient to detect nsp2 and protein interaction partners by western blot and mass spectrometry analysis.

To validate our findings, we cross referenced our data set with previous coronavirus interactomics studies. A prior study of expressed SARS-CoV-1 nsp2 in *trans* with SARS-CoV-1 infection in HEK293 cells stably expressing the ACE2 virus receptor. They identified 11 host interactors, five of which overlap with our SARS-CoV-1 list, including GIGYF2, PHB, PHB2, STOML2, and EIF4E2 (17). We also cross referenced our interactors with a recently published SARS-CoV-2 interactomics data set performed in HEK293T cells (18). Interestingly, we identified 18 new interactors, though several of

these share secondary interactions with the proteins identified by Gordon et al. (**Figure 7-4**). In addition, we cross referenced our host interactor data set with tissue- and cell line-specific protein expression data sets to determine interactor expression levels in tissues associated with primary infection (**Figure 7-5**)(26–28). We find that expression of identified interactors is enriched in lung and upper aerodigestive tissues in multiple proteomics datasets, confirming the relevance of these factors to coronavirus tropism.
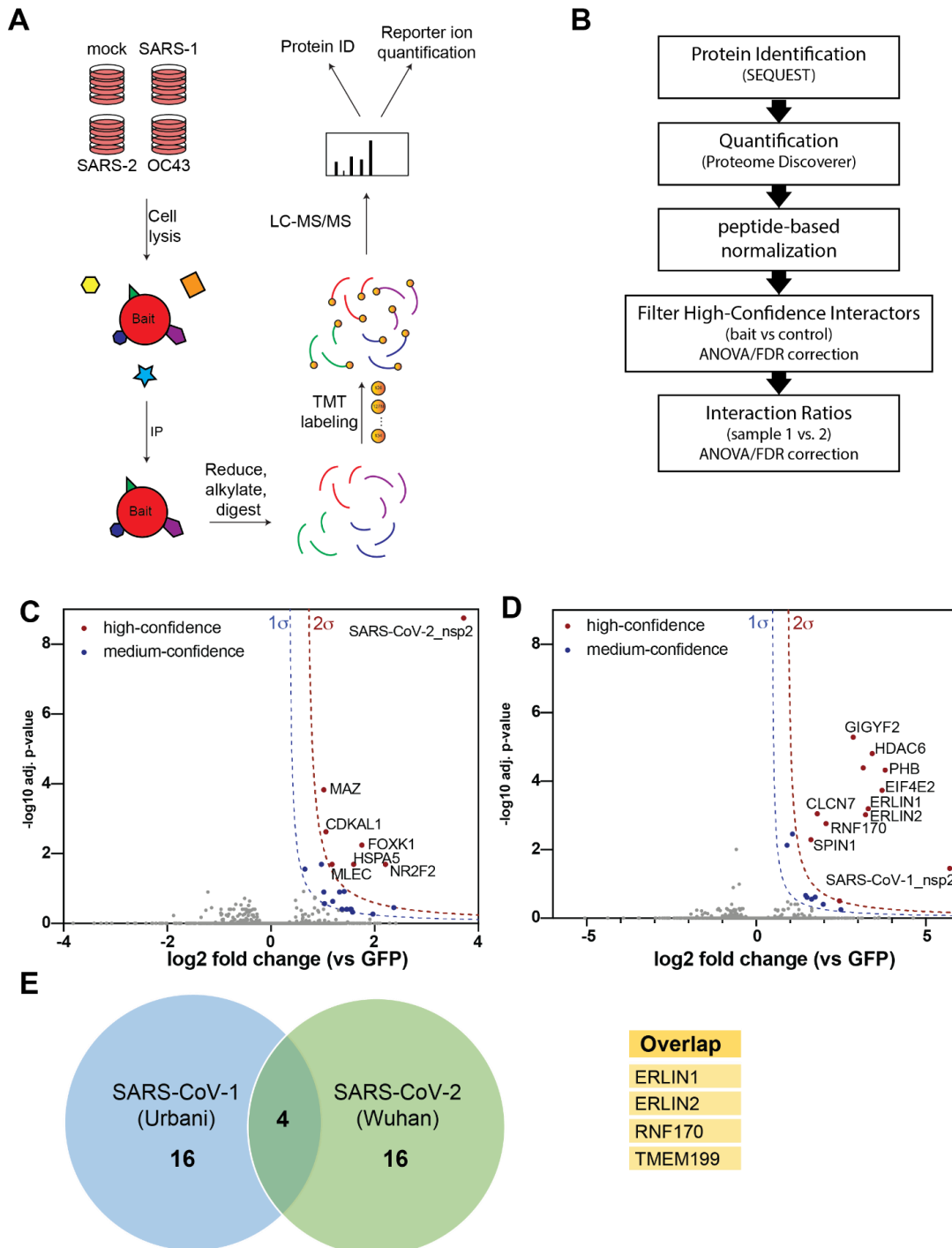
Figure 2-2. Affinity purification-mass spectrometry (AP-MS) identifies nsp2 interactors.
  A. General AP-MS workflow to quantitatively determine interactors of viral nsp homolog. HEK293T cells are transfected with FLAG-tagged expression constructs of nsps as bait or GFP (mock) and lysed. Bait proteins are

immunoprecipitated (IP) along with interacting proteins, reduced, alkylated, and tryptic digested. Peptides are then tandem-mass tag (TMT) labeled, pooled, and analyzed by LC-MS/MS for identification and quantification.

B. Data processing workflow. Peptide spectra are identified and matched to corresponding proteins (SEQUEST HT), then quantified based on TMT reporter ion intensity (Proteome Discoverer 2.4). High confidence interactors are filtered by comparing bait vs control. Interaction ratios between bait homologs are determined (log2 fold change) and adjusted p-value calculated using ANOVA.

C. (C-D) Volcano plot of SARS-CoV-2 nsp2 (C) and SARS-CoV-1 nsp2 (D) datasets to identify medium- and high-confidence interactors. Plotted are log2 TMT intensity fold changes for proteins between nsp2 bait channels and GFP mock transfections versus -log10 adjusted p-values. Curves for the variable cutoffs used to define high-confidence (red) or medium confidence (blue) interactors are shown. $1\sigma = 0.5$

D. Same as in (C), for SARS-CoV-1 nsp2. $1\sigma = 0.43$

E. Venn diagram comparing high-confidence interactors between nsp2 homologs. Sixteen unique proteins were identified each, while four proteins overlapped both data sets (listed in adjacent table).

2.3.3   Quantitative comparison of SARS-CoV-1 and SARS-CoV-2 interactors

Apart from determining nsp2 host cell interactors, we sought to understand to what degree interactions vary between SARS-CoV-1 and SARS-CoV-2. Our multiplexed analysis enabled direct comparison of TMT intensities between the SARS-CoV-1 and SARS-CoV-2 nsp2 Co-IPs (**Figure 2-3A**). We validated that nsp2 bait levels are largely invariable across the replicates, enabling the direct comparison of prey protein intensities (**Figure 7-2D**). We find a subset of interactors are clearly enriched for SARS-CoV-1, including GIGYF2, HDAC8, EIF432, and PHB2 (**Figure 2-3A**). In contrast, several other interactors are enriched more strongly for SARS-CoV-2, for instance FOXK1 and NR2F2.

We performed unbiased hierarchical clustering of the enrichment intensities to group the nsp2 interactors in an unbiased way. This analysis yielded four distinct clusters. Clusters 1 and 2 contained shared interactors between SARS-CoV-1 and SARS-CoV-2 nsp2. On the other hand, clusters 3 and cluster 4 contained proteins that bound exclusively to either SARS-CoV-2 or SARS-CoV-1, respectively (**Figure 2-3B**). To better visualize the relationship between shared and unique nsp2 interactors, we constructed a network plot (**Figure 2-3C**). We also included experimentally validated secondary interactions from the STRING database to group shared and unique interactors into functionally relevant subclusters.

Several of these subclusters are shared between SARS-CoV-1 and SARS-CoV-2 nsp2, for instance one including STOML2, PHB, PHB2, VDAC2. These proteins were previously shown to interact and upregulate formation of metabolically active mitochondrial membranes(29). Another subcluster involves ERLIN1, ERLIN2, and RNF170, which form a known complex regulating ubiquitination and degradation of inositol 1,4,5-triphosphate receptors (IP_3Rs), which in turn are channels regulating Ca^{2+} signaling from the ER to the mitochondria. Consistent with this, we detect mitochondrial calcium ion transmembrane transport as one of the unique biological processes associated with SARS-CoV-1 nsp2, but not SARS-CoV-2 (**Figure 7-3C**). Interestingly, ERLIN1 and ERLIN2 show stronger interactions with SARS-CoV-1 nsp2 than with SARS-CoV-2, indicating some strain-specific preference, which was confirmed by western blot analysis of homolog co-IPs (**Figure 7-6A**). Additional shared interactors include a subunit of the vacuolar ATPase (ATP6AP1) and a regulatory protein (TMEM199), supporting a common role for nsp2 to influence lysosomal processes. Finally, we observe one cytosolic and one ER-resident Hsp70 chaperone (HSPA8, HSPA5) as shared interactors, highlighting their role in nsp2 folding.
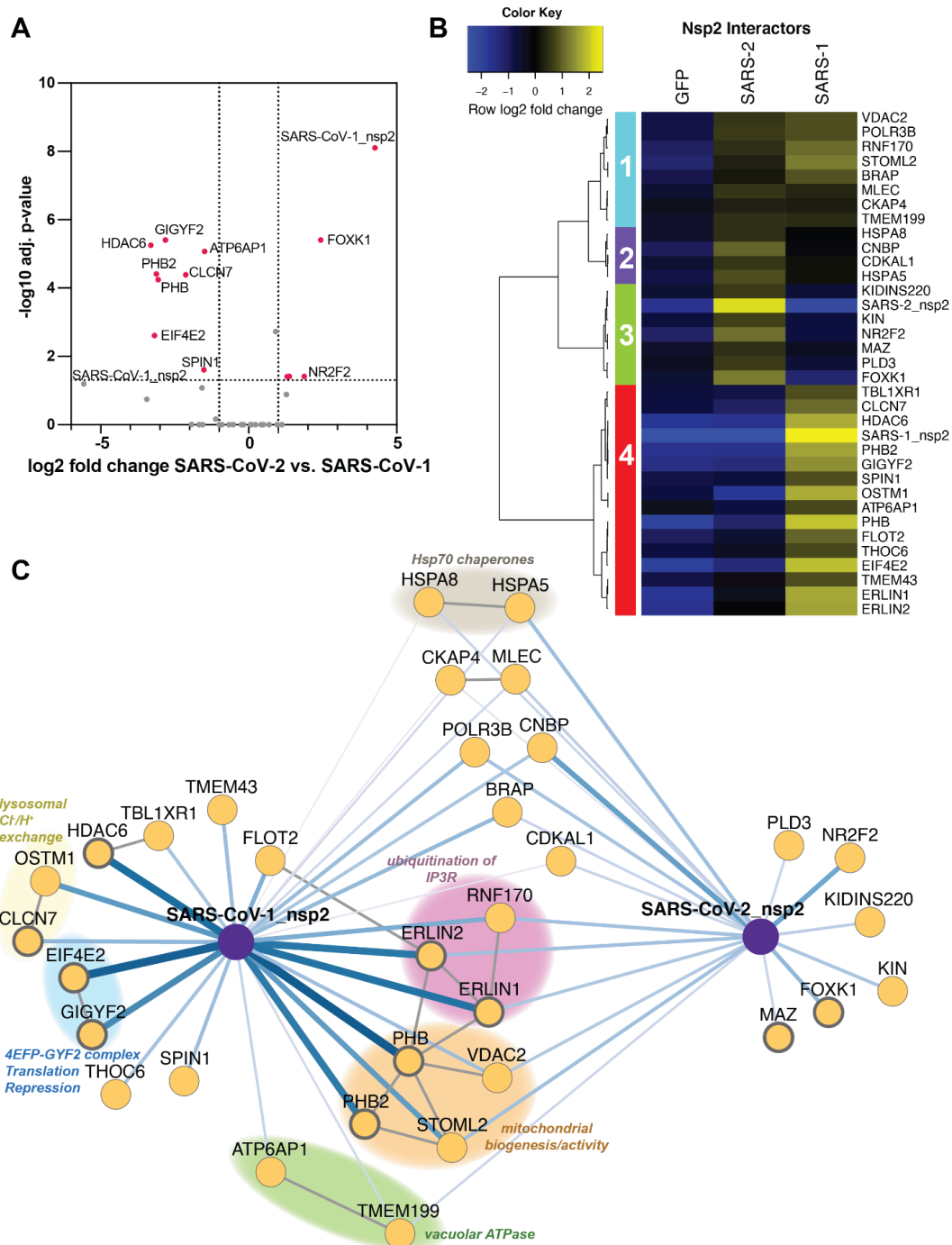
Figure 2-3. Quantitative comparison of SARS-CoV-1 and SARS-CoV-2 nsp2 interactors.

A. Volcano plot comparing interactions between nsp2 homolog from SARS-CoV-1 and SARS-CoV-2. Only high- and medium confidence interactors of nsp2 are shown. Highlighted proteins meet the filter criteria of adjusted p-value < 0.05 and $|\log2$ fold change$| > 1$.

B. Heatmap comparing the enrichment of SARS-CoV-1 and SARS-CoV-2 nsp2 interactors compared to GFP control. log2 fold change is color-coded and centered by row (blue low, yellow high enrichment). Hierarchical clustering using Ward's method shown on the left was carried out on euclidean distances of log2 fold changes scaled by row. Clusters 1 and 2 corresponds to shared interactors of SARS-CoV-1 and -2 nsp2, while cluster 3 and 4 for are unique interactors for SARS-CoV-2 and SARS-CoV-1 nsp2, respectively.

C. Protein-protein interaction (PPI) network map of nsp2 homologs. Blue lines indicate viral-host PPIs, where line width corresponds to fold enrichment compared to the GFP control. Grey lines indicate annotated host-host PPIs in STRING (score > 0.75). Groups of interactors with a common functional role are highlighted.

34

Unique SARS-CoV-2 interactors include FOXK1 and NR2F2, both of which are anti-viral transcription factors induced in response to other viruses(30,31). We also observe an exonuclease regulator of endosomal nucleic acid sensing (PLD3)(32), a transcription factor associated with the influenza humoral response (MAZ)(33,34), and a DNA-binding protein implicated in B cell class switching (KIN or KIN17)(35). In contrast, the list of unique SARS-CoV-1 interactors includes components of the 4EFP-GYF2 translation repression complex (GYGYF2, EIF4E2), lysosomal ion channels involved in chloride/proton ion exchange (CLCN7, OSTM1), and the cytosolic histone deacetylase 6 (HDAC6). While SARS-CoV-1 interactors GIGYF2 and EIF4E2 were also identified in the recent SARS-CoV-2 nsp2 dataset(18), it is clear from our quantitative comparison that enrichment of this complex with SARS-CoV-2 nsp2 is much weaker than with SARS-CoV-1 nsp2.

2.3.4 Comparative profiling of CoV nsp4 interactions

We extended our comparative analysis of host cell interactors to another CoV non-structural protein nsp4 involved in the replication complex. We applied the same AP-MS workflow used to identify nsp2 interactors (**Figure 2-2A**). In addition to SARS-CoV-1 and SARS-CoV-2 nsp4, we also included the hCoV-OC43 nsp4 construct. With this addition, we sought to probe the protein-protein interactions that differentiate strains causing severe pathogenesis versus non-severe. To this end, four co-immunoprecipitation replicates of respective nsp4 homologs were pooled into a single MS run, along with mock GFP transfected cells to differentiate non-specific background proteins (**Figure 2-2B**). The full data set included three individual MS runs, containing 40 Co-IPs (SARS-CoV-2 n = 12; SARS-CoV-1 n = 8; OC43 = 8; GFP (mock) n = 12) (**Figure 7-7A**).

As previously described, we optimized variable cutoffs for high- and medium-confidence interactors based on their magnitude enrichment compared to GFP samples (**Figure 2-4A, Figure 7-7B-C**). We identified 29, 20, and 13 high-confidence interactors for SARS-CoV-2, SARS-CoV-1, and OC43 respectively using the most stringent cutoff (**Figure 2-4A, Figure 7-7B-C**). Including medium-confidence interactors, we identified 86, 126, and 93 nsp4 interactors for SARS-CoV-2, SARS-CoV-1, and OC43 nsp4 homologs respectively. Comparisons of high-confidence interactors yielded 17 shared interactors between all strains (**Figure 2-4B**) or 30 medium-confidence shared interactions (**Figure 7-7D**).

Similarly to our analysis of nsp2, we compared our data set with previously published nsp4 interactomics data, including the recently published study of the SARS-CoV-2 interactome(18), and find there is relatively little overlap between our identified SARS-CoV-2 nsp4 interactors and published nsp4 interactomics data (**Figure 7-8**). This discrepancy could be attributed to the nsp4 constructs in our study including the N-terminal residues of nsp3, which were added to ensure proper localization and prevent the hydrophobic N-terminal region of nsp4 to serve as signal sequence (24). For further validation, we determined interactor expression levels in human tissues and find interactors are enriched in tissues relevant to coronavirus tropism (**Figure 7-5**)(26–28).
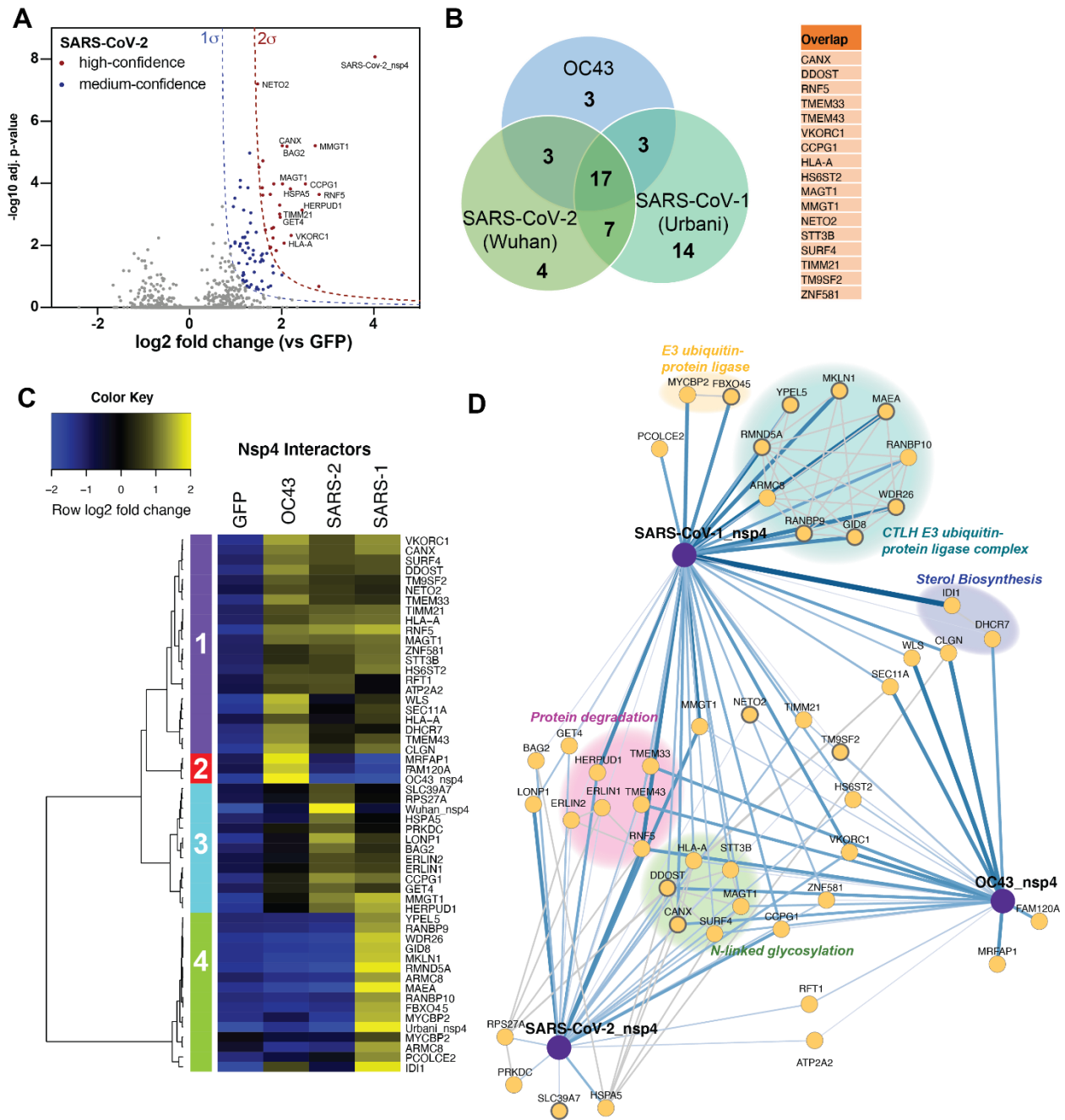
Figure 2-4. Comparative profiling of nsp4 interactions.

A. Volcano plot of the SARS-CoV-2 nsp4 datasets to identify medium- and high-confidence interactors. Plotted are log2 TMT intensity differences for proteins between nsp4 bait channels and GFP mock transfections versus -log10 adjusted p-values. Curves for the variable cutoffs used to define high-confidence (red) or medium confidence (blue) interactors are shown. $1\sigma = 0.66$. Equivalent volcano plot for SARS-CoV-1 and OC43 nsp4 are shown in **Figure 7-5**B-C.

B. Venn diagram of interactors from nsp4 homologs. Overlapping nsp4 interactors between all strains are listed in the adjacent table.

C. Heatmap comparing the enrichment of interactors for the different nsp4 homologs. log2 fold change is color-coded and centered by row (blue low, yellow high enrichment). Hierarchical clustering using Ward's method shown on the left was carried out on euclidean distances of log2 fold changes scaled by row. Clusters 1 corresponds to shared interactors of SARS-CoV-1, -2, and OC43 nsp4. Cluster 2, and 4 contain unique interactors for OC43 and SARS-CoV-1 nsp4, respectively, while cluster 3 contains shared interactors of SARS-CoV-1 and SARS-CoV-2.

D. Protein-protein interaction (PPI) network map of interactors of nsp4 homolog. Blue lines indicate measured viral-host PPIs, where line width corresponds to fold enrichment compared to the GFP control. Grey lines indicate

annotated host-host PPIs in STRING (score > 0.75). Groups of interactors with a common functional role are highlighted.

Analysis of GO-terms associated with the nsp4 interactors showed multiple enriched biological processes, such as cell organization and biogenesis, transport, and metabolic processes (**Figure 7-9A**). Interestingly, several shared SARS nsp4 interactors are associated with cell death, cellular communication, and cell differentiation. Shared interactors of all three strains are predominantly ER-membrane associated proteins, while many SARS-CoV-1 and OC43 specific interactors are annotated as nuclear-localized (**Figure 7-9B**). Comparisons of gene set enrichment analysis between strains indicate the ERAD pathway is significantly enriched for SARS strains, most strongly for SARS-CoV-1 (**Figure 7-9C**). Ubiquitin-dependent protein catabolic processes and ER mannose trimming are also strongly enriched for SARS-CoV-1. In general, processes strongly enriched for SARS-CoV-1 are less enriched for SARS-CoV-2, and to an even lesser extent for OC43.

Our multiplexed analysis of nsp4 homolog Co-IPs enabled direct comparison across strains (**Figure 7-10A-C**). We validated that nsp4 bait levels were mostly similar across replicates, allowing for direct comparison of bait protein intensities (**Figure 7-10D**). Unbiased hierarchical clustering of enrichment intensities to group nsp4 interactors yielded four distinct clusters. Clusters 1 contained common interactors of all nsp4 homologs (**Figure 2-4C**), while cluster 3 contains shared interactors of SARS-CoV-2, SARS-CoV-1 nsp4 that displayed weaker enrichment with OC43. In contrast, cluster 2 and 4 contained unique interactors enriched for OC43 and SARS-CoV-1 nsp4, respectively. To visualize functionally relevant subclusters of shared and unique nsp4 interactors, we constructed a network plot, including high-confidence interactions (score > 0.75) from the String database (**Figure 2-4D**). Inclusion of all median-confidence interactors of the nsp4 homologs yielded a similar clustering and network organization (**Figure 7-10E, Figure 7-11**).

We identified several common interactors across all three nsp4 homologs. These include components of UPR signaling (TMEM33) and ER-phagy (CCPG1). We also identify RNF5, an ER-localized E3 ubiquitin ligase known to modulate anti-viral innate immune signaling(36,37), and VKORC1, which reduces Vitamin K, a key cofactor for several coagulation factor proteins(38). Not surprisingly, given that nsp4 is a glycosylated protein, we also identify several members of the *N*-linked glycosylation machinery (STT3B, MAGT1, CANX, DDOST) (**Figure 2-4D**) in all three strains.

We identified several shared interactors between SARS-CoV-1 and SARS-CoV-2 that were absent in OC43. These include the ERLIN1/2 complex, LONP1, HERPUD1, GET4, and BAG2, all of which are involved in a facet of ER homeostasis, proteostasis or trafficking (**Figure 2-4D**). We validated the interactions of nsp4 constructs with ERLIN2 and CANX by Co-IP, and confirmed that ERLIN2 enriches significantly more strongly with SARS-CoV-1 and -2 compared to OC43, while CANX interacts with all three homologs (**Figure 7-6A**). The ERLIN1/2 complex was also identified in the nsp2 data set (**Figure 2-3B-C**) and shows comparable enrichment values between SARS-CoV-1 and SARS-CoV-2. Interestingly, the other four overlapping interactors all exhibit increased enrichment for SARS-CoV-2 versus SARS-CoV-1. LONP1 is a mitochondrial peptidase responsible for removing the majority of damaged mitochondrial proteins via proteolysis. The unfolded protein response (UPR) induces HERPUD1 expression, which is involved in the ER-associated degradation pathway (ERAD) to maintain ER homeostasis (39). BAG2 serves as a co-chaperone for HSP70 chaperones, acting as a nucleotide exchange factor to regulate chaperone-client interactions through modulating HSP70 ATPase rates (40), while GET4 is part of a complex driving trafficking of tail-anchored proteins to the ER (41).

We observed shared interactors between OC43 and SARS-CoV-2, such as the *N*-glycosylation factor RFT1 (42,43)

and a sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA – ATP2A2) (44). In addition, we identified shared interactors between OC43 and SARS-CoV-1, including a regulator of UPR-mediated apoptosis (WLS or GPR177) (45), a member of the signal peptidase complex (SEC11A) (46), and factors involved in cholesterol synthesis (IDI1, DHCR7) (47–50). WLS, SEC11A, and DHRC7 exhibited higher enrichment for OC43, whereas IDI1 was more greatly enriched for SARS-CoV-1. Consistent with this observation, we identified sterol metabolic process as one of the unique processes enriched for OC43 nsp4.

In addition to shared interactors, we found several unique interactors for SARS-CoV-2, including the monoubiquitin-ribosomal fusion protein (RPS27A), a Golgi/ER-resident zinc receptor that has been shown to regulate TNF receptor trafficking and necroptosis (SLC39A7), and the ER-resident Hsp70 chaperone BiP (HSPA5). The latter two play distinct roles in regulating ER homeostasis and proteostasis. In contrast, only two unique OC43 nsp4 interactors were identified: a target of the NEDD8-Cullin E3 ligase pathway (MRFAP1)(51) and FAM120A, an RNA-binding protein found to serve as a scaffolding protein for the IL13 signaling pathway (**Figure 7-10B-C**) (52,53). Both of these proteins are localized to the nucleus (**Figure 7-9B**). Lastly, we identified a large cluster of unique SARS-CoV-1 nsp4 interactors that compose the CTLH E3 ubiquitin ligase complex (**Figure 2-4D**). This nuclear complex maintains cell proliferation rates, likely through the ubiquitination of the transcription factor Hbp1, a negative regulator of cell proliferation(54). This complex is highly enriched for SARS-CoV-1 specifically, presenting one of the most profound differences in interaction profile (**Figure 7-8A,C**). This specificity of engagement was validated through co-IP and western blot (**Figure 7-6A-B**). We confirmed that only SARS-CoV-1 nsp4 co-purified with several of CTLH components (MKLN1, WDR26, RANBP9).

The fact that both OC43 and SARS-CoV-1 nsp4 displayed prominent interactions with nuclear proteins prompted us to evaluate the cellular localization of the protein by immunofluorescence. We detected perinuclear puncti for all constructs which partially co-localized with the ER marker PDIA4 (**Figure 7-12**), consistent with prior studies(55). However, for SARS-CoV-1 and OC43 nsp4, we also detected measurable signal in the nucleus, supporting a nuclear function and the observed interactions with proteins in the nucleus. As observed with nsp2 expression, immunofluorescence staining indicated that nsp4 expression was limited to a low number of cells within transfected plates, implying that identified interactions originate from a subset of cells.

2.3.5   Enrichment of mitochondria-associated membrane proteins as nsp2 and nsp4 interactors

In our evaluation of cellular compartment GO-terms, we noticed that nsp2 and nsp4 interactors are enriched in membranes of the endoplasmic reticulum and the mitochondria (**Figure 7-3B, Figure 7-9B**). In particular, ERLIN1/2 and RNF170 form an E3 ubiquitin ligase complex known to localize to the interface between the ER and mitochondria, regions termed mitochondria-associated membranes (MAMs). We therefore probed our data set for any other MAMs-associated nsp2 and nsp4 interactors. We cross-referenced our interactor lists with three published data sets that specifically characterized the MAMs proteome(56–58) and identified 17 proteins associated with MAMs (**Figure 2-5A-B**). Seven of these factors solely interact with nsp2, eight proteins solely interact with one or more strains of nsp4, and the ERLIN1/2 complex interacts with both nsp2 and nsp4 (**Figure 2-3C, Figure 2-4D**). Interestingly, the ERLIN1/2 complex only interacts with SARS-CoV-1 and -2 proteins and not OC43. SARS-CoVs may use ERLIN1/2 to regulate ER Ca^{2+} signaling and the

myriad of downstream host processes controlled by this signaling pathway (**Figure 2-5D**).

To determine if nsp2 or nsp4 indeed co-localize to MAMs, we performed subcellular fractionation and probed for the presence of viral proteins in MAMs fractions, as well as various fraction markers (CALX and ERLIN2 for MAMs, MCU for mitochondria) by western blotting (**Figure 2-5C, Figure 7-13**). We find that both SARS-CoV-2 nsp2 and nsp4 are detected in MAMs, though nsp4 is more enriched at MAMs than nsp2 (**Figure 2-5C**). SARS-CoV-1 nsp2 is absent from the MAMs fractions, while both SARS-CoV-1 and hCoV-OC43 nsp4 are strongly enriched in MAMs (**Figure 7-13**).
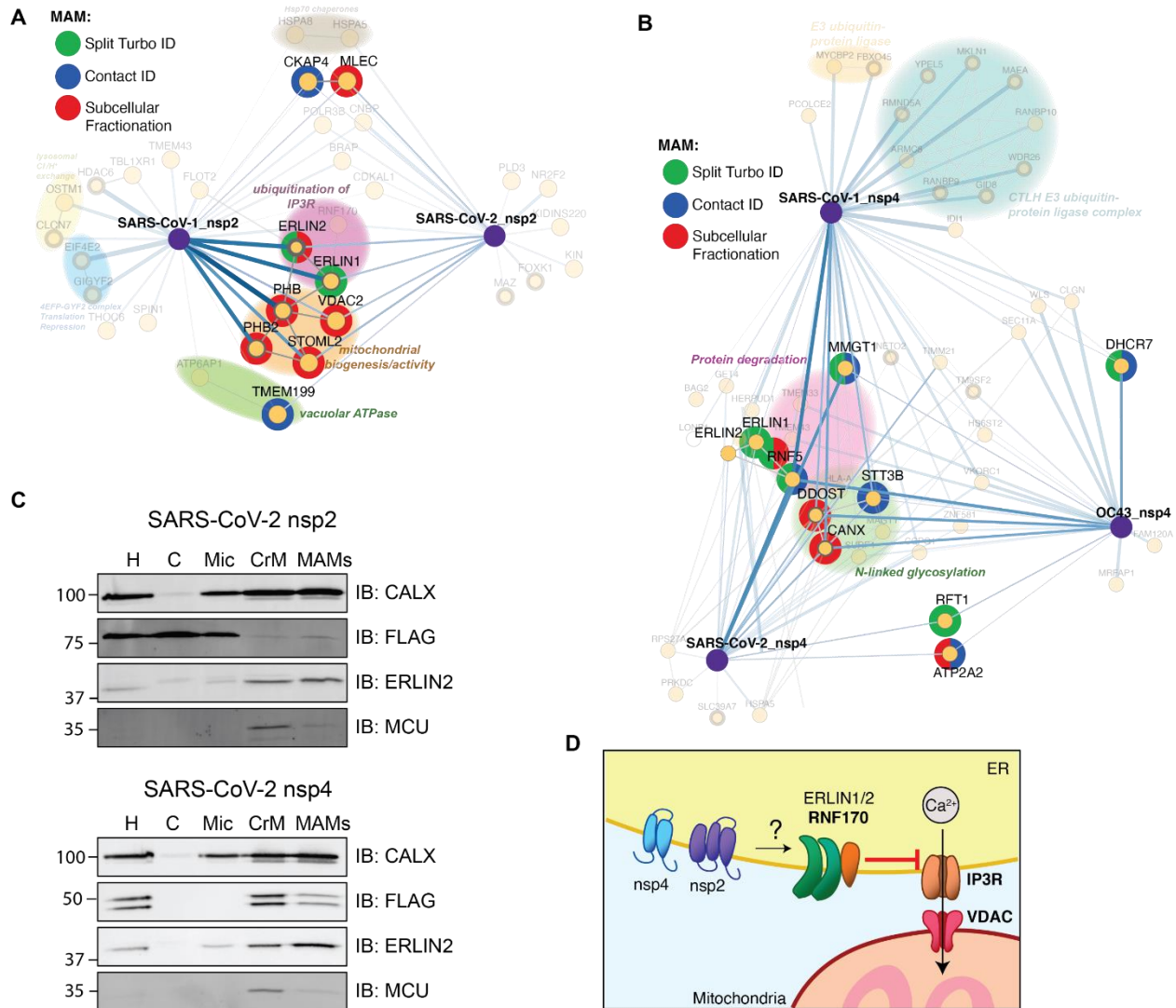
Figure 2-5. Enrichment of mitochondria-associated membrane (MAM) proteins as nsp2 and nsp4 interactors.

A. Interactors of nsp2 homologs annotated for MAM proteins. The lists of interactors were cross-referenced with previous publications profiling the MAM proteome (Split-Turbo ID(56), Contact-ID(57), and subcellular fractionation(58)).

B. Same as in (A), for nsp4 homologs.

C. Subcellular fractions of SARS-CoV-2 nsp2 or nsp4 transfected HEK293T cells to determine localization of viral proteins to MAMs. Homogenate (H), cytosol (C), microsome (Mic), crude mitochondria (CrM), and MAMs fractions were probed via Western blot for subcellular markers (CALX and ERLIN2 for MAMs; MCU for mitochondria) and viral proteins (FLAG). Subcellular fractionation was carried out in triplicate and representative blots shown.

D. Proposed model for how SARS-CoV nsp2 and nsp4 utilize ERLIN1/2 and interacting protein factors to regulate ER Ca^{2+} signaling at MAMs.

2.4 Discussion

Our analysis enables both the identification of interactors for SARS-CoV-1, SARS-CoV-2, and OC43 homologs of nsp2 and nsp4, and comparative quantitative enrichment to differentiate between shared and unique host cell binding partners. We acknowledge the limitations of using transiently transfected viral proteins for AP-MS. Viral infection is a collection of both protein-protein and RNA-protein interactions and our approach of single protein expression may omit direct interactions that would result from the full context of virus replication. In addition, transient and low affinity interactions may not be identified using our current approach. Additional incorporation of chemical crosslinkers may be necessary to capture such transient interactors. However, given the logistical barriers to handling BSL-3 viruses, paired with the urgency of the current pandemic, our workflow is an efficient system to perform comparative analysis and generate a shortlist of interactors to prioritize for further investigation. Furthermore, some direct interactions of CoV proteins may be short in duration and missed by interaction mapping in full virus infection. Thus, relevant but short-lived interactions may be better identified through individual protein expression.

We identify several nsp2 interactors shared across SARS strains, including STOML2, and prohibitins (PHB and PHB2), which were previously identified as interacting with SARS-CoV-1(17). These proteins work in tandem to induce formation of metabolically active mitochondrial membranes to regulate mitochondrial biogenesis. Increased levels of STOML2 are associated with increased ATP production and reduced apoptosis induction(29). This conserved interaction for SARS strains presents an avenue for nsp2 to increase mitochondrial metabolism and stall apoptosis to maintain a pro-viral cellular environment. Additionally, STOML2 has been found to play a key role in stabilizing hepatitis C virus replication complexes(59) and PHB has been shown to promote entry of both Chikungunya virus(60) and enterovirus 71(61). These factors may prove effective pan-RNA virus targets for host-directed therapies. We also attempted to extend the comparative analysis to OC43 nsp2. However, this construct did not express detectable protein, which could be due to much lower homology to SARS-CoV-2 and SARS-CoV-1 than for other non-structural proteins.

For nsp4, we identify multiple unique SARS-CoV-1 interactors, most notably members of the CTLH E3 ubiquitin ligase complex. This complex is known to regulate levels of Hbp1, a negative regulator of proliferative genes(54) and was previously shown to interact with the dengue viral protein NS2B3(62), implicating this complex as a target for RNA viruses to influence cell proliferation. We also identified the FBXO45-MYCBP2 E3 ubiquitin ligase complex, which has been shown to prevent cell death in mitosis (63). Together, this may support a role in SARS-CoV-1 nsp4 co-opting host ubiquitin complexes to extend cell viability during infection to promote viral replication. During resubmission, Gordon et al. published a comparative coronavirus interaction network confirming the SARS-CoV-1 specific interactions with the CTLH E3 ligase complex(64).

Furthermore, we find components of the cholesterol biosynthesis pathway, IDI1 and DHCR7, which were specifically enriched for SARS-CoV-1 and OC43 nsp4 respectively. IDI1 has been shown to be downregulated by host cells in response to CMV-infection-induced interferons(47) and is upregulated by both HIV and HCV during infection(48,49). DHCR7 is downregulated during RNA virus infection in macrophages to promote IRF3 signaling and IFN-1 production. Moreover, inhibition of DHCR7 aids in clearance of multiple RNA viruses(50). These previous findings indicate that interactions with

IDI1 and DHCR7 may provide means for coronaviruses to counteract anti-viral responses. Interestingly, these interactions with the aforementioned E3 ligase complexes and cholesterol biogenesis factors are not enriched for SARS-CoV-2 nsp4, implying that SARS-CoV-2 pathogenesis may not require these interactions.

As a whole, it appears that SARS-CoV-2 homologs differ from SARS-CoV-1 not by gaining new interactions, but rather by losing network nodes. This is emphasized in the gene enrichment analysis of nsp2 and nsp4 (**Figure 7-3C, Figure 7-9C**), in which multiple pathways are more strongly enriched for SARS-CoV-1, as well as in the nsp4 interactome (**Figure 2-4C-D**), particularly with the absence of E3 ligase complex interactions for SARS-CoV-2. It will be important to investigate potential functional implications of the engagement of the E3 ubiquitin ligase, as well innate immune signaling factors on CoV infections and the course of pathogenicity for the divergent strains.

To gain functional insights into which nsp2 and nsp4 interactions may have an impact on CoV infection, we mined recently published data from a genome wide CRISPR knockout screen and a targeted siRNA knockdown/CRISPR knockout screen of SARS-CoV-2 interactors (**Figure 7-14A-B**)(64,65). The comprehensive genome wide dataset by Heaton et al. conducted in A549 lung cancer cells identified that knockout of several of the proteostasis components (BAG2, DDOST), as well as ERLIN1 and ERLIN2 enhanced cell survival in the presence of SARS-CoV-2, suggesting that these factors may have an anti-viral function (**Figure 7-14A**). When comparing the more limited siRNA knockdown data in A549 cells and CRISPR knockout data in Caco2 colorectal cancer cells, ATP6AP1 stood out as hampering SARS-CoV-2 infection in both cell models supporting a pro-viral role (**Figure 7-14B**)(64). ATP6AP1/Ac45 is a critical accessory subunit to facilitate the assembly of the vacuolar ATPase in support of lysosome function and autophagy playing a role in viral infection(66). Future functional genomic screens will be necessary to evaluate the role of other interactors on CoV infection and evaluate differential roles for the distinct strains.

A particularly noteworthy finding is the identification of 17 MAMs factors in the combined nsp2 and nsp4 datasets, based on cross-referencing interactors with previously published proteomics studies of MAMs proteins(56–58). Given the prominence of these interactions, it is tempting to speculate that nsp2 and nsp4 localize to MAMs and influence processes at these important organelle contact sites (**Figure 2-5D**). MAMs are nodes for innate immune signaling and apoptosis pathways, both of which are common targets for viral manipulation.

In particular, we identify the E3 ubiquitin ligase RNF5 interacting with all nsp4 homologs. RNF5 targets STING for degradation, which stabilizes retinoic acid-inducible gene-I (RIG-I) and mitochondrial antiviral-signaling protein (MAVS) interactions at MAMs, thereby inducing inteferon-1 and -3 production via IRF3 and NF-κB signaling(36,37). RIG-I is one of the main viral RNA genome sensors in host cells; therefore, it is possible that nsp4 increases targeting of RNF5 to MAMs to inhibit downstream signaling of RIG-1.

We also identify the ERLIN1/2 complex in both nsp2 and nsp4 data sets. In the nsp2 interaction network, the complex is associated with a different E3 ligase, RNF170. RNF170 has been shown to inhibit innate immune signaling by targeting TLR3 for degradation, thereby blocking IRF3 and NF-κB signaling pathways(67). In addition, ERLIN1/2 acts in concert with RNF170 to target the inositol-1,4,5-triphophate receptor (IP$_3$R) for degradation via polyubiquination(68). IP$_3$R is an ER-resident Ca^{2+} channel integral in the formation of MAMs(69,70). Calcium flux at MAMs has been shown to increase mitochondrial calcium uptake, which increases ATP production, thereby benefitting active viral replication(71). Indeed, several other viruses have been shown to influence ER Ca^{2+} exchange. For instance, the Hepatitis C viral protein, NS5A,

promotes degradation of IP$_3$R3 to limit apoptosis induction triggered by persistent Ca^{2+} signaling at MAMs(72). Previous studies have shown the SARS-CoV-1 E protein acts as a channel to leak ER calcium stores during infection(73), but to our knowledge, no such features have been attributed to either nsp2 or nsp4. Thus, manipulation of ER Ca^{2+} signaling via IP$_3$R regulation may represent a novel method by which coronaviruses manipulate mitochondrial function (**Figure 2-5D**). In support of this, we find that both SARS-CoV-2 nsp2 and nsp4 colocalize to MAMs using subcellular fractionation (**Figure 2-5C**). Additionally, a recent study found that IP$_3$R3 is significantly upregulated during SARS-CoV-2 infection(74) (**Figure 7-14C**). Further studies will be important to evaluate whether ER calcium exchange and mitochondrial metabolism could impact coronavirus infection.

2.5 Acknowledgments

2.6 Methods

Protein expression constructs

Coding sequences for nsp2 and nsp4 were obtained from GenBank (MN908947 SARS-CoV-2 isolate Wuhan-Hu-1; AY278741 SARS-CoV-1 Urbani; NC_006213 hCoV OC43 strain ATCC VR-759). Human codon optimized sequences were designed, genes synthesized, and cloned into pcDNA3.1-(+)-C-DYK (nsp4) to append a C-terminal FLAG tag, or into pcDNA3.1-(+)-N-DYK (nsp2) to append an N-terminal FLAG tag (GenScript).

Cell culture and transfection

HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with high glucose and supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 1% glutamine. Cells were kept at 37°C, 5% CO2. Generally, 2 x 10^6 cells were seeded into 10cm dishes. 24 hours post-seeding, cells were transfected with 5μg nsp2, nsp4, or fluorescent control DNA constructs in pcDNA3.1-(+)-C/N-DYK vectors using a calcium phosphate method. Media was exchanged 16 hours post-transfection, and cells were harvested 24 hours after changing media.

Immunoprecipitation

Cells were collected and washed with PBS. Immunoprecipitation samples were lysed by resuspension in TNI buffer (50 mM Tris pH 7.5, 150mM NaCl, 0.5% IGEPAL-CA-630) with Roche cOmplete protease inhibitor on ice for at least 10 minutes, followed by sonication in a room temperature water bath for 10 minutes. Lysates were cleared by centrifugation at 17,000 x g for 10-20 minutes. Sepharose 4B resin (Sigma) and G1 anti-DYKDDDDK resin (GenScript) were pre-washed 4x with the respective lysis buffer for each sample. Protein concentrations in cleared lysates were normalized using BioRad Protein Assay Dye and added to 15 μL Sepharose 4B resin for 1 hour, rocking at 4°C. Resin was collected by centrifugation

for 5-10 minutes at 400 x g and pre-cleared supernatant was added directly to 15 μL G1 anti-DYKDDDDK resin and rocked at 4°C overnight. The next day, supernatant was removed and resin was washed 4 times with the respective lysis buffer. Bound proteins were eluted with the addition of modified 3x Laemelli buffer (62.5 mM Tris, 6% SDS) for 30 minutes at room temperature followed by 15 minutes at 37°C, followed by a second elution for 5-15 minutes at 37°C. 10% of elution was set aside for SDS-PAGE and silver staining to confirm immunoprecipitation efficiency, and the remainder was prepared for mass spectrometry. Silver staining was performed using a Pierce Silver Stain kit (Thermo Scientific). Separate biological replicates of co-immunoprecipitated lysates were identically processed, in which inputs and elutions were normalized and run on SDS-PAGE gel, transferred to PVDF membrane, and blotted for various host interactors using the following antibodies (1:1000 dilutions): anti-FLAG (Sigma-Aldrich, F1804), anti-calnexin (GeneTex, GTX109669), anti-muskelin (Santa Cruz Biotechnology, SC-398956) anti-ERLIN2 (Sigma-Aldrich, HPA002025), and anti-GAPDH (GeneTex, GTX627408) as a loading control.

Tandem Mass Tag sample preparation

Sample preparation was carried out as described(75). Briefly, eluted proteins were precipitated in methanol/chloroform/water (3:1:3), washed twice in methanol, and protein pellets were air dried. Pellets were resuspended in 1% Rapigest SF (Waters), reduced, and alkylated. Proteins were digested in trypsin-LysC overnight. Digested peptides were labeled using 11-plex or 16-plex tandem mass tag (TMT or TMTpro) reagents (Thermo Scientific), pooled, and acidified using formic acid. Cleaved Rapigest was removed by centrifugation of samples at 17,000 x g for 30 min.

MudPIT liquid chromatography-mass spectrometry analysis

Triphasic MudPIT microcolumn were prepared as described(76). Individual pooled TMT proteomics samples were directly loaded onto the microcolumns using a high-pressure chamber followed by a wash with 5% acetonitrile, 0.1% formic acid in water (v/v) for 30 min. Peptides were analyzed by liquid chromatography-mass spectrometry on an Exploris 480 in line with an Ultimate 3000 nanoLC system (Thermo Fisher). The MudPIT microcolumns were installed on a column switching valve on the nanoLC systems followed by 20 cm fused silica microcapillary column (ID 100μm) ending in a laser-pulled tip filled with Aqua C18, 3μm, 100 Å resin (Phenomenex). MudPIT runs were carried out by 10μL sequential injection of 0, 10, 20, 40, 60, 80, 100 % buffer C (500mM ammonium acetate, 94.9% water, 5% acetonitrile, 0.1% formic acid), followed by a final injection of 90% C, 10% buffer B (99.9% acetonitrile, 0.1% formic acid v/v). Each injection was followed by a 130 min gradient using a flow rate of 500nL/min (0 – 6 min: 2% buffer B, 8 min: 5% B, 100 min: 35% B, 105min: 65% B, 106 – 113 min: 85% B, 113 – 130 min: 2% B). Electrospray ionization was performed directly from the tip of the microcapillary column using a spray voltage of 2.2 kV, ion transfer tube temperature of 275ºC and RF Lens of 40%. MS1 spectra were collected using the following settings: scan range 400 – 1600 m/z, 120,000 resolution, AGC target 300%, and automatic injection times. Data-dependent tandem mass spectra were obtained using the following settings: monoisotopic peak selection mode: peptide, included charge state 2 – 7, TopSpeed method (3s cycle time), isolation window 0.4 m/z, HCD fragmentation using a normalized collision energy of 32, resolution 45,000, AGC target 200%, automatic injection times, and dynamic exclusion (20 ppm window) set to 60 s.

Experimental layout and data analysis

The nsp2 AP-MS experiments included three individual MS runs combining 34 Co-AP samples (SARS-CoV-2 n = 13; SARS-CoV-1 n = 9; GFP (mock) n = 12). Samples were distributed to TMTpro 16plex or TMT11plex channels as outlined in Figure S2A. The nsp4 AP-MS experiments consisted of three individual MS runs, containing 40 Co-IPs (SARS-CoV-2 n = 12; SARS-CoV-1 n = 8; OC43 = 8; GFP (mock) n = 12). Samples were distributed to TMTpro 16plex channels as outlined in Figure S5A. Identification and quantification of peptides and proteins were carried out in Proteome Discoverer 2.4 (Thermo Fisher) using a SwissProt human database (Tax ID 9606, release date 11/23/2019). CoV nsp2 and nsp4 protein sequences were added manually. Searches were conducted in Sequest HT using the following setting: Trypsin cleavage with max. 2 missed cleavage sites, minimum peptide length 6, precursor mass tolerance 20 ppm, fragment mass tolerance 0.02 Da, dynamic modifications: Met oxidation (+15.995 Da), Protein N-terminal Met loss (-131.040 Da), Protein N-terminal acetylation (+42.011 Da), static modifications: Cys carbamidomethylation (+57.021 Da), TMTpro or TMT6plex at Lys and N-termini (+304.207 Da for TMTpro or +229.163 for TMT6plex). Peptide IDs were filtered using the Percolator node using an FDR target of 0.01. Proteins were filtered based on a 0.01 FDR requiring two peptide IDs per protein, and protein groups were created according to a strict parsimony principle. TMT reporter ions were quantified using the reporter ion quantification considering unique and razor peptides, and excluding peptides with co-isolation interference greater than 25%. Peptide abundances were normalized based on total peptide amounts in each channel assuming similar levels of background signal in the APs. Protein quantification roll-up used all quantified peptides. Pairwise ratios between conditions were calculated based on total protein abundances and ANOVA on individual proteins was used to test for changes in abundances and to report adjusted p-values.

To filter high-confidence interactors of individual CoV nsp proteins, we used a variable filter combining log2 fold enrichment and adjusted p-value according to a published method(77). Briefly, the histogram of log2 protein abundance fold changes between nsp-transfected versus mock-transfected groups were fitted to a gaussian curve using a nonlinear least square fit to determine the standard deviation σ (see **Figure 7-2B-C**). Fold change cutoffs for high-confidence and medium-confidence interactors were based on 2 σ, or 1 σ, respectively. For actual cutoffs taking into consideration adjusted p-values, we utilized a hyperbolic curve $y > c / (x - x_0)$, where y is the adj. p-value, x the log2 fold change, x_0 corresponds to the standard deviation cutoff (2 σ or 1 σ), and c is the curvature (c = 0.4 for 1 σ, and 0.8 for 2 σ). (see **Figure 2-2C-D, Figure 2-4A, Figure 7-7B-C**).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE(78) partner repository with the dataset identifier PXD022017.

Gene set enrichment analysis

GO-term categories for biological processes and cellular components for interactors were based on assignment in the Proteome Discoverer Protein Annotation node. Gene set enrichment analysis was conducted in EnrichR(79). The analysis was conducted separately for sets of interactors of individual nsp2 or nsp4 homologs and GO-terms for biological processes were filtered by adjusted p-values < 0.1. Redundant GO-terms were grouped manually based on overlapping genes in related terms.

Network plots and identification of overlapping interactions with published data

Extended and overlapping interactomes between novel interactors identified in this study and previously published interactors(18) were generated by scraping the top n interactors of each primary prey protein on the STRING database using the python API. We established an extended secondary interactome by searching for the top 20 and top 30 STRING db interactors of the nsp4 primary interactors and nsp2 interactors respectively using limit parameter in STRING API and searching against the human proteome (species 9606). We then compared the extended interactomes of our data with the previously published data by dropping any secondary interactors that did not appear in both data sets. Next, we concatenated the primary interactors from our data, the primary interactors from the published data, and the overlapping secondary interactors into a single data frame. Finally, we searched the overlapping secondary interactors against the STRING database human proteome to determine interactors between secondary interactors with a threshold of greater than 50% likelihood in the experimental score category. The results were plotted in Cytoscape.

Immunofluorescence confocal microscopy

HEK293T cells were cultured on glass-bottom culture dishes (MatTek, P35G-0-14-C) and transfected with CoV expression constructs as previously described. Cells were fixed with 4% paraformaldehyde-PBS, washed thrice with PBS, then permeabilized in 0.2% Triton-X (in PBS). After three PBS washes, cells were blocked in PBS with 1% BSA with 0.1% Saponin (blocking buffer). After blocking, cells were incubated with anti-PDIA4 primary antibody (Protein Tech, 14712-1-AP) in blocking buffer (1:1000 dilution) for 1 hour at 37°C. After three PBS washes, cells were incubated with AlexFluor 488-conjugated anti-rabbit goat antibody (ThermoFisher, A-11008) in blocking buffer (1:500 dilution) at room temperature for 30 min. Cells were then stained with M2 FLAG primary antibody (SigmaAldrich, F1804) and AlexFluor 594-conjugated anti-mouse goat antibody (ThermoFisher, A-11005) using the same conditions. Cells were then mounted in Prolong Gold with DAPI stain (ThermoFisher, P36935). Cells were imaged using an LSM-880 confocal microscope (Zeiss) and images were merged using Image J software.

Subcellular fractionation

HEK293T cells were transfected with tdTomato, nsp2, or nsp4 constructs as previously described and harvested via scraping (1.5-2.0 x 10^8 cells per sample). Cells were fractionated based on a previously published protocol(80). In brief, cells were mechanically lysed in 6 mL of sucrose homogenization medium (250 mM sucrose, 10 mM HEPES, pH 7.4) using a douncer homogenizer (Kimbell, DD9063). Homogenate was spun four times (5 min, 600xg, 4°C) to pellet cell debris and intact cells. The supernatant was then centrifuged for 10 min at 10,300xg, 4°C in a fixed angle rotor to pellet crude mitochondria. The resulting supernatant was centrifuged at 10,300xg for 10 min at 4°C twice more to pellet residual crude mitochondria and then ultracentrifuged at 100,000xg for 60 min at 4°C to separate the microsome fraction (pellet) from the cytosolic fraction (supernatant). The crude mitochondrial pellet was resuspended in 0.5 mL of mannitol buffer A (250 mM mannitol, 0.5 mM EGTA, 5 mM HEPES, pH 7.4), layered on top of 8 mL of 30% (w/v) Percoll solution (Sigma, P1644), then ultracentrifuged at 95,000xg for 65 min at 4°C to separate MAMs from crude mitochondria. MAMs were pelleted by centrifugation at 100,000xg for 1 h at 4°C. Fractions were run on SDS-PAGE gel and proteins detected via Western blotting with the following antibodies: anti-calnexin (GeneTex, GTX109669), anti-MCU (ThermoFisher, MA5-24702), anti-FLAG

(Sigma-Aldrich, F1804), and anti-ERLIN2 (Sigma-Aldrich, HPA002025).

2.7   References

1.   Fehr, A. R. & Perlman, S. Coronaviruses: An overview of their replication and pathogenesis. in *Coronaviruses: Methods and Protocols* vol. 1282 1–23 (2015).

2.   *Novel Coronavirus (2019-nCoV) Situation Report 1*. (2020).

3.   Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

4.   Perlman, S. & Netland, J. Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews Microbiology* vol. 7 439–450 at https://doi.org/10.1038/nrmicro2147 (2009).

5.   Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e9 (2020).

6.   Fung, T. S. & Liu, D. X. Coronavirus infection, ER stress, apoptosis and innate immunity. *Frontiers in Microbiology* vol. 5 1–13 at https://doi.org/10.3389/fmicb.2014.00296 (2014).

7.   Tan, Y.-X. *et al.* Induction of Apoptosis by the Severe Acute Respiratory Syndrome Coronavirus 7a Protein Is Dependent on Its Interaction with the Bcl-XL Protein. *J. Virol.* **81**, 6346–6355 (2007).

8.   Yeung, M. L. *et al.* MERS coronavirus induces apoptosis in kidney and lung by upregulating Smad7 and FGF2. *Nat. Microbiol.* **1**, 1–8 (2016).

9.   Yue, Y. *et al.* SARS-Coronavirus Open Reading Frame-3a drives multimodal necrotic cell death. *Cell Death Dis.* **9**, 1–15 (2018).

10.   von Brunn, A. *et al.* Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. *PLoS One* **2**, e459 (2007).

11.   Pfefferle, S. *et al.* The SARS-Coronavirus-host interactome: Identification of cyclophilins as target for pan-Coronavirus inhibitors. *PLoS Pathog.* **7**, e1002331 (2011).

12.   V'kovski, P. *et al.* Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8**, (2019).

13.   Jean Beltran, P. M., Cook, K. C. & Cristea, I. M. Exploring and Exploiting Proteome Organization during Viral Infection. *J. Virol.* **91**, (2017).

14.   Hashimoto, Y., Sheng, X., Murray-Nerger, L. A. & Cristea, I. M. Temporal dynamics of protein complex formation and dissociation during human cytomegalovirus infection. *Nat. Commun.* **11**, (2020).

15.   Shah, P. S. *et al.* Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **175**, 1931-1945.e18 (2018).

16.   Nicod, C., Banaei-Esfahani, A. & Collins, B. C. Elucidation of host–pathogen protein–protein interactions to uncover mechanisms of host cell rewiring. *Curr. Opin. Microbiol.* **39**, 7–15 (2017).

17.   Cornillez-Ty, C. T., Liao, L., Yates, J. R., Kuhn, P. & Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis

and Intracellular Signaling. *J. Virol.* **83**, 10314–10318 (2009).

18. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

19. Brito, A. F. & Pinney, J. W. Protein-Protein Interactions in Virus-Host Systems. *Front. Microbiol.* **8**, 1–11 (2017).

20. Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S. & Denison, M. R. The nsp2 Replicase Proteins of Murine Hepatitis Virus and Severe Acute Respiratory Syndrome Coronavirus Are Dispensable for Viral Replication. *J. Virol.* **79**, 13399–13411 (2005).

21. Angeletti, S. *et al.* COVID-2019: the role of the nsp2 and nsp3 in its pathogenesis. *J. Med. Virol.* 0–3 (2020) doi:10.1002/jmv.25719.

22. Oostra, M. *et al.* Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication. *J. Virol.* **81**, 12323–12336 (2007).

23. Gadlage, M. J. *et al.* Murine Hepatitis Virus Nonstructural Protein 4 Regulates Virus-Induced Membrane Modifications and Replication Complex Function. *J. Virol.* **84**, 280–290 (2010).

24. Oudshoorn, D. *et al.* Expression and cleavage of middle east respiratory syndrome coronavirus nsp3-4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication. *MBio* **8**, e01658-17 (2017).

25. Kanjanahaluethai, A., Chen, Z., Jukneliene, D. & Baker, S. C. Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology* **361**, 391–401 (2007).

26. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387-402.e16 (2020).

27. Jiang, L. *et al.* A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269-283.e19 (2020).

28. Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).

29. Christie, D. A. *et al.* Stomatin-Like Protein 2 Binds Cardiolipin and Regulates Mitochondrial Biogenesis and Function. *Mol. Cell. Biol.* **31**, 3845–3856 (2011).

30. Panda, D. *et al.* The transcription factor FoxK participates with Nup98 to regulate antiviral gene expression. *MBio* **6**, e02509-14 (2015).

31. Yenamandra, S. P. *et al.* Expression profile of nuclear receptors upon Epstein -- Barr virus induced B cell transformation. *Exp. Oncol.* **31**, 92–6 (2009).

32. Gavin, A. L. *et al.* PLD3 and PLD4 are single-stranded acid exonucleases that regulate endosomal nucleic-acid sensing. *Nat. Immunol.* **19**, 942–953 (2018).

33. Parks, C. L. & Shenk, T. Activation of the adenovirus major late promoter by transcription factors MAZ and Sp1. *J. Virol.* **71**, 9600–9607 (1997).

34. Ovsyannikova, I. G. *et al.* Gene signatures associated with adaptive humoral immunity following seasonal influenza A/H1N1 vaccination. *Genes Immun.* **17**, 371–379 (2016).

35. Le, M. X. *et al.* Kin17 facilitates multiple double-strand break repair pathways that govern B cell class switching. *Sci. Rep.* **6**, 1–13 (2016).

36. Zhong, B. *et al.* The Ubiquitin Ligase RNF5 Regulates Antiviral Responses by Mediating Degradation of the

Adaptor Protein MITA. *Immunity* **30**, 397–407 (2009).

37. Fenech, E. J. *et al.* Interaction mapping of endoplasmic reticulum ubiquitin ligases identifies modulators of innate immune signalling. *Elife* **9**, 1–29 (2020).

38. Owen, R. P., Gong, L., Sagreiya, H., Klein, T. E. & Altman, R. B. VKORC1 pharmacogenomics summary. *Pharmacogenetics and Genomics* vol. 20 642–644 at https://doi.org/10.1097/FPC.0b013e32833433b6 (2010).

39. Ho, D. V. & Chan, J. Y. Induction of Herpud1 expression by ER stress is regulated by Nrf1. *FEBS Lett.* **589**, 615–620 (2015).

40. Qin, L., Guo, J., Zheng, Q. & Zhang, H. BAG2 structure, function and involvement in disease. *Cell. Mol. Biol. Lett.* **21**, 1–11 (2016).

41. Chartron, J. W., Suloway, C. J. M., Zaslaver, M. & Clemons, W. M. Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12127–12132 (2010).

42. Gottier, P. *et al.* RFT1 protein affects glycosylphosphatidylinositol (GPI) anchor glycosylation. *J. Biol. Chem.* **292**, 1103–1111 (2017).

43. Haeuptle, M. A. *et al.* Human RFT1 Deficiency Leads to a Disorder of N-Linked Glycosylation. *Am. J. Hum. Genet.* **82**, 600–606 (2008).

44. Papp, B. *et al.* Endoplasmic Reticulum Calcium Pumps and Cancer Cell Differentiation. *Biomolecules* **2**, 165–186 (2012).

45. Seo, J. *et al.* GPR177 promotes gastric cancer proliferation by suppressing endoplasmic reticulum stress-induced cell death. *J. Cell. Biochem.* **120**, 2532–2539 (2019).

46. Cui, J. *et al.* Competitive inhibition of the endoplasmic reticulum signal peptidase by non-cleavable mutant preprotein cargos. *J. Biol. Chem.* **290**, 28131–28140 (2015).

47. Blanc, M. *et al.* Host defense against viral infection involves interferon mediated down-regulation of sterol biosynthesis. *PLoS Biol.* **9**, e1000598 (2011).

48. van 't Wout, A. B. *et al.* Nef Induces Multiple Genes Involved in Cholesterol Synthesis and Uptake in Human Immunodeficiency Virus Type 1-Infected T Cells. *J. Virol.* **79**, 10053–10058 (2005).

49. Diamond, D. L. *et al.* Temporal Proteome and Lipidome Profiles Reveal Hepatitis C Virus-Associated Reprogramming of Hepatocellular Metabolism and Bioenergetics. *PLoS Pathog.* **6**, e1000719 (2010).

50. Xiao, J. *et al.* Targeting 7-Dehydrocholesterol Reductase Integrates Cholesterol Metabolism and IRF3 Activation to Eliminate Infection. *Immunity* **52**, 109-122.e6 (2020).

51. Larance, M. *et al.* Characterization of MRFAP1 turnover and interactions downstream of the NEDD8 pathway. *Mol. Cell. Proteomics* **11**, 1–11 (2012).

52. Kelly, T. J., Suzuki, H. I., Zamudio, J. R., Suzuki, M. & Sharp, P. A. Sequestration of microRNA-mediated target repression by the Ago2-associated RNA-binding protein FAM120A. *RNA* **25**, 1291–1297 (2019).

53. Bartolomé, R. A. *et al.* IL13 receptor α2 signaling requires a scaffold protein, FAM120A, to activate the FAK and PI3K pathways in colon cancer metastasis. *Cancer Res.* **75**, 2434–2444 (2015).

54. Lampert, F. *et al.* The multi-subunit GID/CTLH E3 ubiquitin ligase promotes cell proliferation and targets the transcription factor Hbp1 for degradation. *Elife* **7**, e35528 (2018).

55. Oostra, M. *et al.* Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication. *J. Virol.* **81**, 12323–12336 (2007).

56. Cho, K. F. *et al.* Split-TurboID enables contact-dependent proximity labeling in cells. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12143–12154 (2020).

57. Kwak, C. *et al.* Contact-ID, a tool for profiling organelle contact sites, reveals regulatory proteins of mitochondrial-associated membrane formation. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12109–12120 (2020).

58. Carreras-Sureda, A. *et al.* Non-canonical function of IRE1α determines mitochondria-associated endoplasmic reticulum composition to control calcium transfer and bioenergetics. *Nat. Cell Biol.* **21**, 755–767 (2019).

59. Kim, J. H., Rhee, J. K., Ahn, D. G., Kim, K. P. & Oh, J. W. Interaction of Stomatin with Hepatitis C virus RNA Polymerase stabilizes the viral RNA Replicase complexes on detergent-resistant membranes. *J. Microbiol. Biotechnol.* **24**, 1744–1754 (2014).

60. Wintachai, P. *et al.* Identification of prohibitin as a Chikungunya virus receptor protein. *J. Med. Virol.* **84**, 1757–1770 (2012).

61. Too, I. H. K., Bonne, I., Tan, E. L., Chu, J. J. H. & Alonso, S. Prohibitin plays a critical role in Enterovirus 71 neuropathogenesis. *PLoS Pathog.* **14**, (2018).

62. Shah, P. S. *et al.* Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **175**, 1931-1945.e18 (2018).

63. Richter, K. T., Kschonsak, Y. T., Vodicska, B. & Hoffmann, I. FBXO45-MYCBP2 regulates mitotic cell fate by targeting FBXW7 for degradation. *Cell Death Differ.* **27**, 758–772 (2020).

64. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science (80-.).* **22**, eabe9403 (2020).

65. Heaton, B. E. *et al.* SRSF protein kinases 1 and 2 are essential host factors for human coronaviruses including SARS-CoV-2. *bioRxiv* 2020.08.14.251207 (2020) doi:10.1101/2020.08.14.251207.

66. Abbas, Y. M., Wu, D., Bueler, S. A., Robinson, C. V. & Rubinstein, J. L. Structure of V-ATPase from the mammalian brain. *Science (80-.).* **367**, 1240–1246 (2020).

67. Song, X. *et al.* E3 ubiquitin ligase RNF170 inhibits innate immune responses by targeting and degrading TLR3 in murine cells. *Cell. Mol. Immunol.* (2019) doi:10.1038/s41423-019-0236-y.

68. Lu, J. P., Wang, Y., Sliter, D. A., Pearce, M. M. P. & Wojcikiewicz, R. J. H. RNF170 protein, an endoplasmic reticulum membrane ubiquitin ligase, mediates inositol 1,4,5-trisphosphate receptor ubiquitination and degradation. *J. Biol. Chem.* **286**, 24426–24433 (2011).

69. Berridge, M. J. *Inositol trisphosphate and calcium signalling*. (1993).

70. Bartok, A. *et al.* IP3 receptor isoforms differently regulate ER-mitochondrial contacts and local calcium transfer. *Nat. Commun.* **10**, 1–14 (2019).

71. Missiroli, S. *et al.* Mitochondria-associated membranes (MAMs) and inflammation. (2018) doi:10.1038/s41419-017-0027-2.

72. Kuchay, S. *et al.* NS5A Promotes Constitutive Degradation of IP3R3 to Counteract Apoptosis Induced by Hepatitis C Virus. *Cell Rep.* **25**, 833-840.e3 (2018).

73. Nieto-Torres, J. L. *et al.* Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. (2015) doi:10.1016/j.virol.2015.08.010.

74. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* 1–8 (2020) doi:10.1038/s41586-020-2332-7.

75. Wright, M. T., Kouba, L. & Plate, L. Thyroglobulin Interactome Profiling Uncovers Molecular Mechanisms of 1 Thyroid Dyshormonogenesis 2 3. doi:10.1101/2020.04.08.032482.

76. Fonslow, B. R. *et al.* Single-step inline hydroxyapatite enrichment facilitates identification and quantitation of phosphopeptides from mass-limited proteomes with MudPIT. *J Proteome Res* **11**, 2697–2709 (2012).

77. Keilhauer, E. C., Hein, M. Y. & Mann, M. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). *Mol. Cell. Proteomics* **14**, 120–135 (2015).

78. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

79. Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

80. Bozidis, P., Williamson, C. D. & Colberg-Poley, A. M. Isolation of Endoplasmic Reticulum, Mitochondria, and Mitochondria-Associated Membrane Fractions from Transfected Cells and from Human Cytomegalovirus-Infected Primary Fibroblasts. *Curr. Protoc. Cell Biol.* **37**, 3.27.1-3.27.23 (2007).

81. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

82. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).

CHAPTER 3

COMPARATIVE INTERACTOMICS OF SARS-COV-2 NONSTRUCTURAL PROTEIN 3 AND HUMAN CORONAVIRUS HOMOLOGS

This chapter is adapted from "Comparative Host Interactomes of the SARS-CoV-2 Nonstructural Protein 3 and Human Coronavirus Homologs" published in *Molecular & Cellular Proteomics* and has been reproduced with the permission of the publisher and my co-authors: Katherine M. Almasy, Lars Plate.

3.1 Summary
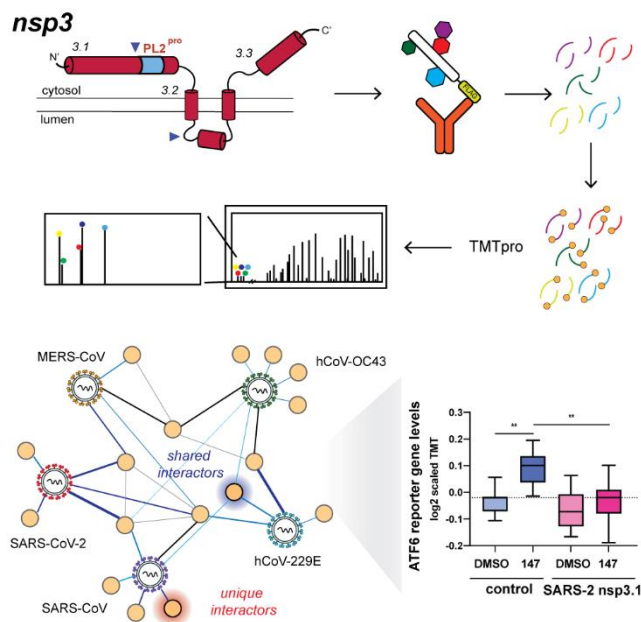
Human coronaviruses have become an increasing threat to global health; three highly pathogenic strains have emerged since the early 2000s, including most recently SARS-CoV-2, the cause of COVID-19. A better understanding of the molecular mechanisms of coronavirus pathogenesis is needed, including how these highly virulent strains differ from those that cause milder, common-cold like disease. While significant progress has been made in understanding how SARS-CoV-2 proteins interact with the host cell, non-structural protein 3 (nsp3) has largely been omitted from the analyses. Nsp3 is a viral protease with important roles in viral protein biogenesis, replication complex formation, and modulation of host ubiquitinylation and ISGylation. Herein, we use affinity purification-mass spectrometry to study the host-viral protein-protein interactome of nsp3 from five coronavirus strains: pathogenic strains SARS-CoV-2, SARS-CoV, and MERS-CoV; and endemic common-cold strains hCoV-229E and hCoV-OC43. We divide each nsp3 into three fragments and use tandem mass tag technology to directly compare the interactors across the five strains for each fragment. We find that few interactors are common across all variants for a particular fragment, but we identify shared patterns between select variants, such as ribosomal proteins enriched in the N-terminal fragment (nsp3.1) dataset for SARS-CoV-2 and SARS-CoV. We also identify

unique biological processes enriched for individual homologs, for instance nuclear protein important for the middle fragment of hCoV-229E, as well as ribosome biogenesis of the MERS nsp3.2 homolog. Lastly, we further investigate the interaction of the SARS-CoV-2 nsp3 N-terminal fragment with ATF6, a regulator of the unfolded protein response. We show that SARS-CoV-2 nsp3.1 directly binds to ATF6 and can suppress the ATF6 stress response. Characterizing the host interactions of nsp3 widens our understanding of how coronaviruses co-opt cellular pathways and presents new avenues for host-targeted antiviral therapeutics.

All supplemental materials can be found in Appendix 2 or accompanying the online publication.

3.2   Introduction

Coronaviruses are a family of positive-sense, single-stranded RNA viruses that typically cause upper respiratory infection in humans. Four endemic strains have been characterized that cause symptoms resembling those of the common cold. However, since 2002, three more pathogenic strains have emerged: SARS-CoV in 2002, MERS-CoV in 2012, and SARS-CoV-2, the causative agent of COVID-19, in 2019(1–5). Some of the differences in pathogenicity can be attributed to differential receptor binding, for example, SARS-CoV and SARS-CoV-2 utilize the angiotensin converting enzyme 2 (ACE2) receptor, while 229E (a common-cold causing strain) uses the human aminopeptidase N receptor(5–7). At the same time, the engagement of viral proteins with different host proteins or complexes within infected cells is equally critical to understand changes in pathogenicity. These engagements alter the native protein-protein interaction (PPI) architecture of the cell and have been shown to perform various pro-viral functions such as suppression of the type I interferon system for immune evasion purposes(8–10).

