JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Identification and validation of the human genes that support human coronavirus infection using functional genomics approaches

Inaugural Dissertation submitted to the Faculty of Medicine in partial fulfillment of the requirements for the PhD-Degree of the Faculties of Veterinary Medicine and Medicine of the Justus Liebig University-Giessen

By

Albert, Benadict Vincent

Of

Kanchipuram-India

Giessen (2023)

From the Institute of “Rudolf-Buchheim Institut für Pharmakologie”
Director / Chairman: Prof. Dr. Michael Kracht
of the Faculty of Medicine of the Justus Liebig University Giessen

First Reviewer / Supervisor: Prof. Dr. Michael Kracht
Second Reviewer: Prof. Dr. Eva Herker
Chair: Prof. Dr. Christa Ewers
Vice-Chair / Co-Supervisor: Prof. Dr. Friedemann Weber

Date of Doctoral Defense: 13.11.2023

Table of Contents

1. Abstract

Human coronaviruses frequently cause respiratory tract infections ranging from mild upper respiratory illness to severe acute respiratory distress syndrome and death. For example, the global spread of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has triggered a public health and economic crisis. Although the development of mRNA vaccines and antiviral drugs has helped reduce the extent and severity of SARS-CoV-2 infections, vaccines and drugs targeting viral proteins have limitations in terms of potential drug resistance or side effects and a limited number of viral targets. CoV replicate in infected host cells using a variety of cellular factors for their replication. Therefore, it is of great importance to identify the set of human genes required and manipulated by coronaviruses for propagation and to elucidate their function in the viral replication cycle.

In this dissertation, the complex relationship between human coronaviruses and the essential host factors responsible for viral replication and cell survival was systematically investigated using modern genome-wide sequencing techniques and loss-of-function approaches. Focusing on human coronavirus 229E (HCoV-229E) and Middle East respiratory syndrome coronavirus (MERS-CoV), CRISPR-Cas9-based genetic tools were used to identify specific host factors and investigate their functional relevance. In addition, the antiviral effects of thapsigargin, a small molecule activator of an ER stress response and effective inhibitor of coronavirus replication, were investigated using genetic screening methods.

Initially, we investigated the hypothesis that host cell genes particularly strongly induced by CoV play a critical role in coronavirus replication and their disruption could represent a potential target for antiviral intervention. From a genome-wide transcriptome analysis of Huh7 cells infected with HCoV-229E, 11 genes induced by HCoV-229E were selected and characterized in more detail. Here, subsequent knockouts of ANKRD1, EDEM1, KLF6, and FICD did not significantly affect viral replication and the host cell transcriptome.

To identify essential host factors in an expanded, genome-wide approach, an unbiased, CRISPR-Cas9-based knockout screening system was employed, which led to the identification of ANPEP, the cellular receptor for HCoV-229E, as the strongest hit. Similarly, the cellular surface proteins DPP4 and TMPRSS2 were identified as the strongest hits in additional screens of MERS-CoV-infected cells. Validation experiments on the functional role of host factors in viral release and cell viability after infection revealed that the absence of host factors SFTA2, HDAC4, and ZDHHC3 (which emerged as significant hits from HCoV-229E selection) each slightly impaired viral release during HCoV-229E infection, providing new insights into the complex mechanisms employed by CoV to ensure successful replication.

In the final part of this work, genetic screens identified over 1000 host factors as significant hits in thapsigargin-treated and / or HCoV-229E-infected cells. For functional follow-up, the intersection of this group of genes with significantly deregulated proteins from comparable studies of the nascent

proteome was determined and 40 thapsigargin-specific factors were identified. Individual knockouts of ARFGAP3, EMC6, and CLU each increased the cell viability of thapsigargin-treated infected cells, and the absence of SEC24A reduced the long-term antiviral effect of thapsigargin on HCoV-229E replication.

In summary, this work provides valuable insights into specific host factors involved in antiviral effects and CoV-induced cell death and points to potential targets of antiviral agents. Based on these findings, the molecular mechanisms involved in the underlying host-virus interactions and their applicability to suppress future emerging CoV infections can be further investigated.

Zussamenfassung

Humane Coronaviren verursachen häufig Infektionen des Respirationstraktes, die von milden Erkrankungen der oberen Atemwege bis hin zum schweren akuten Atemnotsyndrom und Todesfällen reichen. So hat die weltweite Verbreitung des *Severe Acute Respiratory Syndrome Coronavirus* 2 (SARS-CoV-2) eine öffentliche Gesundheits- und Wirtschaftskrise ausgelöst. Obwohl die Entwicklung von mRNA-Impfstoffen und antiviralen Medikamenten dazu beigetragen hat, das Ausmaß und den Schweregrad von SARS-CoV-2-Infektionen zu reduzieren, weisen Impfstoffe und Medikamente, welche auf virale Proteine abzielen, Limitationen hinsichtlich möglicher Medikamentenresistenzen oder Nebenwirkungen und einer begrenzten Anzahl von viralen Zielmolekülen auf. CoV replizieren in den infizierten Wirtszellen, indem sie eine Vielzahl an zellulären Faktoren für ihre Replikation nutzen. Daher ist es von großer Bedeutung, die Gesamtheit an menschlichen Genen zu identifizieren, die von Coronaviren zur Propagierung benötigt und manipuliert werden und ihre Funktion im Replikationszyklus der Viren aufzuklären.

In dieser Doktorarbeit wurde die komplexe Beziehung zwischen humanen Coronaviren und den essenziellen Wirtsfaktoren, die für die virale Replikation und das Überleben der Zelle verantwortlich sind, systematisch mittels moderner genomweiter Sequenzierungstechniken und *Loss of Function* Ansätze untersucht. Unter Fokussierung auf das humane Coronavirus 229E (HCoV-229E) sowie dem *Middle East Respiratory Syndrome Coronavirus* (MERS-CoV) wurden CRISPR-Cas9-basierte genetische Werkzeuge eingesetzt, um spezifische Wirtsfaktoren zu identifizieren und deren funktionelle Relevanz zu untersuchen. Darüber hinaus wurden die antiviralen Effekte von Thapsigargin, einem niedermolekularen Aktivator einer ER-Stressreaktion und effektiven Inhibitor der Coronavirusreplikation, mittels genetischer *Screening* Verfahren untersucht.

Initial wurde die Hypothese untersucht, dass besonders stark durch CoV induzierte Wirtszellgene eine entscheidende Rolle bei der Coronavirus-Replikation spielen und ihre Störung ein potenzielles Ziel für antivirale Eingriffe darstellen könnte. Aus einer genomweiten Transkriptom-Analyse von mit HCoV-229E infizierten Huh7-Zellen wurden 11 durch HCoV-229E induzierte Gene ausgewählt und näher charakterisiert. Hierbei zeigten nachfolgende *Knockouts* von *ANKRD1*, *EDEM1*, *KLF6* und *FICD* keine signifikante Beeinflussung der viralen Replikation und des Wirtszelltranskriptoms.

Um die essentiellen Wirtsfaktoren in einem erweiterten, genomweiten Ansatz zu ermitteln, wurde im Folgenden ein unvoreingenommenes, auf CRISPR-Cas9-basiertes *Knockout-Screening* System eingesetzt, welches zur Identifizierung von *ANPEP*, dem zellulären Rezeptor für HCoV-229E, als stärkstem Hit führte. Analog wurden die zellulären Oberflächenproteine *DPP4* und *TMPRSS2* als stärkste Hits in weiteren *Screens* von mit MERS-CoV-infizierten Zellen identifiziert. In Validierungsexperimenten zur funktionellen Rolle von Wirtsfaktoren bei der viralen Freisetzung und der Zellviabilität nach der Infektion zeigte sich, dass das Fehlen der Wirtsfaktoren SFTA2, HDAC4

und ZDHHC3 (die sich als signifikante Treffer aus der HCoV-229E-Auswahl ergaben) die virale Freisetzung während der HCoV-229E-Infektion jeweils leicht beeinträchtigte, wodurch neue Einblicke in die komplexen Mechanismen gewonnen werden konnten, die von CoV zur Sicherstellung einer erfolgreichen Replikation eingesetzt werden.

Im letzten Teil dieser Arbeit wurden mittels der genetischen *Screens* über 1000 Wirtsfaktoren in mit Thapsigargin behandelten und / oder HCoV-229E-infizierten Zellen als signifikante Hits identifiziert. Für funktionelle Nachuntersuchungen wurde die Schnittmenge dieser Gengruppe mit signifikant deregulierten Proteinen aus vergleichbaren Untersuchungen des naszierenden Proteoms ermittelt und 40 Thapsigargin-spezifische Faktoren identifiziert. Individuelle *Knockouts* von ARFGAP3, EMC6 und CLU erhöhten jeweils die Zellviabilität von Thapsigargin-behandelten infizierten Zellen und das Fehlen von SEC24A reduzierte den langfristigen antiviralen Effekt von Thapsigargin auf die HCoV-229E Replikation.

Zusammenfassend liefert diese Arbeit wertvolle Erkenntnisse über spezifische Wirtsfaktoren, welche an antiviralen Effekten und dem CoV-induzierten Zelltod beteiligt sind und weist auf potenzielle Zielmoleküle antiviraler Wirkstoffe hin. Basierend auf diesen Erkenntnissen können die beteiligten molekularen Mechanismen der zugrundeliegenden Wirt-Virus-Interaktionen und deren Anwendbarkeit zur Suppression zukünftig auftretender CoV Infektionen weiter untersucht werden.

2. Introduction

2.1 Background to human Coronavirus – structure, genome, and viral life cycle

Coronaviruses (CoVs) comprise a larger group of enveloped positive-sense single-stranded RNA viruses with a broad range (humans, other mammals, and avians) of host tropism. Until this date, seven human coronaviruses have been identified. The first human coronavirus (HCoV) was identified in 1965 from the nasal specimens of patients with the common cold. Among the 30 other strains identified, the prototypic strain of human coronavirus 229E (HCoV-229E) was identified from the infected tissue culture experiments [1]. Shortly after the identification of HCoV-229E, another human coronavirus was isolated from the human tracheal organ culture and was termed HCoV-OC43 (Organ Culture 43) [2, 3]. Until the emergence of a novel coronavirus in 2003, these two HCoVs were the primary focus of human coronavirus research. The novel coronavirus (later termed to be SARS-CoV) caused an outbreak in the year 2002-2003 where the infected patients suffered severe acute respiratory syndrome (SARS) [4]. Two other human coronaviruses were identified in consecutive years; HCoV-NL63 (Netherland 63) was isolated from the nasopharyngeal specimen of a 7-month-old child and HCoV-HKU1 (Hong Kong University 1) was isolated from a 71-year old patient with pneumonia [5, 6]. Several years later, in 2012, and 2019 two zoonotic HCoVs were identified; namely the Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) [7, 8].

Coronavirus virions are generally spherical in structure and are about 80-120 nm (diameter) in size. They have a typical fringe on their surface, which resembles solar corona; hence, the name coronavirus was derived. The virions are composed of four main structural proteins: the spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein. However, not all of the four proteins are needed to form the outer structure of the virion. In most cases, the viral envelope is composed of the S, M, and E proteins, except in some coronaviruses like HCoV-HKU1, and bovine coronavirus (BCoV) an additional hemagglutinin esterase (HE) protein is found [9]. The spike (S) protein is a major component of the virion envelope and ranges from 128-160 kDa (non-glycosylated) and 150-200 kDa (glycosylated). The major role of the S protein is to mediate the viral entry by attachment to the host receptor and thereby facilitating the fusion of the host cell and viral membrane. The S protein is functionally divided into two subunits: the S1 subunit plays an important role in the host cell receptor recognition and the S2 subunit is important for membrane fusion events [10, 11]. The membrane (M) protein ranges from 25-30 kDa and plays an important role in the virion assembly. The envelope (E) protein is a very small structural protein ranging from 8-12 kDa and plays an important role in virion assembly, envelope formation, and budding. The nucleocapsid (N) protein ranges from 43-46 kDa and is a structural component of the viral nucleocapsid. The N protein functions to pack the

viral genome into a ribonucleoprotein (RNP) particle to protect the genomic RNA and to incorporate the genome into a new virion [10, 11].

The very distinct feature of the coronavirus is the large size of the viral genome ranging from 26 to 32 kb. The viral genome is a non-segmented, positive sense single-stranded RNA (+ssRNA) which is similar to most eukaryotic mRNAs with 5' caps and 3' poly-adenine tails. The major portion of the viral genome consists of two large overlapping open reading frames (ORF), ORF1a and ORF1b, which encode 16 non-structural viral proteins (nsps). The nsp1-11 are encoded in ORF1a and nsp12-16 are encoded in ORF1b [11]. The minor portion of the viral genome is transcribed into a set of subgenomic RNAs that encodes for the viral S, E, M, and N proteins as well as other accessory proteins which are specific depending on the coronavirus species [10].

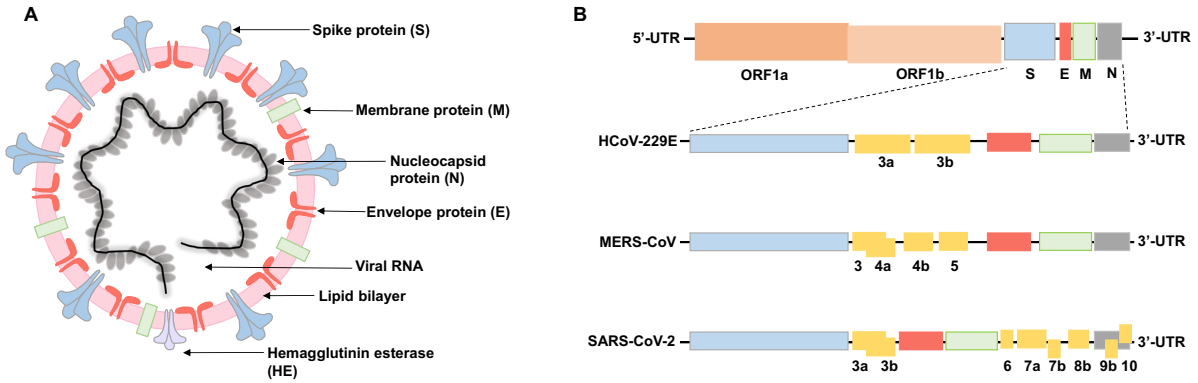


Figure 1: Illustrative figures of the structure and genome organization of coronaviruses.
A) **Typical structure of a mature coronavirus virion. The outer membrane is composed of lipids that are derived from the host cell layer during the viral egress and is assembled with a spike (S), envelope (E), membrane (M), and hemagglutinin esterase (HE) (for some ß-coronaviruses) proteins. The inner region is composed of the viral RNA genome that is surrounded by a protective capsid, which is a lattice of repeated proteins called the nucleocapsid (N) proteins.** B) **The genome encodes two large genes ORF1a (dark orange) and ORF1b (light orange) that encode the viral non-structural proteins (nsp1-16). The nsps are translated and process the translation of sub-genomic RNA that encodes the spike (light blue), envelope (red), membrane (green), and nucleocapsid (grey) protein. Additionally, species-specific accessory proteins that are involved in the regulation of viral replication and the host response are illustrated in the figure with examples for HCoV-229E, MERS, and SARS-CoV-2. This figure was modified from the source [10] (Fig. 1 and 6 from the publication) and re-created using Microsoft PowerPoint for this thesis.**

The coronavirus life cycle is staged into four steps: attachment and entry into the host cell, translation of the open reading frames, replication of the viral genome, and assembly and budding of newly packed virions. In the first step, the spike (S) protein of the coronavirus is attached to the host cell receptors. This is a very crucial and determinant stage of pathogenicity, as the host receptors vary for all HCoVs. HCoV-229E utilizes the host aminopeptidase N (ANPEP) as the receptor [12]. Angiotensin-converting enzyme 2 (ACE2) is used by SARS-CoV / SARS-CoV-2 [13], and HCoV-NL63 [14], while dipeptidyl peptidase 4 (DPP4) by MERS [15], and 9-O-acetylated sialic acid are used by HCoV-OC43 and HCoV-HKU1 [16, 17]. The attachment of the S1 subunit of the spike protein facilitates a conformational change in the S2 subunit of the spike protein and thereby promotes a fusion of the viral and plasma membrane forming an endosome [18]. The viral genome is released into the host cytoplasm following the endosomal uncoating. In the second step, the open reading frames in the viral genome are translated into two polyproteins- ORF1a into polyprotein pp1a, and the downstream ORF1b is translated through ribosomal frameshifting of one nucleotide (nt) in the -1 direction of ORF1a into polyprotein pp1ab [19, 20]. Furthermore, the polyproteins are post-translationally processed into 16 non-structural proteins: pp1a (nsp1-11) and pp1ab (nsp1-10, and 12-16). The nsps play an important role in viral genome replication by forming a viral replication-transcription-complex (RTC) and also recruiting the early endosome host factors to form double-membrane vesicles (DMV) [21]. The viral genome replication is initiated in this secured environment and the full-length negative-sense genomic copies are synthesized to provide a template for the generation of new positive-sense genomic viral RNA [19]. Coronaviruses possess a special feature called the discontinuous transcription process where the RTC produces a subset of negative-sense subgenomic RNA, which is later utilized to produce the positive-sense subgenomic mRNAs, which are translated into structural and accessory proteins [22]. The final step of the coronavirus replication starts with the accumulation and assembly of viral structural proteins at the Endoplasmic Reticulum - Golgi intermediate compartment (ERGIC). The viral nucleocapsid bounded genomic RNA interacts with the assembled structural proteins at the ERGIC and forms a new virion, which is then secreted outside the cytoplasm via exocytosis [21].

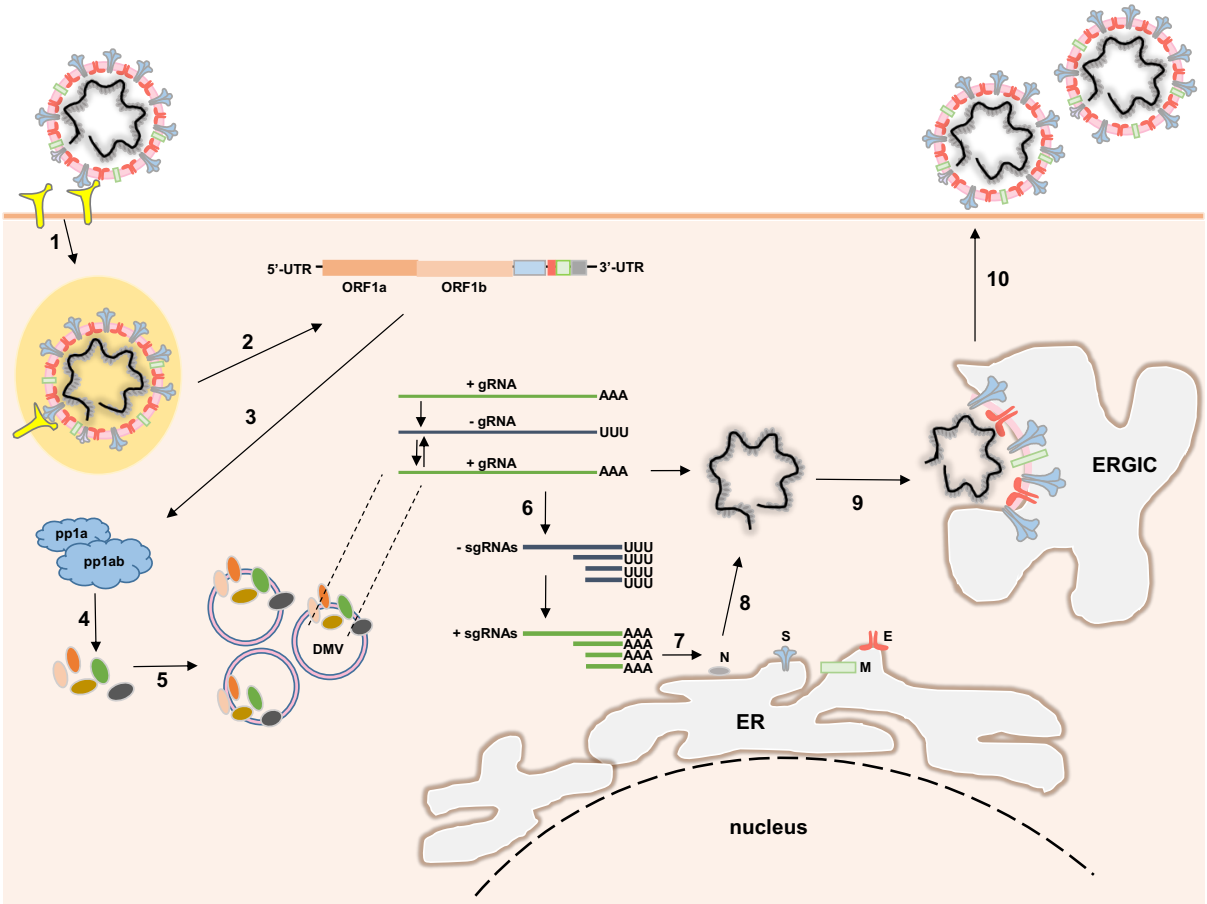


Figure 2: Replication cycle of a typical coronavirus.
1. The coronavirus particle first binds to the specific cellular membrane receptors (such as ANPEP, ACE2, and DPP4) by the spike (S) protein and enters the cell via the endosomes. **2**. By the specific cleavage action of the host-cell receptor and proteases, the viral membrane is uncoated from the endosomes and the viral genome is released into the cytoplasm. **3**. Following the release of the viral genome, the two large open-reading frames ORF1a and ORF1b translates to polyproteins pp1a and pp1b. **4**. The polypeptides are post-translationally processed to individual non-structural proteins (nsps) and form a protective micro-environment called the double-membrane vesicles (DMVs) (**5**) for the viral genome replication and transcription of subgenomic mRNAs (**6**). **7**. The subgenomic mRNAs are translated to form the structural proteins (S, E, M, and N) in the endoplasmic-reticulum (ER). **8**. The nucleocapsid protein wraps up the full-length viral genome to form a capsid and assemble with the structural proteins that are translocated to ER-Golgi intermediate complex (ERGIC) (**9**). **10**. The fully assembled virions are then secreted from the cells via exocytosis. UTR-Untranslated region, gRNA-genomic RNA, sgRNAs-subgenomic RNAs, AAA-poly-adenine tail, UUU-poly-uridine tail**.** The figure was modified from the source [19] (Fig. 1 from the publication) and re-created using Microsoft PowerPoint for this thesis.

2.2 Pathogenesis of human coronaviruses

Human coronaviruses are known to cause mild to severe respiratory tract illness. The outbreak of SARS-CoV-2 on the 8th of December, 2019 changed the perspective of the coronaviruses to cause a global pandemic. As of now, more than 600 million people have been infected with SARS-CoV-2 globally and among them, around 6 million deaths have been reported (WHO reports). The primary route of transmission is through respiratory aerosols from infected individuals. It is also reported to be transmitted through the contact of infected surfaces and fecal excretions of the infected individuals [23]. The infected individual starts to present the symptoms or can be asymptomatic following the incubation period of 2-14 days. The common symptoms are marked by fever, dry cough, fatigue, muscle pain, dyspnoea, and diarrhea [24]. Depending on the age, health status, and comorbidities of the infected individual, the disease (COVID-19) can be worse leading to acute respiratory distress syndrome (ARDS), and pneumonia [23-25]. The mortality rate of COVID-19 is comparatively lesser than that of the disease caused by MERS-CoV with 2494 infected cases and 858 deaths during the period 2012. The MERS-CoV infected individuals suffer from the common symptoms of COVID-19 and in the major cases with ARDS, pneumonia, and renal failure [26, 27].

Unlike the SARS-CoV and MERS-CoV, the other four human coronaviruses are generally not life-threatening and only cause severity in young children, the elderly, and those that are immunocompromised. HCoV-NL63, HCoV-229E, HCoV-OC43, and HCoV-HKU1 following the incubation period of 2-11 days are often associated with mild upper respiratory tract infections and are considered to cause the common cold. However, HCoV-NL63 has been reported to be associated with gastrointestinal tract infections and Kawasaki disease [28]. HCoV-OC43 has been reported to be detected in patients with Parkinson's disease and multiple-sclerosis (MS) [29]. HCoV-HKU1 can also induce severe lower respiratory tract illness in young and immunocompromised patients [30]. HCoV-229E was reported to cause pneumonia in elderly patients with diabetes mellitus [31]. Interestingly, a rare case of HCoV-229E associated with ARDS was reported in 2018 in a healthy adult with previously non-existent underlying diseases [32]. Although these human coronaviruses are restricted to the respiratory system, they can still manifest complications in the heart, kidneys, and brain [33].

2.3 Drug-based strategies to inhibit coronavirus infection

Before the emergence of SARS-CoV in 2003, symptomatic treatment was the only considered option for human coronavirus infections. Previous CoV outbreaks and the recent pandemic situations fastened the research for finding a suitable therapeutic anti-viral agent and preventing any upcoming potential emerging CoV infections. Recent developments in the CoV antiviral strategies include: targeting the CoV proteins for effective inhibition of viral infection and targeting the host cell proteins involved in the viral replication [34].

2.3.1 Inhibition of coronavirus infection by targeting CoV proteins

During the coronavirus replication, the replicase polyproteins translated from ORFs undergo post-translational modification by two viral proteases which reside within the polyproteins called chymotrypsin-like protease ($3CL^{pro}$) and papain-like protease (PL^{pro}). This proteolytic cleavage is very important to release the NSPs from the polyproteins and initiate the viral replication-transcription-complex [21]. **$3CL^{pro}$** is a highly conserved protein among human coronaviruses and non-homologous to humans, which made it an interesting target to inhibit and effectively block viral RNA replication [35]. Based on the chemical structure, the $3CL^{pro}$ inhibitors can be divided into peptide and non-peptidomimetic inhibitors. **Peptidomimetic inhibitors** can naturally bind to the $3CL^{pro}$ forming an initial non-covalent bond. The reactive group also called the warhead, mainly consists of Micheal receptors, aldehydes, and ketones that covalently bind to the cysteine residues in the $3CL^{pro}$ and exhibit an inhibitory effect [35]. During the last decade, several reports have been published which showed the efficacy of many peptidomimetic inhibitors, especially the aldehyde and ketone groups to perform a good inhibitory activity against SARS-CoV, SARS-CoV-2, and MERS-CoV replication in-vitro [35]. Recently, a major bioavailable oral SARS-CoV-2 protease inhibitor called **Nirmatrelvir / Ritonavir (**developed by Pfizer and sold under the brand name - **Paxlovid**) has been conditionally authorized on December 2021, for the treatment of COVID-19 patients in the United Kingdom. Thereafter, in early 2022, this drug has been approved by Food and Drug Administration (FDA) for use in the USA and by European Medicines Agency (EMA) for use in the European Union [36]. Nirmatrelvir is a peptidomimetic inhibitor that acts on the SARS-CoV-2 main protease and blocks it reversibly for further polyprotein processing and viral replication [37]. Ritonavir is a human immunodeficiency virus type1 (HIV-1) protease inhibitor and it prevents the metabolism of nirmatrelvir by the inhibition of cytochrome P450 3A4 (CYP3A4) and increases the bioavailabilty of nirmatrelvir [37, 38].

The coronavirus **nsp13** is a highly conserved non-structural protein, that functions as a helicase to unwind the double-stranded RNA (ds RNA) and DNA (ds DNA) in a 5' – 3' direction for the effective translation of viral proteins, nucleic acid separations in the highly stable viral RNA secondary structures. This feature enabled the viral helicases as an interesting anti-viral target [39]. A non-

competitive inhibitor of nsp13, **SSYA10-001** inhibits the unwinding activity of nsp13 by interfering with its conformational changes during the viral genome replication [40, 41].

Viral genome replication is the core step involved in viral infection. Coronaviruses utilize an RNA-dependent-RNA polymerase (RdRp) enzyme activity which is highly conserved in the **nsp12** to transcribe full-length negative and positive strand viral RNAs, a set of subgenomic viral mRNAs, and set of subgenomic negative-strand RNAs. Nsp12 has two domains: one with nucleotidyl transferase activity and the canonical RdRp domain. This crucial role of nsp12 paved the development of nucleoside and nucleotide analogs to inhibit the polymerase activity and viral RNA synthesis [42]. These inhibitors are the analogs of nucleic acids with a chemical modification of either a sugar moiety or the heterocyclic ring that incorporates mismatching nucleotides in the replicating viral genome disrupting the transcription by chain termination [43].

Ribavirin (1-ß-D-ribofuranosyl-1, 2, 4-triazole-3-carboxamide) is a synthetic nucleoside with a guanosine analog, which when interacts with the host cell inosine monophosphate dehydrogenase (IMPDH) to inhibit the production of guanosine triphosphate (GTP). The depletion of cellular GTP pools leads to poor incorporation of this nucleotide during the viral RNA transcription and thereby blocking viral replication. It is also reported to directly inhibit the viral polymerases and increase the mutation rate in the viral RNA genome by incorporation of its triphosphate form [44]. High doses of ribavirin treatment partially inhibited the replication of SARS-CoV and MERS-CoV in-vitro. Due to its inhibitory efficacy in other RNA viruses in-vitro, Ribavirin was proposed to be used as an emergency drug during the SARS-CoV-2 outbreak [43].

Favipiravir (6-Fluoro-3-oxo-3,4-dihydropyrazine-2-carboxamide) is another purine-based synthetic analog that is activated by intracellular phosphoribosylation and it is incorporated into viral RNA genome by the viral polymerase leading to mutagenesis and chain termination [45]. It has been reported to show a broad-spectrum anti-viral activity against RNA viruses. Recently, a high concentration of favipiravir (EC_{50}: 61.88 μM) has been reported to effectively inhibit the replication of SARS-CoV-2 in Vero cells [46].

Remdesivir (GS-5734) is a prodrug with an adenosine analog (GS-441524) that is converted into active metabolite by the intracellular enzymes to compete with the host Adenosine triphosphates (ATPs) to be incorporated into the active transcribing viral RNA genome by the viral RdRp. This incorporation results in the halting of viral RNA synthesis after three more nucleotides are added to the newly synthesized strand [47]. Remdesivir was first developed to tackle the Ebola virus outbreak and further, it was used as a broad-range RNA virus inhibitor. In-vitro studies of SARS-CoV, MERS-CoV, and SARS-CoV-2 in human airway epithelial cells demonstrated a strong anti-viral activity of Remdesivir. In-vivo studies in transgenic mice and rhesus monkeys treated either prophylactic or after being infected with SARS-CoV and MERS-CoV showed better survival of infected mice and reduced viral titer in

lungs [47, 48]. Remdesivir was the first drug approved by the FDA in the USA under emergency usage for treating young and adult patients with severe COVID-19. However, the clinical trials that were conducted after the initial set of administration showed increased clinical recovery and no promising improvement in the mortality rate of the patients with severe COVID-19 which in turn created a mixed outcome in the efficacy and further administration of this drug [49].

Molnupiravir is a prodrug that is cleaved by the host esterases into an active nucleoside analog ß-D-N4-hydroxycytidine (NHC or EIDD-1931) [37]. The mechanism of action is similar to remdesivir, as the active nucleoside analog of molnupiravir is converted to NHC-triphosphates by the host kinases and thereby competitively inhibits cytidine and uridine for its incorporation into the viral RNA by the action of viral RdRp [37, 50]. This action by RdRp generates a mutated copy of viral RNA resulting in the further formation of mutated transcripts and a halt in the viral protein translation [50]. Molnupiravir was initially developed as an anti-influenza drug and later to treat multiple RNA viruses (including coronaviruses and encephalitic alphaviruses) infections owing to its efficacy and bioavailability [51]. Similar to paxlovid, it is the second oral drug that has been approved by FDA and EMA for emergency use authorization (EUA) in mild to severe COVID-19 patients [37].

2.3.2 Inhibition of coronavirus infection by targeting host cell proteins

Host cellular proteins play an important role in the coronavirus infection at multiple steps including viral entry, viral genome replication, and viral shedding. Angiotensin-converting enzyme-2 (**ACE2**) is a carboxypeptidase that cleaves polypeptides from the renal-angiotensin system and is an entry receptor for SARS-CoV, and SARS-CoV-2 [52]. **Chloroquine** is a medication primarily used to treat malaria and it was repurposed after the SARS-CoV outbreak in 2003. Chloroquine acts by increasing the endosomal pH which is required by the virus for entry / fusion. Moreover, it interferes with the glycosylation of the ACE2 receptor disrupting its availability for the virus to attach. Recent reports have shown that chloroquine at lower concentrations could inhibit pre and post-SARS-CoV-2 entry and infection in-vitro [52, 53].

Transmembrane protease serine 2 (**TMPRSS2**) is an important protein widely expressed in the respiratory tract which is responsible for the cleavage and activation of the S2 subunit of spike protein in SARS-CoV and MERS-CoV for the viral entry / fusion and later spread of the infection to other cells [54]. A serine-protease inhibitor, **Camostat mesylate** effectively inhibits the enzyme activity of TMPRSS2 and has been reported to effectively inhibit the SARS-CoV, SARS-CoV-2, and MERS-CoV infection in the cell culture systems [55, 56].

Cyclophilins (**Cyps**) are highly conserved intracellular proteins with peptidyl-prolyl cis-trans isomerase (PPIase) activity. There are seven types of human cyclophilins (CypA-E, Cyp40, and CypNK), and are shown to be involved in the replication of many RNA viruses. CypA which plays an important role in intracellular protein synthesis, folding, trafficking, and immunomodulation has been shown to closely

interact with SARS-CoV N protein and Nsp1 for its effective viral replication [57, 58]. **Cyclosporin A** (**CsA**) is a non-toxic immunosuppressant that blocks the formation of Cyps and calcineurin (CaN) leading to the inhibition of Nuclear Factor of Activated T cells (NFAT) factor to regulate the transcription of expression of the cytokine interleukin-2 (IL-2) [58]. CsA has been reported to effectively inhibit the replication of coronaviruses including SARS-CoV, MERS-CoV, HCoV-229E, and HCoV-NL63 at lower concentrations in the cell culture experiments [57, 59]. Recently, a non-immunosuppressive cyclophilin inhibitor **alisporivir** (Debio 025) has been used in the treatment of COVID-19 since CsA aggravates the condition of patients due to its strong immunosuppressive properties. Alisporivir has been reported to exhibit strong dose-dependent anti-viral properties against SARS-CoV-2 in Vero E6 cells by reducing the viral RNA and was also shown to specifically inhibit the post-entry step of SARS-CoV-2 [60].

Emerging pieces of evidence have highlighted the role of unfolded protein response (UPR) by endoplasmic reticulum stress (ER stress) in controlling viral infections [61]. Recent studies have shed light on the potential of **thapsigargin** (Tg), a small molecule extracted from a plant *Thaspsia garganica* known to non-competitively inhibit sarcoplasmic / endoplasmic reticulum **Ca^{2+} ATPase pump** as a broad-spectrum anti-viral agent [62]. Tg has been shown to elicit a robust host innate immune response that effectively blocks the replication of influenza A virus, respiratory syncytial virus (RSV), and coronaviruses (HCoV-OC43, and SARS-CoV-2) in both immortalized and primary human cells [63]. Further evidence has revealed the intricate mechanisms underlying the Tg anti-viral activity. In a recent publication by Shaban et al. and his colleagues, Tg was demonstrated to effectively inhibit three coronaviruses (HCoV-229E, MERS-CoV, and SARS-CoV-2) in multiple cell types at lower nanomolar concentrations [62, 64]. Shaban et el. has shown that Tg reverses the virus-induced translational shutdown, enhances the metabolic state and cell viability of infected cells, and upregulates proteins (BiP, HERPUD1, CTH, and IRE1α) involved in ER-associated protein degradation (ERAD), ER-quality control (ERQC), and intracellular membrane fusion / rearrangements [62, 64]. Additionally, Shaban et al. has shown that Tg disrupts the coronavirus-induced selective autophagy flux by blocking the autophagosomes with lysosomes in the late phase of autophagy pathway [62, 64]. These findings suggest that Tg hold a promise as a potential therapeutic option to complement the exisiting anti-viral strategies against coronaviruses.

2.4 Genome editing using CRISPR-Cas9 system

Recent advancements in the field of genome engineering have led to the manipulation and study of DNA sequences and their functional products endogenously in any organism of choice. Genetic perturbations in recent days have eased researchers to link the commonalities between biological phenotypes and pathogenesis. In the past two decades, the development of cutting-edge technology to precisely cut the genome by specialized nucleases has outcasted the use of Zinc finger nucleases, and

Transcription activator-like effector nuclease (TALEN) [65]. CRISPR stands for clustered regularly interspaced short palindromic repeats which were described first in 1987 by Japanese researchers as a series of short sequences in *Escherichia coli* (*E.Coli*), later received special attention in 2005 when three research groups reported it as the spacer sequences derived from plasmids and viral genomes [66]. CRISPR is an adaptive immunity possessed by bacteria and archaea against invading phages and foreign genetic material. The CRISPR loci naturally consist of 20-23 nucleotide repeated elements (CRISPR repeats), variable sequences (CRISPR spacers), and CRISPR-associated (Cas) genes. In the natural environment, CRISPR-Cas immunity works in three steps. In the first step (**adaptation stage**) the short sequences homologous to viral DNA or plasmid sequences are integrated into the CRISPR loci. These short spacers are segregated by CRISPR repeats for the system to recognize between the self and non-self genetic elements. In the next step (**CRISPR RNA biogenesis**) a long primary transcript of the CRISPR locus (pre-crRNA) is generated and further processed into short crRNAs. The processing of crRNA is catalyzed by Cas enzymes. In the third step (**Interference**), viral DNA or RNA is targeted by the crRNA accompanied by the Cas enzyme complex to complement the viral sequences and it is cleaved by the endonuclease activity of the Cas enzymes [65-67].

CRISPR-Cas system consists of two main classes and six types, which comprise a variety of Cas enzymes differing in their endonuclease activity. Cas9 enzyme belongs to the type-II CRISPR system that functions with a trans-activating crRNA (tracrRNA) for the crRNA maturation and crRNA:tracrRNA mediated endonuclease activity was first reported in *Streptococcus pyogenes* as a sequence-specific immunity against parasite genome. This remarkable finding led to the development of synthetic single-guide RNA (sgRNA) which consists of the crRNA and tracrRNA along with Cas9 to mediate a site-specific cleavage in the eukaryotes. CRISPR-Cas9-based DNA target recognition requires a 2-5 nucleotide protospacer adjacent motif (PAM) downstream of the target sequence [68-70].

Cas9 enzymes are characterized by two functional nuclease domains: RuvC-like nuclease domain is responsible for the cleavage of the DNA strand complementary to the target strand (non-target strand) and the HNH-like nuclease domain is responsible for the cleavage of the targeting strand complementing the sgRNA [71]. The Cas9 enzyme is in the inactive state (the apoenzyme) in the absence of guide RNA and has two distinct lobes, the alpha-helical recognition (REC) lobe and the nuclease (NUC) lobe containing the conserved HNH and RuvC domains as well as the C-terminal domain (CTD). The CTD contains PAM-interacting sites required for PAM search and recognition. Once the Cas9 enzyme binds to the sgRNA, the complex is active and starts the search for the complementary target DNA sequences. The PAM sequence plays an important role in the target sequence search and the sgRNA complementing. The PAM sequence for the *S. pyogenes*-derived Cas9 is 5'-NGG-3', where N denotes any of the four DNA bases. After a successful target recognition with an appropriate PAM sequence by the sgRNA-Cas9 enzyme complex, the enzyme melts the bases upstream of the PAM and allows the sgRNA to complement the target sequence. If the sgRNA

completely pairs with the target sequence, the RuvC, and HNH domains will produce a double-strand break (DSB) after the third nucleotide base upstream of the PAM [72, 73].

Upon cleavage by the Cas9 enzyme, the target locus typically undergoes one of the two major DNA damage repair pathways: the non-homologous end joining (NHEJ) or the high-fidelity homology-directed repair (HDR) pathway [74]. In mammalians, the DSBs are often repaired by NHEJ and the classical NHEJ is an error-prone double-strand break repair pathway that usually follows a fairly complicated mechanism. In the first step, the DSBs are recognized by the heterodimeric protein Ku which is composed of two polypeptides namely Ku70 / Ku80 and binds to the ends of dsDNA at the break site. The Ku heterodimer serves as a scaffold to recruit other NHEJ factors importantly, DNA-dependent protein kinase (DNA-PKcs) at the break site to form a synapse which serves as a bridge to hold the two ends of the DNA. If the ends are compatible, the XRCC4 / DNA ligase IV complex completes the NHEJ process by joining the ends of dsDNA. During this process of end-joining, there is usually an insertion or deletion of a few nucleotide bases at the site of DSB called INDEL (insertion / deletion) mutations which subsequently alter the coding sequence in the ORF thereby causing a frame-shift mutation and a premature stop codon for the gene product to be knocked out [75]. HDR on the other hand is an alternative and precise DNA repair pathway. HDR usually occurs at lower frequencies compared to NHEJ, as it is active during the late S and G2 phases of the cell cycle. The DSB is repaired by homologous recombination using a DNA template which is usually provided by the sister chromatids during mitosis. Additionally, an exogenous repair template homologous to the sequences of DSB can be in the form of single-stranded synthetic oligonucleotides (ssODNs) that can be delivered in a cell to insert or change the sequences of a target gene [76].

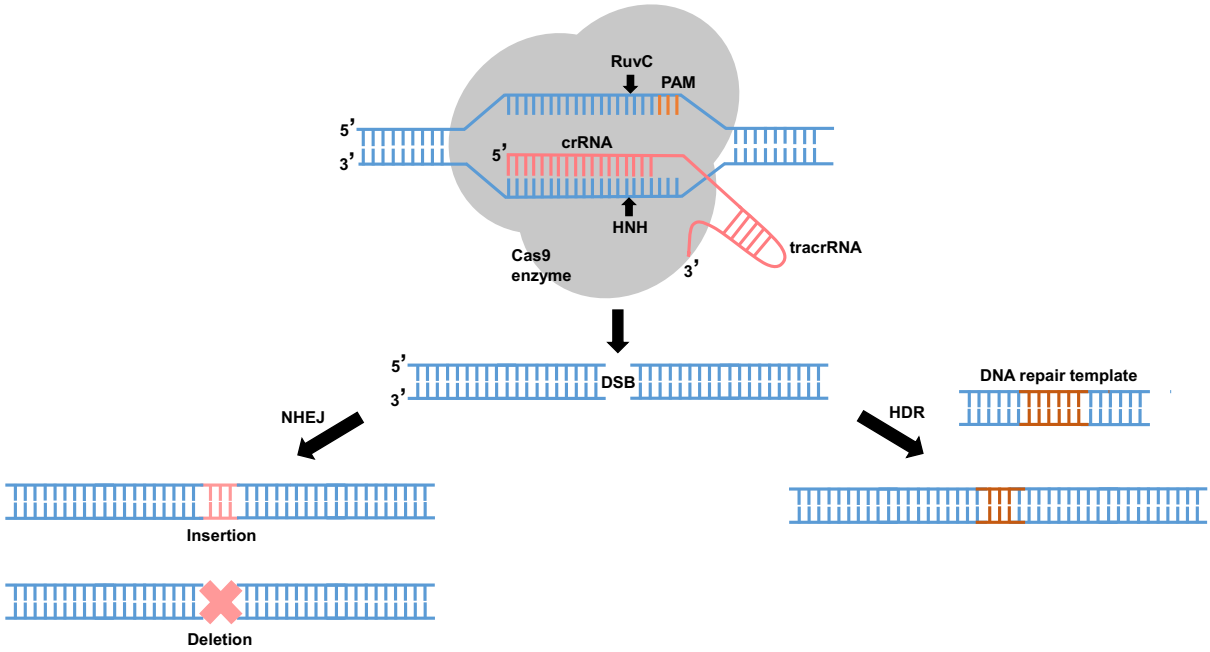


Figure 3: Mechanism of gene editing by CRISPR-Cas9 system.
Cas9 nuclease enzyme (grey) from *S.pyogenes* is targeted to the specific genomic DNA site by the sgRNA consisting of 20 nt guide sequence (pink) and the guide-RNA scaffold (pink). The guide sequence is paired to the DNA target (blue bottom strand) that is directly upstream of a 5'-NGG adjacent motif (PAM) (orange). The Cas9 enzyme using the two nuclease domains: RuvC and HNH, cleaves the DNA by producing a double-strand break (DSB). The DSB can be repaired in one of the two ways as illustrated above. The error-prone non-homologous end-joining (NHEJ) pathway process the ends by endogenous DNA repair machinery and results in the addition or deletion of nucleotides (INDELs). INDEL mutations occurring within the coding region of a gene result in a frame-shift or generation of a premature stop codon, ultimately leading to a gene knock out. Alternatively, the DSBs can be repaired by supplementing the DNA repair template (brown) that is homologous to the cut ends and this results in high fidelity and precise editing. crRNA-CRISPR RNA, RuvC- RuvC-like nuclease domain, HNH-HNH-like nuclease domain, tracrRNA-transacting crRNA, HDR-Homology direct repair pathway. This figure was modified from the source [74] (Fig. 2 from the publication) and re-created using Microsoft PowerPoint for this thesis.

2.5 Advancements in the CRISPR-Cas9 genome editing

In recent years, CRISPR-Cas9 systems have advanced beyond gene editing to offer sequence-specific gene regulation by modifying the nuclease activity of Cas9 which was initially designed to precisely bind and cleave the DNA by the guided sgRNA. Cas9 endonuclease activity was modified by mutating the RuvC and HNH domains to produce a catalytically inactive enzyme termed dead-Cas9 (dCas9). The dCas9 lacks the ability to cleave the DNA but retains the sequence-specific binding to the DNA when guided by the sgRNA. The dCas9 along with the sgRNA complex can be used to interfere with the transcription elongation process by blocking the RNA polymerase II (pol II) activity. Furthermore, the binding of dCas9 can block the recruitment of transcription initiation factors and perturb the expression of genes [77]. The repurposing of Cas9 to interfere with the gene expression was termed CRISPR interference (CRISPRi) and further advanced with the combination of dCas9 with transcription-repressors such as Krüppel-associated box (KRAB), and mSin3 interaction domains (SID4X) for efficient repression of endogenous genes in mammalian cells [78]. Another interesting technology

emerged from dCas9 for enhancing the transcription levels of endogenous genes called CRISPR activation (CRISPRa). This tool combines the dCas9 enzyme with the transcriptional activators such as VP64 or the p65 activation domain (p65AD) along with the sgRNA to activate the transcription levels of targeted genes. Current technology improved the efficacy of transcription activation by synergistic activation mediator (SAM) complex which includes the sgRNA with two stem-loops that can bind to MS2 bacteriophage coat proteins and activation domains of p65 and human heat-shock factor 1 (HSF1) as additional activator factors [79, 80].

2.5.1 Genome-wide CRISPR-Cas9 knock out screen

The robustness of targeting any genomic loci by the sgRNA-guided Cas9 has led the researchers in early 2014 to design a sgRNA library consisting of ten to many thousands of unique sgRNA sequences targeting either a set of a functional genome or the whole genome to either activate or knock out and study the physiological changes involved by genetic perturbations. The fundamental idea behind the genome-wide CRISPR-Cas9 knockout (GeCKO) screening is to target the entire genome by the sgRNA library and ensuing a single knock out per cell in a pool of mixed knock out cell populations. The pooled knock out cell population is then screened for a selection pressure (virus challenge, drug toxicity, and tumor transformation), resulting in an enormous cell depletion or cell survival. Based on the predominance, the sgRNAs in the depleted or surviving cells can be analyzed by deep sequencing by next-generation sequencing (NGS) to identify the genes involved in the phenotype of interest. Although the technology seems to be a straightforward approach, this unbiased screening involves a set of prerequisites to be well-optimized and organized to get the desired results at the end of the screen. A few practical considerations which were also important in this thesis are introduced in the upcoming sections.

2.5.1.1 Genome-wide sgRNA library

The foremost consideration for the setup of a genome-wide CRISPR-Cas9 screen is to choose the sgRNA library based on the types of gene modifications. In this context, there are three types of GeCKO libraries used for gene knock out (KO library), repression (CRISPRi library), and activation (CRISPRa library). In this thesis, it was intended to determine the host factors involved in the CoV replication and virus-induced cell death, and thus it was necessary to use the loss-of-function of genes to determine the potential host factor which could resist the infection and might help the cells to cope better with the survival. Currently, there are many knock out sgRNA libraries available that can be used to perform a genome-wide CRISPR screen. The most common and widely used sgRNA library was synthesized by Zhang *et al*. from the Broad Institute of MIT and Harvard, Cambridge, USA. In 2014, they designed a sgRNA library (GeCKO V1) targeting 18,080 human genes with 64,751 unique guide sequences and used it to identify genes essential for cell viability in tumor cell models [81]. This library was further improved to have a higher knock out efficiency by a high-throughput algorithm to target 5’

constitutively expressed exons. This **GeCKO V2** sgRNA library consists of 123,411 unique sgRNAs targeting 19,050 human genes and has 6 sgRNAs targeting each gene that is split into two half-libraries. Each library consists of 1000 sgRNAs targeting the intron region in the genome (non-targeting controls) and 4 sgRNAs targeting 1,864 micro-RNAs [82]. The Sabatini / Lander lab sgRNA library is also widely used for in-vitro screening and it is split into three sub-libraries consisting of 187,535 unique sgRNAs targeting 18,663 human genes (10 sgRNAs for each gene) [83]. Another common sgRNA library that is widely used is designed and synthesized by David Root / John Doench lab consisting of 76,441 unique sgRNAs targeting 19,114 human genes (4 sgRNAs for each gene) [84]. These libraries can be used depending on the cost and feasibility of the screen. GeCKO V2 and Sabatini library allow greater statistical certainty for hits by the use of (6-10 sgRNAs per gene) that emerge from the screening readout. On the other hand, the use of smaller sgRNA libraries would require fewer cells and reduce the cost compared to screening a larger library.

In recent years, many labs have developed sub-pool libraries targeting the genes with specific functions related to kinase, cell cycle, and transcription factor activity. These are called focussed sgRNA libraries and thus allow the user to screen fewer cells and narrow the hypothesis to identify kinases and phosphatases that are required for cell transformation in tumor development for instance. It also provides a higher coverage for each sgRNA with a smaller-sized library resulting in improved data output [85]. These sgRNA libraries are pre-made, deposited, and readily available at the company addgene (https://www.addgene.org/crispr/libraries/).

2.5.1.2 Delivery methods for sgRNA library screening

The next important consideration is to choose a delivery system that can efficiently transfer and integrate the sgRNA library into the cell models. Lentivirus (LV)-based plasmids have been widely used to deliver the sgRNA library, as they (i) can be titred to control the number of sgRNA copies per cell which is a very crucial factor in genome-wide screening and (ii) have been shown to stably integrate into the host cell genome. Lentiviral vectors, derived from the human immunodeficiency virus (HIV) require three essential genes to function. The *gag* gene encodes structural proteins, the *pol* gene encodes the enzymes that are required for reverse transcription and integration into the host cell genome, and the *env* gene encodes the viral envelope protein [86]. The first generation lentivirus vectors consisted of three separate plasmids: 1) the transfer plasmid encoding the genome with 5' and 3' Long-terminal repeats (LTR), the psi (ψ) sequence, and the rev response element (RRE), 2) the plasmid encoding the viral envelope protein, 3) the plasmid encoding the packaging elements (*gag*, *pol*, and HIV regulatory elements such as *vif*, *vpu*, *Nef*). However, the first-generation plasmids posed a safety concern with the generation of replication-competent lentiviruses (RCL). The second-generation lentivirus plasmids have been developed by removing the regulatory elements (*vif*, *vpu*, *Nef*) in the packaging plasmid without affecting the efficacy of genome transfer. The third-generation plasmids have further increased

biosafety and reduced the possibility of RCL production by splitting the packaging plasmids into two: one encoding the *gag*, *pol,* and the other encoding *rev* element [87, 88].

Currently, the sgRNA library is widely cloned into a lentiviral backbone consisting of Cas9 enzyme and an antibiotic resistance marker called lentiCRISPR V2 plasmid. The lentiCRISPR V2 is a second-generation LV plasmid that can be combined with both second and third-generation packaging and envelope plasmids to produce the lentivirus by transfecting the plasmids in Human embryonic kidney cells (HEK 293). The lentiCRISPR V2 is available as a single-vector system that encodes the sgRNA sequence, the Cas9 enzyme, and the puromycin-resistant cassette in a single plasmid. Additionally, the dual-vector system separates the Cas9 enzyme and the sgRNA encoding plasmid to achieve higher titers compared to the single-vector system. However, this system works only with pre-transduced cells expressing the Cas9 enzyme [89].

2.5.1.3 Choice of the cell line for creating an optimal pool of genome-wide knock out cells

The choice of cell line greatly influences the output of the genome-wide screen depending on the copy number of genes. Primary cells, organoids, and mouse models often limit the ease and scale of genetic perturbations compared to the tumor-derived cell lines. Genetic manipulation works much more efficiently in the haploid or diploid cell lines, which makes the genetic screens yield better data quality in these cells. The tumor-derived cell lines often carry more than two copies of a gene. In some cases, it is unavoidable to perform the genetic screens in the aneuploid cell lines. In such an instance, the genetic screens can be performed in a hyperdiploid cell line and can be compared with a haploid or diploid cell line. Tumor-derived models give the flexibility to scale-up or scale-down the genetic screen based on the biological questions [90-92].

The lentiviral systems often utilize a resistance gene to a drug-selectable marker (such as puromycin, blasticidin, hygromycin, and neomycin) to select the transduced cells from the untransduced ones. Each drug has a different kinetic for the selection of cells and it is also based on the susceptibility of the cell line to the drug. For instance, puromycin kills the non-resistant cells much more quickly than neomycin or blasticidin. It is very important to optimize the concentration of the drug in the cell line of interest before the screening, as a less concentrated drug would result in untransduced cells lingering in the cell pool, whereas a higher concentration of the drug would be too toxic to the transduced cells [91].

Another consideration is the transduction efficiency of a cell line, as this can greatly make impact the success rate of genetic modifications at a genome-scale. The cell line should possess the ability to be infected by the lentiviruses and choosing the cell line with lower transduction efficiency can harness the adequate representation of a genome-wide knock out cell pool. Lentiviral transductions for the generation of GeCKO cells are typically performed at a lower multiplicity of infection (MOI) (usually between 0.3 to 0.5) to ensure that each cell is transduced with a single copy of the sgRNA plasmid only.

It is important to titer the lentivirus in the desired cell line to ensure a lower MOI. Therefore, the amount of virus needed to achieve a lower MOI is primarily dependent on the transduction efficiency of the cell line used [90].

2.5.1.4 Choice of the screen selection pressure

After the generation of GeCKO cells, there are two ways to select the cells with an intended phenotype: positive and negative-screen selection. The positive-screen selection aims to identify the knock out cells that survive the post-selection process. In a positive-screen selection, the GeCKO cells are challenged with a desired selection pressure, such as drug treatment, toxic substances, or virus infection, that potentially leads to cell death under normal cell-type conditions. However, some of the genetically perturbed cells can resist and outgrow the selection pressure and confer resistance to the type of challenge given to the cells. The hypothesis is that some mutations are essential for the cells to survive and proliferate under adverse environments. At the end of the positive-screen selection, the vast majority of the GeCKO cells are lost from the cell pool and the surviving cells contain the information of sgRNAs targeting the genes which confer the resistance to the pressure. It is very important in this type of screen selection to culture the cells long enough after the post-selection, to enrich the sgRNA-integrated cells before analyzing these cells by deep-sequencing by NGS [93-95].

In contrast, the negative-screen selection aims to identify the knock out cells that are sensitive to the selection pressure. This type of screen is a little more complicated than the positive-screen selection since the sgRNAs having a null or actual effect will be lost from the cell population under the pressure. Therefore, the screen is carried out with a control population of GeCKO cells with no selection pressure along with another population of GeCKO cells that are challenged with a desired pressure. At the end of selection pressure, there are sgRNAs in both the populations which are over-represented, however, the lost sgRNAs in the selected populations can then be compared with the control population to identify the sensitive genes [93-95].

For both types of screen selection, it is very important to screen the cells with adequate coverage of sgRNAs per gene. For a positive-screen selection, a minimum coverage of 200-500-fold per sgRNA is required to achieve a higher statistical read-out and to avoid false-positive results at the end of the screen. In such a case, if the screen is conducted by using the GeCKO V2 library of 123,411 sgRNAs, a minimum of 50 million GeCKO cells should be screened. Negative-screen selection requires much higher coverage of sgRNAs of about 500-1,000-fold, to detect the subtle changes in the sgRNA dropout [93].

2.5.1.5 Computational tools to analyze the screen and identify the hits

After screen selection and ferreting out the survivor cells, the enriched sgRNAs that are integrated into the surviving cells are identified by amplifying the 20 nt regions by a polymerase chain reaction (PCR). The amplified libraries encoding typically Illumina barcodes are then deep sequenced by NGS. The level of sequencing depth is an important consideration to achieve a high-quality data output in the analysis. For a positive screen with a strong selection pressure, useful results can be obtained by sequencing the samples at a depth of 2-5*10^7 reads for a complex library of >100,000 sgRNA plasmids. On the other hand, the negative-screen selection may require a much higher sequencing depth of about 1-5*10^8 reads. After deep sequencing, it is also important to have the quality control (QC) metrics of the sequenced raw reads to visualize the sequencing depth of the samples, variability between the replicates, and sgRNA library representation in the controls and the selected samples. The QC metrics provide a strong foundation for the downstream analysis and therefore, careful consideration should be provided during the initial analysis of the sequenced data [90].

In recent years, many benchmark computational tools are developed to analyze genome-wide screens. Each tool works on a different algorithm to provide a high level of data analysis which can be used to identify genes and related pathways that arise from the genetic screens. Some algorithms directly work from the raw reads obtained from the deep sequencing to perform a systematic analysis to generate read counts to gene list tables, while others generally rely on the other existing tools for read counts to obtain the differential changes in the sgRNA representation. A few algorithms which are currently been used for the analysis of genetic screens are listed below [96].

RNA interference (RNAi) gene enrichment ranking (**RIGER**) is based on a statistical analysis methodology, initially designed for large-scale RNAi screens. The RIGER algorithm is designed to determine the statistical score that models the probability of a gene hit based on the multiple siRNAs enrichment / depletion ($\log_2$ fold change) per gene. The algorithm used in the RIGER analysis is termed redundant siRNA activity (RSA) and it ranks the siRNA targeting the gene across the siRNA library and assigns a p-value that indicates the statistical significance. Based on the p-value and rank of siRNA targeting the gene, the RSA provides a ranked gene list table as an output [96-98].

Model-based analysis of genome-wide CRISPR-Cas9 knock out (**MAGeCK**) is the first program specifically designed for CRISPR-based screens and it is based on a robust ranking algorithm (RRA). Unlike RIGER, the RRA works by taking the raw reads, counting the sgRNA frequency, and normalizing the read counts across the samples through a median normalization method to avoid bias in the library size after sequencing. After the normalization method, the RRA applies a negative binomial model to test the significance of sgRNA enrichment / depletion of the experimental samples compared to the controls, which is similar to differential RNA-seq analysis (DESeq). The p-values obtained for each sgRNA targeting a gene are combined at the gene level by a modified robust ranking

aggregation algorithm to rank the genes and provide a statistical MAGeCK score. MAGeCK also uses RRA to provide an enriched pathway analysis report based on the gene ranking list [89, 96, 98, 99]. Bayesian analysis of gene essentiality (**BAGEL**) is a supervised learning algorithm for analyzing genes from the knock out screens. The algorithm is coupled to the list of reference essential and non-essential gene lists to first estimate the differential changes in the sgRNA targeting the reference list and estimates the likelihood of observed changes in the sgRNA fold changes for the remaining test genes. The algorithm provides a final score for each gene called the log Bayes factor that is derived from the ratio of fold change distributions of sgRNA targeting the test gene and the essential / non-essential training distributions. This tool only works with prior information on the essential and non-essential genes for the type of screen selection performed, which limits the broad-range application of this algorithm [96, 98, 100].

2.6 Aims of this study

Human coronavirus 229E is a member of the genus *Alphacoronavirus* associated with mild upper respiratory tract infections. The virus replicates in the host cell cytoplasm and triggers immunomodulatory and ER-associated stress responses by poorly characterized mechanisms. In addition, host factors play a crucial role in multiple stages of HCoV-229E replication. To gain a comprehensive understanding of the host cell factors that support HCoV-229E pathogenesis, the following objectives were aimed in this thesis:

1) to identify and validate the functional relevance of strongly upregulated genes in the HCoV-229E-infected Huh7 cells.

2) to uncover host cell factors required for HCoV-229E compared with MERS-CoV replication and virus-mediated cell death.

3) to analyze the genes that support the anti-viral effect of thapsigargin on HCoV-229E replication.

To this end, in the first part of the study, the genome-wide transcriptome data of HCoV-229E infected Huh7 cells was used to identify the most strongly regulated host cell factors and the functional relevance of the identified genes on the HCoV-229E replication (positively or negatively) was validated by genetic perturbation (by CRISPR-Cas9 knock outs).

In the second part of this study, an unbiased genome-wide knock out screen was used to identify the genes that support HCoV-229E and MERS-CoV replication. The functional relevance of the identified genes was validated in various stages of HCoV-229E replication, including viral RNA replication, budding, and virus-mediated cell death. To examine the biological relevance of the identified host factors, comparative analysis with MERS-CoV and other published genetic screens was perfomred to explore the shared / distinct genes that would provide insights on common and / or novel cellular pathways used by the coronaviruses.

Finally, the genetic screen was used to identify the genes that support thapsigargin-mediated inhibition of HCoV-229E replication. To identify a set of prominent host factors and to obtain a deep understanding of how thapsigargin (positively or negatively) regulates these genes to set out a strong anti-viral effect, the candidate list from the Tg genetic screen was intersected with the nascent proteomics data obtained from thapsigargin-treated infected cells.

3. Materials

3.1 Cell lines

Cell line	Origin/Generated by	Parental or Engineered
Huh7	Human hepatoma cells (Japanese Collection of Research Bioresources (JCRB) cell bank)	Parental, a kind gift from Prof. John Ziebuhr, Institute of Medical Virology, Giessen
HEK293FT	Human kidney cells (Thermo Fisher Scientific, R7000)	Parental
H9C2	Embryonic rat cardiomyocytes (ATCC CRL-1446)	Parental
Huh7 px459	Generated by B.Vincent Albert (B.V.A)	CRISPR-Cas9 engineered, for this study
ANKRD1_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ANKRD1_sg3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ANKRD1_sg1+2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ANKRD1_sg1+3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ERO1LB_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ERO1LB_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ERO1LB_sg1+2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ERO1LB_sg3+4	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ERO1LB_pool	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DEC1_sg4	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DEC1_sg5	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DEC1_sg4+5	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DEC2_sg4	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DEC2_sg5	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DEC2_sg4+5	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
FICD_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
FICD_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
FICD_sg1+2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
FICD_sg1+3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study

FICD_sg2+3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
EDEM1_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
EDEM1_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
EDEM1_sg3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
EDEM1_sg1+2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
EDEM1_sg2+3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
EDEM1_sg1+3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
CTH_sg1+2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
CTH_sg3+4	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
FUT1_sg2+3	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
KLF6_sg1+2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
GECKO_BIO1	Genome-wide CRISPR knock out Huh7 cell pool, by B.V.A	CRISPR-Cas9 engineered, for this study
GECKO_BIO2	Genome-wide CRISPR knock out Huh7 cell pool, by B.V.A	CRISPR-Cas9 engineered, for this study
GECKO_BIO3	Genome-wide CRISPR knock out Huh7 cell pool, by B.V.A	CRISPR-Cas9 engineered, for this study
Empty vector	Huh7 parental transduced with Lentiviral CRISPR V2 empty vector, by B.V.A	CRISPR-Cas9 engineered, for this study
ANPEP_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ANPEP_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
SFTA2_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
HDAC4_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
HDAC4_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ATG7_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ZDHHC3_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ZDHHC3_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
KIF3B	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
CTCFL	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
SYT3_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
SYT3_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study

STY16_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
SYT16_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ORMDL1_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ORMDL1_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
CPB2_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
CPB2_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ARFGAP3_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ARFGAP3_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
EMC6_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DDRGK1_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
DDRGK1_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
ERMP1_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
SEC24A_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
SEC24A_sg2	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
CLU_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
HMOX2_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
SDE2_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
TGFB1_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study
FAM50A_sg1	Generated by B.V.A	CRISPR-Cas9 engineered, for this study

3.2 Bacterial strains

Strain name	Source	Genotype
XL-1 Blue E.coli	Thermo Fisher Scientific	F′ Tn10(Tetr) proA +B + laclqΔ(lacZ)M15I recA1 gyrA96 (Nalr) tbi-1bsdR17 (r-k m-k) glnV44 relA1 lac
One Shot TOP10 chemical competent E.coli	Thermo Fisher Scientific	F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80LacZΔM15 Δ LacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (Str^R) endA1 nupG

3.3 Expression Vectors

Name of the vector	Source
pSpCas9(BB)-2A-Puro (pX459) V2.0	Addgene #48139
px459_ANKRD1_sg1	Cloned by B.V.A, for this study
px459_ANKRD1_sg2	Cloned by B.V.A, for this study
px459_ANKRD1_sg3	Cloned by B.V.A, for this study
px459_ERO1LB_sg1	Cloned by B.V.A, for this study
px459_ERO1LB_sg2	Cloned by B.V.A, for this study
px459_ERO1LB_sg3	Cloned by B.V.A, for this study
px459_ERO1LB_sg4	Cloned by B.V.A, for this study
px459_DEC1_sg4	Cloned by B.V.A, for this study
px459_DEC1_sg5	Cloned by B.V.A, for this study
px459_DEC2_sg4	Cloned by B.V.A, for this study
px459_DEC2_sg5	Cloned by B.V.A, for this study
px459_FICD_sg1	Cloned by B.V.A, for this study
px459_FICD_sg2	Cloned by B.V.A, for this study
px459_FICD_sg3	Cloned by B.V.A, for this study
px459_EDEM1_sg1	Cloned by B.V.A, for this study
px459_EDEM1_sg2	Cloned by B.V.A, for this study
px459_EDEM1_sg3	Cloned by B.V.A, for this study
px459_CTH_sg1	Cloned by B.V.A, for this study
px459_CTH_sg2	Cloned by B.V.A, for this study
px459_CTH_sg3	Cloned by B.V.A, for this study
px459_CTH_sg4	Cloned by B.V.A, for this study
px459_FUT1_sg2	Cloned by B.V.A, for this study
px459_FUT1_sg3	Cloned by B.V.A, for this study
px459_KLF6_sg1	Cloned by B.V.A, for this study
px459_KLF6_sg2	Cloned by B.V.A, for this study
GECKO plasmid Library A	Addgene #1000000048
GECKO plasmid Library B	Addgene #1000000048
GECKO plasmid Library A+B	Cloned by B.V.A, for this study
psPAX2	Addgene #12260
pMD2.G	Addgene #12259
LentiCRISPR V2 Empty vector (LCV2)	Addgene #52961
LCV2_ANPEP_sg1	Cloned by B.V.A, for this study
LCV2_ANPEP_sg2	Cloned by B.V.A, for this study
LCV2_SFTA2_sg1	Cloned by B.V.A, for this study
LCV2_HDAC4_sg1	Cloned by B.V.A, for this study
LCV2_HDAC4_sg2	Cloned by B.V.A, for this study
LCV2_ATG7_sg1	Cloned by B.V.A, for this study
LCV2_ZDHHC3_sg2	Cloned by B.V.A, for this study
LCV2_KIF3B	Cloned by B.V.A, for this study
LCV2_CTCFL	Cloned by B.V.A, for this study
LCV2_SYT3_sg1	Cloned by B.V.A, for this study
LCV2_SYT3_sg2	Cloned by B.V.A, for this study
LCV2_STY16_sg1	Cloned by B.V.A, for this study

LCV2_SYT16_sg2	Cloned by B.V.A, for this study
LCV2_ORMDL1_sg1	Cloned by B.V.A, for this study
LCV2_ORMDL1_sg2	Cloned by B.V.A, for this study
LCV2_CPB2_sg1	Cloned by B.V.A, for this study
LCV2_CPB2_sg2	Cloned by B.V.A, for this study
LCV2_ARFGAP3_sg1	Cloned by B.V.A, for this study
LCV2_ARFGAP3_sg2	Cloned by B.V.A, for this study
LCV2_EMC6_sg2	Cloned by B.V.A, for this study
LCV2_SEC24A_sg1	Cloned by B.V.A, for this study
LCV2_SEC24A_sg2	Cloned by B.V.A, for this study
LCV2_CLU_sg1	Cloned by B.V.A, for this study
LCV2_HMOX2_sg1	Cloned by B.V.A, for this study
LCV2_TGFB1_sg1	Cloned by B.V.A, for this study
LCV2_SDE2_sg1	Cloned by B.V.A, for this study
LCV2_FAM50A_sg1	Cloned by B.V.A, for this study
LCV2_DDRGK1_sg1	Cloned by B.V.A, for this study
LCV2_DDRGK1_sg2	Cloned by B.V.A, for this study
LCV2_ERMP1_sg1	Cloned by B.V.A, for this study

3.4 Oligonucleotides, primers, and vectors

All the oligonucleotides and primer pairs were synthesized from Eurofins Genomics at a synthesis scale of 0.01 μmol, except for sgRNA library preparation (0.05 μmol), and purified using High Purity Salt-Free (HPSF) technology. All the sequences are denoted in the direction of 5' – 3', se – denotes sense, and as – denotes anti-sense. "CACCG" at the start of the sense sequence and "AAAC" at the start & "C" at the end of the anti-sense sequence denotes the cloning overhangs.

