OPEN ACCESS

University of BRISTOL

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Abed, Alaa S A

Title:
Comparative interactomic analysis of MERS-CoV & SARS-CoV-2 proteins in human, bat, and camel cells

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

•Your contact details
•Bibliographic details for the item, including a URL
•An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

University of BRISTOL

Comparative interactomic analysis of MERS-CoV & SARS-CoV-2 proteins in human, bat, and camel cells

Alaa Abed

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the School of Cellular and Molecular Medicine.

March 2023

Abstract

The Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are Betacoronaviruses capable of causing fatal human infections. Both viruses are believed to have emerged from bats *via* an intermediate host (camels for MERS-CoV, unknown for SARS-CoV-2) into the human population. MERS-CoV and SARS-CoV-2 are enveloped positive-sense RNA viruses which encode four structural proteins (envelope (E), nucleocapsid, spike, and membrane (M)). The E and M proteins are involved in virus assembly, budding, envelope formation, and pathogenesis. Finding cellular protein interactors for these viral proteins, conserved across species, will increase our understanding of the coronavirus lifecycle and identify targets for antiviral development.

Initial validation of the results for 11 cellular proteins interacting with the MERS-CoV E and/or M proteins by co-immunoprecipitation (co-IP) /Western blot analysis and immunofluorescence co-localisation, confirmed 10 of the interactions.

Three cell lines (human HEK293, bat *Pteropus alecto* PaKiT and *Camelus dromedarius* Dubca) were used for transient expression of the MERS-CoV and SARS-CoV-2 E and M proteins (FLAG epitope-tagged) followed by co-IP and high-throughput mass spectrometry-based interactomic analysis. Bioinformatic analysis revealed E/M protein interactions with ER, Golgi, mitochondrial and nuclear proteins. There were 32 high-confidence cellular interaction proteins conserved amongst the different cell lines and viruses ($p < 0.05$, > 0 $\log_2$ fold change compared to the controls).

To determine the importance of the 11 cellular proteins interacting with the MERS-CoV E and/or M proteins and 32 cellular proteins conserved across species, in the virus lifecycle, functional validation was done by siRNA depletion in human cells, followed by infection with SARS-CoV-2. An interesting four cellular proteins were shown to be important for SARS-CoV-2 replication. These interesting proteins include (CERS2, LPCAT1, UBA52, and TM9SF2) that when it was knocked down, succeeded in reducing the SARS-CoV-2 replication. This can be followed by working to find if these proteins can be safely targeted by a drug to be reduced in cells.

Dedication and Acknowledgements

In the name of Allah, the infinitely Compassionate and Merciful

A PhD is just a degree, but during these four years, I have learned a lot about life and science. Without the help of many, this learning would not have been possible. I want to thank them for that.

I want to start by expressing my gratitude to Prof. Andrew Davidson, my dear supervisor. Without his constant encouragement, counsel, recommendations, and unending patience, this work would not be possible. I am very fortunate to have him as my supervisor because I know it was a huge challenge to get the PhD science without lab experience and rudimentary science knowledge from the very first step through to the culmination, which is this work achievement.

I would like to acknowledge the Ministry of Health (MOH). Kingdom of Saudi Arabia, for granting my PhD.

I would like to thank my co-advisor, Prof. David Matthews, for his support. I also would like to thank Dr. Phil Lewis for his great effort and help in data analysis and inspiring talks over coffee and his impressive comments. I would like to thank Dr. Kate Heesom and the staff in the proteomic facility for their technical assistance throughout the study.

The E-50 virology lab was a wonderful place for me, I would like to thank the previous member, I'ah Donovan-Banfield, for her guidance and assistance during my first steps in the Lab. Maia Kavanagh Williamson, for her support and sharing her knowledge and expertise. Holly Baum, for advice and help. Dr Abdulaziz Almuqrin for being a wonderful friend.

Thanks to all members of our virology group, Max Erdmann, Shichun Gu, Andreea Nenciu, Dr. Edward Sullivan, Dr.Alexander Walker, and Dr. Katja Klein, for making our laboratory a friendly and pleasant place to work. Thanks Dr. Rachel Milligan and James Bazire for conducting biosafety level 3 experiments. Thanks Dr. Caitlin Simpson for her support and sharing her transfection and the siRNA experiment knowledge and expertise. Furthermore, I would like to thank all support for the E-floor technician team, especially Dr. Natalie Griffiths-Stubbs.

I would like to express my love and deepest gratitude to my brothers Basem Abed and Abdulrahim Abed and My sister Ola Abed for their friendship and continuous support throughout the study period and my difficult time. Their love is always beside me wherever I am and whenever I need it.

Finally, I would like to thank my parents for their invaluable and infinite love. I have no words to express how much I appreciated their love and support. I believe that my father who was my inspiration for PhD studies would be proud of me if he could see my achievement.

Last but not least, special thanks to my wife, Walaa Shibaj for being incredibly supportive, encouraging, and loving during my PhD journey. This thesis is dedicated to you. Not forgetting my beloved son and daughter Abdulrahim and Rama for adding joy and happiness to my life.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed ALAA ABED Date: 30 03 2023

Table of Contents

List of Figures

List of Tables

Abbreviations

+ssRNA	Positive single strand RNA
3' UTR	3' untranslated regions
A549	human adenocarcinoma alveolar basal epithelial cells
A549-A2-T2	A549 cells stably expressing ACE2, and TMPRSS2
A549-ACE2	A549 cells stably expressing ACE2
AA	amino acid
AAV	adeno-associated virus
ACE2	angiotensin-converting enzyme 2
AGC	automatic gain control
ARDS	acute respiratory distress syndrome
ATP	adenosine triphosphate
BTRS	body transcription regulatory sequences
Calu-3	human airway-derived cells
Cas	CRISPR-associated endonuclease protein
CD	conserved domain
CDC	Centers for Disease Control
CFR	case fatality rate
co-IP	co-immunoprecipitation
COVID-19	coronavirus disease 19
CRAPome	Contaminant Repository for Affinity Purification database
CRISPER	clustered regularly interspaced short palindromic repeats
DAVID	Database for Annotation, Visualization and Integrated Discovery
DF1	DharmaFECT 1
DMEM	Dulbecco’s modified Eagle’s medium
DMV	double membrane vesicles
DPP4	dipeptidyl peptidase 4
DUB	deubiquitinating enzyme
Dubca	*Camelus dromedarius* fibroblast cells
E	envelope protein
EASE	Fisher Exact Statistics in DAVID system
EBV	Epstein–Barr virus
EGFR	epidermal growth factor receptor
ELISA	enzyme-linked immunosorbent assay
EM	electron microscopy
ER	endoplasmic reticulum
Erp72	ER protein 72
ERGIC	ER-Golgi intermediate compartment
EUL	Emergency Use Listing
FCS	foetal calf serum
FDA	US Food and Drug Administration

FDR	false discovery rate
FGF	fibroblast growth factor
GCSF	granulocyte colony-stimulating factor
GFP	green fluorescent protein
GO	Gene ontology
GOBP	Gene ontology biological process
GOCC	Gene ontology cellular component
GOMF	Gene ontology molecular function
GRP94	Glucose regulated protein
h	hours
HA	hemagglutinin
HCD	high-energy collision dissociation
HCoV-229E	Human coronavirus 229E
HCoV-HKU1	human coronavirus HKU1
HCoV-NL63	human coronavirus NL63
HCoV-OC43	human coronavirus OC43
HD	hydrophobic domain
HEK293-Env	stable Flp-InTM T-RExTM-293 cells inducibly expressing FLAG epitope-tagged versions of the MERS-CoV E protein.
HEK293-Flp	The parental Flp-InTM T-RExTM-293 cell line
HEK293-M	stable Flp-InTM T-RExTM-293 cells inducibly expressing FLAG epitope-tagged versions of the MERS-CoV M protein.
HEK293T	human embryonic kidney cells
HELaM	Human Cervical cancer cell line M
HMOI	high multiplicity of infection
HIV	human immunodeficiency virus
hpt	hours post-transfection
hpi	hours post-infection
HRP	immunoglobulin/horseradish peroxidase
Huh-7	hepatic carcinoma-derived cells
IBDV	infectious bursal disease virus
IBV	infectious bronchitis virus
ICU	intensive care unit
IFA	immunofluorescence assay
IFN	interferon
IFNs	interferons
IgG	immunoglobulins G
IgM	immunoglobulins M
IL	Interleukin
IMEx	International Molecular Exchange
IP	immunoprecipitation
IRF3	interferon regulatory transcription factor 3
LC	liquid chromatography
LC-MS/MS	liquid chromatography /mass spectrometry
lipoRM	Lipofectamine RNAiMAX
LMOI	low MOI
LRT	lower respiratory tract

LTRS	leader Transcription regulatory sequences
M	membrane protein
M2	matrix-2 protein
MERS	Middle East Respiratory Syndrome
MERS-CoV	Middle East respiratory syndrome - coronavirus
MHV	Mouse Hepatitis Virus
min	minute
MOI	multiplicity of infection
mRNA	Messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
N	Nucleocapsid Protein
NAAT	Nucleic Acid Amplification test
nap	non-structural proteins
NAT	Nucleic acid Testing
NEAA	non-essential amino acids
NGI-1	N-linked glycosylation inhibitor-1
nsp	non-structural protein
Opti-MEM	serum-free medium
ORF	open reading frame
ORF1a	open reading frame 1a gene
ORF1b	open reading frame 1b gene
OST	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase
Pakit	*Pteropus alecto* kidney cells
PBS	phosphate buffered saline
PBST	phosphate buffer saline + 1% Tween
PCC	Pearson's Correlation Coefficient
PI3K	Phosphatidylinositol 3-kinase 3
PLP	papain-like protease
PPE	personal protective equipment
PPI	protein protein interactions
PTGES2	prostaglandin E synthase 2
PVDF	polyvinylidene difluoride
RdRp	RNA-dependent RNA polymerase
RNAi	RNA interference
ROI	region of interest
RPL40	ribosomal protein L40
rRT-PCR	real-time RT-PCR
rSARS-CoV-E	recombinant SARS-CoV
rSARS-CoV-ΔE	The recombinant virus lacking the E gene
RT	room temperature
RT- PCR	real-time reverse transcription-polymerase chain reaction
S	Spike protein
SARS-CoV	severe acute respiratory syndrome coronavirus
SARS-CoV-2	Acute Respiratory Syndrome coronavirus-2

SCAP	Sterol regulatory element-binding protein cleavage-activating protein
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sgRNA	single guide RNA
siGLO	green fluorescein amidite
SILAC	stable isotope labelling of proteins by amino acids in cell culture
siRNA	small interfering RNA
Smad7	Mothers against decapentaplegic homolog 7
SPS	Synchronous Precursor Selection
SREBP	sterol regulatory element-binding protein signalling
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
STT3A	Catalytic subunits A
Tet	tetracycline
TGN	trans-Golgi network
Th1	type I T helper cells
Th2	type II T helper
TM	transmembrane
TMD	transmembrane domain
TMEM41	Transmembrane protein 41B
TMPRSS2	transmembrane protease serine 2
TMT	tandem mass tagging
TRITC	Tetramethylrhodamine
TRIM4	tripartite motif containing 4
UBA52	Ubiquitin-60S ribosomal protein L40
UP	Uniprot
upE	upstream region of the E gene
UPR	Unfolded protein response
UPS	ubiquitin proteasome system
URT	upper respiratory tract
US	United state
UTR	untranslated region
V	volt
VLPs	virus-like particles
VMP1	Vacuole membrane protein 1
WB	Western blot
WHO	World Health Organization

CHAPTER 1 General Introduction

1.1 Coronavirus as a global public health problem

1.1.1 History and current global status

Nearly all viruses that have recently emerged in human populations and cause significant health problems are zoonotic. They include the human immunodeficiency virus (HIV), avian influenza viruses, Ebola virus, and recently, the coronaviruses associated with the Middle East respiratory (MERS-CoV) and severe acute respiratory (SARS-CoV and SARS-CoV-2) syndromes (1). MERS-CoV have emerged in human, avian, and mammalian populations, including pigs, birds, cats, dogs, whales, mice, and camels (2). In most cases, bats have been identified as the natural reservoirs of these emergent coronaviruses (3).

Seven human coronaviruses have been identified to date. SARS-CoV emerged in humans in 2003, causing an outbreak lasting >24 weeks in 30 countries, with 8000 infected individuals and 774 deaths (4, 5). The distribution of SARS-CoV is shown in Figure 1-1. Previously, the two known human coronaviruses caused only mild respiratory disease. Human coronaviruses 229E (HCoV-229E) and OC43 (HCoV-OC43) were identified as causes of the common cold in 1960 (6). Human coronaviruses NL63 (HCoV-NL63) and HKU1 (HCoV-HKU1) are the fourth and fifth discovered (both in 2004) and cause mild to severe respiratory infections (7). HCoV-NL63 was initially identified in the nasopharyngeal aspirate of a seven-month-old child in Amsterdam with symptoms that suggested a respiratory tract infection (8). In contrast, HCoV-HKU1 was identified in a 71-year-old man with pneumonia in Hong Kong in the same year (9). Since 2004, two additional human coronaviruses have been identified, MERS-CoV and SARS-CoV-2.

MERS-CoV was initially named human coronavirus-Erasmus Medical Center/2012 before being revised to Middle East respiratory syndrome coronavirus (2). The first reported human MERS-CoV case was in Saudi Arabia in June 2012, when a male patient died due to severe respiratory illness (10). The initial MERS-CoV infection case was followed by many others, becoming the first MERS-CoV outbreak. Epidemiological data suggest that Saudi Arabia is a

MERS-CoV endemic area (11). Since this initial outbreak, MERS-CoV has appeared in other countries, including those in the Middle East and Gulf (United Arab Emirates, Qatar, Bahrain, Oman, Iraq, Kuwait, Jordan, Syria, Lebanon, Egypt, and Palestine) and Europe, including France and the United Kingdom (12). The distribution of MERS-CoV is shown in Figure 1-2. The number of laboratory-confirmed MERS-CoV infection cases reported by the World Health Organization (WHO) was 2603, causing at least 935 deaths between 2012 and January 2023 (13). In addition, 2195 MERS-CoV cases were reported in Saudi Arabia between 2012 and January 2023 (14). MERS-CoV causes 34.4% mortality in susceptible individuals (15). The second most serious outbreak after Saudi Arabia occurred in South Korea in 2015 (16), with 39 deaths from 186 positive reported cases (17). It should be noted that immunocompromised patients with chronic renal failure after organ transplantation and diabetic patients are most likely to develop fatal MERS-CoV infections (6). MERS-CoV continues to cause disease and mortality today (11).

In December 2019, the novel SARS-CoV-2 emerged in humans in Wuhan, Hubei Province, China, causing a novel coronavirus disease (18). The first 41 cases were connected with exposure to the Huanan seafood market in Wuhan (18). This virus rapidly spread to other regions of China and to other countries, becoming a global pandemic (19, 20). The distribution of SARS-CoV-2 is shown in Figure 1-3. Patients were admitted to hospitals with pneumonia and respiratory distress and four criteria: (i) fever, (ii) radiographic pneumonia evidence, (ii) low or normal white-cell count or low lymphocyte count, and (iv) no improvement in symptoms after three days of antimicrobial treatment (21, 22). SARS-CoV-2 was named on 11 February 2020 by the WHO, who also named the disease it caused as coronavirus disease 19 (COVID-19) (21). This outbreak was recognised as a pandemic on 11 March 2020 (23). The WHO has recorded over 758,390,564confirmed cases and over 6,859,09 confirmed deaths worldwide up to 28 February 2023 (24).

COVID-19 has a 0.47%–1.4% mortality rate (25), influenced by several factors, including health, age, and chronic conditions such as diabetes, obesity, lung disease, and liver disease (26). COVID-19 spreads mostly through respiratory droplets and is known to be transmissible by both asymptomatic and symptomatic individuals (26). While many COVID-19 patients are asymptomatic, they can spread the virus to others (26). The prolonged pandemic has caused social division, travel restrictions, reduced trade, significant unemployment, commodity price declines, and financial hardship, all of which have harmed the global economy (26).

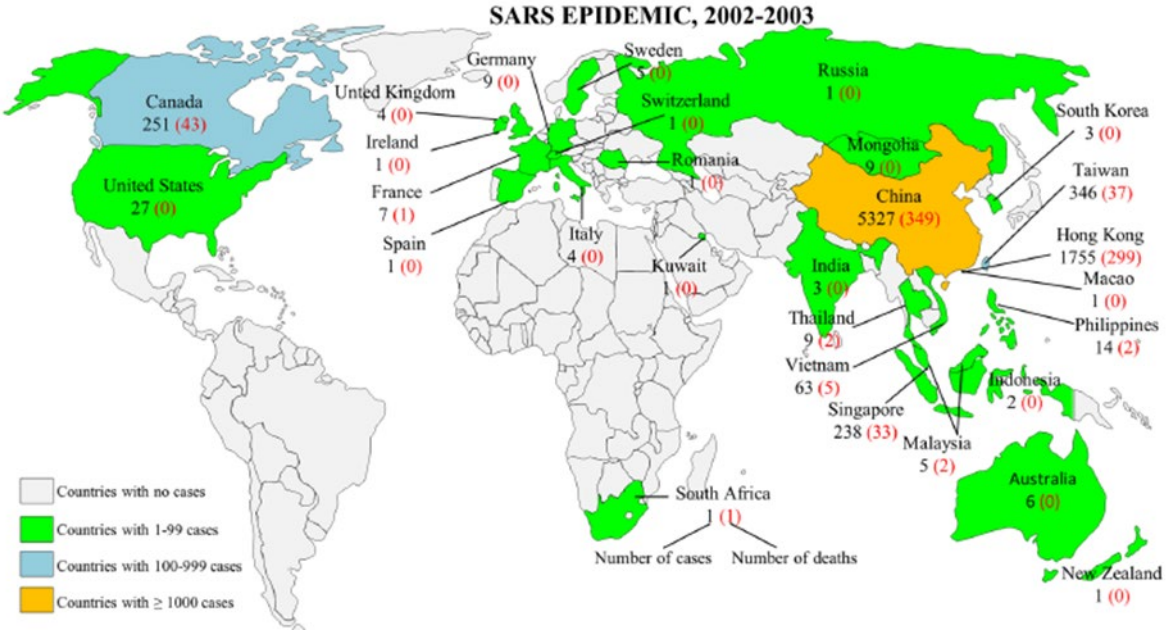


Figure 1-1 The worldwide distribution of SARS-CoV infections and the number of infected individuals.

The figure was taken from (27).

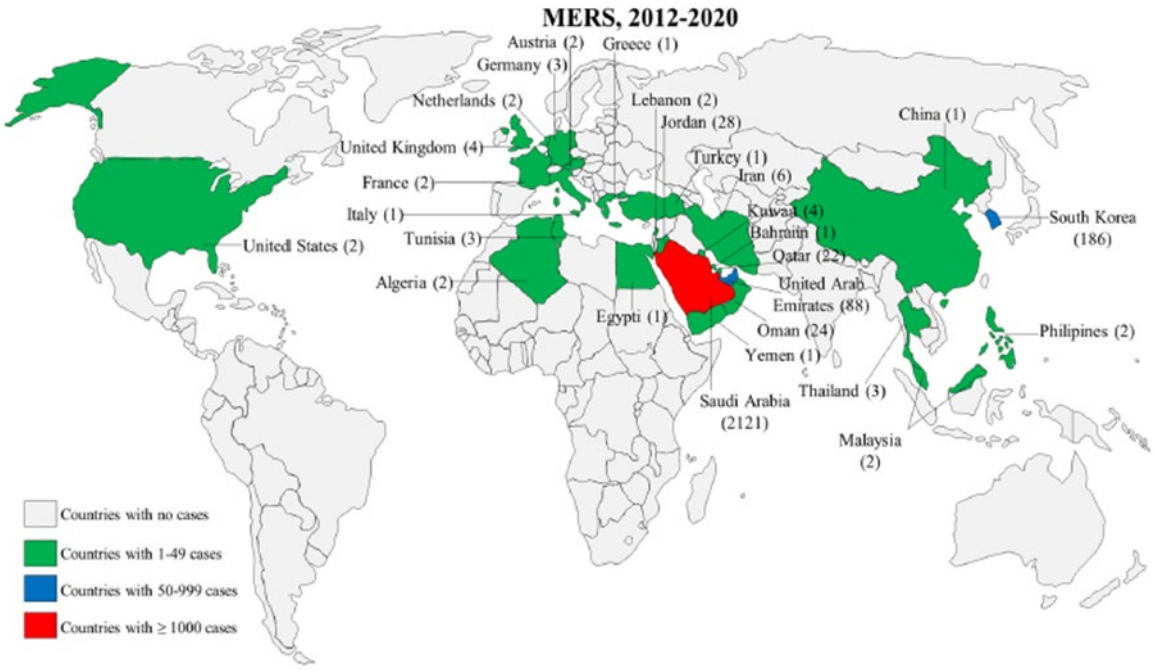


Figure 1-2 The worldwide distribution of MERS-CoV infections and the number of infected individuals.

The figure was taken from (27).

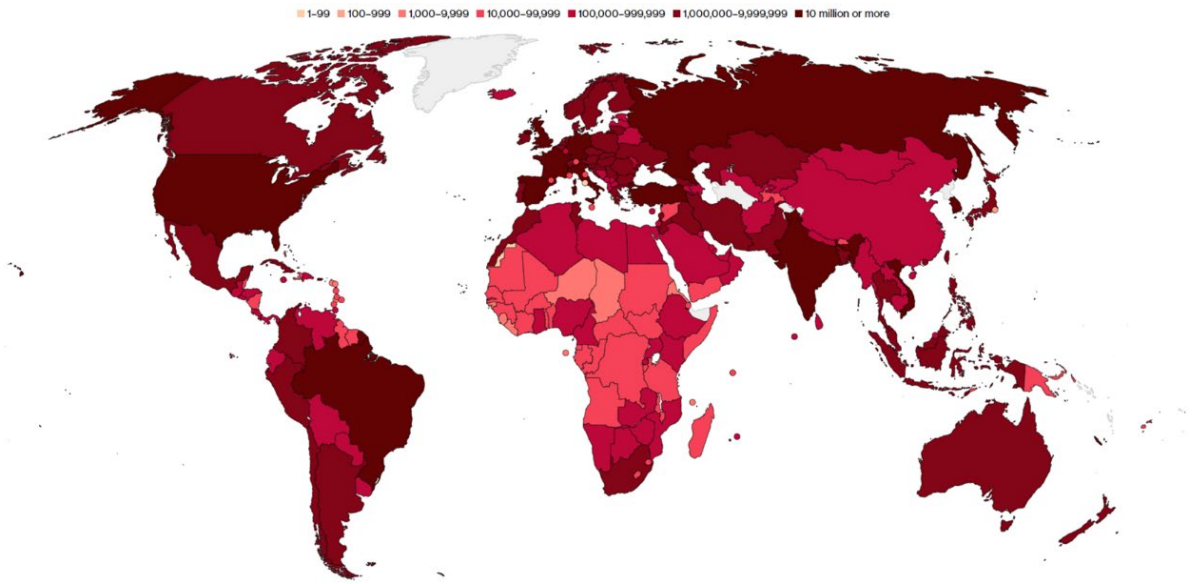


Figure 1-3 The worldwide distribution of SARS-CoV-2 infections and the number of infected individuals on 21 June 2022.

The figure was taken from (28).

1.1.2 Coronaviruses

The *Coronaviridae* family is in the *Nidovirales* order and is divided into two subfamilies: *Torovirinae* and *Coronavirinae*. The subfamily *Coronavirinae* contains four genera: *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus*, and *Gammacoronavirus* (Figure 1-4) (15). The *Coronaviridae* classification depends on antigenic relationships and comparisons between the latest viral genome sequences (16). Members of the *Alphacoronavirus* and *Betacoronavirus* genera infect mammals, including humans, and various animal species (cattle, camels, cats, dogs, rodents, bats, ferrets, civets, mink, snakes, and other wildlife). In contrast, members of the *Deltacoronavirus* and *Gammacoronavirus* genera mostly infect birds (29).

Of the seven coronaviruses that cause human disease, two are *alphacoronaviruses* (HCoV-NL63 and HCoV-229E) and the rest are *betacoronaviruses* (HCoV-OC43, HCoV-HKU1, MERS-CoV, SARS-CoV, and SARS-CoV-2) (17). Coronaviruses are enveloped RNA viruses (6) with characteristic spike-shaped glycoproteins embedded in the envelope (6). The positive-sense single-stranded coronavirus RNA genomes range from 25.5 to 32 kb in size (6), the largest RNA viral genomes known (18).

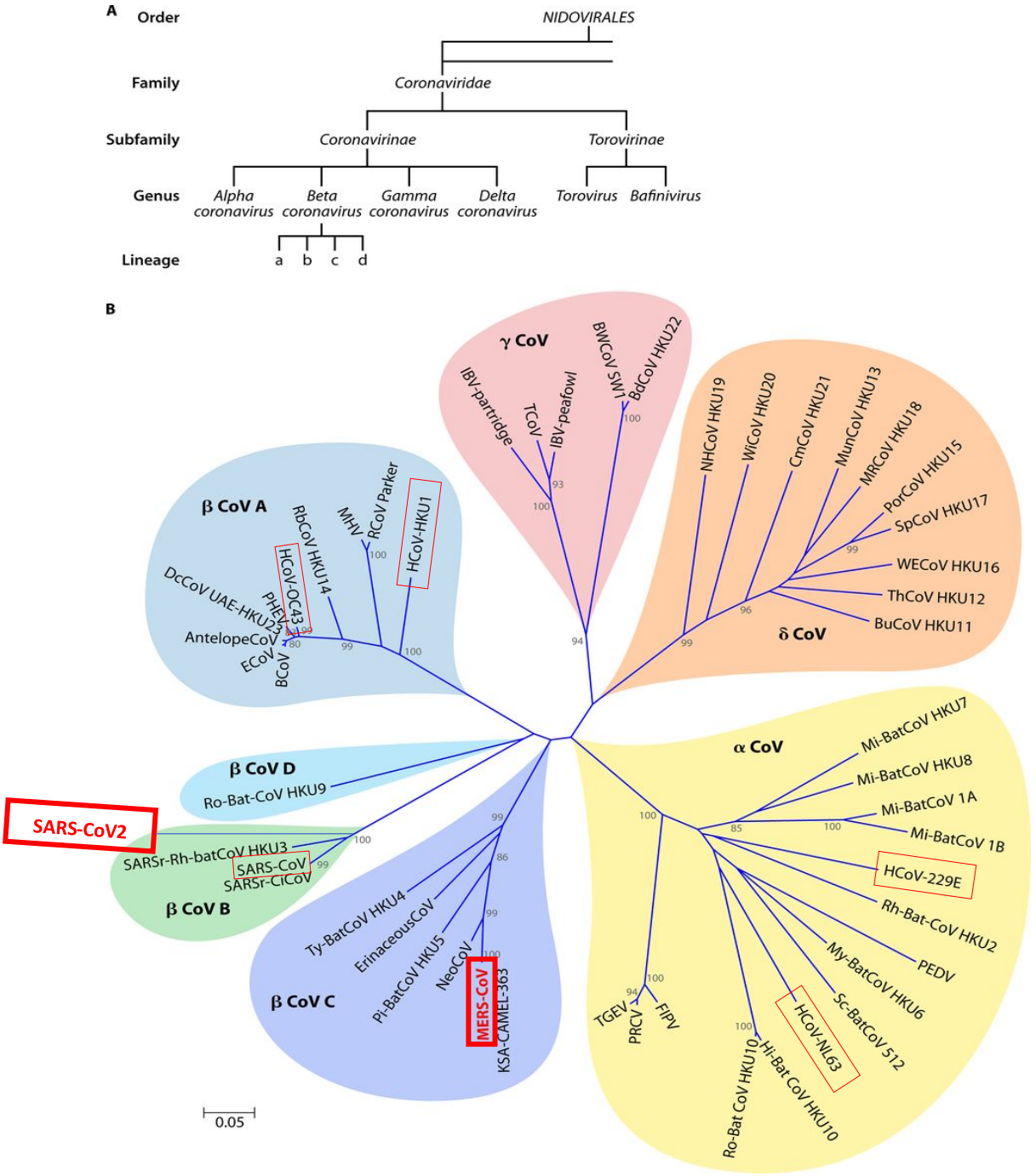


Figure 1-4 Coronavirus classification and phylogenetic relationships according to the International Committee on the Taxonomy of Viruses.

(A) The *Betacoronavirus* genus comprises four lineages (a, b, c, and d). (B) The B lineage of *Betacoronavirus* includes SARS-CoV and SARS-CoV-2, while MERS-CoV comes under *Betacoronavirus* lineage C. The figure was modified from (30, 31).

1.2 Coronaviruses diseases

1.2.1 Viral transmission

1.2.1.1 Animal to human

Coronaviruses are known to be zoonotic (1). Evidence shows that MERS-CoV and other coronaviruses, such as SARS-CoV and SARS-CoV-2, initially emerged in bats and then infected animals and humans (32). *Tylonycteris* bat coronavirus HKU4 and *Pipistrellus* bat coronavirus HKU5 are the two phylogenetically closest bat coronaviruses to MERS-CoV (17). In contrast, SARS-CoV-like coronaviruses have been isolated from bats (33). In addition, SARS-CoV-2 closely resembles the bat SL-CoVZC45 and SL-CoVZXC21 coronaviruses (34). MERS-CoV RNA and anti-MERS-CoV antibodies have been isolated from the serum and milk of infected dromedary camels (35, 36), a major animal reservoir for MERS-CoV and an intermediary host between bats and humans (17). Civet cats are an intermediate host for SARS-CoV (34).

While the collection, preparation, and ingestion of camel milk and meat are the major contacts between camels and humans (37), there is no evidence confirming that the ingestion of camel products leads to MERS-CoV infection (37). Some studies clearly state that the chance of acquiring MERS-CoV infection is 15% and 25% for camel shepherds and slaughterhouse workers, respectively (6). In addition, other studies have isolated identical MERS-CoVs from patients and their camels (38), indicating that direct contact with camels is likely responsible for most primary MERS-CoV infections in Saudi Arabia (11). Moreover, it supports the hypothesis that the primary virus transmission route is inhaling infectious droplets. For example, milking camels or dealing with camel meat can generate aerosols that carry viral particles (39). Moreover, eye contact can also be a transmission route (37).

Instances of animals being infected with SARS-CoV-2 have been observed globally (40). Bats appear to be the SARS-CoV-2 natural reservoir (41). The precise intermediate host responsible for the transmission and evolution of SARS-CoV-2 remains undetermined (42) . Apart from wildlife that may have direct contact with bats in their natural ecological setting and subsequently transmit the virus to humans in wildlife markets, domestic animals have also been identified as potential intermediate hosts subsequent to infection by SARS-CoV-2 (42).

The risk of animals transmitting the SARS-CoV-2 to humans is generally low; however, transmission from humans to animals during close contact is more commonly documented (2). Notably, many of these animals have contracted the SARS-CoV-2 following direct contact infected individuals, such as owners, caretakers. A wide range of animal species have been reported to be infected worldwide, encompassing both domestic pets, including cats, dogs, hamsters, and ferrets, as well as animals housed in zoological facilities, such as lions, tigers, snow leopards, hyenas, hippopotamuses, coatimundis and manatees (40). In addition, minks raised in mink farms have also been susceptible to infection, as have certain wildlife species, including deers, anteaters, and wild minks found in close proximity to mink farms (43).

1.2.1.2 Human to human

MERS-CoV and SARS-CoV-2 are transmitted from human to human by direct contact between or indirectly through contact with contaminated surfaces (Figure 1-5). Direct contact with asymptomatic or symptomatic individuals, contaminated objects or surfaces, and respiratory droplets are the most common methods of human-to-human SARS-CoV-2 transmission (34). While human coronaviruses are typically transmitted through respiratory droplets, direct contact with infected surfaces and faecal-oral transmission were also recorded during the SARS-CoV pandemic (44).

Direct transmission by respiratory droplets is facilitated by high SARS-CoV-2 replication in the upper (URT) and lower (LRT) respiratory tracts. An increasing number of reports indicate human-to-human transmission among close contacts who are actively coughing (44). In contrast, evidence suggests limited direct human-to-human MERS-CoV transmission. In the Korean outbreak, 83% of the viral transmission events were due to only 5 of the 186 MERS-CoV infected patients (45). Healthcare workers with prolonged viral exposure represented most of those infected by human-to-human contact in the Korean outbreak (37).

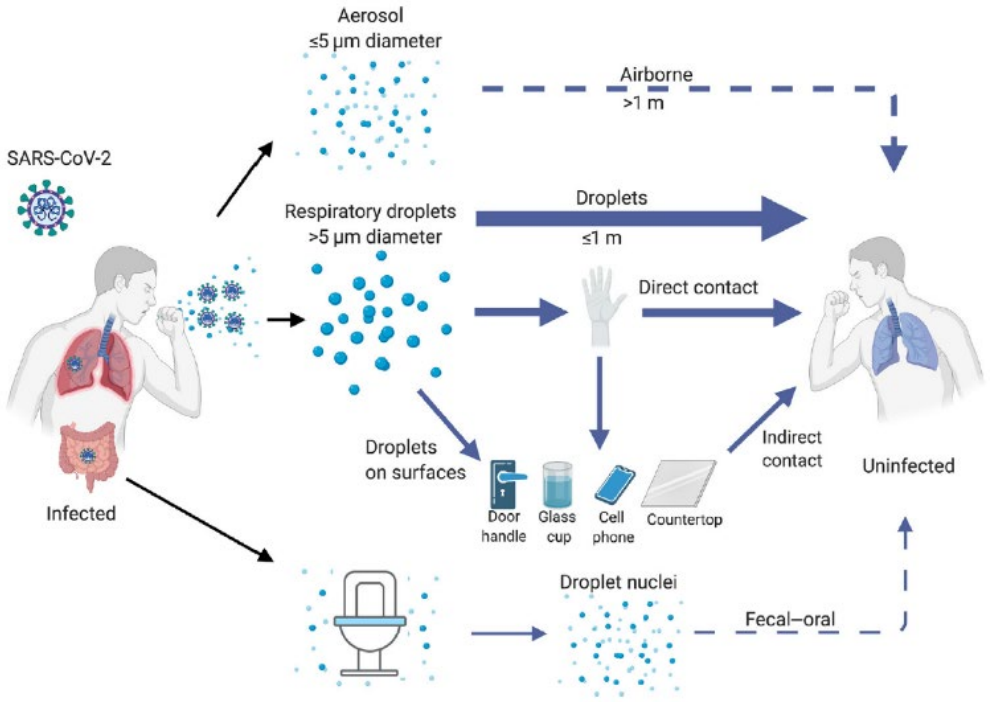


Figure 1-5 The SARS-CoV-2 transmission routes.

The virus is transmitted from human to human directly through airborne droplets (>5 µm), aerosol (≤5 µm), and direct human-to-human contact, especially among family members in households, and indirectly through contaminated surfaces. However, faecal-oral transmission has also been reported. Dark blue arrows denote confirmed transmission routes. The figure was taken from (46).

The possibility of SARS-CoV-2 transmission by aerosol, respiratory droplets, and personal contact has all been proposed since it can remain viable and infectious in aerosols for hours (h) (29, 47). In addition, viral transmission via the faecal-oral route, conjunctival secretions, flatulence, and sexual and vertical transmission from mother to foetus are all suggested viral transmission routes (29). Moreover, respiratory viruses can cause ocular (visual) difficulties in infected individuals, leading to respiratory infections and increasing SARS-CoV-2 transmission risk through unprotected mucous membranes of the eyes, mouth, and nose (48).

SARS-CoV-2 RNA levels in throat swabs and sputum samples were found to peak around 5–6 days after the onset of symptoms (49). During sneezing, coughing, or even chatting, large virus-containing droplets are secreted close to the sick individual, which can infect nearby uninfected individuals (29). The size of the virus-carrying droplets is critical (29). It was discovered that the infectious SARS-CoV-2 particles remained in aerosols for about 3 h, with a steady decline over time (47). The microscopic droplets ejected by an infected individual have been shown to travel 1–10 metres in the air (50).

The presence of viral RNA in faeces on toilet seats, washbasins, and sinks indicates a possible faecal-oral transmission route (51, 52). SARS-CoV-2 viral RNA was found in gastrointestinal tissue samples taken from a COVID-19 patient (52). This is expected since SARS-CoV-2's cell receptor, angiotensin-converting enzyme 2 (ACE2), is highly expressed in the gastrointestinal system (29). The virus can survive for up to four days in faecal samples, a long retention period (53, 54). While rectal swab testing is more useful and reliable in diagnosing and measuring the quantity of viral RNA in a patient's body (29), throat swabs have been used. Nevertheless, it was suggested to estimate the hospital stay length and the quarantine period's end (29). In addition, urine samples from infected individuals have also been found to contain the live SARS-CoV-2 virus (29).

Mathematical and animal models and intervention research have indicated that contact transmission is the primary transmission route for coronaviruses (55). While SARS-CoV-2 is associated with lower mortality than SARS-CoV or MERS-CoV, it is more transmissible (46). Its transmission routes include via droplets (>5 μl, travelling <1 m), airborne particles (≤5 μl, travelling >1 m), direct contact with an infected individual and, less commonly, direct contact with contaminated surfaces (55). Coronaviruses can survive on porous and non-porous materials such as metal, plastic (e.g. telephones, light switches, latex, rubber, and polystyrene), woven and non-woven fabrics (e.g. cotton, polyester, handkerchiefs, and disposable tissues), paper, wood, glass, and Formica, and objects such as stethoscopes, tissues, bank notes, tiles, eggs, feathers, and soft toys (55).

Coronaviruses can remain infectious for long periods on dry surfaces, especially when deposited in human secretions (55). Existing research suggests that SARS-CoV-2 can survive on plastic and stainless steel for up to 72 h before its stability and titer deteriorate (29). In addition, MERS-CoV and SARS-CoV can persist in an aerosol for more than 24 hours, which has serious implications for infection control (55). One study examined the survival of MERS-CoV in aerosols and found a 7% reduction in survival after 10 minutes (56). Table 1-1 shows virus viability on different materials. Despite the similarities between SARA-CoV-2 and SARS-CoV in genome structure, tissue tropism and viral pathogenesis, SARS-CoV-2 appears more transmissible (46). While studies have detected SARS-CoV-2 viral RNA in air samples for up to 3 h (57) and 16 h (58), various agencies did not previously consider airborne transmission except in restricted indoor environments.

Table 1-1 Viral viability time on different materials

	Virus viability (hours) on different materials					
Virus	Aerosol	Plastic	Stainless steel	Copper	Cardboard	Faeces
SARS-CoV-2	3	72	72	4	24	96
SARS-CoV	3	72	72	8	8–24	-
MERS-CoV	26	48	48	-	-	-
Data were taken from published studies (29, 47, 56)						

1.2.2 Symptoms and disease

Both circulating pathogenic human coronaviruses SARS-CoV-2 and MERS-CoV cause severe respiratory illness with cough, fever, shortness of breath, and gastrointestinal abnormalities and symptoms, including diarrhoea, vomiting, and nausea (17, 59). COVID-19 symptoms are similar to other specific and non-specific URT symptoms, making it difficult to distinguish from other disorders (60). A case report from Eiju General Hospital in Tokyo, Japan, showed the presence of the influenza virus alongside SARS-CoV-2, with similar symptoms making the two infections challenging to distinguish (61). SARS-CoV-2 targets tissues expressing the ACE2 receptor, including the heart, lungs, gastrointestinal tract, and urinary tract (62). In severe infections, organ malfunction, including shock, acute respiratory distress syndrome (ARDS), acute heart injury, acute renal injury, and death, might occur (59). MERS-CoV has a case fatality rate (CFR) of 35%, higher than SARS-CoV (9.6% (17)) and SARS-CoV-2 (0.47%–1.4(25)). A retrospective, observational study of 52 critically ill adult COVID-19 patients admitted to the intensive care unit (ICU) found that 15 (29%) had an acute renal injury, 12 (23%) had a cardiac injury, 15 (29%) had liver dysfunction, and 35 (67%) had ARDS (63).

Immunological differences can be used to distinguish between severe and non-severe COVID-19 patients, including leukocytosis, neutrophilia, lymphopenia, an increased neutrophil to lymphocyte ratio, and increased proinflammatory cytokine levels (e.g. interleukin 6 [IL]-6 and tumour necrosis factor-alpha [TNF-α]) (64). The incubation periods of MERS-CoV and SARS-CoV-2 are 5–6 days and 5–13 days, respectively (6, 21, 37), after which the symptoms described above start to occur (21). In a limited number of cases, infections can be asymptomatic (32, 46).

Moreover, viral transmission is divided into two periods. Transmission that occurs before the onset of symptoms is called presymptomatic. Transmission that occurs after the onset of symptoms is called symptomatic (34). Evidence of nonsymptomatic/presymptomatic SARS-CoV-2 transmission has been reported (65). Asymptomatic MERS-CoV and SARS-CoV-2 transmission occurs throughout the incubation period, which lasts on average 5–6 days but can last up to 14 days (34). Within the first three days after the onset of symptoms, virus shedding is highest in the URT (66). A few cases have been documented in which individuals who tested positive for COVID-19 spread it 1–3 days before they became symptomatic (67). These cases suggest that the viral load is high enough to allow transmission before symptoms occur (34).

Short-term immunity to SARS-CoV-2 has been suggested based on temporary seasonal coronaviruses (68). Antibody expression in SARS-CoV, MERS-CoV, and SARS-CoV-2 was recorded (69). MERS-CoV immunoglobulins M (IgM) and G (IgG) showed similar steady increases for 2–3 weeks after infection before declining. SARS-CoV-2 IgM expression begins 3–6 days after infection and peaks after 9–15 days. In contrast, IgG expression begins ten days after symptoms and peaks after 20–30 days (69).

1.2.3 Diagnosis and treatment

1.2.3.1 Diagnosis

Most human coronaviruses cause a respiratory illness, making differentiation difficult. After the onset of initial symptoms typical of coronavirus disease, the first and most critical step in SARS-CoV-2 and MERS-CoV diagnosis is proper specimen collection and processing. Then, travel history, contact tracing, and laboratory tests are required for diagnosis (37). The viral RNA can be isolated from URT samples, LRT samples, serum, saliva, stool, and urine (17, 52, 70, 71). However, the highest viral load is usually in lower respiratory specimens such as sputum samples (17, 71, 72). Both nucleic acid and serological testing can be performed on blood samples. There are three primary tests. The first is the nucleic acid amplification test

(NAAT), which detects the presence of actual virus genetic material. The reverse transcription polymerase chain reaction (RT-PCR) is the most commonly used NAAT method (73). This test needs a nasal pharyngeal swab, pharyngeal swab, or blood sample (74). The second is the antigen test, which detects one of the outer proteins of the viral envelope (75). The third test searches for antibodies specific to the virus's outer surface. Consequently, it reflects whether the individual has mounted an immune response or developed immunity against a particular virus. This test needs a blood sample.

Early in the outbreak, the SARS-CoV-2 genome was sequenced, and point-of-care RT-PCR assays were rapidly developed and used as diagnostic tests (76, 77). SARS-CoV-2 RNA can be detected using RT-PCR assays targeting its open reading frame 1b (*ORF1b*) and nucleocapsid (*N*) genes. These tests could detect up to ten SARS-CoV-2 copies per reaction, with the *N* gene assay having approximately ten-fold greater sensitivity than the *ORF1b* gene assay (78). Moreover, envelope (*E*), RNA-dependent RNA polymerase (*RdRp*), and spike (*S*) gene assays have been developed (74). Therefore, the two or three regions tested must be positive to identify a positive case (70). Moreover, the WHO recommends real-time quantitative RT-PCR-based assays for MERS-CoV detection (79), including a repeat test (37) and assaying a minimum of two different genomic regions (71). Three real-time RT-PCR (rRT-PCR) assays for the routine detection of MERS-CoV have been developed and published, which target the upstream region of the *E* gene (upE), the *ORF1b* gene, and the open reading frame 1a (*ORF1a*) gene (72, 80, 81). The upE assay is extremely sensitive and recommended for screening (81). The *ORF1a* assay is considered equally sensitive to the upE assay. However, the *ORF1b* assay was less sensitive than the *ORF1a* assay (81). Additionally, an alternative approach involving two rRT-PCR assays targeting the MERS-CoV *N* gene has been published, which can be used in conjunction with the upE and *ORF1a* assays for screening and confirmation (82). To date, these rRT-PCR assays have shown no cross-reactivity with other respiratory viruses, including other human coronaviruses. They can also detect all known MERS-CoV strains in humans and dromedary camels (81).

Antigen tests are immunoassays that detect the presence of a particular viral antigen or protein, which is indicative of an active viral infection. There are currently SARS-CoV-2 authorised point-of-care, laboratory-based, and self-testing antigen tests. However, there are no antigen tests for MERS-CoV (66). While antigen tests provide rapid results (within 30 minutes), they are typically less sensitive than antibody tests (40) . Antibody testing is not used for early diagnosis but is useful for epidemiological inquiry and predicting disease fate (74). The enzyme-linked immunosorbent assay is one of the most commonly used serological tests. Antibody responses develop in most MERS-CoV patients by the third week of illness but can

be delayed further in severely ill patients requiring mechanical ventilation (83). The recommended antibody screening tests included those against N and S proteins for SARS-CoV-2 and MERS-CoV (84).

1.2.3.2 Treatment

MERS-CoV does not have a specific treatment, only general treatments that aim to alleviate symptoms (78). Most individuals who contract COVID-19 will be able to recover at home (85). However, the NHS treats COVID-19 patients at high risk of becoming seriously ill with antivirals, including nirmatrelvir and ritonavir (Paxlovid), remdesivir (Veklury), molnupiravir (Lagevrio), and sotrovimab (Xevudy). Sotrovimab is a neutralising monoclonal antibody that resembles human antibodies in their function within the immune system. They are produced by cloning an antibody that can bind to and neutralise the virus S protein. They adhere to the virus, preventing it from entering the lungs and causing an infection.

1.3 Host immune response and pathogenesis

Like many other viruses, coronaviruses have evolved strategies to antagonise the host's innate immune response on various levels (86). The generation of type I interferons (IFNs; α and β) is the primary response of mammalian cells to viral infection, activating the type I IFN-mediated innate immune response (87). MERS-CoV, SARS-CoV, and SARS-CoV-2 show several similarities in their immune responses and clinical presentation, from asymptomatic infection to severe disease with an immune storm. Figure 1-6 shows the immunological map for SARS-CoV, MERS-CoV, and SARS-CoV-2 infection. SARS-CoV-2 and SARS-CoV use ACE2 as their cell receptor with S processing by transmembrane protease serine 2 (TMPRSS2) for optimal infection efficiency. In contrast, MERS-CoV uses dipeptidyl peptidase 4 (DPP4) to enter cells (64). The ACE2 and DPP4 cell receptors are widely expressed on respiratory epithelial, kidney, alveoli, small intestine, liver, and prostate cells (88). All three coronaviruses affect the human immune cell response and cytokine expression. The response includes an increase in proinflammatory cytokines (Interleukin (IL) -1, IL-6, IL-8, and TNF-α), type I T helper cell cytokines (IL-2, IL-12, and IFN-γ), type II T helper cytokines (IL-4, and IL-5), T helper 17 cell cytokines (IL-17), and regulatory T cell cytokines (IL-10) (89, 90). Moreover, neutrophil and cluster of differentiation 4 $(CD4)^+$ and 8 $(CD8)^+$ T cell numbers were decreased in SARS-CoV-2, SARS-CoV, and MERS-CoV patients (64). When COVID-19 progression reaches critical stages, laboratory tests show increased neutrophil counts, up to 90% of the white blood cell count (18). Moreover, MERS-CoV can infect immune macrophages and dendritic and T cells (90). When MERS-CoV infects dendritic cells and macrophages, they produce high levels of proinflammatory cytokines and chemokines, such as TNF-α, IL-6, IL-8,

C-X-C motif chemokine ligand 10 (CXCL10), and C-C motif chemokine ligands 2 (CCL2), 3 (CCL3), and 5 (CCL5) (91).

SARS-CoV-2 infection may result in an imbalanced host immune response, including decreased type I and type III IFN signalling and strong chemotactic and inflammatory responses in infected cells, animal models, and COVID-19 patients (92). The MERS-CoV membrane (M) and open reading frame 4a (ORF4a), 4b (ORF4b), and 5 (ORF5) proteins were experimentally shown to be potent SARS-CoV-2 interferon antagonists (87). Eight SARS-CoV proteins are interferon antagonists (93, 94): N; M; papain-like protease; non-structural proteins 1 (nsp1), 7 (nsp7), and 15 (nsp15); open reading frame 3b (ORF3b) and 6 (ORF6). It has been hypothesised that the immune proinflammatory and attraction properties of the IL-6, IL-12, and TNF-α cytokines could underlie immune cell infiltration into infected patients' LRTs, resulting in severe inflammation and tissue destruction (91).

T cells infected with MERS-CoV undergo apoptosis mediated by extrinsic and intrinsic apoptotic pathways (95). MERS-CoV evades the T cell response in the peripheral blood and lymphoid organs through this pathway, leading to virus dissemination and severe immunopathology (88). Furthermore, the elevated expression of mothers against decapentaplegic homolog 7 (SMAD7) and fibroblast growth factor (FGF) 2 due to MERS-CoV causes apoptosis in kidney and lung cells (96). Type I IFN signalling increases pathogenic inflammatory monocytes and macrophages, accompanied by strong virus replication, resulting in higher lung cytokine and chemokine levels, virus vascular leakage, and reduced-specific T cell responses (97).

Comparisons between COVID-19 patients and healthy controls showed that some parameters increased, including IL-1ß, IL-1Ra, IL-3, IL-8, IL-9, IL-10, basic FGF, granulocyte colony-stimulating factor (GCSF), granulocyte-macrophage colony-stimulating factor, IFN-γ, CXCL10/IP10, CCL2/MCP1, CCL3/MIP1A, CCL4/MIP1B, platelet-derived growth factor, TNF-α, and vascular endothelial growth factor (18). Furthermore, levels of specific cytokines, including IL-2, IL-7, IL-10, GCSF, CXCL10/IP10, CCL2/MCP1, CCL3/MIP1A, and TNF-α, are higher in ICU than in non-ICU patients (18). These findings indicate that immunopathology may play a role in disease severity progression (90). Therefore, it was proposed that COVID-19 treatment should include restraining the host inflammatory responses using an antibody or cytokine neutraliser to reduce viral antagonism of the immune response (90).

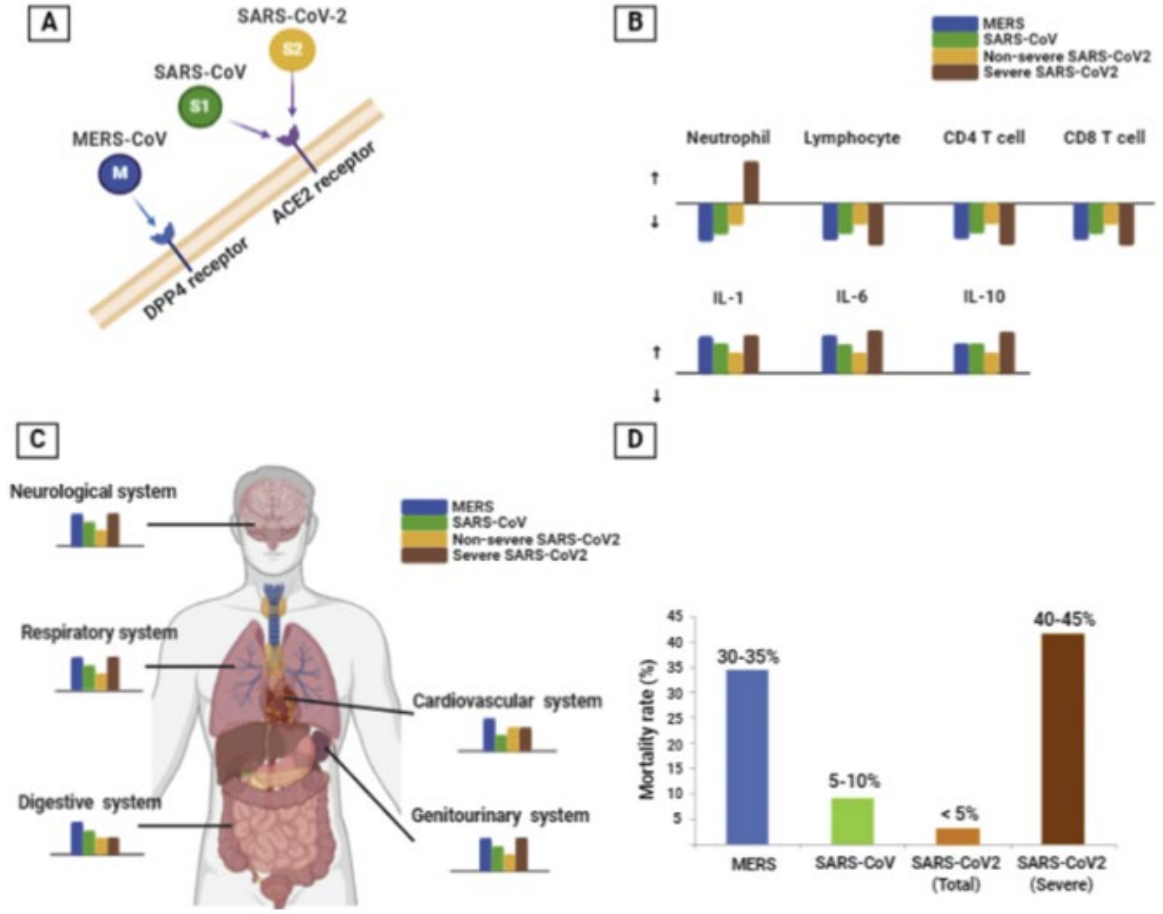


Figure 1-6 The immunological map for MERS-CoV, SARS-CoV, and SARS-CoV-2 infection.

(A) Different coronaviruses use different cell receptors to enter living cells. SARS-CoV and SARS-CoV-2 use the ACE2 receptor, while MERS-CoV uses DPP4. (B) Immune cell and cytokine levels change in MERS-CoV, SARS-CoV, and SARS-CoV-2 (severe and non-severe) infection. The up and arrows denote increases and decreases, respectively. The coloured bar's length corresponds to the magnitude of the infection's effect. (C) Viral pathogenesis affects different human body systems. (D) A diagram of MERS-CoV, SARS-CoV, and SARS-CoV-2 (severe and non-severe) infection fatality rates. The figure was taken from (64).

1.4 Coronavirus virion and genome organisation

The coronaviruses virion is spherical and 100–160 nm in diameter (32, 98, 99) with spike-shaped glycoproteins embedded in the lipid envelope (6). Coronaviruses have the largest single-stranded RNA (27–32 kb) genome among RNA viruses (100) that encodes ≈25 proteins. The SARS-CoV-2 genomic sequence has 80% sequence identity with the SARS-CoV genome and 50% with the MERS-CoV genome (101). A typical coronavirus genome has a 5' untranslated region (UTR), a conserved replicase domain (ORF1ab), four genes (S, E, M, and N) encoding structural proteins, unique genes encoding coronavirus-type-specific accessory proteins, and a 3' UTR (31). The virus-specific accessory genes include six ORFs (3, 4a, 4b, 5, 8b, and 8c) in the MERS-CoV genome but 11 (3a, 3b, 3c, 3d, 6, 7a, 7b, 8, 9b, 9c, and 10) in the SARS-CoV-2 genome (74, 99). Figure 1-7 shows a coronavirus particle and SARS-CoV-2 and MERS-CoV genome organisation.

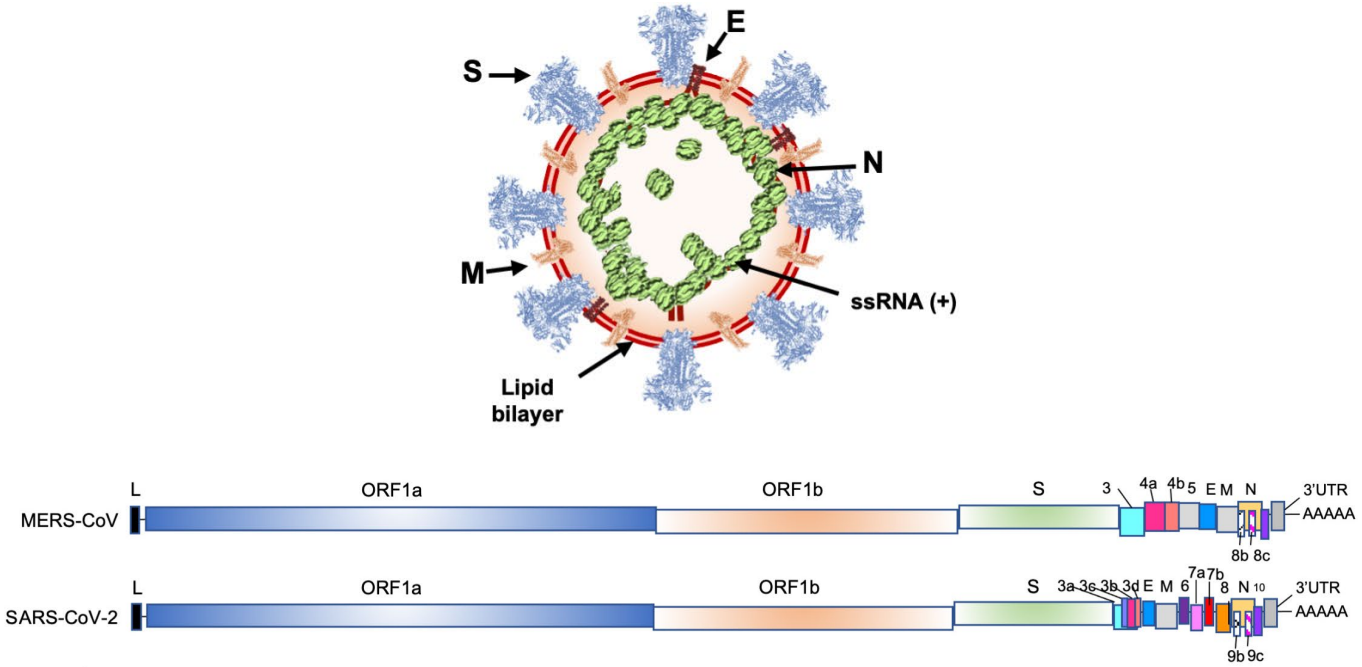


Figure 1-7 A coronavirus particle and SARS-CoV-2 and MERS-CoV genome organisation.

Coronaviruses encode two large overlapping ORFs (1a [blue box] and 1b [orange box]) that are translated to produce polyproteins 1a and 1ab by ribosomal frameshifting. The two polyproteins are proteolytically processed into the non-structural proteins (nsps) used for replication. The spike (S; green box), envelope (E; blue box), membrane (M; grey box), and nucleocapsid (N; yellow box) structural proteins are shown. Accessory proteins (different colour boxes) are ORFs 3a, 3b, 3c, 3d, 6, 7a, 7b, 8, 9b, 9c, and 10 in the SARS-CoV-2 genome (bottom) and ORFs 3, 4a, 4b, 5, 8b, and 8c in the MERS-CoV genome (top). The figure was taken from (100).

1.4.1 Coronavirus non-structural and accessory proteins

The coronavirus' *ORF1a* and *ORF1b* genes span two-thirds of the viral genome and are translated into two polyproteins, pp1a and pp1ab, cleaved by viral proteases into 16 highly conserved nsps (nsp1–16) (102). These proteins are involved in genome replication, transcription, and subgenomic messenger RNA (mRNA) synthesis (100). Nsp5 is one of two critical proteases encoded by SARS-CoV-2. It is referred to as SARS-CoV-2 $3CL^{pro}$ due to its homology to the 3C proteases of picornavirus and some other single positive-stranded RNA (+ssRNA) viruses (103). $3CL^{pro}$-like proteases are considered therapeutically crucial because they are required for the cleavage of large polyprotein products produced by +ssRNA viruses. Importantly, chemical protease inhibitors have a broad spectrum of activity against members of a given virus' family (104). Table 1-2 presents the coronavirus's nsps and their function.

The coronavirus accessory proteins were produced by discontinuous ORF transcription in single negative-strand RNA to produce mRNAs translated as accessory proteins (100). While the accessory proteins do not play a role in viral genome replication, they are thought to play a role in viral pathogenesis (71), viral replication, and subversion of the host cell's innate immune response (17, 105). It is challenging to study them because of their lower molecular weight and low expression levels (17, 105). Moreover, they are not well conserved between coronaviruses (Figure 1-7). The MERS-CoV genome includes six ORFs encoding the accessory proteins 3, 4a, 4b, 5, 8b, and 8c, respectively, while the SARS-CoV-2 genome includes 11 ORFs encoding the accessory proteins 3a, 3b, 3c, 3d, 6, 7a, 7b, 8, 9b, 9c, and 10, respectively (74, 99). Only some of these ORFs are translated. SARS-CoV-2 ORFs 3b and 10 might not be translated (100).

Despite being called accessory proteins, MERS-CoV replication was affected by the deletion of ORF3, ORF4, and ORF5 in human airway-derived Calu-3 cells (71). In addition, deletion of ORF4a and ORF4b also affected viral replication in hepatic carcinoma-derived Huh-7 cells (71). These examples suggest a role for the accessory proteins in MERS-CoV replication (71), making them potential antiviral drug targets, even though their biological function is not fully understood (105).

Numerous SARS-CoV-2 accessory proteins are associated with immune evasion mechanisms, such as cytokine secretion inhibition by ORF9c and the counteraction of type I IFN action by ORF3b, ORF6, ORF7a, ORF8 and ORF9c (106). In addition, these

accessory proteins alter other significant cellular mechanisms, including autophagy and apoptosis by ORF3a, mitochondrial function by ORF3d, and inflammasome activation by ORF9b (103, 106).

Table 1-2 Coronavirus non-structural proteins and their functions.

Protein	Function in viral replication cycle	Effect on host cells
Nsp1	A virulence factor. A genus-specific marker. Highly divergent among coronaviruses.	Prevents host mRNA translation.
Nsp2	Not reported.	Disrupts the intracellular host cell survival signalling pathway by interacting with prohibitin 1 and 2 host proteins.
Nsp3	Has cysteine-type endopeptidase activity at the N-terminus of the replicase polyprotein.	Host membrane rearrangement. Downregulates proinflammatory cytokine expression. An IFN antagonist.
Nsp4	Formation of viral replication complexes through interactions with nsp3 and nsp6.	Forms double-membrane vesicles (DMV)
Nsp5	Cleaves the C-terminus of the replicase polyprotein at 11 sites.	Plays a role in modifying host vacuole intracellular pH.
Nsp6	Plays a role in viral replication.	DMV formation.
Nsp7	A cofactor for RNA-dependent RNA polymerase (RdRP).	Not reported.
Nsp8	Cofactor for RdRP.	Not reported.
Nsp9	It suggested acting as RNA binding protein	Not reported.
Nsp10	Interact with nsp1, nsp7, nsp14, and nsp16. Plays a role in viral transcription.	Not reported.
Nsp11	Unknown	Not reported.
Nsp12	Plays a role in viral RNA genome replication and transcription.	Not reported.
Nsp13	Has magnesium-dependent helicase activity.	Not reported.

Nsp14	Acts as an exoribonuclease on RNA in 3' to 5'direction. Functions as a proofreader. RNA cap formation.	Interacts with DEAD-box helicase 1 via its N-terminus. Modulates the innate immune response.
Nsp15	Manganese cation-dependent, uridylate-specific enzyme.	Evasion of an immune response.
Nsp16	RNA cap formation.	Functions as an IFN antagonist.
Data were taken from published studies (100, 107).		

1.4.2 Coronavirus structural proteins

There are four coronavirus structural proteins (Figure 1-7). The glycosylated trimeric S protein (~600 kDa) has S1 and S2 domains that are necessary for the attachment and fusion of the virus to the host cell (108). The second small protein, E, comprises 76–109 amino acids (109) with a hydrophobic N-terminus, and >60% acts as a transmembrane domain (TMD) (110). The E protein's main functions are viral assembly, viral envelope formation, and viral apoptosis activation (109, 110). In contrast, the third protein, M, comprises three TMDs and is located in the viral envelope. It is present in high quantities in the viral particle and is responsible for virion shape. In addition, it plays a role in the viral assembly process (109). The fourth soluble protein, N, is involved in viral RNA packaging and nucleocapsid formation (109).

1.4.2.1 Coronavirus E protein

The E protein is a minor viral protein with the lowest copy number of the structural proteins found in the lipid envelope of mature virus particles (111). It is a small protein comprising 76–109 amino acids (109) (Figure 1-7) with a molecular weight of 8.4–12 kDa. It has an N-terminus comprising a short run of 7–12 hydrophobic amino acids followed by a large hydrophobic TMD of 25 amino acids and a long hydrophilic C terminal region (112) (113) (Figure 1-8). It is an integral membrane protein embedded in the viral envelope's lipid bilayer (114). Under certain experimental conditions, each protein exists as a homo-oligomer (114). The SARS-CoV-2 E protein is a 75-residue hydrophobic protein with 94.74% amino acid identity to the SARS-CoV E protein. In contrast, it only has 36.00% amino acid similarity with the MERS-CoV E protein (111).

Table 1-3 and Figure 1-9 present the E protein's homology and sequence between SARS-CoV-2 and two other coronaviruses SARS-CoV, and MERS-CoV. The E protein's main functions are viral assembly, using its hydrophobic domain (HD), viral envelope formation, and viroporin production, which is described in more detail below (115). The E protein contributes to disease severity through hyperactive cytokine release (116) . However, it is not involved in the coronavirus RNA packaging process (109, 113).

1.4.2.1.1 Viral assembly and budding

The E protein participates in assembly, budding, and intracellular trafficking and is essential for robust viral production (117). It is localised on cellular membranes required for macromolecule secretion and trafficking from the cell, including the endoplasmic reticulum (ER), Golgi, and the ER-Golgi intermediate compartment (ERGIC), where it participates in coronaviruses assembly and budding (87, 115).

The viral envelope is assembled in the ERGIC. Virions then bud into the lumen, reach the cell surface via the host secretory pathway, and eventually exit the cell (115, 117). It was found that the E protein plays a role in changing the cellular secretory pathway to promote the viral assembly by preserving S protein activity (118).

The viral assembly process is followed by membrane scission to complete the budding process. It was found that enveloped viruses can use the scission machinery of host cells or express their own scission proteins (119). The influenza virus failed to undergo scission when its matrix-2 (M2) protein was mutated (120), which is the viral protein responsible for influenza virus budding and scission (120). There appear to have been no attempts to ascertain whether the coronavirus E protein is responsible for coronavirus virion scission during budding (115). However, limited data is available on E-host protein-protein interactions (PPIs), and evidence suggests that the E protein is involved in the assembly and release of coronaviruses. However, its exact involvement remains unknown (115).

1.4.2.1.2 Envelope formation

While the coronavirus M protein coordinates the building of viral particles, both M and E proteins are required for the production and release of virus-like particles (VLPs) (115) (121). The coronavirus particle has only a low E protein level compared to the M protein, which is much more abundant in the viral envelope (115, 122). The importance of the E protein in virus production and maturation was shown by the abnormal morphology, much lower viral titres, impaired viral maturation, and insufficient progeny for propagation with recombinant coronaviruses missing the E protein (115, 123).

The C-termini of the M and E proteins interact on the ERGIC's cytoplasmic side to form the viral envelope (124). The release of infectious virus particles requires the E protein's HD or TMD (117, 125). In a mouse-adapted infection model, residues in the HD of the SARS-CoV E protein promoted viral fitness and pathogenesis (117, 126).

Moreover, when the mouse hepatitis virus (MHV) E protein's C-terminal residues were changed to alanine, the virions were temperature sensitive and formed pinched, elongated shapes instead of the typical spherical shapes seen with wildtype virions (115, 127). The E protein induces membrane curvature in the viral envelope, allowing coronavirus particles to acquire their typical spherical shape and morphology, even though viral assembly and production are not entirely prevented in its absence (115).

A recombinant SARS-CoV (rSARS-CoV-ΔE) lacking the *E* gene showed decreased viral replication in hamster URTs and LRTs with titres 100- to 1,000-fold lower than the wildtype virus. The lower viral load was accompanied by less lung inflammation in rSARS-CoV-ΔE virus-infected hamsters than in wildtype virus-infected hamsters (115, 128). Because recombinant coronaviruses lacking the *E* gene grow to considerably lower titers or are propagation inept, the E protein must be required for efficient virus production (117).

1.4.2.1.3 Viroporin production

The E protein is a viroporin with ion channel activity (117). Viroporins are virally encoded membrane pore-forming proteins that can influence cellular ion channels and, in the case of the E protein, are formed by the N-terminal domain's transmembrane helix (114). These ion channels have been associated with numerous viral life cycle stages, including viral entry, assembly, and release, and viral pathogenesis (129). Viroporins are mostly small hydrophobic proteins (60–120 amino acids) that oligomerise in the membranes of infected cells, creating hydrophilic holes (115). Hydrophobic residues line the structure's perimeter, while hydrophilic residues line the inside of the pore to form cation-selective ion channels (130). Experiments showing that the known ion channel inhibitor hexamethylene amiloride reduced the viral titer of cultured cells infected with wildtype MHV but not mutant MHV missing the entire *E* gene provide the greatest evidence for the E protein's ion channel activity during infection (117, 131).

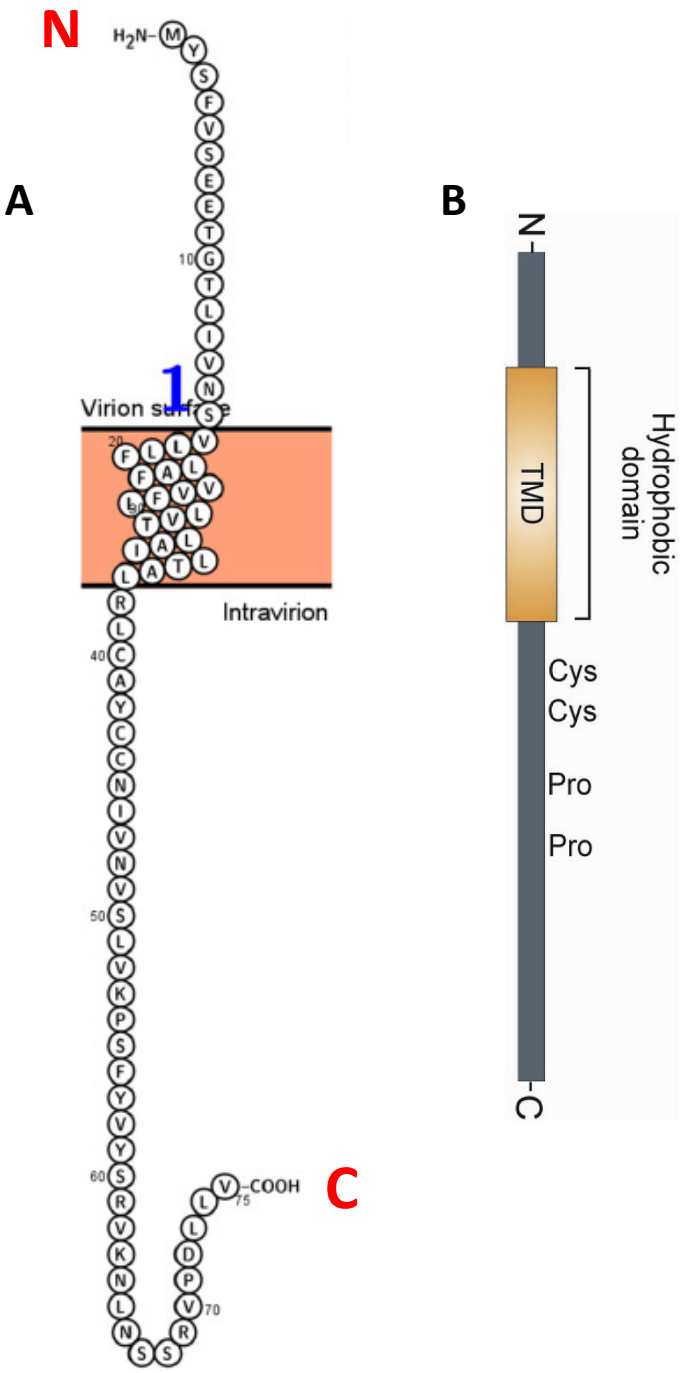


Figure 1-8 Coronavirus E protein.

(A) Modified E protein membrane topology with one TMD, the N terminus on the virion's exterior, and the C terminus on the virion's interior. The snake diagram was designed using Protter. The figure was modified from (26). (B) An E protein schematic domain showing cysteine and proline residue locations. The figure was taken from (109).

Table 1-3 Coronavirus E protein homology between SARS-CoV-2, SARS-CoV, and MERS-CoV.

Coronavirus	Uniprot ID	Length (aa)	Mass (kDa)	% similarity to SARS-CoV-2
SARS-CoV-2	P0DTC4	75	8.365	100%
SARS-CoV	P59637	76	8.361	94.74%
MERS-CoV	K9N5R3	82	9.354	36.00%
Data were taken from (111).				

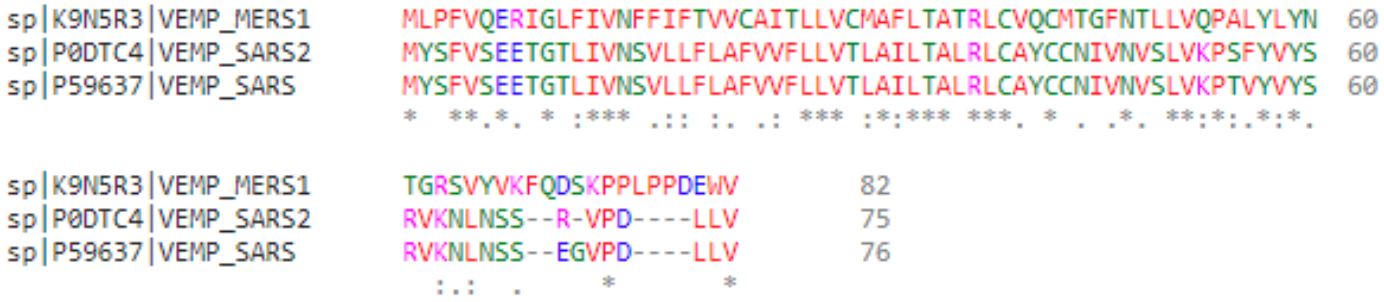


Figure 1-9 Comparison of the E proteins from SARS-COV-2, SARS-CoV, and MERS-CoV.

The protein sequence alignment was created with Uniprot (132)**.** Key: *, a conserved amino acid.

1.4.2.2 Coronavirus M proteins

The M protein is the most abundant in the viral envelope's lipid bilayer (109). It ranges from 219 to 222 amino acids and 24 to 25 kDa (133). It is present in high quantities in the viral envelope, is responsible for virion shape, and interacts with other proteins during the viral assembly process (36). The M protein's structure has three TMDs: a small N-terminal domain located outside the virion or inside the lumen of intracellular organelles, three-membrane spanning HDs, and a large C terminal domain spanning half the protein and located inside the virion or in the infected cells' cytoplasm (Figure 1-10) (134, 135).

A comparison of the M viral proteins for MERS-CoV, SARS-CoV, and SARS-CoV-2 is shown in Table 1-4 and Figure 1-11. The SARS-CoV-2 M protein is 89.14% identical to the SARS-CoV M protein and 38.36% similar to the MERS-CoV M protein (26).

Several M proteins accumulated in the Golgi complex of mammalian host cells (136). The MERS-CoV M protein co-localised with the Golgi apparatus and ERGIC in the perinuclear area, and there were additional discrete puncta within the cytoplasm (87). MHV assembly occurs at the intermediate compartment's smooth membranes between the ER and the Golgi complex (135, 137). SARS-CoV-2 M protein levels in the ER or ER-proximal membranes suggested that it is associated with endocytic pathway and lysosomal membranes (103, 135).The C terminal domains of MERS-CoV and infectious bronchitis virus M proteins were found to be in the trans-Golgi network and ERGIC/cis-Golgi, respectively (138).

The M protein determines the budding site in coordination with additional viral proteins. It migrates beyond the budding compartment and localises to the late-Golgi complex (135) (121). It is plausible to assume that the binding of the M protein to the N protein binding prevents it from migrating to the budding compartment (135). This M and N protein interaction aids nucleocapsid envelopment at the budding compartment (135).

The M protein can interact with other structural proteins (139). It is thought to be the driving force for viral particle assembly (139). The M protein's interaction with the nucleocapsid protein is essential for viral RNA packaging (113). While the E protein is not directly involved in packaging coronavirus RNA, it is involved in viral envelope formation with the M protein (113). Extensive electron microscopy experiments on M proteins from various

coronaviruses have found no evidence that it can induce membrane curvature on its own (115, 122).

In addition, the MERS-CoV M protein has been experimentally shown to be an interferon antagonist (87). A proximity proteomic study identified the tripartite motif containing 4 (TRIM4) interferon signalling regulator as a close interacting partner of SARS-CoV-2 M protein (103).
Additionally, four PPIs were identified between the SARS-CoV-2 M protein and accessory proteins (ORF6, ORF7a, ORF7b, and ORF10) and six PPIs with nsps (nsp2, nsp4, nsp5, nsp8, and nsp16) (92). These structural protein interactions may be critical for viral particle organisation during replication and assembly (92).

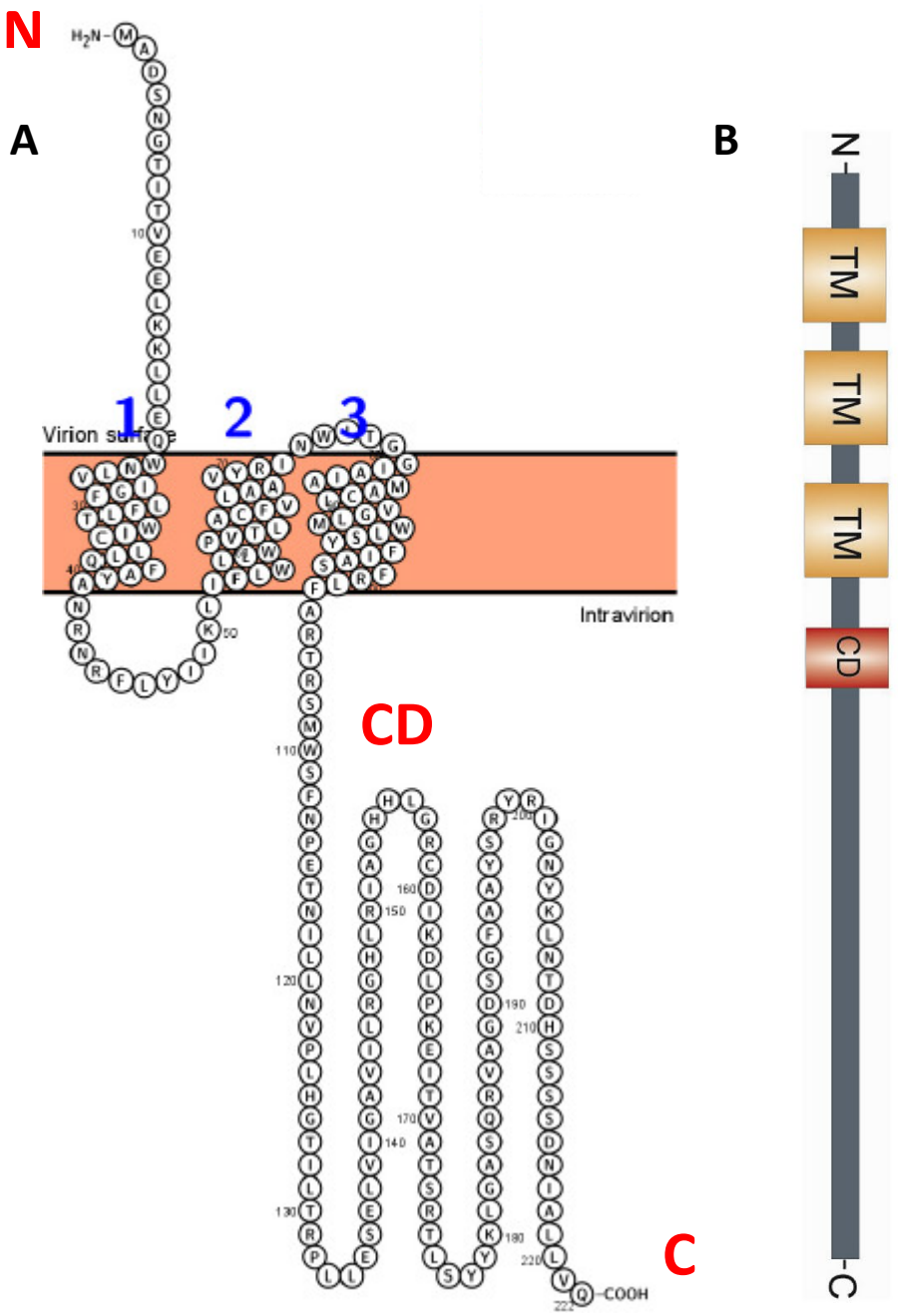


Figure 1-10 Coronavirus M protein.

(A) Modified M protein membrane topology with three TMDs, the N terminus on the virion's exterior, and the C terminus on the virion's interior. The snake diagram was created using Protter. The figure was modified from (26). (B) A schematic of the M protein's domains showing the TMDs and conserved domain (CD). The figure was taken from (139).

Table 1-4 Coronavirus M protein homology between SARS-CoV-2 and Two different coronaviruses.

Coronavirus	Uniprot ID	Length (aa)	Mass (kDa)	% similarity to SARS-CoV-2
SARS-CoV-2	A0A6C0NA72	222	25.147	100%
SARS-CoV	D3KDM9	221	25.089	89.14%
MERS-CoV	A0A455MNB0	219	24.532	38.36%
Data were taken from (26).				

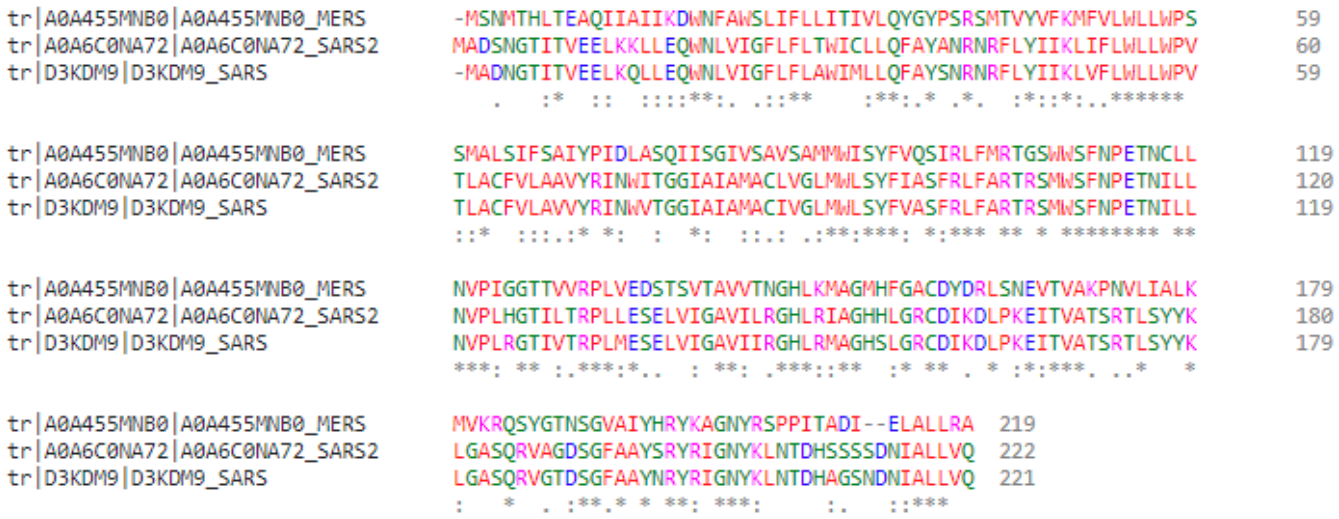


Figure 1-11 Comparison of the M proteins of SARS-COV-2, SARS-CoV, and MERS-CoV.

The protein sequence alignment was created using Uniprot (132). Key: *, a conserved amino acid.

1.5 Coronavirus life cycle and replication complex

1.5.1 Coronavirus attachment and entry

Coronaviruses bind to a specific cell receptor on the cell surface. MERS-CoV uses DPP4 as its multifunctional long transmembrane glycoprotein cell receptor (105) to enter bat and human cells (17). In contrast, SARS-CoV-2 use ACE2 as its cell receptor (140). These receptors appear on the surface of epithelial cells such as alveoli, kidney, small intestine, liver, prostate, type II pneumocyte, and LRT cells (32, 37, 141). Furthermore, MERS-CoV can infect human dendritic cells and macrophages *in vitro*, reducing the immune system's efficiency (141). The viral replication cycle is described clearly and simply in Figure 1-12. The coronavirus entry process begins by bringing the viral S protein into contact with the DPP4 cell receptor in the case of MERS-CoV and the ACE2 cell receptor in the case of SARS-CoV-2 (140). This interaction results in the cleavage of the S protein by host cell surface proteases, such as TMPRSS2, promoting viral uptake and fusion at the cellular membrane or endocytosis by endosomes, where the S protein is processed by cathepsin L (100). Then, the virus enters the cell and releases the viral RNA into the cytoplasm (140).