The coronavirus genome is among the largest RNA virus genomes, at approximately 30 kilo base pairs in length. The 3' third of the genome encodes for the four structural proteins used to construct new virions, as well as several accessory factors shown to be important for pathogenesis. The 5' two thirds of the genome consist of two open reading frames (orf1a and orf1b) that encode for sixteen non-structural proteins (nsps) that perform a number of functions throughout the viral life cycle, including replication and proofreading of the RNA genome and formation of the replication-transcription complex. The largest of these proteins, at approximately 2000 amino acids, is nsp3. Nsp3 is a large multi-domain protein, of which the papain-like-protease (PL2^{Pro}) domain has been most closely studied. In addition to autoproteolysis of the viral polyprotein, the PL2^{Pro} domains possess both deubiquitinase and deISGylation activities(11–13). Additionally, nsp3 in complex with nsp4 and nsp6 has been shown to be sufficient for formation of the double-membraned vesicles (DMVs) implicated in the CoV replication cycle(14,15). Expression of the C-terminus of nsp3 and full-length nsp4, while not enough to induce DMV formation, does cause zippering of the ER membrane(16). However, role(s) of nsp3 outside of the PL2^{Pro} remain less well understood(17).

Herein, we focused our analysis on four nsp3 homologs from the genus betacoronavirus (hCoV-OC43, MERS-CoV, SARS-CoV, SARS-CoV-2) and one homolog from the genus alphacoronavirus (hCoV-229E). Within the betacoronaviruses, hCoV-OC43 is from clade A, SARS-CoV and SARS-CoV-2 are from clade B, and MERS-CoV is from clade C. The domain organization of nsp3 varies widely among coronavirus genera, and even from strain to strain. Despite the differences, ten

regions are conserved across all coronavirus variants: two ubiquitin-like domains (UBLs), a glutamic acid-rich domain, protease domain, two transmembrane regions separated by an ER ectodomain, and two C-terminal Y domains (**Figure 8-1**).

Affinity purification-mass spectrometry (AP-MS) has been used extensively to characterize the coronavirus interactome, including in two large studies by Gordon et al. to characterize the SARS-CoV-2 host protein interactions and compare these to the SARS-CoV and MERS-CoV interactomes[18,19]. Another study compared the interactomes of isolated SARS-CoV and SARS-CoV-2 PL2[pro] domains[11]. However, the remaining parts of nsp3 have been missing from SARS-CoV-2 interactome studies thus far, likely because of their complex topology and large size making expression of the protein difficult. To circumvent this problem, we divided the nsp3 protein into three fragments based on earlier interrogation of the SARS intraviral interactome[14,20]. These fragments are referred to as nsp3.1 for the N-terminal fragment, nsp3.2 for the middle fragment, and nsp3.3 for the C-terminal fragment[21]. We expressed each fragment from each of the five viruses listed above (hCoV-229E, hCoV-OC43, MERS-CoV, SARS-CoV, SARS-CoV-2). We employed tandem mass tag (TMTpro 16plex) isobaric tagging technology, which enables highly multiplexed analysis for direct comparison of interactor abundances across homologs from all strains. Previously, we demonstrated the use of AP-MS and TMT technologies in comparing the interactomes of the coronavirus nsp2 and nsp4 proteins from three strains: SARS-CoV-2, SARS-CoV, and hCoV-OC43[22]. We now extend the analysis to the understudied nsp3 protein across additional viral strains. In particular, comparing host protein interactions for homologs from multiple coronavirus strains and genera may provide insight into how the molecular mechanisms of pathogenic coronaviruses differ from endemic ones, as well as the evolution of functions that nsp3 assumes across these strains.

Our study finds that very few interactors are shared among the five strains, although several interactors are common among the mostly conserved SARS variants. Several previously unknown pathways are discovered to be highly enriched with individual variants, such as ERAD processing for SARS-CoV nsp3.2 and nuclear import for 229E nsp3.2. In addition, we find that SARS-CoV-2 nsp3.1 interacts with the unfolded protein response (UPR) transcription factor ATF6 and suppresses the ATF6 pathway in both basal and activated conditions. These discoveries open the door for further work delineating the role of nsp3 in coronavirus infections, with a particular emphasis on the variation in roles between different CoV strains.

3.3   Results

3.3.1   Expression of nsp3 truncations and AP-MS of CoV nsp3

The orf1a and orf1b open reading frames encode for 16 non-structural proteins which perform crucial roles during the viral life cycle, including replication of the genome and formation of double-membraned vesicles. Non-structural protein 3 (nsp3) is the largest of the 16 non-structural proteins; the full-length protein is 1586 – 1945 amino acids long (177-217 kDa) (**Figure 3-1A**). While the domain organization is different among CoV variants, several domains are shared, including multiple ubiquitin-like-domains, at least one highly conserved papain-like protease (PL2[Pro]) domain, and an ER luminal domain postulated to be important for nsp4 binding and double membrane vesicle (DMV) formation (**Figure 3-1B-C, Figure 8-1**).
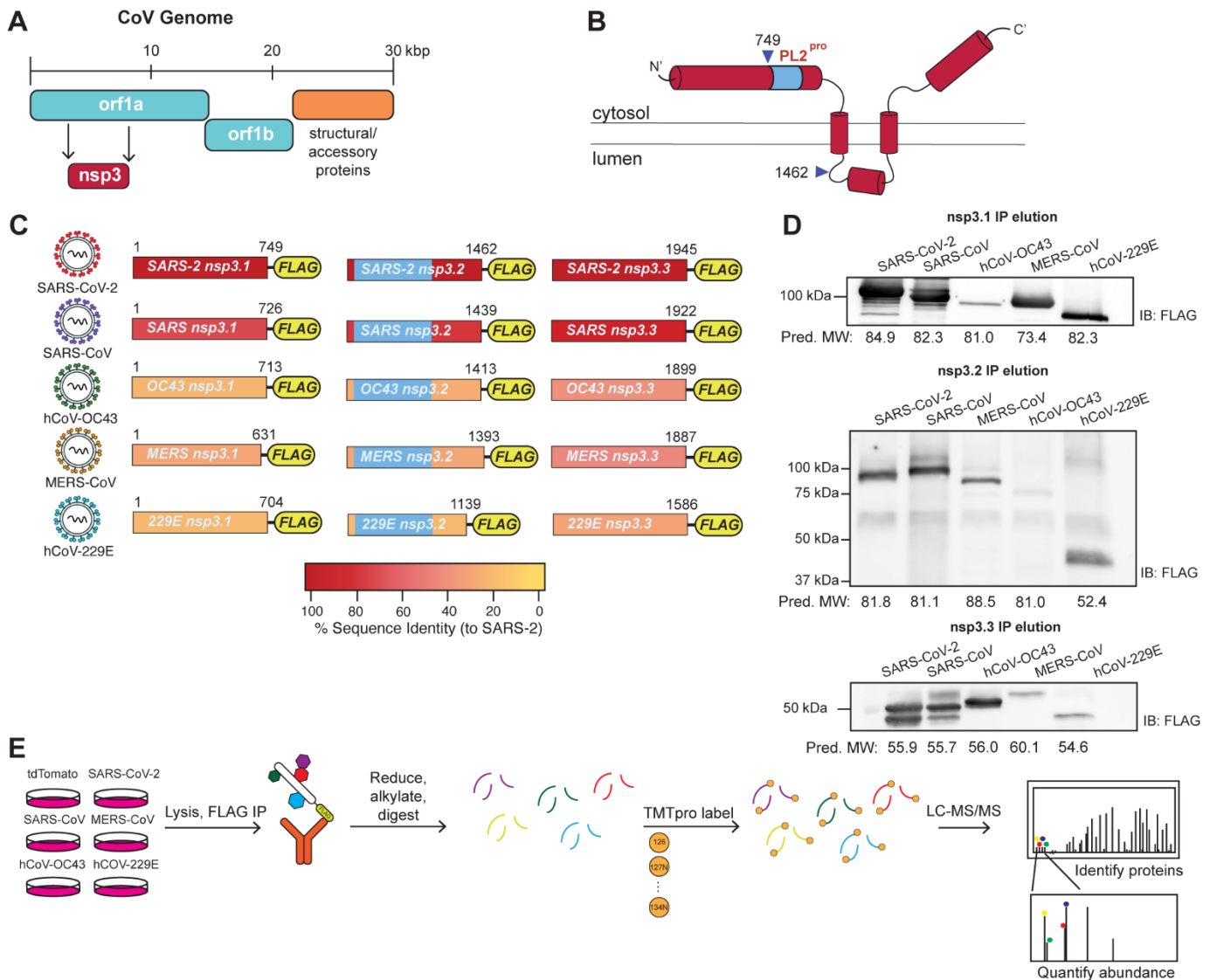
Figure 3-1. Design and expression of CoV nsp3 truncations for affinity-purification mass spectrometry (AP-MS).

A. Coronavirus (CoV) genome schematic indicating the regions encoding orf1a, orf1ab, and structural/accessory proteins. Nsp3 is encoded within orf1a.

B. General schematic of SARS-CoV-2 nsp3 protein topology. Nsp3 has two single-span transmembrane regions anchoring the protein in the ER membrane, with a small luminal domain and both N and C terminal regions in the cytosol. The conserved papain-like protease (PL2^{pro}) domain is on the N-terminal cytosolic portion. For this study, the nsp3 protein was truncated into three fragments: nsp3.1 (1-749), nsp3.2 (750-1462), and nsp3.3 (1463-1945), numbering corresponding to SARS-CoV-2 nsp3.

C. The nsp3 truncations for homologs from all five human coronaviruses used in this study. All fragments contain a C-terminal FLAG tag for affinity-purification. Percent sequence identity compared to SARS-CoV-2 is indicated. The PL2^{pro} domain in nsp3.2 homologs is highlighted in light blue.

D. Western blotting of immunopurified nsp3.1, nsp3.2, and nsp3.3 homologs after transient transfection in HEK293T cells. Predicted MW of proteins is indicated below.

E. AP-MS workflow to identify virus-host protein interactions of nsp3 fragment homologs. HEK293T cells were transfected with corresponding homologs and lysates were immunopurified using anti-FLAG beads to enrich for viral proteins in complex with host interactors. Proteins were reduced, alkylated, and digested with trypsin. Peptides were then labeled with tandem mass tags (TMTpro) and analyzed by tandem mass spectrometry (LC-MS/MS) to both identify and quantify host interactors.

Initial attempts at expression of the full-length construct for SARS-CoV-2 nsp3 were unsuccessful, likely due to the complex structure and topology of the full-length protein. To study the interactome, we thus divided the protein into three portions based on a prior interactome study carried out on similar truncations of the SARS-CoV nsp3 homolog(20). The N-terminal fragment (nsp3.1), comprising the first ubiquitin-like-domain through the SARS unique domain, is expected to localize exclusively to the cytosol (**Figure 3-1C, Figure 8-1**). The second portion (nsp3.2), starting before the second ubiquitin-like domain and ending just after the first transmembrane region, includes the papain-like-protease (PL2^{Pro}) domain. The C-terminal fragment (nsp3.3) starts with the ER-localized 3Ecto-domain, includes the second transmembrane region and ends with the C-terminus of the protein. Based on multiple sequence alignments, similar N-terminal, middle, and C-terminal constructs were created for the SARS-CoV, MERS-CoV, hCoV-OC43, and hCoV-229E nsp3 homologs (**Figure 8-1**). Each construct contains a C-terminal FLAG tag for affinity purification (**Figure 3-1C**).

Constructs were transfected into HEK293T cells, and expression was confirmed by immunoprecipitation and immunoblotting for the FLAG tag and/or detection by mass spectrometry (**Figure 3-1D**). Overall, SDS-PAGE gel patterns matched with the expected molecular weights; the exception was the MERS nsp3.1 fragment, which ran larger than expected. In addition, confocal immunofluorescence microscopy of transfected HEK293T cells shows that nsp3.2 and nsp3.3 homologs colocalize with PDIA4, an ER marker, while nsp3.1 exhibits cytosolic localization as expected (**Figure 8-2A-C**). Although kidney cells do not represent a primary tissue target of the virus, previous work has identified HEK293T cells as appropriate cell lines to recapitulate relevant CoV protein interactions with host factors(18,22). While the constructs showed variable expression levels, all fragments were reproducibly detectable by mass spectrometry, and peptides coverage spanned the length of each fragment (**Figure 3-1D**). For native co-immunoprecipitations with host interactors, protein constructs were expressed, lysed in mild detergent buffer to maintain interactions, and co-immunoprecipitated from lysates using anti-FLAG beads. After confirmation of IP by silver stain, samples were reduced, alkylated, and trypsin digested. Samples were labeled using TMTpro 16plex reagents for MS2 quantification of peptides abundances(23) (**Figure 3-1E**). Between 2-4 co-immunoprecipitation replicates for each construct were pooled into a single TMTpro 16plex run, along with mock IPs from tdTomato transfected cells to establish the background signal. The final datasets consisted of 184 IPs across the 16 constructs (tdTomato & 15 viral protein constructs) (**Figure 8-3**). We quantified the abundance of nsp3 bait proteins across the biological replicates showing consistent enrichment compared to the mock control (**Figure 8-3G-I**).

After identification and quantification of interactors using Proteome Discoverer, a variable cutoff method was used to determine high- and medium- confidence interactors for each construct based on enrichments compared to the tdTomato background control.

3.3.2 Comparison of CoV nsp3.1-host interactors

The N-terminal fragments of nsp3 contain a conserved ubiquitin-like domain (Ubl1), a highly variable or acidic-rich region, a conserved macrodomain (Mac1), followed by strain-specific domains (**Figure 3-2A, Figure 8-1**). In hCoV-229E and hCoV-OC43, this fragment includes a papain-like protease domain (termed PL1^{Pro}). This first PL1^{Pro} domain is absent from all viruses in clades B and C of the betacoronaviruses, hence its absence in MERS-CoV and the two SARS variants. In contrast, SARS strains contain a SARS-unique domain (SUD) consisting of 3 sequential macrodomains. Overall,
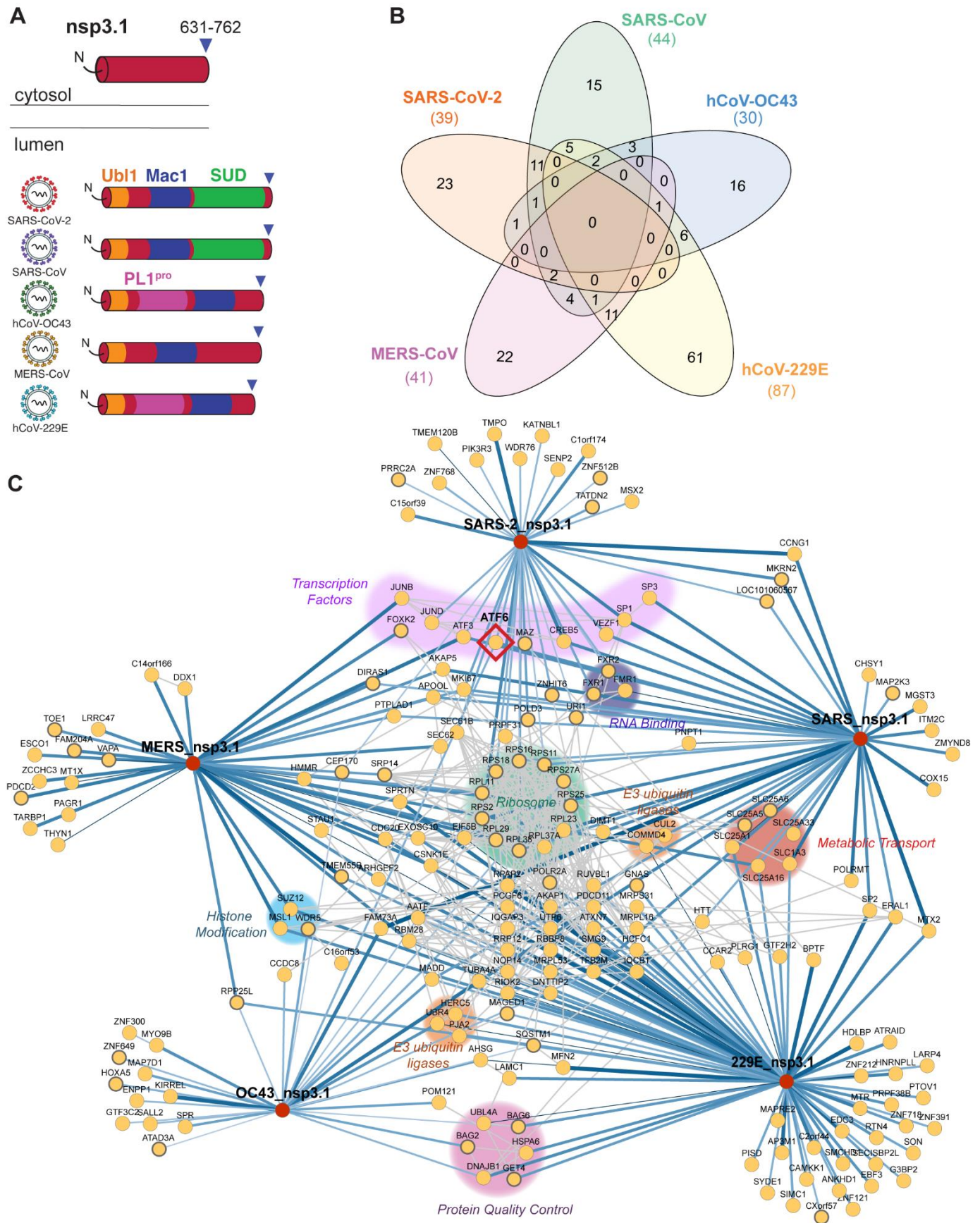
Figure 3-2. Identification of CoV nsp3.1 host interactors.

A. Schematic of nsp3.1 topology for all five CoV homologs. Nsp3.1 is a cytosolic fragment, comprising residues from 1 to 631-762. All fragment homologs contain a ubiquitin-like domain (Ubl1, yellow) and a conserved macrodomain

56

(Mac1, blue). SARS-CoV-2 and SARS-CoV contain a SARS-unique domain (SUD, green), while hCoV-OC43 and hCoV-229E contain a papain-like protease domain (PL1^{pro}, pink).

B. Venn diagram showing the number of unique and shared host interactors amongst all five CoV nsp3.1 homologs. Total interactors for each homolog are shown in parentheses.

C. Network plot of virus-host interactors. Individual nsp3.1 homologs are shown as red circles, while host interactors are shown as yellow circles. Blue lines indicate virus-host protein interactions, where line width and shade are wider/darker for more highly enriched interactions. Grey lines indicate known host-host protein interactions from the STRING database. Notable clusters of host proteins are highlighted. The transcription factor ATF6 (red diamond) interacts with SARS-CoV-2, SARS-CoV, and hCoV-OC43 and was subjected to later functional follow-up (**Figure 3-5**).

this region of nsp3 is most variable from strain to strain, despite several of the predicted domains being conserved (**Figure 3-1B-C, Figure 8-1**).

The final dataset for the N-terminal fragment analysis combined 5 mass spectrometry runs containing 62 co-IP samples (tdTomato 15, SARS-CoV-2 13, SARS-CoV 9, OC43 9, MERS 8, 229E 8) (**Figure 8-2A**). We identified a robust set of high-confidence interactors for each fragment: 39 for SARS-CoV-2, 44 for SARS-CoV, 41 for MERS-CoV, 30 for hCoV-OC43, and 87 for hCoV-229E (**Figure 8-4**).

Of these high-confidence interactors, none were common across all 5 strains, and none were common to all betacoronaviruses (**Figure 3-2B**). Fourteen interactors were common between the two SARS variants, and 2 of these 14 interactors, MKI67 and ATF3, were also shared with the MERS fragment. MKI67 is required to maintain individual chromosomes in the cytoplasm during mitosis, and ATF3 is a cyclic AMP-dependent transcription factor which negatively regulates the cellular antiviral response. Lastly, we identified 25 high confidence interactors unique to SARS-CoV-2, 16 unique to SARS-CoV, 17 unique to hCoV-OC43, 62 unique to hCoV-229E, and 23 unique to MERS-CoV. Notably contained in this fragment is the SARS-unique domain, which is exclusively found in SARS-CoV and SARS-CoV-2 nsp3. As noted, 14 proteins were identified as interactors of both fragments. Eleven of these interactors were exclusive to the SARS variants, including several ribosomal proteins (RPL38, RPS11, RPS2, RPS16) and assorted factors (FMR1, MKRN2, SP1, VEZF1, CREB5, CCNG1, and PEX11B).

We also identify 9 shared interactors between OC43 and 229E, both of which have a PL1^{pro} domain (UBR4, TUBA4A, SQSTM1, RPP25L, MADD, DNAJB1, BAG6, AHSG, AATF). BAG6 and DNAJB1 are co-chaperones components involved in protein quality control (24,25), while UBR4 is an E3 ubiquitin ligase involved in membrane morphogenesis and SQSTM1 is an autophagy receptor(26,27). These shared interactors may represent possible targets of the PL1^{pro} domains in OC43 and 229E.

We performed the same filtering on proteins observed to be less enriched with the viral bait and more enriched with the tdTomato background, interpreting these largely to be proteins which bind non-specifically to the FLAG beads in the absence of viral bait and its interactors (**Figure 8-4**). To determine any overlap between our data set and contaminant interactors, we queried all identified high-confidence interactors in the Contaminant Repository for Affinity Purification (CRAPome) dataset for common contaminant proteins found in control AP-MS experiments(28). Using a frequency threshold of 25%, only 34 of 186 unique high-confidence interactors overlapped with the CRAPome (**Figure 8-5**), emphasizing the specificity of host interactors for viral bait proteins. Many of these overlapping interactors are ribosomal

and chaperone proteins with central roles in protein biogenesis and may be expected to be found in control AP-MS experiments.

As an alternative method to map shared and unique interactors for nsp3.1 from the different strains, we performed hierarchical clustering of the grouped protein abundance Z-scores for all identified high-confidence interactors (**Figure 8-6**). Consistent with the Venn diagram, we observed a cluster of shared interactors for the SARS construct (cluster 4), distinct interactors of 229 (cluster 5) and MERS (cluster 3). A large cluster (7) contained interactors present in four of the strains but absent in SARS-CoV-2.

To ascertain how these proteins cluster into cellular pathways with potential pro- or anti-viral roles, we grouped the high-confidence interactors into a network plot. To highlight connections between the binding partners, known protein interactions based on the STRING database were included (**Figure 3-2C**). Several groups of associated interactors emerged, including a large cluster of eleven ribosomal proteins (RPS2, RPS11, RPS16, RPS18, RPS25, RPS27A, RPL11, RPL23, RPL29, RPL37A, RPL38). Other notable clusters include factors involved in protein quality control (BAG2, BAG6, HSPA6, DNAJB1, UBL4A, GET4), metabolic transport (SLC1A3, SLC25A1, SLC25A5, SLC25A6, SLC25A16, SLC25A33), histone modification (SUZ12, MSL1, WDR5), RNA-binding proteins (FXR1/2, FMR1), and E3 ubiquitin ligases (CUL2, COMMD4, HERC5, UBR4, PJA2). Interactions between nsp3 and FXR1/2, as well as several ribosomal proteins were also prominently observed in a BioID proximity ligation data set[29]. Additionally, ten transcription factors were identified with at least one interaction to severely pathogenic betacoronaviruses. These include VEZF1, a transcription factor for IL-3[30], ATF3, a broad negative regulator of NF-κB, IL-4, IL-5, IL-13 expression[31], and ATF6, which turns on expression of the ATF6 branch of the unfolded protein response (UPR)[32].

As a last method to identified pathways represented by the interactors, we filtered the list of proteins through EnrichR to determine the most common gene ontology terms related to biological processes among each interactome (**Figure 8-7**). For both SARS variants, as well as MERS-CoV, we observed a strong enrichment for several ribosomal-related processes, including SRP-dependent cotranslational protein targeting, ribosome biogenesis, and rRNA metabolic processes. These processes were not observed in the less pathogenic hCoV-OC43 and hCoV-229E strains. For SARS-CoV, but not SARS-CoV-2, we also observed several mitochondrial related processes, such as mitochondrial transport, and mitochondrial RNA metabolic processes. hCoV-OC43 showed enrichment of factors associated with an upregulation of transcription from RNA polymerase II factors. Lastly, despite having the highest overall number of high confidence interactors, no individual pathways showed enrichment at the high confidence interactor level for hCoV-229E.

3.3.3 Comparison of nsp3.2 CoV interactors

Next, we turned to the middle fragment of the nsp3 constructs, which includes the highly conserved PL2^{Pro} domain, the protease responsible for the self-cleavage of several coronavirus non-structural proteins from the orf1a/b polypeptide[33]. This domain also has deubiquitination and de-ISGylation activity, an activity well-studied in the pathogenic coronaviruses, but less studied in the endemic strains[34,35]. The importance of this domain for both viral polypeptide processing and remodeling of host ubiquitination/ISGylation modifications make it an intriguing target for drug development. Also included in this fragment are the second ubiquitin-like domain UBL2, the first transmembrane domain, and the first portion of the ER ectodomain (**Figure 3-3A**).
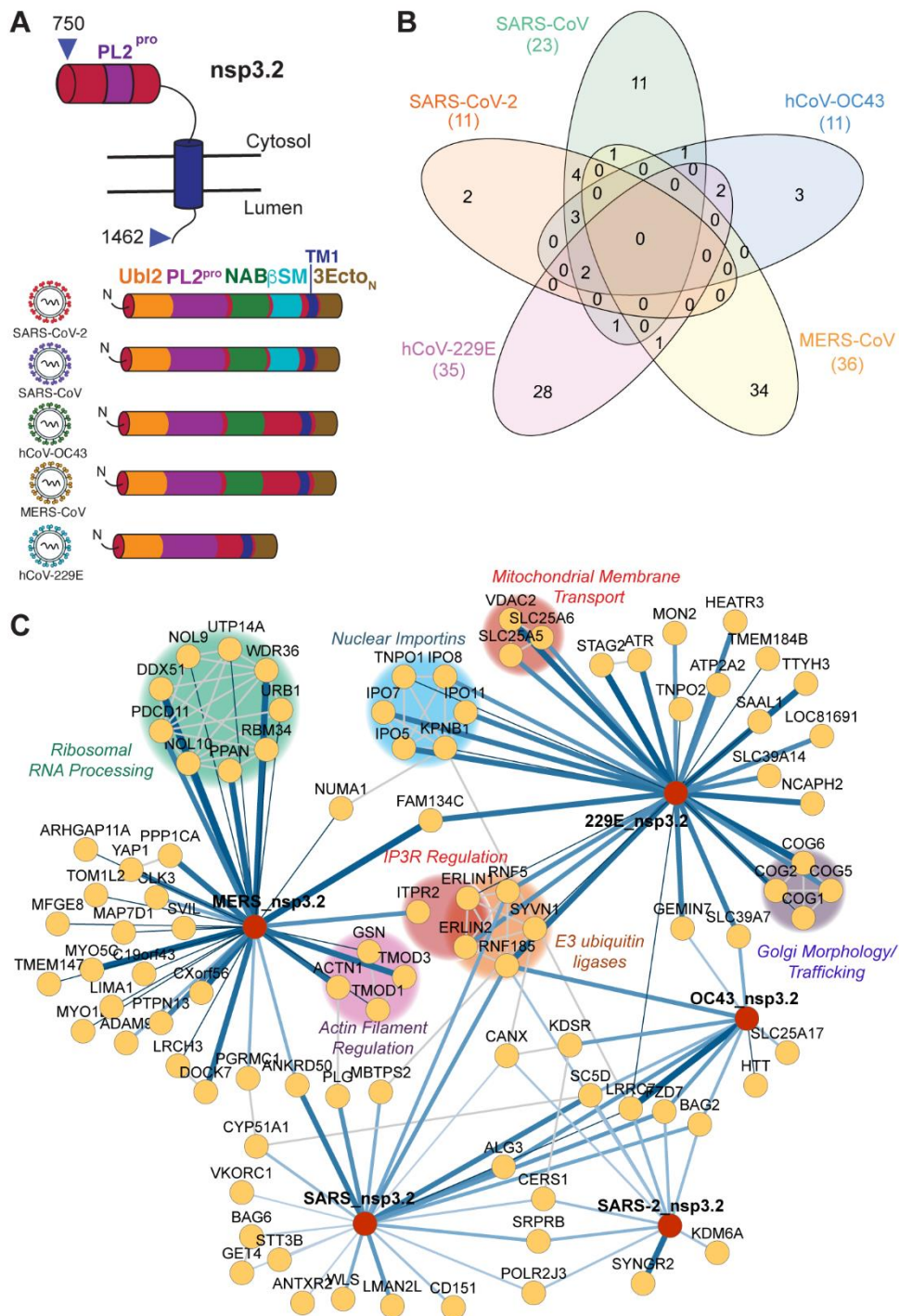
Figure 3-3. Comparative interactomics of CoV nsp3.2 homologs.

A. Schematic of nsp3.2 protein topology for all five CoV homologs. Nsp3.2 contains both cytosolic, transmembrane, and luminal regions. All homologs contain a ubiquitin-like domain (Ubl2, yellow), a papain-like protease domain (PL2^{pro}, purple), a transmembrane region (TM1, blue), and the N-terminal portion of the ectodomain (3Ecto$_N$, brown). Betacoronavirus homologs also contain a nucleic acid binding domain (NAB, green), while SARS strains also have a betacoronavirus-specific marker (βSM, light blue).

B. Venn diagram showing the number of unique and shared host interactors amongst all five CoV nsp3.2 homologs. Total interactors for each homolog are shown in parentheses.

C. Network plot of virus-host interactors. Individual nsp3.2 homologs are shown as red circles, while host interactors are shown as yellow circles. Blue lines indicate virus-host protein interactions, where line width and shade are wider/darker for more highly enriched interactions. Grey lines indicate known host-host protein interactions from the STRING database. Notable clusters of host proteins are highlighted.

59

The final dataset for the middle fragment analysis combined 3 mass spectrometry runs containing 59 co-IP samples (tdTomato 10, SARS-CoV-2 11, SARS-CoV 11, hCoV-OC43 10, MERS-CoV 10, hCoV-229E 7) (**Figure 8-3B,E**). In total, we identified 11 high confidence interactors for SARS-CoV-2, 23 for SARS-CoV, 11 for hCoV-OC43, 36 for MERS-CoV, and 35 for hCoV-229E (**Figure 3-3B, Figure 8-8**). Nine proteins were observed as high confidence interactors for both SARS-CoV and SARS-CoV-2. No proteins were observed to be common among all 5 strains at the high confidence level (6 at the medium confidence level), and no proteins were observed to be exclusive interactors of betacoronaviruses at the high confidence level (2 at the medium confidence level). Lastly, we identified 2 high confidence interactors unique to SARS-CoV-2, 11 unique to SARS-CoV, 3 unique to OC-43, 28 unique to 229E, and 34 unique to MERS. When querying the CRAPome dataset, we only found overlap in 14 of 93 identified proteins, supporting the specificity of the interactors for the nsp3.2 bait proteins (**Figure 8-5**).

We also grouped the high-confidence interactions into a network plot highlighting previously known interactions We carried out hierarchical clustering of the grouped abundance Z-scores for all high-confidence interactors. The resulting heatmap (**Figure 8-9**) more clearly highlights a cluster of shared interactions for all five strains (cluster 2), although intensities vary, which could explain why not all may have passed our stringent high-confidence cutoff. Cluster 1 contains binding partners found in four of the strains, but absent in MERS. A striking observation were two large clusters of strain-specific interactors for 229E (cluster 3) and MERS (cluster 4) respectively.

from String DB (**Figure 3-3C**). We identify several ER-associated degradation (ERAD) components as high confidence interactors of SARS-CoV nsp3.2. While some of these components were identified as high confidence interactors of SARS-CoV-2 as well, the number and magnitude were both much less. Analysis of GO terms associated with both the medium and high confidence interactors confirmed the enrichment of the ERAD machinery, as well as several related processes such as membrane proteolysis and cellular response to ER stress (**Figure 8-10**).

The 229E fragment revealed a high number of unique interactors related to nuclear importins and translocation of proteins into the nucleus (IPO5, IPO7, IPO11, IPO8, KPNB1, TNPO1). A recent report identified a role of MHV nsp3 in forming pores across the DMV membrane in coronavirus infection, potentially for the purpose of dsRNA export from the DMVs(36). The nuclear importins could be additional host factors co-opted by 229E or alphacoronaviruses specifically for a similar purpose. Additional interactors of the 229E fragment include ERLIN1 and ERLIN2, which we previously identified as being interactors of both SARS-CoV-2 and SARS-CoV nsp2 and nsp4. These proteins are also associated with ERAD, most notably in regulating ubiquitination and degradation of the inositol 1,4,5-triphosphate receptor IP3R(37). Lastly, other unique interactor of hCoV-229E nsp3.2 included a cluster of mitochondrial membrane transporters (SLC25A6, SLC25A5, VDAC2), as well as a cluster of subunits of the conserved oligomeric Golgi (COG) complex involved in intra-Golgi mediated vesicle transport. COG6 was identified in a recent CRISPR screen as essential for hCoV-229E replication(38), and the intra-Golgi mediated vesicle transport was a pathway enriched only in our hCoV-229E nsp3.2 dataset, not appearing for the other strains (**Figure 8-10**).

Lastly, for MERS nsp3.2. we identified a complex of ribosomal RNA (rRNA) processing factors as highly unique interactors, including DDX51, NOL9, NOL10, UTP14A, WDR36, URB1, RBM34, PPAN, and PDCD11 (**Figure 3-3C**). Consistent with this finding, RNA processing and ribosome biogenesis were the most highly-enrichment GO terms specific

to the MERS nsp3.2 fragment (**Figure 8-10**). Other viruses have been shown to target ribosomal biogenesis to facilitate infection, such as Human cytomegalovirus (HCMV) and HIV-1(39,40). The specificity of these interactors for MERS nsp3.2 is striking and may represent a unique replication strategy to modulate host protein synthesis.

3.3.4   Comparison of nsp3.3 CoV interactors

The C-terminal fragments begin with the second half of the ER ectodomain, continuing through the second transmembrane region and Y domains, ending with the nsp3/nsp4 cleavage site. The Y domains, although one of the more conserved domains across the 5 strains, remains largely unstudied. In co-expression experiments of individual SARS-CoV-2 nsp3 fragments and nsp4, we observed co-immunoprecipitation of this C-terminal fragment with nsp4 (**Figure 8-11**). In contrast, the N-terminal and middle fragments did not co-immunoprecipitate with nsp4.

The final dataset for the middle fragment analysis combined 5 mass spectrometry runs containing 63 co-IP samples (tdTomato 15, SARS-CoV-2 14, SARS-CoV 9, hCoV-OC43 9, MERS-CoV 8, hCoV-229E 8) (**Figure 8-3C,F**). In total, we identified 75 high confidence interactors for SARS-CoV-2, 66 for SARS-CoV, 52 for hCoV-OC43, 49 for MERS-CoV, and 76 for hCoV-229E (**Figure 3-4B, Figure 8-12**). One protein, SLC39A6, a zinc transporter, was observed as a high confidence interactor for all five strains (68 proteins observed as common medium confidence interactors). Five proteins, DERL3, TMEM33, SC5D, CERS1, and ALG8, were observed as high confidence interactors unique to the betacoronavirus strains. DERL3 is a functional component of the ERAD system, TMEM33 is involved in tubular ER network organization as well as being a component of the IRE1 and PERK stress response pathways, SC5D is involved in cholesterol biosynthesis, CERS1 is involved in lipid biosynthesis, and ALG8 is involved in N-glycan biosynthesis. Lastly, we identify 48 high confidence interactors unique to SARS-CoV-2, 17 unique to SARS-CoV, 20 unique to hCoV-OC43, 41 unique to hCoV-229E, and 18 unique to MERS-CoV (**Figure 3-4B**). As for the other nsp3 fragment, little overlap was identified with the CRAPome dataset (12 of 206 proteins), confirming the specificity of the interactions (**Figure 8-5**).

We carried out hierarchical clustering of the grouped abundance Z-scores for all high-confidence interactors. The resulting heatmap (**Figure 8-13**) highlights 6 clusters of interactors, including a cluster of shared interactors for all five strains (cluster 4), a cluster of shared betacoronavirus interactors (cluster 5), and a cluster of shared SARS interactors (cluster 6).

Despite these C-termini being the most conserved of the fragments across strains, there was a large divergence observed in the enriched pathways when searched for enriched GO-terms (Figure 8-14). In all strains except hCoV-OC43, ERAD was a highly enriched pathway. It has been postulated that nsp3, in tandem with nsp4 and nsp6, hijacks ERAD-tuning vesicles during double-membrane vesicle (DMV) formation, after ERAD factors EDEM1 and OS-9 were shown to co-localize with MHV-derived DMVs in infected cells(14)·(41). Our results may provide insight into a role for nsp3 in modulating this machinery

Unexpectedly, the biological process most enriched for the hCoV-229E fragment was protein N-linked glycosylation. N-linked glycosylation sites are localized at the N-terminus of the 3.3 fragment in the ectodomain. One of these interactors, MGAT1 (unique to the hCoV-229E interactome), was identified in the CRISPR screen as being an essential gene for hCoV-229E infection(38). In addition, we identify STT3B, a catalytic component of the oligosaccharyl transferase complex, as a high confidence hCoV-229E nsp3.3 interactor.
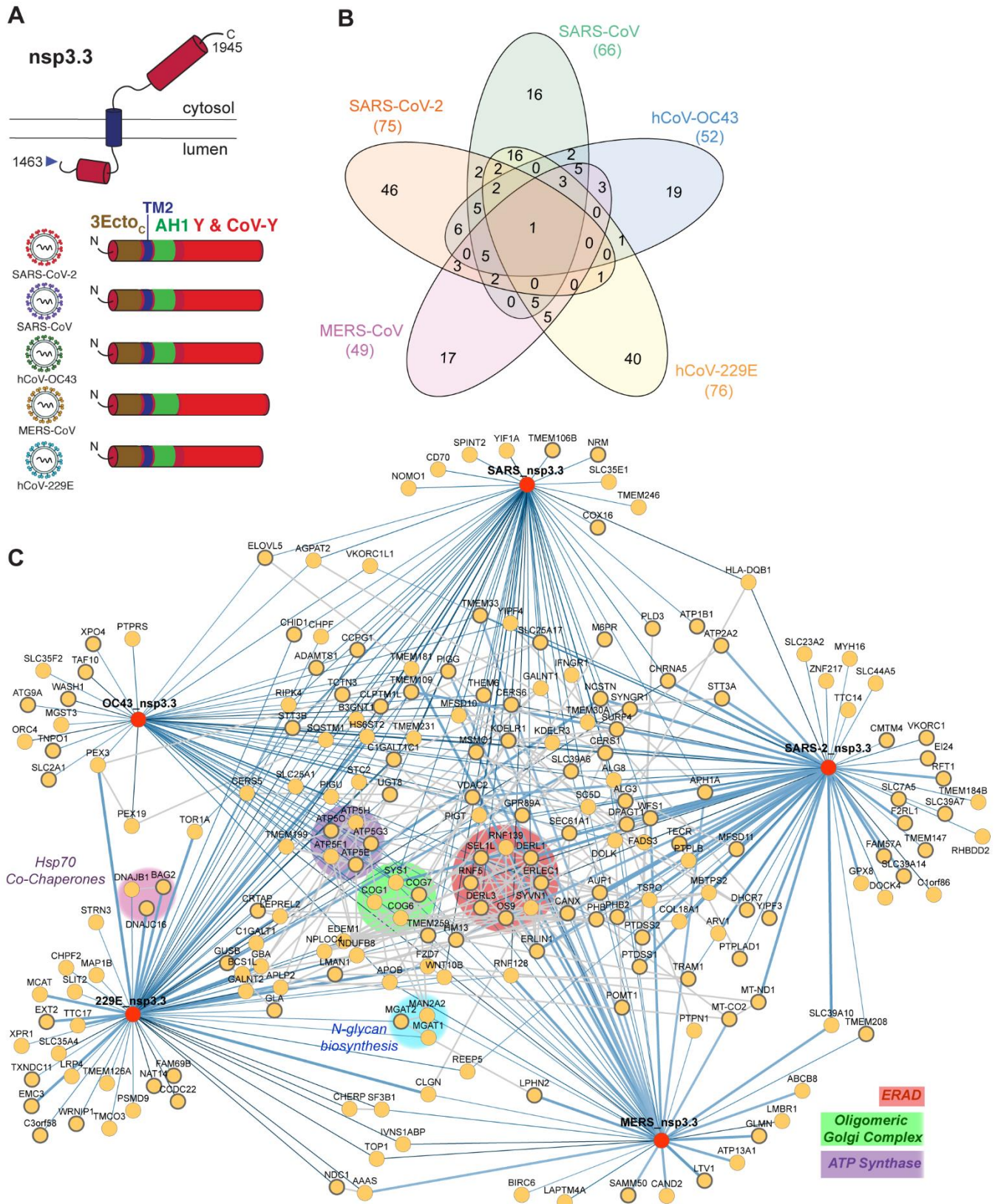
Figure 3-4. Identification of CoV nsp3.3 host interactors.

A. Schematic of nsp3.3 protein topology for all five CoV homologs. Nsp3.3 contains both cytosolic, transmembrane, and luminal regions. All homologs contain the C-terminal portion of the ectodomain (3Ecto_C, brown), the second transmembrane region (TM2, blue), a likely amphipathic helix (AH1, green), and a Y&CoV-Y domain (red).

B. Venn diagram showing the number of unique and shared host interactors amongst all five CoV nsp3.3 homologs. Total interactors for each homolog are shown in parentheses.
C. Network plot of virus-host interactors. Individual nsp3.3 homologs are shown as red circles, while host interactors are shown as yellow circles. Blue lines indicate virus-host protein interactions, where line width and shade are wider/darker for more highly enriched interactions. Grey lines indicate known host-host protein interactions from the STRING database. Notable clusters of host proteins are highlighted.

The patterns observed in our GO term analysis were consistent with network plots of high-confidence interactors (**Figure 3-4C**). Multiple homologs shared interactors involved in ERAD (RNF139, DERL1, DERL3, ERLEC1, SYNV1, OS9, RNF5, SEL1L) and the oligomeric Golgi complex (COG1, COG6, COG7, SYS1). A cluster of five components of the mitochondrial ATP synthase complex (ATP5E, ATP5G3, ATP5H, ATP5F1, and ATP5O) was enriched for hCoV-OC43 nsp3.3, while clusters of Hsp70 co-chaperones (DNAJB1, DNAJC16, BAG2) and *N*-glycan biosynthesis factors (MGAT1, MGAT2, MAN2A2) were enriched for hCoV-229E nsp3.3. Interestingly, a large swath of identified high-confidence interactors are interconnected based on STRING analysis but do not cluster into distinct biological categories outside of those previously mentioned.

3.3.5 SARS-CoV-2 nsp3.1 interacts with ATF6 and suppresses the ATF6 branch of the Unfolded Protein Response (UPR)

An intriguing observation was the shared interaction of the nsp3.1 fragment from SARS-CoV-2, SARS-CoV and hCoV-OC43 with ATF6 (**Figure 3-2C**), a transmembrane protein located across the ER membrane serving as one of the sensors of the ER unfolded protein response (UPR). Upon activation by ER stress, ATF6 is trafficked to the Golgi apparatus, where site-1 and site-2 proteases (SP1 & MBTPS2) cleave the protein, releasing an active transcription factor that serves to upregulate chaperones and other proteostasis factors which aid in relieving ER stress(32). Intriguingly, we also observed MBTPS2 as an interactor of SARS-CoV nsp3.2. MSTPS2 was found in a recent CRISPR to be an essential gene in the replication of SARS-CoV-2(38). Coronaviruses are known to upregulate the UPR. Specifically, the SARS-CoV spike protein was shown to be sufficient to upregulate transcription of factors such as GRP78/BiP and GRP94(42,43), both known to be downstream targets of ATF6(44,45). Furthermore, SARS-1 Orf8 can activate ATF6 and promote UPR induction(46), but a role for nsp3 in UPR modulation during virus infection is not known.

As technical validation of the interaction of the N-terminal fragment of SARS-CoV-2 nsp3 with ATF6, we used a HEK293T cell line expressing a doxycycline-inducible 3xFLAG-ATF6 construct(47). The affinity tag of the nsp3 N-terminal construct was replaced with a 2x strep tag (2xST) to allow for complementary immunopurifications (IP) and detection. We transiently transfected 3xFLAG-ATF6 cells with SARS-CoV-2 nsp3.1-2xST and carried out reciprocal co-IPs to validate the interactions (**Figure 3-5A**). We detected some background nsp3.1-2xST protein in the control (DMSO-treated) FLAG-IP where ATF6 was not induced. This was likely due to some leaked expression. However, the viral protein was noticeably enriched when 3xFLAG-ATF6 was robustly induced with doxycycline. ATF6 was also observed in reciprocal Co-IPs (**Figure 3-5A**), further validating the interaction between SARS-CoV-2 nsp3.1 and ATF6.
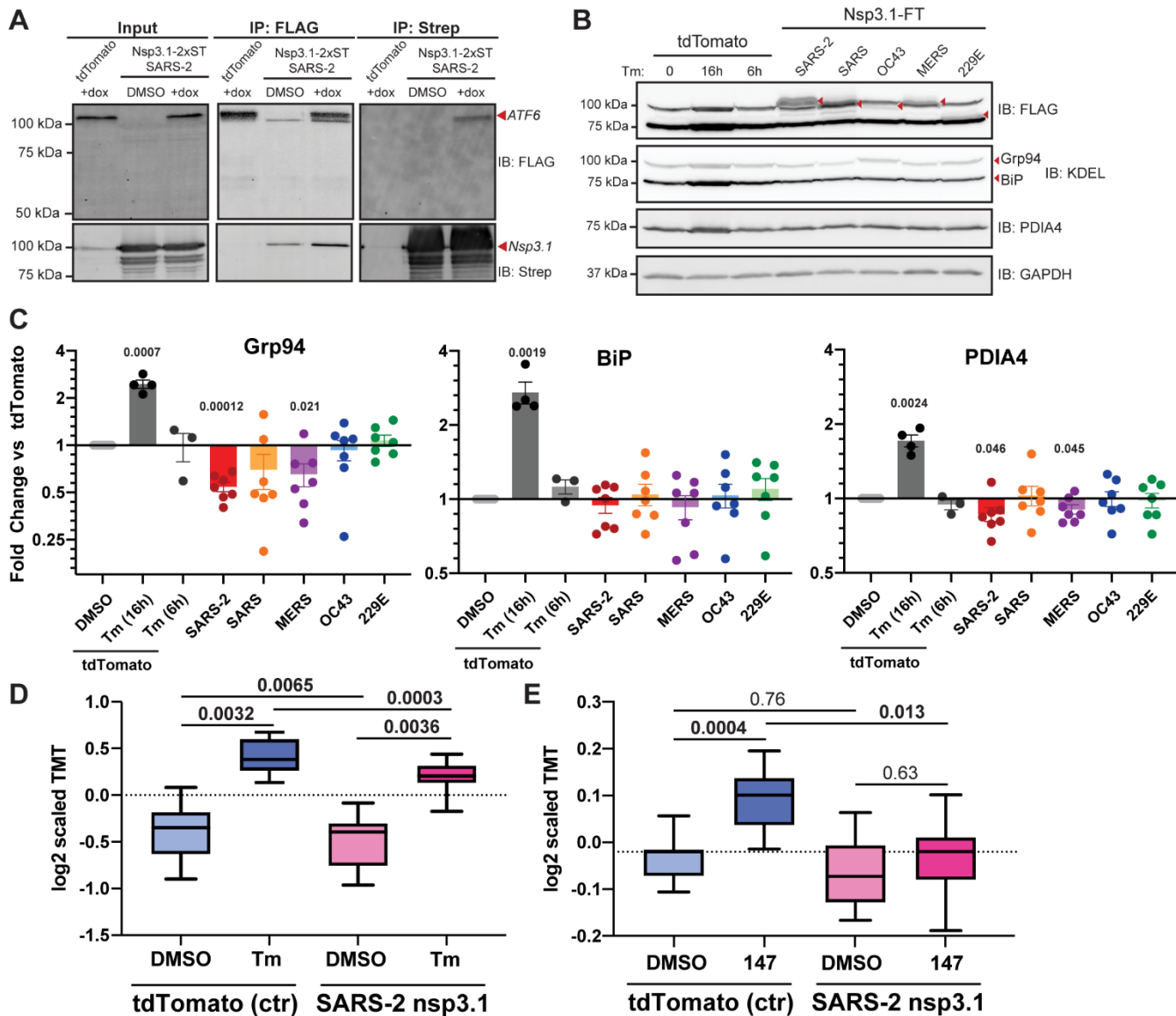
Figure 3-5. SARS-CoV-2 nsp3.1 interacts with ATF6 and suppresses the ATF6 branch of the Unfolded Protein Response (UPR).

A. Representative co-immunopurification (IP) western blots of 3xFT-ATF6 cells transfected with SARS-CoV-2 nsp3.1-2xST or tdTomato (control), lysed, and immunopurified for either FLAG or 2xStrepTag. Cells were treated with 100 nM doxycycline to induce 3xFT-ATF6 expression 24 h pre-harvest. Input and IP elution blots were probed with both anti-FLAG and anti-StrepTag antibodies. $n = 3$

B. Representative western blots of HEK293T cells transfected with nsp3.1-FT homologs or tdTomato (control), lysed, and probed for FLAG, ATF6 branch markers (Grp94, BiP, PDIA4), and GAPDH as a loading control. Control cells were either treated with DMSO or 6 µg/mL Tunicamycin (Tm) at 6 or 16 h pre-harvest to induce an ER stress response. TdTomato 16 h pre-harvest, $n = 4$; TdTomato 6 h pre-harvest, $n = 3$; all others, $n = 7$.

C. Quantification of western blots shown in (B). Error bars indicate average ±SEM. Paired student T-tests were used to test for significance between samples and tdTomato+DMSO control, with p-values < 0.05 shown.

D. Box-and-whisker plots of ATF6-regulated protein abundance measured by quantitative proteomics. HEK293T cells were transfected with tdTomato (control) or SARS-CoV-2 nsp3.1-FT and treated with DMSO or 6 µg/mL tunicamycin for 16 h pre-harvest. Shown are the distribution of scaled log2 TMT intensities for ATF6-regulated proteins based on published genesets(45). A one-way ANOVA with Geisser-Greenhouse correction and post-hoc Tukey's multiple comparison test was used to determine significance. Adjusted p-values are shown. $n = 4$ biological replicates in a single mass spectrometry run.

E. Box-and-whisker plots of ATF6-regulated protein abundance measured by quantitative proteomics. HEK293T cells were transfected with tdTomato (control) or SARS-CoV-2 nsp3.1-FT and treated with DMSO or 10 µM **147**

64

for 16 h pre-harvest. Shown are the distribution of scaled log2 TMT intensities for ATF6-regulated proteins based on published genesets(45). A one-way ANOVA with Geisser-Greenhouse correction and post-hoc Tukey's multiple comparison test was used to determine significance. Adjusted *p*-values are shown. *n* = 3 biological replicates in a single mass spectrometry run.

Given this observation of SARS-CoV-2 nsp3.1 directly binding to ATF6, we sought to determine if this interaction could constitute a mechanism by which the UPR may be modulated. Cells were transfected with all five nsp3.1 homologs and immunoblotting performed for known ATF6-specific target genes (GRP78/BiP, GRP94, PDIA4) to determine if upregulation of these proteins occurred. The nsp3.1 fragments were robustly expressed, but no notable upregulation of GRP78/BiP, GRP94 or PDIA4 was observed (**Figure 3-5B-C**). Instead, we saw a small but significant reduction in protein levels for GRP94 and PDIA4 with SARS-CoV-2 and MERS-CoV nsp3.1. The same experiment was performed to monitor changes of RNA transcript level, but similarly no upregulation of *BiP*, *PDIA4* or *ERDJ4* (IRE1 marker) transcripts was seen (**Figure 8-15**). We also did not observe any site-1/2 protease mediated-cleavage of ATF6 in SARS-CoV-2 nsp3.1-transfected samples (**Figure 3-5A**).

Given the small reduction in GRP94 and PDIA4 protein levels (**Figure 3-5B-C**), an alternative possibility was that the nsp3.1 binding event could suppress ATF6 activation. To further test this possibility, cells were transfected with the respective nsp3.1 constructs and treated with the global UPR activator tunicamycin (Tm). We profiled global expression changes using quantitative proteomics to gain a broader insight into the regulation of UPR target genes based on larger genesets of ATF6-regulated targets, as well as targets of the other UPR branches (IRE1/XBP1s and PERK)(45). SARS-CoV-2 nsp3.1 resulted in a small, but significant downregulation of ATF6 targets compared to a control transfection (**Figure 3-5D**), which was consistent with the data obtained from quantitative Western blot. We observed robust activation of ATF6-target genes with Tm treatment (**Figure 3-5D, Figure 8-16A**), and also observed upregulation of IRE1/XBP1s and PERK target genes (**Figure 8-16B-C**). However, when nsp3.1 was expressed in cells, the Tm-induced expression changes of ATF6 target genes were significantly lower than with a mock control, confirming a suppression of the ATF6 response (**Figure 3-5D, Figure 8-16A**). When measured at the transcript level using qPCR, the same general trend was observed, though statistical significance was not reached (**Figure 8-16I**). In contrast, nsp3.1 had no significant effect on the expression of IRE1/XBP1s or PERK targets (**Figure 8-16B-C**), highlighting that the viral protein does not globally suppress the UPR response.

CoV infection typically activates the UPR to a smaller degree than global UPR stressors, such as tunicamycin or thapsigargin(42,48). For this reason, we set out to test whether nsp3.1 could more fully suppress the ATF6 response during such partial activation. We used compound **147**, a small molecule ER proteostasis regulator that preferentially activates the ATF6 branch of the UPR to mimic the incomplete UPR activation observed during CoV infection(49,50). We carried out the same global quantitative proteomics and geneset analysis to profile UPR target expression. ATF6 target genes were robustly activated when control-transfected cells were treated with **147** (**Figure 3-5E, Figure 8-16D**). However, as expected, the overall activation was reduced compared to Tm treatment, and no IRE1/XBP1s or PERK activation occurred with **147** (**Figure 8-16E-F**). In contrast, nsp3.1 expression completely blocked the upregulation of upregulation of ATF6 targets when cells were treated with **147**, further confirming the suppression of the ATF6 response (**Figure 3-5E, Figure 8-16D**). To ensure that the suppression was specific to ATF6 targets, we also analyzed the response of **147** treatment and SARS-CoV-2

nsp3.1 expression on other UPR branches and the cytosolic heat shock response (**Figure 8-16G-H**). We did not observe any change in expression in response to **147** or nsp3.1 further validating that the regulation is specific to ATF6.

3.4   Discussion

As evidenced by its many domains, nsp3 likely serves a multitude of roles within the coronavirus replication cycle. Some of these roles are well characterized, such as the requirement of nsp3 for formation of double-membraned vesicles and the papain-like protease function in autocleavage of the orf1a polypeptide(14–17,51,52). Other roles remain less defined, and the interactome of the individual nsp3 fragments contained herein may serve to help delineate the role of this protein across several CoV strains. Using tandem mass tags (TMTpro 16plex), we are able to directly identify and compare the abundance of interactors across five coronavirus strains, including two different genera and three different betacoronavirus clades. Given the frequent emergence of severely pathogenic coronavirus strains over the past 20 years, it is becoming increasingly apparent that a deeper knowledge of the molecular mechanisms of coronavirus replication is needed to understand how we may better prepare therapeutics against potential future strains of these viruses.

The N-terminal portion of the protein (nsp3.1) is the only fragment lacking transmembrane domains and exclusively localizes to the cytosol. This could explain the high expression levels seen with this fragment, although there was still some variation among the expression efficiency for the homologs from different viral strains. An interesting divergence observed in the nsp3.1 dataset is the enrichment of pathways related to mitochondrial transport and metabolism in the SARS-CoV dataset that is absent in the SARS-CoV-2 dataset. Given the two fragments share high sequence identity, it is noteworthy when one pathway is so highly enriched in one fragment versus the other, pointing to rapid divergent evolution. Other SARS-CoV-2 proteins have been shown to interact with several components of the mitochondria(18,19), including our earlier nsp2/4 dataset that found both SARS-CoV and SARS-CoV-2 proteins interacting with mitochondria-associated membrane factors involved in controlling calcium flux between ER and mitochondria(22). Interestingly, the most highly enriched pathways in both of the SARS variants are also enriched (to a similar magnitude) in the MERS-CoV nsp3.1 fragment pointing towards conserved functions among the pathogenic variants.

The C-terminal portion of the nsp3.1 fragment for SARS-CoV and SARS-CoV-2 consists of the SARS unique domain (SUD). While a definitive function of this domain has not been established in the coronavirus replication cycle, prior studies described a role for macrodomains within the SUD binding G-quadruplexes, strings of RNA containing multiple guanosines(51,53,54). While no RNA-binding pathways were observed as enriched with either of the SARS nsp3.1 fragments, potentially related processes such as mRNA catabolism, rRNA metabolism, and ribosome biogenesis were significantly enriched. This pathway was also observed in the MERS-CoV nsp3.1 interactor set, potentially pointing to a broader role for pathogenic strains. The singular pathway shown to be enriched only in the two SARS variants was "positive upregulation of transcription in response to ER stress", which prompted us to investigate the interactions and regulation of this fragment with the UPR sensor ATF6, which is discussed further below.

The middle fragment yielded the smallest number of overall interactors of the three fragments, across all five strains. Generally, expression levels of this fragment were also lower than the termini, which may have limited detection of interactors. However, all fragments and interactors were still reproducibly detectable by mass spectrometry. Surprisingly,

we observed highly distinct pathway enrichment from the MERS-CoV and hCoV-229E fragments showing substantial interactions with ribosomal RNA processing proteins and nuclear import proteins, respectively. Interactions of other coronavirus proteins with nuclear transport machinery have been observed, for instance nsp9 with the nuclear pore complex and nsp1 with mRNA export machinery(18,55). Orf6 of SARS-CoV and SARS-CoV-2 inhibits STAT1 signaling by blocking nuclear import of phosphorylated STAT1. Nuclear localization of SARS-CoV and SARS-CoV-2 orf3b has been shown, as well as for N proteins(56–58). In contrast, no strains studied to date have observed interactions of nuclear transport pathways with nsp3. One possibility is the recruitment of nuclear transport proteins to 229E nsp3 to assist in the transport of viral components from double-membraned vesicles, which nsp3 is essential in helping to create(14–16). A recent study showed that nsp3 from murine hepatitis C virus (MHV) not only assists in initial formation of the DMVs, but subsequently forms the core of a pore-like structure that may aid in export of viral RNA(36). The absence of these protein interactions between nsp3.1 homologs from betacoronavirus strains and nuclear importins points to evolution of functional divergence.

We also find that MERS nsp3.2 uniquely interacts with rRNA processing machinery involved in ribosome biogenesis. Several of these factors (WDR36, NOL10, URB1) were previously found to interact with SARS-CoV-2, SARS-CoV, and MERS-CoV nsp8 homologs(18,19). Some viruses, such as HIV-1, have also been shown to down-regulate factors involved in rRNA processing during infection(40). Given the dependence of all coronaviruses on host translation machinery, it is interesting that MERS is the only CoV to maintain these interactions. This may represent a specialized avenue for MERS-CoV to modulate host translation through nsp3.

Viruses often modulate cellular stress responses in order to assist during replication and/or subvert the host immune system. Many groups of viruses, including coronaviruses and flaviviruses, have been shown to upregulate branches of the UPR(42,43,46,59–64), although exact roles of this activation for the viral life cycle are still debated. We were therefore particularly interested to find ATF6 as a high confidence interactor of the SARS-CoV and SARS-CoV-2 nsp3.1 fragments. ATF6 is a transmembrane sensor representing one of the most upstream portions of the ER unfolded stress response (UPR). As noted, coronaviruses replicate near the ER, and nsp3 possesses transmembrane domains that anchor the protein in the ER membrane. Given the localization of nsp3.1 to the cytosol, the protein is likely to interact with the cytosolic basic leucine zipper (bZIP) transcriptional activator domain of ATF6. Therefore, we sought to determine if the fragment could be regulating the ability of ATF6 to respond to stress. We probed the transcriptional and translational level of known ATF6 regulated genes after overexpression of nsp3.1 fragments. Surprisingly, SARS-CoV-2 nsp3.1 led to a significant decrease in ATF6 protein markers, suggesting suppression of the ATF6 pathway. In addition to inhibiting basal ATF6 activity, we showed that SARS-CoV-2 nsp3.1 was also capable of partially suppressing ER-stress induced ATF6 activation, and fully blocked the more modest induction by the pharmacological activator **147**. It is likely that UPR activity has to be finely tuned during infection to prevent detrimental consequences from prolonged activation, such as apoptosis induction. Prior studies found that the IRE/XBP1s UPR branch was inhibited by the SARS-CoV E protein and that ATF6 gene targets were not upregulated during MHV infection despite ATF6 activation and cleavage(60,65). These results suggest that coronaviruses employ strategies to attenuate distinct UPR signaling branches, but little is known about what viral proteins are responsible for tuning UPR activity. Our results indicate that the N-terminal regions of SARS-CoV-2 nsp3 directly acts on ATF6 to suppress activation. Further work is needed to understand the molecular mechanisms of this suppression and its role during SARS-CoV-2 infection.

Many studies have highlighted the utility of viral protein interactomes in identifying roles for host pathways in the life cycle of many viruses(38,66), and large-scale interactome studies with SARS-CoV-2 demonstrated how these interactomes may be useful for identifying existing drugs to be repurposed to fight viral infections(18). Despite the large body of work done in the past year on SARS-CoV-2, interactome studies of nsp3 have been limited to the PL2[pro] papain-like-protease domain(11). In addition to presenting data for the full nsp3 interactome, our studies highlight the utility of using tandem mass tag technology to quantitatively compare the SARS-CoV-2 nsp3 interactome to the interactome of other known coronaviruses, both severely and mildly pathogenic. We find very divergent interactomes, suggesting that while there is a large conservation of domains, there may also be more specific roles for this protein in the context of each individual virus. One caveat is the use nsp3 fragments in our analysis, which may miss host interactions that require different region or domains of the protein to coordinate. Additionally, it would be intriguing to investigate the interactomes of nsp3 (fragmented or in full) with nsp4 and/or nsp6, as coordination between these viral proteins is required to form the double-membraned vesicles characteristic in CoV infection. In the future, it may be important to study the variations in host protein interactions that occur between specific SARS-CoV-2 variants that are rapidly emerging. For instance, the recently isolated B.1.1.7 (alpha) variant, shown to be highly transmissible and responsible for a large uptick in cases in the United Kingdom, possesses four mutations within orf1ab, three of which are in nsp3(67)·(68). Additionally, the B.1.351 (beta) variant contains one nsp3 mutation, and the P.1 (gamma) variant contains two(67). This makes nsp3 protein variants relevant to study for better understanding how such evolutionary adaptations in non-structural proteins may impact virulence.

3.5   Acknowledgements

3.6   Methods

Table 3-1. Primers used for qRT-PCR and generation of constructs

Target	Primer Sequence (5'-3')	Notes
Nsp3_N_term_rem_F	atcaaggtgttcaccacagtg	Forward primer to remove N-terminus of Wuhan nsp3 (nsp3.2, 3.3 generation)
Nsp3_N_term_rem_R	catgctagccagcttggg	Reverse primer to remove N-terminus of Wuhan nsp3 (nsp3.2 generation)
Nsp3_middle_rem_F	tattgtacaggcagcatccc	Forward primer to remove middle fragment of Wuhan nsp3 (nsp3.3 generation)
Nsp3_middle_rem_R	ggttctcacctcccgcag	Reverse primer to remove middle fragment of Wuhan nsp3 (nsp3.1 generation)
Nsp3_C_term_rem_F	ctcgagtctagagggccc	Forward primer to remove C-terminus of Wuhan nsp3 (nsp3.1, 3.2 generation)

Nsp3_C_term_rem_R	ggtggcgattgtcacgttg	Forward primer to remove C-terminus of Wuhan nsp3 (nsp3.2 generation)
SARS2_nsp3.1_xFT_F	atccgcagtttgaaaagtaaacccg ctgatcagcctcg	Removal of FLAG tag from Wuhan nsp3.1 fragment
SARS2_nsp3.1_xFT_R	catcccccgccgccttcgagggttc tcacctcccgcag	Removal of FLAG tag from Wuhan nsp3.1 fragment
2xStrep_F	ctcgaaggcggcggggga	Amplification of 2xStrep tag from pLVX vector
2xStrep_R	ttacttttcaaactgcggatgtgac catgatccac	Amplification of 2xStrep tag from pLVX vector
PDIA4_F	agtggggaggatgtcaatgc	
PDIA4_R	tggctgggatttgatgactg	
BiP_F	gcctgtatttctagacctgcc	
BiP_R	ttcatcttgccagccagttg	
ERDJ4_F	ggaaggaggagcgctaggtc	
ERDJ4_R	atcctgcaccctccgactac	
GAPDH_F	gtcggagtcaacggatt	
GAPDH_R	aagcttcccgttctcag	

Construct design

The coding sequences for full-length nsp3 were obtained from GenBank (SARS-CoV-2 isolate Wuhan-Hu-1 MN908947, SARS-CoV Urbani AY279741, hCoV-OC43 NC006213, MERS-CoV JX869059, hCoV-229E AF304460).

Sequences for creating fragments of the SARS-CoV nsp3 were chosen based on Pan et al, 2008, PLoS One(21). The amino acid sequences of full-length nsp3 for the remaining hCoV strains were aligned using ClustalOmega to the SARS-CoV fragments to determine the corresponding starting/ending positions for each fragment.

Full length SARS-CoV-2 nsp3 was codon optimized and cloned into a pcDNA-(+)-C-DYK vector (Genscript). Truncations were performed using primers listed in Table 1. All other nsp3 fragments were individually codon optimized and cloned into pTwist CMV Hygro vectors (Twist Biosciences).

SARS-CoV and SARS-CoV-2 nsp3.1 strep-tagged constructs were created using primers listed in Table 1. Briefly, nsp3 fragment plasmids were amplified to exclude the FLAG tag. A pLVX vector containing a strep tag was used to amplify the insert. The two fragments were ligated using a DNA HiFi assembly kit (NEB). All plasmid constructs were confirmed by sequencing (Genewiz).

Cell culture and transfection

HEK293T and HEK293T-REx cells were maintained in Dulbeccos' Modified Eagle's Medium (high glucose) and supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, and 1% glutamine. Cells were kept at 37°C, 5% CO_2. For transfections, 2E6 cells were seeded into 10cm tissue culture dishes. 24 hours after seeding, cells were transfected using a calcium phosphate method with 5 µg nsp3 or tdTomato construct. Media was exchanged 16 hours post-transfection and cells were harvested 24 hours after media exchange.

FLAG immunoprecipitations

Immunoprecipitations were performed as reported previously(22). Cells were collected from 10cm dishes via scraping, washed with PBS, and lysed by suspension in TNI buffer (50mM Tris pH 7.5, 150mM NaCl, 0.5% IGEPAL-CA-630) supplemented with Roche cOmplete protease inhibitor. Cells were left to lyse on ice for at least 10 minutes, followed by 10 minute sonication in a room temperature water bath. Lysates were cleared by centrifugation at 21.1xg for 20 minutes. Protein concentrations were normalized using 1x BioRad Protein Assay Dye, and normalized lysates were added to 15 µL pre-washed (4x in lysis buffer) Sepharose 4B beads (Sigma) and rocked at 4°C for 1 hour. Resin was collected by centrifugation at 400xg for 10 minutes, and pre-cleared supernatant was added to 15 µL G1 anti-DYKDDDDK resin (GenScript) and rocked at 4°C overnight. The next day, resin was collected by centrifugation at 400xg for 10 minutes. Resin was washed 4x with lysis buffer. Resin-bound proteins were eluted with the addition of modified 3x Laemelli buffer (6% SDS, 62.5mM Tris) for 30 minutes at room temperature, followed by 15 minutes at 37°C. A second elution was performed for 15 minutes at 37°C.

During the first IP for each construct, immunoprecipitation was confirmed by silver stain using a Pierce Silver Stain Kit (Thermo Fisher) and by western blotting with anti-FLAG antibody (Sigma-Aldrich, F1804)

Immunofluorescence confocal microscopy

HEK293T cells were transfected with nsp3 constructs as described in "Cell culture and transfection". Two hours post-media change, $2x10^4$ transfected cells were seeded into glass-bottom culture dishes (MatTek, P35G-0-14-C). At 40 hours post-transfection, cells were fixed with 4% paraformaldehyde-PBS, washed with PBS three times, then permeabilized in 0.2% Triton-X (in PBS). After three PBS washes, cells were blocked in PBS with 1% BSA and 0.1% Saponin (blocking buffer) for 1 hour at room temperature. After blocking, cells were incubated with anti-PDIA4 primary antibody (Protein Tech, 14712-1-AP) in blocking buffer (1:1000 dilution) overnight at 4°C. After three PBS washes, cells were incubated with AlexFluor 488-conjugated anti-rabbit goat antibody (ThermoFisher, A-11008) in blocking buffer (1:500 dilution) at room temperature for 30 min. Cells were then stained with M2 FLAG primary antibody (SigmaAldrich, F1804) and AlexFluor 594-conjugated anti-mouse goat antibody (ThermoFisher, A-11005) using the same conditions. Cells were then mounted in Prolong Gold with DAPI stain (ThermoFisher, P36935) overnight. Cells were imaged using an LSM-880 or LSM-710 confocal microscope (Zeiss) and images were merged using Image J software.

Nsp3.1-ST & ATF6-FT Co-immunoprecipitation

HEK293T-REx cells expressing doxycycline-inducible 6xHis-3xFLAG-HsATF6α were seeded in 15cm tissue culture dishes(69). Cells were transiently transfected with 15 µg SARS-CoV-2 nsp3.1-ST or tdTomato constructs 24 hours after seeding as previously described. At 16 hours post-transfection media was exchanged and cells were treated with either DMSO or 100 µM doxycycline. Cells were harvested 24 hours later via scraping in cold PBS with 10 µM MG132. Cells were lysed in ATF6 lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 5 mM EDTA, 1x protease inhibitor, and 1% LMNG (Anatrace NG322) for 15 min at 4°C as previously described(70). Lysates were vortexed twice briefly and then cleared by centrifugation at 17,000 x g for 15 min. Lysates were normalized and pre-cleared with 15 µL Sepharose 4B beads in ATF6 wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 5 mM EDTA). Pre-cleared supernatant was then divided in two, with half

immunoprecipitated for FLAG-ATF6 with 15 µL G1 anti-DYKDDDK resin and the other half affinity purified for nsp3.1-ST using 15 µL Strep-Tactin XT Superflow High Capacity resin (IBA Life Sciences, 2-4030-002) and rocked at 4°C overnight. Resin was washed 4x with ATF6 wash buffer. Resin-bound proteins were eluted with modified 3x Laemelli buffer (6% SDS, 62.5mM Tris) as described previously. Inputs and elutions were then diluted with 6x Laemelli buffer (12% SDS, 125 mM Tris pH 6.8, 20% glycerol, bromophenol blue, 100 mM DTT), heated at 37°C for 30 min, and run on SDS-PAGE gel. Proteins were detected via western blotting with M2 anti-FLAG (Sigma-Aldrich, F1804) and THE anti-Strep II tag FITC (Genescript, A01736-100) antibodies.

Western blot and RT-qPCR analysis of UPR activation

HEK293T cells were transfected with nsp3.1-FT homologs or tdTomato (mock) in 6-well plates as previously described. Mock samples were treated with DMSO or 6 µg/mL Tunicamycin (Tm) for 16 h (protein analysis) or 6 h (transcript analysis) prior to harvest by cell scraping. Harvested cells were split into a 2/3 aliquot for protein analysis and a 1/3 aliquot for mRNA analysis.

For protein analysis, cells were lysed as previously described for immunoprecipitations. Lysates were normalized using 1x BioRad Protein Assay Dye and run on SDS-PAGE gel before transfer to PVDF membranes and subsequent western blotting. Blots were probed with 1:1000 dilutions of anti-KDEL (Enzo, ADI-SPA-827-F), anti-PDIA4 (ProteinTech, 14712-1-AP), M2 anti-FLAG (Sigma-Aldrich, F1804), and anti-GAPDH (GeneTex, GTX627408) antibodies in 5% BSA.

For RT-qPCR, cellular RNA was extracted using the Zymo Quick-RNA miniprep kit. Then 500 ng total cellular RNA was synthesized into cDNA using random hexamer primers (IDT), oligo-dT primers (IDT), and Promega M-MLV reverse transcriptase. Subsequent qPCR analysis was carried out using BioRad iTaq Universal SYBR Green Supermix, combined with the primers listed below for target genes. Reactions were run in 96-well plates on a BioRad CFX qPCR instrument. Conditions used for amplification were 95ºC, 2 minutes, 45 repeats of 95ºC, 10s and 60ºC, 30s. A melting curve was generated in 0.5 ºC intervals from 65 ºC to 95 ºC. Cq values were calculated by the BioRad CFX Maestro software. Transcripts were normalized to a housekeeping gene (GAPDH). All measurements were performed in technical duplicate; each of these duplicates was treated as a single measurement for the final average. Data was analyzed using the BioRad CFX Maestro software.