Oligonucleotides for cloning ANKRD1 targeting sgRNAs into px459 vector

Name	Sense sequence	Anti-sense sequence
ANKRD1_sg1	CACCGGGCAGGGGAATTCCTTC CTG	AAACCAGGAAGGAATTCCCCTG CCC
ANKRD1_sg2	CACCGTGCTGTTACTTTAGAGA AGC	AAACGCTTCTCTAAAGTAACAG CAC
ANKRD1_sg3	CACCGTTTTCCACTGTTGCTCCC CC	AAACGGGGGAGCAACAGTGGA AAAC

Oligonucleotides for cloning ERO1LB targeting sgRNAs into px459 vector

Name	Sense sequence	Anti-sense sequence
ERO1LB_sg1	CACCGAGCCAAGGGGTCCGCCG GGC	AAACGCCCGGCGGACCCCTTGG CTC
ERO1LB _sg2	CACCGCGCCTGCCCGGCGGACC CCT	AAACAGGGGTCCGCCGGGCAGG CGC

ERO1LB _sg3	CACCGAACGCTCTGGGTTCAGC AAT	AAACATTGCTGAACCCAGAGCG TTC
ERO1LB _sg4	CACCGCTGAACCCAGAGCGTTA CAC	AAACGTGTAACGCTCTGGGTTC AGC

Oligonucleotides for cloning DEC1 and DEC2 targeting sgRNAs into px459 vector

Name	**Sense sequence**	**Anti-sense sequence**
DEC1_sg4	CACCGGTCCCGAGTGTTCTCGTGC T	AAACAGCACGAGAACACTCGGGA CC
DEC1_sg5	CACCGCGAGACCACCCGGTGGAG GT	AAACACCTCCACCGGGTGGTCTC GC
DEC2_sg4	CACCGATCGCCCATTCAGTCCGAC T	AAACAGTCGGACTGAATGGGCGA TC
DEC2_sg5	CACCGCGTGCAAGTGGTTGATCA GC	AAACGCTGATCAACCACTTGCAC GC

Oligonucleotides for cloning FICD targeting sgRNAs into px459 vector

Name	**Sense sequence**	**Anti-sense sequence**
FICD_sg1	CACCGACCGCCATCACTGAAGCC AT	AAACATGGCTTCAGTGATGGCGG TC
FICD_sg2	CACCGGGTCTGGAGCCGCTTCCTC T	AAACAGAGGAAGCGGCTCCAGAC CC
FICD_sg3	CACCGTGTCCGGTTTGCTCCTGAG C	AAACGCTCAGGAGCAAACCGGAC AC

Oligonucleotides for cloning EDEM1 targeting sgRNAs into px459 vector

Name	**Sense sequence**	**Anti-sense sequence**
EDEM1_sg1	CACCGGACGAGCGCTCGCCATTG CA	AAACTGCAATGGCGAGCGCTCGT CC
EDEM1_sg2	CACCGGGGCTGGTGCTCCTCCGG CT	AAACAGCCGGAGGAGCACCAGC CCC
EDEM1_sg3	CACCGTGTGGCTCGTCTTCGGGC TG	AAACCAGCCCGAAGACGAGCCA CAC

Oligonucleotides for cloning CTH targeting sgRNAs into px459 vector

Name	**Sense sequence**	**Anti-sense sequence**
CTH_sg1	CACCGTCGCCTGCGTGGCGAAATG T	AAACACATTTCGCCACGCAGGCG AC
CTH _sg2	CACCGCGCCACGCAGGCGATCCAT G	AAACCATGGATCGCCTGCGTGGCG C
CTH _sg3	CACCGGGTTTTGAATATAGCCGTT C	AAACGAACGGCTATATTCAAAAC CC

CTH _sg4	CACCGTTTTCAAGGCAATTCCTAG T	AAACACTAGGAATTGCCTTGAAA AC

Oligonucleotides for cloning FUT1 and KLF6 targeting sgRNAs into px459 vector

Name	**Sense sequence**	**Anti-sense sequence**
FUT1_sg2	CACCGGGTCTGGACACAGGATCG AC	AAACGTCGATCCTGTGTCCAGACC C
FUT1_sg3	CACCGTTGGGGCCCATCGCAGTA CC	AAACGGTACTGCGATGGGCCCCA AC
KLF6_sg1	CACCGACGTCCATGTCGGGCCGG GT	AAACACCCGGCCCGACATGGACG TC
KLF6_sg2	CACCGCCCCATGTGCAGCATCTTC C	AAACGGAAGATGCTGCACATGGG GC

Oligonucleotides for cloning sgRNAs targeting genes emerging as hits from the GECKO screens into LCV2 vector for follow-up studies

Name	**Sense sequence**	**Anti-sense sequence**
ANPEP_sg1	CACCGATGGCCGGCTCATCGAA GCA	AAACTGCTTCGATGAGCCGGCCA TC
ANPEP_sg2	CACCGGCTCATGTTTGACCGCT CCG	AAACCGGAGCGGTCAAACATGA GCC
SFTA2_sg1	CACCGTTCAGTTGCAAAGTCAT ACC	AAACGGTATGACTTTGCAACTGA AC
HDAC4_sg1	CACCGACTTACCCGTACCAGTA GCG	AAACCGCTACTGGTACGGGTAAG TC
HDAC4_sg2	CACCGATGAGCTCCCAAAGCC ATCC	AAACGGATGGCTTTGGGAGCTCA TC
ATG7_sg1	CACCGGGTGAACCTCAGTGAAT GTA	AAACTACATTCACTGAGGTTCAC CC
ZDHHC3_sg1	CACCGTAAACAGGACGAAGTA CTTC	AAACGAAGTACTTCGTCCTGTTT AC
ZDHHC3_sg2	CACCGAGTGGAATCCCACCATG ATG	AAACCATCATGGTGGGATTCCAC TC
CTCFL	CACCGTATGTTTATAGCGCCTG TGT	AAACACACAGGCGCTATAAACAT AC
KIF3B	CACCGACTTACTCCAGTCTGGT TAT	AAACATAACCAGACTGGAGTAA GTC
SYT3_sg1	CACCGATGACTTTGACCGCTTC TCG	AAACCGAGAAGCGGTCAAAGTC ATC
SYT3_sg2	CACCGTTCGCCCTGCGGTACCT CTA	AAACTAGAGGTACCGCAGGGCG AAC
SYT16_sg1	CACCGGATTCTCAGAATAACGC CGT	AAACACGGCGTTATTCTGAGAAT CC
SYT16_sg2	CACCGAGCCGTTCGTTGTTGAC ACA	AAACTGTGTCAACAACGAACGGC TC
ORMDL1_sg1	CACCGTTTGTTAAAGTCCAAGC AAC	AAACGTTGCTTGGACTTTAACAA AC

ORMDL1_sg2	CACCGACCTTTCGAAACTCCTGACC	AAACGGTCAGGAGTTTCGAAAGGTC
CPB2_sg1	CACCGATACTGTTCATAGTACGATG	AAACCATCGTACTATGAACAGTATC
CPB2_sg2	CACCGAAAGCCCATTTAAATGTGAG	AAACCTCACATTTAAATGGGCTTTC
ARFGAP3_sg1	CACCGGACGAAAAGATGAACATTAG	AAACCTAATGTTCATCTTTTCGTCC
ARFGAP3_sg2	CACCGATCATTACGATTAGCCTATA	AAACTATAGGCTAATCGTAATGATC
DDRGK1_sg1	CACCGTCTTCCACTCAGCTTCGCGC	AAACGCGCGAAGCTGAGTGGAAGAC
DDRGK1_sg2	CACCGTGTAGTTGATGAACTCTGTC	AAACGACAGAGTTCATCAACTACAC
ERMP1_sg1	CACCGTCTGACCGTGCACTACCTTT	AAACAAAGGTAGTGCACGGTCAGAC
EMC6_sg2	CACCGACGGCCGCCTCGCTGATGAA	AAACTTCATCAGCGAGGCGGCCGTC
SEC24A_sg1	CACCGGGGATGATGCACGAGGACAA	AAACTTGTCCTCGTGCATCATCCCC
SEC24A_sg2	CACCGCAATCTCGTCGTAACTACTG	AAACCAGTAGTTACGACGAGATTGC
CLU_sg1	CACCGTACGCACGCGTCTGCAGAAG	AAACCTTCTGCAGACGCGTGCGTAC
HMOX2_sg1	CACCGTCAGCACAGCCATAGCTGTT	AAACAACAGCTATGGCTGTGCTGAC
TGFB1_sg1	CACCGCGGGTTCAGGTACCGCTTCT	AAACAGAAGCGGTACCTGAACCCGC
SDE2_sg1	CACCGCTGTCTGTTTGAGATTTAGT	AAACACTAAATCTCAAACAGACAGC
FAM50A_sg1	CACCGGCAGATGAAGCAGCGCATCG	AAACCGATGCGCTGCTTCATCTGCC

Vector maps used in this thesis

pSpCas9(BB)-2A-Puro (pX459) V2.0

This vector is derived from the plasmid pX330 and the "pSpCas9(BB)" indicates the presence of the *Streptococcus pyogenes* Cas9 nuclease gene, which is the primary component of the CRISPR-Cas9 system. The "2A-Puro" refers to the presence of a self-cleaving 2A peptide sequence followed by a puromycin resistance gene [101]. The 2A peptide also allows for the co-expression of multiple genes from a single mRNA sequence and the puromycin resistance gene enables the selection of cells after a transient transfection of this vector through puromycin treatment [101]. The vector can be digested using the restriction enzyme BbsI and the annealed oligo pairs can be cloned into the single guide RNA scaffold.

LentiCRISPR V2 Empty vector (LCV2)

This vector is derived from lentiviruses, which belong to the retrovirus family known for their ability to efficiently deliver genetic material into a wide range of cell types. The plasmid contains two expression cassettes, hSpCas9 from the *Streptococcus pyogenes* and the chimeric guide RNA scaffold. The vector can be digested using the restriction enzyme *BsmBI* and the annealed sgRNA oligos can be cloned into the single guide RNA scaffold. The puromycin resistance gene enables the selection of cells after stable transduction of this vector through puromycin treatment [102].

psPAX2 and pMD2.G

psPAX2 is a packaging plasmid that provides the necessary components for the production of lentiviral particles. The plasmid contains a *gag-pol* element that encodes the structural and enzymatic proteins necessary for the assembly and packaging of lentiviruses. The *rev* gene encodes the Rev protein that is required for the export of unspliced and partially spliced viral RNA from the nucleus to the cytoplasm of the host cell. The *tat* gene encodes the Tat protein that is essential for viral RNA transcription and the *psi* element is required for the RNA packaging into newly formed lentiviral particles [86, 88].
pMD2.G is an envelope plasmid that provides the envelope glycoprotein (VSV-G) necessary for the production of pseudotyped lentiviral particles [86, 88]. When used together, these two plasmids provide the necessary components for lentiviral particle production (along with the transfer plasmid lentiCRISPR V2).

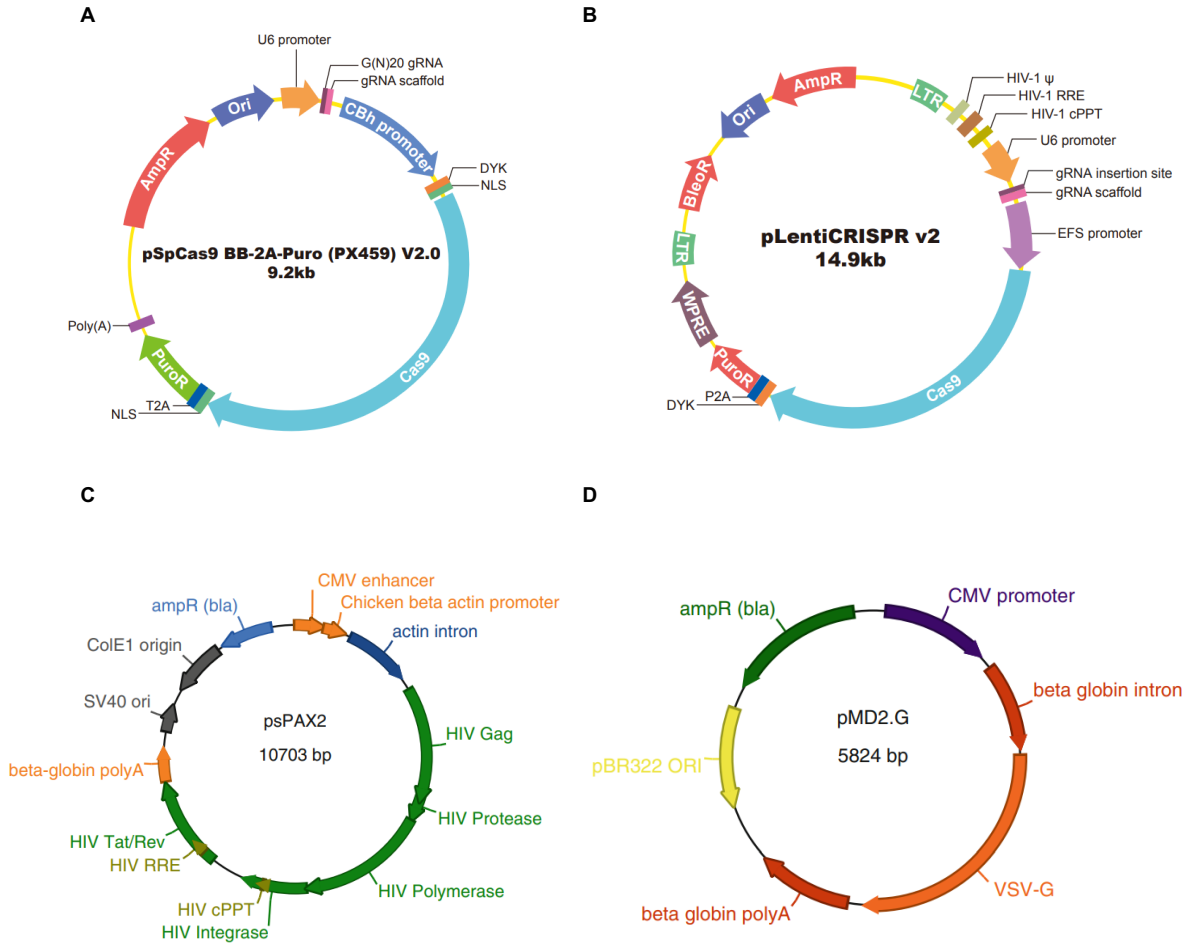


Figure 4: Vector maps used for sgRNA cloning and lentivirus production.
A) pSpCas9(BB)-2A-Puro (pX459) V2.0 is a single vector system for delivering sgRNA along with the guide RNA scaffold and Cas9 for efficient gene editing, along with 2A peptide sequence, and puromycin resistance gene. **B)** LentiCRISPR V2 is a lentivirus-based single vector system with an hSpCas9 gene, chimeric guide RNA scaffold, and puromycin resistance gene. **C)** and **D)** psPAX2 and pMD2.G are the vectors providing the components for lentivirus production, including the *gag-pol* element, *rev* gene, *tat* gene, and *psi* element (psPAX2), and envelope glycoprotein (VSV-G – pMD2.G). The vector maps A) and B) are obtained from GenScript company (https://www.genscript.com/). The vector maps B) and D) are obtained as a kind gift from Didier Trono lab (Addgene plasmid # 12260, # 12259).

Sequencing primer for the verification of sgRNA insertion into px459 vector

U6 primer_se: GCATATACGATACAAGGCTG

Sequencing primer for the verification of sgRNA insertion into LCV2 vector

hU6_se: GAGGGCCTATTTCCCATGATT

Sequencing primers for the verification of indel mutations in the knock out cell lines

Cell line	Sense sequence	Anti-sense sequence
ANKRD1_sg1+3	ACTGCTCAAAGGTGGCTGTT	TCCGTTGCTTGGGATTAACT
FUT1_sg2+3	TCCTCAGCCTCAGAGCATTT	CTCAAGTCCGCGTACTCCTC
EDEM1_sg2+3	GCTAGTTCCTGGTGCGTCA	AAGGCGTGAGCCATGTAGTT
FICD_sg2+3	GTTTCTCCTGCGACTCTTGG	GTAGTCCGCCTGGATGATGT

Primers designed for RT-qPCR

Target	Sense sequence	Anti-sense sequence
HCoV-229E S RNA	TTTCAGGTGATGCTCACATACC	ACAAACTCACGAACTGTCTTAG

Forward primers designed for sgRNA library preparation and Illumina sequencing

Name	Sequence
F1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTTAAGTAGAGGCTTTATATATCTTGTGGAAAGGACGAAACACC
F2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTATCATGCTTAGCTTTATATATCTTGTGGAAAGGACGAAACACC
F3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTGATGCACATCTGCTTTATATATCTTGTGGAAAGGACGAAACACC
F4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTCGATTGCTCGACGCTTTATATATCTTGTGGAAAGGACGAAACACC
F5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTTCGATAGCAATTCGCTTTATATATCTTGTGGAAAGGACGAAACACC
F6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTATCGATAGTTGCTTGCTTTATATATCTTGTGGAAAGGACGAAACACC
F7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTGATCGATCCAGTTAGGCTTTATATATCTTGTGGAAAGGACGAAACACC
F8	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTCGATCGATTTGAGCCTGCTTTATATATCTTGTGGAAAGGACGAAACAC C
F9	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTACGATCGATACACGATCGCTTTATATATCTTGTGGAAAGGACGAAACA CC

F10	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTTACGATCGATGGTCCAGAGCTTTATATATCTTGTGGAAAGGACGAAAC ACC

Reverse primers designed for sgRNA library preparation and Illumina sequencing

Name	Sequence	Barcode
R1	CAAGCAGAAGACGGCATACGAGATTCGCCTTGGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	TCGCCTTG
R2	CAAGCAGAAGACGGCATACGAGATATAGCGTCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	ATAGCGTC
R3	CAAGCAGAAGACGGCATACGAGATGAAGAAGTGTGACTGGA GTTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTT TTCAA	GAAGAAGT
R4	CAAGCAGAAGACGGCATACGAGATATTCTAGGGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	ATTCTAGG
R5	CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	CGTTACCA
R6	CAAGCAGAAGACGGCATACGAGATGTCTGATGGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	GTCTGATG
R7	CAAGCAGAAGACGGCATACGAGATTTACGCACGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	TTACGCAC
R8	CAAGCAGAAGACGGCATACGAGATTTGAATAGGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	TTGAATAG
R9	CAAGCAGAAGACGGCATACGAGATTCCTTGGTGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	TCCTTGGT
R10	CAAGCAGAAGACGGCATACGAGATACAGGTATGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	ACAGGTAT
R11	CAAGCAGAAGACGGCATACGAGATAGGTAAGGGTGACTGGA GTTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTT TTCAA	AGGTAAGG
R12	CAAGCAGAAGACGGCATACGAGATAACAATGGGTGACTGGA GTTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTT TTCAA	AACAATGG
R13	CAAGCAGAAGACGGCATACGAGATACTGTATCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	ACTGTATC
R14	CAAGCAGAAGACGGCATACGAGATAGGTCGCAGTGACTGGA GTTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTT TTCAA	AGGTCGCA
R15	CAAGCAGAAGACGGCATACGAGATAGGTTATCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	AGGTTATC

R16	CAAGCAGAAGACGGCATACGAGATCAACTCTCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	CAACTCTC
R17	CAAGCAGAAGACGGCATACGAGATCCAACATTGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	CCAACATT
R18	CAAGCAGAAGACGGCATACGAGATCTAACTCGGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	CTAACTCG
R19	CAAGCAGAAGACGGCATACGAGATATTCCTCTGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	ATTCCTCT
R20	CAAGCAGAAGACGGCATACGAGATCTACCAGGGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	CTACCAGG
R21	CAAGCAGAAGACGGCATACGAGATGGTACTCAGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	GGTACTCA
R22	CAAGCAGAAGACGGCATACGAGATGCAGTCATGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	GCAGTCAT
R23	CAAGCAGAAGACGGCATACGAGATGTCTAGCAGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	GTCTAGCA
R24	CAAGCAGAAGACGGCATACGAGATGTGCCATCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTT CAA	GTGCCATC

3.5 Antibodies

Primary antibody	Species	Source	Applications and Dilutions
Anti-CARP / ANKRD1	mouse	Santa Cruz Biotechnology sc-365056	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-EDEM1	rabbit	Sigma-Aldrich E8406	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-HYPE / FICD	mouse	Santa Cruz Biotechnology sc-517074	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-FUT1	rabbit	Abcam ab198712	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-SHARP1 / DEC2	rabbit	Novus biologicals NBP1-19613	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-SHARP2 / DEC1	mouse	Santa Cruz Biotechnology sc-101023	Western Blotting 1:1000 in 5% dry-milk + TBST

Anti-CTH F-1	mouse	Santa Cruz Biotechnology sc-374249	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-CHAC1	rabbit	Sigma-Aldrich AV42623	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-CREB2 / ATF4	mouse	Santa Cruz Biotechnology sc-390063	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-KLF6	rabbit	Santa Cruz Biotechnology Sc-7158	Western Blotting 1:1000 in 3% BSA + TBST
Anti-ERO1LB	rabbit	Abclonal A3682	Western Blotting 1:1000 in 3% BSA + TBST
Anti-PERK	rabbit	Abcam (Ab65142)	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-BiP (C50B12)	rabbit	Cell Signaling (3177)	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-HCoV-229E N protein (1E7)	mouse	Eurofins/Ingenasa M.30.HCo.B1E7; Batch 250609	Western Blotting 1:500 in 5% dry-milk + TBST
Anti-HCoV-229E nsp8	rabbit	Prof. John Ziebuhr, Institute of Medical Virology, Giessen	Western Blotting 1:500 in 5% dry-milk + TBST
Anti-HCoV-229E nsp12	mouse	Carsten Grötzinger	Western Blotting 1:500 in 5% dry-milk + TBST
Anti-β-actin (C4)	mouse	Santa Cruz Biotechnology sc-47778	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-Actinin	mouse	Sigma-Aldrich A5044	Western Blotting 1:1000 in 5% dry-milk + TBST
Anti-ARFGAP3	rabbit	Invitrogen PA5-81884	Western Blotting 1:500 in 5% dry-milk + TBST
Anti-EMC6	rabbit	Abcam ab84902	Western Blotting 1:500 in 5% dry-milk + TBST

Secondary antibody	**Species**	**Source**	**Applications and Dilutions**
Anti-mouse Immunoglobulins / HRP	goat	Dako P0447	Western Blotting 1:2000 in 5% dry-milk + TBST
Anti-rabbit Immunoglobulins / HRP	goat	Dako P0448	Western Blotting 1:2000 in 5% dry-milk + TBST

3.6 Chemicals

All the chemicals that were used in this study were purchased from the following companies, otherwise stated: Baker, Roche, Merck, Promega, Biomol, Invitrogen / Thermo Fisher Scientific, Pharmacia, Roth, Serva, and Fluka (Molecular biology or cell culture grade).

3.7 Plastics and other disposable materials

All the plastic and disposable materials were purchased from the following companies: Eppendorf, Greiner, Ansell, Sarstedt, Neolab, Roth, Nerbe-Plus, Nunc, Omnilab, and Brand.

3.8 Enzymes

Name	Source
GoTaq G2 Flexi DNA Polymerase	Promega M7805
RevertAid Reverse Transcriptase (200 U/µl)	Thermo Fisher Scientific EP0441
RNase A (10 mg/ml)	Thermo Fisher Scientific EN0531
T4 DNA Ligase	Thermo Fisher Scientific EL0011
T4 Polynucleotide Kinase	Thermo Fisher Scientific EK0031
Plasmid-Safe ATP-Dependent DNase	Biozym 161010
FastDigest BbsI	Thermo Fisher Scientific FD1014
FastDigest Esp3I / BsmBI	Thermo Fisher Scientific FD0454
Proteinase K	Macherey-Nagel 740506

3.9 Inhibitors

Name	Source
Thapsigargin	Cayman chemicals 10522
Leupeptin hemisulfate	Roth CN33.2
Microcystin	Enzo ALX-350-012-M001
Pepstatin A	Applichem A2205
PMSF	Sigma P7626
Protease inhibitor cocktail tablets	Roche 118735800001

3.10 Ready-to-use kits, materials, and reagents

Name	Source
10X FastDigest buffer	Thermo Fisher Scientific FD1014
4X Roti-Load	Roth K929.2
5X Reaction Buffer	Thermo Fisher Scientific EP0441
6X DNA Loading dye	Lifetechnologies R0661
Agilent DNA chip	Agilent 5067-1504
Amersham ECL Western Blotting Detection reagent	GE healthcare RPN2106
Ampicillin	Roth K029.2
ATP	Sigma A9187
Avicel	FMC Biopolymer RC591
CellTiter 96® Aqueous One Solution Cell Proliferation Assay (MTS)	Promega G3582
CellTiter-Glo Luminescent Cell Viability Assay	Promega G7570
DMEM	PAN-biotech P04-3550
DMSO	Roth 4720.4
dNTPs	Lifetechnologies R0193
DPBS	PAN-biotech P04-36500
Ethidium bromide 1%	Roth (2218.2)
Fast SYBR Green PCR Master Mix	Applied biosystems 4385612
FBS good forte	PAN-biotech (P40-47500)
GeneRuler DNA Ladder Mix	Thermo Fisher Scientific SM0331
Geneticin	Gibco 11811023
Glycerol	Merck 1.04092.1000
GoTaq G2 Flexi buffer	Promega M7805
Immobilon Western Chemiluminescent HRP-Substrate	Millipore WBKL50500
LB broth	BD (244610)
L-Glutamine	PAN-biotech P04-80100
Lipofectamine LTX and PLUS reagent	Invitrogen 15338-100
MEM medium	Gibco P04-08056
$MgCl_2$	Thermo Fisher Scientific R0971
NEB Next High Fidelity PCR Master Mix, 2X	NEB M0541L
Non-Essential amino acids 100X	Gibco 21013024
NucleoBond PC500 kit	Macherey-Nagel 740574.50

NucleoSpin Gel and PCR clean-up kit	Macherey-Nagel 740609.250
NucleoSpin Plasmid kit	Macherey-Nagel 740588.250
NucleoSpin RNA II kit	Macherey-Nagel 740955.250
NucleoSpin Tissue kit	Macherey-Nagel 740952.50
OptiMEM medium	Gibco 51985-026
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific 26617
Penicillin / Streptomycin	PAN-biotech P06-07100
Pierce Detergent compatible Bradford kit	Thermo Fisher Scientific 23200
Polybrene (hexadimethrine bromide)	Sigma 1076898
Ponceau S	Sigma P7170
Puromycin	Sigma P8833
Qubit™ dsDNA HS kit	Thermo Fisher Scientific Q32851
Quick genomic DNA midiprep plus kit	Zymo Research D4075
Random Hexamer Primer	Thermo Fisher Scientific S0142
Rotiphorese Gel 30	Roth 3029.1
Roti-PVDF	Roth T830.1
Roti-Quant Bradford kit	Roth K015.1
Standard nutrient agar	Merck 1.07881.0500
Syringe filter unit, 0.45 µM	Roth KC79.1
T4 PNK buffer A	Thermo Fisher Scientific EK0031
Trypsin / EDTA	PAN-biotech P10-02-1100
Whatman gel blotting paper	Whatman 426892
Zymo-Spin V with Reservoir	Zymo Research C1016-50

3.11 Buffers and medium

Name	Ingredients
"Special lysis buffer" for cell lysis	10 mM Tris pH 7.05 30 mM $Na_4P_2O_7$ 50 mM NaCl 1% Triton X-100 2 mM Na_3VO_3 50 mM NaF 20 mM ß-Glycerophosphate Freshly added to 1 ml lysis buffer : 1 µl Pepstatin (1 µg/ml) 0.5 µl Leupeptin (10 µg/ml)

	2.5 µl PMSF (1 mM) 1 µl Microcystin (1 µM)
TAE buffer for agarose gels	40 mM Tris 1 mM EDTA 0.11% Acetic acid
Stacking gel buffer	1 M Tris (pH 6.8)
Separating gel buffer	1 M Tris (pH 8.8)
TBS buffer	10 mM Tris 150 mM NaCl (pH 7.4) 0.05% Tween 20 for TBS-T
Ponceau S	0.1% Ponceau S in 5% acetic acid
1X Laemmli buffer	25 mM Tris 192 mM Glycine 0.1% SDS
Western Blotting buffer	25 mM Tris 192 mM Glycine 20% Methanol
2X HEBS	280 mM NaCl 50 mM HEPES 1.5 mM Na_2HPO_4
$CaCl_2$ solution	2 M $CaCl_2$
DMEM medium	10% FBS 2 mM L-glutamine, 100 U/ml penicillin 100 µg/ml streptomycin
LB-medium	10 g Tryptone 5 g of yeast extract 10 g NaCl 1 l distilled water pH 7.0 and autoclaved for 20 minutes at 120°C
LB-agar plates	37 g Standard nutrient agar 1 l ddH_2O and autoclaved Ampicillin (100 µg/ml) was added before plating
TSS buffer	For 25 ml : 10% Polyethylene glycol 6000 – 2.5 g 5% DMSO – 1.25 ml 50 mM $MgSO_4.7H_2O$ – 0.31 g Add upto 25 ml with LB medium

3.12 Devices and apparatus

3.12.1 Agarose gel electrophoresis and Western Blotting

Name	Source
Agilent 2100 Bioanalyzer	Agilent
Elektrophorese Power Supply EPS 600, 601, 3500	Pharmacia Biotech
Chemi Doc Touch Imaging System	Bio-Rad
Gel iX Imager UV-Transilluminator	INTAS
Mighty Small II mini electrophoresis system	Hoefer
Mupid-exU electrophoresis system	Takara
Owl P9DS electrophoresis system	Owl Separation Systems
PerfectBlue ′Semi-Dry′-Blotter, Sedec	VWR Peqlab

3.12.2 Photometer

Name	Source
NanoDrop ND-1000	Peqlab
Spectramax Plus Microplate Spectrophotometer	Molecular devices
Qubit fluorometer	Invitrogen
infinite 200	AG Schmitz, Institute of Biochemistry, Giessen

3.12.3 PCR Thermocycler

Name	Source
7500 Fast Real-Time PCR System	Applied Biosystem
Thermocycler T Professional	Biometra

3.12.4 Microscope

Name	Source
CKX41	Olympus

3.12.5 Cell culture

Name	Source
BBD 6220 CO_2 Incubator	Thermo Fisher Scientific
CO_2 Incubator Auto zero	Heraeus
Safe 2020 bench	Thermo Fisher Scientific

3.12.6 Centrifugation, shakers, and mixers

Name	Source
CH-4103 Shaker	Infors HT
Multitron standard	Infors HT
Sorvall RC5S Plus	Thermo Fisher Scientific
Table-centrifuge 5415 R	Eppendorf
Table-centrifuge 5424	Eppendorf
Table-centrifuge 5424 R	Eppendorf
Thermomixer comfort	Eppendorf
TJ-25 centrifuge	Beckman Coulter
Vortex mixer	neoLab

3.13 Software

Name	Version
Conda	v.4.5.4 or newer
DNASTAR SeqBuilder	Version 11.1.0.54
DNASTAR SeqMan Pro	Version 11.1.0.54
Draw Venn	Online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/)
Endnote	Version 20.1
GraphPad Prism	Version 9.0.0
ImageJ	Java 1.8.0 (https://imagej.net/imagej-wiki-static/ImageJ) [103]
ImageLab	Version 6.0.1 build 34
Inkscape	Version 1.0 (https://inkscape.org/)
MAGeCK	v.0.5.6 or newer (https://sourceforge.net/p/mageck/wiki/Home/) [99]

Metascape	v3.5.20230101 (https://metascape.org/gp/index.html#/main/step1) [104]
Microsoft Excel	2016
Python	Version 2.7 and newer

4. Methods

4.1 Working with mammalian cell culture

All the cell culture methods were strictly carried out in sterile environments using cell culture grade plastics, medium, and reagents as mentioned in the upcoming sub-sections.

4.1.1 Maintenance of the cells

Human hepatoma cells – Huh7 (Japanese Collection of Research Bioresources (JCRB) cell bank) were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g/L glucose, L-Glutamine, w/o Sodium pyruvate, 3.7 g/L Sodium bicarbonate. The medium was supplemented with 10% filtered bovine serum (FBS), 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. Huh7 cells are adherent and divide every 24 to 30 h. Once the cells reached a confluency of about 90-95%, the cells were washed once with Dulbecco's phosphate-buffered saline (DPBS) to remove any excess DMEM medium and were trypsinized using trypsin / EDTA. The cells were then re-suspended into new culture dishes at a dilution of 1:20 with the fresh DMEM medium and passaged only up to 20 times. Except for Huh7 parental cells, all the other CRISPR-Cas9-engineered Huh7 cells were maintained with 1 µg/ml puromycin. The cells were grown in the incubator at 37°C.

Human embryonal kidney cells - HEK293FT were cultured in DMEM with 4.5 g/L glucose, L-Glutamine, Sodium pyruvate, and 1.5 g/L Sodium bicarbonate. The medium was supplemented with 10% filtered bovine serum (FBS), 200 mM L-glutamine, 10 mM MEM Non-Essential Amino Acids, and 100 mM MEM Sodium Pyruvate. HEK293FT cells were maintained in the complete medium with 500 µg/ml geneticin and were passaged (only up to five times) at a confluency of 80%. The cells were grown in the incubator at 37°C.

For determining the seeding density for all the experiments done in this thesis, the cells were counted using a Neubauer chamber. In brief, 10 µl of the cell suspension in DMEM medium was loaded onto the Neubauer chamber and was sealed with a glass cover slip. The chamber was placed under a light microscope at 10 X resolution and all the cells in the four big squares (corner) were counted. The concentration of the cells was calculated using the following formula:

Cell concentration / ml = number of cells counted * 10^4 (volume of four big squares) / number of big squares counted

4.1.2 Freezing and thawing the cells

Before freezing the cell lines, particular care was taken to rule out mycoplasma contamination, which was mentioned in the following section. The cell lines were grown to a confluency of 90% and were trypsinized, and resuspended in 10 ml fresh DMEM medium to 15 ml falcon tubes. The cells were centrifuged for 5 minutes at 500 g, and the supernatant was removed carefully. To the cell pellet, 2 ml

DMEM medium containing 10% DMSO (freezing medium) was added and the suspension was transferred to 2 ml cryo-tubes. The cryo-tubes were stored at -80℃ in the styrofoam boxes for a week and later transferred to the liquid nitrogen tanks for longer storage periods.

The cryo-tubes were taken from the liquid nitrogen tanks and were kept in the water bath at 37℃ for a few minutes until the frozen medium was thawed. The contents of the tubes were transferred to 15 ml pre-warmed DMEM medium, resuspended, and centrifuged for 5 minutes at 500g. The supernatant was carefully removed and fresh DMEM medium was added to the cell pellet and transferred to the culture dishes. The cells were incubated at 37℃ until they were adherent to the cell culture dishes and were supplemented with the appropriate antibiotics.

4.1.3 Testing for mycoplasma

Mycoplasma contamination often causes changes in the host cell metabolism, cell morphology, and even cytopathic events. It is therefore highly recommended cell culture practice check and control for mycoplasma contamination in the cultured cells before freezing and after thawing. The ready-to-use mycoplasma test kit (PanReac Applichem A3744) allows the detection of mycoplasma by the amplification of conserved 16s rRNA gene region using specific primers by PCR and agarose gel electrophoresis. In a 2 ml vial, 1 ml of cell culture supernatant was transferred and centrifuged at 250g briefly to pellet any cellular debris. The supernatant was transferred to a fresh sterile tube and centrifuged at 12 – 20,000 x g for 10 minutes to sediment mycoplasma. The supernatant was discarded and the pellet (usually not visible) was re-suspended with 50 µl of the buffer solution and heated to 95℃ for 3 minutes. The samples were PCR amplified using the reaction mix provided in the kit and 20 µl of the amplified samples were loaded onto a 2% agarose gel. The samples were compared to the internal positive control sample that shows a DNA fragment band at 270 bp, which indicates the presence of mycoplasma DNA. The cell lines mentioned in section 3.1 were tested negative for the presence of mycoplasma. If the cell lines were detected with positive mycoplasma, they were safely discarded from the cell culture.

4.1.4 Transient transfection

Stable knock out cell lines that were mentioned in section 5.1.2 were generated using a calcium-phosphate-based transient transfection method. The transfection method would allow the entry of cloned sgRNA + Cas9 enzyme expressing px459 plasmid into Huh7 cells based on the formation of calcium phosphate-DNA precipitates that bind easily to the cell surface. The sgRNA combined with Cas9 would bind the target gene and introduce a DSB, leading to indel mutation. This method only with low frequency and randomly facilitates the stable integration of sgRNA plasmid components in the host cell genome. Nevertheless, the transient expression of sgRNA and Cas9 enzyme is sufficient to generate DSBs that are irreversible and are carried over to the daughter cells during cell division. Stable, transfected cell lines were selected using stably integrated selection markers.

In brief, Huh7 cells were seeded onto 10 cm dishes at a density of 9*10^5 in 10 ml DMEM medium and were incubated overnight at 37°C. The next day, the transfection mixture was prepared as follows:

Component	Amount
2X HEBS	1.5 ml
H_2O	1.35 ml
Plasmid DNA	30 µg
2M CaCl2	189 µl

The mixture was shortly vortexed and incubated at room temperature for 10 minutes. The transfection mixture was slowly added to the Huh7 cells drop-wise. After 5 h of incubation at 37°C, the DMEM medium was completely removed and 3 ml of DMEM + 10% glycerol (glycerol shock medium) was added and incubated for 3 minutes. The medium was removed and washed with DPBS two times. Finally, 10 ml of fresh DMEM medium was added and the cells were incubated overnight at 37°C. Thereafter, the successfully transfected Huh7 cells were selected by the addition of 1 µg/ml puromycin in the fresh DMEM medium every 48 to 72 h, until the untransfected cells were removed from the cell pool.

4.2 Working with bacteria

4.2.1 Bacterial transformation

The bacterial strain XL-1 blue *E.Coli* was made competent (for the bacterial transformation) using a transformation and storage solution (TSS) buffer described by Chung & Miller et al. [105]. In brief, an overnight culture was inoculated from a frozen XL-1 blue cryotube into a 3 ml lysogeny broth (LB) medium without antibiotics (since the bacteria does not yet possess a resistance plasmid). The culture was incubated at 37°C and at 160 rpm on the CH-4103 shaker. In the next morning, the overnight culture was inoculated in 150 ml of LB medium to make them competent, at an optical density (OD) of 0.05, which was measured at 595 nm using LB medium as a reference in the colorimeter. The culture was grown until the OD reached 0.3 to 0.5 (1 – 2 hours) and thereafter, the culture was immediately pelleted in 50 ml falcon tubes at 4°C for 10 minutes at 1500 x g. The bacterial pellet was kept on ice and resuspended in 1/10 volume of TSS buffer. The suspension was aliquoted in sterile 1.5 ml centrifuge tubes at 200 µl / tube and was immediately frozen in liquid nitrogen, before storage at -80°C.

To transfer the plasmid DNA into the chemical-competent E.coli bacteria, a chemical-competent bacterial transformation method was used. Previously prepared in-house chemical competent XL-1 blue or Top 10 E.coli were used for this purpose. The vials containing 200 µl of chemical-competent cells were placed on ice for a few minutes. To the vials, 100 – 150 ng of ready-to-use plasmids or 1.5 to 2 µl of cloned plasmid constructs were added and incubated on the ice for 20 minutes. In the next step, the

vials were kept on the thermoblock mixers at 42°C for 45 seconds. This heat shock procedure would allow the bacterial cell membrane to be more permeable for the plasmid DNA to pass through. After the heat shock, the vials were immediately kept on ice for 2 minutes. To the vials, 600 µl of warm LB-medium was added and placed on the CH-4103 shaker for 1 h at 37°C. The vials were then centrifuged at 3000 rpm for 5 minutes to pellet down the bacteria. The supernatant was removed, 100 µl of new LB-medium was added, and the pellet was resuspended. For plating the bacteria onto the pre-warmed LB-agar plates, 50 – 100 µl of resuspended bacteria were added onto the plates and were spread using the sterile cell spreaders. The plates were then incubated at 37°C incubators for about 16-18 h for the transformed bacterial colonies to grow.

4.2.2 Mini and maxi-bacterial cultures

After the successful bacterial transformation and colony formations on the LB-agar plate, the colonies were picked and were further cultured on the liquid LB-medium depending on the usage of the bacterial cultures – mini cultures (for sanger sequencing verifications) and maxi-cultures (for large-scale extraction of plasmid DNA).

For the production of mini-cultures, 4 ml of pre-warmed LB-medium was added to the 5 ml sterile plastic tubes with the lids. To the LB-medium 100 µg/ml ampicillin was added and the colonies were picked using sterile filter tips and immersed in the medium briefly. The tubes were placed on the CH-4103 shaker at 37°C at 160 rpm and incubated for 16 to 18 h for the bacterial culture to grow.

For the production of maxi-cultures, 150 to 200 ml of pre-warmed LB-medium was added to the 500 ml sterile glass flasks. To the LB-medium 100 µg/ml ampicillin was added and the colonies were picked using sterile filter tips and immersed in the medium briefly. The flasks were placed on the Multitron standard shaker at 37°C at 220 rpm and incubated for 16 to 18 h for the bacterial culture to grow.

4.3 Working with human coronaviruses

4.3.1 HCoV-229E propagation

For the viral propagation, Huh7 cells were seeded to 145 mm dishes at a density of 3*10^6 in 20 ml DMEM medium and were incubated at 37°C overnight. On the day of infection, the dishes were transferred to 33°C for about 1 h and then infected with 100 µl of previous viral stock using an MOI of 0.1. The medium was removed and replaced with 16 ml of fresh DMEM medium after an hour of infection. The cells were incubated at 33°C for about 48 to 50 h until the cytopathic effects were visible under the light microscope. The medium was completely transferred to 50 ml falcon tubes, centrifuged at 500 g for 5 minutes, and then aliquoted in 1.5 ml sterile tubes (500 µl – 1 ml) and stored at -80°C for longer storage periods.

4.3.2 Viral titer determination using the $TCID_{50}$ method

To determine the infectivity of HCoV-229E, the viral titers were determined as described below based on the cytopathic effect on the cells for a given virus stock. The ability of the virus dilution to infect 50% of the cell culture is termed the median tissue culture infectious dose ($TCID_{50}$). Several methods have been developed to calculate the $TCID_{50}$, of which the Reed-Muench method has been used widely [106]. To assay the viral titer of HCoV-229E, Huh7 cells were seeded onto 96-well plates (in technical triplicates) at a density of 12,000 cells / well in 100 µl DMEM medium. After 24 h of seeding and incubation, the plates were incubated for 1 h at 33°C before the infection. The viral stock, for which the viral titer was to be measured was diluted serially from 10^-1 to 10^-8. Each well of a single row was infected with 100 µl of the single dilution. One complete row was added with 100 µl of DMEM medium as an uninfected control. The plates were incubated at 33°C for 6 days. After the infection period of 6 days, the medium from all 96 wells was removed and 100 µl of 0.1% crystal violet solution was added. The plates were incubated for 2 to 5 minutes to allow the crystal violet to stain the cells and the dye was removed by carefully tapping the plates on tissue papers. Thereafter, the stained and unstained wells from each virus dilution were counted and the $TCID_{50}$ was calculated from the representative figure as follows:

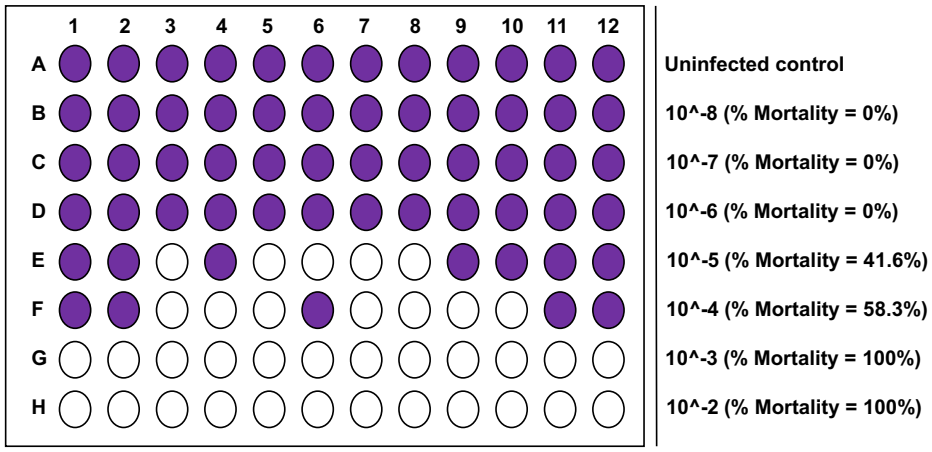


Figure 5: Representative figure showing the crystal violet staining of a 96-well plate to determine the viral titer using the $TCID_{50}$ method.
Purple wells indicate no infections and white wells indicate lysed cells. The percentage of lysis (mortality) for each row is calculated according to the formula: (number of lysed wells) * 100 / 12.

Proportional distance = (% of wells with mortality at dilution above 50% - 50%) / (% of wells with mortality at dilution above 50%) – (% of wells with mortality at dilution below 50%)

$LogID_{50}$ = log (dilution with >50% mortality) + Proportional distance * (-log (dilution factor))

For instance, from the figure above: PD = (58.3 – 50 / 58.3 – 41.6) = 0.49

Using the correct proportional distance value, the $LogID_{50}$ can be calculated as follows:

50% endpoint factor or $LogID_{50}$ = -4 + 0.49 * (-1) = -4 – 0.49 = -4.49

$ID_{50} = 10^{-4.49}$

The viral titer was calculated in terms of infectious doses per unit volume. If the viral inoculum was 100 µl, then the titer of the virus stock would be: 1 / (10^ (-4.49)) / 0.1 = $3*10^5$ $TCID_{50}$/ml

4.3.3 Viral titer determination using Plaque assay

Plaque assay was used to determine the number of infectious particles that lysed the cell monolayer when infected [107]. In this assay, the infectious virus stock was serially diluted and added to the cell monolayer. After a short period of infection (about 1 hour for the virus to enter the cell membrane), the virus-containing medium was removed and the cell layer was covered with semi-solid cellulose along with a medium that provides nutrients to the cells. At the end of one replication cycle, the newly synthesized viral particles could only infect the neighboring cells due to the semi-solid medium. At the end of the assay, the wells with uninfected cells would contain a confluent layer with the crystal violet stain, while the cells that were lysed completely would appear clear. The intermediate number of lysed and uninfected cells would contain circular clear areas called plaques. Each plaque was considered to have originated from one infectious viral particle or a plaque-forming unit (PFU).

In brief, Huh7 cells were seeded onto 12-well plates at a density of 2*10^5 cells / well in 1 ml DMEM medium. After 24 h of seeding and incubation, the plates were incubated for 1 h at 33℃ before the infection. The viral stock, for which the viral titer was to be measured was diluted serially from 10^-2 to 10^-6. For each dilution, two wells (technical duplicates) were infected with 400 µl of the dilutions. Two wells were mock infected with 400 µl of DMEM medium as an uninfected control. The plates were incubated at 33℃ for 2 h and the medium was removed completely and replaced with 500 µl of Avicel (1.25%) + MEM medium (10% FBS, and 100 µg/ml streptomycin + 100 U/ml penicillin). The plates were then incubated at 33℃ for 3 days. At the end of the incubation, 1 ml of water was added to dissolve the overlay. The avicel overlay was carefully removed and stained with 0.15% crystal violet solution for 5 minutes. The stain was completely removed and washed once with water. The plaques were then counted and the viral titer was measured as follows from the illustrative figure:

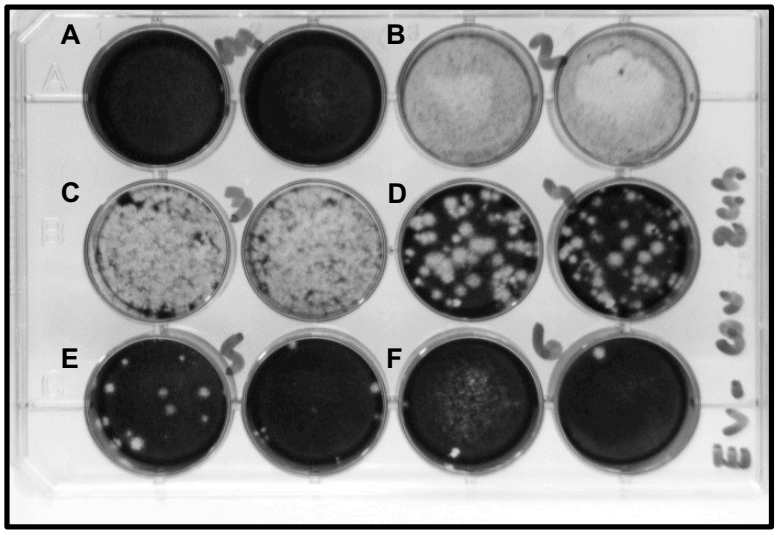


Figure 6: Representative figure showing the crystal violet staining of a 12-well plate to determine the viral titer using plaque assay.
The numbers B) to F) show serial dilution factors.

Virus dilution	Well 1	Well 2	Average plaques	PFU/Well
1E-02	All dead	All dead	-	-
1E-03	All dead	All dead	-	-
1E-04	52	59	55.5	2.8E+05
1E-05	12	5	8.5	4.3E+05
1E-06	All alive	All alive	-	-

The average PFU/well was then calculated to be ((2.8 + 4.3) / 2) * 10^5 = 3.5E+05. The viral titer was calculated in terms of infectious doses per ml. If the viral inoculum was 400 µl, then the titer of the virus stock would be 3.5E+05 * 2.5 = $8.8*10^5$ PFU/ml.

4.4 Molecular biology methods

4.4.1 Designing sgRNAs

The sgRNAs used in the generation of stable knock out cell lines that were mentioned in section 5.1.2 were designed with the help of the Genetic Perturbation Platform (GPP) (https://portals.broadinstitute.org/gpp/public/) [84]. In the first step, the gene sequence for which the sgRNAs were designed was extracted from the UCSC genome browser (Dec.2013 GRch38 / hg38). The complete gene sequence extracted from the browser was then uploaded to the SeqBuilder application of the DNASTAR navigator tool and a gene sequence file was created. The exons were marked in upper case letters and the introns were marked in lower case letters. For picking the sgRNAs, about 1000 nucleotides were marked from the transcriptional start site of the gene and were uploaded

to the GPP web portal. The specialized portal was designed to pick the 20 nt sgRNAs with additional information including the off-target score, target efficiency, and the direction of the sgRNA binding for the user to choose from the list of sgRNAs provided. From the list, 3 sgRNAs that had the maximum efficiency, and a lower off-target score were chosen and the sequences were marked in the already created SeqBuilder gene sequence. In the next step, with the SeqBuilder application, the sgRNAs targeting the gene were virtually tested for efficiency (when combined with the Cas9 to create DSBs). The cut site occurs at the sgRNA binding sequence 3-5 bases upstream from the PAM sequence (NGG). The sgRNAs that produced the change in the downstream translational sequence in the amino acid sequence after the cut site (usually premature stop codon) (also provided by the SeqBuilder application) were then selected as the candidate sgRNAs for cloning into the desired vectors.

4.4.2 Cloning sgRNAs into px459 vector

The sgRNAs which were selected from the above-mentioned designing method were synthesized as oligonucleotides from the Eurofins genomics as stated in section 3.2. The oligonucleotides were cloned into the px459 vector using the following protocol.

Step 1: Annealing the oligonucleotides

Component	Amount
sgRNA se sequence (10 µM)	1 µl
sgRNA as sequence (10 µM)	1 µl
T4 PNK buffer A	5 µl
ddH_2O	43 µl
Total	50 µl

The complete 50 µl mixture was then annealed in a thermocycler using the following conditions:

95°C for 4 minutes

70°C for 10 minutes

37°C for 15 minutes

4°C pause

Step 2: Phosphorylation of the annealed oligonucleotides

Component	Amount
Annealed oligos	2 µl
T4 PNK buffer A	1 µl
ATP (1 mM)	1 µl
T4 PNK (10 U/µl)	1 µl
ddH_2O	5 µl
Total	10 µl

The complete 10 µl mixture was then phosphorylated in a thermocycler using the following conditions:

37°C for 30 minutes

70°C for 10 minutes

4°C pause

After the reaction was completed, the phosphorylated oligos were diluted 1:8 with ddH_2O.

Step 3: Restriction and ligation

Component	Amount
px459 vector	3 µl (100 ng)
Phosphorylated oligos (diluted 1:8)	2 µl
FastDigest buffer	2 µl
ATP (10 mM)	1 µl
FastDigest BbsI (BpiI) (10 U/µl)	1 µl
T4 DNA ligase	1 µl
ddH_2O	10 µl
Total	20 µl

The reaction was set up in a thermocycler using the following conditions:

37°C for 5 minutes

21°C for 5 minutes

The reaction was set up as a loop for 6X.

Step 4: Digestion of remaining linearized plasmid DNA with PlasmidSafe exonuclease

Component	Amount
Ligated plasmid	10 µl
PlasmidSafe buffer	1,5 µl
ATP (10 mM)	1,5 µl
PlasmidSafe DNAse (10 U/µl)	1 µl
Total	14 µl

The reaction was set up in a thermocycler using the following conditions:

37°C for 30 minutes

70°C for 30 minutes

After the reaction was completed, the ligated plasmid DNA can be transformed into bacteria for amplification as mentioned in section 4.2.1

4.4.3 DNA extraction

For the extraction of plasmid DNA from the amplified mini and/or maxi-bacterial cultures, the protocol was followed according to the instructions provided in the kit: NucleoSpin plasmid kit (mini-prep) and NucleoBond Maxi kit (maxi-preps). The eluted plasmid DNA was measured using the Nanodrop for the concentration and was stored at -20°C for longer storage periods.

Genomic DNA from the CRISPR-Cas9 engineered cell lines was isolated using the NuceloSpin Tissue kit according to the instructions provided within the kit. The eluted genomic DNA from the cell lines was measured using the Nanodrop for the concentration and was stored at -20°C for longer storage periods.

4.4.4 Sanger sequencing

To verify the indel mutations in the CRISPR-Cas9 engineered cell lines, the genomic DNA was isolated and the sgRNAs binding regions were PCR amplified using the following conditions:

Component	Amount
Flexi-Buffer	12 µl
$MgCl_2$	9.6 µl
dNTPs	1.2 µl
Forward primer (diluted 1:10)	3 µl
Reverse primer (diluted 1:10)	3 µl

GoTaq G2 Flexi DNA Polymerase (5 U/µl)	0.3 µl
Genomic DNA	1 µg
ddH_2O	Adjusted to 60 µl
Total	60 µl

The reaction was set up in a thermocycler using the following conditions:

Step	Temperature and time	Repeat
1	95°C for 2 minutes	-
2	95°C for 30 seconds	-
3	60°C for 30 seconds	-
4	72°C for 40 seconds	From step 2 for 35X
5	72°C for 5 minutes	-

The PCR amplicons, which were usually around 400 to 500 nucleotides in length, were run on a 2% agarose gel and then visualized under a UV transilluminator Gel iX imager for the amplified DNA fragment. The samples were purified using the NucleoSpin Gel and PCR clean-up kit according to the instructions provided. The purified PCR amplified samples were sent to LGC genomics or MicroSynth according to the instructions provided by the company. The Sanger sequenced fastq files were analyzed with the SeqMan Pro application of the DNASTAR tool.

4.4.5 RNA extraction and cDNA synthesis

For measuring the viral genome transcripts in the infected cell lines, total RNA was extracted using the NucleoSpin RNA kit according to the manual provided. The RNA was eluted with 50 µl RNAse-free-H_2O and the concentration was measured using the Nanodrop. The eluted RNA was stored at -80°C for longer storage periods or used immediately for synthesizing cDNA.

The cDNA was synthesized from the eluted RNA using the RNA-dependent DNA polymerase enzyme using the following conditions:

Component	Amount
Reverse transcription RT buffer (5X)	4 µl
Random hexamer primer	0.5 µl
dNTP mix (10 mM)	0.5 µl
RevertAid Reverse Transcriptase (200 U/µl)	0.5 µl

Total RNA	500 ng
ddH_2O	Adjusted to 20 µl
Total	20 µl

The reaction was set up in a thermocycler using the following conditions:

25°C for 10 minutes

42°C for 60 minutes

70°C for 10 minutes

10°C pause

After the reaction was complete, the cDNA was directly used for RT-qPCR or stored at -20°C for longer storage periods.

4.4.6 RT-qPCR (Absolute quantification method)

Gene expression can be quantified by using RT-qPCR by two methods: relative and absolute quantification. The relative quantification method compares the gene expression levels of treated to untreated samples, in most instances by additionally using a non-regulated "housekeeping" gene for normalization. However, the absolute quantification method is based on the generation of a standard curve from the gene of interest template with a known concentration. Then, the concentration or the copy numbers of the gene of interest in the unknown samples can be determined by interpolating its cycle number into the standard curve.

To determine the copy numbers of the viral-S-mRNA in the infected knock out cell lines, the absolute quantification method was used. The cDNA from the viral-S-mRNA was generated by using a specific primer set that amplified the 151 bp DNA fragment. This served as the template from which the standard curve was generated later on. The concentration of the cDNA was measured using a Nanodrop. To generate a standard curve from the template, a 10-point 10-fold serial-dilution was prepared which started with a dilution of 1:2000. A PCR master mix was prepared for the unknown cDNA samples and the serial-diluted cDNA templates using the following conditions:

Component	**Amount**
SYBR green master mix	10 µl
H_2O	6 µl
HCoV-229E S RNA_se (1:10)	1 µl
HCoV-229E S RNA_as (1:10)	1 µl
cDNA	2 µl

The PCR reaction was set up on an Applied Biosystems 7500 Real-Time PCR instrument in 96-well plates with technical duplicates using the following conditions:

Step	Temperature and time	Repeat
1	95°C for 20 seconds	-
2	95°C for 3 seconds	-
3	60°C for 30 seconds	From step 2 for 40X

The Ct values from the standard dilutions and the unknown cDNA samples were extracted from the 7500 Real-Time PCR software. In the first step, the copy number of 1:2000 dilution of the viral-S-mRNA cDNA template was calculated using the following formula:

copy number = n * Avogadro constant (6.022×10^{23} molecules / mol) / molecular mass of amplicons (g / mol)

n = concentration of the cDNA template (ng / µl) used in the PCR reaction

molecular mass = length of the template (bp) * 1*10^9 * 650 g / mol (average mass of 1 bp dsDNA)

The copy numbers for the 10-fold dilutions were then calculated. The standard curve was generated by plotting the log values of copy numbers against the Ct values. The slope of the line was measured to assess the assay's efficiency. The slope (m) values between -3.1 and -3.6 were considered acceptable efficiency (90 to 100%). The y-intercept (b) was calculated from the line and R^2 (should be > 0.99) was measured as the coefficient of correlation between the data generated and the results expected under the ideal conditions. To determine the copy number of the unknown cDNA samples, the following equation was used:

copy number of unknown sample = $10^{((Ct\ unknown-b)\ (intercept))\ /\ m\ (slope))}$

4.5 Biochemistry methods

4.5.1 Cell lysis

To visualize the protein expression levels, the cells were harvested after the indicated experimental conditions in the results section. The cell pellets were resuspended with 60 µl of Triton-based lysis buffer (called "special lysis buffer") and were incubated on ice for 15 minutes. Thereafter, lysates were centrifuged at 10,000g for 15 minutes. The supernatants (protein lysates) were then transferred to fresh tubes.

Additionally, the cell pellets were also lysed using an SDS lysis buffer. The cell pellets were resuspended with 90 µl of Ca^{2+}/Mg^{2+} free PBS and 10 µl of 10% SDS was added and boiled to 100°C for 10 minutes. Thereafter, the samples were centrifuged at 600g for a minute and the supernatants were

transferred to fresh tubes. The samples were boiled again to 100°C for 10 minutes and centrifuged at 600g for a minute.

4.5.2 Bradford assay

Bradford assay is a colorimetric protein assay that is based on the absorbance shift in the coomassie dye when the red form of the coomassie reagent changes to blue form by the binding of the protein. The bound form has the maximum absorption spectrum at 595 nm. Thus, an increase in the absorbance at 595 nm is proportional to the concentration of protein in the sample. The assay was done to measure the protein concentration in the samples as follows:

This assay depends on the absolute quantification method by comparing the absorbance of known protein concentrations (protein standards) to the absorbance of unknown protein samples. In the first step, the standard concentrations of BSA were prepared (from 5 µg to 35 µg). In a 96-well plate, 100 µl of each BSA standard was added in technical triplicates. The protein lysates for which the concentration to be measured were diluted 1:300 in PBS and from that, 100 µl was added to 96-well plates in technical triplicates. To the standards and unknown protein lysates, 100 µl of Roti-Quant Bradford reagent was added and incubated for 2 minutes. The absorbance was read on a Spectramax Plus Microplate Spectrophotometer at 595 nm and the protein concentrations (µg/ml) of the unknown protein lysates were calculated based on the standard curve.

For the protein lysates prepared using the SDS lysis buffer, the lysates were diluted 1:150 in PBS. The BSA standards were used in the same concentrations as mentioned in the above protocol. To the standards and unknown protein lysates, 100 µl of Pierce Detergent compatible Bradford reagent was added and incubated for 2 minutes. The absorbance was read on a Spectramax Plus Microplate Spectrophotometer at 595 nm and the protein concentrations (µg/ml) of the unknown protein lysates were calculated based on the standard curve. The samples without protein lysates were used to estimate the background absorbance of buffers.

