Genome-wide CRISPR screen to identify lncRNAs affecting SARS-CoV-2 cell entry

Research Thesis
In Partial Fulfillment of The Requirements for the
Degree of Master of Science in Biology

Yuval Ben Eliahu

Submitted to the Senate of the Technion - Israel
Institute of Technology
Elul 5782 Haifa September 2022

The Research Thesis Was Done Under the Supervision of Dr. Assaf Bester in the Faculty of Biology.

The generous financial help of the Technion is gratefully acknowledged.

Table of Contents

Figures and Tables

Abstract

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the causative agent of the Coronavirus disease 2019 (COVID-19) pandemic. SARS-CoV-2 infection is mediated by the binding of the viral spike glycoprotein to the host angiotensin-converting enzyme 2 (ACE2) receptor. Discovering host factors critical for SARS-CoV-2 infection will facilitate the development of novel therapeutic treatments for COVID-19, as well as elucidate SARS-CoV-2 pathogenesis. While most studies focus on identifying protein-coding genes crucial for the infection, the non-coding genome is largely understudied. The majority of our genome is composed of genes that transcribe into non-coding RNA (ncRNA). Long non-coding RNA (lncRNA) is the largest and most diverse group of ncRNA. lncRNAs are defined as transcripts longer than 200 nucleotides and have no open reading frame (ORF). lncRNAs can regulate all levels of gene expression by interacting with DNA, RNA, and proteins. Considering lncRNAs' involvement in many cellular processes, it is safe to assume they also play a role in SARS-CoV-2 infection. Characterizing lncRNAs can be challenging due to their enrichment in the nucleus, their large number, and their lack of ORF making them difficult to target and producing a loss of function. Clustered regularly interspaced short palindromic repeats (CRISPR) based screening is a robust, high-throughput approach to studying lncRNAs. Specifically, CRISPR interference (CRISPRi) can modulate gene expression at the transcription level without the requirement of genome editing. In this study, we performed a CRISPR screen of lncRNAs affecting SARS-CoV-2 infection in SNU449 human cells, originating from liver Carcinoma and modified to express the ACE2 receptor. To imitate SARS-CoV-2 viral entry, we used a system of vesicular stomatitis virus (VSV) pseudovirus displaying SARS-CoV-2 spike. Our screen identified many proviral and antiviral lncRNA genes. We validated RP11-977G19.11 and RP11-314A20.5 as critical for SAR-CoV-2 infection. Both genes were found to regulate the expression of neighboring protein-coding genes. Specifically, RP11-977G19.11 regulate the expression of Citrate synthase, a key member of energy metabolism, and RP11-314A20.5 regulate Chemokine (C-X-C motif) ligand 16 (CXCL16), which is involved in the immunity response. The validated genes could potentially serve as novel targets for COVID-19 therapy and expand our comprehension of SARS-CoV-2 pathogenesis. Our findings will be relevant for the current pandemic, as well as novel viral pathogens sharing the same host factors.

Abbreviations list

SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
COVID-19	Coronavirus disease 2019
ACE2	Angiotensin-converting enzyme 2
lncRNA	Long non-coding RNA
ORF	Open reading frame
CRISPRi	CRISPR interference
VSV	Vesicular stomatitis virus
sc	Single cycle
rc	Replication competent
CoV	Coronavirus
TMPRSS2	Transmembrane protease, serine 2
CTSL	Cathepsin L
ER	Endoplasmic Reticulum
ERGIC	ER-to-Golgi intermediate compartment
CRISPR	Clustered regularly interspaced short palindromic repeats
RAB7A	Ras-Related Protein Rab-7a
TMEM106B	Transmembrane Protein 106B
TGN	Trans-Golgi network
PI3K	Phosphatidylinositol 3-kinase
PI3P	Phosphatidylinositol 3-phosphate
SCAP	Sterol-regulatory-element-binding protein cleavage-activating protein
RNAi	RNA interference
ASO	Antisense oligonucleotide
sgRNA	Single guide RNA
PAM	Protospacer adjacent motif
DSB	Double-strand break
LOF	Loss of function
CRISPRa	CRISPR activation
KRAB	Krüppel associated box
SAM	Synergistic activation mediator
NGS	Next-generation sequencing
KD	Knockdown
GFP	Green Fluorescent Protein
MOI	Multiplicity of infection
FACS	Fluorescence-Activated Cell Sorting
RT-PCR	Real-time PCR
KO	Knockout
G	Glycoprotein
D614G	Aspartic acid-to-glycine substitution at position 614
Δ18	18-amino acid deletion of the cytoplasmic tail
CRiNCL	CRISPRi non-coding library
CS	Citrate synthase
CXCL16	Chemokine (C-X-C motif) ligand 16

Introduction

The COVID-19 pandemic

The Coronavirus disease 2019 (COVID-19) pandemic is caused by the infection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (Lai et al., 2020). SARS-CoV-2 is a novel member of the Coronavirus (CoV) subfamily. CoVs are enveloped, positive-strand RNA viruses that mainly infect mammals' respiratory tract, giving rise to respiratory symptoms (Pal et al., 2020). SARS-CoV-2 functional receptor is the host angiotensin-converting enzyme 2 (ACE2). The ACE2 receptor is mainly found on the surface of the lung and small intestine epithelia cells, which explains why the respiratory tract is the major site of infection (Hamming et al., 2004), and also how enterocytes (intestinal absorptive cells) might serve as cell reservoir for SARS-CoV-2 (Feng et al., 2020). Compared to other CoVs, SARS-CoV-2 has much higher transmissibility (Cevik et al., 2020), leading to its worldwide spread.

SARS-CoV-2 infection and life cycle

The first step of SARS-CoV-2 infection is viral entry into host cells. Viral entry is mediated by the viral spike glycoprotein, which covers the virus surface and is composed of subunits S1 and S2. The spike protein binds to the cell ACE2 receptor, this induces a conformational change in the spike protein, exposing the S2' site in the S2 subunit. Following the binding, the virus can enter the cells by two distinct pathways. The cell surface pathway utilizes the host transmembrane protease, serine 2 (TMPRSS2), which cleaves the S2' site, separating the subunits. This activates the S2 subunit, which mediates the fusion of the viral and cellular membranes, catalyzed by the fusion peptide in the subunit. Viral RNA is then released into the host cell cytoplasm through the fusion pore. When the cell expresses low level of TMPRSS2, the virus enters using the endosomal pathway, and the two entry pathways can be studied separately depending on TMPRSS2 expression in the cell (Koch et al., 2021). Following ACE2 binding, the virus is internalized into the cell by endocytosis mediated by clathrin, which coats transport vesicles to catalyze membrane trafficking. In the endosome, Cathepsin L (CTSL) cleaves the S2' site. This is followed by a similar membrane fusion and viral RNA release. (Jackson et al., 2021, Walls et al., 2020, Hoffmann et al., 2020b). Inside the cytoplasm, the viral RNA is uncoated and subjected to translation of two open reading frames, ORF1a and ORF1b, resulting in polyproteins that are processed into structural proteins and proteins required for viral replication. Synthesis of viral RNA occurs in protective membranes in the endoplasmic reticulum (ER). The structural proteins are translocated to the ER-to-Golgi intermediate compartment (ERGIC), where they assemble with the RNA to form mature virions. The virions are then secreted from the host cell via exocytosis (V'kovski et al., 2020, Prydz & Saraste, 2022).

Host factors for SARS-CoV-2 infection

Like all viruses, SARS-CoV-2 utilize the host factors and cellular pathways for infection and life cycle. Discovering proviral host factors is crucial to advance the development of targeted therapies for COVID-19, as well as to increase our understanding of SARS-CoV-2 biology. The identified host factors could be shared by other viruses, and might be relevant for the treatment of future viral pathogens. Many studies have already been carried out to discover of host factors required for SARS-CoV-2 infection, using high-throughput approaches such as genome-wide genetic screens and interactome analyses (Baggen, Vanstreels, et al., 2021). Clustered regularly interspaced short palindromic repeats (CRISPR) based screening is a robust approach to study virus-host interactions. It has high specificity to the genomic target, and can be scaled up to high-throughput with ease (Shalem et al., 2015). Several

CRISPR screens have been conducted to identify host factors that regulate different aspects of SARS-CoV-2 life cycle (Baggen, Persoons, et al., 2021, Biering et al., 2022, Mac Kain et al., 2022, Samelson et al., 2022, Rebendenne et al., 2022, Flynn et al., 2021, Zhu et al., 2021, (Wei et al., 2020), Daniloski et al., 2021, R. Wang et al., 2020).

There are several parameters to consider when choosing a cell line for a CRISPR screen, including the physiological relevance, in this case susceptibility to SARS-CoV-2 infection, the organism and tissue from which the cell line originate, and the sensitivity to CRISPR based mutagenesis. Five main cell lines were used in the screens mentioned above. The African green monkey kidney derived Vero E6, and the human hepatocellular carcinoma Huh7, both endogenously express the ACE2 receptor, but not the TMPRSS2 protease. The human lung adenocarcinoma A549, which require an ectopic overexpression of ACE2 to achieve permissibility to SARS-CoV-2. And the human colorectal adenocarcinoma Caco-2 and human lung adenocarcinoma Calu-3 cell lines which express both TMPRSS2 and ACE2, and are highly permissive to SARS-CoV-2 infection (Chu et al., 2020). In addition to different cell lines, several viral models for SARS-CoV-2 infection were utilized in the screens. Being the most physiology relevant model, SARS-CoV-2 (USA/WA-1/2020 strain) was used in many of the screens (R. Wang et al., 2020, Wei et al., 2020, Daniloski et al., 2021). Another commonly used model was Vesicular Stomatitis Virus (VSV) displaying SARS-CoV-2 spike. VSV is a negative-strand RNA virus that is frequently used to study virus-host interactions. Unlike SARS-CoV-2, that requires biosafety level 3 (BSL-3) facilities, VSV can be used in BSL-2 containment. VSV possess a small, easily manipulated genome and can be produced at a very high yield (Whitt, 2010). VSV can be pseudotyped with the SARS-CoV-2 spike protein to produce single-cycle VSV (scVSV), or modified to express the gene encoding the protein to produce replication competent VSV (rcVSV) (Farzani et al., 2020). Both rcVSV and scVSV displaying SARS-CoV-2 spike were used in screens or validation assays, as a way to test if the identified genes act at the viral entry level (Wei et al., 2020, Mac Kain et al., 2022, R. Wang et al., 2020, Biering et al., 2022).

Unsurprisingly, the top-ranking host factors identified by the screens were the ACE2 receptor and the TPMRSS2 and CTSL proteases, which are required for SARS-CoV-2 cell entry through the surface and endosomal pathways. Beside these critical genes, several other host factors and cellular processes were discovered to be involved in SARS-CoV-2 infection. Vesicle trafficking plays an important role in promoting SARS-CoV-2 cell entry. Specifically, Ras-Related Protein Rab-7a (RAB7A), which is a GTPase regulator of vesicular transport, was found to promote ACE2 cell-surface expression (Daniloski et al., 2021). Additionally, Transmembrane Protein 106B (TMEM106B), which regulates lysosome function, was found to promote SARS-CoV-2 infection, with high expression in airway cells isolated from COVID-19 patients (Baggen, Persoons, et al., 2021) Two sub-units of clathrin-associated adaptor protein complex 1 (AP1B1 and AP1G1), which play a role in trafficking between the trans-Golgi network (TGN) and endosomes, were found to promote SARS-CoV-2 viral entry (Rebendenne et al., 2022). The phosphatidylinositol 3-kinase (PI3K) signaling pathway was also shown to be promote SARS-CoV-2 infection. Two catalytic subunits of the PI3K complex (PIK3C3 and PIK3R4), that mediate formation of phosphatidylinositol 3-phosphate (PI3P) and is involved in endocytic trafficking and autophagy, were found to be proviral, and also demonstrated drug targets potential (Baggen, Persoons, et al., 2021), (R. Wang et al., 2020). Genes involved in cholesterol homeostasis including sterol-regulatory-element-binding protein cleavage-activating protein (SCAP), endopeptidase S1P and S2P (MBTPS1 and MBTPS2) were also found to promote viral entry. Cellular cholesterol seem to be required for Spike-

meditated entry of SARS-CoV-2 (R. Wang et al., 2020), possibly through cholesterol-rich lipid rafts which recruit ACE2 and TMPRSS2 to interact with the spike protein (Palacios-Rápalo et al., 2021).

Long non-coding RNA

While most studies focus on identifying protein-coding host factors crucial for SARS-CoV-2 infection, the role and therapeutic potential of the vast non-coding genome have been largely overlooked. Less than 3% of our genome is transcribed into mRNA that encode proteins. The remaining majority is composed of non-coding genes that transcribe several groups of RNA, such as miRNA, snRNA and lncRNA. Out of these groups, long non-coding RNA (lncRNA) is the largest and most diverse. lncRNAs are defined as transcripts longer than 200 nucleotides and have no open reading frame (ORF), or a very short one (Liao et al., 2018). Compared to mRNA, lncRNAs have a low and tissue-specific expression levels. (Ma et al., 2013). lncRNAs can regulate gene expression by interacting with DNA, RNA and proteins. Through these interactions, lncRNAs can function at the epigenetic, transcriptional, post-transcriptional, translational and post-translational levels (X. Zhang et al., 2019). lncRNAs mechanisms of action can be separated to regulating the expression of neighboring protein-coding genes in *cis*, or of distant protein-coding genes in *trans* (Yan et al., 2017). LncRNAs gene regulatory activity make them modulators of different cellular processes such as cell cycle, differentiation and development (Yao et al., 2019), and were found to be involved in both health and disease (Aliperti et al., 2021). lncRNAs play a major role in all aspects of cancer development (M.-C. Jiang et al., 2019), and recent studies also revealed their participant in viral infection (W. Liu & Ding, 2017), (Yi et al., 2019). Considering their important roles in many biological processes, and that they are still largely uncharacterized, we believe that lncRNAs have high potential to include novel host factors critical for the SARS-CoV-2 infection.

Challenges and approaches to lncRNA study

There are a number of challenges involved in characterizing lncRNAs. They have a very large number (Evans et al., 2016) and are poorly annotated (Volders et al., 2013), meaning high-throughput tools are required to study them. They also tend to be enriched in the nucleus (Bridges et al., 2021), so methods such as RNA interference (RNAi), which target the cytoplasm, are mostly ineffective. CRISPR screening overcome these challenges and has proven to be a robust approach to identify functional lncRNAs. Compared to other perturbation methods such as Antisense oligonucleotides (ASOs), CRISPR can be scaled to high-throughput at a relatively low cost (Joung et al., 2017). The CRISPR-Cas9 system can effectively target the gene at the nucleus, and with less off-target effects compared to RNAi (Ui-Tei, 2013). The CRISPR-Cas9 tool is based on the *Streptococcus pyogenes* bacteria immune system. It is composed of two components: Cas9 endonuclease and a single guide RNA (sgRNA) consisting of a variable 20 nt sequence. The sgRNA is recognized by Cas9, leading it to a homologues sequence in the genome which is followed by a protospacer adjacent motif (PAM). There Cas9 binds to the DNA and induce a double-strand break (DBS). A break in a gene ORF result in small insertions and deletions which are called indels, this can lead to a frameshift mutation and a loss of function (LOF) (F. Jiang, 2017). Since lncRNAs have no ORF, a frameshift mutation is unlikely to generate a loss of function (Pulido-Quetglas & Johnson, 2021). New CRISPR technologies have been developed to challenge such issues. Specifically, CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) can be used to modulate the gene transcription without permanently editing the genome. CRISPRi and CRISPRa utilize a catalytically dead Cas9 (dCas9), whom endonuclease activity has been removed, but retain is capability to bind DNA. dCas9 is fused to regulatory domains such as Krüppel associated box (KRAB)

(Gilbert et al., 2013) or synergistic activation mediator (SAM) (Y. Zhang et al., 2015) that repress or activate the transcription respectively.

CRISPR screens for identifying functional lncRNAs

CRISPR screens can be conducted in both arrayed and pooled formats. Arrayed screens introduce an individual sgRNA to each cell population in separate culture wells, meaning that a large number of sgRNAs need to be individually cloned, and a lot ofs wells to culture to maintain many isogenic cell lines. In contrast, pooled screen introduces a library of sgRNA into a single population of cells. The cells then undergo a selection step, in which the cells are selected according to a desired phenotype. Such as resistance to drug or sorting by a fluorescent marker. The selected cells are then harvested for genomic DNA, the sgRNA sequences are amplified by PCR, and the amplicons are subjected to next generation sequencing (NGS). Finally, significantly enriched or depleted sgRNAs in the selected population compared to a control population are identified by a statistical analysis. The genes targeted by the identified sgRNAs are candidates to be involved in the phenotype of interest (Joung et al., 2017). CRISPRa screening was used to identify functional lncRNAs affecting drug response (Bester et al., 2018), and CRISPRi screening identified lncRNAs regulating cell growth in 7 distinct cell lines (S. J. Liu et al., 2017).