Global proteomics of UPR activation

HEK293T cells were transfected with SARS-CoV-2 nsp3.1-FT or tdTomato (mock) in 10 cm dishes in triplicate as previously described. Samples were treated with either DMSO, 10 µM **147**, or 6 µg/mL tunicamycin for 16 h prior to harvest by cell scraping. Harvested cells were lysed in TNI lysis buffer and 20 µg of protein (as measured using Bio-Rad protein assay dye reagent) was aliquoted and precipitated via methanol/chloroform precipitation as described below. Pellets were digested with trypsin (Thermo Fisher) and labeled with 16plex TMTpro.

Tandem Mass Tag sample preparation

Sample preparation was carried out as previously described(22). Eluted proteins were precipitated via methanol/chloroform water (3:1:3) and washed thrice with methanol. Each wash was followed by a 5 minute spin at 10,000xg. Protein pellets

were air dried and resuspended in 1% Rapigest SF (Waters). Resuspended proteins were reduced (TCEP) for 30 minutes, alkylated (iodoacetamide) for 30 minutes, and digested with 0.5 µg trypsin/Lys-C (Thermo Fisher) overnight. Digested peptides were labeled using 16plex TMTpro (Thermo Scientific) and quenched with the addition of ammonium bicarbonate. Each TMT channel shown in Figure 8-3**A-F** represents an independent immunoprecipitation and biological replicate. Samples were pooled, acidified, and concentrated. Cleaved Rapigest products were removed by centrifugation at 17,000xg for 45 minutes.

MudPIT LC-MS/MS analysis

Triphasic MudPIT columns were prepared as previously described using alternating layers of 1.5cm C18 resin, 1.5cm SCX resin, and 1.5cm C18 resin(71). Pooled TMT samples (roughly one-third of pooled IP samples and 20 µg of peptide from global UPR activation samples) were loaded onto the microcapillaries using a high-pressure chamber, followed by a 30 minute wash in buffer A (95% water, 5% acetonitrile, 0.1% formic acid). Peptides were fractionated online by liquid chromatography using an Ultimate 3000 nanoLC system and subsequently analyzed using an Exploris480 mass spectrometer (Thermo Fisher). The MudPIT columns were installed on the LC column switching valve and followed by a 20cm fused silica microcapillary column filled with Aqua C18, 3µm, C18 resin (Phenomenex) ending in a laser-pulled tip. Prior to use, columns were washed in the same way as the MudPIT capillaries. MudPIT runs were carried out by 10µL sequential injections of 0, 10, 20, 40, 60, 80, 100 % buffer C (500mM ammonium acetate, 94.9% water, 5% acetonitrile, 0.1% formic acid) for IP samples and 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% buffer C for global UPR activation samples, followed by a final injection of 90% C, 10% buffer B (99.9% acetonitrile, 0.1% formic acid v/v). Each injection was followed by a 130 min gradient using a flow rate of 500nL/min (0-6 min: 2% buffer B, 8 min: 5% B, 100 min: 35% B, 105min: 65% B, 106-113 min: 85% B, 113-130 min: 2% B). ESI was performed directly from the tip of the microcapillary column using a spray voltage of 2.2 kV, an ion transfer tube temperature of 275°C and a RF Lens of 40%. MS1 spectra were collected using a scan range of 400-1600 m/z, 120k resolution, AGC target of 300%, and automatic injection times. Data-dependent MS2 spectra were obtained using a monoisotopic peak selection mode: peptide, including charge state 2-7, TopSpeed method (3s cycle time), isolation window 0.4 m/z, HCD fragmentation using a normalized collision energy of 32 (TMTpro), resolution 45k, AGC target of 200%, automatic injection times, and a dynamic exclusion (20 ppm window) set to 60s.

Data analysis

Identification and quantification of peptides were performed in Proteome Discoverer 2.4 (Thermo Fisher) using the SwissProt human database (TaxID 9606, released 11/23/2019; 42,252 entries searched) with nsp3 fragment sequences (15 entries) manually added (42,267 total entries searched). Searches were conducted with Sequest HT using the following parameters: trypsin cleavage (maximum 2 missed cleavages), minimum peptide length 6 AAs, precursor mass tolerance 20ppm, fragment mass tolerance 0.02Da, dynamic modifications of Met oxidation (+15.995Da), protein N-terminal Met loss (-131.040Da), and protein N-terminal acetylation (+42.011Da), static modifications of TMTpro (+304.207Da) at Lys and N-termini and Cys carbamidomethylation (+57.021Da). Peptide IDs were filtered using Percolator with an FDR target of 0.01. Proteins were filtered based on a 0.01 FDR, and protein groups were created according to a strict parsimony

principle. TMT reporter ions were quantified considering unique and razor peptides, excluding peptides with co-isolation interference greater that 25%. Peptide abundances were normalized based on total peptide amounts in each channel, assuming similar levels of background in the IPs. For global UPR proteomics, protein abundances were also scaled. Protein quantification used all quantified peptides. Post-search filtering was done to include only proteins with two identified peptides.

Experimental Design and Statistical Rationale

Nsp3.1 analysis combined 62 co-IP samples distributed across 5 individual MS runs (biological replicates: 15x tdTomato (negative co-IP control), 13x SARS-CoV-2, 9x SARS-CoV, 8x MERS-CoV, 9x hCoV-OC43, 8x hCoV-229E). Nsp3.2 analysis combined 63 co-IP samples distributed across 4 individual MS runs (biological replicates: 15x tdTomato (negative co-IP control), 14x SARS-CoV-2, 9x SARS-CoV, 9x hCoV-OC43, 8x MERS-CoV, 8x hCoV-229E). Nsp3.3 analysis combined 63 co-IP samples distributed across 5 individual MS runs (biological replicates: 15x tdTomato (negative co-IP control), 14x SARS-CoV-2, 9x SARS-CoV, 8x MERS-CoV, 9x hCoV-OC43, 8x hCoV-229E). Pairwise ratios between conditions were calculated in Proteome Discoverer based on total protein abundance, and ANOVA was performed on individual proteins to test for change in abundances and report adjusted P-values. To filter interactors of individual nsp3 fragments, we used a variable cutoff method combining log2 enrichment and adjusted p-value according to a published method (22,72). The histogram of log2 protein abundance fold changes for each construct vs the tdTomato control was fitted to a gaussian curve with a bin width of 0.1 using a nonlinear least square fit (excluding outliers) to determine the standard deviation σ of the scatter. For medium- and high-confidence interactors, the cutoff values were 1σ and 2σ respectively. To take into consideration the adjusted p-values, we used a hyperbolic curve $y > c/(x-x_0)$ where y is the adjusted p-value, x is the log2 fold change, and x_0 corresponds to the value of the 1σ or 2σ standard deviation. Negative medium- and high-confidence interactors were filtered using cutoff values of -1σ and -2σ respectively and the hyperbolic curve $y > - c/(x+x_0)$.

CRAPome overlap, geneset enrichment analysis, comparative heatmaps, and network plots

High-confidence interactors of all homologs for each nsp3 fragment were queried in the CRAPome "*Homo sapiens* Single Step Epitope tag AP-MS" dataset(28). A frequency threshold of 25% of control experiments found in was used to qualify an interactor as overlapping with the CRAPome. A gene ontology (GO) enrichment analysis for biological processes was conducted in EnrichR. The analysis was conducted separately for the high (and/or medium) confidence interactors for each fragment. GO terms were manually filtered for adjusted p-values <0.1. Redundant GO terms were grouped manually based on overlapping genes in related pathways.

Network plots were generated in Cytoscape(73); human protein interactions were validated based on the STRING database.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD024566. All other necessary data are contained within the manuscript.

3.7   References

1. Ksiazek, T. G. *et al.* A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **348**, 1953–1966 (2003).
2. Drosten, C. *et al.* Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
3. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
4. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
5. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
6. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).
7. Fehr, A. R. & Perlman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. *Methods Mol. Biol.* **1282**, 1 (2015).
8. Chan, Y. K. & Gack, M. U. A phosphomimetic-based mechanism of dengue virus to antagonize innate immunity. *Nat. Immunol.* **17**, 523–530 (2016).
9. Hou, Z. *et al.* Hepatitis B virus inhibits intrinsic RIG-I and RIG-G immune signaling via inducing miR146a. *Sci. Rep.* **6**, 1–12 (2016).
10. Shi, C.-S. *et al.* SARS-Coronavirus Open Reading Frame-9b Suppresses Innate Immunity by Targeting Mitochondria and the MAVS/TRAF3/TRAF6 Signalosome. *J. Immunol.* **193**, 3080–3089 (2014).
11. Shin, D. *et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662 (2020).
12. Clementz, M. A. *et al.* Deubiquitinating and Interferon Antagonism Activities of Coronavirus Papain-Like Proteases. *J. Virol.* **84**, 4619–4629 (2010).
13. Báez-Santos, Y. M., St. John, S. E. & Mesecar, A. D. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antiviral Res.* **115**, 21–38 (2015).
14. Angelini, M. M., Akhlaghpour, M., Neuman, B. W. & Buchmeier, M. J. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio* **4**, (2013).
15. Oostra, M. *et al.* Topology and Membrane Anchoring of the Coronavirus Replication Complex: Not All Hydrophobic Domains of nsp3 and nsp6 Are Membrane Spanning †. *J. Virol.* **82**, 12392–12405 (2008).
16. Oudshoorn, D. *et al.* Expression and cleavage of middle east respiratory syndrome coronavirus nsp3-4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication. *MBio* **8**, (2017).
17. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res.* **149**, 58–74 (2018).
18. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
19. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science (80-.).* **370**, eabe9403 (2020).
20. Pan, J. *et al.* Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. *PLoS One* **3**, (2008).
21. von Brunn, A. *et al.* Analysis of Intraviral Protein-Protein Interactions of the SARS Coronavirus ORFeome. *PLoS One* **2**, e459 (2007).
22. Davies, J. P., Almasy, K. M., McDonald, E. F. & Plate, L. Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *ACS Infect. Dis.* **6**, 3174–3189 (2020).
23. Li, J. *et al.* TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat. Methods* **17**, 399–404 (2020).
24. Mock, J. Y. *et al.* Bag6 complex contains a minimal tail-anchor-targeting module and a mock BAG domain. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 106–111 (2015).
25. Shi, Y., Mosser, D. D. & Morimoto, R. I. Molecular Chaperones as HSF-1 specific transcriptional repressors. *Genes Dev* **12**, 654–666 (1998).
26. Nakatani, Y. *et al.* p600, a unique protein required for membrane morphogenesis and cell survival. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15093–15098 (2005).
27. Bjørkøy, G. *et al.* p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *J. Cell Biol.* **171**, 603–614 (2005).

28. Mellacheruvu, D. *et al.* The CRAPome: A contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).

29. Samavarchi-Tehrani, P. *et al.* A SARS-CoV-2 – host proximity interactome. *bioRxiv* 2020.09.03.282103 (2020) doi:10.1101/2020.09.03.282103.

30. Koyano-Nakagawa, N., Nishida, J., Baldwin, D., Arai, K. & Yokota, T. Molecular cloning of a novel human cDNA encoding a zinc finger protein that binds to the interleukin-3 promoter. *Mol. Cell. Biol.* **14**, 5099–5107 (1994).

31. Jadhav, K. & Zhang, Y. Activating transcription factor 3 in immune response and metabolic regulation. *Liver Research* vol. 1 96–102 at https://doi.org/10.1016/j.livres.2017.08.001 (2017).

32. Walter, P. & Ron, D. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science (80-.).* **334**, 1081–1086 (2011).

33. Baez-Santos, Y. M., Mielech, A. M., Deng, X., Baker, S. & Mesecar, A. D. Catalytic Function and Substrate Specificity of the Papain-Like Protease Domain of nsp3 from the Middle East Respiratory Syndrome Coronavirus. *J. Virol.* **88**, 12511–12527 (2014).

34. Barretto, N. *et al.* The Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus Has Deubiquitinating Activity. *J. Virol.* **79**, 15189–15198 (2005).

35. Mielech, A. M., Kilianski, A., Baez-Santos, Y. M., Mesecar, A. D. & Baker, S. C. MERS-CoV papain-like protease has deISGylating and deubiquitinating activities. *Virology* **450**–**451**, 64–70 (2014).

36. Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science (80-.).* **369**, 1395–1398 (2020).

37. Wright, F. A., Bonzerato, C. G., Sliter, D. A. & Wojcikiewicz, R. J. H. The erlin2 T65I mutation inhibits erlin1/2 complex-mediated inositol 1,4,5-trisphosphate receptor ubiquitination and phosphatidylinositol 3-phosphate binding. *J. Biol. Chem.* **293**, 15706–15714 (2018).

38. Wang, R. *et al.* Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses. *Cell* **184**, 1–14 (2020).

39. Bianco, C. & Mohr, I. Ribosome biogenesis restricts innate immune responses to virus infection and DNA. *Elife* **8**, (2019).

40. Kleinman, C. L. *et al.* HIV-1 Infection Causes a Down-Regulation of Genes Involved in Ribosome Biogenesis. *PLoS One* **9**, e113908 (2014).

41. Reggiori, F. *et al.* Coronaviruses hijack the LC3-I-positive EDEMosomes, ER-derived vesicles exporting short-lived ERAD regulators, for replication. *Cell Host Microbe* **7**, 500–508 (2010).

42. Chan, C.-P. *et al.* Modulation of the Unfolded Protein Response by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. *J. Virol.* **80**, 9279–9287 (2006).

43. Versteeg, G. A., van de Nes, P. S., Bredenbeek, P. J. & Spaan, W. J. M. The Coronavirus Spike Protein Induces Endoplasmic Reticulum Stress and Upregulation of Intracellular Chemokine mRNA Concentrations. *J. Virol.* **81**, 10981–10990 (2007).

44. Shoulders, M. D. *et al.* Stress-Independent Activation of XBP1s and/or ATF6 Reveals Three Functionally Diverse ER Proteostasis Environments. *Cell Rep.* **3**, 1279–1292 (2013).

45. Grandjean, J. M. D. *et al.* Deconvoluting Stress-Responsive Proteostasis Signaling Pathways for Pharmacologic Activation Using Targeted RNA Sequencing. *ACS Chem. Biol.* **14**, 784–795 (2019).

46. Sung, S. C., Chao, C. Y., Jeng, K. S., Yang, J. Y. & Lai, M. M. C. The 8ab protein of SARS-CoV is a luminal ER membrane-associated protein and induces the activation of ATF6. *Virology* **387**, 402–413 (2009).

47. Gallagher, C. M. *et al.* Ceapins are a new class of unfolded protein response inhibitors, selectively targeting the ATF6α branch. *Elife* **5**, 1–33 (2016).

48. Jiang, X. S. *et al.* Quantitative analysis of Severe Acute Respiratory Syndrome (SARS)-associated coronavirus-infected cells using proteomic approaches: Implications for cellular responses to virus infection. *Mol. Cell. Proteomics* **4**, 902–913 (2005).

49. Plate, L. *et al.* Small molecule proteostasis regulators that reprogram the ER to reduce extracellular protein aggregation. *Elife* **5**, 1–26 (2016).

50. Paxman, R. *et al.* Pharmacologic ATF6 activating compounds are metabolically activated to selectively modify endoplasmic reticulum proteins. *Elife* **7**, 1–23 (2018).

51. Thiel, V. *et al.* Mechanisms and enzymes involved in SARS coronavirus genome expression. *Journal of General Virology* vol. 84 2305–2315 at https://doi.org/10.1099/vir.0.19424-0 (2003).

52. Harcourt, B. H. *et al.* Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity. *J. Virol.* **78**, 13600–13612 (2004).

53. Chang, Y.-S. *et al.* SARS Unique Domain (SUD) of Severe Acute Respiratory Syndrome Coronavirus Induces NLRP3 Inflammasome-Dependent CXCL10-Mediated Pulmonary Inflammation. *Int. J. Mol. Sci.* **21**, 3179 (2020).

54. Tan, J. *et al.* The SARS-Unique Domain (SUD) of SARS Coronavirus Contains Two Macrodomains That Bind G-Quadruplexes. *PLoS Pathog.* **5**, e1000428 (2009).

55. Zhang, K. *et al.* Nsp1 protein of SARS-CoV-2 disrupts the mRNA export machinery to inhibit host gene expression. *Sci. Adv.* **7**, (2021).

56. Kopecky-Bromberg, S. A., Martínez-Sobrido, L., Frieman, M., Baric, R. A. & Palese, P. Severe Acute Respiratory Syndrome Coronavirus Open Reading Frame (ORF) 3b, ORF 6, and Nucleocapsid Proteins Function as Interferon Antagonists. *J. Virol.* **81**, 548–557 (2007).

57. Frieman, M. *et al.* Severe Acute Respiratory Syndrome Coronavirus ORF6 Antagonizes STAT1 Function by Sequestering Nuclear Import Factors on the Rough Endoplasmic Reticulum/Golgi Membrane. *J. Virol.* **81**, 9812–9824 (2007).

58. Miorin, L. *et al.* SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28344–28354 (2020).

59. Peña, J. & Harris, E. Dengue virus modulates the unfolded protein response in a time-dependent manner. *J. Biol. Chem.* **286**, 14226–14236 (2011).

60. Bechill, J., Chen, Z., Brewer, J. W. & Baker, S. C. Coronavirus Infection Modulates the Unfolded Protein Response and Mediates Sustained Translational Repression. *J. Virol.* **82**, 4492–4501 (2008).

61. Fung, T. S. & Liu, D. X. Coronavirus infection, ER stress, apoptosis and innate immunity. *Front. Microbiol.* **5**, (2014).

62. Minakshi, R. *et al.* The SARS Coronavirus 3a Protein Causes Endoplasmic Reticulum Stress and Induces Ligand-Independent Downregulation of the Type 1 Interferon Receptor. *PLoS One* **4**, e8342 (2009).

63. Carletti, T. *et al.* Viral priming of cell intrinsic innate antiviral signaling by the unfolded protein response. *Nat. Commun.* **10**, 1–9 (2019).

64. Ke, P. Y. & Chen, S. S. L. Activation of the unfolded protein response and autophagy after hepatitis C virus infection suppresses innate antiviral immunity in vitro. *J. Clin. Invest.* **121**, 37–56 (2011).

65. DeDiego, M. L. *et al.* Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Regulates Cell Stress Response and Apoptosis. *PLoS Pathog.* **7**, e1002315 (2011).

66. Marceau, C. D. *et al.* Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* **535**, 159–163 (2016).

67. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).

68. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations.

69. Gallagher, C. M. & Walter, P. Ceapins inhibit ATF6α signaling by selectively preventing transport of ATF6α to the Golgi apparatus during ER stress. *Elife* **5**, (2016).

70. Torres, S. E. *et al.* Ceapins block the unfolded protein response sensor atf6α by inducing a neomorphic inter-organelle tether. *Elife* **8**, (2019).

71. Fonslow, B. R. *et al.* Single-Step Inline Hydroxyapatite Enrichment Facilitates Identification and Quantitation of Phosphopeptides from Mass-Limited Proteomes with MudPIT. *J. Proteome Res.* **11**, 2697–2709 (2012).

72. Wright, M. T., Kouba, L. & Plate, L. Thyroglobulin Interactome Profiling Defines Altered Proteostasis Topology Associated With Thyroid Dyshormonogenesis. *Mol. Cell. Proteomics* **20**, 100008 (2021).

73. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

74. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Research* at https://doi.org/10.1016/j.antiviral.2017.11.001 (2018).

CHAPTER 4

THE GLYCOPROTEIN QUALITY CONTROL FACTOR MALECTIN PROMOTES CORONAVIRUS REPLICATION AND VIRAL PROTEIN BIOGENESIS

4.1 Summary
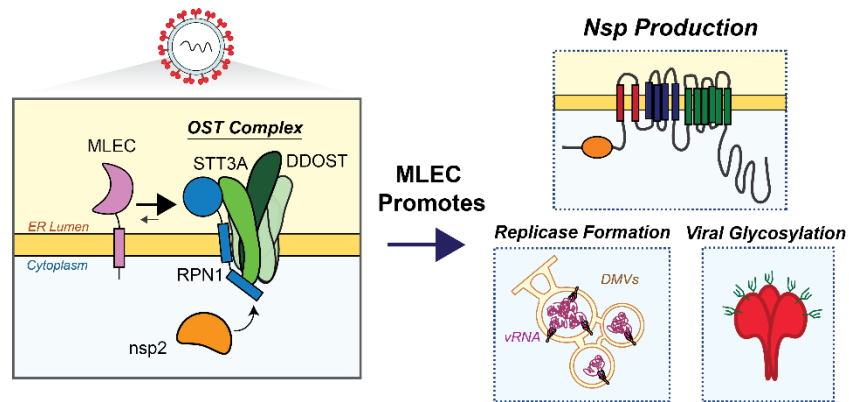
Coronavirus (CoV) modulate host protein production and protein homeostasis (proteostasis) networks to promote infection. Much of this is mediated by the nonstructural proteins (nsps), such as nsp2 and nsp4, which rewire host processes through protein-protein interactions (PPIs). With the emergence of SARS-CoV-2, it is imperative to identify host PPIs that are shared across CoV nsp homologs and understand how these interactions promote infection. Using quantitative proteomics and functional genetic screening, I identify conserved host proteostasis interactors of nsp2 and nsp4 which serve pro-viral roles in infection of a model betacoronavirus, MHV. Of note is the glycoprotein quality control factor, Malectin (MLEC), the knockdown of which leads to a reduction of infectious titers by two $\log_{10}$ PFU/mL. I find that during infection, nsp2 and MLEC maintain a genetic interaction and that nsp2 interacts with MLEC-associated protein complexes. MHV infection also drastically modifies the MLEC-interactome, stabilizing association with the Oligosaccharyltransferase (OST) complex while titrating off basal interactors. Furthermore, I find MLEC promotes early viral replication events, namely nsp production, with knockdown of MLEC leading to defects in viral protein abundance and genome replication. This pro-viral role is dependent on the glycoprotein quality control role of MLEC. Together, these results reveal a new role for the conserved interactor MLEC in promoting CoV infection by enhancing viral glycoprotein production and presenting a potential therapeutic target for pan-CoV antivirals.

All supplemental materials can be found in Appendix 3.

4.2 Introduction

Several highly pathogenic human coronaviruses (CoVs) have emerged in the past few decades and imposed drastic public health burdens, including the betacoronaviruses SARS-CoV, MERS-CoV, and most recently SARS-CoV-2. Given

the pattern of past CoV emergence and the likelihood of future outbreaks, it is imperative to understand the molecular strategies shared by CoVs to infect human cells. CoVs hijack host cellular systems through virus-host protein-protein interactions (PPIs) to mediate infection. In particular, the sixteen nonstructural proteins (nsps) encoded by CoVs alter the cellular environment to enhance infection and create viral replication/transcription complexes (RTCs) to reproduce viral genomes.

Nsps have varying degrees of sequence conservation, topology, and ascribed functions. This work focuses on two contrasting nsps, nsp2 and nsp4. Nsp2 is a soluble protein with an ambiguous role in viral infection and low sequence conservation amongst CoV homologs. Previous work has found that when nsp2 is deleted from the Murine Hepatitis Virus (MHV) and SARS-CoV genome, infectious virus is still produced, albeit with significant stunting in titers (~two $\log_{10}$ PFU/mL)(1). Nsp2 has also been shown to associate with other nsps amid the viral RTC(2–4), indicating that while dispensable, nsp2 promotes viral infection and likely works in coordination with other viral proteins. In contrast, nsp4 is a multipass transmembrane glycoprotein with high sequence conservation amongst CoV homologs(5–7). In conjunction with nsp3 and nsp6, nsp4 plays an indispensable role in host membrane manipulation to form double membrane vesicles (DMVs), which serve as the primary sites of viral genome replication (8–11).

One of the primary cellular systems hijacked by CoVs during infection is the host proteostasis network. The proteostasis network maintains the cellular proteome, from protein folding, addition of post-translational modifications, and trafficking to degrading old or misfolded proteins (12). Key to this network is the coordination of PPIs between quality control factors and client proteins to surveil the state of proteome, which is particularly important in the biogenesis and secretion of glycosylated proteins (glycoproteins)(13–16).

Previous work has harnessed quantitative proteomics to characterize interactions between nsp2 and the host proteostasis system. V'kovski et al. applied proximity-labeling using a BioID-nsp2 encoding MHV strain to identify host factors proximal to the RTC, including autophagy and translation initiation factors(17). Nsp-host interactors of have also been identified using affinity-purification mass spectrometry (AP-MS) in multiple studies(18–23). These studies have consistently identified the host translation repressor complex of 4EHP/GIGYF2 as an interactor of SARS-CoV and SARS-CoV-2 nsp2 homologs, and subsequent work has characterized the structural determinants of this interaction(24–26). There is conflicting evidence on the function of this interaction, as some studies have found that nsp2 interferes with the translation repression activity of the 4EHP/GIGYF2 complex(24), while others report that nsp2 enhances the translation repression, particularly against IFN-β production(25). MERS-CoV nsp2 interacts with components of the Ribosome Quality Control (RQC) pathway(22), including the ASC-1 complex(27) and TCF25(28), in addition to a component of *O*-linked oligosaccharide biosynthesis (GALNT2)(29). The interface between nsp2 and glycoprotein biogenesis pathways has remained unexplored.

Interactions between nsp4 and endoplasmic reticulum (ER) proteostasis factors have also been examined, particularly in the context of DMV formation(30). A variety of ER interactors have been shown to promote formation of DMVs from ER-derived membranes during SARS-CoV-2 infection, including RTN3/4(31), VMP1, and TMEM41B(32). SARS-CoV-2 nsp4 has also been found to activate the Unfolded Protein Response (UPR), a cellular stress response which mitigates accumulation of misfolded proteins in part through ER expansion and ER chaperone upregulation(33).

Previously, we performed a comparative interactomics study of nsp2 and nsp4 homologs from SARS-CoV-2, SARS-CoV, and hCoV-OC43(18). We identified several conserved proteostasis interactors amongst SARS-CoV-2 and SARS-CoV nsp2 homologs, including the chaperones HSPA5 and HSPA8, the glycoprotein quality control factor MLEC(34,35), and the ubiquitination complex ERLIN1/2/RNF170(36,37). We also found conserved nsp4-proteostasis interactors, such as components of the *N*-linked glycosylation machinery (DDOST, STT3B, MAGT1, CANX), the ER Membrane Complex (MMGT1)(38,39), ER-cargo receptors (CCPG1(40), SURF4(41)), and protein degradation factors (TMEM33, TMEM43, HERPUD1, ERLIN1/2). In this work, I expand our interactomics to a wider panel of nsp2 and nsp4 homologs to include MERS-CoV, hCoV-229E, and MHV. I evaluate the functional relevance of PPIs and find that many conserved ER proteostasis interactors serve pro-viral roles in a CoV infection model, most notably the glycoprotein quality control factor Malectin (MLEC). I interrogate the nature of the interaction between nsp2 and MLEC by combining quantitative proteomics with classical virological assays and find that MLEC promotes viral nsp production during infection. This work identifies a new virus-host dependency with potential for pan-CoV therapeutics and furthers our understanding of the basic biology of CoV infection.

4.3   Results

4.3.1   Nsp2 and nsp4 ER proteostasis interactors are conserved across CoV homologs

We previously mapped the comparative interactome of nsp2 (SARS-CoV-2 and SARS-CoV) and nsp4 (SARS-CoV-2, SARS-CoV, hCoV-OC43) homologs(18). To determine the scope of conservation of previously identified interactors, I expanded our panel of nsp homologs to include the alphacoronavirus hCoV-229E and the betacoronaviruses MERS-CoV and Murine Hepatitis Virus (MHV) (**Figure 4-1A**). hCoV-OC43 homologs were omitted due to poor expression. The MHV homologs were included to determine the relevance of using MHV as a BSL-2 infection model for testing the functionality of interactors in active infection. I utilized our previously reported nsp construct design and affinity-purification mass spectrometry (AP-MS) workflow(18,19), in which FLAG-tagged nsp homolog constructs were transiently transfected into HEK293T cells and cells harvested 40 hours post-transfection (Figure 4-1**B**). Cells were lysed and bait proteins immunopurified for FLAG-tag along with accompanying host interactors (Figure 9-1**,** Figure 9-2) before being reduced, alkylated, and trypsin/LysC digested. Peptides were subsequently labeled with TMTpro tags and analyzed by tandem mass spectrometry (LC-MS/MS), enabling both protein identification and relative quantification (**Supplemental Tables S1, S2,** Figure 9-1**,** Figure 9-2)

I cross referenced our previous comparative interactomics dataset with this dataset and identified several proteins of interest that were conserved across most, if not all, homologs analyzed (Figure 9-1**,** Figure 9-2). Several of these conserved interactors are ER proteostasis factors, such as MLEC (nsp2-interactor), CCPG1, and SURF4 (nsp4-interactors), and others are mitochondria-associated proteins, including STOML2, PHB2 (nsp2-interactors) and TIMM21 (nsp4-interactor) (Figure 4-1**C**). These interactors are conserved with MHV homologs to various degrees, enabling us to utilize MHV as an accessible model BSL-2 virus to evaluate the relevance of interactors for active CoV infection.
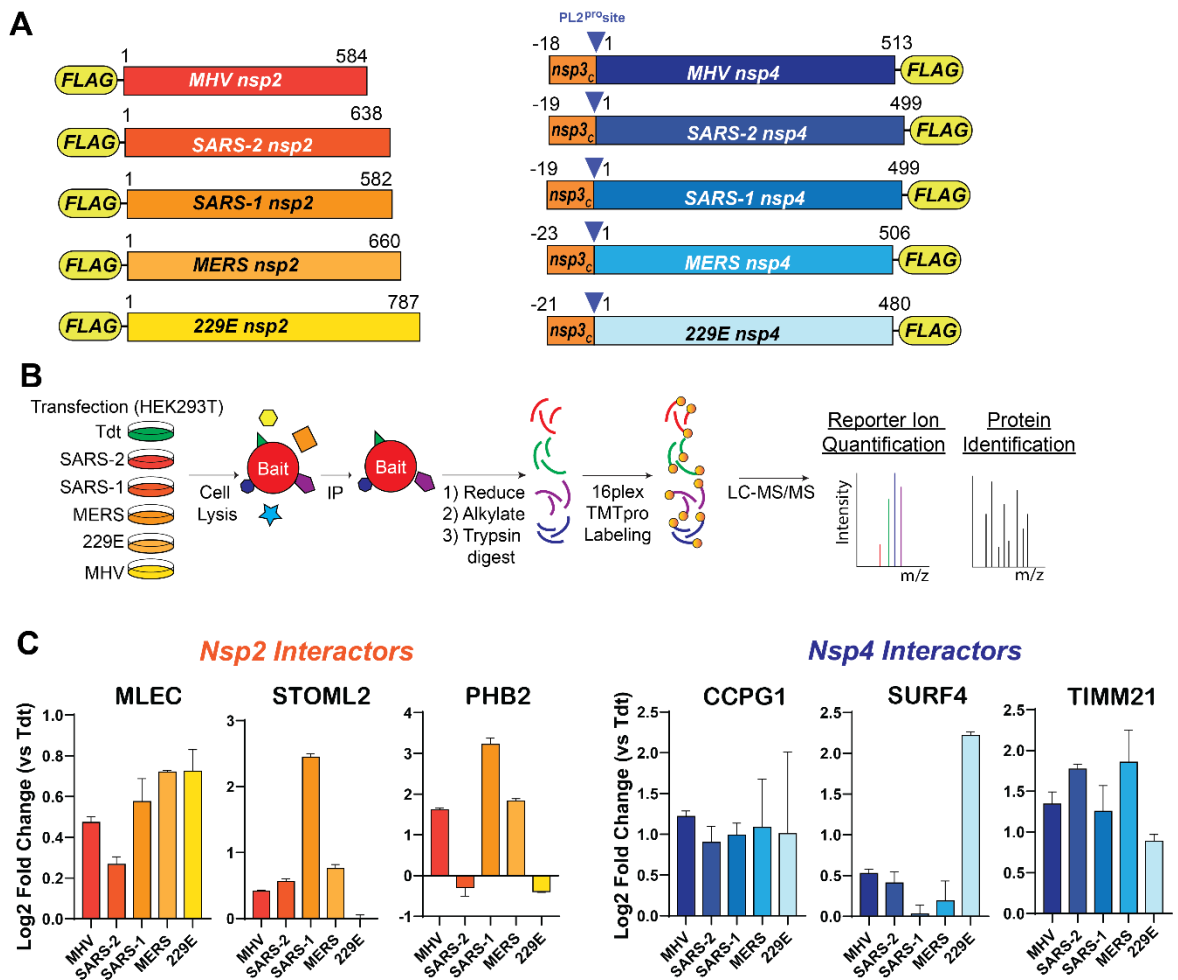
Figure 4-1. Nsp2 and nsp4 interactors are conserved across CoV homologs, including ER proteostasis factors.

A. Nsp2 & nsp4 FLAG-tagged homolog construct design. Placement of N-terminal FLAG-tag for nsp2 and C-terminal FLAG-tag for nsp4 was based on previous construct designs(18). A corresponding 18 - 23 amino acid long leader sequence from the C-terminus of nsp3, including the PL2pro cleavage site, is added on the N-terminus of nsp4 constructs to improve protein biogenesis and insertion, as has been reported previously(10,42).

B. Schematic of comparative affinity-purification mass spectrometry (AP-MS) workflow to identify nsp-host interactors.

C. Enrichment of select nsp2 and nsp4 interactors conserved across multiple CoV homologs versus negative control (Tdtomato) background. n = 3-4 IP replicates/homolog in 1 MS run.

4.3.2 Conserved nsp-proteostasis interactors are pro-viral factors, including malectin

I next sought to test the functional relevance of conserved nsp-interactors and determine their pro- or anti-viral role in active CoV infection. I turned to the BSL-2 model betacoronavirus MHV, which has been used extensively to study CoV biology(43–45), to screen interactors in two phases. In the first phase, priority interactors (10 nsp2 interactors, 16 nsp4 interactors) were knocked down (KD) using pooled siRNA in Delayed Brain Tumor (DBT) cells, a highly permissive cell culture model for MHV, and then infected with a previously characterized replicase reporter virus containing a genome-encoded Firefly Luciferase protein fused to the N-terminus of nsp2 (MHV-FFL2)(2) (Figure 4-2**A**). Interactors were prioritized based on conservation across multiple CoV homologs or categorization in distinct biological clusters of interest. Cells were infected at multiplicity of infection (MOI) of 0.1 and 1 plaque forming units (PFU) per cell to account for MOI-dependent effects. Scramble siRNA was used as a negative control to represent basal infection levels and siCeacam1 was

used as positive control to deplete the cellular receptor for MHV. At 8 hours post infection (hpi) or 10 hpi (for MOI 1 and 0.1 respectively), cells were harvested and luminescence measured as a correlate of viral replicase protein production (Figure 9-3).
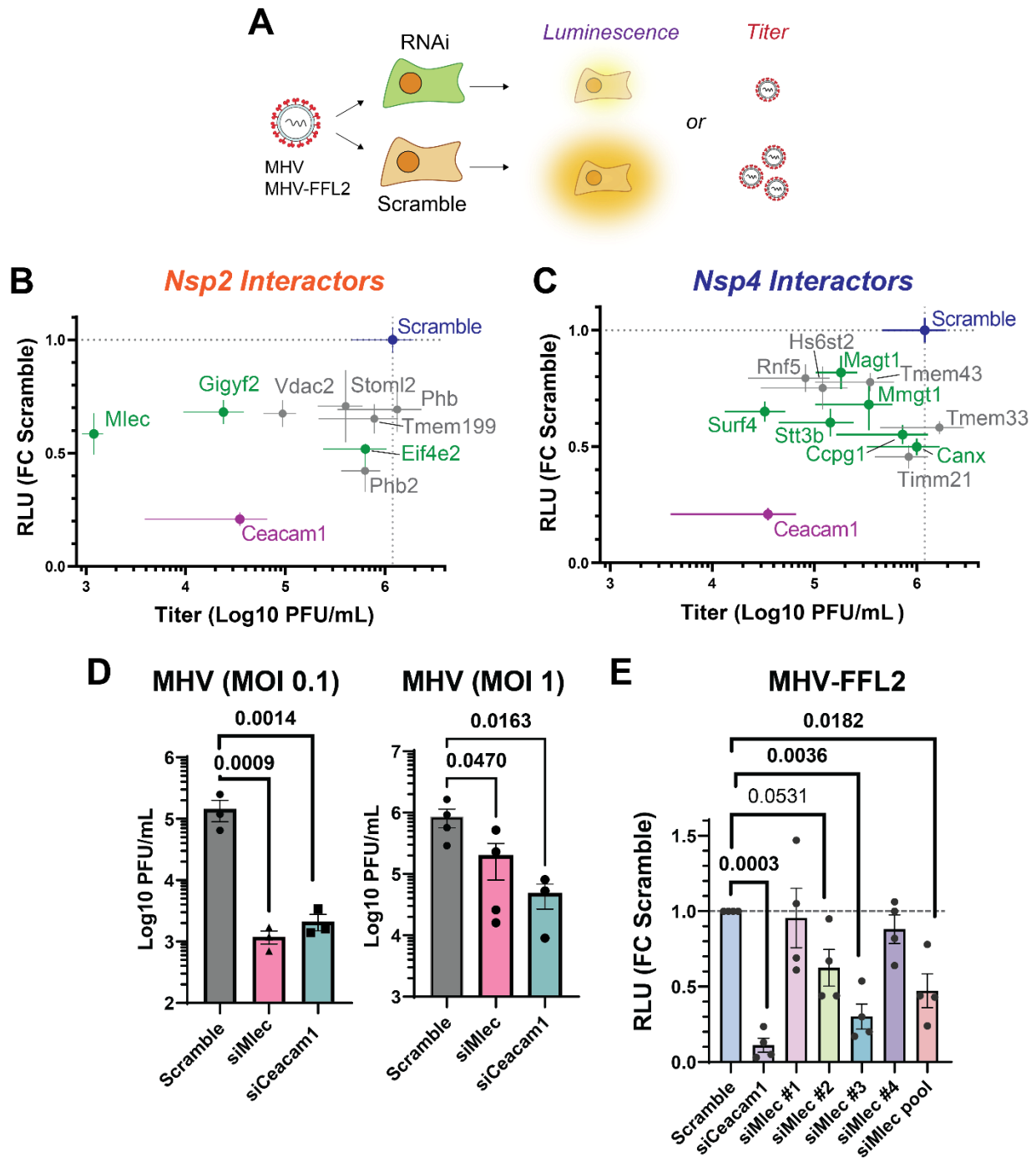
Figure 4-2. Conserved nonstructural protein (nsp) host interactors are pro-viral factors for CoV infection, including malectin (MLEC).

  A. Schematic of RNAi screening of priority interactors using a replicase reporter virus (MHV-FFL2, luminescence) or plaque assay (MHV, titer).

  B. Correlation plot of nsp2 interactor knockdown effect against replicase reporter virus (MHV-FFL2, luminescence) versus infectious titer (MHV, plaque). Scramble siRNA (blue) was used as basal control for comparison. siCeacam1 (MHV cellular receptor) was used as positive control (purple). Proteostasis factors are highlighted in green. Infections were performed at MOI 0.1 for 10 hpi, n=3-5. Dot represents mean, crosshatch indicates SEM.

C. Same as in (B), but of nsp4 interactor knockdowns.
D. MHV titers at MOI 0.1 or 1 in DBT cells treated with Scramble, siCeacam1, or siMlec siRNA (10 hpi), as measured by plaque assay. Paired Student's t-test for significance, $p<0.05$ considered significant, n = 3 (MOI 0.1), n = 3-4 (MOI 1).
E. Viral replicase reporter levels (luminescence) in DBT cells treated with Scramble, siCeacam1, individual siMlec (#1-4), or pooled siMlec siRNA, then infected with MHV-FFL2 (MOI 1, 9 hpi). Student's t-test used for significance, with $p<0.05$ considered significant. n = 4

I identified several proteins with significant effects on the viral replicase reporter. These proteins were then advanced to the second phase of screening in which the effect of interactor KD was assessed on MHV-WT viral titers using the lower throughput plaque assay (Figure 4-2**A**). The MOI and time point at which the significant effect in the luminescence reporter assay was observed were retained for the plaque assay. I identified several interactor KDs with significant reductions in titers (Figure 9-4). To better visualize the effect of interactor KDs on both viral replicase reporter and titers, I graphed these results together as a correlation plot for MOI 0.1 (Figure 4-2**B, C**) and MOI 1 (Figure 9-5). As expected, CEACAM1 KD leads to a large reduction in both replicase reporter and titer levels. I find that almost all interactor KDs lead to a reduction in both replicase reporter and titer levels, indicating that these proteins are pro-viral factors. At MOI 0.1, there are several ER proteostasis factors which lead to substantial reductions in replicase reporter levels but with a range of effects on viral titer. For nsp2 interactors, these include the glycoprotein quality control factor malectin (MLEC), and components of the translation repression complex EIF4E2 and GIGYF2. Among nsp4 interactors were several pro-viral factors that have previously been shown to support infection of other viral families. These include MMGT1 (EMC5, a component of the ER membrane complex) (46–48), SURF4 (ER secretory cargo receptor)(41,49–52), CCPG1 (ER-phagy receptor)(40), and components of the glycoprotein biogenesis factors (STT3B, MAGT1, CANX)(53,54). I also tested the effects of interactor KDs on cell viability and found that none led to more than 17% reduction in viability, indicating that cell toxicity was not driving observed phenotypes (Figure 9-6).

I found that MLEC KD yielded the largest decrease in viral titer, even surpassing the effect of CEACAM1 receptor KD. For this reason, I chose to focus on MLEC and further characterize this interaction. I verified that MLEC KD leads to titer reductions at both MOI 0.1 and 1 (10 hpi), albeit to a lesser extent with the latter, indicating that this is not a fully MOI dependent phenotype (Figure 4-2**D**). An MOI of 1 was used in all subsequent experiments unless otherwise specified.

I further validated if the reduction in MHV infection was MLEC-specific by testing the constituent individual siRNAs which compose the pooled siMlec treatment used previously. I find that at least two of the four siMlec (#2, 3) lead to reductions in viral replicase reporter levels (as measured with MHV-FFL2) and that endogenous MLEC protein levels are reduced by ~80% compared to Scramble (Figure 4-2**E,** Figure 9-7). Together, these results support a MLEC-specific phenotype and show that MHV is highly dependent on MLEC as a pro-viral factor for infection.

4.3.3   Nsp2 interacts with MLEC-associated protein complexes during infection

MLEC is an ER-resident glycoprotein quality control factor which binds with high affinity to diglucosylated moieties on protein-linked oligosaccharides in the ER lumen (34,55). These oligosaccharides are typically rapidly trimmed from triglucosyl to diglucosyl to monoglucosyl moieties by glucosidases (Glucosidase I and II respectively), but diglucosylated proteins may result from mistrimming events likely related to misfolding(56). MLEC expression is

upregulated during ER stress, when misfolded proteins accumulate, and has been found to retain diglucosylated glycoproteins in the ER lumen to prevent secretion of misfolded glycoprotein(35,57). MLEC has also been shown to associate with ribophorin I (RPN1)(58–60), a component of the OST complex, indicating a tight and early linkage between the initial addition of the triglucosylated oligosaccharide by the OST complex and the surveillance function of MLEC (Figure 4-3**A**).
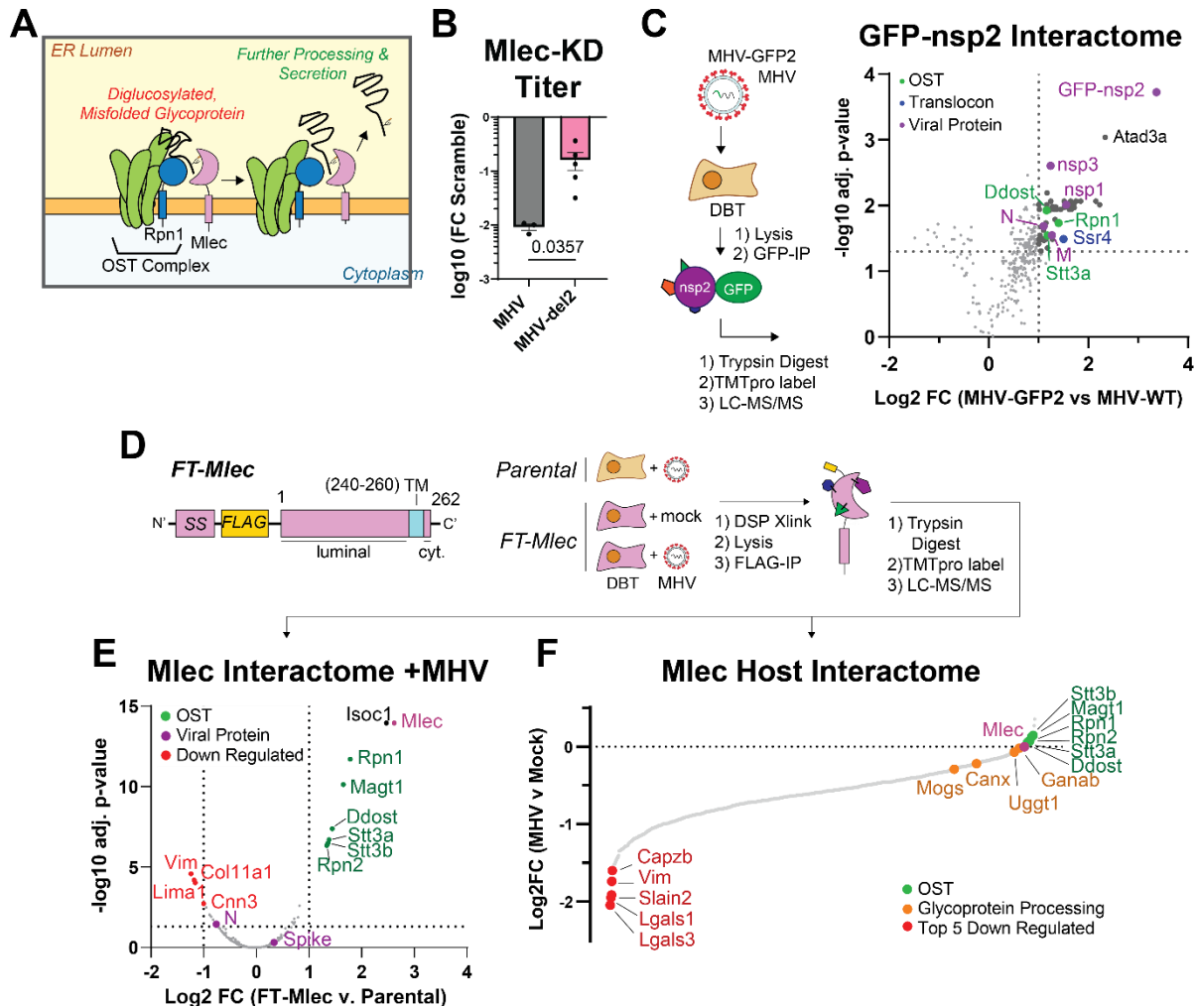

Figure 4-3. Nsp2 maintains a genetic interaction with MLEC and protein-protein interactions with MLEC-associated protein complexes during infection.
   A. Schematic of the role of MLEC in glycoprotein quality control.
   B. Comparison of MHV-WT vs. MHV-del2 (nsp2 deleted) infectious titers in siMlec vs. Scramble-treated DBT cells, expressed as Log10 fold change. MHV-WT (MOI 0.1, 10 hpi) or MHV-del2 (MOI 0.5, 10 hpi). n = 3-5, with Mann-Whitney T-test for significance, p <0.05 considered significant.
   C. Volcano plot of MHV GFP-nsp2 interactome during infection (MOI 1, 9 hpi). Comparison to MHV-WT infected cells was used as a background control for the GFP-IP. Linear filters of Log2 fold change > 1 and adjusted p-value > 0.05 were applied. Viral proteins (purple), OST complex factors (green), and translocon factors (blue) are annotated. MHV-GFP2 (n = 8 IPs), MHV-WT (n = 7 IPs) combined in 1 MS run. One-way ANOVA with multiple testing corrections was used for significance.
   D. Schematic of FT-MLEC construct design and IP-MS experimental design. FT-MLEC construct is composed of an N-terminal signal sequence (SS), followed by a FLAG-tag, and the MLEC luminal domain. There is a 10 residue long transmembrane helix with a short 2 amino acid long cytosolic C-terminal tail. Stably expressing FT-MLEC or WT DBT cells were infected with MHV (MOI 2.5, 9 hpi) and subjected to DSP (0.5 mM) crosslinking, before lysis

and FLAG IP. Elutions were reduced, alkylated, trypsin digested, and analyzed by LC-MS/MS. Parental DBT and MHV (n = 4 IPs), FT-MLEC DBT and mock (n = 6 IPs), FT-MLEC DBT and MHV (n = 6 IPs/condition), pooled into 1 MS run.

E. Volcano plot of FT-MLEC interactors during MHV infection. Samples were normalized to total peptide amounts. Linear filters of Log2 fold change > 1 or < -1 and adjusted p-value > 0.05 were applied. Viral proteins and MLEC (purple), downregulated (red), and OST complex factors (green) are annotated. T-test with multiple testing corrections for significance.

F. Waterfall plot of FT-MLEC interactome comparing MHV versus mock infection. Samples were normalized to FT-MLEC bait levels. Components of the OST complex (green), the glycoprotein processing pathway (orange), and the five most downregulated interactors (red) are annotated.

I found it curious that nsp2, a non-glycosylated, cytoplasmic protein, would interact with MLEC, an integral ER membrane protein with a short two amino acid cytoplasmic tail. I chose to first assess if there is a broader genetic interaction between nsp2 and MLEC. To this end, I compared the effect of MLEC-KD on WT MHV infection versus infection with MHV lacking nsp2 (MHV-del2)(1) (Figure 4-3**B,** Figure 9-8). As expected, infection with MHV-del2 with Scramble siRNA leads to reduced titer compared to WT MHV infection, which has been previously shown (1). I find that MLEC KD leads to only an order of magnitude reduction in titer of MHV-del2, as opposed to the two $\log_{10}$ PFU/mL reduction of MHV-WT, indicating that there is a genetic interaction between nsp2 and MLEC. Absent this interaction, we would have expected a greater decrease in titer with both MLEC KD and MHV-del2 infection.

I next sought to determine if nsp2 forms a PPI with MLEC during infection. I profiled the nsp2 interactome by infecting DBT cells (MOI 1) with an MHV genetically encoding GFP fused to the N-terminus of nsp2 (MHV-GFP2), which has been previously characterized(2). Infection with WT MHV was used as a background control for the GFP IP. At 9 hpi, I harvested and lysed cells with gentle lysis buffer to maintain native interactions, then immunoprecipitated GFP-nsp2. Elutions were reduced, alkylated, trypsin digested, and TMTpro labeled prior to LC-MS/MS analysis as described previously (Figure 4-3**C, Supplemental Table S3,** Figure 9-9).

I found that nsp2 interacts with several viral proteins, including the structural nucleocapsid (N) and membrane (M) proteins, and the nonstructural proteins nsp1 and nsp3. I also observe several mitochondrial interactors, most notably ATAD3A, which acts as a scaffold protein on the surface of the mitochondria to tether the outer and inner mitochondrial membranes together(61,62). This interaction may suggest a role for nsp2 in modifying mitochondrial activity, as previously suggested(18).

Excitingly, I identify several interactions between nsp2 and the OST complex (RPN1, STT3A, DDOST) and SSR4 (a component of the translocon complex). While MLEC itself was not identified, these interactors are previously known to associate with MLEC. These proteins themselves are integral ER membrane proteins with minimal cytosolic domain presence, likely indicating there may be some intermediary protein facilitating these interactions. Nsp2 has been shown to localize to RTC sites from the cytoplasmic face(2,3), so by extension these interactions between the OST complex and MLEC may be concentrated at these replication sites. Regardless, these interactions with known MLEC-associated proteins point towards a larger connection between nsp2 and glycosylated membrane protein processing machinery during infection.

To complement our GFP-nsp2 IP-MS results, I pursued a host-centric interactomics approach and sought to determine whether MLEC interacts with nsp2 or any other viral proteins during CoV infection. Additionally, I asked whether MHV infection alters the interactome of MLEC. To this end, I made DBT cells stably expressing an N-terminal FLAG-tagged construct of MLEC (FT-MLEC) (Figure 4-3**D**). I validated that the signal sequence of this construct is properly

cleaved (as evident by the immunoblotting reactivity to M1 FLAG antibody) and that FT-MLEC localizes to the ER, as shown by colocalization IF with PDIA4, an ER marker (Figure 9-10). I further tested whether overexpression of FT-MLEC in these cells alters MHV infectivity and find that viral titers are similar to infection in WT DBT cells (Figure 9-10).

To characterize the interactome of MLEC during infection, I infected FT-MLEC DBT cells with MHV (MOI 2.5) for 10 hpi (**Figure 4-3D**). Parental DBT cells were infected to serve as a background control for FLAG-IP. Mock infected FT-MLEC DBT cells served as a basal control to differentiate how MHV infection alters the MLEC interactome. Prior to harvest, infected cells were subjected to crosslinking with dithiobis(succinimidyl propionate) (DSP, 0.5 mM) to capture potentially transient interactions characteristic of lectin-client interactions and then lysed and immunoprecipitated for FLAG-tag. Elutions were subsequently trypsin digested, labeled with TMTpro, pooled, and analyzed by tandem mass spectrometry (LC-MS/MS) as described previously.

I compared FT-MLEC versus Parental DBT cells to identify MLEC-interacting proteins during infection (**Figure 4-3E, Supplemental Table S4, Figure 9-11**). Several members of the OST complex (RPN1, RPN2, MAGT1, DDOST, STT3A, STT3B) were significantly enriched as expected, given they are canonical MLEC interactors. The MHV Spike protein was positively but only modestly enriched. No nsp proteins were identified as interactors, contrary to my initial expectations.

I also examined the effect of MHV infection on the host interactome of MLEC (**Figure 4-3F**). Comparing FT-MLEC IPs from cells either mock or MHV infected (normalized to FT-MLEC levels, **Figure 9-11**), I found that there is a large loss of the basal MLEC interactome during infection. Some of the most downregulated proteins include CAPZB, VIM, SLAIN2, and LGALS1/3. LGALS proteins (galectin) are a group of lectins which bind beta-galactosides and typically reside in the cystol, though they can be secreted through transport into the ER-Golgi intermediate compartment (ERGIC)(63). There is also a more modest downregulation of interactions with glycoprotein processing factors, including MOGS (glucosidase I, GluI) and to a lesser extent GANAB (glucosidase II, GluII). GluI acts upstream of MLEC to trim the triglucosyl oligosaccharide on glycoproteins to a diglucosyl moiety, whereas GluII acts downstream to trim the diglucosyl oligosaccharide to a monoglucosyl group(64–66). From there, client glycoproteins enter the calnexin (CANX) folding cycle, which is partially regulated by UGGT1(13,16,67,68). Both CANX and UGGT1 are modestly downregulated interactors during MHV infection, suggesting a possible decoupling of early and late glycoprotein quality control systems.

In contrast, OST complex interactors (RPN1/2, STT3A/B, MAGT1, DDOST) are maintained and even slightly upregulated during infection. These results indicate that MHV infection induces a reorganization of the MLEC host interactome in which some canonical interactions are retained (the OST complex), while a slew of others is downregulated, including with other components of the glycoprotein biogenesis pathway and other lectin family proteins.

In summary, I find that nsp2 maintains a genetic interaction with MLEC and interacts with MLEC-associated proteins during infection. Furthermore, MHV infection seems to drastically alter the interactome of MLEC, leading to a stabilization of OST interactions while titrating off most basal interactors.

4.3.4   MLEC promotes early CoV replication events and viral protein biogenesis

I next sought to understand at which stage in the viral replication cycle MLEC promotes infection, namely viral entry and genome release, viral polyprotein biogenesis, replicase formation and genome replication, or virion assembly and release (Figure 4-4**A**).

To test if MLEC promotes viral entry, I measured abundance of viral positive-sense genomic RNA, *(+)gRNA*, present within the cell early in MHV infection (MOI 1, 2 hpi) via quantitative reverse transcription PCR (RT-qPCR) (Figure 4-4**B, Table 4-1**). I find that KD of the MHV cellular receptor CEACAM1 leads to a reduction in *(+)gRNA* levels compared to Scramble control, whereas siMlec actually leads to an increase in intracellular *(+)gRNA*. This finding indicates MLEC does not promote viral entry and genome release.
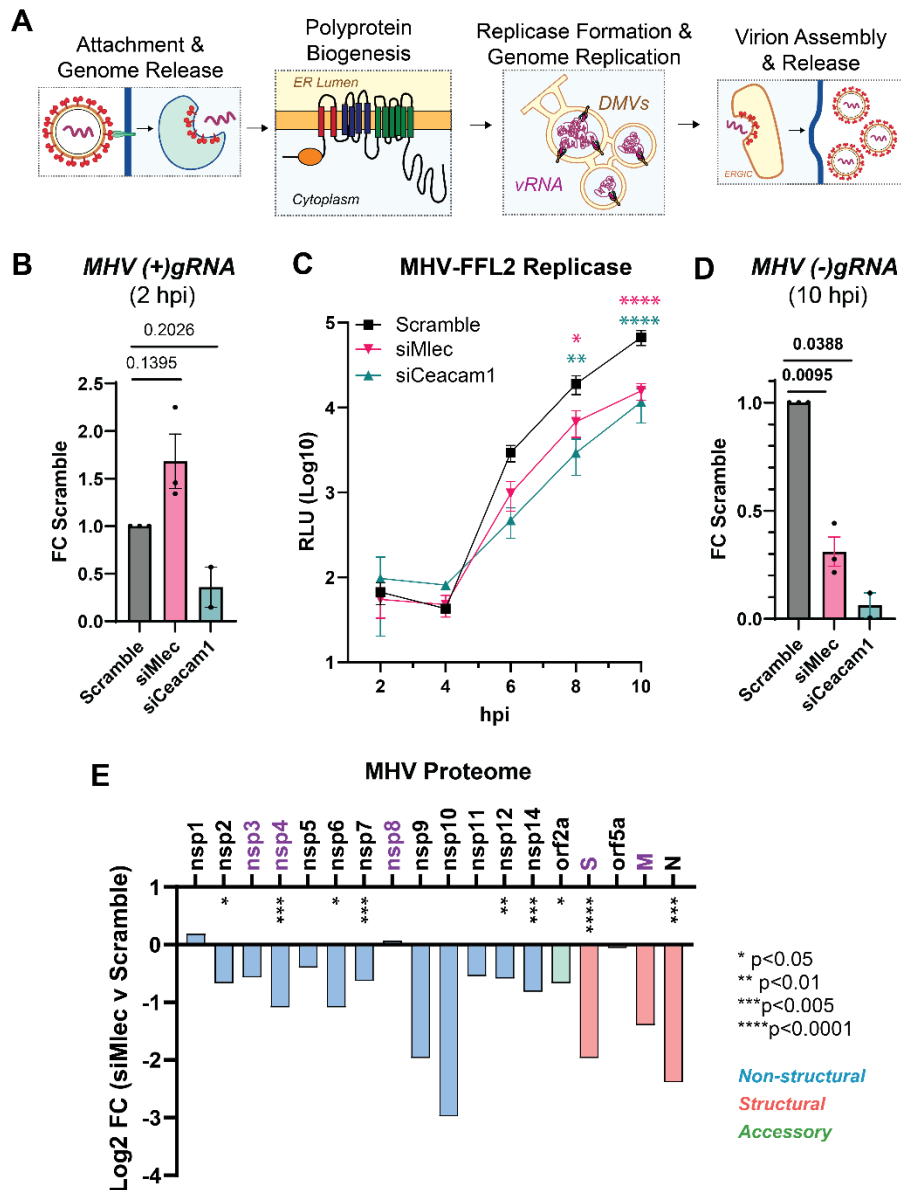
Figure 4-4. MLEC promotes early CoV replication events and viral protein biogenesis.
  A. Schematic of main events during the CoV replication cycle. Viral RNA (vRNA), double membrane vesicle (DMV), ER-Golgi Intermediate Compartment (ERGIC).
  B. Abundance of positive-sense MHV genomic RNA (*(+)gRNA*) as measured by RT-qPCR (MHV WT, MOI 1, 2 hpi) and normalized to *Rpl13a* transcript levels. Paired Student's T-test for significance, p<0.05 considered significant, n = 2-3.

C. MHV-FFL2 replicase reporter time course during infection from 2 – 10 hpi (MOI 1). Mock infected cells used as background luminescence control. Two-way ANOVA, Benjamini, Krieger, Yekutieli multiple comparisons, n = 3, * p<0.05, ** p<0.01, ***p<0.005, ****p<0.0001, color coding represents results of siMlec (pink) or siCeacam1 (teal) compared to Scramble control (black).

D. Abundance of negative-sense MHV genomic RNA (*(-)gRNA*) as measured by RT-qPCR (MHV WT, MOI 1, 10 hpi). Paired Student's T-test, p<0.05 considered significant, n = 2-3.

E. Relative abundance of intracellular viral proteins during MHV infection in DBT cells treated with Scramble or siMlec siRNA (MOI 1, 10 hpi), measured by quantitative proteomics. Purple protein names denote viral glycoproteins. Spike protein (S), membrane protein (M), nucleocapsid protein (N). One-way ANOVA with multiple testing correction. * p<0.05, ** p<0.01, ***p<0.005, ****p<0.0001, n = 6-7 biological replicates/condition across 2 MS runs.

Next, I asked whether MLEC promotes viral polyprotein production using the MHV-FFL2 virus, which reports on the abundance of nsp2 as a correlate of replicase protein production. DBT cells were treated with Scramble, siCeacam1, or siMlec and then infected with MHV-FFL2 (MOI 1) and lysates harvested every two hours, from 2 to 10 hpi (Figure 4-4**C**). I find that early on at 2 and 4 hpi luminescence levels are low and relatively equal amongst conditions. At 6 hpi, siMlec and siCeacam1 levels were stunted compared to Scramble, and this difference continued to significantly widen at 8 and 10 hpi. These results indicate MLEC likely promotes early replicase protein production.

I hypothesized then that a reduction in replicase proteins would in turn result in a loss of viral genomic RNA production. To test this outcome, I used RT-qPCR to measure the abundance of negative-sense viral gRNA, *(-)gRNA*, which is the biochemical intermediate of *(+)gRNA* replication, at 10 hpi (Figure 4-4**D, Table 4-1**). As expected, MLEC-KD leads to a significant reduction in *(-)gRNA* levels compared to Scramble control. I also find that *(+)gRNA* and *Nucleocapsid* transcripts are significantly reduced at 10 hpi, corroborating this result (Figure 9-12).

Since MLEC KD reduces nsp2 levels as seen in the MHV-FFL2 timecourse, I wondered how MLEC KD might affect the abundance of the rest of the MHV proteome. To this end, I turned again to quantitative proteomics to measure the abundance of intracellular viral proteins. DBT cells were treated with Scramble or siMlec and infected with MHV-WT (MOI 1, 10 hpi). Lysates were reduced, alkylated, trypsin digested, and TMTpro labeled for LC-MS/MS analysis as before. Comparing the siMlec vs. Scramble conditions, I find there is a global reduction in the total viral proteome, with significant decreases in structural, non-structural, and accessory proteins (Figure 4-4**E,** Figure 9-13). There are significant reductions in several viral glycoproteins, including Spike, nsp4, and nsp6, however MLEC KD also affects non-glycosylated protein levels, such as nucleocapsid and nsp2. Given that nsps are translated first in infection and that structural proteins are translated from subgenomic RNAs (sgRNAs) produced during viral genomic replication, the reduction in nsp levels with MLEC KD likely spurs the subsequent drop in structural protein abundance. As such, I conclude that nsp protein production is dependent on MLEC, which cascades into loss of viral genome replication and subsequent defect in viral titers.

4.3.5 MLEC mediates viral replication through the glycoprotein biogenesis pathway

Since MLEC acts as a quality control factor in the *N*-linked glycoprotein biogenesis pathway, I hypothesized that MLEC promotes viral protein production through this function. This would then predict that upstream inhibition of the pathway would not have a combinatorial effect with MLEC KD in suppressing viral protein production. In other words, if

MLEC promotes MHV replication through a mechanism outside of the glycoprotein biogenesis, then pathway inhibition and MLEC KD would suppress viral replication from two separate avenues, leading to synergistic infection suppression.

To test this hypothesis, I used the small molecule NGI-1, which inhibits the glycan transfer activity of the OST complex, upstream of GluI(69) (Figure 4-5**A**). DBT cells were transfected with Scramble siRNA (negative control), siCeacam1 (positive control), or siMlec and then treated with 5 µM NGI-1 or DMSO at the start of infection with MHV-FFL2 (MOI 1, 10 hpi). This dosage and timing were chosen to partially inhibit the OST complex without fully ablating viral infection, as NGI-1 has been shown previously to be a potent positive-sense RNA virus inhibitor(53) (Figure 9-14). Lysates of infected cells were analyzed by Western blot to verify MLEC KD and probe for ERDJ3, an ER-resident glycosylated chaperone, to measure *N*-linked glycosylation inhibition (Figure 4-5**B,** Figure 9-14). I found that the reduction in viral replicase reporter levels upon NGI-1 treatment were comparable between Scramble and MLEC KD cells, indicating that there is not a synergistic effect of upstream OST inhibition and MLEC KD and that MLEC promotes viral glycoprotein biogenesis through the canonical role (Figure 4-5**C**). In contrast, while not significant, CEACAM1 KD combined with NGI-1 treatment leads to increased suppression of MHV-FFL2 replicase reporter levels, as compared to Scramble-treated background.
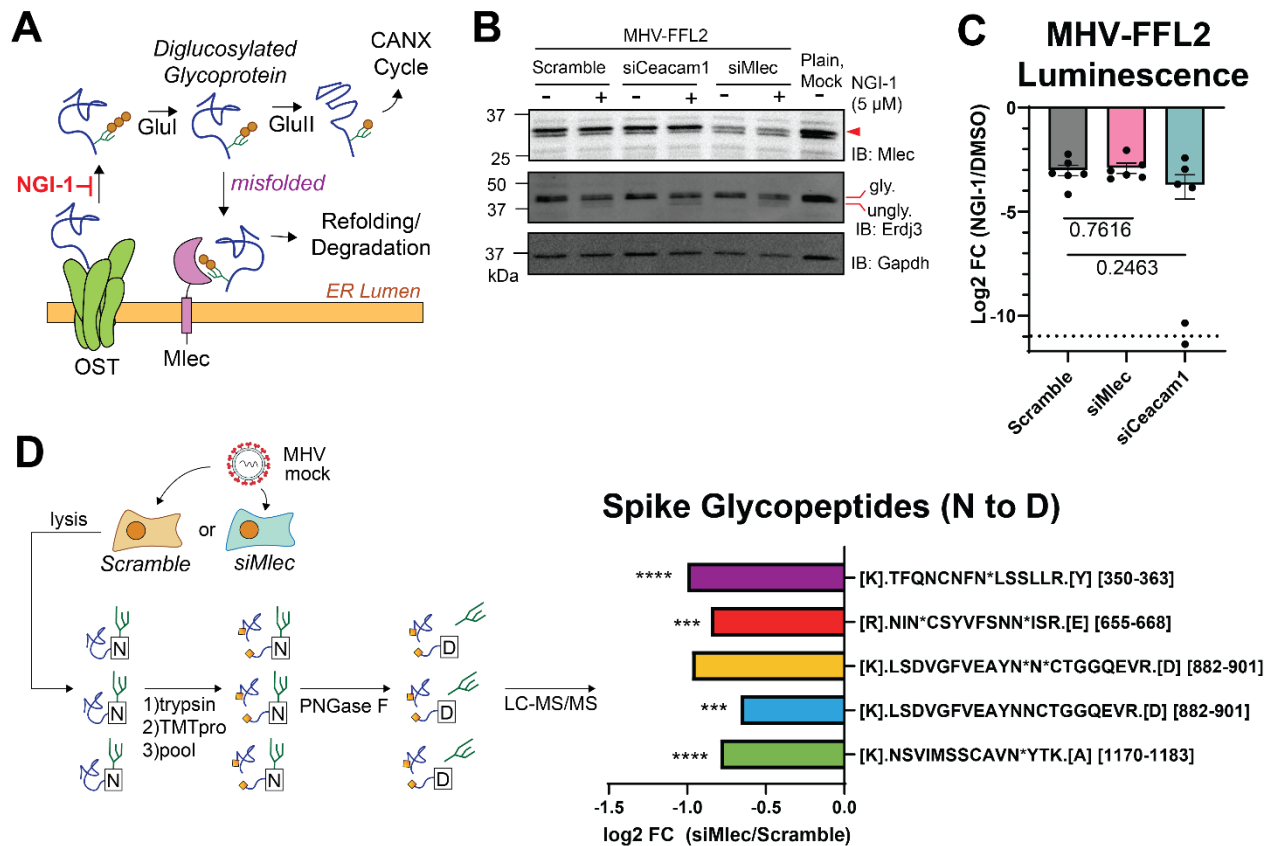
Figure 4-5. MLEC mediates replication through the glycoprotein quality control pathway.
  A. Schematic of glycoprotein biogenesis and quality control pathway. NGI-1 is an OST inhibitor that blocks *N*-linked glycosylation(69).
  B. Representative Western blot of DBT lysates after OST inhibition and MLEC KD during MHV-FFL2 infection (MOI 1, 10 hpi). Blotting for glycosylated and non-glycosylated ERDJ3 served as a marker of *N*-linked glycosylation inhibition. Lysates were matched with the corresponding luminescence reading.
  C. Levels of MHV-FFL2 replicase reporter with NGI-1 or DMSO treatment at the time of infection in cells with Scramble, CEACAM1, or MLEC KD in the same lysates analyzed in (B). One replicate of siCeacam1 fell below

the limit of detection (dotted line). Mock infected cells used as control for background luminescence. Student's T-test for significance, with p<0.05 considered significant, n = 6.

D. Abundance of *N*-linked glycosylated Spike peptides in siMlec versus Scramble-treated DBT cells. Cells were treated with siRNA and infected with mock or MHV (MOI 1, 10 hpi). Lysates were trypsinized, TMTpro labeled, pooled, then treated with PNGase F to cleave oligosaccharides, resulting in a deamidation mass shift (N to D) detectable by LC-MS/MS. Peptide sequence is listed, with "N*" denoting a deamidation modified Asn in a canonical glycosylation sequon (N-X-S/T, where X is not Proline). Peptides where the modified Asn could not be resolved lack the "N*" annotation. Peptide position within the protein is in brackets. One-way ANOVA, adjusted p-value. * p<0.05, ** p<0.01, ***p<0.005, ****p<0.0001, n = 4/condition, 1 MS run.

I then asked how knockdown of MLEC might affect the *N*-linked glycosylation status of viral proteins during infection. Based on our previous results, I would predict that MLEC KD would lead to disruptions in the glycoprotein biogenesis pathway and decreased glycosylation of viral glycoproteins. To assess the glycosylation status of viral proteins, I treated DBT cells with Scramble or siMlec siRNA and then infected cells with MHV (MOI 1, 10 hpi). Lysates were trypsin digested, TMTpro labeled, and pooled as before. Labeled peptides were then treated with PNGase F to trim off *N*-linked oligosaccharides, leading to an Asn to Asp conversion and a mass shift (deamidation, +0.984 Da) detectable by LC-MS/MS analysis(16,68) (Figure 4-5**D, Supplemental Table S5,** Figure 9-15). Viral peptides were then filtered for deamidation modifications and verified to contain canonical *N*-linked glycosylation site sequons (N-X-S/T, where X is not Proline; glycopeptides). I was unable to detect glycopeptides from nsps but identified five glycopeptides from the Spike protein. All five showed a decrease in abundance in siMlec versus Scramble siRNA treated cells, with four being statistically significant. While we cannot ascertain the effect on nsp glycopeptides with this data, these results further support a role for MLEC promoting viral protein biogenesis through the glycoprotein quality control pathway.

With this mass spectrometry data set, I also examined general alterations to the cellular proteome with MLEC KD and MHV infection (Figure 9-15). I find that in uninfected cells, MLEC KD leads to relatively little proteome-wide changes, with MLEC being the only protein significantly downregulated and no other proteins significantly upregulated, supporting the specificity of MLEC KD in MHV suppression (Figure 9-15**B**). In Scramble-treated cells, MHV infection leads to large alterations in the cellular proteome (Figure 9-15**C**), and MLEC KD attenuates these changes and viral protein abundance (Figure 9-15**D**).

4.4   Discussion

In this study, we build off our previous comparative interactomics study(18) to characterize conserved protein-protein interactions (PPIs) between CoV nsps and host proteostasis factors. By expanding the panel of nsp homologs tested, I was able to identify interactors conserved with additional viruses, including MERS-CoV, hCoV-229E, and MHV (Figure 4-1). MHV was then used as model virus to determine the pro- or anti-viral roles of conserved interactors. I identified several conserved ER proteostasis factors with pro-viral roles, most notably the glycoprotein quality control factor Malectin (MLEC), which leads to a drastic reduction in viral titers (Figure 4-2). Interrogating the nsp2-MLEC interaction, I found that nsp2 and MLEC maintain a genetic interaction and that nsp2 interacts with MLEC-associated proteins such as the OST complex during infection. MHV infection also remodels the MLEC interactome, enriching MLEC-OST complex interactions while titrating off basal MLEC interactors (Figure 4-3). I show that MLEC KD leads to defects in viral nsp production, which likely induces subsequent reductions in viral genome replication and structural protein production (Figure

4-4). Lastly, I find that MLEC promotes viral replication through the glycoprotein biogenesis pathway and MLEC KD leads to a decrease in viral glycopeptides during infection (Figure 4-5).