4.5.3 SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis is a method used to separate the proteins based on the molecular mass between 5 to 250 kDa. Polyacrylamide forms a mesh-like matrix to allow the separation of proteins based on their size. The separation takes place by the migration of negatively charged SDS-bound proteins towards the positive pole during the electrophoresis. In general, the low molecular weight proteins are separated much faster than the proteins with high molecular weight. The SDS-gel is divided into two polyacrylamide gels: stacking and separating gel. The stacking gel has a lower percentage of acrylamide and a pH (of 6.8) than the bottom separating gel (with varied higher acrylamide concentration and a pH (of 8.8)). The Laemmli buffer contains glycine (pH of 8.3), which is predominately negative charged, forming glycinate anions. When an electric field is applied, the

anions pass through the stacking gel and change to neutrally charged zwitterions. On the other hand, the Cl- ions (from Tris-HCl in the gel) move faster toward the anode. When the Cl- and glycine zwitterions pass through the wells with protein samples in stacking gel, they form a steep voltage gradient with the protein samples moving in between these boundaries. This results in the formation of a tight protein band in the stacking gel and it is resolved in the separating gel based on the molecular weight of the protein [108]. In this thesis, 8 and 10% separating gel was used to resolve proteins with molecular weights of 30 kDa – 200 kDa. The following recipes were used in the preparation of stacking (10 ml) and separating gels (27 ml):

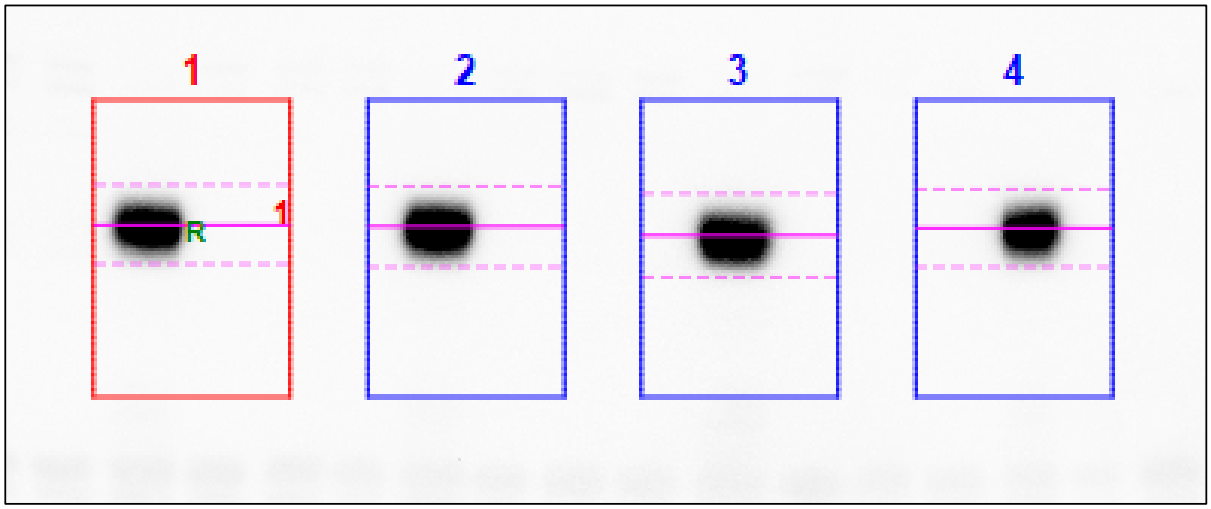
Component	For 8% separating gel	For 10% separating gel	For stacking gel
H_2O	5.6 ml	3.8 ml	5.6 ml
Separating gel buffer	10.3 ml	10.3 ml	-
Stacking gel buffer	-	-	1.26 ml
50% glycerol	2.4 ml	2.4 ml	0.9 ml
Rotiphoresis gel 30	7.2 ml	9 ml	1.68 ml
2% SDS	1.35 ml	1.35 ml	50 µl
40% APS	41 µl	41 µl	30 µl
TEMED	27 µl	27 µl	20 µl

The separating gel was cast first and was allowed to solidify for 30 minutes and thereafter the stacking gel was cast along with a plastic comb with the desired number of lanes for loading. The comb was removed carefully after the stacking gel was solidified. The protein lysates were mixed with 4X Roti-load dye and were boiled to 95°C for 5 minutes and were loaded to the SDS gels alongside a PageRuler prestained protein ladder. The electrophoresis was done in an Owl P9DS system using 1X Laemmli buffer at a constant 45 mA per gel until the loading dye reached the very end of the gel plate.

4.5.4 Western Blotting

Western Blotting is a method used to transfer the proteins from the PAGE-separated SDS gels to a PVDF membrane based on the electrophoretic mobility of the proteins when an electric field is applied. The semi-dry blotting method was used in this thesis and the transfer sandwich consisted of two filter papers soaked in the blotting buffer, followed by a PVDF membrane (facing the positive electrode), the SDS gel, and then two filter papers soaked in the blotting buffer was placed between the two electrode plates of the PerfectBlue ′Semi-Dry′-Blotter. Before making the sandwich, the PVDF membrane was activated with 100% methanol for a few seconds. The transfer was done for 2 hours with a constant current of 0.75 mA per cm^2. After the transfer, the PVDF membrane was soaked in Ponceau S stain for 5 to 10 minutes as a control for the transfer process. The Ponceau S Stain was then documented and the stain was removed by soaking the membrane in TBS-T buffer for 10 minutes. Thereafter, to visualize

the protein of interest on the membranes, the membranes were soaked in 5% dry-milk + TBS-T or 5% BSA + TBS-T solution with continuous shaking for 1 hour at room temperature to block the non-specific areas. The blocking solution was removed and the membranes were added with the primary antibodies of interest at 4°C with an overnight continuous shaking. The primary antibodies were removed and were washed with TBS-T buffer for 30 minutes with a periodic change of buffer every 10 minutes to remove any unbound antibodies. Thereafter, the membranes were incubated with HRP-coupled secondary antibody for 1 h at room temperature. The membranes were washed with TBS-T buffer for 30 minutes with a periodic change of buffer every 10 minutes and were incubated with ECL or Immobilon chemiluminescent reagent for the detection of protein bands using the ChemiDoc Touch Imaging System. The protein expression levels were analyzed using the ImageLab software. In brief, the protein bands were quantified in the software using a relative quantification method. Lanes were adjusted and the protein band of interest was selected manually within a small region of interest (ROI) and raw values, background, and adjusted (i.e. background subtracted) protein intensity values were calculated by the software. If appropriate, a reference band was selected (e.g., a control) to generate values for relative changes in protein band intensity between two conditions. Based on the relative quantity of the reference band, the protein levels of treated samples were then calculated.



Lane	Band No.	Relative Front	Adj. Volume (Int)	Volume (Int)	**Relative quantification (using Adj. Volume)**	Band %	Lane %
1	1	0.518987	6276624	11538104	**1**	100	96.297665
2	1	0.56962	5769052	10873324	**0.919133**	100	95.354655
3	1	0.582278	6686844	11916848	**1.065357**	100	96.959796
4	1	0.468354	5549400	12110444	**0.880061**	100	98.697331

Figure 7: An illustrative figure showing the quantification of protein bands using ImageLab software.
Total number of lanes (blue boxes) and bands (violet lines) were adjusted manually in an immunoblot using the ImageLab software. The protein band of interest was selected manually within the small region of interest (discontinuous pink boxes). A reference band was selected (indicated by R in lane 1) and the software will generate the relative changes in the protein band intensities for the other conditions / lanes (as shown in the table).

4.5.5 Cell viability assay

The cell viability of the host factor knock outs upon the HCoV-229E infection was assessed by using a cell proliferation assay. The CellTiter 96® Aqueous One Solution Cell Proliferation Assay is a colorimetric method for determining the number of viable cells in a proliferation assay. The reagent consists of a tetrazolium compound (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt; MTS) which is bio reduced by cells into a soluble colored formazan product in the cell culture medium. The cellular dehydrogenase enzymes in metabolically active, healthy cells facilitate this reduction. The amount of formazan produced is measured by the absorbance at 490 nm and is directly proportional to the number of living cells in the cell culture. Briefly, Huh7 cells were seeded onto 96-well plates at a density of 5000 cells / well in technical triplicates or quadruplicates. The cells were infected and / or treated with thapsigargin and were incubated for 72 h and / or 96 h as mentioned in the results section. After the incubation period, the medium was completely removed from the wells and replaced with 96 µl of DMEM + 4 µl CellTiter 96® Aqueous One Solution reagent and incubated at 37°C for 1 h. The absorbance was measured on the Spectramax Plus Microplate Spectrophotometer at 490 nm. The absorbance values for all the measured samples including the technical replicates were background subtracted (corrected absorbance) from the average absorbance value of the blank (medium only). The cell viability of the infected conditions was calculated as follows:

Cell viability of infected sample (%) = (corrected absorbance of infected sample / corrected absorbance of the uninfected sample) * 100

The mean cell viability value of the infected sample was calculated from all the technical replicates. The mean cell viability value of the uninfected control was set to 100%.

4.6 Genome-wide CRISPR-Cas9 knock out screening

A detailed explanation of the setup of genome-wide CRISPR-Cas9 knock out screening which included plasmid library amplification, lentivirus production for transduction, generation of stable genome-wide knock out cell pool, screen selection, analysis of the screen, and validation of genes from the screen was described systematically in this section.

4.6.1 sgRNA plasmid library amplification

The pre-made sgRNA library was purchased from Addgene (1000000058) as two pooled DNA half-libraries that consisted of 20 µl of plasmid DNA with a concentration of 50 ng/µl. Library A consisted of 65,383 sgRNAs and library B of 58,028 sgRNAs. The full sgRNA library consisted of 123,411 sgRNAs with 6 unique sgRNAs targeting each gene (19,050 humans genes targeted), including 1,864 sgRNAs targeting human microRNAs (miRNAs), and 1000 non-targeting control sgRNAs (designed to bind at the non-coding regions). The sgRNAs were cloned into a lentiCRISPR V2 (one-vector system), which integrates into the host cell genome and stably expresses the sgRNAs + Cas9 enzyme and the puromycin N-acetyl-transferase for puromycin resistance.

The quantity of pre-made library was not sufficient for the lentivirus production, and hence, the library was amplified using One Shot TOP10 chemical competent cells. To avoid any potential bias in the sgRNA plasmid distribution, each half-library was amplified separately. For one half-library, a total of 6X bacterial transformations (1X transformation for 10,000 sgRNAs) were done. To each vial of TOP10 bacteria (kept on ice), 2 µl (100 ng) of sgRNA plasmid library A or B was added and incubated on ice for 30 minutes. The vials were then kept on thermoblock mixers at 42°C for 30 seconds and then immediately placed on ice for 2 minutes. To the vials, 250 µl of pre-warmed SOC medium was added and placed on a CH-4103 shaker at 37°C at 225 rpm for 1 h. During the recovery period, 30 to 40 standard 100 mm LB agar plates for each half-library transformation were pre-warmed at 37°C. After 1 h of the recovery period, the 6X vials were pooled together and in addition, 1.5 ml of pre-warmed SOC was added and mixed by inverting the tubes 10 times. To estimate the transformation efficiency, a 100-fold dilution was prepared by adding 10 µl of the pooled transformed cells with 990 µl LB medium. Then, 100 µl of the 100-fold dilution was added to 900 µl of LB medium for a 1,000-fold dilution. Finally, 100 µl of the 1,000-fold dilution was plated onto an LB-agar plate (control plate) that gave a 10,000 fold-dilution, from which the transformation efficiency was estimated. The remaining pooled transformed cells were plated onto 30 LB-agar plates equally (100µl per dish) and were incubated at 37°C for only 14 h to avoid any inter-colony competition or differences in colony growth rates. After the incubation period, the colonies that were grown on the control plate were counted. The number of colonies on the control plate was multiplied by 10,000 and the number of transformations was used. The outcome should be in the range of 6*10^6, to ensure each sgRNA was amplified about 100 times and to avoid a huge drop out of sgRNAs during the amplification. In the next step, the colonies were harvested from all the LB-agar plates with 1 ml LB medium and were scraped off using a sterile cell spreader. The LB-medium with the harvested colonies was collected in a 50 ml falcon tube and a maxi-prep was done to isolate the plasmid DNA using a NucleoBond Maxi-prep kit. The concentration of the isolated sgRNA plasmid library was measured using a NanoDrop.

In the next step, the sgRNA library was PCR amplified to determine the sgRNA distribution by NGS. The following conditions were used to prepare the sgRNA library for NGS. 10X PCR reactions were set up that consisted of one forward primer (F1-F10) and the same reverse primer.

Component	Amount	Final concentration
NEB High Fidelity PCR master mix, 2X	25 µl	1X
sgRNA plasmid library	1 µl	0.4 ng/µl
NGS forward primers (F1-F10) (Section 3.4)	1.25 µl	0.25 µM
NGS reverse primer (unique)	1.25 µl	0.25 µM
ddH_2O	21.5 µl	-
Total	50 µl	-

The prepared mixture was PCR amplified on a thermocycler using the following conditions:

Step	Temperature and time	Repeat
1	95°C for 5 minutes	-
2	98°C for 20 seconds	-
3	60°C for 15 seconds	-
4	72°C for 15 seconds	From step 2 for 12X
5	72°C for 1 minute	-

After the PCR reaction, 10X reactions were pooled together and purified using the NucleoSpin Gel and PCR purification kit. The purified PCR samples (2 µg) were loaded onto a 2% agarose gel and separated for 30 minutes. The 270 bp PCR product was excised from the gel and was purified using the NucleoSpin Gel and PCR purification kit. The purified gel extracted samples were quantified using the Qubit dsDNA High-sensitivity Assay kit according to the instructions provided in the kit. The full sgRNA library (A+B) was then sequenced on the Illumina MiSeq according to the user manual with 80 cycles of read 1 (forward) and 8 cycles of index 1. As a control for sgRNA library diversity, 5% PhiX was spiked in and the read coverage was aimed for >100 reads per sgRNA. PhiX is a small non-tailed bacteriophage with a single-stranded DNA and has a genome of 5,386 nucleotides. Due to its small genome, PhiX has been commonly used as a control for Illumina sequencing runs to increase the library diversity at the beginning of the run and as a quality control for cluster generation [109].

The fastq files generated from the sequencing run were then used to analyze the read count distributions of sgRNAs. To analyze the fastq files, a python script that was already created by the Broad Institute of MIT and Harvard was used. In a Linux environment, python 2.7 and biopython were installed and a

.csv file containing the guide RNA sequences (20 nt) with each row corresponding to one sequence was created. To determine the sgRNA distribution, the following parameters were used:

Flag	Description	Default
-f	Fastq file containing NGS data for analysis	1.fastq
-o	Output .csv file with guide RNA sequences in the first column and respective read counts in the second column	Output.csv
-i	Input .csv file with sgRNA sequences	2.csv
-no-g	Indicate the absence of guanine before the guide spacer sequence	Guanine is present

The python script was run by entering the following command:

```
"C:\Python27\python.exe" "C:\Users\gk4325\Desktop\NGS\Screening protocols\count_spacers.py" -f="C:\Users\gk4325\Desktop\NGS\Inputs\1.fastq" -i="C:\Users\gk4325\Desktop\NGS\Inputs\2.csv" -o="C:\Users\gk4325\Desktop\NGS\Output\Output.csv" -no-g
```

The run generated an excel file as an output that consisted of sgRNA sequences (123,411 sgRNAs) in the first column and their corresponding read counts in the second column. In addition, the output file consisted of relevant statistical information including the number of perfect guide matches, non-perfect guide matches, the total number of reads processed, the percentage of perfectly matching guides, the undetected guide percentage, and a skew ratio.

4.6.2 Lentivirus production and titer measurement

The lentivirus was produced using the second-generation packaging plasmids: pMD2.G, which encodes the viral envelope protein, and psPAX2, which encodes the packaging proteins for the viral assembly and formation (detailed description of the plasmids were mentioned in section 3.4). Along with the transfer plasmid (sgRNA plasmid), the packaging and envelope plasmids were transfected in the HEK293FT cell line that served as an ideal platform for lentivirus production.

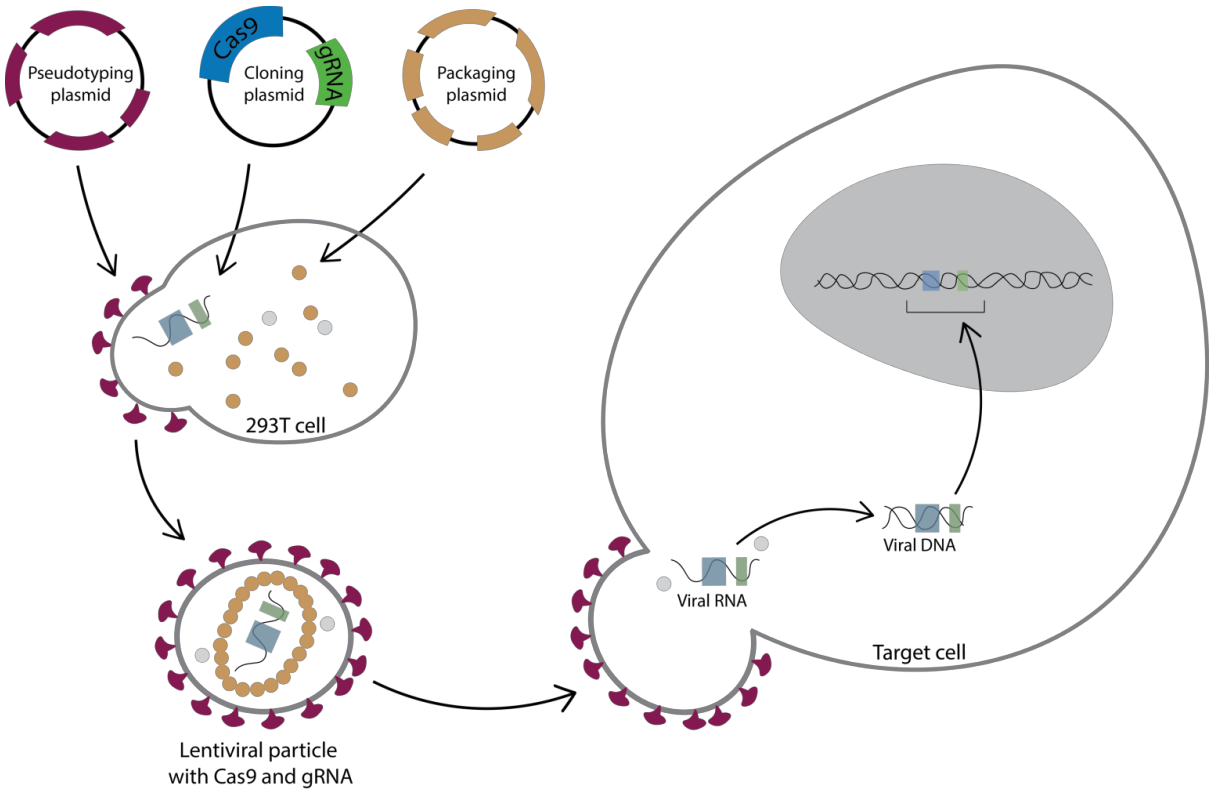


Figure 8: Schematic representation of lentivirus production and transduction.
Lentivirus particles are produced by transfecting three plasmids into HEK293FT cells: 1) a pseudotyping plasmid pMD2.G encodes an envelope glycoprotein (purple), 2) a transfer plasmid that encodes a stable Cas9 gene (blue) and sgRNA of interest (green), and 3) a packaging plasmid psPAX2 that encodes capsid proteins (brown) and reverse transcriptase enzyme (grey). Following the transfection of three plasmids, the 293FT cells express the components of three plasmids and assemble the viral proteins along with the sgRNA to produce lentivirus particles. The lentivirus is transduced into a cell line of interest and the viral genome is released into the cytoplasm of the cell. The viral RNA genome (sgRNA) is reverse transcribed into viral DNA using the reverse transcriptase enzyme (grey) and it is imported and integrated into the genome of the cell. This figure is adapted and re-created from the tutorial series titled "How to create knockouts using CRISPR" (https://www.benchling.com/blog/how-to-express-crispr-in-your-target-cells) by Roger Pellegrini et al. March 23, 2016.

In brief, HEK293FT cells were seeded onto 4X T175 flasks at a density of 1.8*10^7 cells per flask in a total volume of 45 ml DMEM medium (high-glucose) and were incubated at 37°C for 20-24 h before the day of transfection. The next day, the following lentiviral target mixture was prepared for the lentivirus plasmid transfection:

Component	Amount per T175 flask
Opti-MEM medium	2250 µl
pMD2.G plasmid	15.3 µg
psPAX2 plasmid	23.4 µg
sgRNA plasmid library (A+B)	30.6 µg

The mixture was prepared in a 50-ml falcon tube and was vortexed briefly. In the next step, the PLUS reagent mix was prepared as follows:

Component	Amount per T175 flask
Opti-MEM medium	2250 µl
PLUS reagent	297 µl

The prepared PLUS reagent mix was combined with the lentiviral target mixture and incubated at room temperature for 5 minutes. Thereafter, the transfection reagent mixture was prepared as follows:

Component	Amount per T175 flask
Opti-MEM medium	4500 µl
Lipofectamine LTX	270 µl

The transfection reagent mixture was combined with the lentiviral target and PLUS reagent mixture and was incubated at room temperature for 5 minutes. To each T175 flask, 9 ml of the lentiviral transfection mixture was added very carefully without dissociating the cells, and the flasks were incubated at 37°C for 4 h. Thereafter, the medium was removed carefully and replaced with 45 ml of pre-warmed DMEM medium. After 48 h of transfection, the supernatant was collected from all the flasks in 50 ml falcon tubes and was centrifuged at 500g for 5 minutes to pellet any cell debris. The lentivirus supernatant was then filtered using 0.45 µm sterile syringe filters, aliquoted in fresh tubes, and stored at -80°C for storage up to one year.

Before the generation of the genome-wide knock out cell pool, the lentivirus was titered in Huh7 cells through transduction and puromycin selection. Huh7 cells were seeded onto 6-well plates at a density of 3*10^6 cells / well in 3 ml DMEM medium + 10% FBS and were incubated overnight at 37°C. This seeding density was very important later on for the generation of genome-wide knock out cells through transduction with a selective lentivirus titer. To each well, 8 µg/ml polybrene was added shortly before the transduction. Polybrene is a cationic polymer that enhances lentiviral transduction by neutralizing the effect on viral particles and facilitating their attachment to the host cells [110]. To determine the viral titer, various volumes (25 µl to 1 ml) of lentivirus supernatant were added to each well of the 6-well plate and one well was left uninfected. The plates were incubated at 37°C for 24 h. After 24 h of transduction, the cells from all virus volume conditions and uninfected control were trypsinized, resuspended in a 2 ml DMEM medium, and counted. From each virus volume and uninfected condition, 4*10^3 cells in 100 µl DMEM medium were seeded onto a 96-well plate in quadruplicates. Two wells of each condition were added with an addition of 100 µl of DMEM medium + 1µg/ml puromycin and the other two wells were added with 100 µl DMEM medium only. After 96 h of puromycin selection,

the cell viability was measured for all virus volumes and uninfected conditions using the CellTiter Glo kit according to the manufacturer's protocol. The luminescence was measured using a fluorescent plate reader - infinite 200 (application - Tecan i-control). The raw luminescence values from all the conditions were subtracted from the background luminescence value (medium only). The cell viability for each virus volume condition was calculated using the following formula:

Cell viability = corrected luminescence value of puromycin selected well / corrected luminescence value of medium-only well

The mean cell viability (for each virus volume condition) was calculated from the technical duplicates of puromycin-selected wells and medium-only wells

The viral titer (MOI) was calculated based on the ratio of the mean cell viability of puromycin-selected wells to the mean cell viability of medium-only wells.

4.6.3 Generation of genome-wide CRISPR-Cas9 knock out cell pool

To generate a genome-wide CRISPR-Cas9 knock out cell pool, the sgRNA lentivirus library was used to transduce Huh7 cells at an MOI of 0.3. The rationale behind the lower MOI was to ensure that, statistically, most cells would receive only one genetic perturbation on average. To ensure that each sgRNA was represented by ~500 cells, about 2*10^8 Huh7 cells were transduced at an MOI of 0.3. To achieve this cell number for transduction, the following formula was used:

Cell number for transduction = (123,411 (number of sgRNAs in the library) * 500 (coverage of sgRNA per cell)) / 0.3 (MOI used for transduction)

In brief, 3*10^6 Huh7 cells were seeded onto each well of a 6-well plate in a 2 ml DMEM medium. In total, 11X 6-well plates were used to seed 2*10^8 Huh7 cells. To each well, 8 µg/ml polybrene was added shortly before the transduction and thereafter, 250 µl (corresponds to MOI 0.3 – section 5.2.2) of sgRNA lentivirus library was added to each well. The plates were incubated at 37°C for 24 h and after 24 h of transduction, all the cells from 11X 6-well plates were pooled together and counted. The cells were re-plated onto 40X 145 mm dishes at a density of 5*10^6 cells / dish in 20 ml fresh DMEM medium + 1 µg/ml puromycin. To achieve a maximal knock out efficiency, the cells were selected for puromycin resistance for 7 -10 days by freshly adding DMEM medium + 1 µg/ml puromycin every 48 h. After 10 days of puromycin selection, the cell monolayers were confluent again for the screen selection. If the cells were to be passaged, care was taken to maintain the cell population at a density of 80*10^6 (123,411 * 650) cells to maintain the sgRNA coverage (about 650 cells / sgRNA). The transduction with the sgRNA lentivirus library was performed three times to generate a biologically independent genome-wide knock out cell pool for screen selection replicates.

4.6.4 Screen selection

For HCoV-229E, MERS, and thapsigargin screen selection, 80*10^6 (650 cells / sgRNA – library coverage) library transduced cells were used as an unselected control that represented the full sgRNA library. For 7 days of HCoV-229E screen selection, 80*10^6 library transduced cells were challenged with HCoV-229E at an MOI 1 and were incubated at 33°C for 7 days. In addition, 40*10^6 library transduced cells were infected with heat-inactivated HCoV-229E (50°C for 20 minutes) and were incubated at 33°C for 48 h (as a negative control). Thereafter, the heat-inactivated HCoV-229E challenged library transduced cells were harvested for the isolation of genomic DNA. The surviving cells from the 7 days of the HCoV-229E challenge were pooled together and counted using a Neubauer chamber to determine the survival percentage. The cells were further expanded up to four weeks until they reached a cell density of about 80*10^6 in total. Thereafter, the genomic DNA was isolated from the surviving cells.

For 72 h of HCoV-229E screen selection, 2*10^8 (1,600 cells / sgRNA – library coverage) library transduced cells were challenged with HCoV-229E at an MOI 1 and were incubated at 33°C for 72 h. The surviving cells from 72 h of the HCoV-229E challenge were pooled together and counted to determine the survival percentage. The cells were further expanded up to two weeks until they reached a cell density of about 80*10^6 in total. Thereafter, the genomic DNA was isolated from the surviving cells.

For MERS screen selection, 2*10^8 (1,600 cells / sgRNA – library coverage) library transduced cells were challenged with MERS at an MOI of 0.01 and were incubated at 37°C for 40 h. The surviving cells were pooled together and counted to determine the survival percentage. The cells were further expanded up to two weeks until they reached a cell density of about 80*10^6 in total. Thereafter, the genomic DNA was isolated from the surviving cells.

For thapsigargin screen selection, 2*10^8 (1,600 cells / sgRNA – library coverage) library transduced cells were challenged with HCoV-229E (MOI 1) and/or thapsigargin (1 µM) and were selected for 72 h (without expansion) and 96 h (surviving cells were expanded). The genomic DNA was isolated from all the surviving cells.

4.6.5 Preparation of genomic DNA from the surviving cells for NGS

The genomic DNA from all surviving cells was isolated using the Quick genomic DNA midiprep plus kit according to the manufacturer's instructions. The gDNA was eluted in a total of 300 µl ddH_2O and the concentration was measured using the Nanodrop. In the next step, the gDNA was PCR amplified to determine the sgRNA distribution by NGS. For each PCR reaction, 2 µg of gDNA was used as an input and the reactions were scaled up, such that all the gDNA isolated from the surviving cells were amplified. The following conditions were used to amplify the gDNA:

Component	Volume	Final concentration
NEB High Fidelity PCR master mix, 2X	25 µl	1X
gDNA	varies	2 µg
Pooled NGS forward primers (F1-F10) (Section 3.4)	1.25 µl	0.25 µM
NGS reverse primer (unique)	1.25 µl	0.25 µM
ddH_2O	Adjusted to 50 µl	-
Total	50 µl	-

The prepared mixture was PCR amplified on a thermocycler using the following conditions:

Step	Temperature and time	Repeat
1	95°C for 5 minutes	-
2	98°C for 20 seconds	-
3	60°C for 15 seconds	-
4	72°C for 15 seconds	From step 2 for 25X
5	72°C for 1 minute	-

The PCR reactions were pooled together after the amplification in a 50 ml falcon tube and were purified using the Zymo-Spin V with reservoir. To the pooled PCR reactions, 5 volumes of DNA-binding buffer were added and mixed well. Each Zymo-Spin V column was loaded with 12 ml of the mixture and was scaled up according to the volume to be purified. The Spin columns were centrifuged at 500g for 5 minutes at room temperature and the supernatant was discarded. The columns were washed twice with 2 ml of DNA wash buffer and were centrifuged at 500g for 5 minutes at room temperature. The reservoir from the spin columns was removed and transferred to 2 ml collection tubes and centrifuged at 12,000g for 2 minutes to remove the residual wash buffer. Thereafter, the purified PCR product from the reservoirs was eluted in 150 µl / reservoir of ddH_2O and the concentration was measured using the Nanodrop.

Sample ID	Amount of cells harvested	Concentration of gDNA (ng/µl)	Total concentration of gDNA (µg)	No. of PCRs used for library preparation	Vol of purified PCR prep (ml)	Concentration of purified PCR product (ng/µl)
Unselected control_1	80*10^6	241.87	532.1	266	0.75	459
HCoV-229E selection for 72 h	38*10^6	167.53	331.7	166	0.4	476
Unselected control_2	80*10^6	170.91	376.0	188	0.5	592.9
HCoV-229E selection for 96 h	58.5*10^6	203.53	386.7	193	0.9	207

Figure 9: Illustrative figure showing a detailed view of the yield of gDNA from the control and / or selected samples.
The table shows the cell numbers, the yield of total gDNA, and the number of PCR reactions needed for sgRNA library preparation for the unselected control samples and the surviving cells after selection.

The purified PCR product (3 µg) was loaded on a 2% agarose gel and was separated for 30 minutes. The 270 bp PCR product was excised from the gel and was purified using the NucleoSpin Gel and PCR purification kit. The purified gel extracted samples were quantified using the Qubit dsDNA High-sensitivity Assay kit according to the instructions provided in the kit. In addition, as a quality control, the purified samples were loaded on a high-sensitivity DNA chip and were run on the Agilent 2100 Bioanalyzer, before the samples were pooled and sent to the NGS for deep sequencing.

The samples were run on an Illumina NextSeq 550 (Institute of Medical Microbiology, Giessen) and / or the NovaSeq 660 platform (Novogene) according to the user manual instructions. The paired-end sequencing (PE150) run was spiked with 20% PhiX as an internal control and was aimed for 50 – 60 million reads as a sequencing depth per sample.

4.6.6 Analysis of the NGS data from the screen using the MAGeCK tool

Model-based Analysis of Genome-wide CRISPR-Cas9 knock out (MAGeCK) is a computational tool that was used to analyze the NGS data from the screen to determine the sgRNA read distribution, sgRNA enrichment in the survivor cells, and to identify the important genes based on the sgRNA enrichment [99, 111, 112]. The following steps were systematically followed to setup MAGeCK:

In the first step, Miniconda 3 was installed in a Linux environment using Python (version 3.7). Miniconda is a mini variant of Anaconda, which is used as a package manager (collection of scientific software) and deployment. It consists of python, conda, and a few basic packages (pip, zlib, and a few others for easy installation of packages). Using the conda / bioconda channel, MAGeCK was installed using the following command line:

```
conda install –c bioconda mageck
```

In the next step, the environment was activated using the following command line:

```
source activate magecknv
```

Optionally, the mageck software can be updated and deactivated using the command lines:

```
conda update mageck

source deactivate
```

After the successful installation of mageck environment, the sgRNA read counts were generated using the fastq files from the NGS run as follows:

```
mageck count –l sgRNA libraray.csv –n output --sample-label HCoV-229E selected,control --fastq_HCoV-229E selected.fq.gz, fastq_control.fq.gz
```

Parameters	Meaning
mageck	The main portal of the MAGeCK program
count	A sub-command used by MAGeCK to generate sgRNA read count table
-l sgRNA library.csv	The input file created with sgRNA ID, the sequence, and the gene it is targeting in separate columns
-n Output	The prefix of the output file which contains the sgRNA read count tables
--sample-label HCoV-229E selected,control	The labels of the two samples (fastq files from the NGS run)
--fastq_HCoV-229E selected.fq.gz, fastq_control.fq.gz	The provided fastq files. Technical replicates of the same sample were indicated using a comma as a separator

The output of the MAGeCK count command run was generated as a .txt file and this file was imported to an excel sheet to visualize the sgRNA raw read and normalized read count tables as shown below (example):

Raw read counts table

sgRNA_ID	Gene	Control_1	Control_2	HCoV-229E_1	HCoV-229E_2
HGLibA_02209	ANPEP	28236	125845	10559862	1345017
HGLibA_02210	ANPEP	29783	105664	7724456	4841709
HGLibA_02211	ANPEP	18314	27401	1590001	6378989
HGLibB_02207	ANPEP	808	878	38166	4592
HGLibB_02208	ANPEP	16070	44350	3438383	3464483
HGLibB_02209	ANPEP	804	1247	75691	40034

Normalized read counts table

sgRNA_ID	Gene	Control_1	Control_2	HCoV-229E_1	HCoV-229E_2
HGLibA_02209	ANPEP	3799.930346	20799.89154	72363990.52	13396553.96
HGLibA_02210	ANPEP	4008.121741	17464.33898	52933689.93	48224086.3
HGLibA_02211	ANPEP	2464.652371	4528.887346	10895863.72	63535606.13
HGLibB_02207	ANPEP	108.7386216	145.1174442	261541.6814	45736.95038
HGLibB_02208	ANPEP	2162.660457	7330.249034	23562345.31	34506726.27
HGLibB_02209	ANPEP	108.2003116	206.1064385	518690.7563	398744.1358

In the next step, with the generated normalized read count table, the MAGeCK test command analyzed the enrichment / depletion of sgRNAs in the selected samples compared to the unselected controls. Finally, based on the significance of sgRNA enrichment / depletion, the genes corresponding to the sgRNAs were ranked from top to bottom order.

The MAGeCK test was run using the following command line:

```
mageck test –k sgRNA_readcount.txt –t HCoV-229E_1,HCoV-229E_2 –c Control_1,Control_2 –n HCoV-229EvsControl --pdf-report
```

Parameters	Meaning
mageck	The main portal of the MAGeCK program
test	A sub-command used by MAGeCK to analyze and generate sgRNA and gene summary files
-k sgRNA_readcount.txt	The input file containing the raw and normalized sgRNA read counts
-t HCoV-229E_1,HCoV-229E_2	The labels of the two experimental samples
-c Control_1,Control_2	The labels of the two control samples
-n HCoV-229EvsControl	The prefix of the output file
--pdf-report	To generate the analyzed file as a PDF report

The output of the mageck test command run was generated as a .txt file. This file was imported to an excel sheet to visualize the sgRNA and gene summary as shown below (example):

sgRNA summary file

sgrna	Gene	control_count	treatment_count	control_mean	treat_mean	LFC	control_var	adj_var	score	p.low	p.high	p.twosided	FDR	High in treatment
HGLibA_02209	ANPEP	3799.9/20800	7.2364e+07/1.3397e+07	12300	4.29E+07	11.767	1.45E+08	38576	2.18E+05	1	0	0	0	True
HGLibA_02210	ANPEP	4008.1/17464	5.2934e+07/4.8224e+07	10736	5.06E+07	12.202	9.05E+07	33422	2.77E+05	1	0	0	0	True
HGLibA_02211	ANPEP	2464.7/4528.9	1.0896e+07/6.3536e+07	3496.8	3.72E+07	13.377	2.13E+06	10247	3.68E+05	1	0	0	0	True
HGLibB_02207	ANPEP	108.74/145.12	2.6154e+05/45737	126.93	1.54E+05	10.23	661.71	314.48	8656.6	1	0	0	0	True
HGLibB_02208	ANPEP	2162.7/7330.2	2.3562e+07/3.4507e+07	4746.5	2.90E+07	12.578	1.34E+07	14137	2.44E+05	1	0	0	0	True
HGLibB_02209	ANPEP	108.2/206.11	5.1869e+05/3.9874e+05	157.15	4.59E+05	11.502	4792.8	393.39	23120	1	0	0	0	True

gene summary file

id	num	neg|score	neg|p-value	neg|fdr	neg|rank	neg|goodsgrna	neg|lfc	pos|score	pos|p-value	pos|fdr	pos|rank	pos|goodsgrna	pos|lfc
ANPEP	6	1	1	1	21705	0	11.985	7.33E-17	2.28E-07	0.00495	1	6	11.985
NOL12	6	0.81998	0.83395	0.99885	17929	1	4.188	6.35E-06	3.76E-05	0.311881	2	5	4.188
KAT7	6	0.27343	0.47021	0.99885	9583	1	4.0864	7.01E-06	4.31E-05	0.311881	3	5	4.0864
SFTA2	6	0.79099	0.81597	0.99885	17487	1	0.30655	2.56E-05	0.00014802	0.75884	4	4	0.30655
GOLGA3	6	0.36709	0.54125	0.99885	11225	1	4.0315	3.19E-05	0.00018269	0.75884	5	5	4.0315
ACE	6	0.99996	0.99995	1	21704	0	1.0134	4.30E-05	0.00024108	0.75884	6	6	1.0134
TIMELESS	6	0.99996	0.99995	1	21703	0	1.3337	4.40E-05	0.00024473	0.75884	7	6	1.3337
HDAC4	6	0.91184	0.91186	0.99885	19774	0	0.94647	6.73E-05	0.00036014	0.861333	8	5	0.94647
CYP4Z1	6	0.99984	0.99983	0.999964	21702	0	0.72769	0.00016202	0.00088564	0.861333	9	6	0.72769
SLC5A1	6	0.60924	0.73575	0.99885	15371	1	1.3657	0.00017844	0.0009714	0.861333	10	5	1.3657

Figure 10: Representative figure showing the sgRNA and gene summary excel file as an output file generated from the mageck test command run.

The sgRNA summary file consists of columns with the information on the sgRNA ID, the targeting gene, the normalized read counts from the unselected and selected samples (biological replicates), and the enrichment scores (LFC, score, and variance). Based on the sgRNA score, the gene ranking is provided in the table below with the columns containing the gene ID, the number of sgRNAs targeting the gene and the rank, the MAGeCK score, and the statistics (p-value and FDR).

4.6.7 Generation of individual gene knock out cell lines to validate the screen

The knock out cell lines that were mentioned in sections 5.3 and 5.6 were generated by picking the top enriched sgRNA sequences from the sgRNA summary file of the respective screen selection. The sgRNAs were ordered as oligonucleotides as mentioned in section 3.4 and were cloned into the lentiCRISPR V2 vector using the protocol mentioned in section 4.4.2. In step 3 of restriction and ligation, instead of the px459 vector, the lentiCRISPR V2 vector was used and FastDigest Esp3I was used as the restriction enzyme.

The successfully cloned sgRNA plasmids were used to produce lentivirus for transduction. The protocol mentioned in section 4.6.2 was used in the production of lentivirus, except the reagents and materials were scaled down (7X) to use with the T25 flasks instead of T175 flasks. The lentivirus was titered and transduced in Huh7 cells at MOI 0.5 to 1. The transduced cell lines were selected for puromycin resistance by maintaining them with 1 µg/ml puromycin for 7 to 10 days to generate the individual sgRNA-carrying knock out cell lines for validation.

5. Results

5.1 Identification and validation of induced gene sets from the microarray data of HCoV-229E infected Huh7 cells

In the previous finding by our working group, HCoV-229E infected Huh7 cells induced the mRNA expression of 1,073 genes at the genome-wide level as measured by four independent microarray experiments [113]. Based on the microarray analysis, a specific set of genes was identified to be interesting to validate their functional relevance in virus and / or host response. The gene sets were sorted out by the following selection criteria:

1. Genes with mRNA expression levels of $\log_2$ fold change > 2
2. No previously reported putative functions related to coronavirus infections
3. Genes that have a metabolic enzyme or a transcription factor activity
4. Commercial available antibodies or reagents to validate the protein / gene function

The list of 11 genes that were selected based on the above criteria was shown in the table below. These genes showed a strong induction in the mRNA expression upon HCoV-229E infection after 24 h in all of the four independent microarray experiments performed. The role of these gene sets in coronavirus infections remains elusive to this date. Moreover, the selected gene sets have their putative protein function as a metabolic enzyme or a transcription factor, which raised the question of whether HCoV-229E depends on these factors for its replication and/or virus-host response.

		HCoV-229E / untreated				HCoV-229E / untreated
gene	**gene name**	**exp 1**	**exp 2**	**exp 3**	**exp 4**	**average ratios**
ANKRD1	ankyrin repeat domain 1 (cardiac muscle)	**5.7**	**2.8**	**4.7**	**5.9**	**4.8**
FICD	FIC domain containing	**5**	**4.7**	**4.2**	**5**	**4.7**
FUT1	fucosyltransferase 1 (galactoside 2-alpha-L-fucosyltransferase, H blood group)	**3.5**	**4.1**	**4.8**	**6.1**	**4.6**
KLF6	Kruppel-like factor 6	**4.9**	**3.3**	**4.5**	**5.4**	**4.5**
CHAC1	ChaC, cation transport regulator homolog 1 (E.coli)	**4.9**	**3.7**	**3.8**	**4.1**	**4.1**
BHLHE41 /DEC2	basic helix-loop-helix family, member E41	**3.3**	**1.9**	**4.1**	**4.3**	**3.4**
ERO1LB	ERO1-like beta (S.cerevisiae)	**2.7**	**2.2**	**3.2**	**3.7**	**3**
BHLHE40 /DEC1	basic helix-loop-helix family, member E40	**3.3**	**1.8**	**2.4**	**2.9**	**2.6**
CTH	cystathionase (cystathionine gamma-lyase)	**2.1**	**1.8**	**2.3**	**3.5**	**2.4**
EIF2AK3	eukaryotic translation initiation factor 2-alpha kinase 3	**2.7**	**1.0**	**2.8**	**3.2**	**2.4**
EDEM1	ER degradation enhancer, mannosidase alpha-like1	**2.7**	**2.3**	**1.8**	**2.8**	**2.4**

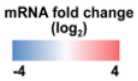


Table 1: List of selected genes with log_2 transformed x-fold changes and the average ratios in the mRNA expression from four independent microarray experiments of infected Huh7 cells.

5.1.1 Assessing the protein expression levels of selected gene sets upon HCoV-229E infection

It was very evident from the microarray analysis that the mRNA expression of the selected gene sets was induced upon HCoV-229E infection. To understand whether HCoV-229E induced transcriptional induction of these genes was also reflected at the translational level, a time kinetic infection experiment was performed in Huh7 cells for 3, 6, 9, 12, 16, 20, and 24 h. As a comparison, Huh7 cells were treated with IL-1 (10 ng/ml) for 1 h to understand if these factors were also regulated by proinflammatory cytokine signaling. After the time kinetics, total cell lysates were prepared from each individual time point of infections and IL-1 treatment. To visualize the protein expression levels of the gene sets, Western blotting was performed according to the protocol mentioned in the methods section. The immunoblots were probed with the following antibodies: anti-CARP / ANKRD1, anti-EDEM1, anti-KLF6, anti-CHAC1, anti-PERK, anti-DEC1 / SHARP2, anti-CTH, and anti-CREB2 / ATF4. At this point, the antibodies for other factors DEC2, FICD, ERO1LB, and FUT1 did not produce a distinct and interpretable band at the desired size on the immunoblots. As shown in Fig. 11, the protein levels of N protein, nsp 12, and 8 indicated that these viral components were detected first after 9 h p.i. and gradually accumulated until 24 h p.i. In contrast, there were three different scenarios observed for the protein levels of the selected factors. As indicated in red in Fig. 11, ANKRD1 and KLF6 showed an increased expression at 24 h p.i. compared to the uninfected control. In contrast, EDEM1, PERK, DEC1 / SHARP2, and BiP / GRP78 (previously reported to be specifically downregulated by HCoV-229E [64] and used in this experiment as a control) marked in blue in Fig. 11 showed decreased protein levels at 24 h p.i. when compared to the uninfected control. The protein levels of CHAC1, CTH, and CREB2 / ATF4 showed a steady state of expression until 24 h p.i. There were no changes in the protein levels of the host factors upon IL-1 treatment except for KLF6, where the treatment caused an increase in KLF6 protein levels when compared to the untreated control.

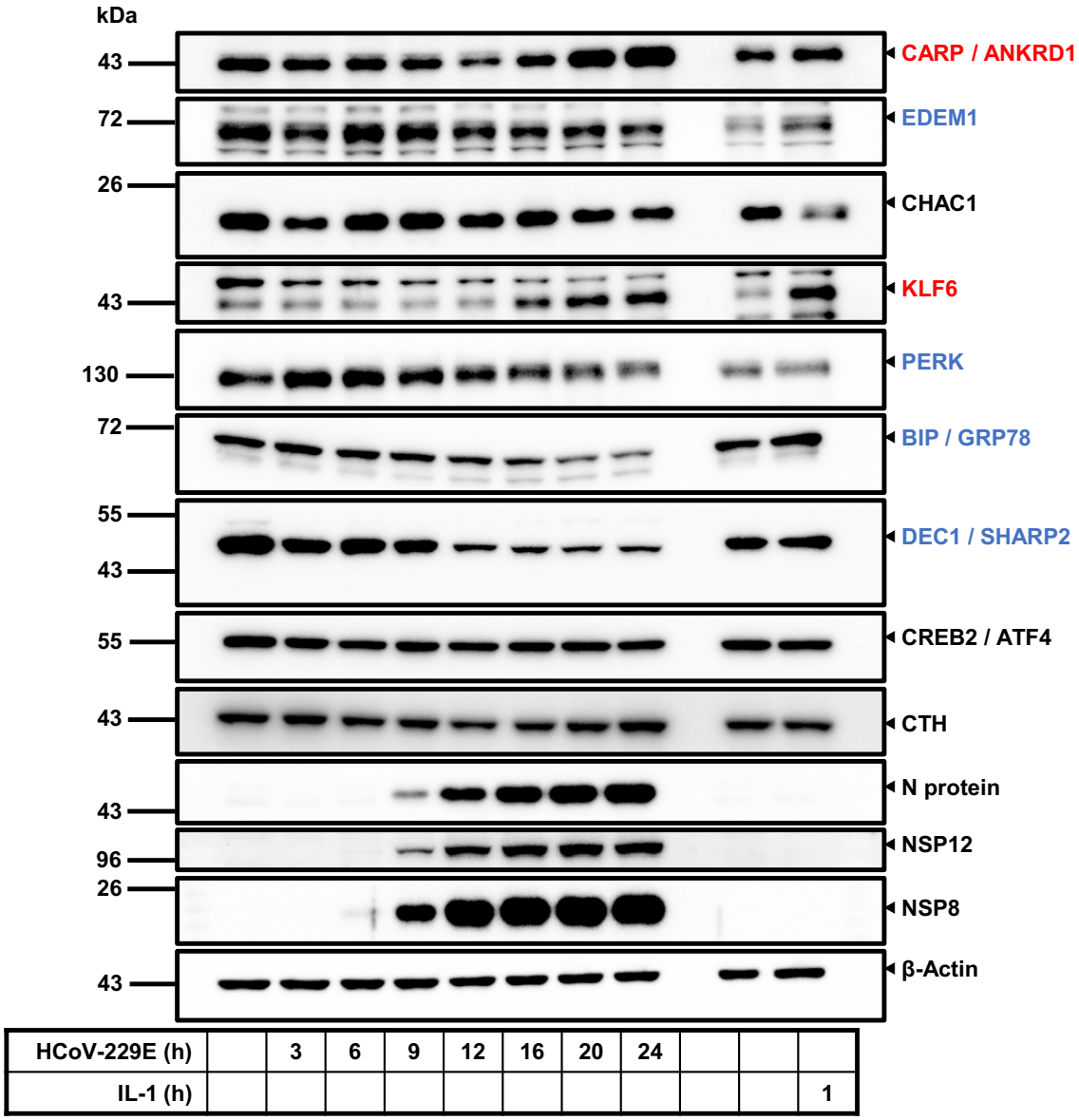


Figure 11: Differential protein expression levels of HCoV-229E induced gene sets.
Huh7 cells were infected with HCoV-229E (MOI = 1) for indicated time points or left uninfected (control). IL-1 (10 ng/ml) was treated separately to Huh7 cells for 1 h or cells were left untreated (control). Total cell extracts were analyzed by immunoblotting for the indicated antibodies. ß-Actin was used as a loading control for all the samples.

5.1.2 Generation and characterization of stable knock out cell lines for the selected gene sets

To obtain more insights into the functional relevance of the host factors in HCoV-229E viral replication and / or virus-host response, protein loss-of-function studies of the host factors were selected as a preferential approach using a genetic system. CRISPR-Cas9 gene-editing technique was used to create stable knock out cell lines of these host factors. For each of the factors, many different single-guide RNAs targeting the gene at multiple / same exon regions were designed. In addition, to achieve maximum efficiency in gene editing, two sgRNAs binding to the same exon at nearby regions (between 50 to 100 nucleotides) were also designed. The sgRNAs were cloned into the plasmid px459 v2.0 vector that carried a stable Cas9 expressing cassette and an antibiotic (Puromycin) resistance cassette for selection. For creating the stable knock out cell lines, a chemical-based transfection technique using calcium phosphate ($Ca_3(PO_4)_2$) was carried out according to the detailed protocol stated in the methods section. After a successful transient transfection of sgRNA encoding px459 v2.0 plasmids in Huh7 cells, the cells were selected for puromycin resistance by adding fresh 1 µg/ml puromycin to the culture medium every 48 h until the non-transfected cells were removed from the cell pool. To verify the loss of protein expression mediated by CRISPR-Cas9 gene editing, total cell lysates were prepared from the cell lines selected for puromycin for about three to four weeks. The lysates were SDS-PAGE separated and Western blotting was performed to visualize the protein levels.

For the generation of the ANKRD1 knock out cell line, four different cell lines each carrying a single or multiple sgRNAs were tested for the total loss / reduction in the ANKRD1 protein levels. As shown in the immunoblots from Fig. 12A, none of the sgRNA encoding cell lines showed a complete loss of ANKRD1 protein levels. Nevertheless, ANKRD1_sg1 and ANKRD1_sg1+3 showed a reduction in the ANKRD1 protein levels (Fig. 12A) as compared to the H9C2 cell line (embryonic rat cardiomyocytes which had higher expression of ANKRD1 protein) and empty vector-transfected Huh7 cell line. The H9C2 cell line was chosen as a positive control because it was known that heart cells express high levels of Ankrd1.

In the case of ERO1LB, four different single and/or multiple sgRNAs (in addition, all sgRNAs were pooled together) were used to create a stable knock out cell line. It was difficult to unambiguously visualize the levels of ERO1LB protein, as the antibody probed against ERO1LB produced multiple bands at the expected size in the immunoblots as shown in the Fig. 12B. The expected band size of ERO1LB was 40-50 kDa according to the datasheets provided from the antibody manufacturers. At this point, there were no appropriate positive controls such as ERO1LB over-expressing cell line or an ERO1LB protein lysate to overcome the artefacts produced by this antibody.

To generate knock out cell lines for BHLHE40 and BHLHE41, two different sgRNAs (sg4, and sg5) and a combination of sg4+5 were used. In the case of BHLHE40 (DEC1 / SHARP2), the cell lines sg4 and sg4+5 showed a partial reduction in the levels of DEC1 protein as compared to the Huh7 parental control (Fig. 12C). There was no noticeable reduction in the DEC2 / SHARP1 protein levels for all three-cell lines generated for BHLHE41 (Fig. 12D).

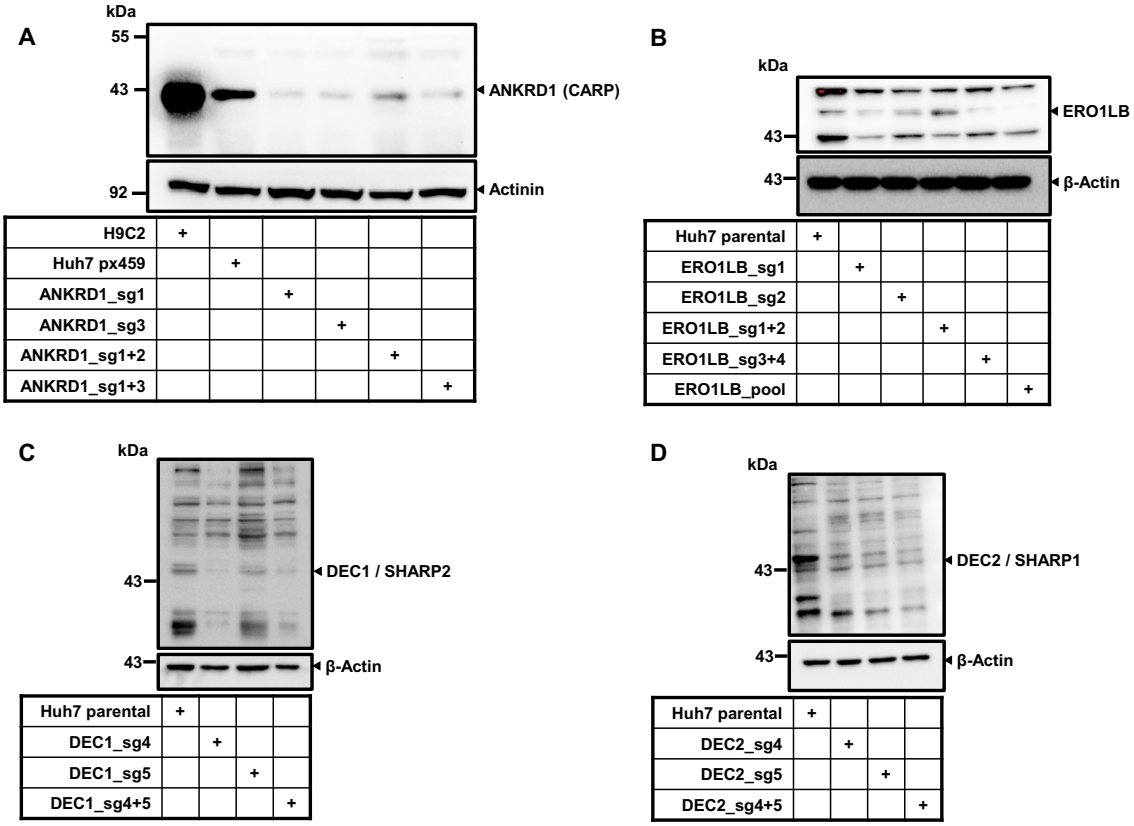


Figure 12: Validation of stable knock out cell lines (ANKRD1, ERO1LB, BHLHE40 / 41) by Western blotting.

A) Huh7 cells were transfected either with px459 empty vector or sgRNAs as indicated. After three to four weeks of puromycin (1 µg/ml) selection, total cell lysates were used to detect the ANKRD1 protein levels. Total cell lysates from the H9C2 cell line were used as a control for ANKRD1 protein expression. Actinin was used here as a loading control. **B, C, D)** Huh7 cells were transfected with the sgRNAs as indicated. Total cell lysates collected after three to four weeks of puromycin selection were analyzed by immunoblotting for the indicated antibodies. ß-Actin was used here as a loading control.

To generate FICD and EDEM1 knock out cell lines, three different sgRNAs (sg1, 2, 3) and their dual combinations (sg1+2, 2+3, 1+3) were used. As shown in Fig. 13A, out of all the single sgRNAs and their combinations, FICD_sg2+3 showed a clear reduction in the HYPE (FICD) protein levels as compared to the Huh7 parental and px459 empty-vector control. In the case of EDEM1, except for sg1+2, all the single sgRNAs and the combinations showed highly reduced EDEM1 protein levels. In particular, sg3 and sg 2+3 showed a total loss of EDEM1 protein levels as shown in the Fig. 13B. The combinations sg1+2, and 3+4 resulted in a reduction of CTH protein levels (in the case of sg1+2) and almost a total loss of protein level (in case of sg3+4) as shown in the Fig. 13C.

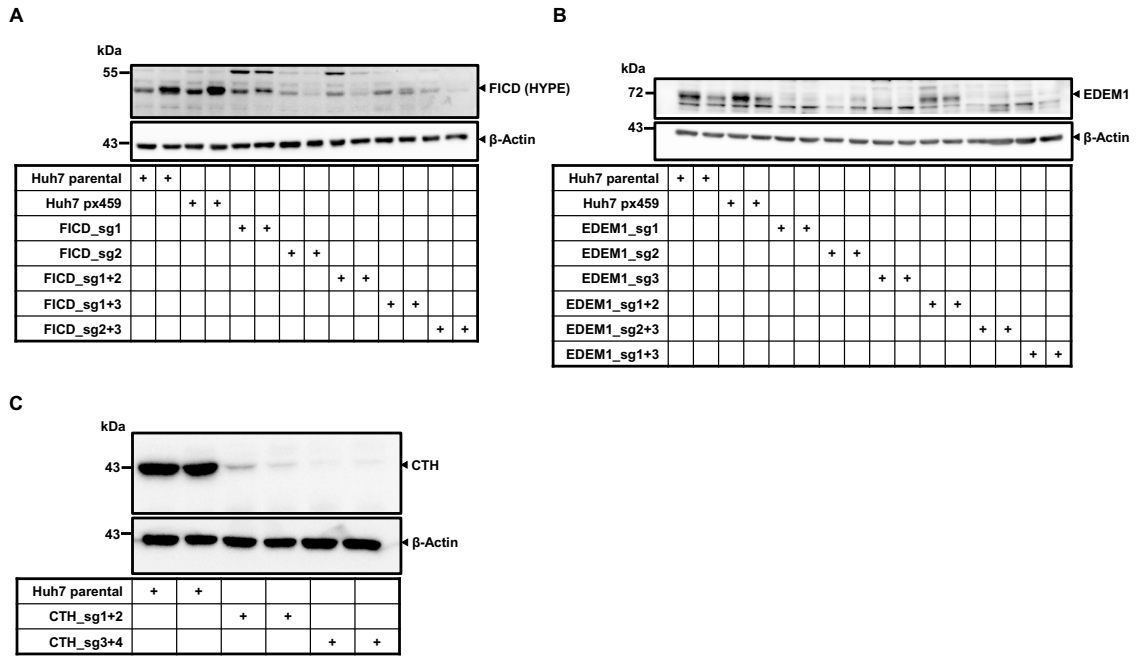
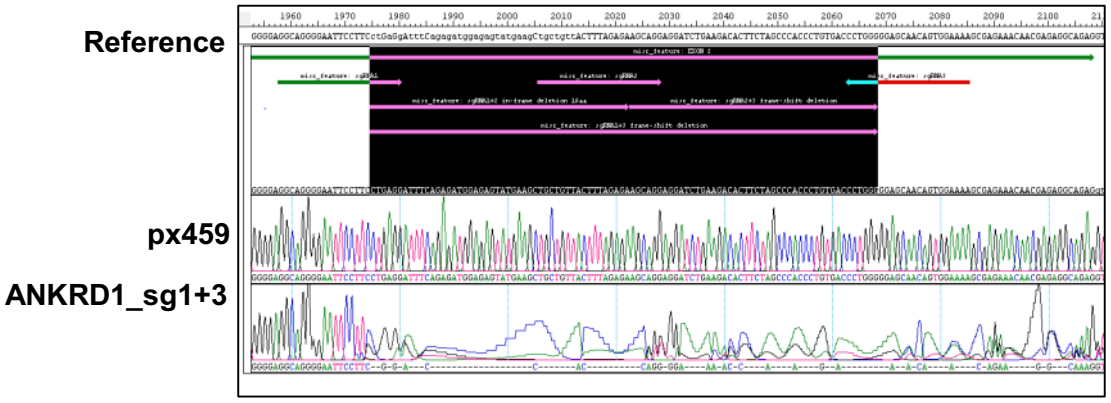


Figure 13: Validation of stable knock out cell lines (FICD, EDEM1, and CTH) by Western blotting.
Huh7 cells were transfected either with px459 empty vector or sgRNAs as indicated. Total cell lysates prepared from the cell lines after three to four weeks of puromycin selection were analyzed by immunoblotting for the indicated antibodies **A)** FICD / HYPE **B)** EDEM1 **C)** CTH. ß-Actin was used here as a loading control.

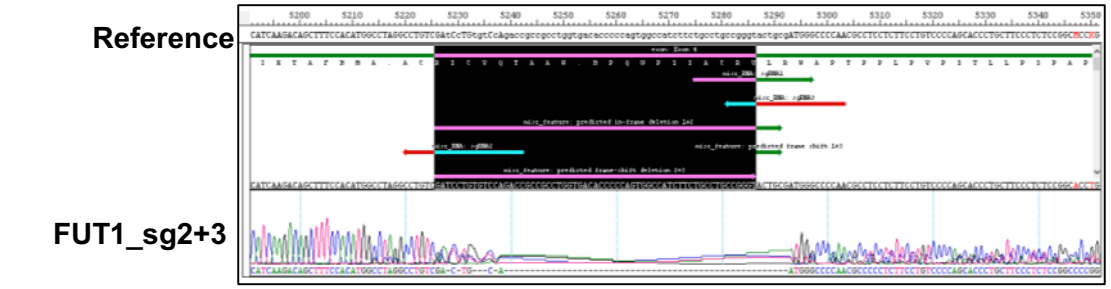
To verify the indel mutations caused by CRISPR-Cas9 gene editing in the knock out cell lines, PCR-based amplicons containing the sgRNA binding site were sequenced by Sanger sequencing. Based on the immunoblots, a specific sgRNA cell line showing the most reduction in their protein levels was selected to verify the indel mutations. In this context, amplicons sequenced from the genomic DNA of the ANKRD1_sg1+3 cell line showed a deletion of around 90 nucleotides and random insertions of nucleotides in between the binding site of sg1 and 3 (Fig. 14A). The same effect was observed in the amplicons sequenced from the FUT1_sg2+3 cell line (deletion of around 60 nucleotides) and for the EDEM1_sg2+3 cell line (deletion of around 37 nucleotides) (Fig. 14B and C). There were no indel

mutations observed in the sequenced amplicons of the FICD_sg2+3 cell line (Fig. 14D), in contrast to the reduction of FICD / HYPE protein levels as observed in the immunoblots.

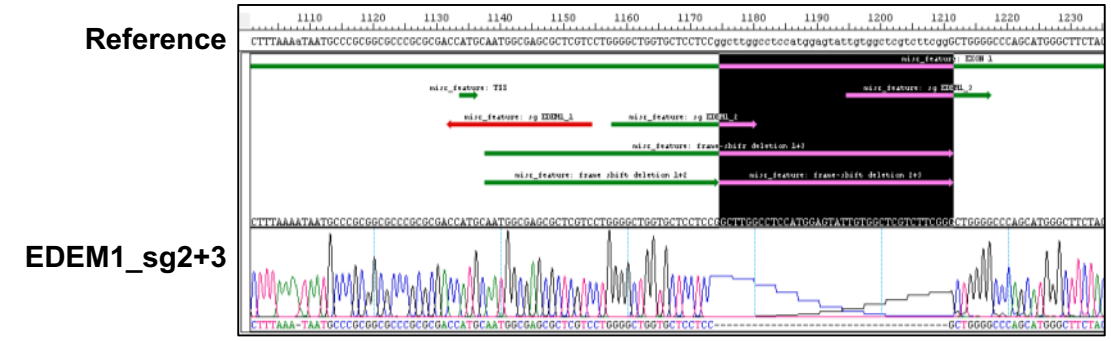
A



B



C



D

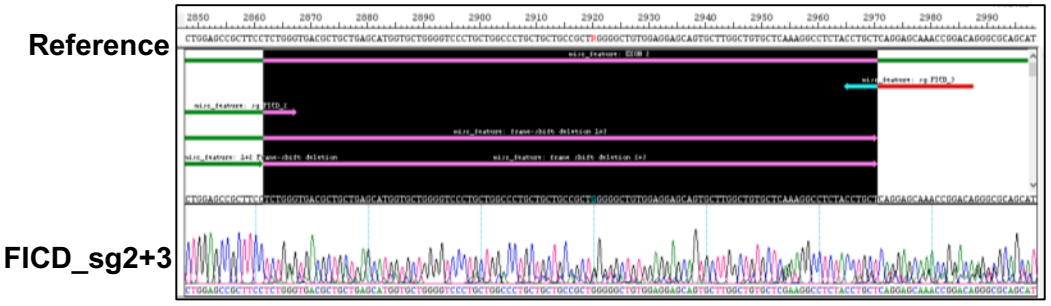


Figure 14: Verification of the indel mutations caused by CRISPR-Cas9 gene editing using Sanger sequencing.
Genomic DNA isolated from the sgRNA encoding cell lines (**A.** ANKRD1 sg1+3, **B.** FUT1 sg2+3, **C.** EDEM1 sg2+3, **D.** FICD sg2+3) was PCR amplified using specific primers to amplify the sgRNA binding sites. The amplicons were sequenced by Sanger sequencing and aligned with the reference gene sequence using DNA STAR Seq-Man Pro to verify the indel mutations.

5.1.3 Analysing the functional relevance of gene knock outs on HCoV-229E replication and virus-host response

Based on the characterization of knock outs by Sanger sequencing and Western blotting, five factors (ANKRD1, EDEM1, FUT1, KLF6, and FICD) were selected to validate their functional relevance in virus replication and / or virus-host response. From the data shown in section 5.1.2, a specific cell line that showed a stable knock out/knock down and indel mutations were tested to see if these factors could affect the HCoV-229E replication. For this observation, the cell lines were either infected with HCoV-229E at a MOI=1 for 24 h or left untreated. Huh7 parental cell line and empty-vector transfected cell line were used as a positive control for this experiment. The total cell lysates were prepared after 24 h of virus infection and the protein levels of the respective factors and HCoV-229E replication components were visualized by Western blotting. As shown in Fig. 15A, the immunoblots probed with the antibodies anti-CARP / ANKRD1, EDEM1, FUT1, and KLF6 indicated that the protein levels of these factors were strongly reduced or completely knocked out for the respective cell lines as compared to the Huh7 parental and empty-vector cell line. However, the reduction in the protein levels of these host factors did not influence the protein levels of HCoV-229E replication component Nsp 8 in any of the knock out cell lines (Fig. 15A). In the case of FICD, the protein levels of HYPE / FICD were strongly induced in Huh7 parental and empty-vector cell line infected with HCoV-229E after 24 h, but a reduction in the HYPE protein levels was observed in the infected FICD_sg 2+3 cell line which indicated a strong knock down (Fig. 15B). Nevertheless, the reduction in HYPE levels did not influence the protein levels of HCoV-229E major structural component (N protein) in FICD_sg2+3 when compared to the infected controls.