Gap of knowledge

Whereas many protein-coding genes crucial for SARS-Cov2 infection have been identified, there are no genome-wide studies of essential lncRNAs to date, and the mechanisms by which lncRNAs mediate the infection are largely unknown.

Hypothesis

We hypothesize that lncRNAs may play role in SARS-Cov2 infection, and that these lncRNAs could be clinically relevant as cell-type specific targets or biomarkers.

Significance

Discovering lncRNA genes critical for the SARS-CoV-2 infection will shed light on the virus pathogenesis mechanisms, and could potentially serve as novel therapeutic targets for the treatment of COVID-19 as well as future viral originated pandemics. Furthermore, the study will grow our knowledge on the function and biology on lncRNAs and will increase our understanding of this important group of genes.

Research objective

The objective of our research is to identify new genes involved in the SARS-CoV-2 cell entry using unbiased CRISPR screen targeting lncRNA genes. To this end, we will establish a SARS-CoV-2-spike displaying pseudovirus system to measure cell entry, and use it to perform high a parallel CRISPR knockdown (KD) using a library of sgRNAs targeting commonly expressed lncRNAs. Following identification of significantly enriched genes, we will validate the screen results using individual sgRNA plasmids. Finally, we will characterize the validated genes and study the mechanism by which they affect the infection.

Aim 1. Identify lncRNA genes critical for the SARS-CoV-2 infection using a CRISPR screen

Our first aim is to establish a system to identify lncRNAs that impact SARS-CoV-2 infection. Toward this aim, we designed a CRISPRi screening approach targeting non-coding genes in a model of pseudovirus SARS-CoV-2 infection.

Aim 1.1. Establish a system to measure SARS-CoV-2 infection. To simulate SARS-CoV-2 cell entry, we plan to generate Vesicular Stomatitis Virus (VSV) pseudotyped to display the SARS-CoV-2 Spike protein (Hoffmann et al., 2020). The VSV we will use in our experiments have the gene encoding the native G glycoprotein deleted (VSVΔG), making them capable of a single cycle of infection (scVSV) (Farzani et al., 2020), and replaced with a Green Fluorescent Protein (GFP) reporter gene to quantify the entry of Spike displaying VSV (VSVΔG-Spike) particles into the cells by flow cytometry (Dieterle et al., 2020).

Aim 1.2. Establish a stable cell line for CRISPRi screening. For our CRISPR screening we require cells stably express the CRISPRi system (dCas9-KRAB) (Alerasool et al., 2020) to modulate the expression of genes targeted by our sgRNA library, as well as expressing the ACE2 receptor and the TMPRSS2 protease to make the cells susceptible to SARS-CoV-2 infection. We will introduce the systems into cell lines proven to show an effective repression when challenged with CRISPR. Such as the human embryonic kidney HEK-293T, or the human liver Carcinoma SNU-449.

Aim 1.3. CRISPR screen to identify lncRNAs critical for the SARS-CoV-2 infection. To discover lncRNA genes altering SARS-CoV-2 infection, we will use a lentiviral transduction to introduce the CRISPRi non-coding library, targeting 1329 genes with 10 sgRNAs/TSS (S. J. Liu et al., 2017), into cells stably expressing CRISPRi and ACE2-TMPRSS2 (established in Aim 1.2). To maintain the complexity of the library, the infection will be done at a low multiplicity of infection (MOI) with 1000 cells per sgRNA. Then, the cells will be infected with VSV displaying SARS-CoV-2 spike and encoding GFP. Next, we will use Fluorescence-Activated Cell Sorting (FACS) to collect the top and bottom 10% GFP fluorescent cells, which represent cells with high or low susceptibility for SARS-CoV-2 infection respectively. We will also collect the 80% cells with medium GFP to use as control (Sharon et al., 2020). Four biological replicates will be performed.

Aim 1.4. Analysis of enriched sgRNAs targeting genes that affect SARS-CoV-2 infection. To identify sgRNAs enriched in the GFP high and low populations relative to the control, we will construct sequencing libraries as previously described (Horlbeck et al., 2016). The sgRNA sequences will be amplified by PCR with primers that will serve as unique barcodes. The libraries will be pooled and the pool subjected to next-generation sequencing (NGS), with an average of 250 reads per sgRNA. Following the sequencing, we will perform a bioinformatic analysis to test the screen quality, align the sgRNA reads to the non-coding library to generate read-count tables and identify significantly enriched genes in the experimental groups relative to the control.

Aim 2. Validation of the identified genes

Following the screen, we will seek to validate the identified genes in an arrayed format with individual sgRNAs. Initially we will confirm the results in the same cell line generated to conduct the screen. Next, we will seek to test the efficiency and specificity of the hits KD. Finally, we will validate the hits in more physiologically relevant cell lines.

Aim 2.1. Validation of screening hits. To validate our screen results, we will construct individual plasmids containing the top enriched sgRNAs. The sgRNAs will be introduced into the same cell line generated for the screen, stably expressing CRISPRi and ACE2-TMPRSS2. Then, the cells will be infected with VSV displaying SARS-CoV-2 spike and encoding GFP. Finally, we will use flow cytometry to measure the infection efficiency, and test if the genes knockdown regulated the infection in accordance to the cell population in which the sgRNAs were enriched (GFP high or low).

Aim 2.2. Validation of hits KD efficiency and specificity. To test the efficiency and specificity of the validated genes KD by individual plasmids we will use Real-time PCR (RT-PCR) to measure the relative RNA expression of the validated gene following KD. In case the lncRNA overlap with protein-coding genes, we will locate sequence regions unique to the lncRNA, and design primers amplifying the unique regions. The primers will also span an exon-exon junction to avoid amplifying genomic DNA (gDNA) (Dhamija & Menon, 2021).

Aim 2.3. Validation of hits in other cell lines. Further validation of the genes will be done in more physiologically relevant cells such Calu-3, which derive from lung adenocarcinoma, or Caco-2 which derive from colorectal adenocarcinoma. Calu-3 and Caco-2 are the most permissive cells to SARS-CoV-2 infection, and promote rapid replicating of the virus compared to other cell lines (Cagno, 2020) (Zupin et al., 2022). The validation will be done as described in Aim 2.1.

Aim 3. Characterization of the validated genes

Next, I aim to study the molecular and cellular functions of the validated lncRNAs. To this end, I will test if the lncRNAs may function in *cis* or in *trans*. And shed slight on the cellular function that affect SARS-CoV-2 infection.

Aim 3.1. Studying the cis regulatory role of the validated hits. To test if the validated genes regulate the expression of neighboring protein-coding genes, we will first use the sgRNAs designed in aim 2.1 to KD the validated genes. Then, we will use RT-PCR to measure the expression of neighboring genes in KD cells. Following identification of protein-coding genes regulated by the validated genes, we will confirm the role of the coding genes in SARS-CoV-2 infection. To this end, we will use CRISPR-Cas9 to knockout (KO) the coding genes. Then we will infect the KO cells with VSV displaying spike and measure the infection efficiency.

Aim 3.2. Characterization of the validated hits function. To test if the validated genes regulate ACE2 expression, we will use conjugated antibodies specific to ACE2 in both normal and fixed cells to measure ACE2 protein expression in the whole cell and on the surface. To test if the validated genes effect is specific to SARS-CoV-2-spike displaying VSV, we will infect KD cells with different virus types such as the HIV based lentivirus or VSV displaying the native glycoprotein.

Although important, further molecular and cellular analysis are beyond the scope of my research due to time limits and will be performed in a continues studies

Research description

In this study, we performed a CRISPR screen using VSV pseudotyped to display SARS-CoV-2-spike in the SNU449 cell line modified to overexpress the ACE2 receptor and TMPRSS2 protease. The screen identified over 30 non-coding genes that were both proviral and antiviral using two distinct analysis tools. Specifically, RP11-977G19.11 and RP11-314A20.5 were validated to be critical for the infection. We found that RP11-977G19.11 and RP11-314A20.5 regulate the expression of the neighboring genes Citrate synthase and Chemokine (C-X-C motif) ligand 16 (CXCL16) respectively, implying a potential mechanism by which RP11-977G19.11 and RP11-314A20.5 affect SARS-CoV-2 infection. Summary of the work can be found at Figure 20. Our results add to the growing knowledge of essential host factors for the SARS-CoV-2 infection, with a unique focus on lncRNAs. The identified genes could represent novel therapeutic targets for the treatment of COVID-19, and further elucidate SARS-COV-2 pathogenesis.

Results

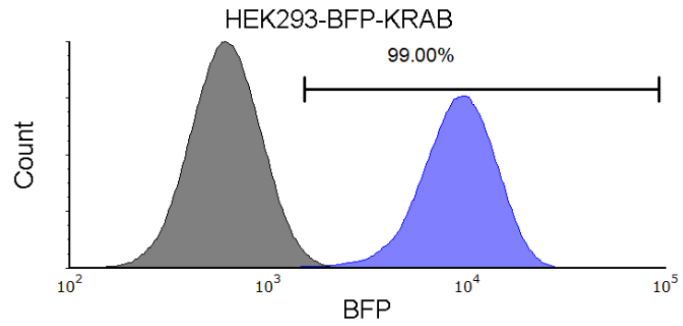
Establish CRISPR and pseudovirus systems

As described in aim 1, we first sought to establish a system to identify lncRNAs that affect SARS-CoV-2 infection. To this end, we required a cell line that will demonstrate both robust repression by CRISPRi, as well as efficient infection by VSV displaying SARS-CoV-2 spike.

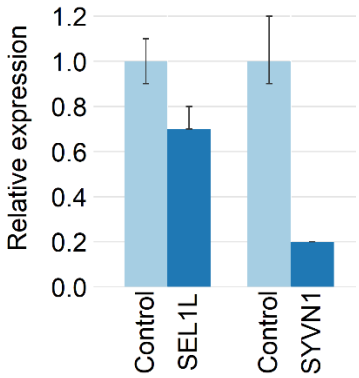
Generating CRISPRi cell lines

We tested two cell lines as candidate target cells for our CRISPR screen. The human embryonic kidney HEK293 cells, which are frequently utilized due to their easy maintenance and susceptibility to transfection (Petiot et al., 2015), and the human liver Carcinoma SNU449 cells, which are highly permissive to viral infection (Fournier et al., 2018). To generate a stable cell line expressing CRISPRi, we transduced HEK293 cells with lentivirus packaging a dCas9-BFP-KRAB plasmid (Gilbert et al., 2013), or SNU449 cells with lentivirus packaging dCas9-BFP-KRAB or dCas9-mCherry-KRAB fused to a ZIM3 domain, which was shown to increase the repression efficiency (Alerasool et al., 2020). The transduced populations were enriched by sorting until the majority of the cells expressed the plasmid (Fig. 1A, C, E). To assess the repression efficiency, we transfected the cells with validated sgRNA sequences that served as positive control, or with non-targeting sgRNAs for negative control. We then measured the relative RNA expression using RT-qPCR. While in the HEK293 cells repression efficiency topped at 5-fold, in the SNU449 cells up to 30-fold repression was measured (Fig. 1D, F), with the best results observed in cells harboring the dCas9-KRAB-ZIM3 system. These results suggest that SNU449 cells expressing dCas9-KRAB-ZIM3 are the better option to use in our screen.

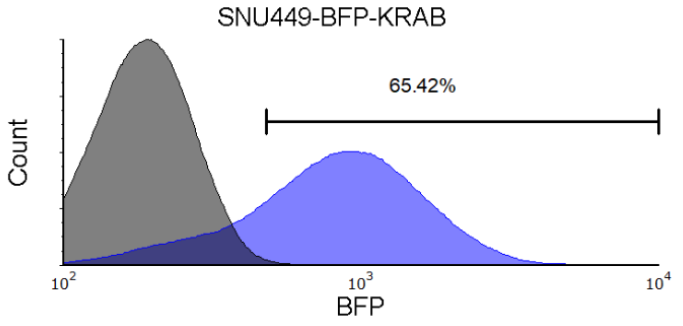
A



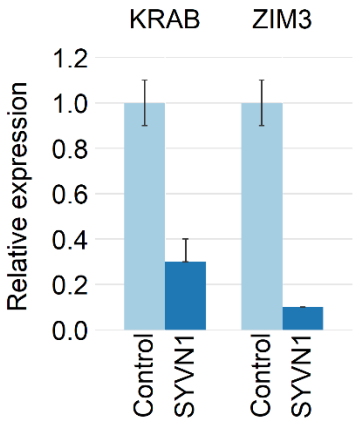
B



C



D



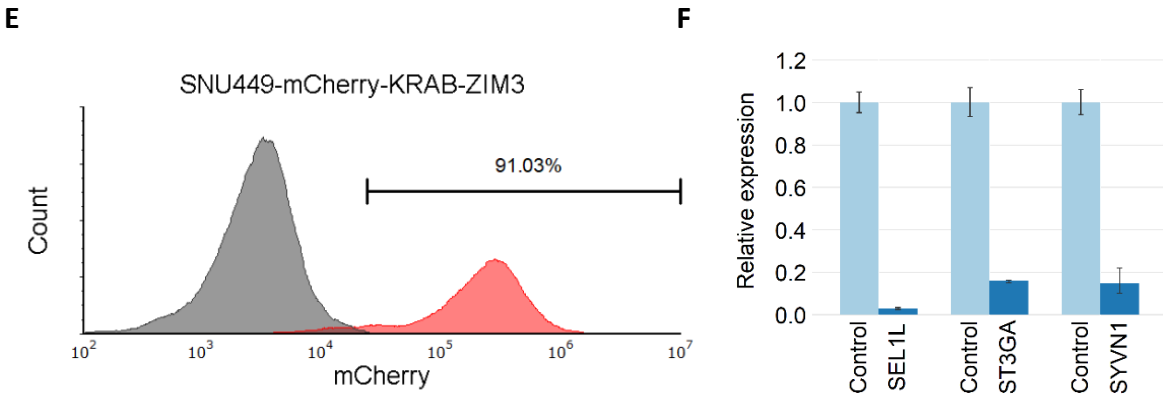


Figure 1. Generation of HEK293 and SNU449 expressing CRISPRi. HEK293T cells (A) or SNU449 (C, E) were transduced with dCas9-BFP-KRAB (A, C) or dCas9-mCherry-KRAB-ZIM3 (E), and positives went through sorting using Fluorescence-activated Cell Sorting (FACS). Relative RNA expression of targeted genes was quantified with RT-qPCR in the following cell lines: HEK293-KRAB (B), SNU449-KRAB (D), and SNU449-KRAB-ZIM3 (D, F). The expression was normalized to cells transfected with a non-targeting sgRNA sequence (10010 or 10057). n=1.

Production of VSV-ΔG-G

The VSV used in our experiments presents the native glycoprotein (G) on its surface, but the gene encoding it has been deleted (VSV-ΔG-G). The VSV genome also contained a sequence encoding a green fluorescent protein (GFP) to allow detection and measurement of the infection (Fig. 2A). To amplify our stock of VSVΔG-G, we transfected BHK-21 cells with a VSV-G plasmid. The transfected cells were infected using VSVΔG-G, and viral particles were harvested and aliquoted. The viral titer of the stock was measured by infecting BHK-21/WI-2 at 10-fold serial dilutions of the virus and manually counting cells using a fluorescence microscope (Fig. 2B).

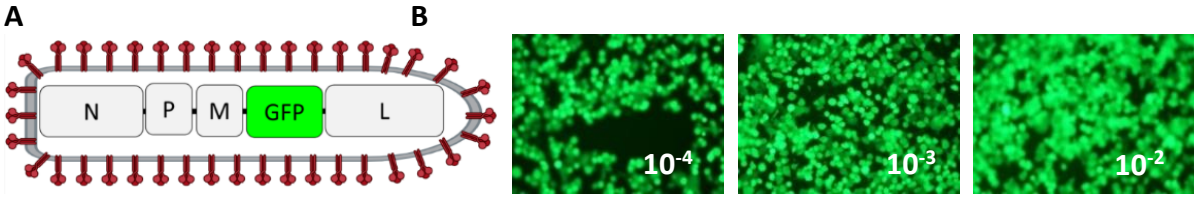


Figure 2. Generation of Vesicular Stomatitis Virus presenting native glycoprotein. (A) Schematic representation of the VSV genome in which the native glycoprotein gene has been deleted. The genome was further modified to encode a green fluorescence protein (GFP) to measure infection efficiency. The genome is composed of Nucleoprotein (N), Phosphoprotein (P), Matrix protein (M), and Large polymerase protein (L). (B) Representative images of BHK21 cells infected with VSV-ΔG-G at 10^{-2} to 10^{-4} serial dilutions. Recorded 24h post-transduction at magnification X20 using a fluorescence microscope.