There are several potential mechanisms by which MLEC mediates CoV infection. Initially, I thought nsp2 might recruit MLEC to act as a retention factor that interacts with the pp1a/ab polyprotein to enhance nsp biogenesis and possibly promote the structural formation of RTCs. However, I was unable to detect a direct PPI between MLEC and nsps during MHV infection, so this may not be mediated through a direct interaction. A recent AP-MS study profiling the interactome of co-expressed SARS-CoV-2 nsp3 and nsp4 did identify MLEC as an interactor (70). Therefore, while I did not detect MHV nsp3 and nsp4 as MLEC interactors, the SARS-CoV-2 homologs may interact instead. It is possible that the set of conditions used here were not conducive to detecting these PPIs or that interactions are more prominent in earlier timepoints of infection as opposed to the one tested (10 hpi). Further work should be done to test alternative conditions and timepoints. This also points to some of the limitations of using individual protein overexpression for AP-MS, as the greater infection context is absent and interaction networks may differ. Additionally, it cannot be ruled out that MLEC also interacts with the other diglucosylated moiety present in cells, specifically the Lipid-Linked Oligosaccharide (LLO) precursor during glycan synthesis prior to addition of the third terminal glucosyl group by ALG10.

Alternatively, nsp2 and MLEC may be connected through an indirect mechanism. I show that nsp2 does interact with several OST complex factors, including DDOST, STT3A, and RPN1, the latter being a bonified interactor of MLEC(58,59). Further, MHV infection enhances the association of MLEC with the OST complex, potentially leading to more efficient glycoprotein biogenesis. Given that nsp2 has previously been shown to localize with RTCs during infection(2,3) and I detect nsp2 interactions with nsp3 (an RTC marker), I propose that the nsp2-OST-MLEC interaction axis leads to enhanced viral glycoprotein production around RTCs, most prominently being the biogenesis of nsps. This is in line with previous proximity-labeling work that showed host cell translational machinery locate near nsp2 and RTCs(17). Concurrently, knockdown of MLEC leads to impediment of nsp production and aberrant glycosylation of other viral proteins like Spike.

In parallel, MLEC KD may handicap the glycoprotein quality control network and lead to accumulation of misfolded diglucosylated proteins, thereby perturbing ER proteostasis. CoVs rely on the host ER for infection, from building DMVs from ER membranes to using host ER chaperones (many of which are glycosylated) to fold a high influx of nascent viral proteins. Imbalances in ER proteostasis would lead to less-than-amenable environment for viral replication and indeed previous work has shown that potent ER stressor small molecules lead to hampered CoV infection(71). Therefore, MLEC may also promote CoV replication through maintenance of ER proteostasis. Future studies should explore the impact of MLEC KD on basic cell biology and ER proteostasis.

Additional work is needed to characterize the spatial organization of the nsp2-OST-MLEC interaction complex during infection and to determine whether MLEC and OST complexes localize to RTCs. MLEC localization has been shown to be regulated by its association with RPN1, where under basal conditions MLEC localizes to the ER but under ER stress conditions can traffic to the Golgi (72). RPN1 overexpression restores MLEC localization to the ER. My FT-MLEC interactomics show that during MHV infection, MLEC retains its interactions with RPN1 and other OST interactors. Additionally, I showed that nsp2 interacts with RPN1 during infection. This suggests that nsp2 may use RPN1 to retain MLEC in the ER to assist in viral protein production.

It will also be important to identify the structural determinants of these interactions, including resolving the binding interfaces by creating truncations and point mutations, such as abolishing the lectin binding residues in MLEC(58). Profiling the kinetics of viral glycoprotein production in the absence of MLEC and developing targeted MS workflows to detect nsp glycopeptides will also aid our understanding of the role of MLEC in viral glycoprotein biogenesis. Lastly, the dependency of MLEC for infection will need to be tested with other CoVs, including SARS-CoV-2, as well as other RNA viruses.

To my knowledge, this is the first time MLEC has been implicated in CoV infection. Comparative interactomics approach suggests MLEC may play a role in multiple CoVs. Prior work has also indicated a role for MLEC in influenza protein folding(35), indicating that MLEC may serve a pro-viral function in a variety of RNA virus infections and make an attractive broad spectrum anti-viral target. Indeed, the strategy of drugging host proteostasis factors to inhibit RNA virus replication is being actively pursued(73–79). Additionally, MLEC remains an understudied component of the glycoprotein biogenesis pathway and our interactomics data sets can be of additional value to further interrogate its function.

In conclusion, this chapter combines quantitative proteomics with traditional virological assays to identify the conserved nsp2 interactor MLEC and reveal its pro-viral function in mediating nsp protein biogenesis.

4.5   Acknowledgements

4.6   Methods

Plasmids

Expression constructs for SARS-CoV-2, SARS-CoV, and hCoV-OC43 nsp2 and nsp4 homologs were used as previously reported(18). Expression sequences for nsp2 and nsp4 homologs of MERS-CoV, hCoV-2293, and MHV were pulled from GenBank (NC_019843 MERS-CoV EMC/2012; AF304460 hCoV-229E; KF268337 MHV-icA549). For nsp4, a corresponding 18 (MHV), 23 (MERS) or 21 (229E) aa leader sequence from the C-terminus of nsp3 was added to the N-terminus, as performed previously (10,42). Sequences were human optimized, synthesized, and cloned into a pTwist-CMV-Hygro vector with an N-terminal 3x-FLAG tag (nsp2) or C-terminal 3x-FLAG tag (nsp4) (Twist Biosciences).

Expression construct for FT-MLEC was made using the mouse coding sequence (GenBank, NP_780612.2). The endogenous signal sequence (1-29 aa) was replaced with the pro-trypsin signal sequence to improve expression efficiency (MSALLILALVGAAVA), followed by a 3x-FLAG tag prior to the remaining MLEC sequence. Construct was mouse codon optimized, synthesized and cloned into a pTwist-CMV-Hygro vector (Twist Biosciences).

Primers

Table 4-1. Primers used in this study.

ID	Name	Sequence	Notes

1	(+)gRNA_F	TGCGGTTTTATCACGCAGTT	To detect positive-sense MHV genomes ((+)gRNA), targeting nsp2 sequence from cDNA by RT-qPCR(80)
2	(+)gRNA_R	GGCAACCCACCCACGAA	To detect positive-sense MHV genomes ((+)gRNA) targeting nsp2 sequence from cDNA by RT-qPCR (80)
3	(-)gRNA_cDNA	GTGGATACACATCGTTATCG	To synthesize cDNA from negative-sense MHV genomes ((-)gRNA)(81)
4	(-)gRNA_F	CATGTTTACGGGCTAGCACC	To detect (-)gRNA, targeting nsp12 reverse sequence from cDNA by RT-qPCR (81)
5	(-)gRNA_R	CGAAGTTCTATGGCGGTTGG	To detect (-)gRNA, targeting nsp12 reverse sequence from cDNA by RT-qPCR (81)
6	Nucleocapsid_F	GCCTCGCCAAAAGAGGACT	To detect subgenomic viral RNA (Nucleocapsid) from cDNA by RT-qPCR(17)
7	Nucleocapsid_R	GGGCCTCTCTTTCCAAAACAC	To detect subgenomic viral RNA (Nucleocapsid) from cDNA by RT-qPCR (17)
8	Gapdh_F	ATGTGTCCGTCGTGGATCTG	To detect *Gapdh* transcripts by RT-qPCR
9	Gapdh_R	GGTGGAAGAGTGGGAGTTGC	To detect *Gapdh* transcripts by RT-qPCR
10	Rpl13a_F	GGCCGAGGCCATCTTCTT	To detect *Rpl13a* transcripts by RT-qPCR (17)

11	Rpl13a_R	TTCCGGAGAAAGGCCAGATA	To detect *Rpl13a* transcripts by RT-qPCR (17)

Cell Lines

HEK293T cells were a generous gift from Dr. Joseph Genereux (University of California, Riverside) and grown in high glucose Dulbecco's Modified Eagle Growth Medium with 10% fetal bovine serum, 1% glutamine, and 1% penicillin/streptomycin (DMEM-10). Murine Delayed Brain Tumor (DBT) cells were a generous gift from Dr. Mark Denison (Vanderbilt University Medical Center) and grown in DMEM-10 with 10 mM HEPES (DMEM-10-HEPES). HEK293T and DBT cells were maintained at 37°C with 5% CO2 and passaged with 0.25% (HEK293T) or 0.05% (DBT) Trypsin-EDTA. Cells were tested regularly for mycoplasma.

To generate stable expressing FT-MLEC DBT cells, parental DBT cells were transfected with 1.5 µg pTwist-CMV-Hygro-FT-MLEC plasmid in 6-well plates using 4.5 uL FuGENE 4K transfection reagent (#E5911) in DMEM-10-HEPES media (antibiotic-free) according to the manufacturer's directions. After two days, cells were expanded into 10 cm plates and selected with 200 µg/mL Hygromycin in DMEM-10-HEPES media. After an additional three days, cells were selected with 100 µg/mL Hygromycin in DMEM-10-HEPES media for another eleven days, with periodic media changes. Surviving polyclonal cells were expanded and frozen down, with a portion of cells lysed and tested for FT-MLEC expression by Western blot.

Virus Stocks

MHV-A59 (MHV-WT), MHV-FFL2 (Firefly luciferase fused to the N-terminus of nsp2)(2), MHV-GFP2 (GFP fused to the N-terminus of nsp2)(2), and MHV-del2 (deletion of nsp2)(1) were also gifted from Dr. Mark Denison (Vanderbilt University Medical Center). Virus was propagated in 10 cm dishes of DBT cells by infecting at MOI 0.01 for 24 hpi in DMEM-2-HEPES (same recipe as DMEM-10-HEPES, with only 2% FBS), freezing plates at -80°C, then thawing, centrifuging media (4,000 rpm, 10 min, 4°C), and storing supernatant in -80°C. Low passage stocks were used for infection experiments.

Transient Transfection

HEK293T cells were seeded into 10 cm dishes at 1.5×10^6 cells/plate in DMEM-10 media and a day later transfected with 5 µg of respective nsp plasmids or Tdtomato control via drop-wise addition of a calcium phosphate mixture (0.25 M CaCl2, 1x HBS (137 mM NaCl, 5 mM KCl, 0.7 mM Na2HPO4, 7.5 mM D-glucose, 21 mM HEPES)). Media was exchanged for fresh DMEM-10 16 h post-transfection, and cells were harvested by scraping an additional 24 h later using cold PBS (containing 1 mM EDTA) on ice, followed by x2 PBS washes. Cells were lysed in TNI lysis buffer (50mM Tris pH 7.5, 150mM NaCl, 0.5% IGEPAL-CA-630) containing 1x Roche c0mplete EDTA-free protease inhibitor (#4693132001) on ice for 15 min, sonicated 10 min at room temperature (RT), and then spun at 21,100xg (15 min, 4°C) to clear lysates of cell

debris. Supernatants were transferred to new tubes. Protein concentration was measured by Pierce BCA kit (Thermo Fisher Scientific, #23225) and normalized.

SDS-PAGE and Western Blotting Analysis

Samples were combined with a 6x Laemelli buffer (12% SDS, 125 mM Tris, pH 6.8, 20% glycerol, bromophenol blue, 100 mM DTT) at a 5:1 ratio, then heated at 37°C for 30 min. Samples were run on an SDS-PAGE gel before being transferred to Immobilon-FL PVDF membrane paper (Millipore Sigma, 0.45 um, #IPFL00010). Blots were blocked with 5% milk in TBS-T (Tris-buffered saline with 0.1% Tween) for 1 h, RT then probed with primary antibody. Anti-M2 FLAG (Sigma Aldrich, #F1804), anti-M1 FLAG (Sigma Aldrich, #F3040), anti-GAPDH (Genetex, #GTX627408), anti-ERDJ3 (Santa Cruz Biotechnology, #SC-271240), or anti-GFP (produced in-house) at 1:1000 in blotting buffer (5% BSA with 0.01% Sodium Azide in TBS-T). Anti-MLEC antibody (ProteinTech, # 26655-1-AP) was used at 1:800 in blotting buffer. Blots were then treated with either 1:10,000 dilutions in 5% milk/TBS-T of goat anti-mouse IgG StarBright700 (BioRad, #12004158) or goat anti-rabbit IgG IRDye800CW (Fischer, #926-32211) secondary antibody or 1:10,000 dilutions in TBS-T of anti-mouse or anti-rabbit IgG HRP (Promega, #W4021 or #W4011). All blots were imaged by ChemiDoc and analyzed on BioRad ImageLab.

siRNA Reverse Transfection

siGENOME SMARTpool or individual mouse siRNAs for interactors, siGENOME Non-targeting Control (Scramble; # D-001206-13), and siTOX (#D-001500-01-05) were purchased from Horizon. In 96-well plate experiments, plates were pre-treated with poly-D-lysine (0.1 mg/mL) for 1 min at 37°C and washed twice with PBS. DBT cells ($1x10^4$, 96-well; $7.5x10^4$, 12-well; $1.5x10^5$, 6-well) were reverse transfected with siRNA (25 nM final concentration) using DharmaFECT 1 Transfection Reagent (Horizon/Dharmacon, #T-2001-02) according to the manufacturer's instructions and incubated at 37°C for 40 h.

CellTiter-Glo Assay

DBT cells were seeded in 96-well plates and reverse transfected with siRNAs in triplicate as detailed above. After 40 h, media was removed, cells washed with PBS, and analyzed by Cell TiterGlo kit (Promega, #G7572) according to the manufacturer's instructions.

MHV Infection

Viral stocks were thawed on ice prior to use. DBT cells were counted, and the amount of virus added to cells was calculated based on respective MOI. Virus stock was mixed thoroughly with 25 µL (96-well), 100 µL (12-well), or 200 µL of DMEM10-HEPES or antibiotic-free DMEM-10-HEPES (for siRNA transfected samples) per well. Cells were washed with PBS, then inoculum added, and plates incubated for 30 min at 37°C, with rocking every 10 min. Inoculum was removed and cells washed twice with PBS before addition of respective media. Cells were incubated at 37°C for the prescribed infection duration. At harvest, supernatant was removed and saved for plaque assay. Cells were washed with PBS and then harvested by 0.05% Trypsin-EDTA.

Plaque Assay

DBT cells were plated in technical duplicate/dilution at 2.5×10^5 cells/well in 12-well plates. After 24 h, serial 1:10 dilutions of virus samples were made in gel saline (3 mg/mL gelatin from porcine skin (Sigma Aldrich, #G2625) in DPBS (Gibco, #14040-133). 50 µL of respective viral dilutions were added to cells and incubated for 30 min at 37°C (rocking every 10 min). After, 1 mL of overlay solution of 1:1 2% agar:2xDMEM-10-HEPES was added to wells and allowed to solidify (5 min, RT), before plates were incubated at 37°C for 20 hpi or until visible plaques formed. The assay was stopped with addition of 4% formaldehyde in PBS (20 min, RT) and solid overlay removed. Plates were air dried, and plaques counted manually.

MHV-FFL2 Replicase Reporter Assay

DBT cells were infected with MHV-FFL2 as detailed above. At harvest, supernatant was removed, and cells washed with PBS. Assays in 96-well plates (RNAi interactor screen) were then lysed directly with 100 µL/well of SteadyGlo substrate (Promega, # E2510) and 5 min of low rpm shaking. Assays in 12-well plates (all others) were lysed with 200 µL of Glo Lysis Buffer (Promega, #E2661) for 5 min (RT), spun at 13,000xg, 2 min (RT), and mixed in a 1:1 ratio with SteadyGlo substrate in three technical replicates/sample in a 96-well plate. Plates were measured for luminescence. A portion of lysates was retained for Western blot analysis.

Immunofluorescence

FT-MLEC DBT cells were plated on glass-bottom culture dishes (MatTek, #P35G-0-14-C) and grown for 24 h. Cells were fixed with 4% paraformaldehyde in PBS, washed x3 with PBS, then permeabilized in 0.2% Triton-X/PBS. After three PBS washes, cells were treated with blocking buffer (1% BSA/PBS with 0.1% Saponin). After blocking, cells were incubated with a 1:750 dilution of anti-PDIA4 primary antibody (Protein Tech, 14712-1-AP) in blocking buffer for 1 hour at 37°C. After three PBS washes, cells were incubated with AlexFluor 488-conjugated anti-rabbit goat antibody (ThermoFisher, A-11008) in blocking buffer (1:500 dilution) for 30 min at RT. Cells were then stained with 1:1000 dilution of M2 FLAG primary antibody (SigmaAldrich, F1804) and 1:500 AlexFluor 594-conjugated anti-mouse goat antibody (ThermoFisher, A-11005) in blocking buffer. Cells were then mounted in Prolong Gold with DAPI stain (ThermoFisher, P36935). Cells were imaged using a Zeiss LSM-880 confocal microscope and analyzed in ImageJ.

Quantitative Reverse Transcriptase Polymerase Chain Reaction (RT-qPCR)

Cellular RNA was extracted with a QuickRNA miniprep kit (Zymo, #R1055). Then, 500 ng total cellular RNA was reverse transcribed into cDNA using M-MLV reverse transcriptase (Promega, #M1701) with random hexamer primers (IDT, #51-01-18-25) and oligo-dT 15mer primers (IDT, #51-01-15-05). For negative-sense gRNA measurement, primer #3 (**Table 1**) was used instead to generate cDNA. iTaq Universal SYBR Green Supermix (BioRad, #1725120) was then used with respective primers (primers #1-2, 4-11) for target genes and reactions run in technical duplicate in 96-well plates on a BioRad CFX qPCR instrument. Amplification was carried out in the following sequence: 95°C, 2 min; 45 cycles of 95°C,

10 s and 60°C, 30 s. A melting curve was generated in 0.5°C intervals from 65°C to 95°C. Cq values were calculated in the BioRad CFX Maestro software and transcripts were normalized to a housekeeping gene (*Gapdh* or *Rpl13a*).

NGI-1 Treatment

To determine dosage, DBT cells were seeded in 12-well plates, then 4 h later treated with a range of doses (0.05, 0.5, 5, 10, 20 µM) of NGI-1 (SigmaAldrich, #SML1620), Tunicamycin (Tm, 1 µg/mL; Abcam, #ab120296), or DMSO. After 20 h, cells were harvested and lysed in TNI lysis buffer as described in "Transient Transfection." DMSO-treated lysates were split in half, with one half deglycosylated by PNGase F (Promega, #V4831) treatment for 1 h at 37°C. Lysates were then analyzed by Western blot.

To test MLEC KD and NGI-1 effects on MHV-FFL2, 40 h post-siRNA reverse transfection DBT cells were infected MHV-FFL2 as described above. After inoculum incubation and PBS washes, media containing 5 µM NGI-1 or DMSO was added to cells. Infected cells were measured for luminescence as detailed above.

FLAG Immunoprecipitations (IPs)

Sepharose 4B resin (SigmaAldrich, #4B200) and G1 anti-DYKDDDDK resin (GenScript, #L00432-10) were pre-washed four times with wash buffer (lysis buffer without protease inhibitor) for each sample. Cell lysates were normalized in wash buffer and added to 15 µL Sepharose 4B resin for 1 h, rocking at 4°C. Resin was collected by centrifugation for 10 minutes at 400xg (4°C) and pre-cleared supernatant was added directly to 15 µL G1 anti-DYKDDDDK resin and rocked at 4°C overnight. Supernatant was then removed, and resin was washed x4 with wash buffer. Bound proteins were eluted with elution buffer (62.5 mM Tris, pH 6.8, 6% SDS) for 30 min at RT then 15 min at 37°C, followed by a second elution for 15 minutes at 37°C.

GFP IP

DBT cells were infected with MHV-GFP2 or MHV-WT in 10 cm plates (MOI 1, 9 hpi) as detailed above. Cells were harvested by 0.05% trypsin-EDTA and lysed in 200 µL TNI lysis buffer with protease inhibitor (Roche, #4693132001) as detailed in "Transient Transcription." Normalized lysates were pre-cleared with 15 µL Sepharose 4B resin as in "FLAG IPs" and then cleared supernatant was added to 15 µL GFP-Trap magnetic agarose beads (Chromotek, #GTMA-20) and incubated for 2 h rocking (4°C). Beads were separated with a magnet and rinsed four times with wash buffer. Proteins were eluted with elution buffer as in "FLAG IPs."

FT-MLEC IP and DSP Crosslinking

Parental or FT-MLEC DBT cells were infected with MHV (MOI 2.5, 9 hpi) in 10 cm plates as detailed above. At harvest, cells were washed x3 with PBS, then incubated with 10 mL of 0.5 mM dithiobis(succinimidyl propionate) (DSP) crosslinker (ThermoScientific, #PI22585) in PBS at 37°C for 10 min. Crosslinker was quenched with 1 mL of 1 M Tris (pH 7.5) for 5 min at 37°C, after which cells were washed x3 with PBS and trypsinized to remove from plate. Cells were lysed in RIPA lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 0.5% deoxycholate) with protease

inhibitor for 15 min on ice, followed by two 10 sec vortex pulses and spun at 21,100xg (15 min, 4°C). Cleared supernatant was normalized and immunoprecipitated for FLAG as in "FLAG IP."

Global Viral Proteomics Samples

DBT cells were reverse transfected with siRNA in 6-well plates and after 40 h infected with MHV (MOI 1, 10 hpi). Cells were harvested by 0.05% trypsin-EDTA and then lysed in TNI lysis buffer and prepared for mass spectrometry as follows.

General Mass Spectrometry Sample Preparation

Elutions or 20 µg of protein/lysate were prepared for mass spectrometry. Proteins were precipitated with LC/MS-grade methanol:chloroform:water (in a 3:1:3 ratio), then washed three times with methanol, with each wash followed by a 2-minute spin at 10,000xg at RT. Protein pellets were air dried and resuspended in 5 µL of 1% Rapigest SF (Waters, #186002122). Resuspended proteins were diluted with 32.5 µL water and 10 µL 0.5 M HEPES (pH 8.0), then reduced with 0.5 µL of fresh 0.5 M TCEP for 30 min at RT. Samples were alkylated with 1 µL of fresh 0.5 M iodoacetamide for 30 min in the dark at RT and then digested with 0.5 µg Pierce Trypsin/Lys-C (Thermo Scientific, # A40007) overnight at 37°C shaking. Peptides were diluted to 60 µL with LC/MS-grade water and labeled using TMTpro labels (Thermo Scientific, # A44520) for 1 h at RT, followed by quenching with fresh ammonium bicarbonate (0.4% v/v final) for 1 h at RT. Samples were then pooled, acidified to pH < 2.0 with formic acid, concentrated to 1/6th original volume via Speed-vac, and diluted back to the original volume with buffer A (95% water, 5% acetonitrile, 0.1% formic acid). Cleaved Rapigest products were removed by centrifugation at 17,000xg for 30 min (RT) and supernatant transferred to fresh tubes.

PNGase F Mass Spectrometry Sample Preparation

DBT cells were seeded in 12-well plates and reverse transfected with siRNA as previously described, then infected with MHV (MOI 1) after 40 h. At 9 hpi, cells were harvested by 0.05% trypsin-EDTA, washed twice with PBS, then lysed in RIPA lysis buffer with protease inhibitor. Protein concentrations were measured by BCA kit and 20 µg protein/sample was carried forward to MS preparation on a BioMek i7 automated workstation. Proteins were reduced with 5 mM DTT (60°C, 30 min, 1,000 rpm shaking), followed by alkylation with 20 mM Iodoacetamide (RT, 30 min, in dark) and quenched with additional 5 mM DTT (15 min, RT). Hydrophilic and hydrophobic Sera-Mag Carboxylate-modified SpeedBeads (Cytiva, #09-981-121, #09-981-123) were prepared in a 1:1 ratio and washed x4 with LC/MS-grade water. Eight (8) µL of beads mixture was added to sample to bind proteins and mixed. Pure ethanol (90 µL) was added to samples and incubated 5 min, RT, at 1,000 rpm. Beads were separated by a magnet and washed x3 with 180 µL 80% ethanol. Proteins were then on-bead digested with 0.5 µg Trypsin/LysC (Thermo Scientific, # A40007) in 0.5 M HEPES buffer (pH 8.0) for 10 h, shaking at 1,000 rpm at 37°C. After digestion, samples were labeled with TMTpro labels (1 h, RT) and quenched with 10% w/v ammonium bicarbonate (1 h, RT), then pooled together and reduced to 1/5th volume by Speed-vac. Approximately 40 µg of labeled peptides were then treated with PNGase F (Promega, #V4831) for 2 h at 37°C and taken forward to LC-MS/MS analysis.

Tandem Mass Spectrometry (LC-MS/MS) Analysis

Triphasic MuDPIT columns(82) were packed with 1.5cm Aqua 5 μm C18 resin (Phenomenex, #04A-4299), 1.5cm Luna 5 μm SCX resin (Phenomenex # 04A-4398), and 1.5cm Aqua 5 μm C18 resin. TMT-labeled samples (20 μg of lysates or 1/3rd of IP elutions) were loaded onto the microcapillaries via a high-pressure chamber. Samples were washed, with buffer A (95% water, 5% acetonitrile, and 0.1% formic acid) for 30 min. MudPIT columns were installed on the LC column switching valve and followed by a 20cm fused silica microcapillary analytical column packed with 3 μm Aqua C18 resin (Phenomenex, # 04A-4311) ending in a laser-pulled tip. Analytical columns were washed with buffer A for 30 min prior to use. Peptides were fractionated online using liquid chromatography (Ultimate 3000 nanoLC system) and then analyzed with an Exploris480 mass spectrometer (Thermo Fisher). MudPIT runs were carried out by 10μL sequential injections of 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% buffer C (500mM ammonium acetate, 94.9% water, 5% acetonitrile, 0.1% formic acid), followed by a final injection of 90% C, 10% buffer B (99.9% acetonitrile, 0.1% formic acid v/v). Each injection was followed by a 130 min gradient using a flow rate of 500nL/min (0-6 min: 2% buffer B, 8 min: 5% B, 100 min: 35% B, 105min: 65% B, 106-113 min: 85% B, 113-130 min: 2% B). ESI was performed directly from the tip of the microcapillary column using a spray voltage of 2.2 kV, an ion transfer tube temperature of 275°C and an RF Lens of 40%. MS1 spectra were collected using a scan range of 400-1600 m/z, 120k resolution, AGC target of 300%, and automatic injection times. Data-dependent MS2 spectra were obtained using a monoisotopic peak selection mode: peptide, including charge state 2-7, TopSpeed method (3s cycle time), isolation window 0.4 m/z, HCD fragmentation using a normalized collision energy of 36%, 45k resolution, AGC target of 200%, automatic (whole cell lysates) or 150 ms (IP elutions) maximum injection times, and a dynamic exclusion (20 ppm window) set to 60s.

Mass Spectrometry Data Processing

Identification and quantification of peptides were performed in Proteome Discoverer 2.4 (Thermo Fisher). For viral protein interactomics in HEK293T cells, the SwissProt human database (TaxID 9606, released November 23rd, 2019; 42,252 entries searched) with FT-nsp2 and nsp4-FT sequences (11 entries) manually added (42,263 total entries searched) was used. For peptides from experiments in murine DBT cells, the SwissProt mouse database (TaxID 10090, released November 23rd, 2019; 25,097 entries searched) with MHV-WT (GenBank: KF268337, 25 entries) or MHV-GFP2 sequences (GenBank: KF268337, with addition of eGFP sequence to the N-terminus of nsp2 as previously detailed(2), 25 entries) or the FT-MLEC sequence (1 entry) manually added (25,122 - 25,123 total entries searched) were used. Searches were conducted with Sequest HT using the following parameters: trypsin cleavage (maximum two missed cleavages), minimum peptide length 6 AAs, precursor mass tolerance 20 ppm, fragment mass tolerance 0.02 Da, dynamic modifications of Met oxidation (+15.995 Da), protein N-terminal Met loss (−131.040 Da), and protein N-terminal acetylation (+42.011 Da), static modifications of TMTpro (+304.207 Da) at Lys, and N-termini and Cys carbamidomethylation (+57.021 Da). For PNGaseF-treated samples, an additional dynamic modification of Asn deamidation (+0.984 Da) was used. Peptide IDs were filtered using Percolator with an FDR target of 0.01. Proteins were filtered based on a 0.01 FDR, and protein groups were created according to a strict parsimony principle. TMTpro reporter ions were quantified considering unique and razor peptides, excluding peptides with co-isolation interference greater than 25%. Peptide abundances were normalized based on total peptide amounts in each channel, assuming similar levels of background, or based on bait protein for IPs where noted.

Protein quantification used all quantified peptides. Post-search filtering was carried out to include only proteins with two or more identified peptides.

Data Availability

The mass spectrometry proteomics raw data and **Supplementary Tables S1-S5**, which contain the complete protein and peptide identification list, are deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PXD049130).

4.7   References

1. Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S. & Denison, M. R. The nsp2 Replicase Proteins of Murine Hepatitis Virus and Severe Acute Respiratory Syndrome Coronavirus Are Dispensable for Viral Replication. *Journal of Virology* 13399–13411 (2005). doi:10.1128/jvi.79.21.13399-13411.2005

2. Freeman, M. C., Graham, R. L., Lu, X., Peek, C. T. & Denison, M. R. Coronavirus Replicase-Reporter Fusions Provide Quantitative Analysis of Replication and Replication Complex Formation. *J Virol* **88,** 5319–5327 (2014).

3. Hagemeijer, M. C. *et al.* Dynamics of Coronavirus Replication-Transcription Complexes. *J Virol* **84,** 2134–2149 (2009).

4. Prentice, E., McAuliffe, J., Lu, X., Subbarao, K. & Denison, M. R. Identification and Characterization of Severe Acute Respiratory Syndrome Coronavirus Replicase Proteins. *J. Virol.* **78,** 9977–9986 (2004).

5. Oostra, M. *et al.* Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication ▿ . *J Virol* **81,** 12323–12336 (2007).

6. Gadlage, M. J. *et al.* Murine Hepatitis Virus Nonstructural Protein 4 Regulates Virus-Induced Membrane Modifications and Replication Complex Function ▿ . *J Virol* **84,** 280–290 (2010).

7. Beachboard, D. C., Anderson-Daniels, J. M. & Denison, M. R. Mutations across Murine Hepatitis Virus nsp4 Alter Virus Fitness and Membrane Modifications. *J Virol* **89,** 2080–2089 (2015).

8. Sparks, J. S., Lu, X. & Denison, M. R. Genetic Analysis of Murine Hepatitis Virus nsp4 in Virus Replication. *J Virol* **81,** 12554–12563 (2007).

9. Angelini, M. M., Akhlaghpour, M., Neuman, B. W. & Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. *Mbio* **4,** e00524-13 (2013).

10. Oudshoorn, D. *et al.* Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication. *Mbio* **8,** e01658-17 (2017).

11. Tabata, K. *et al.* Convergent use of phosphatidic acid for hepatitis C virus and SARS-CoV-2 replication organelle formation. *Nat Commun* **12,** 7276 (2021).

12. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting Proteostasis for Disease Intervention. *Science* **319,** 916–919 (2008).

13. Hammond, C., Braakman, I. & Helenius, A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl. Acad. Sci.* **91,** 913–917 (1994).

14. Wright, M. T., Kouba, L. & Plate, L. Thyroglobulin Interactome Profiling Defines Altered Proteostasis Topology Associated With Thyroid Dyshormonogenesis. *Mol. Cell. Proteom.* **20,** 100008 (2021).

15. Wright, M. T., Timalsina, B., Lopez, V. G. & Plate, L. Time-Resolved Interactome Profiling Deconvolutes Secretory Protein Quality Control Dynamics. *bioRxiv* 2022.09.04.506558 (2022). doi:10.1101/2022.09.04.506558

16. Guay, K. P. *et al.* ER chaperones use a protein folding and quality control glyco-code. *Mol. Cell* (2023). doi:10.1016/j.molcel.2023.11.006

17. V'kovski, P. *et al.* Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8,** e42037 (2019).

18. Davies, J. P., Almasy, K. M., McDonald, E. F. & Plate, L. Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *Acs Infect Dis* **6,** 3174–3189 (2020).

19. Almasy, K. M., Davies, J. P. & Plate, L. Comparative Host Interactomes of the SARS-CoV-2 Nonstructural Protein 3 and Human Coronavirus Homologs. *Mol Cell Proteomics* **20,** 100120 (2021).

20. Cornillez-Ty, C. T., Liao, L., Yates, J. R., Kuhn, P. & Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling ▽. *J Virol* **83,** 10314–10318 (2009).

21. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583,** 459–468 (2020).

22. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370,** eabe9403 (2020).

23. Stukalov, A. *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594,** 246–252 (2021).

24. Zou, L., Moch, C., Graille, M. & Chapat, C. The SARS-CoV-2 protein NSP2 impairs the silencing capacity of the human 4EHP-GIGYF2 complex. *Iscience* **25,** 104646 (2022).

25. Xu, Z. *et al.* SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc National Acad Sci* **119,** e2204539119 (2022).

26. Gupta, M. *et al.* CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. *bioRxiv* 2021.05.10.443524 (2021). doi:10.1101/2021.05.10.443524

27. Juszkiewicz, S., Speldewinde, S. H., Wan, L., Svejstrup, J. Q. & Hegde, R. S. The ASC-1 Complex Disassembles Collided Ribosomes. *Mol. Cell* **79,** 603-614.e8 (2020).

28. Kuroha, K., Zinoviev, A., Hellen, C. U. T. & Pestova, T. V. Release of Ubiquitinated and Non-ubiquitinated Nascent Chains from Stalled Mammalian Ribosomal Complexes by ANKZF1 and Ptrh1. *Mol. Cell* **72,** 286-302.e8 (2018).

29. Zilmer, M. *et al.* Novel congenital disorder of O -linked glycosylation caused by GALNT2 loss of function. *Brain* **143,** 1114–1126 (2020).

30. Pahmeier, F. *et al.* Identification of host dependency factors involved in SARS-CoV-2 replication organelle formation through proteomics and ultrastructural analysis. *J. Virol.* **97,** e00878-23 (2023).

31. Williams, J. M., Chen, Y.-J., Cho, W. J., Tai, A. W. & Tsai, B. Reticulons promote formation of ER-derived double-membrane vesicles that facilitate SARS-CoV-2 replication. *J Cell Biol* **222,** e202203060 (2023).

32. Ji, M. *et al.* VMP1 and TMEM41B are essential for DMV formation during β-coronavirus infection. *J Cell Biol* **221,** e202112081 (2022).

33. Davies, J. P. *et al.* Expression of SARS-CoV-2 Nonstructural Proteins 3 and 4 Can Tune the Unfolded Protein Response in Cell Culture. *J. Proteome Res.* (2023). doi:10.1021/acs.jproteome.3c00600

34. Schallus, T. *et al.* Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N-Glycosylation. *Mol Biol Cell* **19,** 3404–3414 (2008).

35. Galli, C., Bernasconi, R., Soldà, T., Calanca, V. & Molinari, M. Malectin Participates in a Backup Glycoprotein Quality Control Pathway in the Mammalian ER. *Plos One* **6,** e16304 (2011).

36. Wright, F. A. *et al.* A Point Mutation in the Ubiquitin Ligase RNF170 That Causes Autosomal Dominant Sensory Ataxia Destabilizes the Protein and Impairs Inositol 1,4,5-Trisphosphate Receptor-mediated Ca2+ Signaling*. *J Biol Chem* **290,** 13948–13957 (2015).

37. Lu, J. P., Wang, Y., Sliter, D. A., Pearce, M. M. P. & Wojcikiewicz, R. J. H. RNF170 Protein, an Endoplasmic Reticulum Membrane Ubiquitin Ligase, Mediates Inositol 1,4,5-Trisphosphate Receptor Ubiquitination and Degradation*. *J Biol Chem* **286,** 24426–24433 (2011).

38. Goytain, A. & Quamme, G. A. Identification and characterization of a novel family of membrane magnesium transporters, MMgT1 and MMgT2. *Am. J. Physiol.-Cell Physiol.* **294,** C495–C502 (2008).

39. Bai, L., You, Q., Feng, X., Kovach, A. & Li, H. Structure of the ER membrane complex, a transmembrane-domain insertase. *Nature* **584,** 475–478 (2020).

40. Smith, M. D. *et al.* CCPG1 Is a Non-canonical Autophagy Cargo Receptor Essential for ER-Phagy and Pancreatic ER Proteostasis. *Dev Cell* **44,** 217-232.e11 (2018).

41. Emmer, B. T. *et al.* The cargo receptor SURF4 promotes the efficient cellular secretion of PCSK9. *Elife* **7,** e38839 (2018).

42. Kanjanahaluethai, A., Chen, Z., Jukneliene, D. & Baker, S. C. Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology* **361,** 391–401 (2007).

43. Gosert, R., Kanjanahaluethai, A., Egger, D., Bienz, K. & Baker, S. C. RNA Replication of Mouse Hepatitis Virus Takes Place at Double-Membrane Vesicles. *J Virol* **76,** 3697–3708 (2002).

44. Prentice, E., Jerome, W. G., Yoshimori, T., Mizushima, N. & Denison, M. R. Coronavirus Replication Complex Formation Utilizes Components of Cellular Autophagy*. *J Biol Chem* **279,** 10136–10141 (2004).

45. Sims, A. C., Ostermann, J. & Denison, M. R. Mouse Hepatitis Virus Replicase Proteins Associate with Two Distinct Populations of Intracellular Membranes. *J Virol* **74,** 5647–5654 (2000).

46. Barrows, N. J. *et al.* Dual roles for the ER membrane protein complex in flavivirus infection: viral entry and protein biogenesis. *Sci Rep-uk* **9,** 9711 (2019).

47. Ngo, A. M. *et al.* The ER membrane protein complex is required to ensure correct topology and stable expression of flavivirus polyproteins. *Elife* **8,** e48469 (2019).

48. Lin, D. L. *et al.* The ER Membrane Protein Complex Promotes Biogenesis of Dengue and Zika Virus Non-structural Multi-pass Transmembrane Proteins to Support Infection. *Cell Reports* **27,** 1666-1674.e4 (2019).

49. Tang, X. *et al.* A SURF4-to-proteoglycan relay mechanism that mediates the sorting and secretion of a tagged variant of sonic hedgehog. *P Natl Acad Sci Usa* **119,** e2113991119 (2022).

50. Saegusa, K. *et al.* Cargo receptor Surf4 regulates endoplasmic reticulum export of proinsulin in pancreatic β-cells. *Commun Biology* **5,** 458 (2022).

51. Yan, R., Chen, K., Wang, B. & Xu, K. SURF4-induced tubular ERGIC selectively expedites ER-to-Golgi transport. *Dev Cell* (2022). doi:10.1016/j.devcel.2021.12.018

52. Kong, L. *et al.* Surfeit 4 Contributes to the Replication of Hepatitis C Virus Using Double-Membrane Vesicles. *J Virol* **94,** (2020).

53. Puschnik, A. S. *et al.* A Small-Molecule Oligosaccharyltransferase Inhibitor with Pan-flaviviral Activity. *Cell Reports* **21,** 3032–3039 (2017).

54. Marceau, C. D. *et al.* Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* **535,** 159–163 (2016).

55. Schallus, T., Fehér, K., Sternberg, U., Rybin, V. & Muhle-Goll, C. Analysis of the specific interactions between the lectin domain of malectin and diglucosides. *Glycobiology* **20,** 1010–1020 (2010).

56. Tannous, A., Pisoni, G. B., Hebert, D. N. & Molinari, M. N-linked sugar-regulated protein folding and quality control in the ER. *Semin Cell Dev Biol* **41,** 79–89 (2015).

57. Chen, Y. *et al.* Role of malectin in Glc2Man9GlcNAc2-dependent quality control of α1-antitrypsin. *Mol Biol Cell* **22,** 3559–3570 (2011).

58. Takeda, K., Qin, S.-Y., Matsumoto, N. & Yamamoto, K. Association of malectin with ribophorin I is crucial for attenuation of misfolded glycoprotein secretion. *Biochem Bioph Res Co* **454,** 436–440 (2014).

59. Qin, S.-Y. *et al.* Malectin Forms a Complex with Ribophorin I for Enhanced Association with Misfolded Glycoproteins*. *J Biol Chem* **287,** 38080–38089 (2012).

60. Ramírez, A. S., Kowal, J. & Locher, K. P. Cryo–electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science* **366,** 1372–1375 (2019).

61. Arguello, T. *et al.* ATAD3A has a scaffolding role regulating mitochondria inner membrane structure and protein assembly. *Cell Reports* **37,** 110139 (2021).

62. Zhao, Y. *et al.* ATAD3A oligomerization causes neurodegeneration by coupling mitochondrial fragmentation and bioenergetics defects. *Nat Commun* **10,** 1371 (2019).

63. Zhang, M. *et al.* A Translocation Pathway for Vesicle-Mediated Unconventional Protein Secretion. *Cell* **181,** 637-652.e15 (2020).

64. Totani, K., Ihara, Y., Matsuo, I. & Ito, Y. Substrate Specificity Analysis of Endoplasmic Reticulum Glucosidase II Using Synthetic High Mannose-type Glycans*. *J. Biol. Chem.* **281,** 31502–31508 (2006).

65. Deprez, P., Gautschi, M. & Helenius, A. More Than One Glycan Is Needed for ER Glucosidase II to Allow Entry of Glycoproteins into the Calnexin/Calreticulin Cycle. *Mol. Cell* **19,** 183–195 (2005).

66. D'Alessio, C., Caramelo, J. J. & Parodi, A. J. UDP-GlC:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin. Cell Dev. Biol.* **21,** 491–499 (2010).

67. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4,** 181–191 (2003).

68. Adams, B. M., Canniff, N. P., Guay, K. P., Larsen, I. S. B. & Hebert, D. N. Quantitative glycoproteomics reveals cellular substrate selectivity of the ER protein quality control sensors UGGT1 and UGGT2. *eLife* **9,** e63997 (2020).

69. Lopez-Sambrooks, C. *et al.* Oligosaccharyltransferase inhibition induces senescence in RTK-driven tumor cells. *Nat Chem Biol* **12,** 1023–1030 (2016).

70. Li, J. *et al.* The REEP5/TRAM1 complex binds SARS-CoV-2 NSP3 and promotes virus replication. *J. Virol.* **97,** e00507-23 (2023).

71. Shaban, M. S. *et al.* Multi-level inhibition of coronavirus replication by chemical ER stress. *Nat Commun* **12,** 5536 (2021).

72. Yang, Q.-P. *et al.* Subcellular distribution of endogenous malectin under rest and stress conditions is regulated by ribophorin I. *Glycobiology* **28,** 374–381 (2018).

73. Almasy, K. M. *et al.* Small-molecule endoplasmic reticulum proteostasis regulator acts as a broad-spectrum inhibitor of dengue and Zika virus infections. *Proc National Acad Sci* **118,** e2012209118 (2021).

74. Aviner, R., Li, K. H., Frydman, J. & Andino, R. Cotranslational prolyl hydroxylation is essential for flavivirus biogenesis. *Nature* 558–564 (2021). doi:10.1038/s41586-021-03851-2

75. Aviner, R. *et al.* SARS-CoV-2 Nsp1 regulates translation start site fidelity to promote infection. *bioRxiv* 2023.07.05.547902 (2023). doi:10.1101/2023.07.05.547902

76. Taguwa, S. *et al.* Zika Virus Dependence on Host Hsp70 Provides a Protective Strategy against Infection and Disease. *Cell Reports* **26,** 906-920.e3 (2019).

77. Sims, A. C. *et al.* Unfolded Protein Response Inhibition Reduces Middle East Respiratory Syndrome Coronavirus-Induced Acute Lung Injury. *Mbio* **12,** e01572-21 (2021).

78. Echavarría-Consuegra, L. *et al.* Manipulation of the unfolded protein response: A pharmacological strategy against coronavirus infection. *Plos Pathog* **17,** e1009644 (2021).

79. Heaton, N. S. *et al.* Targeting Viral Proteostasis Limits Influenza Virus, HIV, and Dengue Virus Infection. *Immunity* **44,** 46–58 (2016).

80. Anderson-Daniels, J., Gribble, J. & Denison, M. Proteolytic Processing of the Coronavirus Replicase Nonstructural Protein 14 Exonuclease Is Not Required for Virus Replication but Alters RNA Synthesis and Viral Fitness. *J Virol* **96,** e00841-22 (2022).

81. Johnson, R. F. *et al.* Effect of Mutations in the Mouse Hepatitis Virus 3′(+)42 Protein Binding Element on RNA Replication. *J Virol* **79,** 14570–14585 (2005).

82. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19,** 242–247 (2001).

CHAPTER 5

SARS-COV-2 NONSTRUCTURAL PROTEINS 3 AND 4 TUNE THE UNFOLDED PROTEIN RESPONSE

This chapter is adapted from "Expression of SARS-CoV-2 Nonstructural Proteins 3 and 4 Can Tune the Unfolded Protein Response in Cell Culture" published in *Journal of Proteome Research* and has been reproduced with the permission of the publisher and my co-authors: Athira Sivadas, Katherine R. Keller, Brynn K. Roman, Richard J.H. Wojcikiewicz, Lars Plate.

5.1 Summary
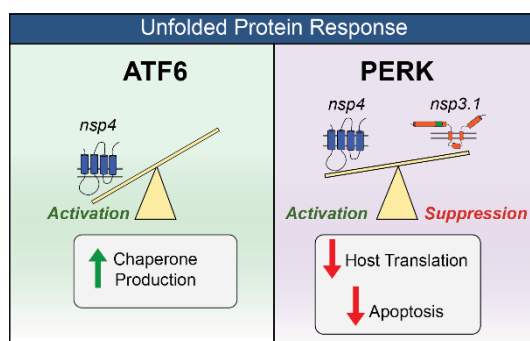
Coronaviruses (CoV), including SARS-CoV-2, modulate host proteostasis through activation of stress-responsive signaling pathways such as the Unfolded Protein Response (UPR), which remedies misfolded protein accumulation by attenuating translation and increasing protein folding capacity. While CoV nonstructural proteins (nsps) are essential for infection, little is known about the role of nsps in modulating the UPR. We characterized the impact of overexpression of SARS-CoV-2 nsp4, a key driver of replication, on the UPR in cell culture using quantitative proteomics to sensitively detect pathway-wide upregulation of effector proteins. We find nsp4 preferentially activates the ATF6 and PERK branches of the UPR. Previously, we found an N-terminal truncation of nsp3 (nsp3.1) can suppress pharmacological ATF6 activation. To determine how nsp3.1 and nsp4 tune the UPR, their co-expression demonstrated that nsp3.1 suppresses nsp4-mediated PERK, but not ATF6 activation. Re-analysis of SARS-CoV-2 infection proteomics data revealed time-dependent activation of PERK targets early in infection, which subsequently fades. This temporal regulation suggests a role for nsp3 and nsp4 in tuning the PERK pathway to attenuate host translation beneficial for viral replication while avoiding later apoptotic signaling caused by chronic activation. This work furthers our understanding of CoV-host proteostasis interactions and highlights the power of proteomic methods for systems-level analysis of the UPR.

All supplemental materials can be found in Appendix 4 or accompanying the online publication.

5.2 Introduction

The coronavirus SARS-CoV-2 has caused the COVID-19 pandemic, resulting in more than 6.8 million deaths to date(1). Coronaviruses (CoV) require host cell translation machinery and endoplasmic reticulum (ER) derived membranes for replication. CoV infection is known to induce ER stress and activate the Unfolded Protein Response (UPR)(2–4). The UPR alleviates stress from accumulation of misfolded proteins in the ER lumen by temporarily suppressing global protein translation, increasing the production of ER chaperones, expanding ER membrane synthesis, and if stress persists, triggering apoptosis. The UPR is composed of three branches that signal downstream of their respective ER-membrane localized stress sensors: 1. protein kinase R-like ER kinase (PERK), 2. Inositol requiring enzyme 1α (IRE1α), and 3. activating transcription factor 6 (ATF6) pathways (5,6).

Activation of PERK leads to the phosphorylation of eukaryotic initiation factor 2α (eIF2α) and attenuation of global protein translation to prevent further accumulation of misfolded proteins in the ER. A select group of proteins are translated under these conditions, including activating transcription factor 4 (ATF4), which upregulates expression of various genes involved in protein folding, antioxidant response, and the pro-apoptotic transcription factor C/EBP Homologous Protein (CHOP)(7,8). IRE1α has endoribonuclease activity which, upon ER stress, cleaves X-box-binding protein 1 (*XBP1*) transcripts, leading to splicing and translation of the XBP1s transcription factor and increased gene expression of ER protein chaperones, translocation and secretion factors, and components of ER-associated degradation (ERAD), as well as ER biogenesis(9). Lastly, ATF6 is translocated to the Golgi upon ER stress, where it is cleaved by site-1 and site-2-proteases. The N-terminal fragment (ATF6p50) is a transcription factor which initiates upregulation of various protein folding, secretion, and degradation factors, as well as expansion of the ER(5,6,10,11).

Coronavirus replication requires extensive production, folding, and modification of viral proteins(12), as well as alteration of ER-derived membranes to form double-membrane vesicles (DMVs) for replication sites. These processes can trigger ER stress and activate the UPR(13). Indeed, a previous study found that all three branches of the UPR are activated during SARS-CoV-2 infection and that expression of SARS-CoV-2 Spike or orf8 protein is sufficient to activate the UPR(3). SARS-CoV proteins orf3a and orf8ab were also found to activate the PERK and ATF6 pathways respectively(14,15). A related betacoronavirus, mouse hepatitis virus (MHV), triggers activation of XBP1 and PERK pathways while also hindering production of certain UPR-responsive genes such as CHOP(4). This previous work shows that coronaviruses can modulate stress responses at multiple phases and emphasizes the need for downstream proteome measurements to elucidate the consequences in the host cell.

Surprisingly, relatively little is known about the impact of CoV nonstructural proteins (nsps) on the UPR. Nsps are the first viral proteins translated during infection, rewire the host cell, and replicate the viral genome. In particular, three transmembrane nsps (nsp3, nsp4, and nsp6) are responsible for DMV formation from ER-derived membranes(16). Nsp3 contains a papain-like protease domain (PL2pro), which cleaves the orf1a/b polypeptide and also possesses deubiquination/de-ISGylation activity(17,18). Nsp4 is a glycoprotein containing four transmembrane domains and plays a key role in membrane reorganization(19,20).

We have previously characterized the interactomes of nsp3 and nsp4 CoV homologs and found that both proteins have evolutionary conserved interactions with several ER proteostasis factors(21,22). We identified an interaction between an N-terminal fragment of SARS-CoV-2 nsp3 (nsp3.1) and ATF6 and showed that nsp3.1 can suppress pharmacologic activation of the ATF6 pathway(22). Given the known role of nsp4 in host membrane alteration, we sought to elucidate if nsp4 activates or suppresses the UPR and whether nsp4 may act in concert with nsp3 to tune UPR activation. In particular, we leveraged a quantitative proteomics approach to characterize the upregulation of genes known to be transcriptionally activated by the UPR(6,23,24). Importantly, this approach enables sensitive measurement of pathway-wide changes in effector proteins and accounts for viral protein-mediated regulation that may occur downstream of transcriptional activation. A more precise understanding of the role of CoV nsps in modulating the UPR will further our knowledge of how coronavirus replication manipulates host proteostasis pathways during infection.

5.3   Results

5.3.1   SARS-CoV-2 nsp4 upregulates expression of UPR reporter proteins

Given the role of nsp4 in modulating host ER membranes during infection, we hypothesized that nsp4 may induce ER stress and subsequently activate the UPR pathway. To test this, we exogenously expressed a C-terminally FLAG-tagged SARS-CoV-2 nsp4 construct (nsp4-FT, **Figure 5-1A**) in HEK293T cells as previously reported(21) and measured transcript and protein expression of several UPR branch markers. Tdtomato (Tdt) expression was used as a negative control to account for any transfection-induced cell stress and benchmark the basal state of UPR marker levels. Importantly, previous work has shown transient expression of fluorescent, cytosolic proteins does not induce ER stress and the UPR(25). Treatment with tunicamycin (Tm, 1 µg/mL, 6 h) was used as a positive control for strong, general UPR activation. Samples were harvested 40 hours post-transfection, to ensure robust expression of viral proteins. Using RT-qPCR, we found that nsp4-FT expression led to a moderate but signification upregulation of *HSPA5* transcripts and upregulation, though not significantly, of *PDIA4* transcripts, both markers for ATF6 branch-activation (6,11,26) (**Figure 5-1B**).

To probe downstream protein expression of ATF6 UPR markers, we measured BiP, GRP94, and PDIA4 levels by Western blot in the presence of nsp4-FT (**Figure 5-1C-F, Figure 10-1, Figure 10-2**). These proteins have been extensively used in prior studies as Western blot markers for ATF6(27). Quantification of the control and nsp4-FT lanes show that nsp4 induces a weak upregulation of ATF6 reporter proteins (**Figure 5-1D,F**). We also found that nsp4-FT induces upregulation of a PERK pathway marker, CHOP(28) (**Figure 5-1E,F, Figure 10-2**), to a greater extent than SARS-CoV-2 FT-nsp2 (a cytosolic viral protein).

To determine whether nsp4-induced CHOP upregulation is mediated through the PERK pathway as opposed to other stress pathways, such as the Integrated Stress Response (ISR) pathway, we utilized the PERK inhibitor GSK2656157 (PERKi)(29,30). In HEK293T cells, GFP transfection with DMSO treatment served as basal control and Tm (1 µg/mL, 16 h) with PERKi treatment (1 µM, 16 h) was used as positive control for PERK inhibition activity. We found that treating cells expressing nsp4-FT with PERKi reduces induction of an upstream transcriptional regulator of CHOP, ATF4, (**Figure 10-3**) indicating that nsp4 upregulation of the ATF4/CHOP axis is dependent on PERK.

Together, these results indicate that SARS-CoV-2 nsp4 activates the UPR, albeit at milder levels than the toxic ER stressor tunicamycin, but still at biologically-relevant levels.
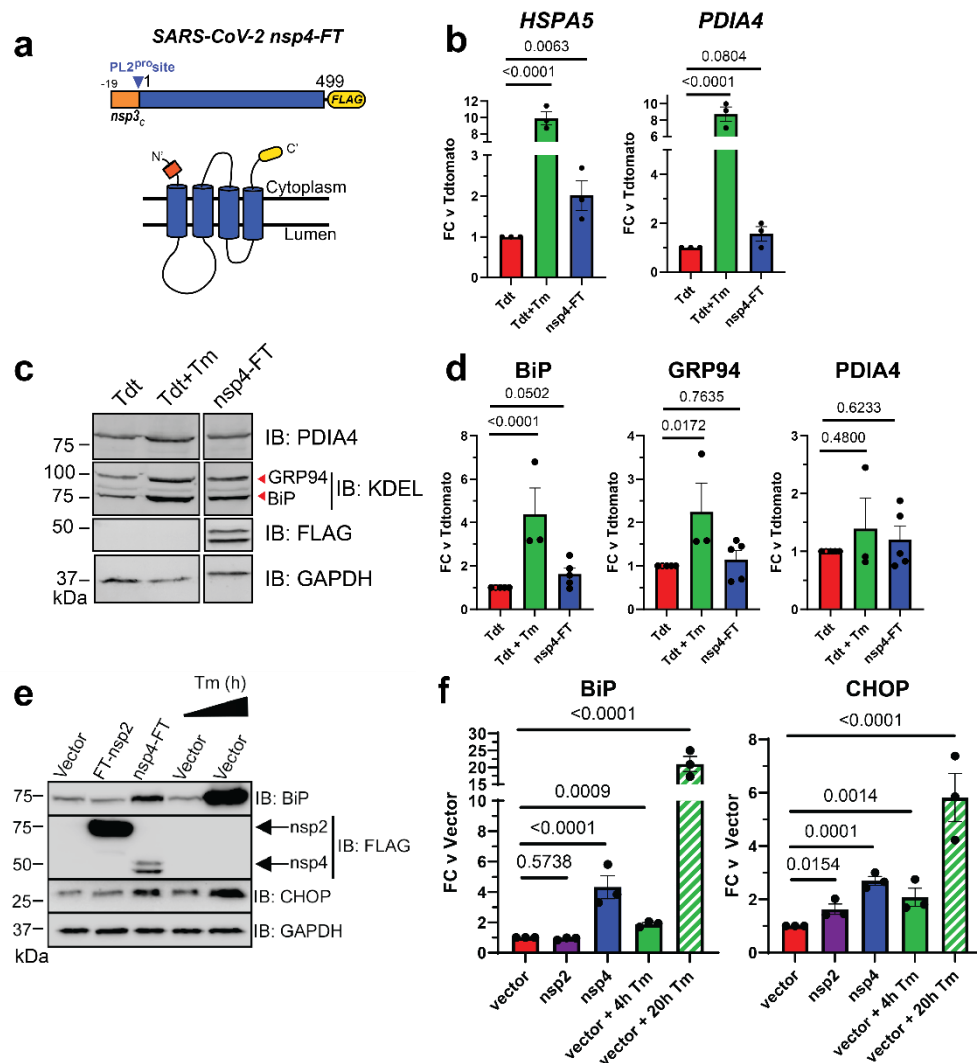
Figure 5-1. SARS-CoV-2 nsp4 upregulates expression of UPR reporter proteins.

A. SARS-CoV-2 nsp4 expression construct and membrane topology, containing a C-terminal FLAG-tag (nsp4-FT) and an N-terminal segment of the last 19 C-terminal amino acids in nsp3 for optimal membrane insertion(21). Construct contains the native PL2^{pro} protease cleavage site.

B. RT-qPCR of ATF6 pathway activation reporters *HSPA5* and *PDIA4* in the presence of SARS-CoV-2 nsp4-FT, normalized to *GAPDH* transcripts and compared to basal levels (GFP). Treatment with 1 µg/mL Tunacimycin (Tm) for 6 h was used as a positive control. n = 3, mean ±SEM, One-way ANOVA with Benjamini, Krieger and Yekutieli multiple testing correction, p<0.05 considered statistically significant.

C. Representative Western blot of ATF6 protein markers (PDIA4, GRP94, BiP) in the presence of Tdtomato (Tdt), SARS-CoV-2 nsp4-FT, or treatment with 1 µg/mL Tm (16 h), with GAPDH as a housekeeping gene for loading control. Displayed blot sections are from the same blot image and exposure settings (see **Figure 10-1** for all quantified, original blots).

D. Quantification of Tdtomato (Tdt), and nsp4-FT in Western blots in **(c)**, normalized to GAPDH band intensities, and compared to basal levels (Tdt). n = 3-5, mean ±SEM, One-way ANOVA with Benjamini, Krieger and Yekutieli multiple testing correction, p<0.05 considered statistically significant.

E. Representative Western blot of CHOP and BiP in the presence of SARS-CoV-2 FT-nsp2, nsp4-FT, or treatment with 5 µg/mL Tm for short (4h) or long (20h) time points. See **Figure 10-2** for all quantified, original blots.

F. Quantification of Western blots in **(e)**, normalized to GAPDH band intensities and compared to basal control (vector). n = 3, mean ±SEM. One-way ANOVA with Benjamini, Krieger and Yekutieli multiple testing correction, p<0.05 considered statistically significant.

5.3.2   Nsp4 upregulates the ATF6 and PERK pathways as measured by quantitative proteomics

While measuring a select few protein markers by Western blot provides some indication of UPR activation, a more precise and pathway-wide approach is required to comprehensively characterize downstream modulation of the cellular proteome. To this end, we analyzed the global proteome of HEK293T cells transfected with SARS-CoV-2 nsp4-StrepTag (nsp4-ST) using tandem mass spectrometry (LC-MS/MS) with TMTpro isobaric tags to quantify protein abundance (**Figure 5-2A, Supplemental Tables S1, S2, S3**). Transfection of GFP was used as a control to indicate the basal state of the UPR (-nsp4). Samples were normalized based on global peptide abundance (**Figure 10-4**). Over 4,200 proteins were identified, with HSPA5 being the most upregulated host protein (**Figure 5-2B**). A ranked list enrichment analysis of Gene Ontology (GO) Biological Processes terms identified "Cytoplasmic Translation" (GO:0002181) as the only GO term significantly enriched, with the majority of these categorized proteins positively enriched in the presence of nsp4 (**Figure 10-5**).

To measure pathway-wide changes of each UPR branch, we used a previously defined set of genes that have been shown to be transcriptionally upregulated upon stress-independent activation by RNA-Seq analysis(6,24,31) (**Supplemental Table S1**). By quantifying proteome changes in these pathway reporters, we can account for viral protein-mediated alterations in translation or degradation of proteostasis factors that may occur downstream of transcriptional activation.

We compared UPR pathway upregulation in the absence or presence of nsp4 by examining the distribution of target upregulation for each pathway (**Figure 5-2C, Figure 10-6**). We found that both the ATF6 and PERK pathways were significantly upregulated.

We also examined the abundances of individual proteins identified within the ATF6 and PERK pathways (**Figure 5-2D**). There was a heterogeneous ATF6 response to nsp4 expression, with some proteins (DNAJB11 and MANF) displaying relatively small change while others (HSPA5, PDIA4, HYOU1, HSP90B1) showed much higher upregulation. Comparing to previously published proteomics dataset(22), we find nsp4-ST induces higher expression of ATF6 markers over basal levels compared to a specific ATF6 pharmacological activator, compound **147**, but to a lesser extent than Tm treatment (**Figure 10-7**). We found that the PERK pathway upregulation by nsp4 was largely dominated by changes in ASNS and WARS abundance (**Figure 5-2D**).

To determine whether UPR activation by nsp4 was simply a result of overexpressing a multi-pass transmembrane protein, we expressed a panel of proteins with similar topology properties in HEK293T cells and measured effects on UPR markers by TMTpro LC/MS-MS compared to nsp4 and a GFP control (**Figure 10-8, Supplemental Tables S4, S5**). We chose multi-pass transmembrane proteins with previously characterized roles in human disease, specifically wild-type Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)(32), HA-tagged human Ether-à-go-go-Related Gene (hERG-HA)(33), and Myc-tagged Peripheral Myelin Protein 22 (PMP22-Myc)(34). None of these transmembrane proteins activated the UPR to a similar extent as nsp4, most notably the ATF6 pathway, indicating a nsp4-specific phenotype.
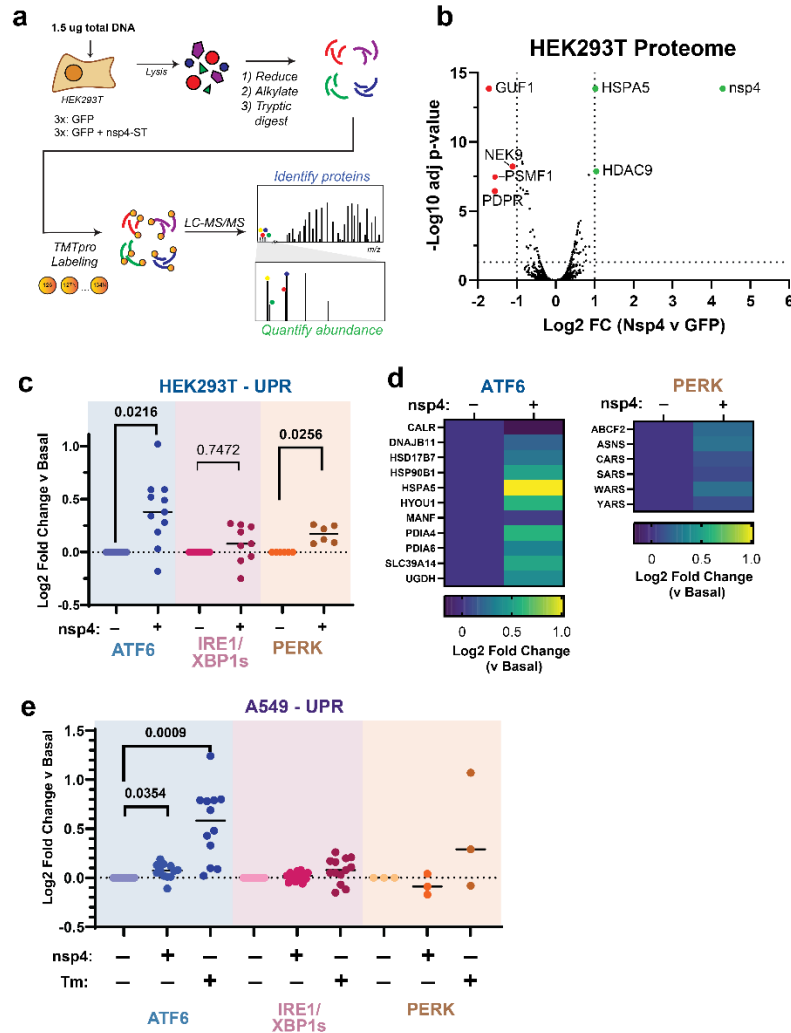
Figure 5-2. SARS-CoV-2 nsp4 upregulates the ATF6 and PERK pathways as measured by quantitative proteomics.

A. Experimental schematic to measure UPR induction by SARS-CoV-2 nsp4-ST expression in HEK293T cells using tandem mass spectrometry (LC/MS-MS) and TMTpro-based quantification.

B. Volcano plot of global proteome changes, with cut-offs indicating Log2 FC (vs GFP) < -1 or > 1, adjusted p-value < 0.05. Proteins passing cut-offs are annotated. Adjusted p-value calculated using a t-test, controlling for FDR by the Benjamini-Hochberg method. N = 3, 1 MS run.

C. Log2 fold change of UPR branch protein markers in the absence (GFP) or presence of SARS-CoV-2 nsp4-ST, as outlined in **(a)** assessed by TMT reporter ion intensities. Dots represent individual protein markers previously defined by RNA-seq analysis to be transcriptionally upregulated upon stress-independent activation of respective UPR branches (6,24,31). Swarm plot shows median, n = 3 biological replicates, 1 MS run, one-way ANOVA with Geisser–Greenhouse correction and post-hoc Tukey's multiple comparison test was used to test significance, *p*<0.05 considered significant. See **Supplemental Tables S2, S3** for mass spectrometry data set.

D. Heatmap of individual log2 fold change for ATF6 and PERK protein markers in the absence or presence of SARS-CoV-2 nsp4-ST.

E. Log2 fold change of UPR branch protein markers in A549 lung epithelial cells the presence of SARS-CoV-2 nsp4-FT or Tunicamycin (1 µg/mL, 16 h) compared to basal control (GFP). UPR branch protein markers were defined as in **(b)**. Swarm plot shows median, n = 4 biological replicates, 1 MS run, one-way ANOVA with Geisser–Greenhouse correction and post-hoc Tukey's multiple comparison test was used to test significance, *p*<0.05 considered significant. See **Supplemental Tables S6, S7** for mass spectrometry data set.

To determine if this phenotype extends to disease-relevant cell models, we expressed SARS-CoV-2 nsp4-FT in A549 lung epithelial cells and measured changes in UPR markers via TMTpro LC/MS-MS (**Figure 5-2E, Figure 10-4, Figure 10-9, Figure 10-10, Supplemental Tables S6, S7**). As a positive control for UPR activation, we included samples treated with Tm (1 µg/mL, 16 h). In our proteomics data set, we identified close to 4,000 proteins, with only one host protein surpassing a Log2 fold change > 1 threshold (AHSG) (**Figure 10-9**). A ranked list enrichment GO term analysis identified several terms significantly enriched, including "Nuclear DNA replication" and "Regulation of chromosome separation" (**Figure 10-9**).

ATF6 protein markers are moderately but statistically significantly increased in the presence of nsp4, though to a lesser extent than in HEK293T cells, while the PERK pathway was not upregulated but rather downregulated, indicative of some cell-type specific effects. Together, these results indicate SARS-CoV-2 nsp4 activates an ATF6 and PERK response characterized by substantial ER chaperone upregulation, though to a lesser extent compared to the potent ER stressor, tunicamycin.

5.3.3   SARS-CoV-2 nsp3.1 suppresses nsp4-induced PERK activation, but not ATF6 activation

Nsp4 and nsp3 are key inducers of double-membrane vesicle (DMV) formation during CoV replication(16,19), which, along with viral protein production, may induce ER stress and the UPR. We previously showed an N-terminal fragment of nsp3 (nsp3.1) suppresses pharmacological ATF6 activation(22). Therefore, we sought to test whether nsp3.1 might dampen nsp4-induced UPR activation, providing a form of viral tuning of the host UPR. To this end, we co-transfected HEK293T cells with SARS-CoV-2 nsp4-ST and nsp3.1-FT plasmids in equal amounts (0.75 µg) and measured the effect on ATF6 protein reporters by Western blot and UPR markers by mass spectrometry (**Figure 5-3A, Supplemental Tables S2, S3**). In addition, we tested if nsp3.1 could suppress UPR activation induced by SARS-CoV-2 orf8-ST, another viral protein known to activate the UPR(3). A GFP transfection served as a control for basal UPR state. To ensure that all samples were expressing similar amounts of protein, individual viral protein transfections were supplemented with equal amounts of GFP-expression plasmid (0.75 µg each). All transfections used 1.5 µg DNA plasmid in total.

As expected, we found that transfection with nsp4 and orf8 individually induced increased protein expression of the ATF6 markers BiP and GRP94 (though not PDIA4) compared to basal conditions, as measured by Western blot (**Figure 5-3B,C, Figure 10-11**). However, nsp3.1 did not lead to any notable decrease in the expression of ATF6 markers when co-expressed with nsp4 or orf8.

To more comprehensively quantify UPR protein markers, we analyzed these same lysates by TMTpro-quantitative LC-MS/MS. We found that nsp3.1 does not suppress nsp4- or orf8-induced ATF6 activation (**Figure 5-3D**). Interestingly, we found that co-expression of nsp3.1 with nsp4 significantly lowers PERK marker levels compared to nsp4 alone (**Figure 5-3E**), while IRE1/XBP1s activation was not significantly decreased with co-expression (**Figure 10-12**). These results demonstrate that nsp3.1 is not capable of suppressing ATF6 activation induced by other viral proteins, such as nsp4 or orf8, but can suppress PERK activation.
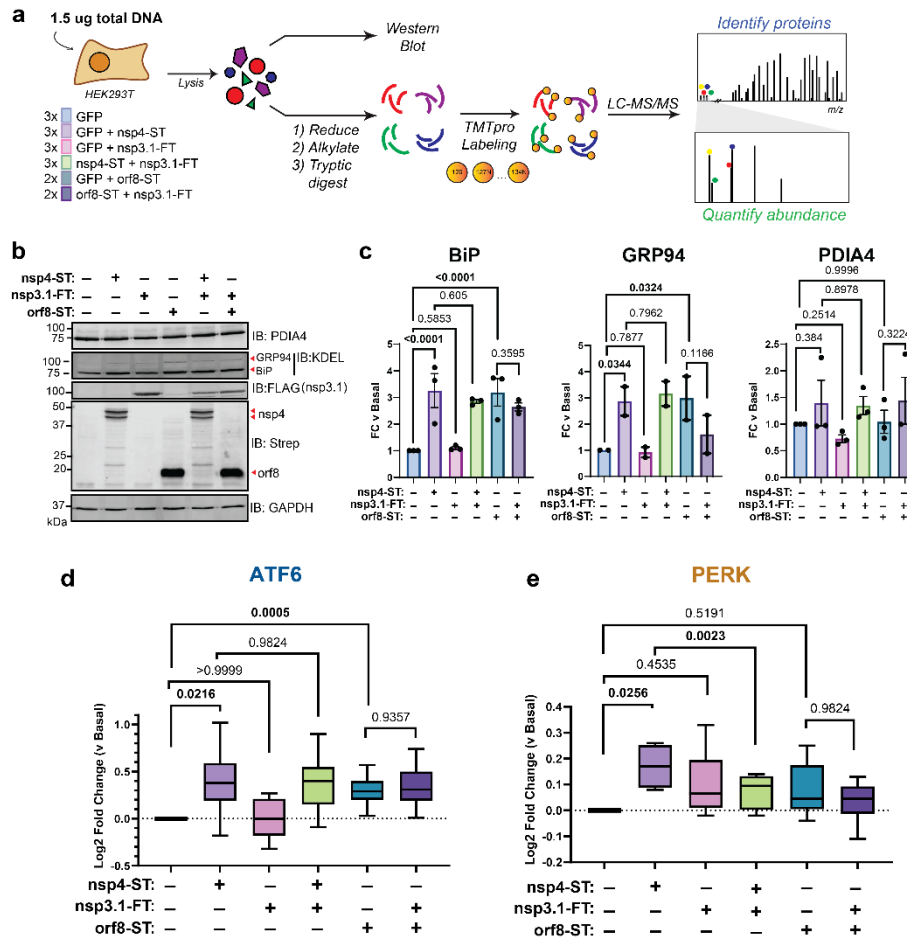
Figure 5-3. SARS-CoV-2 nsp3.1 suppresses nsp4-induced PERK, but not ATF6 activation.

A. Experimental schematic to test the effect of SARS-CoV-2 nsp3.1 on nsp4- or orf8-induced activation of the UPR in HEK293T cells by Western blot and tandem mass spectrometry (LC-MS/MS).

B. Representative Western blot of ATF6 protein markers (PDIA4, BiP, GRP94) and viral proteins (SARS-CoV-2 nsp4-ST, nsp3.1-FT, and orf8-ST), with GAPDH as a housekeeping gene for a loading control. See **Figure 10-11** for all quantified, original blots.

C. Quantification of Western blots in **(B)**, normalized to GAPDH band intensities. n = 2-3, mean ±SEM, One-way ANOVA with Benjamini, Krieger and Yekutieli multiple testing correction, p<0.05 considered statistically significant.

D. Global proteomics analysis of ATF6 protein marker levels in the presence of SARS-CoV-2 nsp4-ST, nsp3.1-FT, orf8-ST, or specified combinations. Box-and-whisker plot shows median, 25th and 75th quartiles, and minimum and maximum values. One-way ANOVA with Geisser–Greenhouse correction and post-hoc Tukey's multiple comparison test was used to test significance; 1 MS run, n = 2-3 biological replicates. See **Supplemental Tables S2, S3** for mass spectrometry data set.