1.5.2 Replicase protein expression

After the virus enters the cell and the nucleocapsid-enclosed RNA genome is uncoated and released into the cytoplasm, the pp1a and pp1ab polyproteins are produced by translation of *ORF1a* and *ORF1ab* by ribosomal frameshifting (141). The viral encoded protease (M^{pro}, also called $3CL^{pro}$) is used to proteolytically cleave the pp1a and pp1ab polyproteins and produce 16 nsps (17, 141). These proteins play a significant role in building the RNA replicase transcriptase complex (17, 141). In addition, they are essential for both RNA replication and non-replicative functions, such as targeting the host cell translation and blocking the innate immune response, a function reported for nsp1 (142). This protein complex activates RNA replication to produce viral RNA genomes for new viral particles (100). In addition, it activates the discontinuous transcription process to create subgenomic mRNAs for structural and accessory proteins (100). This discontinuous transcription starts from the 3' end of the RNA genome until it reaches the body transcription regulatory sequences (BTRS) for a specific protein and then jumps to the leader transcription regulatory sequences (LTRS) at the 5' end of the RNA genome. The reason for calling this a discontinuous transcription process is that the sequence between the BTRS and LTRS is not transcribed (100).

The virion progeny are made via the secretory pathway, including the translation and accumulation of viral structural proteins S, E, M, and N in the ER before their transfer to the ERGIC (141, 142).

The localisation of E and M viral proteins in the ERGIC enables them to participate in coronavirus assembly, budding, and intracellular trafficking (115). The M protein acts as a central organiser of coronavirus assembly in the ERGIC, interacting with itself and all other major coronaviral structural proteins (143). In addition, the M protein helps condense the N protein that interacts with genomic RNA to form the nucleocapsid (100). Moreover, the M and E proteins are required for viral envelope formation (100). The E protein was found to affect the S protein's ability to promote the assembly of SARS-CoV-2 virus-like particles (144). Virion formation is followed by budding into the lumen and exocytosis (100).

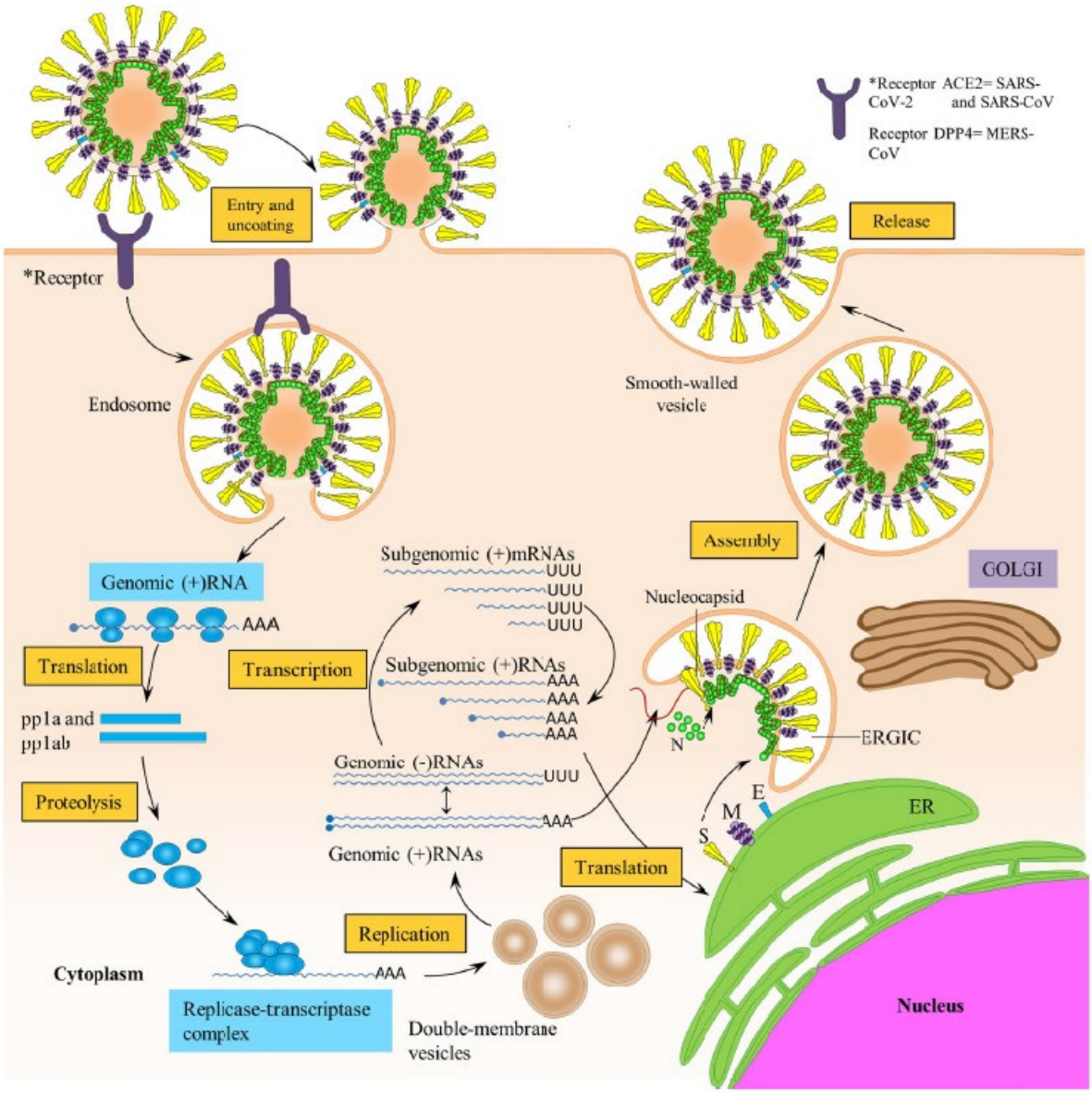


Figure 1-12 Simply schematic of coronavirus entry, replication, and exocytosis.

The virus binds to a cellular receptor, enters the cell, and is uncoated. Then, the viral genome is released into the cytoplasm, followed by RNA translation and polyprotein pp1a and pp1ab production. The two polyproteins are cleaved into functional proteins that form a replicase transcriptase complex. Finally, the virus is reproduced, assembled, and released see section 1.5. The figure was taken from (27).

1.6 Coronavirus prevention and control

Coronavirus human-to-human transmission can occur (6), especially in healthcare facilities and communities (17). As MERS-CoV and SARS-CoV-2 remain infectious for at least 48 and 72 h, respectively, on steel and plastic surfaces (145), a robust infection control system is critical in controlling their transmission, especially in healthcare facilities (37). This system should include using personal protective equipment, avoiding crowding, good ventilation, and cleaning and disinfecting surfaces and rooms (37). Moreover, because human coronaviruses spread through droplets and aerosols, the repercussions for containment will be severe, especially in crowded areas with poor ventilation (29). Therefore, recirculation of the potentially infected air in the same environment should be avoided (29). Since the WHO has recognised that coronavirus can be transmitted through the air, sick individuals must wear masks indoors and outside, even in distant locations, to prevent the virus from spreading further (29).

The final recommendations to control viral spread include: (1) adequate ventilation of high-risk areas such as toilets, which should be sterilised as a potential source of viral transmission; (2) masks should be worn, and social distancing of ≥2 metres should be used to reduce the risk of airborne viruses; (3) the quarantining and treatment of patients; (4) finding a suitable treatment and massively vaccinating individuals worldwide (146).

1.6.1 Vaccine development

While numerous vaccines against SARS-CoV and MERS-CoV have been developed and tested, only a few have progressed to clinical trials, and none have been approved by the US Food and Drug Administration (FDA) (147). Compared to SARS-CoV and MERS-CoV, which tend to resolve after strict quarantine and testing, the global scope of the COVID-19 pandemic has required vaccine development (147). This critical need has resulted in various approaches to vaccine development. Initially, novel vaccine platforms, such as nucleic acid and viral vector vaccines, were successfully and widely deployed. Their ability to be developed based only on sequence information has risen to prominence in the race to establish a COVID-19 vaccine (148). They are highly adaptable to emerging pathogens, and their safety profiles have already been determined. Nucleic acid vaccines have been extensively studied in the context of recent influenza, Ebola, and Zika outbreaks (149). In addition, the clinical development process for COVID-19 vaccines was accelerated by conducting trials concurrently rather than sequentially. For example, several COVID-19 vaccine candidates entered clinical trials without preclinical data in animal models, and many vaccine trials used an integrated phase I/II or II/III approach to save time (150). Companies are expanding their manufacturing capacity to meet global demand, and governments are playing a significant role in financing. As a result of this

extreme effort, the WHO has approved several COVID-19 vaccines and given them an emergency use listing (151) (Table 1-5).

Table 1-5 The emergency use vaccine list for SARS-CoV-2.

Company	Vaccine name	Approval date	Platform
Pfizer/BioNTech	BNT162b2	31 December 2020	RNA
SII/COVISHIELD and AstraZeneca	AZD1222	16 February 2021	Viral vector
Johnson & Johnson	Janssen/Ad26.COV 2.S	12 March 2021	Viral vector
Moderna	mRNA 1273	30 April 2021	RNA
Sinopharm	COVID-19 vaccine	7 May 2021	Inactivated
Sinovac	CoronaVac	1 June 2021	Adsorbed COVID-19 (inactivated)
Bharat Biotech	BBV152 COVAXIN vaccine	3 November 2021	Inactivated
Data were taken from published studies (147, 151).			

1.7 Proteomic analysis of coronavirus protein expressing cells

PPIs in response to viral infection can advance our knowledge about disease pathogenesis and identify potential druggable targets after identifying and mapping the alterations in cellular processes (152). In addition to directly targeting viral proteins, one antiviral development strategy is attempting to target host factors required for the virus to complete its lifecycle. Viruses have significantly less coding space in their genomes than their hosts. Consequently, they rely on host proteins to supplement the activities of their proteins (153). There are two primary advantages to indirectly inhibiting a virus via an essential host factor. First, since multiple viruses may use the same host protein, host-directed therapeutics have the potential to be broadly acting. Second, while direct virus targeting can rapidly select for resistant viral mutants, it is believed to be much more difficult for a viral mutation to overcome the inhibition of a co-opted host protein (153). Most coronaviral (SARS-CoV-2, SARS-CoV, and MERS-CoV) proteins have the same localisation pattern, supporting this hypothesis (153). Understanding which host factors coronaviruses require to infect and replicate in cells is crucial to identify targets for antiviral drug development.

As described above, the coronavirus replication cycle involves receptor binding, fusion/endocytosis, nucleocapsid entry into the cytoplasm, viral replication protein translation, genome replication, sub-genomic RNA production and translation into structural and accessory proteins, and virion assembly and release. In principle, each step may serve as a target for antiviral intervention. However, while the coronavirus entry step is relatively well understood, the host-virus interaction during subsequent viral life cycle stages is less well understood.

Most studies examining host cell changes in response to coronavirus proteins use a protein expression approach. Previous interactomic studies have identified the cellular interactome for various coronavirus proteins via their immunoprecipitation (often using an affinity tag) and high-throughput quantitative proteomic analysis to identify host proteins interacting with them. The stable isotope labelling of proteins by amino acids in cell culture (SILAC) and tandem mass tagging (TMT) are frequently used as differential labelling techniques. These techniques are followed by liquid chromatography (LC) and tandem mass spectrometry LC-MS/MS analyses, which is the most frequently used strategy for detecting changes in cellular protein levels during coronavirus infection. This approach was first used in a study examining 26 SARS-CoV-2 tagged viral proteins (of the 29 total viral proteins) that identified 332 high-confidence protein-protein interactions between SARS-CoV-2 and human proteins Gordon et al (154). These findings included 66 druggable human proteins by 69 compounds, of which 29 are US FDA approved, 12 are in clinical trials, and 28 are in preclinical trials (154). A second

study using the same approach identified 295 host proteins as SARS-CoV-2 protein interactors Li et al (92). The latter study highlighted interactions between viral proteins, including ORF9b, N, and ORF3a with SARS-CoV-2 E protein and nsp5, OFR10, ORF7a, nsp16, ORF6, S, nsp2, nsp8, N, S, and ORF7b with SARS-CoV-2 M protein (92). Moreover, an atlas of 2422 human proteins proximal to 17 SARS-CoV-2 viral proteins was created by a third study Meyers et al (103).

In addition, a fourth study identified interesting interactions between 1086 cellular proteins and 27 SARS-CoV-2 and 24 SARS-CoV viral proteins Stukalov et al (155). Unlike previous studies, this study covered the proteome of two coronaviruses. Moreover, three lethal coronaviruses were examined by a fifth study Gordon et al (153), which reported 389 high-confidence SARS-CoV-2 interactors, 366 SARS-CoV interactors, and 296 MERS-CoV interactors (153). Functional enrichment analyses showed that all three viruses target the same host processes, including ribosome biogenesis and RNA metabolism regulation (138). In addition, most of the proteins common to SARS-CoV-2, SARS-CoV, and MERS-CoV showed similar localisation patterns, supporting the hypothesis that conserved proteins share functional similarities (153).

1.8 Project aims

As previously stated, coronavirus diseases are a current global health concern. While considerable effort has been made to identify host and viral factors involved in viral growth, its underlying physiological mechanisms remain unknown. The project's overarching goal is to use high-throughput quantitative proteomics to study viral protein interactions in human, bat, and camel cells to identify host proteins that may modulate coronavirus replication. The identification of host proteins that regulate coronavirus replication could provide new antiviral targets against which strategies can be developed to prevent this serious disease.

The project's specific objectives are:

1) Bioinformatically analyse an existing proteomic data set created by high-throughput pull-down experiments with MERS-CoV E and M proteins and human proteins and validate high-confidence interactions via WB and colocalisation.
2) Transiently express the MERS-CoV and SARS-CoV-2 *E* and *M* genes in selected human, bat, and camel cells and perform a high-throughput interatomic analysis using mass spectrometry.
3) Bioinformatically compare the human, bat, and camel cell data sets to identify highly conserved interactions between MERS-CoV and SARS-CoV-2 E and M viral proteins and host cell proteins.

4) Confirm the relevance of the selected interactions by depleting the host proteins of interest via small interfering RNA (siRNA) knockdown.

At the conclusion of this study, it is expected that highly conserved interactions between MERS-CoV and SARS-CoV-2 viral proteins and host cell proteins will have been identified, which have the potential to be developed as antiviral targets in the future.

CHAPTER 2 Materials and Methods

2.1 Cells and culture conditions

Seven cell lines were used in this study. Three cell lines were derived from human embryonal kidney cells (HEK293). The parental Flp-In™ T-REx™-293 cell line (Invitrogen™, ThermoFisher Scientific, MA, USA, termed HEK293-Flp) and stable Flp-In™ T-REx™-293 cells inducibly expressing FLAG epitope-tagged versions of either the MERS-CoV E or M proteins (termed HEK293-E or HEK293-M, respectively). These stable cell lines were previously produced by Mr Lee in the laboratory. The stable Flp-In™ T-REx™-293 cell lines were selected in a medium containing hygromycin (Sigma-Aldrich) (100 µg/ml) and blasticidin (Sigma-Aldrich) (150 µg/ml), and transcription of the genes of interest was induced by the addition of tetracycline/doxycycline (Clontech, Saint-Germain-en-Laye, France) to a final concentration of 1 µg/ml. The other four cell lines were: human embryonal kidney cells (HEK293), *Camelus dromedarius* fibroblast cells (Dubca (ATCC® CRL2276™)), *Pteropus alecto* kidney cells (Pakit cell) (156), and human adenocarcinoma alveolar basal epithelial cells (A549) transduced to express both angiotensin-converting enzyme 2 (ACE2) and transmembrane protease serine 2 (TMPRSS2) (A549 A2 T2) cells (157). The A549 A2 T2 cells were selected in medium containing Geneticin (G418) (2mg/ml) and Hygromycin B (200 µg/ml).

The cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing GlutaMAX™-1, 4.5g/L D-glucose, and pyruvate (Gibco™, ThermoFisher Scientific, Paisley, UK) supplemented with 0.1 mM non-essential amino acids (NEAA) (Gibco™, ThermoFisher Scientific), 10% heat-inactivated foetal calf serum (FCS) (Gibco™, ThermoFisher Scientific) and 10,000 U/ml penicillin / 10,000 µg/ml streptomycin (Gibco™, ThermoFisher Scientific). The cells were grown at 37 °C in a humidified incubator with 5% CO_2. The cells were tested for mycoplasma and found to be negative.

Adherent cell lines were detached from the culture vessel and passaged as follows. The growth medium was removed, and cells were washed once with warm phosphate buffered saline (PBS) (Gibco™, ThermoFisher Scientific). The PBS was removed, and sufficient 0.05% trypsin/EDTA (Gibco™, ThermoFisher Scientific) was added to cover the surface of the flask. The cells were incubated at 37 °C with periodic observation until they detached from the surface of the flask. After adding of the appropriate media and centrifuging of the cells at 200

g for 5 minutes, the cell pellet was resuspended in the appropriate media. The desired number of cells (depending on the growth rate of the cells) was then added to a new flask containing fresh media.

The cell number was counted before passaging, and a defined number of cells was transferred to a new culture vessel as follows. An aliquot of culture medium containing the detached cells was sampled and the cell number determined by counting using a haemocytometer. Either a volume of culture medium containing the required number of cells was used directly for cell seeding or the cells were concentrated by centrifugation at 100 *g* for 10 min, and the cell pellet was then resuspended in fresh medium at the required cell density.

2.2 Cell transfection

Cells in a T75 flask (average number of 4.1 x 10^6 cells) were washed once with PBS. After the washing, 13 ml of an improved minimal essential medium (Opti-MEM) containing GlutaMAX™-1 and reduced serum medium (Gibco™, ThermoFisher Scientific, Paisley, UK) were added and the cells were incubated for 2h. Following the incubation, 2ml from Opti media, 2.5µg from DNA samples (plasmid), and 16µl form Polyethylenimine were added to a sample tube, vortexed and incubated for 15 min before being added to the cells. The cells were then grown at 37 °C in a humidified incubator with 5% CO_2 for 48h. Cells were harvested as described in Section 2.1, and the immunoprecipitation (IP) protocol was used to prepare a sample.

2.3 Cell viability assay

Cell viability was determined using a Vybrant MTT assay (Invitrogen™, Thermo Fisher Scientific). Cells were seeded in standard 96 well plates at a density of 3000 cells per well and cultivated in DMEM with 10% FBS for 48 h. The supernatant was removed, and the cells were washed three times with warm PBS, followed by the addition of 100 µl of a 1mg/ml MTT stock solution (Appendix A). The plate was incubated for 3 h. The MTT solution was removed, and 100 µl of lab-grade ethanol was added and mixed well by pipetting. The plate was covered in foil and left on a shaker for 30 min before the absorbance was immediately measured by a Promega plate reader (Promega Glomax Explorer, Promega, UK) at 600 nm.

2.4 Protein analysis

2.4.1 Sample preparation

To prepare a protein sample for Western blot analysis, cells were grown and harvested as mentioned in Section 2.1, lysed with 2X sample buffer (Appendix A), and then heated at 95 °C for 5 min. Either the samples were then stored at – 20 °C or analysed directly.

2.4.2 SDS Polyacrylamide electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (158) was used throughout the study using a Bio-Rad mini-protein apparatus and power pack (Bio-Rad, CA, USA). 15% SDS resolving gels and 5% SDS stacking gels were prepared as described in Appendix A. Each protein sample was mixed with 4X Laemmli buffer (Appendix A) in a final volume containing 10% β-mercaptoethanol as a reducing agent. Separation of proteins by SDS-PAGE was done using 1X SDS-PAGE running buffer (Appendix A) with a constant voltage of 80 V for the first 10 min, followed by 120 V for the next 60 min. The PageRuler™ Plus Prestained Protein Ladder (Thermo Fisher Scientific, MA) was used as a protein molecular weight marker. Finally, proteins were detected by Western blot.

2.4.3 Western blot (WB)

After proteins were separated by SDS-PAGE, they were transferred to a polyvinylidene difluoride (PVDF) membrane by Western blot using a Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (Bio Rad, CA, USA) for 60 min at 15 V. The PVDF membrane was soaked in methanol for 1 min, then in water for 2 min, and lastly in transfer buffer (Appendix A) for 15 min before protein transfer. Either the membrane was then blocked overnight at 4 °C or for 1 h at room temperature (RT) using a 5% (w/v) skim milk powder solution in PBS (blocking buffer).

In the next step, the blot was washed once with 1X phosphate buffered saline +1% Tween (PBST) and probed with a primary antibody (Table 2-1) diluted appropriately in blocking buffer (Appendix A) either for 1 h at RT or overnight at 4 °C on a shaker. To remove excess primary antibody, the membrane was then washed for 5 min three times with 1X PBST, then incubated with an HRP-conjugated secondary antibody (Table 2-2) diluted appropriately in blocking buffer for 1 h at RT on a shaker. Next, the membrane was washed as above, drained, and covered with the KPL LumiGLO® Chemiluminescent Substrate (Kirkegaard & Perry

Laboratories, 37 MD, USA) for 1 min. Finally, the membrane was exposed to X-ray film (Amersham HyperfilmTM ECL, GE Healthcare Life Sciences, Buckinghamshire, UK) and developed in a Compact X4 Automatic X-ray Film Processor (Xograph Healthcare, Gloucestershire, UK).

Table 2-1 Primary antibodies used for WB analysis in this study

Name	MW (kDa)	Source	Working dilution	Catalogue number	Clone No.	Manufacturer
Anti-FLAG epitope	----	Mouse	1:1000	F1804	Polyclonal	Sigma-Aldrich (Saint Louis, USA)
Anti-β-tubulin	55	Rabbit	1:1000	CST2146	EPR16778	Cell Signaling Technology (London, UK)
Anti-TM9SF2	75.7	Rabbit	1:200	ab121227	Polyclonal	Abcam (Cambridge, UK)
Anti-TMEM43	45	Rabbit	1:750	ab184164	EPR15378(B)	Abcam
Anti-CERS2	45	Rabbit	1:500	ab227501	Polyclonal	Abcam
Anti-YIPF5	28	Rabbit	1:100	HPA073622	Polyclonal	Atlas Antibodies AB (Stockholm, Sweden)
Anti-ERGIC1	32.6	Rabbit	1:250	NBP1-83962	Polyclonal	Novus Biological Europe (Abingdon, UK)
Anti-SLC44A2	80.1	Mouse	1:500	H00057153-M01	Polyclonal	Novus Biological Europe
Anti-IPO11	113	Rabbit	1:400	ab221615	Polyclonal	Abcam
Anti-VDAC1	31	Rabbit	1:250	ab154856	EPR10852(B)	Abcam
Anti-RAB10	22.5	Rabbit	1:1000	ab237703	MJF-R23	Abcam
Anti-LPCAT1	59	Rabbit	1:250	ab214034	EPR19882	Abcam
Anti-SCAMP3	38	Rabbit	1:250	HPA071167	Polyclonal	Atlas Antibodies AB
Anti-GAPDH	36	Mouse	1:200	A85271	GA1R	Antibodies

Anti-Lamin A/C	74/63	Mouse	1:100	MA3-1000	mab636	ThermoFisher

Table 2-2 Secondary antibodies for WB used in this study

Name	Source	Working dilution	Catalogue number	Manufacturer
Anti-mouse IgG, HRP conjugated	Goat	1:3000	12-349	Sigma-Aldrich
Anti-rabbit immunoglobulin/horseradish peroxidase (HRP) conjugated	Goat	1:5000	P0448	Dako (Santa Clara, US)

2.4.4 Immunoprecipitation (IP) protocol

Cells were grown, transfected, and harvested as mentioned in Section 2.1 before being lysed in lysis buffer (Appendix A) and kept for 30 min on ice with pipetting every 10 min, followed by sonication (4X – pulse of 5 sec,10 sec pause, amplitude 60%: Sonics, Newtown, US). The cell lysates were then centrifuged at 17,000 *g* for 20 min at 4 °C. The supernatant was removed to a fresh tube, and a small aliquot removed and stored at −20 °C (input sample). The rest of the lysate was incubated at 4 °C with rotation overnight with equilibrated FLAG-tagged magnetic beads (Sigma Aldrich) with equilibration buffer (Appendix A), (40µl /T75 flask) (50µl /T225 flask). A magnetic rack was used to separate the beads, and the supernatant (non-bound protein sample) was collected and stored at −20 °C. The bead pellet was then washed twice using wash buffer (Appendix A). During the last step, two approaches were employed.

For a proteomic analysis, out of the 500 µl of washing buffer with the bead pellets, 450 µl were drawn off and a minimal amount 50 µl of the washing buffer retained to prevent the beads from drying out before the sample was sent for the proteomic analysis. That remainder of 50 µl, with the bead pellets, was taken for WB after elution, heated in 2X sample buffer (Appendix A), and stored at −20 °C.

For the WB analysis, the 500 µl washing buffer was taken at the last washing step, and a 2X sample buffer (Appendix A) was added to the beads before the sample was heated at 95 °C for 3 min. The magnetic beads were separated using a magnetic rack and the eluted proteins in the 2X sample buffer removed to a new tube and stored at −20 °C. The input, non-bound, and pulldown samples were analysed by WB using an anti-FLAG antibody to ensure that the pulldown was successful.

2.5 Immunofluorescence assay

An immunofluorescence assay (IFA) was used to detect proteins of interest in cells as follows. An appropriate number of cells was seeded onto either a cell culture 96-well imaging microplate (section 2.8) or a glass cover slip coated with poly-D-lysine (Sigma-Aldrich) diluted in PBS at a concentration of 0.1 mg/ml in a 24-well plate for 5 min before being washed three times with PBS. At an appropriate time after seeding/transfection/infection, the cells were washed with ice-cold PBS, fixed with ice-cold 4% paraformaldehyde in PBS for 5 min, and then washed twice with PBS. To permeabilise the cells, they were incubated in 1% Triton X-100 in PBS for 5 min, followed by two PBS washes. The cells were then blocked with a blocking solution (appendix A) for 1 h at RT. The blocking solution was replaced with a primary antibody (Table 2-3) diluted in blocking solution for 60 mins at RT.

The coverslips were then washed four times with PBS for 5 min and stained with an appropriate secondary antibody (Table 2-4) diluted in blocking solution for 30–60 mins at RT. The cells were then washed four times with PBS for 5 min and mounted onto slides using VectaShield mounting medium containing DAPI (Vector Labs, CA, USA). The cells labelled by immunofluorescence were analysed using a Nikon Eclipse TS100 microscope or a widefield imaging system (Leica DM 16000 inverted epifluorescence microscope) in the Wolfson Bioimaging Facility in the Faculty of Life Sciences. Cell imaging and data were collected and analysed by the Fiji software (Version 1.8.0).

The 96 well plates were then washed three times with PBS and stained with DAPI and an appropriate secondary antibody (Table 2-4) diluted in blocking solution for 30 mins at RT. The cells were then washed three times with PBS and imaged immediately or sealed with a silver foil and stored at 4°C. The cells labelled by immunofluorescence were analysed using an Image Xpress Pico (Molecular Devices). A 2-channel assay was used for scoring cells based on DAPI and Tetramethylrhodamine (TRITC), before the positive cells number and percentage were quantified and used to statistical analysis To avoid the variability in a cell number (Hi MOI ≈ 3200 cells and Low MOI ≈ 9200 cells) , we used the percentage of positive cells for the statistical analysis. Statistical analysis was carried out using one-way ANOVA in GraphPad Prism 9.4.1, and if a p-value is less than 0.05, it is flagged with (*), if a p-value is less than 0.01, it is flagged with (**), and if a p-value is less than 0.001, it is flagged with (***).

Table 2-3 Primary antibodies used for IFA in this study

Name	Mw (kDa)	Source	Working dilution	Catalogue number	Clone No.	Manufacturer
Anti-FLAG epitope	---	Mouse	1:100	F1804	Polyclonal	Sigma-Aldrich
Anti-FLAG epitope	---	Rabbit	1:100	F7425	Polyclonal	Sigma-Aldrich
Anti-TM9SF2	75.7	Rabbit	1:200	ab121227	Polyclonal	Abcam
Anti-TMEM43	45	Rabbit	1:750	ab184164	EPR15378(B)	Abcam
Anti-CERS2	45	Rabbit	1:500	ab227501	Polyclonal	Abcam
Anti-YIPF5	28	Rabbit	1:100	HPA073622	Polyclonal	Atlas Antibodies AB
Anti-ERGIC1	32.6	Rabbit	1:250	NBP1-83962	Polyclonal	Novus Biological Europe
Anti-SLC44A2	80.1	Mouse	1:500	H00057153-M01	Polyclonal	Novus Biological Europe
Anti-IPO11	113	Rabbit	1:400	ab221615	Polyclonal	Abcam
Anti-VDAC1	31	Rabbit	1:250	ab154856	EPR10852(B)	Abcam
Anti-RAB10	22.5	Rabbit	1:1000	ab237703	MJF-R23	Abcam
Anti-LPCAT1	59	Rabbit	1:250	ab214034	EPR19882	Abcam
Anti-SCAMP3	38	Rabbit	1:250	HPA071167	Polyclonal	Atlas Antibodies AB
Anti-GAPDH	36	Mouse	1:200	A85271	GA1R	Antibodies
Anti-Lamin A/C	74/63	Mouse	1:100	MA3-1000	mab636	ThermoFisher
Anti-SARS-CoV-2 N	45.6	Rabbit	1:1000	200-401-A50	Polyclonal	ThermoFisher

Table 2-4 Secondary antibodies for IFA used in this study

Name	Working dilution	Catalogue number	Manufacturer
Alexa Fluor® 488 Goat anti-Mouse IgG (H+L) Antibody	1:2000	A11029	Thermo Fisher
Alexa Fluor®568 Goat anti-Rabbit IgG (H+L) Antibody	1:2000	A11011	Thermo Fisher
Alexa Fluor® 555 Goat anti-Mouse IgG (H+L) Antibody	1:1000	A32727	ThermoFisher

2.6 Quantitative measurement of protein-protein colocalisation in cells

Measuring colocalisation between two fluorescently labelled proteins is important, especially with a high chance of random error and bias when using a visually based colocalisation technique (1). The level of colocalisation of each cell was quantified by calculating the Pearson's Correlation Coefficient (PCC) using the Coloc 2 plugin from Fiji software. An intact cell was selected as a region of interest (ROI) for the PCC analysis, which examines the relationship between the pixel intensities of two channels within the same image. Fifty cells were analysed, and a PCC mean was obtained. The PCC values range from -1 to 1, whereby a PCC value of 1 indicates perfect correlation, 0 no correlation, and -1 a perfect anti-correlation (138). The fixed and stained cells (section 2.5) were imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility in the Faculty of Life Sciences before analysis with the Fiji software (version 1.8.0).

2.7 Nucleic acid methods

2.7.1 Agarose gel electrophoresis

Agarose gels were prepared using: 1X TBE buffer (Appendix A), 0.5 µg/ml ethidium bromide, and agarose gel (Invitrogen, USA) (Appendix A). A 0.2 volume of 6X DNA loading buffer (Appendix A) was mixed with DNA samples before these samples were run at 100 V for nearly 1 h by electrophoresis through a 0.8–1.0% (w/v) agarose gel in 1X TBE buffer. A Gene Ruler 1kb Plus DNA Ladder (Life Technologies, CA, USA) was used to estimate the size and

concentration of the DNA. The BioDoc-ITTM System Ultraviolet transilluminator (UVP, CA, USA) was then used to visualise and photograph DNA bands before the concentration of DNA samples was determined using a Nanodrop spectrophotometer (Thermo Fisher Scientific, MA, USA).

2.7.2 Bacterial strains and media

The *Escherichia coli* α-Select Silver Efficiency strain DH5 α (Bioline, UK) was used and routinely cultured on Luria-Bertani (LB) agar plates (containing 1.2% (w/v) agar) or LB broth (Appendix A) (both containing 100 µg/ml ampicillin) and were incubated aerobically at 37 °C overnight. Table 2-5 lists the plasmids used in this study which contained commercially synthesised (GeneArt, Invitrogen™, ThermoFisher Scientific) genes (Appendix B). These plasmids were comprised of pcDNA3.3-TOPO as a backbone and contained specific optimised coronavirus gene sequences fused to a 3′ terminal FLAG epitope tag sequence, such that the encoded proteins contained a C-terminal FLAG epitope tag.

Table 2-5 Plasmids used in this study

Plasmid Name	Selection marker	Source
MERS CoV – E (Envelope)	Ampicillin	Invitrogen by ThermoFisher Scientific
MERS CoV – M (Membrane)	Ampicillin	
SARS CoV-2 E (Envelope)	Ampicillin	
SARS2 CoV-2 M (Membrane)	Ampicillin	

2.7.3 Bacterial transformation

Bacterial transformation was performed by adding 50 µl of chemically competent *E. coli* strain DH5 α cells onto 10–25 ng of plasmid DNA followed by incubation on ice for 30 min. The bacteria were heat-shocked at 42 °C for 20 sec and immediately placed on ice for 2 min. Approximately 950 µl of pre-warmed LB broth was added to the cells, followed by shaking at 225 rpm for 1 h at 37 °C. An LB agar plate containing ampicillin was then used to cultivate the transformed cells.

2.7.4 Plasmid propagation

After overnight incubation of LB agar plate of the transform cells at 37 °C, single colonies were selected and inoculated into LB broth containing ampicillin and grown overnight at 37°C with shaking at 225 rpm. Plasmid DNA was extracted from the bacterial culture before the desired plasmids were identified by restriction enzyme digestion.

2.7.5 Preparation of plasmid DNA

The *E. coli* strain DH5α was used to maintain plasmids as grown in a LB agar plate containing ampicillin. For plasmid extraction, a single colony was picked and inoculated in 5 ml of LB broth before a miniprep procedure using a GeneJET™ Plasmid Miniprep Kit (Thermo Fisher Scientific, MA, USA) was used according to the manufacturer's instructions. For a large plasmid DNA propagation, a single colony was picked and inoculated in 100 ml of LB broth before the Pure Link™ Hi Pure Plasmid Midiprep Kit (Thermo Fisher Scientific, UAB) was used according to the manufacturer's instructions. Plasma DNA in midi preparations was then concentrated by mixing with 0.1 volumes of sodium acetate (pH 5.5) and 2.5 volumes of cold 100% ethanol before incubation at −20 °C to precipitate. The sample was then centrifuged at 13000 *g* for 20 min at 4 °C, and the pellet washed with cold 70% (v/v) ethanol followed by further centrifugation at 13000 *g* for 20 min at 4 °C. The pellet was air dried, resuspended in nuclease-free water, and stored at −20 °C.

2.7.6 Restriction endonuclease reactions

Restriction digestions were done using 5 U of restriction endonuclease per µg of DNA. Restriction digest screening reactions were carried out in the buffer supplied by the manufacturer, in a total volume of 15 µl at 37 °C for 60 min. Table 2-6 shows the restriction enzymes used.

Table 2-6 Restriction endonucleases used in this study

Restriction endonuclease	Source
BamHI	New England Biolabs
XhoI	New England Biolabs

2.8 siRNA transfection conditions

All siRNAs were designed and synthesised by Dharmacon (GE Dharmacon, CO, USA). On the day of transfection, the transfection reagent Lipofectamine RNAiMAX (Invitrogen, Thermo

Fisher Scientific, Lithuania) was used to transfect A549 A2 T2 cells. Then, 0.15 µl/well of Lipofectamine RNAiMAX was mixed with 14.85 µl/well of serum-free medium (Opti-MEM) to a final volume of 15 µl/well and incubated for 5 min at RT. Each siRNA was diluted in 1X siRNA buffer (GE Dharmacon, CO, USA) to make a 1 µM stock. Each siRNA was then mixed with Opti-MEM to reach an appropriate final concentration of 50 nM in a total volume of 15 µl (Table 2-7). The siRNA mixture was added to the Lipofectamine RNAiMAX mix and incubated for 20 min at RT to allow a transfection complex to form. After that, the transfection complexes were added to a cell culture 96-well imaging microplate (Greiner bio one, Germany) (30µl/well) and gently mixed to ensure even distribution. A549 A2 T2 cells were counted and seeded at a density of 3000 cells/well in 70 µl of DMEM medium containing 10% FBS and no antibiotics (siRNA growth media) and then incubated at 37 °C. Untreated cells (control cells) were transfected with transfection reagent only. At 48 h post-transfection, the cells were checked to ensure they were healthy before being infected. The cell infection was done by Prof. Andrew, who started by removing cultural medium before infection and washed the cells with PBS before adding MEM containing 0.1 mM NEAAs and 2% FBS with no antibiotics (infection medium). After adding the SARS-CoV-2, the cells were incubated for 24 h at 37 °C. The virus titre was: VTN7 titre: 1.3 x 10e8 (VTN); 3 x 10e7 (Vero) and the multiplicity of infection (MOI) of 1 = 5000 virus particles = 0.2 ul /well, this was the lower MOI (LMOI). To get a higher multiplicity of infection (HMOI =5) ,1-2 ul /well (from the titration) was added. The MOI was obtained by diluting the virus inoculum in the infection media. 24 h after infection, the culture medium was removed, and cells were fixed, stained, labelled by immunofluorescence, imaged, and analysed as described in Section 2.5.

Table 2-7 siRNA used in this study

Name	Catalogue Number	Name	Catalogue Number
VDAC1	L-019764-00	SAAL1	L-015583-01
SLC25A6	L-007487-00	SAMM50	L-017871-02
ABCB7	L-007305-00	SAP18	L-012140-00
ACAT1	L-009408-00	SCAMP3	L-013442-00
CDIPT	L-009631-01	SGPL1	L-008747-00
CERS2	L-010282-00	SLC1A5	L-007429-00
CSE1L	L-004413-00	SLC38A10	L-007359-03
CWC22	L-023101-02	SLC44A2	L-018034-01
DNAJB12	L-020585-01	TECR	L-009537-01
DYNC1I2	L-012574-00	TM9SF2	L-010221-01
ELOVL5	L-009260-01	TMEM43	L-014342-02
ERGIC1	L-010722-00	TNPO1	L-011308-00
FLOT2	L-003666-01	TRIP4	L-009632-00
IMMT	L-019832-01	TUBG1	L-005160-00
IPO11	L-015397-00	UBA52	L-011794-00
JUP	L-011708-02	UNC45A	L-017653-01
LPCAT1	L-010289-00	USP34	L-006082-00
NUP35	L-018998-01	YIPF5	L-018962-01
PELO	L-019068-01	ZMPSTE24	L-006104-00
PNN	L-019228-00	Non-targeting Control	D-001206-13
RAB10	L-010823-00	Lamin A/C Control	D-001050-01
RAB1A	L-008283-00	KIF11	L-003317-00-0010
RBM17	L-005158-01	GAPD Control	D-001140-01
RBM39	L-011965-00		

2.9 Quantitative mass spectrometry analysis

2.9.1 TMT labelling and high pH reversed phase chromatography

Immuno-isolated samples were prepared as in Section 2.4.4 before it reduced (10 mM tris (2-carboxyethyl) phosphine (TCEP)), 55°C for 1h), alkylated (18.75 mM iodoacetamide, at room temperature for 30 min) and then digested from the beads with trypsin (2.5µg trypsin; 37°C, overnight). The resulting peptides were then labelled with TMT ten-plex reagents according to the manufacturer's protocol (Thermo Fisher Scientific, Loughborough, LE11 5RG, UK), and the labelled samples pooled and desalted using a SepPak cartridge according to the manufacturer's instructions (Waters, Milford, Massachusetts, USA). Eluate from the SepPak cartridge was evaporated to dryness and resuspended in buffer A (20 mM ammonium hydroxide, pH 10) prior to fractionation by high pH reversed-phase chromatography using an Ultimate 3000 liquid chromatography system (Thermo Scientific). In brief, the sample was loaded onto an XBridge BEH C18 Column (130Å, 3.5 µm, 2.1 mm X 150 mm, Waters, UK) in buffer A and peptides eluted with an increasing gradient of buffer B (20 mM ammonium hydroxide in acetonitrile, pH 10) from 0–95% over 60 minutes. The resulting fractions (4 in total) were evaporated to dryness and resuspended in 1% formic acid prior to analysis by nano-LC MSMS using an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific).

2.9.2 Nano-LC mass spectrometry

High pH RP fractions were further fractionated using an Ultimate 3000 nano-LC system in line with an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). In brief, peptides in 1% (vol/vol) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (vol/vol) acetonitrile 0.1% (vol/vol) formic acid, peptides were resolved on a 250 mm × 75 µm Acclaim PepMap C18 reversed-phase analytical column (Thermo Scientific) over a 150 min organic gradient, using 7 gradient segments (1–6% solvent B over 1 min., 6–15% B over 58 min., 15-32%B over 58 min., 32–40%B over 5 min., 40–90%B over 1 min., held at 90%B for 6 min and then reduced to 1%B over 1 min.) with a flow rate of 300 nl min−1. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionised by nano-electrospray ionisation at 2.0kV using a stainless-steel emitter with an internal diameter of 30 µm (Thermo Scientific) and a capillary temperature of 300 °C.

All spectra were acquired using an Orbitrap Fusion Lumos mass spectrometer controlled by Xcalibur 3.0 software (Thermo Scientific) and operated in data-dependent acquisition mode using an SPS-MS3 workflow. FTMS1 spectra were collected at a resolution of 120,000, with

automatic gain control (AGC) target of 200,000 and a max. injection time of 50ms. Precursors were filtered with an intensity threshold of 5000, according to charge state (to include charge states 2–7) and with monoisotopic peak determination set to peptide. Previously interrogated precursors were excluded using a dynamic window (60s +/-10ppm). The MS2 precursors were isolated with a quadrupole isolation window of 0.7m/z. ITMS2 spectra were collected with an AGC target of 10,000, a max. injection time of 70ms and a CID collision energy of 35%.

For FTMS3 analysis, the Orbitrap was operated at 50,000 resolution with an AGC target of 50,000 and a max. injection time of 105ms. Precursors were fragmented by high-energy collision dissociation (HCD) at a normalised collision energy of 60% to ensure maximal TMT reporter ion yield. Synchronous Precursor Selection (SPS) was enabled to include up to 10 MS2 fragment ions in the FTMS3 scan.

2.9.3 Data Analysis

The raw data files were processed and quantified using Proteome Discoverer software v2.1 (Thermo Scientific) and searched against Uniprot/SwissProt human (UP000005640), Bat (UP000010552), and Camel (UP000299084) databases and FASTA files containing the SARS-CoV-2 E (GenBank Protein Identification: QHD43418.1), MERS-CoV E (GenBank Protein Identification: YP_009047209.1), SARS-CoV-2 M (GenBank Protein Identification:QHD43419.1), and MERS-CoV M (GenBank Protein Identification: YP_009047210.1) databases using the SEQUEST HT algorithm. Peptide precursor mass tolerance was set at 10ppm, and tandem mass spectrometry (MS/MS) tolerance was set at 0.6Da. Search criteria included oxidation of methionine (+15.995Da), acetylation of the protein N-terminus (+42.011Da) and methionine loss plus acetylation of the protein N-terminus (-89.03Da) as variable modifications and carbamidomethylation of cysteine (+57.021Da) and the addition of the TMT mass tag (+229.163Da) to peptide N-termini and lysine as fixed modifications. Searches were performed with full tryptic digestion, and a maximum of 2 missed cleavages were allowed. The reverse database search option was enabled, and all data was filtered to satisfy a false discovery rate (FDR) of 5%. The main search was done by Dr Kate Heesom and the ThermoFischer msf file and excel spreadsheet were generated for further analysis. The above proteomics analysis, and as such the protocols listed in section 2.9 were provided as a service from the University of Bristol, Faculty of Life Sciences Proteomics facility.

2.10 Quantification and bioinformatics analysis

The initial bioinformatics and statistical analyses were done by Dr Philip Lewis (Faculty Proteomics Facility) using the statistical computing and graphics program R. Protein groupings were determined by PD2.1. However, the master protein selection was improved with an in-house script which enabled inference of biological trends more effectively in the data set with no loss in the quality of identification or quantification.

The raw abundance data were Log_2 transformed to bring the data closer to a normal distribution. Welch's t-test was used to calculate p-values where appropriate and adjusted for multiple testing using the Benjamini Hochberg Procedure. These data were displayed alongside mean log_2 fold changes and counts of the number of samples used to calculate these statistics. The Galaxy web site (usegalaxy.org) was used to annotate the Dubca cells database by using a camel protein accession number to get a human protein name, while an online version of the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (159) was used to annotate the whole Pakit cells database. The functional enrichment analysis program FunRich version 3.1.3 (160) was used to compare data sets and identify common proteins in interaction lists and produce Venn diagrams. Gene ontology (GO) analysis was done on the top 10% of proteins that showed > 0 log_2 fold change in HEK293, Dubca, and Bat transfected cells with SARS-CoV-2 E, MERS-CoV E, MERS-CoV M, SARS-CoV-2 M proteins compared to negative non-transfected control cells. The GO analysis was done by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (161). STRING was used to generate an interaction network of host proteins that was represented as a node in the interaction network (labelled with the gene symbol). Nodes representing proteins that were most significantly enriched in the GO term are shaded in red.

CHAPTER 3 Comparative interactome analysis of MERS-CoV E and M proteins using stably transformed HEK293 cells

3.1 Introduction

Understanding which host factors coronaviruses require to infect a cell is crucial to better understand coronavirus replication and identify targets for developing antiviral drugs. As described in Section 1.5, the coronavirus replication cycle entails receptor binding, fusion/endocytosis, nucleocapsid entry into the cytoplasm, translation of viral replication proteins, genome replication, the production of sub-genomic RNAs and their translation into structural and accessory proteins, and virion assembly and release. Each step may serve as a target for an antiviral intervention in principle. While the coronaviruses' viral entry step is relatively well understood, the host-virus interaction during subsequent stages of the viral life cycle is less well understood. Several approaches have been used to investigate the cellular interactome of coronavirus proteins.

The most relevant approach to investigate a viral protein's cellular interactome is to pull down the unmodified viral protein from lysates produced from virus-infected cells using highly specific antibodies. To date, no published studies have reported using this approach to study the interaction partners of the coronavirus E and M proteins. However, coronavirus RNA-protein interactions have been identified using comparative RNA interactome capture followed by MS/MS (162). Four studies have identified RNA protein interactors for SARS-CoV-2 (163-166). Another approach to studying the viral-host cell interactome is to express a modified version of the viral protein with an affinity tag. This approach does not require viral protein-specific antibodies. It can be performed using individual recombinant viral proteins expressed exogenously in different cell types or in the viral genome context if the tag does not inhibit virus replication (167).

These tags include FLAG, hemagglutinin (HA), protein A, and hexahistidine (168). These tags can be as small as two amino acids or as large as complete proteins, some of which may comprise several subunits. Nucleic acid sequences encoding the tags can be engineered to either proceed or follow the sequences encoding either the N- or C-terminus of the target protein or, in some cases, both termini (169). An important feature of the selected tag is that it should not interfere with the native folding of the proteins to which it is attached. In addition, it should be water-soluble and highly exposed on the targeted protein's surface (169). One such tag is the HA tag, which is small, well-characterised, highly immunoreactive, and unlikely

to affect protein function (170). The HA tag was used to identify host cell proteins interacting with SARS-CoV-2 and SARS-CoV proteins by introducing individual lentiviruses expressing viral proteins with HA tags into A549 cells (155).

Our study used the FLAG tag. It is an interesting tag comprising eight amino acids (169). It is highly specific (171) and, as a peptide fusion system, comprises a novel and widely applicable method for identifying and purifying proteins (169). FLAG epitope tags were engineered at the C-terminus of the proteins, enabling their affinity pull-down of the target proteins using specific antibodies and the analysis of co-immunoprecipitates to identify their cellular interaction partners (169).

This approach is more powerful when a viral fusion protein containing the tag is tolerated during viral infection since the viral interactome can be analysed during the infection process in the context of other viral proteins. This approach was first demonstrated using a mutant Sindbis virus expressing the viral non-structural protein 3 tagged with a green fluorescent protein (GFP) (172). The tagged viral protein could be localization during virus infection, and host cell interactors identified (172). Moreover, this approach was used to identify >500 host proteins interacting with the replication complex of the coronavirus MHV after a biotin ligase tag was engineered into the nsp2 protein (173). This study conducted another experiment using GFP-tagged nsp2 (173). In addition, using a fluorescent tag permitted the simultaneous tracking of protein localization in living cells with the temporal regulation of virus-host interactions (174).

As described above, individual viral proteins engineered to express with an affinity tag can also be used for interaction studies when expressed stably or transiently in appropriate cells. Cellular overexpression of single epitope-tagged viral proteins had different biological effects on the host cell than infection with a recombinant virus, limiting the interpretation of the data (167). Specifically, this strategy lacked the viral cofactors and protein interactions that might influence viral replication and spread (167). In addition, the tagged viral protein may be expressed at markedly different levels than in viral infection, and its localisation may differ. Therefore, the virus-host protein interactions during infection might not be completely recapitulated, and false negative associations may arise (167). Nevertheless, this method has several advantages and can help determine the host cell interaction partners for a given viral protein (167). Transient expression of tagged coronavirus proteins is a commonly used method in interaction studies. Recently, 26 of 29 predicted SARS-CoV-2 proteins were transiently expressed with a Strep-tag in HEK-293T cells, identifying a genome-wide viral interactome (154). Another study transiently expressed MERS-CoV and SARS-CoV proteins

with the same Strep-tag in the same cells to identify their cellular interaction partners, identifying shared coronavirus interaction partners (153).

The identification of interactions between virus and host cell proteins has been revolutionised by high-throughput quantitative proteomic analysis, which allows both sensitive detection and quantitation of proteins interacting with the specific 'bait' protein compared to a generic protein control. Generally, one of two approaches has been used for protein quantitation: SILAC or TMT. Then, the isotopically labelled proteins/peptides are analysed by LC and LC-MS/MS. SILAC results in the labelling of cellular proteins with stable carbon (^{13}C) and/or nitrogen (^{15}N) isotopes. After approximately eight cell divisions, isotopically labelled amino acids contained in the cell culture medium are completely incorporated into the cellular proteins, and then cells are used for experiments. After collecting cell lysates, equal amounts of protein in the samples are combined and subjected to tryptic digestion before LC-MS/MS analysis. The SILAC method has been used to investigate the cellular response to coronavirus infection (175, 176) and (177). In the TMT method, protein samples are prepared individually and digested with trypsin. The resulting tryptic peptides are labelled with tandem mass tags that bind to amino groups before being combined in equal amounts for LC-MS/MS analysis. The labelled peptides are first visualised as single peaks and then fragmented to release the reporter ions. The released reporter ions indicate each peptide's quantity and protein (178). A distinct advantage of TMT is its multiplex capability, combining ≤15 different samples. This method has been applied to determine the host cell interaction partners for a given viral protein by Gordon et al. (154), Li et al. (92), Meyers et al. (103), Gordon et al. (153), and Stukalov et al. (155).

The coronavirus E and M proteins are located in the viral envelope (113). Relatively little is known about their function. Identifying interactions between these two proteins and cellular proteins will potentially increase our understanding of the virus lifecycle and identify host cell proteins that could be targeted in antiviral therapies. Stable HEK293T cell lines inducibly expressing FLAG-tagged MERS-CoV E (HEK293-E) and MERS-CoV M proteins (HEK293-M) were previously created by Mr Lee (a research technician in the Davidson/Matthews laboratory). Mr Lee used them for co-immunoprecipitation (co-IP) analysis, followed by MS/MS identification and protein quantitation to identify potential cellular interaction partners. This work was completed before the emergence of SARS-CoV-2.

This chapter reports a validation of the proteomic analysis previously performed by Mr Lee as follows: (i) the inducible expression of MERS-CoV E and M proteins in the HEK293-E and HEK293-M cell lines was validated by WB and IFA to confirm the cell lines' identity; (ii) co-IP

analysis was performed in triplicate using these cell lines, and WB was used to confirm the presence of the MERS-CoV E and M proteins in pull-down samples; (iii) the original MS/MS-based co-IP analysis was reanalysed by downstream bioinformatic analysis to identify proteins interacting with MERS-CoV E and M proteins with the highest confidence. Finally, the most interesting cellular interaction partners were selected and validated by WB and colocalisation analysis.

3.2 Results

3.2.1 Validation of cell lines expressing MERS-CoV E and M proteins

Before beginning to validate the proteomic data produced by Mr Lee, it was necessary to validate the inducible expression of the MERS-CoV E and M FLAG-tagged proteins in the stable HEK293-E and HEK293-M cell lines. The parental HEK293-Flp cell line was used as a negative control. For validation, the HEK293-E, HEK293-M, and HEK293-Flp cell lines were grown in the presence and absence of tetracycline (Tet) for 48 h to induce MERS-CoV E and M gene expression before MERS-CoV E and M protein levels and localisation was examined by WB and IFA, respectively.

The WB analysis presented in Figure 3-1 shows that the MERS-CoV E and M proteins were expressed by HEK293-E and HEK293-M cells, respectively, in response to Tet but not in HEK293-Flp cells. In addition, MERS-CoV E and M protein expression were detectable in the uninduced HEK293-E and HEK293-M cells, respectively, suggesting their expression may be leaky. Therefore, the HEK293-Flp cells were the most suitable negative control for further experiments rather than uninduced HEK293-E and HEK293-M cells.

The molecular masses expected for the FLAG-tagged MERS-CoV E and M proteins were 17.0 and 27.8 kDa, respectively (113, 179). A single band was detected at 17 kDa for the MERS-CoV E protein, consistent with its predicted molecular mass. While a major band was detected at ~28 kDa for the MERS-CoV M protein, other expected bands were also seen in the lysates from HEK293-M cells, suggesting it might form oligomers or be modified by ubiquitination or phosphorylation respectively (138). These MERS-CoV M protein bands are marked in Figure 3-1 to clarify the WB's MERS-CoV M protein band pattern. The β-tubulin loading control indicated approximately equal loading of the samples.

The E and M coronavirus proteins are known to localise to the ERGIC and participate in coronavirus assembly, budding, and intracellular trafficking (115). IFA was used to confirm the expression and cellular localisation of the FLAG-tagged MERS-CoV E and M proteins (Figure 3-2). Cells were stained with an anti-FLAG antibody and an antibody against ER protein 72 (Erp72), an ER protein (180). The MERS-CoV E and M proteins were produced by the HEK293-E and -M cells, respectively, but not from the HEK293-Flp control cells. However, the non-induced HEK293-E and -M cells also had a weak signal. The FLAG-tagged MERS-CoV E and M proteins (green) showed a clear colocalisation (yellow) with Erp72 (red), while there was a weak localisation in non-induced HEK293-E and -M cells. The colocalisation of the FLAG-tagged MERS-CoV M and E proteins with Erp72 clearly showed their ER localisation.

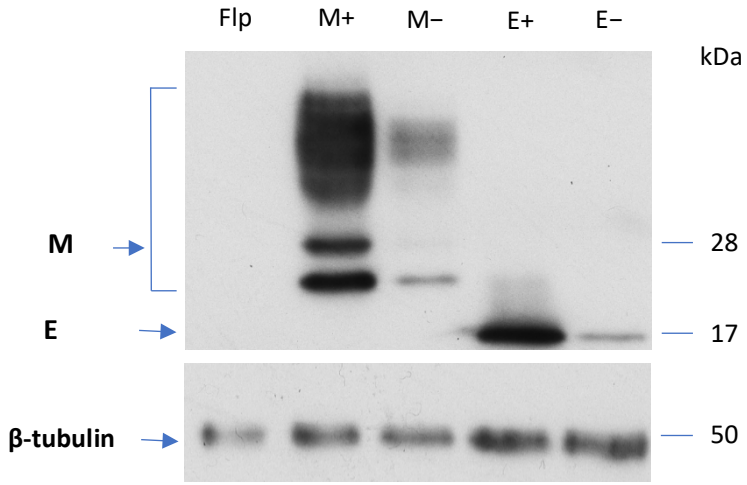


Figure 3-1 Validation of tetracycline-inducible overexpression of FLAG-tagged MERS-CoV E and M proteins in HEK293-E, -M, and -Flp cells.

HEK293-Flp (Flp, control cells), HEK293-E (E), and HEK293-M (M) cells were incubated either with (+) or without (−) Tet at 1 µg/ml for 48 h in culture flasks. Cells were harvested and lysed with 2× SDS-PAGE sample buffer. For each lysate, 10 µg of protein was separated using a 15% SDS-PAGE gel and transferred to a PVDF membrane for WB with anti-FLAG and anti-β-tubulin antibodies as indicated. β-tubulin was used as a loading control. The MERS-CoV E and M protein bands are indicated on the left of the image, and the positions of relevant molecular mass markers (in kDa) are indicated on the right. The molecular masses were as expected for FLAG-tagged MERS-CoV E and M proteins: (tagged with the FLAG epitope) 17 and 27.8 kDa, respectively. In addition to the expected M protein band with a molecular mass of ~27.8 kDa, other bands were detected, suggesting it might form oligomers or be modified by ubiquitination or phosphorylation.

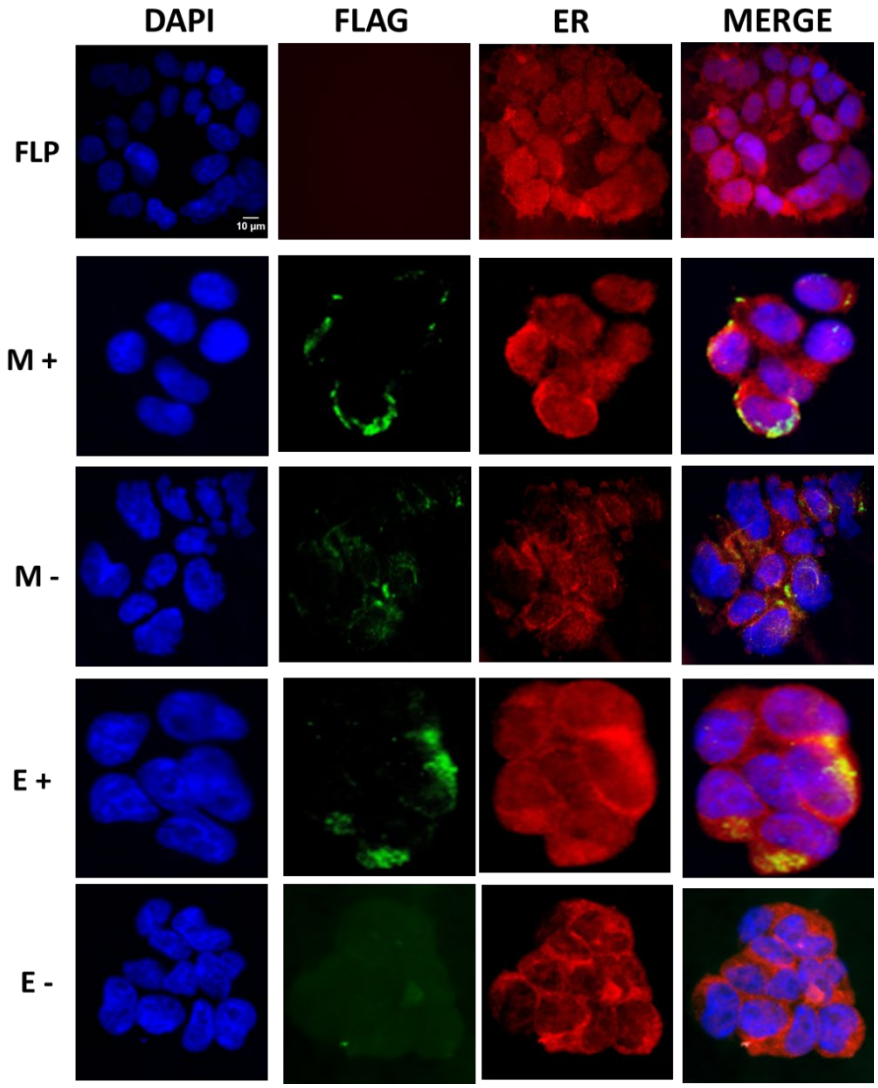


Figure 3-2 Validation of Tetracycline-inducible overexpression of FLAG-tagged MERS-CoV E and M proteins.

HEK293-Flp (Flp, control cells), HEK293-E (E), and HEK293-M (M) cells were incubated with (+) or without (−) Tet at 1 µg/ml for 48 h. Cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100 before immunostaining with anti-FLAG and anti-Erp72 (ER marker) antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-Erp72 (red) primary antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were imaged using a widefield imaging system (Leica DM 16000 inverted epifluorescence microscope) in the Wolfson Bioimaging Facility in the Life Sciences Faculty before being treated and analysed with Fiji software.

3.2.2 Validation of MERS-CoV E and M proteins in immunoprecipitates from HEK293-E and HEK293-M cell lysates

Once MERS-CoV E and M protein expression was confirmed in HEK293-E and HEK293-M, respectively, it was necessary to show they could be immunoprecipitated from cell lysates to validate the high-throughput co-IP analysis previously conducted in the laboratory. Flasks (T225) containing HEK293-Flp, -M and, -E cells (3×10^7) were grown to 90% confluent and induced with Tet for 48 h, harvested, and lysed with an immunoprecipitation lysis buffer. Then, immunocomplexes were captured with anti-FLAG magnetic beads, washed, and eluted by heating the beads at 95°C for 5 min (Figure 3-3). This process was performed in triplicate. To ensure that similar protein lysate amounts were used in all analyses, the non-bound protein lysate samples from the three different replicates were analysed by WB using an anti-GAPDH antibody (Figure 3-4). The results showed similar amounts of GAPDH in all cell lysates. Then input, non-bound, and pull-down samples were analysed by WB using anti-FLAG antibodies (Figure 3-5, Figure 3-6, and Figure 3-7). The MERS-CoV E and M proteins were detected in the input and pull-down samples but not in the non-bound samples. As expected, the molecular masses of the FLAG-tagged MERS-CoV E and M proteins (tagged with the FLAG epitope) were 17 and 27.8 kDa, respectively. As in the previous WB analysis (Figure 3-1), several additional bands were observed in the input and pull-down samples containing the M protein, providing further evidence that the FLAG-tagged M protein might form oligomers or be post-translationally modified.

In conclusion, these results showed that the FLAG-tagged MERS-CoV E and M proteins were expressed in the HEK293-E and -M cells, respectively, and could be successfully immunoprecipitated from their lysates. Therefore, immunoprecipitates could be used to validate the high-throughput co-IP/bioinformatics analysis conducted previously in the laboratory.

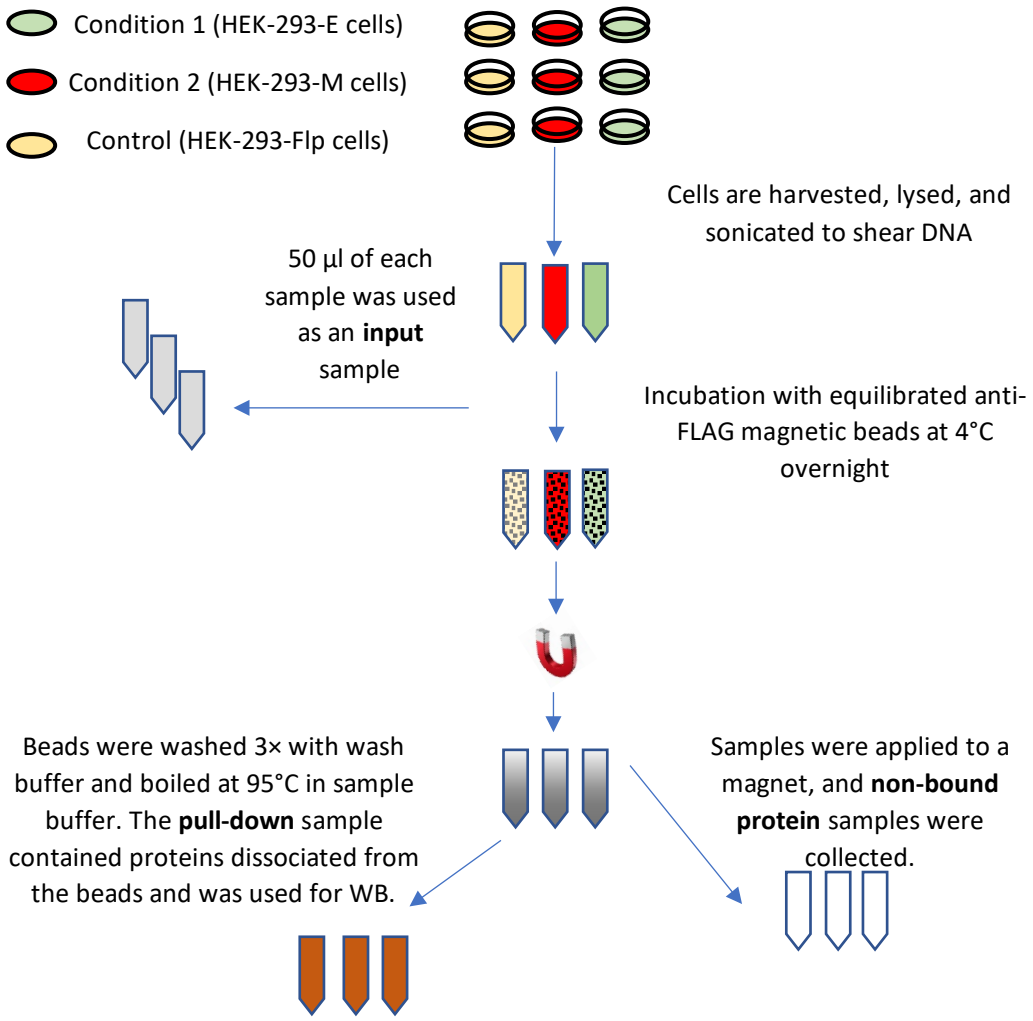


Figure 3-3 Schematic representation of the immunoprecipitation procedure.

HEK293-Flp, -M, and -E cells were treated with 1 µg/ml Tet for 48 h. Cells were harvested, washed with PBS, and lysed with immunoprecipitation lysis buffer. Next, each sample was centrifuged at 17,000 *g* and 4°C for 20 min, and an aliquot of lysis supernatant was transferred to a new tube (input sample). Then, the supernatant containing solubilised MERS-CoV E and M proteins was incubated with anti-FLAG conjugated beads at 4°C overnight. The beads were captured magnetically, the supernatant was removed (non-bound sample), and the beads were washed three times with wash buffer before being heated at 95°C in sample buffer to elute the proteins in the (pull-down) sample that was used for WB analysis.

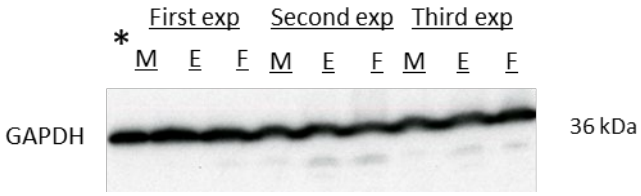


Figure 3-4 Validation of GAPDH expression in non-bound fractions from HEK293-Flp, -E, and -M cell lysates from three replicates.

Non-bound samples were prepared from HEK293-M, -E, and - Flp cell lysates following the protocol in Figure 3-3. Next, 10% of each sample was size separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for WB with an anti-GAPDH antibody. As expected, the molecular mass of the GAPDH band was 36 kDa. *Key: M, HEK293-M; E, HEK293-E; F, HEK293-Flp.

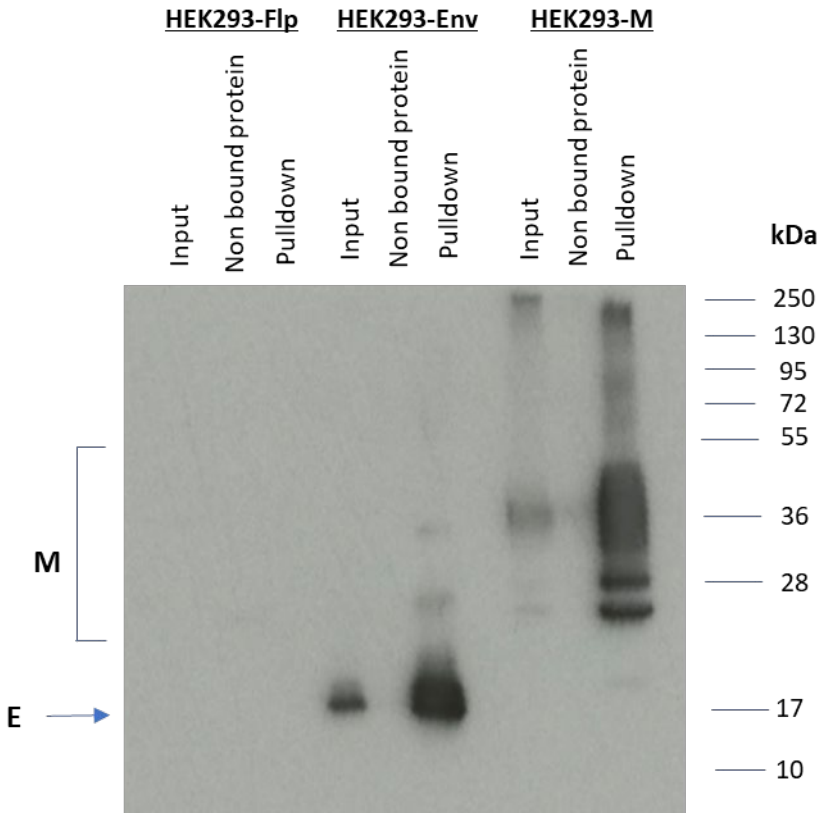


Figure 3-5 Validation of the MERS-CoV E and M proteins in immunoprecipitates from HEK293-E and HEK293-M cell lysates (first replicant).

Lysates from HEK293-Flp, -E, and -M cells incubated with Tet at 1 µg/ml for 48 h were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were separated according to size and transferred to a PVDF membrane for WB with an anti-FLAG antibody. The positions of relevant molecular mass markers (in kDa) are shown on the right of the image. Arrows on the left of the image indicate bands corresponding to the expected molecular masses of the MERS-CoV E and M proteins.

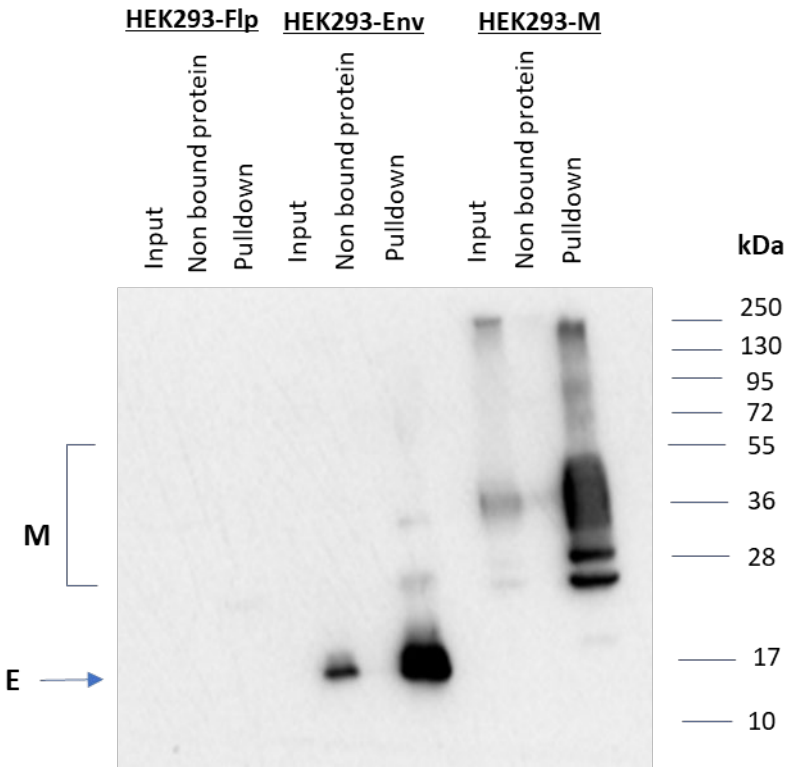


Figure 3-6 Validation of the MERS-CoV E and M proteins in immunoprecipitates from HEK293-E and HEK293-M cell lysates (second replicant).

Lysates from HEK293-Flp, -E, and -M cells incubated with Tet at 1 µg/ml for 48 h were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were separated according to size and transferred to a PVDF membrane for WB with an anti-FLAG antibody**.** The positions of relevant molecular mass markers (in kDa) are shown on the right of the image. Arrows on the left of the image indicate bands corresponding to the expected molecular masses of the MERS-CoV E and M proteins.

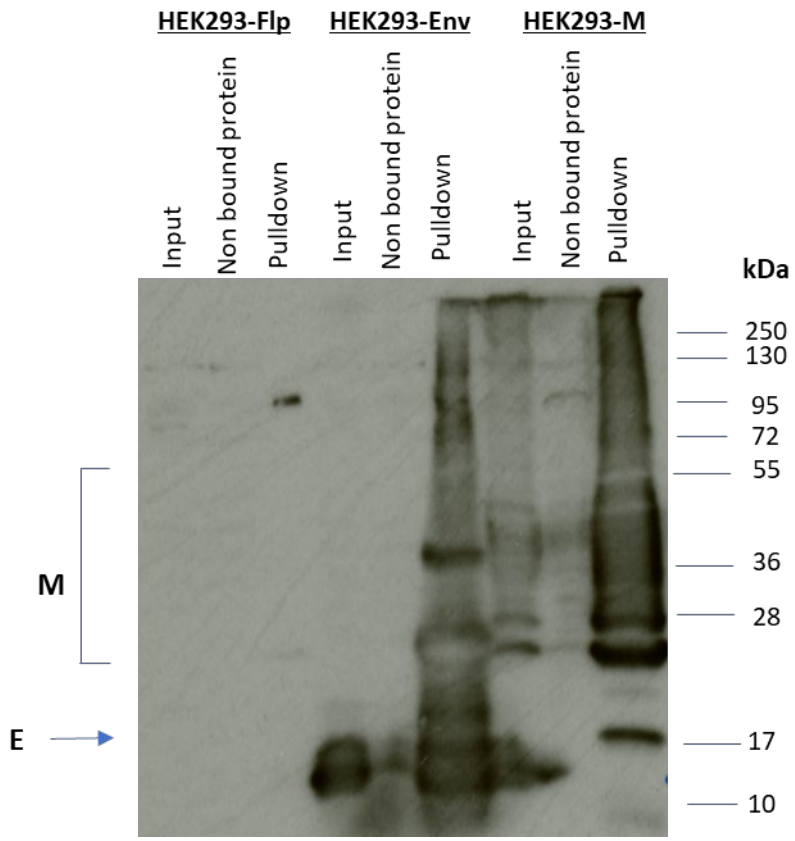


Figure 3-7 Validation of the MERS-CoV E and M proteins in immunoprecipitates from HEK293-E and HEK293-M cell lysates (third replicant).

Lysates from HEK293-Flp, -E, and -M cells incubated with Tet at 1 µg/ml for 48 h were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were separated according to size and transferred to a PVDF membrane for WB with an anti-FLAG antibody**.** The positions of relevant molecular mass markers (in kDa) are shown on the right of the image. Arrows on the left of the image indicate bands corresponding to the expected molecular masses of the MERS-CoV E and M proteins.

3.2.3 Quantitative analysis of co-IP analysis of pulldown samples of MERS-CoV E and M protein

The HEK293-E, -M, and -Flp cells analysed above had previously been used for high-throughput co-IP analysis to identify cellular interaction partners for MERS-CoV E and M proteins (experimental work performed by Mr Lee with data processing by Dr Phil Lewis; University of Bristol Proteomics Facility). To validate the bioinformatic analysis using complementary techniques, one aim of this investigation, the bioinformatic data set was first analysed (Appendix C) to identify cellular proteins interacting with the MERS-CoV E and M proteins with high confidence. As the first selection step, we determined the numbers of human proteins interacting with MERS-CoV E and M proteins that had >1.5 fold higher amounts in immunoprecipitates from HEK293-E/-M cells than from HEK293-Flp cells in ≥2 replicate experiments and a *p*-value of <0.05. One hundred and 136 out of 2126 and 494 proteins met these criteria for the MERS-CoV E and M interactomic data sets, respectively. In contrast, only 61 and 112 proteins showed a >1.5 fold increase in co-IPs from lysates of HEK293-E/-M cells compared to HEK293-Flp cells and had a *p*-value of <0.05 in ≥3 replicate. These proteins were selected for further analysis. Table 3-1 summarises these results.

Table 3-1 Number of proteins identified from proteomic analyses of three pull-down samples from MERS-CoV E and M proteins

MERS-CoV protein:	E	M
Total proteins detected	2126	494
Significant $p < 0.05$ and >1.5 fold change compared to HEK-Flp cells in ≥2 experiments	100	136
Significant $p < 0.05$ and >1.5 fold change compared to HEK-Flp cells in ≥3 experiments	61	112

3.2.4 Downstream bioinformatic analysis of significantly enriched MERS-CoV E and M protein interaction partners

The filtered protein data set from the bioinformatic analysis was subjected to downstream analyses using the DAVID database for GO-based enrichment and the STRING database for protein interaction networks. The DAVID analysis of proteins with amounts >1.5-fold higher in co-IPs from HEK293-E than HEK293-Flp cells in ≥3 experiments showed significant enrichment in proteins clustering with functional groups associated with the terms ER, membrane, mitochondrion, lipid-binding, and nucleocytoplasmic transport (Table **3-2**A). The STRING analysis of the same proteins showed significant enrichment of proteins associated with the biological process term response to proton transmembrane transport (Figure 3-8).

The DAVID analysis of proteins with amounts >1.5 fold higher in co-IPs from HEK293-M than HEK293-Flp cells in ≥3 experiments showed significant enrichment of proteins clustering with functional groups associated with the terms mitochondrion, adenosine triphosphate (ATP) synthesis, ER, oxidative phosphorylation, and protein transport (Table **3-2**B). The STRING analysis of the same proteins showed significant enrichment of proteins associated with the biological process term response to mitochondrial transmembrane transport (Figure 3-9).

Table 3-2A The DAVID analysis of cellular proteins with amounts >1.5 fold higher in co-IPs from HEK293-E than HEK293-Flp cells ($p < 0.05$).

Term	EASE score	Category	Count	%	*p*-value
GO:0005783~endoplasmic reticulum	10.66	GOTERM_CC_DIRECT	22	34.4	8.14E-12
KW-0472~membrane	8.47	UP_KW_CELLULAR_COMPONENT	52	81.3	2.59E-09
GO:0005739~mitochondrion	5.14	GOTERM_CC_DIRECT	16	25.0	1.85E-05
KW-0446~lipid-binding	3.14	UP_KW_LIGAND	3	4.7	0.099373
hsa03013~nucleocytoplasmic transport	2.93	KEGG_PATHWAY	7	10.9	1.25E-05

There were five annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table along the Gene ontology (GO) Category; biological process (GOBP), cellular component (GOCC), or molecular function (GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term. EASE: Fisher Exact Statistics in DAVID system.

Table 3-3B The DAVID analysis of cellular proteins with amounts >1.5 fold higher in co-IPs from HEK293-M than HEK293-Flp cells ($p < 0.05$).

Term	EASE Score	Category	Count	%	*p*-value
TRANSIT ~ mitochondrion	7.3	UP_SEQ_FEATURE	16	15.0	7.08E-08
KW-0066~ATP synthesis	7.2	UP_KW_BIOLOGICAL_PROCESS	12	11.2	1.23E-10
GO:0005783~endoplasmic reticulum	6.5	GOTERM_CC_DIRECT	24	22.4	1.29E-08
hsa00190 ~ oxidative phosphorylation	6.0	KEGG_PATHWAY	11	10.3	4.41E-07
KW-0653~protein transport	3.9	UP_KW_BIOLOGICAL_PROCESS	14	13.1	1.59E-04

There were five annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table along the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.

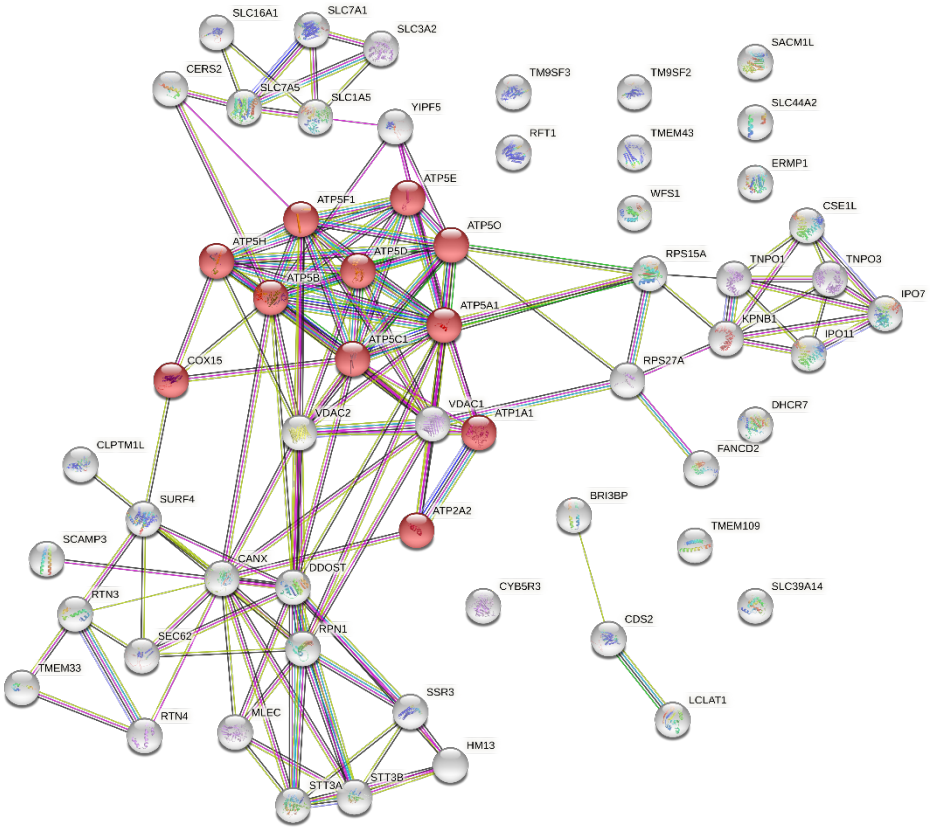


Figure 3-8 The STRING analysis of proteins with amounts >1.5 fold higher in co-IPs from HEK293-E than HEK293-Flp cells ($p < 0.05$).

The STRING database was searched to analyse the interaction network of host proteins whose amount was >1.5 fold higher in co-IPs from HEK293-E than HEK293-Flp cells ($p < 0.05$). Nodes representing the proteins most significantly enriched in the GO term proton transmembrane transport are shaded in red ($p = 2.06 \times 10^{-9}$).

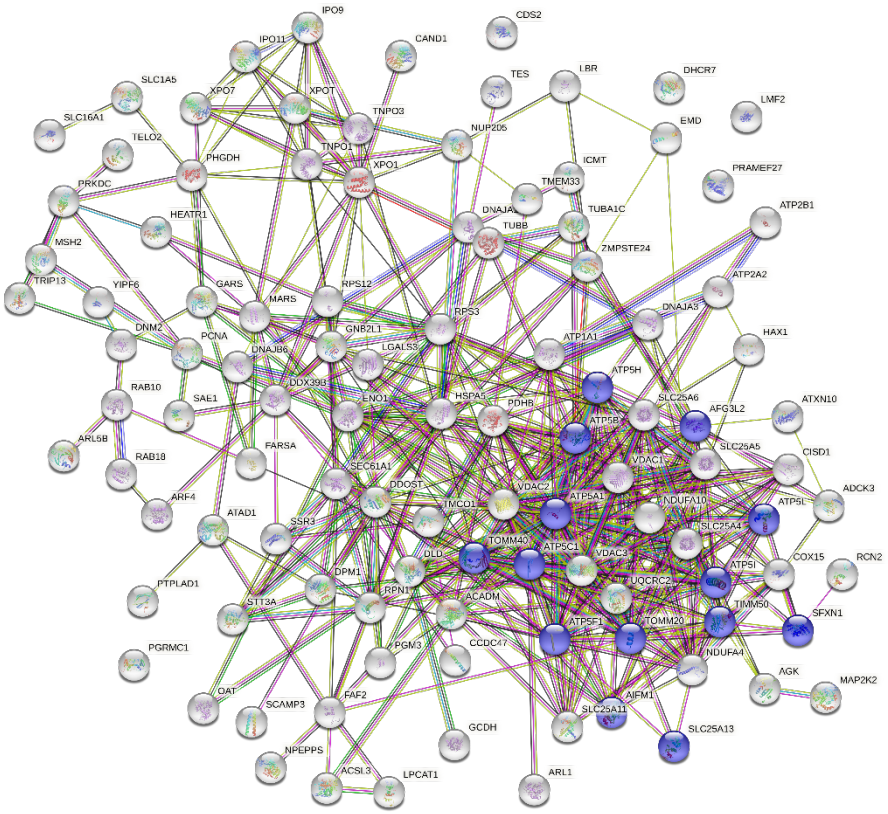


Figure 3-9 The STRING analysis of proteins with amounts >1.5 fold higher in co-IPs from HEK293-M than HEK293-Flp cells ($p < 0.05$).

The STRING database was searched to analyse the interaction network of host pertains with amounts >1.5-fold higher in co-IPs from HEK293-M than HEK293-Flp cells ($p < 0.05$). Nodes representing the proteins most significantly enriched in the GO term mitochondrial transmembrane transport are shaded in blue ($p = 5.15 \times 10^{-12}$)

3.2.5 Target protein selection for downstream validation

The top 30 proteins with >1.5-fold higher amounts in co-IPs from lysates of HEK293-E/-M than HEK293-Flp cells in ≥3 experiments and a $p < 0.05$ were manually filtered according to their cellular functions, localisation, high peptide coverage, high Uniprot annotation score for existence, antibody availability, and low likelihood of being non-specific interactors.

The selection criteria for cellular protein functions included mitochondrial, nuclear, membrane, glycosylation, and lipid synthesis. The criteria for cellular protein localisation included ER, Golgi, mitochondria, and nucleus. The MERS-CoV E and M proteins have been shown to localise to these cellular organelles (115, 136).