VSV-ΔG-Spike infection

To construct a system for studying SARS-CoV-2 spike mediated viral entry, we pseudotyped VSV-ΔG particles with the SARS-CoV-2 spike protein (Fig. 3). The pseudovirus was used to infect HEK293 cells that either stably expresses the human ACE2 receptor, or transiently via plasmid transfection (Fig. 4A). Vero-E6 cells, which originate from the African green monkey and are highly susceptible to SARS-CoV-2 infection (Matsuyama et al., 2020), were also infected for control (Fig. 4B). Infection efficiency was measured using flow cytometry. Efficiency ranged from ~2-3% in the stable and Vero-E6 cell lines to ~6% in the transient cell line. Since our CRISPR screen is based on sorting, we require a high infection efficiency to have distinguished cell populations sorted into each bin. To improve our system, we tested two variants of the spike proteins: 18-amino acid deletion of the cytoplasmic tail (Δ18), which Promotes spike cell surface expression and Protects from lysosomal degradation (Schwegmann-Weßels et al., 2009), and Δ18 combined with an aspartic acid-to-glycine substitution at position 614 (D614G). This variant first emerged in Europe and Increases spike density and binding affinity to the ACE2 receptor (Ozono et al., 2021). We infected 293T-ACE2 cells with pseudovirus harboring the 3 variants (Fig. 4C). The variants showed 5-fold increased efficiency compared to the wild-type spike, with infection efficiency being ~8% with the wild-type spike, and ~40-45% with the variants. Up to this point, we used BHK-21/WI-2 cells for the production of the pseudovirus following a modified protocol of Michael A Whitt (Whitt, 2010). To test if VSV production for downstream experiments could be done in the more commonly used HEK293T cells, we produced VSV-ΔG harboring the Δ18-D614G variant in both HEK293T and BHK-21 cells and infected 293T-ACE2 cells with the virus (Fig. 4D). Slightly higher efficiency was observed when BHK-21 cells were used for virus production.

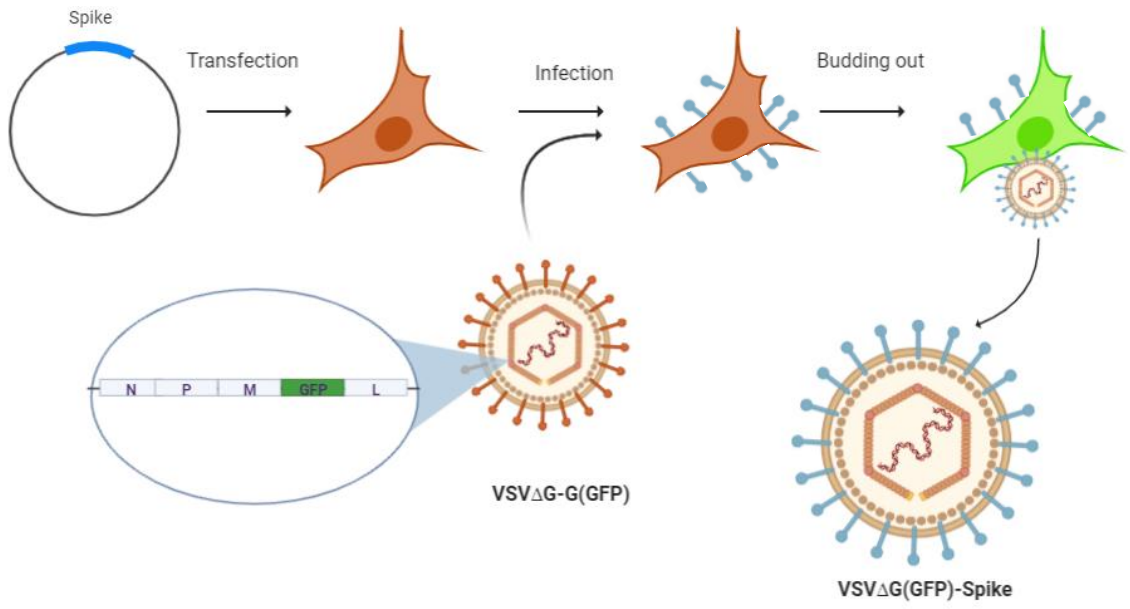


Figure 3. Schematic of VSV presenting SARS-CoV-2 spike protein preparation. To pseudotype the SARS-CoV-2 spike protein on the surface of VSV particles, HEK293 or BHK-21 cells are transfected with a plasmid carrying the SARS-CoV-2 spike. The cells are infected by VSV-ΔG-G encoding a GFP reporter gene. The viral particles that will bud out will harbor display the Spike protein on their surface

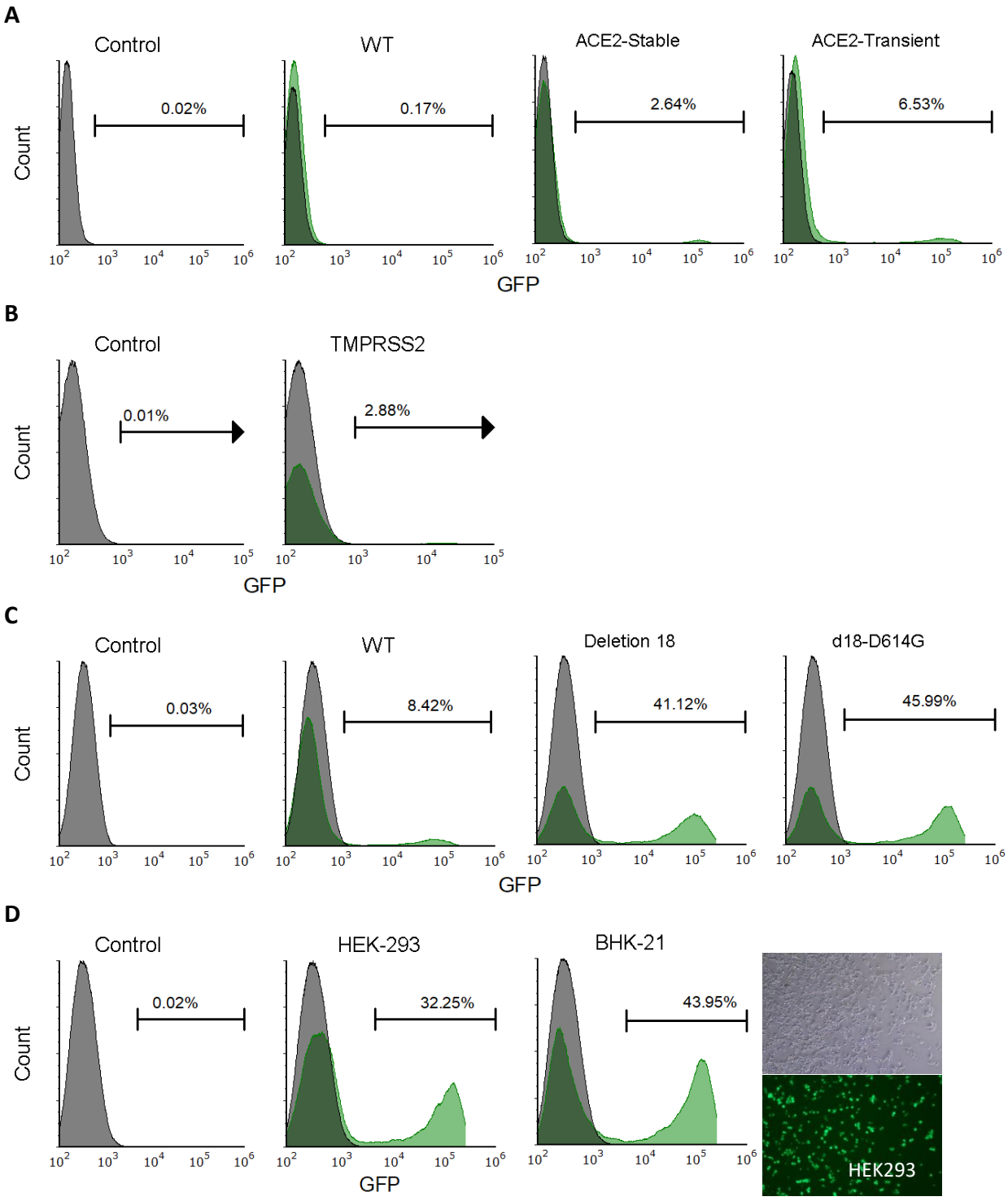


Figure 4. VSV-ΔG-Spike infection of HEK293 and Vero-E6 cells. (A, B) VSV-ΔG-Spike encoding GFP was used to infect HEK293T cells with a stable or transient expression of ACE2 (A), or Vero-E6 cells expressing the TMPRSS2 protease (B). wild-type cells were infected for control. GFP intensity was measured using flow cytometry with unstained cells for gating. (C) VSV-ΔG pseudotyped with wild-type, Δ18, and Δ18-D614G spike variants was used to infect 293T-ACE2 (transient). (D) Left: VSV-ΔG pseudotyped with the Δ18-D614G spike variant was produced in either HEK293T or BHK-21/WI-2 cells and used to infect 293T-ACE2 (transient). Right: Representative images of the cells were recorded 24h post-infection at a magnification of X10. n=1

Following the generation of the SNU449 cell line stably expressing the KRAB-ZIM3 system, we sought to test if they can be effectively infected by our pseudovirus. We generated a stable expression of hACE2 in our SNU449-KRAB-ZIM3 cell line (SNU-KRAB-ACE2) and infected the cells with pseudovirus harboring the Δ18-D614G spike variant produced in HEK293 cells. The infection efficiency was measured using flow cytometry (Fig. 5).

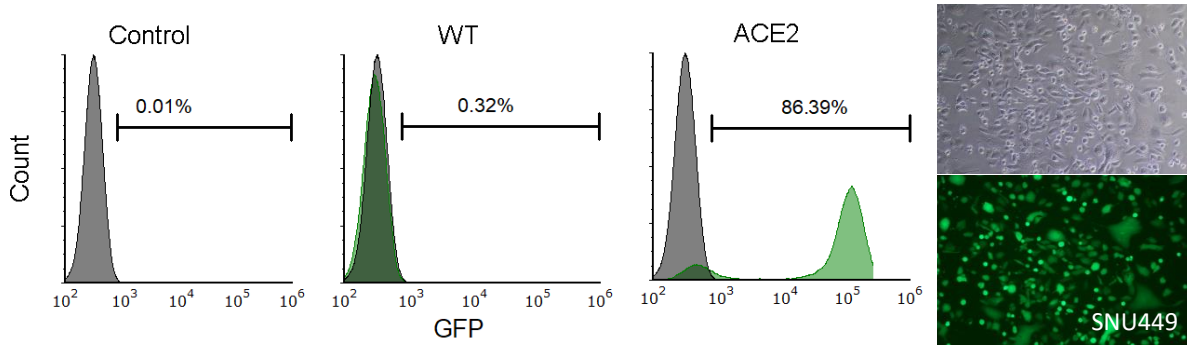


Figure 5. VSV-ΔG-Spike infection of SNU449 cells. VSV-ΔG pseudotyped with the Δ18-D614G spike variant was used to infect SNU449 cells stably expressing ACE2. wild-type cells were infected for control. Left: GFP intensity was measured using flow cytometry with unstained cells for gating. Right: Representative images of the cells were recorded 24h post-infection at a magnification of X10. n=1

ACE2 antibody-binding assay

To evaluate the expression of ACE2 in our SNU449-ACE2 cell line, we performed indirect labeling using an anti-ACE2 santibody and a fluorophore-conjugated secondary antibody. The expression was measured by flow cytometry. ACE2 was highly expressed in the modified cells compared to the wild-type (Fig. 6A). To validate our results, we measured ACE2 relative mRNA expression by RT-qPCR (Fig. 6B).

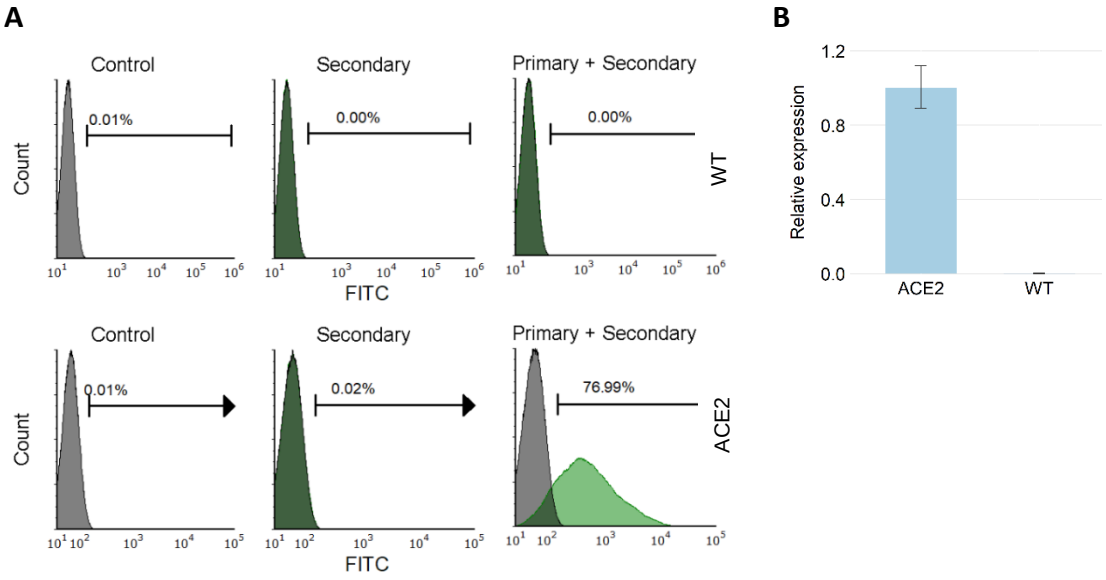


Figure 6. ACE2 cell-surface expression. (A) SNU449-ACE2 and wild-type cells were stained using an anti-ACE2 antibody and a FITC-conjugated secondary antibody. FITC intensity was measured by flow cytometry. Cells incubated only with the secondary antibody were used for control and unstained cells were used for gating. N=1. (B) RNA expression of ACE2 was quantified with RT-qPCR in SNU449-ACE2 and wild-type cells. n=1.

Overall, in line with Aims 1.1 and 1.2, I established a novel infection system and CRISPRi system that can use for CRISPR screening and validation.

CRISPR screen

Using the described system, we performed a pooled CRISPR screen to identify host lncRNA genes affecting SARS-CoV-2 infection at the viral entry step (Fig. 7). SNU449 cells expressing both KRAB-ZIM3 and human ACE2 were transduced with the CRISPRi non-coding library (S. J. Liu et al., 2017). To balance between the high-throughput and a signal to noise ratio, we chose to use a library that target 1329 lncRNA genes that are expressed in 7 different cell lines (IPSC, U87, MDAMB231, HEK293T, K562, MCF7 and HeLa), and thus are likely to express functional lncRNAs. After the transduction of the library at a low MOI (0.3) and high coverage (1000 cells/sgRNA), the cells were infected with single-cycle VSV presenting the Δ18-D614G spike variant and encoding a GFP reporter. Next, cells were sorted into three fractions of top 10% high and low GFP expression (GFP^h and GFP^l respectively), and the 80% between the two extremes for control (Fig. S1). sgRNAs enriched in the GFP^h and GFP^l fractions are expected to target candidate genes that are both antiviral and proviral respectively. The fractions were subjected to DNA extraction, sgRNA amplification, library pooling (Fig. S2) and next-generation sequencing. We perform the full screening in four biological repeats.