E. Global proteomics analysis of PERK protein marker levels, as in **(D)**.

Curiously, we also noted nsp3.1 protein levels are lower in co-transfections with nsp4 versus GFP, as seen by both Western blotting (Figure 5-3**B,** quantified in Figure 10-13) and proteomics (Figure 10-13). Examining global changes in the proteome with nsp4 expression showed few proteins are substantially down- or up-regulated (Figure 10-5). Therefore, it is likely that this is a specific effect on nsp3.1 protein levels.

5.3.4　Re-analysis shows time-dependent PERK activation during SARS-CoV-2 infection

Previous work has shown that the UPR is activated during SARS-CoV-2 infection(3). A re-analysis of global proteomics during a SARS-CoV-2 infection time-course in Caco-2 cells(35) shows a modest upregulation of ATF6 and IRE1/XBP1s pathways by 10 hpi (**Figure 5-4A-D**). Additionally, a moderate upregulation of PERK-induced protein markers at 6 hpi that is downregulated by 24 hpi (**Figure 5-4E,F**), particularly for ABCF2, ASNS, CEBPB, and SLC1A4. This time-dependent regulation of the PERK UPR pathway induction and subsequent suppression may be explained by nsps, such as nsp3.1 and nsp4. Early activation of the PERK pathway may be important to interfere with host translation through eIF2α phosphorylation, while the later deactivation would avoid ATF4/CHOP-mediated cell death typically associated with prolonged PERK activation.

5.4   Discussion

Previous studies show that SARS-CoV-2, SARS-CoV, and MHV activate branches of the UPR to varying degrees(2–4,14,15), however the role of nsps in this process has remained largely unexplored. Of particular interest are the nsps involved in host membrane alteration to promote infection by DMV formation, such as nsp3 and nsp4. In this work, we find that overexpression of SARS-CoV-2 nsp4 activates the ATF6 and PERK branches of the UPR in HEK293T cells (Figure 5-1**,** Figure 5-2). This is evident by both Western blotting for ATF6 markers such as BiP and GRP94 and the PERK marker CHOP (Figure 5-1).

We further harnessed the capabilities of TMTpro-based quantitative proteomics to measure pathway changes in these UPR branches. The grouped analysis of co-regulated protein sets for the individual UPR pathways enables a comparative assessment of the degree of UPR branch activation and can even detect very small levels of induction (Figure 5-2). We also find that nsp4 activates the ATF6 pathway in A549 cells, though more moderately compared to HEK293T cells (Figure 5-2**E**), indicating cell-specific effects that may be tied to variable nsp4 protein expression in different cells, or divergent regulation mechanism of the UPR. We also acknowledge that transient transfection of individual or pairs of viral proteins cannot completely reflect the complete infection milieu, however this reductionist approach allows for characterization of specific viral protein contributions to UPR modulation, as has been conducted in prior studies(3,14,15).

In both HEK293T and A549 cells, the overall degree of UPR activation by nsp4 is lower than with the highly-toxic, global ER stressor tunicamycin, but higher than treatment with branch specific pharmacologic activators (Figure 5-2**,** Figure 10-7). This is in line with the moderate UPR activation observed in viral infection, where strong, chronic UPR activation may trigger apoptosis and would likely be counter-productive for replication. Additionally, prior work has found that modest upregulation of individual UPR branches can substantially improve quality control and secretion of disease-relevant proteins, such as in light-chain amyloidosis(23,24). Furthermore, our use of TMTpro-based quantitative proteomics using reporter ion intensity from MS2 spectra is subject to ratio compression, which may reduce the magnitude of observed fold changes(36). Thus, the moderate UPR activation measured here may be biologically relevant to CoV-reprogramming of host cells during infection.
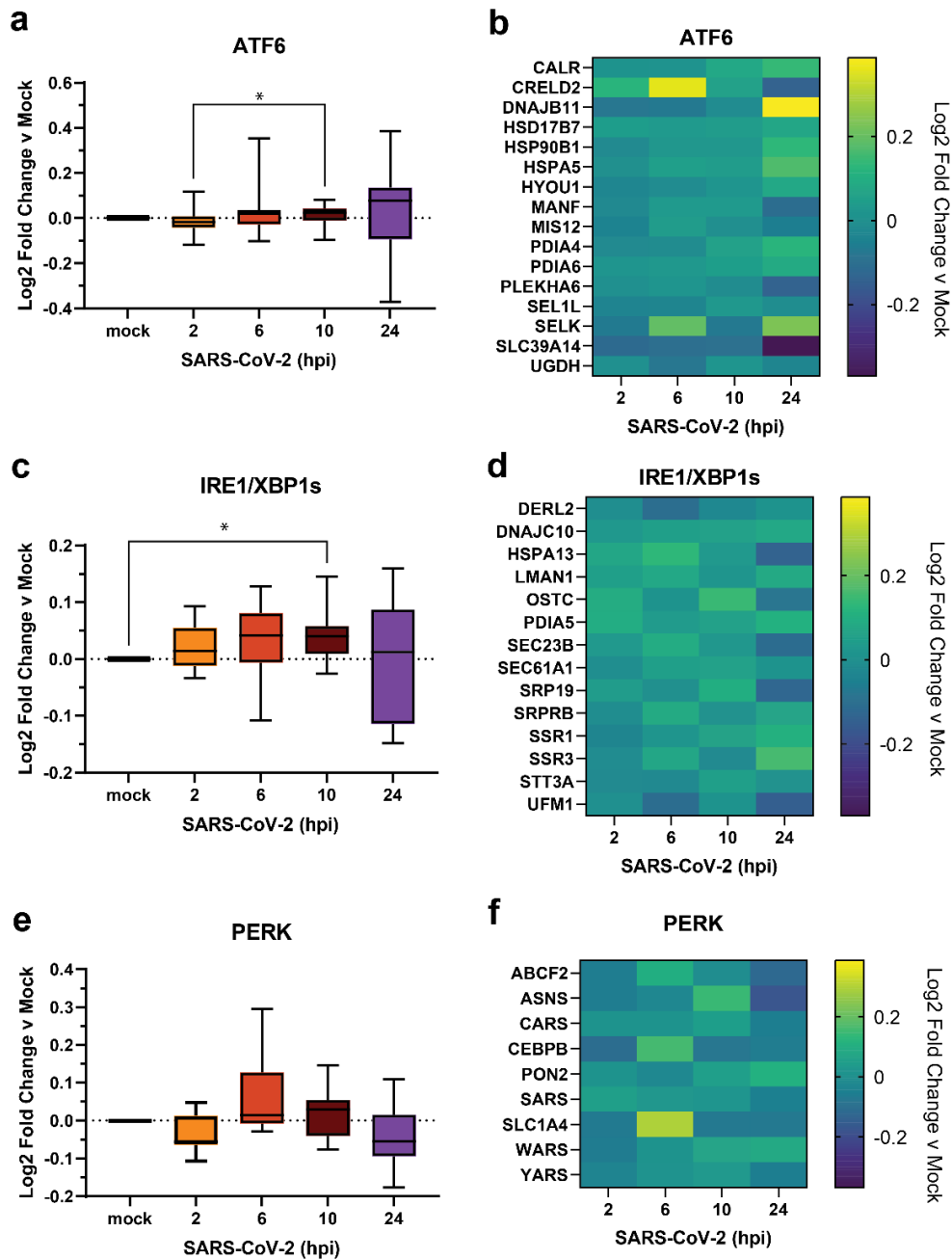
Figure 5-4. Re-analysis of UPR activation in global proteomics data of Caco-2 cells infected with SARS-CoV-2.
(A,C,E) Box-and-whisker plots of ATF6 **(A)**, IRE1/XBP1s **(C)**, and PERK **(E)** protein markers during SARS-CoV-2 infection (2-24 hours post infection (hpi)) compared to mock infected samples. Box-and-whisker plot shows median, 25th and 75th quartiles, and minimum and maximum values. One-way ANOVA with Geisser–Greenhouse correction and post-hoc Tukey's multiple comparison test was used to test significance. *p-value <0.05 annotated.
(B,D,F) Heatmaps of individual ATF6 **(B)**, IRE1/XBP1s **(D)**, and PERK **(F)** protein markers as in **(A), (C), (E)** respectively. Dataset published by Bojkova, et al. 2020(35).

It is pertinent to further study the molecular mechanisms by which nsp4 trigger UPR activation. We have previously profiled a comparative interactome of nsp4 homologs(21) and did not identify protein interactions with the ATF6 sensor protein as seen with nsp3(22), making it unlikely that signaling occurs through a direct interaction. As a multi-pass transmembrane protein, nsp4 may be prone to misfolding or may modulate host ER membranes, which can trigger ATF6

activation in some cases(37). Further work will be needed to define the precise molecular mechanisms by which nsp4 activates the ATF6 and PERK branches.

We previously showed that an N-terminal fragment of SARS-CoV-2 nsp3, nsp3.1, interacts with ATF6 and can suppress pharmacological activation of the ATF6 pathway(22). We hypothesized that nsp3 and nsp4 may act in concert to tune the UPR to increase protein folding capacity while minimizing apoptotic effects of chronic activation. We found that co-expression of nsp3.1 with nsp4 or orf8 does not repress ATF6 activation but does suppress nsp4-induced PERK activation (Figure 5-3). Time-dependent PERK activation was evident from global proteomics data during SARS-CoV-2 infection(35) showing an increase in target protein levels during the first 6 hpi, followed by a later decline (Figure 5-4). Though the magnitude of pathway change was more subdued than our data set, partially due to noisier data, there is still a trend of tight temporal regulation. This suggests a role for nsp3 and nsp4 in moderating PERK signaling to permit the early signaling events (eIF2α-mediated host translational attenuation) that are beneficial for viral propagation while preventing ATF4/CHOP-mediated induction of apoptosis.

The precise mechanism by which nsp3.1 suppresses nsp4-induced PERK activation will require further investigation. SARS-CoV-2 nsp3.1 directly interacts with ATF3 (22), a PERK protein marker which is upregulated by ATF4 and promotes pro-apoptotic signaling(38). This protein interaction may represent one avenue by which the PERK pathway is tuned by nsp3.1 to limit apoptosis in infected cells. We also acknowledge that other kinases, such as those involved in the Integrated Stress Response (ISR), may be activated, contributing to upregulation of protein markers categorized under the PERK pathway. We do find that nsp4-mediated upregulation of these proteins is dependent on PERK, using a small molecule PERK inhibitor (GSK2656157) (Figure 10-3). However, the contributions of other kinases, particularly in the context of our re-analysis of infection-based proteomics, cannot be ruled out.

The absence of ATF6 suppression in co-expression is surprising, given that we previously showed nsp3.1 could suppress ATF6 activation by tunicamycin treatment, which potently activates the global UPR through inhibition of protein glycosylation. This may suggest that the mechanisms by which nsp4 activates the ATF6 pathway are distinct from a general tunicamycin stressor and can overcome the opposing effects of nsp3.1. Alternatively, nsp4-induced ATF6 activation may be too moderate for nsp3.1 to have any measurable effect (Figure 10-7). In the context of viral infection, there are multiple points of UPR induction via nsp4, orf8, Spike, ER membrane perturbation, etc. which in combination may require partial suppression by nsp3.1 to tune the host UPR and prevent host cell apoptosis.

Relatedly, co-expression of nsp4 with nsp3.1 leads to a noticeable decrease of nsp3.1 levels compared to control co-expression of GFP with nsp3.1 (Figure 5-3**B,** Figure 10-13). This drop in protein levels may in part explain the lack of ATF6 suppression by nsp3.1. It is unclear how nsp4 affects nsp3.1 levels. While ATF6 activation increases production of ERAD factors to clear misfolded proteins, nsp3.1 is a cytosolic fragment of nsp3 and should not be subject to increased ERAD.

Additional questions remain regarding nsps and the UPR, such as how conserved is UPR activation across nsp4 homologs? Does nsp6, another key protein in DMV formation(16), have a similar effect on the UPR? And how might nsp3, nsp4, and nsp6 affect the UPR in concert? Using quantitative proteomics to measure changes in UPR-induced protein expression, as we have done here, should prove a powerful and comprehensive tool in answering these questions.

Lastly, previous work has shown that UPR inhibition can attenuate SARS-CoV-2(3) and MERS-CoV(39) infection. Viral families beyond coronaviruses, such as flaviviruses, have been shown to rely on the UPR and can be inhibited using UPR modulators(40–42). These opposing phenotypes, in which either UPR activation or inhibition can disrupt infection, highlight the important and diverse functions the UPR plays in viral replication. Continued efforts to delineate the roles of individual viral proteins in modifying the UPR will be critical to the further development of UPR-targeting anti-virals. Our work contributes to this goal by identifying a new role for SARS-CoV-2 nsp4 as a potent UPR activator and the coordination of nsp3.1 with nsp4 to tune the PERK pathway.

5.5   Acknowledgements

5.6   Methods

DNA Constructs

SARS-CoV-2 nsp4-FT, nsp3-FT, FT-nsp2, and orf8-FT (Wuhan-Hu-1 MN908947) were codon-optimized and cloned into a pcDNA-(+)-C-DYK vector (Genscript) or pcDNA-(+)-N-DYK vector for nsp2 (Genscript). An N-terminal truncation nsp3.1-FT (residues 1-749) was generated as previously described (21). Nsp4-ST and orf8-ST constructs were generated through the NEB HiFi Assembly system. In brief, the FLAG-tag from nsp4-FT was removed through amplification with primers 3 & 4 respectively. A 2xStrepTag was amplified from a SARS-CoV-2 nsp2-ST in a pLVX-EF1alpha plasmid construct (kind gift from Dr. Nevan Krogan, University of California, San Francisco) using primers 1 & 2. The linear nsp4 product was then combined with the 2xStrepTag fragment via HiFi assembly (1:2 vector to insert ratio). The orf8-ST construct was made by amplifying out the orf8 gene from the pcDNA-(+)-C-DYK vector using primers 5 & 6. A pLVX-EF1alpha-SARS-CoV-2-nsp2-2xStrepTag plasmid was linearized using primers 7 & 8, retaining the vector backbone and 2xStrepTag while removing the nsp2 gene. These were then combined via HiFi assembly (1:2 vector to insert ratio). A plasmid expressing human WT CFTR in the pcDNA5 vector was gifted from Drs. Eric Sorcher and Jeong Hong (Emory University)(43). Expression plasmid for human ERG (KCNH2) with an HA-tag in the first extracellular loop (residues 444-460) in a pAG293-pIRES2-GFP backbone was provided by Dr. Brett Kronke (Vanderbilt University Medical Center)(33).

Expression plasmid for human PMP22 with a Myc-tag in the second extracellular loop (residues 126-135) in a pIRES2-EGFP vector was provided by Dr. Charles Sanders (Vanderbilt University)(34). Plasmids were verified by sequencing (Genewiz).

Primers

Table 5-1. Description of primers used in this study. Primers are referenced by ID number in the text.

ID	Primer	Sequence	Notes
1	2xStrepTag_rem_F	CTCGAAGGCGGCGGGGGA	Amplification of 2xStrepTag from pLVX vector
2	2xStrepTag_rem_R	TTACTTTTCAAACTGCGGATGTGACCATGATCCAC	Amplification of 2xStrepTag from pLVX vector
3	nsp4_Wuhan_xFT_F	ATCCGCAGTTTGAAAAGTAATAAACCCGCTGATCAGCC	Removal of FLAG tag from nsp4
4	nsp4_Wuhan_xFT_R	CATCCCCCGCCGCCTTCGAGCAGGACTGCGGAAGTAATG	Removal of FLAG tag from nsp4
5	Wuhan-orf8_F	CCGGTGAATTCGCCGCCACCATGAAGTTCCTGGTATTTC	Amplification of orf8 gene from pcDNA vector
6	Wuhan-orf8_R	CATCCCCCGCCGCCTTCGAGAATAAAGTCCAAGACCAC	Amplification of orf8 gene from pcDNA vector
7	pLVX_F	CTCGAAGGCGGCGGGGGA	Amplication of pLVX vector with 2xStrepTag
8	pLVX_R	GGTGGCGGCGAATTCACCG	Amplication of pLVX vector with 2xStrepTag

9	GAPDH-qPCR_F	GTCGGAGTCAACGGATT	
10	GAPDH-qPCR_R	AAGCTTCCCGTTCTCAG	
11	HSPA5-qPCR_F	GCCTGTATTTCTAGACCTGCC	
12	HSPA5-qPCR_R	TTCATCTTGCCAGCCAGTTG	
13	PDIA4-qPCR_F	AGTGGGGAGGATGTCAATGC	
14	PDIA4-qPCR_R	TGGCTGGGATTTGATGACTG	

Cell Culture and Transfection

HEK293T cells were generously provided by Dr. Joseph Genereux (University of California, Riverside). A549 lung epithelial cells were obtained from ATCC (CCL-185). Cell lines were tested regularly for mycoplasma. HEK293T and A549 cells were maintained at 37°C, 5% CO2 in high glucose Dulbecco's Modified Eagle Growth Medium with 10% fetal bovine serum, 1% glutamine, and 1% penicillin/streptomycin (DMEM-10). Cells were seeded into 6-well plates at a concentration of 4×10^5 cells/well. HEK293T cells were transfected dropwise, twenty-four hours after seeding, using a calcium phosphate cocktail (1.5 µg total DNA, 0.25 M CaCl$_2$, 1x HBS (137 mM NaCl, 5 mM KCl, 0.7 mM Na$_2$HPO$_4$, 7.5 mM D-glucose, 21 mM HEPES)). For PMP22-Myc expression, a lower DNA amount of 0.3 µg was used, as previously described(34). The media was changed 18 hours later with fresh DMEM-10 media. A549 cells were seeded in antibiotic-free DMEM-10 and 24 h later were transfected with FuGENE 4K transfection reagent (FuGENE #E5911) (1.5 µg DNA with 4.5 µL FuGENE 4K) in plain Optimem media. Samples were treated with DMSO or Tunicamycin (Tm, 1 µg/mL) for 16 h (protein analysis) or 6 h (RNA analysis) prior to harvest in described experiments. GFP controls were used to assess transfection efficiency, which averaged 70-80% in HEK293T cells and 40-50% in A549 cells. Transfected cells were harvested by cell scraping in cold 1 mM EDTA in PBS over ice, 40 h post-transfection to enable robust expression of viral proteins. To measure CHOP levels, HEK293T cells were transfected twenty-four hours after seeding, using 1µg cDNAs and 6 µl of 1mg/ml PEI (pre-mixed in 50 µl of serum-free culture medium). Samples were treated with Tunicamycin (Tm, 5 µg/mL) for 4h or 20h prior to harvest.

RT-qPCR

HEK293T cells were transfected and harvested as described previously. Cellular RNA was extracted using the Zymo QuickRNA miniprep kit (#R1055) and 500 ng total cellular RNA was reverse transcribed into cDNA using random hexamer primers (IDT #51-01-18-25), oligo-dT 15mer primers (IDT # 51-01-15-05), and Promega M-MLV reverse transcriptase (#M1701). qPCR analysis was carried out using BioRad iTaq Universal SYBR Green Supermix (#1725120), added to respective primers (primers 9-14) for target genes and reactions were run in 96-well plates on a BioRad CFX qPCR instrument. Conditions used for amplification were 95°C, 2 min, 45 repeats of 95°C, 10 s and 60°C, 30 s. A melting curve was generated in 0.5°C intervals from 65 to 95°C. Cq values were calculated by the BioRad CFX Maestro software and

transcripts were normalized to a housekeeping gene (*GAPDH*). All measurements were performed in technical duplicate; technical duplicates were averaged to form a single biological replicate.

PERK Inhibition Assay

HEK293T cells were seeded in 6-well plates at 4 x 10^5 cells/well and transfected with GFP or SARS-CoV-2 nsp4-FT as previously described. GFP samples treated with either Tunicamycin (Tm, 1 μg/mL) or Tm and PERK inhibitor II GSK2656157 (PERKi, 1 μM) for 16h and nsp4-FT samples treated with PERK inhibitor II GSK2656157 (PERKi, 1 μM) for 36h prior to harvest. Samples were harvested 40 h post transfection by scraping and lysed in RIPA buffer as described in *Western Blot Analysis*.

Western Blot Analysis

HEK293T cells were lysed on ice in TNI buffer (50mM Tris pH 7.5, 150mM NaCl, 0.5% IGEPAL-CA-630) with Roche c0mplete EDTA-free protease inhibitor (#4693132001) for 15 minutes and then sonicated for 10 minutes in a water bath at room temperature. A549 cells were lysed on ice in RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 0.5% deoxycholate) with Roche c0mplete EDTA-free protease inhibitor (#4693132001) for 15 minutes. HEK293T cells for transmembrane protein panel (CFTR, hERG-HA, PMP22-Myc, nsp4-ST) and PERK inhibition assay were also lysed in RIPA buffer. Lysates were spun down at 21.1k xg for 15 minutes at 4°C. Samples were added to 6x Laemelli buffer (12% SDS, 125 mM Tris, pH 6.8, 20% glycerol, bromophenol blue, 100 mM DTT) and heated at 37°C for thirty minutes. The samples were then run on an SDS-PAGE gel and transferred to a PVDF membrane for Western blotting. M2 anti-FLAG (Sigma Aldrich, F1804), anti-KDEL (Enzo ADI-SPA-827-F), anti-PDIA4 (ProteinTech 14712-1-AP), THE anti-Strep II tag FITC (Genescript, A01736-100), anti-CFTR (217) (purified from B lymphocyte hybridoma cells as described in(32)), anti-HA.11 (Biolegend, 901513), anti-Myc (9B11) (Cell Signaling Technology, 2276S), anti-GFP (ThermoFisher, A6455), anti-ATF4 (Cell Signaling Technology, 11815) or anti-GAPDH (GeneTex, GTX627408) antibodies were used to probe Western blots at a 1:1000 dilution in TBS blocking buffer (0.1% Tween, 5% BSA, 0.1% Sodium Azide).

To measure CHOP protein levels, HEK293T cells were lysed on ice in 1% CHAPS lysis buffer (50 mM Tris H-Cl, 150 mM NaCl, 1mM EDTA, 1% CHAPS, 10 μM pepstatin, 0.2 mM phenylmethylsulfonyl fluoride, 0.2 μM soybean trypsin inhibitor, and 1mM DTT; pH 8.0) for 30 minutes. Lysates were spun down at 16,000 xg for 10 minutes at 4°C. Samples were added to 4x gel-loading buffer (50mM Tris/HCl (pH 6.8), 100mM DTT, 2% SDS, 0.1% Bromophenol Blue, 10% glycerol, 100 mM DTT) and heated at 37°C for thirty minutes. The samples were then run on an 13% SDS-PAGE gel and transferred to a nitrocellulose membrane for Western blotting. M2 anti-FLAG (Sigma Aldrich, F1804), anti-BiP #3177 (Cell Signaling Technology), anti-CHOP #7351, and anti-GAPDH #365062 (Santa Cruz Biotechnology Inc.), antibodies were used to probe Western blots in TBS blocking buffer (0.1% Tween, 4% BSA, 0.1% Sodium Azide). GAPDH signal was used to normalize band intensities. Protein expression on Western blots was quantified using ImageLab.

Proteomics Experimental Design

HEK293T UPR analysis combined 2-3 biological replicates into 1 individual MS run (3x GFP, 3x nsp3.1-FT+GFP, 3x nsp4-ST+GFP, 3x nsp3.1-FT+nsp4-ST, 2x orf8-ST+GFP, 2x orf8-ST+nsp3.1-FT). A549 UPR analysis combined 4 biological replicates into 1 individual MS run (4x GFP, 4x GFP+Tm, 4x nsp4-FT). UPR analysis of transmembrane protein panel expression in HEK293T cells combined 3-4 biological replicates into 1 MS run (3x GFP, 4x WT-CFTR, 3x hERG-HA, 3x PMP22-Myc, 3x nsp4-ST).

Sample Preparation for Mass Spectrometry

Samples were harvested and lysed as described in *Western Blot Analysis*. Protein concentration was quantified using 1x BioRad Protein Assay Dye (#5000006) and 20 µg of protein from each sample was prepared for mass spectrometry. Proteins were precipitated via mass spectrometry grade methanol:chloroform:water (in a 3:1:3 ratio) and washed three times with methanol. Each wash was followed by a 2-minute spin at 10,000xg at room temperature. Protein pellets were air dried 30-45 min and resuspended in 5 µL of 1% Rapigest SF (Waters #186002122). Resuspended proteins were diluted with 32.5 µL water and 10 µL 0.5 M HEPES (pH 8.0), then reduced with 0.5 µL of freshly made 0.5 M TCEP for 30 minutes at room temperature. Samples were then alkylated with 1 µL of fresh 0.5 M iodoacetamide (freshly made) for 30 minutes at room temperature in the dark and digested with 0.5 µg Pierce Trypsin/Lys-C (Thermo Fisher # A40007) overnight at 37°C shaking. Peptides were diluted to 60 µL with LC/MS-grade water and labeled using TMTpro labels (Thermo Scientific # A44520) for 1 h at room temperature. Labeling was quenched with the addition of fresh ammonium bicarbonate (0.4% v/v final) for 1 h at room temperature. Samples were then pooled, acidified to pH < 2.0 using formic acid, concentrated to 1/6th original volume via Speed-vac, and diluted back to the original volume with buffer A (95% water, 5% acetonitrile, 0.1% formic acid). Cleaved Rapigest products were removed by centrifugation at 17,000xg for 30 minutes and supernatant transferred to fresh tubes.

MudPIT LC-MS/MS Analysis

Alternating layers of 1.5cm Aqua 5 µm C18 resin (Phenomenex # 04A-4299), 1.5cm Luna 5 µm SCX resin (Phenomenex # 04A-4398), and 1.5cm Aqua 5 µm C18 resin were packed to make triphasic MudPIT columns as described previously(44). TMT-labeled samples (20 µg) were loaded onto the microcapillaries via a high-pressure chamber, followed by a 30-minute wash in buffer A (95% water, 5% acetonitrile, and 0.1% formic acid). The MudPIT columns were installed on the LC column switching valve and followed by a 20cm fused silica microcapillary column filled with Aqua C18, 3µm resin (Phenomenex # 04A-4311) ending in a laser-pulled tip. Columns were washed in the same way as the MudPIT capillaries prior to use. Liquid chromatography (Ultimate 3000 nanoLC system) was used to fractionate the peptides online and then analyzed via an Exploris480 mass spectrometer (Thermo Fisher). MudPIT runs were carried out by 10µL sequential injections of 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% buffer C (500mM ammonium acetate, 94.9% water, 5% acetonitrile, 0.1% formic acid), followed by a final injection of 90% C, 10% buffer B (99.9% acetonitrile, 0.1% formic acid v/v). Each injection was followed by a 130 min gradient using a flow rate of 500nL/min (0-6 min: 2% buffer B, 8 min: 5% B, 100 min: 35% B,

105min: 65% B, 106-113 min: 85% B, 113-130 min: 2% B). ESI was performed directly from the tip of the microcapillary column using a spray voltage of 2.2 kV, an ion transfer tube temperature of 275°C and an RF Lens of 40%. MS1 spectra were collected using a scan range of 400-1600 m/z, 120k resolution, AGC target of 300%, and automatic injection times. Data-dependent MS2 spectra were obtained using a monoisotopic peak selection mode: peptide, including charge state 2-7, TopSpeed method (3s cycle time), isolation window 0.4 m/z, HCD fragmentation using a normalized collision energy of 36%, 45k resolution, AGC target of 200%, automatic maximum injection times, and a dynamic exclusion (20 ppm window) set to 60s.

Peptide identification and quantification

Identification and quantification of peptides were performed in Proteome Discoverer 2.4 (Thermo Fisher) using the SwissProt human database (TaxID 9606, released 11/23/2019; 42,252 entries searched) with nsp3.1, nsp4, orf8, CFTR, hERG-HA, and PMP22-Myc sequences (6 entries) manually added (42,258 total entries searched). Searches were conducted with Sequest HT using the following parameters: trypsin cleavage (maximum two missed cleavages), minimum peptide length 6 AAs, precursor mass tolerance 20 ppm, fragment mass tolerance 0.02 Da, dynamic modifications of Met oxidation (+15.995 Da), protein N-terminal Met loss (−131.040 Da), and protein N-terminal acetylation (+42.011 Da), static modifications of TMTpro (+304.207 Da) at Lys, and N-termini and Cys carbamidomethylation (+57.021 Da). Peptide IDs were filtered using Percolator with an FDR target of 0.01. Proteins were filtered based on a 0.01 FDR, and protein groups were created according to a strict parsimony principle. TMT reporter ions were quantified considering unique and razor peptides, excluding peptides with co-isolation interference greater than 25%. Peptide abundances were normalized based on total peptide amounts in each channel, assuming similar levels of background. Protein quantification used all quantified peptides. Post-search filtering was carried out to include only proteins with two or more identified peptides. Common contaminants (e.g., albumin & keratins) were flagged and not included in downstream analyses.

Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD039797. All other necessary data are contained within the manuscript.

5.7   References

1.   WHO. Weekly Epidemiological Update on COVID-19. *World Heal. Organ.* 1;4 (2020).

2.   Xue, M. & Feng, L. The Role of Unfolded Protein Response in Coronavirus Infection and Its Implications for Drug Design. *Front. Microbiol.* **12**, (2021).

3.   Echavarría-Consuegra, L. *et al.* Manipulation of the unfolded protein response: A pharmacological strategy against coronavirus infection. *PLOS Pathog.* **17**, e1009644 (2021).

4.   Bechill, J., Chen, Z., Brewer, J. W. & Baker, S. C. Coronavirus Infection Modulates the Unfolded Protein

Response and Mediates Sustained Translational Repression. *J. Virol.* **82**, 4492–4501 (2008).

5. Walter, P. & Ron, D. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science (80-.).* **334**, 1081–1086 (2011).

6. Shoulders, M. D. *et al.* Stress-Independent Activation of XBP1s and/or ATF6 Reveals Three Functionally Diverse ER Proteostasis Environments. *Cell Rep.* **3**, 1279–1292 (2013).

7. Szegezdi, E., Logue, S. E., Gorman, A. M. & Samali, A. Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Rep.* **7**, 880–885 (2006).

8. Harding, H. P., Zhang, Y., Bertolotti, A., Zeng, H. & Ron, D. Perk Is Essential for Translational Regulation and Cell Survival during the Unfolded Protein Response. *Mol. Cell* **5**, 897–904 (2000).

9. Lee, A.-H., Iwakoshi, N. N. & Glimcher, L. H. XBP-1 Regulates a Subset of Endoplasmic Reticulum Resident Chaperone Genes in the Unfolded Protein Response. *Mol. Cell. Biol.* **23**, 7448–7459 (2003).

10. Haze, K., Yoshida, H., Yanagi, H., Yura, T. & Mori, K. Mammalian Transcription Factor ATF6 Is Synthesized as a Transmembrane Protein and Activated by Proteolysis in Response to Endoplasmic Reticulum Stress. *Mol. Biol. Cell* **10**, 3787–3799 (1999).

11. Adachi, Y. *et al.* ATF6 Is a Transcription Factor Specializing in the Regulation of Quality Control Proteins in the Endoplasmic Reticulum. *Cell Struct. Funct.* **33**, 75–89 (2008).

12. Aviner, R. & Frydman, J. Proteostasis in Viral Infection: Unfolding the Complex Virus–Chaperone Interplay. *Cold Spring Harb. Perspect. Biol.* **12**, a034090 (2020).

13. Fung, T. S. & Liu, D. X. Coronavirus infection, ER stress, apoptosis and innate immunity. *Front. Microbiol.* **5**, (2014).

14. Minakshi, R. *et al.* The SARS Coronavirus 3a Protein Causes Endoplasmic Reticulum Stress and Induces Ligand-Independent Downregulation of the Type 1 Interferon Receptor. *PLoS One* **4**, e8342 (2009).

15. Sung, S.-C., Chao, C.-Y., Jeng, K.-S., Yang, J.-Y. & Lai, M. M. C. The 8ab protein of SARS-CoV is a luminal ER membrane-associated protein and induces the activation of ATF6. *Virology* **387**, 402–413 (2009).

16. Angelini, M. M., Akhlaghpour, M., Neuman, B. W. & Buchmeier, M. J. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio* **4**, 1–10 (2013).

17. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res.* **149**, 58–74 (2018).

18. Shin, D. *et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662 (2020).

19. Oostra, M. *et al.* Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication. *J. Virol.* **81**, 12323–12336 (2007).

20. Beachboard, D. C., Anderson-Daniels, J. M. & Denison, M. R. Mutations across Murine Hepatitis Virus nsp4 Alter Virus Fitness and Membrane Modifications. *J. Virol.* **89**, 2080–2089 (2015).

21. Davies, J. P., Almasy, K. M., McDonald, E. F. & Plate, L. Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *ACS Infect. Dis.* **6**, 3174–3189 (2020).

22. Almasy, K. M., Davies, J. P. & Plate, L. Comparative Host Interactomes of the SARS-CoV-2 Nonstructural Protein 3 and Human Coronavirus Homologs. *Mol. Cell. Proteomics* **20**, 100120 (2021).

23. Plate, L. *et al.* Quantitative Interactome Proteomics Reveals a Molecular Basis for ATF6-Dependent Regulation of a Destabilized Amyloidogenic Protein. *Cell Chem. Biol.* **26**, 913-925.e4 (2019).

24. Plate, L. *et al.* Small molecule proteostasis regulators that reprogram the ER to reduce extracellular protein aggregation. *Elife* **5**, 1–26 (2016).

25. Simic, M. S. *et al.* Transient activation of the UPR ER is an essential step in the acquisition of pluripotency during reprogramming. *Sci. Adv.* **5**, 1–13 (2019).

26. Ye, J. *et al.* ER Stress Induces Cleavage of Membrane-Bound ATF6 by the Same Proteases that Process SREBPs. *Mol. Cell* **6**, 1355–1364 (2000).

27. Schröder, M. & Kaufman, R. J. THE MAMMALIAN UNFOLDED PROTEIN RESPONSE. *Annu. Rev. Biochem.* **74**, 739–789 (2005).

28. Marciniak, S. J. *et al.* CHOP induces death by promoting protein synthesis and oxidation in the stressed endoplasmic reticulum. *Genes Dev.* **18**, 3066–3077 (2004).

29. Axten, J. M. *et al.* Discovery of GSK2656157: An Optimized PERK Inhibitor Selected for Preclinical Development. *ACS Med. Chem. Lett.* **4**, 964–968 (2013).

30. Atkins, C. *et al.* Characterization of a Novel PERK Kinase Inhibitor with Antitumor and Antiangiogenic Activity. *Cancer Res.* **73**, 1993–2002 (2013).

31. Grandjean, J. M. D. *et al.* Deconvoluting Stress-Responsive Proteostasis Signaling Pathways for Pharmacologic Activation Using Targeted RNA Sequencing. *ACS Chem. Biol.* **14**, 784–795 (2019).

32. McDonald, E. F., Sabusap, C. M. P., Kim, M. & Plate, L. Distinct proteostasis states drive pharmacologic chaperone susceptibility for cystic fibrosis transmembrane conductance regulator misfolding mutants. *Mol. Biol. Cell* **33**, 1–17 (2022).

33. Kozek, K. A. *et al.* High-throughput discovery of trafficking-deficient variants in the cardiac potassium channel KV11.1. *Hear. Rhythm* **17**, 2180–2189 (2020).

34. Marinko, J. T., Carter, B. D. & Sanders, C. R. Direct relationship between increased expression and mistrafficking of the Charcot–Marie–Tooth–associated protein PMP22. *J. Biol. Chem.* **295**, 11963–11970 (2020).

35. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).

36. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8**, 937–940 (2011).

37. Volmer, R., van der Ploeg, K. & Ron, D. Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. *Proc. Natl. Acad. Sci.* **110**, 4628–4633 (2013).

38. Edagawa, M. *et al.* Role of Activating Transcription Factor 3 (ATF3) in Endoplasmic Reticulum (ER) Stress-induced Sensitization of p53-deficient Human Colon Cancer Cells to Tumor Necrosis Factor (TNF)-related Apoptosis-inducing Ligand (TRAIL)-mediated Apoptosis through Up-regulation of Death Receptor 5 (DR5) by Zerumbone and Celecoxib. *J. Biol. Chem.* **289**, 21544–21561 (2014).

39.   Sims, A. C. *et al.* Unfolded Protein Response Inhibition Reduces Middle East Respiratory Syndrome Coronavirus-Induced Acute Lung Injury. *MBio* **12**, 1–22 (2021).

40.   Peña, J. & Harris, E. Dengue virus modulates the unfolded protein response in a time-dependent manner. *J. Biol. Chem.* **286**, 14226–14236 (2011).

41.   Almasy, K. M. *et al.* Small-molecule endoplasmic reticulum proteostasis regulator acts as a broad-spectrum inhibitor of dengue and Zika virus infections. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–12 (2021).

42.   Kolpikova, E. P. *et al.* IRE1α Promotes Zika Virus Infection via XBP1. *Viruses* **12**, 278 (2020).

43.   Sabusap, C. M. *et al.* Analysis of cystic fibrosis–associated P67L CFTR illustrates barriers to personalized therapeutics for orphan diseases. *JCI Insight* **1**, 1–10 (2016).

44.   Fonslow, B. R. *et al.* Single-Step Inline Hydroxyapatite Enrichment Facilitates Identification and Quantitation of Phosphopeptides from Mass-Limited Proteomes with MudPIT. *J. Proteome Res.* **11**, 2697–2709 (2012).

45.   Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

CHAPTER 6

CONCLUSIONS

6.1 Dissertation Summary

In this dissertation, I explore the mechanisms through which coronaviruses (CoVs) take over a host cell for infection, with a focus on the interface between nonstructural proteins (nsps) and the ER proteostasis network. This work centers on the role of nsp2, nsp3, and nsp4 in mediating CoV infection through virus-host protein-protein interactions (PPIs) with both individual proteostasis factors and broader stress response systems.

In **Chapter 1**, I introduced the emergence of SARS-CoV-2 and the overall impact of CoVs on human health, before providing a brief synopsis on the role of nsps in CoV replication and the basics of the ER proteostasis network. I then review some of the more recent findings in the scientific literature related to the intersection of nsps and the host proteostasis network, highlighting the use of AP-MS in discovering interactions and subsequent virological and biochemical functional characterization. Lastly, I laid out the pressing questions that my work aimed to address, namely: what are the conserved or unique protein interactions between SARS-CoV-2 and homologous nsps? Are these interactions supporting or suppressing infection? How do nsps influence individual proteostasis interactors? And how do nsps tune larger stress response systems, like the Unfolded Protein Response (UPR)?

Chapter 2 showcases the rapid deployment of quantitative proteomics to comparatively profile the interactomes of nsp2 and nsp4 homologs in the early months of the COVID-19 pandemic. In a collaborative effort with Katherine M. Almasy, we performed comparative AP-MS analysis of tagged nsp2/4 homologs from SARS-CoV-2, SARS-CoV, and hCoV-OC43 expressed in human cells. We identified clusters of shared interactors for SARS-1 and SARS-2, such as mitochondrial biogenesis and ER calcium transfer factors, as well as unique interactors for SARS-1, including a translation repressor complex. For nsp4 homologs, we found conserved glycosylation machinery and protein quality control factors, as well as large E3-ubiquitin ligase complex for SARS-CoV. Among both data sets there were several proteins localized to ER-mitochondria contact sites (MAMs) and I showed that both nsps, and in particular nsp4, can localize to these MAMs using subcellular fractionation.

Chapter 3 applies our comparative interactomics workflow to nsp3, a protein missing from other CoV interactome maps. We expanded the panel of viruses to include MERS-CoV and hCoV-229E homologs and divide nsp3 into three fragments (nsp3.1, 3.2, 3.3) for comparative interactome profiling. Our AP-MS yielded large interactome networks, including shared factors involved in ERAD and transcriptions factors, and unique clusters, such as nuclear importins (229E nsp3.2) and ribosomal RNA processing factors (MERS nsp3.2). We found a shared interaction between SARS-2, SARS-1, and OC43 nsp3.1 with the UPR sensor and transcription factor ATF6 and postulated this interaction may alter ATF6 activity. Using quantitative proteomics, I found that SARS-CoV-2 nsp3.1 suppressed protein levels of genes induced by ATF6, both under basal conditions and when ATF6 is pharmacologically activated.

In **Chapter 4**, I follow-up on the functional role of conserved nsp2 and nsp4 interactors we identified by AP-MS in Chapter 2. Using the model betacoronavirus MHV, I applied an RNAi screen against priority interactors to categorize them

as either pro- or anti-viral. I found that several conserved proteostasis interactors are pro-viral, most prominently the nsp2-interactor Malectin (MLEC), which is a glycoprotein quality control factor. Applying AP-MS to analyze nsp2 and MLEC interactions during CoV infection, I show that nsp2 interacted with several MLEC-associated components of the OST complex and that MHV infection drastically remodels\ed the interactome of MLEC. Furthermore, MLEC promoted early infection events like viral protein biogenesis and this seems to be mediated through the role of MLEC in glycoprotein quality control.

Lastly, **Chapter 5** details a collaborative project with my undergraduate mentee, Athira Sivadas, and others, in understanding the role of nsps in modulating the UPR at a systems level. This work builds off our finding that SARS-CoV-2 suppresses the ATF6 branch of the UPR by profiling how nsp4, another transmembrane protein involved in DMV formation, modulates the UPR. Applying our quantitative proteomics method, we find that nsp4 activates both an adaptive arm of the UPR (ATF6) and the pro-apoptotic arm (PERK). Further, co-expression of nsp3 and nsp4 retains ATF6 activation but suppresses PERK activation and we find that this trend is also found in a published SARS-CoV-2 infection proteomics study, where there is initially early PERK activation that is quickly tamped down. We predict that this activity is a viral tuning strategy to increase chaperones and folding capacity (ATF6) and suppress host translation (PERK), while avoiding potential apoptosis driven by chronic PERK activation.

6.2   Impact

Pandemic Preparedness
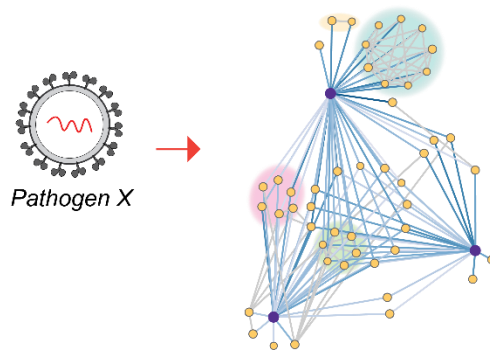
Figure 6-1. AP-MS workflows can assist in pandemic preparedness.

The work presented herein has contributed to the fields of proteomics, proteostasis biology, and virology through application of cutting-edge mass spectrometry methods and classical virology techniques. In response to the COVID-19 pandemic, Katherine Almasy and I worked quickly to identify protein interactors of SARS-CoV-2 proteins, generating the comparative proteomics data by mid-April, 2020 and posting an online preprint article detailing our findings by mid-July, 2020. Our rich nsp interactome datasets have been mined and used as basis for hypotheses-driven functional studies by other researchers around the globe, as reflected by over 120 citations in the scientific literature. We have revealed conserved interactors of CoV homologs, aiding our understanding of the infection strategies utilized across the CoV family. The

comparative interactomics profiling of nsp3 filled an important gap in the existing literature, as it was absent from several large-scale AP-MS studies at the time (1–3). From a methods perspective, our streamlined comparative AP-MS pipeline can be applied to rapidly assess the protein interactome of emerging viruses (**Figure 6-1**). This can save valuable time in identifying potential host dependencies for candidate anti-viral targets and aid in pandemic preparedness for future viral outbreaks.

This work also unveils a previously unreported CoV-host dependency factor with the identification of the nsp2-interactor, Malectin (MLEC). MLEC is a still relatively understudied component of the glycoprotein quality control pathway, having only been discovered in 2008 (4). Much remains to be explored regarding the role of MLEC in glycobiology. For instance, how does MLEC coordinate the refolding of diglucosylated proteins? If the client protein is properly folded, how is the information relayed to release the client from MLEC? Can GluII access the diglucosyl moiety while the client protein is bound to MLEC? And what signal(s) communicate that a retained client protein should be shuttled to ERAD for degradation? The interactomics and glycoproteomics data in Chapter 4 should be a rich source for answering these questions related to the fundamental biology of MLEC.

Additionally, this dissertation furthers our knowledge on the role of nsp2 in CoV infection, which has long had an ill-defined function in replication. Our comparative AP-MS reveals a potential role for mitochondria-associated ER membrane (MAMs) factors with nsp2. I further present a potential intersection between nsp2 and the glycoprotein biogenesis pathway, which has not been previously known. As a conserved interactor of hCoV nsp2 homologs, MLEC also represents a potential target for host-directed therapies against multiple CoVs.

This work also helps to fill a gap in our understanding of how CoVs induce and modulate host stress response pathways like the Unfolded Protein Response (UPR). We show that SARS-CoV-2 nsps can act in concert to tune the UPR in a manner that is potentially advantageous for replication. As multiple studies have shown that small molecule modulation of the UPR can be antiviral (5–7), our findings offer further clarity on the CoV-UPR interface that may be exploited for improved therapeutic intervention. Additionally, our technique of using quantitative proteomics to measure the functional activation of UPR branches is highly sensitive and comprehensive, enabling detection of pathway changes at the proteome level that may be overlooked by transcript-level methods. This methodology can be applied to other biological fields where nuanced UPR activation is relevant, such as in cancer, aging, and neurobiology (8).

6.3   Limitations

As with all research endeavors, there are some limitations to the work presented here. One of the main limitations of the comparative AP-MS workflow is that we use expression of single viral proteins to profile interactomes. Viral infection is a complex, dynamic conglomerate of host interactions that can dramatically alter a cell, and single bait protein overexpression omits this context. There is some irony in this paradox, as we are most interested in studying the mechanisms by which viruses overtake a host cell. However, this practice continues to be extremely common in the proteomics field, as a reductionist approach is logistically more accessible, faster, and versatile, particular in comparative applications like ours. I do partially overcome some of these limitations with the GFP-nsp2 MHV AP-MS experiments presented in Chapter 4.

This is only possible because of the genetic tolerance of nsp2 for fusions which different CoV research groups have exploited for creating tool viruses (9,10).

An accompanying limitation is the use of HEK293T cells as the human cell model for our comparative proteomics. This was used because of ease of transfection to different affinity-tagged constructs, which enabled rapid proteome profiling. However, this brings the caveat that the mapped interactome may not be identical to the lung cells that are natively infected in humans, as HEK293T cells are immortalized and have undergone genetic changes. An immortalized lung cell line, such as A549 cells, or primary human airway epithelial (HAE) cells would be more relevant, though these bring increasing levels of difficulty for experimentation.

Continuing with limitations in model systems, the work presented in Chapter 4 is all performed with a mouse CoV (MHV) in murine cells. MHV has long served as a workhorse model virus for studying CoVs but has limits on the transferability of observations to human CoVs. For instance, nsps in MHV have different glycosylation patterns, additional or missing domains in nsp3, and low sequence conservation of nsp2. Additionally, there are differences in mouse vs. human cellular proteomes. With that acknowledged, there is a large repertoire in the literature of phenotypes and mechanisms being fully or in part conserved with SARS-CoV, MERS-CoV, and SARS-CoV-2. The relevance of interactors I identified, like MLEC, necessitates further testing in human CoV infection.

6.4   Future Directions

There remains much additional work to do and many more research questions to pursue. One of the primary directions is moving towards more infection-relevant models for AP-MS analysis to enhance the power of interactomics. If feasible, performing IPs on viral proteins with endogenous antibodies would be ideal, though this requires a validated, IP-grade antibody for every target of interest. Incorporating an affinity-tag to be expressed endogenously by the virus would circumvent this drawback, though finding tolerated sites of insertion in a polyprotein is typically challenging. Additionally, performing comparative studies with multiple, highly pathogenic CoVs in BSL-3 brings many more logistical challenges. A potential solution is to use viral replicons as an intermediary model. Replicons can encode the full suite of nonstructural proteins as well as other accessory and structural proteins without producing viable virions. These can be transfected or electroporated into cells, as well as be stably expressed in cells in an inducible manner. This strategy incorporates far more of the ensemble of the viral proteome and improves the relevance of interactors. Because replicons are encoded on plasmid vectors, they are also more readily amenable to modification, such as tag or mutation insertion, and panels of replicons can be assembled to compare various virus interactomes. There are several SARS-CoV-2 replicon systems that can generate single-round infectious virions, enabling studies of viral assembly and entry processes in BSL-2+ conditions (11–14). These replicon systems are well-suited for interactomics studies and are being actively pursued by others in the Plate group, which will improve the relevance of future work.

It will also be interesting to determine how mutations observed in SARS-CoV-2 nsps during the progression of the COVID-19 pandemic affect nsp-host interactions. There have been several recent reports showcasing the impact of point mutations in nsps on SARS-CoV-2 infection, such as nsp2 (15), nsp3 (16), nsp4 (17), and nsp6 (13). These mutations likely

affect other virus-host PPIs and further exploration of their effects, as is being pursued by others in the Plate group, will likely lend insight into the roles of nsps in replication.

Building off the findings in Chapter 4, it will be necessary to determine whether MLEC and other conserved proteostasis interactors are dependency-factors for human CoVs like SARS-CoV-2, SARS-CoV, and MERS-CoV. Many genome-wide CRISPR screens of SARS-CoV-2 and other CoVs have been performed (18–24), however some of the limitations of large-scale genetic screens, such as quality of guide RNAs, warrant a direct testing of priority interactors. In the absence of immediate access to BSL-3 facilities, the aforementioned replicon system can help assess this question.
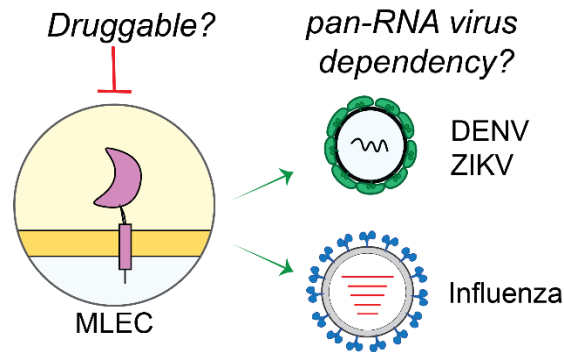
Figure 6-2. MLEC as a host-targeted, pan-RNA virus antiviral target.

Furthermore, ER proteostasis systems tend to be widely used by RNA viruses for replication and so it is of interest to test whether MLEC is a dependency-factor for other pathogens such as flaviviruses. As human-infecting flaviviruses like dengue (DENV) and Zika virus (ZIKV) can be cultured in BSL-2(+) conditions, these are accessible experiments to perform. If MLEC or other interactors are widely used by RNA viruses, these will be exciting targets to develop inhibitors against as host-directed therapies (**Figure 6-2**).

The role of MLEC in CoV infection is just beginning to be explored and further characterization of this interaction is needed. Moving forward, it would be beneficial to develop a method to track CoV glycoprotein biogenesis, folding, and interactions with MLEC and other protein quality control factors during an infection time-course (**Figure 6-3**). This is an area where using chemical biology tools, such as non-canonical amino acid incorporation, could be applied to pulse-label viral proteins and track their progression through the proteostasis network via time-resolved interactomics. This methodology would also be useful in elucidating other mechanisms of basic CoV biology, such as the coordination of proteostasis factor in co- and post-translationally processing a ~6,600 amino acid-long polyprotein. Additionally, it will be of interest to determine the localization and structural determinants of these interactions and how these characteristics contribute to viral infection.

Lastly, there were several pro-viral, proteostasis interactors identified in Chapter 4 that warrant further investigation. These include the ER-phagy receptor, CCPG1, the ER membrane protein complex component, MMGT1, and the cargo receptor, SURF4. Each may be the basis for whole separate projects and should shed light both on fundamental principles of the ER proteostasis network and CoV infection mechanisms.
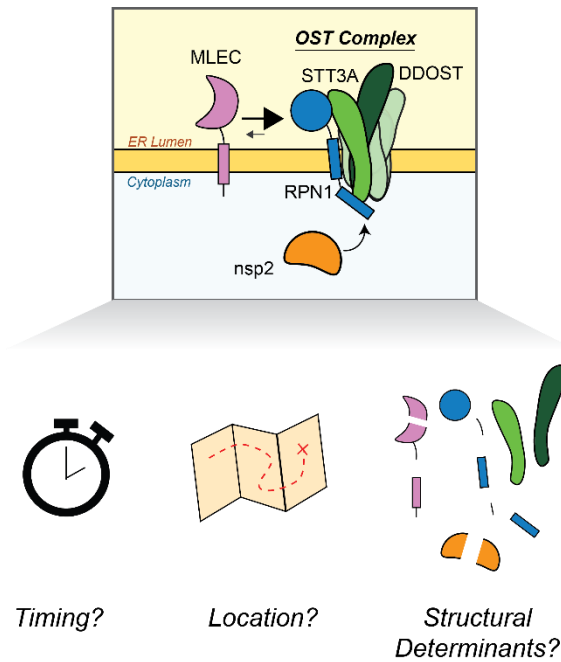
Figure 6-3. Future work to define the nsp2-MLEC-OST interface.

In closing, this dissertation deepens our knowledge of the fundamental basic biology of CoV infection, identifies conserved host dependencies that can be exploited for therapeutic development, and develops a quantitative viral proteomics that can further pandemic preparedness. My hope is that this work can be continued by future graduate students and scientists, as there are many more leads and hypotheses to be pursued.

6.5   References

1. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583,** 459–468 (2020).

2. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370,** eabe9403 (2020).

3. Stukalov, A. *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594,** 246–252 (2021).

4. Schallus, T. *et al.* Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N-Glycosylation. *Mol Biol Cell* **19,** 3404–3414 (2008).

5. Almasy, K. M. *et al.* Small-molecule endoplasmic reticulum proteostasis regulator acts as a broad-spectrum inhibitor of dengue and Zika virus infections. *Proc National Acad Sci* **118,** e2012209118 (2021).

6. Sims, A. C. *et al.* Unfolded Protein Response Inhibition Reduces Middle East Respiratory Syndrome Coronavirus-Induced Acute Lung Injury. *Mbio* **12,** e01572-21 (2021).

7. Echavarría-Consuegra, L. *et al.* Manipulation of the unfolded protein response: A pharmacological strategy against coronavirus infection. *Plos Pathog* **17,** e1009644 (2021).

8. Hetz, C., Zhang, K. & Kaufman, R. J. Mechanisms, regulation and functions of the unfolded protein response. *Nat Rev Mol Cell Bio* **21,** 421–438 (2020).

9. Freeman, M. C., Graham, R. L., Lu, X., Peek, C. T. & Denison, M. R. Coronavirus Replicase-Reporter Fusions Provide Quantitative Analysis of Replication and Replication Complex Formation. *J Virol* **88,** 5319–5327 (2014).

10. V'kovski, P. *et al.* Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8,** e42037 (2019).

11. He, X. *et al.* Generation of SARS-CoV-2 reporter replicon for high-throughput antiviral screening and testing. *Proc National Acad Sci* **118,** e2025866118 (2021).

12. Tanaka, T. *et al.* Establishment of a stable SARS-CoV-2 replicon system for application in high-throughput screening. *Antivir Res* **199,** 105268 (2022).

13. Taha, T. Y. *et al.* Rapid assembly of SARS-CoV-2 genomes reveals attenuation of the Omicron BA.1 variant through NSP6. *Nat Commun* **14,** 2308 (2023).

14. Ricardo-Lax, I. *et al.* Replication and single-cycle delivery of SARS-CoV-2 replicons. *Science* **374,** 1099–1106 (2021).

15. Gupta, M. *et al.* CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. *bioRxiv* 2021.05.10.443524 (2021). doi:10.1101/2021.05.10.443524

16. Taha, T. Y. *et al.* A single inactivating amino acid change in the SARS-CoV-2 NSP3 Mac1 domain attenuates viral replication and pathogenesis in vivo. (2023). doi:10.1101/2023.04.18.537104

17. Lin, X. *et al.* The NSP4 T492I mutation increases SARS-CoV-2 infectivity by altering non-structural protein cleavage. *Cell Host Microbe* **31,** 1170-1184.e7 (2023).

18. Hoffmann, H.-H. *et al.* Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors. *Cell Host Microbe* **29,** 267-280.e5 (2021).

19. Wang, R. *et al.* Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses. *Cell* **184,** 106-119.e14 (2021).

20. Schneider, W. M. *et al.* Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks. *Cell* **184,** 120-132.e14 (2021).

21. Biering, S. B. *et al.* Genome-wide bidirectional CRISPR screens identify mucins as host factors modulating SARS-CoV-2 infection. *Nat Genet* **54,** 1078–1089 (2022).

22. Daniloski, Z. *et al.* Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells. *Cell* **184,** 92-105.e16 (2021).

23. Wei, J. *et al.* Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection. *Cell* **184,** 76-91.e13 (2021).

24. Kratzel, A. *et al.* A genome-wide CRISPR screen identifies interactors of the autophagy pathway as conserved coronavirus targets. *Plos Biol* **19,** e3001490 (2021).

Supplementary figures for Chapter 2.

A

```
Wuhan_nsp2/1-638    1  AYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQLDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP  96
SARS_nsp2/1-638     1  AVTRYVDNNFCGPDGYPLDCIKDFLARAGKSMCTLSEQLDYIESKRGVYCCRDHEHEIAWFTERSDKSYEHQTPFEIKSAKKFDTFKGECPKFVFP  96

Wuhan_nsp2/1-638   97  LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTGDFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPA  192
SARS_nsp2/1-638    97  LNSKVKVIQPRVEKKKTEGFMGRIRSVYPVASPQECNNMHLSTLMKCNHGDEVSWQTCDFLKATCEHCGTENLVIEGPTTCGYLPTNAVVKMPCPA  192

Wuhan_nsp2/1-638  193  CHNSEVGPEHSLAEYHNESGLKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNLLEILQKEKVNINIVGDFK  288
SARS_nsp2/1-638   193  CQDPEIGPEHSVADYHNHSNIETRLRKGGRTRCFGGCVFAYVGCYNKRAYWVPRASADIGSGHTGITGDNVETLNEDLLEILSRERVNINIVGDFH  288

Wuhan_nsp2/1-638  289  LNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGNFKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQK  384
SARS_nsp2/1-638   289  LNEEVAIILASFSASTSAFIDTIKSLDYKSFKTIVESCGNYKVTKGKPVKGAWNIGQQRSVLTPLCGFPSQAAGVIRSIFARTLDAANHSIPDLQR  384

Wuhan_nsp2/1-638  385  AAITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKLKPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIV  480
SARS_nsp2/1-638   385  AAVTILDGISEQSLRLVDAMVYTSDLLTNSVIIMAYVTGGLVQQTSQWLSNLLGTTVEKLRPIFEWIEAKLSAGVEFLKDAWEILKFLIITGVFDIV  480

Wuhan_nsp2/1-638  481  GGQIVTCAKEIKESVQTFFKLVNKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEIIFLEGETLPTEVLTEEV  576
SARS_nsp2/1-638   481  KGGIQVASDNIKDCVKCFIDVVNKALEMCIDQVTIAGAKLRSLNLGEVFIAQSKGLYRQCIRGKEQLQLLMPLKAPKEVTFLEGDSHDITVLTSEEV  576

Wuhan_nsp2/1-638  577  VLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEKYCALAPNMMVTNNTFTLKGG                                      638
SARS_nsp2/1-638   577  VLKNGELEALETPVDSFTNGAIVGTPVCVNGLMLLEIKDKEQYCALSPGLLATNNVFRLKGG                                      638
```

B

```
OC43_nsp4/1-496     1  --------AVFSYFVYVCFVLSLVCFIGLWCLMPTYTVHKSDFQLPVYASYKVLDNGVIRDVSVEDVCFANKFEQFDQWYESTFGLSYYSNSMACPIVVAVI  94
Wuhan_nsp4/1-501    1  -KIVNNWLKQLIKVTLVFLFVAAIFY----LITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFANKHADFDTWFSQRGG--SYTNDKACPLIAAVI  96
SARS_nsp4/1-500     1  -KIVSTCFKLMLKATLLCVLAALVCY----IVMPVHTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFDAWFSQRGG--SYKNDKSCPVVAAII  95

OC43_nsp4/1-496    95  DQDFGSTVFNVPTKVLRY-GYHVLHFITHALSADGVQCYTPHSQISYSNFYASGCVLSSACTMFTMADGSPQPYCYTEGLMQNASLYSSLVPHVRYNLANAK  195
Wuhan_nsp4/1-501   97  TREVGFVVPGLPGTILRTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAYESLRPDTRYVLMD-G  197
SARS_nsp4/1-500    96  TREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNICYTPSKLIEYSDFATSACVLAAECTIFKDAMGKPVPYCYDTNLLEGSISYSELRPDTRYVLMD-G  196

OC43_nsp4/1-496   196  GFIRFPEVLREGLVRIVRTRSMSYCRVGLCEEADEGICFNFNGSWVLNNDYYRSLPGTFCGRDVFDLIYQLFKGLAGPVDFLALTASSIAGAILAVIVVLVF  297
Wuhan_nsp4/1-501  198  SIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLA  299
SARS_nsp4/1-500   197  SIIQFPNTYLEGSVRVVTTFDAEYCRHGTCERSEVGICLSTSGRWVLNNEHYRALSGVFCGVDAMNLIANIFTPLVGPVGALDVSASVVAGGIIAILVTCAA  298

OC43_nsp4/1-496   298  YYLIKLKRAFGDYTSVVFVNVIVWCVNFMMLFVFQVVPILSCVYAICYFYATLYFPSEISVIMHLQWLVMYGTIMPLWFCLLYIAVVVSNHAFWVFSYCRKL  399
Wuhan_nsp4/1-501  300  YYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKR  401
SARS_nsp4/1-500   299  YYFMKFRRVFGEYNHVVAANALLFLMSFTILCLVPAYSFLPGVYSVFYLYLTFYFTNDVSFLAHLQWFAMFSPIVPFWITAIYVFCISLKHCHWFFNNYLRK  400

OC43_nsp4/1-496   400  G--TSVRSDGTFEEMALTTFMITKDSYCKLKNS--LSDVAFNRYLSLYNKYRYYSGKMDTAAYREAACSQLAKAMDTFTNNNGSDVLYQPPTASVSTSFLQ  496
Wuhan_nsp4/1-501  402  RVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKALNDFSN-SGSDVLYQPPQTSITSAVLQ  501
SARS_nsp4/1-500   401  RVMFNGVTFSTFEEAALCTFLLNKEMYLKLRSETLLPLTQYNRYLALYNKYKYFSGALDTTSYREAACCHLAKALNDFSN-SGADVLYQPPQTSITSAVLQ  500
```

C SARS-CoV-2 nsp4

MATTRQVVNV VTTKIALKGG KIVNNWLKQL IKVTLVFLFV AAIFYLITPV HVMSKHTDFS SEIIGYKAID GGVTRDIAST DTCFANKHAD FDTWFSQRGG SYTNDKACPL
IAAVITREVG FVVPGLPGTI LRTTNGDFLH FLPRVFSAVG NICYTPSKLI EYTDFATSAC VLAAECTIFK DASGKPVPYC YDTNVLEGSV AYESLRPDTR YVLMDGSIIQ
FPNTYLEGSV RVVTTFDSEY CRHGTCERSE AGVCVSTSGR WVLNNDYYRS LPGVFCGVDA VNLLTNMFTP LIQPIGALDI SASIVAGGIV AIVVTCLAYY FMRFRRAFGE
YSHVVAFNTL LFLMSFTVLC LTPVYSFLPG VYSVIYLYLT FYLTNDVSFL AHIQWMVMFT PLVPFWITIA YIICISTKHF YWFFSNYLKR RVVFNGVSFS TFEEAALCTF
LLNKEMYLKL RSDVLLPLTQ YNRYLALYNK YKYFSGAMDT TSYREAACCH LAKALNDFSN SGSDVLYQPP QTSITSAVLL ESRGPDYKDD DDK

D SARS-CoV-1 nsp4

MATTRQVVNV ITTKISLKGG KIVSTCFKLM LKATLLCVLA ALVCYIVMPV HTLSIHDGYT NEIIGYKAIQ DGVTRDIIST DDCFANKHAG FDAWFSQRGG SYKNDKSCPV
VAAIITREIG FIVPGLPGTV LRAINGDFLH FLPRVFSAVG NICYTPSKLI EYSDFATSAC VLAAECTIFK DAMGKPVPYC YDTNLLEGSI SYSELRPDTR YVLMDGSIIQ
FPNTYLEGSV RVVTTFDAEY CRHGTCERSE VGICLSTSGR WVLNNEHYRA LSGVFCGVDA MNLIANIFTP LVQPVGALDV SASVVAGGII AILVTCAAYY FMKFRRVFGE
YNHVVAANAL LFLMSFTIIC LVPAYSFLPG VYSVFYLYLT FYFTNDVSFL AHLQWFAMFS PIVPFWITAI YVFCISLKHC HWFFNNYLRK RVMFNGVTFS TFEEAALCTF
LLNKEMYLKL RSETLLPLTQ YNRYLALYNK YKYFSGALDT TSYREAACCH LAKALNDFSN SGADVLYQPP QTSITSAVLL ESRGPDYKDD DDK

E OC43 nsp4

MNKQMANVSV LTTPFSLKGG AVFSYFVYVC FVLSLVCFIG LWCLMPTYTV HKSDFQLPVY ASYKVLDNGV IRDVSVEDVC FANKFEQFDQ WYESTFGLSY YSNSMACPIV
VAVIDQDFGS TVFNVPTKVL RYGYHVLHFI THALSADGVQ CYTPHSQISY SNFYASGCVL SSACTMFTMA DGSPQPYCYT EGLMQNASLY SSLVPHVRYN LANAKGFIRF
PEVLREGLVR IVRTRSMSYC RVGLCEEADE GICFNFNGSW VLNNDYYRSL PGTFCGRDVF DLIYQLFKGL AQPVDFLALT ASSIAGAILA VIVVLVFYYL IKLKRAFGDY
TSVVFVNVIV WCVNFMMLFV FQVYPILSCV YAICYFYATL YFPSEISVIM HLQWLVMYGT IMPLWFCLLY IAVVVSNHAF WVFSYCRKLG TSVRSDGTFE EMALTTFMIT
KDSYCKLKNS LSDVAFNRYL SLYNKYRYYS GKMDTAAYRE AACSQLAKAM DTFTNNNGSD VLYQPPTASV STSFLLESRG PDYKDDDDK

Figure 7-1. Amino acid sequence comparison and MS sequence coverage of CoV non- structural protein homologs.

A. Amino acid sequence alignment of nsp2 homologs from SARS-CoV-1 (SARS) and SARS-CoV-2 (Wuhan). Identical or similar residues are highlighted and color scheme corresponds to amino acid properties.

B. Amino acid sequence alignment of nsp4 homologs from SARS-CoV-1 (SARS), SARS-CoV-2 (Wuhan) and hCoV-OC43. Identical or similar residues are highlighted and color scheme corresponds to amino acid properties.

C. Tandem MS sequence coverage of SARS-CoV-2 (Wuhan) nsp4. Detected peptides are indicated in green.

D. Tandem MS sequence coverage of SARS-CoV-1 (Urbani) nsp4. Detected peptides are indicated in green.

E. Tandem MS sequence coverage of hCoV-OC43 nsp4. Detected peptides are indicated in green.
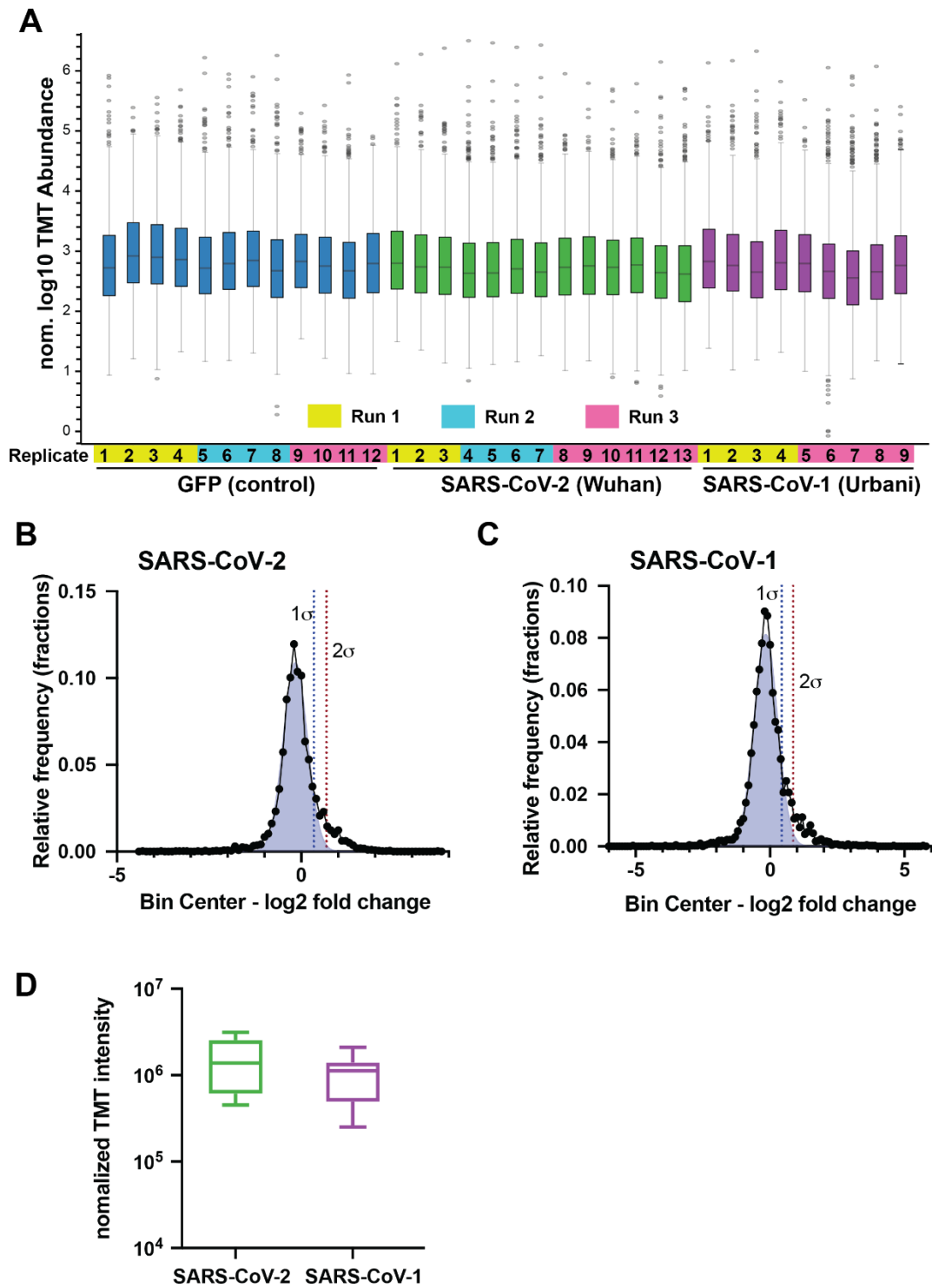
Figure 7-2. TMT normalization and filtering of nsp2 interactors.

    A. Normalized log10 TMT abundances of all proteins in the affinity purification samples expressing GFP control or nsp2 homologs as bait proteins. Pairing of individual samples into three separate mass spectrometry runs is indicated by the color bars

B. Histogram of log2 fold change enrichment of proteins in SARS-CoV-2 nsp2 affinity purificatlllllllllkkion samples compared to GFP controls. A gaussian non-least square fit of the distribution is shown in light blue. The standard deviation of the distribution (σ) is indicated and 1 σ and 2 σ (dotted lines) were used for the variable cutoffs to define medium- and high-confidence interactors, 1 σ = 0.5.

C. Same as in (B), for SARS-CoV-1 nsp2, with 1 σ = 0.43.

D. Normalized TMT intensities comparing the abundances of SARS-CoV-2 and SARS-CoV-1 nsp2 homologs in the replicate samples.
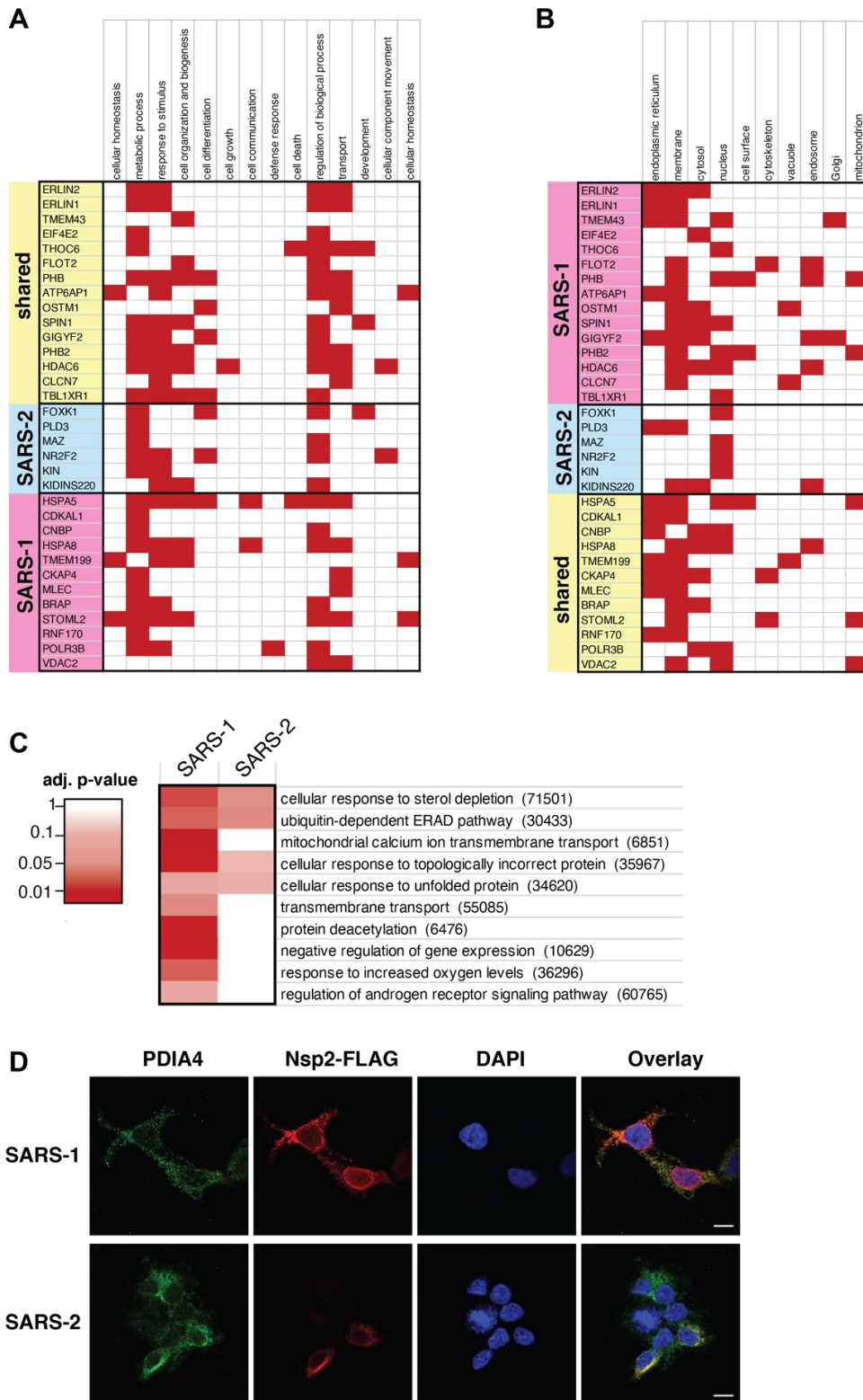
Figure 7-3. Gene ontology (GO) pathway analysis of interactors and cellular localization of nsp2 homologs.