The MERS-CoV E protein was localised at sites of intracellular secretion and trafficking, including the ER, Golgi, and ERGIC, where it participates in coronavirus assembly and budding (115). Several coronavirus M proteins have been found to accumulate in the Golgi complex of mammalian host cells (136). Specifically, the MERS-CoV M protein colocalised with proteins in the Golgi apparatus and ERGIC, and there were additional discrete puncta within the cytoplasm (87). While the MERS-CoV E and M proteins are believed to localise to the ER, Golgi, and ERGIC, some mitochondrial proteins were identified as interaction partners. For example, the MERS-CoV E and M proteins interacted with the mitochondrial voltage-dependent anion-selective channel proteins 1 (VDAC1) and 2 (VDAC2). In addition, three nuclear proteins were present in the MERS-CoV E protein interaction list: transmembrane protein 43 (TMEM43), importin 11 (IPO11), and reticulon 4 (RTN4). Moreover, the nuclear protein IPO11 interacted with MERS-CoV E and M proteins. The protein-protein interactions of MERS-CoV E and M proteins with mitochondrial and nuclear proteins could occur naturally inside HEK293-E and/or -M cells or be artefacts following cell lysis during the co-IP analysis. Therefore, the Contaminant Repository for Affinity Purification (CRAPome) database (181) was used to ensure that the selected proteins were unlikely to be non-specific interactors.

Finally, six and five proteins were selected from the top 30 protein interactors for the MERS-CoV E and M proteins, respectively. The MERS-CoV E viral protein interactors were: transmembrane protein 43 (TMEM43), transmembrane 9 superfamily member 2 (TM9SF2), Yip1 domain family member 5 (YIPF5), ceramide synthase 2 (CERS2), choline transporter-like protein 2 (SLC44A2), and ER-Golgi intermediate compartment protein 1 (ERGIC1). The MERS-CoV M protein interactors were: IPO11, VDAC1, Ras-related

protein Rab-10 (RAB10), lysophosphatidylcholine acyltransferase 1 (LPCAT1), and secretory carrier-associated membrane protein 3 (SCAMP3). Only three protein interactors (SCAMP3, IPO11, and VDAC1) were shown in both MERS-CoV E and M proteins data set. Table **3-4** and Table **3-5** summarise key properties associated with the selected proteins, including their fold changes between HEK293-E/-M and HEK293-Flp cell lysates in ≥3 experiments. While only three interactors were shared between the MERS-CoV E and M proteins, the 11 selected protein interactors were tested in the next section to confirm whether they interact with MERS-CoV E or M or both.

Table 3-4 Description of human proteins selected for biological validation among MERS-CoV E protein interactors.

Gene ID	Uniprot no.	Peptides	Fold change*	Fold change**	Localisation	Function (132)	Number of experiments (found/total)***
TM9SF2	Q9HD45	9	3.5	---	Endosome	• It might act as a channel or transporter for small molecules in intracellular compartments.	7/411
TMEM43	Q9BTV4	23	3.8	---	ER and nuclear inner membrane	• It helps to maintain the nuclear envelope structure by organising protein complexes at the inner nuclear membrane.	4/411
YIPF5	Q969M3	2	3.4	---	ER membrane and cis-Golgi network membrane	• It plays a role in transport between the ER and Golgi.	7/411
CERS2	Q96G23	3	3.1	---	ER membrane	• It is an integral ER membrane protein.	1/411
SLC44A2	Q8IWA5	3	2.9	---	Extracelular exosmose, lysosomal membrane, and plasma membrane	• It enables the transfer of choline from one side of a membrane to the other. Choline is an amino alcohol that occurs widely in living organisms as a constituent of certain phospholipid types and the neurotransmitter acetylcholine.	Unavailable
ERGIC1	Q969X5	2	2.4	---	Golgi apparatus and ER	• Transportation between the ER and Golgi.	58/411

*Fold increase in co-IPs from lysates of HEK293-E cells compared to HEK293-Flp cells in ≥3 experiments.
**Fold increase in co-IPs from lysates of HEK293-M cells compared to HEK293-Flp cells in ≥3 experiments.
***The CRAPome database (181) was used to determine the number of experiments in which the protein was identified (found) /(total) experiments.

Table 3-5 Description of human proteins selected for biological validation among MERS-CoV M protein interactors.

Gene ID	Uniprot no.	Peptides	Fold change*	Fold change**	Localisation	Function (132)	Number of experiments (found/total)***
LPCAT1	Q8NF37	6	3.5	---	ER membrane and Golgi apparatus membrane	• It is involved in the pathway phospholipid metabolism and might synthesise phosphatidylcholine in pulmonary surfactant, playing a pivotal role in respiratory physiology.	7/411
SCAMP3	O14828	5	3.2	2.1	Golgi apparatus	• It functions in post-Golgi recycling pathways.	23/411
IPO11	Q9UI26	10	3.1	1.2	Nucleus and cytoplasm	• It functions in nuclear protein import as nuclear transport receptor.	7/411
VDAC1	P21796	11	2.4	2.8	Cell membrane and mitochondria	• It forms a channel through the mitochondrial outer membrane and the plasma membrane. The channel at the outer mitochondrial membrane allows diffusion of small hydrophilic molecules; at the plasma membrane, it is involved in cell volume regulation and apoptosis.	58/411
RAB10	P61026	3	2.4	---	ER and Golgi apparatus	• The small Rab GTPases are key regulators of intracellular membrane trafficking. • Multiple biological and molecular functions.	56/411

*Fold increase in co-IPs from lysates of HEK293-M cells compared to HEK293-Flp cells in ≥3 experiments.
**Fold increase in co-IPs from lysates of HEK293-E cells compared to HEK293-Flp cells in ≥3 experiments.
***The CRAPome database (181) was used to determine the number of experiments in which the protein was identified (found) /(total) experiments.

3.2.6 Validation of MERS-CoV E and M protein cellular interaction partners identified by high-throughput co-IP and bioinformatic analyses

Next, the candidate MERS-CoV E and M protein interaction partners identified above were validated by co-IP/WB analysis and IF colocalisation analysis to confirm the interaction.

For the WB analysis, HEK293 cells expressing the MERS-CoV E (HEK293-E) or M (HEK293-M) proteins or the parental cell line (HEK293-Flp) were used for co-IP analysis using an antibody against the FLAG epitope tag present on the E and M proteins. Three samples were prepared as described in Section 3.2.2: input, non-bound, and pull-down. They were analysed by WB, with 10% of each sample loaded into a lane on a 15% SDS-PAGE gel, size separated, and transferred to a PVDF membrane for detection with a specific antibody recognising each target protein.

For the IF colocalisation analysis, HEK293 cells expressing the MERS-CoV E (HEK293-E) or M (HEK293-M) proteins or the parental cell line (HEK293-Flp) were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24 well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with an anti-FLAG antibody and a specific antibody recognising each target protein. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG and target protein antibodies, respectively. Nuclear DNA was stained with DAPI. The cells were fixed and stained. The immunostained cells were imaged using a confocal imaging system, followed by analysis with Fiji software to determine protein colocalization by calculating the PCC. The PCC examines the relationship between the intensities of pixels in two different channels within the same image. At least 50 images were analysed for each calculation to obtain a mean PCC. A PCC of 1 indicates a perfect positive correlation, a PCC of 0 indicates no correlation, and a PCC of −1 indicates a perfect negative correlation. The findings of these analyses are shown in

Figure 3-10,

Figure 3-11, Figure 3-12,

Figure 3-13,

Figure 3-14, Figure 3-15, Figure 3-16,

Figure 3-17, Figure 3-18, Figure 3-19, and Figure 3-20; summarised in Table 3-6 for each protein; and discussed in more detail below.

3.2.6.1 Candidate proteins interacting with the MERS-CoV E protein

TM9SF2

The original proteomic analysis showed a 3.5-fold increase in the abundance of TM9SF2 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. However, the confirmatory analysis showed no interaction between the TM9SF2 protein and the MERS-CoV E or M proteins (

Figure 3-10). The co-IP and subsequent WB analyses detected similar amounts of TM9SF2 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV E or M proteins and Flp cells. However, TM9SF2 was not detected in any pull-down samples. However, the amount of TM9SF2 in the pull-down sample could be below the detection limit. The immunofluorescence colocalisation analysis results confirmed the WB results. Neither the FLAG-tagged E nor M proteins (green) colocalised with TM9SF2 (red). The PCC values were 0.04 and 0.03 between TM9SF2 and the MERS-CoV E and M proteins, respectively.

TMEM43

The original proteomic analysis showed a 3.8-fold increase in the abundance of TMEM43 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. Moreover, in the confirmatory analysis, TMEM43 appeared to interact more strongly with the MERS-CoV E protein than the MERS-CoV M protein (

Figure 3-11). This result was expected since TMEM43 is a MERS-CoV E protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of TMEM43 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV E or M proteins and Flp cells. In addition, TMEM43 was detected in the pull-down samples from lysates of cells expressing MERS-CoV E and M but not the Flp cells. Compared to the original proteomic results that identified TMEM43 as a MERS-CoV E interactor, the amount of TMEM43 in the MERS-CoV E pull-down samples was greater than in the MERS-CoV M pull-down samples. The immunofluorescence colocalisation analysis results shown that the FLAG-tagged MERS-CoV E protein (green) showed a clear colocalisation (yellow) with TMEM43 (red), while there was no colocalisation of TMEM43 with the FLAG-tagged MERS-CoV M protein. The PCC values were 0.49 and 0.08 between TMEM43 and the MERS-CoV E and M proteins, respectively. While the co-IP/WB analysis provided evidence that TMEM43 interacted with MERS-CoV M protein,

the co-IP/IFA analysis did not support this. An interaction between TMEM43 and the M protein after cell lysis or an indirect interaction could explain these results.

CERS2

The original proteomic analysis showed a 3-1 fold increase in the abundance of CERS2 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. Moreover, in the confirmatory analysis, CERS2 appeared to interact more strongly with the MERS-CoV E protein than the MERS-CoV M protein (

Figure 3-12). This result was expected since CERS2 is a MERS-CoV E protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of CERS2 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV E or M proteins and Flp cells. CERS2 was detected in the pull-down samples from lysates of cells expressing MERS-CoV E and M proteins and Flp cells (faint band). Compared to the original proteomic results, the amount of CERS2 in the MERS-CoV E pull-down samples was greater than in the MERS-CoV M pull-down samples. The immunofluorescence colocalisation analysis results differed slightly from the WB results. The FLAG-tagged MERS-CoV E protein (green) showed a clear colocalisation (yellow) with CERS2 (red), while there was a stronger localisation of CERS2 with the FLAG-tagged MERS-CoV M protein. The PCC values were 0.36 and 0.45 between CERS2 and the MERS-CoV E and M proteins, respectively.

YIPF5

The original proteomic analysis showed a 3.4-fold increase in the abundance of YIPF5 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. However, in the confirmatory analysis, YIPF5 appeared to interact less strongly with the MERS-CoV E protein than with the MERS-CoV M protein (

Figure **3-13**). This result was unexpected since YIPF5 is a MERS-CoV E protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of YIPF5 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV E or M proteins and Flp cells. In addition, YIPF5 was detected in the pull-down samples from lysates of cells expressing MERS-CoV E and M proteins (faint band) and Flp cells (faint band). Compared to the original proteomic results, the amount of YIPF5 in the MERS-CoV E pull-down sample was lower than in the MERS-CoV M pull-down sample. Moreover, a second band was detected in the YIPF5 WB analysis, suggesting it might

form oligomers or be modified by ubiquitination or phosphorylation (182, 183). The immunofluorescence colocalisation analysis results differed slightly from the WB results. The FLAG-tagged MERS-CoV E protein (green) showed a clear colocalisation (yellow) with YIPF5 (red), while there was a weaker localisation of YIPF5 with the FLAG-tagged MERS-CoV M protein. The PCC values were 0.69 and 0.36 between YIPF5 and the MERS-CoV E and M proteins, respectively.

ERGIC1

The original proteomic analysis showed a 2.4-fold increase in the abundance of ERGIC1 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. However, in the confirmatory analysis, ERGIC1 appeared to interact more with the MERS-CoV M protein than the MERS-CoV E protein (Figure 3-14). This result was unexpected since ERGIC1 is a MERS-CoV E protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of ERGIC1 in the input and non-bound fractions from lysates of cells expressing either MERS-CoV E or M proteins and Flp cells. In addition, ERGIC1 was detected in both the pull-down samples from lysates of cells expressing MERS-CoV E and M proteins (faint band) but not the Flp cells. Compared to the original proteomic results, the amount of ERGIC1 in the MERS-CoV E pull-down sample was lower than in the MERS-CoV M pull-down sample. The FLAG-tagged MERS-CoV M protein (green) showed possible colocalisation (yellow) with ERGIC1 (red), while there was a weak colocalisation of ERGIC1 with the FLAG-tagged MERS-CoV E protein. The PCC values were 0.11 and 0.23 between ERGIC1 and the MERS-CoV E and M proteins, respectively. Despite that the ERGIC1 was detected in both the pull-down samples from lysates of cells expressing MERS-CoV E and M proteins showing a faint band, the ERGIC1 shown a strong band in the input and non-bound fractions from lysates of cells expressing either MERS-CoV E or M proteins and Flp cells. Moreover, the PCC values dose not shown a strong colocalisation between ERGIC1 and the MERS-CoV E and M proteins. These finding does not confirm the protein protein interaction and it needs to be confirmed by redoing the co-IP and subsequent WB analyses, which was done only once.

SLC44A2

The original proteomic analysis showed a 2.9-fold increase in the abundance of SLC44A2 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. However, in the confirmatory analysis, SLC44A2 appeared to interact with both MERS-CoV E and M proteins (Figure 3-15). This result was unexpected since SLC44A2 is only a

MERS-CoV E protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of SLC44A2 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV E or M proteins and Flp cells. In addition, SLC44A2 was detected in both pull-down samples from lysates of cells expressing MERS-CoV E or M proteins and Flp (faint band) cells. The SLC44A2 WB analysis had a dark background that might make it hard to read. The immunofluorescence colocalisation analysis results differed from the WB results. The FLAG-tagged MERS-CoV E protein (green) showed weaker colocalisation (yellow) with SLC44A2 (red), while there was a stronger localisation of SLC44A2 with the FLAG-tagged MERS-CoV M protein. The PCC values were 0.10 and 0.20 between SLC44A2 and the MERS-CoV E and M proteins, respectively. Despite that the SLC44A2 was detected in both the pull-down samples from lysates of cells expressing MERS-CoV E and M proteins, the SLC44A2 shown a faint band in the pull-down from lysates of cells expressing Flp cells. Moreover, the PCC values dose not shown a strong colocalisation between SLC44A2 and the MERS-CoV E proteins. These finding does not confirm the protein protein interaction and it needs to be confirmed by redoing the co-IP and subsequent WB analyses, which was done only once.

3.2.6.2 Candidate proteins interacting with the MERS-CoV M protein

LPCAT1

The original proteomic analysis showed a 3.5-fold increase in the abundance of LPCAT1 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. In the confirmatory analysis, LPCAT1 appeared to interact more strongly with the MERS-CoV M protein than the MERS-CoV E protein (Figure 3-19). This result was expected since LPCAT1 is a MERS-CoV M protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of LPCAT1 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV M or E proteins and Flp cells. In addition, LPCAT1 was detected in the pull-down samples from lysates of cells expressing M and E (faint band) proteins and Flp cells (faint band). Compared to the original proteomic results, the amount of LPCAT1 in the MERS-CoV M pull-down sample was greater than in the MERS-CoV E pull-down sample. The immunofluorescence colocalisation analysis results largely confirmed the WB results. The FLAG-tagged MERS-CoV M protein (green) showed a clear colocalisation (yellow) with LPCAT1 (red), while there was a similar or slightly stronger localisation of LPCAT1 with the FLAG-tagged MERS-CoV E protein. The PCC values were 0.33 and 0.48 between LPCAT1 and the MERS-CoV M and E proteins, respectively.

SCAMP3

The original proteomic analysis showed a 3-2 fold increase in the abundance of SCAMP3 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. In the confirmatory analysis, SCAMP3 appeared to interact more with the MERS-CoV M protein than the MERS-CoV E protein (Figure 3-20). This result was expected since SCAMP3 is a MERS-CoV M protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of SCAMP3 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV M or E proteins and Flp cells. In addition, SCAMP3 was detected in the pull-down samples from lysates of cells expressing MERS-CoV M and E proteins but not Flp cells. Compared to the original proteomic results, the amount of SCAMP3 in the MERS-CoV M pull-down sample was greater than in the MERS-CoV E pull-down sample. The immunofluorescence colocalisation analysis results largely confirmed the WB results. The FLAG-tagged MERS-CoV M and E proteins (green) showed nearly the same clear colocalisation (yellow) with SCAMP3 (red), while the WB analysis showed that SCAMP3 interacted more with the MERS-CoV M protein than the MERS-CoV E protein. The PCC values were 0.50 and 0.53 between SCAMP3 and the MERS-CoV M and E proteins, respectively.

IPO11

The original proteomic analysis showed a 3.0-fold increase in the abundance of IPO11 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. In the confirmatory analysis, IPO11 appeared to interact more with the MERS-CoV M protein than the MERS-CoV E protein (Figure 3-16). The co-IP and subsequent WB analyses detected relatively similar amounts of IPO11 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV M or E proteins and Flp cells. While IPO11 was detected in the pull-down samples from lysates of cells expressing the MERS-CoV M protein, those of cells expressing the MERS-CoV E protein and Flp cells showed a faint band. Compared to the original proteomic results, the amount of IPO11 in the MERS-CoV M pull-down sample was greater than in the MERS-CoV E pull-down sample. The immunofluorescence colocalisation analysis results confirmed the WB results. The FLAG-tagged MERS-CoV M protein (green) showed a clear colocalisation (yellow) with IPO11 (red), while there was a weaker localisation of IPO11 with the FLAG-tagged MERS-CoV E protein. The PCC values were 0.47 and 0.29 between IPO11 and the MERS-CoV M and E proteins, respectively.

VDAC1

The original proteomic analysis showed a 2.4-fold increase in the abundance of VDAC1 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. In the confirmatory analysis, VDAC1 appeared to interact more strongly with the MERS-CoV M protein than the MERS-CoV E protein (

Figure 3-17). This result was expected since VDAC1 is a MERS-CoV M protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of VDAC1 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV M or E proteins and Flp cells. In addition, VDAC1 was detected in the pull-down samples from lysates of cells expressing M protein (strong band) and E protein (faint band) but not Flp cells. Compared to the original proteomic results, the amount of VDAC1 in the MERS-CoV M pull-down sample was greater than in the MERS-CoV E pull-down sample. The FLAG-tagged MERS-CoV M protein (green) showed a clear colocalisation (yellow) with VDAC1 (red), while there was no colocalisation of VDAC1 with the FLAG-tagged MERS-CoV E protein. The PCC values were 0.44 and 0.01 between VDAC1 and the MERS-CoV M and E proteins, respectively.

While VDAC1 was detected in MERS-CoV-E cell lysates using WB, the IFA results did not support colocalisation between VDAC1 and MERS-CoV E. A protein-protein interaction occurring after cell lysis or an indirect interaction could explain these results. Moreover, VDAC1 could be important for this viral protein's function but did not colocalise with it.

RAB10

The original proteomic analysis showed a 2.4-fold increase in the abundance of RAB10 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. However, in the confirmatory analysis, RAB10 appeared to interact more strongly with the MERS-CoV M protein than the MERS-CoV E protein (Figure 3-18). This result was expected since RAB10 is a MERS-CoV M protein interactor. The co-IP and subsequent WB analyses detected relatively similar amounts of RAB10 in the input and non-bound fractions from lysates of cells expressing either the MERS-CoV M or E proteins and Flp cells. RAB10 was detected in the pull-down sample from lysates of cells expressing the

MERS-CoV M protein but not cells expressing the MERS-CoV E protein or Flp cells. Compared to the original proteomic results, the amount of RAB10 in the MERS-CoV M pull-down sample was greater than in the MERS-CoV E pull-down sample (there may be a faint band). The FLAG-tagged MERS-CoV M protein (green) showed a clear colocalisation (yellow) with RAB10 (red), while there was slightly weaker colocalisation of RAB10 with the FLAG-tagged MERS-CoV E protein. The PCC values were 0.29 and 0.28 between RAB10 and the MERS-CoV M and E proteins. respectively. While RAB10 was not detected in MERS-CoV E cell lysate by WB, the IFA results did show colocalisation between RAB10 and MERS-CoV E. A protein-protein interaction occurring after cell lysis or an indirect interaction could explain this result. Moreover, the IFA picture that represent the FLAG-tagged MERS-CoV M protein (green) colocalisation with RAB10 (red) was not the ideal picture to represent the colocalisation, and the IFA could be done again to reconfirm the colocalisation.

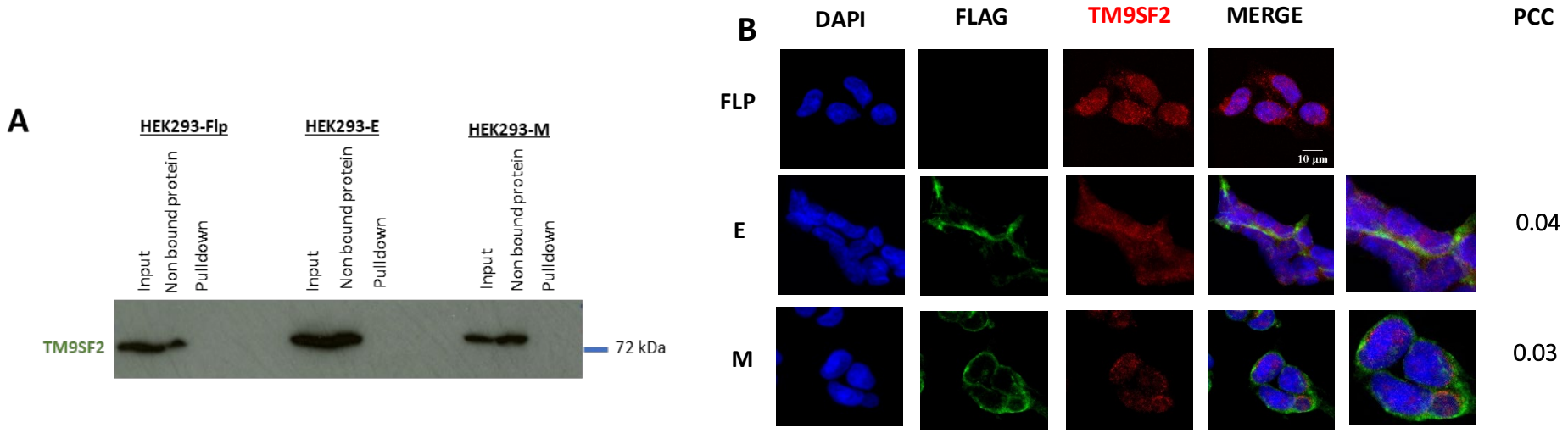


Figure 3-10 Validation of TM9SF2 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-pPAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-TM9SF2. The expected molecular mass for TM9SF2 (72 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-TM9SF2 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-TM9SF2 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

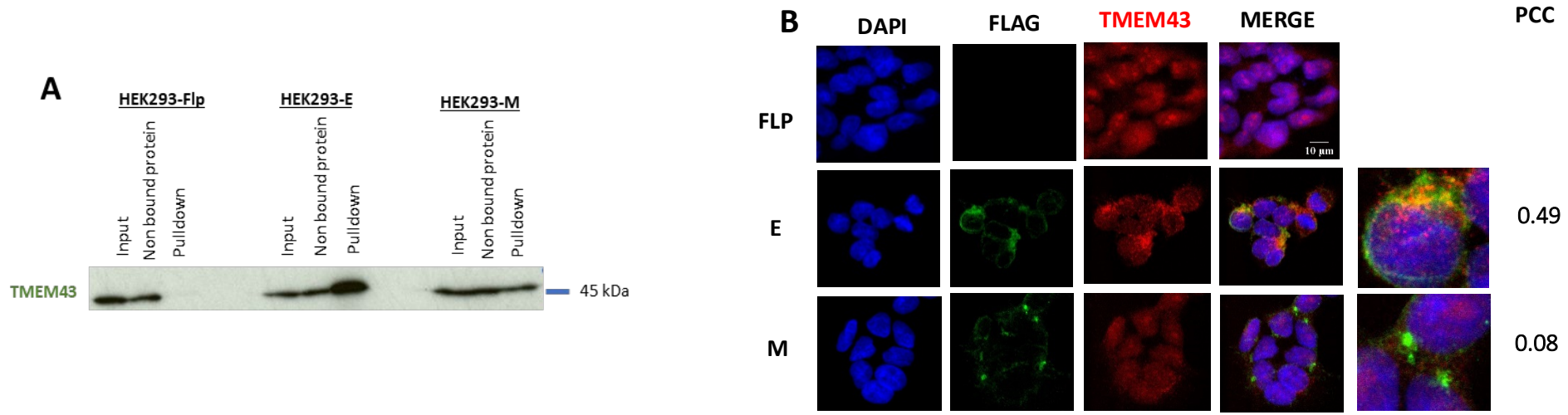


Figure 3-11 Validation of TMEM43 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-TMEM43. The expected molecular mass for TMEM43 (45 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-TMEM43 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-TMEM43 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

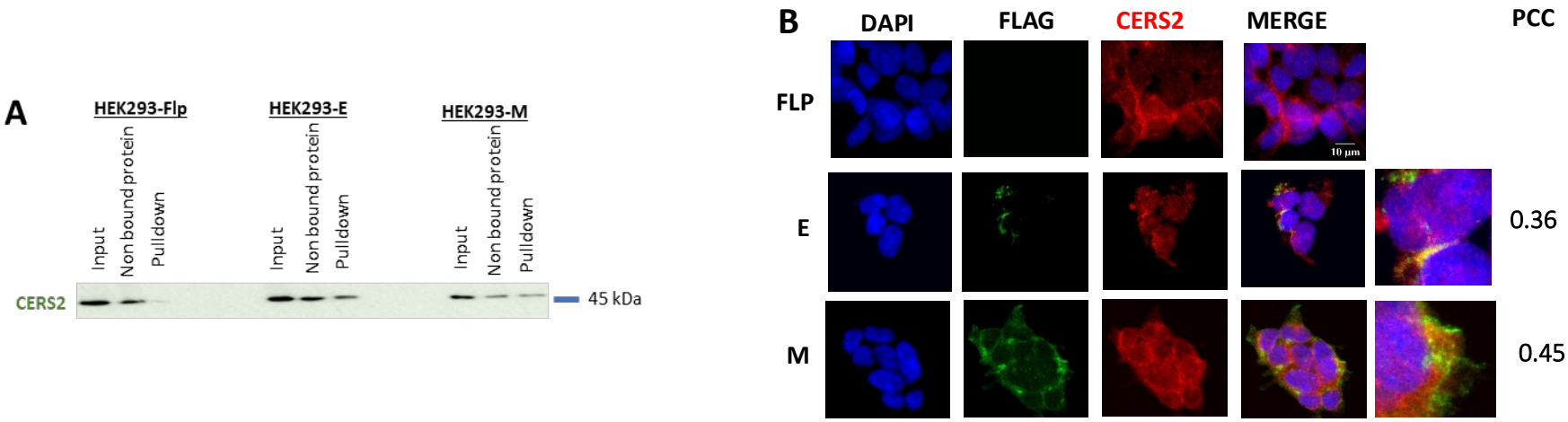


Figure 3-12 Validation of CERS2 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-CERS2. The expected molecular mass for CERS2 (45 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-CERS2 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-CERS2 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

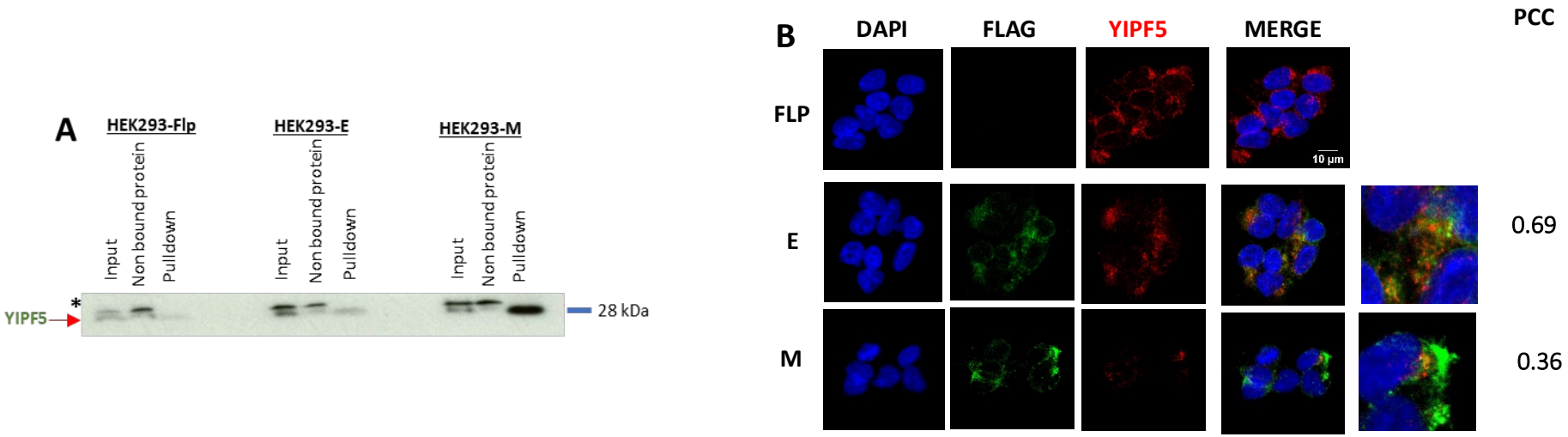


Figure 3-13 Validation of YIPF5 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-YIPF5. The expected molecular mass for YIPF5 (28 kDa) is shown; *, a nonspecific band (182). **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-YIPF5 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-YIPF5 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

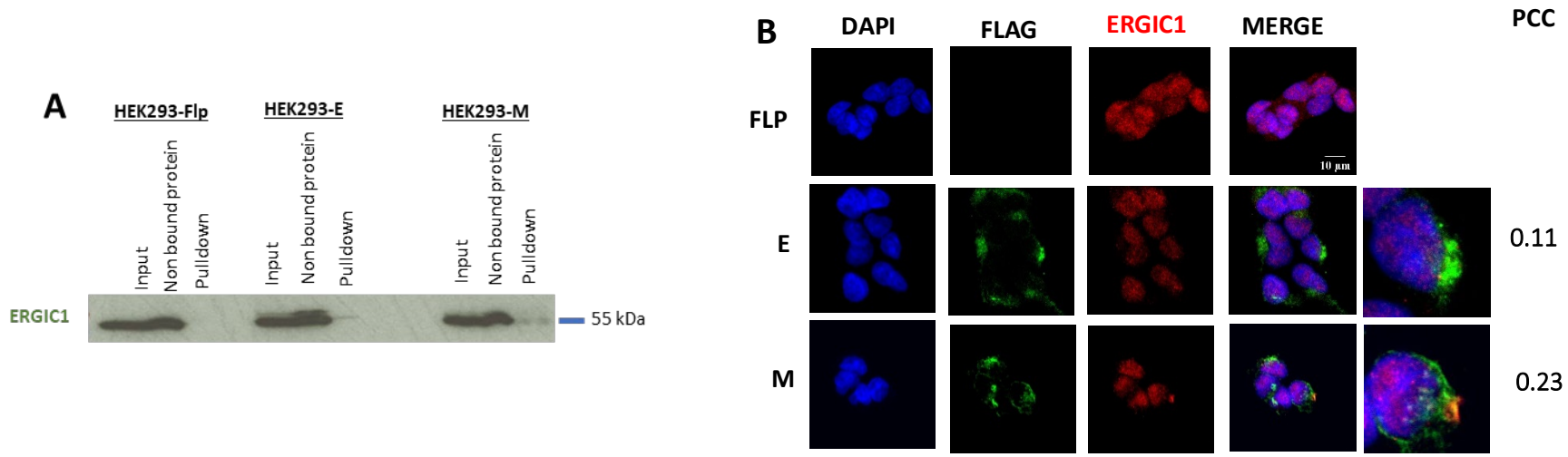


Figure 3-14 Validation of ERGIC1 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-ERGIC1. The expected molecular mass for ERGIC1 (55 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-ERGIC1 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-ERGIC1 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

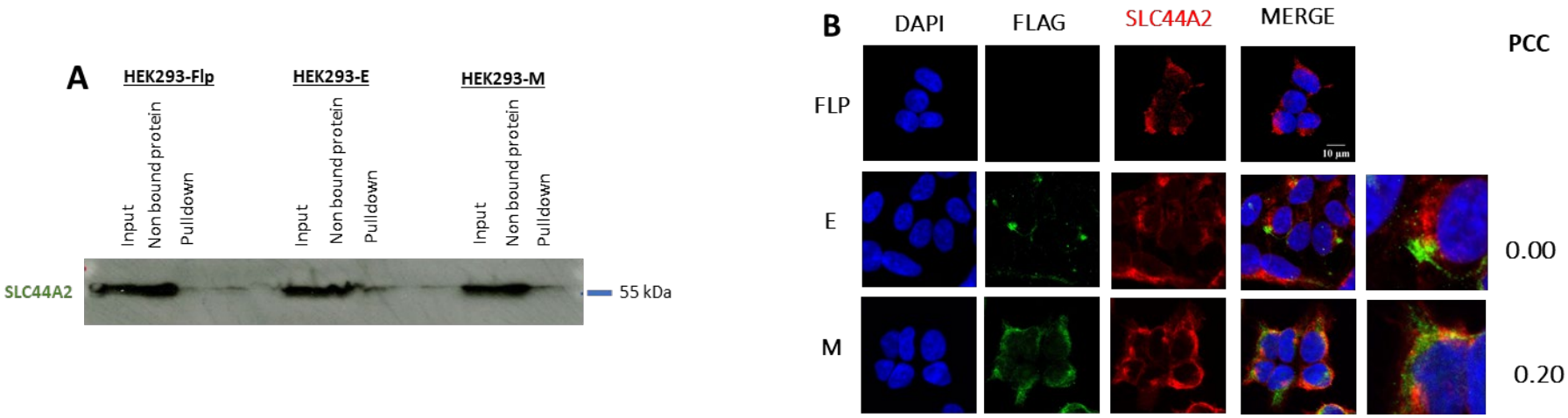


Figure 3-15 Validation of SLC44A2 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-SLC44A2. The expected molecular mass for SLC44A2 (55 kDa) is shown; the SLC44A2 band was shown as 50 kDa in Ref. (184). **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-SLC44A2 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-SLC44A2 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of *n* = 50 cells).

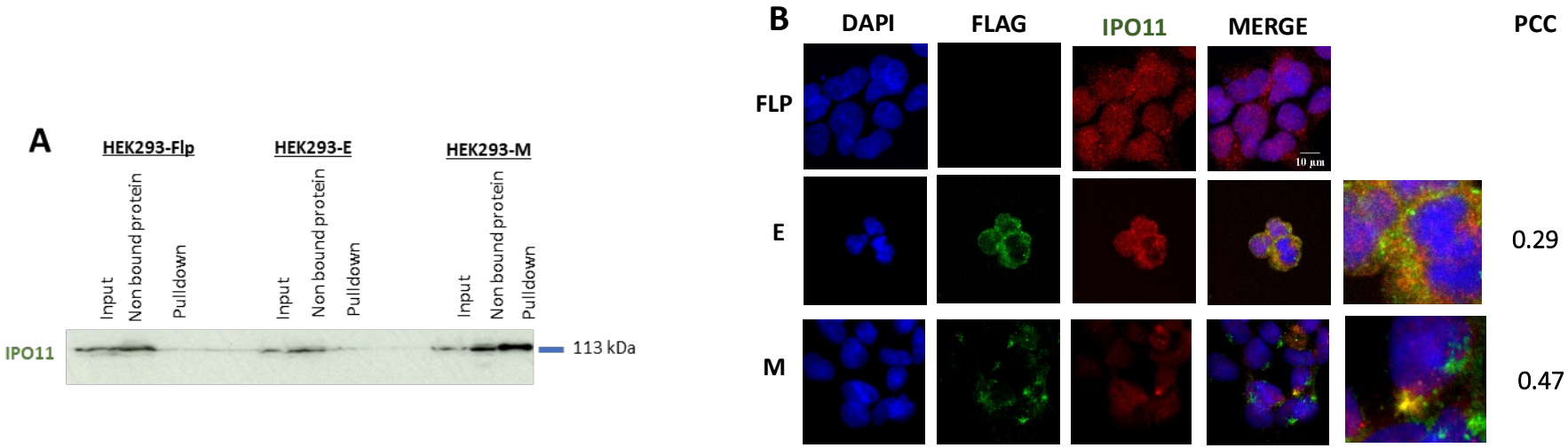


Figure 3-16 Validation of IPO11 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-IPO11. The expected molecular mass for IPO11 (113 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-IPO11 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-IPO11 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

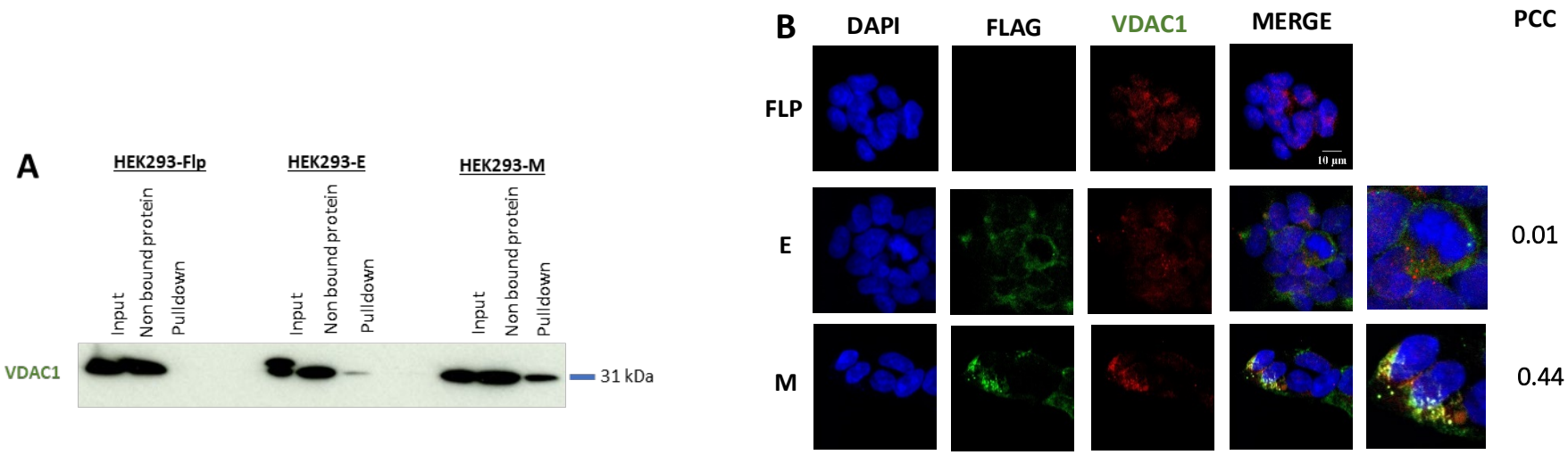


Figure 3-17 Validation of VDAC1 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-VDAC1. The expected molecular mass for VDAC1 (31 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-VDAC1 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-VDAC1 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

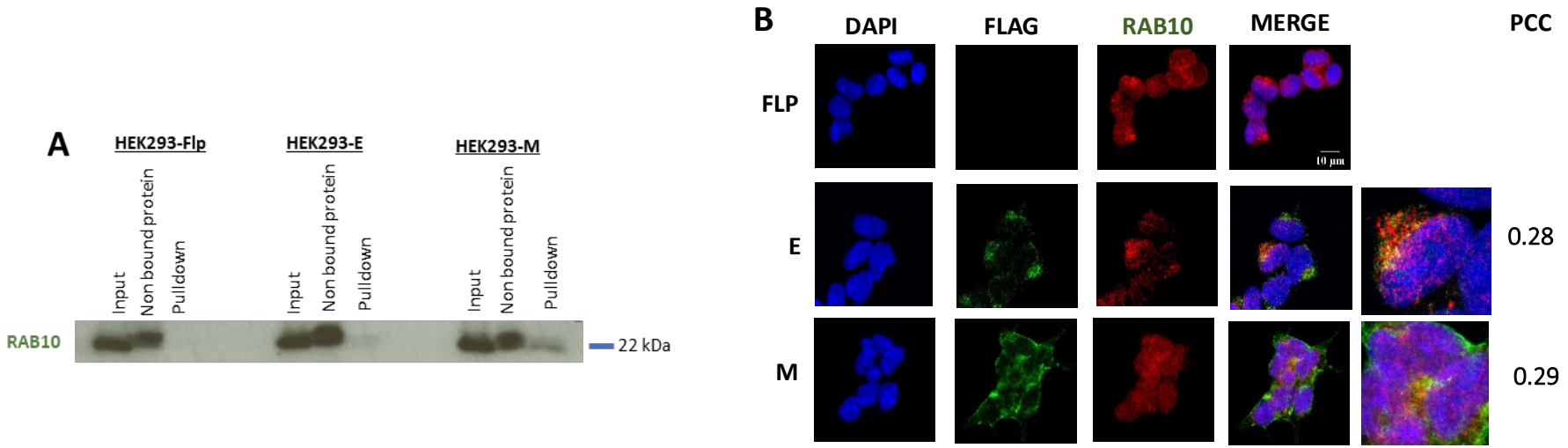


Figure 3-18 Validation of RAB10 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAG gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-RAB10. The expected molecular mass for RAB10 (22 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-RAB10 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-RAB10 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

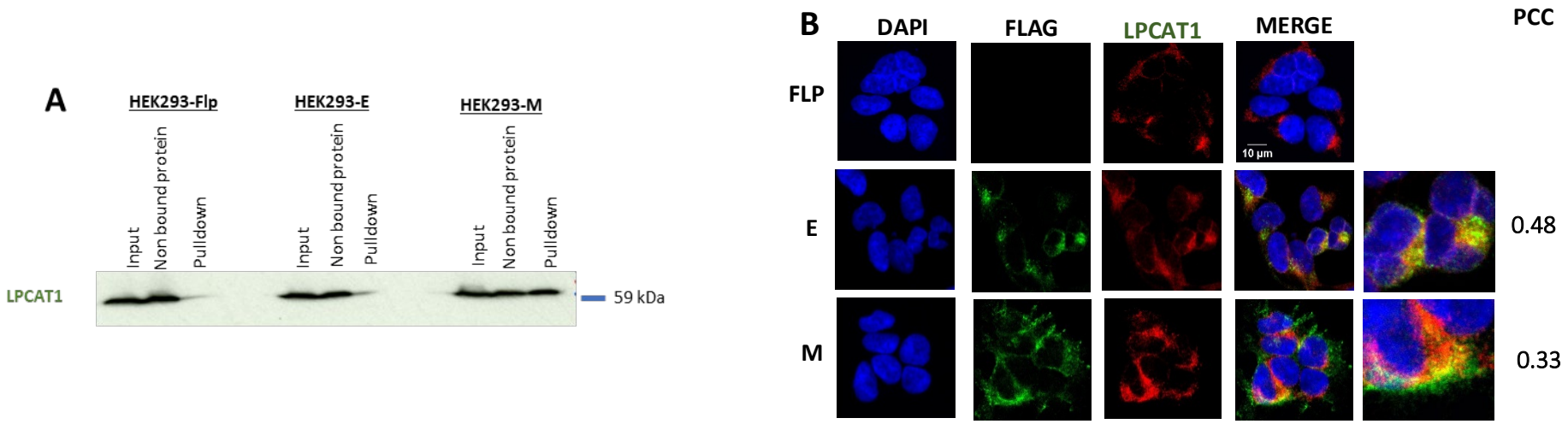


Figure 3-19 Validation of LPCAT1 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-LPCAT1. The expected molecular mass for LPCAT1 (59 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-LPCAT1 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-LPCAT1 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

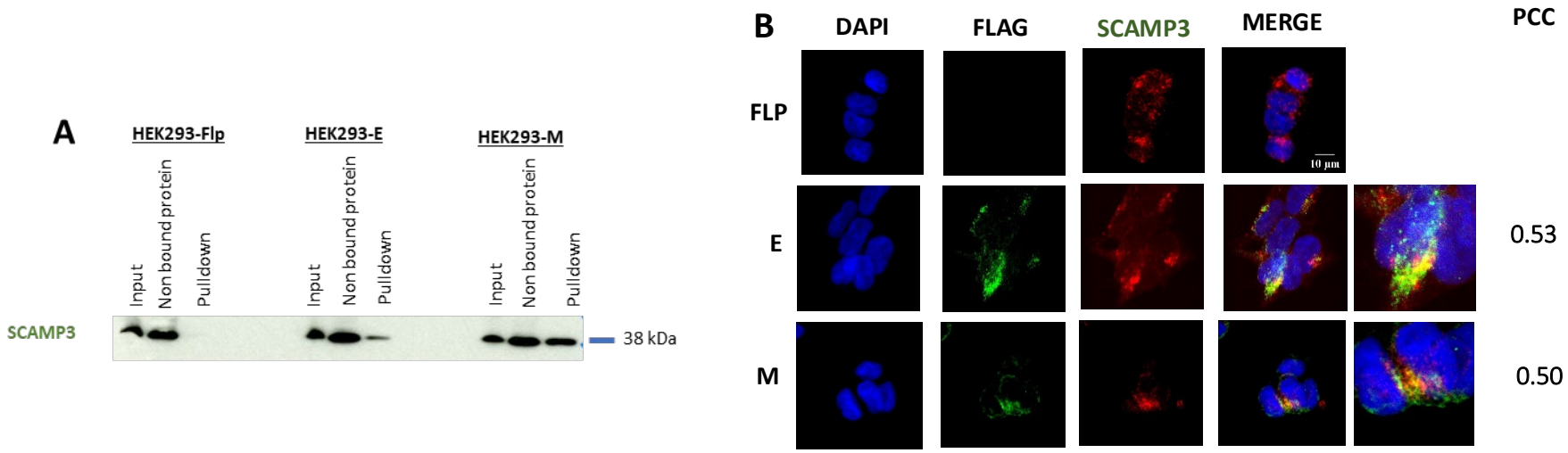


Figure 3-20 Validation of SCAMP3 as a potential interactor of the MERS-CoV E and M proteins.

(A) HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells were used for co-IP analysis using an anti-FLAG antibody. Three samples were analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for WB with anti-SCAMP3. The expected molecular mass for SCAMP3 (38 kDa) is shown. **(B)** IFA analysis of HEK293-E and HEK293-M cells expressing MERS-CoV E and M proteins, respectively, and HEK293-Flp cells. The cells were incubated with Tet at 1 µg/ml for 72 h on coverslips in a 24-well tray. Next, cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, cells were probed with anti-FLAG and anti-SCAMP3 primary antibodies. Alexa Fluor 488 and 568 conjugated secondary antibodies were used to detect the anti-FLAG (green) and anti-SCAMP3 (red) antibodies, respectively. Nuclear DNA was stained with DAPI (blue). The cells were fixed, stained, and imaged using a confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty before analysis with Fiji software (PCC values were calculated from the analysis of n = 50 cells).

Table 3-6 Summary of the WB and IFA validation of MERS-CoV E and M protein interaction partners.

Protein	E	E	M	M
	WB	IFA	WB	IFA
MERS-CoV E protein				
TM9SF2	–	–	–	–
TMEM43	++	+	+	-
CERS2	+	+	+	+
YIPF5	+	++	++	+
ERGIC1	+/-	+/-	+/-	+/-
SLC44A2	+/-	-	+/-	+
MERS-CoV M protein				
IPO11	+/-	+	++	+
VDAC1	+	–	++	+
RAB10	–	+	+	+
LPCAT1	+/-	+	+	+
SCAMP3	+	++	++	++
WB experiments done only once (n=1), and the imaging analysis experiments used 50 cells to calculate the PCC values.				

3.3 Discussion

The laboratory previously conducted a high-throughput co-IP analysis to identify cellular proteins interacting with the MERS-CoV E and M proteins. This experiment used HEK293-E and HEK293-M cells expressing FLAG-tagged MERS-CoV E and M proteins, followed by a high-throughput co-IP analysis. This study has validated the results of this previous analysis using complementary experimental approaches. Initially, MERS-CoV E and M protein expression was validated in these cells by co-IP/WB and IFA analyses to ensure they were suitable for further downstream validation of the bioinformatic data. Then, the interactomic data set was analysed to identify cellular proteins interacting with the MERS-CoV E and M proteins with high confidence. A small number of target proteins were selected for downstream validation. Finally, the interaction of the target proteins with the MERS-CoV E and M proteins was investigated by co-IP/WB and IFA analyses to support the bioinformatics analysis. The result of co-IP/WB and IFA analyses must be regarded as preliminary since the WB Was done only once.

Coronaviruses depend heavily on the host machinery for replication (185). Therefore, identifying the cellular proteins that directly interact with coronavirus proteins can highlight proteins and cellular pathways necessary for coronavirus replication or involved in host cell defence against infection. This study does not indicate whether these interactions are direct or indirect, and false positives may have been detected. Proteins showing ≥1.5-fold enrichment in HEK293-E and -M cells compared to HEK293-Flp (control) cells were analysed using DAVID (Table **3-2** A and B) and STRNG (Figure 3-8 and Figure 3-9), identifying proteins associated with functional annotation terms enriched in the pull-down samples.

The rationale for identifying high-confidence protein interaction partners for the MERS-CoV E or M proteins is described in Section 3.2.5. Protein selection focused on selecting host proteins from an interactomic data set that bound most specifically to the E and M proteins (compared to the control) but also had functional properties and cellular localisation patterns potentially relevant to coronavirus replication. The MERS-CoV E and M proteins are known to be ER localised (115), as confirmed in this study (Figure 3-2). This study selected potential interaction partners known to be ER localised for downstream validation, including TMEM43, YIPF5, CERS2, ERGIC1, LPCAT1, and RAB10.

In contrast, potential interaction partners not localised to the ER were selected to justify the importance of these partners and their localisation for the MERS-CoV E and M proteins. The MERS-CoV E interaction partners SLC44A2 and TM9SF2 (localised to the plasma membrane

and endosome, respectively) and the MERS-CoV M interaction partners VDAC1, SCAPMP3, and IPO11 (localised to the mitochondrion, nucleus, and Golgi respectively) were also selected for validation.

While only three protein interactors were shared between MERS-CoV E and M proteins, the 11 selected protein interactors were tested to confirm whether they interacted with MERS-CoV E or M proteins or both. Table 3-6 shows that most selected host proteins were confirmed as protein interactors for both MERS-CoV E and M proteins. For further confirmation, siRNA knockdown was performed, the results of which will be described in Chapter 5. In the sections below, each cellular protein interactor selected for validation in this study is discussed regarding the outcomes of its validation process and its cellular localisation, function, and potential importance for virus replication, including known interactions with other viral proteins.

3.3.1 ER protein interactors

3.3.1.1 TMEM43

TMEM43 has been previously shown to localise to the ER and nuclear inner membranes and contribute to innate immune signalling regulation (186). The IFN response is enhanced by the loss of TMEM43 (186). Moreover, TMEM43 was identified as a host protein interactor of the M glycoprotein of herpes simplex virus 1 but did not affect viral growth when depleted by siRNA knockdown (187). The results show a 3.8-fold increase in the abundance of TMEM43 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. Moreover, TMEM43 was identified as a MERS-CoV E interactor in HEK293T cells transiently expressing MERS-CoV E (153). The IFA and WB confirmation analyses (

Figure 3-11) confirmed the interaction of TMEM43 with both the MERS-CoV E and M proteins, the colocalisation was confirmed only with MERS-CoV E protein.

3.3.1.2 YIPF5

In a steady state, YIPF5 was observed to localise to the ER exit sites, ERGIC, and cis-Golgi apparatus (183). This localization pattern suggests that YIPF5 undergoes recycling between the ER and Golgi apparatus (183). YIPF5 knockdown caused the partial disassembly of the Golgi apparatus, suggesting that it is involved in maintaining the Golgi structure (183). Moreover, YIPF5 knockdown impaired cellular antiviral responses to DNA viruses (182). These studies show that YIPF5 could be important in viral infections and Golgi function. This study showed a 3.4-fold increase in the abundance of YIPF5 in co-IPs from lysates of cells

expressing the MERS-CoV E protein compared to Flp cells. Moreover, the IFA and WB confirmation analyses (

Figure 3-13) confirmed the YIPF5 interaction and colocalisation with both MERS-CoV E and M proteins. This confirmation indicates that the YIPF5 cellular protein could be important for coronavirus replication.

3.3.1.3 CERS2

CERS2 is localised to the ER membrane and plays a role in lipid biosynthesis (188). CERS2 is one of six enzymes known as ceramide syntheses (CERS) (189). Ceramide is a key lipid signalling molecule (188). CERS2 deficiency altered the lipid composition of cells and inhibited human immunodeficiency virus (HIV-1) envelope receptor binding and/or fusion processes (190). These studies indicate that CERS2 could be important in viral infection. In this study, the proteomic analysis showed a 3.1-fold increase in the abundance of CERS2 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. This information, including its role in lipid biosynthesis, makes CERS2 an interesting protein interactor for further investigation. Moreover, the IFA and WB confirmation analyses (Figure 3-12) confirmed the interaction and colocalisation of CERS2 with both MERS-CoV E and M proteins. This confirmation indicates that CERS2 could be important for coronavirus replication.

3.3.1.4 ERGIC1

ERGIC1 is a membrane-bound protein in the ERGIC between the ER and the Golgi (191). ERGIC1 is known to mediate membrane trafficking and selective cargo transport (192). A previous study examining gene expression profiles in COVID-19 patients found that females expressed ERGIC1 more highly than males (191). Another study found that ERGIC1 interacted with SARS-CoV-2 protein nsp10 (193). In this study, the proteomic analysis showed a 2.4-fold increase in the abundance of ERGIC1 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. Together with ERGIC1's known localization to the ERGIC, where the MERS-CoV E and M proteins are known to localise, this finding underscores the potential significance of ERGIC1 as a host interactor for these viral proteins. Further investigation to validate this interaction is warranted. Despite that, the IFA was shown colocalisation between ERGIC1 and the MERS-CoV E and M proteins, the WB need to be redone since the finding in (Figure 3-14) are not convincing.

3.3.1.5 LPCAT1

LPCAT1 is known to localise in the ER membrane, lipid droplets, and cell membrane (194). It was suggested to play a role in synthesising phosphatidylcholine in pulmonary surfactants and a crucial role in the respiratory system's physiology. In addition, it contributed to regulating the

number and size of lipid droplets (195). Moreover, LPCAT1 is highly expressed in alveolar type II epithelial cells in the lung and has an anti-inflammatory effect (196, 197). This study showed a 3.5-fold increase in the abundance of LPCAT1 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. LPCAT1 was previously identified as a cellular protein interactor for 3A protein for enterovirus A71 (198). Another study found that hepatitis B virus infection caused an increase in LPCAT1 protein expression (199). These two studies suggest that LPCAT1 could be important in multiple virus infections. Moreover, the IFA and WB confirmation analyses (Figure 3-19) confirmed the LPCAT1 interaction and colocalisation with MERS-CoV E and M proteins. This confirmation indicates that the LPCAT1 cellular protein could be important for coronavirus replication.

3.3.1.6 RAB10

RAB10 is known to localise in the Golgi apparatus membrane, ER membrane, endosome membrane, and cytoplasm (200, 201). Interestingly, RAB10 knockdown promoted the accumulation of Epstein–Barr virus (EBV) viral structural proteins within the cell by inhibiting their transport to the plasma membrane, inhibiting the release of infectious virions (202). It was hypothesised that mature EBV virions are released from infected cells to the extracellular environment via the secretory pathway (202), which is also known to be an important cellular pathway for the coronavirus E and M proteins (118). Moreover, it was hypothesised that RAB GTPases Rab8a, Rab10, and Rab11a are crucial for viral release via fusion with the plasma membrane and transport mechanisms enabling mature viruses to reach the plasma membrane (202). The RAB GTPases are essential for intracellular membrane trafficking, from forming transport vesicles to their fusion with membranes (200). Moreover, RAB10 was identified as part of the SARS-CoV-2 high-confidence interactome (203). These studies support our selection of RAB10 for confirmation as a cellular protein interactor for MERS-CoV E and M proteins. Indeed, it was a top cellular protein interactor with the MERS-CoV M protein, showing a 2.4-fold increase in co-IPs from lysates of cells expressing MERS-CoV M protein compared to Flp cells. Moreover, the IFA and WB confirmation analyses (Figure 3-18) only confirmed the RAB10 a weak interaction with MERS-CoV M protein, while it colocalised with MERS-CoV E and M proteins. The MERS-CoV M protein IFA colocalisation represented picture was not an Ideal suggesting than could be done again to reconfirm the colocalisation.

3.3.2 Other cellular organelle protein interactors

3.3.2.1 SCAMP3

SCAMP3 is involved in ER-Golgi trafficking. Following its discovery in 1997, it was associated with epidermal growth factor receptor (EGFR) recycling. It stimulates cell proliferation and motility and is directly associated with EGFR redistribution and degradation (204, 205). EGFR is redistributed from the cytoplasm to the perinucleus by SCAMP3 (206). Moreover, by activating innate immunity, SCAMP3 negatively affected an avian influenza infection (205). SCAMP3 contributed to cellular immunity by reducing IFN-induced transmembrane (IFITM) protein 3 lysosomal degradation, which stops when it is depleted, making IFITM3 subject to further lysosomal degradation (205). Diverse viruses are severely constrained by the IFITM family of proteins (205). In other studies, SCAMP3 was identified as a host protein interactor of glycoprotein M herpes simplex virus 1 protein (187). A recent study identified SCAMP3 as a cellular protein interactor for the enterovirus A71 3A protein using an immunoprecipitation assay. SCAMP3 was also found to interact and colocalise with the viral 3A protein during virus infection, with SCAMP3 knockdown reducing enterovirus A71 RNA synthesis and viral growth (198). Moreover, the same study found the 3A protein-SCAMP3 interaction essential for coxsackievirus but not dengue virus type 2 replication (198). Recently, it was found that reducing SCAMP3 expression modulated SARS-CoV-2 assembly (207). These studies support our selection of SCAMP3 for confirmation as a cellular protein interactor for MERS-CoV E and M proteins. Indeed, it was at the top of the MERS-CoV M protein interactor list with a 3.2-fold increase in abundance between co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. And showed a 2.1-fold increase in the abundance of SCAMP3 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. Moreover, the IFA and WB confirmation analyses (Figure 3-20) confirmed the SCAMP3 interaction and colocalisation with MERS-CoV E and M proteins. This confirmation indicates that the SCAMP3 cellular protein could be important for coronavirus replication.

3.3.2.2 TM9SF2

It has been reported that human TM9SF2 is a Golgi complex-resident transmembrane protein necessary for Golgi complex localisation (208). Several coronavirus M proteins have been shown to accumulate in the Golgi complex of mammalian host cells (136). In addition, TM9SF2 is an endosomal protein functioning as a channel or small molecule transporter (209). The observed localization and function of TM9SF2 are of particular interest given its 3.5-fold greater abundance in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to control Flp cells. Nevertheless, it is an important adeno-associated virus (AAV) transduction factor (210). A pooled clustered regularly interspaced short palindromic repeats

(CRISPR)/CRISPR-associated 9 (Cas9) screen was performed in non-haploid U-2 OS cells using an AAV vector encoding enhanced GFP (EGFP) (210). They found that TM9SF2 significantly reduced the number of EGFP-positive cells compared to parental U-2 OS control cells (210). Another study used a pseudotyped vesicular stomatitis virus expressing the CHIKV E2/E1 envelope proteins to challenge HAP1 cells and examined the surviving cells (211). The CRISPR/Cas9 system was used to create knockout cells for each gene, which were then examined to determine whether these genes were responsible for CHIKV infection. TM9SF2 was identified as an important and critical protein for CHIKV infection (211). The IFA and WB analyses showed no evidence that TM9SF2 interacts or colocalises with the MERS-CoV E or M proteins (

Figure 3-10), contrasting with the results of the proteomic analysis.

3.3.2.3 SLC44A2

In addition to ER localised proteins, coronavirus replication could depend on many integral membrane proteins. This study identified SLC44A2 as a MERS-CoV E protein interactor. It is a transmembrane glycoprotein belonging to the CTL15 family of choline transporter-like proteins and is strongly expressed in lung tissue (212). Recent research implicates SLC44A2 in the regulation of choline transport into mitochondria, platelet activation, and thrombosis (213). Additional experiments showed that the function of the SLC44A2 protein on the surface of neutrophils is to interact with platelets (214). As soon as the neutrophils interacted with the platelets, they began to produce traps that enlarged blood clots by capturing other blood cells and proteins (214). Moreover, examination of gene expression in bovine epithelial cells infected with bovine alpha herpesvirus type 1 showed that *SLC44A2* gene expression was decreased in infected compared to uninfected cells (215). The previous two studies could suggest that the viruses' proteins can reduce *SLC44A2* expression to avoid host immunity. The original proteomic analysis showed a 2.9-fold increase in the abundance of SLC44A2 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. Moreover, the IFA and WB confirmation analyses finding was not convincing (Figure 3-15) and need to be done again since that was done only once.

3.3.2.4 VDAC1

VDAC1 is localised in the mitochondrial outer membrane and mediates the transport of ions and metabolites, regulating communication between mitochondria and the rest of the cell (216). At the outer mitochondrial membrane, VDAC1 is known to interact with >100 proteins and control cellular activities via several signalling pathways (217). It was reported that VDAC1

and the ER IP3 receptor are involved in interactions between the ER and mitochondria (218). ER-mitochondria crosstalk regulates Ca^{2+} transfer, signalling, lipid synthesis, cellular metabolism, autophagy, and apoptosis progression (219). It was reported that the EBV controls Ca^{2+} release in the cellular cytoplasm by controlling VDAC1 (220). VDAC1 also plays a key role in mitochondria-mediated apoptosis (216). The influenza virus PB1-F2 protein was identified as an interactor with VDAC1, controlling apoptosis (221). Moreover, VDAC1 was found to be upregulated during infectious bursal disease virus (IBDV) infection and to regulate IBDV polymerase activity (222). This study infected DF-1 and 293T cells with IBDV and detected VDAC1 protein expression by WB (222). Despite the known ER/ERGIC localisation of the MERS-CoV E and M proteins, VDAC1 was identified as a MERS-CoV M protein interactor. Our experiment showed a 2.4-fold increase in the abundance of VDAC1 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. And 2.8- fold increase in the abundance of VDAC1 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. Moreover, while the IFA and WB confirmation analyses (

Figure 3-17) confirmed the VDAC1 interaction with MERS-CoV M protein more than MERS-CoV E protein, the colocalisation was confirmed only with MERS-CoV M protein. This confirmation indicates that the VDAC1 cellular protein could be important for coronavirus replication.

3.3.2.5 IPO11

IPO11 localises to the cytoplasm and nucleus and functions as a nuclear transport receptor involved in nucleocytoplasmic shuttling (223). One experiment used HEK293T cells to express the SARS-CoV-2 M protein, showing that the SARS-CoV-2 M protein was present in the cytosol and nucleus membranes (103). Moreover, IPO11 was identified as a member of the MERS CoV, SARS-CoV, and SARS-CoV-2 M protein interactomes (153). Numerous viral factors interact with host nuclear transport receptors to access the host nucleus (224). Viruses could use IPO11 to access the host nucleus (225). IPO11 was reported as a predicted Ns2a Zika virus protein interactor without being validated (226). Moreover, in an RNAi study aiming to identify nuclear transport factors important for efficient herpes simplex virus replication, siRNAs targeting IPO11 significantly increased herpes simplex virus gene expression in HeLa cells, indicating that IPO11 may regulate herpes simplex virus replication (227). Our experiment showed a 3.0-fold increase in the abundance of IPO11 in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells. And 1.2-fold increase in the abundance of IPO11 in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. The IFA and WB confirmation analyses (Figure 3-16) confirmed the

IPO11 interaction and colocalisation with both MERS-CoV E and M proteins. This confirmation indicates that the IPO11 cellular protein could be important for coronavirus replication.

CHAPTER 4 Comparative high-throughput interactomic analysis of the MERS-CoV and SARS-CoV-2 E and M proteins in HEK293, PaKiT, and Dubca cells

4.1 Introduction

It is well known that viruses manipulate the host cellular environment during infection to evade the host immune response and replicate. This modulation changes the cellular proteome, which can be investigated using either a stable or transient transfection with a plasmid to make these cells produce a viral protein interacting with host cellular proteins. Previous interactomic studies have identified the cellular interactome and important cellular processes of various coronavirus proteins using a transiently transfected cell line. Most interactomic studies have used human HEK293 and A549 lung carcinoma cells to express SARS-CoV-2, SARS-CoV, and MERS-CoV proteins (153-155). Some of these interactomic studies were mentioned briefly in Chapter 1. Here, we focus on their identified cellular protein interactions and cellular processes for the SARS-CoV-2, SARS-CoV, and MERS-CoV E and M proteins. Gordon et al. (154) identified six and 30 high-confidence protein interactors for the SARS-CoV-2 E and M proteins, respectively (154). Moreover, Li et al. (92) identified 25 host proteins as SARS-CoV-2 M protein interactors (92). The latter study highlighted interactions between viral proteins, including ORF9b, N, and ORF3a with the SARS-CoV-2 E protein and nsp5, OFR10, ORF7a, nsp16, ORF6, S, nsp2, nsp8, N, S, and ORF7b with the SARS-CoV-2 M protein (92).

In addition, Meyers et al. (103) found that the SARS-CoV-2 M protein was concentrated in the ER or ER proximal membranes (103). This study also predicted that the SARS-CoV-2 M protein was associated with endocytic pathway-associated and lysosomal membranes (103). Moreover, the promoter of the interferon signalling regulator TRIM4 discovered proximal to the SARS-CoV-2 M protein was associated with increased COVID severity in patients (103).

In addition, Stukalov et al. (155) identified five important cellular processes for the SARS-CoV-2 and SARS-CoV M proteins (155). This study used A549 lung carcinoma cells transduced with lentiviruses encoding individual SARS-CoV-2 and SARS-CoV viral proteins with haemagglutinin tags (155).

Finally, Gordon et al. (153) reported 389 high-confidence interactors for SARS-CoV-2, 366 for SARS-CoV, and 296 for MERS-CoV, including those shared among the SARS-CoV-2, SARS-CoV and MERS-CoV E and M proteins (153). A cellular compartment GO enrichment analysis was used to compare the localisations of the expressed viral proteins to their interaction partners (153). Functional enrichment analyses showed that all three viruses target the same host processes, including ribosome biogenesis and RNA metabolism regulation (138). This experiment identified seven clusters of viral-host interactions that were either unique to each virus or shared among them. The SARS-CoV-2, SARS-CoV, and MERS-CoV proteins showed similar localization patterns for most shared protein homologs in human cervical cancer cell line M (HeLaM) cells, supporting the hypothesis that conserved proteins share functional similarities (153). In addition, the results showed that protein localizations determined using expressed tags were rarely altered (153). Therefore, they are unlikely to be a significant source of host targeting mechanism variation. Table 4-1 summarises the E and M viral protein interactomic studies, including the cell line used and the identified cellular processes, pathways, and cellular protein interactors.

Table 4-1 Summary of SARS-CoV-2, SARS-CoV, and MERS-CoV E and M proteins interactomic studies, including the cell line used, cellular processes, pathways, and cellular protein interactors.

Virus	Viral protein	Cell line	Cellular processes, pathways, and cellular protein interactors	Ref.
SARS-CoV-2	E	HEK-293T/17	• CWC27, AP3B1, ZC3H18, SLC44A2, BRD2, and BRD4.	(154)
SARS-CoV-2 and SARS-CoV	E	A549 lung carcinoma	• ER and cytosol trafficking.	(155)
SARS-CoV-2, SARS-CoV, and MERS-CoV	E	HEK293T	• Vesicle transport (AP3B1 and SCFD1). • Aminoglycan metabolism (B4G4T1 and CHPF). • CWC27, BRD2, SPTLC2, DCP1A, DNAJB9, TMEM43, ATL3, TNPO3, BRI3BP, WLS, IGHG4, and ATP6AP2.	(153)
SARS-CoV-2	M	HEK-293T/17	• Mitochondrial metabolism (ATP6V1A, ATP1B1, ACADM, AASS, PMPCB, PITRM1, COQ8B, and PMPCA). • Solute transport (SLC25A21, SLC30A7, and SLC30A9). • ER morphology (REEP6, REEP5, YIF1A, and RTN4). • GGCX, INTS4, PSMD8, STOM, TARS2, AAR2, SAAL1, FAM8A1, ETFA, BZW2, AKAP8L, ANO6, and FASTKD5. • TUBGCP3 and TUBGCP2.	(154)
SARS-CoV-2 and SARS-CoV	M	A549 lung carcinoma	• Mitochondrial metalloproteases. • Lipid oxidation. • ER and Golgi trafficking. • Ions and transport by ATPases. • Condensin II complex. • TNFAIP2, IRAK1, JAK2, and TRIM7.	(155)
SARS-CoV-2	M	HEK293T	• IFN signalling regulation (TRIM4).	(103)
SARS-CoV-2	M	HEK293	• SARS-CoV-2 nuclear proteins (STUB1, NUP85, IPO11, XPO4, UTP20, LMNB1, TIMELESS, NUP155, BTAF1, and NUP133). • Plasma membrane proteins (ATP1B1, DNAJB4, and SLC1A5). • Proteins with unknown localisations (TPC1D9, ARFGEF2, ATG9A,DNAJB5, ANKHD1, and SIGMAR1). • Cytosol proteins (EXOC7, EXOC6, EXOC5, and COG4).	(92)

			• Mitochondrial proteins (ATP5PD and ATP5PC).	
SARS-CoV-2, MERS-CoV, and SARS-CoV	M	HEK293T	• Lipid metabolism (PIGG, CDS2, LPGAT1, GPAT4, FADS2, CHPT1, ACADM, and ETFA). • Solute transport (SLC30A7, SLC30A9, SLC19A2, SLC39A14, SLC39A6, SLC1A3, and SLC25A21). • Vesicle transport (YIF1B, YIF1A, KDELR1, DYNC1I2, MON2, COPG2, VAMP8, and LMAN2L). • Mitochondrial RNA processing (FASTKD1, FASTKD5, and TARS2) • Integrator complex (INTS4 and INTS7). • Ion transport (ATP2B1, ATP13A3, ATP1B1, and ATP2C1). • Mitochondrial metabolism (COQ8B, ATP6V1A, ATP6V1B2, and NDUFAF3). • Pericentrin-gamma complex-associated protein (TUBGCP2 and TUBGCP3). • RNA nuclear export (NUP210). • UBE3C, ABCB7, FAM8A1, SAAL1, MTX1, RTN4, RFT1, PSMD8, TM9SF1, GGCX, STOM, AMFR, AASS, AKAP8L, CAPN1, YIPF3, ANO6, FLVCR1, AAR2, YIPF6, FANCI, IPO11, SUCLA2, IPO8, TELO2, LTN1, LMBRD2, MARCH5, FUT8, and DIPK2A.	(153)

Stable cell lines inducibly expressing FLAG-tagged MERS-CoV E and M proteins (HEK293-E and HEK293-M, respectively) were previously created by Mr Lee. These cells were used for LC-MS/MS-based co-IP analyses to create a proteomic data set containing the cellular interactors of MERS-CoV E and M proteins. Mr Lee's proteomic data was used to select eleven cellular protein interactors for validation in Chapter 3.

This study first aimed to conduct high-throughput co-IP proteomic analyses of the MERS-CoV E and M proteins in three transfected host cells and bioinformatically analyse the resulting data sets to identify host pathways and proteins modulated in response to these transfections. The hosts were human HEK293 cells, *Pteropus alecto* PaKiT cells, and *Camelus dromedarius* Dubca cells. As mentioned in Chapter 1, bats are considered reservoirs for various coronaviruses, and camels are known as a major animal reservoir for MERS-CoV and an intermediary host between bats and humans (17). MERS-CoV and other coronaviruses, such as SARS-CoV and SARS-CoV-2, are known to have initially emerged in bats, infecting animals and then humans (32). In addition, SARS-CoV-like coronaviruses have been isolated from bats (33). Moreover, other studies have isolated identical MERS-CoVs from patients and their camels (38).

Dubca cells were successfully infected with recombinant adenoviral vectors encoding the full-length MERS-CoV S protein to develop a MERS-CoV vaccine for camels (228). This finding will be helpful for camel immunization but does not help in understanding the camel's cellular mechanism to adapt to MERS-CoV infection. In addition, Australian bat lyssavirus replication was inhibited in PaKiT cells compared to human cells, suggesting that the autophagy pathway functions as an antiviral mechanism and has evolved as a cellular homeostatic adaptation in bat cells. These antiviral mechanisms help bats to be natural hosts for various pathogenic viruses (229).

The investigation on the MERS-CoV E and M proteins was ongoing when the SARS-CoV-2 emerged, leading to the global pandemic. Therefore, it was decided to include the SARS-CoV-2 E and M proteins in the analysis using the same cell lines. Their inclusion would make the study even more robust in identifying cellular proteins from diverse species interacting with the E and M proteins from two different coronaviruses.

This chapter aimed to identify highly conserved interactions in host cells after transient transfection with four plasmids: MERS-CoV E, SARS-CoV-2 E, MERS-CoV M, and SARS-CoV-2 M. The resulting human, bat, and camel cell data sets were bioinformatically compared to identify highly conserved interactions between SARS-CoV-2 and MERS-CoV E and M proteins and cellular proteins. While the SARS-CoV-2 and MERS-CoV E and M

proteins have low similarities, a shared cellular protein interactor is likely important for most coronavirus E and M proteins and coronavirus replication.

4.2 Results

4.2.1 Plasmid preparation for transient expression of the MERS-CoV and SARS-CoV-2 E and M genes

Plasmids encoding the MERS-CoV and SARS-CoV-2 E and M proteins were commercially synthesised and comprised a pcDNA3.3-TOPO backbone with the appropriate E and M ORFs codon optimised for expression in human cells followed by a sequence encoding a 3′ terminal FLAG epitope tag. The plasmids' sequences are shown in Appendix B.

The four plasmids were transformed into chemically competent *E.coli* strain DH5α cells. The plasmids were propagated and extracted by plasmid miniprep, followed by *Bam*HI and *Xho*I digestion to validate the plasmid constructs. The results are shown in Figure 4-1. The MERS-CoV and SARS-CoV-2 E and M plasmids met expectations: two bands (6000 and 350 bp) for the MERS-CoV and SARS-CoV-2 E plasmids and two bands (6000 and 750 bp) for the MERS-CoV and SARS-CoV-2 M plasmids. Then, the plasmids were prepared on a larger scale for transient gene expression and interactomic analysis of the respective E and M proteins.

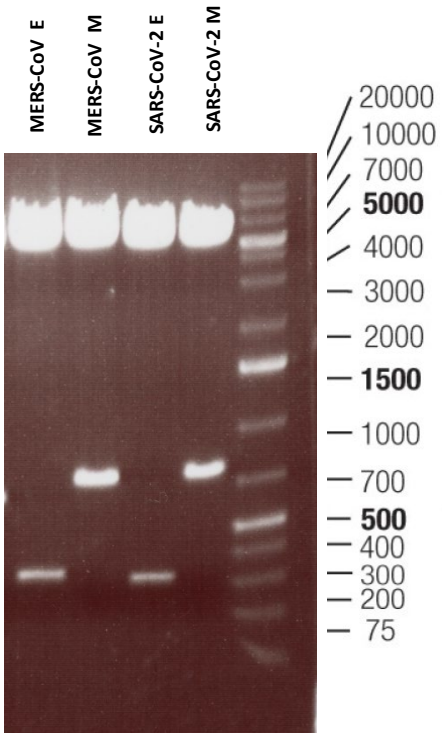


Figure 4-1 Validation of MERS-CoV and SARS-CoV-2 E and M plasmid isolation.

Plasmid mini-preparations were digested with *Bam*HI and *Xho*I, and 5 µg of DNA was loaded per lane on an agarose gel. DNA bands were stained with ethidium bromide and visualised and photographed using a BioDoc-ITTM System Ultraviolet transilluminator.

4.2.2 MERS-CoV and SARS-CoV-2 E and M proteins expression in HEK293, Dubca, and PaKiT cells

Before the interactomic analysis was conducted, it was first necessary to analyse the expression of the MERS-CoV and SARS-CoV-2 E and M proteins in HEK293, Dubca, and PaKiT cells. Previous interactomic studies used HEK293 and human A549 lung carcinoma cells to express SARS-CoV-2, SARS-CoV, and MERS-CoV proteins (153-155). However, no studies have reported expressing coronavirus proteins in Dubca or PaKiT cells.

HEK293, Dubca, and PaKiT cells were grown in flasks (T75) until ~70% confluence (~2.0×10^6 cells) and transfected using a protocol optimised by Dr Abdulaziz Almuqurin (Section 2.2) with the MERS-CoV and SARS-CoV-2 E and M encoding plasmids for 24 h. Then, all three cell types were harvested and lysed with 2× SDS-PAGE sample buffer. Samples were analysed by WB using anti-FLAG antibodies to confirm the expression of the E and M proteins in the different cell lines. In addition, the expression of the respective proteins in Dubca cells was examined by IFA. The plan was to examine the four viral proteins expression in HEK293, PaKiT, and Dubca cells, but due to work restrictions during the pandemic, it was not possible to justify the expression using HEK293 and PaKiT cells.

The WB analyses shown in Figure 4-2 and Figure 4-3 confirmed that the MERS-CoV E, MERS-CoV M, SARS-CoV-2 E, and SARS-CoV-2 M proteins were expressed in HEK293, Dubca, and PaKiT cells. The molecular masses expected for the FLAG-tagged E and M proteins were 17.0 and 27.8 kDa, respectively (154, 179). A single band was detected at 17 kDa for the SARS-CoV and MERS-CoV E proteins, consistent with their predicted molecular mass. While a major band was detected at ~28 kDa for SARS-CoV and MERS-CoV M proteins, other expected bands were seen, suggesting these proteins may form oligomers or be modified by ubiquitination or phosphorylation as previously suggested (138). The results of an IFA analysis using Dubca cells confirmed the WB result. Figure 4-4 shows that the E and M proteins were localised to the cytoplasm as expected.

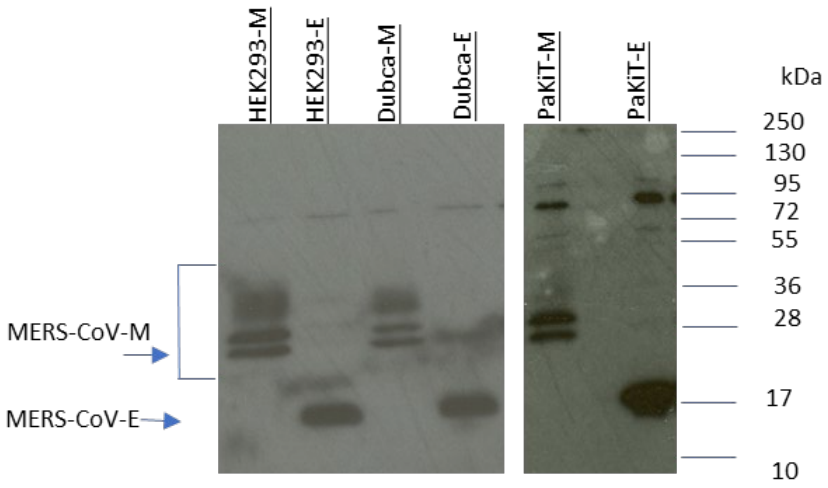


Figure 4-2 Validation of MERS-CoV E and M expression.

HEK293, Dubca, and PaKiT cells were transfected with the MERS-CoV E or MERS-CoV M plasmids for 24 h. Cells were harvested and lysed with 2× SDS-PAGE sample buffer. For each sample, 10 µg of protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting using an anti-FLAG antibody. The expected positions of the MERS-CoV E and M proteins are shown on the left of the figure, and the positions of relevant molecular mass markers (kDa) are shown on the right.

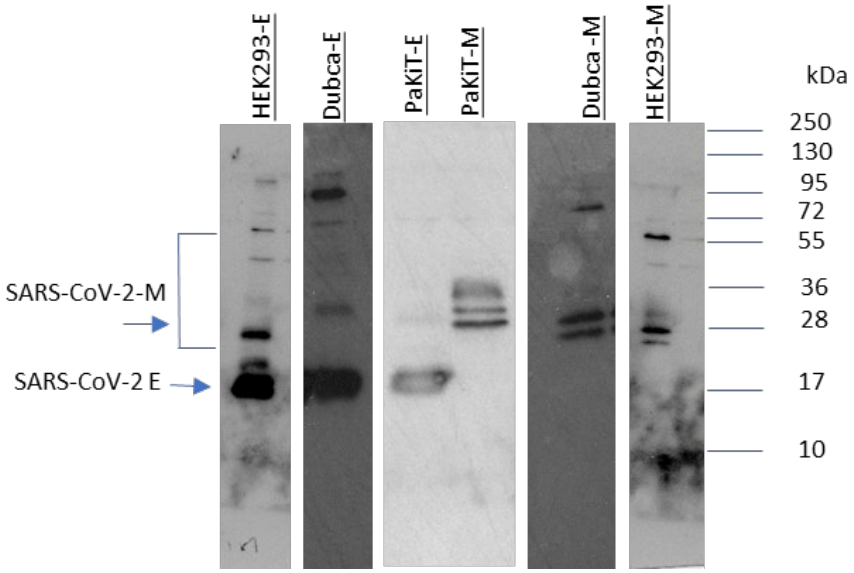


Figure 4-3 Validation of SARS-CoV-2 E and M expression.

HEK293, Dubca, and PaKiT cells were transfected with the SARS-CoV-2 E or SARS-CoV-2 M plasmids for 24 h. Cells were harvested and lysed with 2× SDS-PAGE sample buffer. For each sample, 10 µg of protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting using an anti-FLAG antibody. The expected positions of the SARS-CoV-2 E and M proteins are shown on the left of the figure, and the positions of relevant molecular mass markers (kDa) are shown on the right.

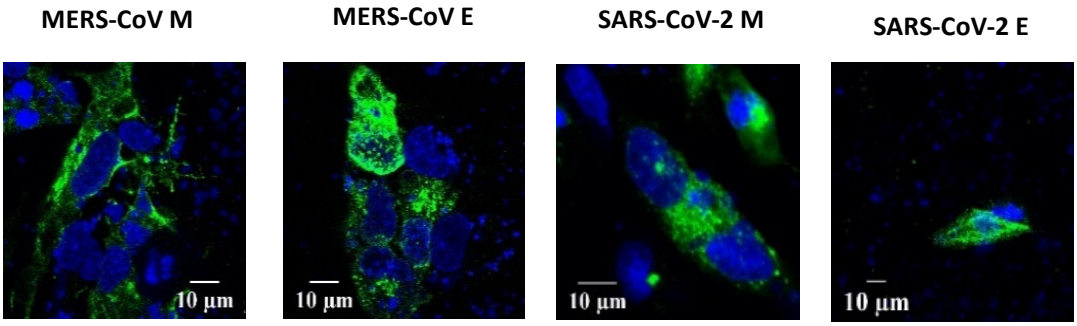


Figure 4-4 IFA analysis of overexpressed FLAG-tagged MERS-CoV and SARS-CoV-2 E and M proteins in Dubca cells.

The cells were grown on coverslips in a 24-well plate and transfected with 2.5 µg of the indicated plasmid for 24 h. Next, the cells were fixed with 4% paraformaldehyde and permeabilised with 1% (v/v) Triton X-100. Then, the fixed cells were probed with an anti-FLAG primary antibody. An Alexa Fluor 488 conjugated secondary antibody was used to detect the primary antibody. A confocal imaging system (Leica SP5 Multi-laser CLSM microscope) in the Wolfson Bioimaging Facility of the Life Sciences Faculty was used to capture the images and visualise DAPI (blue) and the FLAG-tagged M and E proteins (green).

4.2.3 Sample preparation and validation for a high-throughput co-IP analysis

Since the MERS-CoV and SARS-CoV-2 E and M encoding plasmids successfully expressed the proteins in the three cell lines, they were used for immunoprecipitation analysis following the scheme in Figure 4-5. HEK293, Dubca, and PaKiT (4.1×10^6) cells were transfected with the four plasmids. At 24 h post-transfection, the three cell types were harvested and lysed with a lysis buffer. Then, immunocomplexes formed with the FLAG-tagged E and M proteins were captured with anti-FLAG magnetic beads, which were washed to remove proteins bound non-specifically. Finally, the beads were resuspended in 500 µl of wash buffer. A 450 µl volume of resuspended beads was taken for proteomic analysis (after beads were captured and resuspended in a 50 µl volume). The beads in the remaining 50 µl volume were captured and heated at 95°C for 5 min in 2× SDS sample buffer to elute proteins for WB analysis. In addition to the pull-down samples, input and non-bound samples were collected during the procedure. All samples were analysed by WB using anti-FLAG antibodies to confirm protein expression and the pull-down of target proteins before sending the samples for proteomic analysis. The pull-down analysis was repeated in triplicate for each cell line/plasmid combination. The WB analysis showed that the MERS-CoV and SARS-CoV-2 E and M proteins were successfully expressed and pulled down from cell lysates produced from the transfected HEK293, Dubca, and PaKiT cells (Figure 4-6, Figure 4-7, Figure 4-8, Figure 4-9, Figure 4-10, and Figure 4-11). The expected molecular masses of the FLAG-tagged E and M proteins were 17.0 and 27.8 kDa, respectively (113, 179). A single band was detected at 17 kDa for the MERS-CoV and SARS-CoV-2 E proteins, consistent with their predicted molecular mass. While a major band was detected at ~28 kDa for the MERS-CoV and SARS-CoV-2 M proteins, other expected bands were seen, suggesting these proteins may form oligomers or be modified by ubiquitination or phosphorylation (138).