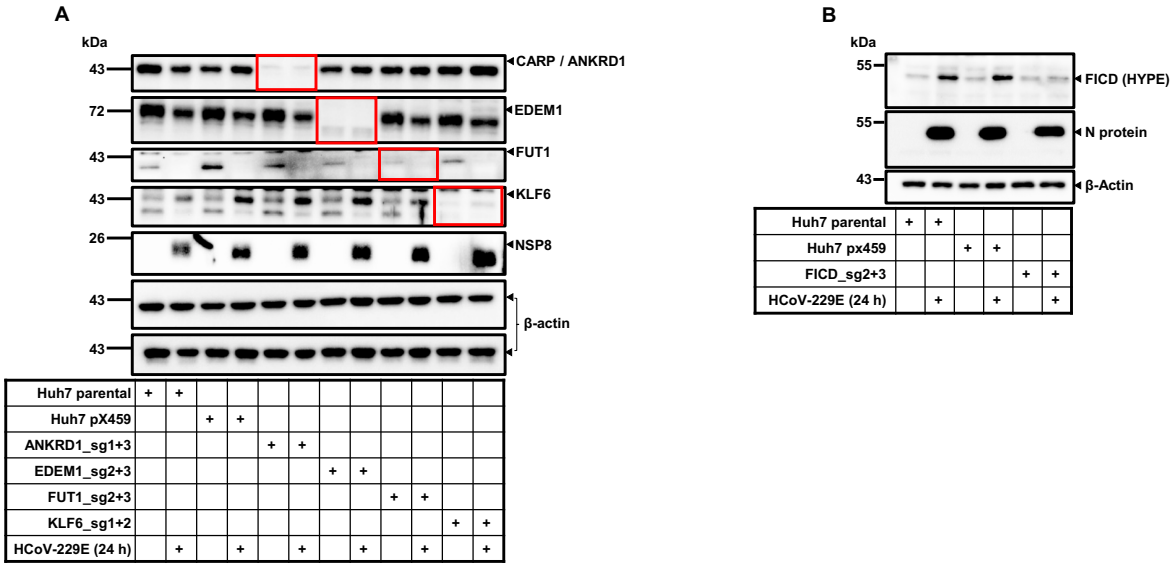


Figure 15: Functional relevance of the gene knock outs on HCoV-229E replication.
After a successful characterization of stable knock outs / knock downs, the cell lines were tested for their role in viral replication. The cell lines were either infected with HCoV-229E at MOI =1 for 24 h or left uninfected along with Huh7 parental and empty vector cell lines as positive controls. Total cell lysates were prepared after 24 h of infection and the immunoblots were probed against the indicated antibodies. **A)** As indicated in red blocks, the antibodies probed against ANKRD1, EDEM1, FUT1, and KLF6 showed a clear reduction or a total loss of their protein levels. **B)** Immunoblot probed against HYPE / FICD showed a strong reduction in the protein levels of HYPE in the infected FICD sg_2+3 cell line as compared to the controls. ß-Actin was used as a loading control. Data shown here were the representative immunoblot from two biologically independent experiments.

These data clearly showed that HCoV-229E does not seem to depend on these five host factors for its viral replication. To further validate the functional relevance of these host factors on virus-host response, an RNA-Seq experiment was performed to understand the changes at the global transcriptome level upon the infection. As shown in Fig. 15A, from the same experiment, total RNA was prepared from all the samples and sent to Novogene for deep sequencing. After the deep sequencing, the sequenced raw data were aligned to the reference human genome using a tool called STAR to determine the read counts for each of the human genes. The median normalized raw reads were then analyzed for differential changes in the gene expression by a tool called DEseq2. To verify if all the samples have a comparable distribution of read counts, a box plot was plotted with the $\log_2$ transformed read counts across all the samples. The box plots indicated that the $\log_2$ transformed mapped raw read counts had a slight variable median distribution across all the samples. However, after median normalization, the samples showed a more even distribution of read counts as expected (Fig. 16).

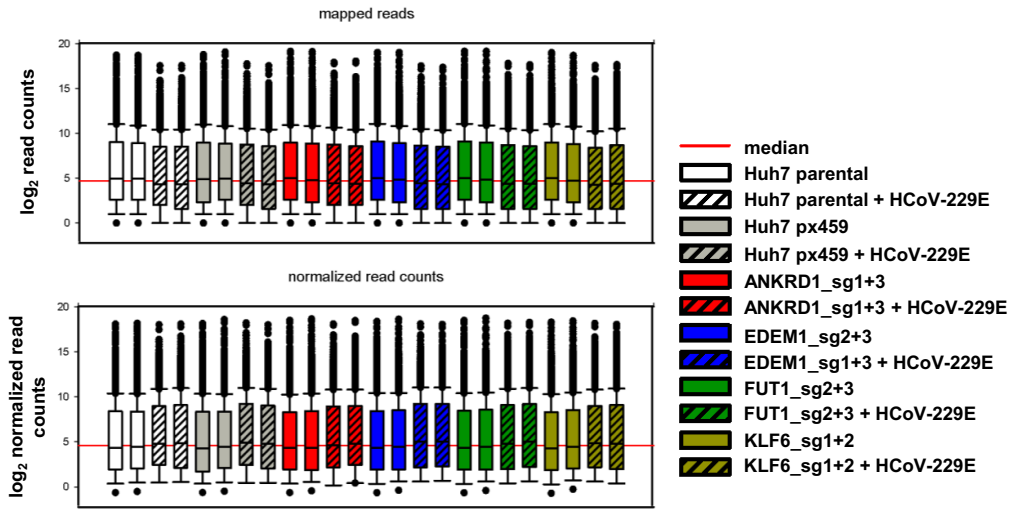
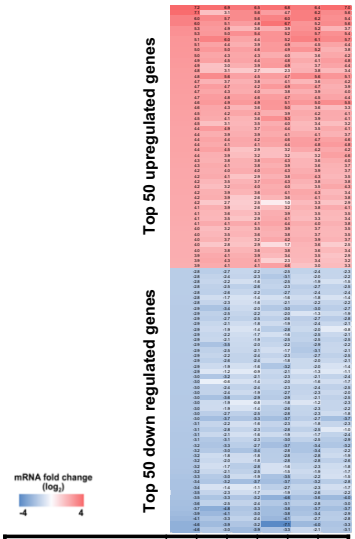


Figure 16: Quality control data for the samples processed after the RNA-Seq.
Using the STAR tool the raw reads were mapped to the human genome sequence and read counts were assigned based on their alignment with the gene. Box plots indicate the median distribution of log_2 transformed mapped read counts before and after normalization across all the samples. Sebastian Werner (Prof. Kracht's working group) kindly generated the quality control plots using the STAR tool.

The normalized read counts across all the samples were log_2 transformed and their ratios for log_2 fold change were pair-wise calculated. To understand the differential changes in the gene expression upon the infection in the absence of host factors, the log_2 fold changes in the global gene expression were compared to the fold changes in the Huh7 parental sample infected with HCoV-229E. The heat map was generated with the top 50 and bottom 50 deregulated genes with a log_2 fold change (< and / or > = 2, p-value = <0.001) in the infected knock out cell lines as compared to the infected Huh7 parental cell line (Fig. 17A). The heat map indicated that there were no obvious significant differences in the regulation of gene expression for the top and bottom selected genes for any of these host factors upon infection as compared to the control.

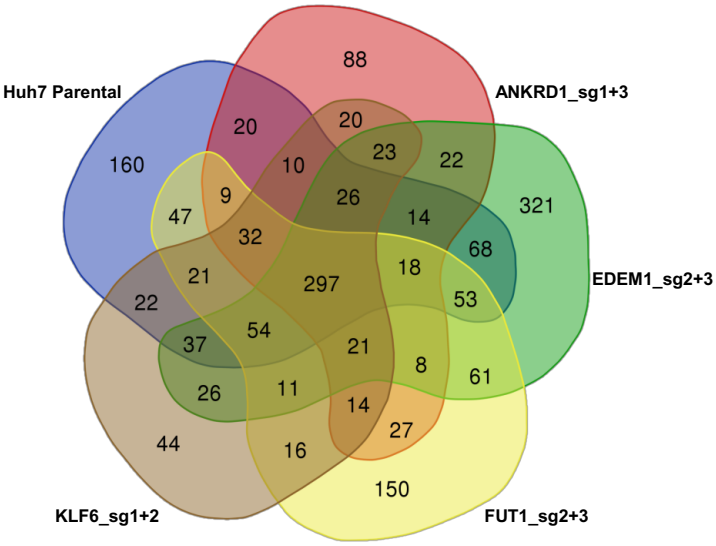
A



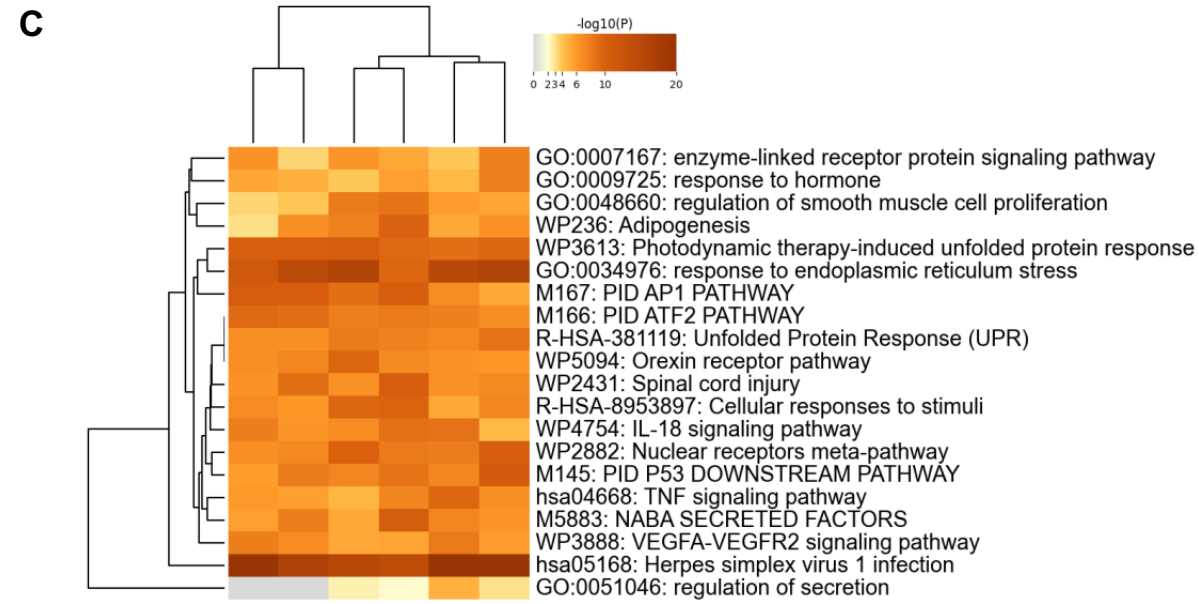
Huh7 parental	+					
Huh7 pX459		+				
ANKRD1_sg1+3			+			
EDEM1_sg2+3				+		
FUT1_sg2+3					+	
KLF6_sg1+2						+
HCoV-229E (24 h)	+	+	+	+	+	+

B

Sorted by genes with fold change < -2 or > 2; p <= 0.001 across all conditions



List names	number of elements	number of unique elements
Huh7 parental	888	888
ANKRD1_sg1+3	649	649
EDEM1_sg2+3	1060	1060
FUT1_sg2+3	839	839
KLF6_sg1+2	674	674
Overall number of unique elements		1740



Huh7 parental	+					
Huh7 pX459		+				
ANKRD1_sg1+3			+			
EDEM1_sg2+3				+		
FUT1_sg2+3					+	
KLF6_sg1+2						+
HCoV-229E (24 h)	+	+	+	+	+	+

Figure 17: RNA-Seq analysis to validate the effect of knock outs in the virus-host response.
Normalized read counts generated by DEseq2 were $\log_2$ transformed and the ratios for the difference in the gene expression were calculated pair-wise. **A)** Heat-map shows the top 50 and bottom 50 deregulated genes with a $\log_2$ fold change < and / or > = 2, p-value = <0.001 across all the cell lines infected with HCoV-229E for 24 h. **B)** Venn diagram comparison shows the filtered genes ($\log_2$ fold change < and / or > = 2, p-value = <0.001) from all the cell lines infected with HCoV-229E and the table shows the values used for each condition as an input. **C)** Heat-map generated by Metascape shows the pathway enrichment analysis for all of the filtered multiple gene sets used for Venn diagram comparison. Sebastian Werner kindly assisted in generating the heat map, Venn diagram comparisons, and Metascape analysis.

In the next step, genes were filtered out from the infected conditions with a $\log_2$ fold change < and / or > = 2, p-value = <0.001, and were compared using a Venn diagram. More than 1700 genes were strongly deregulated across all the infected samples (Fig. 17B). Out of which only 17% (297) of the deregulated genes were found to be common among all the groups. Nevertheless, more than 50% of the deregulated genes from the infected knock out cells were commonly found in the infected parental control. Out of four knock out cell lines, only EDEM1_sg2+3 showed a strong response in the regulation of gene expression upon infection. Out of 1060 strongly deregulated genes, 30% (321) of the genes were unique to EDEM1_sg2+3 upon infection with HCoV-229E (Fig. 17B).

The 1741 filtered gene sets were then analyzed by over-representation analysis using Metascape. As shown in Fig. 17C, the heat map represented the pair-wise comparison of the top 20 enriched pathways across all the cell lines infected with HCoV-229E. The filtered genes were enriched for the pathways, Unfolded Protein Response (UPR) (R-HSA-381119), and response to endoplasmic reticulum stress (GO:0034976) in all the infected cell lines, which has been extensively studied in the context of coronavirus infections. In addition, there was moderate enrichment of genes in the pathway, regulation of secretion (GO:0051046) (according to the $-\log_{10}$ p-value) in the knock out cells as compared to the Huh7 parental and px459 control (Fig. 17C). Overall, the Metascape analysis clearly stated that none of these host factors shared and / or had many unique deregulated pathways during HCoV-229E infection.

In conclusion, despite many efforts, the above-validated host factors did not provide any concrete evidence of their functional role in virus replication. In the context of virus-host response, the four (ANKRD1, EDEM1, KLF6, and FUT1) host factors, did not show any significant changes in the regulation of the top 50 and bottom 50 genes when compared to the parental control. However, there were a unique number of genes that were found to be regulated in the absence of these four host factors during the HCoV-229E infection.

5.2 Identification of human genes supporting HCoV-229E, and MERS-CoV replication using genome-wide CRISPR-Cas9 knock out screening

Forward genetic approaches offer an unbiased selection of human genes associated with a phenotype of interest. Recent developments in the forward genetic screens allow the perturbation of multiple genes at once in the desired cell lines for a selective phenotype and then sequencing of the genes involved in the phenotypic change [78]. In the following sections, to identify the potential genes that might play an important role in the viral replication and / or virus-induced cell death, an unbiased positive screen selection by a genome-wide CRISPR-Cas9 knock out was systematically performed. The rationale behind the approach was that the coronaviruses (in this thesis HCoV-229E, and MERS) exploit many of the host cell factors for their viral replication and / or virus-induced cell death [114]. In the absence of these vital host factors, the host cells can either resist the ongoing viral replication and / or virus-induced cell damage upon the virus infection. By applying a positive screen selection, which means selecting the genetically perturbed cells that survive the HCoV-229E and / or MERS-CoV infection after the infection period of 40 h (MERS-CoV) and 72 h / 7 days (HCoV-229E), genes that resist the virus infection can be identified and further validated for their virus-host relationships.

5.2.1 Next-generation sequencing of the amplified sgRNA plasmid library to determine the sgRNA distribution

The human GeCKO pooled library consists of 123,411 sgRNAs that were cloned individually into a lentiviral vector (LCV2), which is designed to target approximately 19,000 human genes and 1,000 non-targeting controls (designed to randomly target at intronic region). The initial library that was purchased from the Addgene company has to be amplified to maximize the concentration for further downstream steps in the genome-wide screen. It was very crucial to control for the sgRNA distribution after the sgRNA amplification to avoid an unnecessary drop out of sgRNAs. To determine the sgRNA distribution in the amplified sgRNA plasmid library, an Illumina library PCR was performed and the PCR amplified sgRNA plasmid library was deep sequenced by NGS. In the initial step, the bacterial amplified sgRNA plasmid was prepared for NGS by a PCR reaction consisting of forward and reverse primers that can amplify the sgRNA target regions and adds the Illumina adaptors for library preparation.

The detailed conditions of the PCR setup with the information on Illumina primers were described in the materials and methods section. After the PCR reaction, the plasmid libraries were either non-purified or purified using a PCR purification kit as described in the methods section, and 2 µg of the PCR product was separated on a 2% agarose gel. The successful PCR reactions showed a clear band at ~ 260 - 270 bp on the agarose gel. To compare the quality of the PCR amplification before and after the PCR purification, both purified and non-purified libraries were loaded on the agarose gel and separated. The gel doc image (Fig. 18A), showed a very faint band around 300 bp for the non-purified plasmid libraries.

On the other hand, a clear and intense single band was visible around 300 bp for the PCR-purified plasmid libraries. This data indicated that the library PCR amplification was very successful and the amplified libraries were sent to NGS for deep sequencing. For further setup of the genetic screen, the purified plasmid library A+B that represented the full sgRNA plasmid library was selected for NGS to determine the sgRNA distribution.

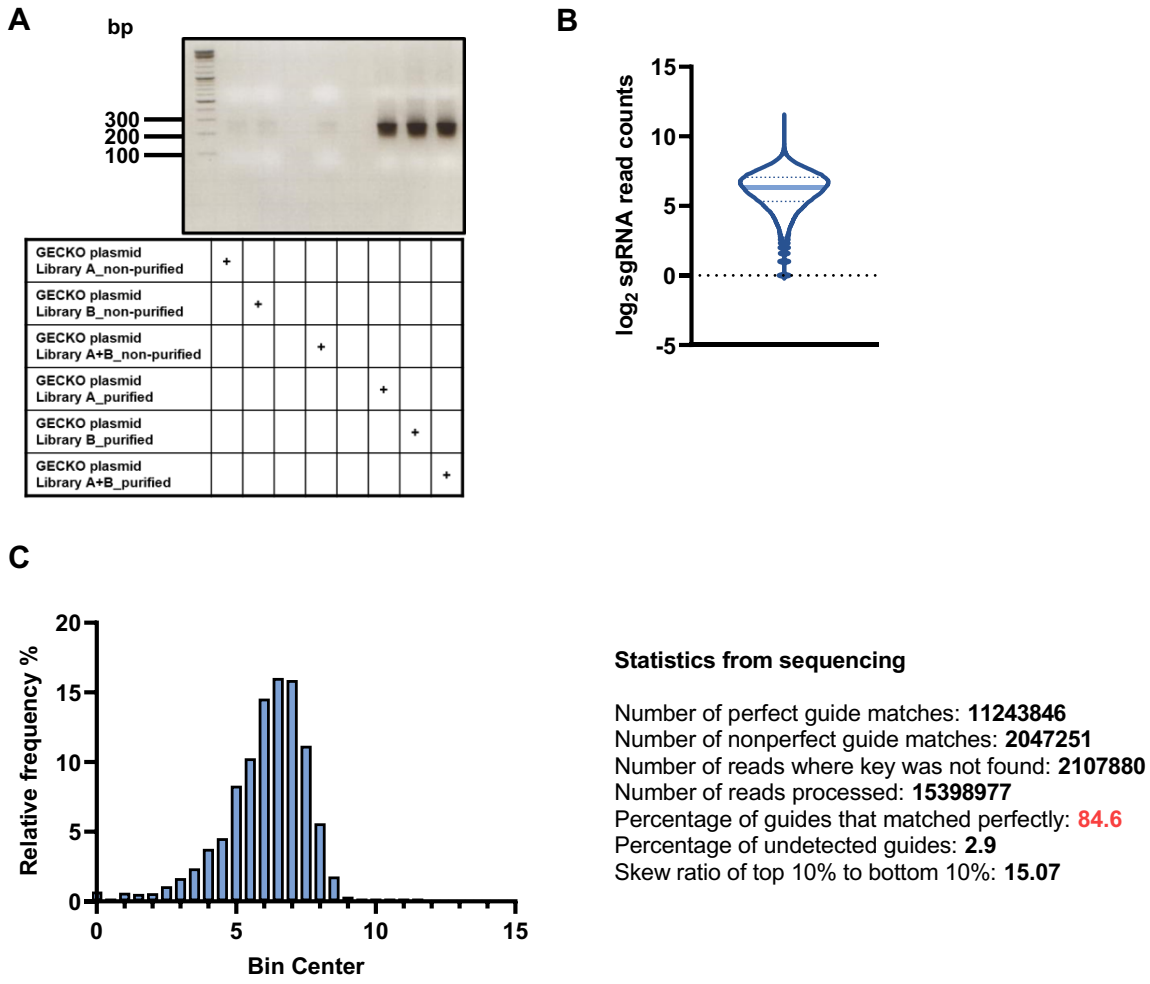


Figure 18: Analysis of the amplified sgRNA plasmid library by PCR amplification and deep sequencing by NGS.
A) Gel doc image shows the loading of non-purified and purified single (A, B) and combined (A+B) sgRNA plasmid libraries on a 2% agarose gel.10 kb gene ruler DNA ladder mix was used as a reference for the band size. **B)** Violin plot showing the $\log_2$ transformed read counts for the sgRNA sequences for the plasmid library A+B after NGS and analysis by the python script. The blue solid line indicates the median. **C)** Histogram shows the relative frequency of the sgRNA reads (%) plotted against the intervals/bins. The descriptive statistics depict the output of sequencing analysis done with the python script. The python scripts were written and run on a Linux environment together with Sebastian Werner.

The PCR-purified plasmid library A+B was sequenced on an Illumina Miseq (Dr. Andrea Nist, Core Facility for Genomics, University of Marburg) according to the user manual with an intended coverage of >100 reads per sgRNA. After the NGS, the sequenced data was run on a python script as described in the methods section to analyze the sgRNA representation. The python script matched the raw reads from the NGS data to the input sgRNA library sequences and provided the read counts for each of the sgRNA sequences in the library. The script also provided descriptive statistics on the sgRNA guide matches, the percentages of the sgRNAs that matched the input sgRNA library sequences, and a skew ratio (Fig. 18C). From the descriptive statistics, more than 84% of the raw reads matched the input library sequence, which indicated a successful sgRNA plasmid library amplification and a thorough deep sequencing (Fig. 18C). The read counts for each of the sgRNA sequences were $\log_2$ transformed to visualize the sgRNA distribution on a violin plot as shown in the Fig. 18B. The violin plot indicated a clear and normal distribution of sgRNA reads throughout the full library and the median (dark blue, Fig. 18B) showed that the vast majority of the sgRNA sequences in the plasmid library have the intended coverage of >100 reads (Median at 6.6 $\log_2$ scale). Additionally, a frequency distribution curve was performed on the read counts of the sgRNA sequences to determine the read count distribution. The histogram (Fig. 18C) represented as bars, indicated that the frequency distribution of the read counts for the plasmid library had a normal distribution curve.

5.2.2 Lentivirus production and titration for the generation of genome-wide knock out Huh7 cells

The lentivirus was successfully produced using the sgRNA plasmid library in HEK293FT cells as described in the methods section. Before the generation of genome-wide knock out Huh7 cell pool by lentiviral transduction, it was very crucial to determine the titer of the produced lentivirus to ensure that most cells have a single genetic perturbation. The sgRNA plasmid encodes a puromycin selection antibiotic cassette that allowed the selection of the transduced cells from the non-transduced cells. Huh7 cells were seeded onto a 6-well plate at a seeding density of 3*10^6 cells in 3 ml DMEM medium and were infected with various volumes of lentivirus supernatants (25 µl to 1 ml), including an uninfected control. After successful transduction for 24 h, the cells from each virus volume condition were trypsinized and re-seeded onto a 96-well plate at a seeding density of 4*10^3 cells in 100 µl DMEM medium in technical quadruplicates. To the re-seeded cells from each virus condition, two of the wells were added with an additional 100 µl DMEM medium with puromycin (final concentration 1 µg/ml) and the other two wells with normal 100 µl DMEM medium and incubated for 72 h. The cell viability was measured using CellTitre Glo according to the manufacturer's protocol.

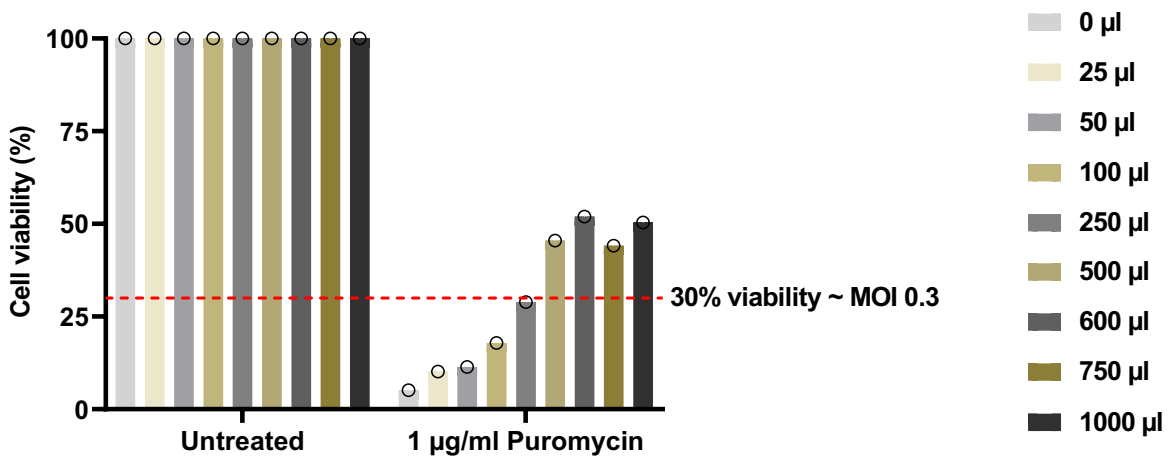


Figure 19: Determination of lentivirus titer through cell viability measurement using CellTitre Glo.
For each of the transduced virus volume conditions, the cells were re-seeded in a 96-well plate and were treated with puromycin (1 µg/ml) or left untreated for 72 h. After 72 h, CellTiter Glo was added to each of the wells and the luminescence was measured on a device called Infinite 200 (Prof. Dr. Tikkanen, Institute of Biochemistry, Giessen). The luminescence values were background subtracted and the cell viability was calculated. The bar graph with a scatter dot represents the mean value generated from technical duplicates of one biological independent experiment.

The cell viability was calculated as the average luminescence of the two wells with puromycin treatment divided by the average luminescence of the two wells without the puromycin. As indicated in the bar graph, (Fig. 19) the untreated virus volume conditions showed 100% cell viability. The linear increase in the cell viability with the increasing volumes of the virus supernatant was observed in the puromycin treatment conditions. In this context, a 30% cell viability was observed for the 250 µl virus supernatant condition under puromycin and the cell viability gets saturated at 50% with the higher volumes of virus supernatant. From this data, the multiplicity of infection (MOI) was calculated to be 0.3 when 250 µl lentivirus supernatant was added to the cells (red dotted lines indicating the cell viability / MOI at 30%). The rationale behind selecting an MOI 0.3 is to ensure that most cells have a single genetic perturbation (i.e. single integration of the sgRNA expression cassette) during the generation of genome-wide knock out cells.

5.2.3 Screening setup for identification of host factors required for HCoV-229E and MERS-CoV replication and virus-induced cell death using genome-wide CRISPR-Cas9 knock out screen

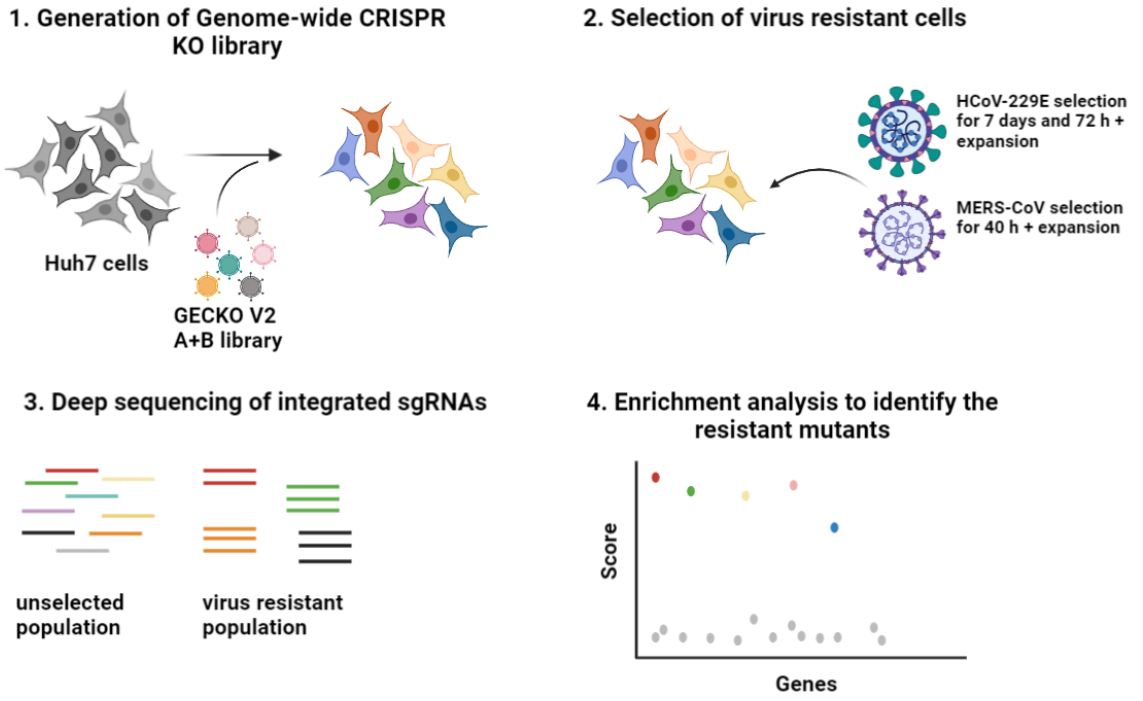


Figure 20: Schematic overview of the genetic screen to identify virus-resistant genes.
The representative figure above shows the experimental setup for the screen selection with the library transduced Huh7 cells challenged with HCoV-229E and MERS. For details, see the text. This image was created using BioRender https://www.biorender.com/.

After the successful determination of lentivirus titer from the above-mentioned experiment, a pool of Huh7 cells harboring a single perturbed gene per cell on average was generated. To maintain the full library sgRNA diversity and adequate coverage of >500 cells expressing each individual sgRNA, 200 million Huh7 cells were transduced to represent the sgRNA library of 123,411 unique sgRNAs at an of MOI 0.3. To achieve this, 3*10^6 Huh7 cells were seeded onto a 6-well plate in a 2 ml DMEM medium. In total, ten 6-well plates were used to transduce 200 million Huh7 cells. As established in the experiment to determine the lentivirus titer, to each well of the already seeded 6-well plate, 250 µl of lentivirus supernatant (MOI 0.3) was added and transduced for 24 h. After 24 h of transduction, the cells were pooled all together and re-seeded into 40 145 mm culture dishes at a density of 5*10^6 cells with 20 ml DMEM medium containing 1 µg/ml puromycin. The pool of cells was selected for 10-14 days (medium change with fresh puromycin every 2 days) until all the untransduced cells were removed from the cell culture dishes. After the puromycin selection, the library transduced Huh7 cells were

maintained at the cell density of ~80 million cells to represent the coverage of 650 cells / sgRNA and were ready for the screen selection.

In the case of HCoV-229E selection, the library transduced Huh7 cells were challenged with HCoV-229E at an MOI 1 for 7 days and 72 h at 33°C and the surviving cells were further expanded to 14 days up to 30 days to select and deep sequence the enriched sgRNA integrated cells. On the other hand, the cells were challenged with MERS-CoV at an MOI of 0.01 for 40 h at 37°C, and the surviving cells were further expanded for 14 days.

5.2.4 Screen selection for 7 days of HCoV-229E infection

Library transduced Huh7 cells were seeded onto 16 145 mm cell culture dishes at a density of 5*10^6 cells in a 20 ml DMEM medium. In total, 80 million library transduced cells were challenged with HCoV-229E at an MOI 1 and were incubated at 33°C for 7 days. As an unselected control, genomic DNA from 80 million library transduced cells was isolated at this time point which served as a full library representation for NGS. In addition, 40 million library transduced cells were challenged with heat-inactivated HCoV-229E (50°C for 20 min) which served as a negative control and were incubated at 33°C for 48 h. As a positive control for virus-induced cytopathic effects, Huh7 parental cells were seeded onto a 145 mm dish at a density of 5*10^6 cells and were infected with HCoV-229E at an MOI 1 for 7 days only to compare the cytopathic effects between the library transduced cells and Huh7 parental cells.

The cytopathic effect was visualized under a light microscope every 24 h p.i. until 144 h p.i. as indicated in the slides from Fig. 21. After the virus challenge of 48 h, cytopathic effects were visible in Huh7 parental cells and library transduced cells (Fig. 21). The library transduced cells infected with the heat-inactivated virus showed no sign of cytopathic effect and hence, the surviving cells were harvested for the genomic DNA at this point. At 144 h p.i., nearly all of the Huh7 parental cells were dead and were floating in the culture medium due to the cytopathic effect of virus infection. In contrast, many colonies of healthy surviving cells could be visualized in library transduced cells after 144 h p.i. (Fig. 21). After 7 days of virus challenge, the surviving library transduced cells were pooled together in a fresh DMEM medium and were expanded for up to four weeks until the cell density reached confluency and 80 million cells in total. After the expansion of surviving cells to this confluency, genomic DNA was isolated from the cells for deep sequencing by NGS. The screen selection with HCoV-229E for 7 days was performed in two independent biological library transduced Huh7 cells.

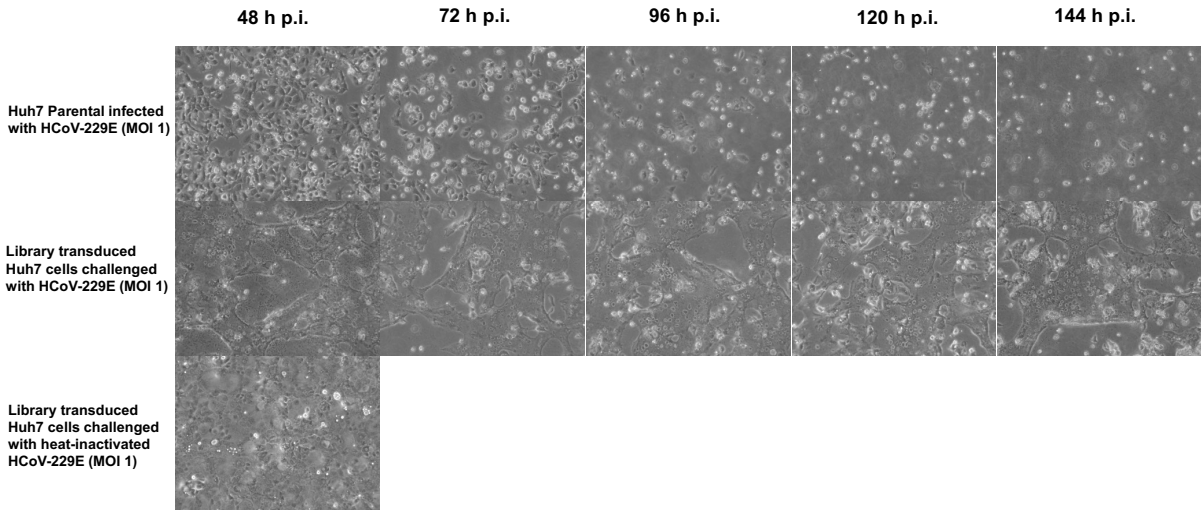


Figure 21: Microscopic images from the 7 day screen selection of library transduced cells challenged with HCoV-229E.
In the first horizontal panel, Huh7 parental cells were infected with HCoV-229E at an MOI 1 for 7 days and were visualized for cytopathic effect from 48 h p.i. until 144 h p.i. The second horizontal panel indicates the library transduced cells challenged with HCoV-229E at an MOI 1 and were visualized for surviving cells from 48 h p.i. until 144 h p.i. The third horizontal panel depicts the library transduced cells challenged with heat-inactivated HCoV-229E (50°C for 20 min) for 48 h p.i. as a negative control. Images shown here were representative microscopic images from two biologically independent experiments.

Genomic DNA isolated from the unselected and selected library transduced cells challenged with HCoV-229E and / or heat-inactivated HCoV-229E was PCR amplified with the Illumina forward and reverse primers to amplify the sgRNA sequences as mentioned in section 5.2.1 for deep sequencing by NGS. To differentiate the samples during the NGS multiplexing, each sample was barcoded by a unique 8 bp barcode in the reverse primer during the library PCR amplification. The samples were deep sequenced on an Illumina Nextseq platform with a coverage of 500 reads / sgRNA and a 20% PhiX control for the library diversity. In the next step, the demultiplexed sequencing data were analyzed for the enrichment of sgRNAs in the selected samples and the corresponding candidate genes from the genome-wide CRISPR-Cas9 screen. For this analysis, a recently developed computational tool called MAGeCK has been used. The detailed protocol on how the MAGeCK tool was used to identify the enriched sgRNAs and the essential genes was described in the methods section. To understand if the genome-wide CRISPR-Cas9 screen at all worked, control metrics were done on the read counts for all the samples before and after the normalization, to verify if the read count distribution behaved normally in the unselected controls and to see the dropout / enrichment of sgRNAs in the samples selected for HCoV-229E infection. The table below summarizes the demultiplexed fastq files that were analyzed by the MAGeCK tool using the software R [99] to generate an overview table consisting of read counts that were sequenced and mapped to the input sgRNA library sequences.

As shown in Table 2, for all the samples the read counts were around 20 – 50 million reads and around 60 to 80% of the sequenced read counts were matched with the input library sgRNA sequences.

Moreover, the unselected control samples and HCoV-229E heat-inactivated samples showed only around 3000 – 4700 sgRNAs that had zero read counts. The Gini index (values from 0 to 1), is defined as a measure of evenness of read coverage across different regions of interest. It is calculated by the ratio of the cumulative proportion of reads to the proportion of target regions and a lower Gini index (0) indicates a perfect distribution of read coverage [115]. The gini index for the unselected controls was around 0.14 to 0.15 for both of the biological replicates, which indicated that the samples had equally distributed individual sgRNA read counts.

Label	Reads	Mapped	Percentage	Total sgRNAs	Zero counts	Gini Index
Control_BIO1	20194891	13642899	0.6756	119461	3802	0.1576
Control_BIO2	33996119	23275667	0.6847	119461	3414	0.1411
HCoV-229E 7 days_BIO1	27764691	19222097	0.6923	119461	52687	0.5744
HCoV-229E 7 days_BIO2	35201549	28166049	0.8001	119461	76256	0.7189
HCoV-229E heat-inactivated 48 h_BIO1	43643295	29843134	0.6838	119461	3601	0.1414
HCoV-229E heat-inactivated 48 h_BIO2	34476109	23556590	0.6833	119461	4708	0.1514

Table 2: Overview table showing the statistics of DNA sequencing results of the HCoV-229E 7 days screen as analyzed by MAGeCK software.
The table shows the overall sequencing depth for all the samples subjected to NGS sequenced. In addition, the percentage of mapped read counts to the input sgRNA library sequences, the total number of sgRNAs used for the analysis, the number of sgRNAs that were undetected, and their gini indices are indicated. Sebastian Werner kindly generated the output from the NGS sequenced samples listed in the table using the MAGeCK tool.

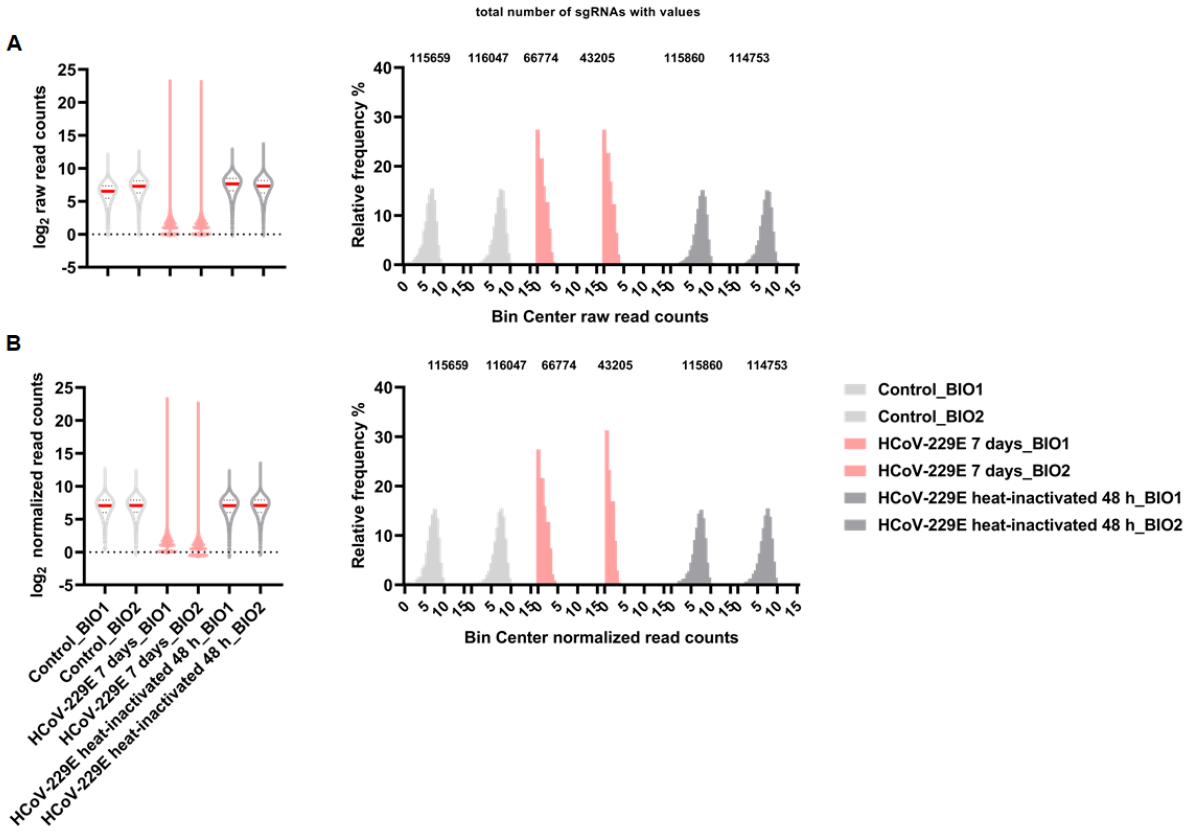


Figure 22: Quality control metrics for the NGS sequenced samples in the 7 days HCoV-229E selection.
MAGeCK tool was used to determine the read counts in the selected and unselected samples. The read counts from all the samples were median normalized and quality control metrics were done to determine the read count distribution in selected and unselected samples. **A)** Violin plot (left) shows the median distribution in all the samples before normalization. The histogram (right) depicts the frequency distribution of raw read counts in the samples before normalization with the values counted for each sample. **B)** Violin plot (left) shows the median distribution in all the samples after median normalization. The histogram (right) depicts the frequency distribution of raw read counts in the samples after normalization with the values counted for each sample. The python scripts for the MAGeCK analysis were written and run on a Linux environment, together with Sebastian Werner, which provided the output as an excel file for quality control metrics.

For all the samples, the raw read counts before and after median normalization were $\log_2$ transformed and plotted in violin plot / histograms to visualize the read count distribution. As shown in the violin plot Fig. 22A and B, the unselected control samples from both biological replicates had an equal read count distribution. Moreover, from the frequency distribution histograms (Fig. 22A and B), it was found that there were only around 3,400 to 3,800 sgRNAs (out of 115659, and 116047 sgRNAs with values), which had zero read counts in both of the biological replicates respectively. This indicated that the unselected control samples showed an unbiased representation of the full sgRNA library. The samples treated with heat-inactivated HCoV-229E for 48 h also showed an equal read count distribution as compared to the controls. This strongly indicated that the heat inactivation did not influence the selection pressure and served as a good negative control for the later analysis. Interestingly, the samples infected with HCoV-229E for 7 days showed a strong shift down in the median distribution of read counts (Violin plots) and a strong shift of the bins / intervals to the left (Histogram bar) which indicated an enormous depletion of integrated sgRNAs. About 52,000 to 76,000 sgRNAs were completely depleted in the HCoV-229E infected samples (both of the biological replicates respectively) due to a strong selection pressure (Fig. 22B, histogram).

In the next step, the MAGeCK tool was used to determine the differences in the enrichment of sgRNAs in the HCoV-229E selected sample compared to the unselected control. Each sgRNA was ranked based on a positive score, p-value, and $\log_2$ fold change value for the enrichment. The tool assigned a positive or negative rank for a gene based on the sgRNA enrichment ranking. If a gene affected the selection pressure, then the sgRNAs targeting this gene would be ranked in the top order across the list of all other sgRNAs. The output of the gene ranking list contained the rank list of 21,694 genes, the number of corresponding sgRNAs which showed an enrichment / depletion compared to the control, and the MAGeCK score (for statistics), all of which were used to filter and identify the candidate genes for further validation experiments.

From the MAGeCK gene ranking list, 104 genes were filtered out with a MAGeCK score <= 0.005 out of which the gene *ANPEP*, the known entry receptor for HCoV-229E, had the highest ranking with the highest MAGeCK score as indicated in the multi-variable plot (Fig. 23). This data indicated that the

positive screen selection had successfully worked and all the six sgRNAs targeting the ANPEP gene were highly enriched in the cells (bubble color, Fig. 23) selected for HCoV-229E infection in accordance with ANPEP as a well-established entry receptor for HCoV-229E. Despite the higher ranking, the other filtered genes showed only a single sgRNA enrichment out of six, which was shown in the multi-variable plot with bubble color (Fig. 23). Even though the genes were statistically significant with a lower MAGeCK score, their sgRNA enrichment was very low (Fig. 23). This unanticipated result made it difficult to interpret if these genes had a major role during the selection or were randomly enriched, which caused a bias in the selection. Overall, the outlier candidate gene *ANPEP* provided a piece of concrete evidence that the positive screen selection had successfully worked but could not provide more meaningful information on the other genes.

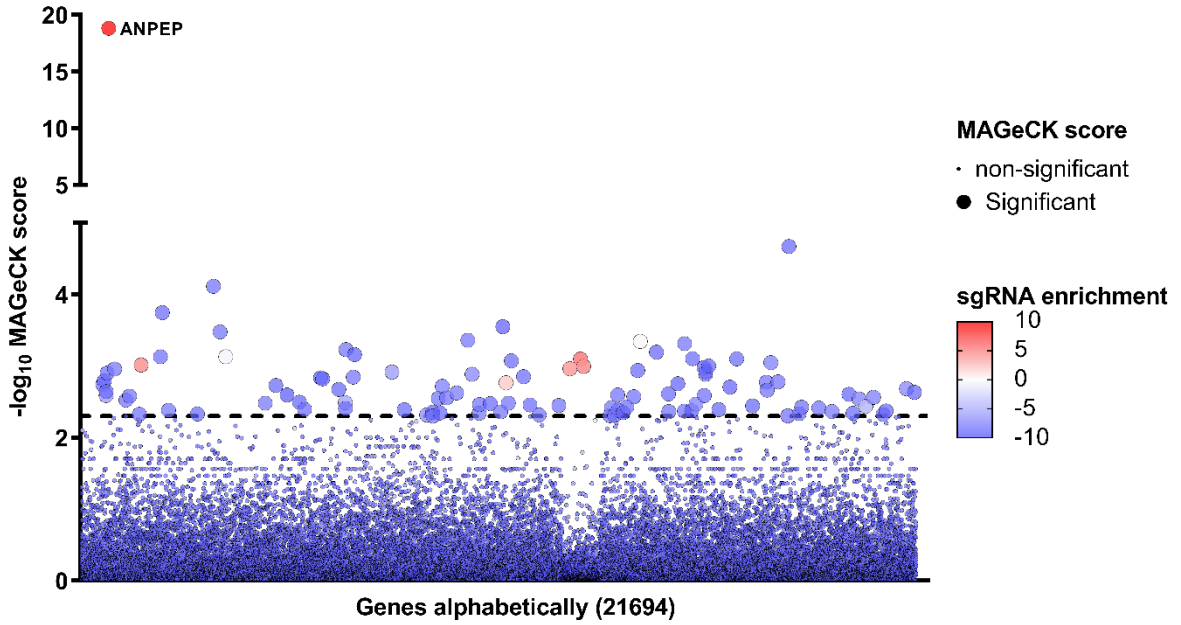


Figure 23: Ranked gene list (HCoV-229E – 7 day selection) from the MAGeCK analysis.
The gene rank list provided by the MAGeCK tool was used to plot the genes (21694) against their MAGeCK score ($-\log_{10}$). Genes were sorted alphabetically. The dotted line represents the threshold at $-\log_{10}=2.3$ (MAGeCK score 0.005). 104 genes were filtered out with a MAGeCK score <=0.005. The bubble size represents the MAGeCK score. The bubble color shows the sgRNA enrichment in $\log_2$ fold change. The heat map in the legend shows the scale from -10 to 10 $\log_2$ fold change. Important candidate genes were represented in red dots.

5.2.5 Screen selection for 72 h of HCoV-229E infection

To overcome the challenge posed by the earlier screen selection for 7 days of HCoV-229E selection, a new strategy was designed. It was sought to reduce the selection pressure by challenging the library transduced Huh7 cells with HCoV-229E for 72 h and then shortly selecting the surviving cells for deep sequencing to identify the enriched sgRNAs. In this attempt, 5*10^6 library transduced Huh7 cells were seeded into 40 145 mm cell culture dishes in 20 ml DMEM medium, in total 200 million cells were challenged with HCoV-229E at an MOI 1 for 72 h at 33℃. As an unselected control, genomic DNA from 80 million library transduced cells was isolated to serve as a full library representation for NGS. As a positive control for virus-induced cytopathic effects, Huh7 parental cells were seeded onto a 145 mm dish at a density of 5*10^6 cells and were infected with HCoV-229E at an MOI 1 for a period of 72 h only to compare the cytopathic effects between the library transduced cells and Huh7 parental cells. The screen selection was performed in two independent biological library transduced Huh7 cells.

The cytopathic effect was visualized under a light microscope every 24 h p.i. until 72 h p.i. as indicated in the slides from Fig. 24. After the virus challenge of 48 h, cytopathic effects were visible in Huh7 parental cells and library transduced cells (Fig. 24). At 72 h p.i., the Huh7 parental cells showed a severe cytopathic effect (Fig. 24); however, the library transduced cells showed a less cytopathic effect, which indicated the surviving cells upon HCoV-229E infection.

After the challenge for 72 h, the surviving cells were pooled together from all the cell culture dishes and were re-seeded into 10 145 mm dishes for 14 days to reach a 90% confluency. Once the surviving cells reached confluence, the genomic DNA was isolated from the cells, the integrated sgRNAs were library PCR amplified, and the amplicons were sent to deep sequencing by NGS.

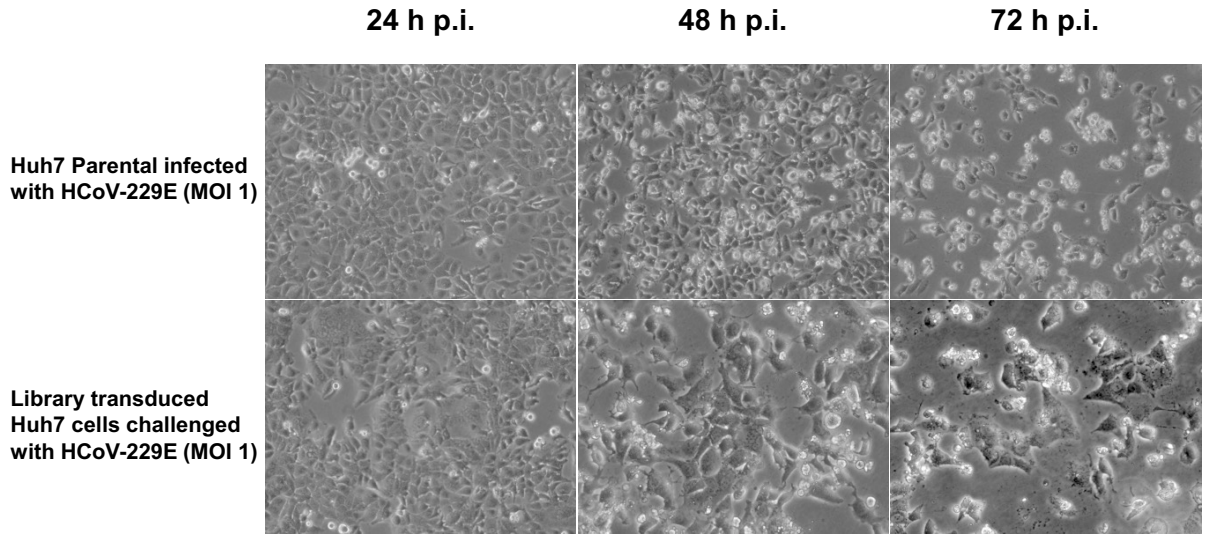


Figure 24: Microscopic images from the 72 h screen selection of library transduced cells challenged with HCoV-229E.
In the first horizontal panel, Huh7 parental cells were infected with HCoV-229E at an MOI 1 for 72 h and were visualized for cytopathic effect from 48 h p.i. until 72 h p.i. The second horizontal panel indicates the library transduced cells challenged with HCoV-229E at an MOI 1 and were visualized for surviving cells from 48 h p.i. until 72 h p.i. Images shown here were representative microscopic images from two biological independent experiments.

To have more sequencing depth and the sgRNA read coverage compared to the previous screen, the amplicons were sequenced on an Illumina Novaseq 550 (Novagene) platform with a minimum sequencing depth of 80 million reads / sample. The deep sequenced and demultiplexed samples were analyzed by the MAGeCK tool to have the read counts for the individual sgRNAs in the full library, the sgRNA enrichment in the surviving cells compared to the unselected control, and to have the corresponding candidate gene lists. Table 3, summarizes the output generated from the demultiplexed fastq NGS files using the MAGeCK tool. The samples had a read coverage of 60 – 80 million reads and had a mapping percentage of 70% to the input library sequence. The unselected control samples had only 2,800 to 3,000 zero counts and a gini index of 0.12 – 0.13.

Label	Reads	Mapped	Percentage	Total sgRNAs	Zero counts	Gini Index
Control_BIO1	75913211	52358210	0.6897	119461	2833	0.124
Control_BIO2	62938308	43397879	0.6895	119461	3064	0.1295
HCoV-229E 72 h_BIO1	83585411	61504371	0.7358	119461	11433	0.3136
HCoV-229E 72 h_BIO2	79129671	57593344	0.7278	119461	15695	0.383

Table 3: Overview table showing the statistics of DNA sequencing results of the HCoV-229E 72 h screen as analyzed by MAGeCK software.
The table shows the overall sequencing depth for all the samples subjected to NGS sequenced. In addition, the percentage of mapped read counts to the input sgRNA library sequences, the total number of sgRNAs used for the analysis, the number of sgRNAs that were undetected, and their gini indices are indicated.

From the MAGeCK analysis, the raw read counts before and after median normalization were $\log_2$ transformed and plotted in violin plot / histograms to visualize the read count distribution. As shown in the violin plot Fig. 25A and B, the unselected control samples from both biological replicates had an equal read count distribution. Moreover, from the frequency distribution histograms (Fig. 25A and B), it was found that there were only around 2,800 to 3,000 sgRNAs (out of 116661 and 1166429 sgRNAs with values), which had zero read counts in both of the biological replicates respectively. This result indicated that the unselected control samples showed an unbiased representation of the full sgRNA library. In contrast, the samples infected with HCoV-229E for 72 h, showed a strong shift down in the median distribution of read counts (Violin plots) and a strong shift of the bin / intervals to the left, (Histogram bar) which indicated a strong depletion of integrated sgRNAs. About 11,000 to 15,000 sgRNAs were completely depleted in the HCoV-229E infected samples (both of the biological replicates respectively) due to the selection pressure (Fig. 25B, histogram).

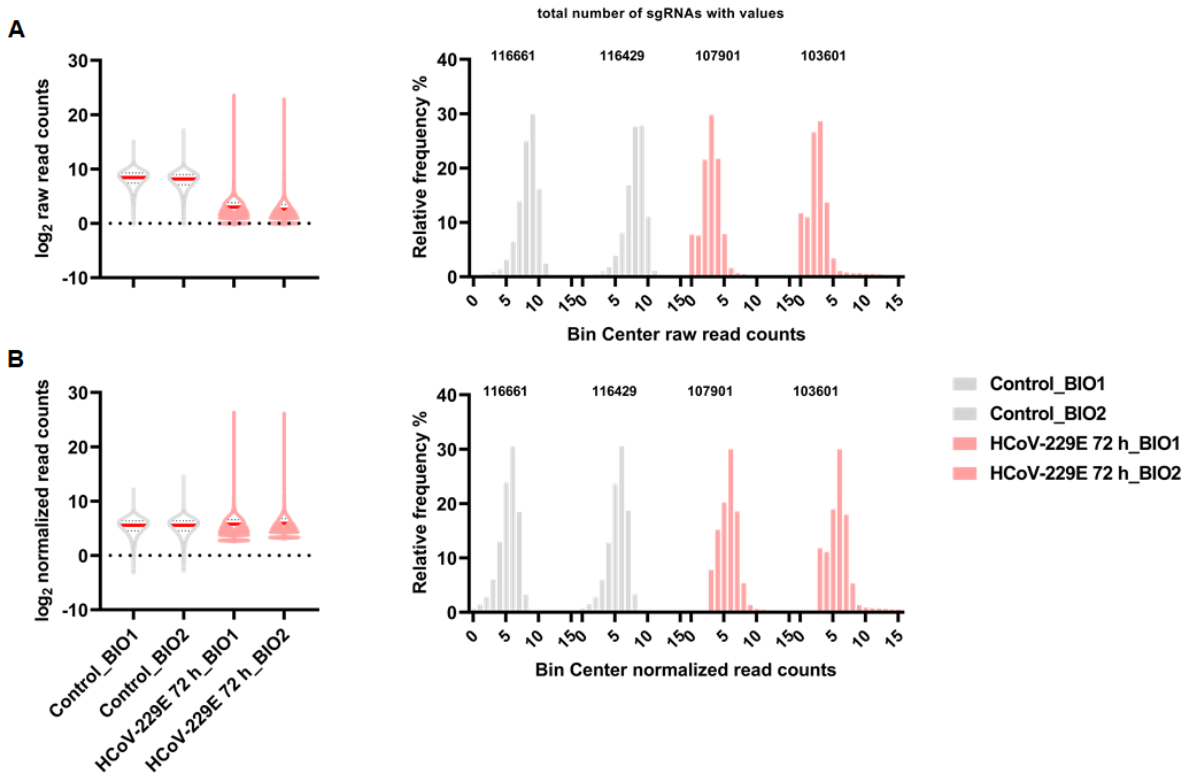


Figure 25: Quality control metrics for the NGS sequenced samples in the 72 h HCoV-229E selection.
MAGeCK tool was used to determine the read counts in the selected and unselected samples. The read counts from all the samples were median normalized and quality control metrics were done to determine the read count distribution in selected and unselected samples. **A)** Violin plot (left) shows the median distribution in all the samples before normalization. The histogram (right) depicts the frequency distribution of raw read counts in the samples before normalization with the values counted for each sample. **B)** Violin plot (left) shows the median distribution in all the samples after median normalization. The histogram (right) depicts the frequency distribution of raw read counts in the samples after normalization with the values counted for each sample.

In the next steps, the MAGeCK tool was used to determine the differences in the enrichment of sgRNAs in the HCoV-229E selected sample compared to the unselected control. Each sgRNA was ranked based on a positive score, p-value, and $\log_2$ fold change value for the enrichment. The tool assigned a positive or negative rank for a gene based on the sgRNA enrichment ranking. If a gene affected the selection pressure, then the sgRNAs targeting this gene would be ranked in the top order across the list of all other sgRNAs. The output of the gene ranking list contained the rank list of 21,705 genes, the number of corresponding sgRNAs which showed an enrichment / depletion compared to the control, MAGeCK score (for statistics) which were used to filter and identify the candidate genes for further validation experiments.

From the MAGeCK gene ranking list, 492 genes were filtered out with a MAGeCK score <= 0.005 out of which *ANPEP* had the highest ranking with the highest MAGeCK score as indicated in the multi-

variable plot (Fig. 26). This data indicated that the positive screen selection has worked and all the six sgRNAs targeting the ANPEP gene were highly enriched in the cells selected for HCoV-229E infection. In addition, the less stringent selection pressure provided additional candidate gene hits as compared to the 7 day selection pressure. For further validation studies, 11 interesting genes were picked which included *ANPEP*, *ATG7*, *CPB2*, *CTCFL*, *HDAC4*, *KIF3B*, *ORMDL1*, *SFTA2*, *SYT3*, *SYT16*, and *ZDHHC3*. These genes were among the top-ranking order and two or more sgRNAs targeting these genes had a $\log_2$ FC >= 2.

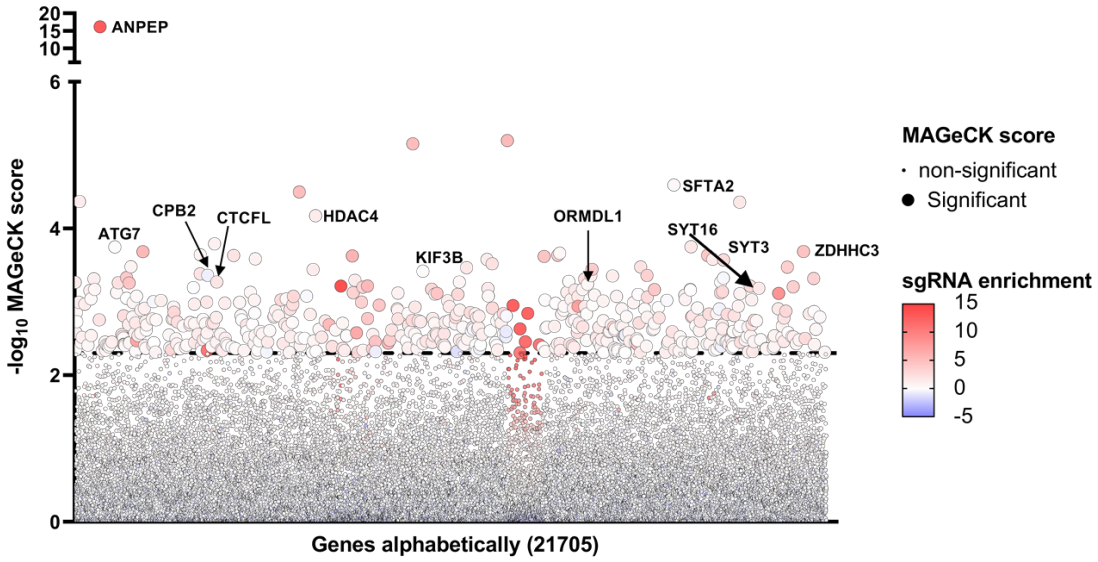


Figure 26: Ranked gene list from the 72 h HCoV-229E selection using the MAGeCK analysis.
The gene rank list provided by the MAGeCK tool was used to plot the genes (21705) against their MAGeCK score ($-\log_{10}$). Genes were sorted alphabetically. The dotted line represents the threshold at $-\log_{10}$=2.3 (MAGeCK score 0.005). 492 genes were plotted with a MAGeCK score <=0.005. The bubble size represents the MAGeCK score and the bubble color indicates the sgRNA enrichment in $\log_2$ fold change. The legend shows the heat map of sgRNA enrichment from $\log_2$ -5 to 15. Important candidate genes were marked with their gene names.

5.2.6 Screen selection for MERS-CoV infection

Since the positive screen selection worked well with HCoV-229E, it was interesting to apply the same genome-wide screen selection to obtain more insights into the genes and pathways involved in the more severe coronavirus (MERS-CoV) infection. Due to the necessity of a Bio Safety Level 3 (BSL-3) lab to perform MERS-CoV infections, the genome-wide screen on MERS-CoV infection was conducted by our co-worker and experienced virologist Dr. Christin Müller at the BSL-3 lab, Institute of Medical Virology, Gießen.

Library transduced cells (200 million in total) were challenged with MERS-CoV at an MOI of 0.01 for 40 h at 37°C. As a positive control for the severe cytopathic effect caused by the MERS-CoV, Huh7 parental cells were infected and incubated for 40 h. The microscopic images from Fig. 27, showed that the Huh7 parental cells infected with MERS-CoV at 40 h suffered from a severe cytopathic effect and most of the cells formed syncytia (red arrows, Fig. 27) as an indication of infection with MERS-CoV. In contrast, the library transduced cells at 40 h p.i., did not show any syncytia formation and indicated that the cells were resistant to the infection (Fig. 27).

After 40 h p.i., the surviving library transduced cells were further expanded until they reached confluency, and the genomic DNA was isolated from the cells to amplify the enriched integrated sgRNAs for deep sequencing by NGS as mentioned in the previous sections.