GO terms associated with the individual interactors of nsp2. Terms were assigned in the Protein Annotation node in Proteome Discoverer 2.4. Proteins were grouped according to the hierarchical clustering in Fig. 3B to distinguish shared and distinct interactors of SARS- CoV-1 and SARS-CoV-2 nsp2.

A. GO terms for biological processes.
B. GO terms for cellular components.
C. Comparisons of pathways identified in gene set enrichment analysis of interactors of nsp2 homologs. Gene set enrichment analysis of high-and medium-confidence interactors was performed in Enrichr. GO terms for biological processes with adjusted p-values < 0.1 were included in the analysis and filtered manually to remove redundant terms containing similar genes. Non-redundant pathways are shown and color scheme indicates confidence of enrichment as represented by adjusted p-values of the Enrichr analysis.
D. Confocal microscopy images of nsp2 homologs expressed in HEK293T cells, stained for PDIA4 (ER marker, green), FLAG-nsp2(red), and DAPI (blue). Scale bar is 10 µm.
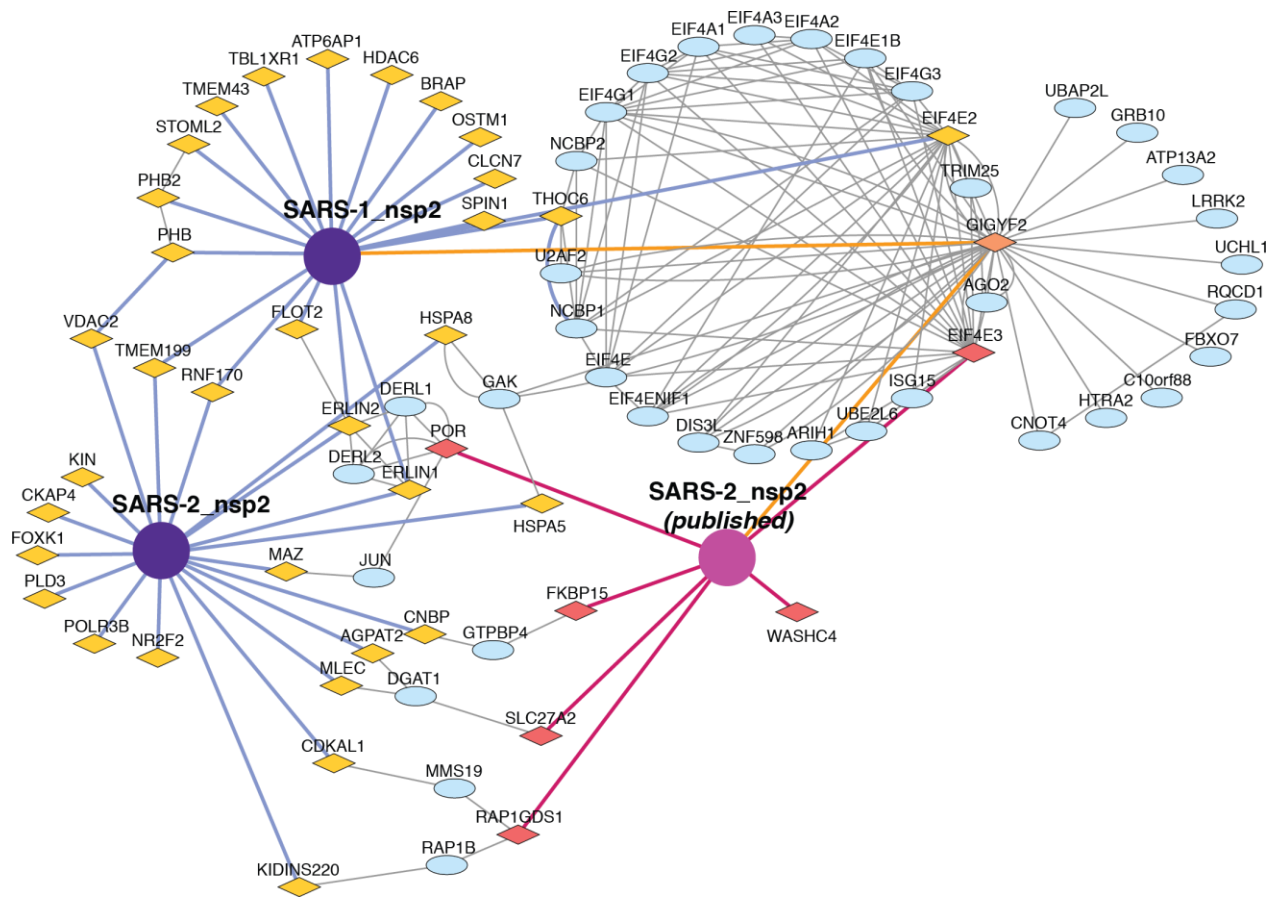
Figure 7-4. Nsp2 interactome overlap with published dataset.
Extended and overlapping interactors between our nsp2 dataset and previously published interactors of SARS-CoV-2 nsp2(1). Nsp2 proteins are shown as purple circles (our dataset) or pink circle (published dataset). Previously published primary interactors are shown as dark pink diamonds, novel primary interactors identified in this study are shown as yellow diamonds, overlapping primary interactors (GIGYF2) are shown as orange diamonds, and overlapping secondary interactors scraped from the STRING database are shown as blue ellipses. Previously identified primary interactions are shown with red edges, novel primary interactions identified in this study are shown with blue edges, overlapping primary interactors are shown with orange edges, and secondary interactions scraped from the STRING database are shown as grey edges between nodes.
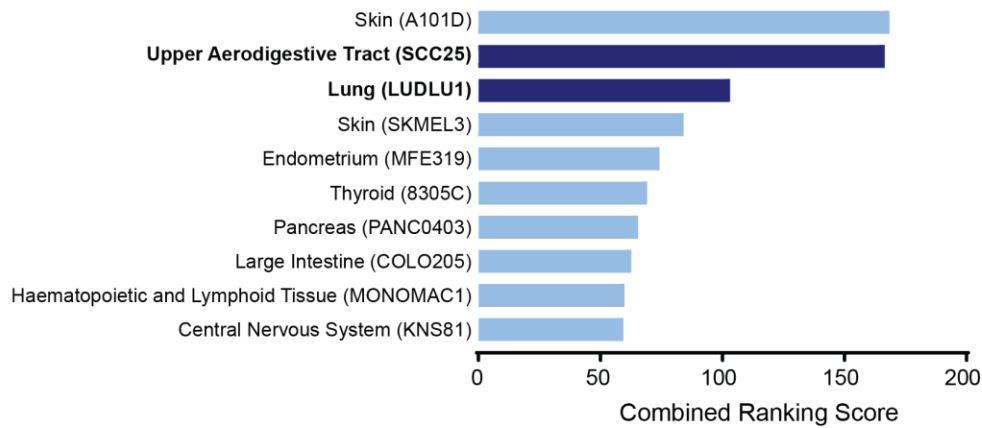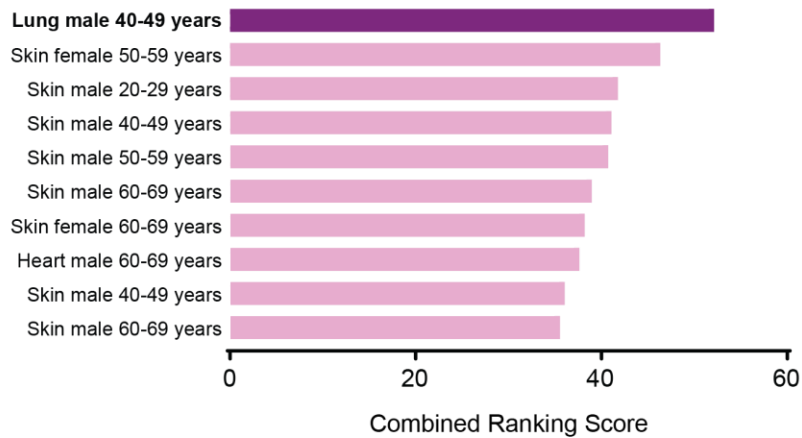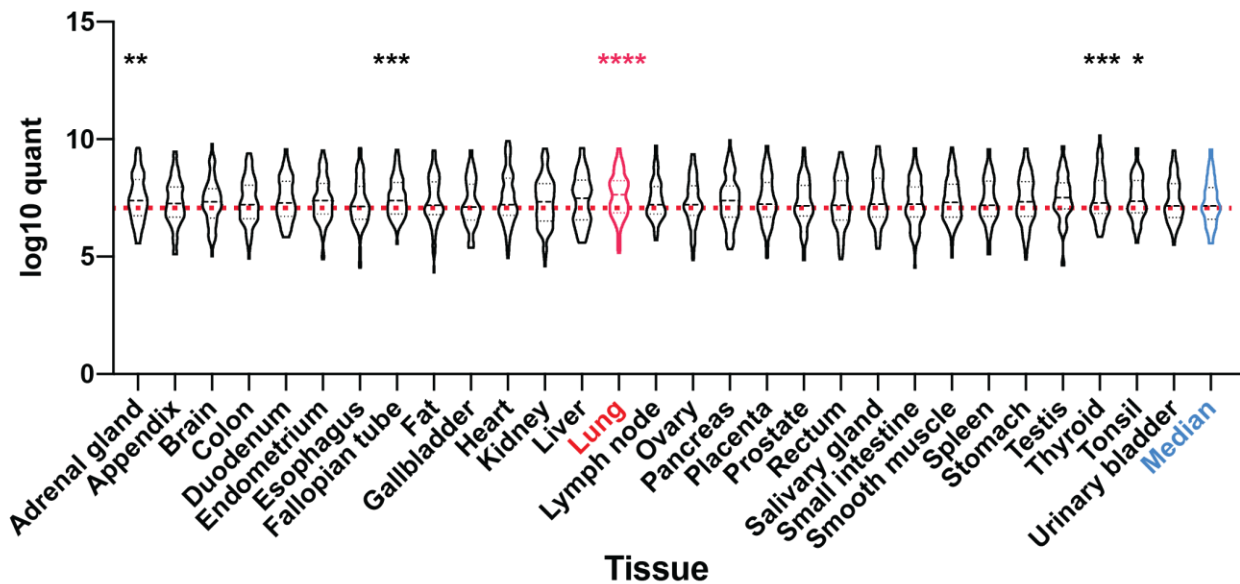
Figure 7-5. Tissue-specific expression of nsp2 and nsp4 interactors.
Nsp2 and nsp4 interactors were cross-referenced with existing data sets on protein expression in cell lines and tissue-samples(2–4) to identify enrichment levels of interactors.
  A. Top ten cell lines enriched for interactors in the CCLE data set(2) based on analysis of combined ranking score using Enrichr(5).

140

B. Top ten tissue samples enriched for interactors in the GTeX data set(3) based on analysis of combined ranking score using Enrichr(5).

C. Violin plots showing $\log_{10}$ protein quantification values for the nsp2 and nsp4 interactions in 29 healthy human tissues based on the dataset published by Wang et al.[4] The median quantification across all tissues is shown on the right. Significance of expression difference was assessed in Graphpad Prism by mixed-effects analysis with the Geisser-Greenhouse correction and Holm-Sidak's multiple comparison test for each tissue compared to the median. * $p<0.05$; **$p<0.01$, ***$p<0.001$; ****$p<0.0001$
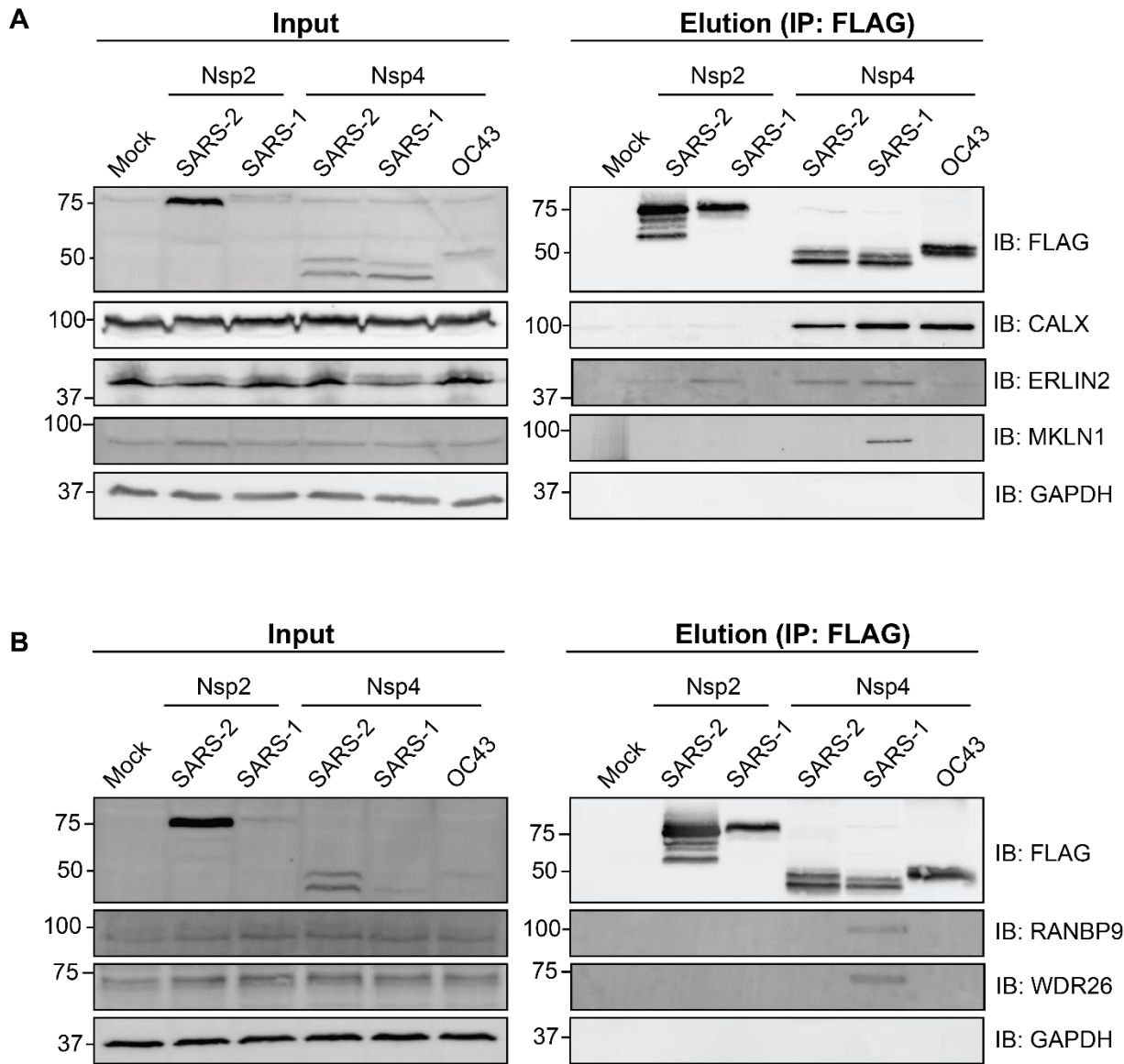
Figure 7-6. Co-immunoprecipitation of MAMs and proteostasis factors with CoV non-structural proteins.
Lysates from HEK293T cells transfected with respective viral homologs of nsp2 or nsp2 were immunopurified using anti-FLAG sepharose beads.

A. Blots probed for MAMs factors (CANX, ERLIN2). Representative blots for input and co-immunoprecipitations of are shown, n=3.

B. Blots probed for CTLH E3 ligase complex interactors (MKLN1, RANBP9, WDR26). Same as in (A).
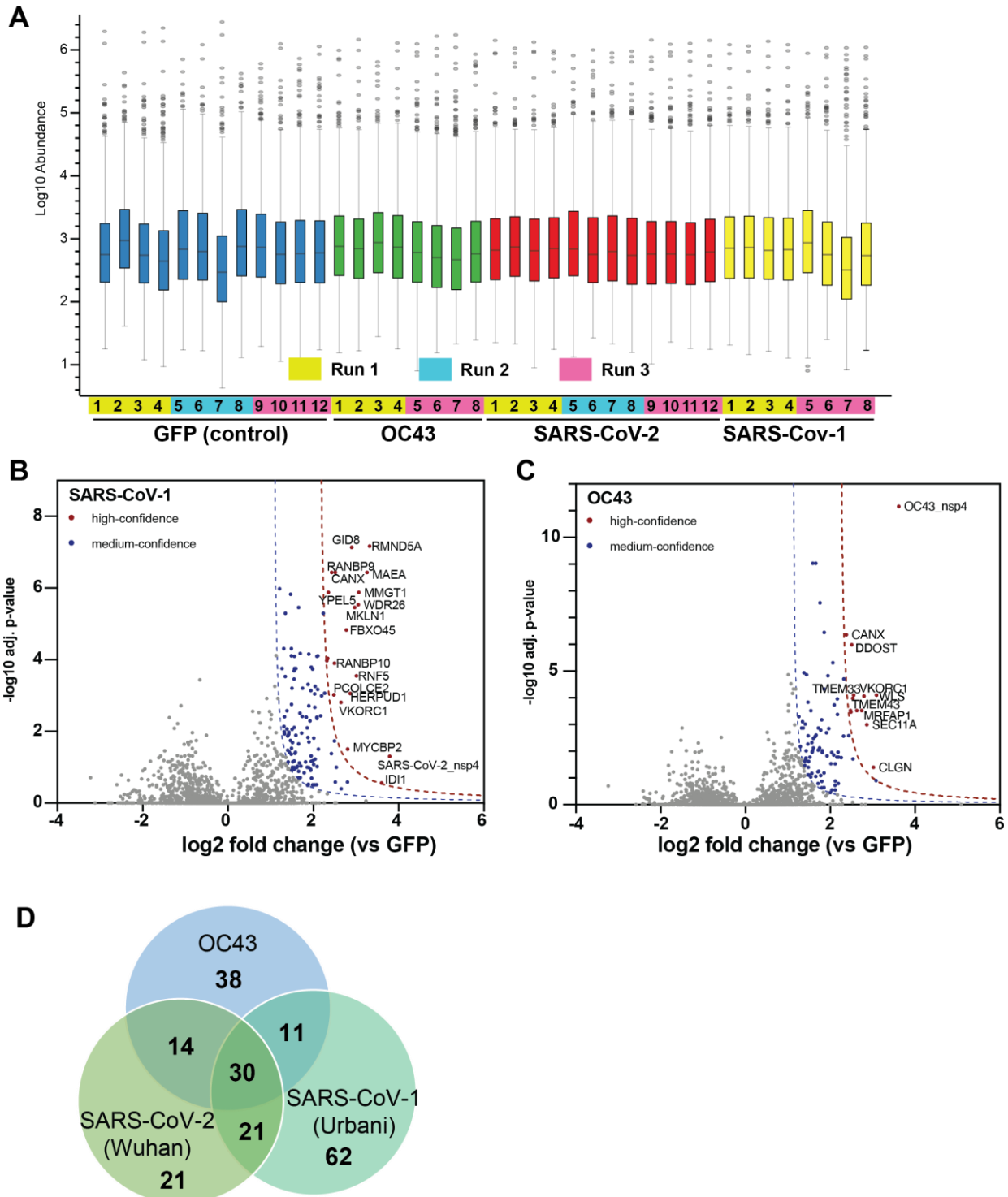
Figure 7-7. TMT normalization and filtering of nsp4 interactors.

A. Normalized log10 TMT abundances of all proteins in the affinity purification samples expressing GFP control or nsp4 homologs as bait proteins. Pairing of individual samples into three separate mass spectrometry runs is indicated by the color bars.

B. Volcano plot of SARS-CoV-1 nsp4 interactors to identify medium- and high-confidence interactors. Plotted are log2 TMT intensity fold changes for proteins between nsp2 bait channels and GFP mock transfections versus -log10 adjusted p-values. Curves for the variable cutoffs used to define high-confidence (red) or medium confidence (blue) interactors are shown. $1\sigma = 0.4$

C. Same as in (B), for hCoV-OC43 nsp4. $1\sigma = 0.40$

D. Venn diagram comparing medium-confidence interactors of nsp4 homologs.
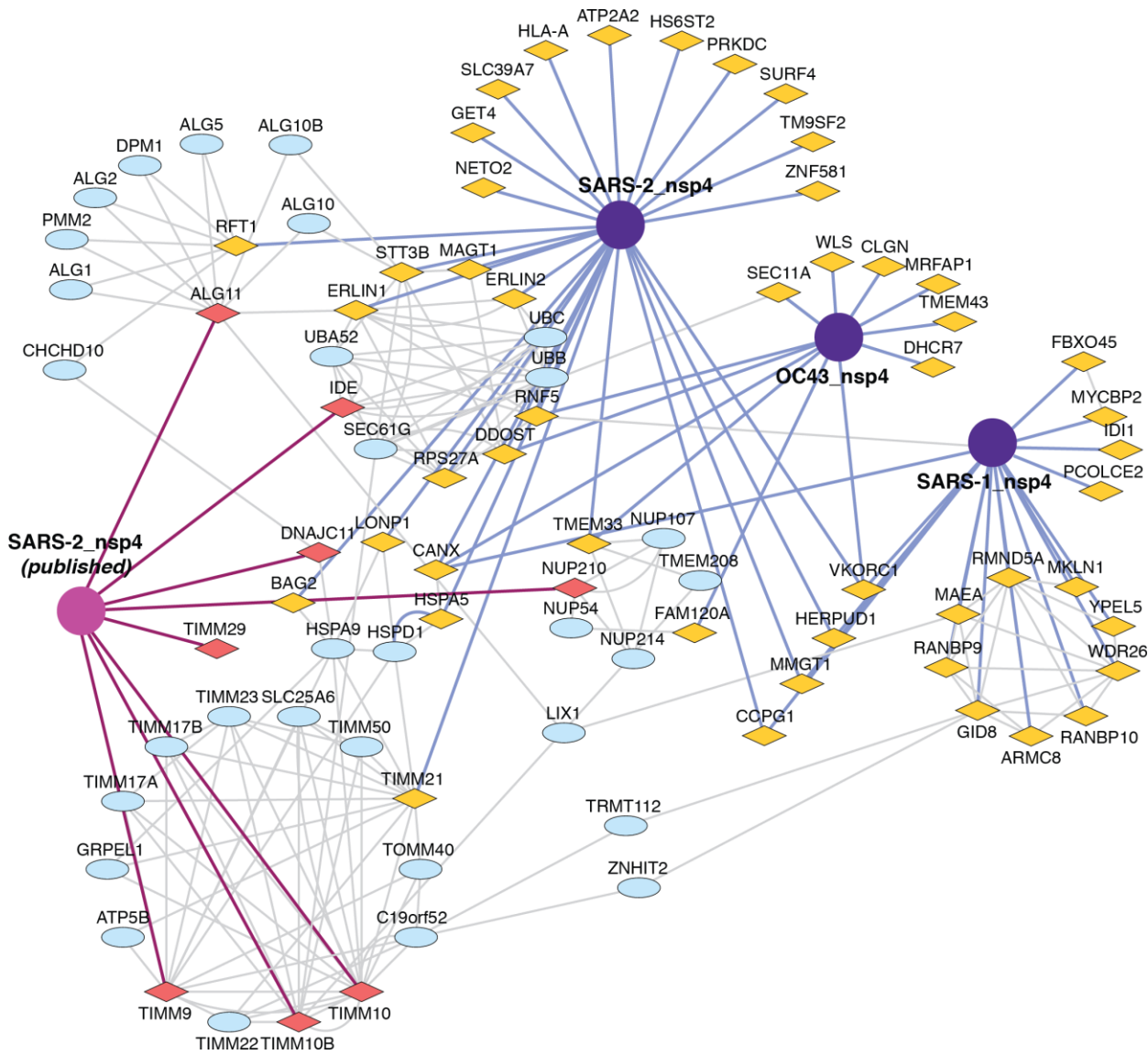
Figure 7-8. Nsp4 interactome overlap with published dataset.

Extended and overlapping interactors between our nsp4 dataset and previously published interactors of SARS-CoV-2 nsp4(1). Nsp4 proteins are shown as purple circles (our dataset) or pink circle (published dataset). Previously published primary interactors are shown as dark pink diamonds, novel primary interactors identified in this study are shown as yellow diamonds, overlapping primary interactors are shown as orange diamonds, and overlapping secondary interactors scraped from the STRING database are shown as blue ellipses. Previously identified primary interactions are shown with red edges, novel primary interactions identified in this study are shown with blue edges, and secondary interactions scraped from the STRING database are shown as grey edges between nodes. The extended overlapping interactome reveals each Nsp4 protein uniquely plugs into clusters of proteins involved in the same pathway, which appear as circles of nodes such as the TIMM pathway involved in protein import into the mitochondria on the bottom left.
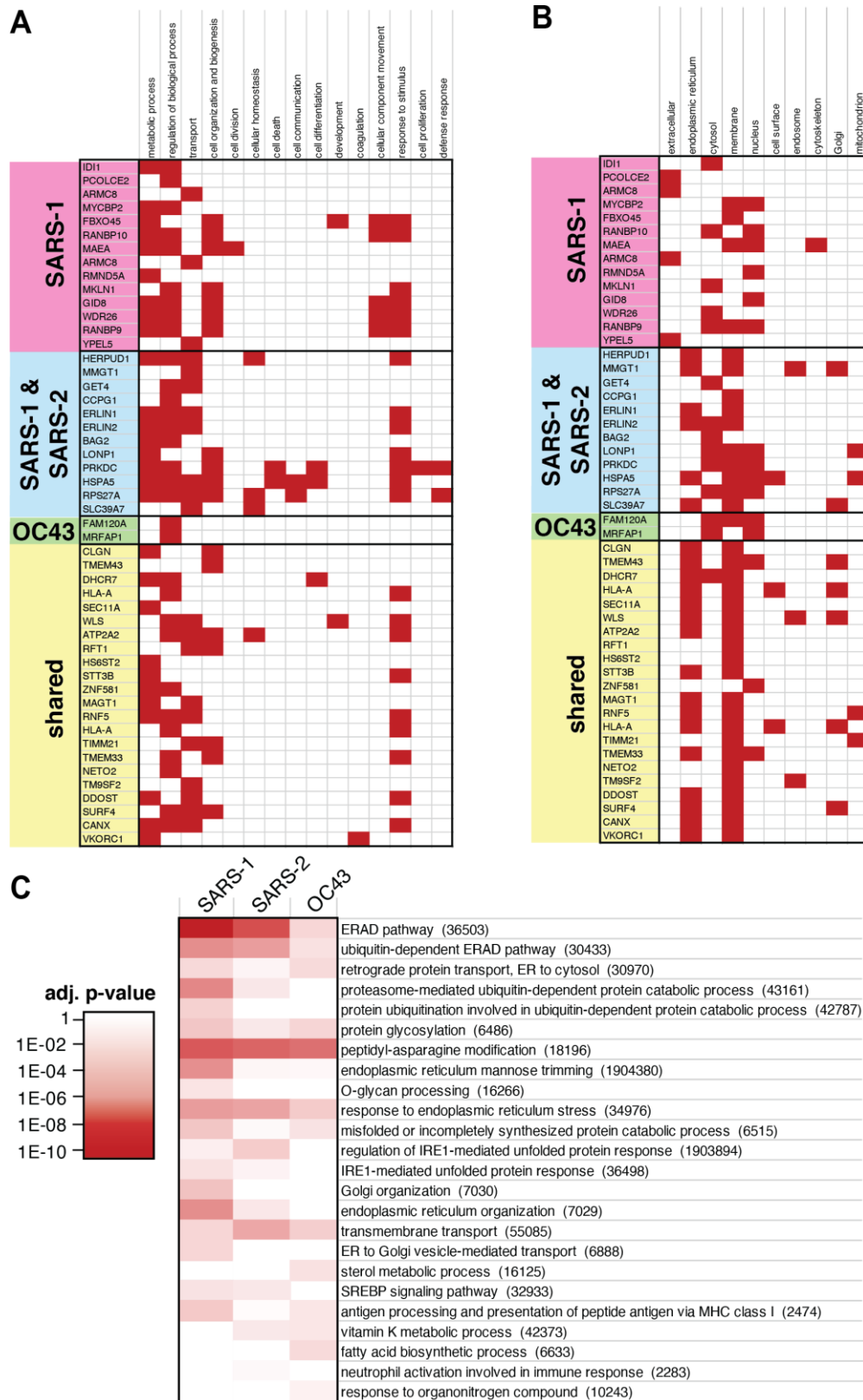
Figure 7-9. Gene ontology (GO) pathway analysis of nsp4 homolog interactors.

GO terms associated with the individual interactors of nsp4. Terms were assigned in the Protein Annotation node in Proteome Discoverer 2.4. Proteins were grouped according to the hierarchical clustering in Fig. 4C to distinguish shared and distinct interactors of SARS- CoV-1, SARS-CoV-2, and OC43 nsp4.

  A. GO terms for biological processes.
  B. GO terms for cellular components.
  C. Comparisons of pathways identified in gene set enrichment analysis of interactors of nsp4 homologs. Gene set enrichment analysis of high-and medium-confidence interactors was performed in Enrichr. GO terms for biological processes with adjusted p-values < 0.1 were included in the analysis and filtered manually to remove redundant terms containing similar genes. Non-redundant pathways are shown and color scheme indicates confidence of enrichment as represented by adjusted p-values of the Enrichr analysis.
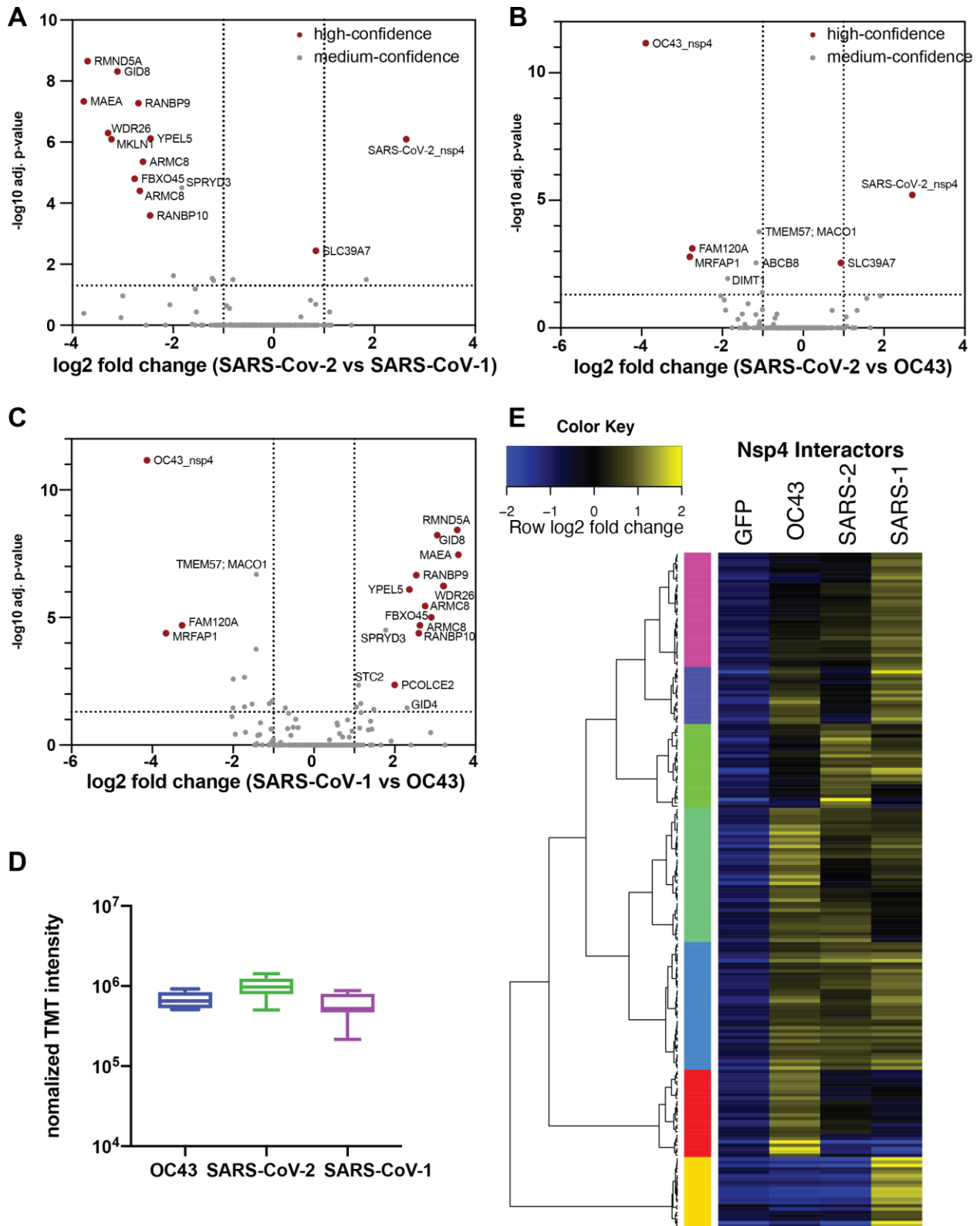
Figure 7-10. Comparative analysis of nsp4 homologs.

148

Volcano plot comparing interactions between nsp4 homolog from SARS-CoV-1, SARS- CoV-2, and OC43. Only high- and medium confidence interactors of nsp4 are shown and high confidence interactors are highlighted in red.

A. Comparison of SARS-CoV-1 and SARS-CoV-2.
B. Comparison of SARS-CoV-2 and OC43.
C. Comparison of SARS-CoV-1 and OC43.
D. Normalized TMT intensities comparing the abundances of SARS-CoV-2, SARS-CoV-1, and OC43 nsp4 homologs in the replicate affinity purification samples.
E. Heatmap of high- and medium-confidence nsp4 homolog interactors compared to GFP control. log2 fold change is color-coded and centered by row (blue low, yellow high enrichment). Hierarchical clustering using Ward's method shown on the left was carried out on euclideanan distances of log2 fold changes scaled by row.
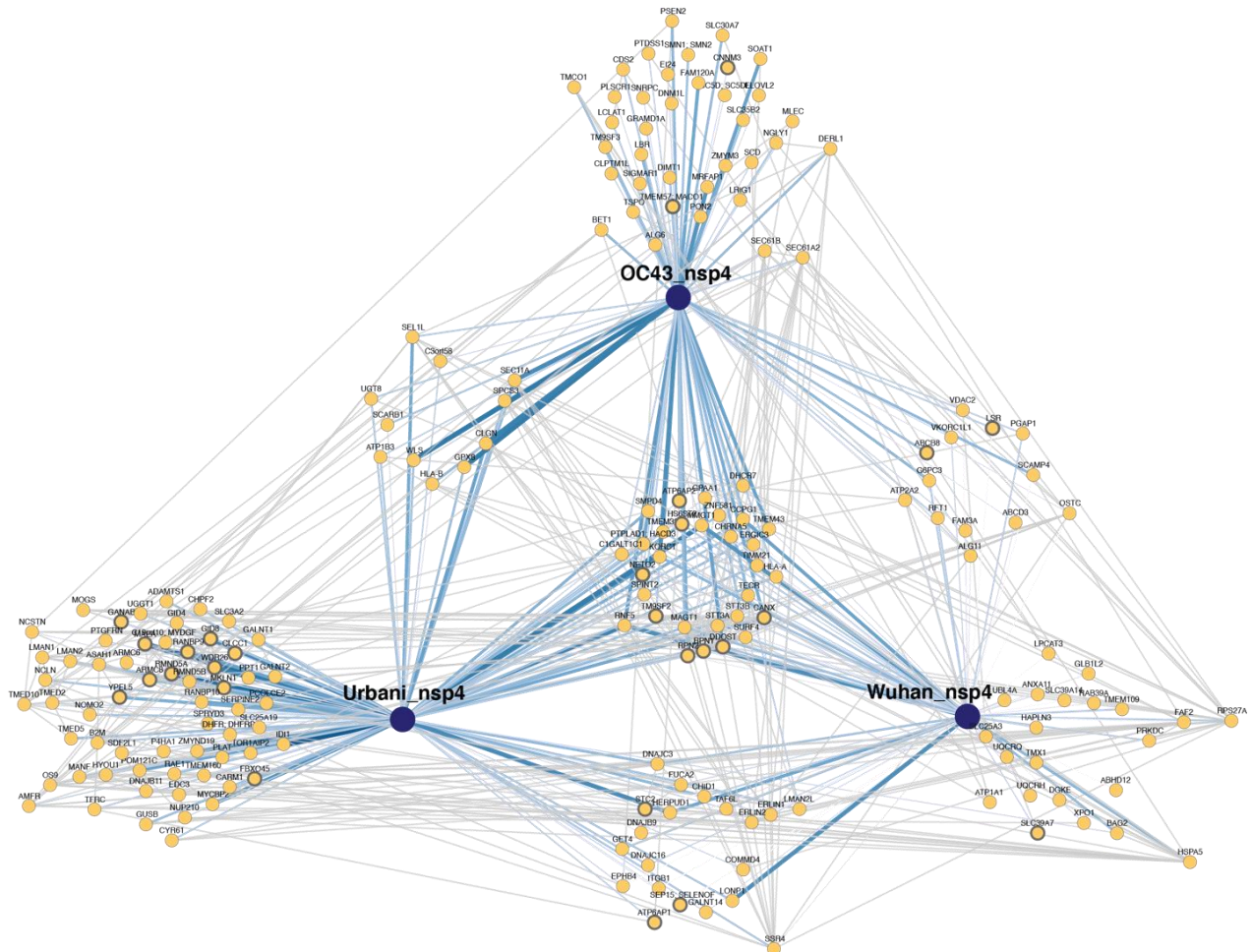
Figure 7-11. Network map of high- and medium-confidence nsp4 homolog interactors.
PPI network map of high- and medium-confidence interactors of nsp4 homolog. Blue lines indicate viral-host PPIs, where line width corresponds to fold enrichment compared to the GFP control. Grey lines indicate annotated host-host PPIs in STRING (score > 0.75).
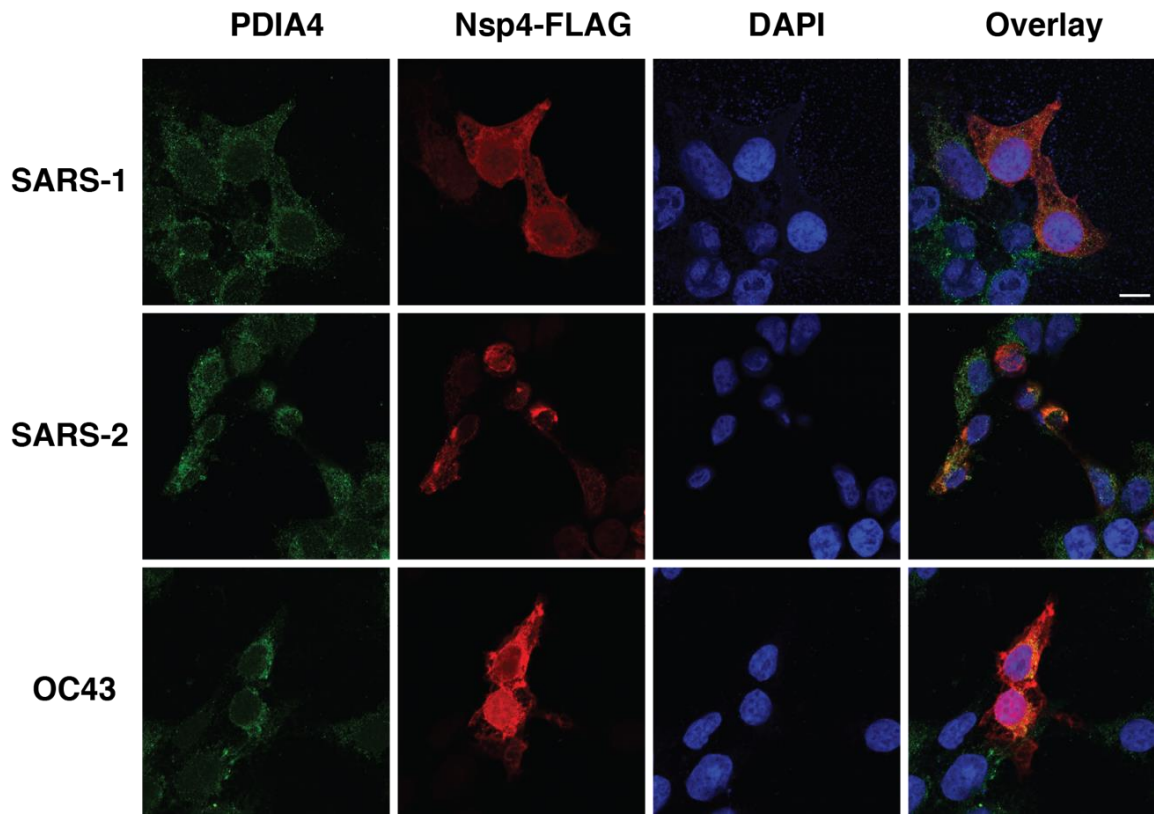
Figure 7-12. Subcellular localization of nsp4 homologs.
Confocal immunofluorescence microscopy images of nsp4 homologs expressed in HEK293T cells, stained for PDIA4 (ER marker, green), FLAG-tag (red), and DAPI (blue). Scale bar is 10 µm.
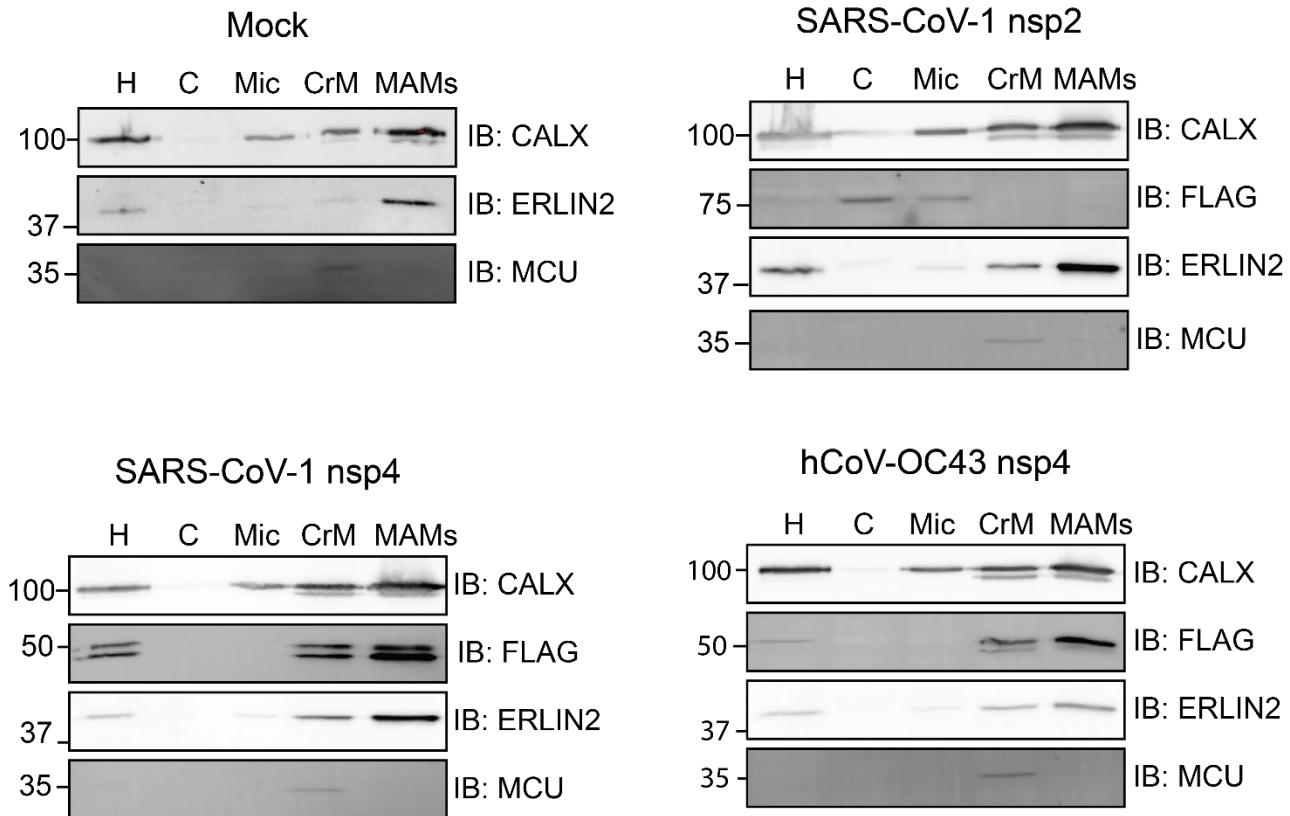
Figure 7-13. Localization of SARS-CoV-1 nsp2, nps4, and hCoV-OC43 nsp4 to MAMs.
SARS-CoV-1 nsp2, nsp4, hCoV-OC43, or mock transfected HEK293T cells were fractionated to determine localization of viral proteins to MAMs. Homogenate (H), cytosol (C), microsome (Mic), crude mitochondria (CrM), and MAMs fractions were probed via Western blot for subcellular markers (CALX and ERLIN2 for MAMs; MCU for mitochondria) and viral proteins (FLAG). (Mock, n=2; all others, n=1)
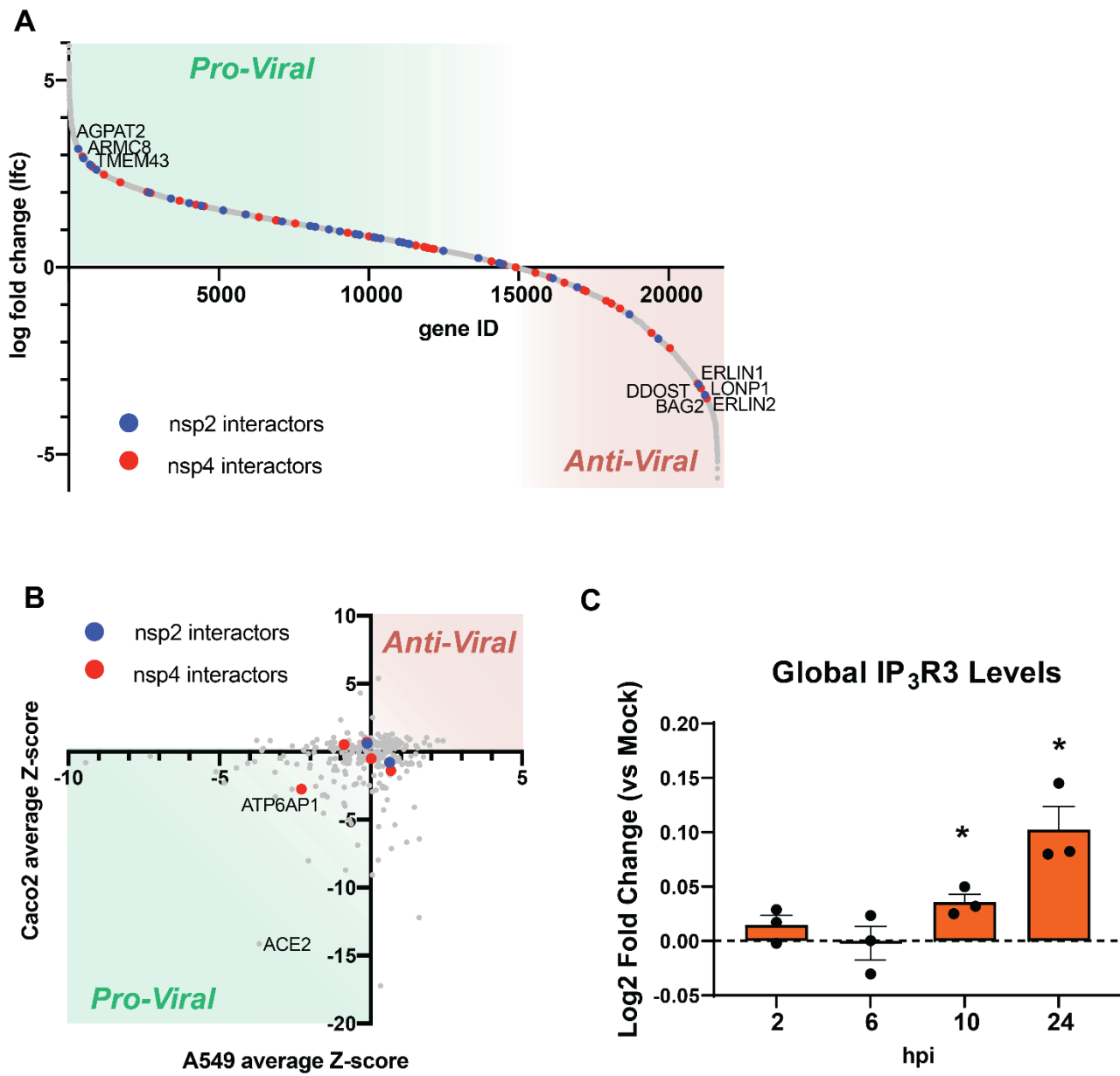
Figure 7-14. Functional genomic screens of interactors and global IP$_3$R3 protein levels in SARS-CoV-2 infection.

A. Sorted enrichment of sgRNAs (lfc) in the genome-wide CRISPR screen of the GeCKO V2 library in A549-Cas9-ACE2 lung cancer cells infected with SARS-CoV-2 published by Heaton et al. (6) Interactors of nsp2 and nsp4 found in the dataset are highlighted in blue and red respectively.

B. Correlation of Z-scores from functional genomics screens in A549-ACE2 and Caco-2 cells infected with SARS-Cov-2 published by Gordon et al. (7) The A549 cells were transduced with siRNA pools of prioritized SARS-CoV interactors and Caco2 cells contained CRISPR knockouts of the prioritized interactors. Z-score < 0 represents decreased infectivity upon knockdown, while Z-score > 0 represents increased infectivity. Interactors of nsp2 and nsp4 found in the dataset are highlighted in blue and red respectively.

C. Changes in the global proteome of Caco-2 cells during SARS-CoV-2 infection were measured via quantitative proteomics up to 24 hpi by Bojkova et al. (8) Mass spectrometry data was mined for log$_2$-fold changes in global IP$_3$R3 levels during infection. Significance was tested using unpaired two-sided students t-tests between infected and mock samples, *p<0.05.

References

1. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

2.   Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387-402.e16 (2020).

3.   Jiang, L. *et al.* A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269-283.e19 (2020).

4.   Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).

5.   Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

6.   Heaton, B. E. *et al.* SRSF protein kinases 1 and 2 are essential host factors for human coronaviruses including SARS-CoV-2. *bioRxiv* 2020.08.14.251207 (2020) doi:10.1101/2020.08.14.251207.

7.   Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science (80-.).* **22**, eabe9403 (2020).

8.   Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).

Supplementary figures for Chapter 3.

Figure 8-1. Multiple-sequence alignment and domain organization of amino acid sequences for nsp3 truncations from different coronavirus strains.

A-C. Amino acid sequences of nsp3.1 (A), nsp3.2 (B), and nsp3.3 (C) homologs were aligned using Clustal Omega. Domain organization is noted by colored boxes(1).

B.

nsp3.2

	DAPI	PDIA4	FLAG	Overlay
SARS-CoV-2				
SARS-CoV				
MERS-CoV				
hCoV-OC43				
hCoV-229E				

C.

nsp3.3

	DAPI	PDIA4	FLAG	Overlay
SARS-CoV-2				
SARS-CoV				
MERS-CoV				
hCoV-OC43				
hCoV-229E				
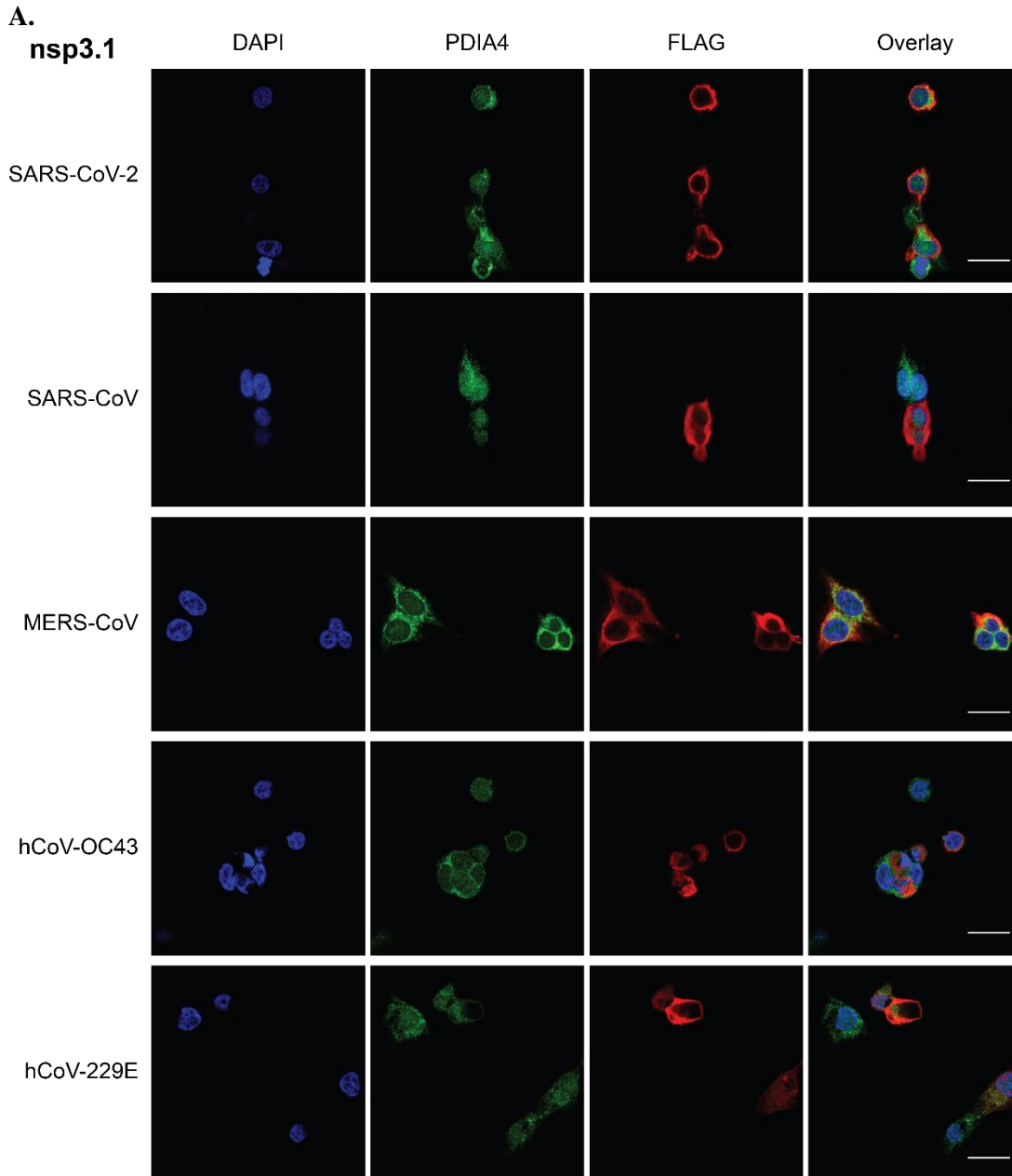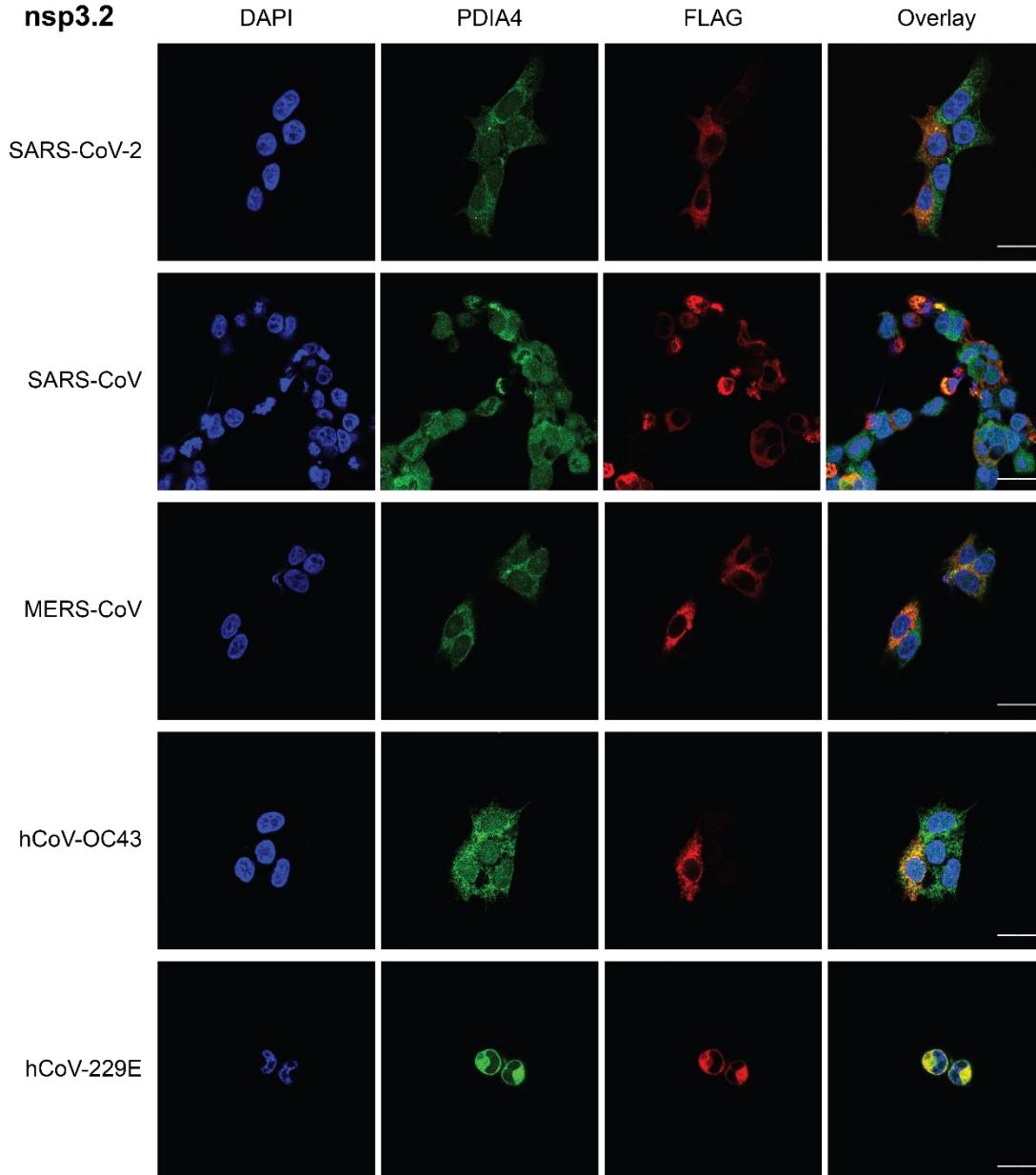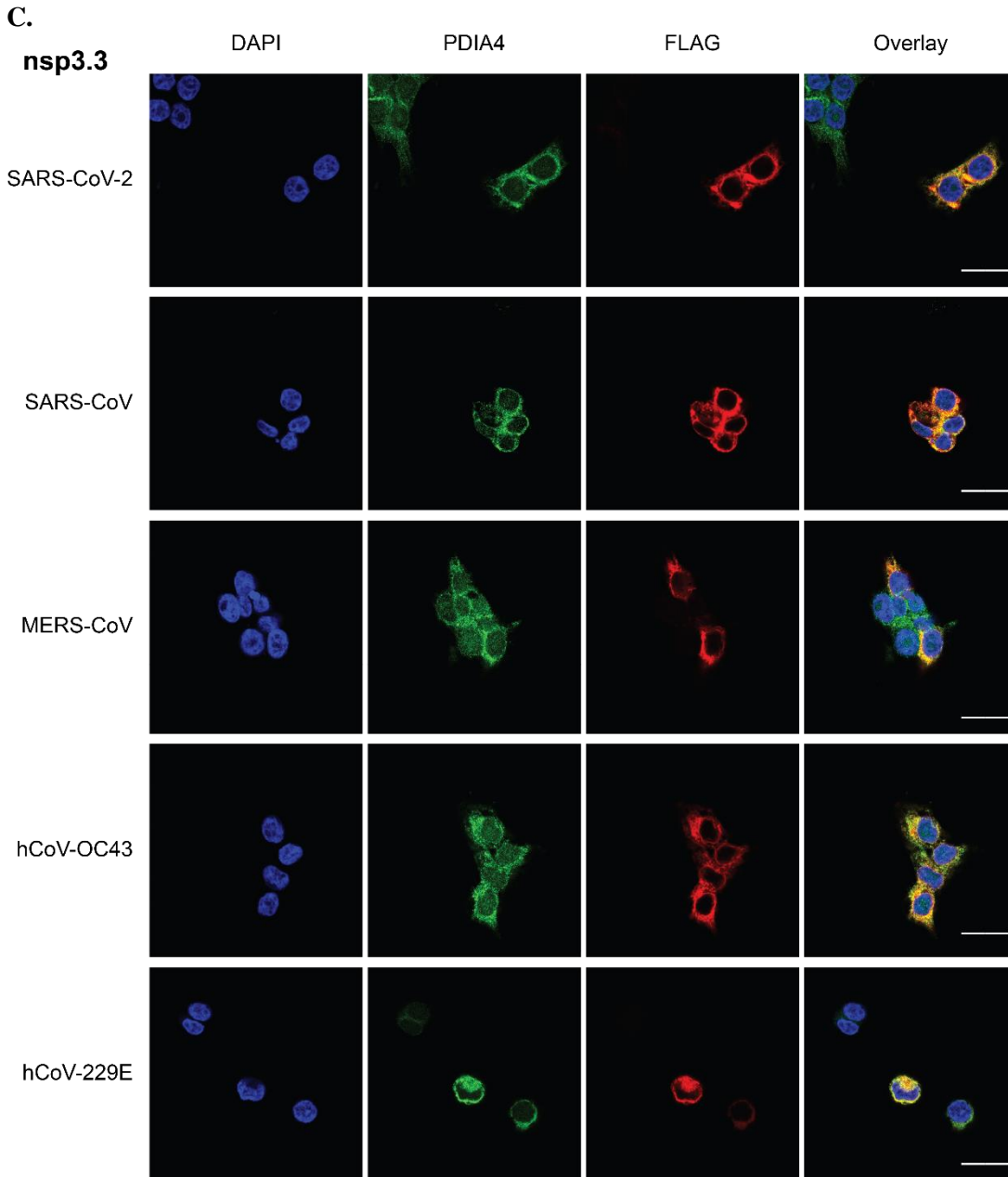
Figure 8-2. Immunofluorescence confocal imaging of nsp3 fragments.

A.-C. Colocalization of nsp3 fragment FLAG-tagged homologs (red) with the ER marker, PDIA4 (green), in HEK293T cells. Representative images shown. Scale bar is 20 µm. (**A**) nsp3.1 is predominantly localized diffusely cytosolic. (**B**) nsp3.2 and (**C**) nsp3.3 show greater co-localization with the ER maker PDIA4 confirming ER localization.
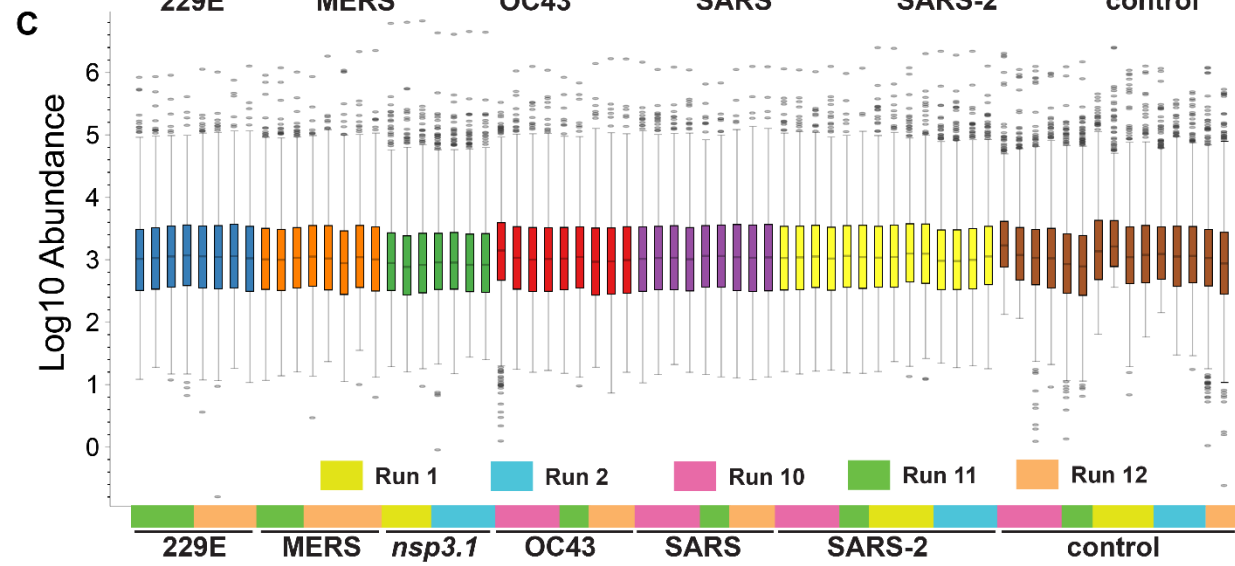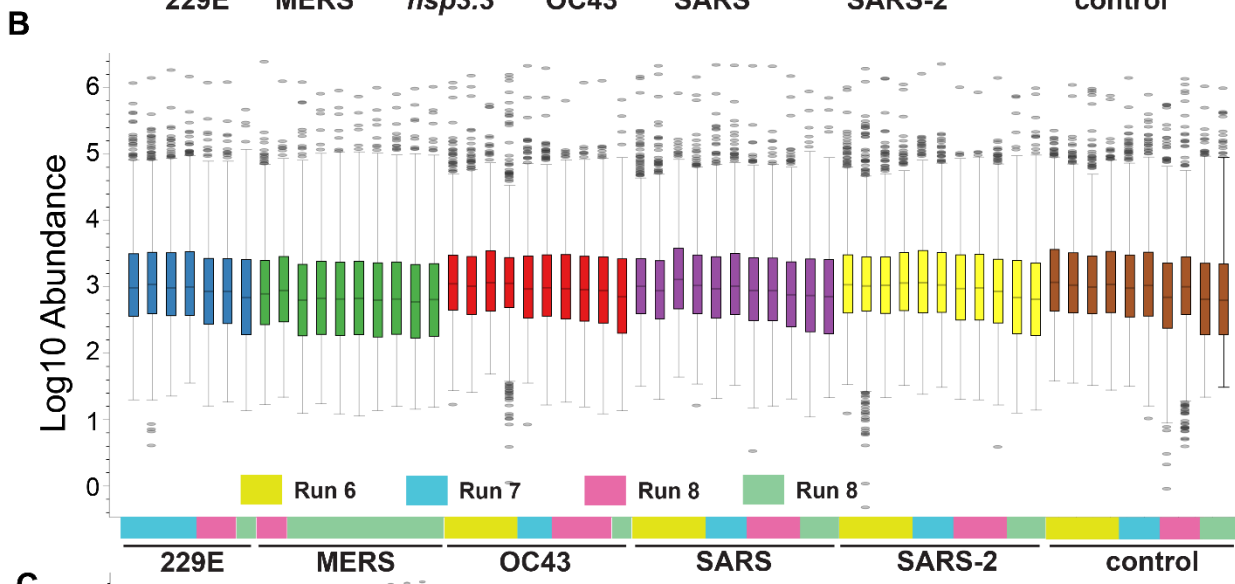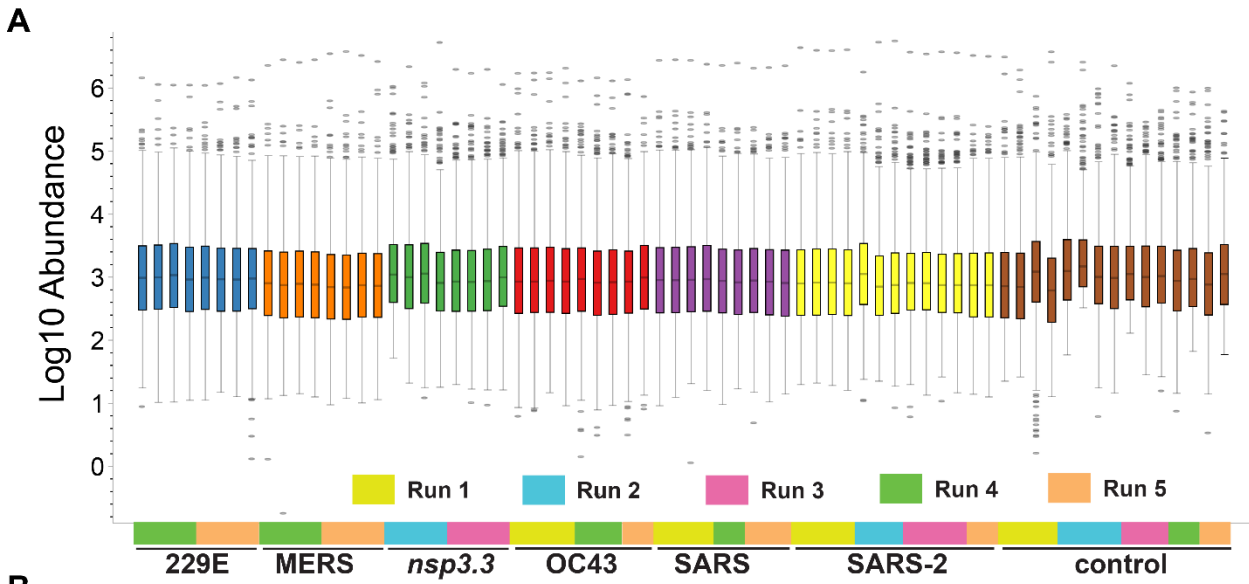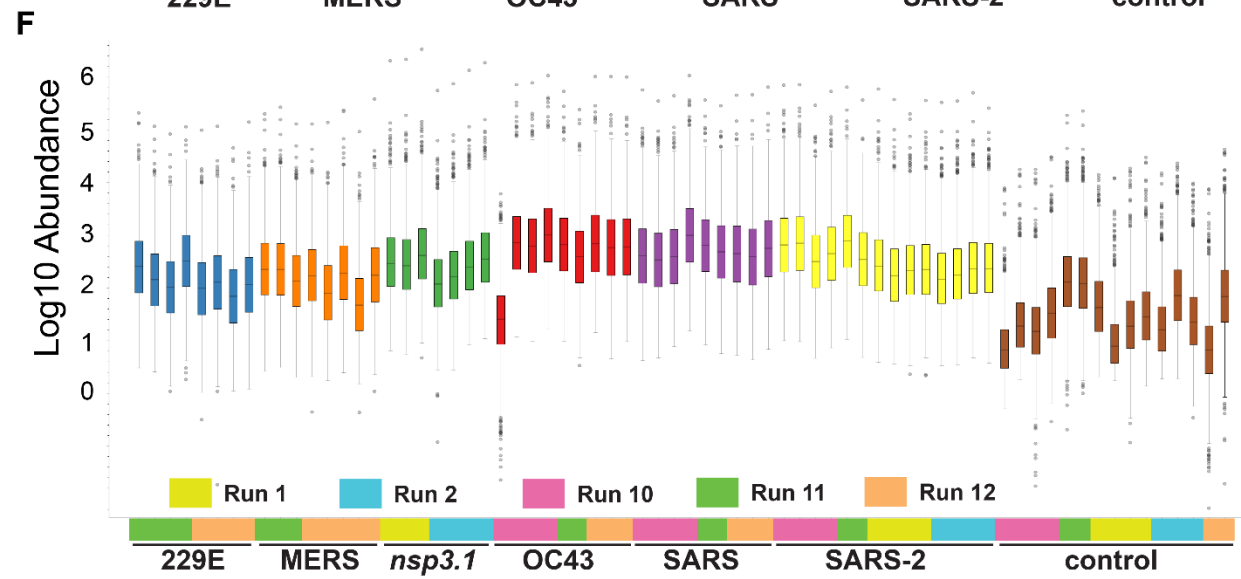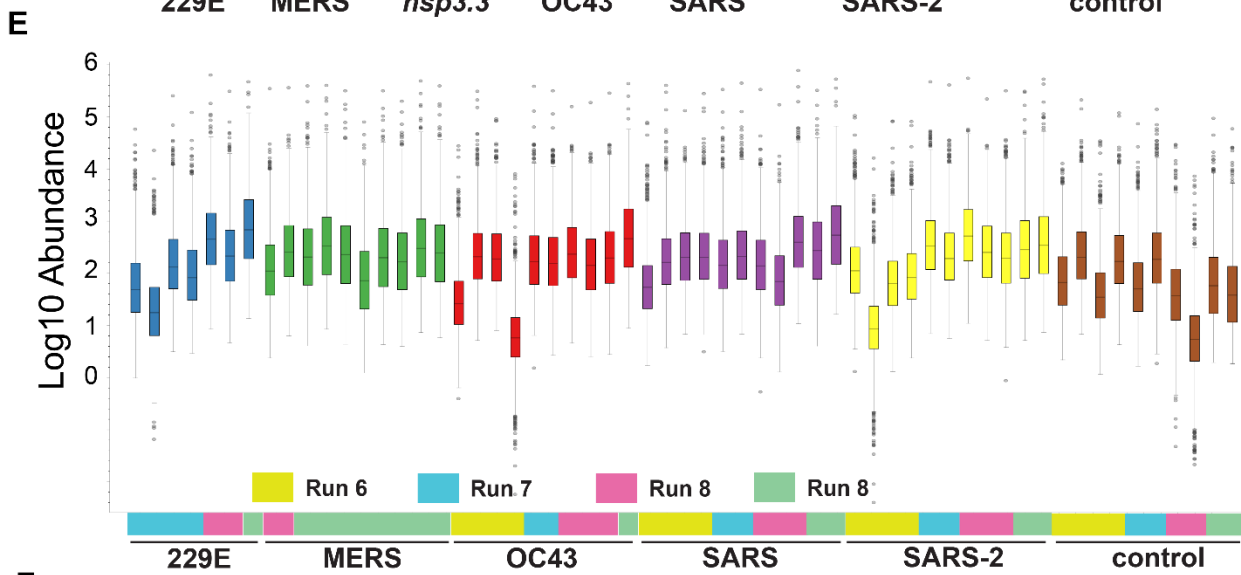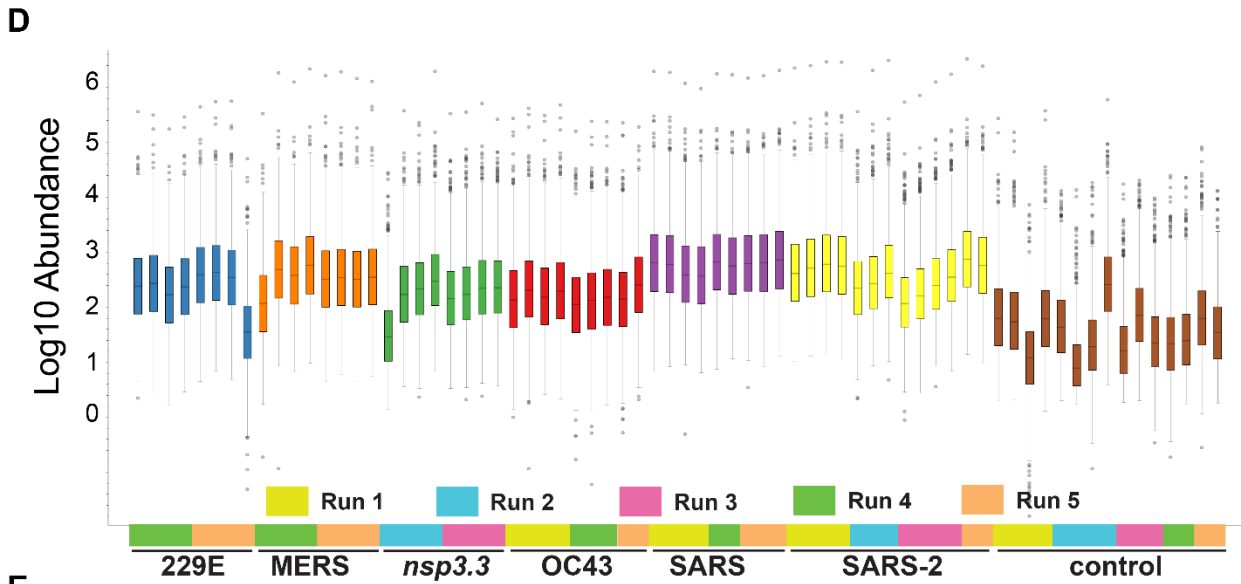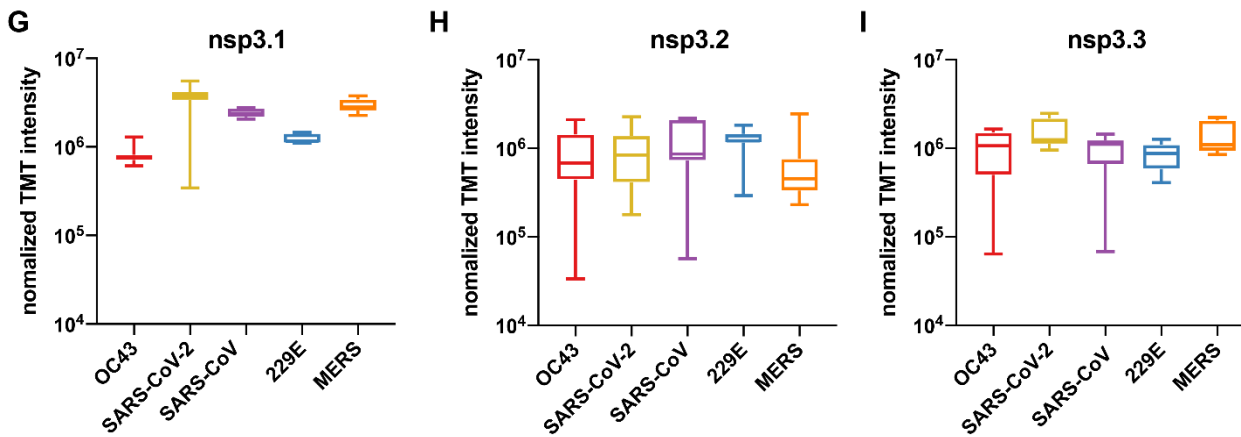
Figure 8-3. Nsp3 TMT intensity distribution.