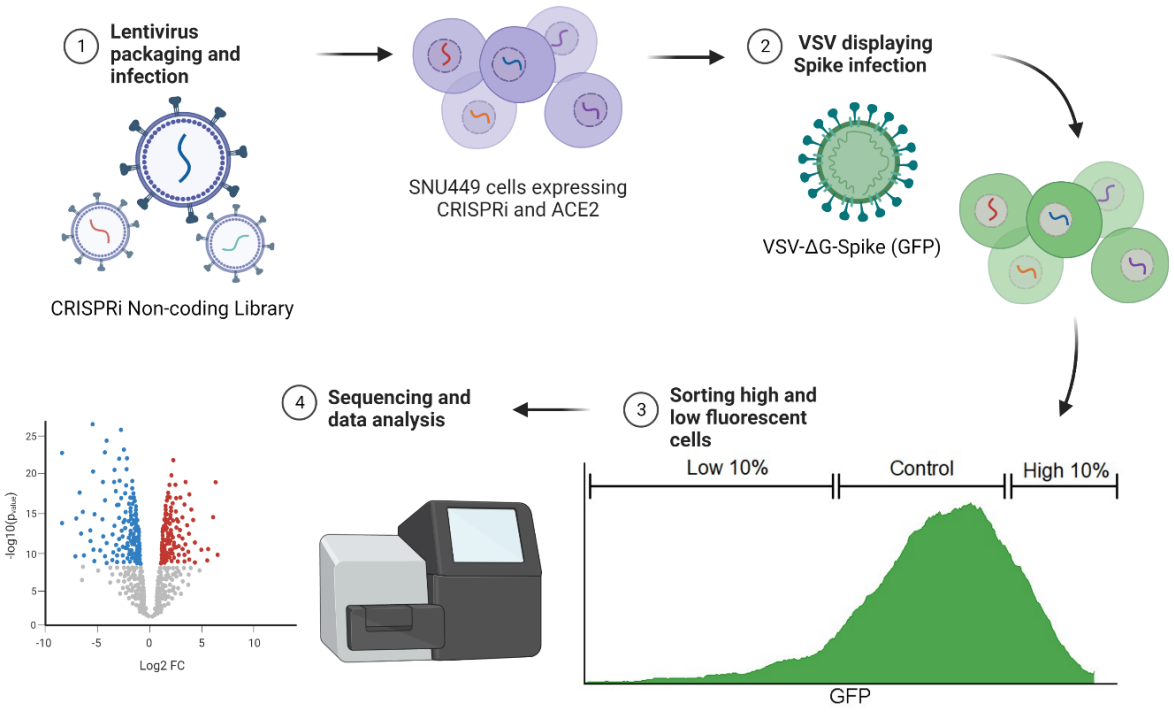


Figure 7. Schematic of the pooled screen. SNU449-KRAB-ZIM3 cells expressing hACE2 were transduced with the CRISPRi non-coding library (CRiNCL), which targets 1329 genes with 10 sgRNAs/TSS as well as 258 non-targeting sgRNAs (1). The cells were infected with VSV-ΔG-Spike encoding GFP (2) and sorted into the high and low 10% GFP-expression populations. A control population consisting of the middle GFP-expressing cells was also collected (3). The cells were harvested for genomic DNA extraction and sgRNA sequences were amplified by PCR. The amplicons were subjected to next-generation sequencing and were analyzed to identify enriched sgRNAs in the high and low fractions relative to the control (4). n=4

Screen Analysis

Sequence quality control

Quality control, read alignment, and sgRNA enrichment analysis were done using the PinAPL-Py web application. The FastQC tool was used to assess the sequence quality. Per base quality shows the entirety of the read to be in the green area, indicating high quality (Fig. 8A). Bowtie 2 was used to align the sequencing reads to the sgRNA reference library, as well as score the alignment quality (Fig. 8B). Most of the reads yield a high mapping score, meaning they uniquely align to a single library sequence. To assess if the sgRNA representation is changed between the different conditions and replicates, we checked the distribution of normalized read counts (Fig. 8C). The representation was relatively unchanged between the samples.

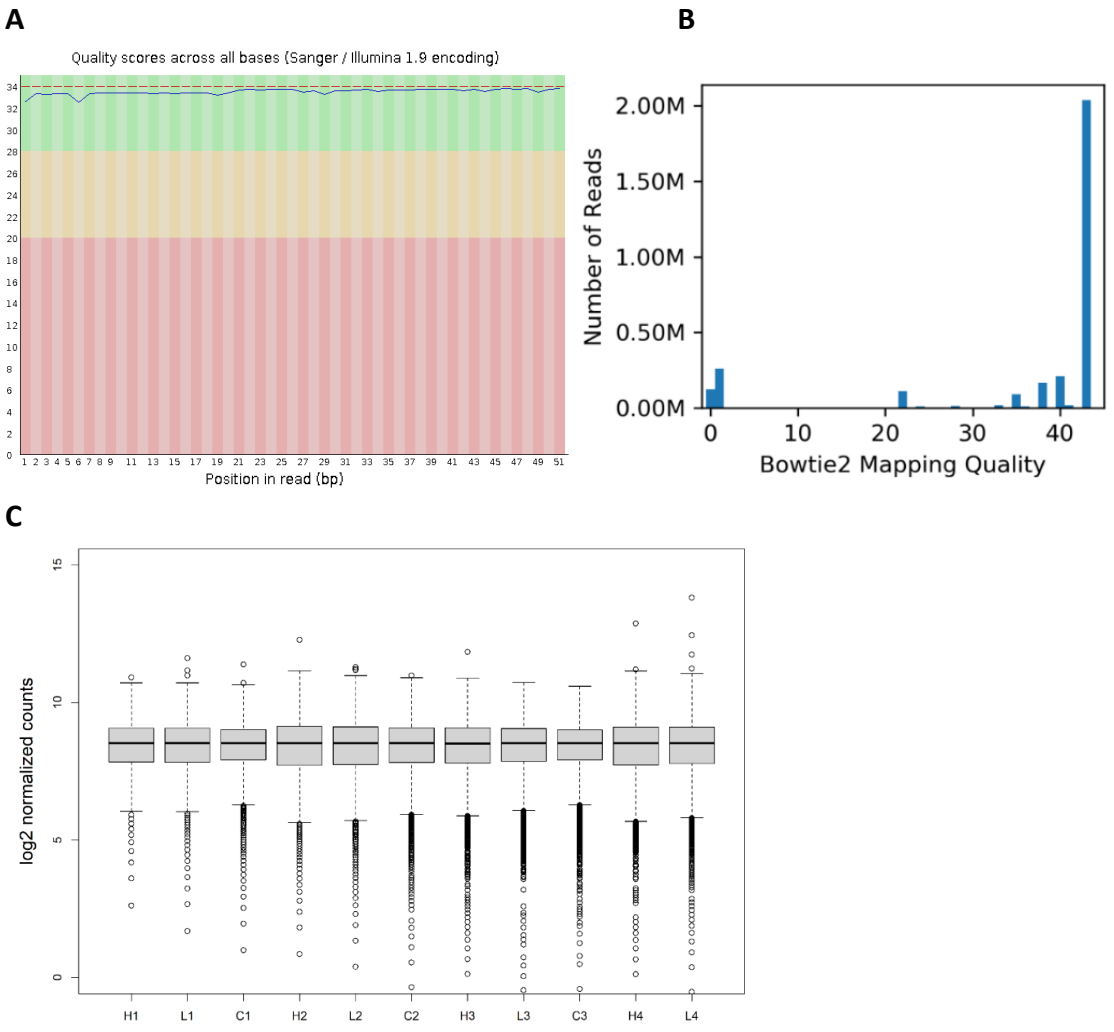


Figure 8. Sequencing and alignment quality control. (A) per-base sequence quality plot produced by FastQC. X-axis: base position in the read. Y-axis: sequence quality control defined as by Phred measure. (B) Mapping quality of reads aligned to the reference library, produced by Bowtie2. A high score indicates reads that uniquely align to a single sequence while a low score indicates reads that align to multiple or no sequences. The reads are from the control sample (replicate 1) but represent the entirety of the reads. (C) Box plot of log2 normalized counts distribution. Conditions: GFP^h – H, HGP^l – L, Control – C. Replicates are numbered from 1 to 4.

To explore the correlation between the samples, we performed a principal component analysis based on the top variable sgRNAs (Fig. 9A). Further cluster analysis was produced by PinAPL-Py (Fig. 9B). The strongest correlations are seen between the screening replicates. With replicates 1 and 3 forming a cluster. This batch effect is a result of performing each biological replicate of the screen at a different time point. To circumvent the effect, the biological replicates, as well as the different stages of the screening, should be done simultaneously whenever possible. The sgRNA read counts were highly correlated among the replicates (Fig. 9C). Overall, we conclude that following our experimental design we maintained a high coverage and low enrichment bias during cell culturing and analytic pipeline. therefore, we expect to be able to validate the function of enriched sgRNA.

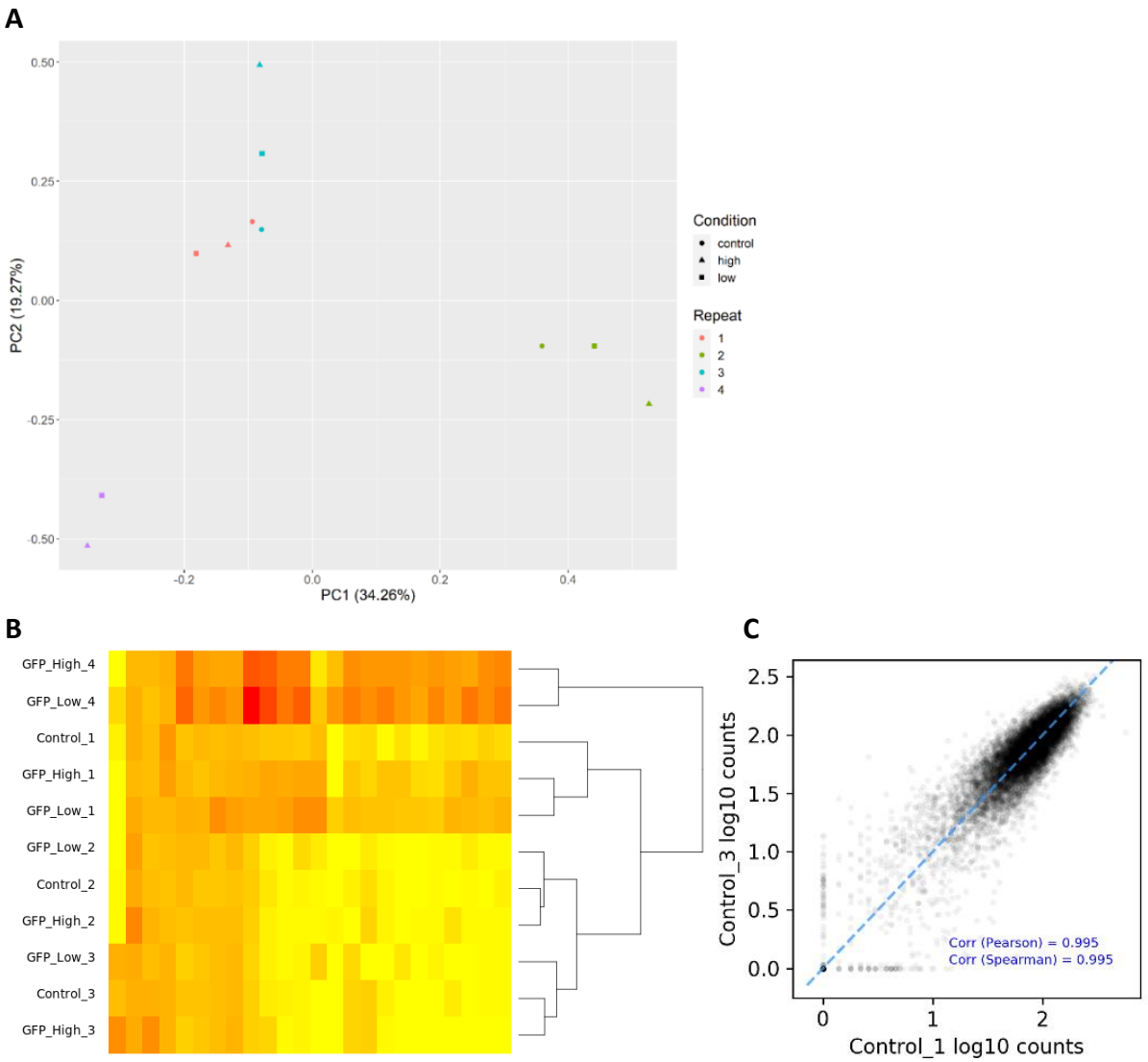


Figure 9. Correlation and clustering analysis. (A) Principal component analysis (PCA) of the samples based on the top 500 variable sgRNAs. Shapes represent condition and color for replicates. (B) Heatmap showing samples clustering and correlation based on the top 25 variable sgRNAs. Log10 normalized read counts are color-coded from lowest (yellow) to highest (red). (C) Scatter plot showing the normalized sgRNA read counts in one replicate of each condition versus another. Pearson and Spearman correlation coefficients are reported. Replicates 1 and 3 of the control are used for representation.

sgRNA enrichment analysis

Next, we performed statistical analysis to identified enriched sgRNAs in the GFP^{h} and GFP^{l} fractions relative to the control. The PinAPL-PY analysis used non-targeting sgRNAs provided by CRiNCL library for read counts normalization. Importantly the analysis identified all non-targeting sgRNAs as low scored, non-enriched genes. Which further support the conclusion that the experimental design was successful. Out of the 13,548 sgRNAs in the library, a total of 33 were significantly enriched with a threshold of $p < 0.01$, with 20 from the GFP^{l} fraction and 13 from the GFP^{h} fraction (Fig. 10A-C; Table S1). Interestedly, out of the 10 sgRNAs targeting the same gene, only one was statistically significant in the enrichment analysis (Fig. 10D). This result may be due the sensitivity of CRISPRi to the specific position of sgRNA in the genome, leading to high variability in KD efficiency.

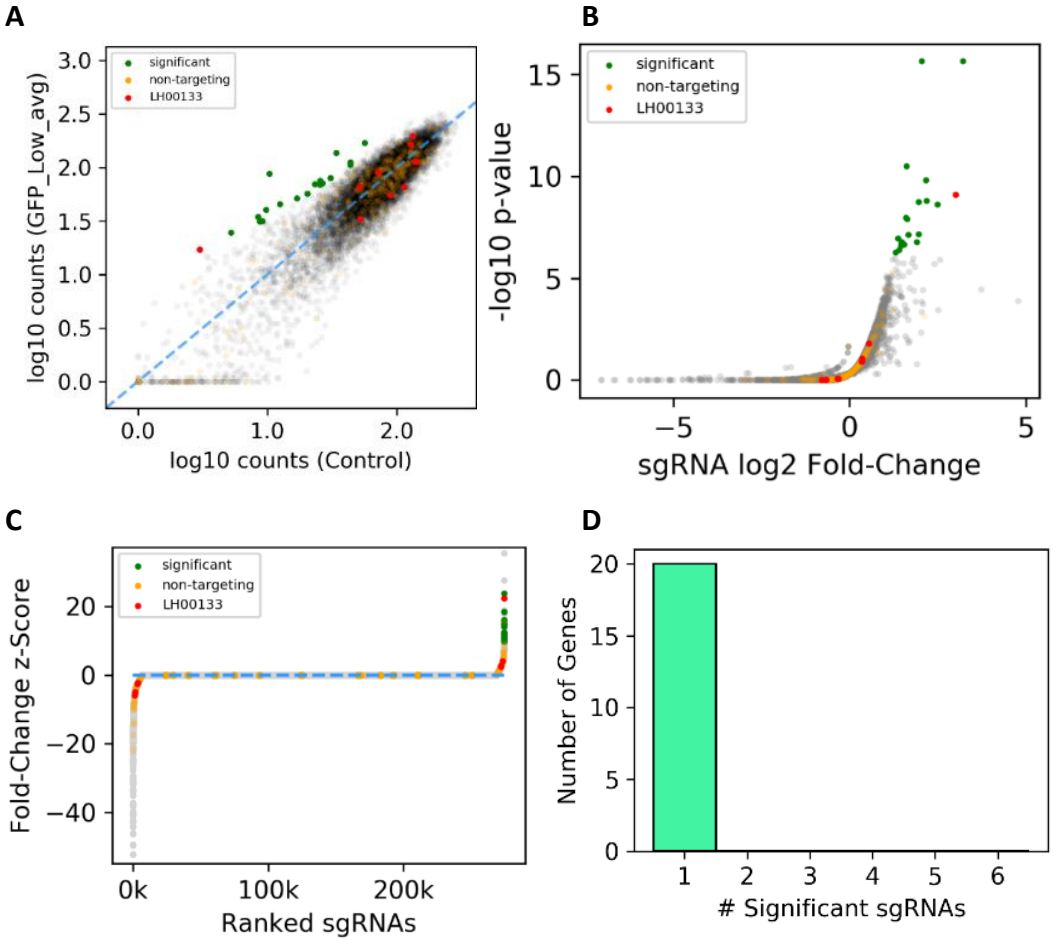


Figure 10. sgRNA enrichment analysis. (A-C) Significantly enriched sgRNAs are plotted in green, non-targeting controls are plotted in orange, and sgRNAs targeting the gene LH00133 (for example) are plotted in red. figures are taken from the GFP^{l} analysis, but also represent the GFP^{h} analysis. (A) Scatter plot of log10 normalized sgRNA read counts in the sample versus the average normalized count in the control. (B) Volcano plot of sgRNA -log10 p-value against log2 fold-change. (C) z-Score plot of the fold-change z-score for each sgRNA ranked from lowest to highest. The z-Score is the normalized deviation from the mean read count. (D) Bar plot showing the efficacy of sgRNAs targeting the same gene. Genes are categorized by the number of targeting sgRNAs that reached statistically significant enrichment.