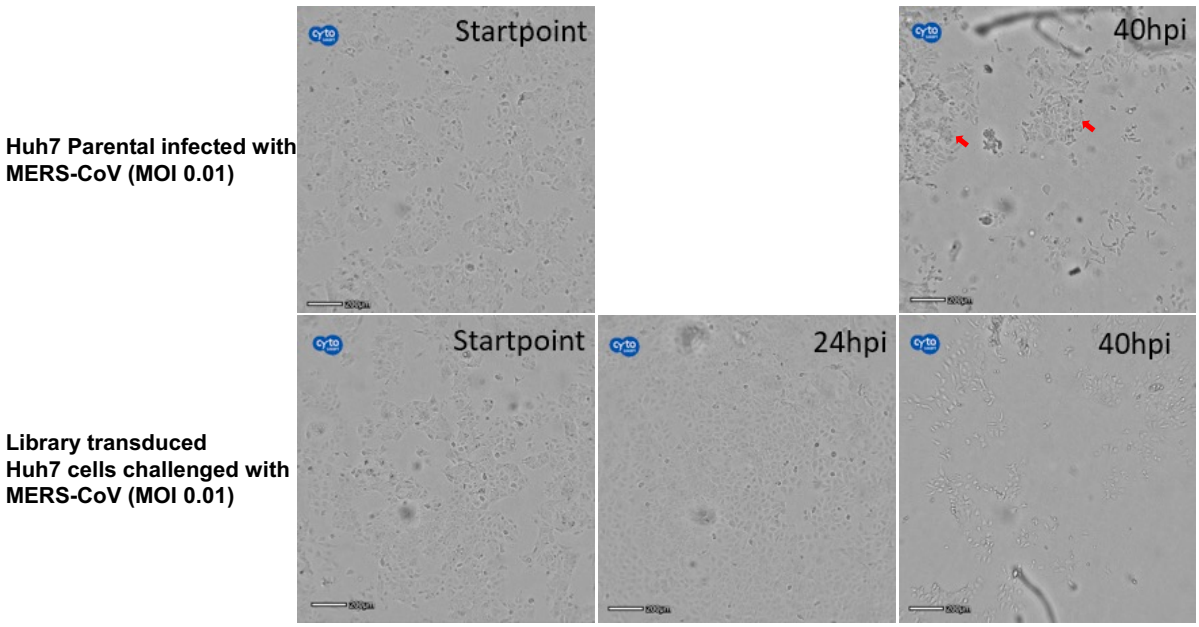


Figure 27: Microscopic images from the 40 h selection of library transduced cells challenged with MERS-CoV.
In the first horizontal panel, Huh7 parental cells were infected with MERS-CoV at an MOI 0.01 for 40 h and were examined for cytopathic effects from the start point of seeding until 40 h p.i. Red arrows show syncytia formation due to severe MERS-CoV infection. The second horizontal panel indicates the library transduced cells challenged with MERS-CoV at an MOI 0.01, which were inspected for surviving cells at 24 h p.i. or 40 h p.i., respectively Images shown here were representative microscopic images from two biologically independent experiments. Dr. Christin Müller (Institute of Medical Virology, Giessen) kindly performed the screen selection and generated the microscopic images.

Table 4, summarizes the output generated from the demultiplexed fastq NGS files using the MAGeCK tool. The samples had a read coverage of 60 – 80 million reads and had a mapping percentage of around 70% to the input library sequence. The unselected control samples had only 4,000 to 4,600 zero counts and a gini index of 0.13 – 0.14.

Label	Reads	Mapped	Percentage	Total sgRNAs	Zero counts	Gini Index
Control_BIO1	66125918	45522576	0.6884	119461	4001	0.1392
Control_BIO2	69687777	48109145	0.6904	119461	4603	0.1466
MERS-CoV 40 h_BIO1	63242988	43388827	0.6861	119461	21778	0.5222
MERS-CoV 40 h_BIO2	83503556	61888095	0.7411	119461	21313	0.4231

Table 4: Overview table showing the statistics of DNA sequencing results of the MERS-CoV 40 h screen as analyzed by MAGeCK software.
The table shows the overall sequencing depth for all the samples subjected to NGS sequenced. In addition, the percentage of mapped read counts to the input sgRNA library sequences, the total number of sgRNAs used for the analysis, the number of sgRNAs that were undetected, and their gini indices are indicated.

From the MAGeCK analysis pipeline, the raw read counts before and after median normalization were $\log_2$ transformed and plotted in violin plot / histograms to visualize the read count distribution. As shown in the violin plots (Fig. 28A and B), the unselected control samples from both biological replicates had an equal read count distribution. Moreover, the frequency distribution histogram bars (Fig. 28A and B) showed that there were only around 4,000 to 4,600 sgRNAs (out of 115460, and 114858 sgRNAs with values), which had zero read counts in both of the biological replicates respectively. This result indicated that the unselected control samples showed an unbiased representation of the full sgRNA library. In contrast, the samples infected with MERS-CoV for 40 h showed a strong shift down in the median distribution of read counts (see violin plots) and a strong shift of the bin / intervals to the left (see histogram bar) which indicated a strong depletion of integrated sgRNAs. About 21,000 sgRNAs were completely depleted in the MERS-CoV infected samples (both biological replicates) due to the selection pressure (Fig. 28B, histogram).

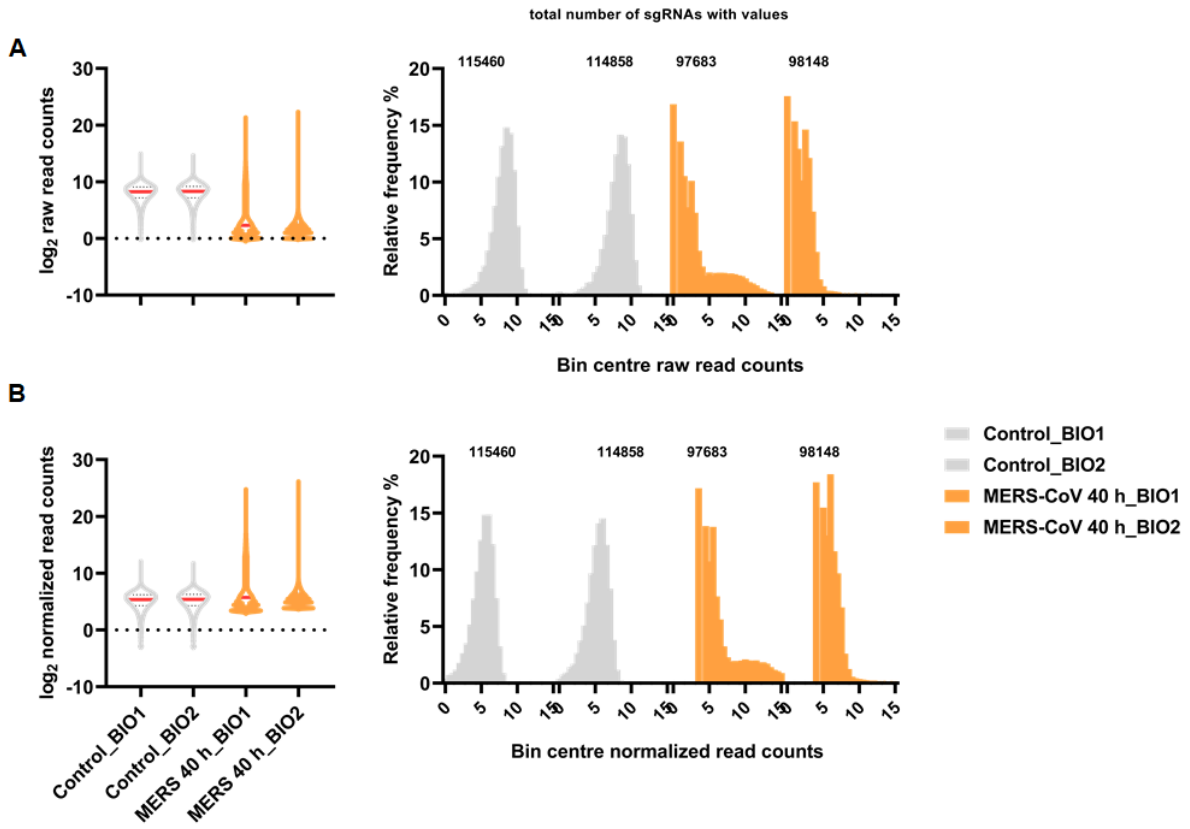


Figure 28: Quality control metrics for the NGS sequenced samples in the 40 h MERS-CoV selection.
The MAGeCK tool was used to determine the mappable read counts in the selected and unselected samples. The read counts from all the samples were median normalized and quality control metrics were applied to determine the read count distribution in selected and unselected samples. **A)** Violin plot (left) showing the median distribution in all the samples before normalization. The histogram (right) depicts the frequency distribution of raw read counts in the samples before normalization with the values counted for each sample. **B)** Violin plot (left) showing the median distribution in all the samples after median normalization. The histogram (right) depicts the frequency distribution of raw read counts in the samples after normalization with the values counted for each sample.

From the final MAGeCK output, i.e. the gene ranking list, 486 genes were filtered out with a MAGeCK score <= 0.005 out of which *DPP4*, the known entry receptor for MERS-CoV, had a high ranking with a significant MAGeCK score as displayed in the multi-variable plot (Fig. 29). These data indicated that the positive screen selection had worked and all the six sgRNAs targeting the *DPP4* gene were highly enriched in the selected cell population surviving the MERS-CoV infection. In addition, the gene-ranking list provided additional candidate hits including *ANPEP*, an unexpected observation, as this is the HCoV-229E entry receptor as described before, as well as *TMPRSS2* (a well-studied serine protease involved in the entry of some coronaviruses).

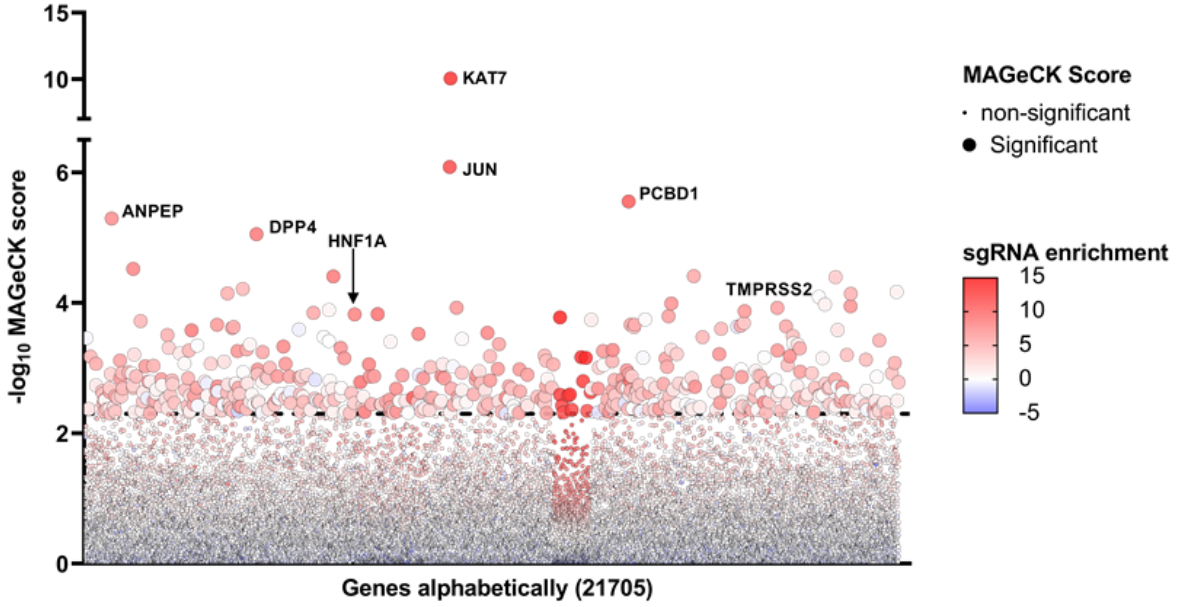


Figure 29: Ranked gene list from the genome-wide sgRNA screen with MERS-CoV-infected cells as retrieved by applying the MAGeCK analysis pipeline.
The gene rank list provided by the MAGeCK tool was used to plot all annotated 21705 genes) against their MAGeCK score ($-\log_{10}$). Genes were sorted alphabetically. The dotted line represents the threshold at $-\log_{10}=2.3$ (corresponding to a MAGeCK score of 0.005). 486 genes were above the threshold of a MAGeCK score <=0.005. The bubble size represents the MAGeCK score and the bubble color indicates the sgRNA enrichment or depletion as $\log_2$ fold change according to the legend on the right. Interesting candidate genes were marked with their gene names.

5.3 Validation of genes identified by the genome-wide sgRNA screens for their functional relevance on virus replication and virus-induced cell death

To validate the functional relevance of the candidate host factors on HCoV-229E viral replication and virus-induced cell death, individual knock out cell lines for 11 host factors (selected from the HCoV-229E 72 h screen) were tested for viral resistance (using plaque assays or viral RNA measurements by RT-qPCR) and cell survival upon infection (using MTS-based cell viability assays). These genes were among the top-ranking order and two or more sgRNAs targeting these genes had a $\log_2$ FC >= 2. Each sgRNA targeting the 11 host factors was selected from the sgRNA library according to the sgRNA screen enrichment analysis and was cloned into a lentiCRIPSR V2 vector that incorporates a stable Cas9 expressing enzyme together with the target sgRNA. This single vector system was transduced in Huh7 cells to achieve the desired gene deletion or a frame-shift mutation, which should both result in the loss of function of the encoded protein.

5.3.1 Effects of the candidate host factor knock outs on virus-induced cell death

To assess the cell viability of the host factor knock outs upon virus infection, each sgRNA-carrying cell line for the 11 host factors was infected with HCoV-229E at an MOI 1. After 72 h p.i., cell viability was measured in all the uninfected controls and the infected cells by MTS assay. In detail, each sgRNA-carrying cell line was seeded onto a 96-well plate at a density of 5000 cells / well in quadruplicates. Huh7 parental and empty vector (LentiCRISPR V2 only lacking a single guide RNA) cell lines were included as a positive control in this experiment. The cells were left either uninfected (control) or infected with HCoV-229E and incubated at 33℃ for 72 h. After the incubation period, the cell viability was measured using an MTS assay as described in the methods section.

The raw absorbance values were background subtracted and the cell viability for the uninfected knock out cell lines was calculated relative to the uninfected Huh7 parental control and for the infected cell lines respective to their uninfected controls. As shown in the bar graphs from Fig. 30A, after 72 h, the knock out host of either of the 11 factors did not show any growth deficiencies in basal conditions. As compared to the uninfected Huh7 parental control, all the sgRNA-carrying cell lines showed almost 100% cell survival as measured by the MTS assay. This data indicated that, individually, the selected 11 host factors were not essential for cell growth or survival.

In contrast, Huh7 parental or empty vector control cells infected with HCoV-229E for 72 h showed that the virus-induced cytopathic effect destroyed almost 60% of the cells with the remaining 40% cells still having metabolic activity (Fig. 30B). Interestingly, the cell lines carrying sgRNAs (sg1 and 2) targeting the ANPEP gene showed almost 100% cell viability after 72 h p.i. (Fig. 30B). This expected effect indicated that the knock out cells lacking ANPEP resisted the entry of HCoV-229E and thereby the cells survived the virus-induced cell death. Nevertheless, the other host factors did not show such an increase in cell viability. Rather, the cell viability was as comparable to the empty vector control cells or the Huh7 parental cells at approximately 40% viability (Fig. 30B).

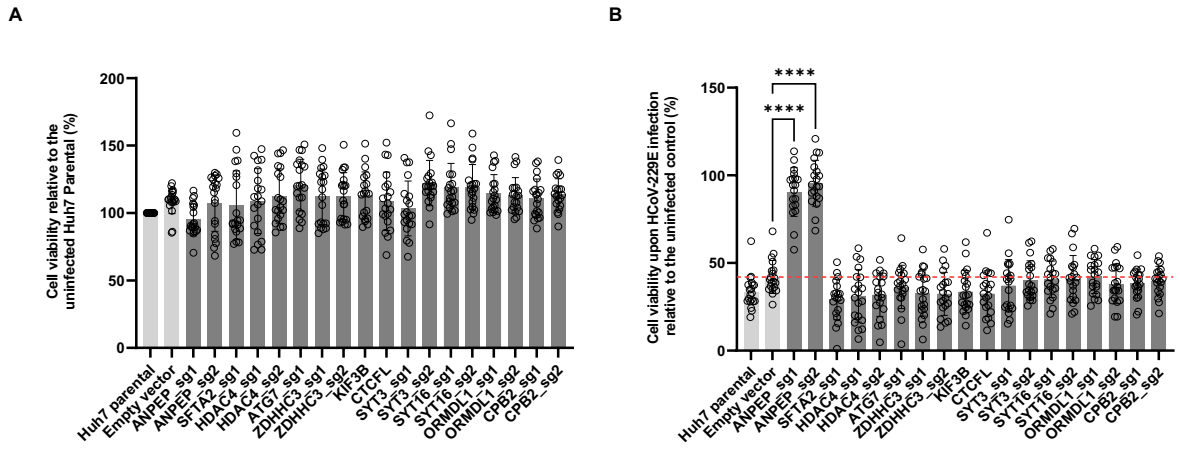


Figure 30: Effects of individual knock outs of hits emerging from genome-wide sgRNA screens on the cell viability of HCoV-229E infected cells.

Each sgRNA-carrying cell line targeting the 11 candidate host factors was seeded on a 96-well plate at a density of 5000 cells / well in quadruplicates and was either left uninfected (control) or infected with HCoV-229E at an MOI 1 for 72 h at 33°C. Huh7 parental and empty vector cell lines were used in this experiment as a positive control to compare the cell viability upon infection with the other cell lines. After 72 h p.i., the MTS reagent was added to each well and incubated at 37°C for 1 h and the absorbance was measured in the colorimeter at 495 nm (readout). **A)** Cell viability of the uninfected sgRNA cell lines was calculated in reference to the uninfected Huh7 parental cell line. **B)** Cell viability of the infected cell lines was calculated in reference to their uninfected controls. The red dotted line indicates the mean cell viability of the empty vector cell line upon infection = 42%. Bar graphs with scatter dots represent the mean data ± SD calculated from five biological independent experiments. **** represents the statistical p-value ($p \leq 0.0001$) obtained by ordinary one-way ANOVA analysis.

5.3.2 Effects of the candidate host factor knock outs on viral replication

To understand whether these host factors contribute to ongoing viral replication, a direct measurement of viral progeny release was performed using plaque assay. For this purpose, sgRNA-carrying cell lines targeting 8 candidate host factors (KIF3B, CTCFL, and CPB2 were excluded in this validation due to unexpected slower growth rate in the cell culture) in addition Huh7 parental and empty vector were seeded onto a 6-well plate at a density of 50,000 cells / well in 2 ml DMEM medium. After 24 h of seeding, the cells were infected with HCoV-229E at an MOI 1 and incubated at 33°C. Every 24 h p.i. until 72 h p.i., 100 µl cell culture supernatant containing viral progeny was collected from all the cell lines and the viral titer was measured by plaque assay as described in detail in the methods section.

The plaques were counted and the viral titers were plotted as bar graphs for each time point as described in the legend of Fig. 31A-C. As shown in Fig. 31A, the viral titer after 24 h p.i., for the infected empty vector cell line was 2.3*10^5 PFU/ml. In contrast, the viral titer was significantly reduced in the ANPEP_sg2 cell line after 24 h p.i. (1.2*10^4) (Fig. 31A). These data indicated that the loss of ANPEP impaired the entry of HCoV-229E and thereby substantially reduced viral replication. However, there were no significant changes in the viral titers for the other host factors at 24 h p.i., when compared to the viral titer of an empty vector cell line (Fig. 31A).

The viral titer for the infected empty vector cell line at 48 h p.i was measured to be 1.7*10^6 PFU/ml (Fig. 31B). As compared to 24 h p.i., time point, the viral titer increased one log scale which indicated that there was no deviation in the ongoing viral replication in the control cells. However, the viral titer was still significantly reduced in the infected ANPEP_sg2 cell line at 48 h p.i., and was measured to be 7*10^5 PFU/ml (Fig. 31B). Interestingly, the viral titers for some of the infected host factors were reduced at 48 h p.i. (when compared to 24 h p.i). Specifically, the viral titers for the infected SFTA2_sg1, HDAC4_sg1, ATG7_sg1, and ZDHHC3_sg2 were slightly reduced as compared to the empty vector. Among these host factors, the most significant reduction in the viral titer was observed to occur in the infected ZDHHC3_sg2 cell line. The viral titer at 48 h p.i. was 4*10^5 PFU/ml, which was around a log scale lower than the titer observed in the empty vector cell line (Fig. 31B).

The viral titers in the infected empty vector control dropped to 2.7*10^5 PFU/ml at 72 h p.i. (Fig. 31C). This was most likely due to the reduction in the viable cells, which suffered from the cytopathic effect. This reduction in the viral titers was also observed in all other cell lines, but there was no significant difference when compared to the viral titer of the infected empty vector cell line (Fig. 31C).

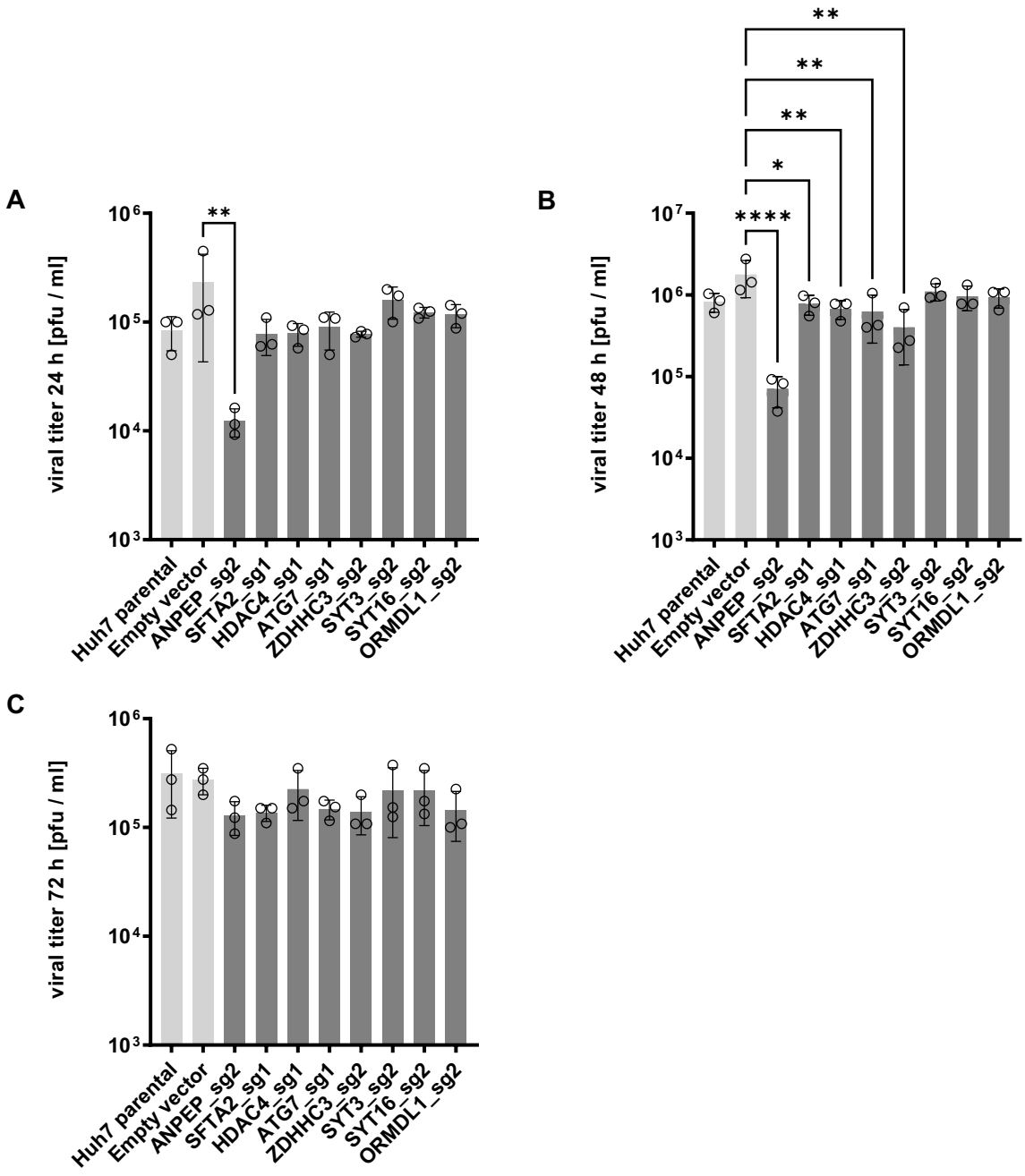


Figure 31: Transient and weak effect of some host factor knock outs on the HCoV-229E viral titers.

All the indicated cell lines were infected with HCoV-229E at an MOI 1 for a period of 24 to 72 h. The cell culture supernatants (100 µl) were taken at 24 h time intervals and were serially diluted 1:10 to determine the viral titer by plaque assays. The plaques were stained with 0.1% crystal violet solution and counted per well for determining the viral titer. **A)** Viral titer measured for the supernatants collected at 24 h p.i. **B)** Viral titer measured for the supernatants collected at 48 h p.i. **C)** Viral titers measured for the supernatants collected at 72 h p.i. Bar graphs with scatter dots represent the mean ± SD measured from three biologically independent experiments and asterisks indicate p values (*$p \leq 0.05$, **$p \leq 0.01$, ****$p \leq 0.0001$) obtained from ordinary one-way ANOVA analysis.

To further reveal the rate of viral RNA replication within the host cells, an absolute viral RNA quantification was performed from the total RNA prepared from the host factor knock out cells infected with HCoV-229E for 48 h. Based on the results from the 48 h p.i. experiments (Fig. 31B), it was most interesting to infect the cells for 48 h at an MOI of 1. The total RNA prepared from all the infected samples was reverse-transcribed to cDNA and absolute quantification of the viral S-RNA transcripts in the infected samples was done using a RT-qPCR. The detailed protocol for the absolute quantification of viral RNA transcripts was described in the methods section.

The viral S-RNA copy numbers were normalized to the total input of intracellular RNA used for cDNA synthesis. The viral S-RNA copy number for the infected empty vector cell line at 48 h p.i., was observed to be at 1.5*10^8 copies / 500 ng of total RNA. This was an interesting observation, as the viral titer observed in this cell line at 48 h p.i. was 1.7*10^6 PFU/ml (Fig. 32). Assuming that one copy of RNA corresponds to one plaque later on, this result would indicate that the rate of viral RNA replication did not exactly correlate to the rate of the release of viral progeny as measured for the viral titers by plaque assays. On the other hand, the viral S-RNA copy number for the infected ANPEP_sg2 was measured to be 8.5*10^6 copies / 500 ng of total RNA (Fig. 32). The viral titer measured for this cell line at 48 h p.i. was 7*10^5 PFU/ml and did not correlate with the viral S-RNA copy numbers measured at 48 h p.i. Nevertheless, as shown in Fig. 32, the viral S-RNA transcription was reduced in the ANPEP_sg2 cell line as compared to the empty vector cell line. However, the viral S-RNA transcription rates for the other host factors measured at 48 h p.i. showed no significant reductions as compared to the empty vector cell line (Fig. 32).

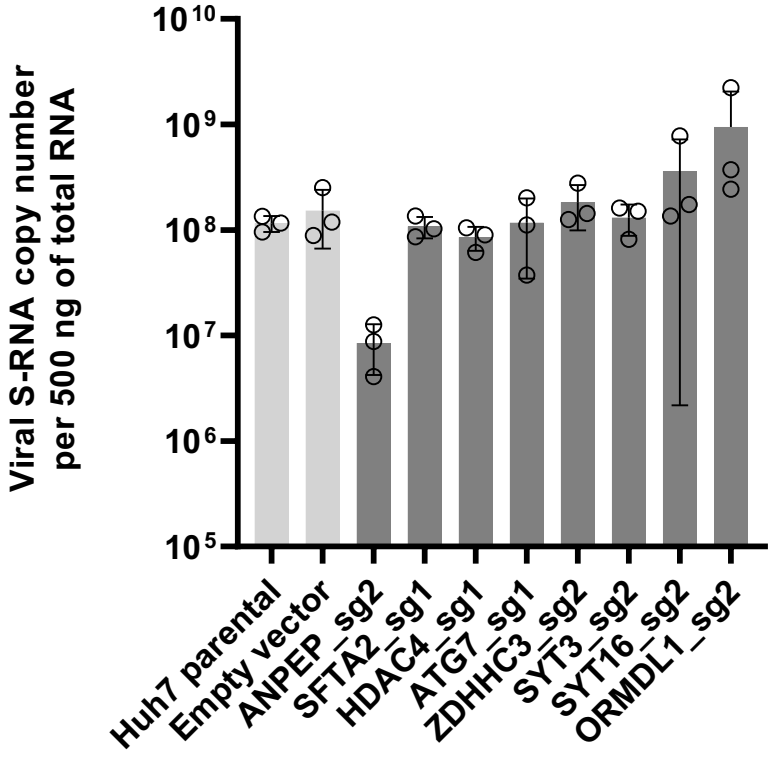


Figure 32: Effect of the host factor knock outs on the viral S-RNA transcription rates as determined by RT-qPCR.

All the indicated cell lines were infected at an MOI 1 for a period of 48 h. Total RNA was prepared from the cells and 500 ng of total RNA was used for cDNA synthesis with random hexamer primers. HCoV-229E S-RNA targeting primers were used for absolute quantification of viral S-RNA by RT-qPCR (see Methods section for details) and the Ct values obtained for the infected samples were compared to the standard curve (purified viral S-RNA PCR product) to obtain the copy numbers. The obtained copy number values were then normalized to the total input of RNA and the final copy numbers for all the infected samples were visualized by bar graph plots. The data shown here represent the mean ± SD obtained from three biologically independent experiments.

5.4 Comparison of the hit lists from the genome-wide CRISPR screens with the other coronavirus genetic screens

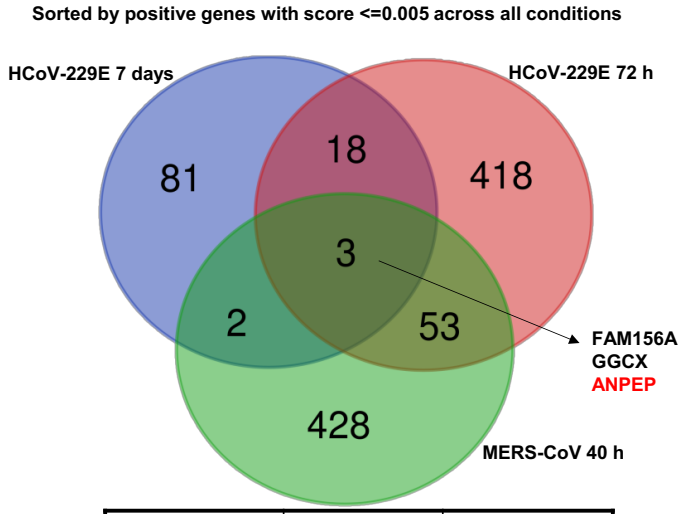
To get more insights into the biological relevance of candidate genes that were selected from all three functional genomic screens, the gene sets across all conditions were first compared to each other to visualize the commonalities among all the screens. In the next step, each screen hit list was compared to the publically available already published coronavirus genetic screens to assess unique or common results from each screen and to identify the shared gene sets and pathways that might represent the most promising candidates, that can be validated in future experiments.

5.4.1 Overlap analysis of the hit lists from all three genome-wide screens performed in this thesis

Gene sets were collected from all three genome-wide screens based on a MAGeCK score <= 0.005. This resulted in gene lists of 104 (HCoV-229E 7 day selection), 492 (HCoV-229E 72 h selection), and 486 hits (MERS-CoV 40 h selection). The gene sets were compared with each other using Venn diagrams and as shown in Fig. 33A, out of 1003 unique genes, only 3 genes were found in all three genetic screens. These genes were identified to be *ANPEP*, *GGCX*, and *FAM156A* (Fig. 33A). Interestingly, the two genetic screens performed for HCoV-229E did not show many overlaps. Only 18 genes were found to be shared between the two groups, indicating strong differences in the requirement of host factors during the kinetic of the infection. The most interesting group was between the HCoV-229E 72 h selection and MERS 40 h selection, where 53 genes were found to be shared between both groups, which included the entry receptor for MERS-CoV, *DPP4* (Fig. 33A).

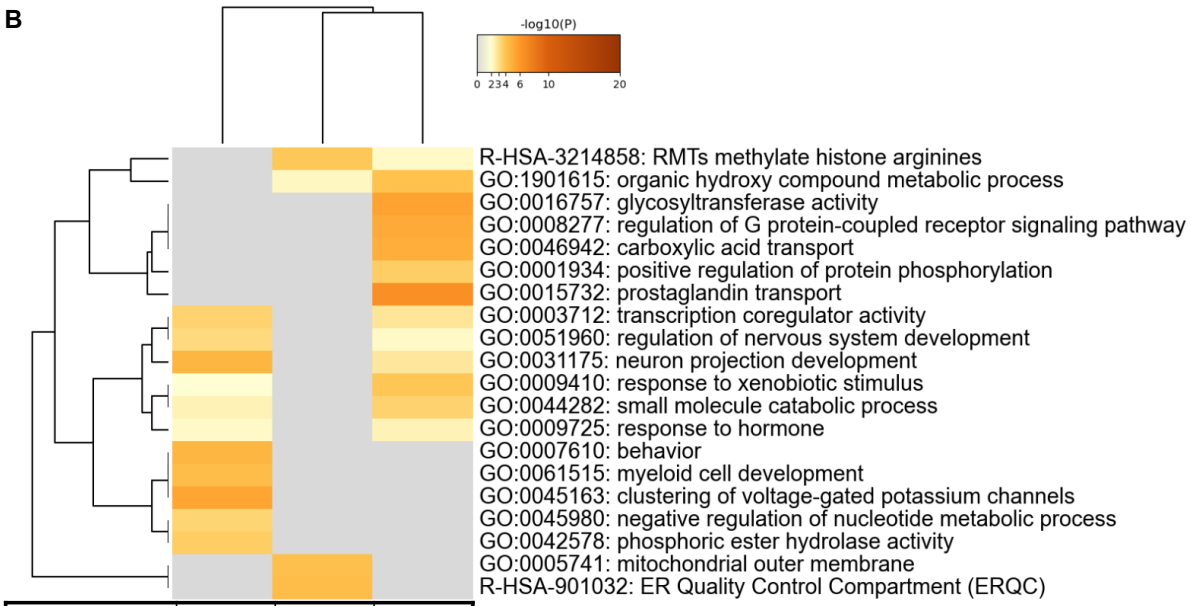
Over-representation analysis for enriched pathway terms using the software Metascape revealed the top 20 most significantly enriched pathways across all three conditions (Fig. 33B). Notable were the relatively weak enriched p-values (Fig. 33B). This indicated that only small numbers of well-annotated genes were contained in the input gene lists from the three sgRNA screens as evidenced by the in term / in list numbers in Fig. 33B. Some shared pathways for HCoV-229E 72 h and MERS 40 h screen selection included pathway GO:0003712: transcription coregulatory activity (34 / 495 genes) and GO:0044282: small molecule catabolic process (29 / 354 genes) that were found to be interesting in the context of coronavirus replication (Fig. 33B). Nevertheless, there were not many significant enriched pathways, which were shared among all the three genetic screens across the top 100 pathways (map not shown in the thesis).

A



List names	number of elements	number of unique elements
HCoV-229E 7 days	104	104
HCov-229E 72 h	492	492
MERS-CoV 40 h	486	486
Overall number of unique elements		1003

B



HCoV-229E 72 h (Score <=0.005)	+		
HCoV-229E 7 days (Score <=0.005)		+	
MERS 40 h (Score <=0.005)			+

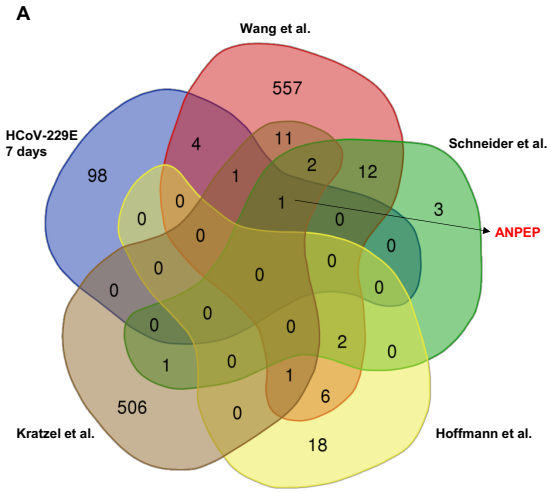
Figure 33: Biological pathways annotated to the gene sets identified in two HCoV-229E and MERS-CoV sgRNA screens.
Genes were sorted out from all three genetic screens using a MAGeCK score <=0.005. **A)** The filtered gene sets were uploaded to an interactive online Venn diagram creator to generate a Venn diagram comparing the overlapping and distinct gene sets from all three conditions. **B)** The filtered gene sets from all three conditions were additionally examined for overrepresented pathway terms using the multiple gene list function of the software Metascape [104].

5.4.2 Comparison of the hit lists from HCoV-229E screen with other, published HCoV-229E genetic screens

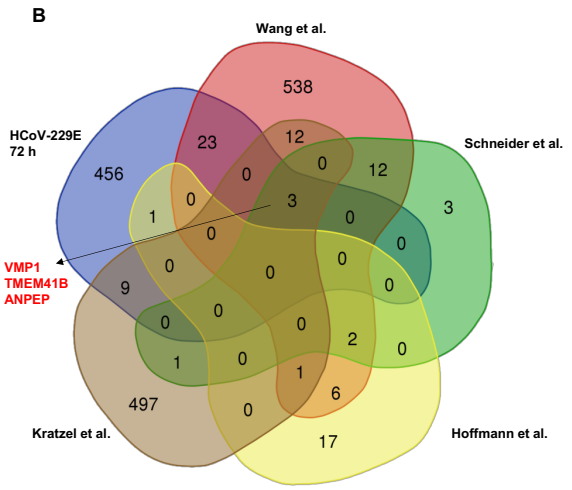
During the completion of our HCoV-229E screens, similar CRISPR-based sgRNA screen studies in Huh7 cells were published. It was therefore intriguing to compare the gene lists obtained from both HCoV-229E screens performed during this thesis to the other published data. Gene lists were obtained from the following three studies (Wang et al., 2021 [116], Schneider et al., 2021 [117], Hoffmann et al., 2021 [118], and Kratzel et al., 2021 [119]). The genes were sorted out from these datasets using the same stringency (MAGeCK score <= 0.005) as applied for the HCoV-229E 7 or 3 days selections. This stringency factor could not be applied to the results from Hoffmann et al., 2021, or Schneider et al., 2021 since in these studies, the top hits were calculated using a rank Z-score matrix, and the criterion for hits was a Z-score below or above 2.4. Altogether, 27 (Hoffmann et al.), 21 (Schneider et al.), 597 (Wang et al.), and 523 (Kratzel et al.) genes were collected as significant hits in these screens. Their overlaps were compared to the gene sets identified from HCoV-229E 7 day (104 genes) and 3-day selection (492 genes) using a 5 groups Venn diagram visualization. In the first comparison, 492 genes from HCoV-229E 72 h selection were compared to the other datasets. The Venn diagram showed no genes which were found to be shared across all the groups (Fig. 34B). There were 3 genes (*ANPEP*, *VMP1*, and *TMEM41B*) that were found in 4 groups except for Hoffmann et al. (Fig. 34B). However, there were only 23 genes found from the HCoV-229E 3 days screen, which were shared with Wang et al. and 9 genes which were shared with Kratzel et al. (Fig. 34B). In the second comparison, 104 genes from HCoV-229E 7 day selection were compared with 4 other groups. *ANPEP* was the only gene, which was shared among 4 groups except Hoffmann et al. (Fig. 34A). There were no interesting commonalities between the HCoV-229E 7 days group with the other datasets (Fig. 34A).

In the next step, the filtered gene sets from all the groups were used to generate a clustered heat map containing the top 20 differentially enriched pathways by overrepresentation analysis using Metascape. The heat map showed (Fig. 34C), many strongly enriched pathways for the gene list from Wang et al., related to vesicle-mediated transport (R-HAS-5653656), vacuole (GO:0007033) and membrane organization (GO:0061024), autophagy (GO:0006914), and glycosylation (GO:0070085). Some of the pathways were also commonly enriched in the gene lists from Schneider et al., Hoffmann et al., and Kratzel et al. (Fig. 34C). However, only the autophagy pathway was found to be enriched among all the

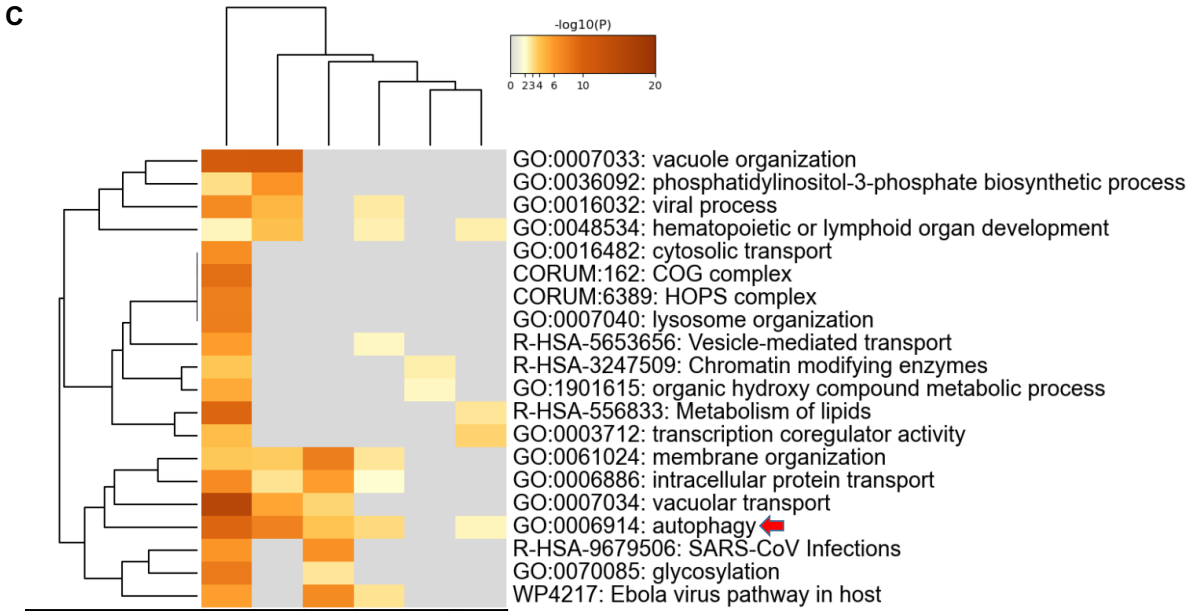
groups except the HCoV-229E 7 days group (Fig. 34C). Gene sets from both HCoV-229E screens and Kratzel et al. showed only weak enriched pathways as compared to the other three groups (Fig. 34C).



List names	number of elements	number of unique elements
HCoV-229E 7 days	104	104
Hoffmann et al.	27	27
Kratzel et al.	523	523
Schneider et al.	21	21
Wang et al.	597	597
Overall number of unique elements		1223



List names	number of elements	number of unique elements
HCoV-229E 72 h	492	492
Hoffmann et al.	27	27
Kratzel et al.	523	523
Schneider et al.	21	21
Wang et al.	597	597
Overall number of unique elements		1581



Wang et al. (Score <=0.005)	+					
Schneider et al. (Z-Score)		+				
Hoffmann et al. (Z-Score)			+			
Kratzel et al. (Score <=0.005)				+		
HCoV-229E 7 days (Score <=0.005)					+	
HCoV-229E 72 h (Score <=0.005)						+

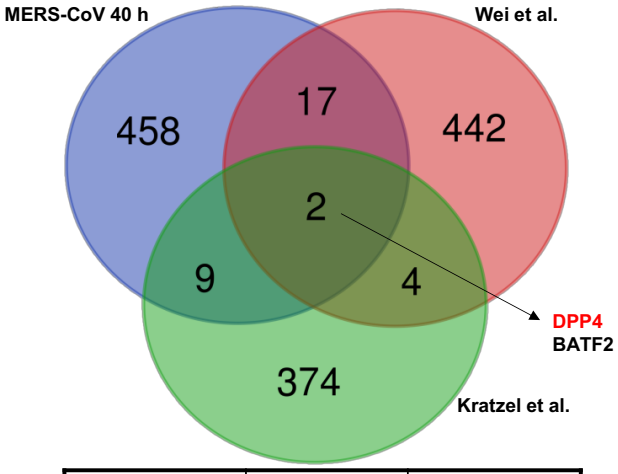
Figure 34: Overlapping and distinct sets of genes and pathways identified across published HCoV-229E sgRNA screens and the screens of this thesis.
Genes were retrieved from all available genetic screens based on a MAGeCK score <= 0.005, except for Hoffmann et al., and Schneider et al. (where hits were selected based on ranked Z-Scores). **A)** The filtered gene sets were uploaded onto an interactive online Venn diagram creator to generate a Venn diagram comparing the gene sets from HCoV-229E 7 day selection to the other 4 groups. **B)** The gene sets from HCoV-229E 3 day selection were compared to the other groups **C)** The filtered gene sets from all the groups were used to generate a clustered heat map containing the top 20 enriched pathway terms according to Metascape analysis.

5.4.3 Comparison of the hit lists from the MERS-CoV 40 h screen with the other MERS genetic screens

The gene sets filtered from MERS-CoV 40 h screen selection were compared to gene sets filtered out from two recently published MERS-CoV genetic screens (Wei et al., 2021 [120] and Kratzel et al., 2021 [119]). The gene sets were filtered out from Kratzel et al. based on a MAGeCK score <= 0.005 and for Wei et al., the genes were sorted based on a Z-score > 2.7. This resulted in a gene list of 389 genes for Kratzel et al., and 465 genes for Wei et al., which were compared to the 486 genes filtered out from the MERS-CoV 40 h sgRNA screen of this thesis. Venn diagram comparison showed that there were only two genes (*DPP4*, and *BATF2*) that were found in all three groups (Fig. 35A). MERS-CoV 40 h screen selection shared only 17 genes with Wei et al., and only 9 genes with Kratzel et al. There were only 4 genes that were shared between Wei and Kratzel et al. (Fig. 35A).

In the next step, the filtered gene sets from all the groups were used to generate a clustered heat map containing the top 20 differentially enriched pathway terms by overrepresentation analysis using Metascape. The heat map shown in Fig. 35B revealed many weakly enriched pathways, which were shared among all three groups. These were related to chromatin-modifying enzymes (R-HAS-3247509), positive regulation of protein phosphorylation (GO:0001934), and transcriptional regulation by RUNX1 (R-HSA-8878171). Pathways related to carboxylic acid transport (GO:0046942), glycosyltransferase activity (GO:0016757), and G protein-coupled receptor signaling (GO:0008277) were only found to be enriched in the MERS-CoV 40 h group (Fig. 35B).

A



List names	number of elements	number of unique elements
MERS-CoV 40 h	486	486
Kratzel et al.	389	389
Wei et al.	465	465
Overall number of unique elements		1306

B

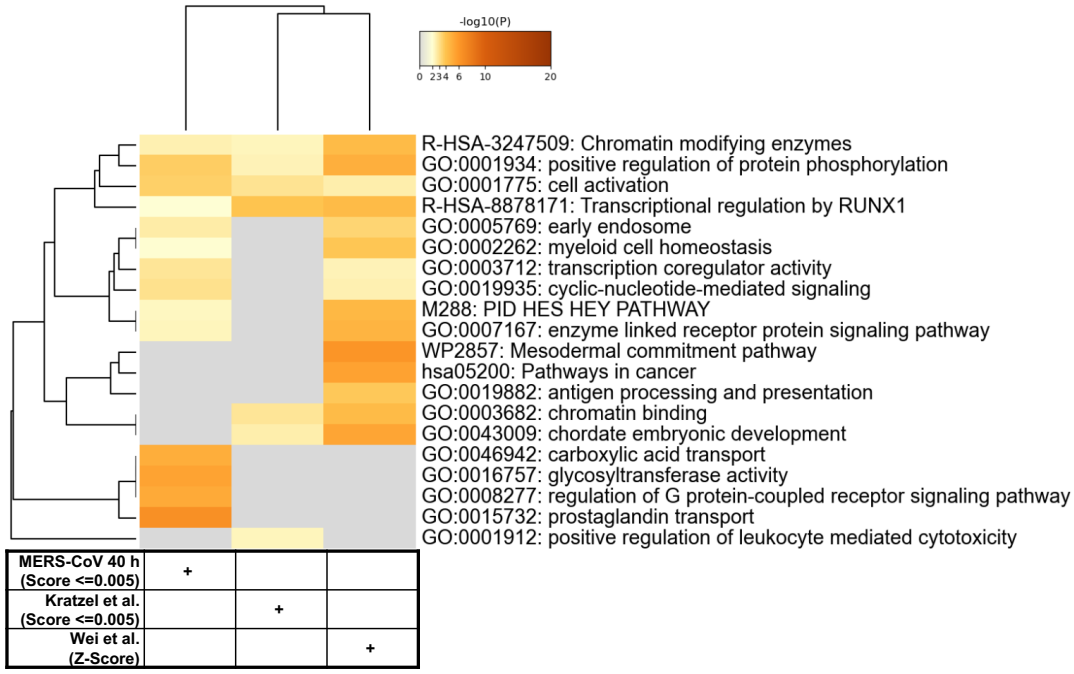


Figure 35: Comparison of the gene sets identified from sgRNA screens of MERS-CoV-infected cells during this thesis with other published MERS-CoV sgRNA screen datasets.
Genes were sorted out from all the genetic screens using a MAGeCK score <= 0.005, except for Wei et al. (which used a Z-score > 2.7 as a criterion). **A)** The filtered gene sets were uploaded onto an interactive online Venn diagram creator to generate a Venn diagram comparison. **B)** The filtered gene sets from all the groups were used to generate a clustered heat map containing the top 20 enriched pathway terms according to Metascape analysis.

5.5 Identification of human genes involved in cell death and anti-coronavirus activity mediated by the small compound thapsigargin using genome-wide CRISPR-Cas9 knock out screening

Recently published data from our working group showed that the replication of three human coronaviruses (HCoV-229E, SARS-CoV-2, and MERS-CoV) was inhibited by ER-stress inducer thapsigargin (Tg) in multiple cell types [64]. In these models, Tg also enhanced the cell viability upon infection by unknown mechanisms. These results raised the question if an unbiased genome-scale CRISPR-Cas9 knockout screen could be used to identify the host factors, which contribute to the Tg-mediated antiviral effects.

5.5.1 Determination of optimal conditions for the thapsigargin treatment and selection before setting up the genome-wide sgRNA screens

Before the screen selection, it was very important to determine the time points of infection / Tg treatment and the optimal, antiviral concentration of Tg with the lowest cytotoxic effect to the cells. A small-scale experiment was conducted in Huh7 cells infected with HCoV-229E at an MOI 1 for 72 h and / or 96 h, treated in the presence or absence of different concentrations (100 nM, 300 nM, and 1 µM) of Tg. In brief, 5*10^3 Huh7 cells were seeded onto a 96-well plate in quintuplicates for uninfected controls and for HCoV-229E infection. For the Tg treatment or in combination of Tg with HCoV-229E infection the cells were seeded in triplicates. In addition, Huh7 cells were treated with DMSO at the same concentrations used for Tg treatment (as vehicle control). The cells were infected and / or Tg treated for 72 h and / or 96 h. Thereafter, the cell viability from all the conditions and time points was assessed using MTS assays.

At the 72 h time point, DMSO treatment did not affect cell viability, even at the highest concentration used (1 µM) (Fig. 36A). In contrast, the cell viability upon 72 h of Tg treatment was greatly reduced to under 50% (Fig. 36A). This effect was comparable for all the different Tg concentrations used (Fig. 36A). The cell viability upon 72 h of HCoV-229E was strongly reduced to under 20% (Fig. 36A). Interestingly, the Tg treatment rescued the cell viability in the infected cells to almost 50% (Fig. 36A).

Likewise, DMSO did not affect cell viability after 96 h of treatment (Fig. 36B). However, the Tg treatment for 96 h strongly reduced cell viability under 20% (Fig. 36B). The cell viability was reduced to nearly 10% only in the infected cells at 96 h p.i. (Fig. 36B). Nevertheless, the Tg-treated infected cells showed increased cell viability to 20% (Fig. 36B).

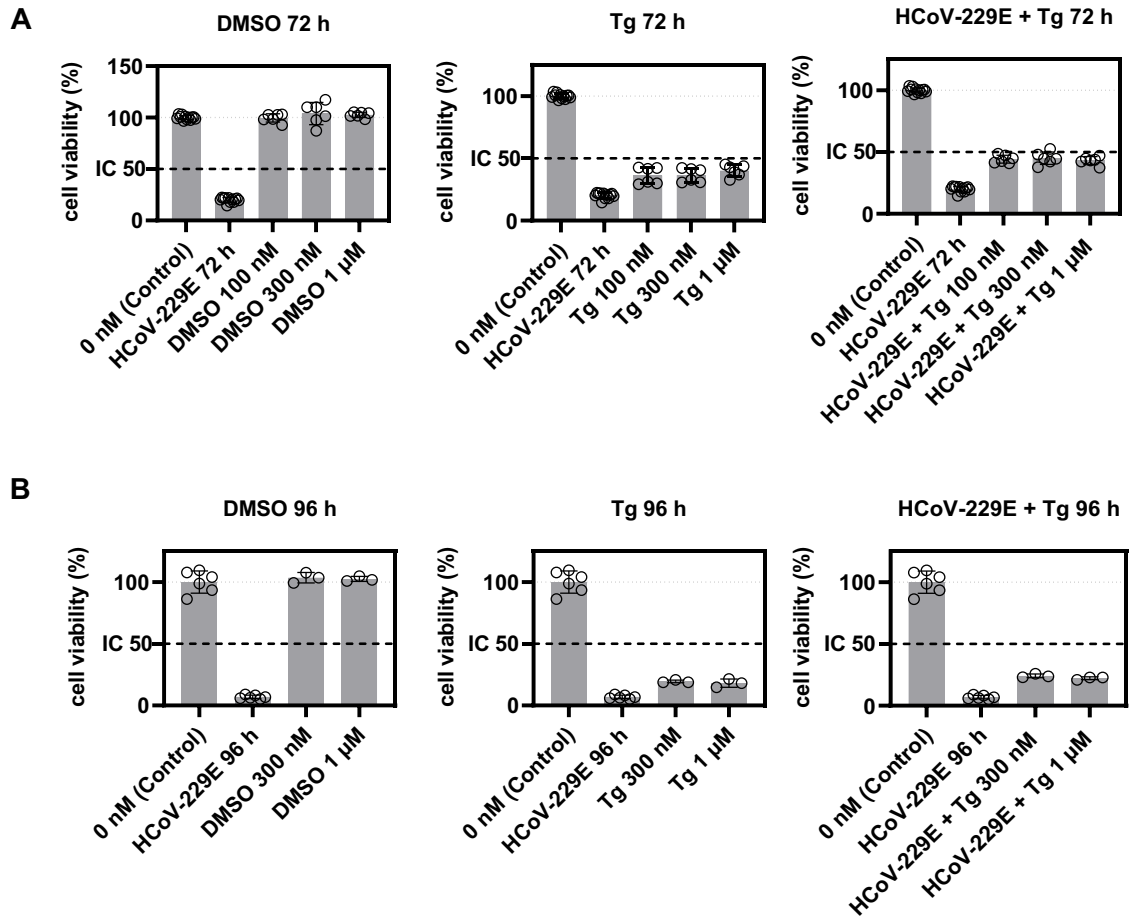


Figure 36: Cell viability assays to determine the optimal Tg concentration and treatment durations for subsequent genetic screens of HCoV-229E-infected cells.
Huh7 cells were infected with HCoV-229E in the presence or absence of various concentrations of Tg as indicated for 72 h and 96 h. Thereafter, cell viability was measured by MTS assay. **A)** Cell viability at 72 h for DMSO controls, Tg treatment, HCoV-229E infection, or a combination of both, Tg and infection. The bar graph represents mean ± SD from two biologically independent experiments, performed in technical triplicates. **B)** Cell viability at 96 h for DMSO controls, Tg treatment, and HCoV-229E infected Tg treated cells. The bar graph represents mean ± SD from the technical replicates of one biological experiment. Dr. Christin Mayr-Buro (Prof. Kracht's working group) kindly performed the MTS assays, evaluated the raw data, and generated the graphs.

5.5.2 Genome-wide screen setup for selection of virus and / or Tg-resistant cells

Based on the MTS assay results, it was decided to use 1 µM Tg (for high selection pressure against cell death) and two-time points of treatment and / or infection. The overall design of the genome-wide screen for HCoV-229E and / or Tg treatment conditions is illustrated in the figure below. 200 million library transduced cells were infected with HCoV-229E (at an MOI of 1) in the presence or absence of 1 µM Tg. 80 million non-treated, library transduced cells served as an unselected control population. Genomic DNA from surviving cells was harvested at one early time point, i.e. 72 h after selection, or, 96 h after treatment / infection and further expansion of the small surviving cell population for about two to four weeks (to obtain enough cells for sequencing). The PCR amplified sgRNA libraries from three independent experiments of control and selected populations were subjected to DNA sequencing. Normalized read counts were analyzed by MAGeCK bioinformatics tools to identify statistically significant differences in the abundance of integrated sgRNAs in the selected compared to the control populations.

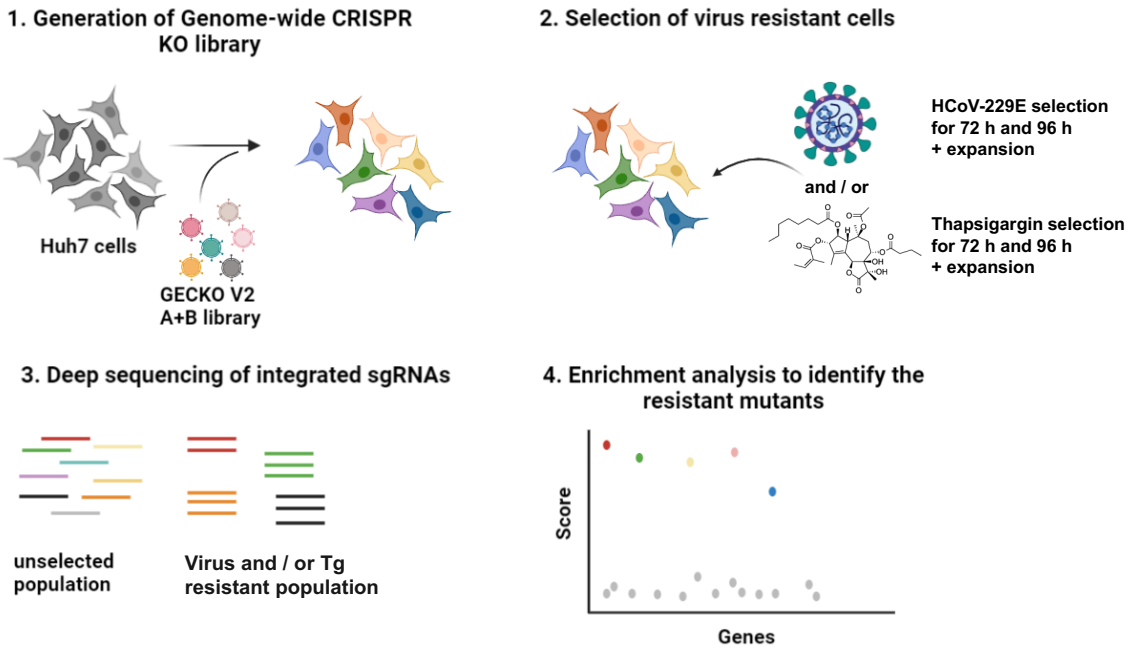


Figure 37: Schematic overview of the genetic screen to identify genes that contribute to Tg-mediated anti-viral or cell death effects.
The scheme shows the experimental setup for the screen selection with the library transduced Huh7 cells challenged with HCoV-229E and / or Tg treatment for 72 h and / or 96 h. This image was created from the BioRender app (https://app.biorender.com/).

Library transduced Huh7 cells were seeded onto 40 145 mm cell culture dishes at a density of 5*10^6 cells in 20 ml DMEM medium for each of the three treatment / infection conditions. The library transduced cells were challenged with HCoV-229E (MOI 1) and / or Tg (1 µM) for 72 h. As an unselected control, genomic DNA from 80 million library transduced cells was isolated to serve as a full library representation of all transduced sgRNAs for NGS. As a positive control for virus-induced cytopathic effects and Tg-mediated cytotoxicity, Huh7 parental cells were seeded onto a 145 mm dish at a density of 5*10^6 cells and were infected with HCoV-229E (MOI 1) and / or Tg (1 µM) for 72 h in an identical manner.

The cytopathic / cytotoxic effects were visible by light microscopy at the end of the selection time (72 h) in both, the parental as well as the library transduced cells exposed to HCoV-229E infection and / or Tg-treatment (Fig. 38). At this point, the first series of genomic DNA was isolated from infected and / or Tg-treated library cells in three biologically independent experiments.

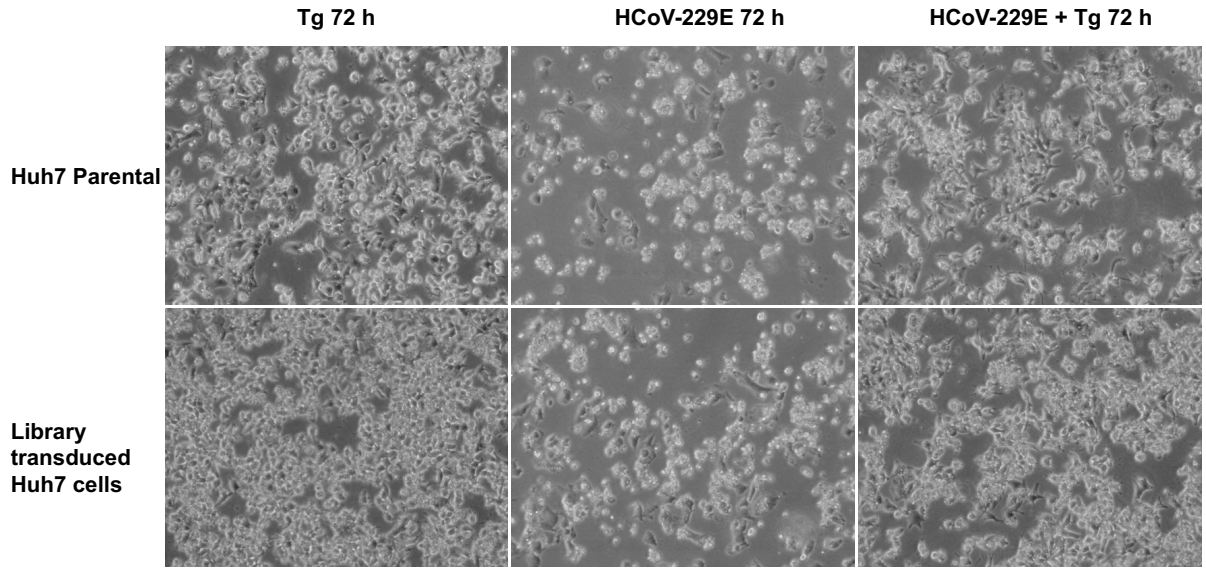


Figure 38: Microscopic images from the 72 h screen selection of parental or library transduced cells challenged with HCoV-229E and / or Tg treatment.
Huh7 parental cells (first horizontal panel) or library transduced Huh7 cells (second horizontal panel) were infected with HCoV-229E (MOI 1) and / or Tg (1 µM) for 72 h and were examined for cytopathic / cytotoxic effects after 72 h of selection. Images shown here were representative microscopic images from three biologically independent experiments. Dr. Christin Mayr-Buro and Dr. Samer Shaban (Prof. Kracht's working group) performed one or more biological replicates of the screen selection. The microscopic images were generated together with Dr. Christin Mayr-Buro.

In the second experiment, infection with HCoV-229E and / or Tg treatment were continued for 96 h. According to the optimization experiments shown in Fig. 36, most cells died under these conditions. Therefore, the surviving cells were further expanded for about two to three weeks (without renewed infection or Tg treatment) and the genomic DNA was isolated from the expanded surviving cell populations to amplify the integrated sgRNAs. As an unselected control, genomic DNA from 80 million library transduced cells was isolated to serve as a full library representation for NGS. As a positive control for virus-induced cytopathic effects and Tg-mediated cytotoxicity, Huh7 parental cells were seeded onto a 145 mm dish at a density of 5*10^6 cells and were infected with HCoV-229E (MOI 1) and / or treated with Tg (1 µM) for 96 h.

As before, the cytopathic / cytotoxic effects were visualized under a light microscope at the end of the selection time of 96 h (Fig. 39). Strong cytopathic effects were visible in Huh7 parental cells 96 h p.i. (Fig. 39, upper panels). In contrast, library transduced cells showed more healthy and surviving cells at 96 h p.i. upon HCoV-229E infection, and there were more surviving cells in Tg alone or Tg combined with infection conditions (Fig. 39, lower panels). After the selection time of 96 h, the surviving cells from all three conditions were further expanded for two to three weeks to reach a confluency of about 80 million cells. Genomic DNA was isolated from infected and / or Tg-treated library cells after further expansion. Screen selection for 96 h was performed in three biologically independent experiments.

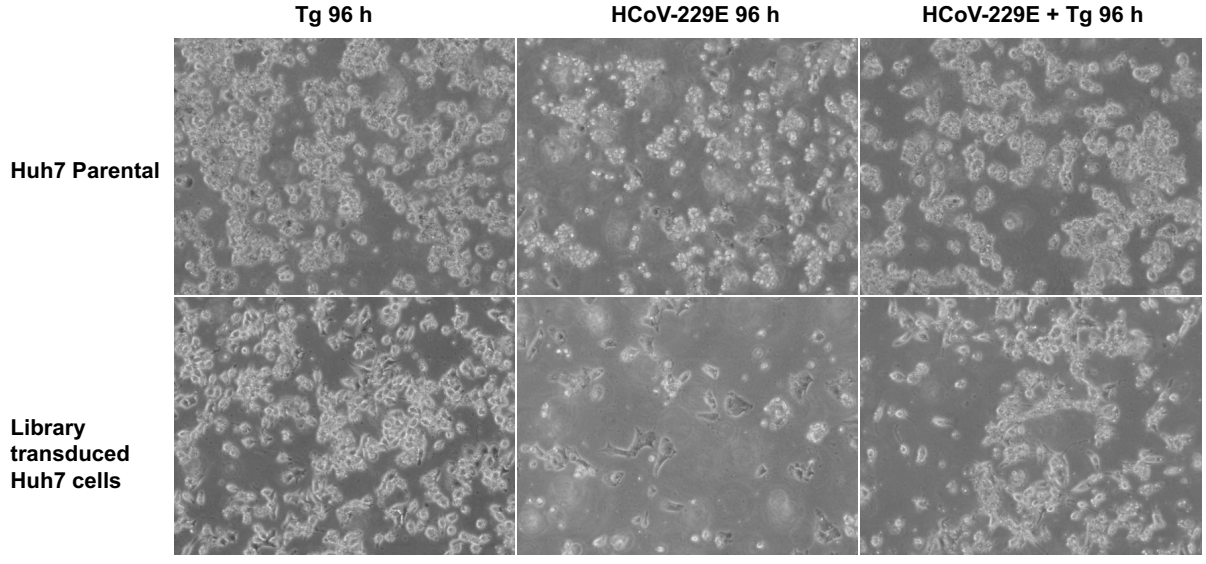


Figure 39: Microscopic images from the 96 h screen selection of parental or library transduced cells challenged with HCoV-229E and / or Tg treatment.
Huh7 parental cells (first horizontal panel) or library transduced Huh7 cells (second horizontal panel) were infected with HCoV-229E (MOI 1) and / or Tg (1 µM) for 96 h and were examined for cytopathic / cytotoxic effects after 96 h of selection. Images shown here were representative microscopic images from three biologically independent experiments. Dr. Christin Mayr-Buro and Dr. Samer Shaban (Prof. Kracht's working group) performed one or more biological replicates of the screen selection. The microscopic images were generated together with Dr. Christin Mayr-Buro.