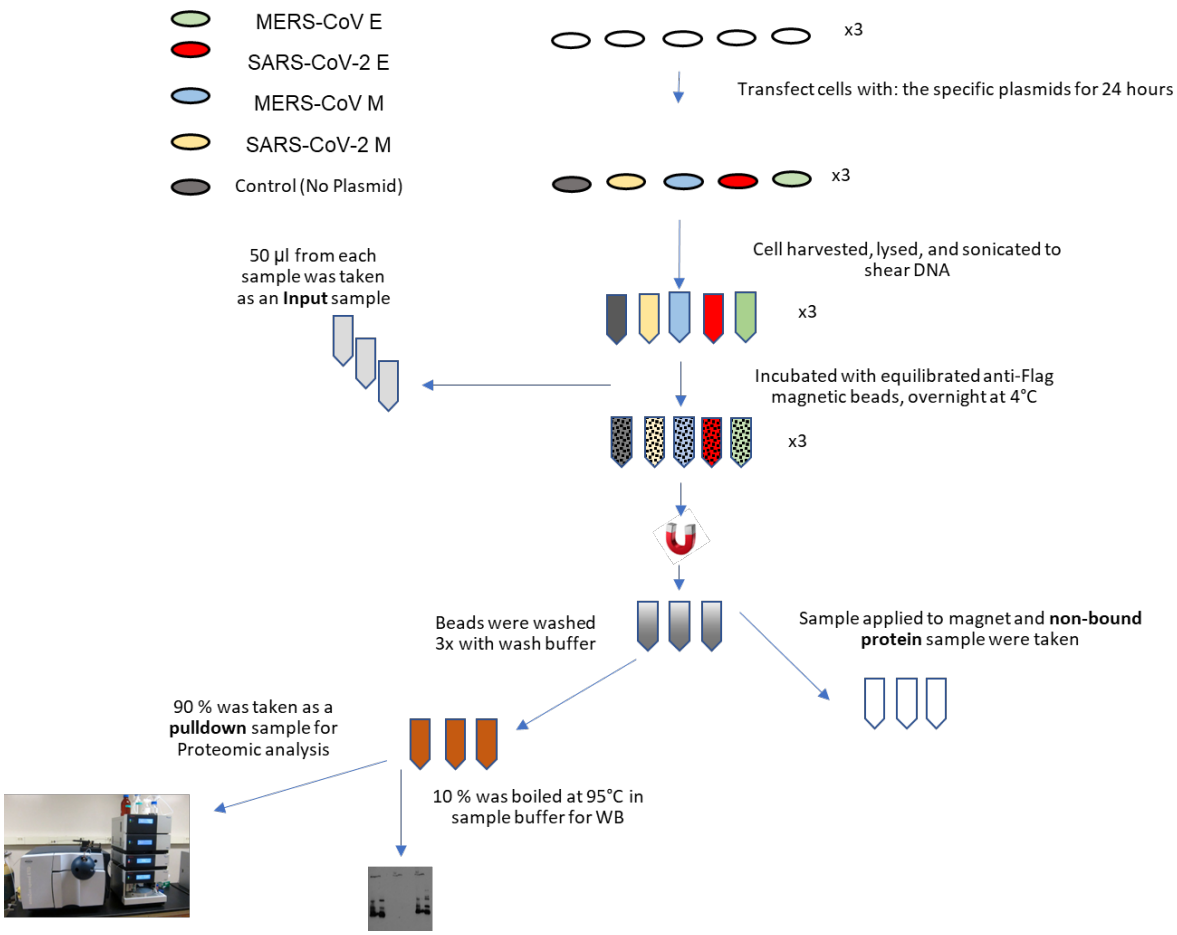


Figure 4-5 A schematic representation of the immunoprecipitation procedure.

HEK293, Dubca, and PaKiT cells were transfected with plasmids encoding the MERS-CoV and SARS-CoV-2 E and M proteins for 24 h. The cells were harvested, washed with PBS, and lysed with lysis buffer. Next, the samples were centrifuged at 17,000 g and 4°C for 20 min, and an aliquot of the lysate supernatant was transferred to a new tube (input samples). Then, the supernatants containing the solubilised M and E proteins were incubated with anti-FLAG conjugated beads overnight at 4°C. Finally, the beads were magnetically captured, the supernatant removed (non-bound sample), and the beads washed three times with wash buffer before being resuspended in 500 µl. A 450 µl volume that was then used for proteomic analysis (after the beads were recaptured and resuspended in 50 µl of wash buffer). In the latter case, the remaining 50 µl of beads were captured and heated at 95°C in 2× SDS sample buffer to elute the bound proteins (pull-down samples). Then, the input, non-bound, and pulldown samples were used for WB analysis.

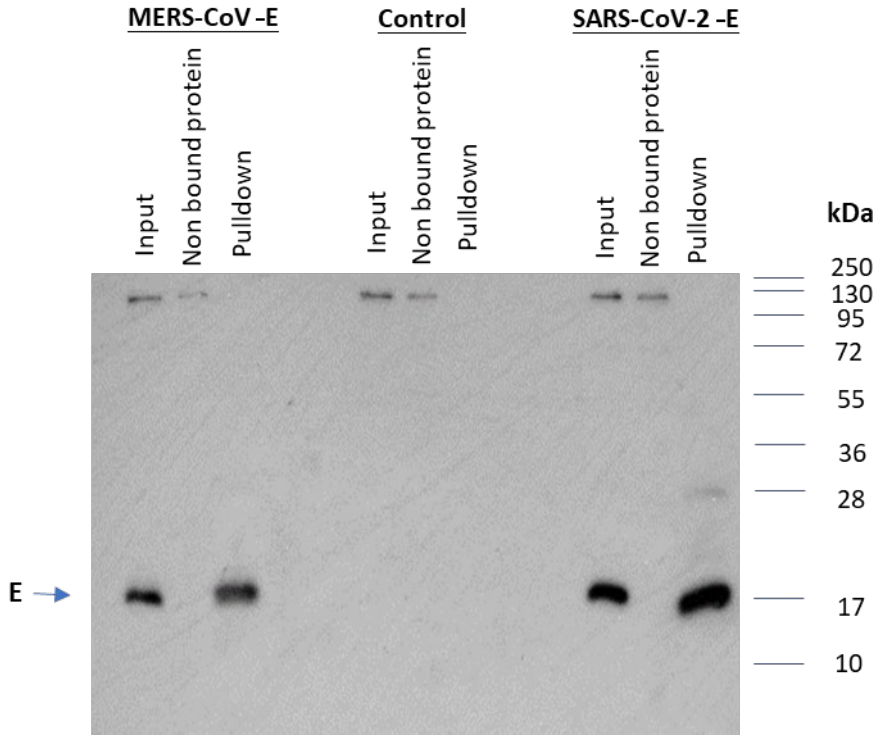


Figure 4-6 Validation of MERS-CoV and SARS-CoV-2 E protein pull-down samples.

Lysates from Dubca cells transfected with plasmids expressing either the MERS-CoV or SARS-CoV-2 E proteins or a no plasmid control were used for co-IP analysis using anti-FLAG peptide magnetic beads. Three samples were collected during the procedure and analysed by WB; input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for immunoblotting with an anti-FLAG antibody. An arrow indicates the position of the E protein, and the positions of relevant molecular mass markers (kDa) are shown on the right of the figure.

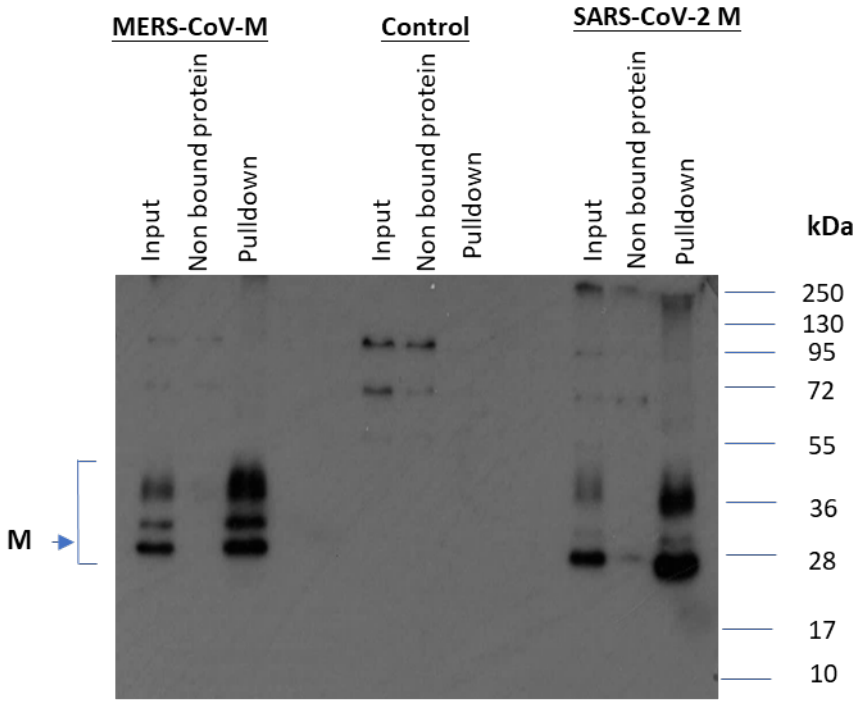


Figure 4-7 Validation of MERS-CoV and SARS-CoV-2 M protein pull-down samples.

Lysates from Dubca cells transfected with plasmids expressing either the MERS-CoV or SARS-CoV-2 M proteins or a no plasmid control were used for co-IP analysis using anti-FLAG peptide magnetic beads. Three samples were collected during the procedure and analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for immunoblotting with an anti-FLAG antibody. An arrow indicates the position of the M protein, and the positions of relevant molecular mass markers (kDa) are shown on the right of the figure.

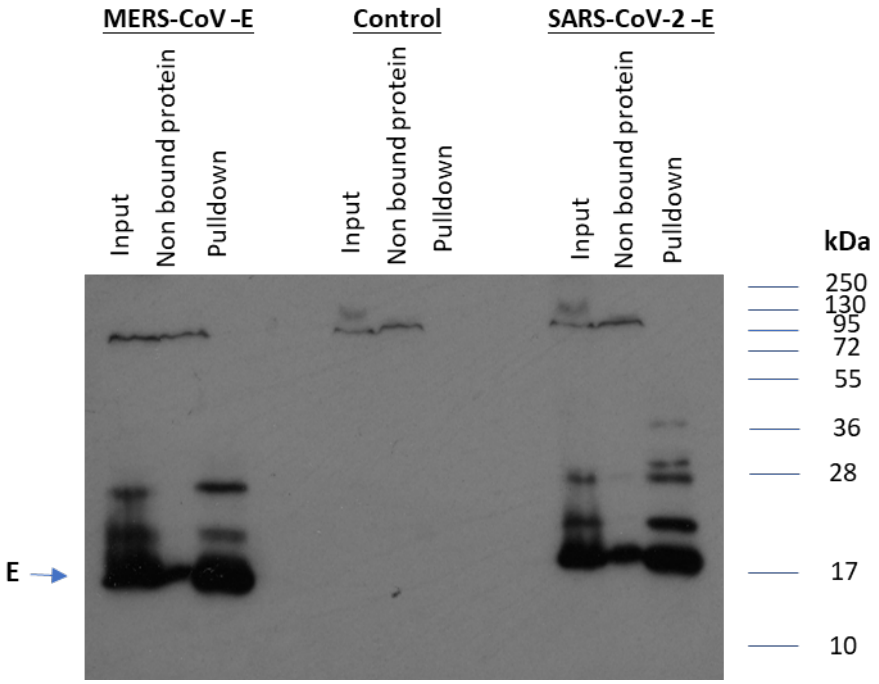


Figure 4-8 Validation of MERS-CoV and SARS-CoV-2 E protein pull-down samples.

Lysates from HEK293 cells transfected with plasmids expressing either the MERS-CoV or SARS-CoV-2 E proteins or a no plasmid control were used for co-IP analysis using anti-FLAG peptide magnetic beads. Three samples were collected during the procedure and analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for immunoblotting with an anti-FLAG antibody. An arrow indicates the position of the E protein, and the positions of relevant molecular mass markers (kDa) are shown on the right of the figure.

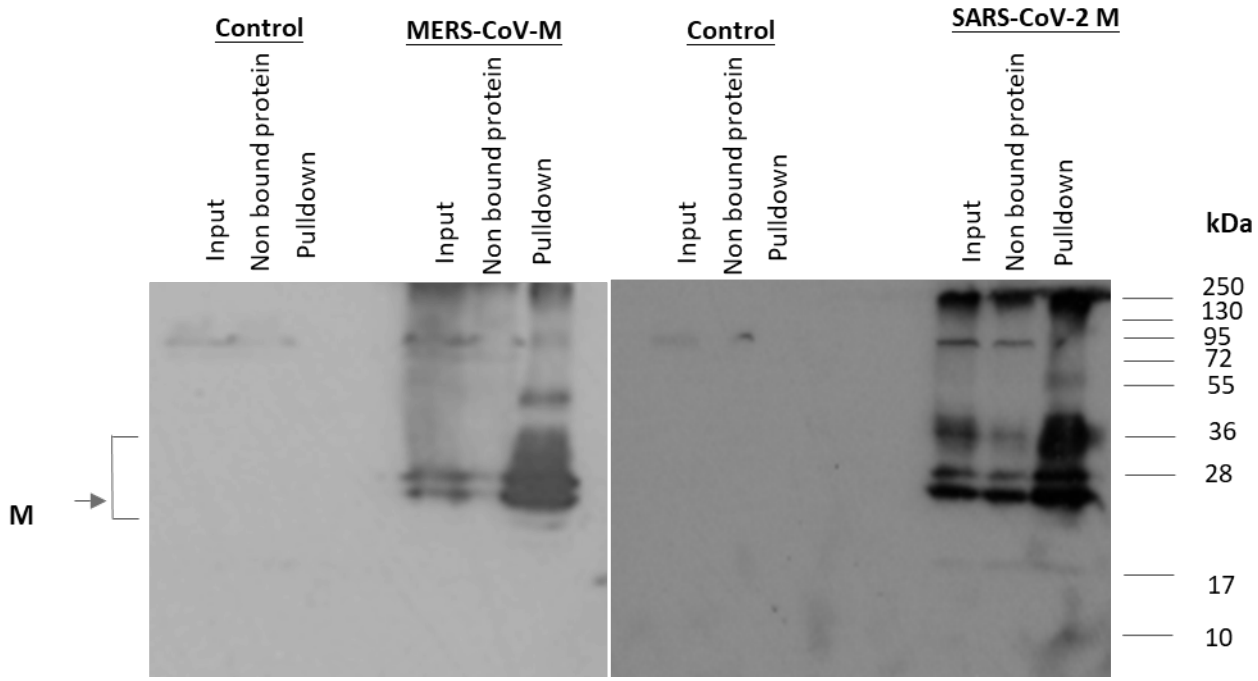


Figure 4-9 Validation of MERS-CoV and SARS-CoV-2 M protein pull-down samples.

Lysates from HEK293 cells transfected with plasmids expressing either the MERS-CoV or SARS-CoV-2 M proteins or a no plasmid control were used for co-IP analysis using anti-FLAG peptide magnetic beads. Three samples were collected during the procedure and analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for immunoblotting with an anti-FLAG antibody. An arrow indicates the position of the M protein, and the positions of relevant molecular mass markers (kDa) are shown on the right of the figure.

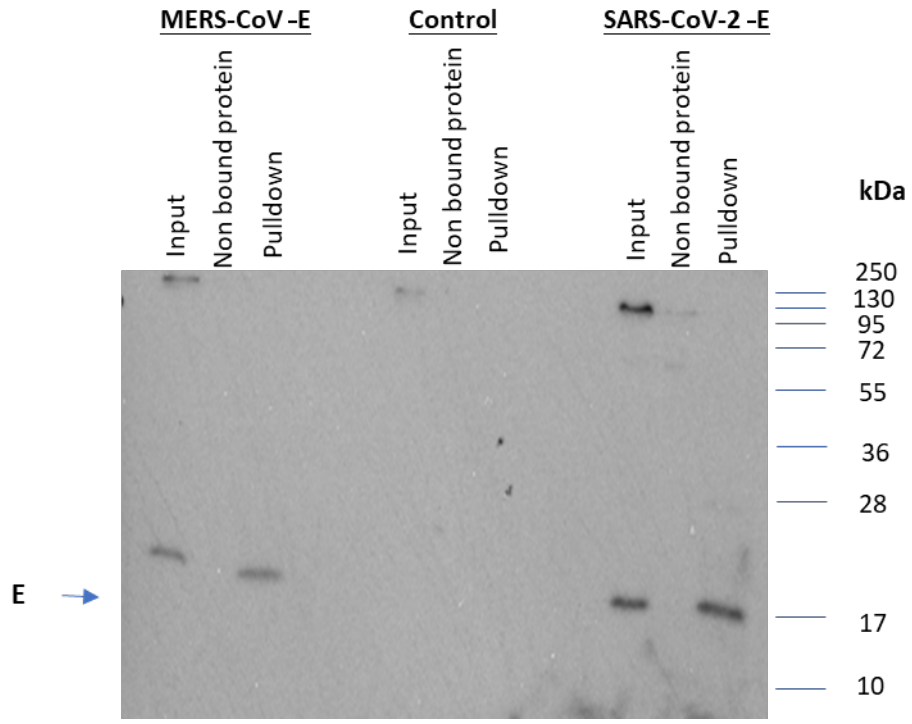


Figure 4-10 Validation of MERS-CoV and SARS-CoV-2 E protein pull-down samples.

Lysates from PaKiT cells transfected with plasmids expressing either the MERS-CoV or SARS-CoV-2 E proteins or a no plasmid control were used for co-IP analysis using anti-FLAG peptide magnetic beads. Three samples were collected during the procedure and analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for immunoblotting with an anti-FLAG antibody. An arrow indicates the position of the E protein, and the positions of relevant molecular mass markers (kDa) are shown on the right of the figure.

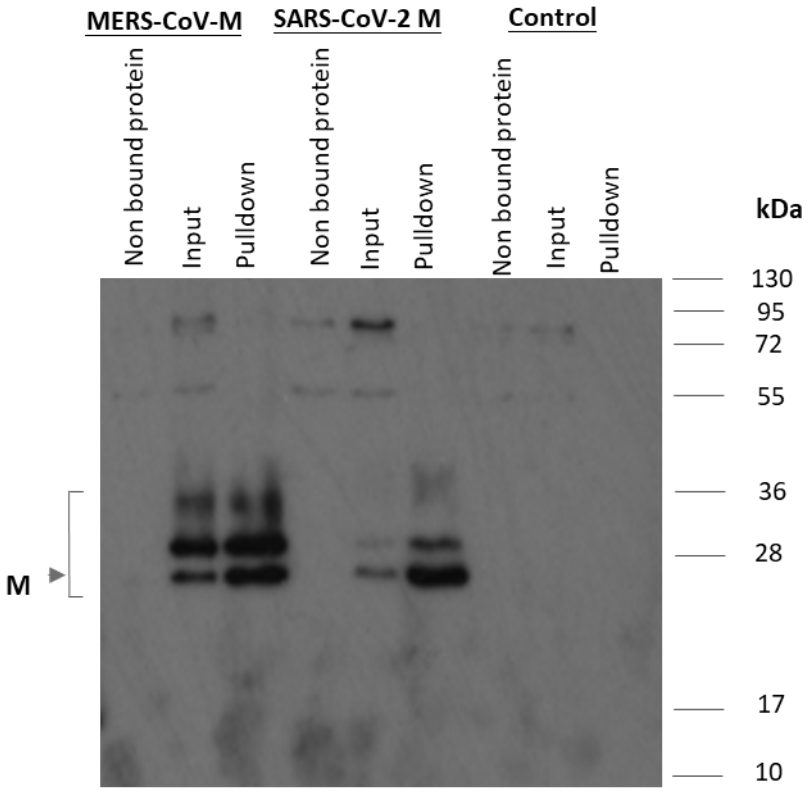


Figure 4-11 Validation of MERS-CoV and SARS-CoV-2 M protein pull-down samples.

Lysates from PaKiT cells transfected with plasmids expressing either the MERS-CoV or SARS-CoV-2 M proteins or a no plasmid control were used for co-IP analysis using anti-FLAG peptide magnetic beads. Three samples were collected during the procedure and analysed by WB: input, non-bound, and pull-down. Each lane of the 15% SDS-PAGE gel was loaded with 10% of the respective sample. Proteins were size separated and transferred to a PVDF membrane for immunoblotting with an anti-FLAG antibody. An arrow indicates the position of the M protein, and the positions of relevant molecular mass markers (kDa) are shown on the right of the figure.

4.2.4 Quantitative LC-MS/MS analysis of three hosts transfected cell lysates

The MERS-CoV and SARS-CoV-2 E and M protein and no plasmid control pull-down samples prepared from the HEK293, PaKiT, and Dubca cells were used for high-throughput co-IP analysis to determine the interactome of the E and M proteins in the different cell types. Each data set (five per cell type) was repeated in triplicate. Figure 4-12 shows the process for HEK293 cells; the same process was used for PaKiT and Dubca cells. All final terms are summarised in Table 4-2.

For high-throughput co-IP analysis, the samples were digested and labelled with TMT ten-plex reagents before high pH reversed-phase chromatography fractionation using an Ultimate 3000 liquid chromatography system. The resulting fractions (four in total) were analysed by nano-LC-MS/MS using an Orbitrap Fusion Lumos mass spectrometer to identify and quantify the proteins. The details of the complete analysis are described in Section 2.9. The raw data files were processed and quantified using Proteome Discoverer software v2.1 and searched against the UniProt/SwissProt human, *Pteropus alecto*, and *Camelus dromedarius* databases and FASTA files containing the protein identification details for SARS-CoV-2 E (GenBank protein identifier: QHD43418.1), MERS-CoV E (GenBank protein identifier: YP_009047209.1), SARS-CoV-2 M (GenBank protein identifier: QHD43419.1), and MERS-CoV M (GenBank protein identifier: YP_009047210.1).

The raw abundance data were Log_2-transformed to bring the data closer to a normal distribution. Welch's *t*-test was used to calculate *p*-values where appropriate and adjusted for multiple testing using the Benjamini–Hochberg procedure. These data were visualised alongside mean log_2 fold changes and counts of the number of samples used to calculate these statistics. While all proteomic samples had been prepared in triplicate, only two replicates were used for HEK SARS2 M and Dubca SARS2 E since one was discarded after a sample was identified as an outlier by the proteomics facility's bioinformatician (Dr Phil), confirmed by the original sample's BCA analysis. Discarding one sample reduced the significant protein number used for comparison with the other data sets, which may affect the enrichment analysis for shared proteins.

The Galaxy website (usegalaxy.org) was used to annotate the Dubca cells database using a camel protein accession number to obtain a human protein name. In contrast, an online version of STRING (159) was used to annotate the whole PaKiT cell data set.

Analyses of the three transfected cell lines (HEK293, Dubca, and PaKiT) expressing SARS-CoV-2 or MERS-CoV E and M proteins identified interactors for each viral protein. Each data set of viral protein interactors was filtered and processed to present the $\log_2$ fold change for a specific host protein interacting with the viral target protein of interest compared to non-transfected (negative control) cells with a *t*-test <0.05. Host proteins showing a $\log_2$ fold change >0 with a specific viral protein compared to non-transfected (negative control) cells with a *t*-test <0.05 were selected for the comparative analysis. The reason for focusing on $\log_2$ fold change differences of >0, with a *t*-test <0.05, was to identify as many shared proteins as possible between the MERS-CoV and SARS-CoV-2 E and M data sets.

Often in proteomic studies, a particular threshold is selected. However, this was difficult with the different data sets since a threshold in one data set could be either too generous or restrictive for another one. Therefore, the data were analysed using several approaches to select proteins for downstream analysis. Initially, histograms of the total number of proteins in each data set, proteins with a $\log_2$ fold change >0 and a *t*-test <0.05 on each data set, and normalised proteins with a $\log_2$ fold change >0 and a *t*-test <0.05 were examined (see Appendix E). Finally, it was decided that the best way was to select the top 10% of proteins from each data set for further bioinformatics analysis. If the top 10% represented <20 proteins, all proteins showing significant increases ($\log_2$ fold change >0 with a *t*-test <0.05) were selected.

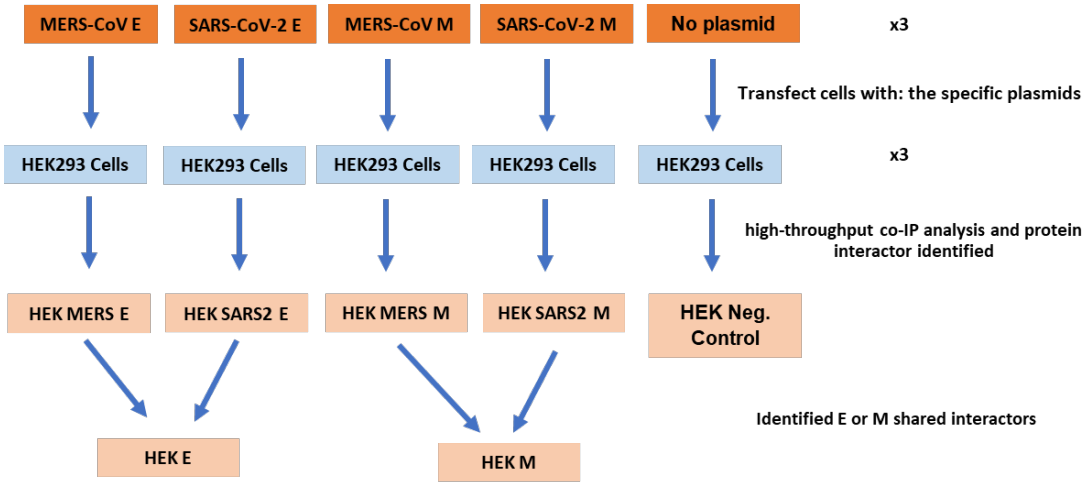


Figure 4-12 A schematic representation of the proteomic data set resulting from a high-throughput pull-down analysis of MERS-CoV and SARS-CoV-2 E and M proteins expressed in HEK293 cells.

HEK293 cells were transfected with plasmids expressing MERS-CoV E, SARS CoV 2 E, MERS-CoV M, or SARS CoV 2 M for 24 h. High throughput co-IP analysis was performed, and protein interactions were identified, quantified, and filtered (log_2 fold change of >0 and *t*-test <0.05). Data were compared to select shared HEK M and HEK E interactomes.

Table 4-2 The data sets used in this study.

Viral protein	Host cell	Data set	Merged data set
MERS-CoV E	HEK293	HEK MERS E	HEK E
SARS-CoV-2 E	HEK293	HEK SARS2 E	
MERS-CoV M	HEK293	HEK MERS M	HEK M
SARS-CoV-2 M	HEK293	HEK SARS2 M	
Negative control (no transfected plasmid)	HEK293	HEK Neg. Control	
MERS-CoV E	Dubca	Dubca MERS E	Dubca E
SARS-CoV-2 E	Dubca	Dubca SARS2 E	
MERS-CoV M	Dubca	Dubca MERS M	Dubca M
SARS-CoV-2 M	Dubca	Dubca SARS2 M	
Negative control (no transfected plasmid)	Dubca	Dubca Neg. Control	
MERS-CoV E	PaKiT	PaKiT MERS E	PaKiT E
SARS-CoV-2 E	PaKiT	PaKiT SARS2 E	
MERS-CoV M	PaKiT	PaKiT MERS M	PaKiT M
SARS-CoV-2 M	PaKiT	PaKiT SARS2 M	
Negative control (no transfected plasmid)	PaKiT	PaKiT Neg. Control	

4.2.4.1 *Quantitative LC-MS/MS analysis of HEK293 transfected cell lysates*

Comparing HEK293 transfected cells expressing SARS-CoV-2 E (HEK SARS2 E) or MERS-CoV E (HEK MERS E) to non-transfected HEK293 cells (HEK Neg. Control) identified 4626 protein interactors (Table 4-3). In the HEK SARS2 E and HEK MERS E data sets, 250 and 536 proteins showed a $\log_2$ fold change >2 compared to the HEK Neg. Control data set with a *t*-test <0.05, respectively. In addition, 2078 and 2288 proteins showed a $\log_2$ fold change >1 with a *t*-test <0.05, respectively. Moreover, 2496 and 2676 proteins showed a $\log_2$ fold change >0 with a *t*-test <0.05, respectively. There were 2271 proteins shared between the HEK SARS2 E and HEK MERS E data sets ($\log_2$ fold change >0 with a *t*-test <0.05). The top 10% of proteins from each data set (250 from HEK SARS2 E, 268 from HEK MERS E, and 227 from the shared proteomic data set) were selected for further bioinformatics analysis.

Comparing HEK293 transfected cells expressing SARS-CoV-2 M (HEK SARS2 M) or MERS-CoV M (HEK MERS M) to non-transfected HEK293 cells (HEK Neg. Control) identified 650 and 5113 protein interactors, respectively (Table 4-3). In the HEK SARS2 M and HEK MERS M data sets, 17 and 3041 proteins showed a $\log_2$ fold change >2 compared to the HEK Neg. Control data set with a *t*-test <0.05, respectively. In addition, 156 and 4793 proteins showed a $\log_2$ fold change >1 with a *t*-test <0.05, respectively. Moreover, 628 and 4996 proteins showed a $\log_2$ fold change >0 with a *t*-test <0.05, respectively. There were 456 shared proteins between the HEK SARS2 M and HEK MERS M data sets with a $\log_2$ fold change >0 with a *t*-test <0.05 compared to the HEK Neg. Control data set. The top 10% of proteins from each data set (63 from HEK SARS2 M, 500 from HEK MERS M, and 46 from the shared proteomic data set) were selected for further bioinformatics analysis.

Table 4-3 The number of proteins identified in proteomic analyses of pull-down samples from transfected HEK293 cells.

Data set	Replicate number	Total number of proteins quantified	Number of proteins with a $\log_2$ fold change >2 (*t*-test <0.05)	Number of proteins with a $\log_2$ fold change >1 (*t*-test <0.05)	Number of proteins with a $\log_2$ fold change >0 (*t*-test <0.05)	Number of proteins (10%) selected for enrichment analysis with the cut-off of the bottom protein	Number of proteins with a $\log_2$ fold change >0 (*t*-test <0.05) for both viruses	Number of proteins (10%) selected for enrichment analysis with the cut-off of the bottom protein
HEK SARS2 E	3	4626	250	2078	2496	250	2271	227
HEK MERS E	3	4626	536	2288	2676	268		
HEK SARS2 M	2	650	17	156	628	63	456	46
HEK MERS M	3	5113	3041	4793	4996	500		

4.2.4.2 Quantitative LC-MS/MS analysis of Dubca transfected cell lysates

Comparing Dubca transfected cells expressing SARS-CoV-2 E (Dubca SARS2 E) or MERS-CoV E (Dubca MERS E) to non-transfected Dubca cells (Dubca Neg. Control) identified 1930 protein interactors (Table 4-4). In the Dubca SARS2 E and Dubca MERS E data sets, 0 and 1 protein showed a $\log_2$ fold change >2 compared to the Dubca Neg. Control data set with a *t*-test <0.05, respectively. In addition, 14 and 18 proteins showed a $\log_2$ fold change >1 with a *t*-test <0.05, respectively. Moreover, 167 and 1208 proteins showed a $\log_2$ fold change >0 with a *t*-test <0.05, respectively. There were 137 proteins shared between the Dubca SARS2 E and Dubca MERS E data sets ($\log_2$ fold change >0 with a *t*-test <0.05). The top 10% of proteins from each data set (17 from Dubca SARS2 E, 121 from Dubca MERS E, and 137 from the shared proteomic data set) were selected for further bioinformatics analysis.

Comparing Dubca transfected cells expressing SARS-CoV-2 M (Dubca SARS2 M) or MERS-CoV M (Dubca MERS M) to non-transfected Dubca cells (Dubca Neg. Control) identified 2670 protein interactors (Table 4-4). In the Dubca SARS2 M and Dubca MERS M data sets, 66 and 203 proteins showed a $\log_2$ fold change >2 compared to the Dubca Neg. Control data set with a *t*-test <0.05, respectively. In addition, 576 and 1077 proteins showed a $\log_2$ fold change >1 with a *t*-test <0.05, respectively. Moreover, 1214 and 1911 proteins showed a $\log_2$ fold change >0 with a *t*-test <0.05, respectively. There were 1165 proteins shared between the Dubca SARS2 M and Dubca MERS M data sets ($\log_2$ fold change >0 with a *t*-test <0.05). The top 10% of proteins from each data set (122 from Dubca SARS2 M, 192 from Dubca MERS M, and 116 from the shared proteomic data set) were selected for further bioinformatics analysis.

Table 4-4 The number of proteins identified in proteomic analyses of pull-down samples from transfected Dubca cells.

Data set	Replicate number	Total number of proteins quantified	Number of proteins with a $\log_2$ fold change >2 (*t*-test <0.05)	Number of proteins with a $\log_2$ fold change >1 (*t*-test <0.05)	Number of proteins with a $\log_2$ fold change >0 (*t*-test <0.05)	Number of proteins (10%) selected for enrichment analysis with the cut-off of the bottom protein	Number of proteins with a $\log_2$ fold >0 (*t*-test <0.05) for both viruses	Number of proteins (10%) selected for enrichment analysis with the cut-off of the bottom protein
Dubca SARS2 E	2	1930	0	14	167	17	137	137
Dubca MERS E	3	1930	1	18	1208	121		
Dubca SARS2 M	3	2670	66	576	1214	122	1165	117
Dubca MERS M	3	2670	203	1077	1911	192		

*4.2.4.3 Quantitative LC-MS/MS analysis of **PaKiT** transfected cell lysates*

Comparing PaKiT cells expressing SARS-CoV-2 E (PaKiT SARS2 E) or MERS-CoV E (PaKiT MERS E) to non-transfected PaKiT cells (PaKiT Neg. Control) identified 2488 protein interactors (Table 4-5). In the PaKiT SARS2 E and PaKiT MERS E data sets, 14 and 35 proteins showed a $\log_2$ fold change >2 compared to the PaKiT Neg. Control data set with a *t*-test <0.05, respectively. In addition, 319 and 1295 proteins showed a $\log_2$ fold change >1 with a *t*-test <0.05, respectively. Moreover, 940 and 2072 proteins showed a $\log_2$ fold change >0 with a *t*-test <0.05, respectively. There were 930 proteins shared between the PaKiT SARS2 E and PaKiT MERS E data sets ($\log_2$ fold change >0 with a *t*-test <0.05). The top 10% of proteins from each data set (94 from PaKiT SARS2 E, 207 from PaKiT MERS E, and 93 from the shared proteomic data set) were selected for further bioinformatics analysis.

Comparing PaKiT transfected cells expressing SARS-CoV-2 M (PaKiT SARS2 M) and MERS-CoV M (PaKiT MERS M) to non-transfected PaKiT cells (PaKiT Neg. Control) identified 3480 proteins interactors (Table 4-5). In the PaKiT SARS2 M and PaKiT MERS M data sets, 54 and 222 proteins showed a $\log_2$ fold change >2 compared to the PaKiT Neg. Control data set with a *t*-test <0.05, respectively. In addition, 129 and 1384 proteins showed a $\log_2$ fold change >1 with a *t*-test <0.05, respectively. Moreover, 132 and 1624 proteins showed a $\log_2$ fold change >0 with a *t*-test <0.05, respectively. There were 127 proteins shared between the PaKiT SARS2 M and PaKiT MERS M data sets ($\log_2$ fold change >0 with a *t*-test <0.05). The top 10% of proteins from each data set (54 from PaKiT SARS2 M, 162 from PaKiT MERS M, and 127 from the shared proteomic data set) were selected for further bioinformatics analysis.

Table 4-5 The number of proteins identified in proteomic analyses of pull-down samples from transfected PaKiT cells.

Data set	Replicate number	Total number of proteins quantified	Number of proteins with a $\log_2$ fold change >2 (*t*-test <0.05)	Number of proteins with a $\log_2$ fold change >1 (*t*-test <0.05)	Number of proteins with a $\log_2$ fold change >0 (*t*-test <0.05)	Number of proteins (10%) selected for enrichment analysis with the cut-off of the bottom protein	Number of proteins with a $\log_2$ fold change >0 (*t*-test <0.05) for both viruses	Number of proteins (10%) selected for enrichment analysis with the cut-off of the bottom protein
PaKiT SARS2 E	3	2488	14	319	940	94	930	93
PaKiT MERS E	3	2488	35	1295	2072	207		
PaKiT SARS M	3	3480	54	129	132	132	127	127
PaKiT MERS M	3	3480	222	1384	1624	162		

4.2.5 Protein selection from among the high-confidence interaction partners for SARS-CoV-2 and MERS-CoV E and M proteins

It is important to select promising proteins for validation and determine whether they are important in the viral life cycle. The first comparison level included cellular proteins showing significant increases (log_2 fold change >0 with a *t*-test <0.05) in co-IPs with the MERS-CoV and SARS-CoV-2 E proteins compared to the negative control. There were 2271, 137, and 930 proteins commonly interacting with the MERS-CoV and SARS-CoV E proteins in HEK293, Dubca and PaKiT cells, respectively. A similar comparison included cellular proteins showing significant increases (log_2 fold change >0 with a *t*-test <0.05) in co-IPs with the MERS-CoV and SARS-CoV-2 M proteins compared to the negative control. There were 456, 1165, and 127 proteins commonly interacting with the MERS-CoV and SARS-CoV E proteins in HEK293, Dubca, and PaKiT cells, respectively (Figure 4-13).

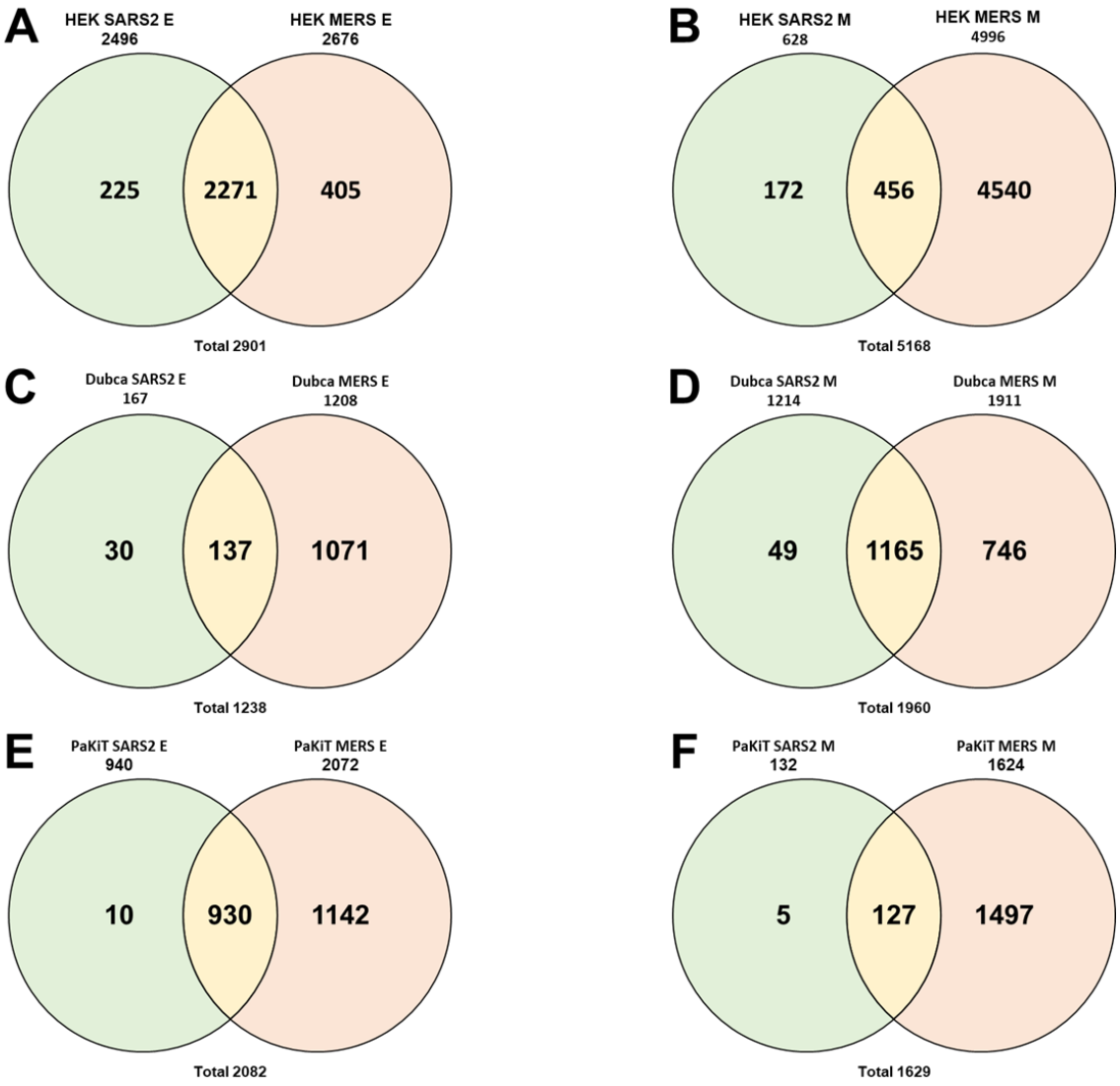


Figure 4-13 Cellular interaction partners common to the MERS-CoV and SARS-CoV-2 E and M proteins in HEK293, Dubca, and PaKiT cells.

Venn diagrams show the number of cellular proteins uniquely and commonly interacting with the MERS-CoV and SARS-CoV-2 E and M proteins compared to the negative control (log_2 fold change >0 with a *t*-test <0.05).

4.2.5.1 *Comparison of the proteomic analyses for SARS-CoV-2 and MERS-CoV E proteins in the three transfected host cell types:*

The second comparison level involved studying and comparing protein data sets between different cell types. The comparison included proteins showing significant increases in the HEK MERS E and HEK SARS2 E (HEK E), Dubca MERS E and Dubca SARS2 E (Dubca E), and PaKiT MERS E and PaKiT SARS2 E (PaKiT E) data sets. The Fun Rich software was used to mark shared proteins between the three cellular data sets (HEK E, Dubca E, and PaKiT E; Figure 4-14). The results showed that 379 proteins commonly interacted with the SARS-CoV-2 and MERS-CoV E proteins in HEK and PaKiT cells, 40 in HEK and Dubca cells, and six in Dubca and PaKiT cells. Only 21 proteins commonly interacted with the SARS-CoV-2 and MERS-CoV E proteins in all three cell lines, which are summarised in Table 4-6 and Table 4-7. The lowest $\log_2$ fold change (0.17) was for nucleoporin 35 (NUP35) in the Dubca SARS2 E data set, and the highest $\log_2$ fold change (3.85) was for the spliceosome-associated protein homolog (CWC22) in the HEK SARS2 E data set.

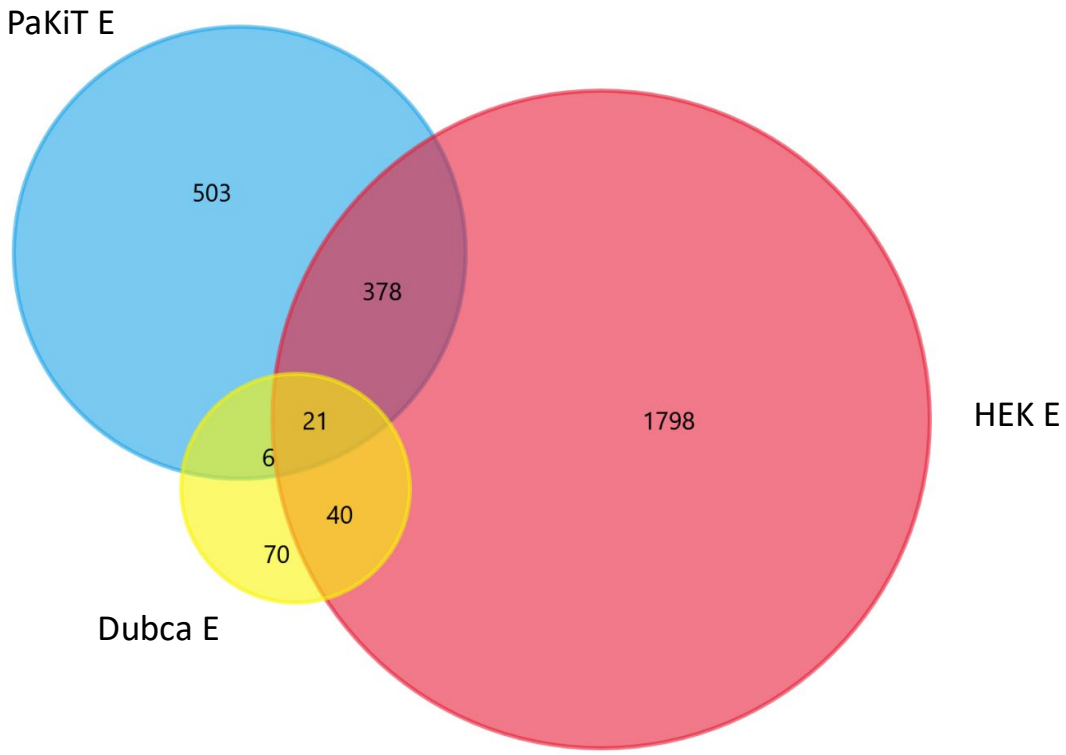


Figure 4-14 Proteins commonly interacting with the MERS-CoV and SARS-CoV-2 E proteins in HEK293, PaKiT, and Dubca cells.

The Venn diagram shows the number of shared cellular proteins interacting with the E protein in HEK293, PaKiT, and Dubca cells compared to the negative control (log_2 fold change >0 with a *t*-test <0.05). The Fun Rich software was used to construct the Venn diagram.

Table 4-6 Shared cellular proteins showing significant increases in co-IPs from lysates of HEK293, PaKiT, and Dubca cells expressing the MERS-CoV and SARS-CoV-2 E proteins compared to negative control cells.

Uniprot ID	Protein symbol	Protein name
P24752	ACAT1	Acetyl-CoA acetyltransferase, mitochondrial
P14923	JUP	Junction plakoglobin
Q9NXW2	DNAJB12	DnaJ homolog subfamily B member 12
P23258	TUBG1	Tubulin gamma-1 chain (Gamma-1-tubulin)
O14735	CDIPT	CDP-diacylglycerol—inositol 3 phosphatidyltransferase
O75027	ABCB7	Iron-sulfur clusters transporter
Q9H307	PNN	Pinin
Q9H3U1	UNC45A	Protein unc-45 homolog A
Q14254	FLOT2	Flotillin-2
O00422	SAP18	Histone deacetylase complex subunit
Q9Y512	SAMM50	Sorting and assembly machinery component 50
P12236	SLC25A6	ADP/ATP translocase 3
Q8NFH5	NUP35	Nucleoporin
Q9BRX2	PELO	Protein pelota
Q9NZ01	TECR	Very-long-chain enoyl-CoA reductase
O95470	SGPL1	Sphingosine-1-phosphate lyase 1
Q15650	TRIP4	Activating signal cointegrator 1
Q96I25	RBM17	Splicing factor 45
Q9HCG8	CWC22	Pre-mRNA-splicing factor
Q14498	RBM39	RNA-binding protein 39
P62987	UBA52	Ubiquitin-60S ribosomal protein L40

Table 4-7 Log_2 fold changes of shared cellular proteins showing significant increases in co-IPs from lysates of HEK293, PaKiT, and Dubca cells expressing the MERS-CoV and SARS-CoV-2 E proteins compared to negative control cells.

Protein symbol	Log_2 HEK MERS E	Log_2 HEK SARS2 E	Log_2 PaKiT MERS E	Log_2 PaKiT SARS2 E	Log_2 Dubca MERS E	Log_2 Dubca SARS2 E
ACAT1	1.53	1.63	1.38	0.95	0.54	0.43
JUP	1.09	0.77	1.35	0.75	0.66	0.62
DNAJB12	2.84	1.49	1.17	1.05	0.85	0.75
TUBG1	1.32	1.50	1.16	0.82	0.46	0.34
CDIPT	1.77	1.64	1.14	0.82	0.52	0.43
ABCB7	1.59	1.58	1.10	0.85	0.67	0.65
PNN	1.19	3.10	1.09	1.01	0.62	1.11
UNC45A	1.52	1.44	1.07	0.83	0.46	0.51
FLOT2	1.44	1.14	1.06	0.86	0.68	0.72
SAP18	1.10	2.31	1.05	0.71	0.42	1.03
SAMM50	3.17	2.65	1.02	0.64	0.44	0.33
SLC25A6	1.42	1.75	1.00	0.59	0.58	0.39
NUP35	2.29	1.95	0.96	0.69	0.33	0.17
PELO	1.61	1.69	0.87	0.59	0.46	0.38
TECR	1.95	2.95	0.85	0.73	0.52	0.69
SGPL1	1.15	1.26	0.82	0.67	0.31	0.48
TRIP4	0.96	0.96	0.81	0.60	0.38	0.37
RBM17	1.10	0.89	0.80	0.77	0.52	0.50
CWC22	1.59	3.85	0.76	0.59	1.01	1.03

RBM39	1.53	1.27	0.69	0.46	0.26	0.24
UBA52	1.21	1.39	0.36	0.37	0.61	1.15

4.2.5.2 *Comparison of the proteomic analyses for SARS-CoV-2 M and MERS-CoV-M proteins in the three transfected host cell types:*

The second comparison level involved studying and comparing protein data sets between different cells. The comparison included proteins showing significant increases in the HEK SARS2 M and HEK MERS M (HEK M), Dubca SARS2 M and Dubca MERS M (Dubca M), and PaKiT SARS2 M and PaKiT MERS M (PaKiT M) data sets. The Fun Rich software was used to mark shared proteins between the three cellular data sets (HEK M, Dubca M, and PaKiT M; Figure 4-15). The results showed that one protein commonly interacted with the SARS-CoV-2 and MERS-CoV M proteins in HEK and PaKiT cells, 156 in HEK and Dubca cells, and 76 in Dubca and PaKiT cells. Only 12 proteins commonly interacted with the SARS-CoV-2 and MERS-CoV M proteins in all three cell lines, which are summarised in Table 4-8 and Table 4-9. The lowest $\log_2$ fold change (0.4) was for ubiquitin-specific peptidase 34 (USP34) in the HEK MERS M data set, and the highest $\log_2$ fold change (3.77) was for SLC38A10 in the HEK MERS M data set. Only one cellular protein was a shared interactor with SARS-CoV-2 and MERS-CoV E and M proteins in HEK293, Dubca, and PaKiT cells: sorting and assembly machinery component 50 (SAMM50).

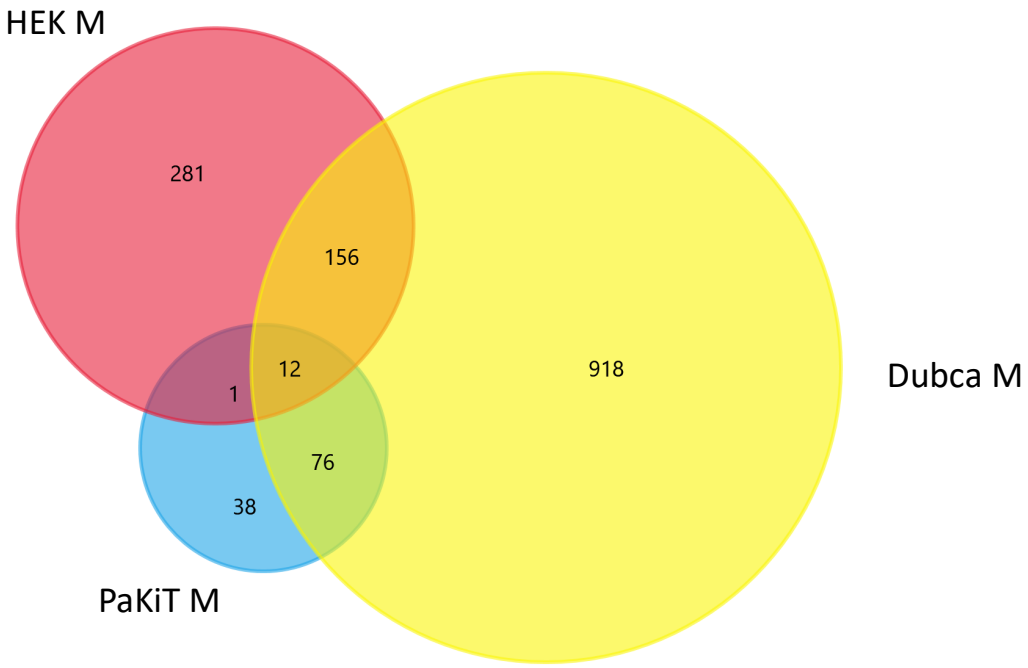


Figure 4-15 Proteins commonly interacting with the MERS-CoV and SARS-CoV-2 M proteins in HEK293, PaKiT, and Dubca cells.

The Venn diagram shows the number of shared cellular proteins interacting with the M protein in HEK293, PaKiT, and Dubca cells compared to the negative control (log_2 fold change >0 with a *t*-test <0.05). The Fun Rich software was used to construct the Venn diagram.

Table 4-8 Shared cellular proteins showing significant increases in co-IPs from lysates of HEK293, PaKiT, and Dubca cells expressing the MERS-CoV and SARS-CoV-2 M proteins compared to negative control cells.

Uniprot ID	Protein symbol	Protein name
O75844	ZMPSTE24	CAAX prenyl protease 1 homolog
Q16891	IMMT	MICOS complex subunit MIC60
Q15758	SLC1A5	Neutral amino acid transporter B
Q9HBR0	SLC38A10	Putative sodium-coupled neutral amino acid transporter 10
P55060	CSE1L	Exportin-2
Q9Y512	SAMM50	Sorting and assembly machinery component 50
Q9UI26	IPO11	Importin-11
Q13409	DYNC1I2	Cytoplasmic dynein 1 intermediate chain 2
Q5U0I6	RAB1A	H.sapiens ras-related Hrab1A protein
Q70CQ2	USP34	Ubiquitin carboxyl-terminal hydrolase 34
A0A024RD35	ELOVL5	Elongation of very long chain fatty acids protein 5
Q96ER3	SAAL1	Protein SAAL1

Table 4-9 Log_2 fold changes of shared cellular proteins showing significant increases in co-IPs from lysates of HEK293, PaKiT, and Dubca cells expressing the MERS-CoV and SARS-CoV-2 M proteins compared to negative control cells.

Protein symbol	Log_2 HEK MERS M	Log_2 HEK SARS2 M	Log_2 PaKiT MERS M	Log_2 PaKiT SARS2 M	Log_2 Dubca MERS M	Log_2 Dubca SARS2 M
ZMPSTE24	3.53	2.18	2.73	2.22	2.39	2.08
IMMT	1.62	1.19	2.70	1.91	2.08	1.86
SLC1A5	3.76	2.10	2.70	2.31	2.33	1.80
SLC38A10	3.77	0.76	2.68	2.30	2.23	1.86
CSE1L	3.45	1.95	2.58	2.40	2.55	2.00
SAMM50	1.47	1.38	2.50	1.71	1.83	1.55
IPO11	3.25	1.61	2.36	1.83	2.12	1.63
DYNC1I2	2.31	0.41	2.32	1.74	1.91	1.58
RAB1A	2.50	0.39	2.31	1.79	1.08	0.83
USP34	2.87	0.40	2.29	1.88	1.55	1.12
ELOVL5	3.22	1.30	2.28	1.83	1.90	1.40
SAAL1	3.14	2.05	1.95	1.90	2.07	1.54

4.2.6 Bioinformatic analysis of proteomic data sets

The data sets described in Section 4.2.4, containing the top 10% of proteins significantly increased in co-IPs using lysates from HEK293, Dubca, and PaKiT cells expressing the MERS-CoV and SARS-CoV-2 E and M proteins (Table 4-2) compared to the negative control cells were subjected to further downstream bioinformatic analysis. Appendix F list proteins that significantly increased in anti-FLAG co-IPs using lysates from HEK293, Dubca, and PaKiT transfected cells with SARS-CoV-2 E, MERS-CoV E, SARS-CoV-2 M, and MERS-CoV M proteins compared to a negative non-transfected control. If the 10% was less than 20 proteins, proteins that significantly increased (>0 $\log_2$ fold change with a *t*-test < 0.05) was selected. The top 10% was only 13 proteins in PaKiT SARS2 M data set, while the significantly increased proteins were 132 proteins, which was selected. Moreover, the top 10% was less than 20 proteins in Dubca E and PaKiT M data sets, while the significantly increased proteins were 137, and 127, respectively, which were selected.

The high-confidence interaction partners (top 10% of significantly increased proteins) were subject to gene ontology (G.O.) enrichment analysis using DAVID (161) to generate functionally related groups of enriched proteins and examine their significance using a modified Fisher's exact test (EASE score) (Appendix F). An EASE score of 1.3 was set as the threshold for significance. The KEGG category of the DAVID pathway viewer was used to determine which cellular pathways were significantly altered. Moreover, analysing data by DAVID was used to reveal protein group enrichment and proteins involved in specific host pathways that could be important for viral replication. The STRING database was used to generate interaction networks to determine whether there were virus or cell type specific differences for the E and M protein interactomes. The outcomes of these analyses are described below.

4.2.6.1 *Proteins significantly increased in anti-FLAG co-IPs using lysates from HEK SARS2 E and HEK MERS E transfected cells compared to HEK Neg control*

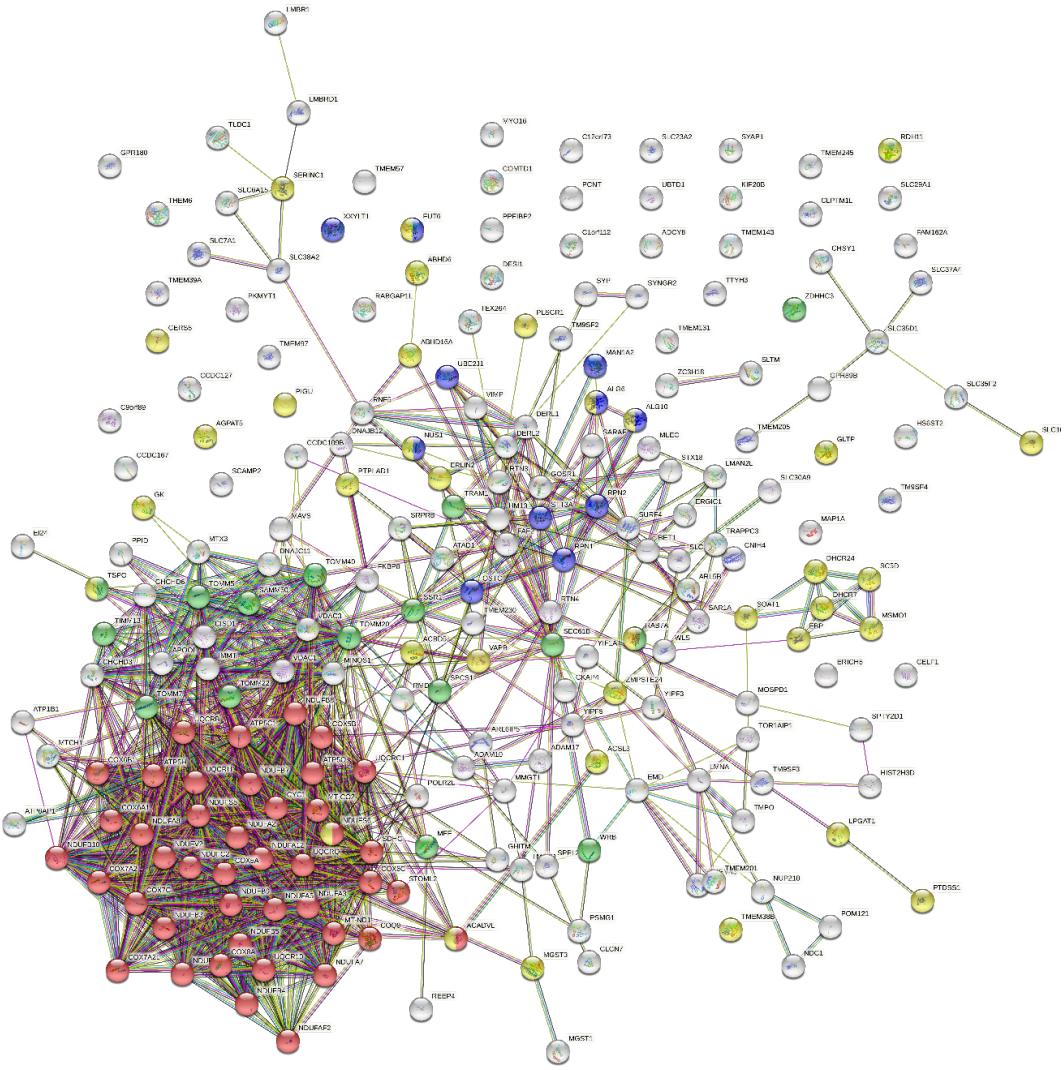
DAVID analysis of the high-confidence interaction partners for HEK SARS2 E and HEK MERS E revealed significant enrichment of 15 annotation clusters with an EASE score >1.3 (Table 4-10). Five enrichment clusters were associated with mitochondrial function: “mitochondrion inner membrane”, “oxidative phosphorylation”, “cellular respiration”, “inner mitochondrial membrane organization”, and “mitochondrial intermembrane space”. Other clusters contain proteins associated with annotation terms such as “membrane”, “protein transmembrane transporter activity” “ER to Golgi vesicle-mediated transport”, “ER-Golgi transport”, and “endoplasmic reticulum unfolded protein response”. Moreover, two enrichment clusters were identified “protein N-linked glycosylation” and “lipid biosynthesis” (proteins under this enrichment cluster were an endoplasmic reticulum membrane protein). STRING analysis of the high-confidence interaction partners for HEK SARS2 E and HEK MERS E showed two significant enrichment GO terms associated with mitochondrial function (Figure 4-16).

DAVID analysis of the high-confidence interaction partners for HEK SARS2 E or HEK MERS E proteins separately revealed significant enrichment of 19 and 18 annotation clusters with an EASE score >1.3, respectively (Table 4-11 & Table 4-12). One unique enrichment cluster was revealed for HEK SARS2 E “Iron-sulfur” (proteins under this enrichment cluster were also associated with the term mitochondrial proteins). Moreover, one unique enrichment cluster was revealed for HEK MERS E “ubiquitin-specific protease binding” (this cluster contains multifunctional proteins also associated with the terms “unfolded protein response” and “protein transportation to cytosol”). STRING analysis of the high-confidence interaction partners for HEK SARS2 E or HEK MERS E proteins separately shown one unique significant enrichment GO terms for HEK SARS2 E “ubiquitin-dependent protein catabolic process” and one unique significant enrichment GO terms for HEK MERS E "protein n-linked glycosylation” (Figure 4-17 & Figure 4-18).

Table 4-10 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV and SARS-CoV-2 E proteins in HEK293 cells

Term	EASE Score	Category	Count	%	*p*-value
KW-0472~Membrane	41.1	UP keywords	209	88.2	1.48E-45
KW-0999~Mitochondrion inner membrane	27.2	UP keywords	54	22.8	9.94E-43
hsa00190: Oxidative phosphorylation	15.1	KEGG pathway	43	18.1	5.76E-45
GO:0045333~cellular respiration	11.4	GOBP	17	7.2	4.44E-21
GO:0007007~inner mitochondrial membrane organization	8.1	GOBP	9	3.8	3.08E-11
GO:0008320~protein transmembrane transporter activity	4.9	GOMF	4	1.7	0.002170636
GO:0006888~ER to Golgi vesicle-mediated transport	3.7	GOBP	11	4.6	2.70E-06
GO:0006487~protein N-linked glycosylation	3.3	GOBP	8	3.4	2.42E-06
KW-0444~Lipid biosynthesis	3.2	UP keywords	11	4.6	1.06E-04
GO:0005758~mitochondrial intermembrane space	3.0	GOCC	6	2.5	0.003603006
KW-0931~ER-Golgi transport	3.0	UP keywords	11	4.6	6.61E-07
GO:0030968~endoplasmic reticulum unfolded protein response	2.6	GOBP	4	1.7	0.022081401
GO:0046933~proton-transporting ATP synthase activity, rotational mechanism	2.3	GOMF	4	1.7	0.001191216
hsa04020: Calcium Signaling pathway	2.3	KEGG pathway	4	1.7	0.737084854
GO:0008654~phospholipid biosynthetic process	1.3	GOBP	4	1.7	0.007540261

There were 15 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term. EASE: Fisher Exact Statistics in DAVID system.



GO term	Description	Count	FDR
GO:0006091JG156:J156	Generation of precursor metabolites and energy	43 of 405	2.36E-24
GO:0006486	Protein glycosylation	11 of 240	0.017
GO:0006605	Protein targeting	16 of 356	0.00091
GO:0006629	Lipid metabolic process	35 of 1190	8.47E-05

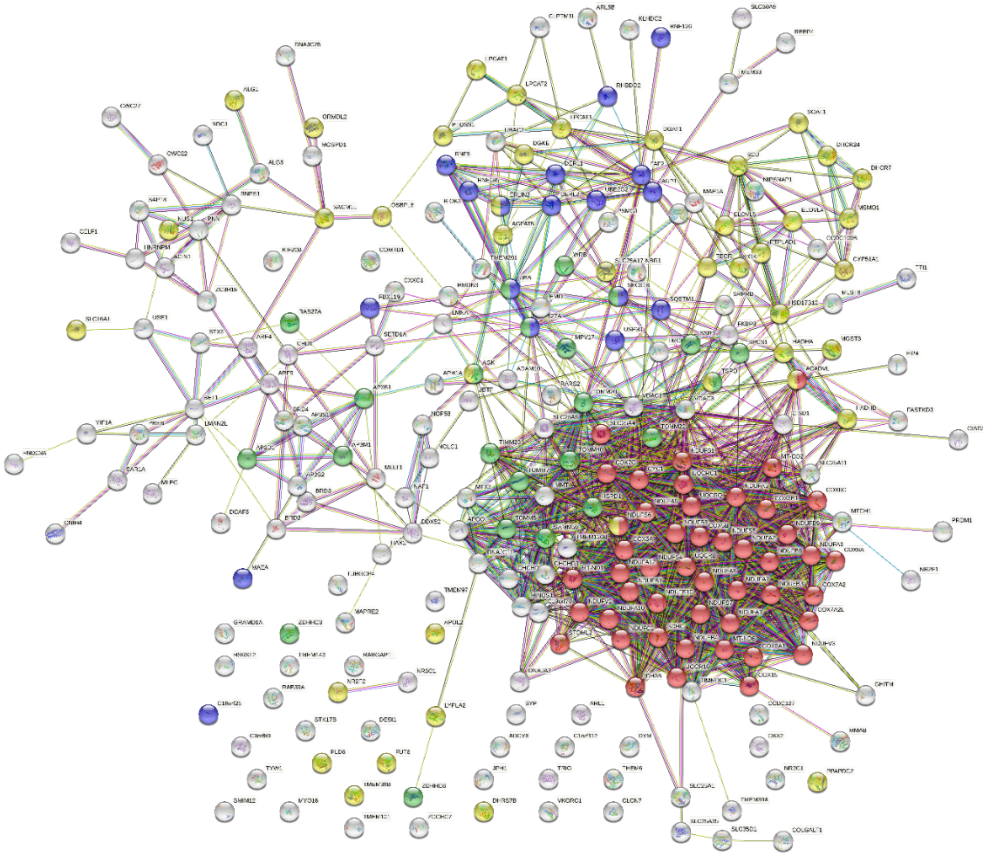
Figure 4-16 STRING analysis of the high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 E proteins in HEK293 cells.

High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 E proteins in HEK293 cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-11 DAVID analysis of the high-confidence interaction partners identified for the SARS-CoV-2 E protein in HEK293 cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0496~Mitochondrion	35.4	UP keywords	103	39.8	9.96E-50
KW-0256~Endoplasmic reticulum	20.5	UP keywords	71	27.4	5.47E-23
KW-0812~Transmembrane	17.5	UP keywords	148	57.1	4.16E-17
GO:0031966~mitochondrial membrane	13.4	GOCC	25	9.7	1.97E-19
GO:0007007~inner mitochondrial membrane organization	9.8	GOBP	10	3.9	1.29808E-12
GO:0006119~oxidative phosphorylation	7.1	GOBP	6	2.3	3.48E-06
GO:0008320~protein transmembrane transporter activity	4.8	GOMF	5	1.9	1.85E-04
GO:0030970~retrograde protein transport, ER to cytosol	4.8	GOBP	5	1.9	3.33E-05
hsa04141: Protein processing in endoplasmic reticulum	3.1	KEGG pathway	9	3.5	0.009856376
hsa01212: Fatty acid metabolism	2.8	KEGG pathway	9	3.5	5.57E-06
hsa04142: Lysosome	2.7	KEGG pathway	5	1.9	0.197812877
KW-0594~Phospholipid biosynthesis	2.5	UP keywords	5	1.9	0.005890892
KW-0411~Iron-sulfur	2.4	UP keywords	6	2.3	0.002252819
GO:0006695~cholesterol biosynthetic process	2.3	GOBP	5	1.9	0.001289059
GO:1990544~mitochondrial ATP transmembrane transport	2.3	GOBP	3	1.2	0.002413989
hsa04979: Cholesterol metabolism	2.1	KEGG pathway	5	1.9	0.010915376
GO:0030968~endoplasmic reticulum unfolded protein response	1.9	GOBP	3	1.2	0.145374564
KW-0449~Lipoprotein	1.5	UP keywords	18	6.9	0.123570483
GO:0006888~ER to Golgi vesicle-mediated transport	1.4	GOBP	7	2.7	0.006931548

There were 19 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), Uniprot (UP) keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0006091	Generation of precursor metabolites and energy	47 of 405	1.71E-26
GO:0006511	Ubiquitin-dependent protein catabolic process	18 of 535	0.0249
GO:0006605	Protein targeting	23 of 356	2.29E-07
GO:0006629	Lipid metabolic process	42 of 1190	1.16E-06

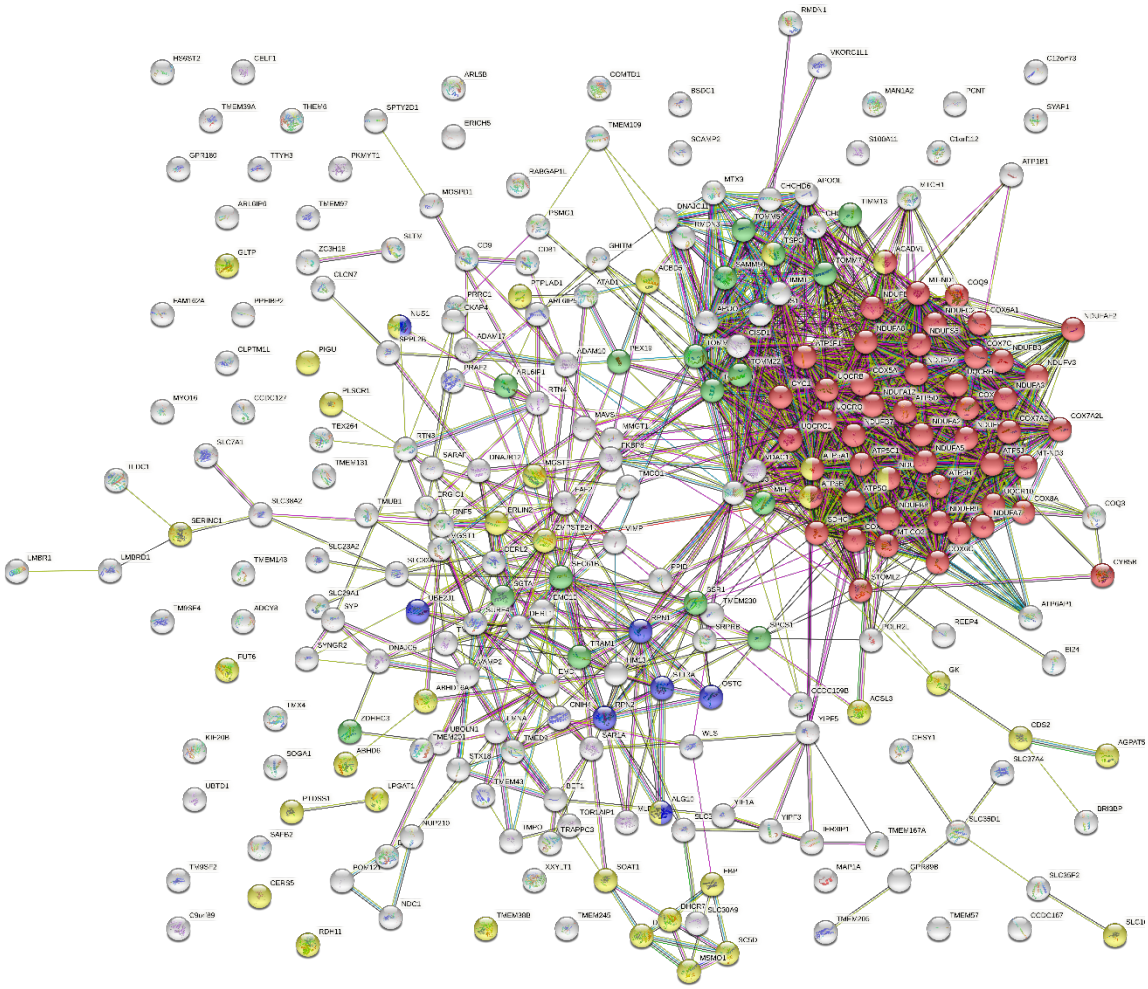
Figure 4-17 STRING analysis of the high-confidence interaction partners identified for the SARS-CoV-2 E protein in HEK293 cells.

High-confidence interaction partners identified for the SARS-CoV-2 E protein in HEK293 cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-12 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV E protein in HEK293 cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0472~Membrane	45.1	UP keywords	238	87.8	8.05E-52
KW-0256~Endoplasmic reticulum	33.8	UP keywords	95	35.1	1.27E-40
hsa00190: Oxidative phosphorylation	29.4	KEGG pathway	48	17.7	8.29E-51
GO:0031966~mitochondrial membrane	13.4	GOCC	26	9.6	2.83E-20
GO:0045333~cellular respiration	10.9	GOBP	17	6.3	3.25E-20
GO:0007007~inner mitochondrial membrane organization	9.7	GOBP	10	3.7	1.40E-12
GO:0046933~proton-transporting ATP synthase activity, rotational mechanism	5.5	GOMF	9	3.3	4.10E-11
GO:0008320~protein transmembrane transporter activity	4.6	GOMF	4	1.5	0.003168007
GO:0006888~ER to Golgi vesicle-mediated transport	3.9	GOBP	13	4.8	1.37E-07
GO:0006487~protein N-linked glycosylation	3.1	GOBP	8	3.0	5.44E-06
KW-0444~Lipid biosynthesis	3.0	UP keywords	12	4.4	7.98E-05
GO:0005758~mitochondrial intermembrane space	2.8	GOCC	6	2.2	0.006294456
GO:0061025~membrane fusion	2.7	GOBP	3	1.1	0.103812224
GO:1990381~ubiquitin-specific protease binding	2.4	GOMF	3	1.1	0.03268307
GO:0009055~electron carrier activity	2.2	GOMF	8	3.0	4.60E-05
GO:0015485~cholesterol binding	2.1	GOMF	7	2.6	7.22E-05
KW-0594~Phospholipid biosynthesis	1.6	UP keywords	5	1.8	0.005890892
GO:0015078~hydrogen ion transmembrane transporter activity	1.6	GOMF	3	1.1	0.073944903

There were 18 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0006091	Generation of precursor metabolites and energy	50 of 405	4.86E-29
GO:0006487	Protein n-linked glycosylation	7 of 77	1.13E-02
GO:0006605	Protein targeting	17 of 356	9.00E-04
GO:0006629	Lipid metabolic process	35 of 1190	8.90E-04

Figure 4-18 STRING analysis of the high-confidence interaction partners identified for the MERS-CoV E protein in HEK293 cells.

High-confidence interaction partners identified for the MERS-CoV E proteins in HEK293 cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.6.2 *Proteins significantly increased in anti-FLAG co-IPs using lysates from HEK SARS2 M and HEK MERS M transfected cells compared to HEK Neg control*

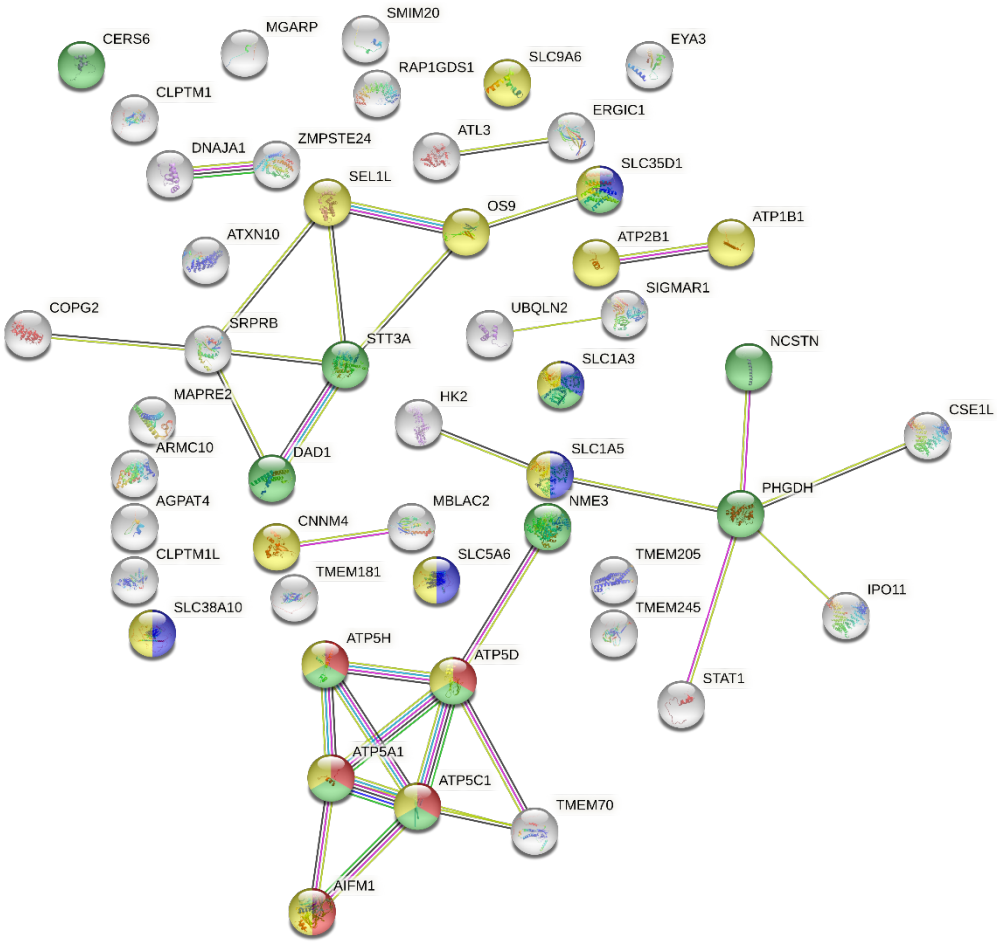
DAVID analysis of the high-confidence interaction partners for HEK SARS2 M and HEK MERS M revealed significant enrichment of 5 annotation clusters with an EASE score >1.3 (Table 4-13). Two enrichment clusters were associated with membrane function: “membrane”, and “integral component of plasma membrane”. Other clusters contain proteins associated with the term’s “mitochondrion” and “intracellular protein transport” (proteins in this cluster were also associated with the term “nucleocytoplasmic carrier activity proteins”, whilst the “sodium transport”, cluster includes proteins that play a role in “ion transmembrane transporter activity”. STRING analysis of the high-confidence interaction partners for HEK SARS2 M and HEK MERS M identified three significant enrichment GO terms associated with “membrane" (Figure **4-19**). While the fourth term was “Organonitrogen compound biosynthetic process”.

DAVID analysis of the high-confidence interaction partners for HEK SARS2 M or HEK MERS M proteins separately revealed significant enrichment of 5 and 24 annotation clusters with an EASE score >1.3, respectively (Table 4-14 & Table 4-15). Unique enrichment clusters revealed for HEK MERS M included “importin-beta, N-terminal” (proteins under this enrichment cluster were also associated with the term “nucleocytoplasmic carrier activity”), “Glycan biosynthesis” (identified an important protein for glycosylation), “lipid metabolism”, “peptidase activity”, “intercellular bridge”, “ATPase activity”, “sugar transporter, conserved site” (proteins identified in this enrichment cluster were also associated with the term “transmembrane transporter activity”). STRING analysis of the high-confidence interaction partners for HEK SARS2 M or HEK MERS M proteins separately showed no unique enrichment clusters (Figure 4-20 & Figure 4-21).

Table 4-13 DAVID analysis of the high-confidence interaction partners identified for MERS-CoV and SARS-CoV-2 M protein in HEK293 cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0472~Membrane	5.2	UP keywords	40	80	3.05E-07
GO:0005739~mitochondrion	3.3	GOCC	14	28	2.30E-05
KW-0739~Sodium transport	2.8	UP keywords	4	8	0.006277694
GO:0005887~integral component of plasma membrane	2.1	GOCC	6	12	0.292234013
GO:0006886~intracellular protein transport	2.1	GOBP	3	6	0.184526061

There were 5 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:1990542	Mitochondrial transmembrane transport	5 of 95	0.0028
GO:1905039	Carboxylic acid transmembrane transport	5 of 189	0.0329
GO:1901566	Organonitrogen compound biosynthetic process	12 of 1346	0.0276
GO:0055085	Transmembrane transport	16 of 1314	0.00093

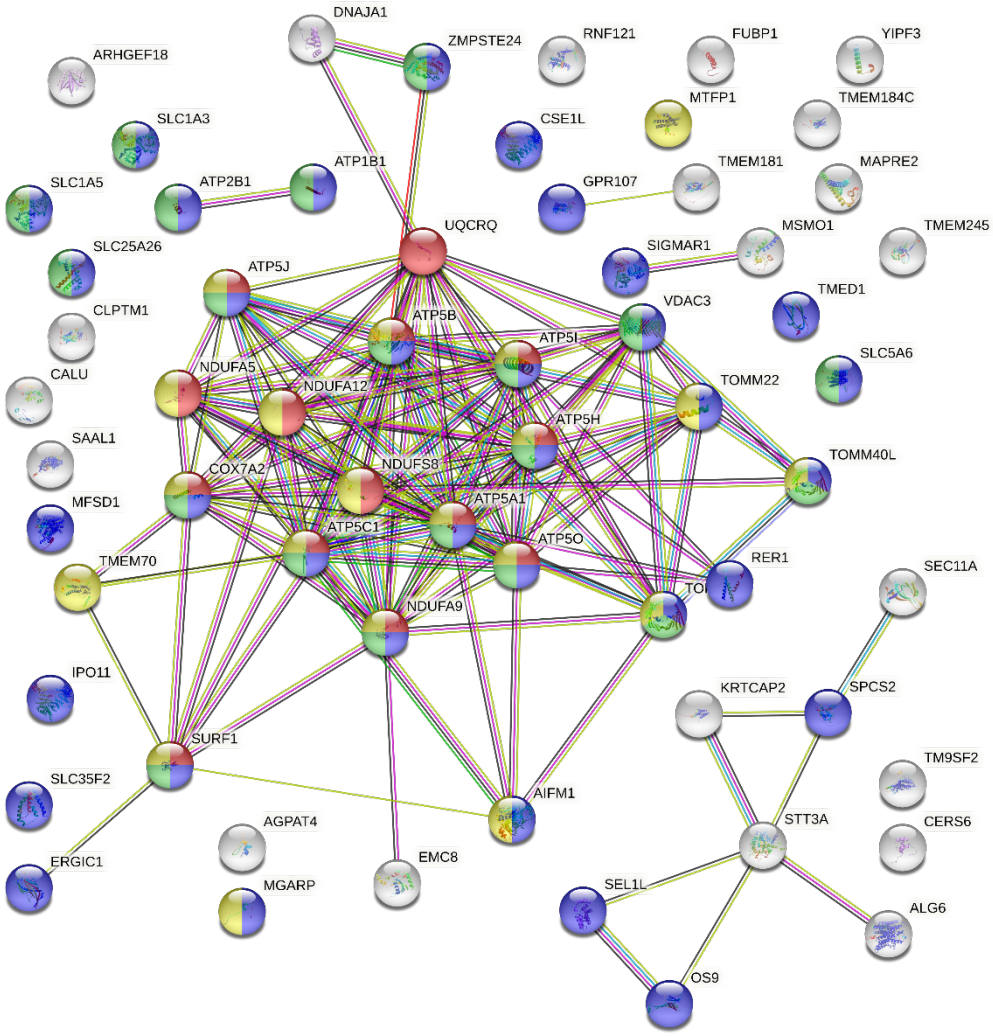
Figure 4-19 STRING analysis of the high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in HEK293 cells.