MAGeCKFlute Analysis

A main limitation using PinAPL-Py is that it calculates the enrichment of each sgRNA as individual. Hence it provides limited indication at the gene level. To overcome this limitation, we furthered analyzed our dataset using MAGeCKFlute (B. Wang et al., 2019), which calculate the significant enrichment based on multiple sgRNAs targeting the same gene. The MAGeCK tool was used to performed reads alignment to the sgRNA library, generate a read-count table and test for screening quality control. The samples had a low Gini index (Fig. 11A), suggesting the sgRNA read counts distributed homogenously. The samples had a high percentage of mapped reads (Fig. 11B), indicating good DNA purity and sequencing. The CRISPR screen was performed in four batches, each done a different time point. This led to unintended variation in experimental conditions and a strong batch effect. The samples are clustered by batches instead of by conditions (Fig. 12). After using the ComBat function to remove the batch effect, the samples were properly clustered by condition (Fig. 12). MAGeCK RRA was used to identify gene hits by comparing the GFP^h and GFP^l conditions to the control. Significantly selected genes were identified and given an FDR score using the α-RRA algorithm. A total of 3 genes from the GFP^l condition were significantly enriched with a threshold of FDR < 0.25 (Fig. 13B ; Table S2). Out of the 10 sgRNAs targeting the top ranked genes, most were enriched in the treatment relative to the control (Fig. 13B). Interestingly, RP11-977G19.11 (gene ID LH03315) was identified by both analysis tools, indicating on high probability for functionality.

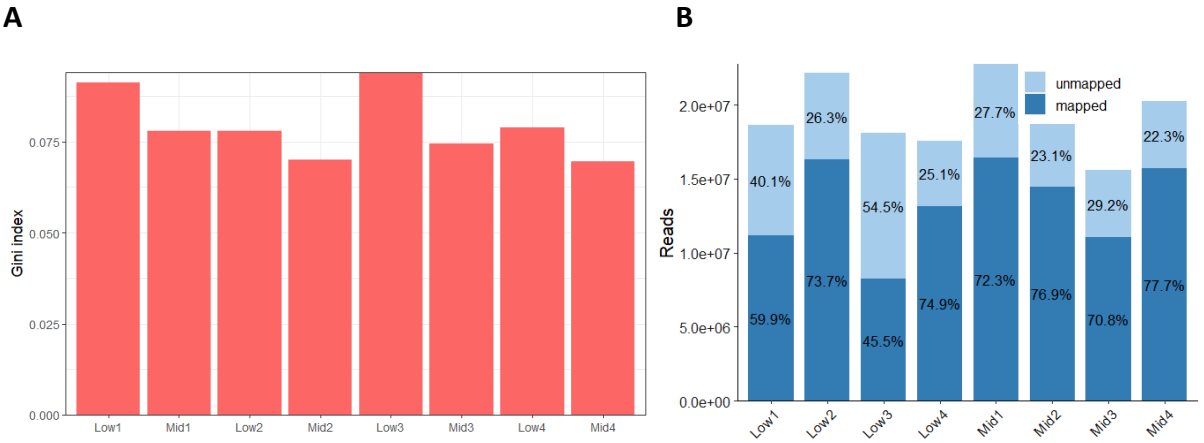


Figure 11. Screen quality control. (A) Evenness of sgRNA read counts in the samples. Measured by Gini index. (B) sgRNA reads count and mapped reads percentage. Read counts were normalized using non-targeting sgRNAs sequences.

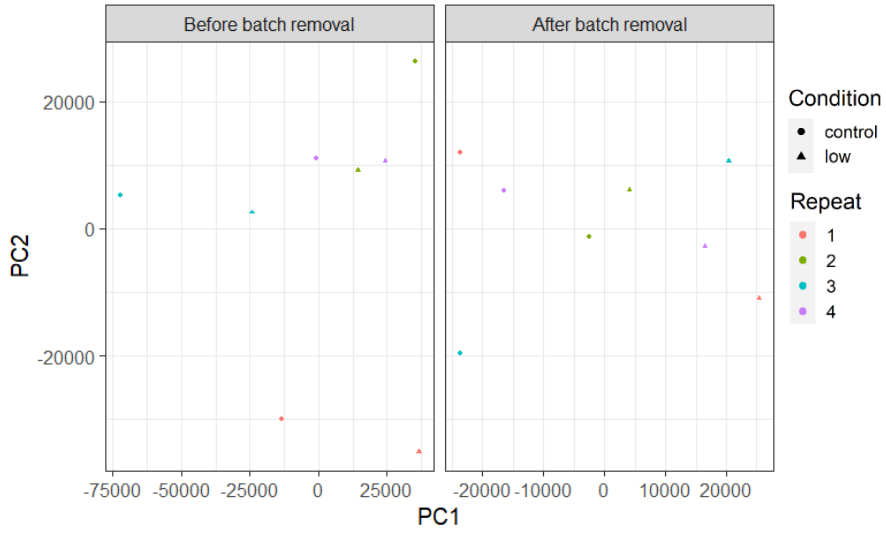
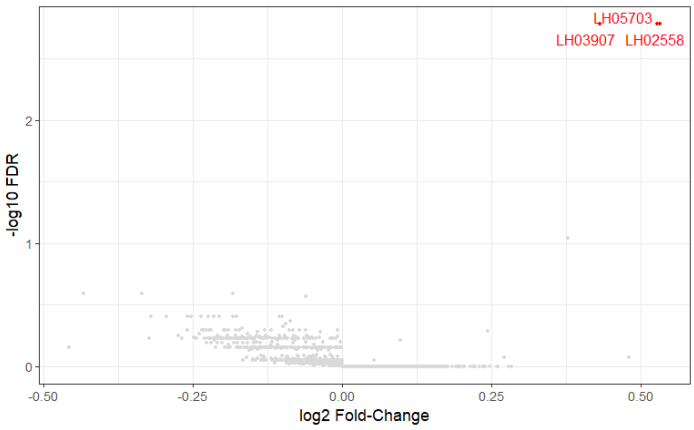


Figure 12. Batch effect removal. Principal component analysis (PCA) of the samples before and after batch removal. based on the topmost variable sgRNAs. Shapes represent condition and color for replicates. Batch effect removal was done using the ComBat function. Figures are taken from the GFP^{l} analysis, but also represent the GFP^{h} analysis.

A



B

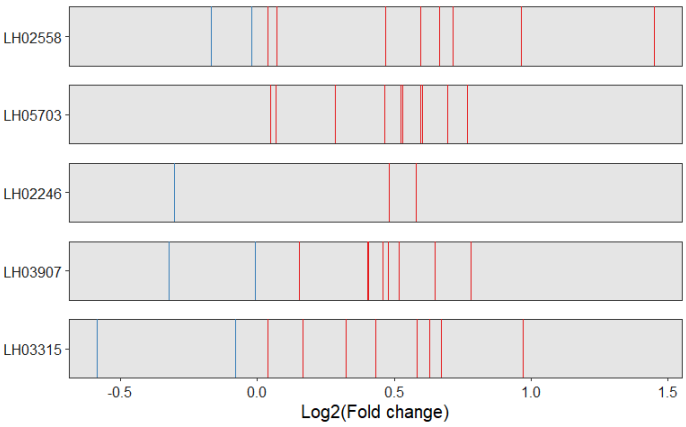


Figure 13. Candidate genes identification. (A) Volcano plot of genes -log10 FDR against log2 fold-change. (B) log2 fold-change of the 10 sgRNAs targeting the top 5 ranked genes. Each line represents one sgRNA. Positively and negatively selected genes are plotted in red and blue respectively.

Validation of candidate genes

To validate the screen in an arrayed format, we constructed individual plasmids containing all of the 33 significantly enriched sgRNAs based on the PinAPL-Py analysis. The plasmids were introduced into SNU449-KRAB-ACE2 cells using lentiviral transduction. The cells were then infected with single-cycle VSV displaying the Δ18-D614G spike variant and encoding a GFP reporter (Fig. 14). Infection efficiency was measured using flow cytometry. The validation was performed in 3 batches (Fig. 15A). The knockdown of RP11-977G19.11 and RP11-314A20.5 led to a 2-fold reduction in infection efficiency relative to the control, validating the genes effect observed in the screening. To further elucidate the effect of RP11-977G19.11 and RP11-314A20.5 on the infection, we separated the infected cells into GFP negative, high and low populations (Fig. 15B). While in the control sample most of the cells were in the GFP high population, in the KD samples the GFP high and low populations were relatively the same, with most of the cells being in the GFP negative population. Meaning that not only fewer cells were infected, the infected cells had less viral particles inside them.

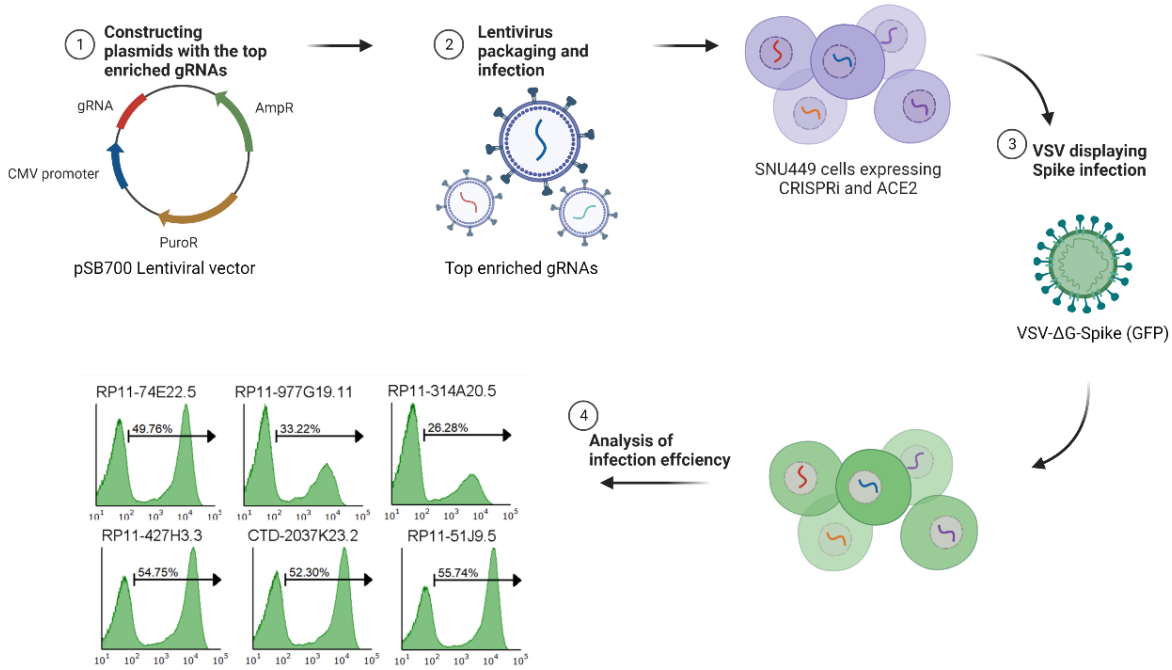


Figure 14. Schematic of the screen validation. The top 33 significantly enriched sgRNAs from the analysis were cloned into a PSB700 vector with a Puro resistance selection marker (1). The plasmids were introduced into SNU449-KRAB-ZIM3 cells by lentiviral transduction (2). The cells were infected with VSV-ΔG-Spike encoding GFP (3). Infection efficiency was measured by flow cytometry (4).

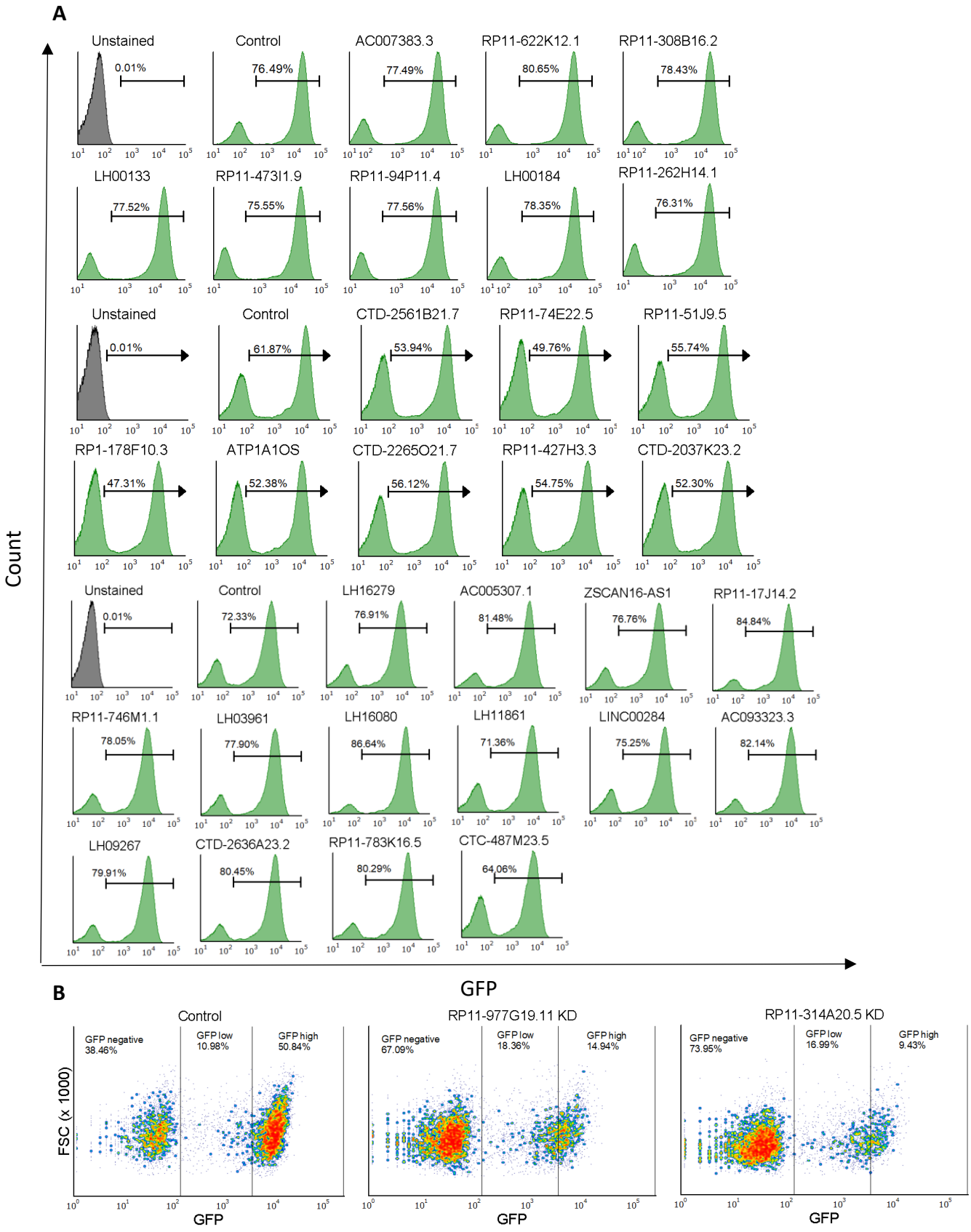


Figure 15. Infection in SNU449 cells with top genes knockdown. (A) VSV-ΔG pseudotyped with the Δ18-D614G spike variant was used to infect SNU449 cells stably expressing ACE2 and transduced with sgRNAs targeting the top genes. Cells transduced with a PSB700 empty vector were used for control. GFP intensity was measured using flow cytometry. The experiment was done in 3 batches. (B) GFP intensity of infected cells measured by flow cytometry. Gates separate the cells into 3 populations. Result represent 4 biological replicates.