A-F. Box-and-whisker plot of the log10 TMT intensity abundance of all 11 mass spectrometry runs used in this study. Nsp3 homologs are grouped by same color box-and-whiskers while the identity of the corresponding mass spectrometry run is denoted by colored blocks below. TMT abundances were normalized based on total peptide amounts (**A-C**). See supporting dataset S8 for the layout of samples across TMT channels.

A,D. Normalized (**A**) and unnormalized (**D**) TMT abundance distribution for nsp3.1 dataset. Channels denoted "*nsp3.3*" were used for the nsp3.3 dataset within the corresponding mass spectrometry run.

B,E. Normalized (**B**) and unnormalized (**E**) TMT abundance distribution for nsp3.2 dataset.

C,F. Normalized (**C**) and unnormalized (**F**) TMT abundance distribution for nsp3.3 dataset. Channels denoted "*nsp3.1*" were used for the nsp3.1 dataset within the corresponding mass spectrometry run.

G-I. Normalized TMT abundances of nsp3 bait proteins to compare the enrichment of fragments from the different strains. Box and whisker blots represent min. to max. values. (**G**) nsp3.1, (**H**) nsp3.2, (**I**) nsp3.3.
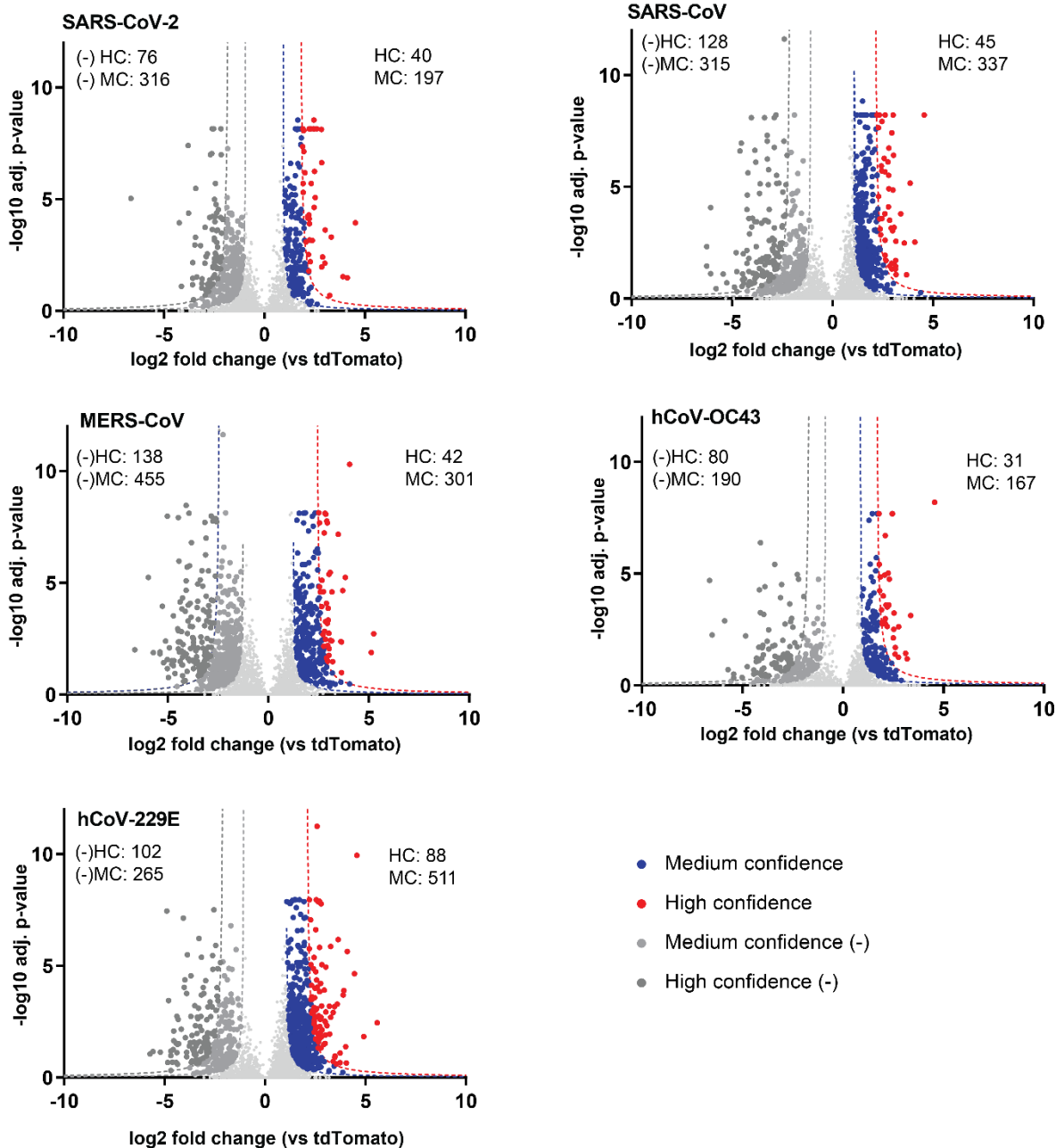
Figure 8-4. Volcano plots of nsp3.1 homolog high-confidence interactors enriched vs tdTomato control.

Host interactors of nsp3.1 homologs were identified by quantitative proteomics and graphed by log2 fold change compared to tdTomato control and -log10 adjusted *p*-value (based on ANOVA). Interactors were filtered for medium (blue) and high (red) confidence interactors based on a hyperbolic curve using 1σ (medium-confidence) or 2σ (high-confidence) standard deviations of the histogram of log2 protein abundance fold changes (refer to "Data analysis" in Methods). Negative medium- and high-confidence (MC, HC) interactors were calculated in a similar manner. Total MC and HC interactors for positive and negative fold change are indicated respectively.
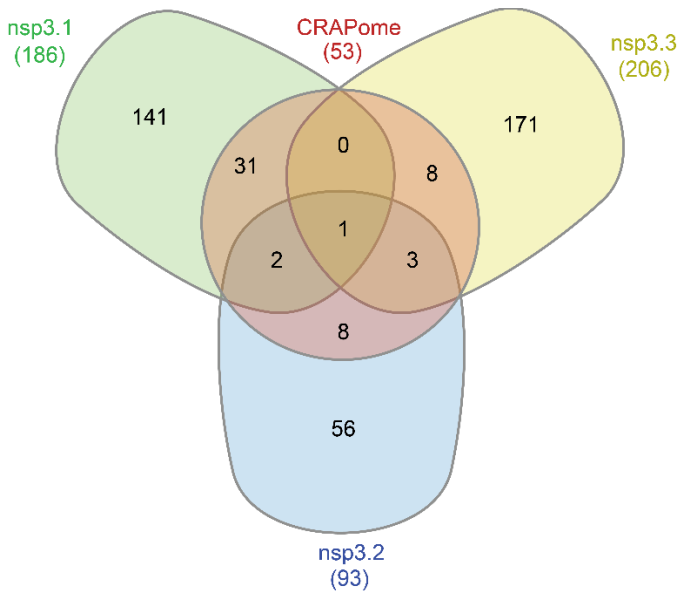
Figure 8-5. CRAPome overlap with nsp3.1, nsp3.2, and nsp3.3 homolog high-confidence interactors.

All high-confidence interactors of nsp3 fragments were queried in the CRAPome Homo sapiens single step epitope tag AP-MS dataset to identify common contaminant proteins(2). Interactors with a frequency of 25% or higher in CRAPome control experiments qualified as overlapping with the CRAPome.
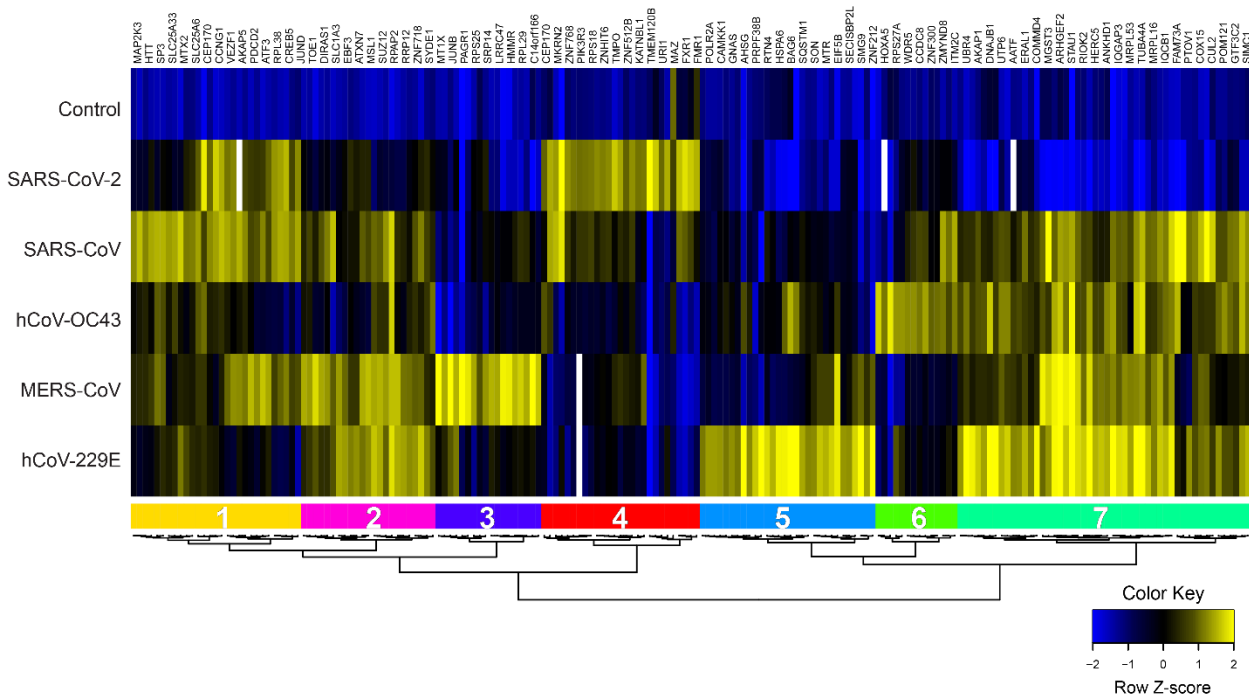
Figure 8-6. Comparative heatmap of nsp3.1 high-confidence interactors.

Unbiased hierarchical clustering of log2 fold change for high-confidence interactors of nsp3.1 homologs yields 7 unique clusters.
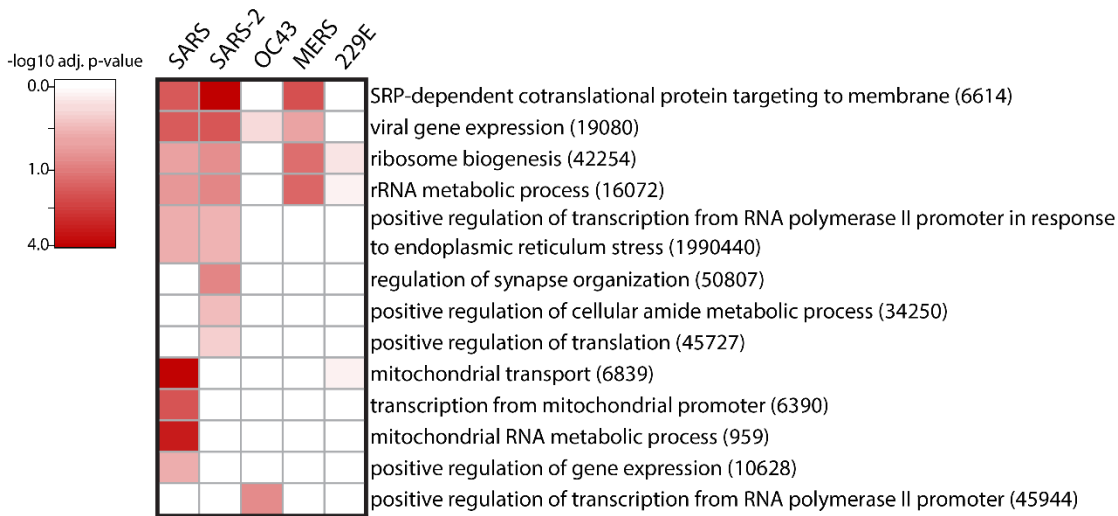
Figure 8-7. GO term analysis of nsp3.1 high-confidence interactors.

Heatmap of gene ontology (GO) term analysis of nsp3.1 high-confidence interactors for enriched biological processes. Selected terms are displayed with corresponding -log10 adjusted *p*-value for each homolog.
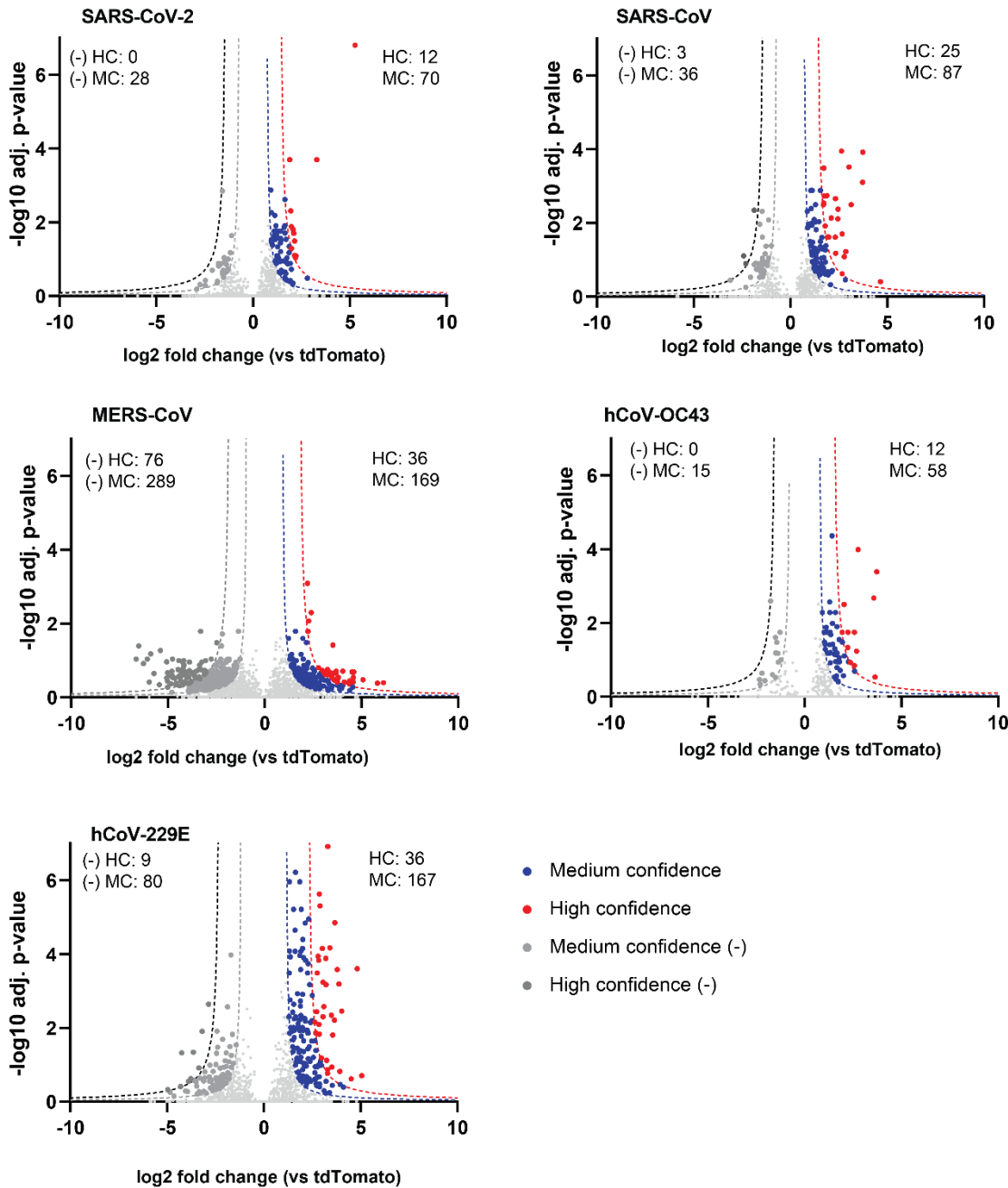
Figure 8-8. Volcano plots of nsp3.2 homolog high-confidence interactors enriched vs tdTomato control.

Host interactors of nsp3.2 homologs were identified by quantitative proteomics and graphed by log2 fold change compared to tdTomato control and -log10 adjusted *p*-value (based on ANOVA). Interactors were filtered for medium (blue) and high (red) confidence interactors based on a hyperbolic curve using 1σ (medium-confidence) or 2σ (high-confidence) standard deviations of the histogram of log2 protein abundance fold changes (refer to "Data analysis" in Methods). Negative medium- and high-confidence (MC, HC) interactors were calculated in a similar manner. Total MC and HC interactors for positive and negative fold change are indicated respectively.
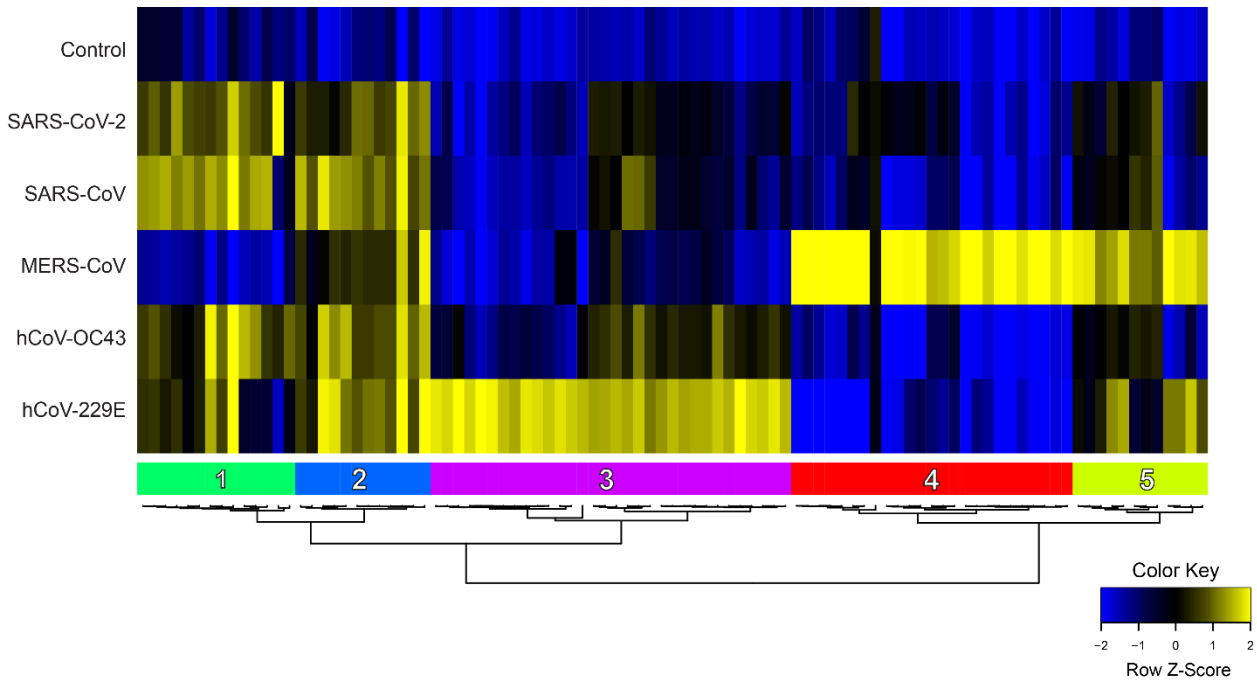
Figure 8-9. Comparative heatmap of nsp3.2 high-confidence interactors.

Unbiased hierarchical clustering of log2 fold change for high-confidence interactors of nsp3.2 homologs yields 5 unique clusters.
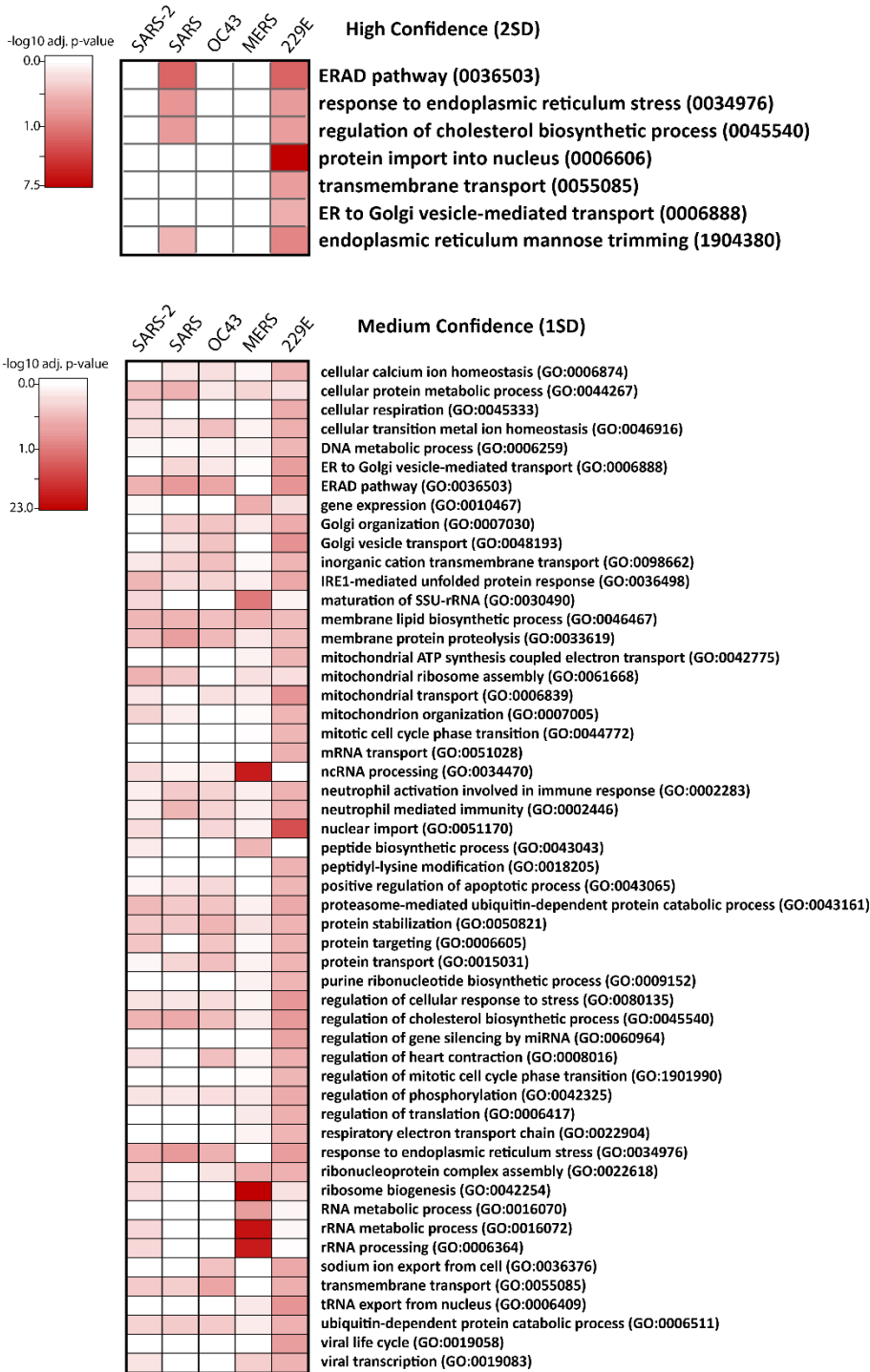
Figure 8-10. GO term analysis of nsp3.2 high-confidence interactors.

Heatmap of gene ontology (GO) term analysis of nsp3.2 high-confidence or medium-confidence interactors for enriched biological processes. Selected terms are displayed with corresponding -log10 adjusted *p*-value for each homolog.
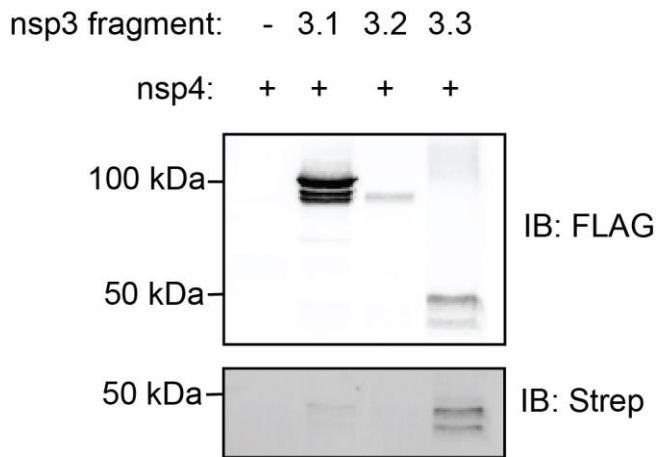
Figure 8-11. Nsp3 fragment co-immunoprecipitations with nsp4.

Samples were co-transfected with 2x-strep tagged SARS-CoV-2 nsp4(3) and individual SARS-CoV-2 FLAG-tagged nsp3 fragments (nsp3.1, nsp3.2, nsp3.3) or tdTomato as a control, respectively. FLAG immunoprecipitations were performed to pull down on nsp3 constructs and immunoblotting was performed with an anti-FLAG antibody to confirm nsp3.3 expression and an FITC anti-Strep antibody to monitor nsp4 co-immunoprecipitation.
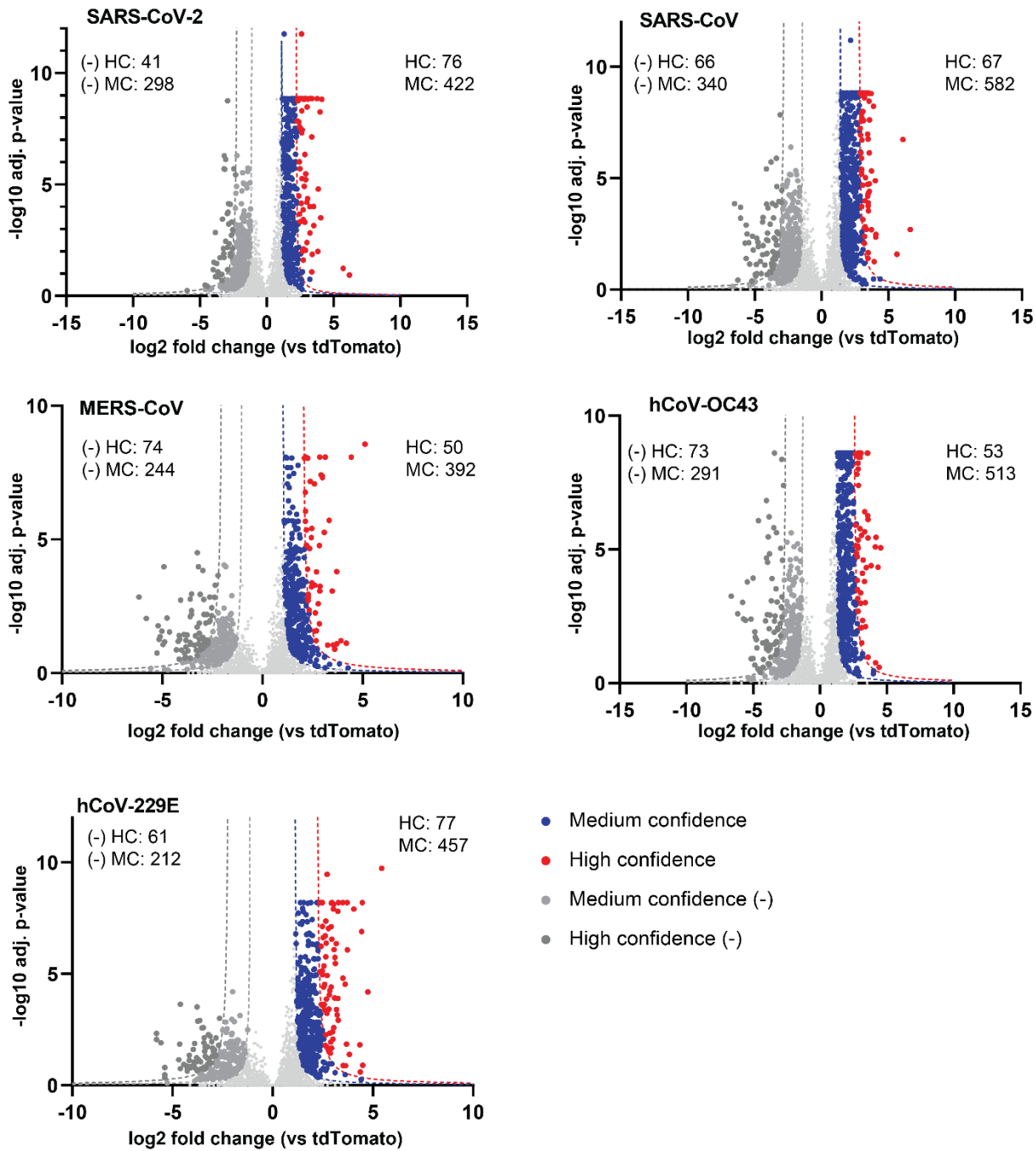
Figure 8-12. Volcano plots of nsp3.3 homolog high-confidence interactors enriched vs tdTomato control.

Host interactors of nsp3.3 homologs were identified by quantitative proteomics and graphed by log2 fold change compared to tdTomato control and -log10 adjusted *p*-value (based on ANOVA). Interactors were filtered for medium (blue) and high (red) confidence interactors based on a hyperbolic curve using 1σ (medium-confidence) or 2σ (high-confidence) standard deviations of the histogram of log2 protein abundance fold changes (refer to "Data analysis" in Methods). Negative medium- and high-confidence (MC, HC) interactors were calculated in a similar manner. Total MC and HC interactors for positive and negative fold change are indicated respectively.
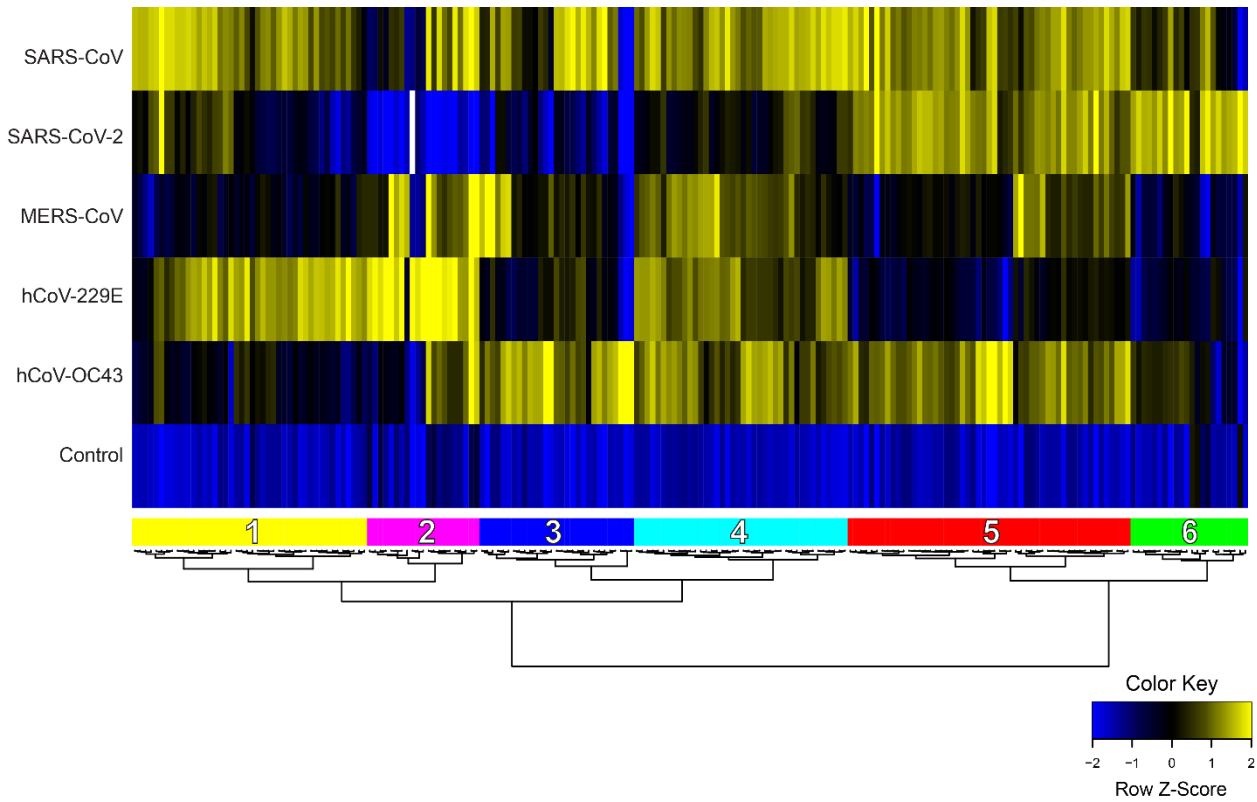
Figure 8-13. Comparative heatmap of nsp3.3 high-confidence interactors.

Unbiased hierarchical clustering of log2 fold change for high-confidence interactors of nsp3.3 homologs yields 6 unique clusters.
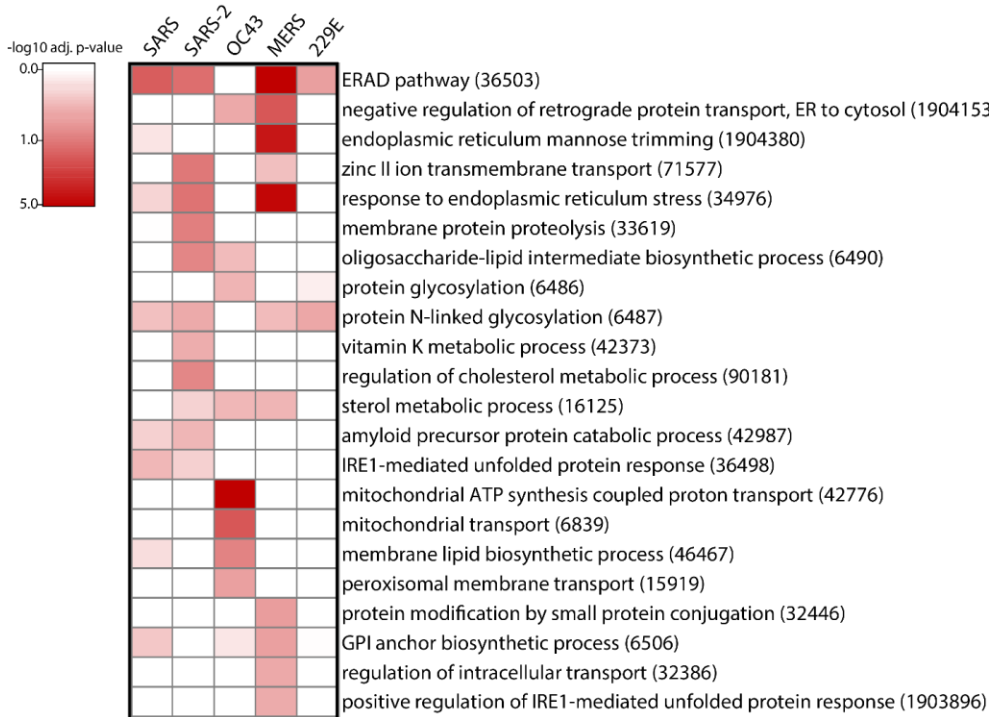
Figure 8-14. GO term analysis of nsp3.3 high-confidence interactors.

Heatmap of gene ontology (GO) term analysis of nsp3.3 high-confidence or medium-confidence interactors for enriched biological processes. Selected terms are displayed with corresponding -log10 adjusted *p*-value for each homolog.
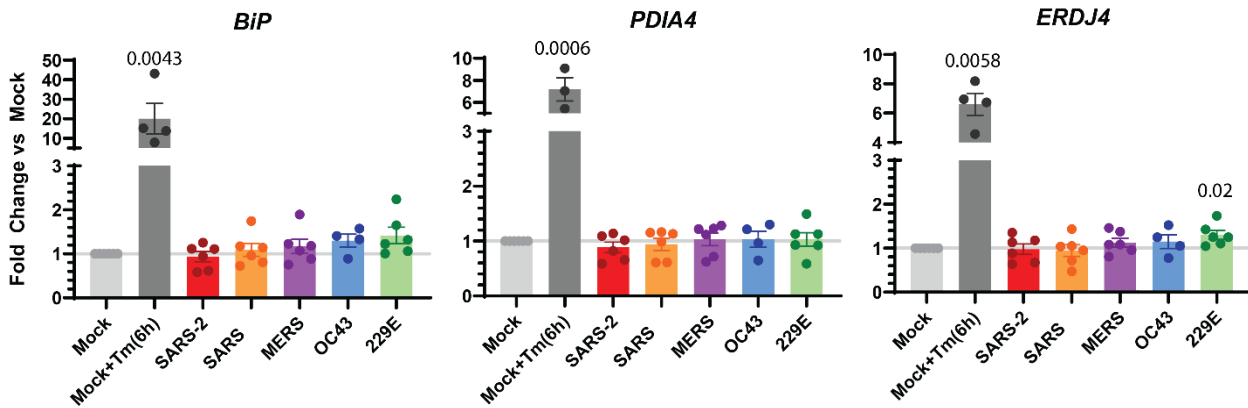
Figure 8-15. Effect of nsp3.1 homologs on UPR transcript expression.

Expression changes of UPR marker transcripts in response to nsp3.1 homolog or mock (tdTomato) transfection measured by reverse transcription qPCR (RT-qPCR). A mock transfected sample was treated with 6 µg/mL Tunicamycin (Tm) 6 h pre-harvest to induce general UPR activation as a positive control. *BiP* and *PDIA4* were measured as ATF6 markers, while *ERDJ4* was measured as a IRE1/XBP1s marker. Bars indicate average, error bars indicate SEM. Paired student T-tests were used to test for significance, with *p*-values < 0.05 shown. Mock+Tm(6h), $n = 4$ biological replicates; all other samples, $n = 6$ biological replicates.
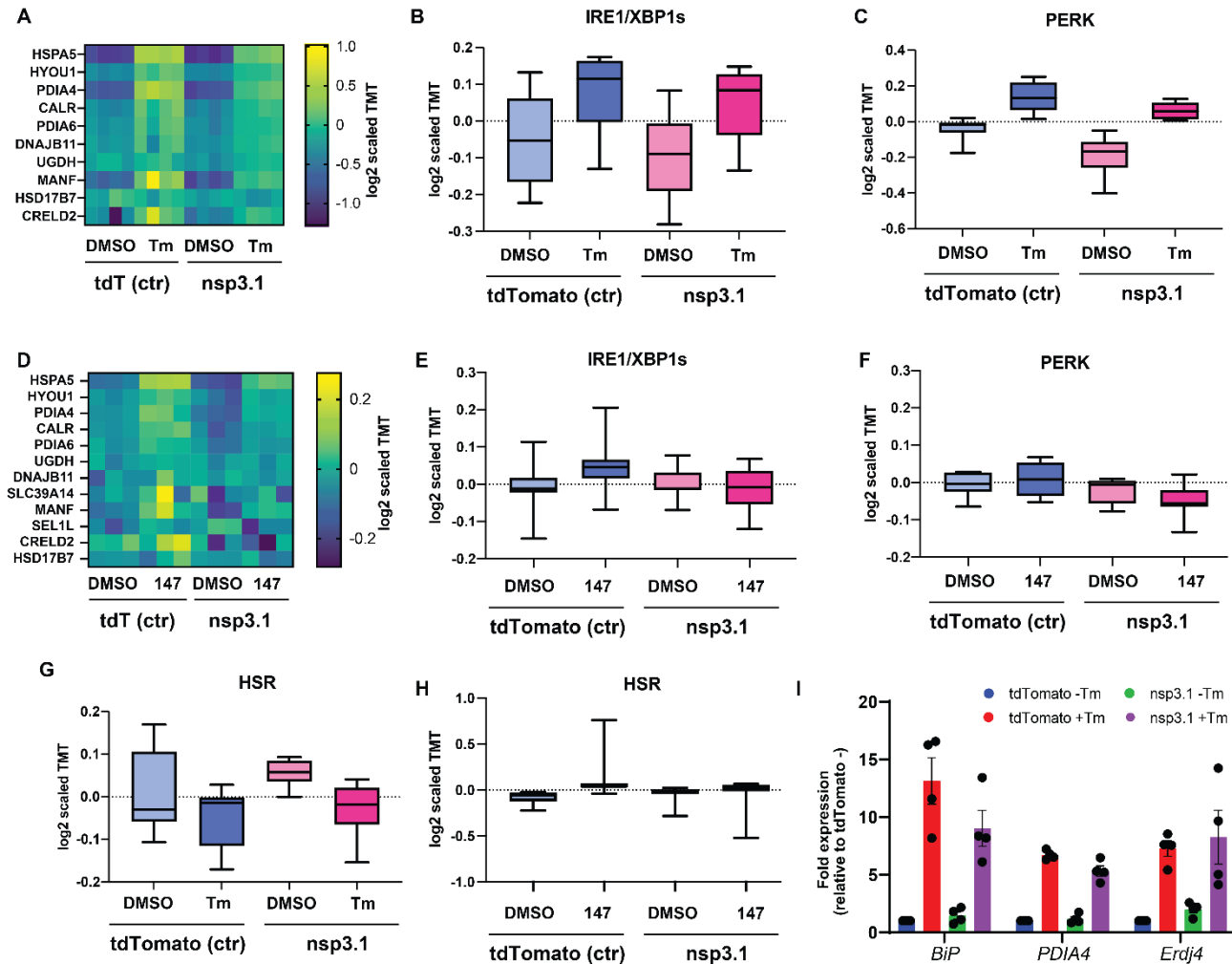
Figure 8-16. IRE1, PERK UPR, and HSR activation by SARS-CoV-2 nsp3.1

A-C, G. Heatmap of ATF6 UPR signaling branch protein markers (**A**), box-and-whisker plots showing distribution of protein makers of IRE1/XBP1s (**B**) and PERK (**C**) UPR signaling branches and the cytosolic heat shock response (HSR) pathway (**G**) as measured by quantitative proteomics. HEK293T cells were transfected with tdTomato (control) or SARS-

171

CoV-2 nsp3.1-FT. Samples were treated with DMSO or 6 µg/mL tunicamycin for 16 h pre-harvest. A one-way ANOVA with Geisser-Greenhouse correction and post-hoc Tukey's multiple comparison test was used to determine significance (adjusted p-values all > 0.05). $n = 4$ biological replicates in a single mass spectrometry run.

D-F, H. Heatmap of ATF6 UPR signaling branch protein markers (**A**) and box-and-whisker plots showing distribution of protein makers of IRE1/XBP1s (**B**) and PERK (**C**) UPR signaling branches and HSR pathway (**H**) as measured by quantitative proteomics. HEK293T cells were transfected with tdTomato (control) or SARS-CoV-2 nsp3.1-FT. Samples were treated with DMSO or 10 µM **147** for 16 h pre-harvest. A one-way ANOVA with Geisser-Greenhouse correction and post-hoc Tukey's multiple comparison test was used to determine significance (adjusted p-values all > 0.05). $n = 3$ biological replicates in a single mass spectrometry run.

I. Expression changes of ATF6 (*BiP*, *PDIA4*) and IRE1/XBP1s (*Erdj4*) markers in response to SARS-CoV-2 nsp3.1 or mock (tdTomato) transfection. Where indicated, tunicamycin (Tm) samples were treated 6 hours prior to harvest. Paired ratio t-tests were used to measure significance ($n = 4$ biological replicates)

References
1. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Research* at https://doi.org/10.1016/j.antiviral.2017.11.001 (2018).
2. Mellacheruvu, D. *et al.* The CRAPome: A contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).
3. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

Supplemental figures for Chapter 4.
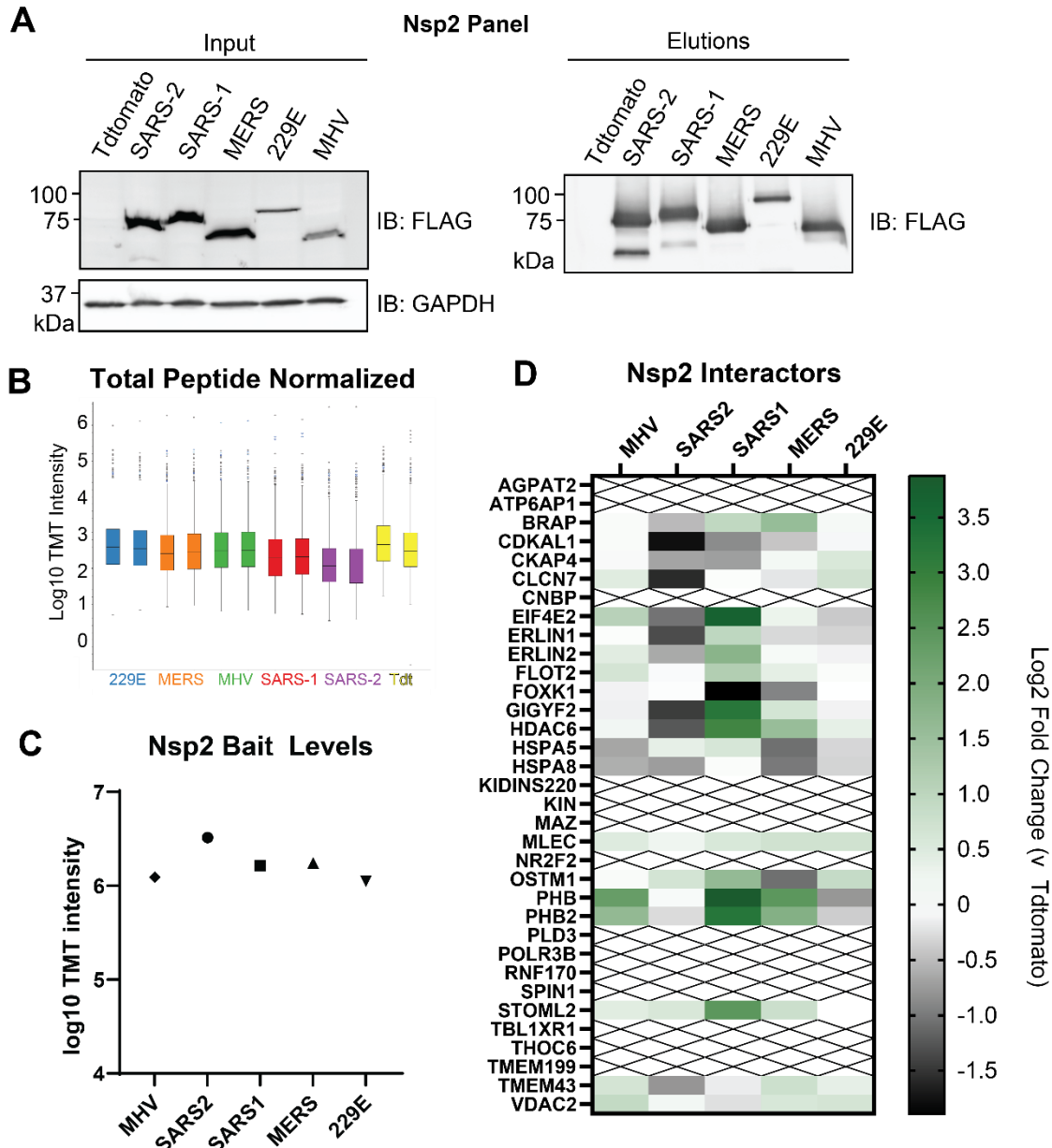
Figure 9-1. Nsp2 homolog panel comparative interactome.

A. Representative western blot of nsp2 homolog panel expressed in HEK293T cells and FLAG immunoprecipitated. Tdtomato serves as negative IP control. Blotting for FLAG and GAPDH as a loading control.
B. Normalized TMT channel abundances of nsp2 panel AP-MS (n = 3-4 IPs/homolog, 1 MS run)
C. Average nsp2 bait levels as measured by AP-MS.
D. Interacting proteins from our previous comparative proteomics study(1) were cross-referenced against this nsp2 AP-MS panel and the log2 fold enrichment against Tdtomato background is displayed. Proteins not identified in this the current AP-MS panel are marked with an "X".
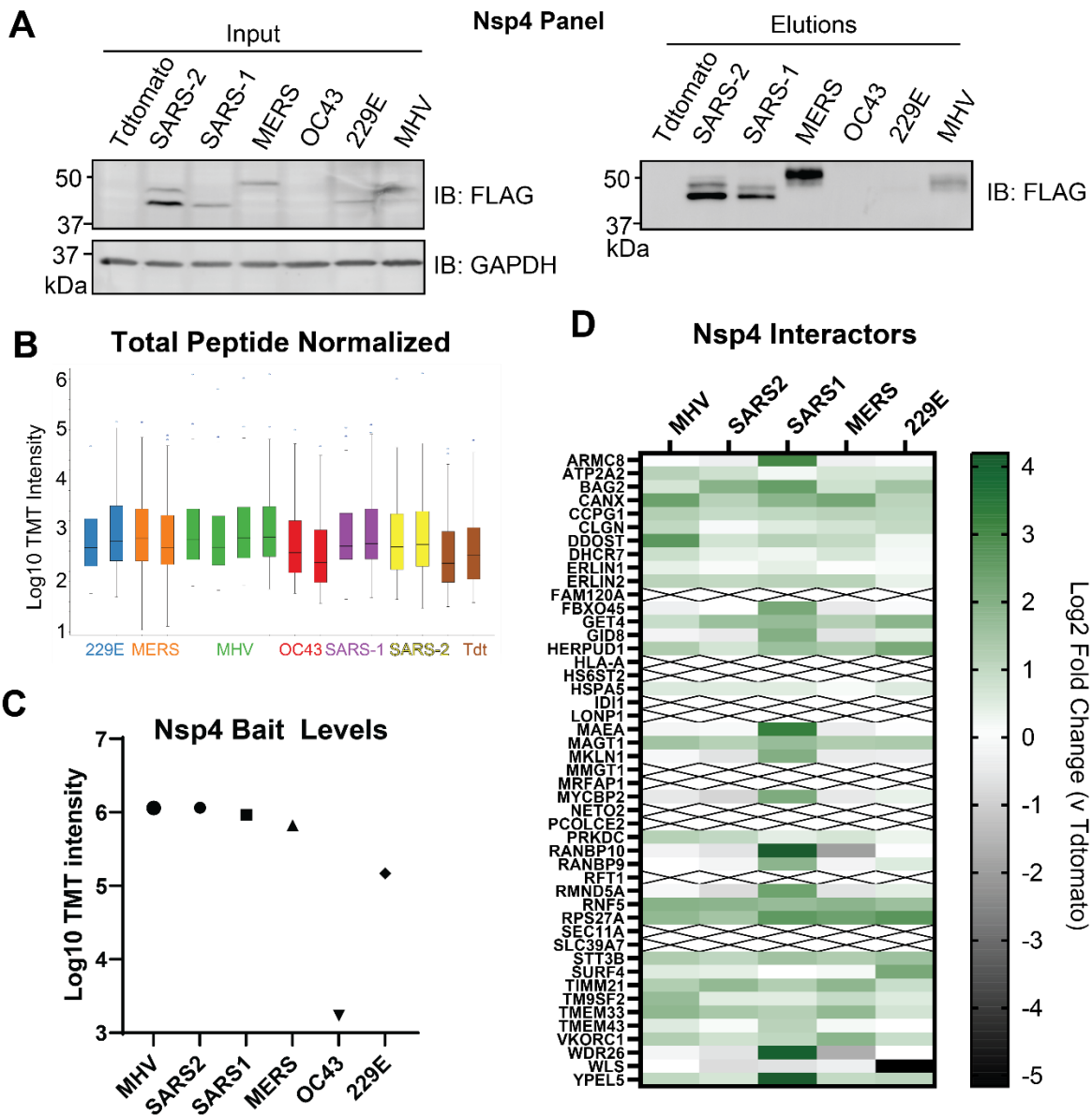
Figure 9-2. Nsp4 homolog panel comparative interactome.

A. Representative western blot of nsp4 homolog panel expressed in HEK293T cells and FLAG immunoprecipitated. Tdtomato serves as negative IP control. Blotting for FLAG and GAPDH as a loading control.

B. Normalized TMT channel abundances of nsp4 panel AP-MS (n = 3-4 IPs/homolog, 1 MS run)

C. Average nsp4 bait levels as measured by AP-MS. OC43 was omitted from analysis due to low expression levels.

D. Interacting proteins from our previous comparative proteomics study(1) were cross-referenced against this nsp4 AP-MS panel and the log2 fold enrichment against Tdtomato background is displayed. Proteins not identified in this the current AP-MS panel are marked with an "X".
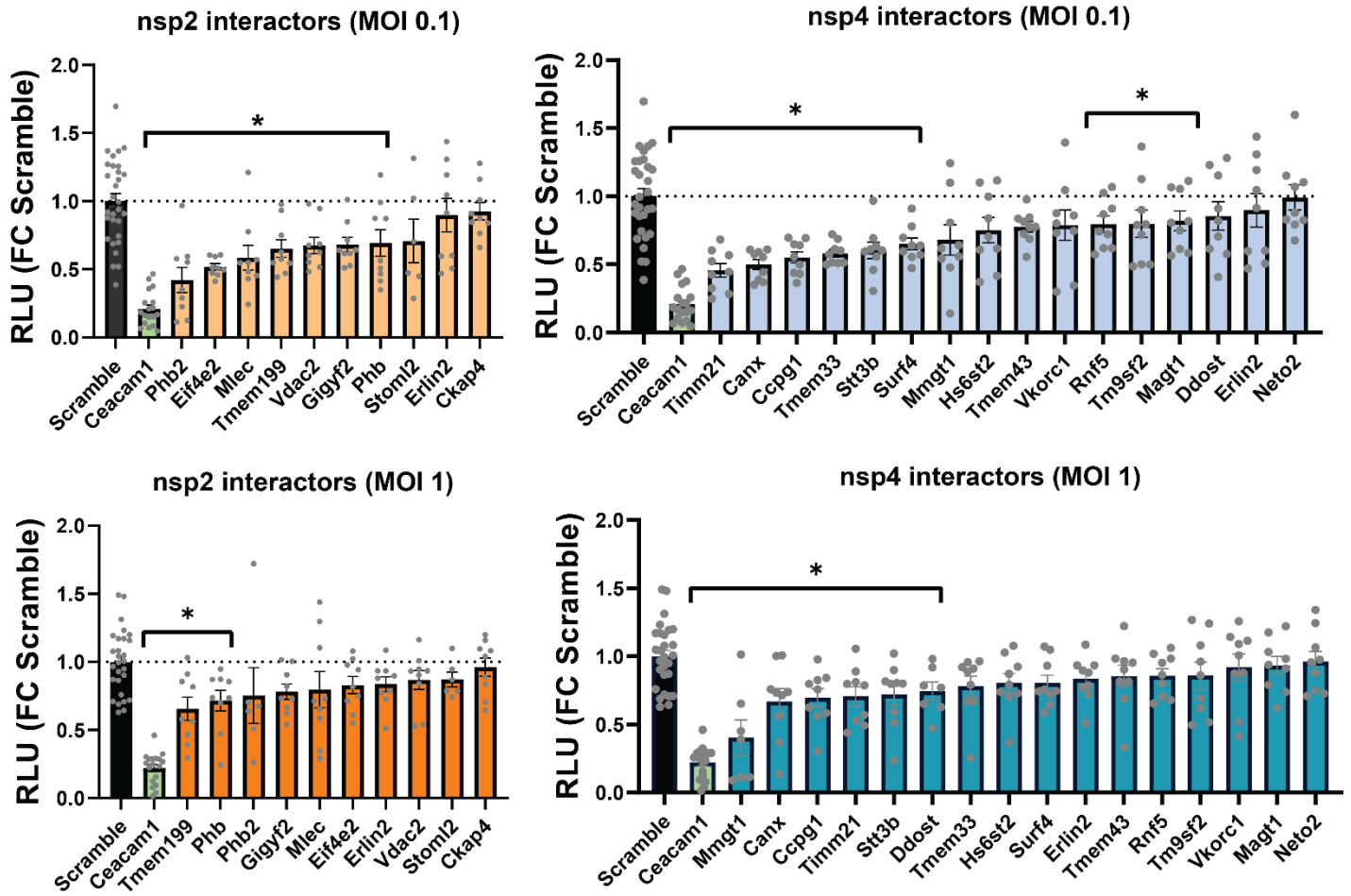
Figure 9-3. MHV-FFL2 replicase reporter screen against RNAi of interactors.

Ten nsp2 interactors and 16 nsp4 interactors were knocked down by siRNA in DBT cells in 96-well plates, then infected with MHV-FFL2 at MOI 0.1 (10 hpi) or MOI 1 (8 hpi). Lysates were harvested and luminescence measured by SteadyGlo assay (Relative Luminescence Units, RLU). Student's T-test for significance, with $p<0.05$ considered significant. KDs that test significant compared to Scramble are annotated by the asterisk bracket. Mean ±SEM; interactor KD, n = 9 across 3 separate assays; Scramble, n = 28 across 10 separate assays; Ceacam1 KD, n = 18 across 6 separate assays.
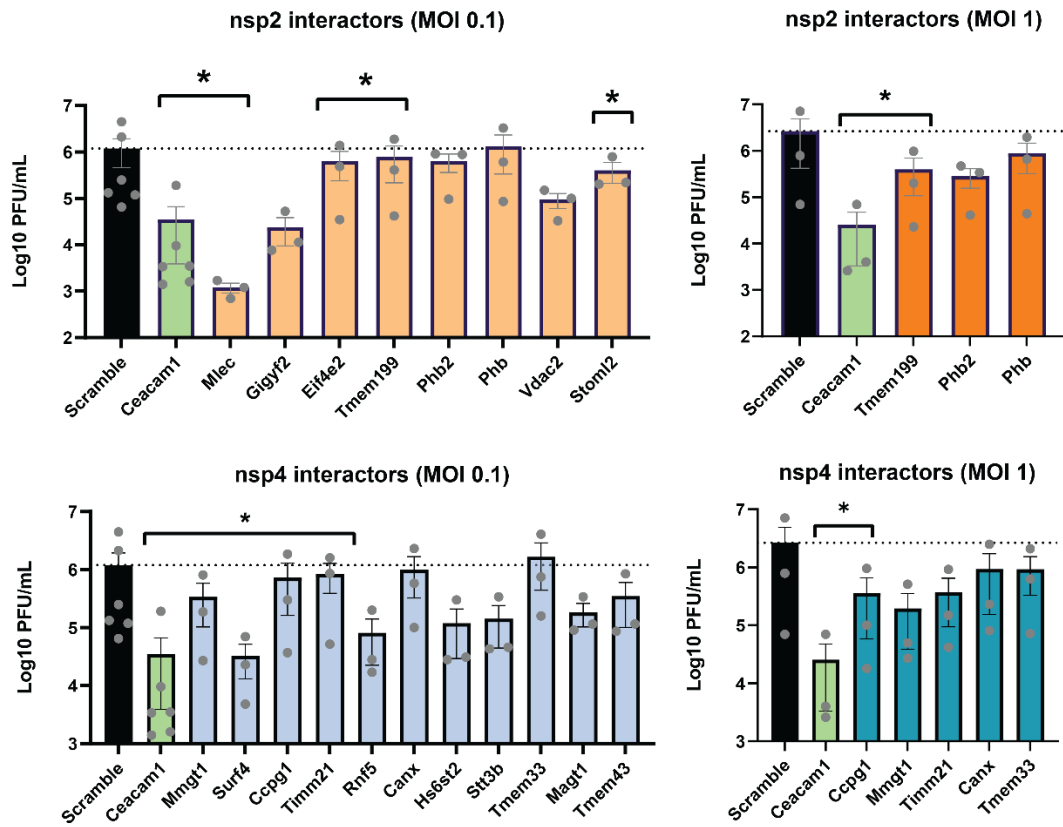
Figure 9-4. Effect of interactor knockdown screen on MHV titer.

Prioritized interactors from the MHV-FFL2 replicase reporter screen were knocked down in DBT cells with siRNA for 40 h, then infected with MHV (MOI 0.1, 10 hpi or MOI 1, 8 hpi as indicated). Scramble siRNA was used as base line control and siCeacam1 as positive control. Supernatant from infected cells was collected and infectious titer measured by plaque assay. Student's t-test for significance, with $p < 0.05$ considered significant. Mean±SEM; Interactors, n = 3; Scramble, n = 6 (MOI 0.1) or 3 (MOI 1); siCeacam1, n = 6 (MOI 0.1) or 3 (MOI 1).
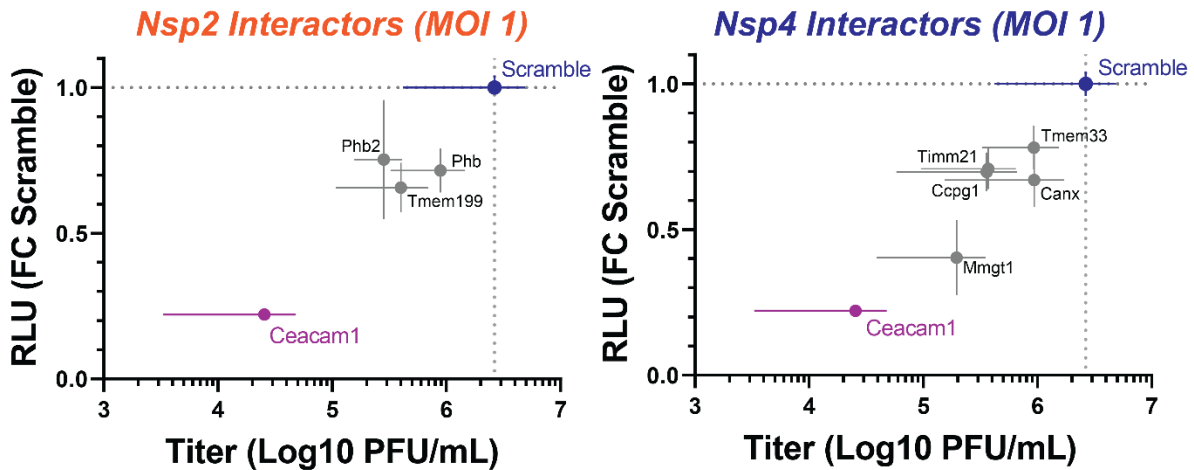
Figure 9-5. Effect of interactor knockdown on MHV-FFL2 replicase reporter vs. MHV titer (MOI 1).

Correlation plot of the effect of prioritized nsp2 and nsp4 interactor knockdowns on MHV-FFL2 replicase reporter vs. MHV titer (MOI 1, 8 hpi). Scramble (blue) siRNA was used as basal control, siCeacam1 (magenta) as a positive control.

Figure 9-6. Cell viability of interactor knockdowns.

DBT cells were treated with respective siRNAs for 40 hpi in 96-well plate and then assayed for cell viability via CellTiter Glo kit. Non-treated cells (Plain) and siTOX (positive control leading to cell death) were included as controls. Mean±SEM; Scramble, n = 12; siCeacam1, n = 6, all others, n = 3.
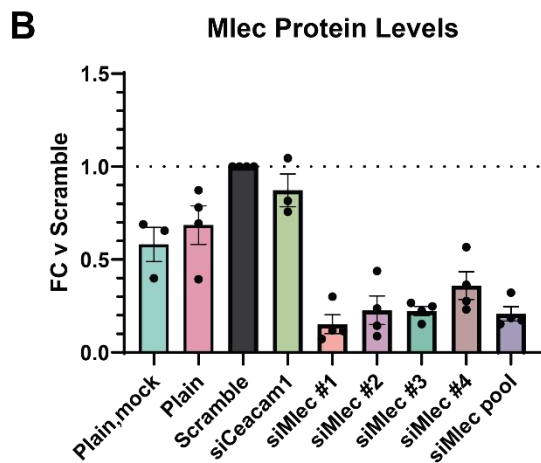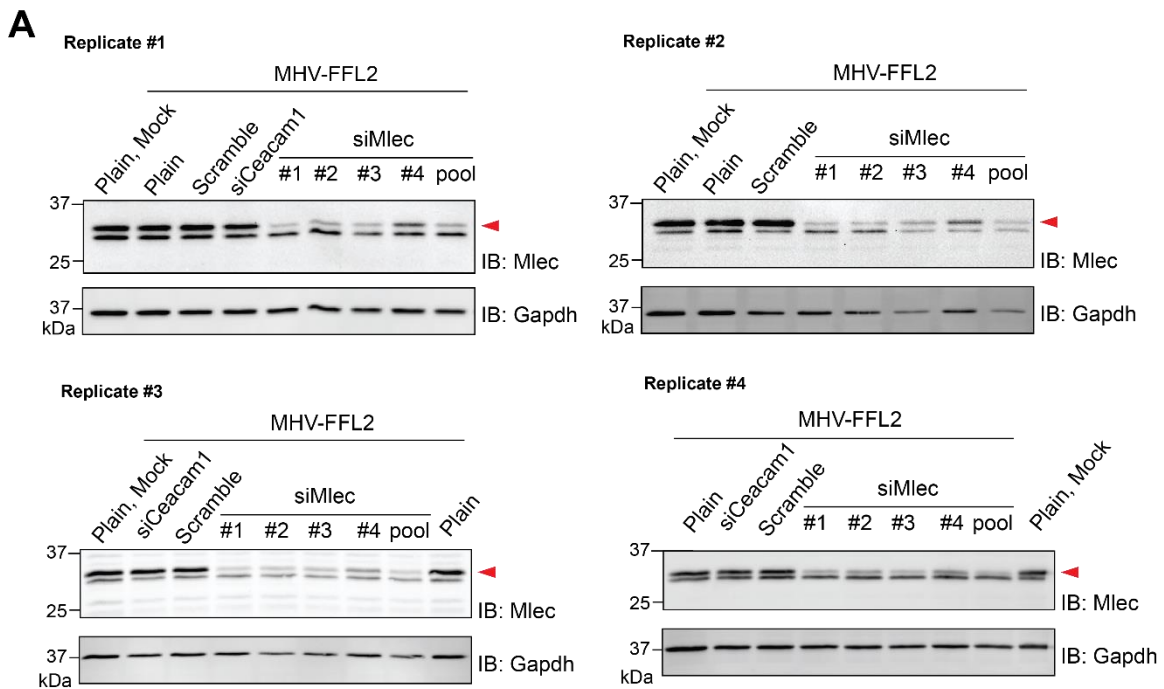
Figure 9-7. Mlec KD efficiency with individual vs. pooled siMlec.

A. Western blots of DBT lysates probed for endogenous Mlec (red arrow) with GAPDH as loading control. Cells were treated with individual siMlec siRNAs (#1-4) or pooled siMlec (all four individual siRNAs combined) for 40 h prior to MHV-FFL2 infection. Plain cells were not treated with siRNA.
B. Quantification of Mlec bands in (A), normalized to GAPDH levels and relative to Scramble control. Mean±SEM, n = 3-4.
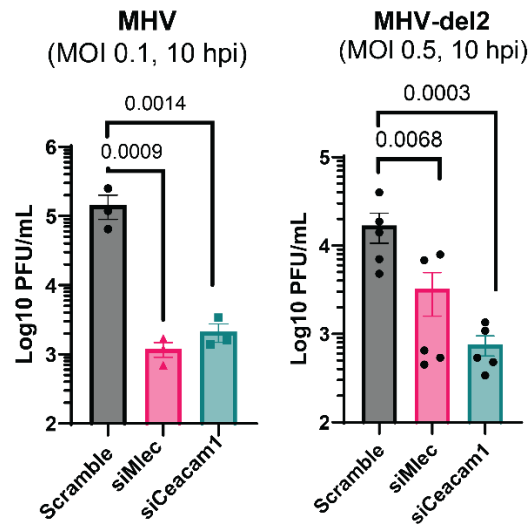
Figure 9-8. Genetic interaction between nsp2 and Mlec.