High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in HEK293 cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-14 DAVID analysis of the high-confidence interaction partners identified for the SARS-CoV-2 M protein in HEK293 cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0812~Transmembrane	7.1	UP keywords	36	75.0	6.09E-08
KW-0496~Mitochondrion	5.5	UP keywords	17	35.4	1.26E-07
KW-0256~Endoplasmic reticulum	3.6	UP keywords	14	29.2	2.49E-05
KW-1003~Cell membrane	1.4	UP keywords	8	16.7	0.914716772
GO:0016491~oxidoreductase activity	1.4	GOMF	4	8.3	0.018697423

There were 5 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0006119	Oxidative phosphorylation	14 of 118	5.43E-14
GO:0006810	Transport	35 of 4353	3.13E-06
GO:0006811	Ion transport	20 of 1344	2.02E-06
GO:0007005	Mitochondrion organization	20 of 452	6.54E-14

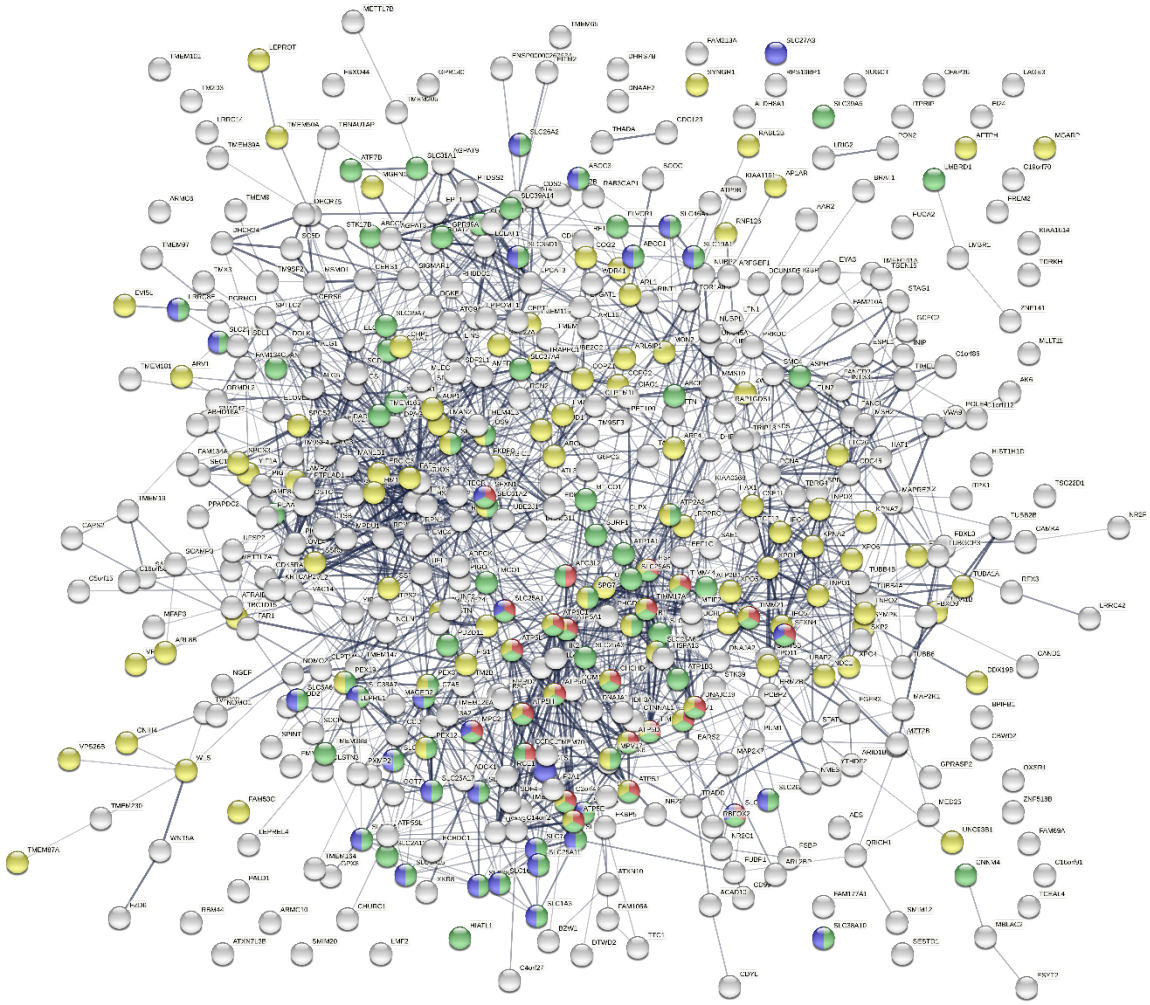
Figure 4-20 STRING analysis of the high-confidence interaction partners identified for the SARS-CoV-2 M proteins in HEK293 cells.

High-confidence interaction partners identified for the SARS-CoV-2 M proteins in HEK293 cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-15 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV M protein in HEK293 cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0472~Membrane	28.0	UP keywords	318	64.9	2.49E-21
GO:0005739~mitochondrion	7.7	GOCC	76	15.5	6.08373E-11
GO:0006606~protein import into nucleus	6.7	GOBP	8	1.6	0.011780725
IPR001494:Importin-beta, N-terminal	6.6	INTERPRO	10	2.0	5.65E-11
hsa03013:Nucleocytoplasmic transport	5.8	KEGG pathway	17	3.5	4.82E-08
hsa00510:N-Glycan biosynthesis	5.3	KEGG pathway	11	2.2	9.99E-07
KW-0443~Lipid metabolism	4.7	UP keywords	40	8.2	5.95E-05
GO:0006890~retrograde vesicle-mediated transport, Golgi to ER	4.3	GOBP	11	2.2	2.92E-07
GO:0015986~ATP synthesis coupled proton transport	4.0	GOBP	11	2.2	3.22E-10
GO:0009925~basal plasma membrane	3.6	GOCC	7	1.4	0.001225165
KW-0029~Amino-acid transport	3.6	UP keywords	12	2.4	1.47E-07
KW-0739~Sodium transport	3.5	UP keywords	11	2.2	0.001757372
KW-0276~Fatty acid metabolism	3.4	UP keywords	10	2.0	0.033901915
GO:0005744~mitochondrial inner membrane presequence translocase complex	3.2	GOCC	6	1.2	2.52E-05
GO:0015171~amino acid transmembrane transporter activity	2.6	GOMF	8	1.6	5.54E-05
GO:0008233~peptidase activity	2.5	GOMF	9	1.8	0.006000808
GO:0005471~ATP:ADP antiporter activity	2.5	GOMF	4	0.8	2.77E-04
GO:0005313~L-glutamate transmembrane transporter activity	2.3	GOMF	4	0.8	0.003482259
hsa00565:Ether lipid metabolism	2.2	KEGG pathway	4	0.8	0.159495448
GO:0015175~neutral amino acid transmembrane transporter activity	2.1	GOMF	6	1.2	4.19E-05
GO:0005778~peroxisomal membrane	2.1	GOCC	11	2.2	4.16E-06
GO:0045171~intercellular bridge	2.0	GOCC	6	1.2	0.066608833
IPR005829:Sugar transporter, conserved site	2.0	INTERPRO	4	0.8	0.045525513
GO:0016887~ATPase activity	1.9	GOMF	17	3.5	0.013260908

There were 24 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:1990542	Mitochondrial transmembrane transport	24 of 95	3.86E-12
GO:0055085	Transmembrane transport	92 of 1314	2.32E-14
GO:0046942	Carboxylic acid transport	32 of 293	2.33E-08
GO:0046907	Intracellular transport	102 of 1520	5.16E-15

Figure 4-21 STRING analysis of the high-confidence interaction partners identified for the MERS-CoV M proteins in HEK293 cells.

High-confidence interaction partners identified for the MERS-CoV M proteins in HEK293 cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.6.3 *Proteins significantly increased in anti-FLAG co-IPs using lysates from Dubca SARS2 E and Dubca MERS E transfected cells compared to Dubca Neg control*

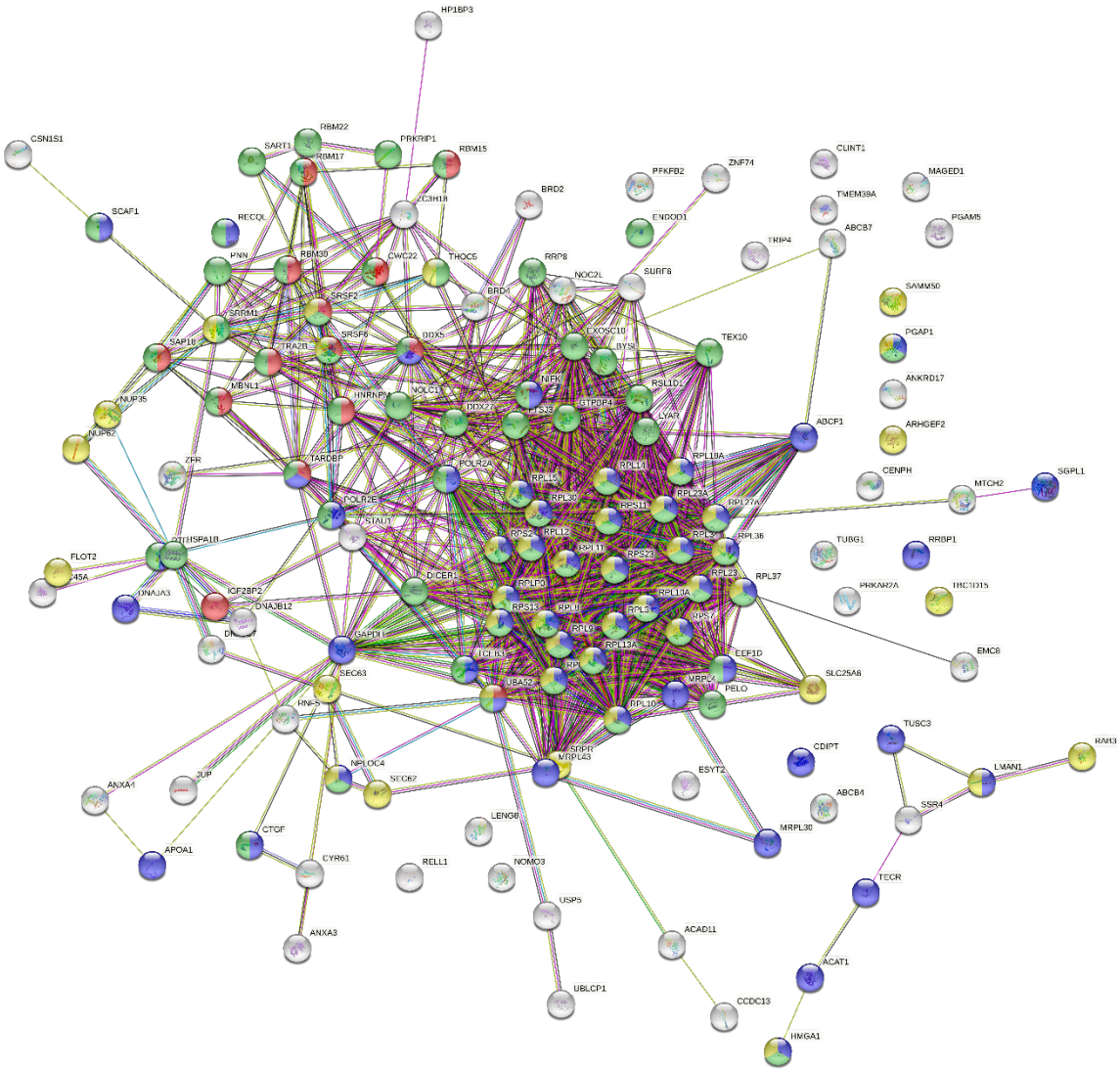
DAVID analysis of the high-confidence interaction partners for Dubca SARS2 E and Dubca MERS E revealed significant enrichment of 5 annotation clusters with an EASE score >1.3 (Table 4-16). Two out of five enrichment clusters were “translation” and “coronavirus disease - COVID-19”, The second cluster identified proteins involved in “nuclear-transcribed mRNA catabolic process”, “nonsense-mediated decay”, “translational initiation”, and “viral transcription”. Moreover, the third enrichment cluster was “isopeptide bond” that identified proteins involved in viral transcription. The last two enrichment clusters were “mRNA processing” and “positive regulation of gene expression”. The last cluster identified proteins involved in “ubiquitin ligase inhibitor activity”. STRING analysis of the high-confidence interaction partners for Dubca SARS2 E and Dubca MERS E identified four significant enrichment GO terms associated with “regulation of mRNA metabolic process”, “organic substance biosynthetic process”, “nucleic acid metabolic process”, and “intracellular transport” (Figure 4-22).

DAVID analysis of the high-confidence interaction partners for Dubca SARS2 E or Dubca MERS E proteins separately revealed significant enrichment of 3 and 22 annotation clusters with an EASE score >1.3, respectively (Table 4-17 & Table 4-18). No unique enrichment clusters were detected for Dubca SARS2 E. Unique enrichment clusters revealed for Dubca MERS E included: “mitochondrion”, “Golgi membrane”, “endoplasmic reticulum”, “protein folding”, and “Chaperone”. STRING analysis of the high-confidence interaction partners for Dubca SARS2 E or Dubca MERS E proteins separately shown no unique significant enrichment GO terms for Dubca SARS2 E. Three unique significant enrichment GO terms for Dubca MERS E “regulation of apoptotic process”, “cellular response to heat”, and “positive regulation of cell death” (Figure 4-23 & Figure 4-24).

Table 4-16 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV and SARS-CoV-2 E protein in Dubca cells

Term	EASE Score	Category	Count	%	*P*-Value
GO:0006412~translation	27.6	GOBP	31	22.5	2.16E-30
hsa05171: Coronavirus disease - COVID-19	27.6	KEGG pathway	26	18.8	7.50E-20
KW-1017~Isopeptide bond	16.5	UP keywords	56	40.6	9.38E-17
KW-0507~mRNA processing	5.2	UP keywords	18	13.0	1.26E-09
GO:0010628~positive regulation of gene expression	2.8	GOBP	5	3.6	0.474632699

There were 4 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:1903311	Regulation of mRNA metabolic process	14 of 338	2.18E-05
GO:1901576	Organic substance biosynthetic process	54 of 2734	4.39E-11
GO:0090304	Nucleic acid metabolic process	70 of 2178	1.87E-27
GO:0046907	Intracellular transport	45 of 1520	1.09E-14

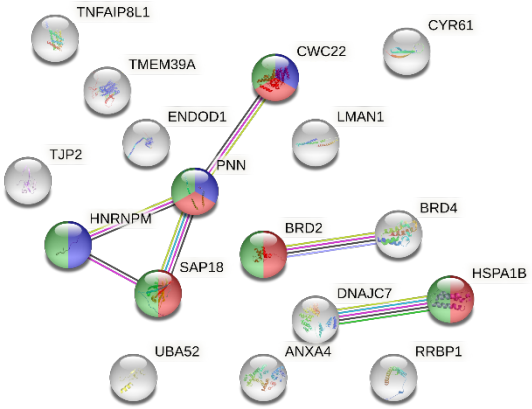
Figure 4-22 STRING analysis of the high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 E proteins in Dubca cells.

High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 E proteins in Dubca cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-17 DAVID analysis of the high-confidence interaction partners identified for the SARS-CoV-2 E protein in Dubca cells

Term	EASE Score	Category	Count	%	*P*-Value
GO:0003723~RNA binding	2.5	GOMF	6	35.3	0.006092377
KW-1017~Isopeptide bond	1.6	UP keywords	6	35.3	0.034455959
GO:0005634~nucleus	1.5	GOCC	9	52.9	0.061673813

There were 3 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0016607	Nuclear speck	5 of 399	0.0312
GO:0071013	Catalytic step 2 spliceosome	3 of 87	0.0345
GO:0016604	Nuclear body	6 of 789	0.0314

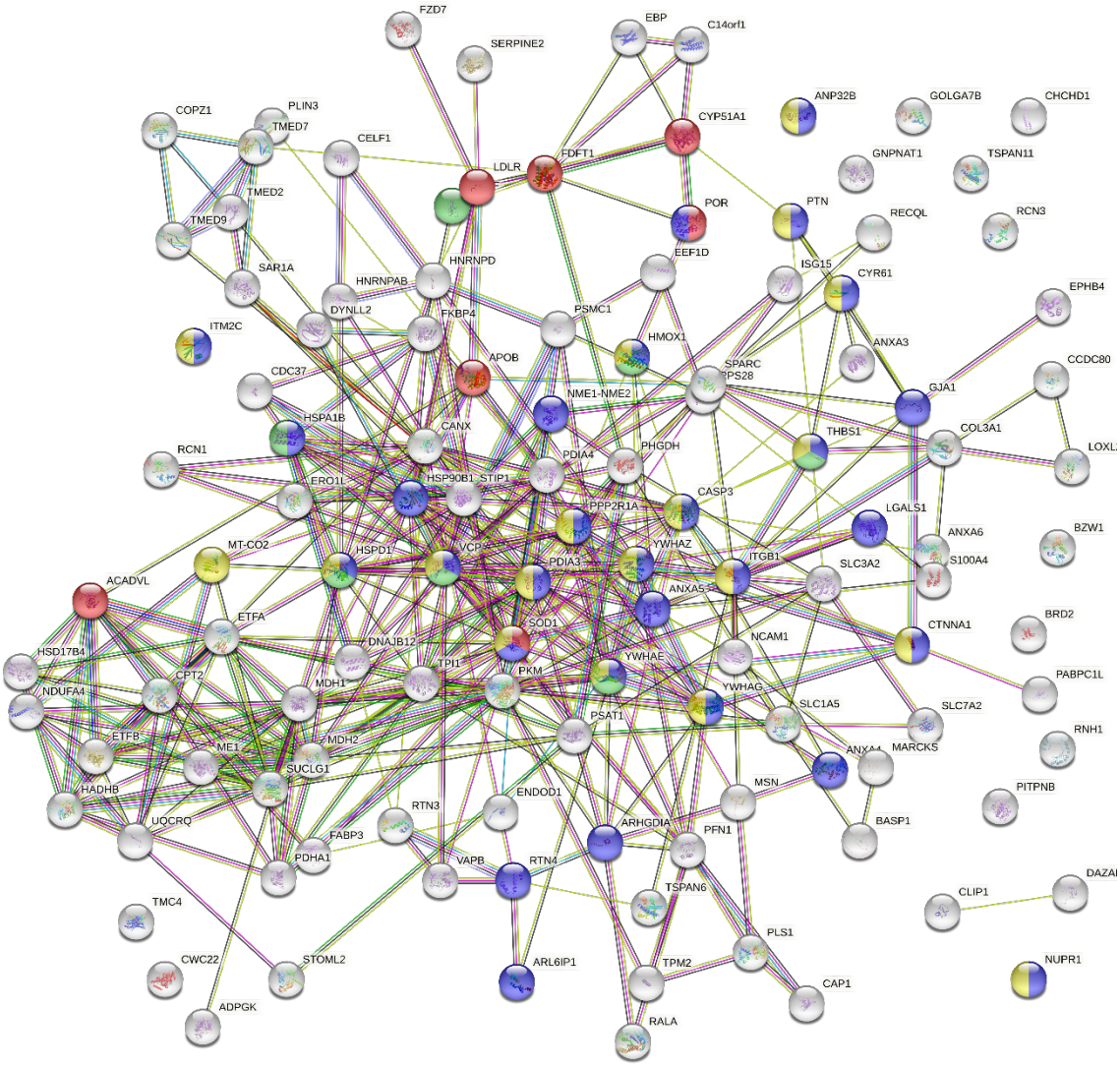
Figure 4-23 STRING analysis of the high-confidence interaction partners identified for the SARS-CoV-2 E proteins in Dubca cells.

High-confidence interaction partners identified for the SARS-CoV-2 E proteins in Dubca cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-18 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV E protein in Dubca cells

Term	EASE Score	Category	Count	%	*P*-Value
GO:0005783~endoplasmic reticulum	8.7	GOTERM_CC_DIRECT	31	25.4	2.85E-12
MOTIF: Prevents secretion from ER	4.3	GOTERM_CC_DIRECT	12	9.8	2.06E-06
GO:0006457~protein folding	4.3	GOTERM_BP_DIRECT	9	7.4	1.01E-05
KW-0143~Chaperone	3.7	UP_KW_MOLECULAR_FUNCTION	9	7.4	7.54E-05
hsa01200: Carbon metabolism	3.6	KEGG_PATHWAY	10	8.2	7.33E-07
GO:0048306~calcium-dependent protein binding	3.6	GOTERM_MF_DIRECT	4	3.3	0.0197289
hsa00620: Pyruvate metabolism	3.0	KEGG_PATHWAY	5	4.1	7.88E-04
GO:0008201~heparin binding	2.3	GOTERM_MF_DIRECT	6	4.9	0.0048850
KW-1207~Sterol metabolism	2.2	UP_KW_BIOLOGICAL_PROCESS	6	4.9	1.91E-04
GO:0034976~response to endoplasmic reticulum stress	2.1	GOTERM_BP_DIRECT	6	4.9	1.87E-04
GO:0043065~positive regulation of apoptotic process	2.0	GOTERM_BP_DIRECT	9	7.4	0.0011061
hsa00270: Cysteine and methionine metabolism	1.9	KEGG_PATHWAY	4	3.3	0.0105355
KW-0496~Mitochondrion	1.9	UP_KW_CELLULAR_COMPONENT	17	13.9	0.0115593
GO:0046718~viral entry into host cell	1.8	GOTERM_BP_DIRECT	6	4.9	4.41E-04
GO:0000139~Golgi membrane	1.8	GOTERM_CC_DIRECT	10	8.2	0.0148012
hsa00010: Glycolysis / Gluconeogenesis	1.6	KEGG_PATHWAY	4	3.3	0.0219005
GO:0016491~oxidoreductase activity	1.5	GOTERM_MF_DIRECT	6	4.9	0.0174184
GO:0045333~cellular respiration	1.4	GOTERM_BP_DIRECT	3	2.5	0.0273768
hsa01212: Fatty acid metabolism	1.4	KEGG_PATHWAY	4	3.3	0.0142500

There were 20 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0090181	Regulation of cholesterol metabolic process	7 of 58	8.75E-05
GO:0042981	Regulation of apoptotic process	29 of 1550	4.44E-05
GO:0034605	Cellular response to heat	7 of 60	9.84E-05
GO:0010942	Positive regulation of cell death	19 of 719	7.18E-05

Figure 4-24 STRING analysis of the high-confidence interaction partners identified for the MERS-CoV E proteins in Dubca cells.

High-confidence interaction partners identified for the MERS-CoV E proteins in Dubca cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.6.4 *Proteins significantly increased in anti-FLAG co-IPs using lysates from Dubca SARS2 M and Dubca MERS M transfected cells compared to Dubca Neg control*

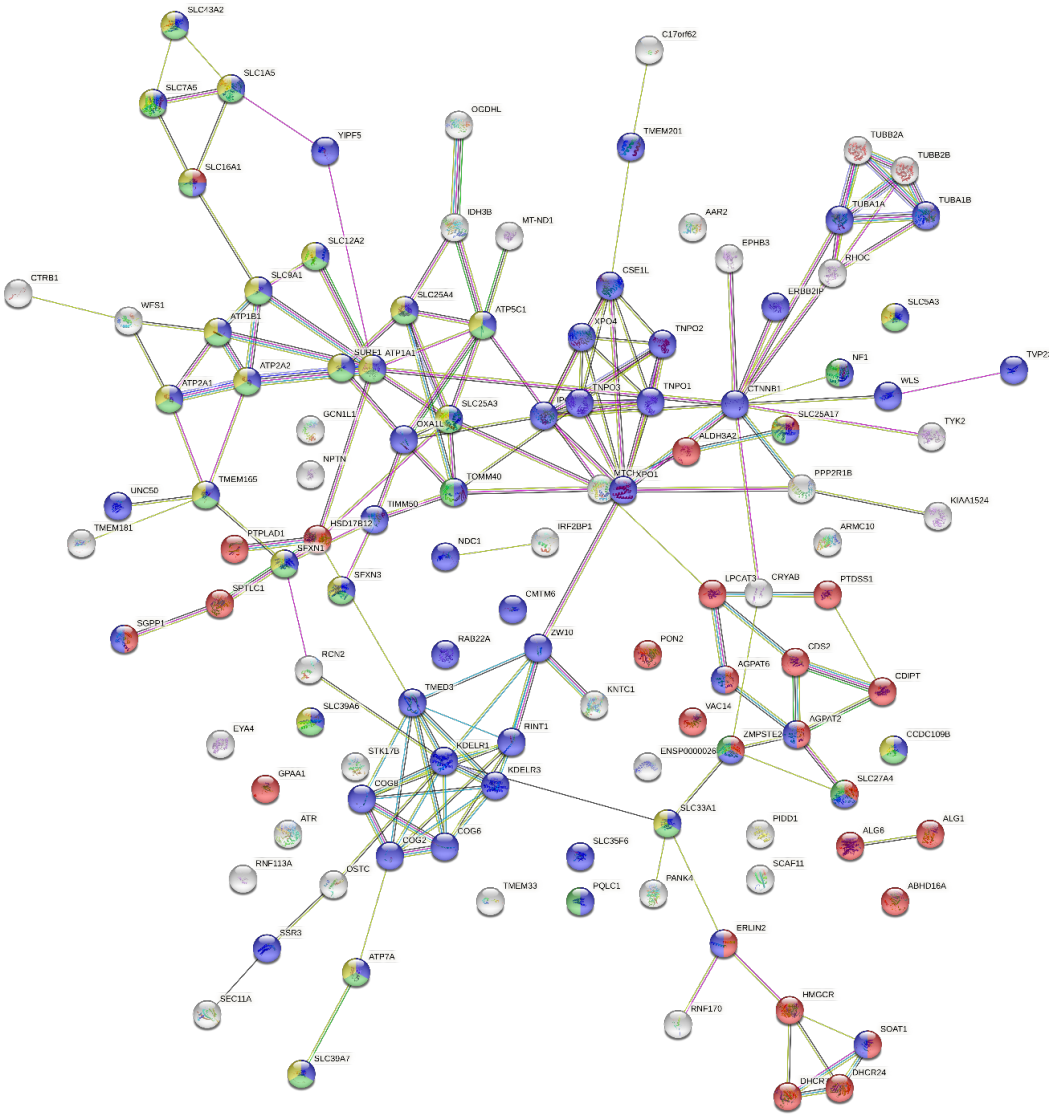
DAVID analysis of the high-confidence interaction partners for Dubca SARS2 M and Dubca MERS M revealed significant enrichment of 15 annotation clusters with an EASE score >1.3 (Table 4-19). Five enrichment clusters were associated with transportation function: “nucleocytoplasmic transport”, “ER-Golgi transport”, “amino-acid transport”, “ion transport”, and “ER to Golgi vesicle-mediated transport”. Other clusters contain proteins associated with expected annotation terms include “cell membrane”, “integral component of membrane”, “trans-Golgi network membrane”, “mitochondrial inner membrane”, “endoplasmic reticulum”, “mitochondrion”, “cytoskeleton”, “lipid metabolism”, “phospholipid metabolism”, and “cholesterol metabolism”. STRING analysis of the high-confidence interaction partners for Dubca SARS2 M and Dubca MERS M identified three significant enrichment GO terms associated with transportation (Figure 4-25).

DAVID analysis of the high-confidence interaction partners for Dubca SARS2 M or Dubca MERS M proteins separately revealed significant enrichment of 18 and 21 annotation clusters with an EASE score >1.3, respectively (Table 4-20 & Table 4-21). Two unique enrichment clusters were revealed for Dubca MERS M “N-glycan biosynthesis” and “calcium signaling pathway”. Moreover, one unique enrichment cluster was revealed for Dubca SARS2 M “methylation”. STRING analysis of the high-confidence interaction partners for Dubca SARS2 M or Dubca MERS M proteins separately shown no unique significant enrichment GO terms (Figure 4-26 & Figure 4-27).

Table 4-19 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV and SARS-CoV-2 M protein in Dubca cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0256~Endoplasmic reticulum	18.0	UP keywords	44	37.6	1.46E-20
GO:0016021~integral component of membrane	13.2	GOCC	71	60.7	7.34E-15
hsa03013:Nucleocytoplasmic transport	5.1	KEGG pathway	9	7.7	3.12E-06
KW-0443~Lipid metabolism	4.0	UP keywords	18	15.4	5.52E-05
KW-1208~Phospholipid metabolism	3.0	UP keywords	6	5.1	1.42E-04
GO:0032588~trans-Golgi network membrane	2.6	GOCC	5	4.3	0.002630309
KW-0931~ER-Golgi transport	2.6	UP keywords	5	4.3	0.00685429
GO:0005743~mitochondrial inner membrane	2.5	GOCC	11	9.4	2.90E-04
KW-0029~Amino-acid transport	2.3	UP keywords	5	4.3	7.80E-04
KW-0406~Ion transport	1.9	UP keywords	13	11.1	0.00528768
KW-0153~Cholesterol metabolism	1.8	UP keywords	5	4.3	0.001777034
KW-1003~Cell membrane	1.8	UP keywords	20	17.1	0.919103715
GO:0005739~mitochondrion	1.8	GOCC	21	17.9	1.39E-04
KW-0206~Cytoskeleton	1.8	UP keywords	8	6.8	0.748583312
GO:0006888~ER to Golgi vesicle-mediated transport	1.5	GOBP	6	5.1	9.78E-04

There were 15 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0006629	Lipid metabolic process	26 of 1190	8.94E-06
GO:0006810	Transport	65 of 4353	4.67E-11
GO:0006811	Ion transport	29 of 1344	2.11E-06
GO:0034220	Ion transmembrane transport	24 of 1010	8.94E-06

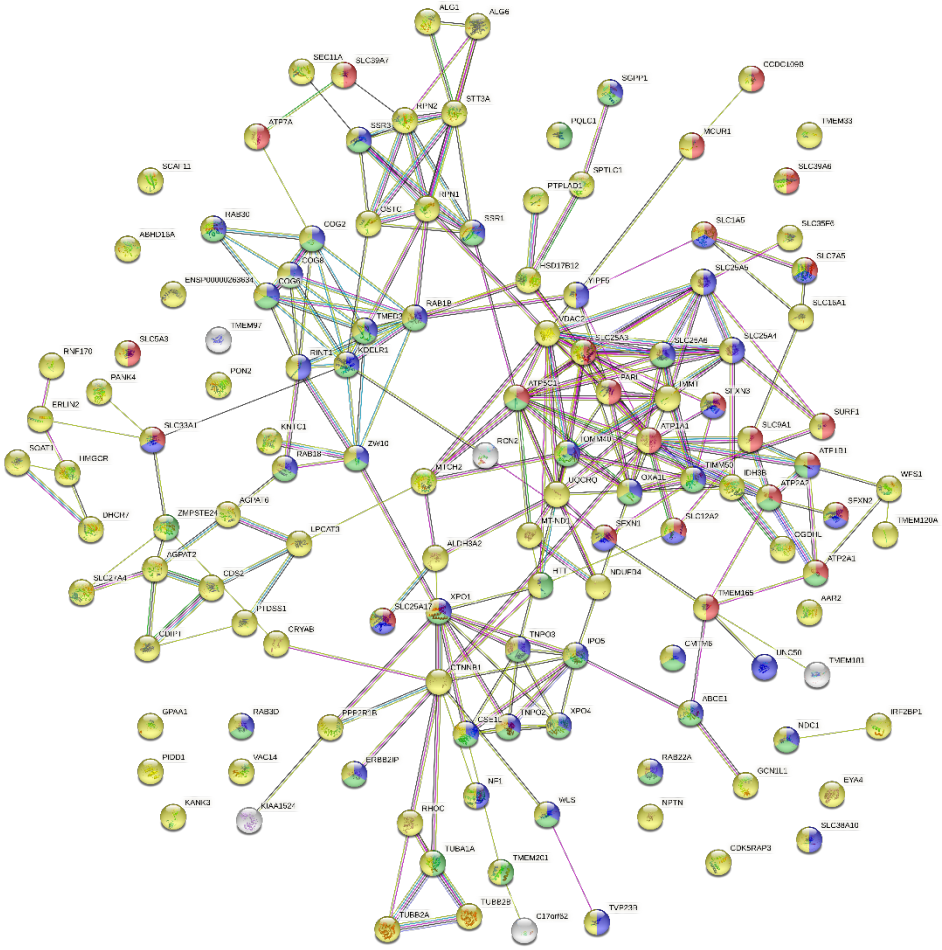
Figure 4-25 STRING analysis of the high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in Dubca cells.

High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in Dubca cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-20 DAVID analysis of the high-confidence interaction partners identified for the SARS-CoV-2 M protein in Dubca cells

Term	EASE Score	Category	Count	%	*P*-Value
GO:0005789~endoplasmic reticulum membrane	17.4	GOCC	43	34.4	7.95E-24
KW-0812~Transmembrane	16.1	UP keywords	84	67.2	1.95E-15
GO:0005739~mitochondrion	7.9	GOCC	29	23.2	2.20E-08
hsa03013: Nucleocytoplasmic transport	3.5	KEGG pathway	8	6.4	3.55E-05
GO:0006865~amino acid transport	3.4	GOBP	4	3.2	0.002117009
hsa00510: N-Glycan biosynthesis	3.4	KEGG pathway	5	4	8.51E-04
KW-0999~Mitochondrion inner membrane	3.0	UP keywords	16	12.8	6.98E-09
hsa03013: Nucleocytoplasmic transport	3.0	KEGG pathway	8	6.4	3.55E-05
GO:0007030~Golgi organization	3.0	GOBP	7	5.6	1.78E-04
KW-0443~Lipid metabolism	2.9	UP keywords	16	12.8	9.94E-04
hsa04020: Calcium signaling pathway	2.9	KEGG pathway	5	4	0.150410596
hsa00190: Oxidative phosphorylation	2.8	KEGG pathway	4	3.2	0.107601679
GO:0006890~retrograde vesicle-mediated transport, Golgi to ER	2.2	GOBP	4	3.2	0.003953127
GO:0006888~ER to Golgi vesicle-mediated transport	1.9	GOBP	6	4.8	0.001331209
GO:0003924~GTPase activity	1.8	GOMF	8	6.4	0.00668551
GO:0000139~Golgi membrane	1.7	GOCC	10	8	0.016321862
GO:0005886~plasma membrane	1.4	GOCC	27	21.6	0.850033054
KW-0488~Methylation	1.4	UP keywords	14	11.2	0.023430357

There were 18 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0098655	Cation transmembrane transport	24 of 725	1.18E-07
GO:0071705	Nitrogen compound transport	46 of 1823	1.48E-12
GO:0046907	Intracellular transport	38 of 1520	9.27E-10
GO:0009987	Cellular process	123 of 15024	5.71E-06

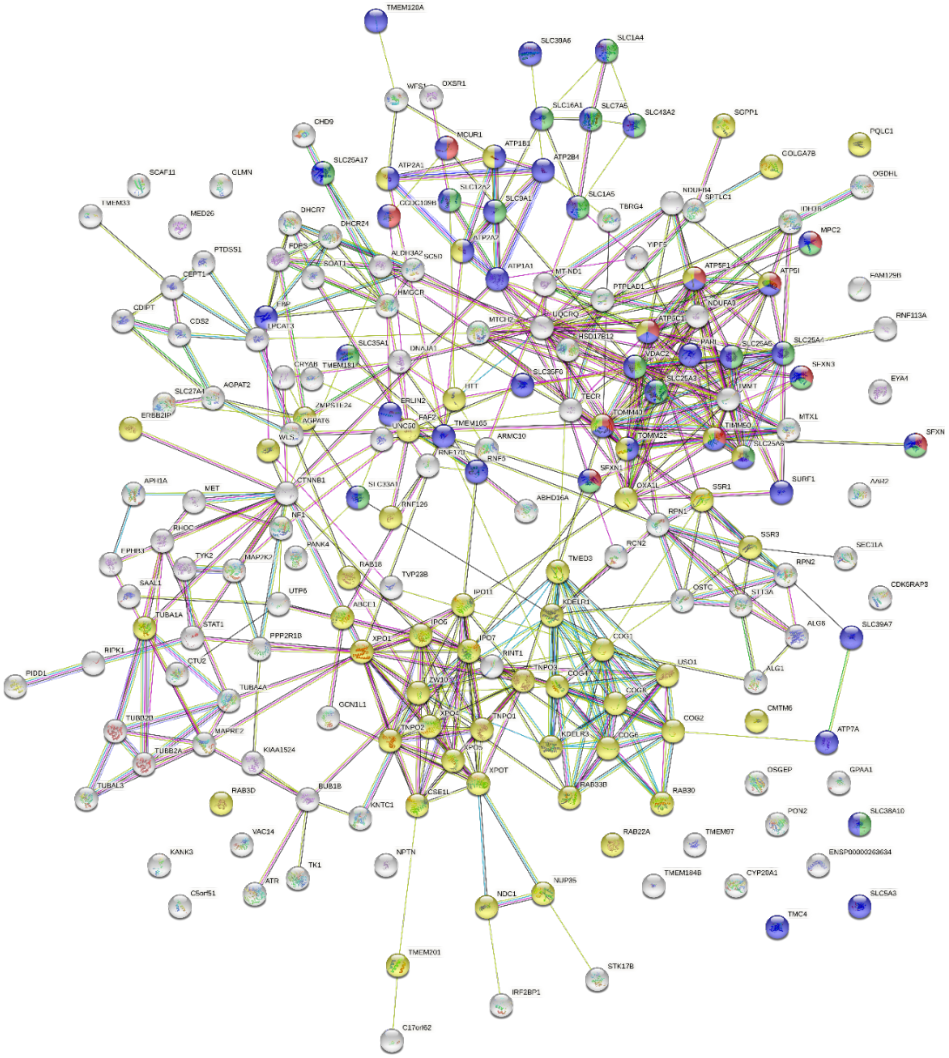
Figure 4-26 STRING analysis of the high-confidence interaction partners identified for the SARS-CoV-2 M proteins in Dubca cells.

High-confidence interaction partners identified for the SARS-CoV-2 M proteins in Dubca cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-21 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV M protein in Dubca cells

Term	EASE Score	Category	Count	%	*P*-Value
GO:0005789~endoplasmic reticulum membrane	19.0	GOCC	54	28.3	4.63E-25
KW-0812~Transmembrane	15.3	UP keywords	110	57.6	4.57E-15
KW-0999~Mitochondrion inner membrane	8.0	UP keywords	19	9.9	1.14E-08
GO:0005739~mitochondrion	7.2	GOCC	41	21.5	2.39E-10
hsa03013: Nucleocytoplasmic transport	6.7	KEGG pathway	14	7.3	1.50E-09
hsa01100: Metabolic pathways	5.0	KEGG pathway	35	18.3	0.001296188
GO:0017119~Golgi transport complex	4.9	GOCC	5	2.6	4.71E-06
hsa00190: Oxidative phosphorylation	4.4	KEGG pathway	7	3.7	0.008951811
GO:0006695~cholesterol biosynthetic process	3.9	GOBP	5	2.6	4.25E-04
GO:0022857~transmembrane transporter activity	3.4	GOMF	5	2.6	0.086524942
GO:0006865~amino acid transport	3.1	GOBP	6	3.1	4.39E-05
KW-0029~Amino-acid transport	3.1	UP keywords	7	3.7	5.29E-05
hsa04020: Calcium signaling pathway	3.0	KEGG pathway	8	4.2	0.041418428
GO:0006486~protein glycosylation	2.9	GOBP	4	2.1	0.140833614
KW-0594~Phospholipid biosynthesis	2.8	UP keywords	7	3.7	2.32E-05
KW-0931~ER-Golgi transport	2.7	UP keywords	6	3.1	0.0074541
GO:0005525~GTP binding	2.5	GOMF	11	5.8	0.006053182
hsa04022: cGMP-PKG signaling pathway	2.3	KEGG pathway	10	5.2	3.77E-04
KW-0560~Oxidoreductase	2.1	UP keywords	11	5.8	0.009910072
GO:0030007~cellular potassium ion homeostasis	1.7	GOBP	3	1.6	0.007743432
KW-0449~Lipoprotein	1.7	UP keywords	10	5.2	0.674315976

There were 21 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:1990542	Mitochondrial transmembrane transport	11 of 95	3.02E-06
GO:0098656	Anion transmembrane transport	20 of 380	1.01E-06
GO:0098655	Cation transmembrane transport	28 of 725	4.94E-07
GO:0046907	Intracellular transport	54 of 1520	1.08E-13

Figure 4-27 STRING analysis of the high-confidence interaction partners identified for the MERS-CoV M proteins in Dubca cells.

High-confidence interaction partners identified for the MERS-CoV M proteins in Dubca cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of

coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.6.5 *Proteins significantly increased in anti-FLAG co-IPs using lysates from PaKiT SARS2 E and PaKiT MERS E transfected cells compared to PaKiT Negative control*

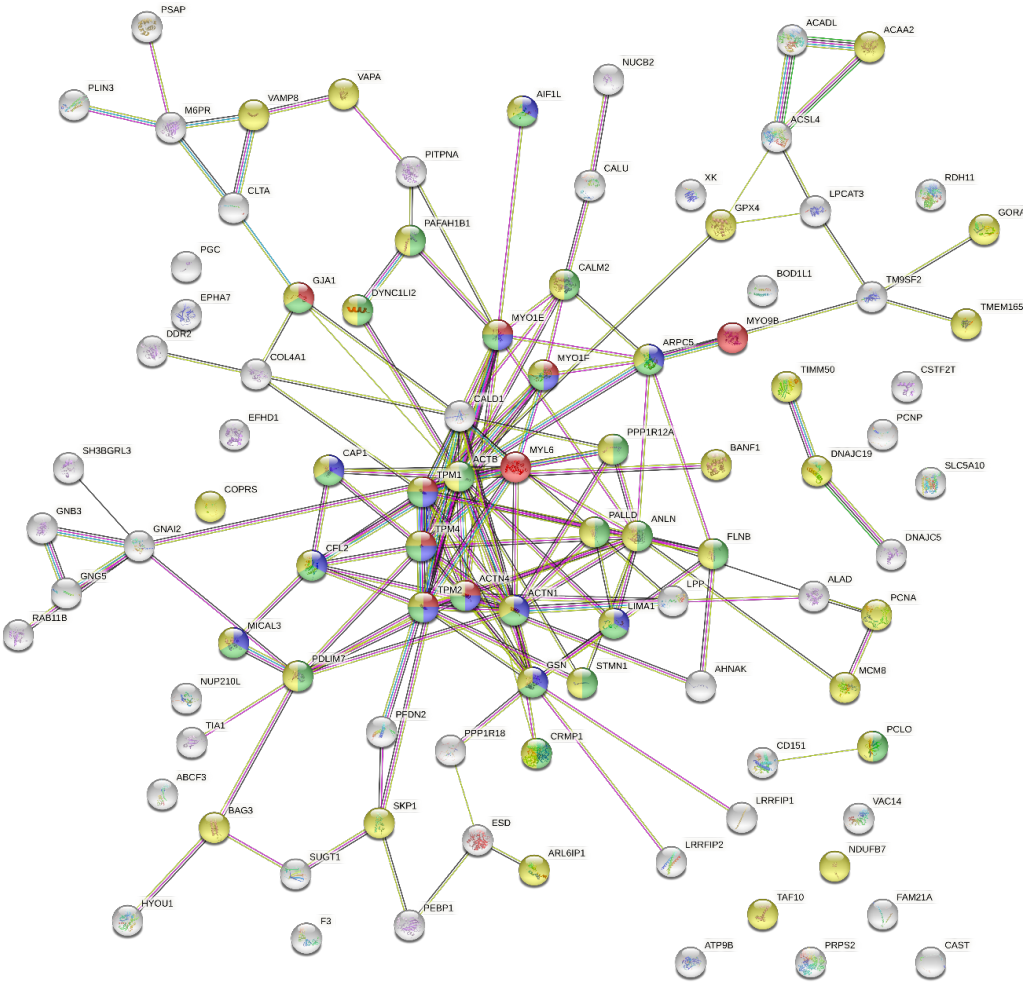
DAVID analysis of the high-confidence interaction partners for PaKiT SARS2 E and PaKiT MERS E revealed significant enrichment of 8 annotation clusters with an EASE score >1.3 (Table 4-22). Five enrichment clusters were associated with filament and skeleton protein function: “cytoskeleton”, “stress fiber”, “myosin complex”, “actin filament”, and “regulation of actin cytoskeleton”. Another three enrichment clusters were associated with “calcium ion binding” and “metal ion binding”, and “cell cycle”. STRING analysis of the high-confidence interaction partners for PaKiT SARS2 E and PaKiT MERS E identified four significant enrichment GO terms associated with “actin filament role in cells” (Figure 4-28).

DAVID analysis of the high-confidence interaction partners for PaKiT SARS2 E or PaKiT MERS E proteins separately revealed significant enrichment of 11 and 29 annotation clusters with an EASE score >1.3, respectively (Table 4-23 & Table 4-24). No unique enrichment cluster was revealed for PaKiT SARS2 E. Multiple unique enrichment clusters were revealed for PaKiT MERS E including: “protein folding”, “extracellular region”, “mitochondrion”, “protein binding involved in protein folding”,” stress response”, “podosome”, “Isomerase”, “protein biosynthesis”, and “nitrosylation”. Moreover, some enrichment cluster had more details like “Zinc finger, LIM-type” include a “stress fiber protein,” “endopeptidase activity” including proteins involved in proteasomal ubiquitin-independent protein catabolic process, “nucleotide-binding” including 5 proteins involved in protein refolding, “thioredoxin” domain including proteins involved in Hydrogen peroxide catabolic process, ” estrogen signaling pathway” involved in Regulation of cellular response to heat, “clathrin-coated pit” includes four endosomal proteins. One Unique enrichment cluster was revealed for PaKiT MERS E “Chaperone-mediated protein folding”. STRING analysis of the high-confidence interaction partners for PaKiT SARS2 E or PaKiT MERS E proteins separately shown no unique significant enrichment GO term for PaKiT SARS2 E and one unique significant enrichment GO term for PaKiT MERS E “cellular response to cytokine stimulus” (Figure 4-29 & Figure 4-30).

Table 4-22 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV and SARS-CoV-2 E protein in PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0206~Cytoskeleton	14.1	UP keywords	30	31.6	1.46E-12
GO:0001725~stress fiber	3.6	GOCC	10	10.5	4.24E-11
GO:0016459~myosin complex	2.8	GOCC	5	5.3	7.23E-05
GO:0005884~actin filament	2.8	GOCC	8	8.4	2.36E-07
GO:0005509~calcium ion binding	2.4	GOMF	11	11.6	0.003685914
hsa04810: Regulation of actin cytoskeleton	1.5	KEGG PATHWAY	7	7.4	0.003013571
GO:0046872~metal ion binding	1.4	GOMF	7	7.4	0.992573547
GO:0007049~cell cycle	1.3	GOBP	7	7.4	0.006748686

There were 8 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0030048	Actin filament-based movement	9 of 121	2.08E-05
GO:0007015	Actin filament organization	14 of 254	1.03E-07
GO:0030036	Actin cytoskeleton organization	20 of 516	7.34E-09
GO:0006996	Organelle organization	44 of 3450	3.68E-07

Figure 4-28 STRING analysis of the high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 E proteins in PaKiT cells.

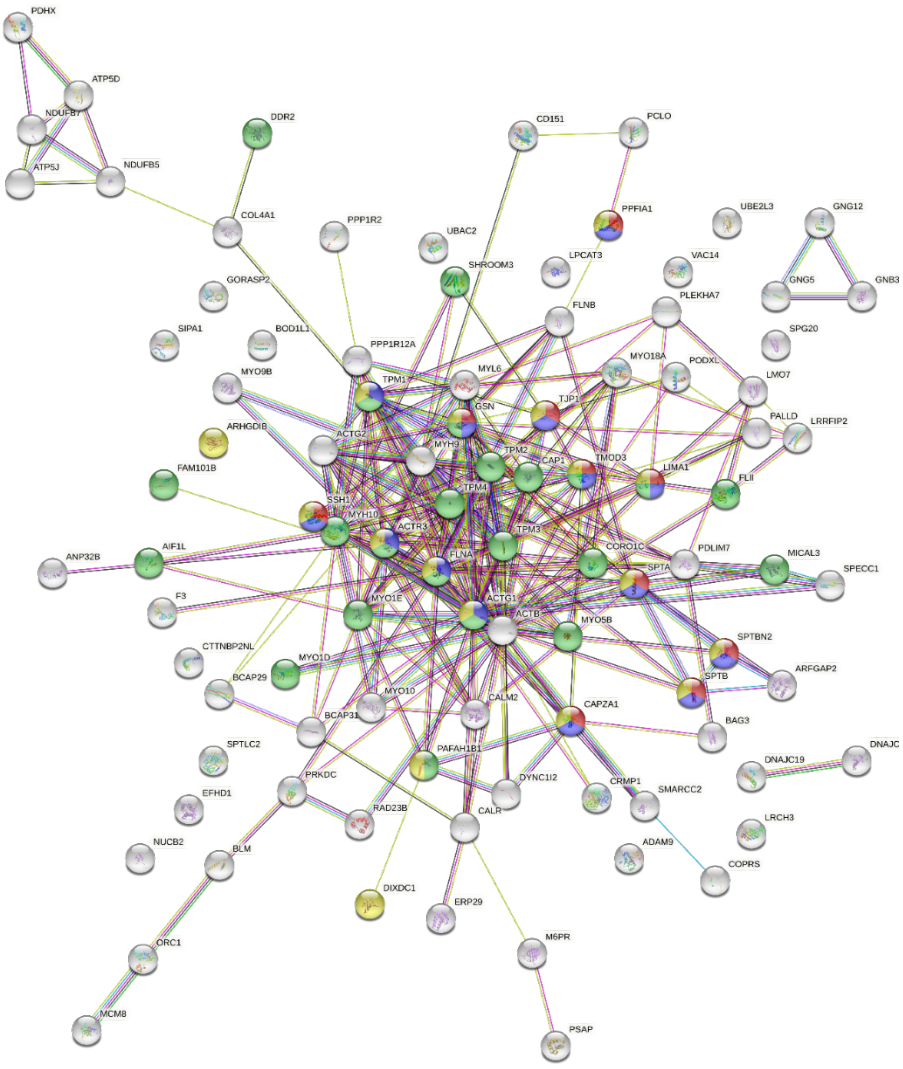
High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 E proteins in PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-23 DAVID analysis of the high-confidence interaction partners identified for the SARS-CoV-2 E protein in PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0206~Cytoskeleton	24.7	UP keywords	39	42.4	2.06E-20
DOMAIN:Calponin-homology (CH)	6.1	UP keywords	9	9.8	1.70E-09
GO:0005524~ATP binding	5.2	GOMF	18	19.6	8.21E-04
GO:0005856~cytoskeleton	4.5	GOCC	16	17.4	1.11E-08
IPR001478: PDZ domain	3.1	INTERPRO	7	7.6	1.05E-04
GO:0005200~structural constituent of cytoskeleton	3.0	GOMF	7	7.6	1.17E-05
hsa04530: Tight junction	2.4	KEGG PATHWAY	7	7.6	9.06E-04
GO:0005911~cell-cell junction	2.4	GOCC	5	5.4	0.009856256
GO:0005509~calcium ion binding	1.9	GOMF	10	10.9	0.009137333
hsa04921: Oxytocin signaling pathway	1.8	KEGG PATHWAY	5	5.4	0.019789218
GO:0046872~metal ion binding	1.3	GOMF	8	8.7	0.974466329

There were 11 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.

-



GO term	Description	Count	FDR
GO:1902904	Negative regulation of supramolecular fiber organization	10 of 143	2.91E-06
GO:0110053	Regulation of actin filament organization	14 of 273	1.18E-07
GO:0097435	Supramolecular fiber organization	23 of 480	3.58E-13
GO:0051493	Regulation of cytoskeleton organization	17 of 546	1.42E-06

Figure 4-29 STRING analysis of the high-confidence interaction partners identified for the SARS-CoV-2 E proteins in PaKiT cells.

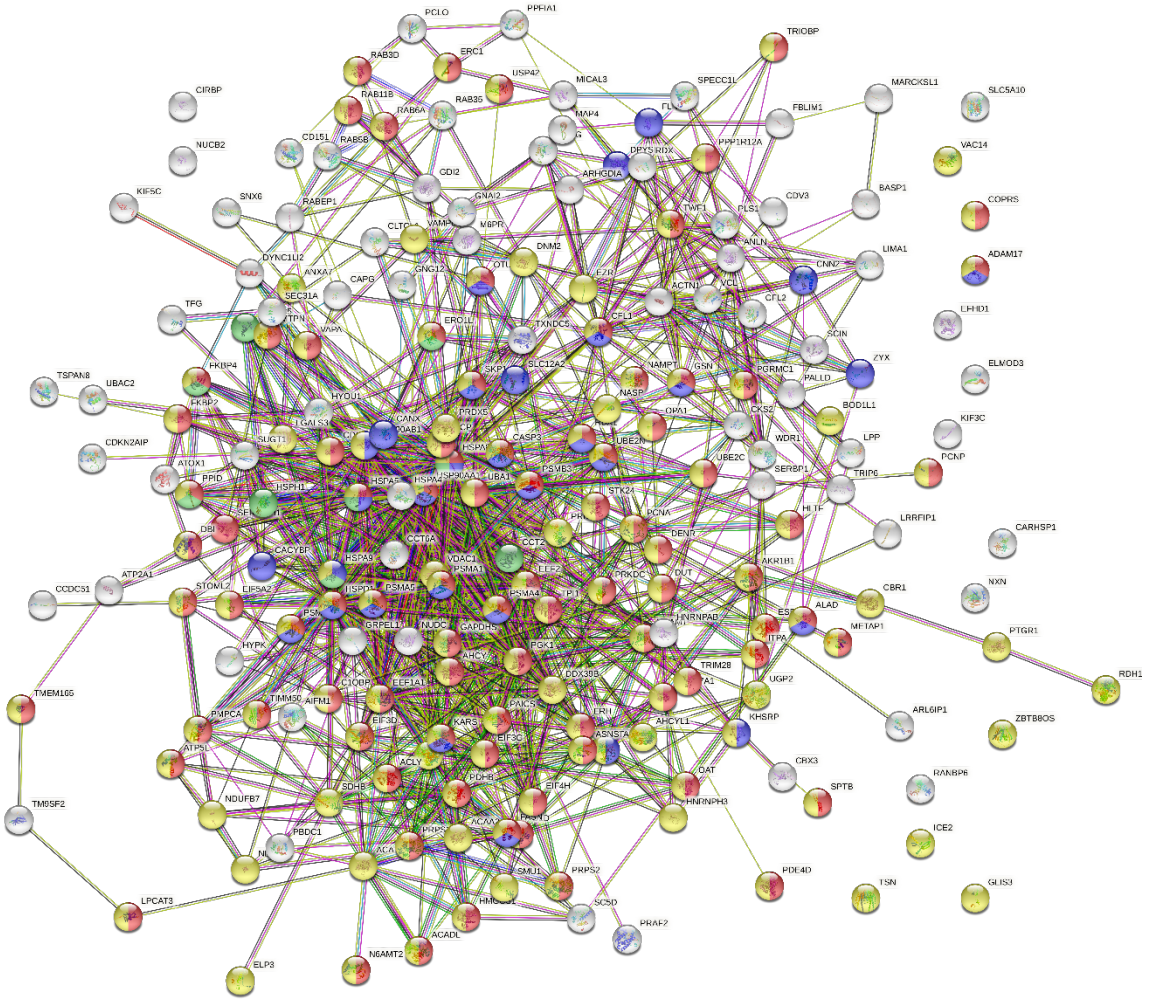
High-confidence interaction partners identified for the SARS-CoV-2 E proteins in PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-24 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV E protein in PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0206~Cytoskeleton	8.3	UP keywords	37	18.5	9.82E-08
GO:0006457~protein folding	6.1	GOBP	16	8	1.70E-10
GO:0005576~extracellular region	5.2	GOCC	21	10.5	0.578397688
GO:0005739~mitochondrion	4.8	GOCC	34	17	2.64E-06
GO:0044183~protein binding involved in protein folding	3.8	GOMF	7	3.5	5.24E-06
GO:0030042~actin filament depolymerization	3.5	GOBP	6	3	1.28E-07
GO:0030036~actin cytoskeleton organization	3.0	GOBP	7	3.5	0.011191558
KW-0346~Stress response	2.9	UP keywords	10	5	8.87E-06
IPR001781:Zinc finger, LIM-type	2.9	INTERPRO	6	3	9.93E-04
GO:0002102~podosome	2.9	GOCC	6	3	9.19E-06
GO:0004175~endopeptidase activity	2.9	GOMF	6	3	0.002160959
hsa00010:Glycolysis / Gluconeogenesis	2.3	KEGG PATHWAY	7	3.5	3.75E-04
hsa00280:Valine, leucine and isoleucine degradation	2.2	KEGG PATHWAY	5	2.5	0.004813244
KW-0547~Nucleotide-binding	2.1	UP keywords	43	21.5	3.16E-04
IPR027417:P-loop containing nucleoside triphosphate hydrolase	2.1	INTERPRO	18	9	0.013952082
IPR013766:Thioredoxin domain	2.0	INTERPRO	4	2	0.007716821
GO:0050821~protein stabilization	2.0	GOBP	8	4	0.007653575
GO:0051287~NAD binding	2.0	GOMF	3	1.5	0.065086864
KW-0413~Isomerase	2.0	UP keywords	7	3.5	0.008527945
hsa00020:Citrate cycle (TCA cycle)	1.9	KEGG PATHWAY	5	2.5	8.26E-04
IPR007052:CS-like domain	1.7	INTERPRO	3	1.5	0.014088797
GO:0005874~microtubule	1.7	GOCC	11	5.5	0.001278348
KW-0648~Protein biosynthesis	1.7	UP keywords	7	3.5	0.011634633
GO:0006414~translational elongation	1.6	GOBP	3	1.5	0.019246095
KW-0702~S-nitrosylation	1.6	UP_KW_PTM	5	2.5	0.007903889
hsa04915:Estrogen signaling pathway	1.5	KEGG PATHWAY	5	2.5	0.136311841
GO:0005905~clathrin-coated pit	1.4	GOCC	4	2	0.020997346
KW-0967~Endosome	1.4	UP keywords	10	5	0.229074262
hsa00280:Valine, leucine and isoleucine degradation	1.4	KEGG PATHWAY	5	2.5	0.004813244

There were 29 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the

table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:1901564	Organonitrogen compound metabolic process	89 of 5244	4.76E-05
GO:0071345	Cellular response to cytokine stimulus	31 of 1013	2.58E-05
GO:0061077	Chaperone-mediated protein folding	9 of 55	5.34E-06
GO:0044237	Cellular metabolic process	122 of 7513	5.14E-07

Figure 4-30 STRING analysis of the high-confidence interaction partners identified for the MERS-CoV E proteins in PaKiT cells.

High-confidence interaction partners identified for the MERS-CoV E proteins in PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.6.6 *Proteins significantly increased in anti-FLAG co-IPs using lysates from PaKiT SARS2 M and PaKiT MERS M transfected cells compared to PaKiT Neg control*

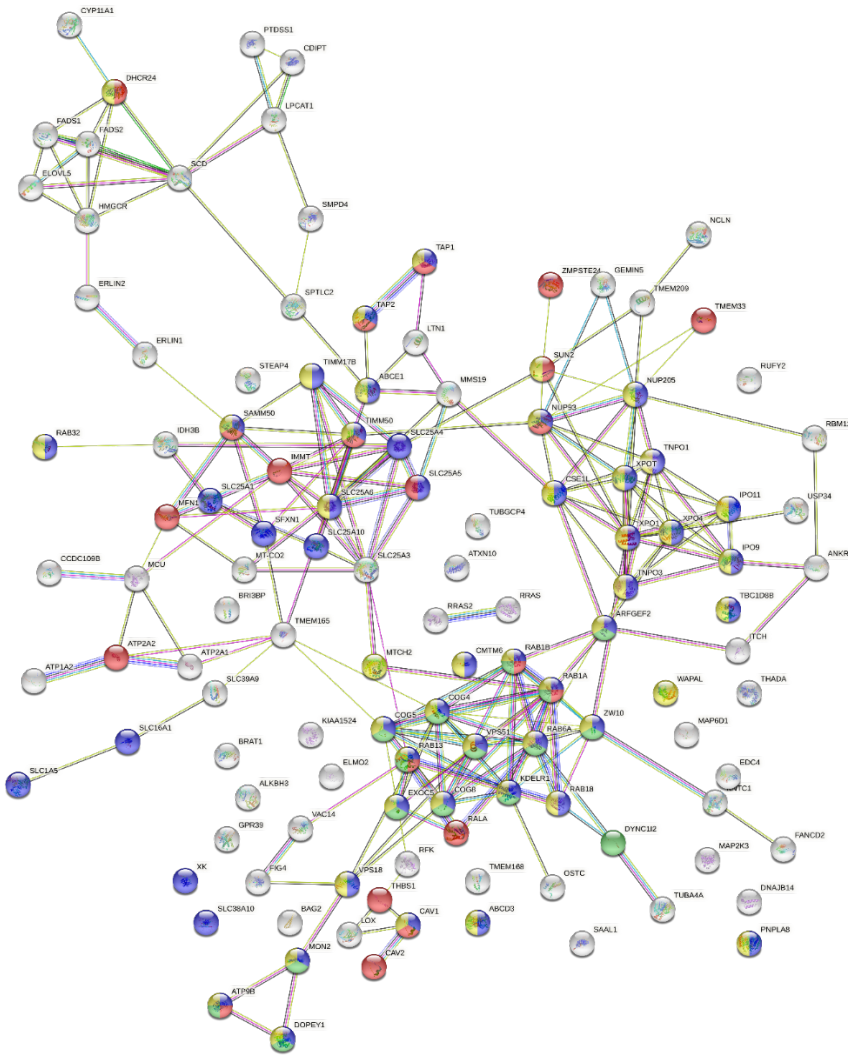
DAVID analysis of the high-confidence interaction partners for PaKiT SARS2 M and PaKiT MERS M revealed significant enrichment of 13 annotation clusters with an EASE score >1.3 (Table 4-25). Multi enrichment clusters were associated with multicellular function and localisation: “ATP binding” “Mitochondrion”, “Golgi apparatus”, “Golgi organisation”, “cholesterol metabolism”, “lipid biosynthesis”, “endoplasmic reticulum membrane”, “membrane”, “nucleocytoplasmic transport”, and “ion transport”. Other enrichment clusters: “nucleotide-binding” (include proteins involved in Intracellular signal transduction), “metabolic pathways” (include proteins related to lipid biosynthesis) and “heat 2” (include proteins involved in nucleocytoplasmic carrier activity). STRING analysis of the high-confidence interaction partners for PaKiT SARS2 M and PaKiT MERS M identified four significant enrichment GO terms “organic substance transport”, "membrane organization”, “Golgi vesicle transport”, and “macromolecule localization” (Figure 4-31).

DAVID analysis of the high-confidence interaction partners for PaKiT SARS2 M or PaKiT MERS M proteins separately revealed significant enrichment of 7 and 20 annotation clusters with an EASE score >1.3, respectively (Table 4-26 & Table 4-27). Two unique enrichment cluster was revealed for PaKiT SARS2 M “Armadillo-like helical” (include proteins involved in Intracellular transportation) and “macromolecular complex” (include proteins associated with the term “endoplasmic reticulum membrane protein”). Multi unique enrichment clusters were revealed for PaKiT MERS M “cristae formation”, “calcium signaling pathway”, “glycan biosynthesis”, “oxidoreductase”, microtubule-based process”, “heat 10” (include proteins associated with the term “nucleus proteins”), “ceramide biosynthetic process” (proteins in this enrichment cluster was involved in cristae formation in inner mitochondrial membrane), “spinocerebellar ataxia” (include proteins involved in Ion transmembrane transport). STRING analysis of the high-confidence interaction partners for PaKiT SARS2 M or PaKiT MERS M proteins separately shown no unique significant enrichment GO term (Figure 4-32 & Figure 4-33).

Table 4-25 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV and SARS-CoV-2 M protein in PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
GO:0005789~endoplasmic reticulum membrane	6.5	GOCC	29	24.6	4.36E-12
KW-0472~Membrane	6.5	UP keywords	82	69.5	3.12E-08
hsa03013: Nucleocytoplasmic transport	6.1	KEGG_PATHWAY	11	9.3	1.24E-08
REPEAT: HEAT 2	4.7	UP keywords	7	5.9	1.80E-06
KW-0333~Golgi apparatus	4.6	UP keywords	19	16.1	3.20E-05
KW-0547~Nucleotide-binding	3.1	UP keywords	22	18.6	0.00386643
hsa01100: Metabolic pathways	2.9	KEGG_PATHWAY	16	13.6	0.253045621
KW-0496~Mitochondrion	2.9	UP keywords	23	19.5	2.12E-05
KW-0444~Lipid biosynthesis	2.9	UP keywords	10	8.5	4.85E-06
GO:0007030~Golgi organization	2.8	GOBP	7	5.9	1.28E-04
KW-0153~Cholesterol metabolism	1.9	UP keywords	5	4.2	0.001496506
KW-0406~Ion transport	1.8	UP keywords	8	6.8	0.224938402
GO:0005524~ATP binding	1.6	GOMF	11	9.3	0.461610518

There were 13 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0071702	Organic substance transport	49 of 2173	3.38E-13
GO:0061024	Membrane organization	21 of 796	1.59E-05
GO:0048193	Golgi vesicle transport	16 of 359	1.09E-06
GO:0033036	Macromolecule localization	44 of 2473	2.81E-08

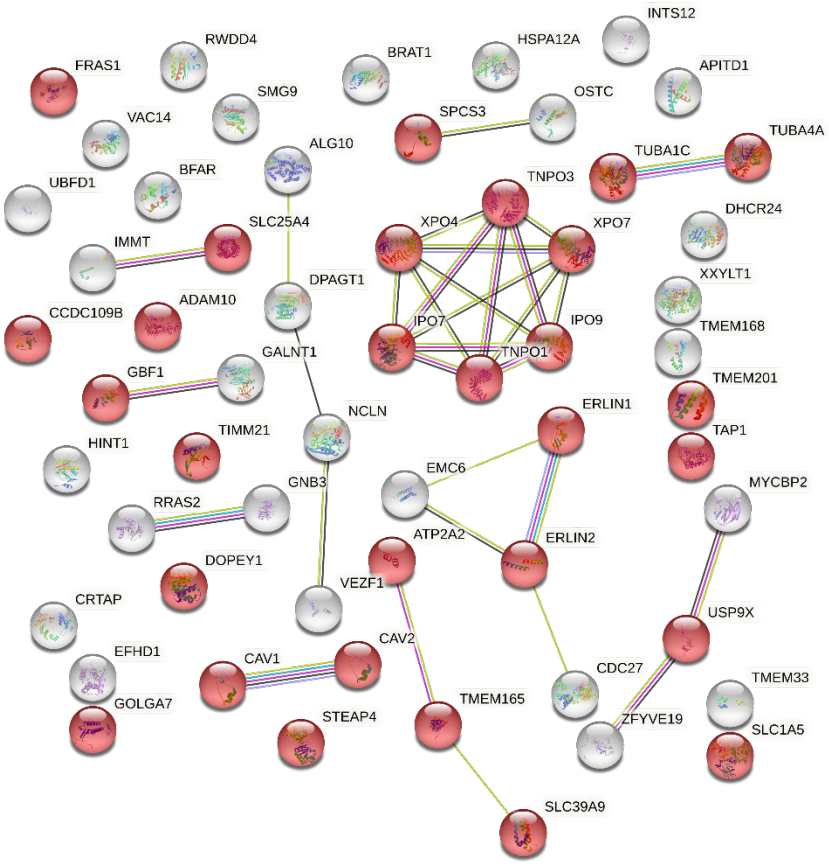
Figure 4-31 STRING analysis of the high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in PaKiT cells.

High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-26 DAVID analysis of the high-confidence interaction partners identified for the SARS-CoV-2 M protein in PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
KW-0256~Endoplasmic reticulum	5.3	UP keywords	15	27.3	1.04E-05
KW-0812~Transmembrane	3.9	UP keywords	30	54.5	2.18E-05
KW-0653~Protein transport	3.8	UP keywords	9	16.4	2.59E-04
IPR011989: Armadillo-like helical	2.7	INTERPRO	7	12.7	3.87E-05
GO:0000139~Golgi membrane	2.5	GOCC_DIRECT	11	20.0	4.49E-06
GO:0032991~macromolecular complex	1.9	GOCC_DIRECT	7	12.7	0.00853768
KW-0443~Lipid metabolism	1.5	UP keywords	3	5.5	0.618879445

There were 7 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0006810	Transport	29 of 4353	4.40E-02

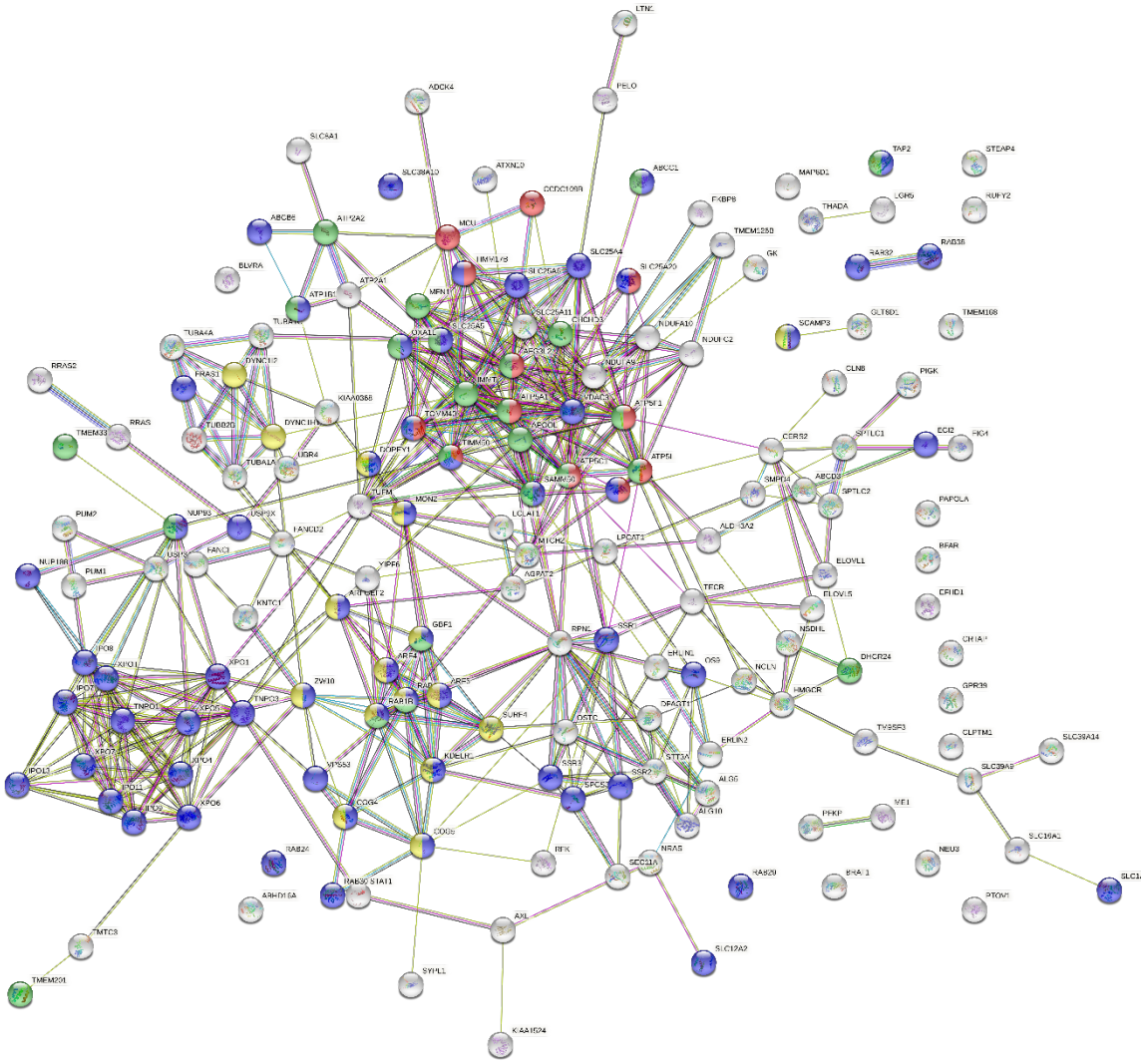
Figure 4-32 STRING analysis of the high-confidence interaction partners identified for the SARS-CoV-2 M proteins in PaKiT cells.

High-confidence interaction partners identified for the SARS-CoV-2 M proteins in PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

Table 4-27 DAVID analysis of the high-confidence interaction partners identified for the MERS-CoV M protein in PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
GO:0005789~endoplasmic reticulum membrane	13.9	GOCC	41	26.3	7.63E-18
KW-0812~Transmembrane	12.6	UP keywords	92	59.0	2.90E-14
hsa03013: Nucleocytoplasmic transport	10.7	KEGG_PATHWAY	16	10.3	3.89E-12
GO:0005739~mitochondrion	7.4	GOCC	37	23.7	1.22E-10
Nuclear structure / Intracellular trafficking and secretion	5.7	INTERPRO	6	3.8	1.77E-10
GO:0003924~GTPase activity	4.5	UP keywords	17	10.9	2.09E-09
REPEAT: HEAT 10	3.7	UP keywords	5	3.2	2.82E-05
GO:0042407~cristae formation	3.6	GOBP	5	3.2	6.47E-06
KW-0333~Golgi apparatus	3.4	UP keywords	21	13.5	2.56E-04
GO:0006888~ER to Golgi vesicle-mediated transport	3.1	GOBP	7	4.5	6.05E-04
GO:0046513~ceramide biosynthetic process	3.0	GOBP	6	3.8	5.88E-06
hsa04020: Calcium signaling pathway	2.9	KEGG_PATHWAY	8	5.1	0.031862889
KW-1134~Transmembrane beta strand	2.7	UP keywords	3	1.9	0.005931009
hsa00510: N-Glycan biosynthesis	2.5	KEGG_PATHWAY	5	3.2	0.003538723
hsa05017: Spinocerebellar ataxia	2.2	KEGG_PATHWAY	11	7.1	1.18E-05
GO:0007017~microtubule-based process	2.0	GOBP	4	2.6	0.003918102
KW-0276~Fatty acid metabolism	1.7	UP keywords	4	2.6	0.198532578
GO:0016887~ATPase activity	1.7	GOMF	5	3.2	0.315608396
KW-0560~Oxidoreductase	1.6	UP keywords	8	5.1	0.116601076
KW-0594~Phospholipid biosynthesis	1.5	UP keywords	3	1.9	0.07032267

There were 20 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:1990542	Mitochondrial transmembrane transport	12 of 95	6.11E-08
GO:0071705	Nitrogen compound transport	60 of 1823	2.44E-17
GO:0061024	Membrane organization	24 of 796	4.04E-05
GO:0048193	Golgi vesicle transport	16 of 359	4.64E-05

Figure 4-33 STRING analysis of the high-confidence interaction partners identified for the MERS-CoV M proteins in PaKiT cells.

High-confidence interaction partners identified for the MERS-CoV M proteins in PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of

coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.6.7 *Proteins significantly increased in anti-FLAG co-IPs using lysates from HEK293 E, Dubca E, and PaKiT E transfected cells compared to Neg control*

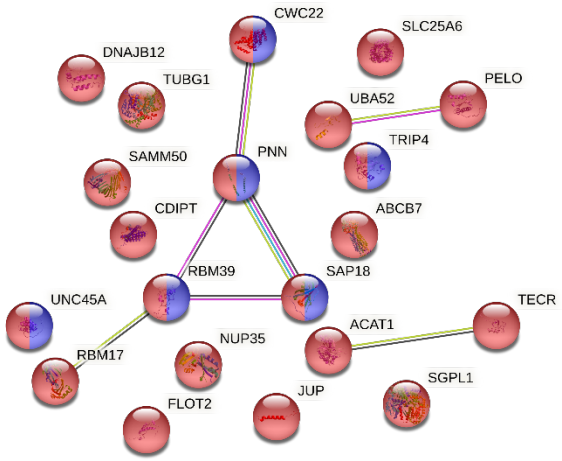
DAVID analysis of the shared high-confidence interaction partners for the MERS-CoV and SARS-CoV-2 E protein in HEK293, Dubca, and PaKiT cells. There was only 1 annotation cluster with an EASE score >1.3. This one cluster contains proteins associated with the annotation term: “Nucleus” (with proteins involved in mRNA processing) (Table 4-28).

STRING analysis of the shared high-confidence interaction partners for the MERS-CoV and SARS-CoV-2 E protein in HEK293, Dubca, and PaKiT cells showed significant enrichment of proteins associated with “Intracellular membrane-bounded organelle” (Figure 4-34).

Table 4-28 DAVID analysis of the shared high-confidence interaction partners identified for the MERS-CoV and SARS-CoV-2 E protein in HEK293, Dubca, and PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
Nucleus	1.4	UP keywords	10	47.6	0.045715391

There was 1 annotation cluster with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
GO:0016604	Nuclear body	6 of 789	0.0479
GO:0043231	Intracellular membrane-bounded organelle	21 of 10761	0.006

Figure 4-34 STRING analysis of the shared high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 E proteins in HEK293, Dubca, and PaKiT cells.

High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in HEK293, Dubca, and PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.6.8 *Proteins significantly increased in anti-FLAG co-IPs using lysates from HEK293 M, Dubca M, and PaKiT M transfected cells compared to Neg control*

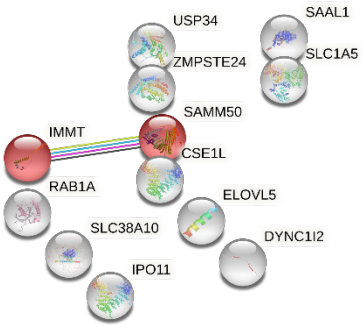
DAVID analysis of the shared high-confidence interaction partners for the MERS-CoV and SARS-CoV-2 M protein in HEK293, Dubca, and PaKiT cells. There were 2 annotation clusters with an EASE score >1.3. These clusters contain proteins associated with the annotation terms: “Transport” and “cytosol” (Table 4-29).

STRING analysis of the shared high-confidence interaction partners for the MERS-CoV and SARS-CoV-2 M protein in HEK293, Dubca, and PaKiT cells identified significant enrichment of proteins associated with SAM complex (Figure 4-35).

Table 4-29 DAVID analysis of the shared high-confidence interaction partners identified for the MERS-CoV and SARS-CoV-2 M protein in HEK293, Dubca, and PaKiT cells

Term	EASE Score	Category	Count	%	*P*-Value
Transport	1.8	UP keywords	6	50.0	0.002295396
GO:0005829~cytosol	1.4	GOCC	5	41.7	0.123360832

There were 2 annotation clusters with an EASE score >1.3. Representative GO terms associated with each cluster are shown in this table with the GO Category (GOBP, GOCC, or GOMF), UP keywords, INTERPRO term, or KEGG pathway. Moreover, the table includes the number of proteins in each cluster (count), the number of proteins associated with each GO term as a percentage of the total number of proteins in the data set (%), and the *p*-value for each annotation term.



GO term	Description	Count	FDR
CL:13630	SAM complex	2 to 5	0.0328

Figure 4-35 STRING analysis of the shared high-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in HEK293, Dubca, and PaKiT cells.

High-confidence interaction partners common to the MERS-CoV and SARS-CoV-2 M proteins in HEK293, Dubca, and PaKiT cells were analysed using the STRING database to identify interaction networks and proteins associated with significantly enriched GO terms. Coloured nodes represent proteins that were associated with significantly enriched GO terms. The number of coloured nodes/total proteins involved for each term and the FDR of each GO term are listed in the table.

4.2.7 Enrichment for proteomic data of HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV -E and-M proteins

The bioinformatics analysis of the proteomic data set in Section 4.2.6 identified several cellular processes that might be important for viral proteins or even for viral replication. In this section, viral protein levels are compared, including between data sets for each viral protein in the three host cells. Specifically, we examined: (1) SARS-CoV-2 E enrichment in the three data sets HEK SARS2 E, Dubca SARS2 E, and PaKiT SARS2 E; (2) MERS-CoV E enrichment in the three data sets HEK MERS E, Dubca MERS E, and PaKiT MERS E; (3) SARS-CoV-2 M enrichment in the three data sets HEK SARS2 M, Dubca SARS2 M, and PaKiT SARS2 M; (4) MERS-CoV M enrichment in the three data sets HEK MERS M, Dubca MERS M, and PaKiT MERS M.

4.2.7.1 *SARS-CoV-2 E enrichment in the HEK SARS2 E, Dubca SARS2 E, and PaKiT SARS2 E data sets:*

The enriched cellular process and pathways differed among SARS-CoV-2 E-expressing cells. This difference was likely due to the smaller numbers of cellular protein interactors in the Dubca SARS2 E and PaKiT SARS2 E data sets compared to the HEK293 SARS2 E data set. Consequently, they did not show as broad an enrichment as the HEK293 SARS2 E data set. Alternatively, this difference could reflect the ability of SARS-CoV-2 E to interact more with HEK293 cellular proteins than with Dubca and PaKiT cellular proteins. Table 4-30 lists the cellular process or pathways enriched in SARS-CoV-2 E expressing HEK293, Dubca, and PaKiT cells. Notably, no cellular processes were shared between two or three of these data sets. The Dubca SARS2 E data set was enriched for proteins involved in RNA binding, isopeptide bond, and nucleus proteins. In addition, the PaKiT SARS2 E data set was enriched for proteins related to cell structures, such as the cytoskeleton and cell-cell junctions. Moreover, the HEK SARS2 E data set was enriched for proteins related to cellular processes such as mitochondrial, ER, fatty acid metabolism, unfolding protein response, and the ubiquitin-dependent protein catabolic process. Consistent with the HEK SARS2 E data set, ER and cytosol trafficking were identified in a previous study (155).

Table 4-30 Cellular process or pathways enriched in the SARS-CoV-2 E protein data sets.

HEK SARS2 E	Dubca SARS2 E	PaKiT SARS2 E
Mitochondrial protein	---	---
ER	---	---
Fatty acid metabolism	---	---
Lysosome	---	---
Iron-sulfur	---	---
ER to Golgi transport	---	---
Protein targeting	---	---
Unfolding protein response	---	---
Ubiquitin-dependent protein catabolic process	---	---
Protein targeting	---	---
---	RNA binding	---
---	Isopeptide bond	---
---	Nucleus	---
---	---	Calponin-homology
---	---	ATP binding
---	---	Cell-cell junction
---	---	Oxytocin signalling pathway
---	---	Metal ion binding
---	---	Cytoskeleton

DAVID and STRING analyses were performed on a proteomic data set for the SARS-CoV-2 E protein. They included three data sets: HEK SARS2 E, Dubca SARS2 E, and PaKiT SARS2 E. A shared localisation, cellular process, and pathway were summarised under one name. For example, ‘mitochondrial’ and ‘oxidative phosphorylation’ were summarised as a ‘mitochondrial protein’.

4.2.7.2 MERS-CoV E enrichment in the HEK MERS E, Dubca MERS E, and PaKiT MERS E data sets:

The cellular process and pathways enriched in MERS-CoV E expressing cells were more enriched compared to those enriched in SARS-CoV-2 E expressing cells. This observation could be due to the differing ability of MERS-CoV E to interact with the three cell types or due to unknown reasons. Table 4-31 lists the cellular processes or pathways enriched in MERS-CoV E expressing HEK293, Dubca, and PaKiT cells. The HEK MERS E data set was enriched for proteins involved in cellular processes such as mitochondrial, ER, lipid biosynthesis, unfolding protein response, and ubiquitin-specific protease binding. In addition, ER to Golgi vesicle-mediated transport was identified in a previous study (153), consistent with the HEK SARS2 E data set. More than one cellular process was shared between two or three of the data sets. These cellular processes included lipid biosynthesis, glycolysis/gluconeogenesis, protein folding and stabilization, and mitochondrial. Some specific cellular processes were enriched in the Dubca MERS E data set, including viral entry into the host cell. In contrast, the PaKiT MERS E data set was enriched for proteins related to cell structures, such as the cytoskeleton and actin cytoskeleton organization.

Table 4-31 Cellular processes or pathways enriched in the MERS-CoV E protein data sets.

HEK MERS E	Dubca MERS E	PaKiT MERS E
Membrane	---	
ER	ER	---
---	Protein folding	Protein folding
Mitochondrial membrane	Mitochondrion	mitochondrion
ER to Golgi vesicle-mediated transport	Golgi membrane	---
Protein N-linked glycosylation	---	---
Lipid biosynthesis	Fatty acid metabolism	---
---	Glycolysis/gluconeogenesis	Glycolysis/gluconeogenesis
Ubiquitin-specific protease binding	---	---
Phospholipid biosynthesis	---	---
Hydrogen ion transmembrane transporter activity	---	---
---	Response to ER stress	Stress response
---	Cellular response to heat	---
---	Calcium-dependent protein binding	---
---	Pyruvate metabolism	---
---	Heparin-binding	---
---	Sterol metabolism	---
---	Positive regulation of the apoptotic process	---
---	Cysteine and methionine metabolism	---
---	Viral entry into the host cell	---
---	Carbon metabolism	---
---	---	Actin cytoskeleton organisation
---	---	Zinc finger, LIM-type
---	---	Protein binding involved in protein folding
---	---	Translational elongation
---	---	Podosome
---	---	Endopeptidase activity
---	---	Valine, leucine, and isoleucine degradation
---	---	Nucleotide-binding
---	---	P-loop containing nucleoside triphosphate hydrolase
---	---	Thioredoxin domain
---	---	Protein stabilization
---	---	NAD binding
---	---	Isomerase
---	---	Citrate (TCA) cycle

---	---	S-nitrosylation
---	---	Clathrin-coated pit
---	---	Extracellular region
---	---	Protein biosynthesis
---	---	Cytoskeleton

DAVID and STRING analyses were performed on a proteomic data set for the MERS-CoV E protein. They included three data sets: HEK MERS E, Dubca MERS E, and PaKiT MERS E. A shared localisation, cellular process, and pathway were summarised under one name. For example, 'mitochondrial' and 'oxidative phosphorylation' were summarised as a 'mitochondrial protein'.

4.2.7.3 *SARS-CoV-2 M enrichment in the HEK SARS2 M, Dubca SARS2 M, and PaKiT SARS2 M data sets:*

The enriched cellular process and pathways differed between SARS-CoV-2 M expressing cells. This difference was likely due to the smaller numbers of cellular protein interactors in the HEK SARS2 M and PaKiT SARS2 M data sets compared to the Dubca SARS2 M data set. Table 4-32 lists the cellular process or pathways enriched in SARS-CoV-2 M expressing HEK293, Dubca, and PaKiT cells. The three data sets shared enrichment in transmembrane and ER proteins, consistent with previous studies (92, 154). In addition, mitochondrial protein enrichment was identified in previous studies (92, 155), consistent with the HEK SARS2 M and Dubca SARS2 M but not the PaKiT SARS2 M data sets. However, while nuclear proteins were previously identified as enriched with SARS-CoV-2 M (92), they were not enriched in our data sets.