5.5.3 Quality control metrics revealed that the unselected controls showed an unbiased sgRNA representation

After the treatments / infections for 72 h and 96 h and further selection as described above, genomic DNA from all conditions and the unselected controls was isolated. This resulted in 24 genomic DNA samples isolated from all the conditions and biological replicates. The integrated sgRNAs were amplified by the PCR library amplification protocol as mentioned in the methods section. The amplicons were barcoded separately during the library PCR amplification and the samples were pooled together and sent to the company Novogene for deep sequencing by NGS. Paired-end sequencing of 150 bp of the samples was performed on a Novaseq 6000 PE150 platform to achieve a sequencing depth of ~ 50 - 60 million reads / sample. The demultiplexed fastq files were analyzed by the MAGeCK tool using the software R [99] to generate an overview table consisting of read counts that were sequenced and mapped to the input sgRNA library sequences.

As shown in Table 2, for all the samples the read counts were around 50 – 60 million reads, which indicated that the NGS sequencing gave the intended sequencing depth. For all the samples, around 60 to 70% of the sequenced read counts were matched with the input library sgRNA sequences. Moreover, the unselected control samples for 72 h and 96 h selection showed only around 3000 – 4000 sgRNAs that had zero read counts. In addition, only around 5500 – 7000 sgRNAs had a read count <= 10. The gini index for the unselected controls was between 0.13 to 0.14, which indicated that the samples had equally distributed individual sgRNA read counts.

Label	Reads	Mapped	Percentage	Total sgRNAs	Zero counts	Counts <= 10	Gini Index
Control for 72 h_BIO1	55,832,636	38,050,049	0.68	119461	3263	5814	0.1343
Control for 72 h_BIO2	55,127,098	37,624,158	0.68	119461	3478	6142	0.1378
Control for 72 h_BIO3	56,947,097	33,123,961	0.58	119461	4553	7061	0.1658
Control for 96 h_BIO1	53,574,457	33,320,709	0.62	119461	3517	6431	0.1994
Control for 96 h_BIO2	56,027,166	38,681,032	0.69	119461	3699	6316	0.1393
Control for 96 h_BIO3	52,827,652	36,731,761	0.70	119461	4274	7097	0.1465
HCoV-229E 72 h_BIO1	50,562,855	34,731,657	0.69	119461	3484	6199	0.1378
HCoV-229E 72 h_BIO2	50,564,839	34,505,112	0.68	119461	4566	7250	0.1474
HCoV-229E 72 h_BIO3	54,973,643	30,994,414	0.56	119461	5119	8119	0.1641
Tg 72 h_BIO1	69,730,821	42,733,476	0.61	119461	3284	5646	0.1332
Tg 72 h_BIO2	59,313,869	36,488,542	0.62	119461	3548	6462	0.1403
Tg 72 h_BIO3	51,144,918	32,445,976	0.63	119461	3697	6824	0.1472
HCoV-229E + Tg 72 h_BIO1	64,388,903	43,970,009	0.68	119461	3382	5675	0.1335
HCoV-229E + Tg 72 h_BIO2	54,215,548	36,019,234	0.66	119461	3852	6797	0.1434
HCoV-229E + Tg 72 h_BIO3	54,800,977	30,512,889	0.56	119461	4144	7385	0.155
HCoV-229E 96 h_BIO1	58,212,222	39,191,930	0.67	119461	22221	109579	0.4028
HCoV-229E 96 h_BIO3	51,107,766	30,725,967	0.60	119461	38126	115252	0.4977
Tg 96 h_BIO1	50,527,169	37,884,084	0.75	119461	19490	72592	0.3717
Tg 96 h_BIO2	55,794,015	31,488,071	0.56	119461	56994	89143	0.7245
Tg 96 h_BIO3	50,197,917	34,803,190	0.69	119461	35161	74258	0.5799
HCoV-229E + Tg 96 h_BIO2	79,580,640	46,255,696	0.58	119461	7204	46381	0.2685
HCoV-229E + Tg 96 h_BIO3	53,558,931	42,860,491	0.80	119461	41413	214502	0.5033

Table 5: Overview table showing the statistics of DNA sequencing results of the sgRNA screen as analyzed by MAGeCK software.
The table shows the overall sequencing depth for all the samples subjected to NGS sequenced. In addition, the percentage of mapped read counts to the input sgRNA library sequences, the total number of sgRNAs used for the analysis, the number of sgRNAs that were undetected, and their gini indices are indicated. Sebastian Werner kindly generated the output from the NGS sequenced samples listed in the table using the MAGeCK tool.

From the MAGeCK analysis, the raw read counts before and after median normalization were $\log_2$ transformed and plotted in histograms to visualize the read count distribution. As shown in the histogram bars (Fig. 40), there were only around 6000 to 7000 sgRNAs (from > 100,000), which had very low relative frequencies in all the control samples, which indicated that the unselected control samples showed an unbiased representation of the full sgRNA library.

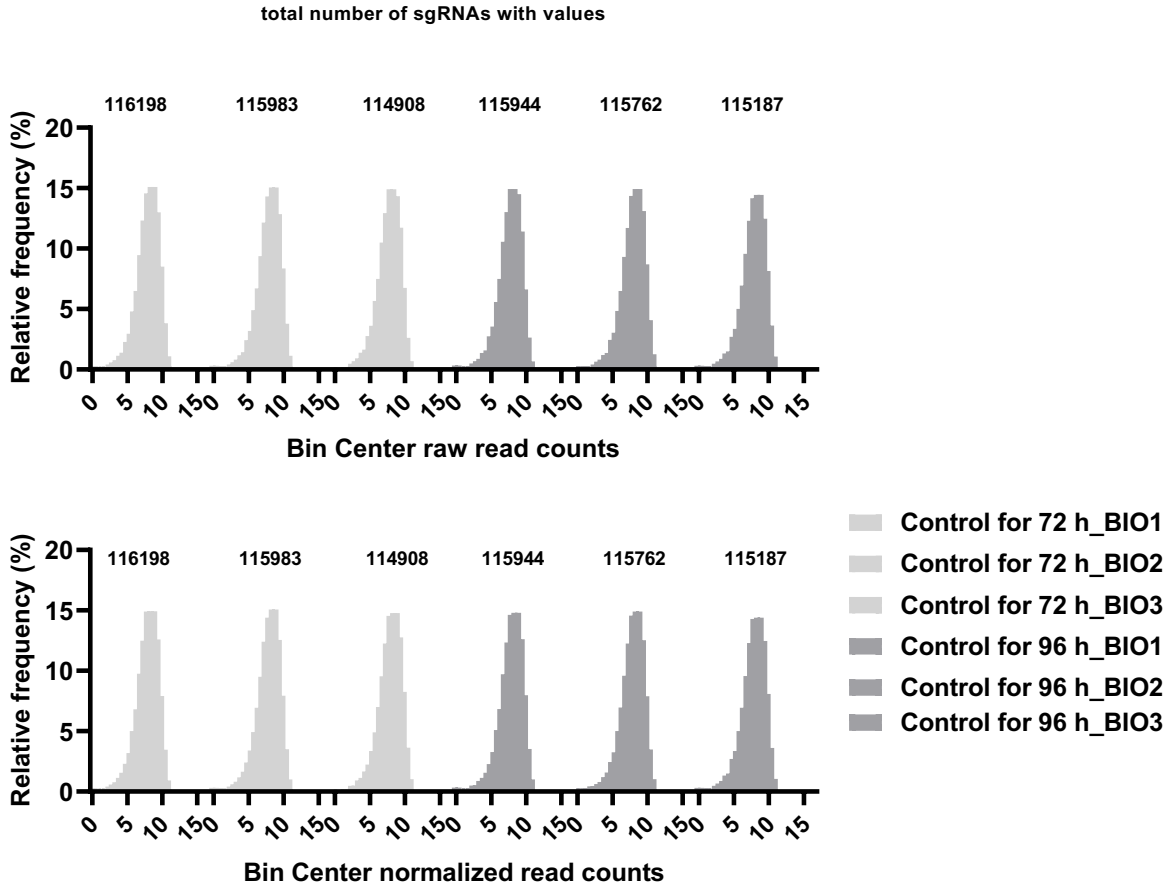


Figure 40: Quality control metrics for the deep sequenced unselected controls for 72 h and 96 h screen selection.
MAGeCK tool was used to determine the read counts in the unselected samples. The read counts from all the samples were $\log_2$ transformed and median normalized and quality control metrics were calculated to determine the read count distribution in unselected samples. The histogram depicts the frequency distribution of raw read counts in the samples before (upper) and after (below) median normalization with the values counted for each sample. The figures were generated and provided by Prof. Kracht.

5.5.4 Quality control metrics for the sgRNA screen results for samples from thapsigargin-treated and / or infected cells

Tg-treated and selected samples from the 72 h time point showed an equal and normal read count distribution as shown in the histograms (Fig. 41) before and after median normalization. On the other hand, after 96 h of Tg treatment and selection, the samples showed a strong shift in the histogram peaks towards the left. This indicated that there was an enormous depletion and / or enrichment of sgRNAs (Fig. 41). About 30,000 to 60,000 sgRNAs were either depleted or had a very low relative frequency as shown in the histograms (Fig. 41). This indicated that Tg treatment had a very strong selection pressure on the library transduced cells in all of the three biologically independent experiments.

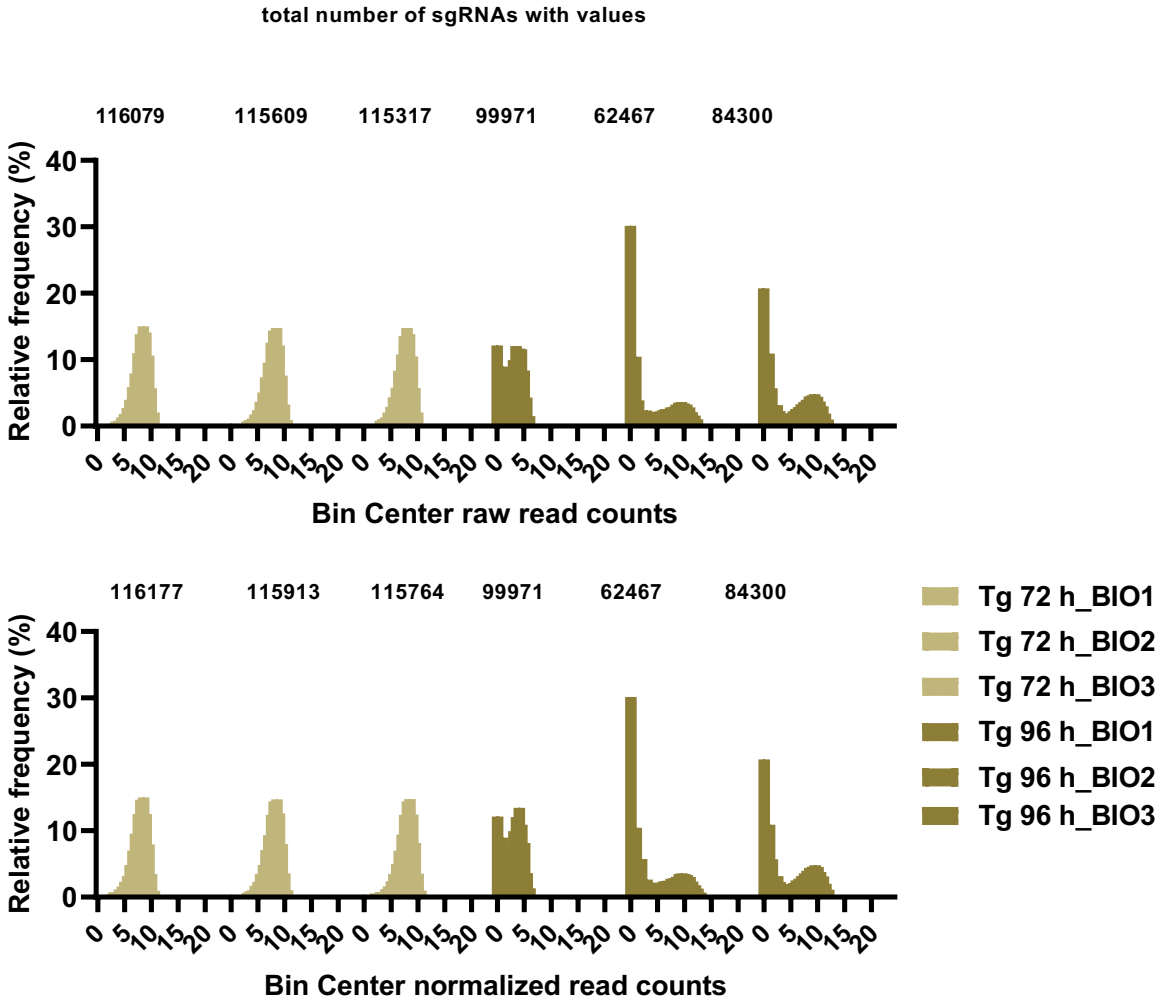


Figure 41: Quality control metrics for the deep sequenced Tg selected samples for the 72 h and 96 h time points.
MAGeCK tool was used to determine the read counts in the Tg-treated and selected samples for 72 h and 96 h. The read counts from all the samples were $\log_2$ transformed and median normalized and quality control metrics were calculated to determine the read count distribution in all the samples. The histogram depicts the frequency distribution of read counts in the samples before (upper) and after (below) median normalization with the values counted for each sample analyzed in three biologically independent experiments. The figures were generated and provided by Prof. Kracht.

HCoV-229E infected samples, selected for 72 h, showed an equal and normal read count distribution as shown in the histograms (Fig. 42) before and after median normalization. On the other hand, HCoV-229E infected samples, kept for 96 h and then further selected showed a strong shift in the histogram peaks towards the left. This indicated that there was a large depletion and / or enrichment of sgRNAs (Fig. 42). About 30,000 to 40,000 sgRNAs were either depleted or had a very low relative frequency as shown in the histograms (Fig. 42). This indicated that the prolonged infection with HCoV-229E 96 h provided a very strong selection pressure on the library transduced cells in the two available biologically independent experiments (in the third experiment, not enough viable cells could be recovered to perform DNA sequencing).

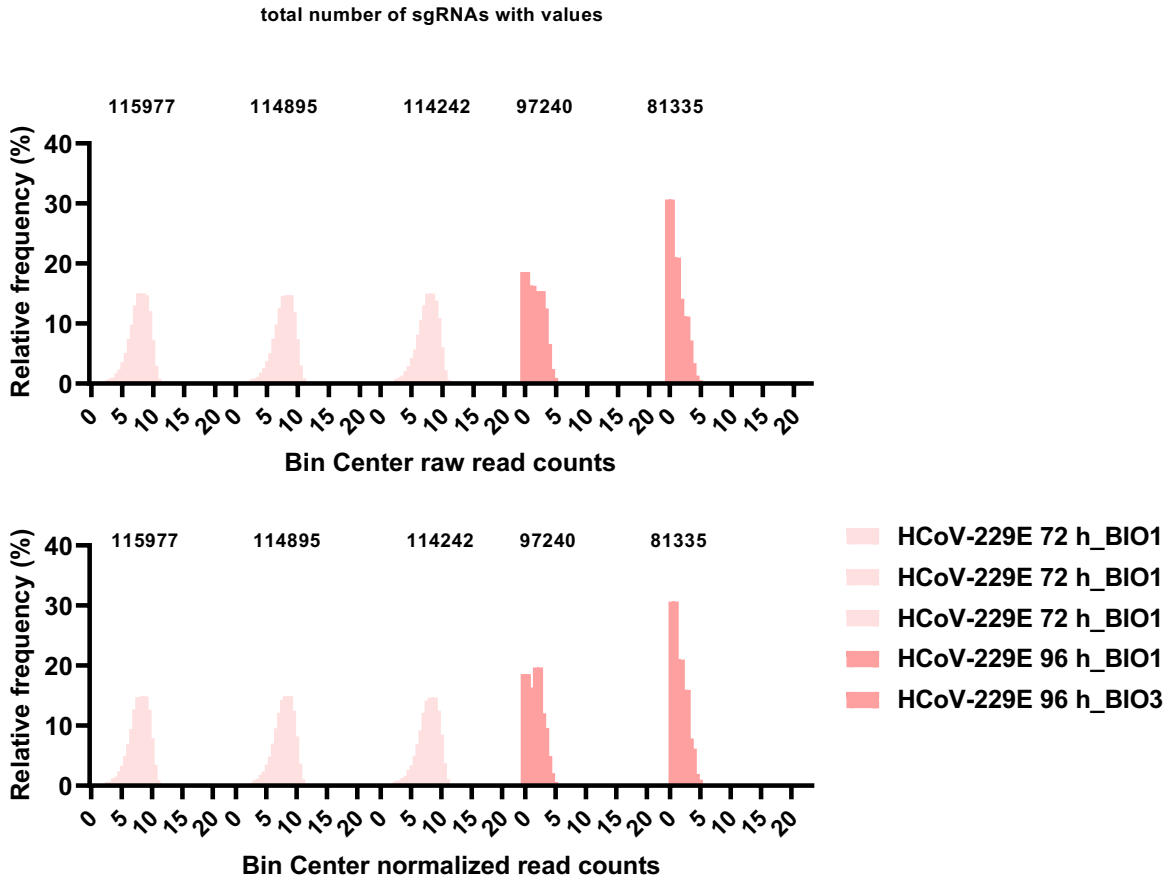


Figure 42: Quality control metrics for the deep sequenced HCoV-229E selected samples for the 72 h and 96 h time points.
MAGeCK tool was used to determine the read counts in the HCoV-229E infected and selected samples for 72 h and 96 h. The read counts from all the samples were $\log_2$ transformed and median normalized and quality control metrics were calculated to determine the read count distribution in all the samples. The histogram depicts the frequency distribution of read counts in the samples before (upper) and after (below) median normalization with the values counted for each sample analyzed in three biologically independent experiments. The figures were generated and provided by Prof. Kracht.

Likewise, HCoV-229E infected and Tg-treated samples from the 72 h time points, showed an equal and normal read count distribution as shown in the histograms (Fig. 43) before and after median normalization. For the 96 h time points, only two samples could be recovered which yielded enough DNA for deep sequencing (96 h BIO2 and BIO3). These two samples differed substantially in the overall read count distributions. While the 96 h BIO2 sample did not show a particularly prominent shift in the histogram peaks towards the left as compared to the BIO3 sample. However, only in the 96 h BIO3 sample, there were about 45,000 sgRNAs that were either depleted or had a very low relative read count distribution as shown in the histograms (Fig. 43).

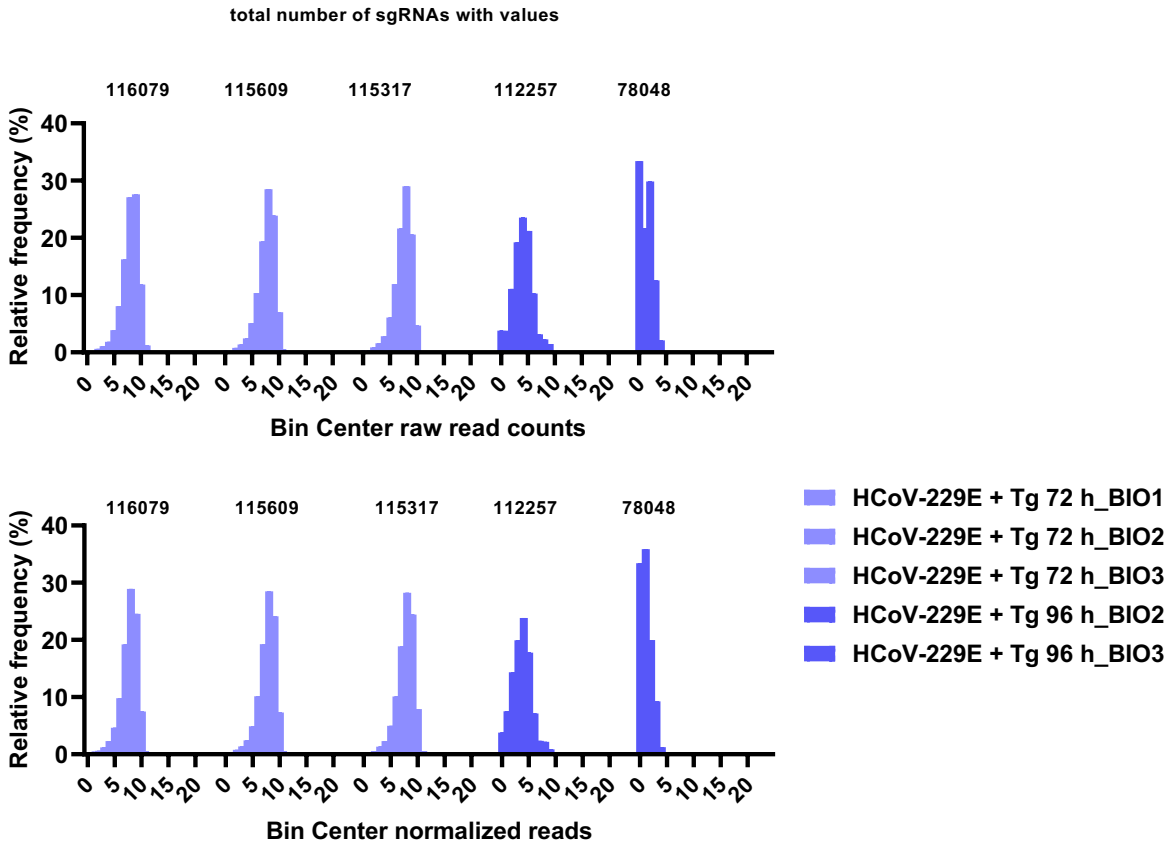


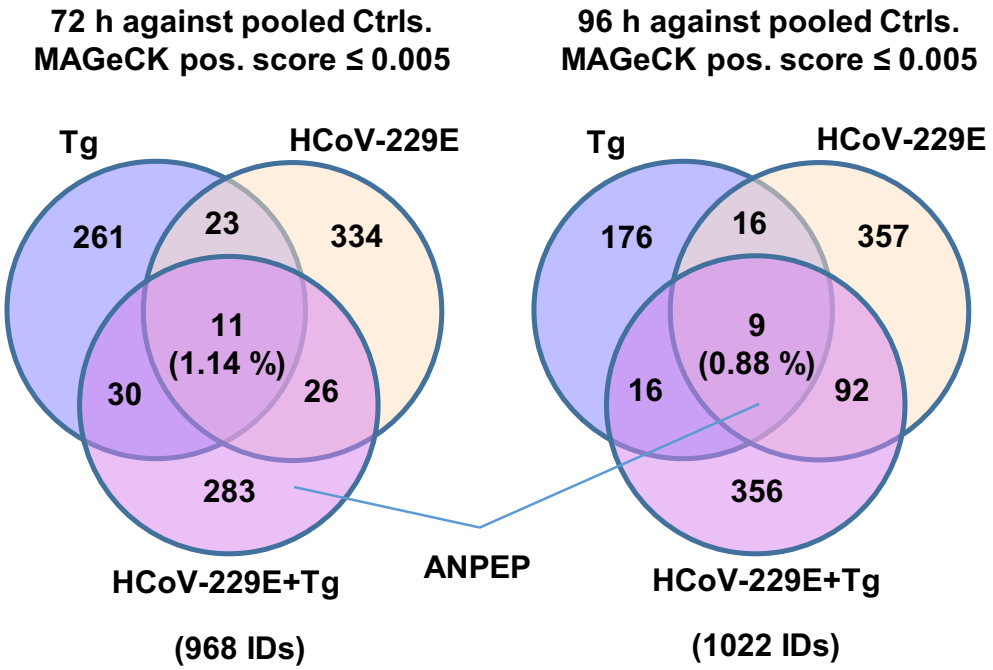
Figure 43: Quality control metrics for the deep sequenced HCoV-229E + Tg selected samples for the 72 h and 96 h time points.
MAGeCK tool was used to determine the read counts in the HCoV-229E + Tg treated and selected samples for 72 h and 96 h. The read counts from all the samples were $\log_2$ transformed and median normalized and quality control metrics were calculated to determine the read count distribution in all the samples. The histogram depicts the frequency distribution of read counts in the samples before (upper) and after (below) median normalization with the values counted for each sample analyzed in three biologically independent experiments. The figures were generated and provided by Prof. Kracht.

5.5.5 Summarizing identification and comparison of all hits from the 72 h and 96 h screen selections

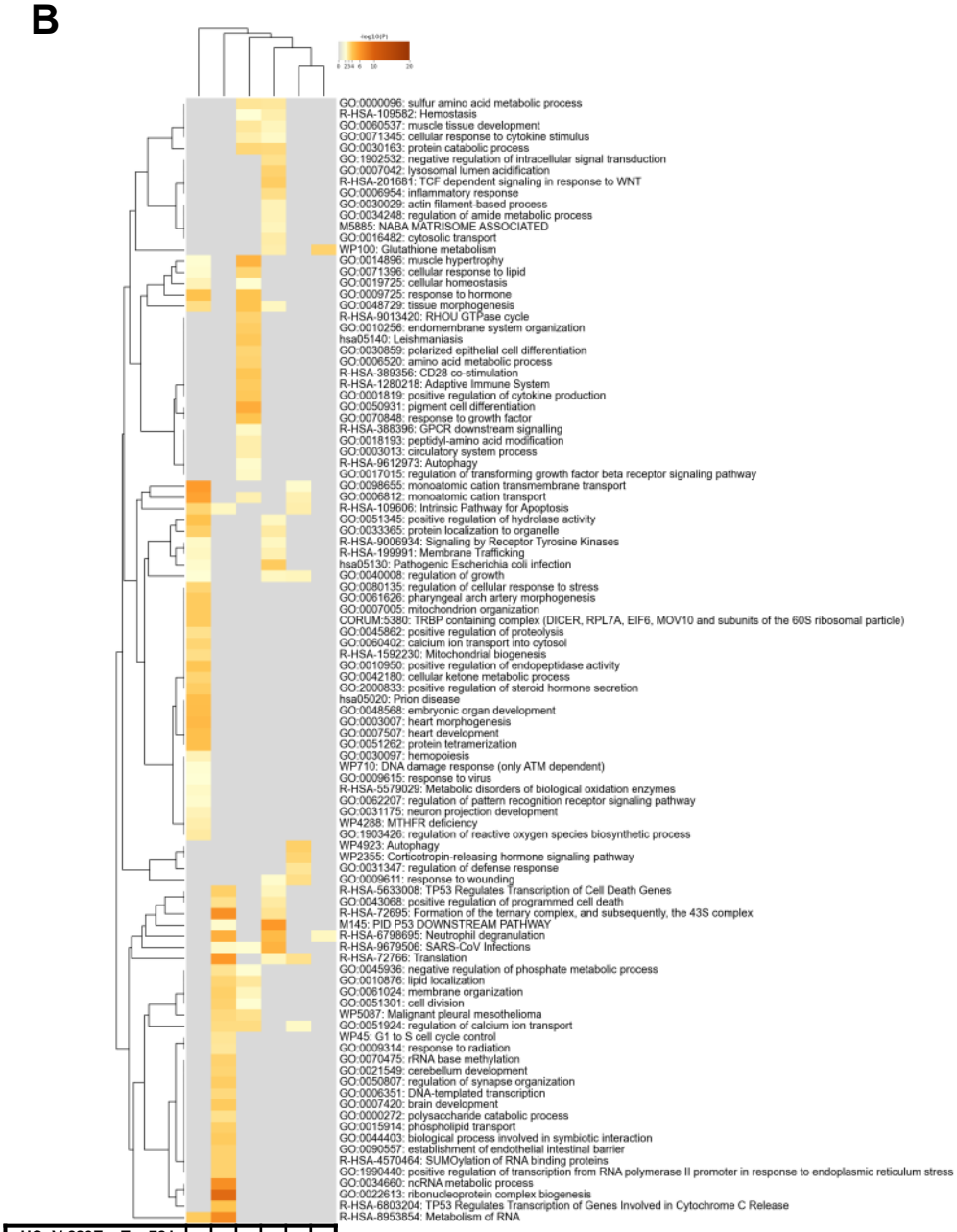
Because there were little differences in the overall distributions of the integrated sgRNAs in the triplicate 72 h and 96 h control sample sets, these six data sets were combined, and used as pooled controls for all further pairwise comparisons to infected or Tg-treated conditions. Based on the MAGeCK analysis of available biological replicates, in total 968 genes were filtered out with a MAGeCK score <= 0.005 in the 72 h screen for all the conditions. Specifically, sgRNAs for 325 (Tg), 394 (HCoV-229E), or 350 (HCoV-229E + Tg) genes were enriched across all conditions. Next, the overlap of these gene sets was evaluated by Venn diagrams and it was found that, interestingly, only 11 (1.14%) out of 968 genes were shared among each other (Fig. 44A). Using the same procedures, in the case of the 96 h screen selection there were 1022 genes filtered out with a MAGeCK score <=0.005 across all conditions. Again, there was only a small overlap of 9 (0.88%) out of 1022 genes found to be shared among all three conditions (Fig. 44A). In conclusion, while sgRNAs for hundreds of genes were significantly enriched in both 72 h and 96 h screen selections, there was very little overlap of these hits across all conditions, suggesting that Tg, HCoV-229E, and Tg + HCoV-229E largely affect different sets of genes required for the short or long-term survival of the cells.

All genes from the six groups were further examined for biological relevance (in terms of pathway representation) by overrepresentation analysis using the Metascape software. The clustered heatmap showed the top 100 most significantly enriched pathways for the genes belonging to the six groups (Fig. 44B). Altogether the pathway analyses of the significantly enriched sgRNAs showed multiple, often not very strongly enriched pathway categories in line with Tg, HCoV-229E and both conditions affecting many cellular processes (Fig. 44B). From the Metascape analysis, the most significantly enriched pathway terms from the individual groups were intersected using upset plots. There was also relatively little overlap between these pathway terms, as at most, three terms were shared between any of the conditions as revealed by upset plot analysis of all pathway terms (Fig. 44C).

A


72 h against pooled Ctrls.
MAGeCK pos. score ≤ 0.005
Tg
HCoV-229E
261
23
334
11
(1.14 %)
30
26
283
HCoV-229E+Tg
(968 IDs)
96 h against pooled Ctrls.
MAGeCK pos. score ≤ 0.005
Tg
HCoV-229E
176
16
357
9
(0.88 %)
16
92
356
HCoV-229E+Tg
(1022 IDs)
ANPEP

B



HCoV-229E + Tg_72 h	+					
HCoV-229E_72 h		+				
HCoV-229E_96 h			+			
HCoV-229E + Tg_96 h				+		
Tg_72 h					+	
Tg_96 h						+

C

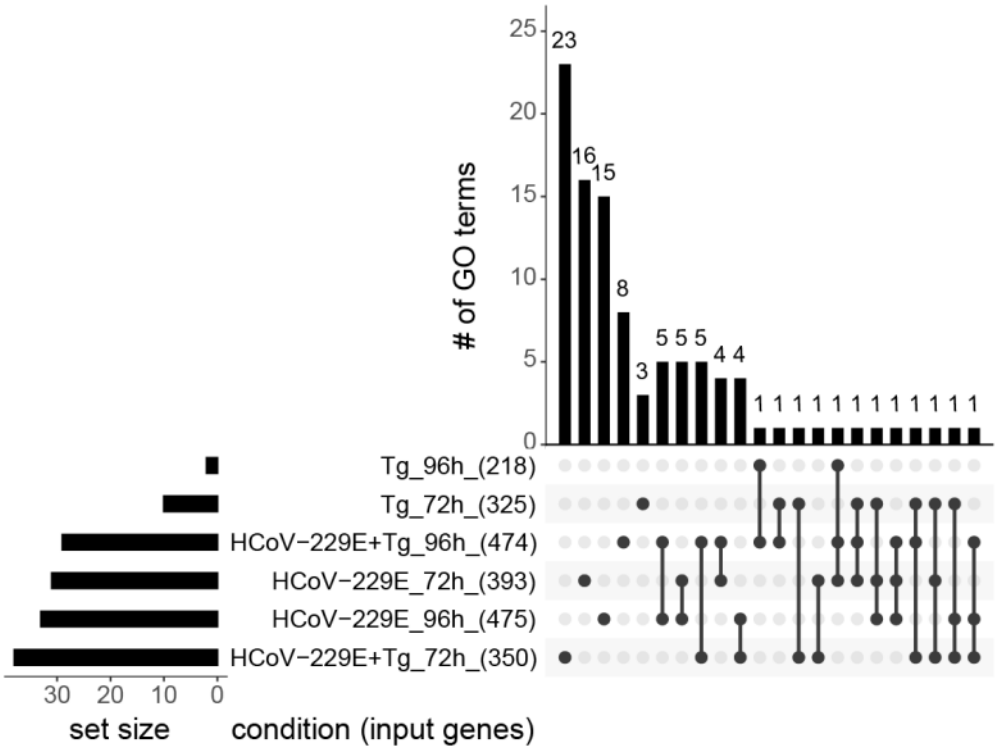


Figure 44: Gene-based Venn diagram comparison of all significant hits emerging from the sgRNA screens according to the MAGeCK analysis.
(A) Genes were sorted out from all the genetic screens with a MAGeCK score <= 0.005. The filtered gene sets were uploaded onto an interactive online Venn diagram creator to generate a Venn diagram comparison. **(B)** All genes of the six groups of the Venn diagrams were examined for biological functions by overrepresentation analysis using the software Metascape [104]. The clustered heatmap shows the top 100 most significantly enriched pathway terms. **(C)** Overlap analysis of shared pathway terms by Upset diagram (plotted with https://www.bioinformatics.com). The top 100 pathway terms as shown in **(B)** per condition used as input. The figures were generated and provided by Prof. Kracht. Connected black dots highlight shared pathways.

5.5.6 Intersection of all sgRNA screen hits with proteomics data sets

To move forward with the selection of the most interesting hits from the genome-wide sgRNA screen for further functional validation, a decision was made to filter the gene sets from both 72 h and 96 h experimental data sets with proteomics data available in the Kracht working group. The idea behind this combined analysis was to increase the likelihood of following biologically important genes and to narrow down the number of potential candidates for follow-up analysis. For this purpose, data from nascent proteomics experiments performed by Dr. Johanna Meier-Soelch in the Kracht working group were used. In these experiments, Huh7 cells were treated with Tg (1 µM), infected with HCoV-229E (MOI of 1), or exposed to both infection and Tg treatment for 24 h. Then, 2 hours before harvesting, the cells received 30 µM O-propargyl-puromycin (OPP) to label nascent polypeptides at the translating ribosomes [121]. The entire set of nascent polypeptides is herein also called translatome. OPP allows the CLICK-chemistry-based covalent linkage of biotin to all nascent polypeptides and their subsequent purification from cell extracts [122]. This approach was interesting because RNA viruses, including coronaviruses, lead to global translational shut-down [123] and it was reasoned that proteins that were still produced against this phenomenon would be probably important for the viral life cycle (or part of the cell-intrinsic host defense). In this sense, nascent proteomics would be better suited to discover regulated host factors compared to “normal” proteomics which assesses steady-state protein levels. To test this hypothesis, both proteomics and nascent proteomic data sets were obtained from the same set of two independent experiments that were subject to LC-MS/MS analysis in three technical replicates. The aggregate results from the combined analyses of these data sets by Prof. Michael Kracht was shown in Fig. 45.

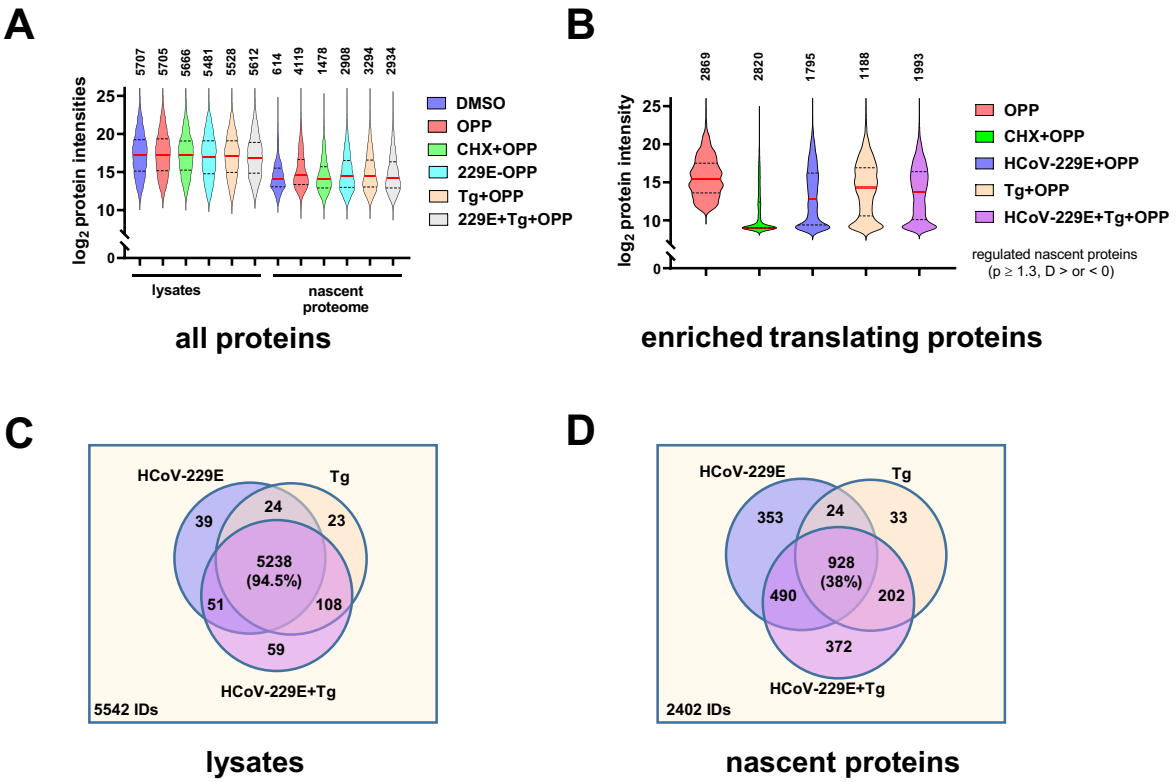


Figure 45: Overview of proteomics and nascent proteomics data sets used for follow-up data analyses.

Huh7 cells were left untreated, infected with HCoV-229E (MOI of 1), treated with Tg (1 µM), or with a combination of both. After 22 h, cells received 30 µM O-propargyl-puromycin (OPP) as indicated for an additional 2 h. Additionally, two samples received solvent only (0.3 % DMSO for 2 h) or 10 µg/ml of the translational inhibitor cycloheximide (CHX) for 21 h. Thereafter, all samples were harvested after 24 h, protein extracts were prepared, and divided, and aliquots were subjected to CLICK-chemistry-based biotinylation. Biotinylated nascent polypeptides were purified from labeled extracts by streptavidin beads. Both, total unmodified protein extracts and purified OPP-labelled nascent polypeptides were identified and quantified by label-free LC-MS/MS from two biological replicates and three technical replicates. **A**) Violin plots showing the raw data as distribution of $\log_2$ transformed means of protein intensities and total numbers of identified proteins in all samples. Red lines show medians and dashed lines show 1st and 3rd quartiles. **B**) To derive nascent polypeptides from these data, the basic set of actively translating proteins (red violin) in untreated cells was identified by the enrichment of protein intensities in the OPP only samples compared to DMSO solvent controls (fold change OPP / DMSO > 0, $-\log_{10}$ p value ≥ 1.3). For the calculation of ratio values, missing data were imputed by a $\log_2$ value of 9, which was below the lowest value measured in the entire data set. For all remaining samples (CHX, Tg + OPP, HCoV-229E + OPP, and HCoV-229E + Tg + OPP), the changes of translatomes were calculated by dividing their intensities by the OPP only values (fold change "sample" / OPP only > 0, $-\log_{10}$ p value ≥ 1.3). The Violin plots show $\log_2$ transformed means of intensities and total numbers of identified nascent polypeptides across all samples. Red lines show medians and dashed lines show 1st and 3rd quartiles. Of note is the strong reduction in protein intensities in the CHX control demonstrating the specificity of this approach. **C**) Venn diagram showing the overlap of identified "steady state" proteins (IDs) in cell extracts. **D**) Venn diagram showing overlapping and distinct protein sets of the nascent proteins in infected and Tg-treated conditions. Dr. Johanna Meier-Soelch performed the experiments and raw data analyses (in Perseus software). Dr. Uwe Linne performed LC-MS/MS identification at the proteomics core facility, at the University of Marburg. Prof. Michael Kracht performed data analysis and graphical visualization.

Out of 968 genes (72 h) and 1022 (96 h) genes detected as hits in the sgRNA screens, 302 and 257 were represented as protein IDs in the proteomics data sets, respectively. For the integrative visualization of key parameters of the intersected genetic and proteomic data sets, multiple variable plots were chosen to visualize sgRNA enrichment, MAGeCK score significance, and the fold changes of either steady-state protein or protein synthesis levels. In the first step, multiple variable plots were generated for the intersected gene sets from the 72 h screen selection and compared with the protein levels and protein synthesis rates of HCoV-229E and / or Tg conditions of the proteomic data. The 302 protein IDs were first compared for their protein levels and synthesis rates in the HCoV-229E + OPP / OPP condition. As indicated in Fig. 46A, the protein intensity levels for the 302 IDs in the infected conditions remained largely at a steady state and the significant sgRNAs targeting these genes were not strongly enriched in the 72 h HCoV-229E selection. In the case of active protein synthesis during the infection, the majority of these IDs were shifted to the left in the multi-variable plot (Fig. 46B). This indicated that there was a substantial translational shutdown of these protein IDs during HCoV-229E infection (Fig. 46B), which did, however, only mildly affect the total levels of these proteins (Fig. 46A), indicating that many of these proteins are relatively stable.

The same scenario was observed for protein intensity levels when comparing HCoV-229E + Tg with HCoV-229E. As indicated in Fig. 46C, the protein intensity levels for the 302 IDs in the infected and Tg-treated conditions showed only small changes at steady state, and the significant sgRNAs targeting these genes were not strongly enriched in the 72 h HCoV-229E + Tg selection. Interestingly, unlike in the comparison described above, the protein synthesis levels for these IDs were equally spread to left and right (log_2 FC in protein synthesis) in the multi-variable plot (Fig. 46D). This indicated that Tg possibly regulated the protein synthesis rate of these IDs in the infected conditions.

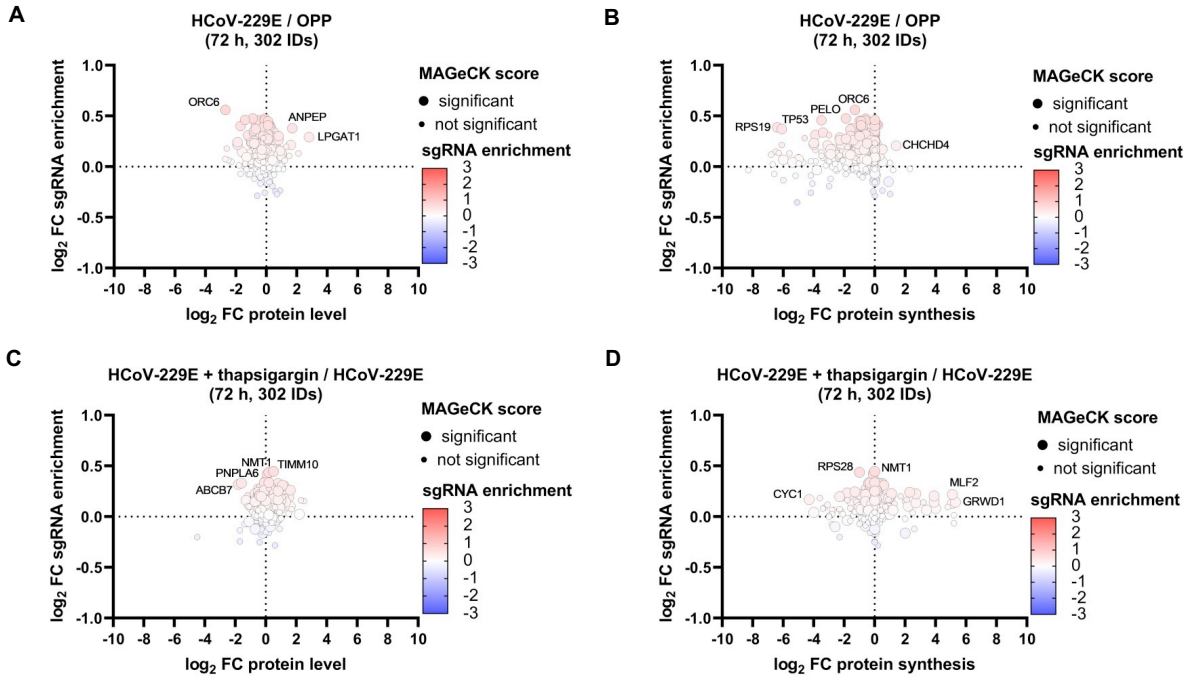


Figure 46: Multi-variable plot showing the intersection of hits from the 72 h sgRNA screen along with their changes at the "steady state" protein and protein synthesis levels in the nascent proteomics data.

The sgRNA enrichment values (compared to the pooled control samples) for 302 protein IDs that overlapped between sgRNA and proteomics screens were plotted against changes in normalized protein intensities **(A, C)** or protein synthesis rates **(B, D)**. **(A, B)** show changes of proteins comparing HCoV-229E-infected (+ OPP) with the uninfected (+ OPP) samples. **(C, D)** show changes of OPP-labelled, nascent proteins comparing HCoV-229E + Tg (+ OPP) with HCoV-229E (+ OPP) samples. The bubble size shows the functional significance of genes from the MAGeCK analysis (based on a MAGeCK score <= 0.005). The bubble color indicates the $\log_2$ FC sgRNA enrichment as shown in the legend. Prof. Michael Kracht performed data analysis and graphical visualization.

Next, multiple variable plots were generated for the hits from 96 h sgRNA selection and compared with the protein levels and synthesis rates of HCoV-229E-infected and / or Tg-treated conditions of the proteomic data. The 256 protein IDs were first compared for their protein levels and synthesis rates in the HCoV-229E-infected (+ OPP) compared with the uninfected (+ OPP) condition. As indicated in Fig. 47A, the protein intensity levels in the infected conditions changed less than twofold for most proteins at steady state, and multiple sgRNAs targeting many of the according genes were significantly enriched in the 96 h HCoV-229E selection. Most notable here was the ANPEP protein, which showed strong sgRNA enrichment and about four-fold enrichment at the protein level (Fig. 47A). In contrast, the synthesis rates of most proteins in this set decreased (Fig. 47B). This was also the case for ANPEP, indicating that the HCoV-229E receptor together with multiple other functionally relevant proteins might be stabilized at the protein level during infection (Fig. 47B).

This situation was different when the sgRNA data sets were plotted against protein levels or synthesis rates that changed between infected plus Tg conditions compared to the infected condition only. This specific comparison was chosen, because it focused on the relative changes induced by Tg in infected cells, independent of the absolute intensity levels determined by mass spectrometry. The intensity levels for many proteins during infection and Tg treatment changed by two to four-fold as indicated in the multi-variable plot (Fig. 47C). In the case of ANPEP, Tg treatment reduced protein intensity level was lower (Fig. 47C), but did not affect the synthesis rate (Fig. 47D). Interestingly, this analysis identified several proteins with moderate changes in steady-state protein levels, but strong increases (ARFGAP3, CLU, and SDE2) or decreases (HMOX2, ERMP1, and EMC6) in protein synthesis rates (Fig. 47D).

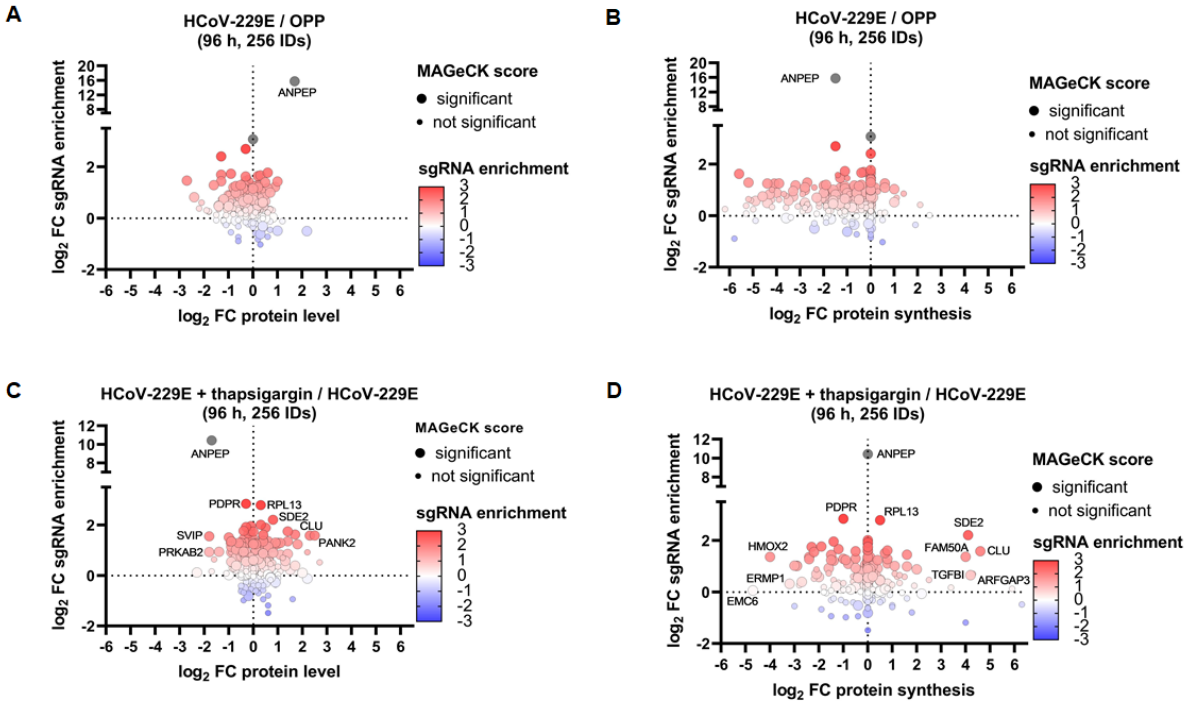


Figure 47: Multi-variable plot showing the intersection of hits from the 96 h sgRNA screen along with their changes at the "steady state" protein and protein synthesis levels in the nascent proteomics data.
The sgRNA enrichment values (compared to the pooled control samples) for 256 protein IDs that overlapped between sgRNA and proteomics screens were plotted against changes in normalized protein intensities (**A**, **C**) or protein synthesis rates (**B**, **D**). (**A**, **B**) show changes of proteins comparing HCoV-229E-infected (+ OPP) with the uninfected (+ OPP) samples. (**C**, **D**) show changes of OPP-labelled, nascent proteins comparing HCoV-229E+Tg (+ OPP) with HCoV-229E (+ OPP) samples. The bubble size shows the functional significance of genes from the MAGeCK analysis (based on a MAGeCK score <= 0.005). The bubble color indicates the $\log_2$ FC sgRNA enrichment as shown in the legend. Gray colors indicate $\log_2$ sgRNA enrichment > 3. The figures were generated and provided by Prof. Kracht.

In summary, the visualization of all data by multi-variable plots confirmed that the intersection of hits from the 96 h sgRNA screen with the protein levels / synthesis rates was suited to provide a refined list of Tg-regulated candidate factors, which were characterized by high sgRNA enrichment and differential protein synthesis rates in infected cells. By focusing on those proteins which had a significant change in synthesis rates ($-\log_{10}$ p value HCoV-229E + Tg OPP / HCoV-229E + OPP), a final list of 40 factors was selected for further analysis. Table 6 summarizes the mean intensity levels of purified nascent polypeptides and their changes from basal levels in Tg-treated, HCoV-229E-infected, or Tg-treated and HCoV-229E-infected conditions. Based on the results the top 9 up or downregulated factors (*DDRGK1*, *ARFGAP3*, *CLU*, *HMOX2*, *ERMP1*, *SDE2*, *TGFB1*, *FAM50A*, and *EMC6*) were chosen for follow-up validation. Additionally, *SEC24A* was included in this selection, because it was previously reported to mediate Tg effects on cell death [124]. The synthesis rates of all proteins encoded by these genes were highly deregulated in HCoV-229E + Tg conditions as compared to the HCoV-229E infection alone as shown in the differential $\log_2$ fold change ratios in Table 6.

In the following section, individual knock out cell lines were generated for the above-mentioned 10 host factors and in addition to ANPEP (as a control) to study the functional relevance of these genes in the Tg-mediated anti-viral effect.

-	-	CHX	Tg	HCoV-229E	HCoV-229E+Tg	HCoV-229E+Tg / HCoV-229E		Gene name
DMSO	OPP	OPP	OPP	OPP	OPP	OPP		
9.0	10.3	9.0	16.7	10.1	16.5	5.0	6.3	DDRGK1
9.0	12.8	9.0	16.5	10.0	15.9	5.3	5.9	ARFGAP3
9.0	18.5	13.5	17.1	12.2	16.8	2.9	4.6	CLU
9.0	14.2	9.0	14.0	9.0	13.3	7.6	4.3	RNF114
9.0	14.3	9.0	13.2	9.0	13.2	8.1	4.2	TGFBI
9.0	12.6	9.0	14.3	9.0	13.1	3.1	4.1	SDE2
9.0	14.8	9.0	14.6	9.0	13.0	3.2	4.0	WDR70
9.0	14.0	9.0	13.5	9.0	13.0	6.3	4.0	FAM50A
9.0	13.9	9.0	13.6	9.0	12.4	3.2	3.4	DCAF4L2
9.0	13.3	9.0	13.4	9.0	11.5	2.0	2.5	PDE3A
9.0	14.9	9.0	14.8	12.0	14.3	1.3	2.2	SEC24A
9.0	18.1	9.0	16.4	14.0	15.5	1.6	1.6	STAU1
9.0	17.0	11.1	16.5	15.1	16.2	1.8	1.1	DIS3
9.0	17.2	11.7	16.0	14.4	15.4	1.7	1.0	YME1L1
15.4	20.2	18.0	20.6	19.4	20.2	2.0	0.8	AARS
16.5	21.7	19.5	21.5	20.8	21.3	1.5	0.5	RPL13
9.0	18.1	13.4	18.2	17.3	17.8	1.5	0.5	RPS27L
9.0	18.8	13.7	18.7	18.6	18.2	1.7	-0.4	SCFD1
14.7	20.2	17.1	19.3	19.4	18.8	2.3	-0.6	RBBP7
9.0	13.7	9.0	13.7	13.5	12.8	1.6	-0.7	NIF3L1
11.1	18.2	14.1	17.7	17.9	17.0	2.7	-0.9	TST
9.8	15.9	11.0	15.7	16.3	14.9	3.6	-1.4	ALDH4A1
16.6	18.8	17.6	18.8	19.8	18.3	3.5	-1.4	ATP5H
9.0	18.3	12.8	17.1	17.7	16.2	1.6	-1.4	MATR3
9.0	11.3	9.0	10.5	10.5	9.0	1.7	-1.5	TAOK1
9.0	17.4	12.6	15.9	16.9	15.3	2.2	-1.5	ATL3
9.0	15.1	10.1	14.3	15.1	13.6	4.1	-1.6	CHPT1
9.0	15.7	9.0	15.4	15.8	14.1	2.4	-1.7	ME2
9.0	17.2	11.3	15.4	16.5	14.7	2.4	-1.8	GGCX
9.0	17.1	12.0	16.8	18.0	16.1	4.9	-1.9	ZMPSTE24
9.0	14.6	9.0	13.4	14.2	12.1	1.9	-2.1	RAB27A
9.0	14.3	9.0	13.1	13.6	11.4	1.7	-2.2	SVIP
9.0	11.8	9.4	10.0	11.7	9.3	1.6	-2.4	CRK
9.0	13.8	9.0	10.8	11.9	9.0	2.8	-2.9	DHFR
9.0	15.1	9.0	13.9	13.7	10.8	2.2	-2.9	ARPC1A
9.0	13.8	9.0	11.4	13.5	10.6	1.6	-3.0	PECR
9.0	14.6	9.0	12.7	13.9	10.9	1.9	-3.0	RTN3
9.0	11.3	9.0	10.9	12.2	9.0	2.9	-3.2	ERMP1
9.0	14.1	9.0	13.5	14.6	10.6	3.4	-4.0	HMOX2
9.0	11.2	9.0	10.7	13.7	9.0	3.2	-4.7	EMC6

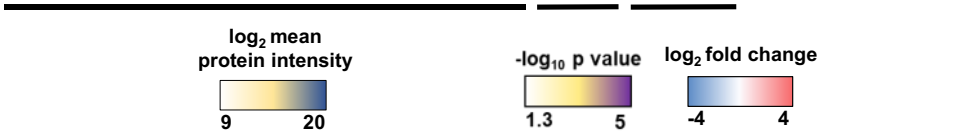


Table 6: Protein synthesis rates for the top 40 candidates that mediate Tg antiviral effects as emerged from intersecting sgRNA screens with proteomics data.

The table shows the $\log_2$ mean protein intensity values of OPP-labelled nascent polypeptides after purification (lanes 1 – 6). The relative changes and their significance ($-\log_{10}$ p value) in the comparison of HCoV-229E + Tg + OPP with HCoV-229E + OPP are shown in lanes 7 to 8, respectively. As an indication of the quality and specificity of detection of nascent polypeptides by this method, almost all intensity values are reduced in samples with CHX treatment (due to inhibition of translation) and are very low or absent in DMSO controls, in which cells did not receive OPP to label nascent polypeptides. Red arrows mark candidate IDs selected for further validation. The table was generated and provided by Prof. Kracht.

5.6 Validation of the selected host factors for Tg-mediated anti-viral effect

To validate the functional relevance of the candidate host factors on Tg-mediated inhibition of HCoV-229E viral replication and virus-induced cell death, knock out cell lines for the selected host factors were tested for viral resistance (plaque assays) and cell survival upon 96 h of infection (cell viability by MTS). Each sgRNA targeting the 10 host factors was selected based on the results of the sgRNA enrichment analysis and was cloned into a lentiCRIPSR V2 vector that enables the stable expression of Cas9 and the target sgRNA when transduced in Huh7 cells, resulting in either deletions or frame-shift mutations for protein loss of function.

5.6.1 Effect of ARFGAP3 and EMC6 knock out on cell viability and viral replication in the presence or absence of Thapsigargin

In the first step, the cell lines carrying sgRNAs targeting ARFGAP3 and EMC6 were tested for the loss of protein levels of ARFGAP3 and EMC6 respectively. In brief, sgRNA-carrying cell lines along with Huh7 parental and empty vector were infected with HCoV-229E at an MOI 1 for 24 h or left untreated in the presence / absence of Tg at 1 µM concentration. The cell lysates were prepared from all the conditions and the protein levels of ARFGAP3, EMC6, and viral N protein were visualized by Western blotting.

Huh7 parental and empty vector cell lines infected and treated with Tg or only Tg alone showed an increased level of ARFGAP3 as compared to the untreated conditions (Fig. 48A). The relative quantification of ARFGAP3 protein levels from the immunoblots showed a four-fold increase in the Tg and HCoV-229E + Tg conditions (Fig. 48B). However, in the ARFGAP3_sg1 and sg2 cell lines, the protein levels were greatly reduced in all the conditions (Fig. 48A). The ARFGAP3 protein levels in Tg conditions was significantly reduced in both of the sgRNA carrying cell lines (Fig. 48B). Interestingly, the ARFGAP3_sg2 cell line showed a slightly increased protein level of viral N protein (Fig. 48B).

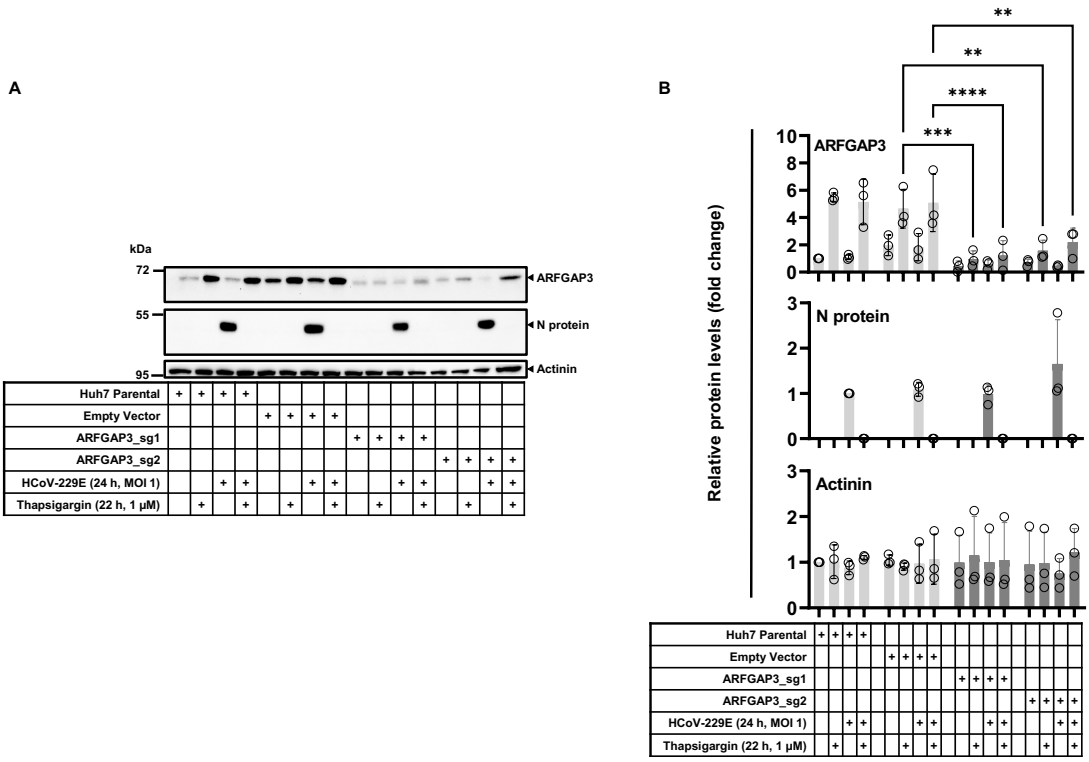


Figure 48: Validation of Tg-mediated regulation of ARFGAP3, ARFGAP3 knock out, and its effect on Tg-mediated HCoV-229E viral replication.

Huh7 parental, empty vector, and sgRNA-carrying ARFGAP3 cell lines were infected with HCoV-229E at an MOI 1 for 24 h or left untreated in the presence or absence of 1 µM Thapsigargin. The cell lysates were prepared and subjected to Western blotting using the indicated antibodies. **A)** Representative immunoblots showing the protein levels of ARFGAP3, viral N protein, and ß-Actin from three biologically independent experiments. **B)** The immunoblots were quantified for the respective protein levels using Image Lab software. The raw values were background subtracted and normalized to the untreated Huh7 parental control. Graphs show relative quantification of ARFGAP3, viral N protein, and ß-Actin levels. Bar graphs represented here were the mean ± SD measured from three biologically independent experiments and asterisks indicate p values (**$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$) obtained from ordinary two-way ANOVA analysis.

In the case of EMC6, the infected empty vector cell line in the presence or absence of Tg showed reduced EMC6 protein levels as compared to the untreated controls (Fig. 49A). The relative protein levels of EMC6 were significantly reduced in all of the conditions for EMC6_sg2 cell line (Fig. 49B). However, the viral N protein level was slightly reduced in the EMC6_sg2 cell line but this was not significant when compared to the infected empty vector control (Fig. 49A and B).