Infectious titers (PFU/mL), as measured by plaque assay, from DBT cells treated with Scramble, siCeacam1, or siMlec siRNA for 40 h prior to infection with MHV (MOI 0.1, 10 hpi) or MHV-del2 (nsp2 deletion, MOI 0.5, 10 hpi). Mean±SEM; MHV, n = 3; MHV-del2, n = 5. Student's t-test for significance, $p < 0.05$ considered significant.
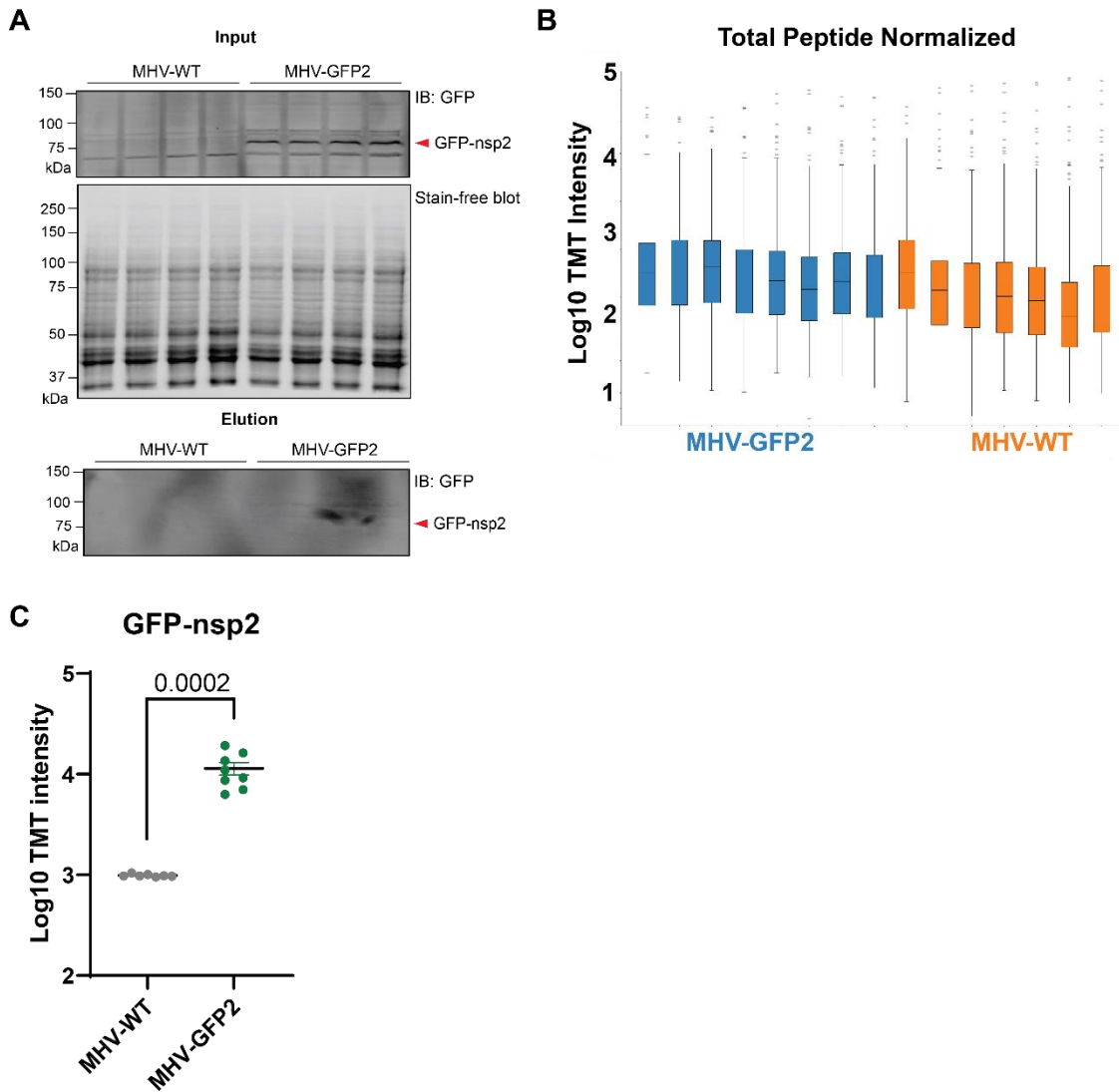
Figure 9-9. MHV-GFP2 IP and AP-MS metrics.

A. Western blots of GFP-IP inputs and elutions for AP-MS (four representative samples per condition), probing for GFP. DBT cells were infected with MHV-WT or MHV-GFP2 (MOI 1, 9 hpi), lysed, and immunoprecipitated for GFP. Stain-free blot of inputs as loading control. Red arrow indicates GFP-nsp2. While we did not detect a GFP-nsp2 band in elution, subsequent MS analysis showed enriched levels of nsp2 over MHV-WT (C), indicating successful enrichment. We attribute lack of blot detection due to poor antibody performance.
B. Total peptide normalized TMT abundance from MHV-GFP2 AP-MS. MHV-WT, n = 7; MHV-GFP2, n = 8 IPs, in 1 MS run.
C. GFP-nsp2 levels measured by AP-MS in MHV-WT and MHV-GFP2 IP samples. One-way ANOVA with multiple testing corrections to test significance.
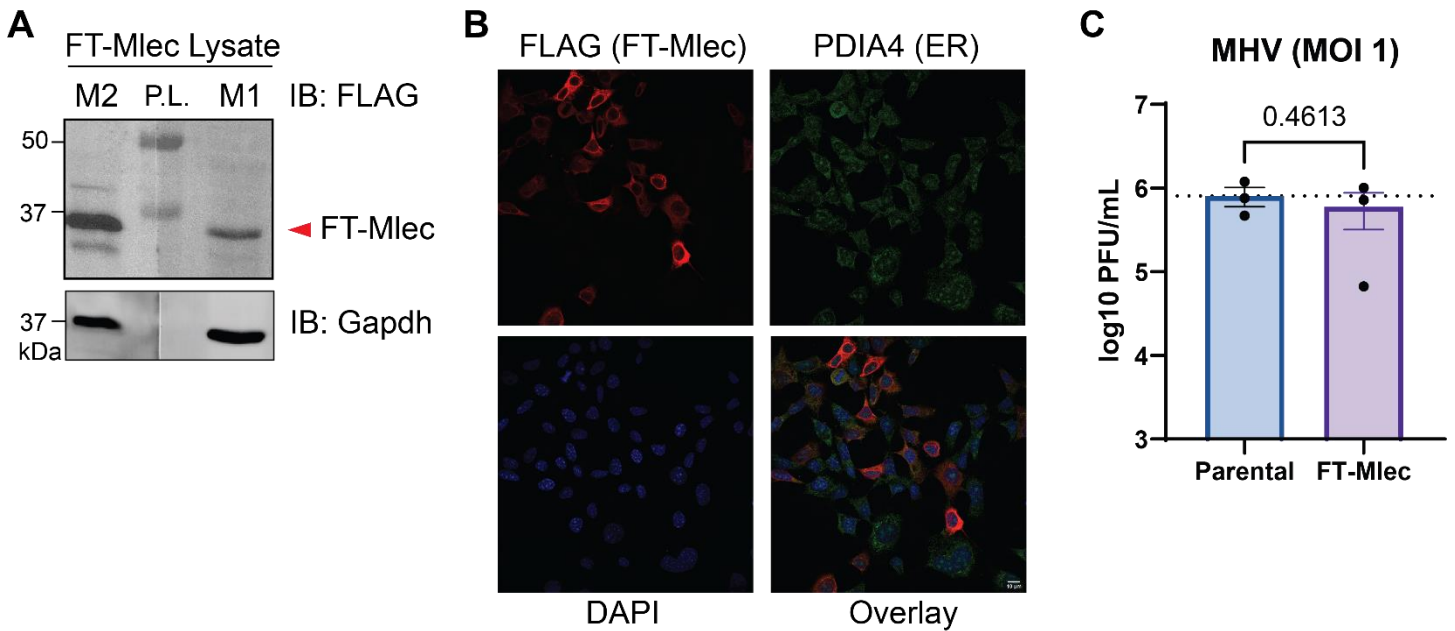
Figure 9-10. Validation of FT-Mlec stable DBT cell line.

A. Western blot of FT-Mlec lysate, probed with either anti-M2 or M1 FLAG, to test for cleavage of signal sequence and reveal of N-terminal FLAG tag (M1) on Mlec construct. The same lysate was split in half and run on an SDS-PAGE gel. After transfer, the blot was cut in two halves and probed separately with respective FLAG antibodies, then imaged together. P.L., protein ladder.

B. Colocalization immunofluorescence (IF) of FT-Mlec DBT cells, staining for FLAG (FT-Mlec), PDIA4 (ER marker) FT-Mlec, and DAPI, with overlay. Scalebar = 10 µm

C. Titers of MHV infection in FT-Mlec or Parental DBT cells (MOI 1, 9 hpi), measured by plaque assay. Paired Student's t-test for significance, $p < 0.05$ considered significant; n = 3.
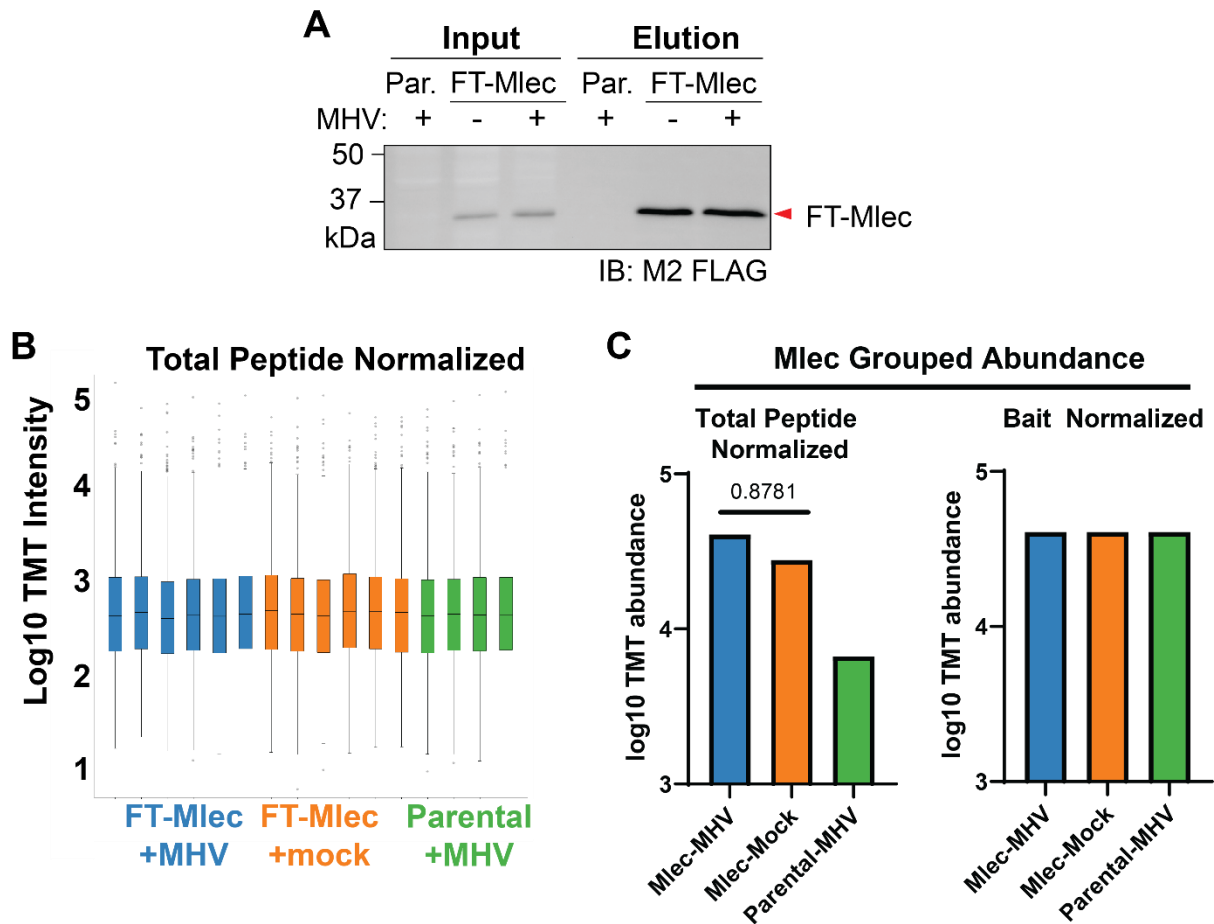
Figure 9-11. FT-Mlec IP and AP-MS metrics.

A. Western blot of representative samples from FLAG IP input and elution in Parental (Par.) or FT-Mlec cells infected with MHV or mock (MOI 2.5, 9 hpi), probed for M2 FLAG. Red arrow indicates FT-Mlec.

B. Total peptide normalized TMT channel abundances of FT-Mlec AP-MS. FT-Mlec + MHV, n = 6; FT-Mlec + mock, n = 6; Parental+MHV, n = 4 IPs.

C. Mlec grouped TMT abundance with total peptide normalization or bait (FT-Mlec) normalization. Bait normalization was used for FT-Mlec interactome comparison between MHV vs. Mock conditions (**Fig. 3F**). T-test with multiple testing corrections for significance.
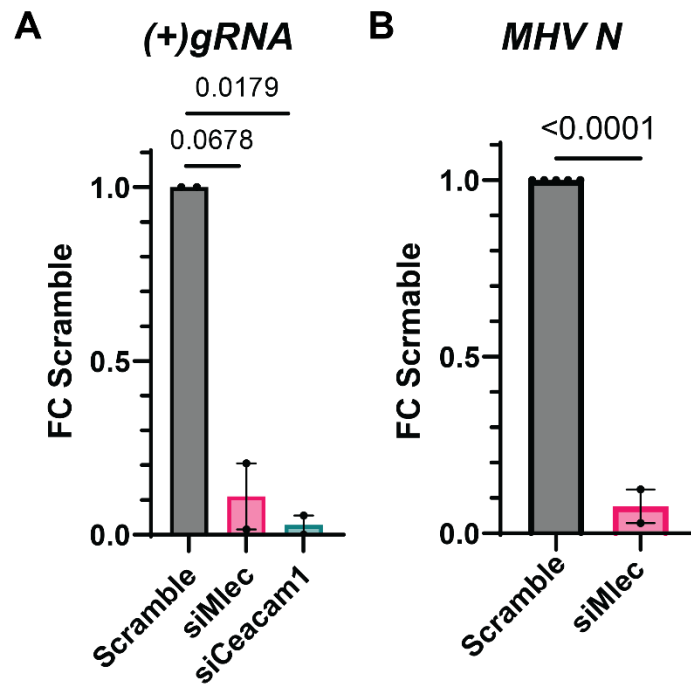
Figure 9-12. Effect of Mlec KD on viral genomic and subgenomic RNA levels.

A. Positive-sense genomic RNA (*(+)gRNA*) levels in Scramble, siCeacam1, or siMlec treated DBT cells infected with MHV (MOI 1, 10 hpi), as measured by RT-qPCR and normalized to *Rpl13a* transcript levels (**Table 1**). n = 2; Student's T-test for significance, $p < 0.05$ considered significant.

B. Subgenomic (*Nucleocapsid*) transcript levels in Scramble vs siMlec siRNA treated DBT cells infected with MHV (MOI 0.1, 10 hpi), as measured by RT-qPCR and normalized to *Gapdh* transcript levels (**Table 1**). Scramble, n = 5; siMlec, n = 2; Student's T-test for significance, $p < 0.05$ considered significant.
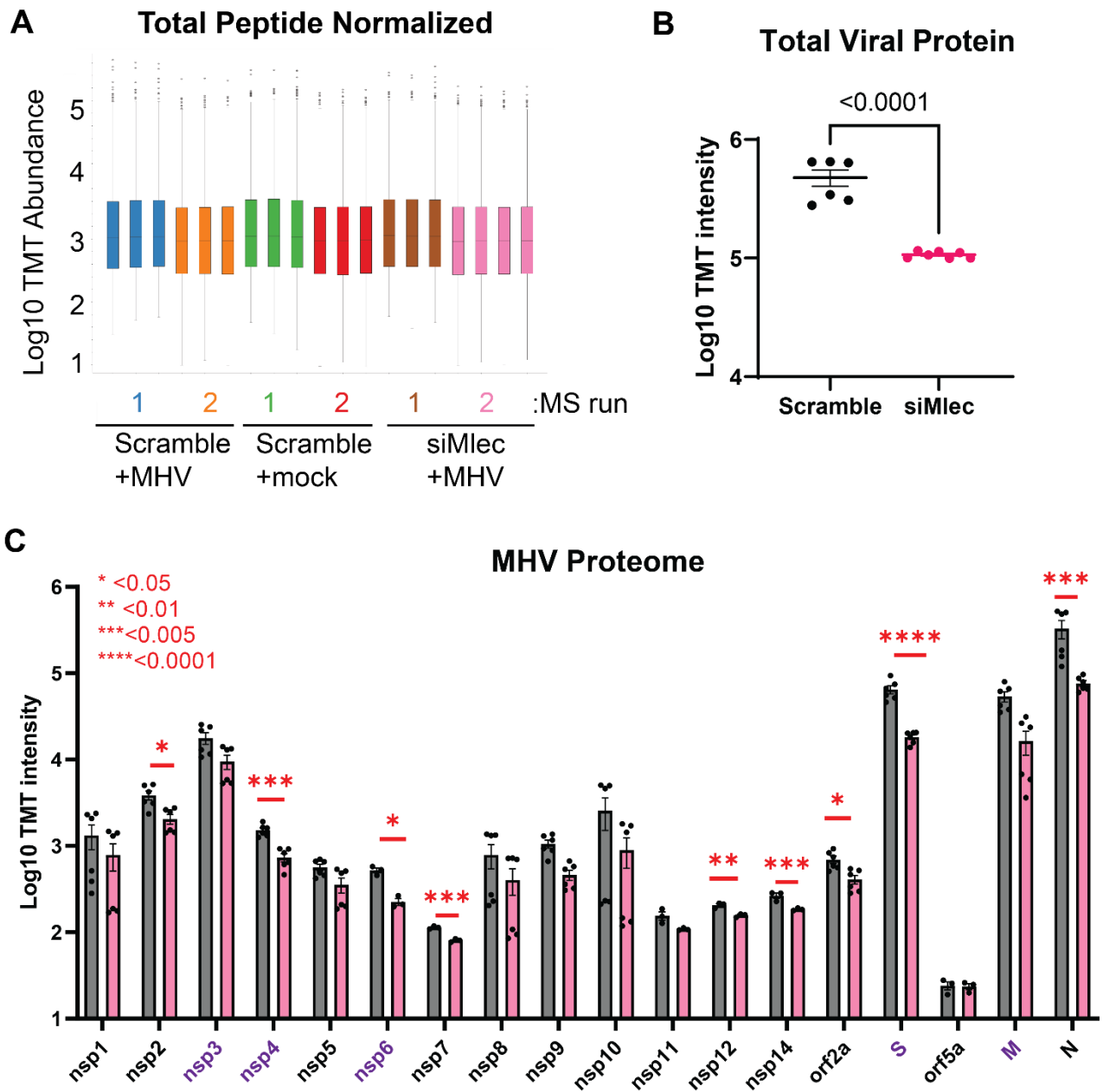
Figure 9-13. Effect of Mlec KD on intracellular MHV proteome.

A. Total normalized TMT channel intensity for global MHV proteome in DBT cells (MOI 1, 10 hpi). Scramble, n = 6; siMlec, n = 7 biological replicates across 2 MS runs.

B. Total viral protein levels in Scramble vs. siMlec treated DBT cells. One-way ANOVA with multiple testing correction.

C. Raw TMT abundance of intracellular viral proteins during MHV infection in DBT cells treated with Scramble or siMlec siRNA (MOI 1, 10 hpi), measured by quantitative proteomics. Purple protein names denote viral glycoproteins. Spike protein (S), membrane protein (M), nucleocapsid protein (N). One-way ANOVA with multiple testing correction. * p<0.05, ** p<0.01, ***p<0.005, ****p<0.0001.
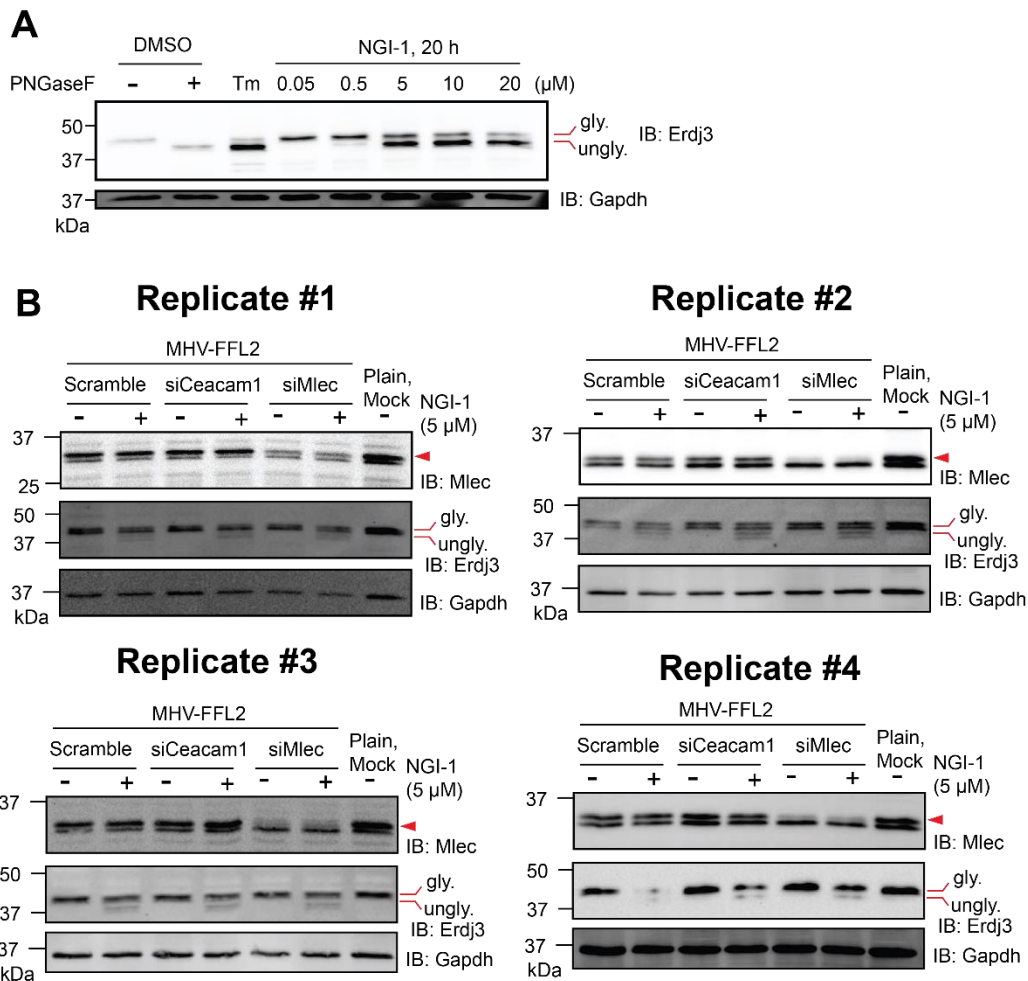
Figure 9-14. NGI-1 inhibition of the oligosaccharide transfer complex (OST).

A. Western blot of DBT cell lysates treated with a range of NGI-1 doses (0.05, 0.5, 5, 10, 20 µM) for 20 h prior to harvest. Treatment with DMSO or Tunicamycin (Tm, 1 µg/mL, inhibitor of upstream oligosaccharide synthesis) were used as controls. DMSO treated lysates were split in half, with one half treated with PNGase F to remove *N*-linked glycans and induce a gel shift on glycosylated proteins. Probing for Erdj3 (a glycosylated ER chaperone) was used to measure changes in glycosylation. Gapdh blotting used as loading control.

B. Western blots of DBT lysates treated with siRNA (40 h, Scramble, siCeacam1, or siMlec), then drugged with NGI-1 (5 µM) or DMSO and infected with MHV-FFL2 (MOI 1, 10 hpi). Plain cells were untreated, uninfected to serve as background luminescence control. Lysates were probed for Mlec to verify knockdown, Erdj3 to verify glycosylation inhibition, and Gapdh as loading control.
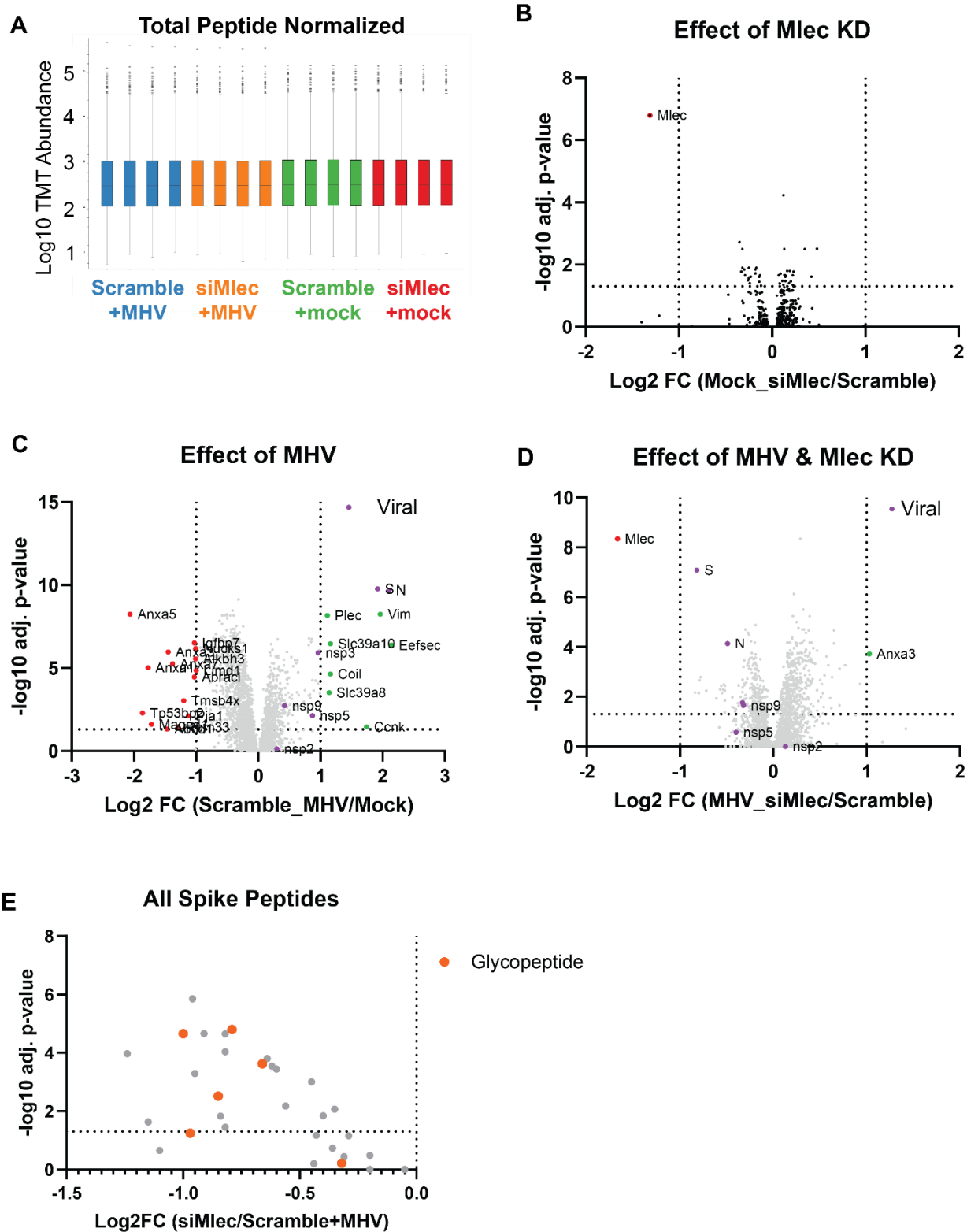
Figure 9-15. Glycoproteomics metrics for MHV infection.

A. TMT channel abundance normalized to total peptide abundance. n = 4/condition, 1 MS run
B. Volcano plot showing the effect of Mlec KD on the cellular proteome in uninfected DBT cells (Mock-siMlec/Mock-Scramble). Dotted lines indicate log2 FC < -1 or > 1 and adj. p-value <0.05. ANOVA test with multiple testing corrections for significance. Downregulated (red), upregulated (green), and viral (purple) proteins are annotated.

C. Same as in (B), but showing the effect of MHV infection on Scramble-treated DBT cells (MHV-Scramble/Mock-Scramble).
D. Same as in (B), but showing the effect of Mlec-KD on MHV infection (MHV-siMlec/Mock-Scramble).
E. Volcano plot showing the effect of Mlec-KD on Spike peptides in MHV infection (MHV-siMlec/MHV-Scramble). Glycopeptides are highlighted (orange). ANOVA test with multiple testing corrections for significance.

References

1. Davies, J. P., Almasy, K. M., McDonald, E. F. & Plate, L. Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *Acs Infect Dis* **6,** 3174–3189 (2020).

Supplemental figures for Chapter 5.
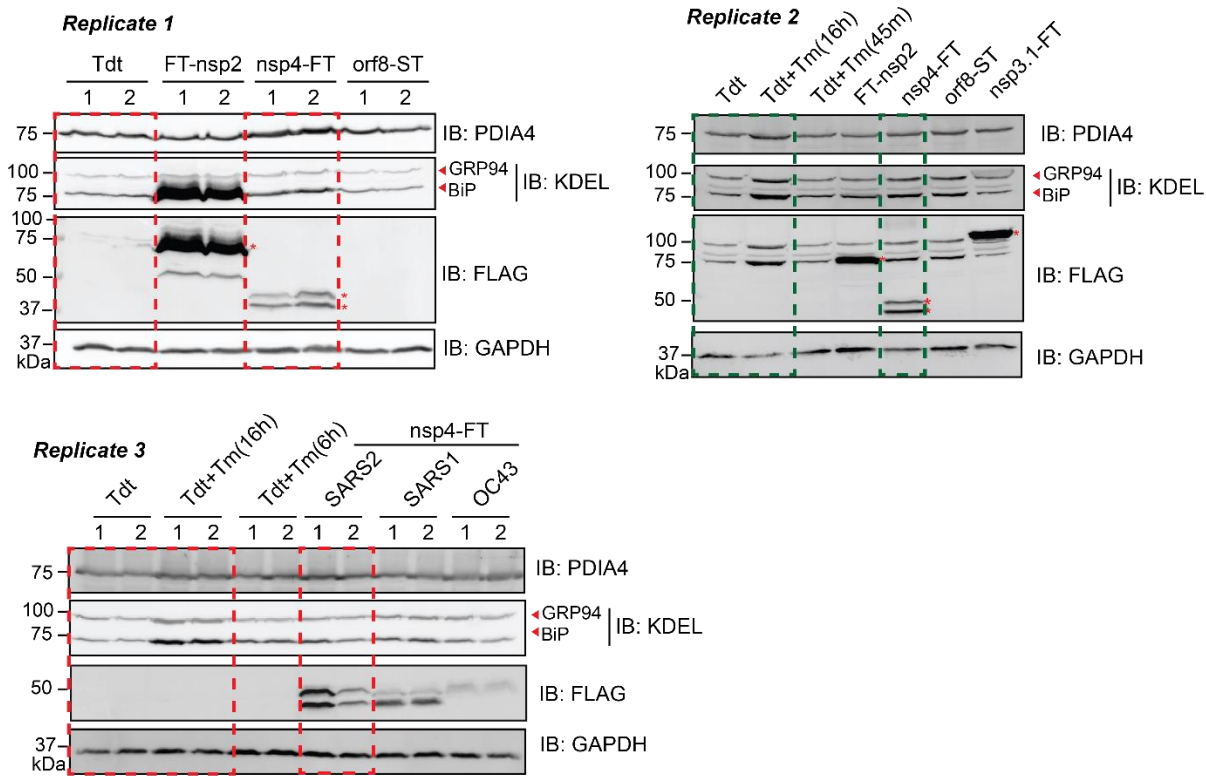
Figure 10-1. Western blot analysis of UPR markers.

Western blot analysis of HEK293T cells transfected with specified viral proteins or Tdtomato as basal control. Tm treatment was used as a positive control. Green boxed lanes were shown for Fig. 1c. Green and red boxed lanes were quantified for Fig. 1d. Red asterisks indicate viral proteins. Tdt, tdTomato; Tm, Tunicamycin (1 µg/mL); FT, FLAG-Tag; ST, StrepTag. Corresponding full blots are shown below.
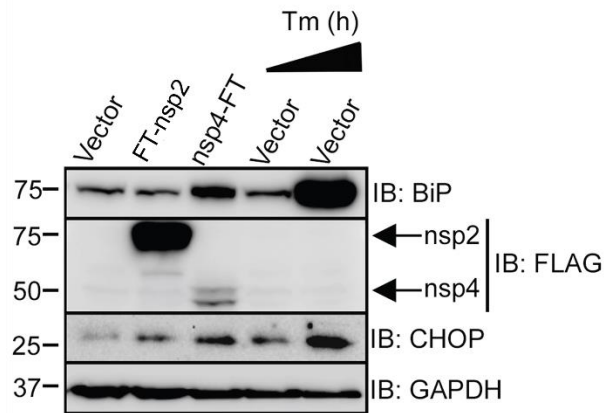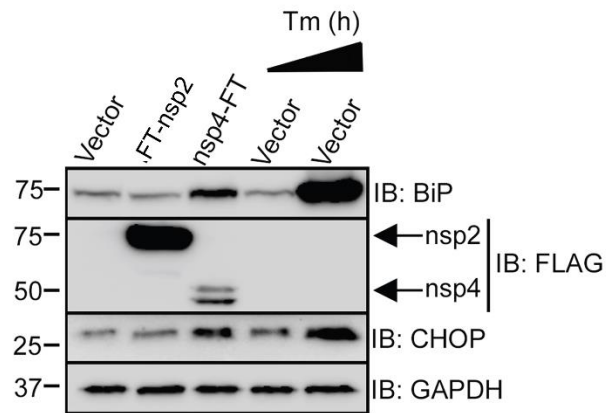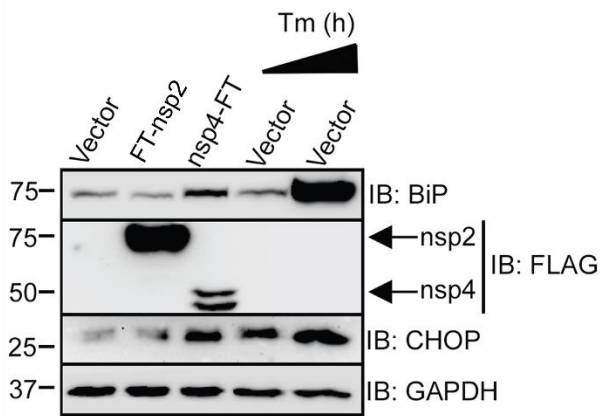
Figure 10-2. Western blots of CHOP and BiP upregulation.

Blots quantified for **Fig. 1f**. Replicate 2 blot was shown in **Fig. 1e**. WT-HEK293T cells were transfected with empty vector or SARS-CoV-2 FT-nsp2 or nsp4-FT, with control vector samples treated for 4 or 20 h Tunicamycin (5 µg/mL). Western blotting for M2-FLAG, BiP, and CHOP shows marked upregulation of BiP and CHOP in nsp4-FT expressing cells. Corresponding full blots are shown below.
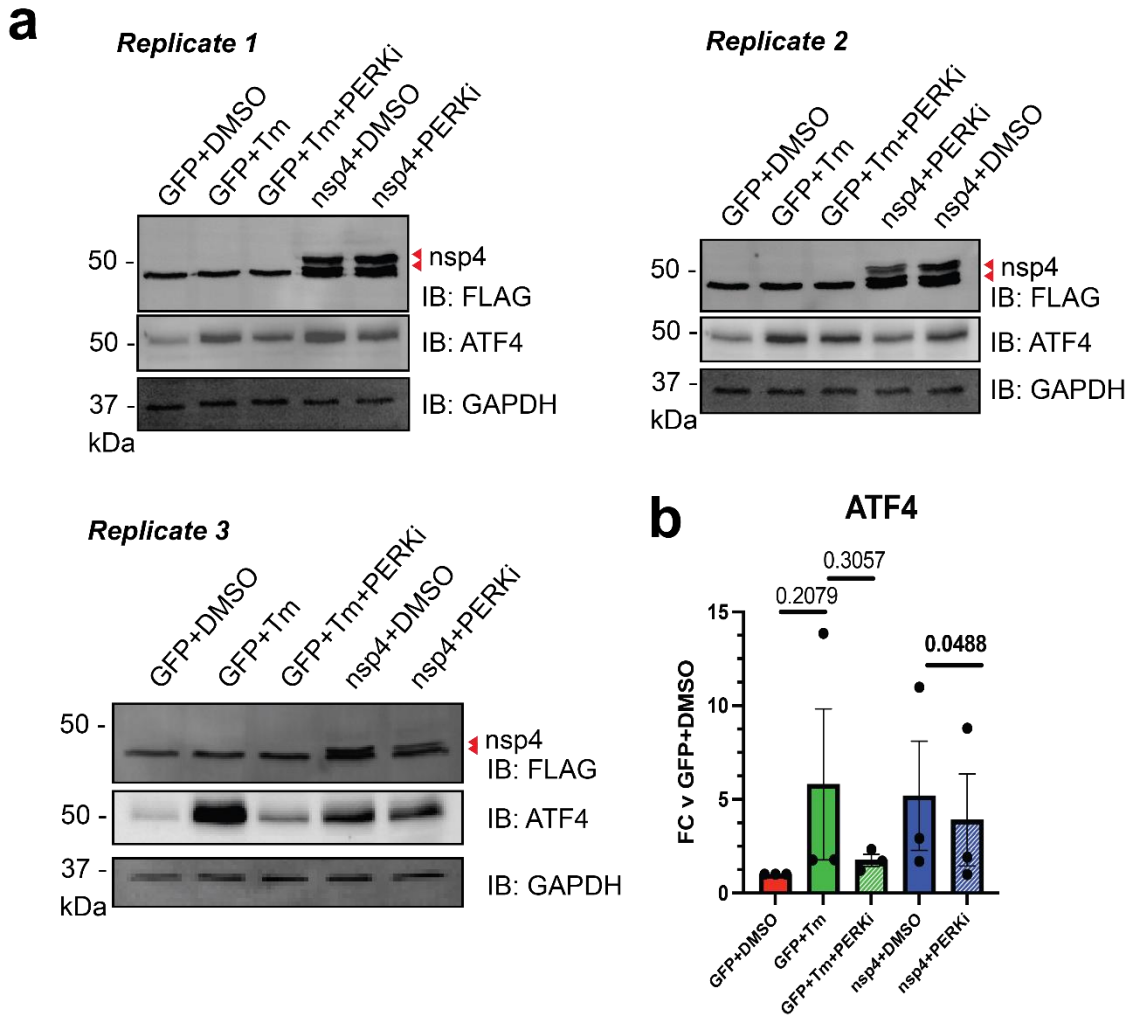
Figure 10-3. Effect of PERK inhibition on ATF4 upregulation mediated by nsp4.

HEK293T cells were transfected with GFP or SARS-CoV-2 nsp4-FT, with GFP samples treated with either Tunicamycin (Tm, 1 µg/mL) or Tunicamycin and PERK inhibitor II GSK2656157 (PERKi, 1 µM) for 16h and nsp4-FT samples treated with PERK inhibitor II GSK2656157 (PERKi, 1 µM) for 36h.

    a)   Western blots were probed for FLAG, ATF4, and GAPDH. Corresponding full blots are shown below.

    b)   Quantification of Western blots in **(a)**, normalized to GAPDH band intensities and relative to basal control (GFP+DMSO). n = 3, mean ±SEM. Paired T-test was used to test significance, p<0.05 considered statistically significant.
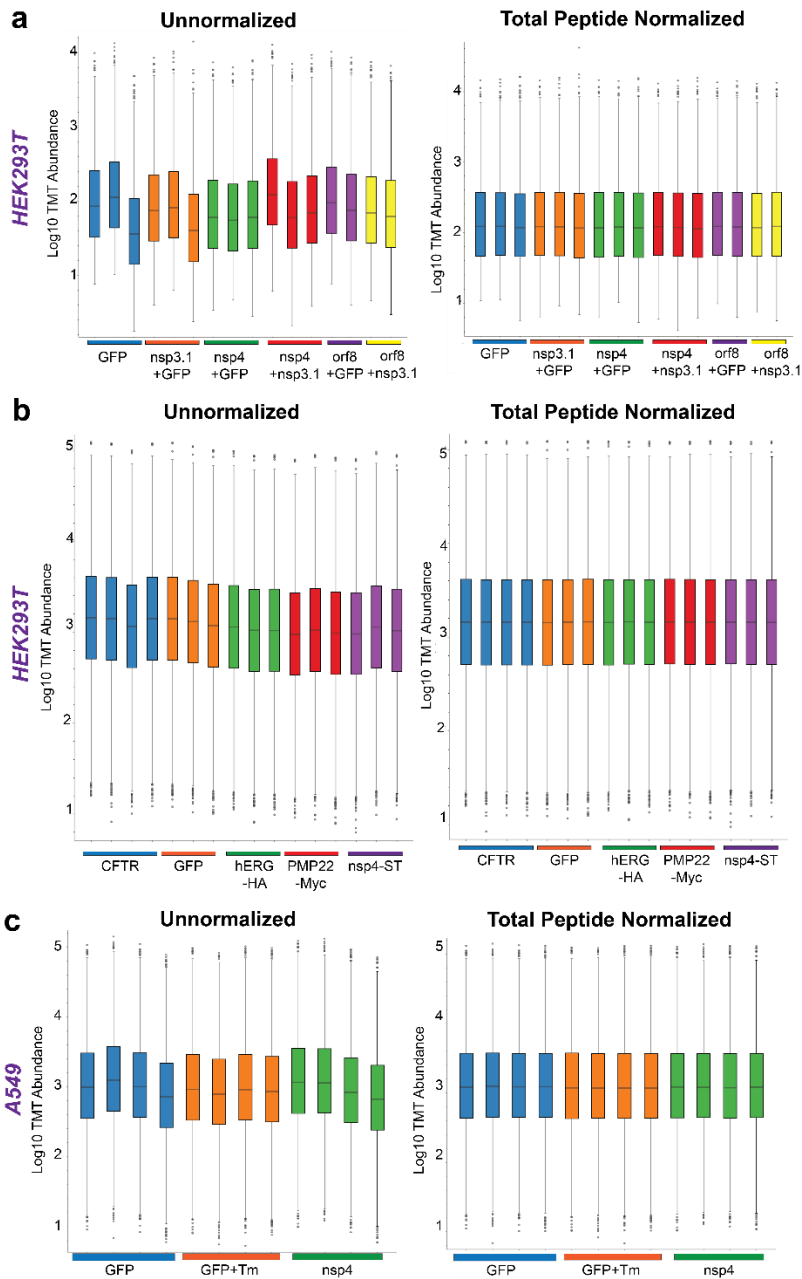
Figure 10-4. TMT channel abundances

HEK293T **(a,b)** or A549 **(c)** cells were transfected with GFP (basal control) or corresponding proteins, lysates harvested and lysed at 40 h post-transfection, labeled with TMTpro isobaric labels, and analyzed by tandem mass spectrometry (LC/MS-MS). Identified proteins were normalized by total peptide amount.

a) Unnormalized and normalized log10 TMT abundances for MS experiment analyzing HEK293T UPR, corresponding to **Fig. 2,3**. See **Supplemental Tables S2, S3**.

b) Unnormalized and normalized log10 TMT abundances for HEK293T cells expressing transmembrane protein panel, corresponding to **Supplemental Fig. S8**. See **Supplemental Tables S4, S5**.

c) Unnormalized and normalized log10 TMT abundances for MS experiment analyzing A549 UPR, corresponding to **Fig. 2**. See **Supplemental Tables S6, S7**.
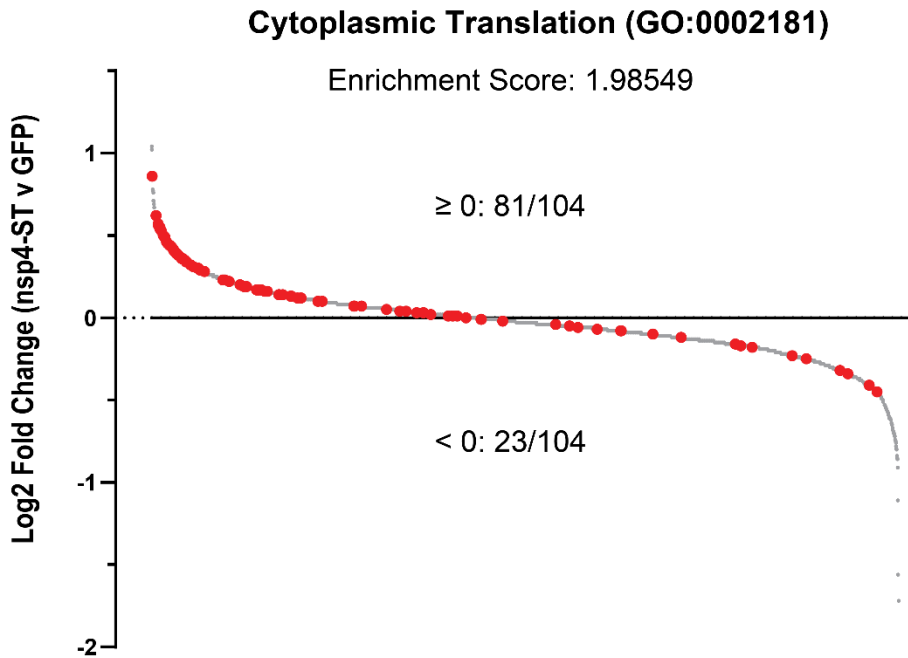
Figure 10-5. Waterfall plot of proteome in HEK293T cells expressing SARS-CoV-2 nsp4-ST vs GFP control

A ranked list enrichment analysis was applied to identified proteins based on Log2 Fold Change (nsp4-ST vs GFP) (using the STRING database(1)). Only one Gene Ontology (GO) Biological Processes term was significantly enriched (Cytoplasmic Translation, GO:0002181), with 81/104 genes having a positive Log2 Fold Change and 23/104 genes having a negative Log2 Fold Change. Proteins within the Cytoplasmic Translation GO Term are highlighted in red.
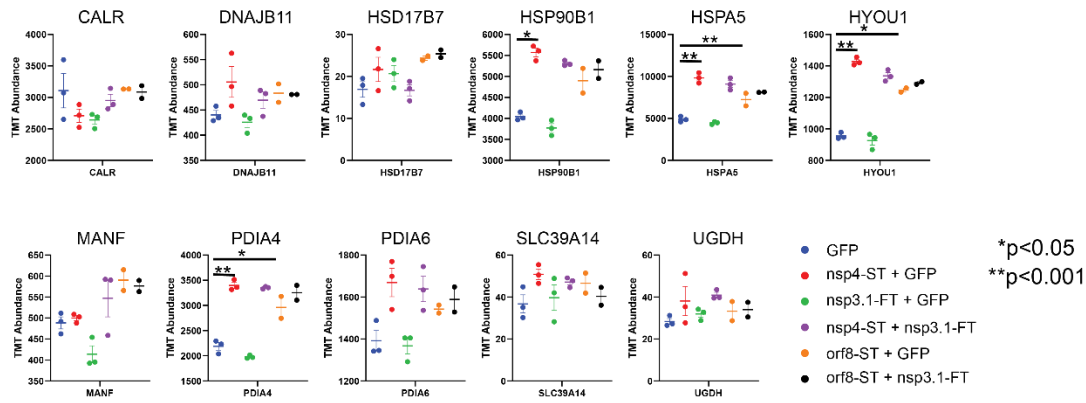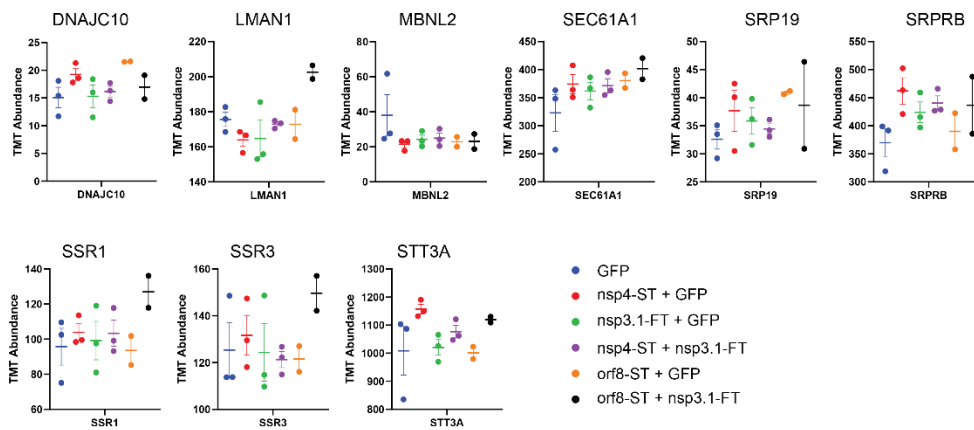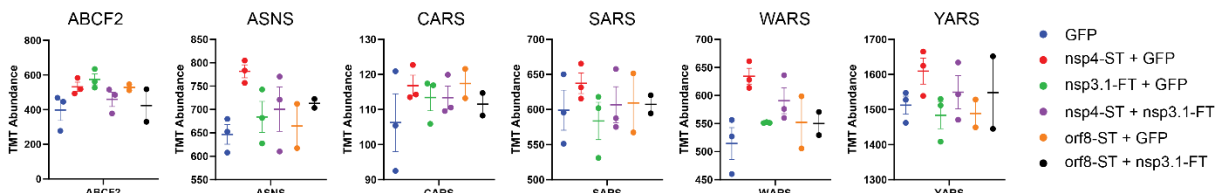
Figure 10-6. Normalized TMT abundance levels of individual Unfolded Protein Response (UPR) pathway protein markers

UPR markers in HEK293T cells expressing various viral proteins (corresponding to Fig. 2, 3). Adjusted p-value calculated using a T-test, controlling for FDR by the Benjamini-Hochberg method. *$p<0.05$, **$p<0.001$
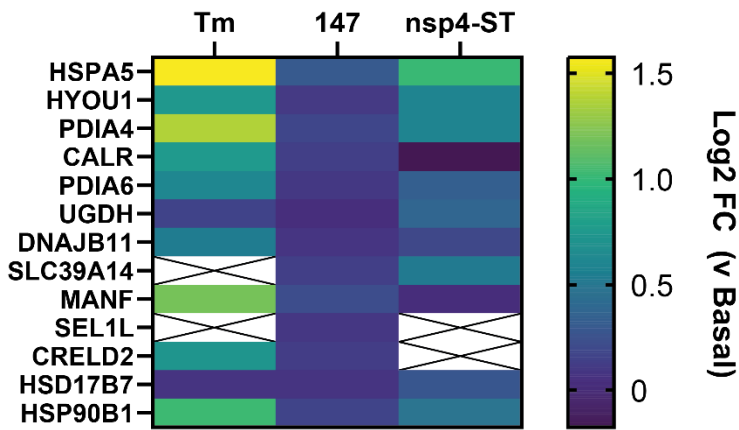
Figure 10-7. Comparison of ATF6 markers with other studies

ATF6 marker log2 fold change enrichment over basal conditions in HEK293T cells with treatment of Tm (1 µg/mL, 16 h), **147** (10 µM, 16 h), or nsp4-ST expression. Samples were quantified by TMTpro-based LC/MS-MS. Basal conditions are Tdtomato transfection with DMSO for drug treatment or GFP transfection for nsp4-ST treatment. Previously published data from Almasy, Davies, Plate *MCP* 2021 is annotated(2).
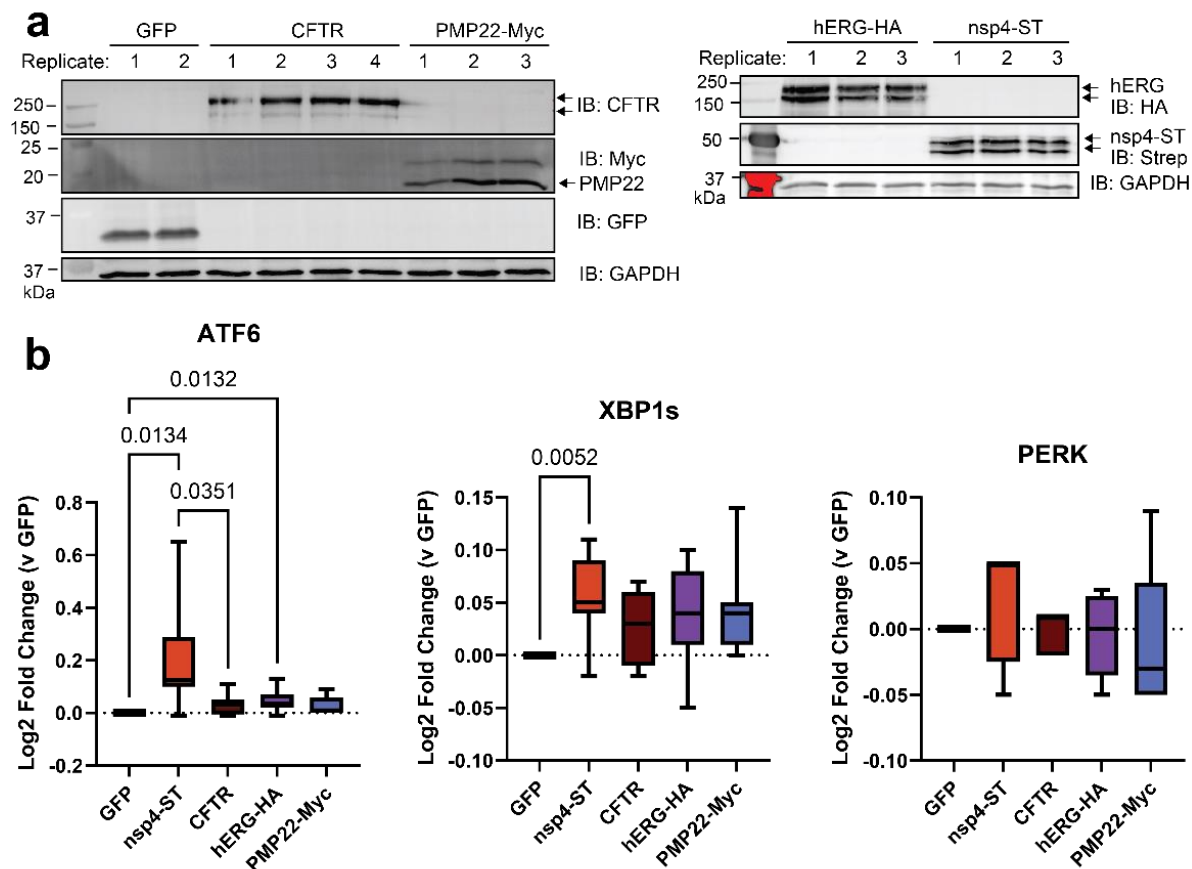

Figure 10-8. Effect of various multi-pass transmembrane proteins on the UPR as compared to nsp4.

Nsp4-ST, WT Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), HA-tagged human Ether-a-go-go Related Gene (hERG-HA), or myc-tagged peripheral myelin protein 22 (PMP22-Myc) were expressed in HEK293T cells and harvested 40 h post-transfection.

a) Western blot confirmation of protein expression in HEK293T lysates.

b) UPR branch reporter protein levels were measured by global proteomics using tandem mass spectrometry (LC/MS-MS) and compared to GFP negative control. One-way ANOVA with Geisser–Greenhouse correction and post-hoc

Tukey's multiple comparison test was used to determine significance, $p<0.05$ considered significant, significant p values shown; 1 MS run, n = 3-4 biological replicates. See **Supplemental Tables S4, S5** for mass spectrometry data set.
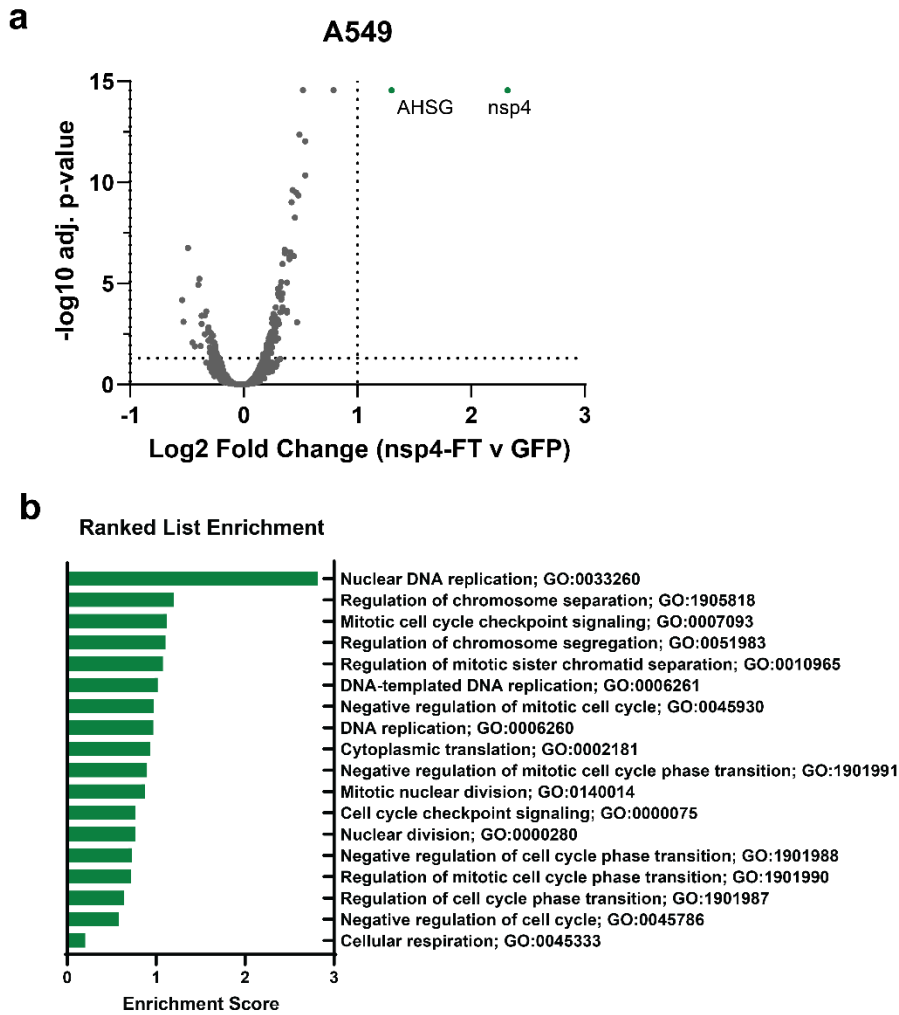
a

b

Figure 10-9. Global proteome analysis of A549 cells expressing SARS-CoV-2 nsp4-FT vs GFP control (corresponding to Fig. 2).

a) Volcano plot of global proteome, with cut-offs indicating Log2 FC (vs GFP) < -1 or > 1, adjusted p-value < 0.05. Common contaminant proteins were filtered out. Proteins passing cut-offs are annotated. Adjusted p-value calculated using a T-test, controlling for FDR by the Benjamini-Hochberg method. N = 4, 1 MS run.
b) A ranked list enrichment analysis was applied to identified proteins based on Log2 Fold Change (nsp4-FT vs GFP) (using the STRING database(1)).
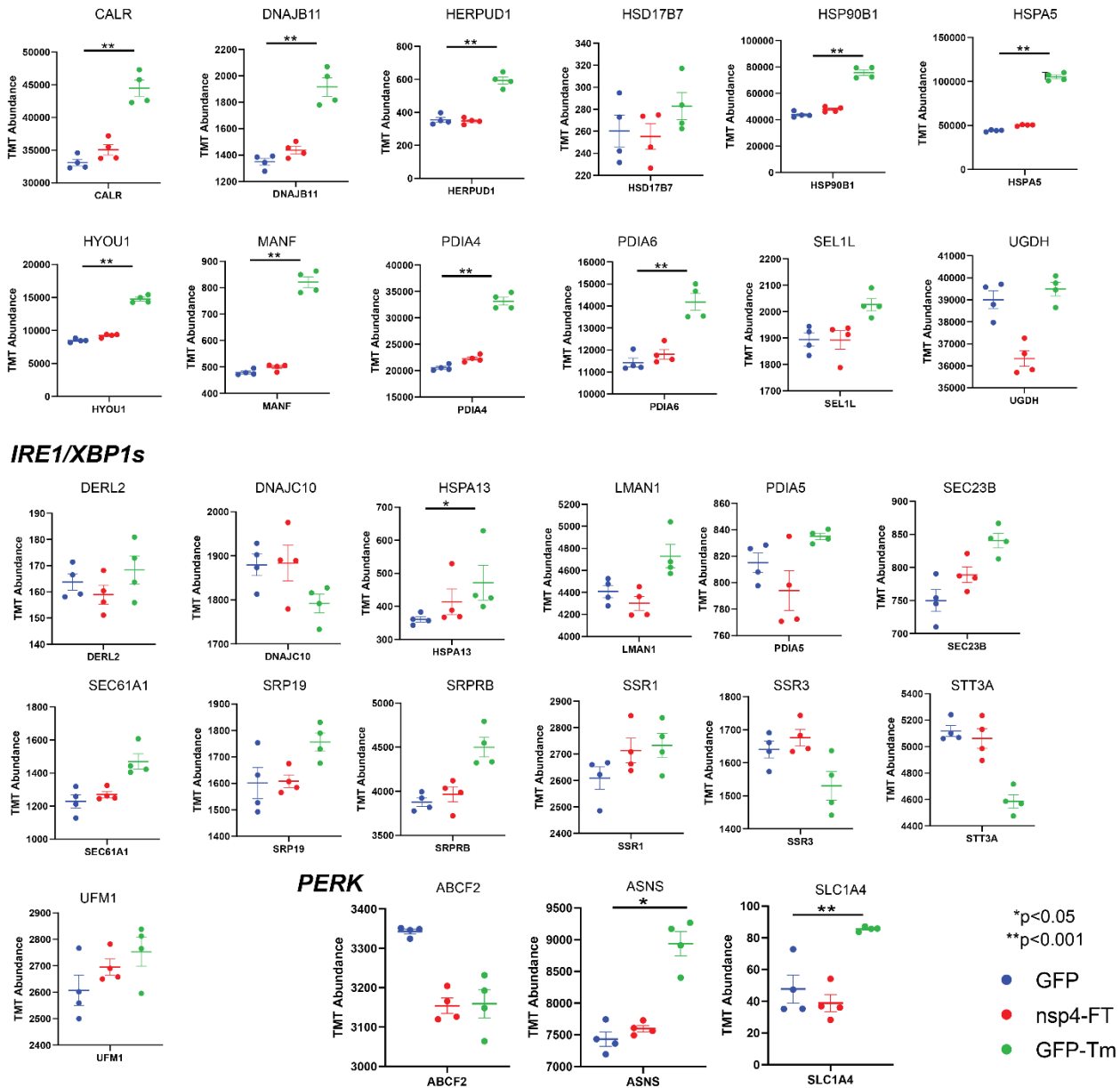
Figure 10-10. Normalized TMT abundance levels of individual Unfolded Protein Response (UPR) pathway protein markers

UPR markers in A549 cells expressing nsp4-FT or treated with Tm (corresponding to Fig. 2). Adjusted p-value calculated using a T-test, controlling for FDR by the Benjamini-Hochberg method. $*p<0.05$, $**p<0.001$
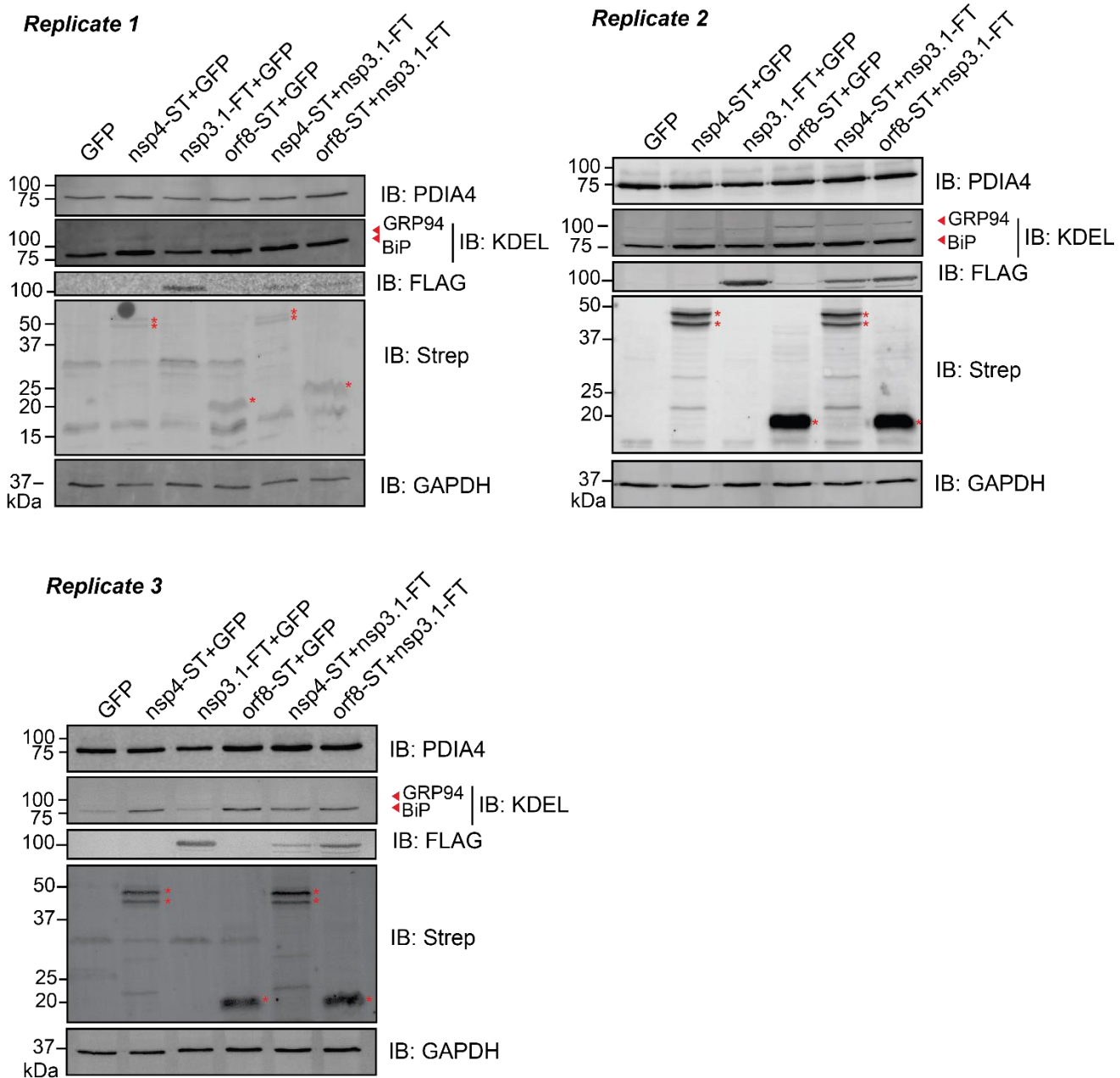
196

Figure 10-11. Western blots of nsp3.1, nsp4, and orf8 expression combinations.

Blots quantified for **Fig. 3c**, replicate 2 was shown in **Fig. 3b**. Red asterisks indicate viral proteins. Corresponding full blots are shown below.
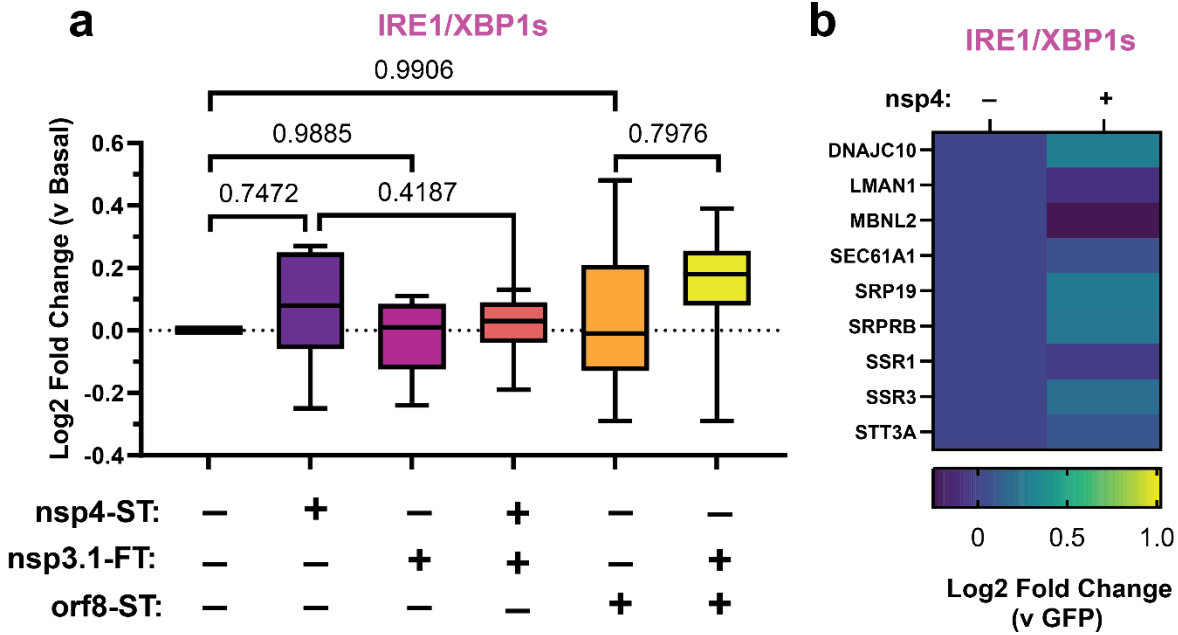
Figure 10-12. Upregulation of IRE/XBP1s proteomics protein markers.

a) Global proteomics analysis of IRE1/XBP1s protein marker levels in the presence of SARS-CoV-2 nsp4-ST, nsp3.1-FT, orf8-ST, or specified combinations. Box-and-whisker plot shows median, 25th and 75th quartiles, and minimum and maximum values. One-way ANOVA with Geisser–Greenhouse correction and post-hoc Tukey's multiple comparison test was used to determine significance, $p<0.05$ considered significant; 1 MS run, n = 2-3 biological replicates. See **Supplemental Tables S2, S3** for mass spectrometry data set.

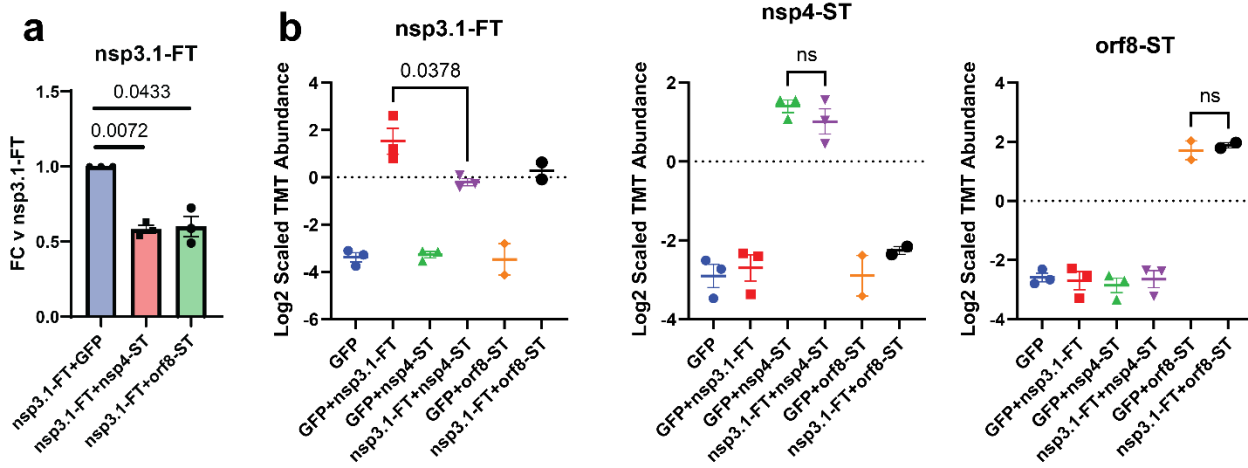b) Individual IRE1/XBP1s pathway protein markers in the absence or presence of nsp4-ST, from **(a)**.

Figure 10-13. Viral protein abundance levels.

a) Quantification of the nsp3.1-FT bands shown in **Fig. 3, Supplemental Fig. S11**. n = 3. T-test for significance between samples, *p*-value annotated.

b) SARS-CoV-2 nsp3.1-FT, nsp4-ST, and orf8-ST protein abundances (log2 scaled TMT abundances) in experiment corresponding to **Fig. 3.** n = 2-3 biological replicates, 1 MS run, T-test for significance between indicated samples, *p*-value annotated.

References

1. Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
2. Almasy, K. M., Davies, J. P. & Plate, L. Comparative Host Interactomes of the SARS-CoV-2 Nonstructural Protein 3 and Human Coronavirus Homologs. *Mol. Cell. Proteomics* **20**, 100120 (2021).