Table 4-32 Cellular processes or pathways enriched in the SARS-CoV-2 M protein data sets.

HEK SARS2 M	Dubca SARS2 M	PaKiT SARS2 M
ER	ER membrane	ER
Transmembrane	Transmembrane	Transmembrane
Mitochondrial	Mitochondrial	---
Cell membrane	---	---
Oxidoreductase activity	---	---
---	Golgi membrane	Golgi membrane
---	Lipid metabolism	Lipid metabolism
---	Nucleocytoplasmic transport	---
---	Amino acid transport	---
---	N-glycan biosynthesis	---
---	Nucleocytoplasmic transport	---
---	Calcium signalling pathway	---
---	Retrograde vesicle-mediated transport, Golgi to ER	---
---	ER to Golgi vesicle-mediated transport	---
---	GTPase activity	---
---	Plasma membrane	---
---	Methylation	---
---	---	Protein transport
---	---	Macromolecular complex

DAVID and STRING analyses were performed on a proteomic data set for the SARS-CoV-2 M protein. They included three data sets: HEK SARS2 M, Dubca SARS2 M, and PaKiT SARS2 M. A shared localisation, cellular process, and pathway were summarised under one name. For example, 'mitochondrial' and 'oxidative phosphorylation' were summarised as a 'mitochondrial protein'.

4.2.7.4 MERS-CoV M enrichment in the HEK MERS M, Dubca MERS M, and PaKiT MERS M data sets:

The enriched cellular process and pathways differed between MERS-CoV M expressing cells. This difference was likely due to the smaller numbers of cellular protein interactors in the Dubca MERS M and PaKiT MERS M data sets compared to the HEK MERS M data set. Table 4-33 lists the cellular processes or pathways enriched in MERS-CoV M expressing HEK293, Dubca, and PaKiT cells, which shared cellular proteins involved in mitochondria, nucleus, protein glycosylation, and ER-Golgi transportation. Both mitochondrial and nuclear cellular processes were also identified in the only study examining MERS-CoV M cellular interactors (153).

Table 4-33 Cellular processes or pathways enriched in the MERS-CoV M protein data sets.

HEK MERS M	Dubca MERS M	PaKiT MERS M
Mitochondrial	Mitochondrial	Mitochondrial
Membrane	---	---
Amino acid transport	Amino acid transport	---
Nucleocytoplasmic transport	Nucleocytoplasmic transport	Nuclear structure/Intracellular trafficking and secretion
N-glycan biosynthesis	Protein glycosylation	N-glycan biosynthesis
Retrograde vesicle-mediated transport, Golgi to ER	ER-Golgi transport	ER to Golgi vesicle-mediated transport
Fatty acid metabolism	---	Fatty acid metabolism
Importin-beta, N-terminal	---	---
Sodium transport	---	---
Amino acid transmembrane transporter activity	Amino-acid transport	---
ATPase activity	---	ATPase activity
Peptidase activity	---	---
L-glutamate transmembrane transporter activity	---	---
Neutral amino acid transmembrane transporter activity	---	---
Peroxisomal membrane	---	---
Intercellular bridge	---	---
Sugar transporter, conserved site	---	---
---	ER membrane	ER membrane
---	Transmembrane	Transmembrane
---	Calcium signalling pathway	Calcium signalling pathway
---	Phospholipid biosynthesis	Phospholipid biosynthesis
---	Cholesterol biosynthetic process	---
---	GTP binding	---
---	cGMP-PKG signalling pathway	---
---	Oxidoreductase	---
---	Cellular potassium ion homeostasis	---
---	Lipoprotein	---
---	Golgi transport complex	---
---	---	Ceramide biosynthetic process
---	---	HEAT 10
---	---	Transmembrane beta strand
---	---	Spinocerebellar ataxia
---	---	Microtubule-based process
---	---	Cristae formation
---	---	Oxidoreductase

DAVID and STRING analyses were performed on a proteomic data set for the MERS-CoV M protein. The included three data sets: HEK MERS M, Dubca MERS M, and PaKiT MERS M. A shared localisation, cellular process, and pathway were summarised under one name. For example, ‘mitochondrial’ and ‘oxidative phosphorylation’ were summarised as a ‘mitochondrial protein’.

4.2.8 Comparison between stable and transient HEK293 cell transfections with MERS-CoV-E and -M proteins

It was interesting to compare the two cell transfection methods (constant and transient) used to identify the MERS-CoV E and M cellular interactors. This analysis compared MERS-CoV E and M proteomic data from transiently transfected cells with the previous proteomic data produced by Mr Lee for stably transfected cell lines inducibly expressing FLAG-tagged MERS-CoV E and M proteins (HEK293-E and HEK293-M, respectively). These cells were used for LC-MS/MS-based co-IP analyses to provide proteomic data representing a cellular interactor for MERS-CoV E and M proteins (see Chapter 3). The total numbers of human interacting proteins were 2126 and 494 proteins for MERS-CoV E and M, respectively (Table 3-1). However, the proteomic data produced in this chapter represent transiently transfected cells inducibly expressing FLAG-tagged MERS-CoV E and M proteins (Section 4.2.4). The Fun Rich software was used to mark shared interactor proteins between transiently and stably transfected MERS-CoV E and M proteins in HEK293 cells (Figure 4-36 and Figure 4-37).

Regarding the MERS-CoV E protein, 4617 proteins were reproducibly detected in cell lysates from transiently transfected HEK293 compared to HEK293-Flp (control) cells. In addition, 2127 proteins were reproducibly detected in cell lysates from stably transfected HEK293 compared to HEK293-Flp (control) cell lysates. There were 1057 proteins shared between these two data sets. DAVID and STRING analyses confirmed the similarity between the two transfection methods (stable and transient) used to produce MERS-CoV -E in HEK293 cells. They shared four enrichment clusters: ‘mitochondrial’, ‘endoplasmic reticulum’, ‘membrane’, and ‘lipid biosynthesis’. The results were described in Sections 3.2.4 and 4.2.6.

Regarding the MERS-CoV M protein, 5113 proteins were reproducibly detected in cell lysates from transiently transfected HEK293 cells compared to HEK293-Flp (control) cells. In addition, 494 proteins were reproducibly detected in cell lysates from stably transfected HEK293 cells compared to HEK293-Flp (control) cells. There were 234 proteins shared between these two data sets. DAVID and STRING analyses confirmed the similarity between the two transfection methods (stable and transient) used to produce MERS-CoV M in HEK293 cells. They shared three enrichment clusters, two associated with mitochondrial function (‘mitochondrion’ and ‘oxidative phosphorylation’) and one associated with ‘protein transportation’. The results were described in Sections 3.2.4 and 4.2.6.

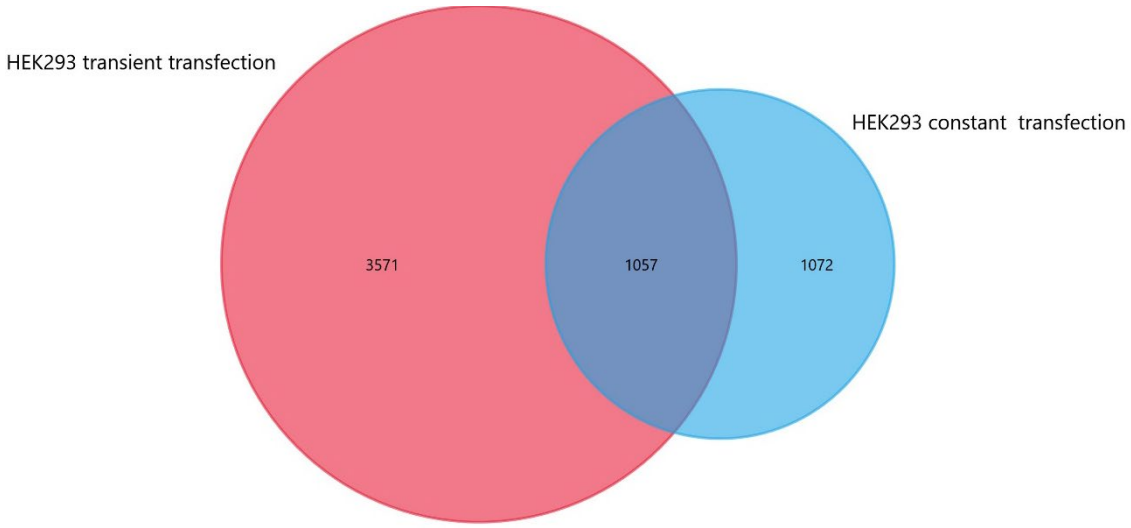


Figure 4-36 Proteins shared between HEK293 cells stably and transiently transfected with MERS-CoV E.

The Venn diagram shows the number of shared cellular proteins interacting with the MERS-CoV E protein in transiently (red) and stably (blue) transfected HEK293 cells. The Fun Rich software was used to construct the Venn diagram.

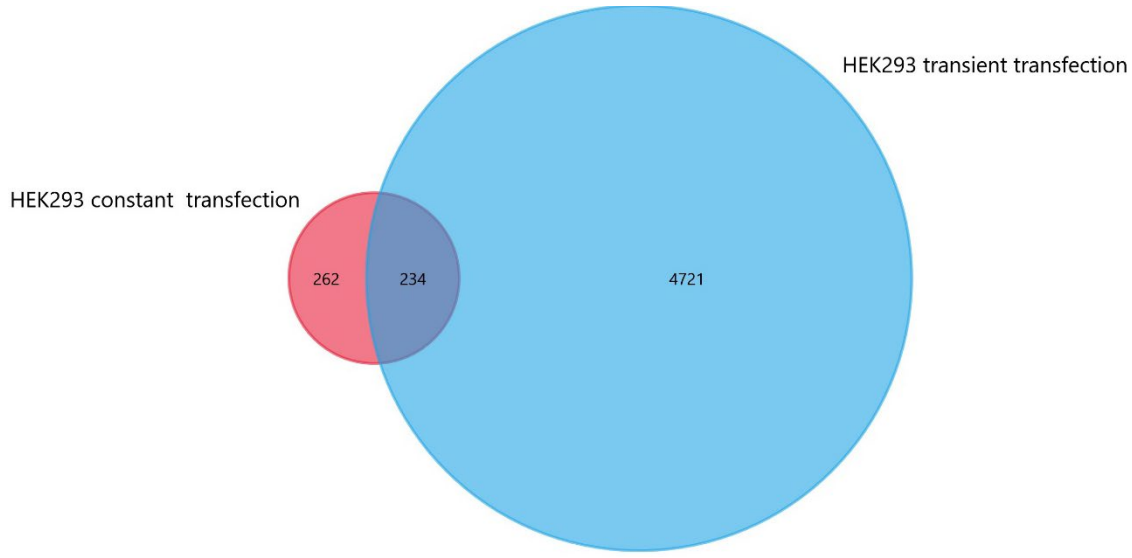


Figure 4-37 Proteins shared between HEK293 cells stably and transiently transfected with MERS-CoV M.

The Venn diagram shows the number of shared cellular proteins interacting with the MERS-CoV M protein in stably (red) and transiently (blue) transfected HEK293 cells. The Fun Rich software was used to construct the Venn diagram.

4.2.8.1 *Comparison of the 11 selected interactors for MERS-CoV E and M proteins in stably transfected HEK293 cells with enrichment in proteomic data for HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV E and M proteins*

Previous proteomic data produced by Mr Lee represented stably transfected cell lines inducibly expressing FLAG-tagged MERS-CoV E and M proteins (see Chapter 3). The numbers of interacting human proteins showing a *p*-value <0.05 and fold increase >1.5 in ≥3 experiments comparing co-IPs for these cell lines to HEK293-Flp cells were 61 and 112 out of 2126 and 783 detected proteins for MERS-CoV E and M, respectively (Table 3-1). Eleven cellular interactors were selected and validated using WB and IFA (Section 3.2.6). This section determined whether these 11 cellular interactors for the MERS-CoV E and M proteins were recognised by the SARS-CoV-2 and MERS-CoV E and M proteins or showed significant increases in anti-FLAG co-IPs from lysates of transiently transfected HEK293, Dubca, and PaKiT cells compared to negative control cells (Section 4.2.4). The comparison results are reported using the names of the 12 data sets (see Table 4-2).

The six selected MERS-CoV E interactors were TM9SF2, TMEM43, YIPF5, CERS2, SLC44A2, and ERGIC1 (Table **3-4**). Comparing the six selected MERS-CoV E interactors with the 12 data sets showed that all except SLC44A2 were in the HEK MERS E and HEK SARS2 E data sets. However, only TM9SF2 was in the PaKiT MERS E and PaKiT SARS2 E data sets. Moreover, none were in the Dubca MERS E or Dubca SARS2 E data sets. See Table 4-34.

In addition, these six MERS-CoV E interactors were in the HEK MERS M data set, except for SLC44A2. Moreover, only TM9SF2, CERS2, and ERGIC1 were in the HEK SARS2 M data set. However, none were in the PaKiT MERS M and PaKiT SARS2 M data sets, while all except SLC44A2 were in the Dubca MERS M and Dubca SARS2 M data sets. See Table 4-35.

The five selected MERS-CoV M interactors were LPCAT1, SCAMP3, IPO11, VDAC1, and RAB10 (Table **3-5**). Comparing the five MERS-CoV M interactors with the 12 data sets showed that all were in the HEK MERS E and HEK SARS2 E data sets. However, only LPCAT1, SCAMP3, and IPO11 were in the PaKiT MERS E and PaKiT SARS2 E data sets. Moreover, only VDAC1 and RAB10 were in the Dubca MERS E data set, and none were in the Dubca SARS2 E data set. See Table 4-36.

In addition, while all five MERS-CoV M interactors were in the HEK MERS M data set, only IPO11 and VDAC1 were in the HEK SARS2 M data set. In addition, only LPCAT1 and IPO11 were in the PaKiT MERS M and PaKiT SARS2 M data sets. Moreover, all except RAB10 were in the Dubca MERS M and Dubca SARS2 M data sets. See Table 4-37.

Table 4-34 The log_2 fold changes of the selected MERS-CoV E protein interactors in stably transfected HEK293 cells in the data sets containing proteins with a log_2 fold change >0 (*t*-test <0.05) in anti-FLAG co-IPs from lysates of HEK293, PaKiT, and Dubca cells transfected with plasmids encoding two viral proteins (SARS-CoV-2 and MERS-CoV E) compared to negative control cells.

Protein symbol	Log_2 HEK MERS E	Log_2 HEK SARS2 E	Log_2 PaKiT MERS E	Log_2 PaKiT SARS2 E	Log_2 Dubca MERS E	Log_2 Dubca SARS2 E
TM9SF2	2.96	1.45	1.56	1.23	---	---
TMEM43	3.25	0.95	---	---	---	---
YIPF5	3.13	1.78	---	---	---	---
CERS2	1.28	0.93	---	---	---	---
SLC44A2	---	---	---	----	---	---
ERGIC1	2.97	1.80	----	----	---	---

Table 4-35 The log_2 fold changes of the selected MERS-CoV E protein interactors in stably transfected HEK293 cells in the datasets containing proteins with a log_2 fold change >0 (*t*-test <0.05) in anti-FLAG co-IPs from lysates of HEK293, PaKiT, and Dubca cells transfected with plasmids encoding two viral proteins (SARS-CoV-2 and MERS-CoV M) compared to negative control cells.

Protein symbol	Log_2 HEK MERS M	Log_2 HEK SARS2 M	Log_2 PaKiT MERS M	Log_2 PaKiT SARS2 M	Log_2 Dubca MERS M	Log_2 Dubca SARS2 M
TM9SF2	3.52	1.41	---	---	1.90	1.43
TMEM43	2.21	---	---	---	1.97	1.62
YIPF5	2.87	---	---	---	2.63	2.10
CERS2	3.77	1.79	---	---	1.66	1.28
SLC44A2	---	---	---	---	---	---
ERGIC1	3.15	1.88	---	---	1.12	0.55

Table 4-36 The log_2 fold changes of the selected MERS-CoV M protein interactors in stably transfected HEK293 cells in the data sets containing proteins with a log_2 fold change >0 (*t*-test <0.05) in anti-FLAG co-IPs from lysates of HEK293, PaKiT, and Dubca cells transfected with plasmids encoding two viral proteins (SARS-CoV-2 and MERS-CoV E) compared to negative control cells.

Protein symbol	Log_2 HEK MERS E	Log_2 HEK SARS2 E	Log_2 PaKiT MERS E	Log_2 PaKiT SARS2 E	Log_2 Dubca MERS E	Log_2 Dubca SARS2 E
LPCAT1	2.09	2.09	1.13	0.97	---	---
SCAMP3	2.18	1.66	1.35	0.90	---	---
IPO11	1.58	1.35	1.28	0.79	---	---
VDAC1	3.19	2.15	---	---	0.71	---
RAB10	1.61	1.28	---	---	0.66	---

Table 4-37 The log_2 fold changes of the selected MERS-CoV M protein interactors in stably transfected HEK293 cells in the data sets containing proteins with a log_2 fold change >0 (*t*-test <0.05) in anti-FLAG co-IPs from lysates of HEK293, PaKiT, and Dubca cells transfected with plasmids encoding two viral proteins (SARS-CoV-2 and MERS-CoV M) compared to negative control cells.

Protein symbol	Log_2 HEK MERS M	Log_2 HEK SARS2 M	Log_2 PaKiT MERS M	Log_2 PaKiT SARS2 M	Log_2 Dubca MERS M	Log_2 Dubca SARS2 M
LPCAT1	3.54	---	2.14	1.84	1.69	1.13
SCAMP3	3.07	---	---	---	1.66	1.25
IPO11	3.25	1.61	2.36	1.83	2.12	1.63
VDAC1	2.53	1.29	---	---	1.99	1.69
RAB10	2.66	---	---	---	---	---

4.3 Discussion

4.3.1 Difference between transiently and stably transfected cells

The results described in Chapter 3 were produced using HEK293 cells stably expressing the MERS-CoV E and M proteins. In this chapter, the interactomes of the MERS-CoV and SARS-CoV-2 E and M proteins were investigated using cell lines from three species. However, producing cell lines for the three species that stably expressed the two viruses' E and M proteins was unrealistic. Therefore, as in some previous studies investigating the interactomes of coronavirus proteins (92, 103, 153-155), the interactomic analysis used cells transiently expressing the viral proteins. Nevertheless, it was important to compare the findings of interactomic analyses using HEK293 cells stably and transiently expressing the MERS-CoV E and M proteins to determine the limitations of the two methods (described in Section 4.2.8). Approximately 50% (1057 proteins) and 47% (234 proteins) of cellular interactors identified using HEK293 cells stably expressing the MERS-CoV E and M proteins, respectively, were identified using HEK293 cells transiently expressing them. Importantly, downstream proteomic analysis of the interactomes generated using HEK293 cells stably or transiently expressing the MERS-CoV E and M proteins using DAVID and STRING analyses showed high similarities in the enriched GO terms.

Regarding the 11 selected and validated cellular interactors identified in HEK293 cells stably expressing the MERS-CoV E and M proteins, all except SLC44A2 were identified as interactors in HEK293 cells transiently expressing the MERS-CoV E and M proteins (Section 4.2.8.1). Some of these 11 interactors had been previously reported as coronavirus protein interactors, including TMEM43 protein with the MERS-CoV E protein (153) and IPO11 with the MERS CoV, SARS-CoV, and SARS-CoV-2 M proteins (153).

4.3.2 Bioinformatic analysis of proteomic data sets

While the WB analysis successfully confirmed the production of the FLAG-tagged SARS-CoV-2 and MERS-CoV E and M proteins in HEK293, Dubca, and PaKiT cells, the IFA was performed only in Dubca cells (Section 4.2.2) due to the COVID-19 pandemic and confirmed their expression. Moreover, the localization of MERS-CoV E and M proteins stably expressed in HEK293 cells was determined (Section 3.2.1). A previous study found that the SARS-CoV-2 M protein was party concentrated in ER (103).

Most interactomic studies investigating the SARS-CoV-2 and MERS-CoV E and M proteins used HEK293 and A549 lung carcinoma cells (153-155). Table 4-1 summarises the key findings of these studies, including the cell line used and the identified cellular processes, pathways, and cellular protein interactors. Unlike this study, no previous studies reported using Dubca or PaKiT cells for coronavirus protein expression.

The bioinformatics analysis of the interactomic data sets described in Section 4.2.6 identified several cellular processes that might be important for viral replication. Moreover, a comparison was made to viral protein levels in Section 4.2.7, including comparing data sets for each viral protein in three host cells. Sections 4.3.2.1 and 4.3.2.2 below highlight interesting and shared pathways or cellular processes that include proteins identified as interactors of the SARS-CoV-2 and MERS-CoV E and M proteins. Then, interesting or shared cellular processes are discussed in more detail in Section 4.3.2.3.

4.3.2.1 Interesting cellular processes in the proteomic analyses of HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV E proteins:

This section discusses DAVID and STRING analyses of interactomic data generated using HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV E proteins.

The enrichment analyses of the HEK SARS2 E and HEK MERS E data sets revealed shared and potentially interesting cellular processes, including the ‘ER unfolded protein response”, ‘lipid biosynthesis’, ‘calcium signalling pathway’, and ‘protein targeting’. One unique enrichment cluster for the HEK MERS E data set was ‘ubiquitin-specific protease binding’. This cluster comprised multifunctional proteins, including those involved in the unfolded protein response and protein transportation to the cytosol.

The enrichment analyses of the Dubca SARS2 E and Dubca MERS E data sets revealed shared and potentially interesting cellular processes, including ‘coronavirus disease-COVID-19’. This enrichment cluster comprised proteins involved in the nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, translational initiation, viral transcription, and ubiquitin ligase inhibitor activity. Another shared cellular process was the ‘positive regulation of gene expression’. This enrichment cluster also comprised proteins involved in ubiquitin ligase inhibitor activity. Moreover, two cellular processes were identified for the Dubca MERS E data set that included enrichment clusters related to the protein folding process: ‘protein folding’ and ‘chaperone’.

The enrichment analyses of the PaKiT SARS2 E and MERS E data sets revealed significant shared and potentially interesting cellular processes, including 'calcium ion binding' and other cellular processes involved in cell structural formation such as 'cytoskeleton' and 'actin filament'. The latter two cellular processes were only identified for the PaKiT E data set. Moreover, two interesting cellular processes were identified for the PaKiT MERS E data set: 'protein folding' and 'chaperone-mediated protein folding'.

A previous proteomic study on the SARS-CoV-2 and SARS-CoV E proteins identified important cellular processes, including ER and cytosol trafficking (155). In addition, vesicle transport and aminoglycan metabolism were identified as important cellular processes for SARS-CoV-2, MERS-CoV, and SARS-CoV E proteins (153). This cellular process was identified with most of the proteomic data sets for HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV E proteins. The exceptions were the PaKiT SARS2 E and Dubca SARS2 E data sets, possibly due to their small numbers of proteins.

Interesting and/or shared cellular processes for the SARS-CoV-2 and MERS-CoV E protein data sets were selected for more detailed discussion in Section 4.3.2.3, including folded and unfolded proteins and lipid biosynthesis; ubiquitination; and coronavirus disease-COVID-19.

4.3.2.2 Interesting cellular processes in the proteomic analysis of HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV M proteins:

This section discusses DAVID and STRING analyses of significantly increased proteins in the proteomic data for HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV M proteins compared to negative control cells.

The enrichment analyses of the HEK SARS2 M and HEK MERS M data sets revealed shared and expected cellular processes, including 'intracellular protein transport'. This cluster included nucleocytoplasmic carrier activity proteins such as IPO11, which had been identified as a SARS-CoV-2 and MERS-CoV M protein interactor in another study (153). Another shared cellular process was 'sodium transport', including the ATPase Na^+/K^+ transporting subunit beta 1 (ATP1B1) protein that plays a role in ion transmembrane transport. A previous study had identified ATP1B1 as a SARS-CoV-2 mitochondrial protein interactor (154). Moreover, the shared cellular 'organonitrogen compound biosynthetic process' included mitochondrial protein STT3 oligosaccharyltransferase complex catalytic subunit A (STT3A), which has been identified as important for the SARS-CoV-2 S protein glycosylation process since inhibiting STT3A reduced SARS-CoV-2 infectivity (230). One unique enrichment cluster was revealed

for the HEK MERS M data set, ‘glycan biosynthesis’, which contained a protein involved in the glycosylation process (STT3A).

The enrichment analyses of the Dubca SARS2 M and Dubca MERS M data sets revealed shared and expected cellular processes, including ‘n-glycan biosynthesis’, which included an important protein for the glycosylation process (STT3A), and the ‘calcium signalling pathway’.

The enrichment analyses of the PaKiT SARS2 M and MERS M data sets revealed shared and expected cellular processes, including ‘mitochondrial’, ‘Golgi apparatus’, ‘Golgi organisation’, and ‘ER membrane’. Moreover, other cellular processes were identified for the PaKiT MERS M data set, including the ‘calcium signalling pathway’, ‘glycan biosynthesis’, and the ‘ceramide biosynthetic process’. Proteins in the last cellular processes were involved in cristae formation in the inner mitochondrial membrane.

A previous proteomic study on the SARS-CoV-2 and MERS-CoV M proteins identified shared cellular processes, including ‘lipid metabolism’, ‘solute transport’, ‘vesicle transport’, ‘mitochondrial RNA processing’, ‘integrator complex’, ‘ion transport’, and ‘RNA nuclear export’ (153). Moreover, another study on the SARS-CoV-2 and SARS-CoV M proteins identified other shared cellular processes, including ‘lipid oxidation’, ‘ER and Golgi trafficking’, ‘ions and transport by ATPases’, and the ‘condensin II complex’ (155). Other studies on the SARS-CoV-2 M protein identified cellular processes such as ‘mitochondrial metabolism’, ‘ER morphology’, ‘plasma membrane proteins’, and ‘cytosol proteins’ (92, 154). These cellular processes were identified in most of the proteomic data sets for HEK293, Dubca, and PaKiT cells transiently expressing SARS-CoV-2 and MERS-CoV M proteins. The exceptions were the PaKiT SARS2 M and HEK SARS2 M data sets, possibly due to their small numbers of proteins. One interesting and/or shared cellular process for the SARS-CoV-2 and MERS-CoV M proteins (‘glycosylation’) was selected for more detailed discussion below.

4.3.2.3 The interesting cellular processes for SARS-CoV-2 and MERS-CoV E and M proteins:

4.3.2.3.1 Ubiquitination:

The ubiquitin-proteasome system (UPS) breaks down most cytosolic short-lived and misfolded proteins (231). Numerous processes, such as cell-cycle progression, signal transduction, transcriptional regulation, and endocytosis, are controlled by the degradation of regulatory proteins (232). During ubiquitination, a target protein’s lysine residue is attached covalently to

the C terminus of ubiquitin, a small protein comprising 76 amino acids (233). One (monoubiquitination) or more (polyubiquitination) ubiquitin molecules can be attached. A group of enzymes performs ubiquitination: ubiquitin ligase, ubiquitin-activating enzyme (E1), ubiquitin transferring enzyme (E2), ubiquitin ligase (E3), and ubiquitin chain elongation/ubiquitination factor enzyme E (234). Then, the 26S proteasome can degrade the ubiquitinated protein (235).

During viral infection, the UPS is crucial for maintaining cellular environmental homeostasis. The UPS is crucial in various stages of the viral life cycle, including capsid uncoating, replication, transcription, envelopment, and viral progeny release (25). It is crucial at different points in the coronavirus life cycle (236). Moreover, certain viruses encode their own E3 ligases, and others can recruit cellular E3 ligases to target and destroy anti-viral proteins (234). We hypothesise that the UPS is used to promote coronavirus infection in three ways. Firstly, viruses can reprogram the UPS to meet their needs and use it to degrade cellular restriction factors instead of viral proteins (234). Secondly, the UPS system can be used to target viral proteins (234). Numerous viral proteins are ubiquitinated and subsequently degraded by the UPS. For example, dengue virus capsid protein degradation facilitates the viral genome transfer into host cells (237). Thirdly, the UPS system can be used to avoid protein accumulation during viral infection. According to one study, the accumulation of cellular protein-ubiquitin conjugates during viral infection reflects the importance of increasing the ubiquitination process for viral replication (238). This situation can explain the accumulation of proteins with proteasome activity inhibition, causing ER stress followed by cell death and inhibiting viral replication (239). MG132 was used as a proteasome inhibitor to reduce the viral protein expression of some coronaviruses, including SARS-CoV (236, 240). Several proteasome inhibitors have shown an effect on viral replication, pneumonia, and acute respiratory distress syndrome, making them a promising viral infection treatment (241).

UBA52

Ubiquitin-encoding genes include monomeric ubiquitin-60S ribosomal protein L40 (*UBA52*) and ubiquitin-40S ribosomal protein S27a (*RPS27A*) (233). These genes encode one ubiquitin unit fused to a ribosomal protein. *UBA52* is known as a housekeeping gene (242). However, ubiquitins B (*UBB*), C (*UBC*) and D (*UBD*) are known as polyubiquitin genes (233, 242). UBA52 was shown to be a ubiquitin supplier and ribosomal protein complex regulator (243). The *UBA52*-encoded fusion protein comprises ubiquitin at the N terminus and ribosomal protein L40 (RPL40) at the C terminus (244). The siRNA knockdown of *UBA52* significantly reduced polyubiquitinated protein and free ubiquitin levels (236). While the ubiquitination process was enriched only in the HEK E data set, UBA52 was a shared cellular protein

interactor in the HEK293 E, PaKiT E, and Dubca E data sets. Two previous interactome studies did not identify ubiquitination as an important biological process for SARS-CoV-2, MERS-CoV, and SARS-CoV (153, 154). Moreover, UBA52 was identified as an influenza A virus-host protein interactor using chicken embryonic fibroblast cells, and the H5N1 titter significantly decreased after *UBA52* knockdown (244).

Deubiquitylation

Another way to use the UPS to benefit viral infection is using the deubiquitinating enzyme (DUB). DUB removes ubiquitin from tagged proteins as a reverse pathway (245). Each Coronaviridae family members encode a DUB, named papain-like proteases (PLPs), which remove ubiquitin from target proteins and alter cellular pathways important for infection (246). The antiviral pathways include producing ubiquitin-dependent cytokines, such as IL-6, IL-8, and TNF-α (247). The innate immune response cascade includes ubiquitinated factors (247). DUB removed ubiquitin and prevented excessive inflammation (247). During MERS-CoV and SARS-CoV infections, DUB activities suppressed the host antiviral pathways (248). In addition, SARS-CoV PLPs deubiquitinated interferon regulatory transcription factor 3 (IRF3), reducing type I interferon production (249).

4.3.2.3.2 Glycosylation:

Protein glycosylation is essential for mammal cells and is crucial in determining the structure, function, and stability of cellular proteins (250). N-glycans are common, structurally diverse, and clinically relevant molecules attached to secretory and membrane proteins (251). N-linked glycosylation is performed in the ER lumen by dolichyl-diphosphooligosaccharide protein glycosyltransferase (OST), a membrane-associated enzyme complex. It has two distinct OST catalytic subunits, A (STT3A) and B (STT3B) (252). When the STT3A isoform enters the ER lumen, it is primarily responsible for the N-linked glycosylation of most target protein sites (230). In contrast, complexes containing STT3B are necessary for effective post-translational glycosylation and mediate glycosylation of sites missed by STT3A (252). Viruses can hijack the glycosylation machinery in the ER-Golgi system of host cells, producing progeny virions with the host cell's characteristic glycosylation pattern (253). Importantly, glycosylation may enhance a virus's ability to evade immune surveillance by covering its proteins and preventing antibody recognition (230). Inhibition of N-glycosylation by the OST inhibitor NGI-1 disrupted virion formation and significantly reduced the infectiousness of herpes simplex and SARS-CoV-2 viruses (254).

While SARS-CoV-2 S, E, and M proteins are glycosylated (44), not much is known about E and M proteins compared to the S protein. The glycosylation sites for the SARS-CoV-2 M protein are N5, N21, N41, N43, N117, N121, N203, and N216 (255). Six of these eight sites are shared by SARS-CoV-2 and SARS-CoV in humans (255). The two possible E protein glycosylated sites are N66 and N48, located in its C-terminus (255). The SARS-CoV-2 S protein is heavily N-glycosylated with glycans (256). Glycans on viral surface proteins aid viral binding to host cells for entry, viral fusion, shielding of specific epitopes, folding, stability, and facilitating immune evasion by inhibiting the humoral and cellular innate immune systems (254, 257). One study showed the effectiveness of inhibitors of the ER α-glucosidases, which are essential for the maturation of N-glycan structures, in reducing viral infection (258). In addition, siRNA was used to inhibit the oligosaccharyltransferase catalytic subunit STT3 in both A and B isoforms to prevent the transfer of N-glycan precursors to ER proteins, reducing SARS-CoV-2 infection in HEK293 cells by 54.2% (256). Moreover, another study used the N-linked glycosylation inhibitor-1 (NGI-1) as an STT3A inhibitor, which reduced SARS-CoV-2 infectivity (230).

N-glycan biosynthesis or protein glycosylation was enriched in all SARS-CoV-2 and MERS-CoV M data sets except HEK SARS2 M and PaKiT SARS2 M. These two data sets contain small numbers of proteins, possibly leading to STT3A not being identified as a shared cellular protein interactor of the SARS-CoV-2 and MERS-CoV M proteins. This issue reflects one limitation of this study. While glycosylation is important for the viral S protein, our data shows that it is also important for the SARS-CoV-2 and MERS-CoV M proteins. Our results show that proteins involved in N-glycan biosynthesis and protein glycosylation cellular processes are also ER proteins (STT3A, RPN2, RPN1, DPAGT1, and ALG1), which could be studied to determine how they are important for coronaviral replication.

The cross-species transmission and properties of the virus could be explained by the structural similarity of the human SARS-CoV-2 M and E proteins to their cross-specific SARS-CoV proteins (259).

4.3.2.3.3 Unfolded proteins, folded proteins, and lipid biosynthesis:

Viruses appear to use all cell machinery to build and produce their proteins. The main location for viral protein synthesis and maturation is the ER (260). The ER and coronavirus replication are physically and functionally related (261). The ER is associated with protein glycosylation, lipid synthesis, and ubiquitination. Moreover, unfolded protein response (UPR) signalling

pathways are triggered in response to ER stress, which is induced by coronavirus infection (261).

The virion budding process changes ER morphology and membrane depletion by depleting phosphatidylcholine, the ER membrane's lipid component (262), inducing ER stress, activating the UPR, and increasing lipid biosynthesis and ER membrane biogenesis (263).

In this study, the identified HEK E, Dubca E, and PaKiT E data set enrichments shared or had interesting cellular processes, including 'unfolded protein', 'folded protein', 'chaperones and lipid biosynthesis'. Moreover, an interaction between the S protein and chaperone proteins, such as calreticulin and calnexin, was confirmed, leading to a hypothesis that the S protein was responsible for ER stress activation according to massive protein production (261). The overexpression of SARS-CoV, MHV, and HCoV-HKU1 S proteins induced ER stress in cell culture (264). In addition, glucose-regulated protein 94 (GRP94) is a known ER stress indicator (265) that was found in cells infected with SARS-CoV (264). Based on these coronavirus S protein functions and the E protein data set enrichments, we hypothesised that the virus uses protein folding machinery and chaperones in addition to protein production machinery at the start of infection. Viral protein accumulation induces cell stress and activates the UPR, followed by ER membrane depletion. In the late stage of viral infection, the virus benefits from apoptosis. The virus could be engulfed by the nearest cells or phagocytes, allowing it to spread without activating the immune response (266).

Regarding the lipid biosynthesis cellular process, a previous study reported that the SARS-CoV E protein controls X-box binding protein 1 (XBP1) splicing. XBP1 is an important protein connected with lipid biosynthesis and the UPR (261). Moreover, a strong association was found between intracellular cholesterol levels and SARS-CoV-2, HCoV-229E, and HCoV-OC43 infections (267).

4.3.2.3.4 COVID-19 enrichment cluster in the Dubca E data set:

This enrichment cluster included 26 proteins identified as ribosomal proteins involved in nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, translational initiation, and viral transcription. The STRING analysis for this enrichment indicated that it involves the translational initiation process, with three proteins classified as ubiquitin ligase inhibitors (RPS7, RPL23, and RPL11). Studying these proteins could be interesting. Moreover, this enrichment included UBA52, which was known as important for the UPS.

CHAPTER 5 Investigation of the role of cellular proteins conserved across three species that interact with the MERS-CoV and SARS-CoV-2 E and M proteins in the virus lifecycle.

5.1 Introduction

To date, most antivirals act by directly targeting viral proteins. However, targeting host factors required for productive virus replication is gaining popularity. Viruses have significantly less coding space in their genomes than their hosts and thus rely on host proteins to supplement the activities of their proteins. Moreover, conventional antivirals target viral enzymes that readily undergo mutation, resulting in drug resistance. Targeting the host proteins required for viral replication can avoid resistance and develop broad-spectrum therapeutics (153). Indeed, families of viruses frequently share cellular pathways and processes (153). It is necessary to characterise the host targets of coronaviruses in human, bat, and camel cells to understand better their roles during the virus life cycle. Most coronaviral (SARS-CoV-2, SARS-CoV, and MERS-CoV) proteins show the same localisation pattern (153), supporting the previous hypothesis. Coronaviruses must encode components capable of modulating cellular signalling pathways and shaping their distinct immune signature.

Recently, several significant studies have used a highly effective gene expression and regulation tool to understand the intricate relationship between the coronavirus and the host cell response. Some studies have used RNA interference or siRNAs (153), while others have used CRISPR screening (267-272). The siRNA experiment begins with designing one or more siRNAs to target the desired genes, followed by their introduction into cells to reduce target gene expression at the mRNA level. One advantage of siRNA is that eukaryotic cells naturally contain the required endogenous machinery. Therefore, only a few components must be delivered into the cells, simplifying the experimental procedure. In contrast, CRISPR-based genome editing needs two components: a guide RNA and a CRISPR-associated endonuclease protein (Cas). The CRISPR experiment starts by designing efficient and specific guide RNAs, followed by transfecting cells with plasmids encoding the guide RNA and Cas protein. After guide RNA transfection into cells, CRISPR generates knockouts at the DNA level. The efficiency of gene knockdown is typically determined at the final step by quantifying protein levels using immunofluorescence experiments.

Various studies have successfully identified host factors required for coronavirus infection. Baggen et al. (268) conducted a series of genome-wide CRISPR-based genetic screens using a human liver cell line (Huh-7) to identify host factors required for SARS-CoV-2 and HCoV-229E infection. Phosphatidylinositol 3-kinase type 3 (PI3KC3) was identified as an important shared host factor for SARS-CoV-2, HCoV-229E, and HCoV-OC43, indicating that small molecules targeting this protein might be useful as a broad-spectrum coronavirus inhibitor (268). In addition, the lysosomal transmembrane protein 106B (TMEM106B) was identified as a critical specific host factor for SARS-CoV-2 infection (268). Similarly, another study reported that the lysosomal protein TMEM106B appeared to be associated with SARS-CoV-2 infection (267). The second study used genome-wide CRISPR-knockout screens in Huh7.5 cells to identify host factors (267). Moreover, along with virus-specific entry factors, the cholesterol homeostasis gene sterol regulatory element-binding protein cleavage-activating protein (SCAP) was identified as a shared host interactor required for infections by SARS-CoV-2 and the common cold coronaviruses OC43 and 229E (267).

A third study by Hoffmann et al. (269) used 332 of the recently published SARS-CoV-2 protein interactors (154) to build a highly focused CRISPR-Cas9 library. CRISPR-based genetic screens were conducted in hepatoma (Huh-7.5) cells to identify host factors required for SARS-CoV-2, HCoV-229E, HCoV-NL63, and HCoV-OC43 infection at two physiologically relevant temperatures (269). Eighty-seven of the 332 (26%) putative SARS-CoV-2 interaction partners were required for SARS-CoV-2 infection (269). This study validated the requirement of seven SARS-CoV-2 M protein interactors for SARS-CoV-2 replication (269).

Another validation study by Schneider et al. (270) selected 19,114 human genes from the Brunello genome-wide library (273) for CRISPR-based knockout in Huh-7.5 cells to identify an absolute requirement for infection by three seasonal coronaviruses: SARS-CoV-2, HCoV-OC43, HCoV-NL63, and HCoV-229E (270). These requirements included some host proteins such as vacuole membrane protein 1 (VMP1) and transmembrane protein 41B (TMEM41) (270). In addition, this study highlighted significant reliances on specific pathways, including glycosaminoglycan biosynthesis, sterol regulatory element-binding protein (SREBP) signalling, bone morphogenetic protein (BMP) signalling, and glycosylphosphatidylinositol biosynthesis (270).

Most host interactor validation studies attempted to highlight important host interactors for two or more coronaviruses to increase their chances of identifying important interactors for most coronaviruses. While the above studies did not validate the most important host interactors with MERS-CoV, Gordon et al. (153) identified and validated a MERS-CoV host interactor with

SARS-CoV-2 and SARS-CoV (153). This study ensured a broad coverage of potential hits by using two distinct cell lines and two gene knockdown-down methods (153). Their siRNA knockdown was performed in A549 cells stably expressing ACE2 (A549-ACE2), and their CRISPR-based knockout was performed in Caco-2 cells (153). For example, the sigma non-opioid intracellular receptor 1 (SIGMAR1) was identified as a functional host-dependency factor, consistent with a previous report of sigma receptor ligand antiviral activity (154). The genetic findings from both cell lines were constructed into a network of PPIs for the SARS-CoV-2, SARS-CoV, and MERS-CoV proteins (153). One common point in this study was identifying shared protein interaction partners required for SARS-CoV-2 and MERS-CoV replication. Notably, they observed enrichment of genetic hits encoding proteins that interact with viral protein Nsp7, which shared a high degree of interaction for all three viruses (153). For example, prostaglandin E synthase 2 (PTGES2) functionally interacted with Nsp7 from SARS-CoV, SARS-CoV-2, and MERS-CoV (153). Table 5-1 lists most published studies.

Table 5-1 Summary of cellular proteome validation studies.

Virus	Cell line	Validation method	Validated protein interactors for E, M, or other viral proteins		Ref.
SARS-CoV-2	Huh7.5	CRISPR	Other	TMEM106B	(267, 268)
SARS-CoV-2	A549-A2-T2 and Huh-7.5	siRNA and CRISPR	M	SLC30A9 TUBGCP2 ATP6V1A TARS2 PITRM1 COQ8B INTS4	(269)
SARS-CoV-2	Huh-7.5	CRISPR	Other	VMP1 TMEM41 TMEM64 TMEM41B	(270)
SARS-CoV-2, SARS-CoV, and MERS-CoV	A549-ACE2 and Caco-2	siRNA and CRISPR	M	ETFA BZW2 SAAL1 FASTKDS	(153)
			E	BRD2 BRD4 ZC3H118	(153)

As described previously in Chapter 4, a high-throughput co-IP proteomic analysis was used to identify proteins with significantly higher levels in the lysates of transfected cells compared to negative control cells. Three host cells were transiently transfected with four plasmids encoding SARS-CoV-2 and MERS-CoV E and M proteins, creating 12 data sets. A bioinformatics comparison of the data sets for human, bat, and camel cells identified 21 and 12 highly conserved interactors shared by SARS-CoV-2 and MERS-CoV E and M proteins, respectively (see Section 4.2.5). Moreover, 11 interesting cellular protein interactors were selected and validated for the MERS-CoV E and M proteins (see Section 3.2.5).

This chapter aimed to validate the importance of the selected proteins that were identified to be significantly increased by proteomic analysis for SARS-CoV-2 replication. These cellular proteins include those that specifically interacted with the MERS-CoV E and M proteins and those that commonly interacted with the SARS-CoV-2 and MERS-CoV E and M proteins in cells from three hosts. This validation used siRNA knockdown targeting selected proteins followed by SARS-CoV-2 infection. Then, the effects of depleting each protein on SARS-CoV-2 replication were investigated by measuring virus production using the IFA. Both IFA and WB analyses were used to confirm the significant depletion of proteins by siRNA knockdown. While this approach could examine MERS-CoV or SARS-CoV-2 or both, the biosafety level 3 lab was used for the SARS-CoV-2 projects after the COVID-19 pandemic, and it was impossible to shut them down to use it for the MERS-CoV experiments. Moreover, the pandemic was affecting the supply chain including shortage in the reagents and also, caused reducing the number of lab worker to reduce the chance of getting infected during the pandemic. Before investigating the effect of selected protein depletion on SARS-CoV-2 replication, siRNA knockdown conditions had to be optimised.

5.2 Results

5.2.1 Optimisation of siRNA knockdown conditions

Before starting the main siRNA knockdown experiments, the experimental conditions had to be optimised to ensure the lowest production of the target protein without affecting cellular function (i.e. without cellular toxicity). As shown in Chapter 3, this study began by identifying MERS-CoV E and M protein interactors. After the COVID-19 pandemic, the SARS-CoV-2 E and M protein interactors were included, and Chapter 4 described shared MERS-CoV and SARS-CoV-2 E and M protein interactors. In this chapter, it was decided to measure SARS-CoV-2 replication after the siRNA knockdown of the interested cellular proteins. Out of all the cells that could be used for the siRNA knockdown experiments, the A549 cell line expressing

ACE2 and *TMPRSS2* (A549-A2-T2) was chosen because it could be transfected with siRNAs, could be infected by SARS-CoV-2, and was available in the lab.

An initial optimisation examined transfection efficiency using an oligonucleotide duplex labelled with green fluorescein amidite (siGLO). The transfection reagent alone was used as the control. Initially, three different A549-A2-T2 cell counts were used (2500, 5000 and 10000 cells/well) with three different siGLO concentrations (25, 50, and 100 nM) in a 96-well imaging microplate, and DharmaFECT 1 (DF1) was used as the transfection reagent. At 48 h post-transfection (hpt), the cells were examined for siGLO signals under a microscope. The results showed that the A549-A2-T2 cells were successfully transfected with siGLO (Figure 5-1), and the best transfection conditions were 100 nM of siGLO, 5000 cells/well, and 0.1 µl of DF1/well. However, this concentration was rather high compared to the standard siRNA amount, typically 10–25 nM, and a concentration of 100 nM is known to induce cell toxicity. Therefore, in the next optimisation stage, different transfection reagents were used to determine whether they allowed lower siRNA amounts to be used since higher amounts often result in cell toxicity.

The next experiment tested knockdown efficiency using four pooled siRNAs targeting LAMIN A/C and non-targeting siRNAs (NON-TAR) at 10, 20, and 50 nM concentrations, with transfection reagent alone used as the control. LAMIN A/C is a control target protein that is easily visualised and does not kill the cells, while the NON-TAR was used to assess non-targeting effects.

The next experiments tested siRNA concentrations of 10, 20, and 50 nM, which are more reasonable than 100 nM. While the manufacturer recommended using the DF1 transfection reagent for Dharmacon siRNAs with A549-A2-T2 cells, the lipoRM transfection reagent was also used to determine whether it increased transfection efficiency. The two transfection reagents were used at three concentrations (0.1, 0.15 and 0.2 µl/well) with two A549-A2-T2 cell counts (2500 and 4000/well).

On the day of transfection, the transfection reagent was mixed with Opti-MEM to a final volume of 15 µl/well and incubated for 5 min at RT. Next, each siRNA was diluted in 1× siRNA buffer to make a 1 µM stock. Then, each siRNA was mixed with Opti-MEM to an appropriate final concentration in a total volume of 15 µl/well. Next, the siRNA mixture was added to the transfection reagent, mixed, and incubated for 20 min at RT to allow transfection complex formation. Then, the transfection complexes were added and mixed into a cell culture 96-well imaging microplate (30µl/well). The A549-A2-T2 cells were counted and seeded at a selected density/well in 70 µl of DMEM medium containing 10% FBS and no antibiotics (siRNA growth media) and then incubated at 37°C. At 48 hpt, cells were checked (no toxic effects were

evident), washed with PBS, fixed, stained, immunofluorescently labelled, and imaged. The full transfection protocol is described in Section 2.8.

Among the different conditions tested, one of the best two conditions used 0.1 µl lipoRM, 20 nM siRNA, and 4000 cells/well to achieve 58.0% knockdown for the gene of interest in the first condition. In contrast, the second condition achieved a higher knockdown percentage (71.7%) by increasing the lipoRM to 0.15 µl and the siRNA concentration to 50 nM and reducing the cell number to 2500/well (Figure 5-1).

WB was used to confirm that siRNA transfections reduced LAMIN A/C protein expression. The transfection procedure was scaled up to a 24-well plate (3 wells/protein) with 10,000 A549-A2-T2 cells, 50 nM siRNA, and 0.6 µl lipoRM/well to ensure sufficient cells were harvested for a successful WB experiment. On the day of transfection, the lipoRM transfection reagent was mixed with Opti-MEM to a final volume of 60 µl/well and incubated for 5 min at RT. Next, each siRNA was diluted in 1× siRNA buffer to make a 1 µM stock. Then, each siRNA was mixed with Opti-MEM to a final concentration of 50 nM in a total volume of 60 µl/well. Next, the siRNA mixture was added to the transfection reagent, mixed, and incubated for 20 min at RT to allow transfection complex formation. Then, the transfection complexes were added to a well in a 24-well imaging microplate (120µl/well) before 10,000 A549-A2-T2 cells in siRNA growth media 280 µl/well were added, followed by incubation at 37°C. At 48 hpt, the cells were checked, washed with PBS, detached, harvested, and lysed with 30 µl of 2× SDS-PAGE sample buffer. For each sample, ⅓ of the protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting. The WB results showed complete LAMIN A/C knockdown (Figure 5-1). Therefore, the best knockdown achieved was 71.7%, which IFA and WB confirmed.

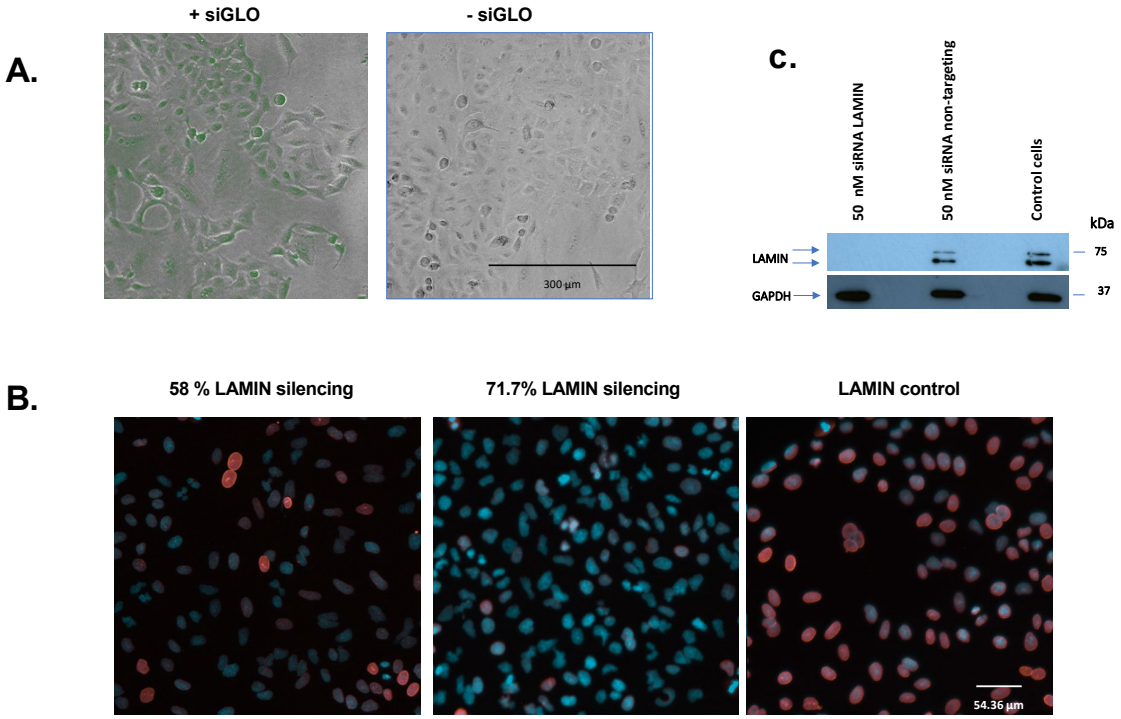


Figure 5-1 Optimisation of siRNA transfection and knockdown efficiency.

A. Microscopic analysis of A549-A2-T2 cells in a 96-well imaging microplate with and without siGLO after 48 h. Cells were fixed with 4% paraformaldehyde, permeabilised with 1% (v/v) Triton X-100, and microscopically examined for a siGLO green fluorescence signal. The figure shows the best transfection conditions: 100 nM of siGLO, 5000 cells/well, and 0.1 µl from DF1/well. The cells were imaged using an Image Xpress Pico automated imaging system.

B. IFA analysis of A549-A2-T2 cells in a 96-well imaging microplate transfected with four pooled siRNAs targeting LAMIN A/C and control at 48 hpt. Next, cells were fixed with 4% paraformaldehyde, permeabilised with 1% (v/v) Triton X-100, and probed with an anti-LAMIN A/C primary antibody. An Alexa Fluor 568 conjugated secondary antibody was used to detect the LAMIN A/C antibody (red). Nuclear DNA was stained with DAPI (blue). Two transfection conditions with different knockdown percentages are shown: 0.1 µl of lipoRM, 20 nM of siRNA, and 4000 cells/well (58.0% knockdown); and 0.15 µl of lipoRM, 50 nM of siRNA, and 2500 cells/well (71.7% knockdown). The cells were imaged using an Image Xpress Pico automated imaging system.

C. A549-A2-T2 cells were transfected with four pooled siRNAs targeting LAMIN A/C, NON-TAR, and control for 48 h. For each sample, ⅓ of the protein lysate was analysed by WB.

The expected positions of the LAMIN A/C and GAPDH proteins are shown on the left of the figure, and the positions of relevant molecular mass markers (kDa) are shown on the right. The GAPDH protein was used as the loading control. The experimentally determined molecular masses for LAMIN A/C and GAPDH proteins were consistent with their expected molecular masses (75 and 37 kDa, respectively).

5.2.2 Analysis of the effect of depleting selected host proteins by siRNA knockdown on SARS-CoV-2 replication in A549-A2-T2 cells

The identification of host proteins in three species that interacted with the MERS-CoV and SARS-CoV-2 E and M proteins with high confidence and the selection of a subset for further investigation was described in Chapters 3 and 4. The 43 selected host proteins comprised four groups: (1) proteins in HEK293, Dubca, and PaKiT cells that commonly interacted with the MERS-CoV and SARS-CoV-2 E proteins in transient transfection experiments (ACAT1, JUP, DNAJB12, TUBG1, CDIPT, ABCB7, PNN, UNC45A, FLOT2, SAP18, SAMM50, SLC25A6, NUP35, PELO, TECR, SGPL1, TRIP4, RBM17, CWC22, RBM39, and UBA52); (2) proteins in HEK293, Dubca, and PaKiT cells that commonly interacted with the MERS-CoV and SARS-CoV-2 M proteins in transient transfection experiments (ZMPSTE24, IMMT, SLC1A5, SLC38A10, CSE1L, SAMM50, IPO11, DYNC1I2, RAB1A, USP34, ELOVL5, and SAAL1; see Section 4.2.5). Two further groups comprised proteins that interacted with either the MERS-CoV E or M proteins stably expressed in HEK293 cells: (3) proteins interacting with the MERS-CoV E protein (TM9SF2, TMEM43, YIPF5, CERS2, SLC44A2, and ERGIC1); and (4) proteins interacting with the MERS-CoV M protein (LPCAT1, SCAMP3, IPO11, VDAC1, and RAB10; see Section 3.2.5).

The optimised siRNA knockdown conditions described in Section 5.2.1 were used to examine the relevance of the selected host proteins for SARS-CoV-2 replication. Initially, a low SARS-CoV-2 MOI was selected for the infection of A549-A2-T2 cells in which the respective proteins had been depleted. A low MOI enabled the examination of SARS-CoV-2 entry, intracellular replication, assembly, and release of infectious viruses. A549-A2-T2 cells (2500/well) were transfected with a pool of four siRNAs (50 nM/well) against each of the 43 selected host proteins in technical triplicate in 96 well plates. In addition, the following controls were included in the experiments: a pool of four siRNAs against LAMIN A/C, NON-TAR, kinesin family member 11 (KIF11; death control), and a no siRNA transfection control. LipoRM (0.15 µl/well) was used as the transfection reagent. The KIF11 (death control) was an easy control for monitoring the siRNA knockdown process since successful knockdown was reflected in cell death.

On the day of transfection, lipoRM was mixed with Opti-MEM to a final volume of 15 µl/well and incubated for 5 min at RT. Next, each siRNA was mixed with Opti-MEM to a final 50 nM concentration in a total volume of 15 µl/well. Then, the siRNA mixture was added to the

transfection reagent, mixed, and incubated for 20 min at RT. Next, the transfection complexes were added to appropriate wells in a 96-well imaging microplate (30 µl/well). Then, A549-A2-T2 cells (2500 cells/well in 70 µl of siRNA growth media) were added to each well, and the plates were incubated at 37°C. At 48 hpt, the cells were checked and washed with PBS before adding the infection medium (Section 2.8) and SARS-CoV-2 at a low MOI of 1 or high MOI of 5 and incubating at 37°C for 24 h. The lower MOI was used in these experiments to examine the effects of siRNA knockdown of the target protein on virus replication and spread. The higher MOI was used to measure cell affinity for infection after siRNA knockdown of the target protein. In addition to the siRNA controls described above, the experiments included control non-transfected cells (control) and virus-infected non-transfected cells (control-V). At 24 h post-infection (hpi), the cells were fixed, and IFA was used to measure virus replication. The effects of target protein siRNA knockdown on SARS-CoV-2 replication were examined in three independent siRNA knockdown experiments with three technical repeats. The plate layout is shown in Appendix G. The IFA results from the three experiments were captured and analysed using the Image Xpress Pico automated imaging system (Figure 5-2). The positive cells number and percentage were quantified and used to statistical analysis. To avoid the variability in a cell number, we used the percentage of positive cells for the statistical analysis. Statistical analysis was carried out using one-way ANOVA, and if a p-value is less than 0.05, it is flagged with (*), if a p-value is less than 0.01, it is flagged with (**), and if a p-value is less than 0.001, it is flagged with (***).

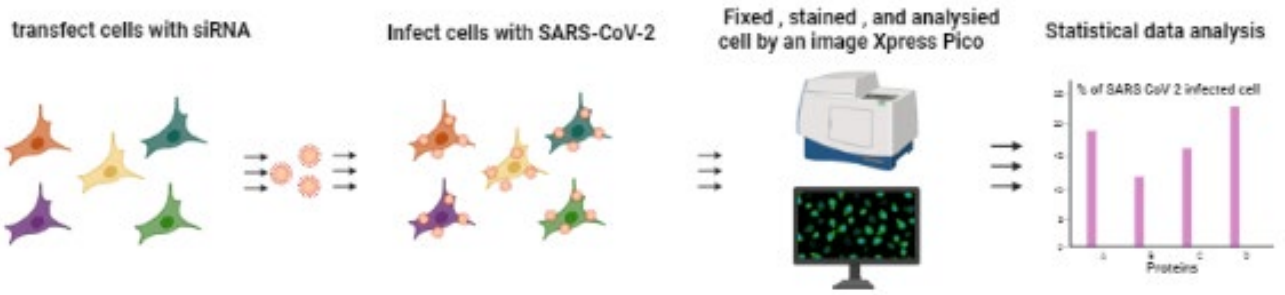


Figure 5-2 Schematic of depleting target proteins by siRNA knockdown and measuring the effect on SARS-CoV-2 replication.

A549-A2-T2 cells were transfected with siRNAs against the target proteins and appropriate controls. At 48 hpt, the cells were infected with SARS-CoV-2 at a low or high MOI. At 24 hpi, the cells were fixed, and IFA was performed and analysed using the Image Xpress Pico automated imaging system. The effects of protein depletion on SARS-CoV-2 infection were examined in three independent siRNA knockdown experiments with three technical repeats. This diagram was created with BioRender.com

5.2.2.1 The Low MOI siRNA experiment

The effect of depleting each target protein by siRNA knockdown on SARS-CoV-2 replication, starting with a low MOI, is shown in Figure 5-3. In each of the three independent experiments, the percentage of cells infected in the virus-only control wells was set at 100%. Then, the relative virus replication in all other wells was expressed as the percentage of the virus-only control. The levels of virus replication varied among the three experiments for most of the siRNA knockdowns. However, siRNA-mediated depletion of UBA52, LPCAT1, TM9SF2, CERS2, and ABCB7 significantly reduced SARS-CoV-2 infection by 58.80% (p = 0.0003), 35.54% (p = 0.0051), 17.32% (p = 0.0395), 11.78% (p = 0.0051), and 8.3% (p = 0.0332), respectively (Figure 5-3B). Interestingly, UBA52 and ABCB7 were SARS-CoV-2 and MERS-CoV E protein interactors shared among the three hosts. In contrast, TM9SF2, TMEM43, and CERS2 were MERS-CoV E interactors, and LPCAT1 was a MERS-CoV M interactor (Chapter 3).

Depletion of the other 38 proteins by siRNA knockdown resulted in a nonsignificant decrease or increase in SARS-CoV-2 replication. Only four proteins correlated with a nonsignificant decrease in viral replication: TUBG1 (19.58%), SAAL1 (11.23%), IPO11 (5.71%), and CWC22 (10.5%). The other 34 proteins correlated with a nonsignificant increase in viral replication. Full details of the results are summarised in Table 5-2.

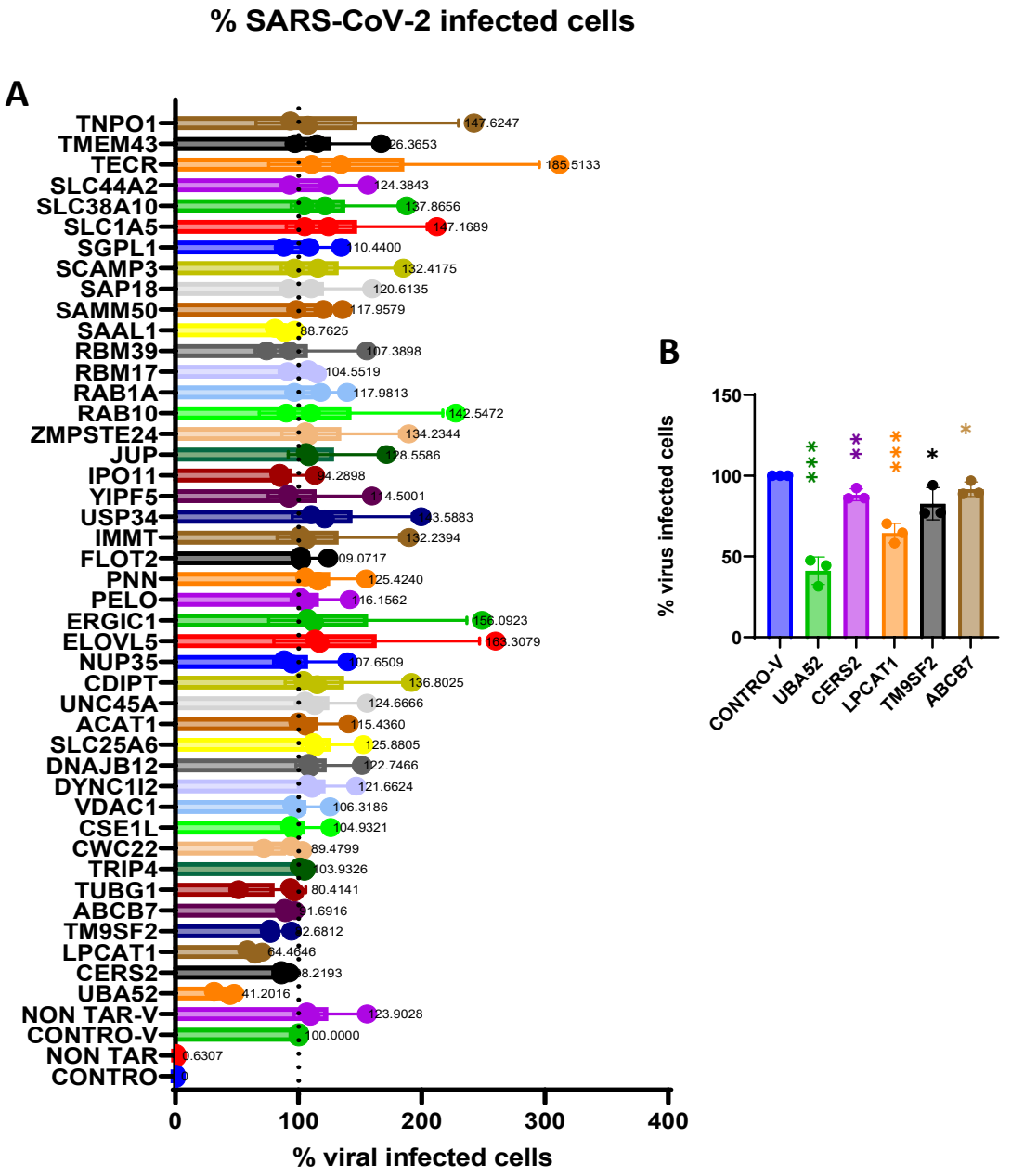


Figure 5-3 Graph of siRNA knockdown levels for the selected host proteins using a low MOI for SARS-CoV-2 infection.

(A) A549-A2-T2 cells were transfected with pooled siRNAs (50 nM/well) against each mRNA corresponding to the target proteins, NON-TAR, and control. At 48 hpt, cells were infected with SARS-CoV-2 at a low MOI, including NON-TAR-V and control wells (control-V). At 24 hpi, the cells were fixed and subjected to IFA using an anti-N antibody. The percentage of virus-infected cells was determined using an Image Xpress Pico automated imaging system. The effects of siRNA knockdown-mediated protein depletion were examined in three independent siRNA knockdown experiments with three technical repeats. The average across technical repeats in each independent experiment is shown as one circle in the graph. **(B)** Graph of

significant SARS-CoV-2 replication reductions in A549-A2-T2 cells in response to UBA52, CERS2, LPCAT1, TM9SF2, and ABCB7 depletion in the low MOI experiments. The positive cells number and percentage were quantified and used to statistical analysis. Statistical analysis was carried out using one-way ANOVA, and if a p-value is less than 0.05, it is flagged with (*), if a p-value is less than 0.01, it is flagged with (**), and if a p-value is less than 0.001, it is flagged with (***).

Table 5-2 Relative percentages of SARS-CoV-2-infected cells after infection at a low MOI.

Protein	Average of the relative % of SARS-CoV-2 infected cells				% of SARS-CoV-2 replication reduction	*p*-value
	Experiment 1	Experiment 2	Experiment 3	Average of the three experiments		
Control	0.00	0.00	0.00	0.00	100	-------
NON-TAR	1.20	0.50	0.20	0.60	99.4	0.00
Control-V	100	100	100	100	0.00	-------
NON-TAR-V	155	106	109	123	−23.9	0.20
UBA52	31.5	44.5	47.6	41.2	58.8	0.0002***
LPCAT1	58.3	64.7	70.4	64.5	35.5	0.0005***
TM9SF2	76.8	77.1	94.2	82.7	17.3	0.03*
CERS2	86.1	92.4	86.2	88.2	11.8	0.005**
ABCB7	89.4	96.9	88.8	91.7	8.30	0.03*
TUBG1	51.2	93.4	96.7	80.4	19.6	0.25
SAAL1	88.7	80.8	96.8	88.8	11.2	0.07
IPO11	113	84.5	85.1	94.3	5.70	0.57
TRIP4	101	106	104	103	−3.93	0.05
CWC22	103	71.8	93.7	89.5	10.5	0.31
CSE1L	126	95.6	93.2	104	−4.93	0.66
VDAC1	125	98.3	95.1	106	−6.31	0.54

DYNC1I2	146	107	110	121	−21.6	0.16
DNAJB12	151	107	108	122	−22.7	0.18
SLC25A6	152	111	113	125	−25.8	0.12
ACAT1	140	105	100	115	−15.4	0.28
UNC45A	155	113	105	124	−24.6	0.19
CDIPT	191	115	103	136	−36.8	0.25
NUP35	139	88.4	94.9	107	−7.65	0.66
ELOVL5	259	113	117	163	−63.3	0.25
ERGIC1	248	107	112	156	−56.0	0.29
PELO	141	105	101	116	−16.1	0.27
PNN	155	105	115	125	−25.4	0.16
FLOT2	124	101	101	109	−9.07	0.29
IMMT	189	101	106	132	−32.2	0.32
USP34	199	110	121	143	−43.5	0.19
YIPF5	159	91.6	92.3	114	−14.5	0.55
JUP	171	106	108	128	−28.5	0.25
ZMPSTE24	189	105	107	134	−34.2	0.28
RAB10	227	90.0	110	142	−42.5	0.37
RAB1A	139	96.7	117	118	−17.9	0.21
RBM17	114	91.1	107	104	−4.55	0.55
RBM39	155	74.1	92.6	107	−7.38	0.77

SAMM50	135	98.2	120	118	−17.9	0.17
SAP18	159	92.0	110	120	−20.6	0.36
SCAMP3	185	96.6	115	132	−32.4	0.29
SGPL1	134	88.0	108	110	−10.4	0.48
SLC1A5	212	105	124	147	−47.1	0.22
SLC38A10	187	104	121	137	−37.8	0.2
SLC44A2	156	92.6	124	124	−24.3	0.25
TECR	311	110	134	185	−85.5	0.24
TMEM43	167	97.1	114	126	−26.3	0.27
TNPO1	242	93.3	107	147	−47.6	0.37

siRNA knockdowns resulting in a significant reduction in viral infection are highlighted in green.

siRNA knockdowns resulting in a nonsignificant reduction in viral infection are highlighted in blue.

The positive cells number and percentage were quantified and used to statistical analysis. Statistical analysis was carried out using one-way ANOVA in GraphPad Prism 9.4.1, and if a p-value is less than 0.05, it is flagged with (*), if a p-value is less than 0.01, it is flagged with (**), and if a p-value is less than 0.001, it is flagged with (***).

5.2.2.2 *The High MOI siRNA experiment*

The effect of depleting each target protein by siRNA knockdown on SARS-CoV-2 replication, starting with a high MOI, is shown in Figure 5-4. Like the experiments using a low MOI, in each of the three independent experiments, the percentage of cells infected in the virus-only control wells was set at 100%. Then, the relative virus replication in all other wells was expressed as a percentage of the virus-only control. Unlike the infection experiments using a low MOI, there was low variation between experiments using a high MOI (Figure 5-4A). siRNA-mediated depletion of the proteins UBA52, LPCAT1, TM9SF2, YIPF5, CERS2, SGPL1, SAAL1, CWC22, SAP18, and RAB10 significantly reduced SARS-CoV-2 infection by 20.35% ($p = 0.0288$), 12.22% ($p = 0.0027$), 5.54% ($p = 0.0009$), 3.91% ($p = 0.0038$), 3.73% ($p = 0.0195$), 3.20% ($p = 0.0020$), 3.12% ($p = 0.0206$), 3.7% ($p = 0.0218$), 1.90% ($p = 0.0451$), and 1.17% ($p = 0.0266$), respectively (Figure 5-4B). UBA52, SGPL1, SAAL1, CWC22, and SAP18 commonly interacted with SARS-CoV-2 and MERS-CoV E proteins in the cells of all three species examined. TM9SF2, YIPF5, and CERS2 interacted with the MERS-CoV E protein in HEK293 cells, while RAB10 and LPCAT1 interacted with the MERS-CoV M protein in HEK293 cells (see Chapter 3).

Depletion of the other 33 proteins by siRNA knockdown resulted in a nonsignificant decrease or increase in viral replication. Of these 33 proteins, 25 caused a nonsignificant decrease in viral replication: SLC25A6 (4.20%), DNAJB12 (5.91%), DYNC1I2 (5.42%), ABCB7 (6.02%), ACAT1 (2.97%), TUBG1 (1.06%), CDIPT (8.38%), UNC45A (3.97%), NUP35 (10.38%), ELOVL5 (3.91%), ERGIC1 (6.27%), PELO (6.73%), FLOT2 (9.64%), IMMT (8.06%), USP34 (15.84%), IPO11 (12.03%), JUP (8.09%), ZMPSTE24 (7.90%), RAB1A (0.89%), SLC1A5 (5.54%), RBM17 (0.17%), RBM39, (5.06%), SAMM50 (4.12%), SCAMP3 (4.00%), and TNPO1 (1.18%). The remaining eight proteins caused a nonsignificant increase in viral replication. Full details of the results are summarised in Table 5-3.

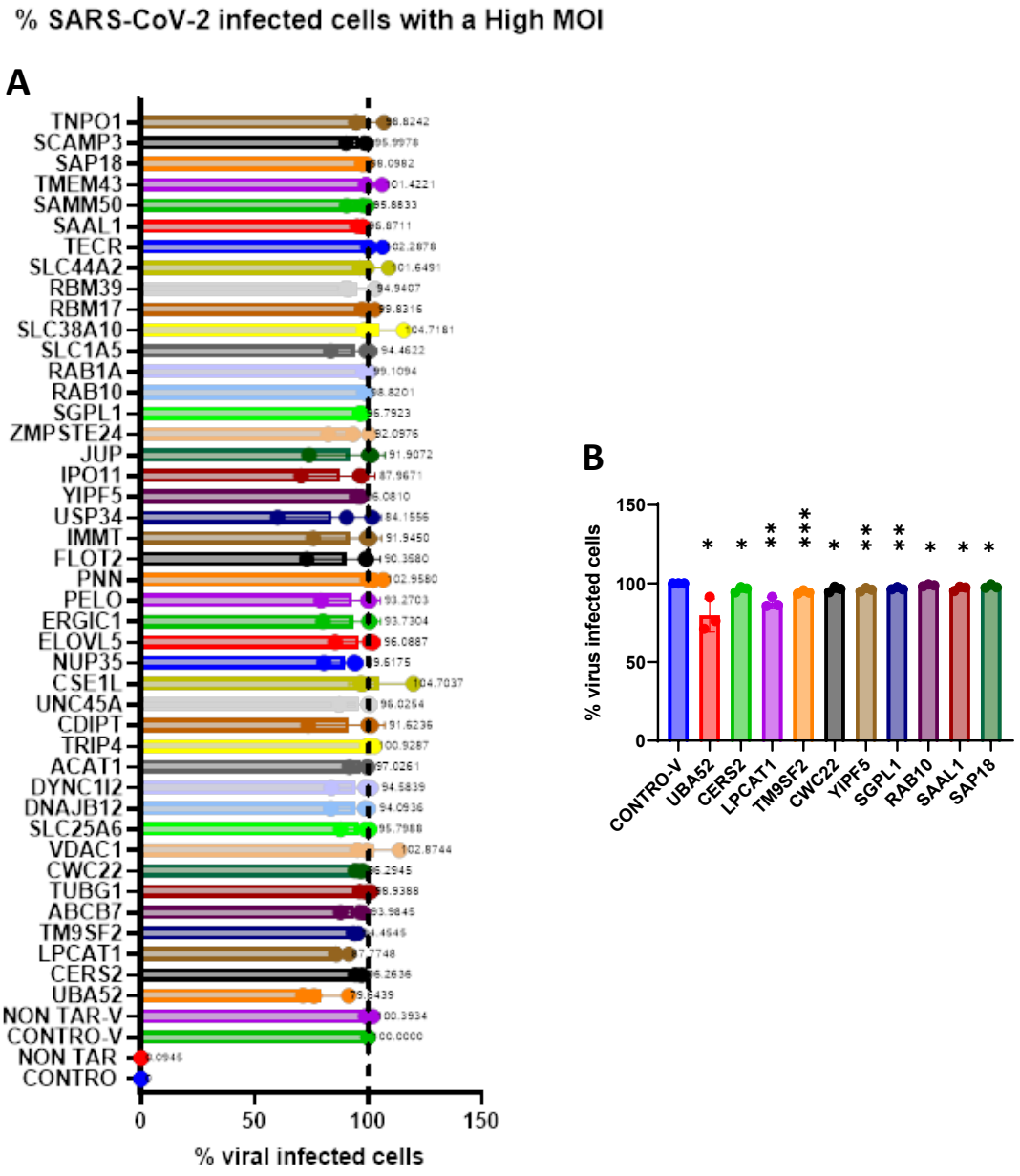


Figure 5-4 Graph of siRNA knockdown levels for the selected host proteins using a high MOI for SARS-CoV-2 infection.

(A) A549-A2-T2 cells were transfected with pooled siRNAs (50 nM/well) against each mRNA corresponding to the target proteins, NON-TAR, and control. At 48 hpt, cells were infected with SARS-CoV-2 at a high MOI, including NON-TAR-V and control wells (control-V). At 24 hpi, the cells were fixed and subjected to IFA using an anti-N antibody. The percentage of virus-infected cells was determined using an Image Xpress Pico automated imaging system. The effects of siRNA knockdown-mediated protein depletion were examined in three independent siRNA knockdown experiments with three technical repeats. The average across the technical repeats in each independent experiment is

shown as one circle in the graph. **(B)** Graph of significant SARS-CoV-2 replication reductions in A549-A2-T2 cells in response to UBA52, LPCAT1, TM9SF2, YIPF5, CERS2, SGPL1, SAAL1, CWC22, SAP18, and RAB10 depletion in the high MOI experiments. The positive cells number and percentage were quantified and used to statistical analysis. Statistical analysis was carried out using one-way ANOVA, and if a p-value is less than 0.05, it is flagged with (*), if a p-value is less than 0.01, it is flagged with (**), and if a p-value is less than 0.001, it is flagged with (***).

Table 5-3 Relative percentages of SARS-CoV-2-infected cells after infection at a high MOI.

Protein	Average % of SARS-CoV-2 infection cells				% of SARS-CoV-2 replication reduction	*p*-value
	Experiment 1	Experiment 2	Experiment 3	Average of the three experiments		
Control	0.00	0.00	0.00	0.00	100	-----
NON-TAR	0.02	0.00	0.30	0.11	99.9	0.00
Control-V	100	100	100	100	0.00	-----
NON-TAR-V	112	98.9	101	104	−4.13	0.36
CWC22	96.6	94.4	97.8	96.2	3.71	0.02*
VDAC1	113	95.2	99.3	102	−2.87	0.64
SLC25A6	87.7	98.8	100	95.8	4.20	0.36
DNAJB12	83.6	98.7	99.9	94.0	5.91	0.32
DYNC1I2	83.7	98.8	101	94.5	5.42	0.38
ABCB7	87.7	96.1	98.0	93.9	6.02	0.13
ACAT1	91.7	99.8	99.4	97.0	2.97	0.32
TRIP4	102	99.6	100	100	−0.92	0.40
TUBG1	101	96.3	99.1	98.9	1.06	0.51
CDIPT	73.5	100	101	91.6	8.38	0.41
CERS2	94.3	97.6	96.7	96.2	3.74	0.02*

UBA52	91.4	71.2	76.2	79.6	20.3	0.03*
UNC45A	87.3	100	100	96.0	3.97	0.41
CSE1L	119	97.6	96.5	104	−4.70	0.57
NUP35	80.5	93.7	94.5	89.6	10.3	0.08
ELOVL5	85.5	100	102	96.0	3.91	0.50
ERGIC1	80.1	100	100	93.7	6.27	0.41
PELO	79.2	100	99.9	93.2	6.73	0.39
PNN	106	99.7	102	102	−2.95	0.22
FLOT2	72.9	99.2	98.8	90.3	9.64	0.33
IMMT	75.7	99.4	100	91.9	8.06	0.38
USP34	60.1	90.5	101	84.1	15.8	0.27
YIPF5	94.9	97.2	96.0	96.0	3.92	0.003**
IPO11	70.4	96.2	97.2	87.9	12.0	0.24
JUP	74.0	100	101	91.9	8.09	0.42
ZMPSTE24	82.4	93.3	100	92.1	7.90	0.21
LPCAT1	91.4	85.7	86.1	87.7	12.2	0.003**
SGPL1	96.3	96.3	97.6	96.7	3.21	0.002**
RAB10	99.4	98.7	98.2	98.8	1.18	0.03*
RAB1A	100	99.1	97.7	99.1	0.89	0.31
SLC1A5	83.5	100	99.0	94.4	5.54	0.37
SLC38A10	115	100	98.3	104	−4.71	0.44

RBM17	102	99.4	97.1	99.8	0.17	0.93
RBM39	103	91.5	90.1	94.9	5.06	0.29
SLC44A2	109	99.5	96.2	101	−1.64	0.69
TECR	106	100	99.7	102	−2.28	0.33
SAAL1	97.9	97.4	95.2	96.8	3.13	0.02*
SAMM50	90.6	97.4	99.5	95.8	4.12	0.20
TM9SF2	94.1	95.6	93.5	94.4	5.55	0.0009***
TMEM43	106	99.1	98.9	101	−1.42	0.58
SAP18	97.5	99.4	97.3	98.1	1.90	0.04*
SCAMP3	90.2	99.4	98.3	96.0	4.00	0.24
TNPO1	107	94.6	94.7	98.8	1.18	0.79

siRNA knockdowns resulting in a significant reduction in viral infection are highlighted in green.

siRNA knockdowns resulting in a nonsignificant reduction in viral infection are highlighted in blue.

The positive cells number and percentage were quantified and used to statistical analysis. Statistical analysis was carried out using one-way ANOVA in GraphPad Prism 9.4.1, and if a p-value is less than 0.05, it is flagged with (*), if a p-value is less than 0.01, it is flagged with (**), and if a p-value is less than 0.001, it is flagged with (***).

5.2.2.3 Comparison between low and high MOI siRNA experiments

The three replicate siRNA experiments using a low MOI for virus infection showed higher variability than those using a high MOI. Analysis of the results of the former experiments indicated that the variability was largely due to one replicate experiment (see Table 5-2, Figure 5-3, and Figure 5-4).

Depleting VDAC1, CSE1L, PNN, SLC38A10, SLC44A2, TECR, and TMEM43 by siRNA knockdown resulted in a nonsignificant increase in SARS-CoV-2 infection in both the low and high MOI experiments.

Depleting TUBG1 (19.58%), SAAL1 (11.23%), IPO11(5.71%), and CWC22 (10.5%) by siRNA knockdown resulted in a nonsignificant decrease in SARS-CoV-2 infection in the low MOI experiment. Similarly, depleting SLC25A6 (4.20%), DNAJB12 (5.91%), DYNC1I2 (5.42%), ABCB7 (6.02%), ACAT1 (2.97%), TUBG1 (1.06%), CDIPT (8.38%), UNC45A (3.97%), NUP35 (10.38%), ELOVL5 (3.91%), ERGIC1 (6.27%), PELO (6.73%), FLOT2 (9.64%), IMMT (8.06%), USP34 (15.84%), IPO11 (12.03%), JUP (8.09%), ZMPSTE24 (7.90%), RAB1A (0.89%), SLC1A5 (5.54%), RBM17 (0.17%), RBM39 (5.06%), SAMM50 (4.12%), SCAMP3 (4.00%), and TNPO1 (1.18%) by siRNA knockdown resulted in a nonsignificant decrease in SARS-CoV-2 infection in the high MOI experiment.

Depleting UBA52, CERS2, LPCAT1 and TM9SF2 by siRNA knockdown significantly reduced SARS-CoV-2 infection in both low and high MOI experiments (Figure 5-5). Therefore, the depletion of these four proteins by siRNA knockdown was selected for validation by WB and IFA.

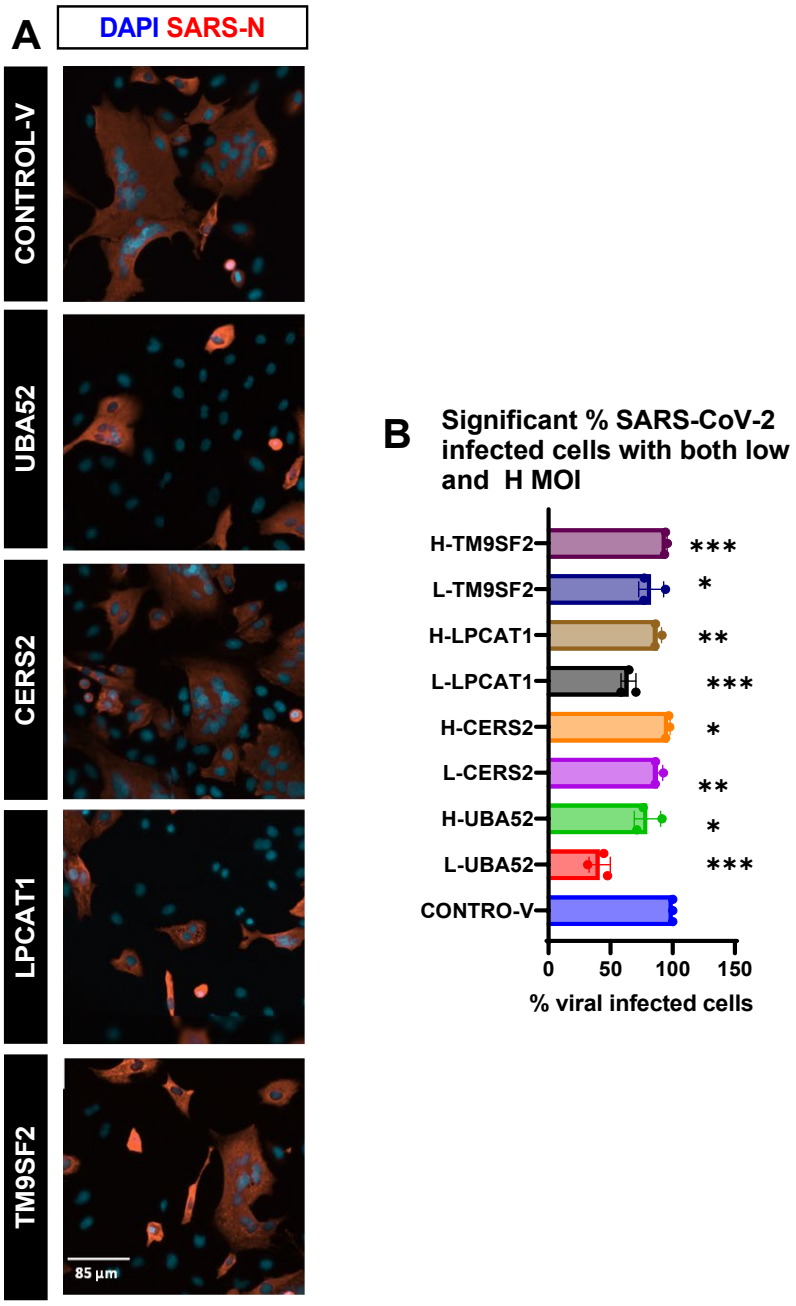


Figure 5-5 Cellular protein depletion by siRNA knockdown significantly reduced SARS-CoV-2 replication after initial infection at low and high MOIs.