Characterization of the validated hits

RP11-977G19.11 and RP11-314A20.5 regulate the expression of neighboring genes

The most common function of lncRNAs is regulate the expression of neighboring protein-coding genes. To test if RP11-977G19.11 and RP11-314A20.5 function in this manner, we used RT-qPCR to measure the relative RNA expression of neighboring genes in RP11-977G19.11 and RP11-314A20.5 KD cells (Fig. 16). RP11-314A20.5 KD led to a 2-fold reduction in the expression of all the tested genes, including CXCL16, whom level is correlated with the severity of COVID-19 (Smieszek et al., 2021). RP11-977G19.11 KD led to 8-fold reduction in the expression of CS (Citrate synthase). Considering CS major involvement in glycolytic energy production (Akram, 2013), this finding suggest a potential role for RP11-977G19.11 in regulating energy metabolism by which it might affect SARS-CoV-2 infection.

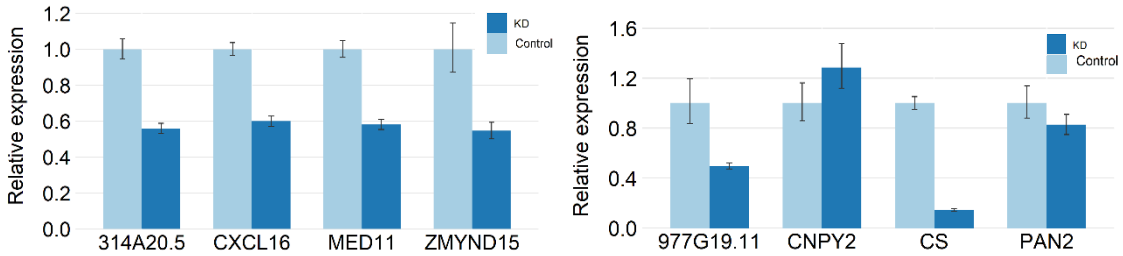


Figure 16. Expression of neighboring genes in KD cells. Relative RNA expression of targeted genes was quantified by RT-qPCR in SNU449-ACE2-KRAB cells with RP11-977G19.11 and RP11-314A20.5 KD. The expression was normalized to cells transduced with a PSB700 empty vector. n=1

RP11-977G19.11 and RP11-314A20.5 are involved in ACE2 regulating

Most of the discovered host factors critical for SARS-CoV-2 infection function at the viral entry level, usually by regulating ACE2 expression (Baggen, Vanstreels, et al., 2021). To test if RP11-977G19.11 and RP11-314A20.5 function in this manner, we performed indirect labeling using an anti-ACE2 antibody and a fluorophore-conjugated secondary antibody in KD cells and measured ACE2 level by flow cytometry. We used fixed cells to measure ACE2 level in the whole-cell (Fig. 17A), and unfixed cells to measure ACE2 level on the cell-surface (Fig. 17B). RP11-977G19.11 KD cells had lower ACE2 levels both in the whole cell and on the surface relative to the control, with a shift from high ACE2 expression to low. Surprisingly, we observed higher ACE2 levels in the RP11-314A20.5 KD cells relative to the control. These results suggest that RP11-977G19.11 might affect SARS-CoV-2 infection through ACE2 regulation, while RP11-314A20.5 affect the infection through a different mechanism.

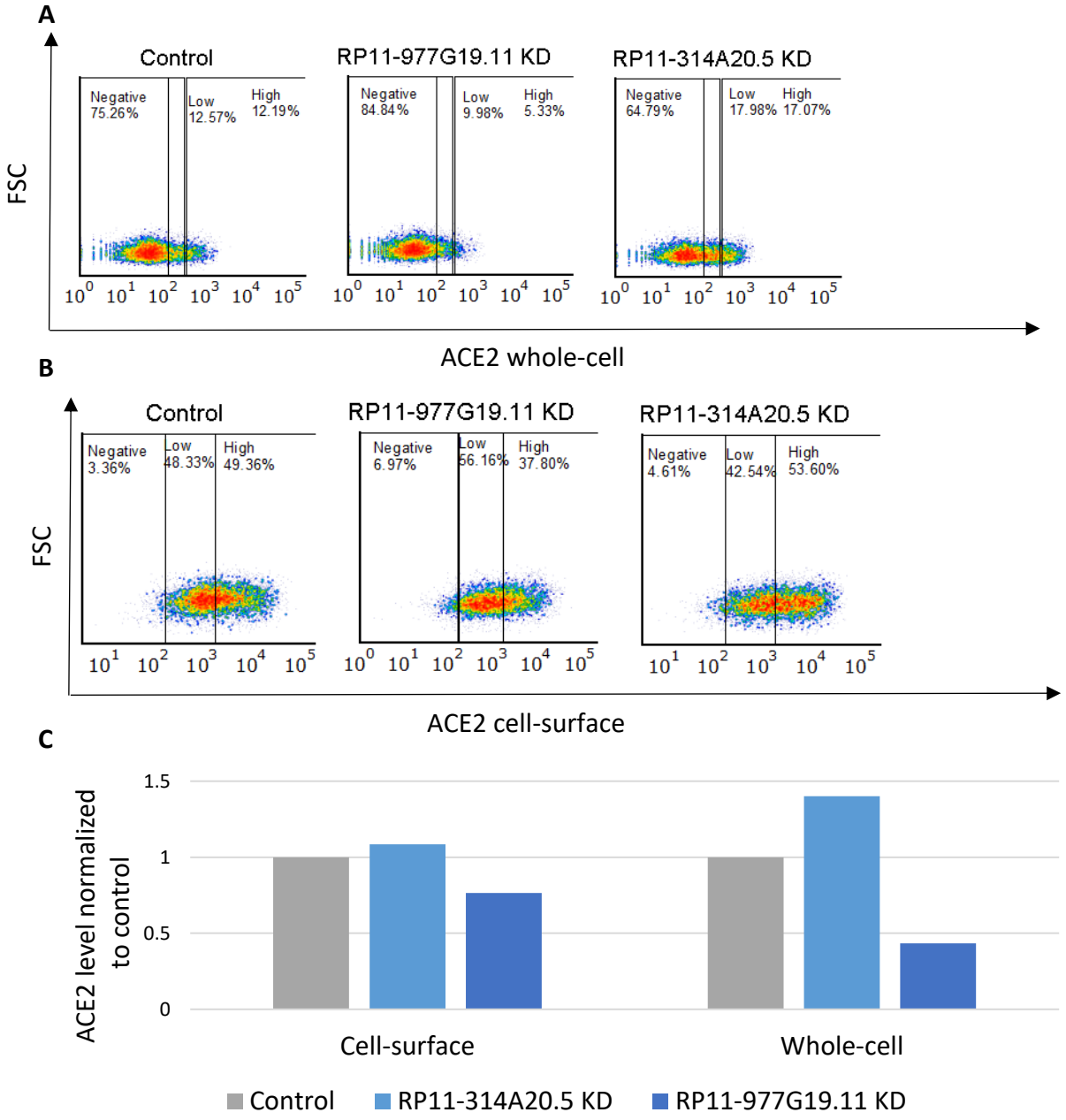


Figure 17. ACE2 expression level in KD cells. SNU449-ACE2-KRAB cells with RP11-977G19.11 and RP11-314A20.5 KD were stained using an anti-ACE2 antibody and an AF488-conjugated secondary antibody. AF488 intensity was measured by flow cytometry. Cell transduced with an empty PSB700 empty vector were used for control, and unstained cells were used for gating. Gates separate the cells into AF448 negative, low and high fractions.
(A) Cells fixed to measure ACE2 in the whole cell. n=1. (B) unfixed cells to measure ACE2 cell-surface expression. n=1. (C) Bar plot representation of the ACE2 high fraction (%) in the samples. The data was normalized to the control sample in the whole-cell and cell-surface experiments.

RP11-977G19.11 is not involved in ATP production

Considering RP11-977G19.11 modulating of CS expression, we hypothesized that RP11-977G19.11 could be an ATP regulator, which could explain its involvement in SARS-CoV-2 infection. To test this hypothesis, we measured the ATP level in RP11-977G19.11 KD cells using a cell viability assay, with the amount of ATP proportional to the luminescent signal (Fig. 18). Surprisingly, the ATP level was similar in the KD cells relative to the control, suggesting a different mechanism for RP11-977G19.11 effect on SARS-CoV-2 infection.

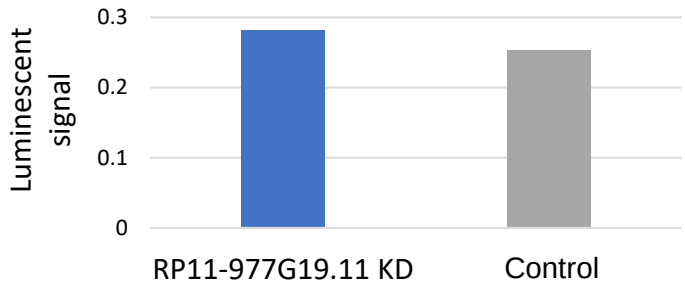


Figure 18. ATP level in KD cells. Luminescent signal proportional to the amount of ATP was measured in SNU449-ACE2-KRAB cells with RP11-977G19.11 KD using CellTiter-Glo® Luminescent Cell Viability Assay. Cells transduced with a non-targeting sgRNA (10010) were used for control. N=4

RP11-977G19.11 and RP11-314A20.5 effect is specific to VSV-ΔG-Spike

To test if RP11-977G19.11 and RP11-314A20 effect is specific to VSV displaying SARS-CoV-2 spike, or is it a general effect on the VSV pseudovirus model, we infected RP11-977G19.11 and RP11-314A20.5 KD cells with VSV displaying the native G protein (VSV-ΔG-G). Infection efficiency was measured by flow cytometry (Fig. 19). Surprisingly, efficiency was higher in RP11-977G19.11 and RP11-314A20.5 KD cells relative to the control. This result suggests that the lncRNAs effect on VSV entry depends on the viral envelope protein.

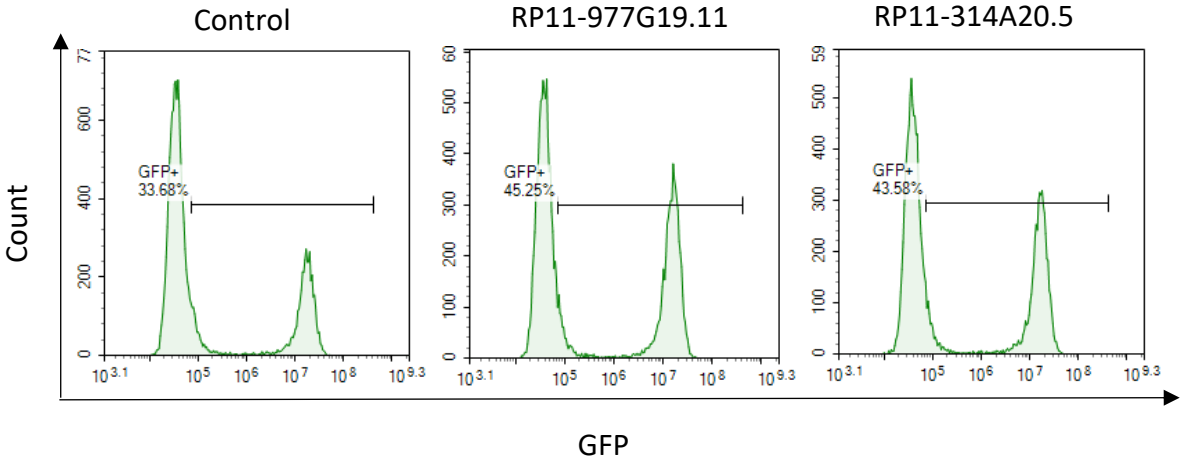


Figure 19. VSV-ΔG-G infection of KD cells. VSV-ΔG-G was used to infect SNU449-ACE2-KRAB cells with RP11-977G19.11 and RP11-314A20.5 KD. GFP intensity was measured using flow cytometry with unstained cells for gating. Cell transduced with an empty PSB700 empty vector were used for control. Result represent 4 biological replicates.

Results summary

In brief, we generated a stable expression of dCas9-KRAB in SNU449 and HEK293 cells. Using control sgRNAs, we observed an efficient repression in the SNU449 cells relative to the HEK293 (Fig. 1). We produced a stock of VSV-ΔG-G and measured the viral titer (Fig. 2). Next, we displayed the SARS-CoV-2 spike protein on the VSV surface (VSV-ΔG-Spike) (Fig. 3), and used the virus to infect HEK293 overexpressing ACE2 and TMPRSS2. The infection efficiency was measured by flow cytometry (Fig. 4A-B). Following a relatively low efficiency, we improved the infection by using a spike variant harboring Δ18 and D614G mutations (Fig. 4C-D), and found that SNU449 overexpressing ACE2 and TMPRSS2 are highly permissive to VSV-ΔG-Spike (Fig. 5). Based on these results, we decided to use SNU449 as the target cells in our screening. We performed a pooled CRISPR screen to identify lncRNA affecting SARS-

CoV-2 infection using the CRISPRi non-coding library. We introduced the library into SNU449-CRISPRi-ACE2, infected them with VSV-ΔG-Spike and sorted the cells into GFP^h, GFP^l and control fractions. Sequencing libraries were constructed from the fractions and subjected to NGS (Fig. 7). Following the sequencing, we performed two separated screen analyses. Both PinAPL-Py and MAGeCKFlute performed a sequencing and alignment quality control (Fig. 8 and 11). Correlation analysis by both tools showed a strong batch effect, which was corrected by the MAGeCK batch effect removal function (Fig. 9 and 12). Next, we performed an enrichment analysis, with PinAPL-PY scoring individual sgRNAs, and MAGeCK scoring genes based on the enrichment of multiple sgRNAs targeting the same gene. The PinAPL-PY analysis identified 33 significantly enriched sgRNAs (Fig. 10), and the MAGeCK analysis identified 3 significantly enriched genes (Fig. 13). We validated the hits in a similar manner to the screening workflow, but using individual plasmids with the top sgRNAs instead of the sgRNA library (Fig. 14). RP11-977G19.11 and RP11-314A20.5 were validated as proviral, having a strong effect on the infection efficiency (Fig. 15). Both validated genes were found to regulate the expression of neighboring protein-coding genes, with RP11-977G19.11 strongly regulating CS, and RP11-314A20.5 regulating CXCL16 (Fig. 16). RP11-977G19.11 had an effect on ACE2 expression both in the whole-cell and on the surface (Fig. 17), suggesting it might be involved in SARS-CoV-2 infection by ACE2 regulation. Considering CS involvement in energy metabolism, we tested if RP11-977G19.11 have an effect on the ATP levels in the cells. interestingly, the ATP level did not change in RP11-977G19.11 KD relative to the control (Fig. 18), suggesting energy production is not affected by CS reduced expression. Finnaly, both RP11-977G19.11 and RP11-314A20.5 had opposite effects on VSV infection depending on the envelope protein displayed on the virus (Fig. 19).

Overall, in my study I established a CRISPRi screening platform to study early stages of SARS-CoV-2 infection, and used it to identify novel lncRNAs specifically affecting SARS-Cov2 cell entry. This platform can be easily modified to study other viruses, hence may pave the way for better understanding of factors affecting viral infection, and potential therapeutic targets.

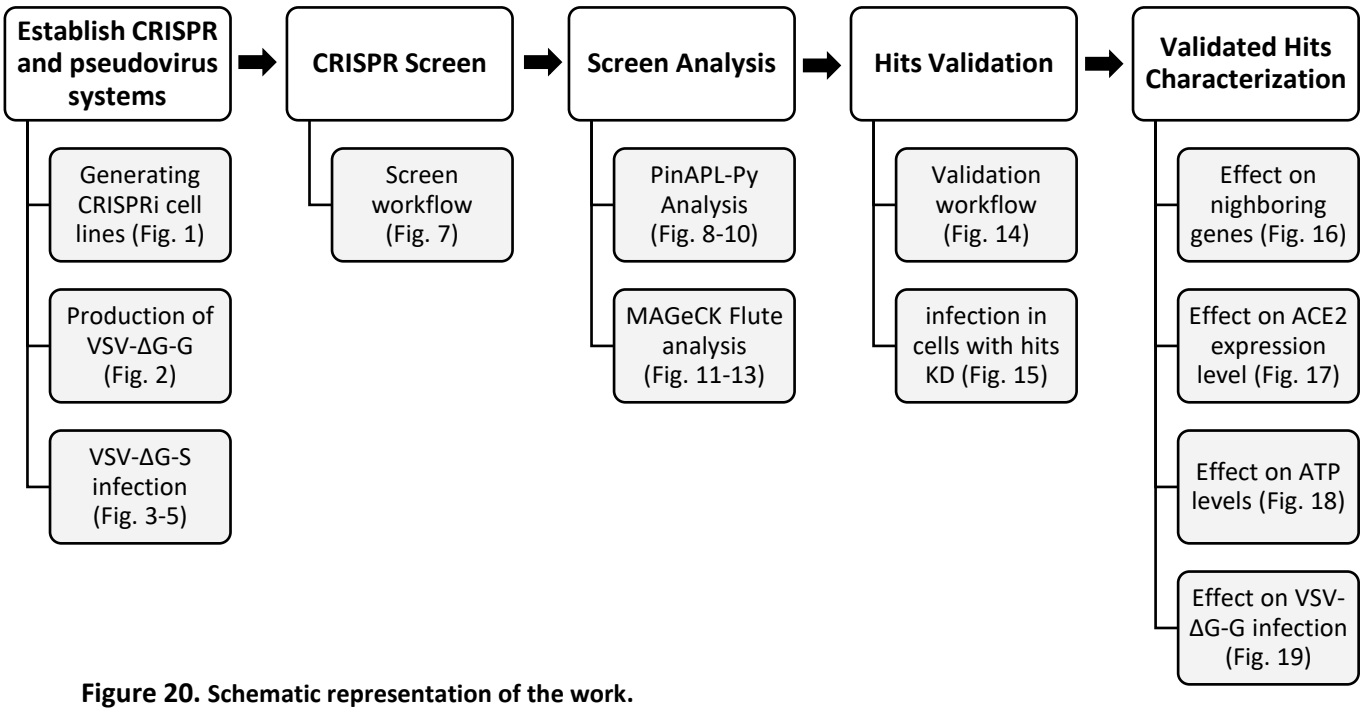


Figure 20. Schematic representation of the work.