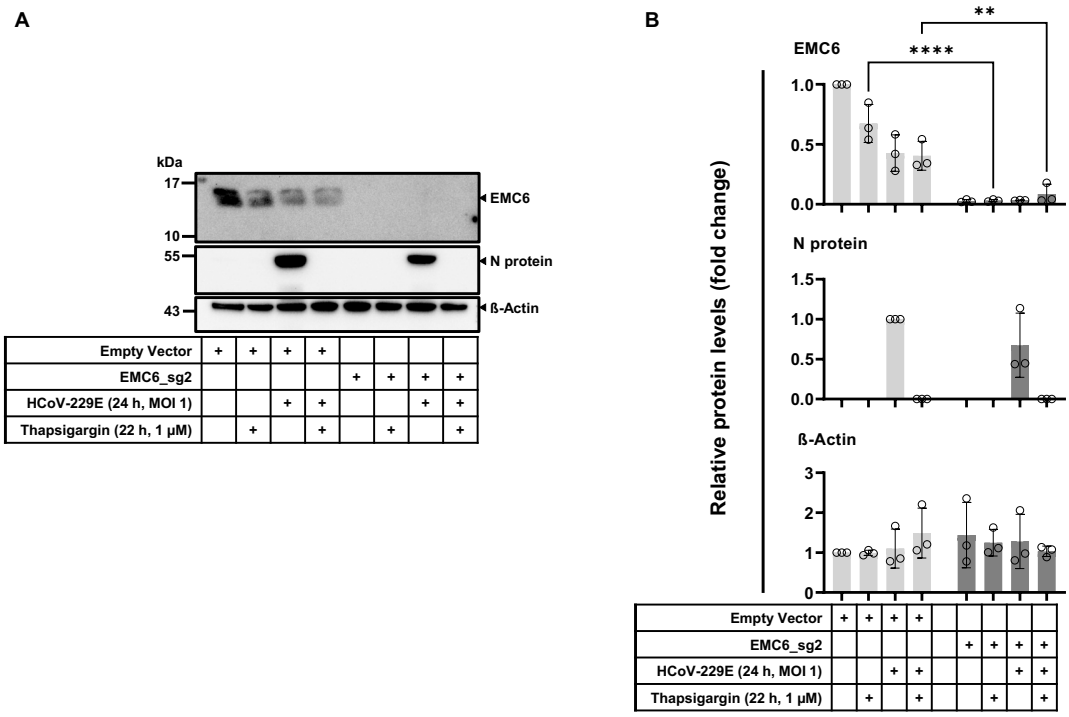
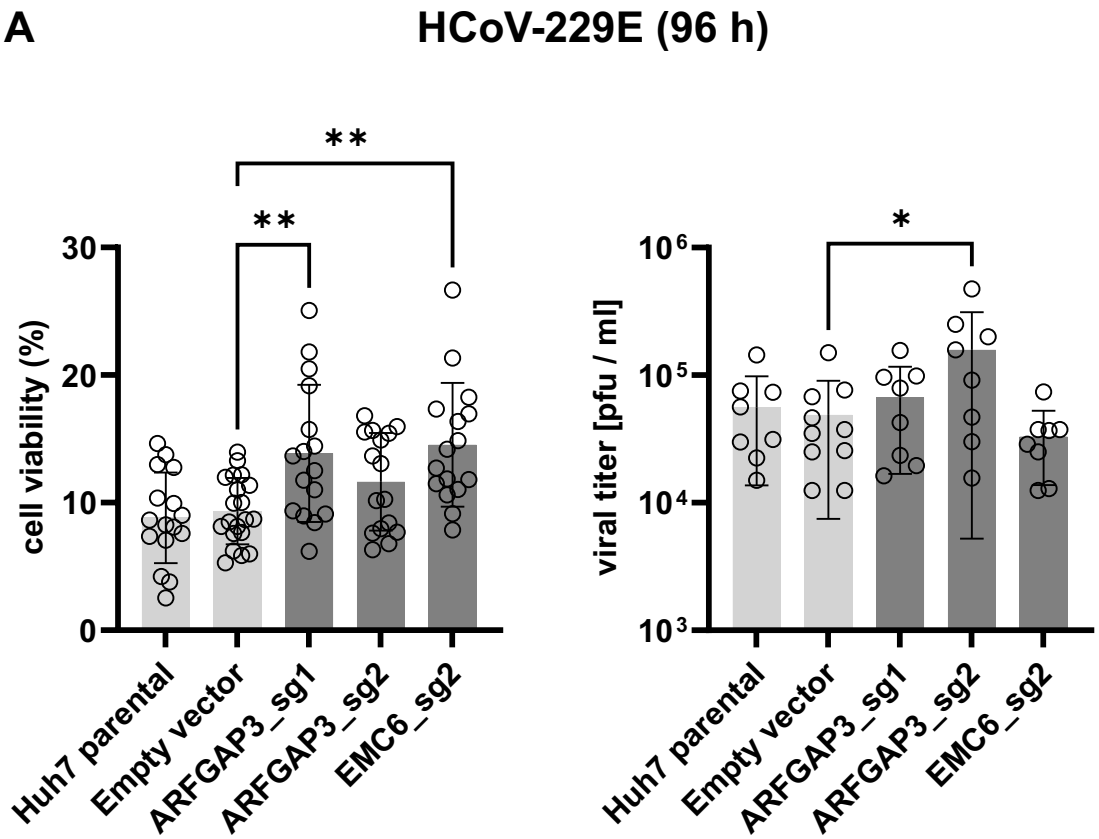


Figure 49: Validation of EMC6 knock out and its effect on the HCoV-229E viral replication.
Empty vector and sgRNA-carrying EMC6 cell lines were infected with HCoV-229E at an MOI 1 for 24 h or left untreated in the presence or absence of 1 µM Thapsigargin. The cell lysates were prepared and subjected to Western blotting for the indicated antibodies. The immunoblots were quantified for the respective protein levels using Image Lab software. The raw values were background subtracted and normalized to the untreated Huh7 parental control. **A)** Representative immunoblots showing the protein levels of EMC6, viral N protein, and ß-Actin from three biologically independent experiments. **B)** Relative quantification of EMC6, viral N protein, and ß-Actin levels. Bar graphs represented here were the mean ± SD measured from three biologically independent experiments and asterisks indicate p values (**$p \leq 0.01$, ****$p \leq 0.0001$) obtained from ordinary two-way ANOVA analysis.

To validate the effect of ARFGAP3 and EMC6 knock out on the cell viability upon HCoV-229E infection and / or Tg treatment, an MTS assay was performed. In brief, Huh7 parental, empty vector, and sgRNA-carrying cell lines were seeded onto a 96-well plate at a density of 5000 cells / well in quadruplicates and were infected and / or treated with Tg for 96 h. The cell viability was measured by MTS assay. ARFGAP3_sg1 and EMC6_sg2 knockdown cells showed increased cell viability after 96 h of HCoV-229E infection compared to the cell viability of the infected empty vector (Fig. 50A). Interestingly, both the infected ARFGAP3 sgRNA cell lines showed a significant increase in the cell viability upon Tg treatment (Fig. 50B). However, the infected EMC6_sg2 cell line showed no improvement in the cell viability upon Tg treatment (Fig. 50B).

In a similar experiment, the cell supernatants were collected from the infected and / or Tg-treated cell lines and the viral titers were measured by plaque assay. Here, the ARFGAP3_sg2 cell line showed an increased viral titer as compared to the other infected cell lines (Fig. 50A). The viral titer of the infected EMC6_sg2 cell line was slightly reduced, but not significant when compared to the infected empty vector control (Fig. 50A). Thapsigargin treatment completely reduced the viral titers in all the infected cell lines (Fig. 50B).



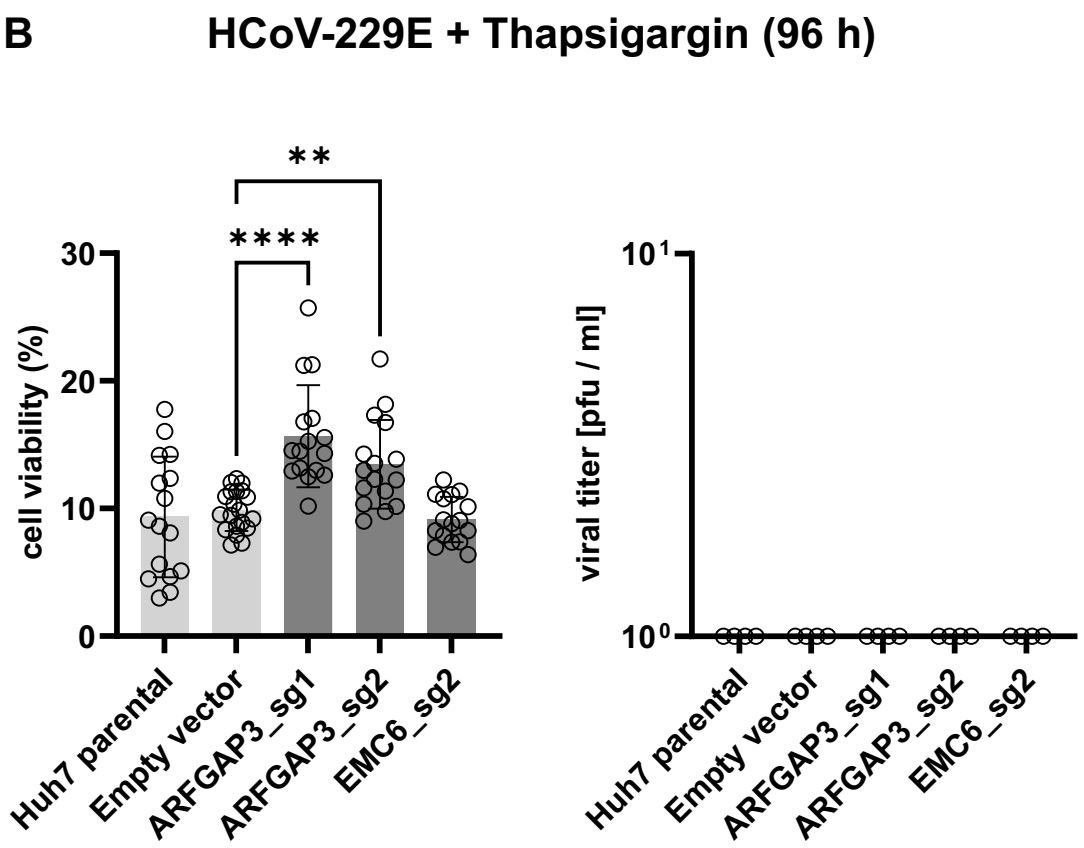
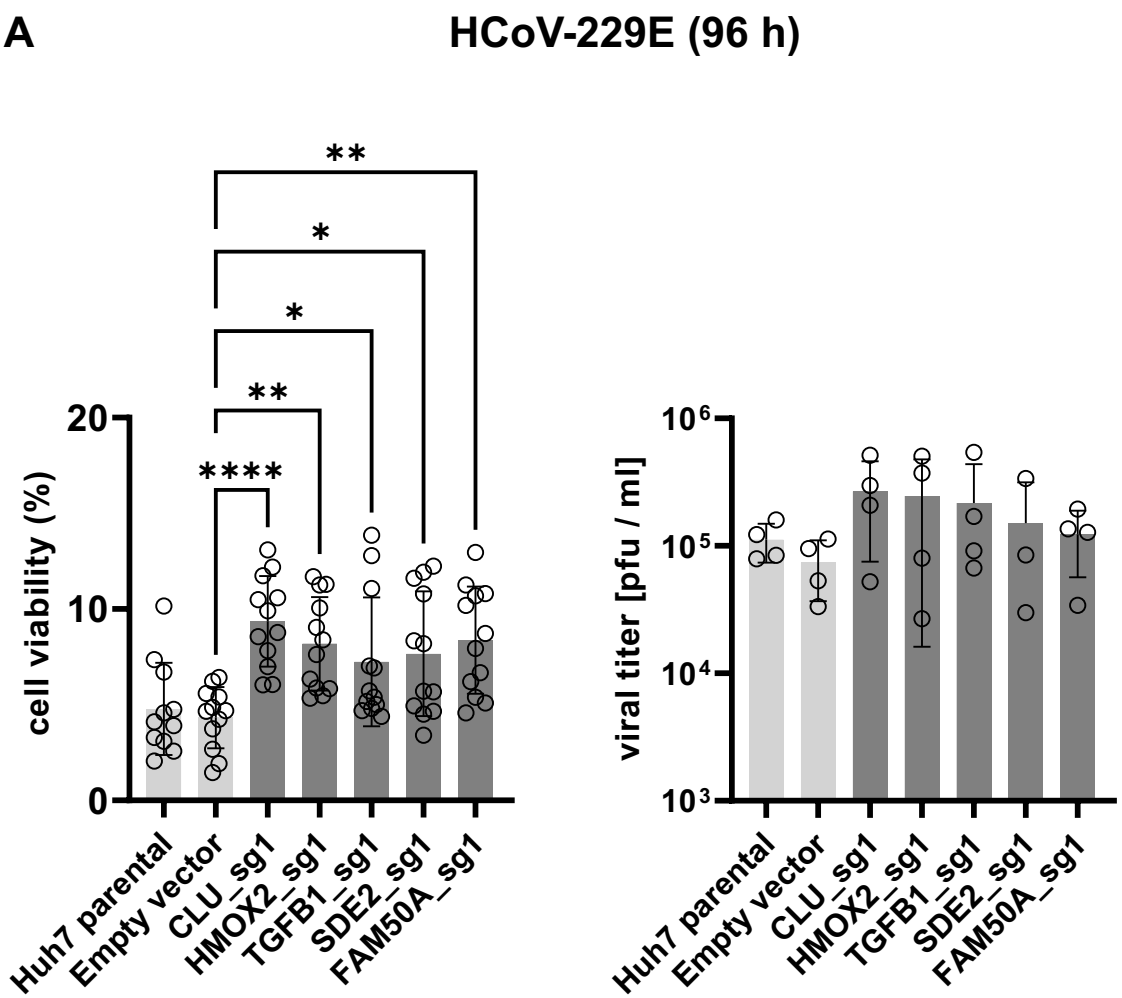


Figure 50: Effect of ARFGAP3 and EMC6 knock out on the cell viability and viral replication upon infection and / or Tg treatment.

Huh7 parental, empty vector, and sgRNA-carrying cell lines were either infected only **(A)** and / or treated with Tg for 96 h **(B)**. Cell viability was measured by MTS assay (left graphs) and viral titers were measured by plaque assays (right graphs). Bar graphs show mean values ± SD and individual values from four biologically independent experiments. MTS assays also include all points from technical replicates. Asterisks show the significance of changes (*$p \leq 0.05$, **$p \leq 0.01$, ****$p \leq 0.0001$) obtained from ordinary one-way ANOVA analysis. Dr. Christin Mayr-Buro kindly performed one or more biological replicates of the MTS assay.

5.6.2 Effect of the host factors CLU, HMOX2, TGFB1, SDE2, and FAM50A on cell viability and viral replication in the presence or absence of thapsigargin.

In a similar experimental setup as described in section 5.6.1, the cell viability and viral titers were measured in the infected and / or Tg-treated CLU, HMOX2, TGFB1, SDE2, and FAM50A cell lines. The sgRNAs targeting DDRGK1 and ERMP1 had a slower growth rate (compared to the parental control and other host factor knock outs) in the cell culture and were decided to exclude from the experimental setup. The cell viability of sgRNA carrying CLU, HMOX2, TGFB1, SDE2, and FAM50A cell lines was significantly enhanced upon 96 h of HCoV-229E as compared to the empty vector (Fig. 51A). However, only infected CLU_sg1 cell line showed a significant increase in the cell viability under Tg treatment (Fig. 51B). The viral titers of the infected knock out cell lines were not affected upon 96 h HCoV-229E infection (Fig. 51A). Thapsigargin treatment completely reduced the viral titers in all the infected cell lines (Fig. 51B).



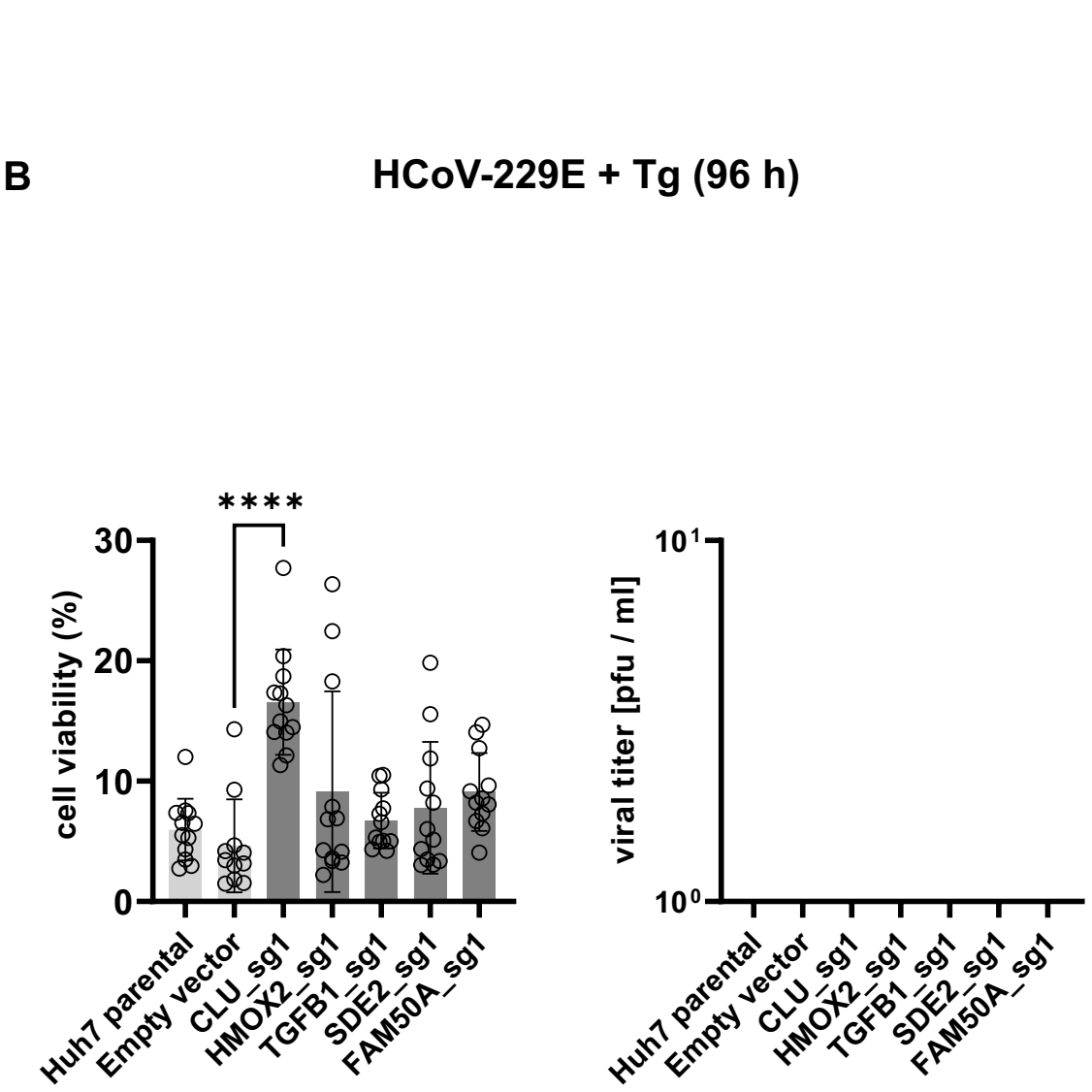
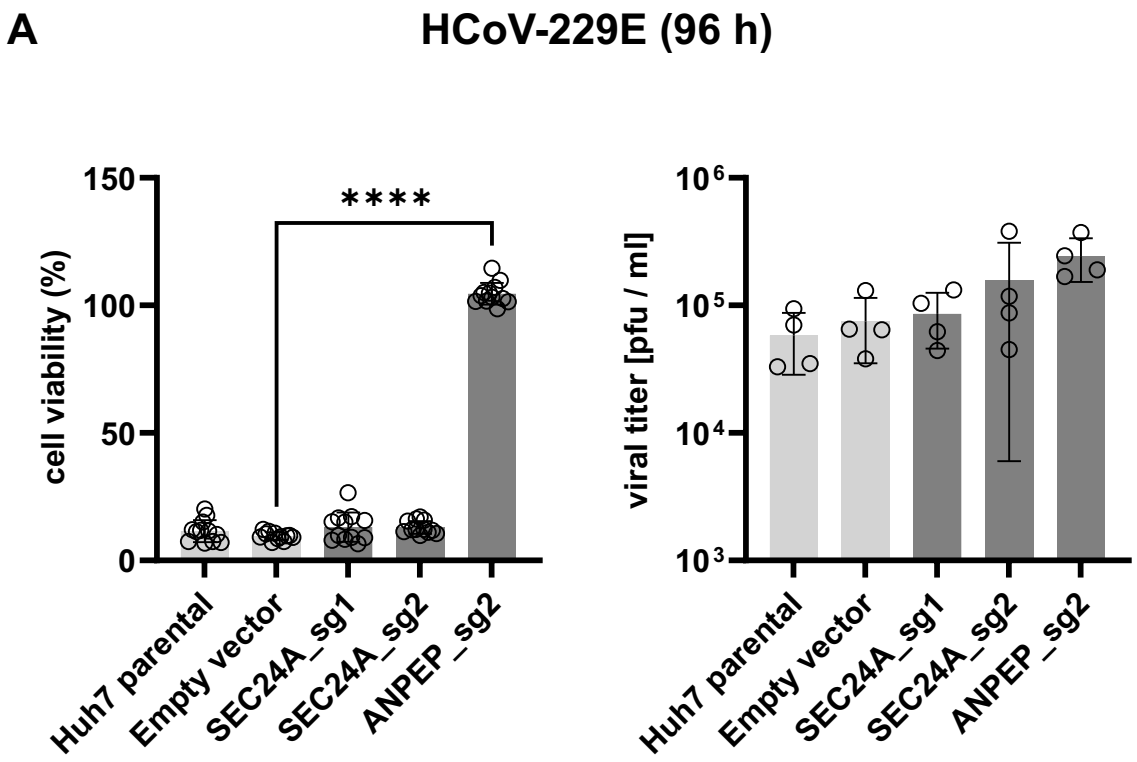


Figure 51: Effect of CLU, HMOX2, TGFB1, SDE2, and FAM50A knock out on the cell viability and viral replication upon infection and / or Tg treatment.
Huh7 parental, empty vector, and sgRNA-carrying cell lines were either infected only **(A)** and / or treated with Tg for 96 h **(B)**. Cell viability was measured by MTS assay (left graphs) and viral titers were measured by plaque assays (right graphs). Bar graphs show mean values ± SD and individual values from four biologically independent experiments. MTS assays also include all points from technical replicates. Asterisks show the significance of changes (*$p \leq 0.05$, **$p \leq 0.01$, ****$p \leq 0.0001$) obtained from ordinary one-way ANOVA analysis.

5.6.3 SEC24A reversed the Tg-mediated inhibition of HCoV-229E viral replication

In a similar experimental setup as described in section 5.6.1, the cell viability and viral titers were measured in the infected and / or Tg-treated SEC24A and ANPEP cell lines (as a control). The cell viability of sgRNA carrying SEC24A cell lines was not improved upon 96 h of HCoV-229E as compared to the empty vector (Fig. 52A). However, the infected SEC24A_sg1 cell line showed a significant increase in the cell viability under Tg treatment (Fig. 52B). The viral titers of the infected SEC24A cell lines were not affected upon 96 h HCoV-229E infection (Fig. 52A). Interestingly, Tg-mediated inhibition of viral replication was partially reversed in the infected SEC24A_sg1 treated with Tg for 96 h (Fig. 52B).



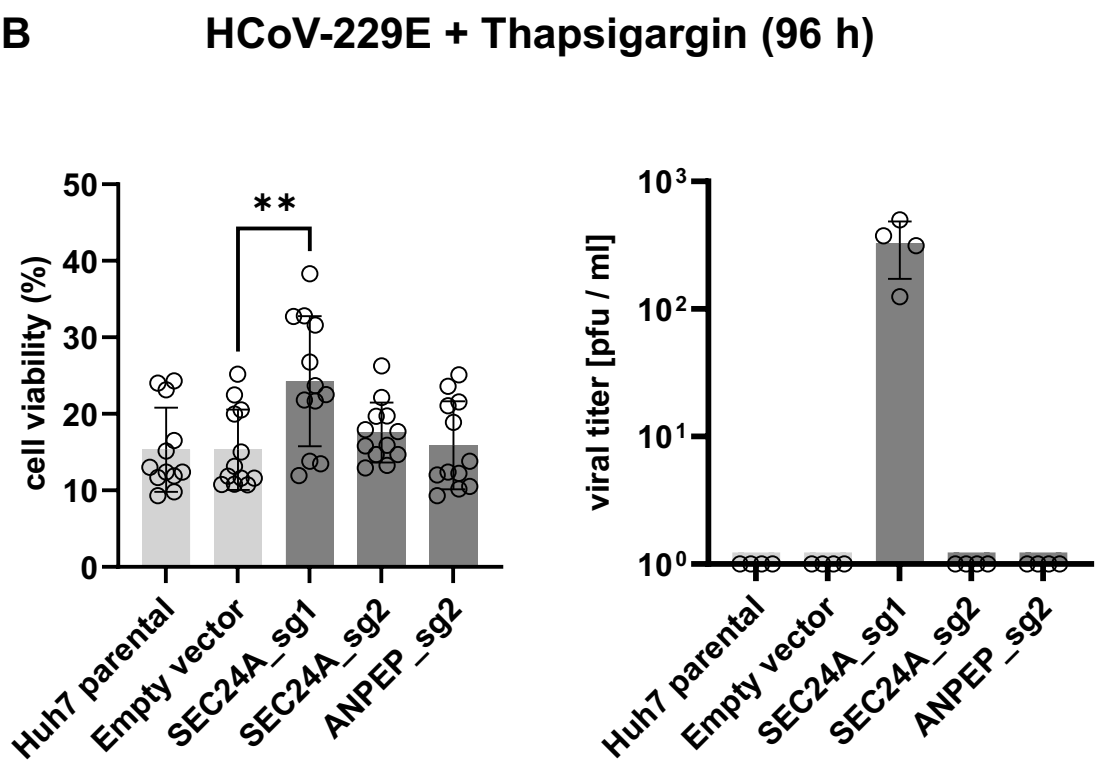


Figure 52: Effect of SEC24A knock out on the cell viability and viral replication upon infection and / or Tg treatment.

Huh7 parental, empty vector, and sgRNA-carrying cell lines were either infected only **(A)** and / or treated with Tg for 96 h **(B)**. Cell viability was measured by MTS assay (left graphs) and viral titers were measured by plaque assays (right graphs). Bar graphs show mean values ± SD and individual values from four biologically independent experiments. MTS assays also include all points from technical replicates. Asterisks show the significance of changes (**$p \leq 0.01$, ****$p \leq 0.0001$) obtained from ordinary one-way ANOVA analysis.

6. Discussion

Coronaviruses, being the virus family with the largest +ssRNA genome encode only a limited number of viral proteins, yet interact with many thousands of host cell factors to establish a complex interaction network to replicate, survive, and evolve [125]. Coronaviruses exploit host cell factors by overturning their functions to benefit the virus in multi-level stages of viral replication including the viral entry, formation of double-membrane vesicles for the viral genome transcription, viral protein translation and processing, and finally, the formation of newly synthesized virions [126-128]. Concurrently, these viruses also reprogram the host factors by modulating (attenuated in most cases by the coronaviruses) the immune response and metabolic pathways to support viral replication. Thus, in the major stages of viral replication, a complex network of host cell factors is involved compared to the viral proteins [20, 21]. It is therefore an emerging and crucial necessity to identify and understand the vital host factors, as they play a key determinant role in pathogenesis as well as viral evolution. However, it has been always a daunting task to identify the many thousands of host factors that are subverted by coronaviruses [20]. The rapid development in the functional genomics approaches has paved the way to overcome the challenges in the identification of host factors at the genome-wide level and also to study the complex interaction networks in the infected environment [126, 128-130]. In this study, several approaches of functional genomics including genome-wide transcriptomics, genome-wide knock out screens, and in combination with genome-wide proteomics have been used to identify the host factors involved in the HCoV-229E (a species of coronavirus that is highly adapted to the humans) and MERS-CoV replication and additionally in the thapsigargin-mediated anti-viral effect. In the following sections, the data obtained in this study using these approaches will be discussed in elaborate.

6.1 Identification of differentially regulated host cell factors in response to HCoV-229E replication and validation of functional relevance to virus replication and / or host response

CoV infection in the host cells triggers direct or indirect dysregulation of the expression levels of genes related to inflammation, stress response, apoptosis, and metabolic responses. Many published reports including the transcriptomic profiling of COVID-19 patients have highlighted more diverse patterns of host gene expression in response to SARS-CoV, MERS-CoV, and HCoV-229E infection [123, 131-133]. However, these studies still lack a comprehensive detail of whether these patterns of gene expression are direct consequences of CoV infection, and have any functional relevance to CoV replication and infection. Therefore, to obtain more insights into the global host response in the coronavirus-infected cells, transcriptome dynamics was studied (already published by our working group) in HCoV-229E-infected Huh7 cells, which revealed numerous signatures of upregulated genes [113]. In the upcoming sections, the data obtained in this study on the specific set of deregulated host

cell factors upon CoV infection were discussed in detail in the context of virus replication and / or host response.

6.1.1 Identification of crucial host cell factors that were upregulated upon HCoV-229E infection

HCoV-229E infection in Huh7 cells induced the upregulation of 1,073 genes at the genome-wide level using four independent microarray experiments. From the gene set enrichment analysis (GSEA) of the microarray data, the top 10 up and down regulated KEGG pathways including cytokine-cytokine receptor interaction, protein processing in the endoplasmic reticulum, Jak-STAT and MAPK signaling, and metabolic pathways were found to be interesting [113]. From these induced gene sets, 11 interesting host cell factors were selected for validating the functional relevance of the virus and/or host response. The genes were selected based on the mRNA expression (log_2 fold change >= 2) from four independent microarray experiments (Section 5.1), and more interestingly the selected genes were connected to ER-stress response pathways (*EIF2AK3*, and *EDEM1*), transcription factor activity (*ANKRD1*, *KLF6*, and *BHLHE41*/*40*), and metabolic pathways (*FUT1*, *FICD*, *CHAC1*, *ERO1LB*, and *CTH*). The replication of coronaviruses occurs in the host cell cytoplasm and the viral proteins are processed at the ER-Golgi compartments [19]. Up-to-date evidence has shown that the replication of coronaviruses causes ER stress and induces unfolded protein response (UPR) in the infected cells [127, 134]. Several genome-wide microarray analyses showed supporting evidence of induction of several genes related to ER stress such as *EIF2AK3*, *ATF3*, *IRE1a*, and *GRP78* or *BiP* [123, 134, 135]. ER-associated degradation (ERAD) is an ER-stress-associated pathway that has also been extensively studied in coronavirus infections [136, 137]. Coronaviruses, including SARS-CoV and mouse-hepatitis virus (MHV), were reported to recruit ER-associated degradation factor EDEM1 for the biogenesis of DMV which is very crucial for its replication [138]. Coronaviruses hijack and reprogram the cellular metabolism for various stages of viral replication [139]. For instance, a study conducted on orientated analysis by overlapping the host proteins targeted by coronaviruses and annotated proteins in the metabolic pathways of the KEGG database has revealed multiple factors involved in nucleotide metabolism, components of DNA and RNA polymerases, and lipid biosynthesis [140-142]. The replication of the viral genome is mediated mainly by the viral replicase enzymes, however, many host cell factors have been implicated in the process [127]. Many RNA-binding proteins and transcription factors have been reported to bind directly to the UTRs of the coronavirus genome to modulate the viral genome transcription [21]. Taken together, these collective published data provided some encouraging evidence supporting the notion of the selection criteria used to identify gene lists in the context of virus replication.

6.1.2 Differential protein expression of the induced gene sets upon HCoV-229E infection

In the eukaryotic cells, translation is initiated by the eukaryotic initiation factor eIF2 complex that is composed of a regulatory α-subunit, a tRNA-binding β-subunit, and a GTP-binding γ-subunit [143]. The eIF2 complex is inactivated by phosphorylation of its α-subunit by eIF2α kinase 4 (GCN2), PERK, and RNA-activated protein kinase (PKR) and subsequently leads to a halt in the translational rate [143]. Coronaviruses depend heavily on the host translational machinery for the synthesis of viral proteins. Therefore, in response to the viral infection, the host cell would shut down the translational rate as a defense mechanism. This translational inhibition by RNA viruses has long been known [123] and has been reported in many coronavirus infections including the SARS-CoV-2 infected cells [144]. This intrigued to study the translational dynamics of the HCoV-229E-induced selected gene sets. As mentioned in section 5.1.1, there were three different scenarios observed for the protein expression of mRNAs whose expression levels changed reproducibly upon HCoV-229E infection. The protein levels of ANKRD1 and KLF6 were induced at 24 h p.i., whereas, reduced protein levels were observed for EDEM1, PERK, DEC1 / SHARP2, and BiP / GRP78. Unchanged steady-state protein expression levels were observed for CHAC1, CTH, and CREB2 / ATF4.

During stress conditions, among the known three UPR sensors, PERK is generally identified to be the first protein to be activated in response to ER stress [21, 61, 134]. The protein chaperone GRP78 / BiP dissociates from the luminal domain of PERK (held inactive by BiP during unstressed conditions), leading to autophosphorylation and activation of PERK. Subsequently, the activated PERK phosphorylates eIF2α to promote global translational shutdown [21, 61, 134]. In line with previous reports on UPR induction by coronaviruses, HCoV-229E infection caused a reduction in the total protein levels of PERK (indirectly referring to the increased phosphorylated levels of PERK).

GRP78 / BiP, one of the most abundant expressed cellular proteins works to function as a molecular chaperone to bind to the misfolded proteins during stress conditions and initiates ER-stress associated ERAD pathway [145, 146]. Elevated expression of GRP78 and co-localization to the cellular membrane in response to the ER stress has been reported in many pathological conditions including cancers and infections [145, 147]. Additionally, the cell-surface GRP78 was reported to facilitate the entry of many viruses including Zika virus, Dengue virus, Hepatitis C Virus, Ebola Virus, MERS, and SARS-CoV-2 [148, 149]. GPR78 acts by binding to the surface spike proteins of MERS-CoV and SARS-CoV-2 to facilitate the entry and in turn, the virus upregulates the localization of GRP78 to the cellular membrane [150, 151]. Nevertheless, the reduced protein levels of BiP and EDEM1 in the HCoV-229E infected cells could lead to the speculation that the attenuation of these important host factors is very much needed for the virus to evade the host response and to maintain the integrity of viral protein synthesis.

It was previously reported that HCoV-229E, by sophisticated mechanisms, fine-tuned the NF-κB pathway to modulate the activity and expression of NF-κB target genes [113]. In this report, the functional relevance of IKK- NF-κB in the HCoV-229E infection was assessed from the knock down of TNFAIP3 protein. TNFAIP3 / A20 was shown to be involved in the negative feedback regulation of NF- κB signaling. Reduced protein levels of TNFAIP3 resulted in the reduction of viral N protein levels and increased mRNA expression levels of *ANKRD1* [113]. Additionally, ANKRD1 was also reported to play a role in feedback inhibition of NF-κB activity in response to the inflammation [152]. In the same line of evidence, KLF6 was also reported to negatively regulate NF-κB signaling in glioblastoma cells. KLF6 expression was shown to activate multiple genes that negatively control the NF-κB pathway and subsequently reduced the nuclear localization and expression of its target genes [153]. These data suggest that the elevated protein levels of ANKRD1 and KLF6 in the HCoV-229E infected cells could be an immunomodulatory effect of HCoV-229E to evade the host response.

During ER stress, if the host cells cannot resolve the problem causing UPR activation, the cells would eventually undergo autophagy and apoptosis to cope with the stress. The phosphorylated eIF2α preferentially activates the translation of UPR-related genes such as the activating transcriptional factor 4 (*ATF4*). In turn, ATF4 activates downstream targets such as CHOP (transcriptional factor C / EBP homologous protein), GADD34 (growth arrest and DNA damage-inducible 34), and ATF3 [154]. The PERK / eIF2α-ATF4-CHOP pathway is considered essential in inducing apoptosis by several mechanisms including the commonly known regulation of apoptotic factors [154-157]. Interestingly, upregulation of ATF4 in transcriptional and translational levels and modulation of apoptosis have been reported in the events of virus infections including Hepatitis C Virus (HCV), human herpesvirus 8, murine cytomegalovirus, infectious bronchitis virus, and recently in SARS-CoV-2 [158-162]. CHAC1 (ChaC glutathione specific gamma-glutamylcyclotransferase 1), a glutathione degrading enzyme was reported to be connected to the UPR and a downstream gene target of the ATF4-ATF3-CHOP cascade. CHAC1 expression was shown to activate apoptosis through the activation of the apoptotic-inducible factor (AIF-PARP) pathway [163]. CTH (Cystathionine gamma-lyase), a major regulator of glutathione levels and cell survival was shown to be induced in the thapsigargin (chemical ER-stress inducer) treated cells [64]. Taken together from these data, the steady-state protein levels of ATF4, CHAC1, and CTH in the HCoV-229E infected Huh7 cells suggest that the virus maintains the balance between the metabolic functions of these host factors and ER-stress-induced cellular apoptosis.

Overall, the HCoV-229E-induced mRNA transcription of the selected genes does not directly reflect the translational rate of these genes (except *ANKRD1* and *KLF6*), which suggest that there could be possible ways of post-transcriptional or translational control posed by HCoV-229E to regulate the host response to the infection.

6.1.3 Functional relevance of the host factor knock outs on the virus and / or host response

To obtain more insights into the functional relevance of HCoV-229E-induced selected gene sets on viral replication and / or host response, a complete loss-of-function approach of these host factors using the CRISPR-Cas9 system was utilized as mentioned in section 5.1.2. Gene knock out provides the advantage to study the function of a specific gene, the pathways that regulate it, and the effect it has on the pathogens (in particular) that depend on it for pathogenesis. Although in this study, a stable knock out cell line was generated for nine of the selected gene sets as mentioned in section 5.1.2, only five of them were tested for their functional relevance on virus replication and / or host response. This was because the cell lines generated for CTH, DEC1 / 2, and ERO1LB loss-of-function could not be maintained as stable knock out cells over a longer period. Because the Huh7 cells carry several copies of a single gene [164], multiple rounds of Cas9-mediated mutations are required for full knock-outs of all active alleles. This was a possible limiting factor when working with a pool of mutant cells, like in this study, rather than with a monoclonal cell line. However, monoclonal cell lines may acquire additional mutations during selection, which is difficult to control. The knock out cells were infected with HCoV-229E at an MOI 1 for 24 h to assess the viral replication by visualizing the viral proteins on the immunoblots. Additionally, the total RNA from the infected knock out cells was analyzed by RNA-Seq to study the global changes in the host transcriptome as the measure of virus and / or host response.

As mentioned in section 5.1.2, the immunoblots showed no changes in the expression of viral proteins in the absence of the five host factors. The RNA-Seq analysis showed that the pattern of deregulated genes in the infected control (Huh7 parental) cells appeared very much closer to that of the infected knock out cells. Moreover, the absence of the host factors did not very much influence the deregulated pathways upon infection when compared to the infected control cells. Despite the notable findings regarding the 229E-specific transcriptional induction and translational regulation of the selected host factors, the validation data did not yield sufficient evidence to establish a direct association between these host factors and HCoV-229E.

Although the previous microarray of our group as well as the RNA-seq data generated during this thesis, revealed numerous HCoV-229E-regulated gene sets, it was (and will be) a challenging task to validate every gene in the list. In these instances, a selection of filter criteria would always ease narrowing down the list with a handful of genes to validate. However, the selection criteria will greatly influence the outcome of the validation data. In this study, the selection criteria were based on the strict threshold of HCoV-229E-induced mRNA expression and the basic functions of the genes, which are very much connected to coronavirus replication. However, it is important to consider that the reverse genetics approach utilized in this study (which aims to establish a link between a gene mutation and a specific phenotype), along with the experimental methods employed to validate the identified genes, may have

imposed certain constraints on attaining a desired phenotype (such as any observable phenotype on viral replication and / or the host response).

In the following, possible explanations why this approach could be a drawback are briefly discussed concerning the host factors examined in detail.

ANKRD1 was found to be upregulated in both transcriptional and translational levels during Hepatitis C Virus infection. Moreover, in this study, the upregulation of ANKRD1 was stated to be mediated through HCV-induced ER stress in the infected cells [165]. Additionally, this study showed that silencing of ANKRD1 disrupted the propagation of HCV. Another study reported that the expression of ANKRD1 was induced upon the stimulation of herpes simplex virus 1 in human peripheral blood mononuclear cells [166]. By silencing the *ANKRD1* gene, the authors showed that the HSV-1 viral load was significantly increased in the antigen-presenting cells. Although the regulation of the *ANKRD1* gene by HCoV-229E was similar to these studies, the knock out of the gene did not show an increase or decrease in the viral protein levels. This could simply mean that HCoV-229E does not entirely depend on this gene for the synthesis of viral proteins and silencing this gene did not favor much for the viral replication as well. To understand the upregulation of ANKRD1 protein despite CoV-induced translational shutdown, a possible experiment could be performed to transiently over-express the ANKRD1 protein expression in Huh7 cells and measure the viral protein synthesis and viral loads by plaque assays in the infected cells. If the desired phenotype (changes in the viral replication) is achieved, then it could be a possibility that the regulation of ANKRD1 was part of an anti-viral host response upon the HCoV-229E infection.

The ER-associated degradation (ERAD) pathway helps in maintaining cellular homeostasis during the ER-stress by targeting the misfolded proteins in the ER for degradation [167, 168]. ER-degradation enhancing α-mannosidase-like (EDEM1) was consistently reported as a crucial regulator for ERAD [169]. Many viruses fine-tune the ERAD activity by dampening by selectively clearing the ERAD components by segregation to ER-derived vesicles called EDEMosomes [168, 170]. For instance, EDEM1 knock out was shown to improve the production of HCV infectious particles and improved the stability of HCV E2 proteins and the over-expression of EDEM1 caused the degradation of HCV E2 proteins. Similarly, the expression of EDEM1 along with the wild-type S, M, and L proteins of the Hepatitis B Virus (HBV) promoted the degradation of the viral proteins [168]. Although it has been already shown that the MHV hijacks EDEM1 for its replication [137, 138], the reduced EDEM1 protein levels in HCoV-229E infected cells could suggest a possible role of HCoV-229E in fine-tuning the ERAD pathway for protecting the viral proteins from degradation. In this case, an experiment with the over-expression of EDEM1 in the Huh7 cells could provide insight if this host factor targets the HCoV-229E viral proteins for degradation.

Taken together, in the process of finding the functional relevance of the HCoV-229E-induced gene sets in the context of virus replication and / or host response, the initial hypothesis was laid on the strong evidence, however, the validation results entirely differed from the hypothesis. It was therefore concluded that validation experiments using unbiased, orthogonal methods for assessing the function of a gene could overcome the limitation faced in this study.

6.2 Identification of host factors required for HCoV-229E and MERS-CoV replication using genome-wide CRISPR-Cas9 knock out screening

Global changes in the host cell transcriptome during the coronavirus infection are generally due to 1) cellular defense responses, 2) virus-induced responses that are essential for viral replication, and 3) an undesired side effect during the infection that is detrimental for both, virus and host [129]. Although the microarray data could reveal multiple interesting patterns of gene expression that belong to any of the three categories, for identification and development of anti-viral drugs both 1) and 2) are interesting [171, 172]. It is therefore essential to use forward genetic or phenotype-based screens instead of gene-driven approaches that depend on any prior assumptions on the gene functions. Recent developments in the functional genomics approaches including the genome-wide RNA interference (RNAi) screen, haploid genetic screen, and genome-wide CRISPR-Cas9 knock out screen have uncovered potential host factors involved in the virus-host interactions [173]. However, the first two mentioned approaches encompass a few limitations including the false positive results due to high off-target effects in the RNAi screen and the inability to produce homozygous mutations in the case of the haploid screen [174, 175]. In recent years, the development of genome-wide CRISPR-Cas9-based genetic screen has outperformed the RNAi and haploid screen by wide-range of genome editing in multiple cell models with minimized off-target effects, more marked phenotypes, and a higher signal-to-noise ratio [94, 176]. Genetic screens using CRISPR-Cas9 aims to elucidate the direct link between the phenotype and genotype by introducing genetic perturbations at a genome-wide level and thereafter analyzing the resulting phenotypic alterations. A successful genetic loss-of-function screen relies on three essential components: the genetic perturbation, the model, and the assay [91]. Recent advancements in the generation of genome-wide sgRNA libraries (also mentioned in section 2.5.1.1) have enabled the investigation of perturbations at a genome-scale. A model and a suitable assay determine the outcome of the genetic screen with a list of genes manipulated by CRISPR-Cas9. This study was intended to identify the host factors that resist virus replication and / or virus-induced cell death. Many viruses including coronaviruses go through sequential steps including viral entry, replication, assembly, and budding that subsequently cause cytopathic effects in the host cells upon infection [177, 178]. By utilizing the virus-induced cytopathic effect as an assay, the genetic screen that has been performed in this study was intended to select cells as a result of the inactivation of a pro-viral factor, which supports viral entry, replication, and in most cases viral egress. The upcoming sections will elaborate on the data obtained using the genetic screen to identify the host factors involved in HCoV-229E and MERS-CoV

replication. Additionally, the data produced in this study were compared to other published genetic screens performed on HCoV-229E and MERS-CoV to state the novelty, advantages, and nuances in the usage of genome-wide knock out screens.

6.2.1 Genome-wide CRISPR-Cas9 knock out screen revealed potential host factors required for HCoV-229E replication

Before delving into the insights on the host factors that were identified to be essential for HCoV-229E replication, it was indispensable in this study to briefly highlight the prerequisites of the genome-wide screen. As mentioned in section 5.2.1, the sgRNA library (GeCKO V2) used in this study was amplified and deep sequenced in very thorough steps which resulted in an 84.6% match to the input sgRNA library sequences after NGS and analysis. The violin plots generated from the $\log_2$ transformed read counts indicated that the amplified sgRNA library had a normal read count distribution and the vast majority of the sgRNA sequences were represented about 100 times (median at 6.6 $\log_2$ scale). There were around 3400 sgRNA sequences that had completely zero read count after the analysis. Although only 2.9% of the sgRNA sequences (out of 123,411) were dropped out from the sgRNA library amplification, it was decisive enough to use the amplified sgRNA library to generate the genome-wide knock out cell pool.

The initial approach for screening the library transduced cells for 7 days of HCoV-229E infection was based on the cell viability assay (data was not shown in this thesis) performed on Huh7 cells upon HCoV-229E infection time-course for 7 days. The MTS assay showed a linear decline in the cell viability of Huh7 cells upon HCoV-229E infection and on day 7, the survival rate was less than 10%. This time point of infection challenge perfectly fits the screen selection, as the positive-screen selection pressure is usually much more stringent in order to ferret out the most resistant survivor cells at the end of the selection.

As expected, at the end of screen selection for 7 days of the HCoV-229E challenge, there were roughly around 1% (out of 80*10^6 cells used for the challenge) of the cells that survived in both of the two biologically independent screen selections. The outcome was much more stringent than observed from the MTS assay data, as nearly 99% of the library transduced cells were dropped out during the selection. In both biological replicates of the screen selection, the surviving population took approximately four weeks to propagate due to the necessary expansion of surviving cells during the survival screen selection. This was crucial to distinguish the enrichment of specific sgRNAs from the unselected population and identify the genes responsible for conferring resistance.

The NGS data from the 7 days screen selection revealed that the unselected controls had a normal distribution of full sgRNA library (section 5.2.4) and the control metrics (histograms) showed that only about 6000 – 7000 sgRNAs had relatively lower frequencies. Only about 3800 sgRNA sequences had zero counts in the unselected controls, which indicated the possibility that the sgRNAs, which target

essential genes for core-fitness of Huh7 cells in culture, were depleted during the generation of genome-wide knock out cell pools.

Almost 50,000 to 76,000 sgRNA-carrying cells were completely depleted from the 7 days of the HCoV-229E selection process in both of the replicates (mentioned in the histograms of section 5.2.4) and the survivor sgRNAs were analyzed for the enrichment and gene identification using MAGeCK. As shown from the multi-variable bubble plots, the screen selection resulted in the identification of 104 genes with a significant score (MAGeCK score <= 0.005). The top candidate gene was *ANPEP* with the MAGeCK score = 1.56E-19 and all of the six targeting sgRNAs were truly enriched compared to the unselected controls, corresponding to an aggregated $\log_2$ fold change = 9.5 at the gene level. Although the MAGeCK score was significant for the other 103 candidate genes, the sgRNA enrichment scores were usually much lower. These genes had only one out of six targeting sgRNAs that were enormously enriched, but the MAGeCK algorithm still placed these genes at the top order across the entire gene list. The rationale approach to finding the host factor mutants that resist the HCoV-229E-induced cell death using the stringent selection pressure, lead to the identification of ANPEP as the most significant hit. This proved that the initial hypothesis and the methodology to achieve the desired result were successful in using a genetic screen. However, from the list of gene hits, it was rather difficult to choose the candidates for further validation of the screen phenotype, making them potentially less robust hits due to the lower sgRNA enrichment values.

A possible explanation for this scenario could be raised from the time point of infection challenge used as the selection pressure. Although the main idea was to identify the strong mutant survivors that could resist the multiple rounds of infection, the screen results showed that only ANPEP mutant cells can withhold this extreme genetic pressure. If there were a possibility for some other genes with a subtle resistant effect, the strong selection pressure would pose a definite limitation to identifying these survivors in this type of study. A similar situation was published in a recent study conducted to identify the host factors that support *Streptococcus agalactiae* ß-Hemolysin / Cytolysisn (ßhc)-induced cell death using a genome-wide CRISPR-Cas9 knock out screen. The positive screen selection with the strong lytic units of ßhc toxin in the HeLa cells uncovered only two significant genes (*VPS13A* and *PLA2G15*). However, the validation experiments proved no effect on ßhc-mediated cell death from knocking out either VPS13A or PLA2G15 [179].

In the following, the screen selection was reduced to 72 h of HCoV-229E infection, which was enough to lose more than 60% of the Huh7 cells (from the time-course MTS assay data for 7 days). Additionally, the number of cells used for the HCoV-229E challenge was increased to 2*10^8 cells (1600 X for each sgRNA), which was twice as compared to the previous screen. This alternative strategy ensured a moderate level of stringency and at the same time provided an opportunity for the potentially resistant mutants to overcome the challenge faced in the previous screen selection process.

As expected, the 72 h screen selection resulted in the enrichment of 492 genes with a MAGeCK score <=0.005, and among them, *ANPEP* was the top ranking gene with a MAGeCK score = 1.43E-16 and all the targeting sgRNAs were truly enriched with an enrichment score = 6.5 (average $\log_2$ fold change). These data validated again the robustness of genetic screening in the identification of viral entry receptors at any given selection pressure. Among the many interesting candidate genes, for further validation of the screen phenotype, 11 top-ranking genes (*ANPEP*, *SFTA2*, *HDAC4*, *ATG7*, *ZDHHC3*, *SYT3*, *SYT16*, *CTCFL*, *KIF3B*, *ORMDL1*, and *CPB2*) were selected based on the sgRNA enrichment score. Single knock out cell lines for these host factors were generated using the top one or two enriched sgRNAs from the sgRNA screen analysis. The validation experiments were conducted to validate the screen phenotype and additionally to check if the absence of host factors affected any of the stages involved in the coronavirus replication.

Therefore, the knock out cell lines were first tested for the screen phenotype, i.e. virus-induced cell death by infecting the cells for 72 h (time point of selection pressure), and the cell viability was measured by MTS assay (Section 5.3.1). Surprisingly, except for ANPEP KO (which showed almost 100% cell viability), none of the KO cell lines showed an improvement in cell viability upon 72 h of HCoV-229E infection. In the next validation experiment, the viral S-mRNA transcripts were measured in the infected knock out cell lines to assess the rate of viral RNA transcription in the absence of the host factors. The viral S-mRNA transcript copy numbers were not greatly affected by knocking the host factors down. An interesting observation in these data was that the viral S-mRNA copy number in the infected ANPEP KO was closer to that of the infected empty vector cell line. Although the absence of ANPEP completely resisted the HCoV-229E-induced cell death, the viral RNA transcription rate was still uninterrupted. This raised speculation on the unknown host factors utilized by HCoV-229E for cellular entry. It has been previously reported that HCoV-229E utilizes type II transmembrane serine proteases TMPRSS2 and HAT to cleave the spike S protein to facilitate cell-cell fusion. In the same report, it was shown that the HCoV-229E infection was significantly higher in the human airway epithelial cells (HAE) that co-expressed TMPRSS2 and CD13 (ANPEP) [180]. In another study, clinical isolates of HCoV-229E were shown to preferentially use TMPRSS2 for the cleavage and activation of the viral spike protein [181]. In the absence of the extra-cellular proteases, HCoV-229E was also shown to use the lysosomal protease CTSL (Cathepsin L) for cellular entry via late endosome [182]. Interestingly, other than cellular proteases, HCoV-229E was shown to use carbohydrate-binding protein CD209L (L-sign) to facilitate the cells [183]. Nevertheless, the in-depth underlying mechanisms by which HCoV-229E bypasses the ANPEP-mediated entry are still to be elucidated and the observation made in this study makes it more intriguing how HCoV-229E manages to facilitate the host cell entry in the absence of a conserved viral receptor.

Finally, the rate of viral egress was assessed from 24 h p.i. up to 72 h p.i. in the absence of the host factors using plaque assays. Except for ANPEP KO, the viral titers measured at 24 h p.i. for the other host factor knock outs were comparable to the controls. Interestingly, the viral titers were reduced at 48 h p.i. in the SFAT2, HDAC4, ATG7, and ZDHHC3 knock out cell lines compared to the controls. SFTA2 (Surfactant Associated 2), a pulmonary surfactant protein is predominantly expressed in lung tissues and predicted to be associated with the secretory pathway. Pulmonary surfactant is an important component that is comprised of phospholipids and proteins for the maintenance of surface tension at the air / liquid interface in the alveoli, lipid homeostasis, and modulation of the immune responses [184, 185]. Although the major pulmonary surfactant proteins are SP-A, B, C, and D, surfactant-associated proteins (SFTAs) were recently reported to have similar properties to that of SP-A and SP-D [186]. Pulmonary surfactant treatment was recently administered to COVID-19 patients to improve lung compliance and less oxygen requirement in severe cases [187].

Histone deacetylases (HDACs) are the enzymes that catalyze the removal of an acetyl group from the lysine residues of histones (and non-histone proteins), allowing them to wrap tightly around the DNA resulting in more condensed chromatin structure and repression of gene expression. HDACs have been previously reported to play a role in type I interferon signaling and modulate the replication of numerous viruses [188, 189]. Recently, the inhibitor of HDACs activity called valproic acid (sodium valproate) has been shown to inhibit the replication of many enveloped viruses including the West Nile Virus (WNV) [190]. Interestingly, valproic acid was recently suggested for the repurposed use of treating COVID-19 patients with ARDS due to the proposed model of valproic acid regulating the expression of pro-inflammatory genes [191, 192].

ATG7 (Autophagy related protein 7) is an essential autophagy effector enzyme that in concert with other autophagy proteins is responsible for the elongation and maturation of autophagosome in the early autophagy pathway [193]. In the past decade, several reports have implicated the role of autophagy in the context of coronavirus replication [194, 195]. Although there are many discrepancies over the interplay between the many different coronaviruses and the autophagy machinery [196], the evidence still suggests the therapeutic potential of targeting the autophagy process for current and future coronavirus infections.

ZDHHC3 (DHHC-type palmitoyltransferase 3) is a Golgi-localised palmitoyltransferase that catalyzes the addition of palmitate onto various protein substrates as a part of post-translational modification (PTM). PTMs have been reported to play an important role in the modification of coronavirus structural proteins that impact viral replication and pathogenesis [197]. In particular, palmitoylation of coronavirus spike protein by DHHC proteins has been recently reported to be critical for viral egress, syncytia formation, and infection progress in the SARS-CoV-2 replication [197, 198]. Moreover, the inhibition of DHHC enzyme palmitoylation activity by a palmitoyl-acyltransferase inhibitor 2-

bromopalmitate (2-BP) has been shown to effectively inhibit the SARS-CoV-2 replication by the reduction of acylation of the viral spike proteins in-vitro [199, 200].

Taken together, the genetic screening conducted in this study to uncover the resistant mutants of HCoV-229E infection led to the identification of many interesting host factors including the well-known entry receptor. Among them, a few host factors which were validated to address the functional effect on the viral replication showed a mild phenotype on the viral budding.

6.2.2 Genome-wide CRISPR-Cas9 knock out screen revealed potential host factors that might be required for MERS-CoV replication

Although the MERS-CoV virus was not a focus of this thesis, it aroused a curiosity to apply the genome-wide screen to identify the host factors involved in the more severe coronavirus (MERS-CoV) infection. As mentioned in section 5.2.6, the genetic screen uncovered 486 genes with a MAGeCK score <= 0.005 that were important for MERS virus-induced cell death. Among the several interesting hits which were filtered out with the significant score, a few genes that stood out at the top ranking order caught the attention to briefly discuss in this section.

DPP4 / CD26 (Dipeptidyl peptidase IV) is a type II transmembrane glycoprotein that is widely expressed in various tissue types and functions to cleave the X-proline dipeptides from the N-terminus of polypeptides such as chemokines, neuropeptides, and peptide hormones [201]. DPP4 acts as the entry receptor for the MERS virus and the identification of this host factor as the top hit in this study proved again the versatility of the genome-wide screen in the identification of entry receptors for any viruses. Interestingly, ANPEP was also among the top order with a significant MAGeCK score (2.85E-05) and a higher sgRNA enrichment score ($\log_2$ FC = 6.3). This raised speculation on the usage of ANPEP by the MERS virus for entry into the Huh7 cells. Since ANPEP also functions by cleaving the peptides from the N-terminal ends of the polypeptides [202] and both DPP4 and ANPEP belong to the exopeptidase family, it might be possible that the MERS virus can utilize this host factor for the cleavage of the spike protein and endocytosis-mediated cellular entry. Nevertheless, more validation experiments have to be performed to reveal this proof of concept.

TMPRSS2, a type II transmembrane serine protease has been extensively reported to function as the entry mediator for many different coronaviruses including the MERS-CoV [203]. MERS-CoV utilizes the TMPRSS2 to cleave the viral spike proteins for the host cell entry and it has been reported to facilitate the syncytium formation and the progression of infection spread in the MERS-CoV virus infection [203, 204]. Additionally, the supporting pieces of evidence on the inhibition of MERS-CoV and recently SARS-CoV-2 replication by blocking the TMPRSS2 function using serine protease inhibitors such as camostat and MM3122 highlights the importance of TMPRSS2 as a potential host factor target for the pan-coronaviruses [205, 206].

HNF1A (Hepatocyte nuclear factor 1α) is a liver-specific transcriptional factor that is shown to play a role in the transcriptional regulation of liver-specific genes. HNF1A was previously shown to interact with histone acetyltransferases (HATs), CREB-binding protein (CBP), and its co-factor PCBD1 (a strong hit in this screen) [207, 208]. PCBD1 selectively interacts with HNF1A and its paralog HNF1B to stabilize their homodimers and enhance their transcriptional activity [207]. Mutations in the HNF1A and PCBD1 have been linked to the early onset of diabetes in the young and hypomagnesemia (low levels of Mg^{2+} in the serum) [209, 210]. Recently, many reports associated the link between hypomagnesemia and cytokine storm (excessive release of pro-inflammatory cytokines) in the more severe COVID-19 patients [211].

6.2.3 Comparison of the hits identified in this study to the multiple genetic screens that identified the host factors promoting HCoV-229E infection

The filtered gene sets from all three genetic screens performed in this study were compared to each other to understand the commonalities between the selection pressure (in the case of HCoV-229E screens) and to identify broad-spectrum coronavirus host factors. For the HCoV-229E screen, 104 genes were filtered out for the high-stringency screen and 492 genes for the moderate-stringency screen. 486 genes were filtered out for the MERS-CoV screen selection. The gene sets were compared to each other using a Venn diagram as mentioned in section 5.4.1. Surprisingly, only 3 genes (*ANPEP*, *GGCX*, and *FAM156A*) were found to be shared by all three genetic screens. 18 genes were shared between the high and moderate stringency HCoV-229E screens and among which, *SYT16*, *KIF3B*, and *CPB2* were found in the top ranking order for 72 h screen selection that was validated for the screen phenotype. 53 genes were found common to both 72 h HCoV-229E and MERS-CoV screen selection. Metascape over-representation analysis revealed that the transcription coregulatory activity and small molecule catabolic process were the only interesting shared pathways for both HCoV-229E and MERS-CoV infection. Taken together from this analysis, the level of stringency used in the selection pressure strongly influenced the outcome of the resistant mutants except for the viral entry host receptor. Although, there were few genes and pathways that were shared among both viruses, arguing that the evolutionary differences in the HCoV-229E and MERS-CoV might explain the possibility of why these viruses utilize many different genes belonging to the same or different pathways for survival and evolution.

Similar genetic screens that identified the host factors promoting the HCoV-229E infection were published in parallel to the time when this study was conducted. To highlight the novelty and commonality between the published data and this study, the gene sets obtained from the HCoV-229E screen were compared to four other genetic screens performed on SARS-CoV-2 and the common cold coronaviruses including the HCoV-229E (as mentioned in the section 5.4.2).

Hoffmann et al. [118] used the 332 hits from the protein-interactome data (SARS-CoV-2 viral proteins interacting with host proteins) and designed a custom-made focused sgRNA library to perform genetic screens on SARS-CoV-2, and common cold coronaviruses (HCoV-229E, NL63, and OC43). The screen selection was performed in the Huh7.5 cells with the challenge of HCoV-229E at an MOI 0.05 at 37℃. Using the Z-scores as the filtering criteria, Hoffmann et al. identified 27 genes that were highly enriched in the HCoV-229E screen. These genes were identified to be the components of Rab GTPases, HOPS complex, and endoplasmic reticulum protein complex (EMC) [118]. Nevertheless, only 1 out of 27 genes were found in the high and moderate-stringency HCoV-229E screen.

Schneider et al. [117] performed the genetic screens on SARS-CoV-2 and the common cold coronaviruses similar to Hoffmann et al. using the Brunello sgRNA library (76,441 sgRNAs targeting 19,114 human genes). The screen was performed on Huh7.5 cells that constitutively expressed Cas9 protein and were challenged with HCoV-229E at an MOI 0.05 at 33℃. Schneider et al. used a high-confidence Z-score and identified 21 genes that were highly enriched in the screen selection. Interestingly, 3 genes (*ANPEP*, *VMP1*, and *TMEM41B*) were found to be shared in the 72 h HCoV-229E screen selection. However, the gene sets identified by Schneider et al. and Hoffmann et al. differ greatly from each other.

Wang et al. [116] also performed a similar genetic screen on SARS-CoV-2 and common cold coronaviruses as the other two studies using the human GeCKO V2 sgRNA full library in Huh7.5.1 cells that constitutively expressed Cas9 protein. The screen selection was performed with HCoV-229E at an MOI 0.05 at 33℃ for 72 h and the surviving cells were expanded for about two weeks. Using the MAGeCK score <= 0.005, 597 genes were filtered from Wang et al. dataset and were compared to the gene sets from this study. Interestingly, 30 genes were found to be shared with the dataset of Wang et al., and among them, *ANPEP*, *VMP1*, and *TMEM41B* were shared between the three studies (including Schneider et al.).

Kratzel et al. [119] performed the genetic screening to identify the host factors involved in HCoV-229E and MERS virus infection, similar to this study. The screen selection was performed in Huh7 cells transduced with the GeCKO V2 library A and was challenged with HCoV-229E (33 ℃, MOI 0.1) and MERS (37℃, MOI 0.05) until the untransduced Huh7 cells were dead due to the virus-induced cytopathic effect. Although the parameters used in the screen selection almost matched the settings used in this study, only 12 out of 523 genes (filtered using MAGeCK score <= 0.005) were shared between the two studies. Interestingly, *ANPEP*, *VMP1*, and *TMEM41B* were the only genes found to be common to this study and the compared published studies (except Hoffmann et al.).

Taken together from the comparative studies, there was a limited overlap in the set of genes identified as hits in the compared genetic screens. This limitation could arise because the parameters used in the genetic screens widely differ in the choice of the cell line, sgRNA library used for the genome-wide knock out, conditions used for the infection challenge, and finally, the bioinformatic tools and filter settings used in the analysis of the NGS data. Although there were few genes (essential host factors for the viral entry) and pathways that were shared among the many different genetic screens and coronaviruses, several other host factors were found to be uniquely interesting from the genetic screens that were published. However, the published genetic screens gave more focus to the well-known host factors such as TMEM41B, TMEM106B, TMPRSS2, and VMP1 to highlight the essence of the study as pan-coronavirus targets [212-216]. To have a deep insight into the host factor utilization of many different coronaviruses, an unbiased compendium of host factor validation studies should be conducted using orthogonal approaches in the future.

6.3 Genome-wide loss-of-function screen revealed host factors that support thapsigargin-mediated inhibition of HCoV-229E virus-induced cell death and replication

A recent study published by our working group reported that the ER-stress inducer thapsigargin effectively inhibits the replication of coronaviruses including the HCoV-229E, MERS-CoV, and SARS-CoV-2 in multiple cell types. The key findings of the study were that the Tg-treated infected cells could partially overcome the virus-induced translational shutdown, enhance the cell viability of the infected cells, and reverse the CoV-mediated downregulation of IRE1α and BiP protein levels [64]. Thus, taken together from this finding it was intriguing and worth using the unbiased genome-wide knock out screen to identify the genes and pathways underlying the mechanism by which thapsigargin mediates an anti-viral effect.

As mentioned in section 5.5.2, a genetic screen was conducted in the HCoV-229E infected cells in the presence or absence of thapsigargin (at a concentration of 1 μM) and the resistant mutants were selected directly after 72 h or after further expansion (for two to four weeks) in case of the 96 h selection protocol. The screen selection resulted in the identification of 968 genes (MAGeCK score <= 0.005) for 72 h selection and 1022 genes for 96 h selection from all three conditions (HCoV-229E and / or Thapsigargin). The pathway analyses of the significantly enriched sgRNAs from the six groups, revealed multiple pathway categories, although not strongly enriched, related to Tg, HCoV-229E, and in both conditions, indicating the involvement of different cellular processes. The comparative pathway term analysis of the screen hits from all six conditions using the upset plots revealed that there was very little overlap with a maximum of three terms being shared between any of the conditions. This suggests the presence of unique molecular mechanisms underlying the respective conditions and highlights the complexity of cellular responses to Tg-induced stress and HCoV-229E infection. Further validation of

these enriched pathways could provide a deeper understanding of the specific pathway supporting the anti-viral effect of thapsigargin.

To select the hits for further validation of screen phenotypes and to elucidate the functional relevance of the gene hits on the anti-viral effect of thapsigargin in the infected cells, an additional filter criterion using the unpublished data from (nascent) proteomics (generated from the Tg-treated infected cells for 24 h by our working group) was used to intersect the hits from the genomic screen and nascent proteomics.