(A) IFA analysis of SARS-CoV-2 infected A549-A2-T2 cells (control-V) and A549-A2-T2 cells depleted of UBA52, CERS2, LPCAT1, and TM9SF2 by siRNA knockdown. The cells were incubated with siRNA for 48 h before being infected with SARS-CoV-2 at a low MOI for 24 h. Then, the cells were fixed and probed with an antibody against the SARS-CoV-2 N protein. An Image Xpress Pico automated imaging system was used to capture the images and visualise DAPI (blue) and SARS-CoV-2 N (red) staining. The effects of protein depletion on virus replication were examined in three independent siRNA knockdown

experiments with three technical repeats. **(B)** Graph of significant SARS-CoV-2 replication reductions in A549-A2-T2 cells in response to UBA52, CERS2, LPCAT1, and TM9SF2 depletion in both low and high MOI experiments.

5.2.3 Validation of protein depletion by siRNA knockdown for proteins relevant to SARS-CoV-2 cellular replication

The siRNA knockdown experiments identified some cellular proteins that could be important for coronavirus replication. This section will present the validation of the selected proteins' depletion by IFA and WB. The knockdown of the selected proteins (UBA52, LPCAT1, TM9SF2, and CERS2) significantly ($p < 0.05$) reduced SARS-CoV-2 replication by almost 10% in both low and high MOI experiments. Moreover, UBA52 was the only protein shared among MERS-CoV and SARS-CoV-2 E proteomic data sets for the three hosts (see Section 4.2.5). TM9SF2 and CERS2 were cellular protein interactors for the MERS-CoV E protein, and LPCAT1 was a cellular protein interactor for the MERS-CoV M protein. The last three proteins were among the 11 selected MERS-CoV E and M proteins interactors (see Section 3.2.5).

Both WB and IFA were used to confirm that the relevant proteins were depleted prior to infection with SARS-CoV-2. The siRNA transfection protocol was performed as described in Section 2.8 using siRNA pools against the relevant proteins and the NON-TAR, KIF11, and transfection-only controls. For IFA, the cells were checked (no toxic effects were observed) at 48 hpt, washed with PBS, fixed, stained, labelled using protein-specific antibodies, and imaged. The automated imaging system was adjusted to identify cells expressing the protein of interest as a positive cell and cells not expressing the protein of interest as a negative cell. The siRNA procedure was scaled up to 24-well plates (3 wells/protein) to isolate sufficient proteins for WB. At 48 hpt, the cells were checked, washed with PBS, harvested, and lysed with 2× SDS-PAGE sample buffer. For each sample, ⅓ of the protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting.

CERS2

As expected, the WB analysis showed that CERS2 could be detected in lysates from cells treated with 50 nM NON-TAR siRNA and control cells but not from cells treated with 50 nM CERS2 siRNA (Figure 5-6A). The WB analysis showed that LAMIN A/C (loading control) could be detected in lysates from cells treated with 50 nM NON-TAR siRNA, control cells, and cells treated with 50 nM CERS2 siRNA. The IFA results confirmed the WB results. As expected, CERS2 (red) was visible in 50 nM NON-TAR siRNA-treated and control cells, while much fewer cells expressed CERS2 among those treated with 50 nM

CERS2 siRNA. Analysis of the IFA images confirmed that CERS2 was depleted in 85% of cells treated with 50 nM CERS2 siRNA compared to control cells. Therefore, only 15% of cells treated with 50 nM CERS2 siRNA and 129% treated with 50 nM NON-TAR siRNA were positive compared to 100% of control cells. See Figure 5-6.

LPCAT1

The WB analysis showed that LPCAT1 could be detected in lysates from cells treated with 50 nM NON-TAR siRNA, control cells, and at a lower level in cells treated with 50 nM LPCAT1 siRNA. LPCAT1 amounts were higher in lysates from control cells than from cells treated with 50 nM NON-TAR siRNA treated cells and, as expected, were lowest in lysates from cells treated with 50 nM LPCAT1 siRNA (Figure 5-7A). The WB analysis showed that GAPDH (loading control) could be detected in lysates from cells treated with 50 nM NON-TAR siRNA, control cells, and cells treated with 50 nM LPCAT1 siRNA. The IFA results confirmed the WB results. As expected, LPCAT1 (red) was visible in 50 nM NON-TAR siRNA-treated and control cells, while much fewer cells expressed LPCAT1 among those treated with 50 nM LPCAT1 siRNA. The IFA confirmed LPCAT1 depletion by 68% in cells treated with 50 nM LPCAT1 siRNA compared to control cells. Therefore, only 32% of cells treated with 50 nM LPCAT1 siRNA and 126% treated with 50 nM NON-TAR siRNA were positive compared to 100% of control cells. See Figure 5-7.

UBA52

The WB analysis showed that UBA52 could be detected in lysates from cells treated with 50 nM NON-TAR siRNA and control cells but not from cells treated with 50 nM UBA52 siRNA. The WB analysis showed that LAMIN A/C (loading control) could be detected in lysates from cells treated with 50 nM NON-TAR siRNA, control cells, and cells treated with 50 nM UBA52 siRNA. The WB results indicated a complete UBA52 knockdown (Figure 5-8A). The IFA results did not confirm the WB results. UBA52 (red) was visible in 50 nM NON-TAR siRNA-treated, control, and 50 nM UBA52 siRNA-treated cells. UBA52 protein depletion was 2% in cells treated with 50 nM UBA52 siRNA compared to control cells. Therefore, 98% of cells treated with 50 nM UBA52 siRNA and 100% treated with 50 nM NON TAR siRNA were positive compared to 100% of control cells. Figure 5-8 shows the unexpected IFA finding, which is discussed in Section 5.3.2.3.

Regarding the IFA experiment, the number of cells treated with 50 nM UBA52 siRNA was lower than those treated with 50 nM NON-TAR siRNA and control cells. The number of cells treated with 50 nM UBA52 siRNA was 32.5% (7632 cells) and treated with 50 nM NON-TAR siRNA was 96.6% (23,448 cells) of the control cells (100%; 22,668 cells). These results indicate that UBA52 could be important for cellular replication and that knocking it down prevented cells from replicating normally. Alternatively, the UBA52 siRNA could be toxic and cause many cells to be near death, making it unsuitable for viral infection.

TM9SF2

The WB analysis showed that TM9SF2 could not be detected in lysates from cells treated with 50 nM NON-TAR siRNA, control cells, and cells treated with 50 nM TM9SF2 siRNA (results not shown). The WB was repeated with a new TM9SF2 protein antibody with the same result (results not shown). It is possible that the amount of TM9SF2 in A549-A2-T2 cells was below the detection limit or that the TM9SF2 protein could be expressed in kidney (HEK293) but not in alveolar (A549-A2-T2) cells. The WB analysis was repeated using HEK293 and A549-A2-T2 cell lysates to test the latter hypothesis. For each cell line, ⅓ and ½ of the protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting. This WB analysis showed that TM9SF2 could be detected in relatively similar amounts at both HEK293 cell lysate quantities, with lower amounts detected at both A549-A2-T2 cell lysate quantities (results not shown). This finding indicates that TM9SF2 is present in low quantities in A549-A2-T2 cells than in HKE293 cells. This finding is consistent with the Human Protein Atlas (274), that showed TM9SF2 was detected in kidney cells but not in alveolar cells.

The IFA results confirmed the WB results. TM9SF2 could not be detected in 50 nM NON-TAR siRNA-treated, control, and 50 nM TM9SF2 siRNA-treated cells. The percentage of TM9SF2-expressing cells was 0% for all three groups (**Error! Reference source not found.**). This unexpected IFA finding could be due to an unspecific antibody or TM9SF2 quantities below the detection limit.

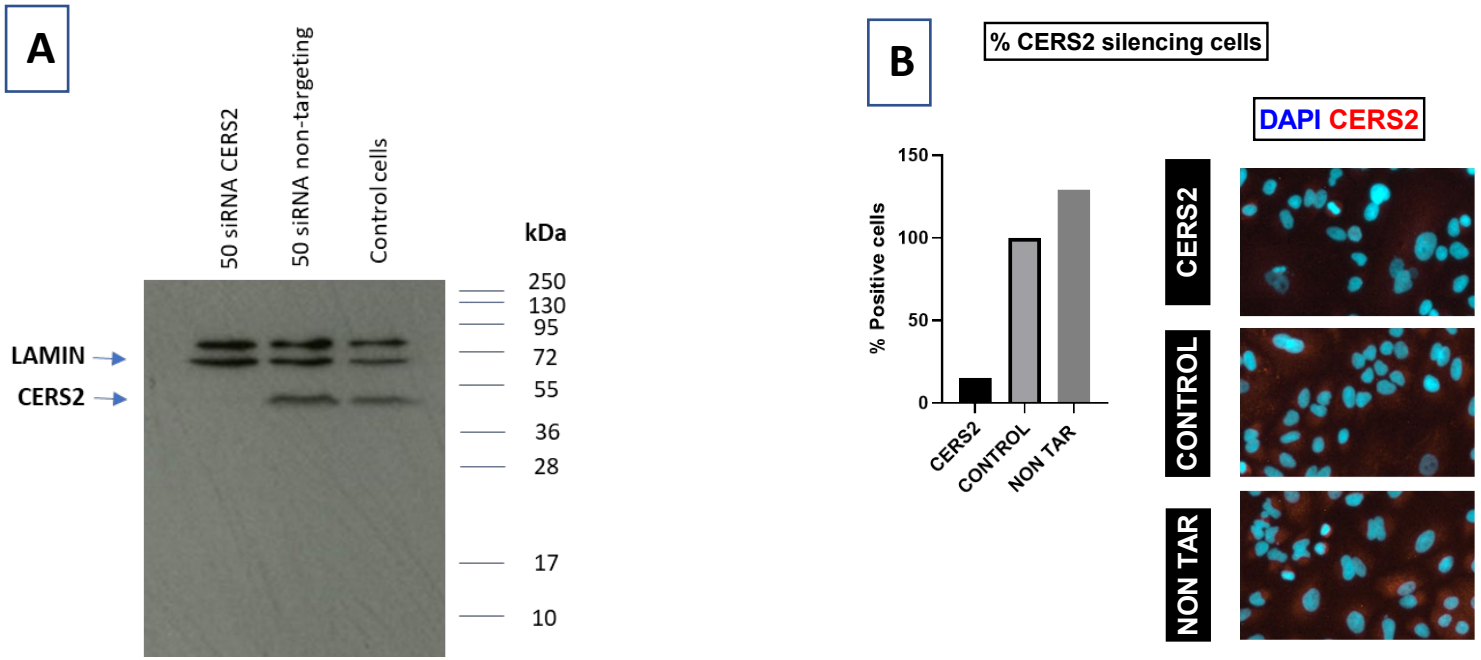


Figure 5-6 Validation of CERS2 depletion by siRNA knockdown.

(A) A549-A2-T2 cells were transfected with four pooled siRNAs targeting CERS2, NON-TAR, and control for 48 h. The cells were harvested and lysed with 2× SDS-PAGE sample buffer. For each sample, ⅓ of the protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting. The positions of CERS2 and LAMIN A/C are shown on the left of the diagram, while the positions of relevant molecular mass markers (kDa) are shown on the right. LAMIN A/C was used as the loading control. The experimentally-determined molecular masses for CERS2 and LAMIN A/C were consistent with expected molecular masses (45 and 75 kDa, respectively). **(B)** IFA analysis of A549-A2-T2 cells in a 96-well imaging microplate with four pooled siRNAs targeting CERS2, NON-TAR, and control after 48 h. Cells were fixed

with 4% paraformaldehyde, permeabilised with 1% (v/v) Triton X-100, and probed with an anti-CERS2 primary antibody. An Alexa Fluor 568 conjugated secondary antibody was used to detect the anti-CERS2 antibody (red). Nuclear DNA was stained with DAPI (blue). The percentage of CERS2-expressing cells was 15%. The cells were imaged using an Image Xpress Pico automated imaging system.

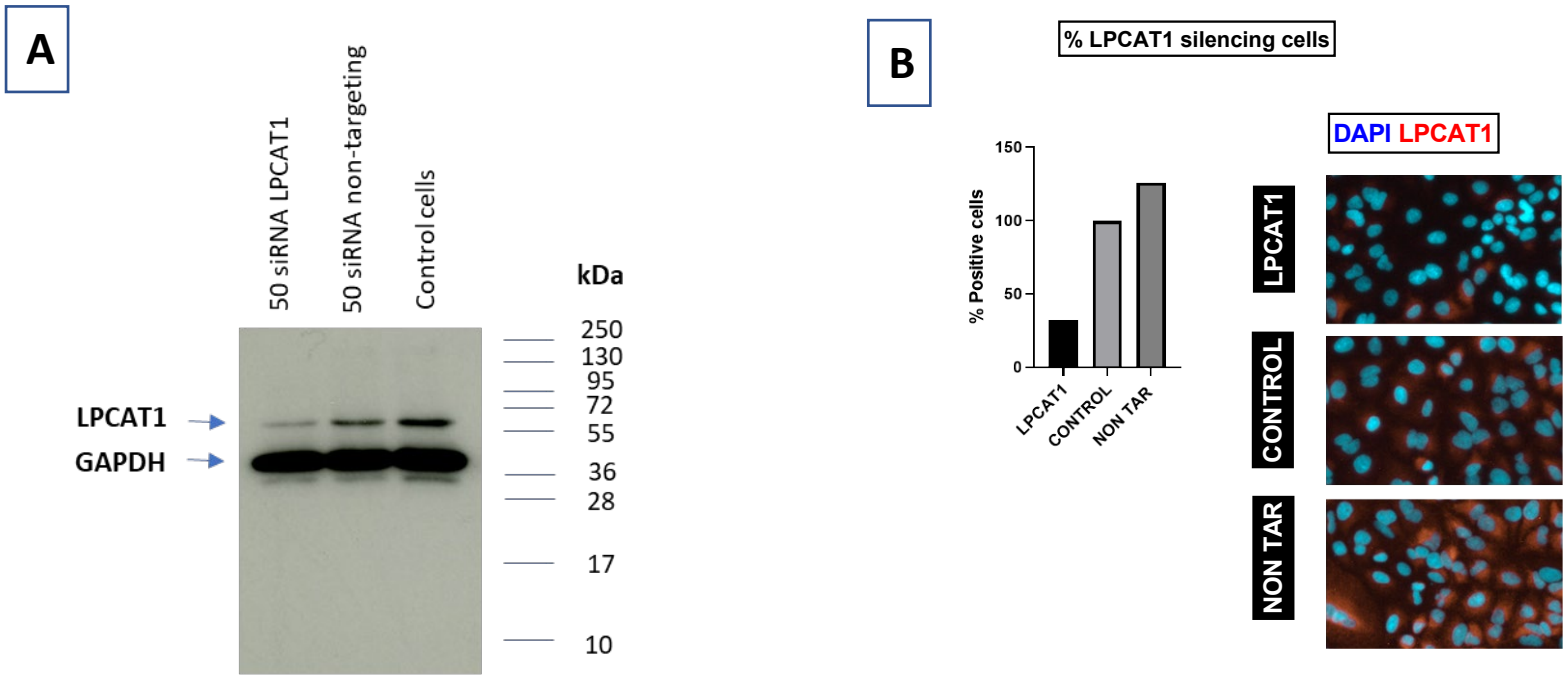


Figure 5-7 Validation of LPCAT1 protein depletion by siRNA knockdown.

(A) A549-A2-T2 cells were transfected with four pooled siRNAs targeting LPCAT1, NON-TAR, and control for 48 h. The cells were harvested and lysed with 2× SDS-PAGE sample buffer. For each sample, ⅓ of the protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting. The positions of LPCAT1 and GAPDH are shown on the left of the diagram, while the positions of relevant molecular mass markers (kDa) are shown on the right. GAPDH was used as the loading control. The experimentally-determined molecular masses for LPCAT1 and GAPDH were consistent with their expected molecular masses (59 and 36 kDa, respectively). **(B)** IFA analysis of A549-A2-T2 cells in a 96-well imaging microplate with four pooled siRNAs targeting LPCAT1, NON-TAR, and control after 48 h. Cells were

fixed with 4% paraformaldehyde, permeabilised with 1% (v/v) Triton X-100, and probed with an anti-LPCAT1 primary antibody. An Alexa Fluor 568 conjugated secondary antibody was used to detect the anti-LPCAT1 antibody (red). Nuclear DNA was stained with DAPI (blue). The percentage of LPCAT1-expressing cells was 32%. The cells were imaged using an Image Xpress Pico automated imaging system.

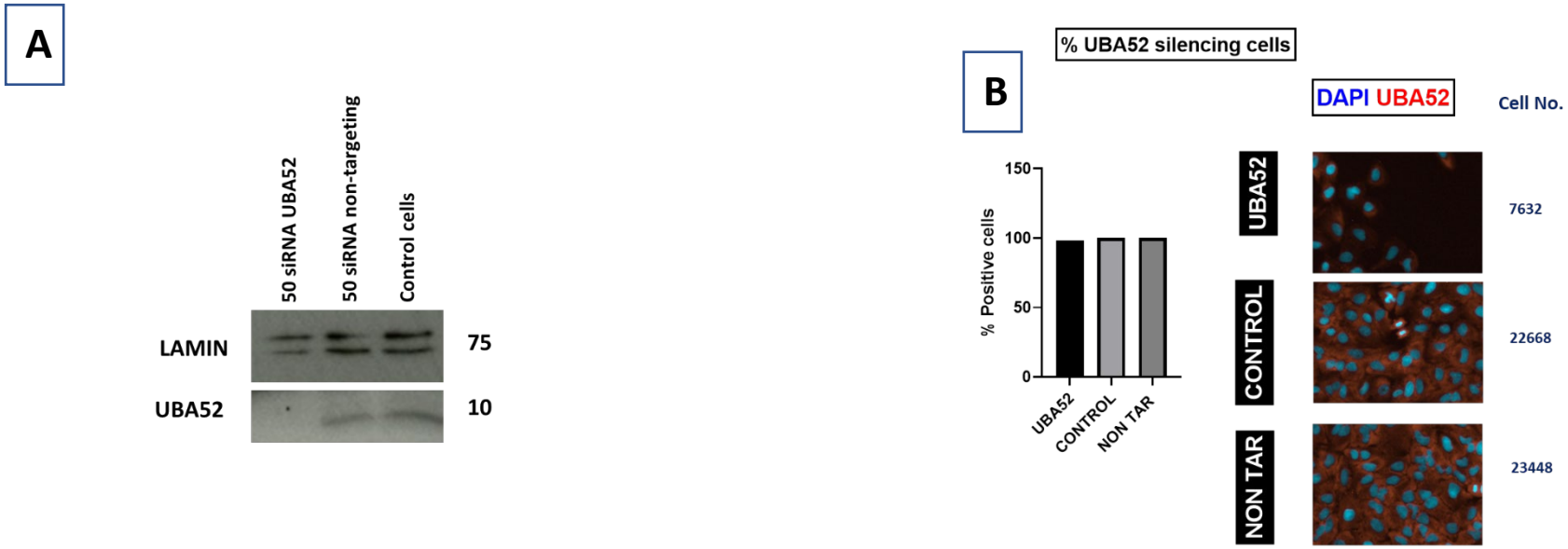


Figure 5-8 Validation of UBA52 protein depletion by siRNA knockdown.

(A) A549-A2-T2 cells were transfected with four pooled siRNAs targeting UBA52, NON-TAR, and control for 48 h. The cells were harvested and lysed with 2× SDS-PAGE sample buffer. For each sample, ⅓ of the protein lysate was separated on a 15% SDS-PAGE gel and transferred to a PVDF membrane for immunoblotting. The positions of UBA52 and LAMIN A/C are shown on the left of the diagram, while the positions of relevant molecular mass markers (kDa) are shown on the right. LAMIN A/C was used as the loading control. The experimentally-determined molecular masses for UBA52 and LAMIN A/C were consistent with their expected molecular masses (10 and 75 kDa, respectively). **(B)** IFA analysis of A549-A2-T2 cells in a 96-well imaging microplate with four pooled siRNAs targeting UBA52, NON-TAR, and control after 48 h. Cells were fixed with 4% paraformaldehyde, permeabilised with 1% (v/v) Triton X-100, and probed with an anti-UBA52 primary antibody. An Alexa Fluor 568 conjugated secondary antibody was used to detect the anti-UBA52 antibody (red). Nuclear DNA was stained with DAPI (blue). The stained cell

number is shown to the right of the cell images. The percentage of UBA52-expressing cells was 98%. The cells were imaged using an Image Xpress Pico automated imaging system.

5.3 Discussion

5.3.1 Comparison with other studies:

siRNA knockdown screens are potent tools for dissecting intricate biological processes such as viral infection. This study conducted three independent siRNA knockdown experiments with three technical repeats. A549-A2-T2 cells were transfected with pooled siRNAs against each mRNA corresponding to the target proteins, followed by SARS-CoV-2 infection using either a low or high MOI. siRNA knockdown and A549-A2-T2 cells were used in validation experiments in two previous studies (153, 269). The findings of these two and other studies were discussed in the introduction of this chapter. Table 5-1 summarises the published studies, including those that used CRISPR in Huh-7.5 or Caco-2 cells. Moreover, most published interactomic studies identified cellular protein interactors for many coronaviral proteins (153, 269, 270), providing a broad data set to compare pathways and interactors shared by different coronavirus proteins. This study focused only on confirmed protein interactors for the SARS-CoV-2, MERS-CoV E and M proteins.

Before the main protein depletion experiment was performed, conditions were optimised to achieve the best knockdown percentage with the least cell toxicity. Both WB and IFA were used during the optimisation step to confirm that the percentage of target protein depletion was rational and much better than using only one method. This step was followed by investigations on four groups of interesting host interactors that were described in Section 5.2.2. The host interactor selection approach limited the number of selected proteins, meaning some other important host interactors were not selected for confirmation. This limitation could impact the findings in more than one way. For example, we did not introduce the shared SARS-CoV-2 E cellular interactor into the three used host cells, which was not necessarily considered when selecting the shared SARS-CoV-2 and MERS-CoV E cellular interactors. Moreover, this approach to filtering host interactors ignored host interactors shared between the E and M viral proteins.

In addition, the potential function of the proteins selected for depletion on SARS-CoV-2 replication was investigated using a low and high MOI. Depleting proteins essential for

virus entry and genome replication/single-guide RNA (sgRNA) production was expected to affect virus replication when cells were infected at both low and high MOI. In contrast, depleting proteins required for virus assembly and release might not affect virus replication (as determined by IFA) when cells were infected at high MOI but would when cells were infected at low MOI due to decreased virus release and spread. One lower MOI experimental replicate was shown to shift results compared to the other two replicates. This issue was one reason for ignoring interactors that increased viral replication, especially since the high MOI experiments did not show a remarkable or significant increase in viral replication.

However, this study did not study the effect of selected host interactor depletion on MERS-CoV replication. In addition, the depletion was not confirmed of all targets, and without this step the data may contain false negative results. Moreover, depletion was not confirmed for proteins causing nonsignificant SARS-CoV-2 replication reductions. The low MOI experiments identified four proteins showing nonsignificant SARS-CoV-2 replication reductions, while the high MOI experiment identified 25 proteins showing nonsignificant SARS-CoV-2 replication reductions. Notably, they shared two proteins: TUBG1, SAAL1, and IPO11.

TUBG1 reduced SARS-CoV-2 replication by 19.58% and 1.06% in low and high MOI experiments, respectively. The other two proteins that nonsignificantly reduced SARS-CoV-2 replication were SAAL1 and IPO11. SAAL1 reduced SARS-CoV-2 replication by 11.23% and 3.1% in low and high MOI experiments, respectively. SAAL1 was validated as a SARS-CoV-2, SARS-CoV, and MERS-CoV M protein interactor (153). IPO11 reduced SARS-CoV-2 replication by 5.71% and 12.03% in low and high MOI experiments, respectively. IPO11 was confirmed as a cellular interactor for the MERS-CoV E and M proteins, and its colocalisation with the MERS-CoV E and M proteins was confirmed (Figure 3-16). IPO11 showed a 3.0-fold higher abundance in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells (see Section 3.2.6).

5.3.2 Selected protein knockdown validation

This section discusses the results related to four proteins that were of most relevance to the coronavirus lifecycle based on the siRNA knockdown experiments, the interactomic analysis results described in Chapter 4, the validation experiment results described in

Chapter 3, and the broader literature regarding protein functions and interactions with other viruses.

5.3.2.1 CERS2

CERS2 was selected for validation after confirmation of its cellular interactions with the MERS-CoV E and M proteins and its colocalisation with the MERS-CoV E and M proteins (Figure 3-12). CERS2 shows a 3.1-fold higher abundance in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells (see Section 3.2.6). Moreover, it was identified as a cellular interactor for MERS-CoV and SARS-CoV-2 E and M proteins transiently expressed in HEK293 and Dubca cells in six data sets: HEK SARS2 E, HEK MERS E, HEK SARS2 M, HEK MERS M, Dubca SARS2 M, and Dubca MERS M. Table 4-34 and Table 4-35 show the $\log_2$ fold changes in CERS2 abundance in co-IPs from lysates of three host cells expressing SARS-CoV-2 and MERS-CoV E and M proteins compared to negative control cells.

CERS2 plays a role in the lipid biosynthesis cellular process, which was reflected in Section 4.2.6.1 by the finding that lipid biosynthesis was an enrichment cluster for proteins significantly increased in two data sets: HEK SARS2 E and HEK MERS E. In these two data sets, the enrichment cluster of lipid biosynthesis proteins was shown as an ER membrane protein, the known localisation of lipid biosynthesis processes and CERS2 (188). CERS2 is one of six enzymes (CERS 1–6) synthesising ceramides (189). Acid ceramidase was identified as a crucial enzyme in the replication of SARS-CoV-2 (275). The primary function of this enzyme is to break down ceramide (275). Fluoxetine prevents viral replication by inhibiting acid ceramidase activity, which increases endo-lysosomal ceramide levels (275). Ceramide is a key lipid signalling molecule (188). CERS2 deficiency was shown to alter the lipid composition of cells and inhibit HIV-1 envelope receptor binding and/or fusion processes (190).

CERS2 has not been previously shown to interact with any coronavirus proteins. Interestingly, CERS2 depletion significantly reduced SARS-CoV-2 cellular replication in low (11.78%, p = 0.0051) and high (3.73%, p = 0.0195) MOI experiments. WB and IFA analyses confirmed a >80% depletion of CERS2. These results suggest that CERS2 could be a therapeutic target to reduce coronavirus replication.

5.3.2.2 LPCAT1

LPCAT1 was selected for validation after confirmation of its cellular interactions with MERS-CoV E and M proteins and its colocalisation with MERS-CoV E and M proteins (Figure 3-19). LPCAT1 showed a 3.5-fold higher abundance in co-IPs from lysates of cells expressing the MERS-CoV M protein compared to Flp cells (see Section 3.2.6). Moreover, it was identified as a cellular interactor for MERS-CoV and SARS-CoV-2 E and M proteins transiently expressed in HEK293, PaKiT, and Dubca cells in nine data sets: HEK SARS2 E, HEK MERS E, PaKiT SARS2 E, PaKiT MERS E, HEK MERS M, PaKiT SARS2 M, PaKiT MERS M, Dubca SARS2 M, and Dubca MERS M. Table 4-36 and Table 4-37 show the LPCAT1 $\log_2$ fold changes in abundance in co-IPs from lysates of three host cells expressing SARS-CoV-2 and MERS-CoV E and M proteins compared to negative control cells.

LPCAT1 plays a role in cellular lipid biosynthesis, which was reflected in Section 4.2.6.1 by the finding that lipid biosynthesis was an enrichment cluster for proteins significantly increased in two data sets: HEK SARS2 E and HEK MERS E. In these two data sets, the enrichment cluster for lipid biosynthesis proteins was shown as an ER membrane protein, the known localisation of lipid biosynthesis processes and LPCAT1. LPCAT1 localised to ER membrane, lipid droplets, and cell membrane (194) and was suggested to play a role in synthesising phosphatidylcholine in pulmonary surfactants, making it crucial in respiratory system physiology. In addition, it contributes to regulating the number and size of lipid droplets (195). Moreover, LPCAT1 was highly expressed in alveolar type II epithelial cells in the lung and showed an anti-inflammatory effect (196, 197). It was interesting to confirm the effectiveness of LPCAT1 depletion on SARS-CoV-2 replication. In addition, LPCAT1 was detected as a cellular protein interactor for the A71 picornavirus protein (198). Moreover, hepatitis virus infection increased *LPCAT1* expression (199). These two studies support our findings and show that studying LPCAT could be promising.

The LPCAT1 protein has not been previously shown to interact with any coronavirus proteins. Interestingly, LPCAT1 depletion significantly reduced SARS-CoV-2 cellular replication in low (35.54%, $p = 0.0051$) and high (12.22%, $p = 0.0027$) MOI experiments. WB and IFA analyses confirmed a >60% LPCAT1 knockdown. These results suggest that LPCAT1 could be a therapeutic target for reducing coronavirus replication.

5.3.2.3 UBA52

UBA52 was selected for validation after confirmation of its cellular interactions with the SARS-CoV-2 and MERS-CoV E proteins. The colocalisation of UBA52 with SARS-CoV-2 and MERS-CoV E proteins was not examined, which is one study limitation. Section 4.2.5 described the $\log_2$ fold increase in UBA52 in anti-FLAG co-IPs from lysates of HEK293, PaKiT, and Dubca cells transfected with two plasmids encoding two viral proteins (SARS-CoV-2 and MERS-CoV E) compared to negative control cells. UBA52 is a ubiquitin supplier and ribosomal protein complex regulator (243). The UBA52-encoded fusion protein comprises ubiquitin at the N terminus and RPL40 at the C terminus (244). It was interesting to confirm the effectiveness of UBA52 depletion on SARS-CoV-2 replication. Indeed, two interactome studies did not identify UBA52 or ubiquitination as important for SARS-CoV-2, MERS-CoV, and SARS-CoV (153, 154).

UBA52 depletion significantly reduced SARS-CoV-2 cellular replication in low (58.80%, p = 0.0003) and high (20.35%, p = 0.0288) MOI experiments. Regarding the knockdown validation experiment, WB but not IFA confirmed almost 100% UBA52 knockdown. Regarding the unexpected IFA finding, UBA52 is a known housekeeping gene (242). It was hypothesised that UBA52 is present in high quantities in the cell and could form oligomers or be modified by ubiquitination or phosphorylation. Moreover, the UBA52 antibody could be nonspecific and mark other proteins.

In addition, the cell number of cells treated with UBA52 siRNA was lower than NON-TAR and control cells. These results might reflect UBA52 being important for cellular replication, and knocking it down prevented cells from replicating normally. One study found that the cell cycle is regulated by UBA52 (243). One study examined cell proliferation to understand the function of UBA52, finding that a UBA52-deficient colon cancer cell line showed lower proliferation than a control cell line (243). This study showed that UBA52 functions as a ubiquitin supplier and a regulator of the ribosomal protein complex (243). Moreover, it was also identified as an influenza A virus-host protein interactor using chicken cells, and H5N1 titters significantly decreased after UBA52 knockdown (244). A previous study confirmed the efficiency of UBA52 knockdown using quantitative real-time PCR but did not examine its cellular effects (244).

5.3.2.4 TM9SF2

While TM9SF2 was not confirmed as a cellular interactor for MERS E and M proteins and did not colocalise with them, it was selected for validation (Figure 3-10). TM9SF2 showed

a 3.1-fold higher abundance in co-IPs from lysates of cells expressing the MERS-CoV E protein compared to Flp cells. It was interesting to confirm the effectiveness of TM9SF2 depletion on SARS-CoV-2 replication (see Section 3.2.6). Moreover, it was identified as a cellular interactor for MERS-CoV and SARS-CoV-2 E and M proteins transiently expressed in HEK293 and Dubca cells in eight data sets: HEK SARS2 E, HEK MERS E, PaKiT SARS2 E, PaKiT MERS E, HEK SARS2 M, HEK MERS M, Dubca SARS2 M, and Dubca MERS M. Table 4-34 and Table 4-35 shown the log_2 fold changes in abundance of TM9SF2 in co-IPs from lysates of three host cells expressing SARS-CoV-2 and MERS-CoV E and M proteins compared to negative control cells.

Human TM9SF2 is a Golgi complex-resident transmembrane protein necessary for Golgi complex localisation (208). Several coronavirus M proteins accumulate in the Golgi complex of mammalian host cells (136). In addition, TM9SF2 is an endosomal protein functioning as a channel or small molecule transporter (209). Moreover, TM9SF2 was identified as an important AAV transduction factor (210). Furthermore, it was an important and critical protein for CHIKV infection (211). It was interesting to confirm the effectiveness of TM9SF2 depletion on SARS-CoV-2 replication.

TM9SF2 depletion significantly reduced SARS-CoV-2 cellular replication in low (17.32%, p = 0.0395) and high (5.54%, p = 0.0009) MOI experiments. Regarding the knockdown validation experiment, it was hypothesised and confirmed by WB that TM9SF2 amounts were lower in A549-A2-T2 cell lysates than in HEK293 cell lysates. Moreover, the IFA confirmation results showed that while TM9SF2 could not be detected in A549-A2-T2 cells (**Error! Reference source not found.**), it could be detected in HEK293 cells (Figure 3-10), confirming that TM9SF2 amounts in A549-A2-T2 cells were below the detection limit.

CHAPTER 6 General Discussion and Future Perspectives

COVID-19, a disease caused by SARS-CoV-2, has become a global health problem since 2019. It distributes all over the world. The high number of cases and the complexity of the disease's progression have prompted significant research into the fundamental causes of COVID-19 pathogenesis, which would aid in creating specific clinical interventions. Targeting viral proteins and vaccination are the currently used antiviral development strategy. For example, for targeting viral proteins, remdesivir inhibits the coronaviral RdRp protein (276). In addition, using vaccines has drawbacks, including continuous reformulation as the virus mutates and viral escape from vaccine-mediated immune protection.

An alternative approach attempts to target host factors required for the virus to complete its lifecycle. No established treatments currently target host cell proteins as an antiviral strategy. The effective targeting of host proteins essential for viral replication can avoid therapeutic resistance and provides broad-spectrum antiviral drugs, especially since viral families often use common cellular pathways (153). Most coronavirus (SARS-CoV-2, SARS-CoV, and MERS-CoV) proteins have the same localisation pattern, supporting the previous hypothesis (153).

Initial validation in Chapter 3 of the results for 11 cellular proteins interacting with the MERS-CoV E and M proteins by co-IP /WB analysis and IFA, confirmed 7 of the interactions. Moreover, 4 cellular proteins shown an unconvincing finding including (TM9SF2, ERGIC1, SLC44A2, and RAB10) suggesting to be reanalysed. Also, the interaction confirmation data that shown are preliminary as WB was done only once.

In Chapter 4 three cell lines (human HEK293, bat *Pteropus alecto* PaKiT and *Camelus dromedarius* Dubca) were used for transient expression of the MERS-CoV and SARS-CoV-2 E and M proteins (FLAG epitope-tagged) followed by co-IP and high-throughput mass spectrometry-based interactomic analysis. Bioinformatic analysis revealed E/M protein interactions with ER, Golgi, mitochondrial and nuclear proteins. There were 32 high-confidence cellular interaction proteins conserved amongst the different cell lines and viruses ($p < 0.05$, > 0 $\log_2$ fold change compared to the controls). Moreover, Chapter 5 determined the importance of the 11 cellular proteins interacting with the MERS-CoV E and/or M proteins and 32 cellular proteins conserved across species, in the virus lifecycle, functional validation

was done by siRNA depletion in human cells, followed by infection with SARS-CoV-2. An interesting four cellular proteins were shown to be important for SARS-CoV-2 replication.

Moreover, our study has investigated interesting cellular proteins that appear important for SARS-CoV-2 replication, including CERS2, LPCAT1, UBA52, and TM9SF2, since their knocked down significantly reduced SARS-CoV-2 replication. Three out of these human cellular interactor proteins (CERS2, LPCAT1, and TM9SF2) were detected for MERS-CoV and SARS-CoV E and M proteins. While UBA52 was detected as a cellular interactor for MERS-CoV and SARS-CoV E protein only. These findings can be followed up by investigating whether drugs can safely target these proteins to reduce viral replication in cells. Up to date I could not find any study that mentioned any therapeutic pathway that can used these proteins. Figure 6-1 was shown a Schematic representation of the project Main aim and followed steps.

In addition, as a future direction, proteins whose knockdown caused a nonsignificant SARS-CoV-2 replication reduction could be tested and validated, including TUBG1, SAAL1, and IPO11. Moreover, MERS-CoV replication could be used to validate the effectiveness of shared cellular interactor depletion. In addition, proteins involved in cellular N-glycan biosynthesis and protein glycosylation (STT3A, RPN2, RPN1, DPAGT1, and ALG1) that are validated coronaviral M protein interactors (153) could be studied and determine how they are important for coronaviral replication.

Moreover, the limitations of cellular interactor selection in Section 4.2.5 could be addressed and shared cellular interactors for each viral protein in three host cells (HEK293, PaKiT, and Dubca) could be identified and validated. This approach would include investigating shared cellular interactors for the SARS-CoV-2 E protein, SARS-CoV-2 M protein, MERS-CoV E protein, and MERS-CoV M protein in the three host cells.

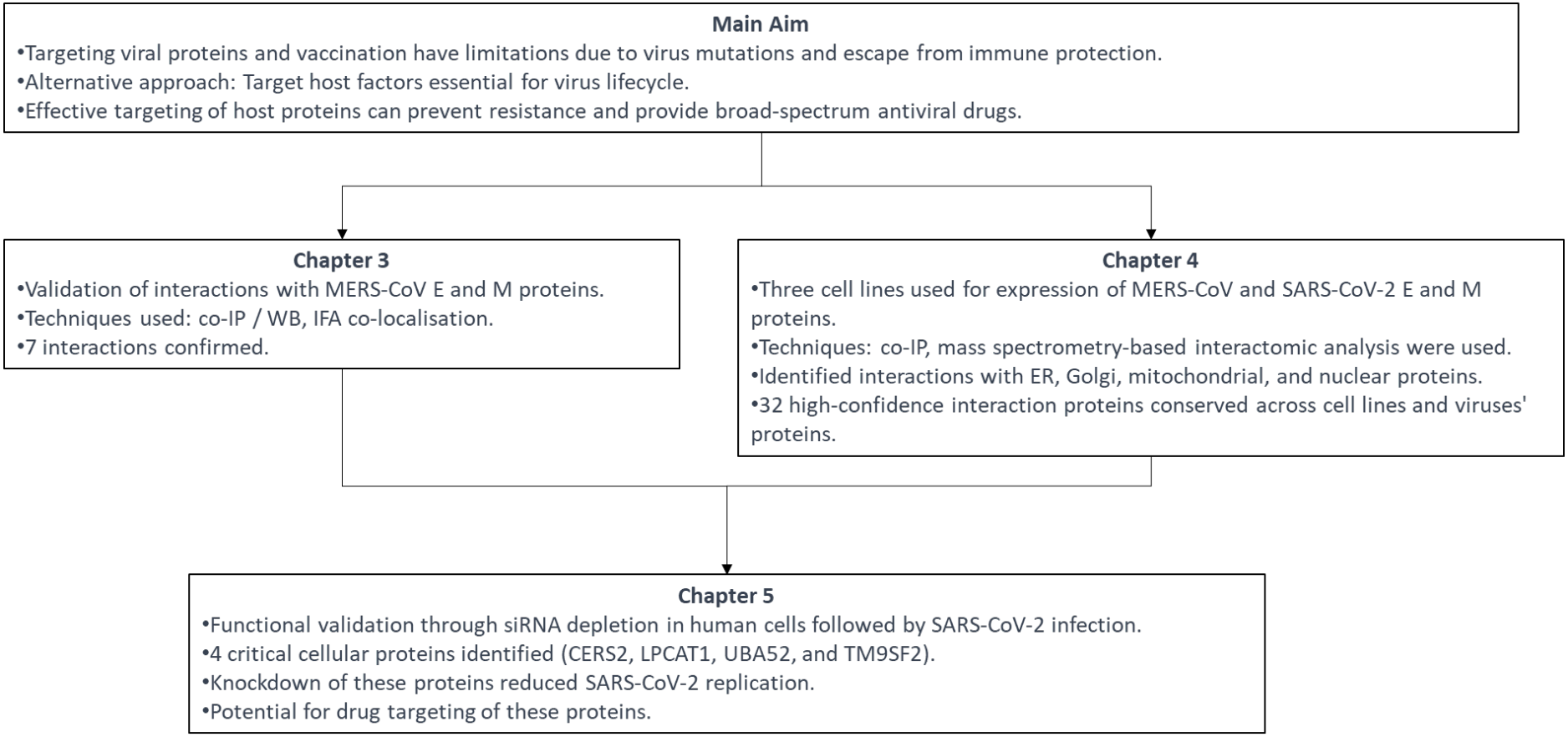


Figure 6-1 Schematic representation of the project Main aim and followed steps.

References

1. Chua KB, Bellini WJ, Rota PA, Harcourt BH, Tamin A, Lam SK, et al. Nipah virus: A recently emergent deadly paramyxovirus. Science. 2000;288(5470):1432-5.
2. Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiology and molecular biology reviews. 2005;69(4):635-64.
3. Luna LKD, Heiser V, Regamey N, Panning M, Drexler JF, Mulangu S, et al. Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. J Clin Microbiol. 2007;45(3):1049-52.
4. Wang LF, Anderson DE. Viruses in bats and potential spillover to animals and humans. Curr Opin Virol. 2019;34:79-89.
5. Ding Y, He L, Zhang Q, Huang Z, Che X, Hou J, et al. Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways. J Pathol. 2004;203(2):622-30.
6. Al-Osail AM, Al-Wazzah MJ. The history and epidemiology of Middle East respiratory syndrome corona virus. Multidisciplinary respiratory medicine. 2017;12:20.
7. Heydari H, Golmohammadi R, Mirnejad R, Tebyanian H, Fasihi-Ramandi M, Moosazadeh Moghaddam M. Antiviral peptides against Coronaviridae family: A review. Peptides. 2021;139:170526.
8. Abdul-Rasool S, Fielding BC. Understanding Human Coronavirus HCoV-NL63. Open Virol J. 2010;4:76-84.
9. Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J Virol. 2005;79(2):884-95.
10. Assiri A, Al-Tawfiq JA, Al-Rabeeah AA, Al-Rabiah FA, Al-Hajjar S, Al-Barrak A, et al. Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. 2013;13(9):752-61.
11. Veit S, Jany S, Fux R, Sutter G, Volz A. CD8+ T Cells Responding to the Middle East Respiratory Syndrome Coronavirus Nucleocapsid Protein Delivered by Vaccinia Virus MVA in Mice. Viruses. 2018;10(12).
12. Buchholz U, Muller MA, Nitsche A, Sanewski A, Wevering N, Bauer-Balci T, et al. Contact investigation of a case of human novel coronavirus infection treated in a German hospital, October-November 2012. Eurosurveillance. 2013;18(8):6-12.
13. World Health Organization. MERS SITUATION UPDATE 2023 [updated January 2023]. Available from: https://www.emro.who.int/health-topics/mers-cov/mers-outbreaks.html.
14. Chatterjee A, Seyfferth J, Lucci J, Gilsbach R, Preissl S, Bottinger L, et al. MOF Acetyl Transferase Regulates Transcription and Respiration in Mitochondria. Cell. 2016;167(3):722-38 e23.
15. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. Asian Pac J Allergy Immunol. 2020;38(1):1-9.
16. Chen X, Chughtai AA, Dyda A, MacIntyre CR. Comparative epidemiology of Middle East respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia and South Korea. Emerging microbes & infections. 2017;6(1):1-6.
17. Shokri S, Mahmoudvand S, Taherkhani R, Farshadpour F. Modulation of the immune response by Middle East respiratory syndrome coronavirus. J Cell Physiol. 2019;234(3):2143-51.
18. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497-506.
19. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581(7807):215-20.

20. Coronaviridae Study Group of the International Committee on Taxonomy of V. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5(4):536-44.
21. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med. 2020;382(13):1199-207.
22. Bogoch, II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. J Travel Med. 2020;27(2).
23. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed. 2020;91(1):157-60.
24. Organization WH. WHO Coronavirus (COVID-19) Dashboard: World Health Organization; 2022 [updated 05 01 2022. Available from: https://covid19.who.int/.
25. Torres-Ibarra L, Basto-Abreu A, Carnalla M, Torres-Alvarez R, Reyes-Sanchez F, Hernández-Ávila JE, et al. SARS-CoV-2 infection fatality rate after the first epidemic wave in Mexico. International journal of epidemiology. 2022;51(2):429-39.
26. Thomas S. The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. Pathog Immun. 2020;5(1):342-63.
27. da Costa VG, Moreli ML, Saivish MV. The emergence of SARS, MERS and novel SARS-2 coronaviruses in the 21st century. Arch Virol. 2020;165(7):1517-26.
28. Bloomberg. Mapping the Coronavirus Outbreak Across the World https://www.bloomberg.com/graphics/2020-coronavirus-cases-world-map/: Bloomberg; 2021 [updated 26 December 2021, 14:21 GMT.
29. Mukhra R, Krishan K, Kanchan T. Possible modes of transmission of Novel coronavirus SARS-CoV-2: a review. Acta Biomed. 2020;91(3):e2020036.
30. Chan JFW, Lau SKP, To KKW, Cheng VCC, Woo PCY, Yuen KY. Middle East Respiratory Syndrome Coronavirus: Another Zoonotic Betacoronavirus Causing SARS-Like Disease. Clin Microbiol Rev. 2015;28(2):465-522.
31. Mittal A, Manjunath K, Ranjan RK, Kaushik S, Kumar S, Verma V. COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. PLoS Pathog. 2020;16(8):e1008762.
32. Graham RL, Donaldson EF, Baric RS. A decade after SARS: strategies for controlling emerging coronaviruses. Nature reviews Microbiology. 2013;11(12):836-48.
33. Li WD, Shi ZL, Yu M, Ren WZ, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310(5748):676-9.
34. Malik YA. Properties of Coronavirus and SARS-CoV-2. Malays J Pathol. 2020;42(1):3-11.
35. Chan RWY, Hemida MG, Kayali G, Chu DKW, Poon LLM, Alnaeem A, et al. Tropism and replication of Middle East respiratory syndrome coronavirus from dromedary camels in the human respiratory tract: an in-vitro and ex-vivo study. Lancet Resp Med. 2014;2(10):813-22.
36. Reusken CB, Farag EA, Jonges M, Godeke GJ, El-Sayed AM, Pas SD, et al. Middle East respiratory syndrome coronavirus (MERS-CoV) RNA and neutralising antibodies in milk collected according to local customs from dromedary camels, Qatar, April 2014. Eurosurveillance. 2014;19(23):8-12.
37. Mackay IM, Arden KE. An Opportunistic Pathogen Afforded Ample Opportunities: Middle East Respiratory Syndrome Coronavirus. Viruses. 2017;9(12).
38. Memish ZA, Cotten M, Meyer B, Watson SJ, Alsahafi AJ, Al Rabeeah AA, et al. Human Infection with MERS Coronavirus after Exposure to Infected Camels, Saudi Arabia, 2013. Emerging Infectious Diseases. 2014;20(6):1012-5.
39. Miguel E, Chevalier V, Ayelet G, Ben Bencheikh MN, Boussini H, Chu DK, et al. Risk factors for MERS coronavirus infection in dromedary camels in Burkina Faso, Ethiopia, and Morocco, 2015. Eurosurveillance. 2017;22(13):15-24.

40. Centers for Disease Control and Prevention. COVID-19 2022 [updated 01 2022. Available from: https://www.cdc.gov/coronavirus/2019-ncov/index.html.
41. Zhao J, Cui W, Tian B-p. The potential intermediate hosts for SARS-CoV-2. Frontiers in microbiology. 2020;11:580137.
42. Mahdy MA, Younis W, Ewaida Z. An overview of SARS-CoV-2 and animal infection. Frontiers in veterinary science. 2020;7:596391.
43. Fenollar F, Mediannikov O, Maurin M, Devaux C, Colson P, Levasseur A, et al. Mink, SARS-CoV-2, and the human-animal interface. Frontiers in Microbiology. 2021;12:663815.
44. Team C-I. Clinical and virologic characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States. Nat Med. 2020;26(6):861-8.
45. Choi S, Jung E, Choi BY, Hur YJ, Ki M. High reproduction number of Middle East respiratory syndrome coronavirus in nosocomial outbreaks: mathematical modelling in Saudi Arabia and South Korea. J Hosp Infect. 2018;99(2):162-8.
46. Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Trends Immunol. 2020;41(12):1100-15.
47. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. N Engl J Med. 2020;382(16):1564-7.
48. Belser JA, Rota PA, Tumpey TM. Ocular tropism of respiratory viruses. Microbiol Mol Biol Rev. 2013;77(1):144-56.
49. Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. Lancet Infect Dis. 2020;20(4):411-2.
50. Morawska L, Cao J. Airborne transmission of SARS-CoV-2: The world should face the reality. Environ Int. 2020;139:105730.
51. Young BE, Ong SWX, Kalimuddin S, Low JG, Tan SY, Loh J, et al. Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. JAMA. 2020;323(15):1488-94.
52. Xiao F, Tang M, Zheng X, Liu Y, Li X, Shan H. Evidence for Gastrointestinal Infection of SARS-CoV-2. Gastroenterology. 2020;158(6):1831-3 e3.
53. Weber DJ, Rutala WA, Fischer WA, Kanamori H, Sickbert-Bennett EE. Emerging infectious diseases: Focus on infection control issues for novel coronaviruses (Severe Acute Respiratory Syndrome-CoV and Middle East Respiratory Syndrome-CoV), hemorrhagic fever viruses (Lassa and Ebola), and highly pathogenic avian influenza viruses, A(H5N1) and A(H7N9). Am J Infect Control. 2016;44(5 Suppl):e91-e100.
54. Wu Y, Guo C, Tang L, Hong Z, Zhou J, Dong X, et al. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. Lancet Gastroenterol Hepatol. 2020;5(5):434-5.
55. Otter JA, Donskey C, Yezli S, Douthwaite S, Goldenberg SD, Weber DJ. Transmission of SARS and MERS coronaviruses and influenza virus in healthcare settings: the possible role of dry surface contamination. J Hosp Infect. 2016;92(3):235-50.
56. van Doremalen N, Bushmaker T, Munster VJ. Stability of Middle East respiratory syndrome coronavirus (MERS-CoV) under different environmental conditions. Euro Surveill. 2013;18(38).
57. Morawska L, Milton DK. It Is Time to Address Airborne Transmission of Coronavirus Disease 2019 (COVID-19). Clin Infect Dis. 2020;71(9):2311-3.
58. Fears AC, Klimstra WB, Duprex P, Hartman A, Weaver SC, Plante KS, et al. Persistence of Severe Acute Respiratory Syndrome Coronavirus 2 in Aerosol Suspensions. Emerg Infect Dis. 2020;26(9).
59. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA. 2020;323(11):1061-9.
60. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, et al. Covid-19 in Critically Ill Patients in the Seattle Region - Case Series. N Engl J Med. 2020;382(21):2012-22.

61. Azekawa S, Namkoong H, Mitamura K, Kawaoka Y, Saito F. Co-infection with SARS-CoV-2 and influenza A virus. IDCases. 2020;20:e00775.
62. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;181(2):271-80 e8.
63. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020;8(5):475-81.
64. Li CH, Chiou HC, Lin MH, Kuo CH, Lin YC, Lin YC, et al. Immunological map in COVID-19. J Microbiol Immunol Infect. 2021;54(4):547-56.
65. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med. 2020;382(22):2081-90.
66. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA. 2020;323(18):1843-4.
67. Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. Presymptomatic Transmission of SARS-CoV-2 - Singapore, January 23-March 16, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(14):411-5.
68. Edridge AWD, Kaczorowska J, Hoste ACR, Bakker M, Klein M, Loens K, et al. Seasonal coronavirus protective immunity is short-lasting. Nat Med. 2020.
69. Augustine R, Das S, Hasan A, S A, Abdul Salam S, Augustine P, et al. Rapid Antibody-Based COVID-19 Mass Surveillance: Relevance, Challenges, and Prospects in a Pandemic and Post-Pandemic World. J Clin Med. 2020;9(10).
70. Zhang W, Du RH, Li B, Zheng XS, Yang XL, Hu B, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. Emerg Microbes Infect. 2020;9(1):386-9.
71. Chafekar A, Fielding BC. MERS-CoV: Understanding the Latest Human Coronavirus Threat. Viruses. 2018;10(2).
72. Memish ZA, Al-Tawfiq JA, Makhdoom HQ, Assiri A, Alhakeem RF, Albarrak A, et al. Respiratory tract samples, viral load, and genome fraction yield in patients with Middle East respiratory syndrome. J Infect Dis. 2014;210(10):1590-4.
73. Grant PR, Garson JA, Tedder RS, Chan PK, Tam JS, Sung JJ. Detection of SARS coronavirus in plasma by real-time RT-PCR. N Engl J Med. 2003;349(25):2468-9.
74. Yan Y, Chang L, Wang L. Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): Current status, challenges, and countermeasures. Rev Med Virol. 2020;30(3):e2106.
75. Organization WH. Antigen-detection in the diagnosis of SARS-CoV-2 infection 2021 [updated 06 10 2021.
76. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
77. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet. 2020;395(10223):470-3.
78. National Health Service . Middle East respiratory syndrome (MERS) [updated 29 September 2021].

]. Available from: https://www.nhs.uk/conditions/middle-east-respiratory-syndrome-mers/.
79. Al-Tawfiq JA, Memish ZA. Diagnosis of SARS-CoV-2 infection based on CT scan vs RT-PCR: reflecting on experience from MERS-CoV. Journal of Hospital Infection. 2020;105(2):154-5.
80. Corman VM, Eckerle I, Bleicker T, Zaki A, Landt O, Eschbach-Bludau M, et al. Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction. Euro Surveill. 2012;17(39).
81. Corman VM, Muller MA, Costabel U, Timm J, Binger T, Meyer B, et al. Assays for laboratory confirmation of novel human coronavirus (hCoV-EMC) infections. Euro Surveill. 2012;17(49).

82. Lu X, Whitaker B, Sakthivel SK, Kamili S, Rose LE, Lowe L, et al. Real-time reverse transcription-PCR assay panel for Middle East respiratory syndrome coronavirus. J Clin Microbiol. 2014;52(1):67-75.
83. Corman VM, Albarrak AM, Omrani AS, Albarrak MM, Farah ME, Almasri M, et al. Viral Shedding and Antibody Response in 37 Patients With Middle East Respiratory Syndrome Coronavirus Infection. Clin Infect Dis. 2016;62(4):477-83.
84. Qiu M, Shi Y, Guo Z, Chen Z, He R, Chen R, et al. Antibody responses to individual proteins of SARS coronavirus and their neutralization activities. Microbes Infect. 2005;7(5-6):882-9.
85. National Health Service. Coronavirus (COVID-19) ,2021 https://www.nhs.uk/conditions/coronavirus-covid-19/self-isolation-and-treatment/when-to-self-isolate-and-what-to-do/ [updated 11 November 2021.
86. Totura AL, Baric RS. SARS coronavirus pathogenesis: host innate immune responses and viral antagonism of interferon. Curr Opin Virol. 2012;2(3):264-75.
87. Yang Y, Zhang L, Geng H, Deng Y, Huang B, Guo Y, et al. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. Protein Cell. 2013;4(12):951-61.
88. Widagdo W, Raj VS, Schipper D, Kolijn K, van Leenders G, Bosch BJ, et al. Differential Expression of the Middle East Respiratory Syndrome Coronavirus Receptor in the Upper Respiratory Tracts of Humans and Dromedary Camels. J Virol. 2016;90(9):4838-42.
89. Sariol A, Perlman S. Lessons for COVID-19 Immunity from Other Coronavirus Infections. Immunity. 2020;53(2):248-63.
90. Liu J, Zheng X, Tong Q, Li W, Wang B, Sutter K, et al. Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. J Med Virol. 2020;92(5):491-4.
91. Zhou J, Chu H, Li C, Wong BH, Cheng ZS, Poon VK, et al. Active replication of Middle East respiratory syndrome coronavirus and aberrant induction of inflammatory cytokines and chemokines in human macrophages: implications for pathogenesis. J Infect Dis. 2014;209(9):1331-42.
92. Li J, Guo M, Tian X, Wang X, Yang X, Wu P, et al. Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis. Med (N Y). 2021;2(1):99-112 e7.
93. Freundt EC, Yu L, Park E, Lenardo MJ, Xu XN. Molecular determinants for subcellular localization of the severe acute respiratory syndrome coronavirus open reading frame 3b protein. J Virol. 2009;83(13):6631-40.
94. Lu X, Pan J, Tao J, Guo D. SARS-CoV nucleocapsid protein antagonizes IFN-beta response by targeting initial step of IFN-beta induction pathway, and its C-terminal region is critical for the antagonism. Virus Genes. 2011;42(1):37-45.
95. Chu H, Zhou J, Wong BH, Li C, Chan JF, Cheng ZS, et al. Middle East Respiratory Syndrome Coronavirus Efficiently Infects Human Primary T Lymphocytes and Activates the Extrinsic and Intrinsic Apoptosis Pathways. J Infect Dis. 2016;213(6):904-14.
96. Yeung ML, Yao Y, Jia L, Chan JF, Chan KH, Cheung KF, et al. MERS coronavirus induces apoptosis in kidney and lung by upregulating Smad7 and FGF2. Nat Microbiol. 2016;1:16004.
97. Channappanavar R, Fehr AR, Vijay R, Mack M, Zhao J, Meyerholz DK, et al. Dysregulated Type I Interferon and Inflammatory Monocyte-Macrophage Responses Cause Lethal Pneumonia in SARS-CoV-Infected Mice. Cell Host Microbe. 2016;19(2):181-93.
98. Bolles M, Donaldson E, Baric R. SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission. Curr Opin Virol. 2011;1(6):624-34.
99. Zhand S, Saghaeian Jazi M, Mohammadi S, Tarighati Rasekhi R, Rostamian G, Kalani MR, et al. COVID-19: The immune responses and clinical therapy candidates. International journal of molecular sciences. 2020;21(15):5559.

100. Chazal N. Coronavirus, the King Who Wanted More Than a Crown: From Common to the Highly Pathogenic SARS-CoV-2, Is the Key in the Accessory Genes? Frontiers in Microbiology. 2021:1970.
101. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3.
102. Zhou Y, Yang Y, Huang J, Jiang S, Du L. Advances in MERS-CoV Vaccines and Therapeutics Based on the Receptor-Binding Domain. Viruses. 2019;11(1).
103. Meyers JM, Ramanathan M, Shanderson RL, Donohue L, Ferguson I, Guo MG, et al. The proximal proteome of 17 SARS-CoV-2 proteins links to disrupted antiviral signaling and host translation. bioRxiv. 2021.
104. He J, Hu L, Huang X, Wang C, Zhang Z, Wang Y, et al. Potential of coronavirus 3C-like protease inhibitors for the development of new anti-SARS-CoV-2 drugs: Insights from structures of protease and inhibitors. 2020;56(2):106055.
105. Mustafa S, Balkhy H, Gabere MN. Current treatment options and the role of peptides as potential therapeutic components for Middle East Respiratory Syndrome(MERS): A review. Journal of Infection and Public Health. 2018;11(1):9-17.
106. Redondo N, Zaldívar-López S, Garrido JJ, Montoya M. SARS-CoV-2 accessory proteins in viral pathogenesis: knowns and unknowns. J Frontiers in Immunology. 2021:2698.
107. Kalhori MR, Saadatpour F, Arefian E, Soleimani M, Farzaei MH, Aneva IY, et al. The potential therapeutic effect of RNA interference and natural products on COVID-19: a review of the coronaviruses infection. Frontiers in Pharmacology. 2021:116.
108. Yang Y, Zhang L, Geng HY, Deng Y, Huang BY, Guo Y, et al. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. Protein & Cell. 2013;4(12):951-61.
109. Ruch TR, Machamer CE. The Coronavirus E Protein: Assembly and Beyond. Viruses-Basel. 2012;4(3):363-82.
110. Maeda J, Repass JF, Maeda A, Makino S. Membrane topology of coronavirus E protein. Virology. 2001;281(2):163-9.
111. Duart G, Garcia-Murria MJ, Grau B, Acosta-Caceres JM, Martinez-Gil L, Mingarro I. SARS-CoV-2 envelope protein topology in eukaryotic membranes. Open Biol. 2020;10(9):200209.
112. Verdia-Baguena C, Nieto-Torres JL, Alcaraz A, Dediego ML, Enjuanes L, Aguilella VM. Analysis of SARS-CoV E protein ion channel activity by tuning the protein and lipid charge. Biochim Biophys Acta. 2013;1828(9):2026-31.
113. Maeda J, Repass JF, Maeda A, Makino S. Membrane topology of coronavirus E protein. Virology. 2001;281(2):163-9.
114. Park SH, Siddiqi H, Castro DV, De Angelis AA, Oom AL, Stoneham CA, et al. Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. PLoS Pathog. 2021;17(5):e1009519.
115. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Virol J. 2019;16(1):69.
116. Hazari R, Pal Chaudhuri P. Analysis of coronavirus envelope protein with cellular automata model. International Journal of Parallel, Emergent and Distributed Systems. 2022;37(6):623-48.
117. Westerbeck JW, Machamer CE. A Coronavirus E Protein Is Present in Two Distinct Pools with Different Effects on Assembly and the Secretory Pathway. J Virol. 2015;89(18):9313-23.
118. Westerbeck JW, Machamer CE. The Infectious Bronchitis Coronavirus Envelope Protein Alters Golgi pH To Protect the Spike Protein and Promote the Release of Infectious Virus. J Virol. 2019;93(11).
119. Rossman JS, Lamb RA. Viral membrane scission. Annu Rev Cell Dev Biol. 2013;29:551-69.
120. Martyna A, Gomez-Llobregat J, Linden M, Rossman JS. Curvature Sensing by a Viral Scission Protein. Biochemistry. 2016;55(25):3493-6.

121. Klumperman J, Locker JK, Meijer A, Horzinek MC, Geuze HJ, Rottier PJ. Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. J Virol. 1994;68(10):6523-34.
122. Machamer CE, Mentone SA, Rose JK, Farquhar MG. The E1 glycoprotein of an avian coronavirus is targeted to the cis Golgi complex. Proc Natl Acad Sci U S A. 1990;87(18):6944-8.
123. Ortego J, Escors D, Laude H, Enjuanes L. Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. J Virol. 2002;76(22):11518-29.
124. Venkatagopalan P, Daskalova SM, Lopez LA, Dolezal KA, Hogue BG. Coronavirus envelope (E) protein remains at the site of assembly. Virology. 2015;478:75-85.
125. Ye Y, Hogue BG. Role of the coronavirus E viroporin protein transmembrane domain in virus assembly. J Virol. 2007;81(7):3597-607.
126. Nieto-Torres JL, DeDiego ML, Verdia-Baguena C, Jimenez-Guardeno JM, Regla-Nava JA, Fernandez-Delgado R, et al. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. PLoS Pathog. 2014;10(5):e1004077.
127. Fischer F, Stegen CF, Masters PS, Samsonoff WA. Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. J Virol. 1998;72(10):7885-94.
128. DeDiego ML, Alvarez E, Almazan F, Rejas MT, Lamirande E, Roberts A, et al. A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. J Virol. 2007;81(4):1701-13.
129. Rossman JS, Jing X, Leser GP, Lamb RA. Influenza virus M2 protein mediates ESCRT-independent membrane scission. Cell. 2010;142(6):902-13.
130. Surya W, Li Y, Verdia-Baguena C, Aguilella VM, Torres J. MERS coronavirus envelope protein has a single transmembrane domain that forms pentameric ion channels. Virus Res. 2015;201:61-6.
131. Wilson L, Gage P, Ewart G. Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. Virology. 2006;353(2):294-306.
132. The Universal Protein Resource ,2021 2021 [updated February 2, 2021]. Available from: https://www.uniprot.org/.
133. Resource TUP. Protein database ,2019 [updated October 15, 2019. Available from: https://www.uniprot.org/.
134. de Haan CA, Rottier PJ. Molecular interactions in the assembly of coronaviruses. Adv Virus Res. 2005;64:165-230.
135. Narayanan K, Maeda A, Maeda J, Makino S. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. J Virol. 2000;74(17):8127-34.
136. Nal B, Chan C, Kien F, Siu L, Tse J, Chu K, et al. Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. J Gen Virol. 2005;86(Pt 5):1423-34.
137. Tooze J, Tooze S, Warren G. Replication of coronavirus MHV-A59 in sac- cells: determination of the first site of budding of progeny virions. Eur J Cell Biol. 1984;33(2):281-93.
138. Perrier A, Bonnin A, Desmarets L, Danneels A, Goffard A, Rouille Y, et al. The C-terminal domain of the MERS coronavirus M protein contains a trans-Golgi network localization signal. J Biol Chem. 2019;294(39):14406-21.
139. Arndt AL, Larson BJ, Hogue BG. A conserved domain in the coronavirus membrane protein tail is important for virus assembly. J Virol. 2010;84(21):11418-28.
140. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res. 2020;24:91-8.
141. Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, et al. From SARS to MERS, Thrusting Coronaviruses into the Spotlight. Viruses. 2019;11(1).
142. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol. 2015;1282:1-23.

143. Kuo L, Masters PS. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. Journal of virology. 2002;76(10):4987-99.
144. Fritz M, Rosolen B, Krafft E, Becquart P, Elguero E, Vratskikh O, et al. High prevalence of SARS-CoV-2 antibodies in pets from COVID-19+ households. One Health. 2020;11:100192.
145. Sun YP, Zhang HD, Shi J, Zhang Z, Gong R. Identification of a Novel Inhibitor against Middle East Respiratory Syndrome Coronavirus. Viruses-Basel. 2017;9(9).
146. Liu Y, Ning Z, Chen Y, Guo M, Liu Y, Gali NK, et al. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. Nature. 2020;582(7813):557-60.
147. Li YD, Chi WY, Su JH, Ferrall L, Hung CF, Wu TC. Coronavirus vaccine development: from SARS and MERS to COVID-19. J Biomed Sci. 2020;27(1):104.
148. van Riel D, de Wit E. Next-generation vaccine platforms for COVID-19. Nat Mater. 2020;19(8):810-2.
149. Rauch S, Jasny E, Schmidt KE, Petsch B. New Vaccine Technologies to Combat Outbreak Situations. Front Immunol. 2018;9:1963.
150. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 Vaccines at Pandemic Speed. N Engl J Med. 2020;382(21):1969-73.
151. Organization WH. Coronavirus disease (COVID-19): Vaccines 2021 [updated 29 November 2021. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-(covid-19)-vaccines?adgroupsurvey={adgroupsurvey}&gclid=Cj0KCQiAqbyNBhC2ARIsALDwAsDZnLu4j6fu_Vlxmu5JqaKcE6NngRbwGY_HjaKeXAhh0zIgmlguPPgaArw8EALw_wcB].
152. Richards AL, Eckhardt M, Krogan NJ. Mass spectrometry-based protein–protein interaction networks for the study of human diseases. Molecular systems biology. 2021;17(1):e8792.
153. Gordon DE, Hiatt J, Bouhaddou M, Rezelj VV, Ulferts S, Braberg H, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science. 2020; 370(6521).
154. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583(7816):459-68.
155. Stukalov A, Girault V, Grass V, Karayel O, Bergant V, Urban C, et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. 2021;594(7862):246-52.
156. Crameri G, Todd S, Grimley S, McEachern JA, Marsh GA, Smith C, et al. Establishment, immortalisation and characterisation of pteropid bat cell lines. PloS one. 2009;4(12):e8266.
157. Rihn SJ, Merits A, Bakshi S, Turnbull ML, Wickenhagen A, Alexander AJ, et al. A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. PLoS biology. 2021;19(2):e3001091.
158. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature. 1970;227(5259):680-5.
159. Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic acids research. 2003;31(1):258-61.
160. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics. 2015;15(15):2597-601.
161. Huang DW, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, et al. Extracting biological meaning from large gene lists with DAVID. Current protocols in bioinformatics. 2009;27(1):13.1. 1-.1. .
162. Koliński M, Kałużna E, Piwecka M. RNA–protein interactomes as invaluable resources to study RNA viruses: Insights from SARS CoV-2 studies. Wiley Interdisciplinary Reviews: RNA. 2022:e1727.

163. Kamel W, Noerenberg M, Cerikan B, Chen H, Järvelin AI, Kammoun M, et al. Global analysis of protein-RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection. Molecular cell. 2021;81(13):2851-67. e7.
164. Flynn RA, Belk JA, Qi Y, Yasumoto Y, Wei J, Alfajaro MM, et al. Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions. Cell. 2021;184(9):2394-411. e16.
165. Lee S, Lee Y-s, Choi Y, Son A, Park Y, Lee K-M, et al. The SARS-CoV-2 rna interactome. Molecular cell. 2021;81(13):2838-50. e6.
166. Schmidt N, Lareau CA, Keshishian H, Ganskih S, Schneider C, Hennig T, et al. The SARS-CoV-2 RNA–protein interactome in infected human cells. Nature microbiology. 2021;6(3):339-53.
167. Lum KK, Cristea IM. Proteomic approaches to uncovering virus–host protein interactions during the progression of viral infection. Expert review of proteomics. 2016;13(3):325-40.
168. Chang IF. Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes. Proteomics. 2006;6(23):6158-66.
169. Einhauer A, Jungbauer A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. Journal of biochemical and biophysical methods. 2001;49(1-3):455-65.
170. Zhao X, Li G, Liang S. Several affinity tags commonly used in chromatographic purification. Journal of analytical methods in chemistry. 2013;2013.
171. Waugh DS. Making the most of affinity tags. Trends in biotechnology. 2005;23(6):316-20.
172. Cristea IM, Carroll J-WN, Rout MP, Rice CM, Chait BT, MacDonald MR. Tracking and elucidating alphavirus-host protein interactions. Journal of Biological Chemistry. 2006;281(40):30269-78.
173. V'kovski P, Gerber M, Kelly J, Pfaender S, Ebert N, Lagache SB, et al. Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. Elife. 2019;8:e42037.
174. Cristea IM, Williams R, Chait BT, Rout MP. Fluorescent proteins as proteomic probes. Molecular & Cellular Proteomics. 2005;4(12):1933-41.
175. Zhang L, Zhang Z-P, Zhang X-E, Lin F-S, Ge FJJov. Quantitative proteomics analysis reveals BAG3 as a potential target to suppress severe acute respiratory syndrome coronavirus replication. 2010;84(12):6050-9.
176. Vogels MW, Van Balkom BW, Kaloyanova DV, Batenburg JJ, Heck AJ, Helms JB, et al. Identification of host factors involved in coronavirus replication by quantitative proteomics analysis. 2011;11(1):64-80.
177. Emmott E, Munday D, Bickerton E, Britton P, Rodgers MA, Whitehouse A, et al. The cellular interactome of the coronavirus infectious bronchitis virus nucleocapsid protein and functional implications for virus biology. 2013;87(17):9486-500.
178. Coombs KMJErop. Quantitative proteomics of complex mixtures. 2011;8(5):659-77.
179. Corse E, Machamer CEJJov. Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. 2000;74(9):4319-26.
180. Basrur V, Yang F, Kushimoto T, Higashimoto Y, Yasumoto K-i, Valencia J, et al. Proteomic analysis of early melanosomes: identification of novel melanosomal proteins. Journal of proteome research. 2003;2(1):69-79.
181. Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. Nature methods. 2013;10(8):730-6.
182. Ran Y, Xiong MG, Xu ZS, Luo WW, Wang SY, Wang YY. YIPF5 Is Essential for Innate Immunity to DNA Virus and Facilitates COPII-Dependent STING Trafficking. J Immunol. 2019.
183. Yoshida Y, Suzuki K, Yamamoto A, Sakai N, Bando M, Tanimoto K, et al. YIPF5 and YIF1A recycle between the ER and the Golgi apparatus and are involved in the maintenance of the Golgi structure. Exp Cell Res. 2008;314(19):3427-43.

184. Bennett JA, Mastrangelo MA, Ture SK, Smith CO, Loelius SG, Berg RA, et al. The choline transporter Slc44a2 controls platelet activation and thrombosis by regulating mitochondrial function. 2020;11(1):1-9.
185. Gatti P, Ilamathi HS, Todkar K, Germain M. Mitochondria targeted viral replication and survival strategies—prospective on SARS-CoV-2. Frontiers in pharmacology. 2020;11:578599.
186. Fenech EJ, Lari F, Charles PD, Fischer R, Laétitia-Thézénas M, Bagola K, et al. Interaction mapping of endoplasmic reticulum ubiquitin ligases identifies modulators of innate immune signalling. Elife. 2020;9:e57306.
187. Boruchowicz H, Hawkins J, Cruz-Palomar K, Lippé R. The XPO6 exportin mediates herpes simplex virus 1 gM nuclear release late in infection. Journal of Virology. 2020;94(21):e00753-20.
188. Laviad EL, Albee L, Pankova-Kholmyansky I, Epstein S, Park H, Merrill AH, et al. Characterization of ceramide synthase 2: tissue distribution, substrate specificity, and inhibition by sphingosine 1-phosphate. Journal of Biological Chemistry. 2008;283(9):5677-84.
189. Pewzner-Jung Y, Ben-Dor S, Futerman AH. When do Lasses (longevity assurance genes) become CerS (ceramide synthases)?: Insights into the regulation of ceramide synthesis. Journal of Biological Chemistry. 2006;281(35):25001-5.
190. Barklis E, Alfadhli A, Kyle JE, Bramer LM, Bloodsworth KJ, Barklis RL, et al. Ceramide synthase 2 deletion decreases the infectivity of HIV-1. Journal of Biological Chemistry. 2021;296.
191. Chen L, Zheng S. Understand variability of COVID-19 through population and tissue variations in expression of SARS-CoV-2 host genes. Informatics in medicine unlocked. 2020;21:100443.
192. Nakamura R, Misawa K, Tohnai G, Nakatochi M, Furuhashi S, Atsuta N, et al. A multi-ethnic meta-analysis identifies novel genes, including ACSL5, associated with amyotrophic lateral sclerosis. Communications biology. 2020;3(1):1-9.
193. Sicari D, Chatziioannou A, Koutsandreas T, Sitia R, Chevet E. Role of the early secretory pathway in SARS-CoV-2 infection. Journal of Cell Biology. 2020;219(9).
194. Moessinger C, Kuerschner L, Spandl J, Shevchenko A, Thiele C. Human lysophosphatidylcholine acyltransferases 1 and 2 are located in lipid droplets where they catalyze the formation of phosphatidylcholine. Journal of Biological Chemistry. 2011;286(24):21330-9.
195. Soupene E, Fyrst H, Kuypers FA. Mammalian acyl-CoA: lysophosphatidylcholine acyltransferase enzymes. Proceedings of the National Academy of Sciences. 2008;105(1):88-93.
196. Cheng S, Chen H, Wang A, Xie M, Xie J, Osanai K, et al. Lentiviral vector-mediated delivery of lysophosphatidylcholine acyltransferase 1 attenuates airway inflammation in ovalbumin-induced allergic asthmatic mice. Asian Pac J Allergy Immunol. 2015;33:320-9.
197. Nakanishi H, Shindou H, Hishikawa D, Harayama T, Ogasawara R, Suwabe A, et al. Cloning and characterization of mouse lung-type acyl-CoA: lysophosphatidylcholine acyltransferase 1 (LPCAT1): expression in alveolar type II cells and possible involvement in surfactant production. Journal of Biological Chemistry. 2006;281(29):20140-7.
198. Lu J-Y, Brewer G, Li M-L, Lin K-Z, Huang C-C, Yen L-C, et al. Secretory Carrier Membrane Protein 3 Interacts with 3A Viral Protein of Enterovirus and Participates in Viral Replication. Microbiology Spectrum. 2021;9(1):e00475-21.
199. Wu T, Zheng X, Yang M, Zhao A, Li M, Chen T, et al. Serum lipid alterations identified in chronic hepatitis B, hepatitis B virus-associated cirrhosis and carcinoma patients. Sci Rep. 2017;7:42710.
200. Hutagalung AH, Novick PJ. Role of Rab GTPases in membrane traffic and cell physiology. Physiol Rev. 2011;91(1):119-49.
201. Babbey CM, Bacallao RL, Dunn KW. Rab10 associates with primary cilia and the exocyst complex in renal epithelial cells. American Journal of Physiology-Renal Physiology. 2010;299(3):F495-F506.
202. Nanbo A. Epstein-barr virus exploits the secretory pathway to release virions. Microorganisms. 2020;8(5):729.

203. Flynn RA, Belk JA, Qi Y, Yasumoto Y, Schmitz CO, Mumbach MR, et al. Systematic discovery and functional interrogation of SARS-CoV-2 viral RNA-host protein interactions during infection. Biorxiv. 2020.
204. Acevedo-Díaz A, Morales-Cabán BM, Zayas-Santiago A, Martínez-Montemayor MM, Suárez-Arroyo IJ. SCAMP3 Regulates EGFR and Promotes Proliferation and Migration of Triple-Negative Breast Cancer Cells through the Modulation of AKT, ERK, and STAT3 Signaling Pathways. Cancers. 2022;14(11):2807.
205. Radic V. A Novel Role for SCAMP3 as an Innate Immune Factor [Doctoral dissertation]: University of California Riverside; 2016.
206. Falguieres T, Castle D, Gruenberg J. Regulation of the MVB pathway by SCAMP3. Traffic. 2012;13(1):131-42.
207. Hu L, Tang Y, Mei L, Liang M, Huang J, Wang X, et al. A new intracellular targeting motif in the cytoplasmic tail of the spike protein may act as a target to inhibit SARS-CoV-2 assembly. Antiviral Research. 2023;209:105509.
208. Pacheco AR, Lazarus JE, Sit B, Schmieder S, Lencer WI, Blondel CJ, et al. CRISPR screen reveals that EHEC's T3SS and Shiga toxin rely on shared host factors for infection. MBio. 2018;9(3):e01003-18.
209. Schimmöller F, Díaz E, Mühlbauer B, Pfeffer SR. Characterization of a 76 kDa endosomal, multispanning membrane protein that is highly conserved throughout evolution. Gene. 1998;216(2):311-8.
210. Meisen WH, Nejad ZB, Hardy M, Zhao H, Oliverio O, Wang S, et al. Pooled screens identify GPR108 and TM9SF2 as host cell factors critical for AAV transduction. Molecular Therapy-Methods & Clinical Development. 2020;17:601-11.
211. Tanaka A, Tumkosit U, Nakamura S, Motooka D, Kishishita N, Priengprom T, et al. Genome-Wide Screening Uncovers the Significance of N-Sulfation of Heparan Sulfate as a Host Cell Factor for Chikungunya Virus Infection. J Virol. 2017;91(13).
212. Nair TS, Kakaraparthi BN, Yang L, Lu L, Thomas TB, Morris AC, et al. Slc44a2 deletion alters tetraspanin and N-cadherin expression: reduced adhesion and enhanced proliferation in cultured mesenchymal lung cells. Tissue and Cell. 2021;73:101599.
213. Bayat B, Tjahjono Y, Berghöfer H, Werth S, Deckmyn H, De Meyer SF, et al. Choline transporter-like protein-2: new von Willebrand factor–binding partner involved in antibody-mediated neutrophil activation and transfusion-related acute lung injury. Arteriosclerosis, Thrombosis, and Vascular Biology. 2015;35(7):1616-22.
214. Constantinescu-Bercu A, Grassi L, Frontini M, Salles-Crawley II, Woollard K, Crawley JT. Activated αIIbβ3 on platelets mediates flow-dependent NETosis via SLC44A2. Elife. 2020;9:e53353.
215. Maróti Z, Tombácz D, Moldován N, Torma G, Jefferson VA, Csabai Z, et al. Time course profiling of host cell response to herpesvirus infection using nanopore and synthetic long-read transcriptome sequencing. Scientific reports. 2021;11(1):1-11.
216. Shoshan-Barmatz V, Mizrachi D, Keinan N. Oligomerization of the mitochondrial protein VDAC1: from structure to function and cancer therapy. Progress in molecular biology and translational science. 2013;117:303-34.
217. Shoshan-Barmatz V, Shteinfer-Kuzmine A, Verma A. VDAC1 at the intersection of cell metabolism, apoptosis, and diseases. Biomolecules. 2020;10(11):1485.
218. Shoshan-Barmatz V, Krelin Y, Shteinfer-Kuzmine A. VDAC1 functions in Ca2+ homeostasis and cell life and death in health and disease. Cell Calcium. 2018;69:81-100.
219. Csordás Gr, Renken C, Várnai P, Walter L, Weaver D, Buttle KF, et al. Structural and functional features and significance of the physical linkage between ER and mitochondria. The Journal of cell biology. 2006;174(7):915-21.
220. Feng X, Ching CB, Chen WN. EBV up-regulates cytochrome c through VDAC1 regulations and decreases the release of cytoplasmic Ca2+ in the NPC cell line. Cell biology international. 2012;36(8):733-8.

221. Zamarin D, García-Sastre A, Xiao X, Wang R, Palese P. Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. PLoS pathogens. 2005;1(1):e4.
222. Han C, Zeng X, Yao S, Gao L, Zhang L, Qi X, et al. Voltage-dependent anion channel 1 interacts with ribonucleoprotein complexes to enhance infectious bursal disease virus polymerase activity. Journal of Virology. 2017;91(16):e00584-17.
223. Chen M, Nowak DG, Narula N, Robinson B, Watrud K, Ambrico A, et al. The nuclear transport receptor Importin-11 is a tumor suppressor that maintains PTEN protein. Journal of Cell Biology. 2017;216(3):641-56.
224. Sajidah ES, Lim K, Wong RW. How SARS-CoV-2 and other viruses build an invasion route to hijack the host nucleocytoplasmic trafficking system. Cells. 2021;10(6):1424.
225. Plafker SM, Macara IG. Importin-11, a nuclear import receptor for the ubiquitin-conjugating enzyme, UbcM2. The EMBO journal. 2000;19(20):5502-13.
226. Sagar SK, Kumar M, Singh P, Sankhwar S, Dohare R. Prediction of Putative Protein Interactions between Zika Virus and Its Hosts Using Computational Techniques. Journal of Communicable Diseases (E-ISSN: 2581-351X & P-ISSN: 0019-5138). 2021;53(4):84-96.
227. Döhner K, Ramos-Nascimento A, Bialy D, Anderson F, Hickford-Martinez A, Rother F, et al. Importin α1 is required for nuclear import of herpes simplex virus proteins and capsid assembly in fibroblasts and neurons. PLoS pathogens. 2018;14(1):e1006823.
228. Kim E, Okada K, Kenniston T, Raj VS, AlHajri MM, Farag EA, et al. Immunogenicity of an adenoviral-based Middle East Respiratory Syndrome coronavirus vaccine in BALB/c mice. Vaccine. 2014;32(45):5975-82.
229. Laing ED, Sterling SL, Weir DL, Beauregard CR, Smith IL, Larsen SE, et al. Enhanced autophagy contributes to reduced viral infection in black flying fox cells. Viruses. 2019;11(3):260.
230. Huang H-C, Lai Y-J, Liao C-C, Yang W-F, Huang K-B, Lee I-J, et al. Targeting conserved N-glycosylation blocks SARS-CoV-2 variant infection in vitro. EBioMedicine. 2021;74:103712.
231. Thrower JS, Hoffman L, Rechsteiner M, Pickart CM. Recognition of the polyubiquitin proteolytic signal. The EMBO journal. 2000;19(1):94-102.
232. Hershko A, Ciechanover A. The ubiquitin system. Annual review of biochemistry. 1998;67(1):425-79.
233. Wang F, Chen X, Yu X, Lin Q. Degradation of CCNB1 mediated by APC11 through UBA52 ubiquitination promotes cell cycle progression and proliferation of non-small cell lung cancer cells. American Journal of Translational Research. 2019;11(11):7166.
234. Tang Q, Wu P, Chen H, Li G. Pleiotropic roles of the ubiquitin-proteasome system during viral propagation. Life sciences. 2018;207:350-4.
235. Kisselev AF, Goldberg AL. Proteasome inhibitors: from research tools to drug candidates. Chemistry & biology. 2001;8(8):739-58.
236. Raaben M, Posthuma CC, Verheije MH, Te Lintelo EG, Kikkert M, Drijfhout JW, et al. The ubiquitin-proteasome system plays an important role during various stages of the coronavirus infection cycle. Journal of virology. 2010;84(15):7869-79.
237. Byk LA, Iglesias NG, De Maio FA, Gebhard LG, Rossi M, Gamarnik AV. Dengue virus genome uncoating requires ubiquitination. MBio. 2016;7(3):e00804-16.
238. Tibullo D, Barbagallo I, Giallongo C, Vanella L, Conticello C, Romano A, et al. Heme oxygenase-1 nuclear translocation regulates bortezomib-induced cytotoxicity and mediates genomic instability in myeloma cells. Oncotarget. 2016;7(20):28868.
239. Gao G, Luo H. The ubiquitin–proteasome pathway in viral infections. Canadian journal of physiology and pharmacology. 2006;84(1):5-14.
240. Yu G-Y, Lai MM. The ubiquitin-proteasome system facilitates the transfer of murine coronavirus from endosome to cytoplasm during virus entry. Journal of virology. 2005;79(1):644-8.
241. Longhitano L, Tibullo D, Giallongo C, Lazzarino G, Tartaglia N, Galimberti S, et al. Proteasome inhibitors as a possible therapy for SARS-CoV-2. International Journal of Molecular Sciences. 2020;21(10):3622.