Discussion

In this work, we designed and conducted a CRISPR screen for lncRNA genes critical for the SARS-CoV-2 infection in a system of VSV displaying SARS-CoV-2-Spike. The screen identified multipath genes with the potential to affect SAR-CoV-2 infection. Two proviral lncRNAs were validated: RP11-977G19.11 and RP11-314A20.5. My results showed that these genes affect infection, and inhibition of these genes reduced not only viral load, but also have a strong effect on the number of viral particles in the infected cells, shifting most of the cells to the GFP low fraction (Fig. 15B).

Specifically, RP11-977G19.11 was found to regulate the expression of the neighboring CS (Citrate synthetase). CS catalyzes the formation of citrate by combining Acetyl CoA and oxaloacetate. This is the first step of the Citric acid cycle, which is followed by the generation of ATP in the electron transport chain (Akram, 2013). Both VSV and SARS-CoV-2 replicate by utilizing the host ATP for RNA synthesis (Testa & Banerjee, 1979), (Codo et al., 2020). We hypothesize that RP11-977G19.11 modulates viral replication by regulating CS expression, leading to altered metabolism in the host cell that results in increased replication. This hypothesis is supported by RP11-977G19.11 effect on the number of viral particles in the cell. Interestingly, genome-scale metabolic models of Huh7 cells infected by SARS-CoV-2 showed downregulation of CS and the mitochondrial electron transport chain, with a shift in energy production by an upregulated glycolysis (Yaneske et al., 2021). This altered glucose metabolism is similar to the Warburg effect observed in cancer cells, and might also be involved in several steps of SARS-CoV-2 infection (Icard et al., 2021). Surprisingly, we did not see a change in ATP levels in RP11-977G19.11 KD cells relative to the control (Fig. 18). This finding suggests that a more complex mechanism for RP11-977G19.11 function. To confirm CS role in SARS-CoV-2 infection, we will use CRISPR-Cas9 to KO the gene in SNU449-ACE2 cells, followed by infection with VSV displaying spike. RP11-977G19.11 was also found to regulate ACE2 expression both on the cell surface and in the whole-cell (Fig. 16), suggesting its cellular function is involved in ACE2 regulation, which in turn affects SARS-CoV-2 infection.

On the other hand, RP11-314A20.5 regulate the expression of the neighboring CXCL16 (Chemokine (C-X-C motif) ligand 16). This chemokine, which belongs to the group of CXC chemokines, is expressed on the surface of antigen-presenting cells and was found to be a biomarker of COVID-19 (Blanco-Melo et al., 2020). The CXCR6/CXCL16 axis is involved in the inflammatory response and is associated with lung disease. Both CXCL16 and its receptor CXCR6 are highly expressed in the lung during inflammation and were shown to correlate with the severity of COVID-19 (Smieszek et al., 2022), (Payne et al., 2021). These recent findings suggest that RP11-314A20.5 have high therapeutic potential as a regulator of CXCL16 expression.

Our screen was carried out in SNU449 cells modified to overexpress ACE2 and TMPRSS2. High expression of TMPRSS2 drives the virus to enter the cells through the faster surface route, rather than the endosomal route (Koch et al., 2021). Both ACE2 and TMPRSS2 are required for SARS-CoV-2 infection, and are highly expressed in the lung tissue. interestingly, the liver tissue, from which the SNU449 cell line originate, also expresses elevated level of ACE2 and TMPRSS2, suggesting the liver may also be infected by SARS-CoV-2 (Baughn et al., 2020). Based on this data, we speculate that the SNU449 cell line, and the genes identified in our screen are physiologically relevant to SARS-CoV-2 infection. Regardless, it will be of great interest to validate our results in the Calu-3 cell line, which originate from the lung tissue and endogenous express ACE2 and TMPRSS2 (Park et al., 2021). Calu-3

cells were shown to be highly susceptible to SARS-CoV-2 infection (Chu et al., 2020), and behave similarly to Human airway epithelial cells upon infection (Rebendenne et al., 2021).

VSVs pseudotyped with envelope glycoproteins of pathogenic viruses are a safe and robust platform to study viral entry of enveloped viruses. Unlike SARS-CoV-2, that require biosafety level 3 (BSL-3), VSV can be used using BSL-2 containment. It possesses a small genome, that can easily be manipulated to replace the gene encoding the glycoprotein with different reporters. Finally, pseudotype viral particles can be produced at a very high yield, which is especially useful when conducting a large-scale screen (Whitt, 2010). The infection efficiency of VSV expressing SARS-CoV-2 spike correlate with ACE2 and TMPRSS2 expression, and it is affected by neutralization assays in a similar manner to SARS-CoV-2 (Dieterle et al., 2020). VSV has already been used in many different studies of SARS-CoV-2, including screening of inhibitors and neutralizing antibodies (Dieterle et al., 2020), vaccine development (Yahalom-Ronen et al., 2020) and several genetic screens for host factors (R. Wang et al., 2020), (Rebendenne et al., 2022), (Wei et al., 2020). Both RP11-314A20.5 and RP11-977G19.11 had opposite effects on VSV infection depending if the virus displayed SARS-CoV-2 spike or the native G protein (Fig. 19), suggesting that the lncRNAs are involved in the spike-mediated viral entry stage of the SARS-CoV-2 infection. An important corroborating of our results will be to infect RP11-314A20.5 and RP11-977G19.11 KD cells with SARS-CoV-2, preferably we will use a variant of concern such as the recently emerged Omicron (DeGrace et al., 2022).

To our knowledge, we performed the first CRISPR screen for lncRNAs host factors critical for the SARS-CoV-2 infection. Our screen discovered and validated 2 proviral lncRNAs. While we initiated the process of characterizing the genes, and discovered their function in regulating the expression of key protein-coding genes, the mechanism underlying how they regulate the infection remain elusive. Our finding will expend the understanding of SARS-CoV-2 pathogenesis, and will be relevant to both emerging variants as well as novel pathogenic CoVs. The validated lncRNAs could aid in the development of novel targeted therapies for COVID-19. lncRNAs are promising Therapeutic targets. Their cell-specific expression can be exploited to design low toxicity treatments (Renganathan & Felley-Bosco, 2017). lncRNAs can be relatively easily targeted using various approaches, such as transcript degradation by siRNA, Gapmers and ASOs, or interaction blocking by small molecules and Aptamers (Pandya et al., 2020). Finally, our pseudovirus and CRISPR systems can be smoothly modified to study different pathogenic viruses, such as HIV and Influenza.

Materials and Methods

Key Resources Table

Virus Strains	
VSV-ΔG-G (GFP)	Laboratory of Prof. Benjamin Podbilewicz
VSV-ΔG-Spike (GFP)	This study
Cell lines	
HEK-293T	Laboratory of Prof. Pandolfi
SNU-449	Laboratory of Prof. Pandolfi
Vero-E6	Laboratory of Dr. Ori Avinoam
BHK-21/WI-2	Laboratory of Prof. Benjamin Podbilewicz
Antibodies	
Anti-ACE2	Abcam #EPR24705-45
Anti-Rabbit FITC	Laboratory of Dr. Ron-Harel
Anti-Rabbit AF488	Jackson #111-545-144
Chemicals	
DMEM media	Biological Industries
RPMI media	Biological Industries
Fetal bovine serum (FBS)	Biological Industries
PolyJet	BioConsult
Polybrene	Sigma #TR-1003-G
TRIzol reagent	Sigma
SYBR Green	Tamar Laboratory Supplies #PB20.15-50
Plasmids	
CRISPRi non-coding library (CRiNCL)	Jonathan Weissman. Addgene #86538
SARS2-Spike	Laboratory of Prof. Benjamin Podbilewicz
hACE2-Blast	Laboratory of Dr. Ori Avinoam
SARS2-Spike-Δ18	Laboratory of Prof. Alon Herschhorn
SARS2-Spike-Δ18-D614G	Laboratory of Prof. Alon Herschhorn
dCas9-KRAB-BFP	Addgene
dCas9-KRAB-mCherry-ZIM3	Addgene
dCas9-KRAB-P2A-mCherry	Addgene
pSB700-10010-mCherry	Addgene
psPAX	Addgene
VSV-G	Addgene
Lenti-ACE2-TMPRSS2-Blast	Addgene
Commercial assays	
qScript cDNA Synthesis Kit	Quantabio #95047-500
Quick-DNA Miniprep Plus Kit	Zymo #D4068
Q5 High-Fidelity 2X Master Mix	NEB #M0492
NucleoSpin® Gel and PCR Clean-up kit	MN #740609
CellTiter-Glo® Luminescent Cell Viability Assay	Promega #G7570
Software	
PinAPL-Py	Spahn et al. 2017
MAGeCKFlute	Wang et al. 2019
R and RStudio	R Core Team
Python	Python Software Foundation
FCS Express	De Novo
BioRender	BioRender

Methods

Cell culture. HEK-293T and SNU-449 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) or Roswell Park Memorial Institute (RPMI), respectively. Supplemented with 10% fetal bovine serum (FBS), 1% Penicillin/Streptomycin, and 1% L-Glutamine. The cells were passaged every 2-3 days using Trypsin at ratios 1:2 to 1:10 and incubated at 37°C with 5% CO2. Cell count and viability were determined by trypan blue using LUNA Automated Cell Counter (Logos Biosystems). All cell lines were tested routinely for mycoplasma contamination.

Generation of a stable cell line expressing ACE2 and CRISPR. Lentiviral vectors packaging the ACE2-TMPRSS2-Blast and dCas9-mCherry-ZIM3-KRAB plasmids were produced in HEK-293T cells using PolyJet transfection reagent with psPAX and VSV-G plasmids according to manufacturer's instructions. At 48 h post-transfection, supernatants were harvested, concentrated by spinning at 1200 rpm for 5 min, and filtered through a 0.45 µm filter. SNU- 449 cells seeded the last day at 3 x 10^5 cells/well in a 6-well plate were transduced with the lentiviral vectors supplemented with 10 µg/ml of Polybrene. The cell lines were enriched using antibiotic selection or FACS on the FACSAria III (BD Biosciences). The expression of ACE2 was confirmed by RT-qPCR and flow cytometry on the BD LSR-II (BD Biosciences).

Pseudovirus production and tittering. VSV-ΔG-G stock was generated by transfecting BHK-21/WI-2 cells with VSV-G plasmid using PolyJet according to the manufacturer's instructions. At 24 h post-transfection, the media was removed and replaced with media supplemented with VSV-ΔG-G encoding GFP. The cells were incubated for 1h and free viruses were washed twice with PBS. At 24 h post-infection, the supernatant was harvested, and cell debris was cleared by spinning at 2500 rpm for 10 min at 4°C. The supernatant was aliquoted for storage at –80°C. To titer the VSV-ΔG-G virions, BHK-21 were seeded at 7.5 x 10^3 cells/well in a 96-well plate. The following day, the media was replaced with 10-fold serial dilutions of the virus ranging from 10^{-2} to 10^{-9}. At 24 h post-infection, cells were manually counted using a fluorescence microscope to calculate viral titer.

SARS-CoV-2 Spike pseudotype infection. VSV-ΔG-Spike virions were produced as described above, apart from transfection HEK-293T cells with Spike-Δ18-D614G plasmid. SNU-449-ACE2 were seeded at 3 x 10^5 cells/well in a 6-well plate. The following day, the media was removed and replaced with the virus. The infection efficiency was measured 24 h post-infection using flow cytometry on the BD LSR-II (BD Biosciences) instrument.

Pooled CRISPR screen. Jonathan Weissman's CRISPRi non-coding library (CRiNCL) contains 10 sgRNAs per TSS, targeting lncRNA genes common to seven cell lines (Addgene #86538). The library was delivered to SNU-449-KRAB-ACE2 cells by lentiviral transduction at ~0.3 MOI. The library representation was kept at a 1000 cells/sgRNA ratio throughout the screening. Four days post-transduction, 2 ug/ml puromycin was added to the media, and the cells were selected for ten days. The cells were infected with the VSV-ΔG-S pseudotype. at 24 h post-infection, the cells were sorted on the FACSAria III Cell Sorter (BD Biosciences). Three fractions of cells were collected: the high and low 10% of GFP expressing cells, and a control fraction of cells with medium GFP expression. For all fractions, the number of cells collected was sufficient to maintain the library representation. The screen was performed in four biological replicates.

Illumina sequencing. Genomic DNA was extracted from the cells using Quick-DNA Miniprep Plus Kit according to manufacturer instructions. To construct Illumina libraries, PCR was performed using Q5 High-Fidelity 2X Master Mix with Illumina compatible Primers to amplify the sgRNA inserts and appending Illumina adaptors and barcodes to the amplicons. The amplicons were loaded on 2% agarose gel to determine the optimal cycle number. Cycling conditions:

Step	Temp	Time	Cycles
Initial Denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	24
Annealing	65°C	75 sec	
Final Extension	65°C	5 min	1
Hold	4°C	∞	

The amplicons were purified using NucleoSpin Gel and PCR Clean-up kit and the purified libraries' quality and concentration were assessed on Qubit and Tapestation. The libraries were then pooled and concentrated. The pooled library was sequenced as single-read 50 bp reads on the NextSeq2000 (Illumina). Jonathan Weissman library preparation protocol suggest 250-500 reads per sgRNA per sample. Our library contained 13K sgRNAs with 12 samples. Thus, we required > 40M reads dedicated for the sgRNAs' sequences, with 30% of the reads dedicated to PhiX sequencing control. We performed the sequencing with 400M reads, which was far above the required number.

PinAPL-Py screen analysis. Sequencing read alignment, read counting, quality control, and sgRNA enrichment analysis of the GFP high, low, and control fractions were carried out using the PinAPL-Py web application (Spahn et al., 2017). All analysis parameters were left in the default setting. Read counts were normalized using 258 non-targeting control sgRNAs provided in the CRiNCL library and the CPM method. p-value adjustment was done using Sidak correction method and the Significance threshold for sgRNA ranking was 0.01. R was used to further assess the sequencing quality by performing Principal Component Analysis (PCA) of the 500 sgRNAs with the top variance and reviewing the read counts distribution using a box plot.

MAGeCKFlute screen analysis. Sequencing read alignment, read counting, quality control, batch effect removal and candidate genes identification was carried out using the MAGeCKFlute tool (B. Wang et al., 2019). The mageck count function was used to generate a read-count table. Read counts were normalized using 258 non-targeting sgRNAs. The ComBat function was used in R to correct the batch effect in the dataset. The mageck test function was used to generate gene and sgRNA ranking using the MAGeCK RRA method. Which rank sgRNAs and genes based on P-value and FDR, and use a modified RRA algorithm to identify positively and negatively selected genes.

Plasmid construction. The 33 significant sgRNAs from the PinAPL-Py analysis were selected for validation. sgRNA oligonucleotides ordered from IDT were annealed and the PSB700-puro plasmid was digested. The sgRNAs were ligated into the digested plasmid. The Ligated plasmids were then transformed into E. coli DH5α competent cells through heat-shock, which were then selected by ampicillin. Colony PCR with backbone-specific primers was used to select positive colonies. DNA from the positive colonies was extracted using Quick-DNA Miniprep Plus Kit according to the manufacturer's instructions.