The intersection of these two types of data (mentioned in section 5.5.6) narrowed down the hit lists to 302 and 257 IDs that were found in the proteomics data for 72 h and 96 h screen selections, respectively. To further scrutinize the list, the matching IDs were plotted on a multi-variable bubble plot to look at their multiple altitudes (sgRNA enrichment, MAGeCK score significance, and protein intensity and synthesis levels). The multi-variable plot for the intersected hits from the 72 h screen selection showed that the protein synthesis levels were significantly lowered upon the HCoV-229E infection, however, upon the treatment with thapsigargin, some of the downregulated IDs were found to be enriched. Nevertheless, the 72 h selection showed weak sgRNA enrichment scores for the intersected IDs due to the lack of further expansion of the surviving sgRNA cells. The multi-variable plot for the intersected hits from the 96 h selection showed a massive downregulation of the protein synthesis levels upon the infection, however, the treatment with thapsigargin enhanced the synthesis rate of some of the protein hits and at the same time diminished the synthesis of some of the protein hits. This data supported the proteome-wide pattern of thapsigargin-specific deregulation of proteins in the coronavirus-infected cells as reported by Shaban et al. [64].

From this specific multi-variable plot analysis, 10 genes (*DDRGK1*, *ARFGAP3*, *EMC6*, *SEC24A*, *HMOX2*, *CLU*, *TGFB1*, *SDE2*, *FAM50A,* and *ERMP1*) that were specifically regulated by thapsigargin upon the infection were picked for validation studies. These genes play an important role in the regulation of protein ufmylation and ER-homeostasis (*DDRGK1*) [217], chaperone activity during induced ER-stress (*CLU*) [218], cargo sorting and membrane fusion events in the golgi (*ARFGAP3*) [219], endoplasmic-reticulum membrane protein biogenesis (*EMC6*) [220], endoplasmic-reticulum metallopeptidase activity (*ERMP1*) [221], and COPII-mediated vesicle formation in the secretory pathway (*SEC24A*) [222]. By combining the collective published data and analysis from the genome-wide screen, it is evident that the host factors identified are strongly linked to ER stress, ER-Phagy, and vesicle transport. This observation aligns with the previous studies that have validated the role of these host factors in coronavirus replication, providing a foundation for understanding the Tg-mediated antiviral effect [134, 135, 137, 138, 141].

6.3.1 Effect of ARFGAP3, EMC6, and CLU on the viral replication and cell viability of thapsigargin-treated infected cells

The treatment with thapsigargin induced the expression level of ARFGAP3 protein in the parental and empty vector cell lines. This is in line with the nascent proteomics data, as the protein synthesis level of ARFGAP3 measured for the thapsigargin-treated infected cells was highly enriched when compared to the infection alone. Interestingly, the reduction in the ARFGAP3 protein level for the sgRNA targeting ARFGAP3_sg2 cell line showed a mild increase in the viral N protein level. The absence of ARFGAP3 enhanced the cell viability of the sgRNA-carrying cell lines upon the long-term (96 h) HCoV-229E infection. On the other hand, the viral titer was slightly increased for the infected ARFGAP3_sg2 cell line. Upon the thapsigargin treatment, the infected ARFGAP3 knock out cell lines enhanced the cell viability but maintained the long-term anti-viral effect of thapsigargin as measured by the viral titers.

ADP-ribosylation factor GTPase-activating proteins (ARFGAPs) play an important role in the hydrolysis of GTP in the adenosine diphosphate ribosylation factor 1 (Arf1), which in its bound-GTP form recruits the COPI-coatomer, a protein complex that forms the membrane-bound transport vesicles and is responsible for cargo sorting and retrograde transport from Golgi to the ER [223-226]. However, COPI coat proteins are also found in the vesicles involved in the other trafficking processes including endocytosis, autophagy, and the anterograde transport from ER to Golgi [227]. Many enveloped viruses utilize these coat protein complexes for the maturation of the envelope and membrane proteins and release into the plasma membrane [227]. It has also been extensively shown that the inhibition of Arf proteins and the coatomer complexes have disrupted the maturation of viral envelope proteins and the viral egress of many enveloped RNA viruses [227-230]. However, the data obtained in this study explains a completely different possibility on the usgae of COPI vesicles by the HCoV-229E. Taken together from this collective validation data, the induction of ARFGAP3 by thapsigargin and the increase in the HCoV-229E viral replication in the absence of ARFGAP3 leads to the speculation that ARFGAP3 might play an anti-viral role mediated by an unknown specific regulation of thapsigargin.

EMC6 protein levels were relatively lower in the infected and / or thapsigargin-treated parental and empty vector cell lines. Interestingly, the complete loss of EMC6 protein mildly reduced the expression of viral N protein and the viral titer and enhanced the cell viability of the infected EMC6_sg2 cell line. However, the thapsigargin treatment did not influence the cell viability of the infected EMC6 knock out cell line, and the anti-viral effect was completely maintained.

Endoplasmic-reticulum membrane protein complex (EMC) plays an important role in the biogenesis of integral membrane protein by providing the insertion, folding, and stability of their transmembrane domains (TMDs) [231, 232]. EMC6 was shown to be the core component for the maturation of the

EMC [232]. Loss of EMC has been reported to cause a range of cellular malfunctions including the aggravation of ER stress and protein-misfolding, activation of ERAD / autophagy, and change in lipid homeostasis [220, 232, 233]. The role of EMCs has been extensively reported in the replication of flaviviruses including the dengue virus, Zika virus, and West Nile Virus using genetic screens and indicated that the EMCs are responsible for the topology and expression of non-structural viral proteins [234-237]. Taken together, the mild reduction in the viral replication in the absence of EMC6 and the complete reduction of EMC6 protein by thapsigargin in the infected cells leads to the possibility of a pro-viral role of EMC6 in the HCoV-229E replication.

Tg treatment significantly enhanced the survival of HCoV-229E-infected cells in the absence of CLU. Clusterin (CLU / Apolipoprotein J) is a stress-induced, and highly-glycosylated chaperone that is involved in various intrinsic cellular responses to ER stress [238]. It has been shown previously that the expression of CLU was upregulated during the ER stress or genetic perturbations disrupting the ER homeostasis [239, 240]. Interestingly, CLU was shown to interact with misfolded proteins, prevent their aggregation, and facilitate their refolding, thereby contributing to ER homeostasis [241]. Further studies have claimed both apoptotic and cytoprotective activity of CLU in response to ER stress [242, 243].

6.3.2 SEC24A mildly attenuated the anti-viral activity of thapsigargin-mediated inhibition of HCoV-229E replication

The absence of SEC24A protein did not influence the cell viability and the viral replication upon the long-term HCoV-229E infection. However, when the infected SEC24A knock out cells were treated with thapsigargin, the cell viability was enhanced. Interestingly, the long-term anti-viral effect of thapsigargin which was observed in the parental and the other knock out cells (ARFGAP3 and EMC6) was slightly diminished in the absence of SEC24A.

Coat protein complex II (COPII) contains a highly conserved set of proteins that are responsible for the production of small membrane vesicles for the cargo transport of many different proteins that are processed in the ER-lumen and are transported to different compartments of the cell via ERGIC [222, 244, 245]. COPII vesicles contain five important core proteins: Sar1, Sec23, Sec24, Sec13, and Sec31 which are sequentially recruited and activated to form the vesicle [222]. Among the other COPII proteins, Sec24 is considered to be the main protein that binds to the cargo proteins for internalization into the vesicles of the ER [246]. Given the nature of Sec24 to associate with many of the cargo proteins, it has been reported that many viruses including HBV and Ebola virus utilize Sec24 for the cargo binding of viral proteins for the transport and viral egress [247, 248]. However, in this study, it was very evident that HCoV-229E maintained lower SEC24A protein synthesis levels and upon the thapsigargin treatment, the SEC24A protein levels were greatly induced, raising a speculation of the role COPII vesicles in the ER-phagy / autophagy and ERAD. Recently, thapsigargin was shown to

effectively inhibit the CoV-induced selective autophagic flux by blocking the lysosomal acidification of autophagosomes and a strong induction in the ERAD factors [62, 64]. Therefore, it is likely a possibility that the thapsigargin-mediated inhibition of HCoV-229E is supported by COPII-mediated cargo recruitment of viral proteins by SEC24A and destruction of viral proteins through ERAD [249].

Another interesting explanation could be derived from the recently published data using the genetic screen that identified SEC24A as an essential mediator of thapsigargin-induced cytotoxicity [124]. In this screen, only SEC24A was identified as a significant hit, with no other paralogues or redundant factors being detected. Furthermore, in the same study, it was postulated that SEC24A plays a role in the recruitment of ATF6 cargo, which is involved in the activation of cell death pathways induced by thapsigargin [124]. Thapsigargin specifically works by inhibiting the SERCA Ca^{2+} flux from the cytoplasm to the ER and therefore, the depletion of ER Ca^{2+} levels activates the UPR. The extension of the genetic screen study by Chidawanyika et al., 2022, identified that SEC24A specifically modulated the Ca^{2+} flux from the ER to mitochondria and the absence of SEC24A reduced the rate of thapsigargin-induced depletion of Ca^{2+} from the ER [250]. These data suggest that the anti-viral effect of thapsigargin observed in this study could be partially mediated via the modulation of Ca^{2+} flux by SEC24A.

6.4 Limitations of this study and outlook

This study strictly focussed on the identification of host factors that are required for the HCoV-229E replication using functional genomics approaches. The reverse genetics approach using the microarray data of the HCoV-229E-induced gene sets identified several interesting host factors involved in metabolic process, ER-stress, and transcription factor activity that in context to virus replication was very well suited. However, the selected gene sets that were validated for the functional relevance of the genes on virus replication and / or host response did not result in any interesting phenotype. This limitation of linking a gene to a particular phenotype was subjugated by the forward genetics approach using the unbiased genome-wide loss-of-function screen. The genome-wide knock out screen identified a set of host factors required for HCoV-229E replication and the identified host factors showed a mild to strong phenotype in the reduction of viral replication and virus-induced cell death. However, the genetic screen had some limitations which were very necessary to discuss to optimize the screen for the betterment in the future.

The genetic screen used in this study could not identify the genes essential for the cell cycle, growth, and metabolism in the Huh7 cells. These genes would have completely dropped off during the generation of a genome-wide pool of knock out cells and hence, if the virus might exploit these host factors for replication, it is unlikely to uncover these genes as hits.

The genetic screens cannot identify the host factors with redundant functions. If multiple genes had a redundant function in the viral replication, it is therefore highly possible that the genetic screen can never identify a single gene that is resistant to the infection. Also, it is almost impossible to confirm the correct permutation of gene hits for the validation at the required scale (< 100.000 sgRNAs). Additionally, this study lacked the screening for HCoV-229E resistant factors in multiple cell types including the human-airway epithelium cells (a cellular model closer to the natural environment for HCoV-229E replication) to understand the uniqueness and versatility of the genetic screen.

And finally, the validation experiments lacked in-depth orthogonal approaches to strongly emphasize the functional relevance of gene hits in the HCoV-229E replication and / or the thapsigargin-mediated anti-viral effect.

Taken together, this study has uncovered some of the critical host factors required for the replication of human coronaviruses, opening a new avenue for the development of potential anti-viral targets. The following summary figure comprises the list of host factors found in this study and a possible role in the coronavirus life cycle.

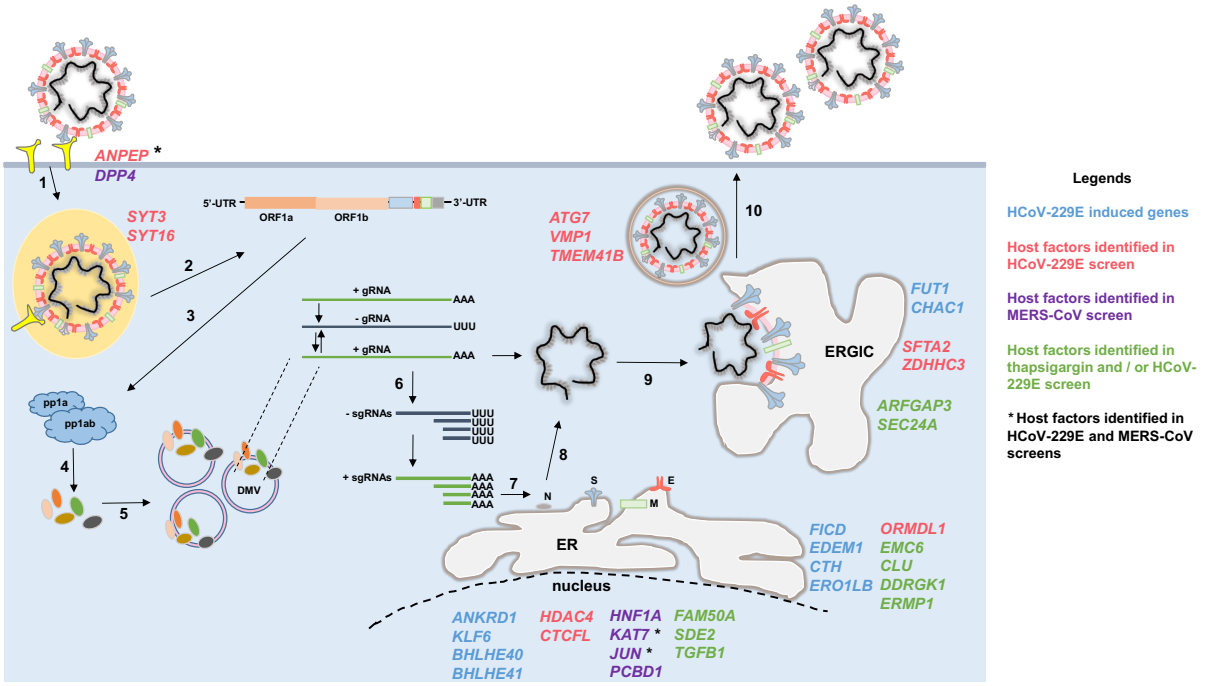


Figure 53: Summary of genes found in this study and their potential roles in the coronavirus life cycle.

The host factors that are identified in this study using the functional genomics techniques are presented adjacent to the stages of viral replication in the cellular compartments where they function. The genes are color-coded based on their identification in this study, as indicated in the legend. **1**. The coronavirus particle first binds to the specific cell membrane receptors (such as ANPEP, ACE2, and DPP4) by the spike (S) protein and enters the cell via the endosomes using the host factors (SYT3 and SYT16). **2**. By the specific cleavage action of the host-cell receptor and proteases, the viral membrane is uncoated from the endosomes and the viral genome is released into the cytoplasm. **3**. Following the release of the viral genome, the two large open-reading frames ORF1a and ORF1b translates to polyproteins pp1a and pp1b. **4**. The polypeptides are post-translationally processed to individual non-structural proteins (nsps) and form a protective micro-environment called the double-membrane vesicles (DMVs) (**5**) for the viral genome replication and transcription of subgenomic mRNAs (**6**). **7**. The subgenomic mRNAs are translated to form the structural proteins (S, E, M, and N) in the endoplasmic reticulum (ER). ER-localized host factors (FICD, EDEM1, CTH, ERO1LB, ORMDL1, EMC6, CLU, DDRGK1, and ERMP1) may play a role in the folding, assembly, and processing of the viral structural proteins. **8**. The nucleocapsid protein wraps up the full-length viral genome to form a capsid and assemble with the structural proteins that are translocated to ER-Golgi intermediate complex (ERGIC) (maybe using the host factors FUT1, CHAC1, SFTA2, ZDHHC3, ARFGAP3, and SEC24A) (**9**). **10**. The fully assembled virions are then secreted from the cells via exocytosis (using the host factors ATG7, VMP1, and TMEM41B for egress). Although the virus replication is localized to the cytoplasm, coronavirus can induce changes in the host cell transcriptional regulation (ANKRD1, KLF6, BHLHE40 / 41, HDAC4, CTCFL, HNF1A, KAT7, JUN, PCBD1, FAM50A, SDE2, and TGFB1) for subverting the host response for successful infection. UTR-Untranslated region, gRNA-genomic RNA, sgRNAs-subgenomic RNAs, AAA-poly-adenine tail, UUU-poly-uridine tail. The figure (coronavirus life cycle) was modified from the source [19] (Fig. 1 from the publication) and re-created using Microsoft PowerPoint for this thesis.

7. References

1. Almeida, J.D. and D.A.J. Tyrrell, *The Morphology of Three Previously Uncharacterized Human Respiratory Viruses that Grow in Organ Culture.* Journal of General Virology, 1967. **1**(2): p. 175-178.
2. Liu, D.X., J.Q. Liang, and T.S. Fung, *Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae)*. Encyclopedia of Virology. 2021:428-40. doi: 10.1016/B978-0-12-809633-8.21501-X. Epub 2021 Mar 1.
3. Hamre, D. and J.J. Procknow, *A new virus isolated from the human respiratory tract.* Proc Soc Exp Biol Med, 1966. **121**(1): p. 190-3.
4. Drosten, C., et al., *Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome.* New England Journal of Medicine, 2003. **348**(20): p. 1967-1976.
5. van der Hoek, L., et al., *Identification of a new human coronavirus.* Nat Med, 2004. **10**(4): p. 368-73.
6. Woo, P.C., et al., *Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia.* J Virol, 2005. **79**(2): p. 884-95.
7. Zaki, A.M., et al., *Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia.* New England Journal of Medicine, 2012. **367**(19): p. 1814-1820.
8. Corman, V.M., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.* Eurosurveillance, 2020. **25**(3): p. 2000045.
9. Vlasak, R., et al., *Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses.* Proc Natl Acad Sci U S A, 1988. **85**(12): p. 4526-9.
10. Artika, I.M., A.K. Dewantari, and A. Wiyatno, *Molecular biology of coronaviruses: current knowledge.* Heliyon, 2020. **6**(8): p. e04743.
11. Masters, P.S., *The Molecular Biology of Coronaviruses*, in *Advances in Virus Research*. 2006, Academic Press. p. 193-292.
12. Yeager, C.L., et al., *Human aminopeptidase N is a receptor for human coronavirus 229E.* Nature, 1992. **357**(6377): p. 420-422.
13. Li, W., et al., *Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus.* Nature, 2003. **426**(6965): p. 450-454.
14. Li, W., et al., *The S proteins of human coronavirus NL63 and severe acute respiratory syndrome coronavirus bind overlapping regions of ACE2.* Virology, 2007. **367**(2): p. 367-374.
15. Doremalen, N.v., et al., *Host Species Restriction of Middle East Respiratory Syndrome Coronavirus through Its Receptor, Dipeptidyl Peptidase 4.* Journal of Virology, 2014. **88**(16): p. 9220-9232.
16. Huang, X., et al., *Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme.* Journal of Virology, 2015. **89**(14): p. 7202-7213.
17. Butler, N., et al., *Murine encephalitis caused by HCoV-OC43, a human coronavirus with broad species specificity, is partly immune-mediated.* Virology, 2006. **347**(2): p. 410-421.
18. Lim, Y.X., et al., *Human Coronaviruses: A Review of Virus–Host Interactions.* Diseases, 2016. **4**(3): p. 26.
19. V'kovski, P., et al., *Coronavirus biology and replication: implications for SARS-CoV-2.* Nature Reviews Microbiology, 2021. **19**(3): p. 155-170.
20. de Wilde, A.H., et al., *Host Factors in Coronavirus Replication*, in *Roles of Host Gene and Non-coding RNA Expression in Virus Infection*, R.A. Tripp and S.M. Tompkins, Editors. 2018, Springer International Publishing: Cham. p. 1-42.
21. Fung, T.S. and D.X. Liu, *Human Coronavirus: Host-Pathogen Interaction.* Annual Review of Microbiology, 2019. **73**(1): p. 529-557.
22. Sola, I., et al., *Continuous and Discontinuous RNA Synthesis in Coronaviruses.* Annu Rev Virol, 2015. **2**(1): p. 265-88.
23. Sharma, A., et al., *Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies.* International Journal of Antimicrobial Agents, 2020. **56**(2): p. 106054.

24. Lamers, M.M. and B.L. Haagmans, *SARS-CoV-2 pathogenesis.* Nature Reviews Microbiology, 2022. **20**(5): p. 270-284.
25. Wu, D., et al., *The SARS-CoV-2 outbreak: What we know.* International Journal of Infectious Diseases, 2020. **94**: p. 44-48.
26. Fani, M., A. Teimoori, and S. Ghafari, *Comparison of the COVID-2019 (SARS-CoV-2) pathogenesis with SARS-CoV and MERS-CoV infections.* Future Virology, 2020. **15**(5): p. 317-323.
27. Guberina, H., et al., *A patient with severe respiratory failure caused by novel human coronavirus.* Infection, 2014. **42**(1): p. 203-206.
28. Dominguez, Samuel R., et al., *Blinded Case-Control Study of the Relationship between Human Coronavirus NL63 and Kawasaki Syndrome.* The Journal of Infectious Diseases, 2006. **194**(12): p. 1697-1701.
29. Murray, R.S., et al., *Detection of coronavirus RNA and antigen in multiple sclerosis brain.* Annals of Neurology, 1992. **31**(5): p. 525-533.
30. Kesheh, M.M., et al., *An overview on the seven pathogenic human coronaviruses.* Reviews in Medical Virology, 2022. **32**(2): p. e2282.
31. Sun, W., et al., *A severe case of human coronavirus 229E pneumonia in an elderly man with diabetes mellitus: a case report.* BMC Infectious Diseases, 2021. **21**(1): p. 524.
32. Vassilara, F., et al., *A Rare Case of Human Coronavirus 229E Associated with Acute Respiratory Distress Syndrome in a Healthy Adult.* Case Reports in Infectious Diseases, 2018. **2018**: p. 6796839.
33. Shah, M.D., et al., *A mini-review on the impact of COVID 19 on vital organs.* Biomedicine & Pharmacotherapy, 2021. **143**: p. 112158.
34. Li, G., et al., *Therapeutic strategies for COVID-19: progress and lessons learned.* Nature Reviews Drug Discovery, 2023. **22**(6): p. 449-475.
35. Liu, Y., et al., *The development of Coronavirus 3C-Like protease (3CLpro) inhibitors from 2010 to 2020.* European Journal of Medicinal Chemistry, 2020. **206**: p. 112711.
36. Lamb, Y.N., *Nirmatrelvir Plus Ritonavir: First Approval.* Drugs, 2022. **82**(5): p. 585-591.
37. Saravolatz, L.D., S. Depcinski, and M. Sharma, *Molnupiravir and Nirmatrelvir-Ritonavir: Oral Coronavirus Disease 2019 Antiviral Drugs.* Clinical Infectious Diseases, 2022. **76**(1): p. 165-171.
38. Marzolini, C., et al., *Recommendations for the Management of Drug–Drug Interactions Between the COVID-19 Antiviral Nirmatrelvir/Ritonavir (Paxlovid) and Comedications.* Clinical Pharmacology & Therapeutics, 2022. **112**(6): p. 1191-1200.
39. Adedeji, A.O. and S.G. Sarafianos, *Antiviral drugs specific for coronaviruses in preclinical development.* Current Opinion in Virology, 2014. **8**: p. 45-53.
40. Adedeji, A.O., et al., *Severe Acute Respiratory Syndrome Coronavirus Replication Inhibitor That Interferes with the Nucleic Acid Unwinding of the Viral Helicase.* Antimicrobial Agents and Chemotherapy, 2012. **56**(9): p. 4718-4728.
41. Adedeji, A.O., et al., *Evaluation of SSYA10-001 as a Replication Inhibitor of Severe Acute Respiratory Syndrome, Mouse Hepatitis, and Middle East Respiratory Syndrome Coronaviruses.* Antimicrobial Agents and Chemotherapy, 2014. **58**(8): p. 4894-4898.
42. Xu, X., et al., *Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design.* Nucleic Acids Research, 2003. **31**(24): p. 7117-7130.
43. Pruijssers, A.J. and M.R. Denison, *Nucleoside analogues for the treatment of coronavirus infections.* Current Opinion in Virology, 2019. **35**: p. 57-62.
44. Graci, J.D. and C.E. Cameron, *Mechanisms of action of ribavirin against distinct viruses.* Reviews in Medical Virology, 2006. **16**(1): p. 37-48.
45. Agrawal, U., R. Raju, and Z.F. Udwadia, *Favipiravir: A new and emerging antiviral option in COVID-19.* Medical Journal Armed Forces India, 2020. **76**(4): p. 370-376.
46. Joshi, S., et al., *Role of favipiravir in the treatment of COVID-19.* International Journal of Infectious Diseases, 2021. **102**: p. 501-508.
47. Cao, Y.-c., Q.-x. Deng, and S.-x. Dai, *Remdesivir for severe acute respiratory syndrome coronavirus 2 causing COVID-19: An evaluation of the evidence.* Travel Medicine and Infectious Disease, 2020. **35**: p. 101647.

48. Amirian, E.S. and J.K. Levy, *Current knowledge about the antivirals remdesivir (GS-5734) and GS-441524 as therapeutic options for coronaviruses.* One Health, 2020. **9**: p. 100128.
49. Singh, A.K., et al., *Remdesivir in COVID-19: A critical review of pharmacology, pre-clinical and clinical studies.* Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020. **14**(4): p. 641-648.
50. Singh, A.K., et al., *Molnupiravir in COVID-19: A systematic review of literature.* Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2021. **15**(6): p. 102329.
51. Painter, W.P., et al., *Human Safety, Tolerability, and Pharmacokinetics of Molnupiravir, a Novel Broad-Spectrum Oral Antiviral Agent with Activity against SARS-CoV-2.* Antimicrobial Agents and Chemotherapy, 2021. **65**(5): p. 10.1128/aac.02428-20.
52. Liu, M., et al., *Potential role of ACE2 in coronavirus disease 2019 (COVID-19) prevention and management.* Journal of Translational Internal Medicine, 2020. **8**(1): p. 9-19.
53. Wang, M., et al., *Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro.* Cell research, 2020. **30**(3): p. 269-271.
54. Shen, L.W., et al., *TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections.* Biochimie, 2017. **142**: p. 1-10.
55. Millet, J.K. and G.R. Whittaker, *Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis.* Virus Research, 2015. **202**: p. 120-134.
56. Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280.e8.
57. Liu, C., A. von Brunn, and D. Zhu, *Cyclophilin A and CD147: novel therapeutic targets for the treatment of COVID-19.* Medicine in Drug Discovery, 2020. **7**: p. 100056.
58. de Wilde, A.H., et al., *Cyclophilins and cyclophilin inhibitors in nidovirus replication.* Virology, 2018. **522**: p. 46-55.
59. de Wilde, A.H., et al., *Cyclosporin A inhibits the replication of diverse coronaviruses.* The Journal of general virology, 2011. **92**(Pt 11): p. 2542-2548.
60. Softic, L., et al., *Inhibition of SARS-CoV-2 Infection by the Cyclophilin Inhibitor Alisporivir (Debio 025).* Antimicrobial Agents and Chemotherapy, 2020. **64**(7): p. e00876-20.
61. He, B., *Viruses, endoplasmic reticulum stress, and interferon responses.* Cell Death & Differentiation, 2006. **13**(3): p. 393-403.
62. Shaban, M.S., et al., *Thapsigargin: key to new host-directed coronavirus antivirals?* Trends in Pharmacological Sciences, 2022. **43**(7): p. 557-568.
63. Al-Beltagi, S., et al., *Thapsigargin Is a Broad-Spectrum Inhibitor of Major Human Respiratory Viruses: Coronavirus, Respiratory Syncytial Virus and Influenza A Virus.* Viruses, 2021. **13**(2): p. 234.
64. Shaban, M.S., et al., *Multi-level inhibition of coronavirus replication by chemical ER stress.* Nature Communications, 2021. **12**(1): p. 5536.
65. Hsu, Patrick D., Eric S. Lander, and F. Zhang, *Development and Applications of CRISPR-Cas9 for Genome Engineering.* Cell, 2014. **157**(6): p. 1262-1278.
66. Barrangou, R. and Luciano A. Marraffini, *CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity.* Molecular Cell, 2014. **54**(2): p. 234-244.
67. Makarova, K.S., et al., *Evolution and classification of the CRISPR–Cas systems.* Nature Reviews Microbiology, 2011. **9**(6): p. 467-477.
68. Chylinski, K., et al., *Classification and evolution of type II CRISPR-Cas systems.* Nucleic Acids Research, 2014. **42**(10): p. 6091-6105.
69. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Current Opinion in Microbiology, 2017. **37**: p. 67-78.
70. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR–Cas systems.* Nature Reviews Microbiology, 2015. **13**(11): p. 722-736.
71. Pickar-Oliver, A. and C.A. Gersbach, *The next generation of CRISPR–Cas technologies and applications.* Nature Reviews Molecular Cell Biology, 2019. **20**(8): p. 490-507.
72. Wang, H., M. La Russa, and L.S. Qi, *CRISPR/Cas9 in Genome Editing and Beyond.* Annu Rev Biochem, 2016. **85**: p. 227-64.
73. Doudna, J.A. and E. Charpentier, *The new frontier of genome engineering with CRISPR-Cas9.* Science, 2014. **346**(6213): p. 1258096.

74. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature Protocols, 2013. **8**(11): p. 2281-2308.
75. Lees-Miller, S.P. and K. Meek, *Repair of DNA double strand breaks by non-homologous end joining.* Biochimie, 2003. **85**(11): p. 1161-1173.
76. Ragu, S., et al., *Chapter 20 - Homologous recombination in mammalian cells: From molecular mechanisms to pathology*, in *Genome Stability (Second Edition)*, I. Kovalchuk and O. Kovalchuk, Editors. 2021, Academic Press: Boston. p. 367-392.
77. Kampmann, M., *CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine.* ACS Chemical Biology, 2018. **13**(2): p. 406-416.
78. Barrangou, R., et al., *Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference.* Nucleic Acids Research, 2015. **43**(7): p. 3407-3419.
79. Dominguez, A.A., W.A. Lim, and L.S. Qi, *Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation.* Nature Reviews Molecular Cell Biology, 2016. **17**(1): p. 5-15.
80. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-588.
81. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science (New York, N.Y.), 2014. **343**(6166): p. 84-87.
82. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nature Methods, 2014. **11**(8): p. 783-784.
83. Wang, T., et al., *Identification and characterization of essential genes in the human genome.* Science, 2015. **350**(6264): p. 1096-1101.
84. Doench, J.G., et al., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.* Nature Biotechnology, 2016. **34**(2): p. 184-191.
85. Peng, J., et al., *High-throughput screens in mammalian cells using the CRISPR-Cas9 system.* The FEBS Journal, 2015. **282**(11): p. 2089-2096.
86. Mátrai, J., M.K.L. Chuah, and T. VandenDriessche, *Recent Advances in Lentiviral Vector Development and Applications.* Molecular Therapy, 2010. **18**(3): p. 477-490.
87. Milone, M.C. and U. O'Doherty, *Clinical use of lentiviral vectors.* Leukemia, 2018. **32**(7): p. 1529-1541.
88. Sakuma, T., Michael A. Barry, and Y. Ikeda, *Lentiviral vectors: basic to translational.* Biochemical Journal, 2012. **443**(3): p. 603-618.
89. Joung, J., et al., *Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening.* Nat Protoc, 2017. **12**(4): p. 828-863.
90. Miles, L.A., R.J. Garippa, and J.T. Poirier, *Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens.* The FEBS Journal, 2016. **283**(17): p. 3170-3180.
91. Doench, J.G., *Am I ready for CRISPR? A user's guide to genetic screens.* Nature Reviews Genetics, 2018. **19**(2): p. 67-80.
92. So, R.W.L., et al., *Application of CRISPR genetic screens to investigate neurological diseases.* Molecular Neurodegeneration, 2019. **14**(1): p. 41.
93. Xue, H.-Y., et al., *CRISPR-Cas9 for medical genetic screens: applications and future perspectives.* Journal of Medical Genetics, 2016. **53**(2): p. 91-97.
94. Bock, C., et al., *High-content CRISPR screening.* Nature Reviews Methods Primers, 2022. **2**(1): p. 8.
95. Sharma, S. and E. Petsalaki, *Application of CRISPR-Cas9 Based Genome-Wide Screening Approaches to Study Cellular Signalling Mechanisms.* International journal of molecular sciences, 2018. **19**(4): p. 933.
96. Bodapati, S., et al., *A benchmark of algorithms for the analysis of pooled CRISPR screens.* Genome Biology, 2020. **21**(1): p. 62.
97. König, R., et al., *A probability-based approach for the analysis of large-scale RNAi screens.* Nature Methods, 2007. **4**(10): p. 847-849.
98. Colic, M. and T. Hart, *Common computational tools for analyzing CRISPR screens.* Emerging topics in life sciences, 2021. **5**(6): p. 779-788.
99. Li, W., et al., *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.* Genome Biology, 2014. **15**(12): p. 554.

100. Hart, T. and J. Moffat, *BAGEL: a computational framework for identifying essential genes from pooled library screens.* BMC Bioinformatics, 2016. **17**(1): p. 164.
101. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-2308.
102. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-784.
103. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nature Methods, 2012. **9**(7): p. 671-675.
104. Zhou, Y., et al., *Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.* Nat Commun, 2019. **10**(1): p. 1523.
105. Chung, C.T. and R.H. Miller, *[43] Preparation and storage of competent Escherichia coli cells*, in *Methods in Enzymology*, R. Wu, Editor. 1993, Academic Press. p. 621-627.
106. Reed, L.J. and H. Muench, *A simple method of estimating fifty per cent endpoints.* American journal of epidemiology, 1938. **27**(3): p. 493-497.
107. Baer, A. and K. Kehn-Hall, *Viral concentration determination through plaque assays: using traditional and novel overlay systems.* JoVE (Journal of Visualized Experiments), 2014(93): p. e52065.
108. Sambrook, J. and D.W. Russell, *SDS-Polyacrylamide Gel Electrophoresis of Proteins.* Cold Spring Harbor Protocols, 2006. **2006**(4): p. pdb.prot4540.
109. Mukherjee, S., et al., *Large-scale contamination of microbial isolate genomes by Illumina PhiX control.* Standards in Genomic Sciences, 2015. **10**(1): p. 18.
110. Davis, H.E., J.R. Morgan, and M.L. Yarmush, *Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes.* Biophysical Chemistry, 2002. **97**(2): p. 159-172.
111. Li, W., et al., *Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR.* Genome Biology, 2015. **16**(1): p. 281.
112. Wang, B., et al., *Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute.* Nature Protocols, 2019. **14**(3): p. 756-780.
113. Poppe, M., et al., *The NF-κB-dependent and -independent transcriptome and chromatin landscapes of human coronavirus 229E-infected cells.* PLOS Pathogens, 2017. **13**(3): p. e1006286.
114. Lee, S., R. Channappanavar, and T.-D. Kanneganti, *Coronaviruses: Innate Immunity, Inflammasome Activation, Inflammatory Cell Death, and Cytokines.* Trends in Immunology, 2020. **41**(12): p. 1083-1099.
115. Wright Muelas, M., et al., *The role and robustness of the Gini coefficient as an unbiased tool for the selection of Gini genes for normalising expression profiling data.* Scientific Reports, 2019. **9**(1): p. 17960.
116. Wang, R., et al., *Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses.* Cell, 2021. **184**(1): p. 106-119 e14.
117. Schneider, W.M., et al., *Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks.* Cell, 2021. **184**(1): p. 120-132 e14.
118. Hoffmann, H.H., et al., *Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors.* Cell Host & Microbe, 2021. **29**(2): p. 267-280.e5.
119. Kratzel, A., et al., *A genome-wide CRISPR screen identifies interactors of the autophagy pathway as conserved coronavirus targets.* bioRxiv, 2021: p. 2021.02.24.432634.
120. Wei, J., et al., *Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection.* Cell, 2021. **184**(1): p. 76-91.e13.
121. Liu, J., et al., *Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin.* Proc Natl Acad Sci U S A, 2012. **109**(2): p. 413-8.
122. Nagelreiter, F., et al., *OPP Labeling Enables Total Protein Synthesis Quantification in CHO Production Cell Lines at the Single-Cell Level.* Biotechnology Journal, 2018. **13**(4): p. 1700492.
123. Fung, T.S. and D.X. Liu, *Coronavirus infection, ER stress, apoptosis and innate immunity.* Frontiers in Microbiology, 2014. **5**.

124. Chidawanyika, T., et al., *SEC24A identified as an essential mediator of thapsigargin-induced cell death in a genome-wide CRISPR/Cas9 screen.* Cell Death Discovery, 2018. **4**(1): p. 115.
125. Vidalain, P.-O. and F. Tangy, *Virus-host protein interactions in RNA viruses.* Microbes and Infection, 2010. **12**(14): p. 1134-1143.
126. Ahlquist, P., et al., *Host Factors in Positive-Strand RNA Virus Genome Replication.* Journal of Virology, 2003. **77**(15): p. 8181-8186.
127. Shi, S.T. and M.M.C. Lai, *Viral and Cellular Proteins Involved in Coronavirus Replication*, in *Coronavirus Replication and Reverse Genetics*, L. Enjuanes, Editor. 2005, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 95-131.
128. Strauss, J.H. and E.G. Strauss, *With a Little Help from the Host.* Science, 1999. **283**(5403): p. 802-804.
129. Nagy, P.D. and J. Pogany, *The dependence of viral RNA replication on co-opted host factors.* Nature Reviews Microbiology, 2012. **10**(2): p. 137-149.
130. Zhang, S., et al., *Comparison of viral RNA–host protein interactomes across pathogenic RNA viruses informs rapid antiviral drug discovery for SARS-CoV-2.* Cell Research, 2022. **32**(1): p. 9-23.
131. Bechill, J., et al., *Coronavirus Infection Modulates the Unfolded Protein Response and Mediates Sustained Translational Repression.* Journal of Virology, 2008. **82**(9): p. 4492-4501.
132. Jain, R., et al., *Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical outcomes.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 153-160.
133. Tang, B.S.F., et al., *Comparative Host Gene Transcription by Microarray Analysis Early after Infection of the Huh7 Cell Line by Severe Acute Respiratory Syndrome Coronavirus and Human Coronavirus 229E.* Journal of Virology, 2005. **79**(10): p. 6180-6193.
134. Fung, T.S., M. Huang, and D.X. Liu, *Coronavirus-induced ER stress response and its involvement in regulation of coronavirus–host interactions.* Virus Research, 2014. **194**: p. 110-123.
135. Schmitz, M.L., et al., *The Crosstalk of Endoplasmic Reticulum (ER) Stress Pathways with NF-κB: Complex Mechanisms Relevant for Cancer, Inflammation and Infection.* Biomedicines, 2018. **6**(2): p. 58.
136. Byun, H., et al., *ERAD and how viruses exploit it.* Frontiers in Microbiology, 2014. **5**.
137. Miller, K., et al., *Coronavirus interactions with the cellular autophagy machinery.* Autophagy, 2020. **16**(12): p. 2131-2139.
138. Reggiori, F., et al., *Coronaviruses Hijack the LC3-I-Positive EDEMosomes, ER-Derived Vesicles Exporting Short-Lived ERAD Regulators, for Replication.* Cell Host & Microbe, 2010. **7**(6): p. 500-508.
139. Perrin-Cocon, L., et al., *The current landscape of coronavirus-host protein–protein interactions.* Journal of Translational Medicine, 2020. **18**(1): p. 319.
140. Sanchez, E.L. and M. Lagunoff, *Viral activation of cellular metabolism.* Virology, 2015. **479-480**: p. 609-18.
141. Xu, L.H., et al., *Coronavirus infection induces DNA replication stress partly through interaction of its nonstructural protein 13 with the p125 subunit of DNA polymerase δ.* J Biol Chem, 2011. **286**(45): p. 39546-59.
142. Yan, B., et al., *Characterization of the Lipidomic Profile of Human Coronavirus-Infected Cells: Implications for Lipid Metabolism Remodeling upon Coronavirus Replication.* Viruses, 2019. **11**(1): p. 73.
143. Jackson, R.J., C.U.T. Hellen, and T.V. Pestova, *The mechanism of eukaryotic translation initiation and principles of its regulation.* Nature reviews. Molecular cell biology, 2010. **11**(2): p. 113-127.
144. Schubert, K., et al., *SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation.* Nature Structural & Molecular Biology, 2020. **27**(10): p. 959-966.
145. Ibrahim, I.M., D.H. Abdelmalek, and A.A. Elfiky, *GRP78: A cell's response to stress.* Life Sciences, 2019. **226**: p. 156-163.
146. Pfaffenbach, K.T. and A.S. Lee, *The critical role of GRP78 in physiologic and pathologic stress.* Current Opinion in Cell Biology, 2011. **23**(2): p. 150-156.

147. Elfiky, A.A., et al., *GRP78 targeting: Hitting two birds with a stone.* Life Sciences, 2020. **260**: p. 118317.
148. Shin, J., et al., *GRP78, a Novel Host Factor for SARS-CoV-2: The Emerging Roles in COVID-19 Related to Metabolic Risk Factors.* Biomedicines, 2022. **10**(8): p. 1995.
149. Shahriari Felordi, M., et al., *Is There any Alternative Receptor for SARS-CoV-2?* Cell journal, 2021. **23**(2): p. 247-250.
150. Ha, D.P., et al., *The stress-inducible molecular chaperone GRP78 as potential therapeutic target for coronavirus infection.* Journal of Infection, 2020. **81**(3): p. 452-482.
151. Chu, H., et al., *Middle East respiratory syndrome coronavirus and bat coronavirus HKU9 both can utilize GRP78 for attachment onto host cells.* Journal of Biological Chemistry, 2018. **293**(30): p. 11709-11726.
152. Liu, X.-H., W.A. Bauman, and C. Cardozo, *ANKRD1 modulates inflammatory responses in C2C12 myoblasts through feedback inhibition of NF-κB signaling activity.* Biochemical and Biophysical Research Communications, 2015. **464**(1): p. 208-213.
153. Masilamani, A.P., et al., *KLF6 depletion promotes NF-κB signaling in glioblastoma.* Oncogene, 2017. **36**(25): p. 3562-3575.
154. Sano, R. and J.C. Reed, *ER stress-induced cell death mechanisms.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2013. **1833**(12): p. 3460-3470.
155. Rozpedek, W., et al., *The Role of the PERK/eIF2α/ATF4/CHOP Signaling Pathway in Tumor Progression During Endoplasmic Reticulum Stress.* Current molecular medicine, 2016. **16**(6): p. 533-544.
156. Liu, Z., et al., *Activating Transcription Factor 4 (ATF4)-ATF3-C/EBP Homologous Protein (CHOP) Cascade Shows an Essential Role in the ER Stress-Induced Sensitization of Tetrachlorobenzoquinone-Challenged PC12 Cells to ROS-Mediated Apoptosis via Death Receptor 5 (DR5) Signaling.* Chemical Research in Toxicology, 2016. **29**(9): p. 1510-1518.
157. Hu, H., et al., *The C/EBP Homologous Protein (CHOP) Transcription Factor Functions in Endoplasmic Reticulum Stress-Induced Apoptosis and Microbial Infection.* Frontiers in Immunology, 2019. **9**.
158. Liao, Y., et al., *Upregulation of CHOP/GADD153 during Coronavirus Infectious Bronchitis Virus Infection Modulates Apoptosis by Restricting Activation of the Extracellular Signal-Regulated Kinase Pathway.* Journal of Virology, 2013. **87**(14): p. 8124-8134.
159. Caselli, E., et al., *Activating transcription factor 4 (ATF4) is upregulated by human herpesvirus 8 infection, increases virus replication and promotes proangiogenic properties.* Archives of Virology, 2012. **157**(1): p. 63-74.
160. Qian, Z., et al., *Murine cytomegalovirus targets transcription factor ATF4 to exploit the unfolded-protein response.* Journal of virology, 2012. **86**(12): p. 6712-6723.
161. Ríos-Ocampo, W.A., et al., *Hepatitis C Virus Proteins Core and NS5A Are Highly Sensitive to Oxidative Stress-Induced Degradation after eIF2α/ATF4 Pathway Activation.* Viruses, 2020. **12**(4): p. 425.
162. Rosa-Fernandes, L., et al., *SARS-CoV-2 activates ER stress and Unfolded protein response.* bioRxiv, 2021: p. 2021.06.21.449284.
163. Mungrue, I.N., et al., *CHAC1/MGC4504 Is a Novel Proapoptotic Component of the Unfolded Protein Response, Downstream of the ATF4-ATF3-CHOP Cascade1.* The Journal of Immunology, 2009. **182**(1): p. 466-476.
164. Kasai, F., et al., *HuH-7 reference genome profile: complex karyotype composed of massive loss of heterozygosity.* Human cell, 2018. **31**(3): p. 261-267.
165. Than, T.T., et al., *Ankyrin Repeat Domain 1 is Up-regulated During Hepatitis C Virus Infection and Regulates Hepatitis C Virus Entry.* Scientific reports, 2016. **6**: p. 20819-20819.
166. Bin, L., et al., *Ankyrin repeat domain 1 regulates innate immune responses against herpes simplex virus 1: A potential role in eczema herpeticum.* The Journal of allergy and clinical immunology, 2018. **141**(6): p. 2085-2093.e1.
167. Bernasconi, R., J. Noack, and M. Molinari, *Unconventional roles of nonlipidated LC3 in ERAD tuning and coronavirus infection.* Autophagy, 2012. **8**(10): p. 1534-1536.
168. Zou, L., et al., *Viruses Hijack ERAD to Regulate Their Replication and Propagation.* International Journal of Molecular Sciences, 2022. **23**(16): p. 9398.

169. Chiritoiu, M., et al., *EDEM1 Drives Misfolded Protein Degradation via ERAD and Exploits ER-Phagy as Back-Up Mechanism When ERAD Is Impaired.* International Journal of Molecular Sciences, 2020. **21**(10): p. 3468.
170. Zhang, H.M., et al., *Cleavage and degradation of EDEM1 promotes coxsackievirus B3 replication via ATF6a-mediated unfolded protein response signalling.* Cellular Microbiology, 2020. **22**(7): p. e13198.
171. Früh, K., et al., *Virogenomics: a novel approach to antiviral drug discovery.* Drug Discovery Today, 2001. **6**(12): p. 621-627.
172. DeFilippis, V., et al., *Functional genomics in virology and antiviral drug discovery.* Trends in Biotechnology, 2003. **21**(10): p. 452-457.
173. Perreira, J.M., P. Meraner, and A.L. Brass, *Functional Genomic Strategies for Elucidating Human-Virus Interactions: Will CRISPR Knockout RNAi and Haploid Cells?* Advances in virus research, 2016. **94**: p. 1-51.
174. Mohr, S.E. and N. Perrimon, *RNAi screening: new approaches, understandings, and organisms.* WIREs RNA, 2012. **3**(2): p. 145-158.
175. Carette, J.E., et al., *Haploid Genetic Screens in Human Cells Identify Host Factors Used by Pathogens.* Science, 2009. **326**(5957): p. 1231-1235.
176. Evers, B., et al., *CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes.* Nature Biotechnology, 2016. **34**(6): p. 631-633.
177. Puschnik, A.S., et al., *A CRISPR toolbox to study virus–host interactions.* Nature Reviews Microbiology, 2017. **15**(6): p. 351-364.
178. Ramage, H. and S. Cherry, *Virus-Host Interactions: From Unbiased Genetic Screens to Function.* Annual Review of Virology, 2015. **2**(1): p. 497-524.
179. Shahi, I., et al., *Genome-Wide CRISPR-Cas9 Screen Does Not Identify Host Factors Modulating Streptococcus agalactiae ßHC;-Hemolysin/Cytolysin-Induced Cell Death.* Microbiology Spectrum, 2022. **10**(1): p. e02186-21.
180. Bertram, S., et al., *TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium.* Journal of virology, 2013. **87**(11): p. 6150-6160.
181. Shirato, K., et al., *Clinical Isolates of Human Coronavirus 229E Bypass the Endosome for Cell Entry.* Journal of Virology, 2017. **91**(1): p. e01387-16.
182. Kawase, M., et al., *Protease-Mediated Entry via the Endosome of Human Coronavirus 229E.* Journal of Virology, 2009. **83**(2): p. 712-721.
183. Jeffers, S.A., E.M. Hemmila, and K.V. Holmes. *Human Coronavirus 229E can Use CD209L (L-Sign) to Enter Cells*. in *The Nidoviruses*. 2006. Boston, MA: Springer US.
184. Mittal, R.A., et al., *SFTA2—A Novel Secretory Peptide Highly Expressed in the Lung—Is Modulated by Lipopolysaccharide but Not Hyperoxia.* PLOS ONE, 2012. **7**(6): p. e40011.
185. Hammel, M., et al., *SFTA2 - A Novel Pulmonary Surfactant Protein Secreted By Type Ii Cells - is Downregulated in LPS-Induced Lung Inflammation.* Pediatric Research, 2011. **70**(5): p. 44-44.
186. Krygier, A., et al., *Molecular Pathogenesis of Fibrosis, Thrombosis and Surfactant Dysfunction in the Lungs of Severe COVID-19 Patients.* Biomolecules, 2022. **12**(12): p. 1845.
187. Li, D., et al., *Insights Gained Into the Treatment of COVID19 by Pulmonary Surfactant and Its Components.* Frontiers in Immunology, 2022. **13**.
188. Herbein, G. and D. Wendling, *Histone deacetylases in viral infections.* Clinical epigenetics, 2010. **1**(1-2): p. 13-24.
189. Nagesh, P.T., et al., *Histone Deacetylase 2 Is a Component of Influenza A Virus-Induced Host Antiviral Response.* Frontiers in Microbiology, 2017. **8**.
190. Vázquez-Calvo, Á., et al., *Inhibition of Enveloped Virus Infection of Cultured Cells by Valproic Acid.* Journal of Virology, 2011. **85**(3): p. 1267-1274.
191. Bhargava, P., et al., *Repurposing valproate to prevent acute respiratory distress syndrome/acute lung injury in COVID-19: A review of immunomodulatory action.* Cancer Research, Statistics, and Treatment, 2020. **3**(Suppl 1).
192. Pitt, B., et al., *Potential repurposing of the HDAC inhibitor valproic acid for patients with COVID-19.* European Journal of Pharmacology, 2021. **898**: p. 173988.

193. Collier, J.J., et al., *Emerging roles of ATG7 in human health and disease.* EMBO Molecular Medicine, 2021. **13**(12): p. e14824.
194. Maier, H.J. and P. Britton, *Involvement of Autophagy in Coronavirus Replication.* Viruses, 2012. **4**(12): p. 3440-3451.
195. Movaqar, A., et al., *Coronaviruses construct an interconnection way with ERAD and autophagy.* Future Microbiology, 2021. **16**(14): p. 1135-1151.
196. Yang, N. and H.-M. Shen, *Targeting the Endocytic Pathway and Autophagy Process as a Novel Therapeutic Strategy in COVID-19.* International journal of biological sciences, 2020. **16**(10): p. 1724-1731.
197. Fung, T.S. and D.X. Liu, *Post-translational modifications of coronavirus proteins: roles and function.* Future Virology, 2018. **13**(6): p. 405-430.
198. Li, D., et al., *Palmitoylation of SARS-CoV-2 S protein is critical for S-mediated syncytia formation and virus entry.* Journal of Medical Virology, 2022. **94**(1): p. 342-348.
199. Williams, C.G., et al., *Inhibitors of VPS34 and fatty-acid metabolism suppress SARS-CoV-2 replication.* Cell reports, 2021. **36**(5): p. 109479-109479.
200. Mekhail, K., et al., *FASN inhibitor TVB-3166 prevents S-acylation of the spike protein of human coronaviruses.* Journal of Lipid Research, 2022. **63**(9): p. 100256.
201. Matteucci, E. and O. Giampietro, *Dipeptidyl Peptidase-4 (CD26): Knowing the Function before Inhibiting the Enzyme.* Current Medicinal Chemistry, 2009. **16**(23): p. 2943-2951.
202. Mina-Osorio, P., *The moonlighting enzyme CD13: old and new functions to target.* Trends in Molecular Medicine, 2008. **14**(8): p. 361-371.
203. Shirato, K., M. Kawase, and S. Matsuyama, *Middle East Respiratory Syndrome Coronavirus Infection Mediated by the Transmembrane Serine Protease TMPRSS2.* Journal of Virology, 2013. **87**(23): p. 12552-12561.
204. Iwata-Yoshikawa, N., et al., *TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection.* Journal of Virology, 2019. **93**(6): p. e01815-18.
205. Yamamoto, M., et al., *Identification of Nafamostat as a Potent Inhibitor of Middle East Respiratory Syndrome Coronavirus S Protein-Mediated Membrane Fusion Using the Split-Protein-Based Cell-Cell Fusion Assay.* Antimicrobial Agents and Chemotherapy, 2016. **60**(11): p. 6532-6539.
206. Mahoney, M., et al., *A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells.* Proceedings of the National Academy of Sciences, 2021. **118**(43): p. e2108728118.
207. Yu, M., et al., *Proteomic screen defines the hepatocyte nuclear factor 1α-binding partners and identifies HMGB1 as a new cofactor of HNF1α.* Nucleic Acids Research, 2007. **36**(4): p. 1209-1219.
208. Bockenhauer, D. and G. Jaureguiberry, *HNF1B-associated clinical phenotypes: the kidney and beyond.* Pediatric Nephrology, 2016. **31**(5): p. 707-714.
209. Viering, D.H.H.M., et al., *Genetic causes of hypomagnesemia, a clinical overview.* Pediatric Nephrology, 2017. **32**(7): p. 1123-1135.
210. Ferrè, S., et al., *Mutations in PCBD1 Cause Hypomagnesemia and Renal Magnesium Wasting.* Journal of the American Society of Nephrology, 2014. **25**(3): p. 574-586.
211. Faa, G., et al., *Association between Hypomagnesemia, COVID-19, Respiratory Tract and Lung Disease.* The open respiratory medicine journal, 2021. **15**: p. 43-45.
212. Baggen, J., et al., *Cellular host factors for SARS-CoV-2 infection.* Nature Microbiology, 2021. **6**(10): p. 1219-1232.
213. Baggen, J., et al., *Genome-wide CRISPR screening identifies TMEM106B as a proviral host factor for SARS-CoV-2.* Nature Genetics, 2021. **53**(4): p. 435-444.
214. Trimarco, J.D., et al., *TMEM41B is a host factor required for the replication of diverse coronaviruses including SARS-CoV-2.* PLOS Pathogens, 2021. **17**(5): p. e1009599.
215. Rebendenne, A., et al., *Bidirectional genome-wide CRISPR screens reveal host factors regulating SARS-CoV-2, MERS-CoV and seasonal HCoVs.* Nature Genetics, 2022. **54**(8): p. 1090-1102.

216. Grodzki, M., et al., *Genome-scale CRISPR screens identify host factors that promote human coronavirus infection.* Genome Medicine, 2022. **14**(1): p. 10.
217. Liu, J., et al., *A critical role of DDRGK1 in endoplasmic reticulum homoeostasis via regulation of IRE1α stability.* Nature Communications, 2017. **8**(1): p. 14186.
218. Trougakos, I.P., *The Molecular Chaperone Apolipoprotein J/Clusterin as a Sensor of Oxidative Stress: Implications in Therapeutic Approaches - A Mini-Review.* Gerontology, 2013. **59**(6): p. 514-523.
219. Shiba, Y. and P.A. Randazzo, *ArfGAPs: key regulators for receptor sorting.* Receptors & clinical investigation, 2014. **1**(5): p. e158-e158.
220. Chitwood, P.J. and R.S. Hegde, *The Role of EMC during Membrane Protein Biogenesis.* Trends in Cell Biology, 2019. **29**(5): p. 371-384.
221. Grandi, A., et al., *ERMP1, a novel potential oncogene involved in UPR and oxidative stress defense, is highly expressed in human cancer.* Oncotarget, 2016. **7**(39): p. 63596-63610.
222. Jensen, D. and R. Schekman, *COPII-mediated vesicle formation at a glance.* Journal of Cell Science, 2011. **124**(1): p. 1-4.
223. Shiba, Y., et al., *ArfGAP3 regulates the transport of cation-independent mannose 6-phosphate receptor in the post-Golgi compartment.* Current biology : CB, 2013. **23**(19): p. 1945-1951.
224. Weimer, C., et al., *Differential roles of ArfGAP1, ArfGAP2, and ArfGAP3 in COPI trafficking.* The Journal of cell biology, 2008. **183**(4): p. 725-735.
225. Saitoh, A., et al., *Three Homologous ArfGAPs Participate in Coat Protein I-mediated Transport*.* Journal of Biological Chemistry, 2009. **284**(20): p. 13948-13957.
226. Béthune, J., F. Wieland, and J. Moelleken, *COPI-mediated Transport.* The Journal of Membrane Biology, 2006. **211**(2): p. 65-79.
227. Robinson, M., et al., *Viral journeys on the intracellular highways.* Cellular and Molecular Life Sciences, 2018. **75**(20): p. 3693-3714.
228. Li, H., et al., *Hepatitis C Virus NS5A Hijacks ARFGAP1 To Maintain a Phosphatidylinositol 4-Phosphate-Enriched Microenvironment.* Journal of Virology, 2014. **88**(11): p. 5956-5966.
229. Zhang, N. and L. Zhang, *Key components of COPI and COPII machineries are required for chikungunya virus replication.* Biochem Biophys Res Commun, 2017. **493**(3): p. 1190-1196.
230. Ghosh, S., et al., *β-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway.* Cell, 2020. **183**(6): p. 1520-1535.e14.
231. Hegde, R.S. and R.J. Keenan, *The mechanisms of integral membrane protein biogenesis.* Nature Reviews Molecular Cell Biology, 2022. **23**(2): p. 107-124.
232. Volkmar, N. and J.C. Christianson, *Squaring the EMC – how promoting membrane protein biogenesis impacts cellular functions and organismal homeostasis.* Journal of Cell Science, 2020. **133**(8).
233. Volkmar, N., et al., *The ER membrane protein complex promotes biogenesis of sterol-related enzymes maintaining cholesterol homeostasis.* Journal of Cell Science, 2019. **132**(2).
234. Savidis, G., et al., *Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics.* Cell Reports, 2016. **16**(1): p. 232-246.
235. Ma, H., et al., *A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death.* Cell Reports, 2015. **12**(4): p. 673-683.
236. Lin, D.L., et al., *The ER Membrane Protein Complex Promotes Biogenesis of Dengue and Zika Virus Non-structural Multi-pass Transmembrane Proteins to Support Infection.* Cell Reports, 2019. **27**(6): p. 1666-1674.e4.
237. Zhao, C., et al., *CRISPR screening of porcine sgRNA library identifies host factors associated with Japanese encephalitis virus replication.* Nature Communications, 2020. **11**(1): p. 5178.
238. Nizard, P., et al., *Stress-Induced Retrotranslocation of Clusterin/ApoJ into the Cytosol.* Traffic, 2007. **8**(5): p. 554-565.
239. Park, J., et al., *Hypoxia inducible factor-1α directly regulates nuclear clusterin transcription by interacting with hypoxia response elements in the clusterin promoter.* Mol Cells, 2014. **37**(2): p. 178-86.
240. Trougakos, I.P., et al., *Intracellular Clusterin Inhibits Mitochondrial Apoptosis by Suppressing p53-Activating Stress Signals and Stabilizing the Cytosolic Ku70-Bax Protein Complex.* Clinical Cancer Research, 2008. **15**(1): p. 48-59.

241. Wyatt, A.R., et al., *Clusterin facilitates in vivo clearance of extracellular misfolded proteins.* Cell Mol Life Sci, 2011. **68**(23): p. 3919-31.
242. Kang, S.-W., et al., *Unglycosylated clusterin variant accumulates in the endoplasmic reticulum and induces cytotoxicity.* The International Journal of Biochemistry & Cell Biology, 2013. **45**(2): p. 221-231.
243. Rodríguez-Rivera, C., et al., *Clusterin: Always protecting. Synthesis, function and potential issues.* Biomedicine & Pharmacotherapy, 2021. **134**: p. 111174.
244. Barlowe, C., et al., *COPII: A membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum.* Cell, 1994. **77**(6): p. 895-907.
245. Pagano, A., et al., *Sec24 Proteins and Sorting at the Endoplasmic Reticulum*.* Journal of Biological Chemistry, 1999. **274**(12): p. 7833-7840.
246. Miller, E., et al., *Cargo selection into COPII vesicles is driven by the Sec24p subunit.* The EMBO Journal, 2002. **21**(22): p. 6105-6113.
247. Zeyen, L., et al., *Hepatitis B subviral envelope particles use the COPII machinery for intracellular transport via selective exploitation of Sec24A and Sec23B.* Cellular Microbiology, 2020. **22**(6): p. e13181.
248. Yamayoshi, S., et al., *Ebola Virus Matrix Protein VP40 Uses the COPII Transport System for Its Intracellular Transport.* Cell Host & Microbe, 2008. **3**(3): p. 168-177.
249. Cui, Y., et al., *A COPII subunit acts with an autophagy receptor to target endoplasmic reticulum for degradation.* Science, 2019. **365**(6448): p. 53-60.
250. Chidawanyika, T., et al., *SEC24A facilitates colocalization and Ca2+ flux between the endoplasmic reticulum and mitochondria.* Journal of Cell Science, 2021. **134**(6).

List of abbreviations

°C	Degree Celsius
ARDS	Acute Respiratory Distress Syndrome
as	Antisense
ATP	Adenosine triphosphate
bp	Base pairs
BSA	Bovine serum albumin
Cas9	CRISPR-associated protein 9
CoV	Coronaviruses
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Da	Dalton
ddH2O	Double distilled water
DMEM	Dulbecco's modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DMV	Double-membrane vesicle
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
DSB	Double-Strand Breaks
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ERAD	Endoplasmic-reticulum-associated protein degradation
ERGIC	Endoplasmic reticulum-Golgi Intermediate Complex
FBS	Fetal bovine serum
Fig.	Figure
GeCKO	Genome-wide CRISPR-Cas9 Knock out
GSEA	Gene Set Enrichment Analysis
HDR	Homology Directed Repair
HE	Hemagglutinin esterase
INDEL	Insertions/Deletions
kDa	kilo Dalton
KO	Knock out
LTR	Long terminal repeats
LCV2	LentiCRISPR Version2
LV	Lentivirus
MAGeCK	Model-based Analysis of Genome-wide CRISPR-Cas9 Knock out
MERS	Middle East Respiratory Syndrome
MHV	Murine hepatitis virus
mM	Millimolar
MOI	Multiplicity of infection
ng	Nanogram
NGS	Next generation sequencing
NHEJ	Non-homologous End Joining
NSP, nsp	Nonstructural protein
ORF	Open reading frame

PAGE	Poly-acrylamide gel electrophoresis
PAM	Protospacer Adjacent Motif
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
QC	Quality Control
qPCR	Quantitative polymerase chain reaction
RBD	Receptor binding domain
RNA	Ribonucleic acid
RO	Replicative organelle
RPM	Rounds per minute
RT	Reverse-transcriptase
RTC	Replication transcription complex
SARS	Severe acute respiratory syndrome
SDS	Sodium Dodecyl Sulfate
se	Sense
seq	sequencing
sgRNA	Single guide RNA
TAE	Tris-Acetate EDTA
TBS	Tris-buffered saline
$TCID_{50}$	Tissue culture infectious dose 50%
TEMED	N,N,N′,N′-Tetramethylethylenediamine
UPR	Unfolded protein response
UTR	Untranslated region
µ	micro
µl	Microliter
µg	Microgram
µM	Micromolar

List of Figures

List of Tables

Declaration

I declare the following:

I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and referenced all text passages that are derived literally from or are based on the content of published or unpublished work of others, and all information that relates to verbal communications. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University of Giessen in carrying out the investigations described in the dissertation.

_____

Benadict Vincent Albert

Acknowledgments

Dear Prof. Michael Kracht, many thanks for this wonderful scientific journey you have given me at the Rudolf-Buchheim Institute of Pharmacology. I express sincere gratitude for steering me to develop as an independent scientific researcher and for being an excellent coach to cultivate a very professional scientific career. You have passed on to me one of the unique talents of yours, “critical thinking”, and I take this with me as a gift in all my future endeavors.

Many thanks, Dr. Johanna Meier-Sölch; you have shown me a path to my future career by introducing me to the CRISPR world. In many instances, you have inspired me to tackle scientific challenges in much more calm and elegant ways.

Dr. Christin Mayr-Buro, I thank the universe for the bond that we shared in the last few years, professionally and personally. You have inspired me to face my fears and difficulties in a supportive manner. I am very grateful for your advice and support in the lab, especially during my tough times.

Dr. Ulas Tenekeci, many thanks for being a jump-start for my independent research. I express my sincere gratitude for extending your unconditional support during my tough times at the lab and on my personal front.

Dear Hendrik Weiser, I could not be more thankful to the universe for providing me with a “go-to” person at the institute. Over these years, you have shaped my lab skills, lab management, and ethics. Many thanks for inspiring and including me in various local activities, by which, I could integrate into the German culture.

Dear Dr. Daniel Heylmann, I enjoyed our time during the last few years both professionally and personally. Many thanks for being there during my toughest times and integrating me into the culture as a “local guy”.

Dear Sebastian Werner, many thanks for being an excellent guide on bioinformatics that I never thought I could learn in my life. In many instances, you have inspired me to self-teach a few things when you lack in-time support.

Dear Dr. Samer Shaban, many thanks for igniting new scientific ideas in many instances during my PhD journey. It was an excellent experience to share the space as a co-PhD student and learn many professional and personal skills from you.

Dear Dr. Axel Weber, Dr. Knut Beuerlein, future Dr. Jana Juli, Petra Kronich, Christiane Zibuschka, Heike Schubert, Jasmine Dörr, and all the other colleagues at RBI, many thanks for supporting me in various aspects in the lab and bureaucratic issues. I express sincere gratitude to Dr. Ramakanth Madugiri, Nadja Karl, and Dr. Christin Müller for their invaluable insights on the technical development of the MERS-CoV screen selection and for providing a good collaborative environment at the SFB1021.

Dear future Dr. Lisa Leib, I thank the universe for providing me with a “constant person” in my life. Many thanks for your love and support in all aspects of my life. You have nurtured me into the person I am now and thanks for being “Lisa” to me always.

Mom, Dad, brother, Nirmal, Sakthi, and Gowtham, thanks for providing me a home in this world and for supporting all the important decisions of my life. Many thanks to my friends in Germany (Vinoth, Nivetha, Arvind, Neeha, Chaitu, and Raghav) for elevating my personal life and balancing out the toughest times during this PhD journey.

Finally, I thank myself for believing in me and to overcome all my fears in life. I dedicate this PhD journey to the future myself.