242. Mao J, O'Gorman C, Sutovsky M, Zigo M, Wells KD, Sutovsky P. Ubiquitin A-52 residue ribosomal protein fusion product 1 (Uba52) is essential for preimplantation embryo development. Biology Open. 2018;7(10):bio035717.
243. Kobayashi M, Oshima S, Maeyashiki C, Nibe Y, Otsubo K, Matsuzawa Y, et al. The ubiquitin hybrid gene UBA52 regulates ubiquitination of ribosome and sustains embryonic development. Scientific reports. 2016;6(1):1-11.
244. Wang Q, Li Q, Liu T, Chang G, Sun Z, Gao Z, et al. Host interaction analysis of PA-N155 and PA-N182 in chicken cells reveals an essential role of UBA52 for replication of H5N1 avian influenza virus. Frontiers in microbiology. 2018:936.
245. Bett JS. Proteostasis regulation by the ubiquitin system. Essays in biochemistry. 2016;60(2):143-51.
246. Mielech AM, Chen Y, Mesecar AD, Baker SC. Nidovirus papain-like proteases: multifunctional enzymes with protease, deubiquitinating and deISGylating activities. Virus research. 2014;194:184-90.
247. Bailey-Elkin BA, Knaap RC, Johnson GG, Dalebout TJ, Ninaber DK, Van Kasteren PB, 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. Journal of Biological Chemistry. 2014;289(50):34667-82.
248. Yang X, Chen X, Bian G, Tu J, Xing Y, Wang Y, et al. Proteolytic processing, deubiquitinase and interferon antagonist activities of Middle East respiratory syndrome coronavirus papain-like protease. Journal of General Virology. 2014;95(3):614-26.
249. Chou C-Y, Lai H-Y, Chen H-Y, Cheng S-C, Cheng K-W, Chou Y-W. Structural basis for catalysis and ubiquitin recognition by the severe acute respiratory syndrome coronavirus papain-like protease. Acta Crystallographica Section D: Biological Crystallography. 2014;70(2):572-81.
250. Ramirez Hernandez E, Hernández-Zimbrón LF, Martínez Zúñiga N, Leal-García JJ, Ignacio Hernández V, Ucharima-Corona LE, et al. The role of the SARS-CoV-2 S-protein glycosylation in the interaction of SARS-CoV-2/ACE2 and immunological responses. Viral immunology. 2021;34(3):165-73.
251. Cummings RD. The repertoire of glycan determinants in the human glycome. Molecular BioSystems. 2009;5(10):1087-104.
252. Ruiz-Canada C, Kelleher DJ, Gilmore R. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. Cell. 2009;136(2):272-83.
253. Watanabe Y, Bowden TA, Wilson IA, Crispin M. Exploitation of glycosylation in enveloped virus pathobiology. Biochimica et Biophysica Acta (BBA)-General Subjects. 2019;1863(10):1480-97.
254. Watanabe Y, Berndsen ZT, Raghwani J, Seabright GE, Allen JD, Pybus OG, et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nature communications. 2020;11(1):1-10.
255. Shajahan A, Pepi LE, Rouhani DS, Heiss C, Azadi P. Glycosylation of SARS-CoV-2: structural and functional insights. Analytical and Bioanalytical Chemistry. 2021;413(29):7179-93.
256. Casas-Sanchez A, Romero-Ramirez A, Hargreaves E, Ellis CC, Grajeda BI, Estevao IL, et al. Inhibition of Protein N-Glycosylation Blocks SARS-CoV-2 Infection. Mbio. 2022;13(1):e03718-21.
257. Helle F, Duverlie G, Dubuisson J. The hepatitis C virus glycan shield and evasion of the humoral immune response. Viruses. 2011;3(10):1909-32.
258. Clarke EC, Nofchissey RA, Ye C, Bradfute SB. The iminosugars celgosivir, castanospermine and UV-4 inhibit SARS-CoV-2 replication. Glycobiology. 2021;31(4):378-84.
259. Bianchi M, Benvenuto D, Giovanetti M, Angeletti S, Ciccozzi M, Pascarella S. Sars-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? BioMed Research International. 2020;2020.
260. Ha DP, Van Krieken R, Carlos AJ, Lee AS. The stress-inducible molecular chaperone GRP78 as potential therapeutic target for coronavirus infection. Journal of Infection. 2020;81(3):452-82.

261. Fung TS, Huang M, Liu DX. Coronavirus-induced ER stress response and its involvement in regulation of coronavirus–host interactions. Virus research. 2014;194:110-23.
262. Testerink N, van der Sanden MH, Houweling M, Helms JB, Vaandrager AB. Depletion of phosphatidylcholine affects endoplasmic reticulum morphology and protein traffic at the Golgi complex. Journal of lipid research. 2009;50(11):2182-92.
263. Sriburi R, Jackowski S, Mori K, Brewer JW. XBP1: a link between the unfolded protein response, lipid biosynthesis, and biogenesis of the endoplasmic reticulum. The Journal of cell biology. 2004;167(1):35-41.
264. Chan C-P, Siu K-L, Chin K-T, Yuen K-Y, Zheng B, Jin D-Y. Modulation of the unfolded protein response by the severe acute respiratory syndrome coronavirus spike protein. Journal of virology. 2006;80(18):9279-87.
265. Samali A, FitzGerald U, Deegan S, Gupta S. Methods for monitoring endoplasmic reticulum stress and the unfolded protein response. International journal of cell biology. 2010;2010.
266. Hay S, Kannourakis G. A time to kill: viral manipulation of the cell death program. Journal of General Virology. 2002;83(7):1547-64.
267. Wang R, Simoneau CR, Kulsuptrakul J, Bouhaddou M, Travisano KA, Hayashi JM, et al. Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses. 2021;184(1):106-19. e14.
268. Baggen J, Persoons L, Jansen S, Vanstreels E, Jacquemyn M, Jochmans D, et al. Identification of TMEM106B as proviral host factor for SARS-CoV-2. J BioRxiv. 2020 Sep 28:2020-09.
269. Hoffmann HH, Sanchez-Rivera FJ, Schneider WM, Luna JM, Soto-Feliciano YM, Ashbrook AW, et al. Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors. Cell Host Microbe. 2021;29(2):267-80 e5.
270. Schneider WM, Luna JM, Hoffmann H-H, Sánchez-Rivera FJ, Leal AA, Ashbrook AW, et al. Genome-scale identification of SARS-CoV-2 and pan-coronavirus host factor networks. J Cell. 2021;184(1):120-32. e14.
271. Wei J, Alfajaro MM, DeWeirdt PC, Hanna RE, Lu-Culligan WJ, Cai WL, et al. Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. J Cell. 2021;184(1):76-91. e13.
272. Daniloski Z, Jordan TX, Wessels H-H, Hoagland DA, Kasela S, Legut M, et al. Identification of required host factors for SARS-CoV-2 infection in human cells. J Cell. 2021;184(1):92-105. e16.
273. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature biotechnology. 2016;34(2):184-91.
274. The Human Protein Atlas. TM9SF2 Information, 2023 [updated January 2023. Available from: https://www.proteinatlas.org/ENSG00000125304-TM9SF2.
275. Geiger N, Kersting L, Schlegel J, Stelz L, Fähr S, Diesendorf V, et al. The Acid Ceramidase Is a SARS-CoV-2 Host Factor. Cells. 2022;11(16):2532.
276. Kokic G, Hillen HS, Tegunov D, Dienemann C, Seitz F, Schmitzova J, et al. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. Nature communications. 2021;12(1):1-7.

Appendix

Appendix A:

- **10 X SDS Running buffer:**

  TrisBase 15g.

  Glycine 72g.

  SDS 5g.

  dH_2O till 500ml.

- **6X Sample buffer:**

  7ml (0.5M) Tris-HCl Ph 6.8.

  3ml Glycerol.

  1.2g SDS.

- **Transfer buffer:**

  Trizma base 1.5g.

  Glycine 7.2g.

  H_2O 250ml.

  Methanol 100ml.

  H_2O to 500ml.

- **Resolving Gel recipes:**

	15%	
	1x	**2x**
H2O	1.7 ml	3.4 ml
40% Acrylamide Mix	1.9 ml	3.8 ml
1.5M Tris pH 8.8	1.3 ml	2.6 ml
10% SDS	50 µl	100 µl
10% APS*	50 µl	100 µl
TEMED*	3 µl	6 µl
Total volume:	**5 ml**	**10 ml**

- **Stacking Gel recipe:**

	5%	
	1x	**2x**
H2O	2.4 ml	4.8 ml
40% Acrylamide Mix	0.5 ml	1 ml
0.5M Tris pH 6.8	1. ml	2 ml
10% SDS	40 µl	80 µl
10% APS*	40 µl	80 µl
TEMED*	4 µl	8 µl
Total volume:	**4 ml**	**8 ml**

- **SDS-PAGE Destain solution**

  Add 50 ml of glacial acetic acid to 350 ml of dH2O.

  Add 100 ml of methanol and mix.

  Store at RT in sealable container.

- **Immunofluorescence blocking solution:**

  Prepare your blocking solution (5% FCS* in PBS) (5ml FCS + 45ml PBS). * Fecal bovin serum.

- **Antibiotic used in this study:**

  Hygromycin (100mg/ml) (Sigma H3274-250MG), 3 ul of Hygromycin /ml of media.

  Tetracycline (10mg/ml)

- **Immunoprecipitation Stock Buffers**

  10X Stock TBS for preparation of other buffers

  25 ml of 1M Tris/Cl.

  15 ml of 5M NaCL.

  10 ml molecular biology grade water.

- **Immunoprecipitation 10% Triton X-100**

  1 ml Dilute Triton X-100.

  9 ml molecular grade water.

- **Immunoprecipitation Protease inhibitor**

  One Tablet cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail – Roche.

  1.5 ml sterile water.

- **Equilibration buffer:**

  1 ml 10X Stock TBS buffer.

  9 ml H2O.

- **Immunoprecipitation Wash buffer**:

  1ml 10X Stock TBS buffer.

  9ml H_2O.

- **Immunoprecipitation Lysis Buffer 5 mls:**

  0.5 ml 10X Stock TBS buffer

  10 ul of 500 mM EDTA

  0.5 ml 10% Triton X-100

  3.75 ml H2O.

  0.7 ml of 7X Protease Inhibitor

5X (5mg/ml) MTT stock solution:

- 12 mM MTT stock solution (5 mg vial of MTT), 1 mL of sterile PBS.

1X(1mg/ml) MTT stock solution:

- 5x (5mg/ml) stock solution, 20 ml of sterile PBS.

10X TBE buffer:

- 108 g Tris base, 55 g Boric acid, 40 ml 0.5M EDTA (pH 8.0).
- Adjust volume to 1 L with dH_2O.
- Autoclave for 20 min at 121 °C.

1X TBE buffer:

- 50 ml 10X TBE buffer.
- Adjust volume to 500 ml with dH_2O.

6X DNA loading dye:

- 10mM Tris-HCL pH 7.6, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanol FF, 60% glycerol and 60Mmedta.

1% (w/v) agarose gel:

- 1 g agarose in 100 ml of 1X TBE.

Luria-Bertani (LB) medium:

- 10 g tryptone, 10 g NaCL, 5 g yeast extract.
- Adjust the pH of the medium to 7 with concentrated NaOH and make up to 1 L with dH_2O.
- Autoclave for 20 min at 121 °C.

LB-agar plate:

- 10 g tryptone, 10 g NaCL, 5 g yeast extract.
- Adjust the pH of the medium to 7 with concentrated NaOH and make up to 1 L with dH2O.
- Add 15 g/L agar and then autoclave for 20 min at 121 °C.
- Cool to 55 °C, add antibiotic if needed and pour into 10 cm plates.
- Let harden , then invert and store at 4 °C in the dark.

Appendix B: Plasmid sequencing

MERS-CoV E:

LOCUS 16AAECCC_pDNA5FRT_IntEnvelope_pcDNA5FrtToIntr 6090 bp DNA circular

FEATURES Location/Qualifiers

promoter 1633..1651

/label="T7_promoter"

promoter complement(5996..6024)

/label="AmpR_promoter"

terminator 2045..2272

/label="bGH_PA_terminator"

CDS complement(5094..5954)

/label="Amp(R)"

CDS 2599..3633

/label="hygroB"

rep_origin complement(4320..4939)

/label="pBR322_origin"

primer complement(3929..3947)

/label="M13_reverse_primer"

primer complement(2042..2059)

/label="BGH\rev\primer"

primer 769..789

/label="CMV_fwd_primer"

primer 3854..3873

/label="EBV_rev_primer"

primer complement(3946..3968)

/label="M13_pUC_rev_primer"

primer 869..893

/label="LNCX_primer"

polyA_site 2048..2272

/label="BGH\pA"

promoter 308..818

/label="CMV_Promoter"

CDS 2608..3633

/label="hygroB"

terminator 3768..3884

/label="SV40_PA_terminator"

CDS 2608..3576

/label="hygroB"

CDS 1666..2004

/label="pDNA5FRT_IntEnvelope"

ORIGIN

1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT CAGTACAATC TGCTCTGATG

61 CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG

121 CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC

181 TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT

241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA

301 TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC

361 CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA GGGACTTTCC

421 ATTGACGTCA ATGGGTGGAG TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT

481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT

541 ATGCCCAGTA CATGACCTTA TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA

601 TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG

661 ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC

721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG

781 GTAGGCGTGT ACGGTGGGAG GTCTATATAA GCAGAGCTCT CCCTATCAGT GATAGAGATC

841 TCCCTATCAG TGATAGAGAT CGTCGACGAG CTCGTTTAGT GAACCGTCAG ATCGCCTGGA

901 GACGCCATCC ACGCTGTTTT GACCTCCATA GAAGACACCG GGACCGATCC AGCCTCCGGA

961 CTCTAGCGTT TAAACTTAAG ACGCGTTGAT CCTGAGAACT TCAGGGTGAG TTTGGGGACC

1021 CTTGATTGTT CTTTCTTTTT CGCTATTGTA AAATTCATGT TATATGGAGG GGGCAAAGTT

1081 TTCAGGGTGT TGTTTAGAAT GGGAAGATGT CCCTTGTATC ACCATGGACC CTCATGATAA

1141 TTTTGTTTCT TTCACTTTCT ACTCTGTTGA CAACCATTGT CTCCTCTTAT TTTCTTTTCA

1201 TTTTCTGTAA CTTTTTCGTT AAACTTTAGC TTGCATTTGT AACGAATTTT TAAATTCACT

1261 TTTGTTTATT TGTCAGATTG TAAGTACTTT CTCTAATCAC TTTTTTTTCA AGGCAATCAG

1321 GGTATATTAT ATTGTACTTC AGCACAGTTT TAGAGAACAA TTGTTATAAT TAAATGATAA

1381 GGTAGAATAT TTCTGCATAT AAATTCTGGC TGGCGTGGAA ATATTCTTAT TGGTAGAAAC

1441 AACTACATCC TGGTCATCAT CCTGCCTTTC TCTTTATGGT TACAATGATA TACACTGTTT

1501 GAGATGAGGA TAAAATACTC TGAGTCCAAA CCGGGCCCCT CTGCTAACCA TGTTCATGCC

1561 TTCTTCTTTT TCCTACAGCT CCTGGGCAAC GTGCTGGTTA TTGTGCTGTC TCATCATTTT

1621 GGCAAAGAAT TGTAATACGA CTCACTATAG GGCGAATTCG GATCCGCCAC CATGCTGCCC

1681 TTCGTGCAGG AACGGATCGG CCTGTTCATC GTGAATTTCT TCATCTTCAC CGTCGTGTGC

1741 GCCATCACCC TGCTCGTGTG CATGGCCTTT CTGACCGCCA CCCGGCTGTG CGTGCAGTGC

1801 ATGACCGGCT TCAACACCCT GCTGGTGCAG CCCGCCCTGT ACCTGTACAA CACCGGCCGC

1861 AGCGTGTACG TGAAGTTCCA GGACAGCAAG CCCCCCCTGC CCCCCGATGA GTGGGTGGCA

1921 GGAAAGCTTG CAGGAGACTA CAAGGACCAC GACGGTGACT ACAAGGACCA CGACATCGAC

1981 TACAAGGACG ACGACGACAA GTGACTCGAG TCTAGAGGGC CCGTTTAAAC CCGCTGATCA

2041 GCCTCGACTG TGCCTTCTAG TTGCCAGCCA TCTGTTGTTT GCCCCTCCCC CGTGCCTTCC

2101 TTGACCCTGG AAGGTGCCAC TCCCACTGTC CTTTCCTAAT AAAATGAGGA AATTGCATCG

2161 CATTGTCTGA GTAGGTGTCA TTCTATTCTG GGGGGTGGGG TGGGGCAGGA CAGCAAGGGG

2221 GAGGATTGGG AAGACAATAG CAGGCATGCT GGGGATGCGG TGGGCTCTAT GGCTTCTGAG

2281 GCGGAAAGAA CCAGCTGGGG CTCTAGGGGG TATCCCCACG CGCCCTGTAG CGGCGCATTA

2341 AGCGCGGCGG GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG

2401 CCCGCTCCTT TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA

2461 GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC

2521 AAAAAACTTG ATTAGGGTGA TGGTTCACGT ACCTAGAAGT TCCTATTCCG AAGTTCCTAT

2581 TCTCTAGAAA GTATAGGAAC TTCCTTGGCC AAAAAGCCTG AACTCACCGC GACGTCTGTC

2641 GAGAAGTTTC TGATCGAAAA GTTCGACAGC GTCTCCGACC TGATGCAGCT CTCGGAGGGC

2701 GAAGAATCTC GTGCTTTCAG CTTCGATGTA GGAGGGCGTG GATATGTCCT GCGGGTAAAT

2761 AGCTGCGCCG ATGGTTTCTA CAAAGATCGT TATGTTTATC GGCACTTTGC ATCGGCCGCG

2821 CTCCCGATTC CGGAAGTGCT TGACATTGGG GAATTCAGCG AGAGCCTGAC CTATTGCATC

2881 TCCCGCCGTG CACAGGGTGT CACGTTGCAA GACCTGCCTG AAACCGAACT GCCCGCTGTT

2941 CTGCAGCCGG TCGCGGAGGC CATGGATGCG ATCGCTGCGG CCGATCTTAG CCAGACGAGC

3001 GGGTTCGGCC CATTCGGACC GCAAGGAATC GGTCAATACA CTACATGGCG TGATTTCATA

3061 TGCGCGATTG CTGATCCCCA TGTGTATCAC TGGCAAACTG TGATGGACGA CACCGTCAGT

3121 GCGTCCGTCG CGCAGGCTCT CGATGAGCTG ATGCTTTGGG CCGAGGACTG CCCCGAAGTC

3181 CGGCACCTCG TGCACGCGGA TTTCGGCTCC AACAATGTCC TGACGGACAA TGGCCGCATA

3241 ACAGCGGTCA TTGACTGGAG CGAGGCGATG TTCGGGGATT CCCAATACGA GGTCGCCAAC

3301 ATCTTCTTCT GGAGGCCGTG GTTGGCTTGT ATGGAGCAGC AGACGCGCTA CTTCGAGCGG

3361 AGGCATCCGG AGCTTGCAGG ATCGCCGCGG CTCCGGGCGT ATATGCTCCG CATTGGTCTT

3421 GACCAACTCT ATCAGAGCTT GGTTGACGGC AATTTCGATG ATGCAGCTTG GGCGCAGGGT

3481 CGATGCGACG CAATCGTCCG ATCCGGAGCC GGGACTGTCG GGCGTACACA AATCGCCCGC

3541 AGAAGCGCGG CCGTCTGGAC CGATGGCTGT GTAGAAGTAC TCGCCGATAG TGGAAACCGA

3601 CGCCCCAGCA CTCGTCCGAG GGCAAAGGAA TAGCACGTAC TACGAGATTT CGATTCCACC

3661 GCCGCCTTCT ATGAAAGGTT GGGCTTCGGA ATCGTTTTCC GGGACGCCGG CTGGATGATC

3721 CTCCAGCGCG GGGATCTCAT GCTGGAGTTC TTCGCCCACC CCAACTTGTT TATTGCAGCT

3781 TATAATGGTT ACAAATAAAG CAATAGCATC ACAAATTTCA CAAATAAAGC ATTTTTTTCA

3841 CTGCATTCTA GTTGTGGTTT GTCCAAACTC ATCAATGTAT CTTATCATGT CTGTATACCG

3901 TCGACCTCTA GCTAGAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT

3961 TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT

4021 GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG

4081 GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG

4141 CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTCGGCTG

4201 CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT

4261 AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC

4321 GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATCGACGC

4381 TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA

4441 AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT

4501 CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG

4561 TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC

4621 GCCTTATCCG GTAACTATCG TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG

4681 GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC

4741 TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGAACAG TATTTGGTAT CTGCGCTCTG

4801 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC

4861 GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT

4921 CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT

4981 TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA

5041 AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA

5101 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC

5161 TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT

5221 GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA

5281 GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT

5341 AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT

5401 GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC

5461 GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC

5521 TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT

5581 ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT

5641 GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC

5701 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT

5761 GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG

5821 ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT

5881 GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA

5941 TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT

6001 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC

6061 ACATTTCCCC GAAAAGTGCC ACCTGACGTC

//

MERS-CoV- M:

LOCUS 16AAEBTC_MembFL_pMK-RQ 3040 bp DNA circular 19-MAY-2008

SOURCE

ORGANISM

FEATURES Location/Qualifiers

rep_origin complement(1300..1967)

/label="Col\E1\origin"

/vntifkey="33"

CDS complement(2115..2912)

/label="KanR"

/vntifkey="4"

CDS 370..1131

/label="MembFL"

ORIGIN

1 CTAAATTGTA AGCGTTAATA TTTTGTTAAA ATTCGCGTTA AATTTTTGTT AAATCAGCTC

61 ATTTTTTAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT AAATCAAAAG AATAGACCGA

121 GATAGGGTTG AGTGGCCGCT ACAGGGCGCT CCCATTCGCC ATTCAGGCTG CGCAACTGTT

181 GGGAAGGGCG TTTCGGTGCG GGCCTCTTCG CTATTACGCC AGCTGGCGAA AGGGGGATGT

241 GCTGCAAGGC GATTAAGTTG GGTAACGCCA GGGTTTTCCC AGTCACGACG TTGTAAAACG

301 ACGGCCAGTG AGCGCGACGT AATACGACTC ACTATAGGGC GAATTGAAGG AAGGCCGTCA

361 AGGCCGCATG GATCCGCCAC CATGAGCAAC ATGACCCAGC TGACCGAGGC CCAGATCATT

421 GCCATCATCA AGGACTGGAA CTTCGCCTGG TCCCTGATCT TCCTGCTGAT CACCATCGTG

481 CTGCAGTACG GCTACCCCAG CCGGTCCATG ACCGTGTACG TGTTCAAGAT GTTCGTCCTG

541 TGGCTGCTGT GGCCCAGCTC CATGGCCCTG AGCATCTTCA GCGCCGTGTA CCCCATCGAC

601 CTGGCCAGCC AGATCATCAG CGGCATCGTG GCCGCCGTGT CCGCCATGAT GTGGATCAGC

661 TACTTCGTGC AGAGCATCCG GCTGTTCATG CGGACCGGCA GCTGGTGGTC CTTCAACCCC

721 GAGACAAACT GCCTGCTGAA CGTGCCCTTC GGCGGCACCA CAGTCGTGCG GCCCCTGGTG

781 GAAGATAGCA CCAGCGTGAC CGCCGTGGTC ACCAACGGCC ACCTGAAGAT GGCCGGCATG

841 CACTTCGGCG CCTGCGACTA CGACCGGCTG CCCAACGAAG TGACCGTGGC CAAGCCCAAC

901 GTGCTGATCG CCCTGAAAAT GGTCAAGCGG CAGAGCTACG GCACCAACAG CGGCGTGGCC

961 ATCTACCACC GGTACAAGGC CGGCAACTAC AGAAGCCCCC CCATCACCGC CGACATCGAG

1021 CTGGCCCTGC TGAGAGCCGC AGGAAAGCTT GCAGGAGACT ACAAGGACCA CGACGGTGAC

1081 TACAAGGACC ACGACATCGA CTACAAGGAC GACGACGACA AGTGACTCGA GCTGGGCCTC

1141 ATGGGCCTTC CTTTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC AGCTGCATTA

1201 ACATGGTCAT AGCTGTTTCC TTGCGTATTG GGCGCTCTCC GCTTCCTCGC TCACTGACTC

1261 GCTGCGCTCG GTCGTTCGGG TAAAGCCTGG GGTGCCTAAT GAGCAAAAGG CCAGCAAAAG

1321 GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC

1381 GAGCATCACA AAAATCGACG CTCAAGTCAG AGGTGGCGAA ACCCGACAGG ACTATAAAGA

1441 TACCAGGCGT TTCCCCCTGG AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT

1501 ACCGGATACC TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA TAGCTCACGC

1561 TGTAGGTATC TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC

1621 CCCGTTCAGC CCGACCGCTG CGCCTTATCC GGTAACTATC GTCTTGAGTC CAACCCGGTA

1681 AGACACGACT TATCGCCACT GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT

1741 GTAGGCGGTG CTACAGAGTT CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGAACA

1801 GTATTTGGTA TCTGCGCTCT GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT

1861 TGATCCGGCA AACAAACCAC CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT

1921 ACGCGCAGAA AAAAAGGATC TCAAGAAGAT CCTTTGATCT TTTCTACGGG GTCTGACGCT

1981 CAGTGGAACG AAAACTCACG TTAAGGGATT TTGGTCATGA GATTATCAAA AAGGATCTTC

2041 ACCTAGATCC TTTTAAATTA AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA

2101 ACTTGGTCTG ACAGTTATTA GAAAAATTCA TCCAGCAGAC GATAAAACGC AATACGCTGG

2161 CTATCCGGTG CCGCAATGCC ATACAGCACC AGAAAACGAT CCGCCCATTC GCCGCCCAGT

2221 TCTTCCGCAA TATCACGGGT GGCCAGCGCA ATATCCTGAT AACGATCCGC CACGCCCAGA

2281 CGGCCGCAAT CAATAAAGCC GCTAAAACGG CCATTTTCCA CCATAATGTT CGGCAGGCAC

2341 GCATCACCAT GGGTCACCAC CAGATCTTCG CCATCCGGCA TGCTCGCTTT CAGACGCGCA

2401 AACAGCTCTG CCGGTGCCAG GCCCTGATGT TCTTCATCCA GATCATCCTG ATCCACCAGG

2461 CCCGCTTCCA TACGGGTACG CGCACGTTCA ATACGATGTT TCGCCTGATG ATCAAACGGA

2521 CAGGTCGCCG GGTCCAGGGT ATGCAGACGA CGCATGGCAT CCGCCATAAT GCTCACTTTT

2581 TCTGCCGGCG CCAGATGGCT AGACAGCAGA TCCTGACCCG GCACTTCGCC CAGCAGCAGC

2641 CAATCACGGC CCGCTTCGGT CACCACATCC AGCACCGCCG CACACGGAAC ACCGGTGGTG

2701 GCCAGCCAGC TCAGACGCGC CGCTTCATCC TGCAGCTCGT TCAGCGCACC GCTCAGATCG

2761 GTTTTCACAA ACAGCACCGG ACGACCCTGC GCGCTCAGAC GAAACACCGC CGCATCAGAG

2821 CAGCCAATGG TCTGCTGCGC CCAATCATAG CCAAACAGAC GTTCCACCCA CGCTGCCGGG

2881 CTACCCGCAT GCAGGCCATC CTGTTCAATC ATACTCTTCC TTTTTCAATA TTATTGAAGC

2941 ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA

3001 CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC

//

SARS-CoV-2 E:

LOCUS 20AD4GEP_SARS-CoV-2_E_pcDNA3.3-TOPO 5756 bp DNA circular 31-JUL-2014

```
FEATURES             Location/Qualifiers
     promoter        1321..1998
                     /label="CMV\promoter"
                     /vntifkey="29"
     primer_bind     1858..1878
                     /label="CMV\forward\primer\binding\site"
                     /vntifkey="28"
     primer_bind     2410..2428
                     /label="TK\pA\reverse\primer\binding\site"
                     /vntifkey="28"
     CDS             2403..2674
                     /label="TK\pA\signal"
                     /vntifkey="87"
     rep_origin      2710..3138
                     /label="fi\ori"
                     /vntifkey="33"
     promoter        3143..3512
                     /label="SV40\early\promoter"
                     /vntifkey="29"
     CDS             3548..4342
                     /label="Neo-R"
                     /vntifkey="22"
     CDS             4518..4648
                     /label="SV40\pA\signal"
                     /vntifkey="87"
     rep_origin      complement(5031..5704)
                     /label="pUC\ori"
                     /vntifkey="33"
     CDS             complement(93..953)
                     /label="Amp-R"
                     /vntifkey="22"
```

```
     promoter        complement(954..1052)
                     /label="bla\promoter"
                     /vntifkey="30"
     CDS             2016..2364
                     /label="SARS-CoV-2_E"
ORIGIN
        1 TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC
       61 TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT
      121 ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA
      181 ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA
      241 CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA
      301 AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG GGAAGCTAGA
      361 GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG
      421 GTGTCACGCT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA
      481 GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT
      541 GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT
      601 CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA
      661 TTCTGAGAAT AGTGTATGCG GCGACCGAGT TGCTCTTGCC CGGCGTCAAT ACGGGATAAT
      721 ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA
      781 AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC
      841 AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG GGTGAGCAAA AACAGGAAGG
      901 CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC
      961 CTTTTTCAAT ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT
     1021 GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG AAAAGTGCCA
     1081 CCTGACGTCG ACGGATCGGG AGATCTCCCG ATCCCCTATG GTGCACTCTC AGTACAATCT
     1141 GCTCTGATGC CGCATAGTTA AGCCAGTATC TGCTCCCTGC TTGTGTGTTG GAGGTCGCTG
     1201 AGTAGTGCGC GAGCAAAATT TAAGCTACAA CAAGGCAAGG CTTGACCGAC AATTGCATGA
     1261 AGAATCTGCT TAGGGTTAGG CGTTTTGCGC TGCTTCGCGA TGTACGGGCC AGATATACGC
     1321 GTTGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT TACGGGGTCA TTAGTTCATA
     1381 GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA TGGCCCGCCT GGCTGACCGC
     1441 CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG
```

1501 GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC

1561 ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT AAATGGCCCG

1621 CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG

1681 TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA GTACATCAAT GGGCGTGGAT

1741 AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT

1801 TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC

1861 AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT TTAGTGAACC

1921 GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT CCATAGAAGA CACCGGGACC

1981 GATCCAGCCT CCGGACTCTA GAGGATCGAA CCCTTGAATT CGGATCCGCC ACCATGTACA

2041 GCTTCGTGTC CGAGGAAACC GGCACACTGA TCGTGAACAG CGTGCTGCTG TTCCTGGCCT

2101 TCGTGGTGTT TCTGCTGGTC ACCCTGGCCA TCCTGACAGC CCTGAGACTG TGCGCCTACT

2161 GCTGCAACAT CGTGAACGTG TCCCTGGTCA AGCCCAGCTT CTACGTGTAC AGCAGAGTGA

2221 AGAACCTGAA CAGCTCCAGA GTGCCCGATC TGCTGGTGGC CGGAAAACTG GCCGGCGACT

2281 ACAAAGACCA CGACGGCGAT TACAAGGATC ACGATATCGA CTACAAGGAC GATGACGACA

2341 AGTGACTCGA GTCTAGAGGG CCCGAAGGGT TCGATCCCTA CCGGTTAGTA ATGAGTTTAA

2401 ACGGGGGAGG CTAACTGAAA CACGGAAGGA GACAATACCG GAAGGAACCC GCGCTATGAC

2461 GGCAATAAAA AGACAGAATA AAACGCACGG GTGTTGGGTC GTTTGTTCAT AAACGCGGGG

2521 TTCGGTCCCA GGGCTGGCAC TCTGTCGATA CCCCACCGAG ACCCCATTGG GGCCAATACG

2581 CCCGCGTTTC TTCCTTTTCC CCACCCCACC CCCCAAGTTC GGGTGAAGGC CCAGGGCTCG

2641 CAGCCAACGT CGGGGCGGCA GGCCCTGCCA TAGCAGATCT GCGCAGCTGG GGCTCTAGGG

2701 GGTATCCCCA CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA

2761 GCGTGACCGC TACACTTGCC AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT

2821 TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC AAGCTCTAAA TCGGGGCATC CCTTTAGGGT

2881 TCCGATTTAG TGCTTTACGG CACCTCGACC CCAAAAAACT TGATTAGGGT GATGGTTCAC

2941 GTAGTGGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT

3001 TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA CCCTATCTCG GTCTATTCTT

3061 TTGATTTATA AGGGATTTTG CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC

3121 AAAAATTTAA CGCGAATTAA TTCTGTGGAA TGTGTGTCAG TTAGGGTGTG GAAAGTCCCC

3181 AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC AATTAGTCAG CAACCAGGTG

3241 TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC

3301 AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC CTAACTCCGC CCAGTTCCGC

3361 CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT GCAGAGGCCG AGGCCGCCTC

3421 TGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA

3481 AAAGCTCCCG GGAGCTTGTA TATCCATTTT CGGATCTGAT CAAGAGACAG GATGAGGATC

3541 GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT GGGTGGAGAG

3601 GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG CCGTGTTCCG

3661 GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG GTGCCCTGAA

3721 TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG TTCCTTGCGC

3781 AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG GCGAAGTGCC

3841 GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA TCATGGCTGA

3901 TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC ACCAAGCGAA

3961 ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC AGGATGATCT

4021 GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA AGGCGCGCAT

4081 GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA ATATCATGGT

4141 GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG CGGACCGCTA

4201 TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG AATGGGCTGA

4261 CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG CCTTCTATCG

4321 CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG CGAAATGACC GACCAAGCGA

4381 CGCCCAACCT GCCATCACGA GATTTCGATT CCACCGCCGC CTTCTATGAA AGGTTGGGCT

4441 TCGGAATCGT TTTCCGGGAC GCCGGCTGGA TGATCCTCCA GCGCGGGGAT CTCATGCTGG

4501 AGTTCTTCGC CCACCCCAAC TTGTTTATTG CAGCTTATAA TGGTTACAAA TAAAGCAATA

4561 GCATCACAAA TTTCACAAAT AAAGCATTTT TTTCACTGCA TTCTAGTTGT GGTTTGTCCA

4621 AACTCATCAA TGTATCTTAT CATGTCTGTA TACCGTCGAC CTCTAGCTAG AGCTTGGCGT

4681 AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT CCACACAACA

4741 TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA ATGAGTGAGC TAACTCACAT

4801 TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT

4861 AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT

4921 CGCTCACTGA CTCGCTGCGC TCGGTCGTTC GGCTGCGGCG AGCGGTATCA GCTCACTCAA

4981 AGGCGGTAAT ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA

5041 AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT TTCCATAGGC

5101 TCCGCCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG CGAAACCCGA

5161 CAGGACTATA AAGATACCAG GCGTTTCCCC CTGGAAGCTC CCTCGTGCGC TCTCCTGTTC

5221 CGACCCTGCC GCTTACCGGA TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT

5281 CTCAATGCTC ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT

5341 GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG

5401 AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT AACAGGATTA

5461 GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA GTGGTGGCCT AACTACGGCT

5521 ACACTAGAAG GACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA

5581 GAGTTGGTAG CTCTTGATCC GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT

5641 GCAAGCAGCA GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG ATCTTTTCTA

5701 CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC ATGAGA

//

SARS-CoV-2 M:

LOCUS 20ABFFVP_M_gene_pcDNA3.3-TOPO 6197 bp DNA circular 31-JUL-2014

FEATURES Location/Qualifiers

promoter 1321..1998

/label="CMV\promoter"

/vntifkey="29"

primer_bind 1858..1878

/label="CMV\forward\primer\binding\site"

/vntifkey="28"

primer_bind 2851..2869

/label="TK\pA\reverse\primer\binding\site"

/vntifkey="28"

CDS 2844..3115

/label="TK\pA\signal"

/vntifkey="87"

rep_origin 3151..3579

/label="fi\ori"

/vntifkey="33"

promoter 3584..3953

/label="SV40\early\promoter"

/vntifkey="29"

CDS 3989..4783

/label="Neo-R"

/vntifkey="22"

CDS 4959..5089

/label="SV40\pA\signal"

/vntifkey="87"

rep_origin complement(5472..6145)

/label="pUC\ori"

/vntifkey="33"

CDS complement(93..953)

/label="Amp-R"

/vntifkey="22"

promoter complement(954..1052)

/label="bla\promoter"

/vntifkey="30"

CDS 2016..2805

/label="M_gene"

ORIGIN

1 TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC

61 TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT

121 ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA

181 ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA

241 CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA

301 AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG GGAAGCTAGA

361 GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG

421 GTGTCACGCT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA

481 GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT

541 GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT

601 CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA

661 TTCTGAGAAT AGTGTATGCG GCGACCGAGT TGCTCTTGCC CGGCGTCAAT ACGGGATAAT

721 ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA

781 AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC

841 AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG GGTGAGCAAA AACAGGAAGG

901 CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC

961 CTTTTTCAAT ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT

1021 GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG AAAAGTGCCA

1081 CCTGACGTCG ACGGATCGGG AGATCTCCCG ATCCCCTATG GTGCACTCTC AGTACAATCT

1141 GCTCTGATGC CGCATAGTTA AGCCAGTATC TGCTCCCTGC TTGTGTGTTG GAGGTCGCTG

1201 AGTAGTGCGC GAGCAAAATT TAAGCTACAA CAAGGCAAGG CTTGACCGAC AATTGCATGA

1261 AGAATCTGCT TAGGGTTAGG CGTTTTGCGC TGCTTCGCGA TGTACGGGCC AGATATACGC

1321 GTTGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT TACGGGGTCA TTAGTTCATA

1381 GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA TGGCCCGCCT GGCTGACCGC

1441 CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG

1501 GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC

1561 ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT AAATGGCCCG

1621 CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG

1681 TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA GTACATCAAT GGGCGTGGAT

1741 AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT

1801 TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC

1861 AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT TTAGTGAACC

1921 GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT CCATAGAAGA CACCGGGACC

1981 GATCCAGCCT CCGGACTCTA GAGGATCGAA CCCTTGAATT CGGATCCGCC ACCATGGCCG

2041 ATAGCAACGG CACAATCACC GTGGAAGAAC TGAAGAAACT GCTGGAACAG TGGAACCTCG

2101 TGATCGGCTT CCTGTTCCTG ACCTGGATCT GCCTGCTGCA GTTCGCCTAC GCCAACCGGA

2161 ACAGATTCCT GTATATTATC AAGCTGATCT TCCTGTGGCT GCTGTGGCCC GTGACACTGG

2221 CCTGTTTTGT GCTGGCCGCC GTGTACCGGA TCAACTGGAT CACAGGCGGA ATCGCCATTG

2281 CCATGGCCTG TCTCGTTGGC CTGATGTGGC TGAGCTACTT TATCGCCAGC TTCCGGCTGT

2341 TCGCCCGGAC CAGATCCATG TGGTCCTTCA ATCCCGAGAC AAACATCCTG CTGAACGTGC

2401 CCCTGCACGG CACCATCCTT ACAAGACCTC TGCTGGAAAG CGAGCTGGTC ATCGGAGCCG

2461 TGATCCTGAG AGGCCACCTG AGAATTGCCG GACACCACCT GGGCAGATGC GACATCAAGG

2521 ACCTGCCTAA AGAAATCACA GTGGCCACCA GCAGAACCCT GTCCTACTAT AAGCTGGGCG

2581 CCAGCCAGAG AGTGGCCGGC GATTCTGGAT TTGCCGCCTA CAGCAGATAC CGGATCGGCA

2641 ACTACAAGCT GAACACCGAC CACAGCTCCA GCAGCGACAA TATCGCACTG CTGGTGCAGG

2701 CCGGAAAACT GGCCGGGGAT TACAAGGATC ACGACGGCGA CTATAAGGAC CACGACATTG

2761 ACTACAAGGA CGACGATGAC AAGTGACTCG AGTCTAGAGG GCCCGAAGGG TTCGATCCCT

2821 ACCGGTTAGT AATGAGTTTA AACGGGGGAG GCTAACTGAA ACACGGAAGG AGACAATACC

2881 GGAAGGAACC CGCGCTATGA CGGCAATAAA AAGACAGAAT AAAACGCACG GGTGTTGGGT

2941 CGTTTGTTCA TAAACGCGGG GTTCGGTCCC AGGGCTGGCA CTCTGTCGAT ACCCCACCGA

3001 GACCCCATTG GGGCCAATAC GCCCGCGTTT CTTCCTTTTC CCCACCCCAC CCCCCAAGTT

3061 CGGGTGAAGG CCCAGGGCTC GCAGCCAACG TCGGGGCGGC AGGCCCTGCC ATAGCAGATC

3121 TGCGCAGCTG GGGCTCTAGG GGGTATCCCC ACGCGCCCTG TAGCGGCGCA TTAAGCGCGG

3181 CGGGTGTGGT GGTTACGCGC AGCGTGACCG CTACACTTGC CAGCGCCCTA GCGCCCGCTC

3241 CTTTCGCTTT CTTCCCTTCC TTTCTCGCCA CGTTCGCCGG CTTTCCCCGT CAAGCTCTAA

3301 ATCGGGGCAT CCCTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAAC

3361 TTGATTAGGG TGATGGTTCA CGTAGTGGGC CATCGCCCTG ATAGACGGTT TTTCGCCCTT

3421 TGACGTTGGA GTCCACGTTC TTTAATAGTG GACTCTTGTT CCAAACTGGA ACAACACTCA

3481 ACCCTATCTC GGTCTATTCT TTTGATTTAT AAGGGATTTT GCCGATTTCG GCCTATTGGT

3541 TAAAAAATGA GCTGATTTAA CAAAAATTTA ACGCGAATTA ATTCTGTGGA ATGTGTGTCA

3601 GTTAGGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT

3661 CAATTAGTCA GCAACCAGGT GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA

3721 AAGCATGCAT CTCAATTAGT CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC

3781 CCTAACTCCG CCCAGTTCCG CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA

3841 TGCAGAGGCC GAGGCCGCCT CTGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT

3901 TGGAGGCCTA GGCTTTTGCA AAAAGCTCCC GGGAGCTTGT ATATCCATTT TCGGATCTGA

3961 TCAAGAGACA GGATGAGGAT CGTTTCGCAT GATTGAACAA GATGGATTGC ACGCAGGTTC

4021 TCCGGCCGCT TGGGTGGAGA GGCTATTCGG CTATGACTGG GCACAACAGA CAATCGGCTG

4081 CTCTGATGCC GCCGTGTTCC GGCTGTCAGC GCAGGGGCGC CCGGTTCTTT TTGTCAAGAC

4141 CGACCTGTCC GGTGCCCTGA ATGAACTGCA GGACGAGGCA GCGCGGCTAT CGTGGCTGGC

4201 CACGACGGGC GTTCCTTGCG CAGCTGTGCT CGACGTTGTC ACTGAAGCGG GAAGGGACTG

4261 GCTGCTATTG GGCGAAGTGC CGGGGCAGGA TCTCCTGTCA TCTCACCTTG CTCCTGCCGA

4321 GAAAGTATCC ATCATGGCTG ATGCAATGCG GCGGCTGCAT ACGCTTGATC CGGCTACCTG

4381 CCCATTCGAC CACCAAGCGA AACATCGCAT CGAGCGAGCA CGTACTCGGA TGGAAGCCGG

4441 TCTTGTCGAT CAGGATGATC TGGACGAAGA GCATCAGGGG CTCGCGCCAG CCGAACTGTT

4501 CGCCAGGCTC AAGGCGCGCA TGCCCGACGG CGAGGATCTC GTCGTGACCC ATGGCGATGC

4561 CTGCTTGCCG AATATCATGG TGGAAAATGG CCGCTTTTCT GGATTCATCG ACTGTGGCCG

4621 GCTGGGTGTG GCGGACCGCT ATCAGGACAT AGCGTTGGCT ACCCGTGATA TTGCTGAAGA

4681 GCTTGGCGGC GAATGGGCTG ACCGCTTCCT CGTGCTTTAC GGTATCGCCG CTCCCGATTC

4741 GCAGCGCATC GCCTTCTATC GCCTTCTTGA CGAGTTCTTC TGAGCGGGAC TCTGGGGTTC

4801 GCGAAATGAC CGACCAAGCG ACGCCCAACC TGCCATCACG AGATTTCGAT TCCACCGCCG

4861 CCTTCTATGA AAGGTTGGGC TTCGGAATCG TTTTCCGGGA CGCCGGCTGG ATGATCCTCC

4921 AGCGCGGGGA TCTCATGCTG GAGTTCTTCG CCCACCCCAA CTTGTTTATT GCAGCTTATA

4981 ATGGTTACAA ATAAAGCAAT AGCATCACAA ATTTCACAAA TAAAGCATTT TTTTCACTGC

5041 ATTCTAGTTG TGGTTTGTCC AAACTCATCA ATGTATCTTA TCATGTCTGT ATACCGTCGA

5101 CCTCTAGCTA GAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC

5161 CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT

5221 AATGAGTGAG CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA

5281 ACCTGTCGTG CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA

5341 TTGGGCGCTC TTCCGCTTCC TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC

5401 GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG

5461 CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT

5521 TGCTGGCGTT TTTCCATAGG CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA

5581 GTCAGAGGTG GCGAAACCCG ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT

5641 CCCTCGTGCG CTCTCCTGTT CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC

5701 CTTCGGGAAG CGTGGCGCTT TCTCAATGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG

5761 TCGTTCGCTC CAAGCTGGGC TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT

5821 TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG

5881 CAGCCACTGG TAACAGGATT AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA

5941 AGTGGTGGCC TAACTACGGC TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA

6001 AGCCAGTTAC CTTCGGAAAA AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG

6061 GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC AGATTACGCG CAGAAAAAAA GGATCTCAAG

6121 AAGATCCTTT GATCTTTTCT ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG

6181 GGATTTTGGT CATGAGA

//

Appendix C: Mr Lee's proteomic data that used to select the eleven cellular protein interactors for MERS-CoV E and M proteins to be validated.

Cellular Proteins that increased >1.5-fold enrichment in MERS- CoV E cells compared to HEK293-Flp (control cells).

Uniprot All	Gene ID	Description
Q9BTV4	TMEM43	Transmembrane protein 43 OS=Homo sapiens GN=TMEM43 PE=1 SV=1
Q9HD45	TM9SF2	Dinucleotide oxidase disulfide thiol exchanger 3 superfamily member 2 OS=Homo sapiens GN=TM9SF2 PE=2 SV=1
Q969M3	YIPF5	Protein YIPF OS=Homo sapiens PE=2 SV=1
Q96G23	CERS2	cDNA FLJ75329, highly similar to Homo sapiens LAG1 longevity assurance homolog 2 (S. cerevisiae), transcript variant 2, mRNA OS=Homo sapiens PE=2 SV=1
Q8IWA5	SLC44A2	Solute carrier family 44, member 2, isoform CRA_a OS=Homo sapiens GN=SLC44A2 PE=4 SV=1
P21796	VDAC1	Voltage-dependent anion-selective channel protein 1 OS=Homo sapiens GN=VDAC1 PE=1 SV=2
A0A024RDQ9	SLC7A1	Solute carrier family 7 (Cationic amino acid transporter, y+ system), member 1, isoform CRA_a OS=Homo sapiens GN=SLC7A1 PE=4 SV=1
Q9NQC3	RTN4	Reticulon-4 OS=Homo sapiens GN=RTN4 PE=1 SV=2
Q96KA5	CLPTM1L	Cleft lip and palate transmembrane protein 1-like protein OS=Homo sapiens GN=CLPTM1L PE=1 SV=1
O15260	SURF4	Surfeit locus protein 4 OS=Homo sapiens GN=SURF4 PE=1 SV=3
O14828	SCAMP3	Secretory carrier-associated membrane protein 3 OS=Homo sapiens GN=SCAMP3 PE=1 SV=3
P62244	RPS15A	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=2
Q99442	SEC62	Translocation protein SEC62 OS=Homo sapiens GN=SEC62 PE=1 SV=1
O95197	RTN3	Reticulon OS=Homo sapiens PE=2 SV=1
Q8WY22	BRI3BP	BRI3-binding protein OS=Homo sapiens GN=BRI3BP PE=1 SV=1
O75947	ATP5H	ATP synthase subunit d, mitochondrial OS=Homo sapiens GN=ATP5H PE=1 SV=3
P53985	SLC16A1	cDNA FLJ53399, highly similar to Monocarboxylate transporter 1 OS=Homo sapiens PE=2 SV=1
P24539	ATP5PB	cDNA FLJ78635, highly similar to Homo sapiens ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b, isoform 1 (ATP5F1), transcript variant 1, mRNA OS=Homo sapiens PE=2 SV=1
Q7Z2K6	ERMP1	Endoplasmic reticulum metallopeptidase 1 OS=Homo sapiens GN=ERMP1 PE=1 SV=2
P45880	VDAC2	Voltage-dependent anion channel 2, isoform CRA_a OS=Homo sapiens GN=VDAC2 PE=4 SV=1
P56381	ATP5E	ATP synthase subunit epsilon, mitochondrial OS=Homo sapiens GN=ATP5E PE=1 SV=2
Q01650	SLC7A5	Large neutral amino acids transporter small subunit 1 OS=Homo sapiens GN=SLC7A5 PE=1 SV=2
P48047	ATP5O	ATP synthase subunit O, mitochondrial OS=Homo sapiens GN=ATP5O PE=1 SV=1

Q8TCJ2	STT3B	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B OS=Homo sapiens GN=STT3B PE=1 SV=1
Q8TCT9	HM13	Minor histocompatibility antigen H13 OS=Homo sapiens GN=HM13 PE=1 SV=1
Q15043	SLC39A14	Zinc transporter ZIP14 OS=Homo sapiens GN=SLC39A14 PE=1 SV=3
P62979	RPS27A	Ubiquitin-40S ribosomal protein S27a OS=Homo sapiens GN=RPS27A PE=1 SV=2
Q6UWP7	LCLAT1	Lysocardiolipin acyltransferase 1 OS=Homo sapiens GN=LCLAT1 PE=1 SV=1
O76024	WFS1	Wolfram syndrome 1 isoform 1 (Fragment) OS=Homo sapiens GN=WFS1 PE=2 SV=1
A0A024QYS2	TM9SF3	Dinucleotide oxidase disulfide thiol exchanger 3 superfamily member 3 OS=Homo sapiens PE=2 SV=1
V9HW26	HEL-S-123m	ATP synthase subunit alpha OS=Homo sapiens GN=HEL-S-123m PE=2 SV=1
P00387	CYB5R3	NADH-cytochrome b5 reductase 3 OS=Homo sapiens GN=CYB5R3 PE=1 SV=3
P36542	ATP5C1	ATP synthase subunit gamma, mitochondrial OS=Homo sapiens GN=ATP5C1 PE=1 SV=1
O95674	CDS2	Phosphatidate cytidylyltransferase OS=Homo sapiens PE=2 SV=1
V9HW31	HEL-S-271	ATP synthase subunit beta OS=Homo sapiens GN=HEL-S-271 PE=1 SV=1
Q9UI26	IPO11	Importin-11 OS=Homo sapiens GN=IPO11 PE=1 SV=1
P46977	STT3A	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A OS=Homo sapiens GN=STT3A PE=1 SV=2
Q14165	MLEC	Malectin OS=Homo sapiens GN=MLEC PE=1 SV=1
P39656	DDOST	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit OS=Homo sapiens GN=DDOST PE=3 SV=1
P04843	RPN1	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Homo sapiens GN=RPN1 PE=1 SV=1
P53985	SLC16A1	Monocarboxylate transporter 1 OS=Homo sapiens GN=SLC16A1 PE=1 SV=3
Q96AA3	RFT1	Protein RFT1 homolog OS=Homo sapiens GN=RFT1 PE=1 SV=1
Q9NTJ5	SACM1L	Phosphatidylinositide phosphatase SAC1 OS=Homo sapiens GN=SACM1L PE=1 SV=2
Q7KZN9	COX15	Cytochrome c oxidase assembly protein COX15 homolog OS=Homo sapiens GN=COX15 PE=1 SV=1
P30049	ATP5D	ATP synthase subunit delta, mitochondrial OS=Homo sapiens GN=ATP5D PE=1 SV=2
A0A024R5F7	DHCR7	7-dehydrocholesterol reductase isoform 1 OS=Homo sapiens GN=DHCR7 PE=2 SV=1
A0A024RAM0	TNPO1	Transportin 1, isoform CRA_a OS=Homo sapiens GN=TNPO1 PE=4 SV=1

P57088	SHINC3	SHINC3 OS=Homo sapiens GN=SHINC3 PE=2 SV=1
Q15758	SLC1A5	Amino acid transporter (Fragment) OS=Homo sapiens PE=2 SV=1
J3KPF3	SLC3A2	4F2 cell-surface antigen heavy chain OS=Homo sapiens GN=SLC3A2 PE=1 SV=1
Q14974	KPNB1	cDNA, FLJ95650, highly similar to Homo sapiens karyopherin (importin) beta 1 (KPNB1), mRNA OS=Homo sapiens PE=2 SV=1
P55060	CSE1L	Exportin-2 OS=Homo sapiens GN=CSE1L PE=1 SV=3
Q9BVC6	TMEM109	cDNA, FLJ94551 OS=Homo sapiens PE=2 SV=1
P16615	ATP2A2	ATPase Ca++ transporting cardiac muscle slow twitch 2 isoform 1 (Fragment) OS=Homo sapiens GN=ATP2A2 PE=2 SV=1
P05023	ATP1A1	Sodium/potassium-transporting ATPase subunit alpha-1 OS=Homo sapiens GN=ATP1A1 PE=1 SV=1
C9J7E5	TNPO3	Transportin-3 OS=Homo sapiens GN=TNPO3 PE=1 SV=1
Q9UNL2	SSR3	Translocon-associated protein subunit gamma OS=Homo sapiens GN=SSR3 PE=1 SV=1
O95373	IPO7	Importin-7 OS=Homo sapiens GN=IPO7 PE=1 SV=1
Q9BXW9	FANCD2	Fanconi anemia group D2 protein OS=Homo sapiens GN=FANCD2 PE=1 SV=2
P27824	CANX	Calnexin OS=Homo sapiens GN=CANX PE=1 SV=2

Cellular Proteins that increased >1.5-fold enrichment in MERS-CoV M cells compared to HEK293-Flp (control cells).

Uniprot All	**GENE ID**	**Description**
K9N7A1	VME1_CVEMC	Membrane protein n=35 RepID=VME1_CVEMC
Q8NF37	LPCAT1	Lysophosphatidylcholine acyltransferase 1 OS=Homo sapiens GN=LPCAT1 PE=1 SV=2
O14828	SCAMP3	Secretory carrier-associated membrane protein 3 OS=Homo sapiens GN=SCAMP3 PE=1 SV=3
P40616	ARL1	ADP-ribosylation factor-like protein 1 OS=Homo sapiens GN=ARL1 PE=1 SV=1
O95394	PGM3	cDNA FLJ13370 fis, clone PLACE1000653, highly similar to Phosphoacetylglucosamine mutase (EC 5.4.2.3) OS=Homo sapiens PE=2 SV=1
Q9UI26	IPO11	Importin-11 OS=Homo sapiens GN=IPO11 PE=1 SV=1
Q9Y4R8	TELO2	Telomere length regulation protein TEL2 homolog OS=Homo sapiens GN=TELO2 PE=1 SV=2
P05023	ATP1A1	Sodium/potassium-transporting ATPase subunit alpha-1 OS=Homo sapiens GN=ATP1A1 PE=1 SV=1
O75844	ZMPSTE24	CAAX prenyl protease 1 homolog OS=Homo sapiens GN=ZMPSTE24 PE=1 SV=2
O95674	CDS2	Phosphatidate cytidylyltransferase OS=Homo sapiens PE=2 SV=1
A0A024QYS2		Dinucleotide oxidase disulfide thiol exchanger 3 superfamily member 3 OS=Homo sapiens PE=2 SV=1
P46977	STT3A	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A OS=Homo sapiens GN=STT3A PE=1 SV=2
O43592	XPOT	cDNA FLJ75831, highly similar to Homo sapiens exportin, tRNA (nuclear export receptor for tRNAs) (XPOT), mRNA OS=Homo sapiens PE=2 SV=1
A0A024R0R4	SAE1	SUMO-1 activating enzyme subunit 1, isoform CRA_b OS=Homo sapiens GN=SAE1 PE=4 SV=1
P53985	SLC16A1	cDNA FLJ53399, highly similar to Monocarboxylate transporter 1 OS=Homo sapiens PE=2 SV=1
P16615	ATP2A2	ATPase Ca++ transporting cardiac muscle slow twitch 2 isoform 1 (Fragment) OS=Homo sapiens GN=ATP2A2 PE=2 SV=1
O75964	ATP5L	ATP synthase subunit g, mitochondrial OS=Homo sapiens GN=ATP5L PE=1 SV=3
P21796	VDAC1	Voltage-dependent anion-selective channel protein 1 OS=Homo sapiens GN=VDAC1 PE=1 SV=2
V9HW31	HEL-S-271	ATP synthase subunit beta OS=Homo sapiens GN=HEL-S-271 PE=1 SV=1
P61026	RAB10	Ras-related protein Rab-10 OS=Homo sapiens GN=RAB10 PE=1 SV=1
Q96CS3	FAF2	FAS-associated factor 2 OS=Homo sapiens GN=FAF2 PE=1 SV=2
P36507	MAP2K2	Dual specificity mitogen-activated protein kinase kinase 2 OS=Homo sapiens GN=MAP2K2 PE=1 SV=1

O75190	DNAJB6	DnaJ homolog subfamily B member 6 OS=Homo sapiens GN=DNAJB6 PE=1 SV=1
Q9BU23	LMF2	Lipase maturation factor 2 OS=Homo sapiens GN=LMF2 PE=1 SV=2
O14980	XPO1	Exportin-1 OS=Homo sapiens GN=XPO1 PE=1 SV=1
Q9Y277	VDAC3	Voltage-dependent anion-selective channel protein 3 OS=Homo sapiens GN=VDAC3 PE=1 SV=1
P11177	PDHB	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial OS=Homo sapiens GN=PDHB PE=1 SV=3
Q9H583	HEATR1	HEAT repeat containing 1 OS=Homo sapiens GN=HEATR1 PE=2 SV=1
P45880	VDAC2	Voltage-dependent anion channel 2, isoform CRA_a OS=Homo sapiens GN=VDAC2 PE=4 SV=1
P06733	ENO1	Alpha-enolase OS=Homo sapiens GN=ENO1 PE=1 SV=2
C9J7E5	TNPO3	Transportin-3 OS=Homo sapiens GN=TNPO3 PE=1 SV=1
P55786	NPEPPS	Puromycin-sensitive aminopeptidase OS=Homo sapiens GN=NPEPPS PE=1 SV=2
B4DR61	SEC61A1	Protein transport protein Sec61 subunit alpha isoform 1 OS=Homo sapiens GN=SEC61A1 PE=1 SV=1
Q9UBB4	ATXN10	Ataxin-10 OS=Homo sapiens GN=ATXN10 PE=1 SV=1
P56192	MARS	cDNA FLJ76789, highly similar to Homo sapiens methionine-tRNA synthetase (MARS), mRNA OS=Homo sapiens PE=2 SV=1
Q9UNL2	SSR3	Translocon-associated protein subunit gamma OS=Homo sapiens GN=SSR3 PE=1 SV=1
Q15645	TRIP13	Pachytene checkpoint protein 2 homolog OS=Homo sapiens GN=TRIP13 PE=1 SV=2
O43175	PHGDH	D-3-phosphoglycerate dehydrogenase OS=Homo sapiens GN=PHGDH PE=1 SV=4
Q86VP6	CAND1	cDNA FLJ77762, highly similar to Homo sapiens cullin-associated and neddylation-dissociated 1 (CAND1), mRNA OS=Homo sapiens PE=2 SV=1
Q53H12	AGK	Acylglycerol kinase, mitochondrial OS=Homo sapiens GN=AGK PE=1 SV=2
Q14739	LBR	Lamin B receptor, isoform CRA_a OS=Homo sapiens GN=LBR PE=4 SV=1
P24539	ATP5F1	cDNA FLJ78635, highly similar to Homo sapiens ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b, isoform 1 (ATP5F1), transcript variant 1, mRNA OS=Homo sapiens PE=2 SV=1
V9HW26	HEL-S-123m	ATP synthase subunit alpha OS=Homo sapiens GN=HEL-S-123m PE=2 SV=1
Q96EY1	DNAJA3	DnaJ (Hsp40) homolog, subfamily A, member 3 variant (Fragment) OS=Homo sapiens GN=DNAJA3 PE=1 SV=1
A0A024RAM0	TNPO1	Transportin 1, isoform CRA_a OS=Homo sapiens GN=TNPO1 PE=4 SV=1
A0A024R5F7	DHCR7	7-dehydrocholesterol reductase isoform 1 OS=Homo sapiens GN=DHCR7 PE=2 SV=1
Q14257	RCN2	Reticulocalbin-2 OS=Homo sapiens GN=RCN2 PE=1 SV=1

P12004	PCNA	Proliferating cell nuclear antigen OS=Homo sapiens GN=PCNA PE=1 SV=1
P06493	DKFZp686L20222	Putative uncharacterized protein DKFZp686L20222 OS=Homo sapiens GN=DKFZp686L20222 PE=4 SV=1
Q9P035	HACD3	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3 OS=Homo sapiens GN=HACD3 PE=1 SV=2
P50570	DNM2	Dynamin-2 OS=Homo sapiens GN=DNM2 PE=1 SV=2
P50402	EMD	Emerin OS=Homo sapiens GN=EMD PE=1 SV=1
Q8NBU5	ATAD1	ATPase family AAA domain-containing protein 1 OS=Homo sapiens GN=ATAD1 PE=1 SV=1
A8K005		cDNA FLJ77896, highly similar to Homo sapiens Ras homolog enriched in brain (RHEB), mRNA OS=Homo sapiens PE=2 SV=1
E7ESC6	XPO7	Exportin-7 OS=Homo sapiens GN=XPO7 PE=1 SV=1
P31689	DNAJA1	DnaJ homolog subfamily A member 1 OS=Homo sapiens GN=DNAJA1 PE=1 SV=2
P20020	ATP2B1	Plasma membrane calcium-transporting ATPase 1 OS=Homo sapiens GN=ATP2B1 PE=1 SV=3
P41250	GARS	Glycine--tRNA ligase OS=Homo sapiens GN=GARS PE=1 SV=3
P12235	SLC25A4	Solute carrier family 25 member 4 isoform 1 (Fragment) OS=Homo sapiens GN=SLC25A4 PE=2 SV=1
P39656	DDOST	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit OS=Homo sapiens GN=DDOST PE=3 SV=1
F5H5D3	TUBA1C	Tubulin alpha-1C chain OS=Homo sapiens GN=TUBA1C PE=1 SV=1
Q92621	NUP205	Nuclear pore complex protein Nup205 OS=Homo sapiens GN=NUP205 PE=1 SV=3
O00483	NDUFA4	Cytochrome c oxidase subunit NDUFA4 OS=Homo sapiens GN=NDUFA4 PE=1 SV=1
Q9H9B4	SFXN1	Sideroflexin-1 OS=Homo sapiens GN=SFXN1 PE=1 SV=4
A0A024R0M6	TIMM50	Translocase of inner mitochondrial membrane 50 homolog (Yeast), isoform CRA_b OS=Homo sapiens GN=TIMM50 PE=4 SV=1
Q6NVC0	SLC25A5	SLC25A5 protein (Fragment) OS=Homo sapiens GN=SLC25A5 PE=2 SV=1
Q9UJS0	SLC25A13	Solute carrier family 25, member 13 (Citrin) variant (Fragment) OS=Homo sapiens PE=2 SV=1
Q96P70	IPO9	Importin-9 OS=Homo sapiens GN=IPO9 PE=1 SV=3
Q15758	SLC1A5	Neutral amino acid transporter B(0) OS=Homo sapiens GN=SLC1A5 PE=1 SV=2
Q59EI9	SLC25A6	ADP,ATP carrier protein, liver isoform T2 variant (Fragment) OS=Homo sapiens PE=2 SV=1
O60725	ICMT	Protein-S-isoprenylcysteine O-methyltransferase OS=Homo sapiens GN=ICMT PE=1 SV=1
D3DUJ0	AFG3L2	Similar to AFG3 ATPase family gene 3-like 2 (Yeast) (Fragment) OS=Homo sapiens PE=2 SV=1
P07437	TUBB	Beta 5-tubulin OS=Homo sapiens GN=TUBB PE=2 SV=1

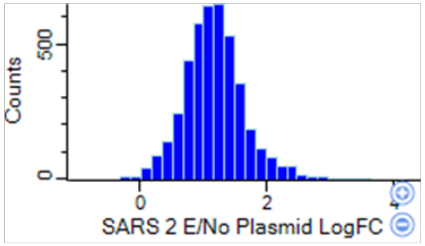
O00264	PGRMC1	Membrane-associated progesterone receptor component 1 OS=Homo sapiens GN=PGRMC1 PE=1 SV=3
P18085	ARF4	ADP-ribosylation factor 4 OS=Homo sapiens GN=ARF4 PE=1 SV=3
P04181	OAT	Ornithine aminotransferase, mitochondrial OS=Homo sapiens GN=OAT PE=1 SV=1
Q02978	SLC25A11	Mitochondrial 2-oxoglutarate/malate carrier protein OS=Homo sapiens GN=SLC25A11 PE=1 SV=3
Q96A33	CCDC47	Coiled-coil domain-containing protein 47 OS=Homo sapiens GN=CCDC47 PE=1 SV=1
Q9NP72	RAB18	Ras-related protein Rab-18 OS=Homo sapiens GN=RAB18 PE=1 SV=1
A0A024R497	ACSL3	Acyl-CoA synthetase long-chain family member 3, isoform CRA_a OS=Homo sapiens GN=ACSL3 PE=4 SV=1
Q7KZN9	COX15	Cytochrome c oxidase assembly protein COX15 homolog OS=Homo sapiens GN=COX15 PE=1 SV=1
A4D0U5	TES	Testis derived transcript (3 LIM domains) OS=Homo sapiens GN=TES PE=4 SV=1
P22695	UQCRC2	Cytochrome b-c1 complex subunit 2, mitochondrial OS=Homo sapiens GN=UQCRC2 PE=1 SV=3
E9KL35	GNB2L1	Epididymis tissue sperm binding protein Li 3a OS=Homo sapiens PE=1 SV=1
K7ER00	FARSA	Phenylalanine--tRNA ligase alpha subunit OS=Homo sapiens GN=FARSA PE=1 SV=1
P23396	RPS3	40S ribosomal protein S3 OS=Homo sapiens GN=RPS3 PE=1 SV=2
O95831	AIFM1	Apoptosis-inducing factor 1, mitochondrial OS=Homo sapiens GN=AIFM1 PE=1 SV=1
P09622	DLD	Dihydrolipoyl dehydrogenase, mitochondrial OS=Homo sapiens GN=DLD PE=1 SV=2
P36542	ATP5C1	ATP synthase subunit gamma, mitochondrial OS=Homo sapiens GN=ATP5C1 PE=1 SV=1
P17931	LGALS3	Galectin (Fragment) OS=Homo sapiens PE=2 SV=1
A0A024R3W2	TOMM20	Translocase of outer mitochondrial membrane 20 homolog (Yeast), isoform CRA_a OS=Homo sapiens GN=TOMM20 PE=3 SV=1
P43246	MSH2	DNA mismatch repair protein Msh2 OS=Homo sapiens GN=MSH2 PE=1 SV=1
P78527	PRKDC	DNA-dependent protein kinase catalytic subunit OS=Homo sapiens GN=PRKDC PE=1 SV=3
O75947	ATP5H	ATP synthase subunit d, mitochondrial OS=Homo sapiens GN=ATP5H PE=1 SV=3
A0A024R0P9	TOMM40	Translocase of outer mitochondrial membrane 40 homolog (Yeast), isoform CRA_c OS=Homo sapiens GN=TOMM40 PE=4 SV=1
V9HWB4	HEL-S-89n	Epididymis secretory sperm binding protein Li 89n OS=Homo sapiens GN=HEL-S-89n PE=2 SV=1
P25398	RPS12	40S ribosomal protein S12 OS=Homo sapiens GN=RPS12 PE=1 SV=3

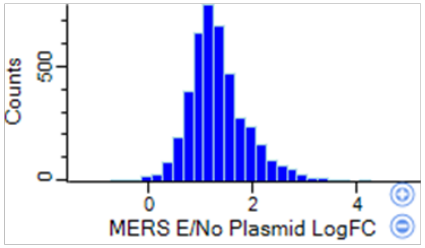
E7ESZ7	NDUFA10	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial OS=Homo sapiens GN=NDUFA10 PE=1 SV=1
Q13838	DDX39B	cDNA FLJ55484, highly similar to ATP-dependent RNA helicase DDX39 (EC 3.6.1.-) OS=Homo sapiens PE=2 SV=1
P04843	RPN1	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Homo sapiens GN=RPN1 PE=1 SV=1
Q8NI60	ADCK3	Atypical kinase ADCK3, mitochondrial OS=Homo sapiens GN=ADCK3 PE=1 SV=1
B4DLN1	Homo sapiens	Uncharacterized protein OS=Homo sapiens PE=2 SV=1
A0A024R7F9	GCDH	Glutaryl-Coenzyme A dehydrogenase, isoform CRA_a OS=Homo sapiens GN=GCDH PE=3 SV=1
Q96EC8	YIPF6	Protein YIPF OS=Homo sapiens PE=2 SV=1
O00165	HAX1	HCLS1-associated protein X-1 OS=Homo sapiens GN=HAX1 PE=1 SV=2
A0A024QZN7	C10orf70	Chromosome 10 open reading frame 70, isoform CRA_b OS=Homo sapiens GN=C10orf70 PE=4 SV=1
B0YIW9	ARL5B	ADP-ribosylation factor-like 5B OS=Homo sapiens GN=ARL5B PE=3 SV=1
Q5QPK2	DPM1	Dolichol-phosphate mannosyltransferase subunit 1 (Fragment) OS=Homo sapiens GN=DPM1 PE=1 SV=1
P57088	TMEM33	SHINC3 OS=Homo sapiens GN=SHINC3 PE=2 SV=1
Q5T4U5	ACADM	Acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain, isoform CRA_a OS=Homo sapiens GN=ACADM PE=1 SV=1
J9JIE6	TMCO1	Transmembrane and coiled-coil domain-containing protein 1 OS=Homo sapiens GN=TMCO1 PE=1 SV=1
P56385	ATP5I	ATP synthase subunit e, mitochondrial OS=Homo sapiens GN=ATP5I PE=1 SV=2

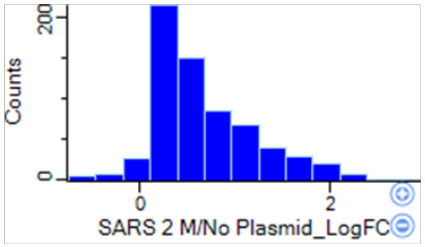
Appendix E : Histogram of proteins distribution on different data sets :

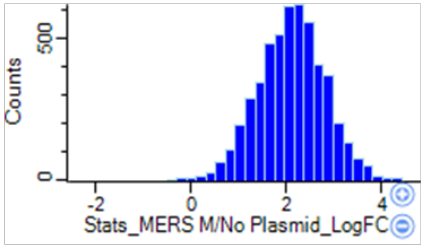
HEK293 cells data sets:

TOTAL number of proteins



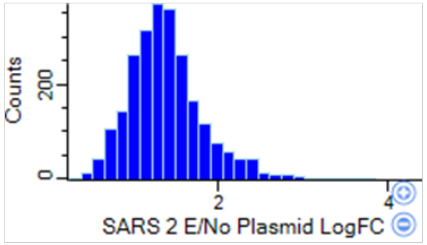


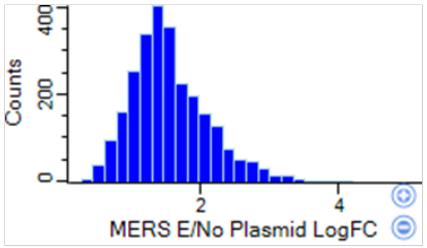


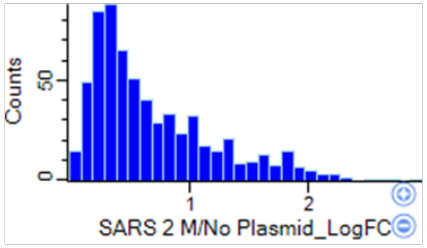


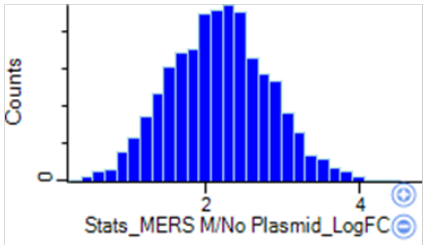
Proteins increased in amount

> 0 log2 fold (T-Test < 0.05)



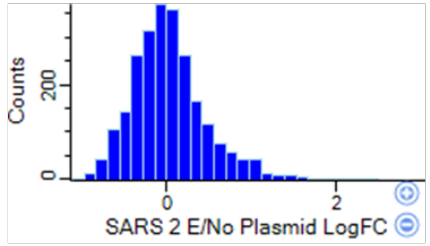


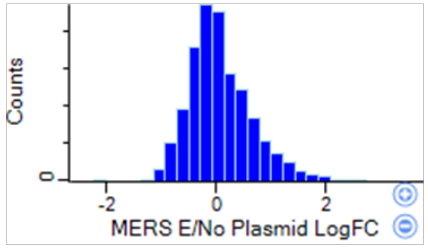


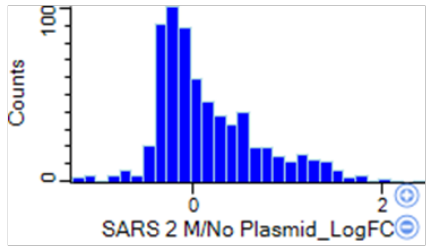


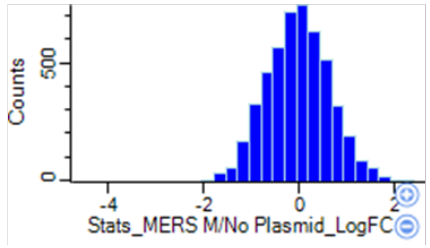
Proteins increased in amount

>0 log2 fold (T-Test < 0.05) and Normalised



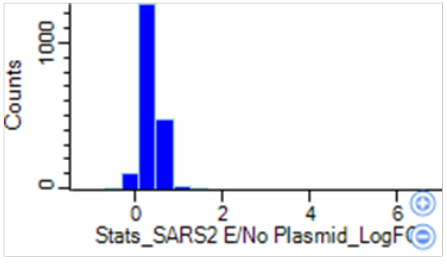


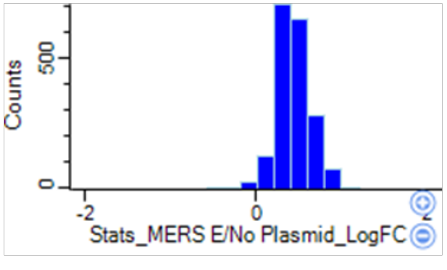


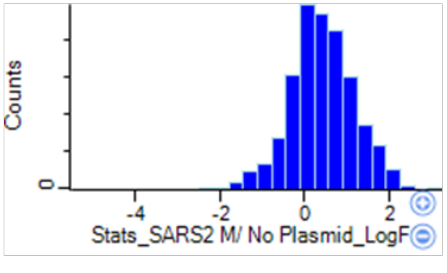


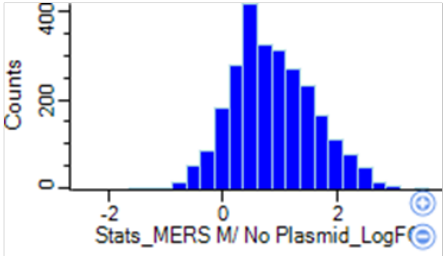
Dubca cells data sets:

TOTAL number of proteins



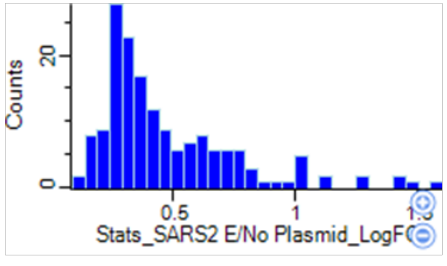


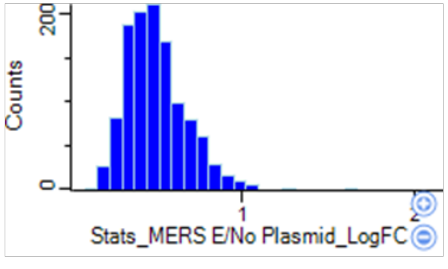


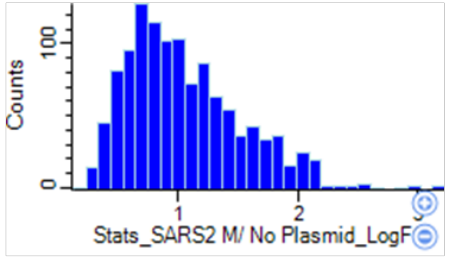


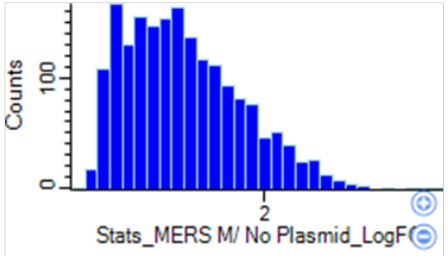
Proteins increased in amount

> 0 log2 fold (T-Test < 0.05)



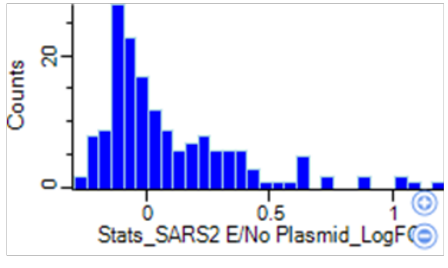


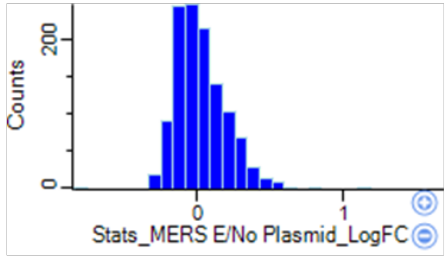


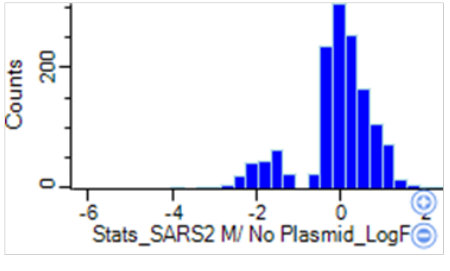


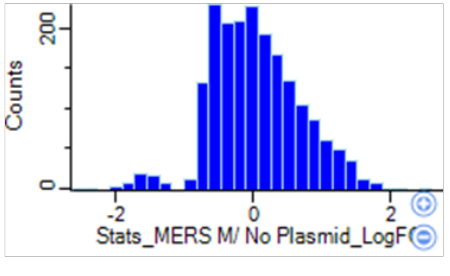
Proteins increased in amount

>0 log2 fold (T-Test < 0.05) and Normalised



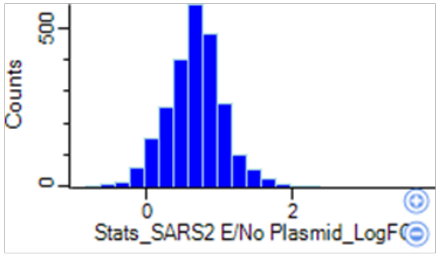


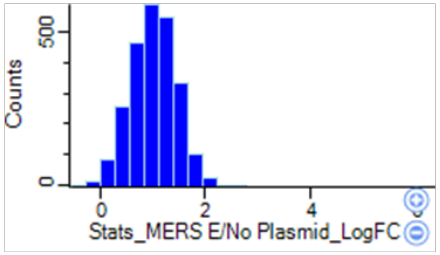


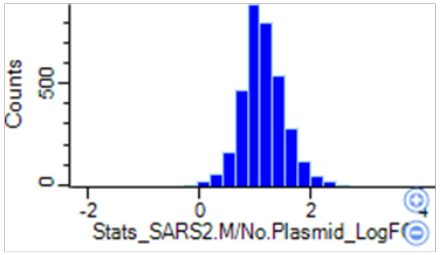


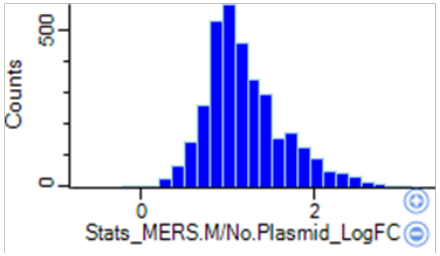
PaKiT cells data sets:

TOTAL number of proteins



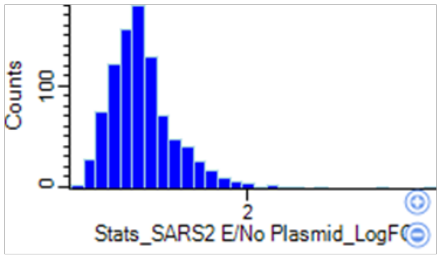


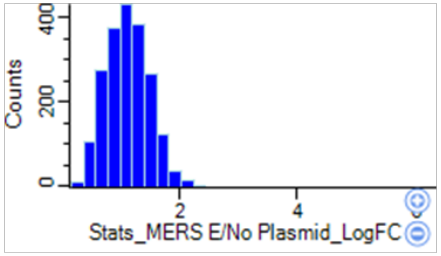


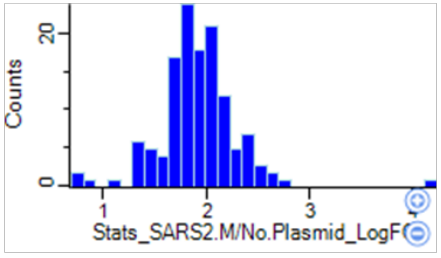


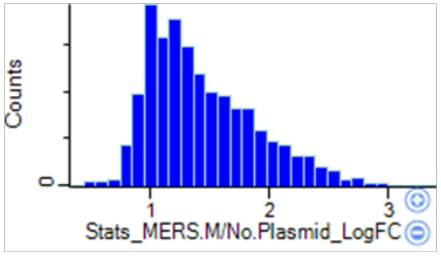
Proteins increased in amount

> 0 log2 fold (T-Test < 0.05)



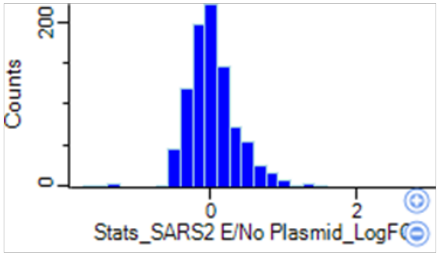


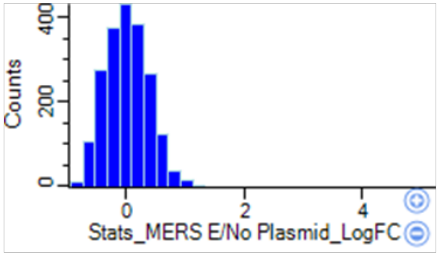


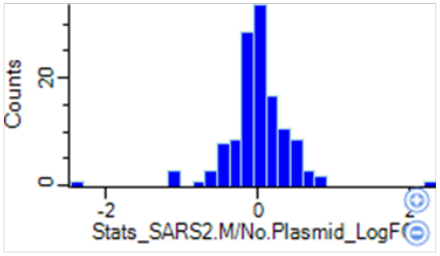


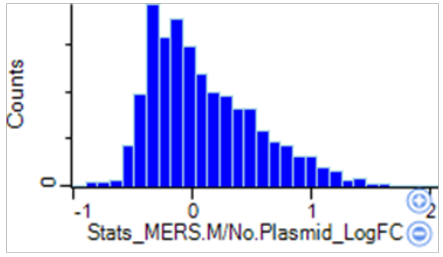
Proteins increased in amount

>0 log2 fold (T-Test < 0.05) and Normalised









Appendix F: the proteomics data including row data, analysed data, and the results of David analysis.

Link for Supplementary Tables: plates plan (for thesis).xlsx

Appendix G : siRNA expriment plat plan

Link for Supplementary Tables:

Thesis HEK293 E Proteomic data set (done).xlsx

Thesis DUBCA E Proteomic data set (done).xlsx

Thesis DUBCA M Proteomic data set (done).xlsx

Thesis HEK293 M Proteomic data set (done).xlsx

Thesis Pakit E Proteomic data set (done).xlsx

Thesis Pakit M Proteomic data set (done).xlsx