Screen validation. Lentiviral vectors packaging the top sgRNAs were produced in HEK-293T cells as described above. SNU-449-KRAB-ACE2 cells were transduced and selected with 2 ug/ml Puromycin for 10 days. The selected cells were infected with VSV-ΔG-S and infection efficiency was measured 24 h post-infection using flow cytometry on the BD LSR-II. Cells transduced by a non-targeting sgRNA were used for control. The infection was performed in two biological replicates.

RT-qPCR. RNA was extracted from the cells using TRIzol reagent according to manufacturer instructions. The RNA was reverse transcribed using qScript cDNA Synthesis Kit and subjected to qPCR to measure the relative expression of tested genes. The relative expression was normalized to that of GAPDH, PGK1, and PPIB.

ACE2 antibody-binding assay. SNU-449-KRAB-ACE2 and naïve cells were collected and incubated with an anti-ACE2 antibody (1:500) at 4 °C for 30 min. The cells were washed with ice-called PBS supplemented with 10% FBS and 0.02% Sodium Azide. The cells were incubated with an Anti-rabbit antibody conjugated with FITC (1:2000) at 4 °C for 30 min. The cells were washed and subjected to flow cytometry analysis on the BD LSR-II instrument.

ATP quantification. ATP was measured using the CellTiter-Glo® Luminescent Cell Viability Assay following manufacturer instructions. An ATP standard curve was generated using ATP at serial dilutions ranging from 1 μM to 10 nM. The amount of ATP is proportional to the luminescent signal.

Supplementary data

Figure S1. FACS gating strategies

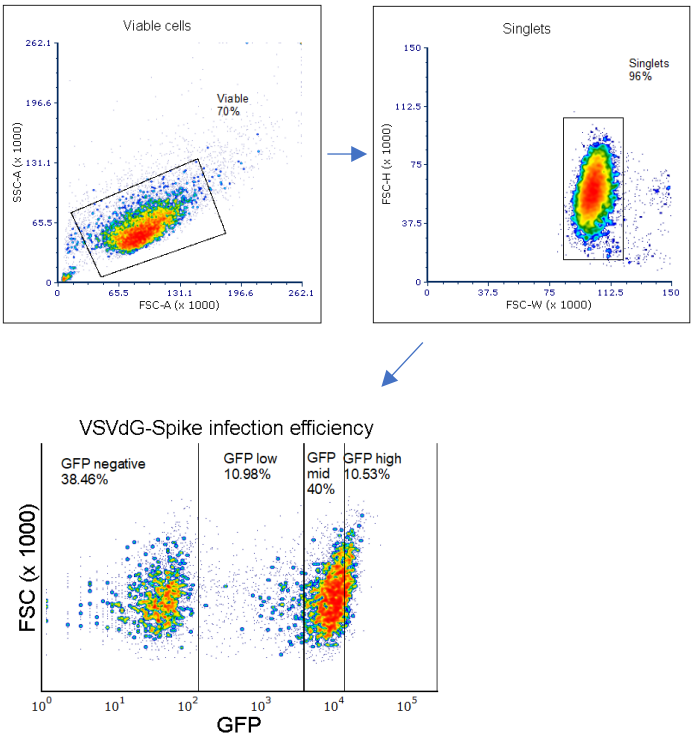
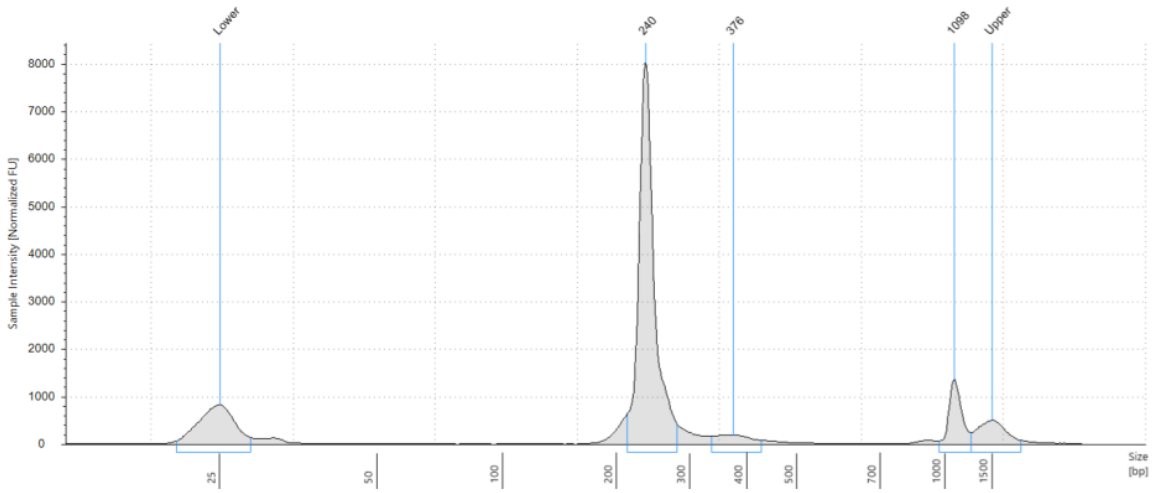


Figure S2. Pooled library Tapestation result



Sample Table

Well	Conc. [pg/µl]	Sample Description	Alert	Observations
A1	2610	Lib		

Peak Table

Size [bp]	Calibrated Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	% Integrated Area	Peak Comment	Observations
25	531	-	32700	-		Lower Marker
240	2180	-	14000	83.53		
376	120	-	490	4.58		
1098	311	-	435	11.88		
1500	250	250	256	-		Upper Marker

Table S1. Significantly enriched sgRNAs identified by the PinAPL-Py analysis

Gene_ID	Gene_Name	sgRNA_Protospacer	Condition	p-value	chromosome
LH15935	RP11-262H14.1	GGAGCAATGAGCCGAGACCC	GFP_Low	1.11E-16	chr9
LH09070	AC007383.3	GCCTAGAGCCCCCCAGGTTG	GFP_Low	1.11E-16	chr2
LH16668	RP11-622K12.1	GATGGCGGCGAGTCCCCTCC	GFP_Low	2.99E-11	chrX
LH12892	RP11-308B16.2	Gcagaggaggcgccgagcga	GFP_Low	1.52E-10	chr5
LH00133	NA	Gctctccaggatgtggaaca	GFP_Low	7.91E-10	chr1
LH05545	RP11-473I1.9	GCCCACAACCCTCTTAATAA	GFP_Low	1.54E-09	chr16
LH03021	RP11-94P11.4	Gccaaggcgggcggacgaag	GFP_Low	1.72E-09	chr11
LH00184	NA	GCTGTGGCACCACTCAGGGT	GFP_Low	2.39E-09	chr1
LH16279	NA	Gggagccaagtcagaggaag	GFP_Low	1.07E-08	chr9
LH07178	CTD-2561B21.7	GCTCACGGCTCTGTCCACTA	GFP_Low	1.20E-08	chr17
LH06720	RP11-74E22.5	GCTCTTGAGAGTGCATGCCA	GFP_Low	6.63E-08	chr17
LH03315	RP11-977G19.11	GGAACCAAGGTGAGCAGCGA	GFP_Low	7.21E-08	chr12
LH16781	NA	Gcccctgcacagagtcatca	GFP_Low	1.11E-07	chrX
LH06726	RP11-314A20.5	GACCGACCTGGGTGAGGTAG	GFP_Low	1.64E-07	chr17
LH11861	NA	GCGCGCCGCCAGCTCGGACT	GFP_Low	1.66E-07	chr4
LH06794	RP1-178F10.3	GGCCTCTGCCATGTTCACAC	GFP_Low	1.87E-07	chr17
LH08290	AC005307.1	Gccacctggaaaggtctcag	GFP_Low	2.21E-07	chr19
LH13802	ZSCAN16-AS1	GATTGGGGGCTCTTTCGAAG	GFP_Low	2.51E-07	chr6
LH07459	RP11-17J14.2	GCCCCTGGAGGGGTACAGTG	GFP_Low	4.00E-07	chr18
LH06355	RP11-746M1.1	GGTGATTACTAGGCAACGGG	GFP_Low	5.35E-07	chr17
LH01196	ATP1A1OS	GCTCACACACCCCACGAGGT	GFP_High	1.09E-13	chr1
LH03961	NA	GCCAGCCCTCTCTAGATTGC	GFP_High	1.59E-09	chr12
LH07682	CTD-2265O21.7	GGAGCCTTGCTGGGTTGAGG	GFP_High	6.56E-09	chr19
LH16080	NA	GCCCAGCCTAAGCACCGTGA	GFP_High	8.08E-09	chr9
LH09289	RP11-427H3.3	GCAGACCCCTCCAGAGTACC	GFP_High	1.21E-08	chr2
LH04048	LINC00284	GTATAAAGGAGCTAAAGATG	GFP_High	8.81E-08	chr13
LH13004	CTD-2037K23.2	GAGCCCGGAGCGTGGGATCC	GFP_High	1.26E-07	chr5
LH15508	RP11-51J9.5	GGCGCACAGGACTGTACTTT	GFP_High	2.71E-07	chr8
LH12558	CTD-2636A23.2	GGCCTGCCCATTGCCGGCAA	GFP_High	3.05E-07	chr5
LH09267	NA	Gtggtacctgaggtaggaca	GFP_High	3.99E-07	chr2
LH11808	AC093323.3	Gtgcctttatgagccgtaac	GFP_High	5.42E-07	chr4
LH02478	RP11-783K16.5	GCACGCCAAGATGGAGCTCC	GFP_High	5.84E-07	chr11
LH12674	CTC-487M23.5	GGTGCTCAGTAAACAAGCTT	GFP_High	7.24E-07	chr5

Table S2. Significantly enriched genes identified by the MAGeCKFlute analysis

Gene_id	Gene_name	score	p-value	fdr	rank	goodsgRNA	lfc
LH05703	RP11-410D17.2	7.73E-09	4.21E-06	0.00165	1	8	0.52799
LH02558	AP001372.2	1.63E-07	4.21E-06	0.00165	2	6	0.53277
LH03907	NA	2.16E-06	4.21E-06	0.00165	3	8	0.43221
LH03315	RP11-977G19.11	1.36E-05	0.000307	0.090347	4	7	0.37894
LH15897	RP11-195F19.9	0.000289	0.002513	0.512376	5	7	0.24343
LH03804	RP11-611O2.5	0.000301	0.002614	0.512376	6	7	0.24373
LH16668	RP11-622K12.1	0.000418	0.003667	0.615983	7	4	0.098021
LH00395	LAMTOR5-AS1	0.000433	0.007556	0.888614	8	8	0.052904
LH05881	RP11-20I23.6	0.000612	0.005931	0.841034	9	6	0.27157

Table S3. qPCR primers

Target gene	Forward primer	Reverse primer
hACE2 primer pair 1	tccattggtcttctgtcacccg	agaccatccacctccacttctc
hACE2 primer pair 2	acagtccacacttgcccaaat	tgagagcactgaagacccatt
RP11-977G19.11	GGAGAAGGGTGCTGATCAGG	TGCAGGCTTATTCCCCATGC
RP11-314A20.5	GTGAGGTAGCGGATCTGAGC	ACTAACGAGCGGCTCCTAGA
CNPY2	GAGGAGCCAGGATCTCCACT	TCATCATGCGATATGTGCAG
Citrate Synthase	CGTTTCCGAGGTCATAGTATCCC	GCTGAGACATAGGGTGTAGGTTGG
PAN2	ACCCTGATGGTAGCAAAAGTGAT	TGTTGCGGGTCTGAATCGTG
MED11	CTACCTCACCCAGGTGGC	TCTCCCTGGCCTAGTTCTCC
ZMYND15	TGGAGAACGAAACAAGAC	CACAAGTAAGAGACAGGAA
CXCL16	GAGCTCACTCGTCCCAATGAA	TCAGGCCCAACTGCCAGA

Reference

Akram, M. (2013). Citric Acid Cycle and Role of its Intermediates in Metabolism. *Cell Biochemistry and Biophysics 2013 68:3*, *68*(3), 475–478. https://doi.org/10.1007/S12013-013-9750-1

Alerasool, N., Segal, D., Lee, H., & Taipale, M. (2020). An efficient KRAB domain for CRISPRi applications in human cells. *Nature Methods*, *17*(11), 1093–1096. https://doi.org/10.1038/s41592-020-0966-x

Aliperti, V., Skonieczna, J., & Cerase, A. (2021). Long Non-Coding RNA (lncRNA) Roles in Cell Biology, Neurodevelopment and Neurological Disorders. *Non-Coding RNA 2021, Vol. 7, Page 36*, *7*(2), 36. https://doi.org/10.3390/NCRNA7020036

Baggen, J., Persoons, L., Vanstreels, E., Jansen, S., Van Looveren, D., Boeckx, B., Geudens, V., De Man, J., Jochmans, D., Wauters, J., Wauters, E., Vanaudenaerde, B. M., Lambrechts, D., Neyts, J., Dallmeier, K., Thibaut, H. J., Jacquemyn, M., Maes, P., & Daelemans, D. (2021). Genome-wide CRISPR screening identifies TMEM106B as a proviral host factor for SARS-CoV-2. *Nature Genetics 2021 53:4*, *53*(4), 435–444. https://doi.org/10.1038/s41588-021-00805-2

Baggen, J., Vanstreels, E., Jansen, S., & Daelemans, D. (2021). Cellular host factors for SARS-CoV-2 infection. *Nature Microbiology 2021 6:10*, *6*(10), 1219–1232. https://doi.org/10.1038/s41564-021-00958-0

Baughn, L. B., Sharma, N., Elhaik, E., Sekulic, A., Bryce, A. H., & Fonseca, R. (2020). Targeting TMPRSS2 in SARS-CoV-2 Infection. *Mayo Clinic Proceedings*, *95*(9), 1989–1999. https://doi.org/10.1016/J.MAYOCP.2020.06.018

Bester, A. C., Lee, J. D., Chavez, A., Lee, Y. R., Nachmani, D., Vora, S., Victor, J., Sauvageau, M., Monteleone, E., Rinn, J. L., Provero, P., Church, G. M., Clohessy, J. G., & Pandolfi, P. P. (2018). An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance. *Cell*, *173*(3), 649-664.e20. https://doi.org/10.1016/j.cell.2018.03.052

Biering, S. B., Sarnik, S. A., Wang, E., Zengel, J. R., Leist, S. R., Schäfer, A., Sathyan, V., Hawkins, P., Okuda, K., Tau, C., Jangid, A. R., Duffy, C. V., Wei, J., Gilmore, R. C., Alfajaro, M. M., Strine, M. S., Nguyenla, X., Van Dis, E., Catamura, C., … Hsu, P. D. (2022). Genome-wide bidirectional CRISPR screens identify mucins as host factors modulating SARS-CoV-2 infection. *Nature Genetics 2022*, 1–12. https://doi.org/10.1038/s41588-022-01131-x

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Møller, R., Jordan, T. X., Oishi, K., Panis, M., Sachs, D., Wang, T. T., Schwartz, R. E., Lim, J. K., Albrecht, R. A., & tenOever, B. R. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, *181*(5), 1036-1045.e9. https://doi.org/10.1016/J.CELL.2020.04.026

Bridges, M. C., Daulagala, A. C., & Kourtidis, A. (2021). LNCcation: lncRNA localization and function. *Journal of Cell Biology*, *220*(2). https://doi.org/10.1083/JCB.202009045/211695

Cagno, V. (2020). SARS-CoV-2 cellular tropism. *The Lancet Microbe*, *1*(1), e2–e3. https://doi.org/10.1016/S2666-5247(20)30008-2

Cevik, M., Kuppalli, K., Kindrachuk, J., & Peiris, M. (2020). Virology, transmission, and pathogenesis of SARS-CoV-2. *BMJ*, *371*. https://doi.org/10.1136/BMJ.M3862

Chu, H., Chan, J. F.-W., Yuen, T. T.-T., Shuai, H., Yuan, S., Wang, Y., Hu, B., Yip, C. C.-Y., Tsang, J. O.-L., Huang, X., Chai, Y., Yang, D., Hou, Y., Chik, K. K.-H., Zhang, X., Fung, A. Y.-F., Tsoi, H.-W., Cai, J.-P., Chan, W.-M., … Yuen, K.-Y. (2020). Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations,

transmissibility, and laboratory studies of COVID-19: an observational study. *The Lancet. Microbe*, *1*(1), e14. https://doi.org/10.1016/S2666-5247(20)30004-5

Codo, A. C., Davanzo, G. G., Monteiro, L. de B., de Souza, G. F., Muraro, S. P., Virgilio-da-Silva, J. V., Prodonoff, J. S., Carregari, V. C., de Biagi Junior, C. A. O., Crunfli, F., Jimenez Restrepo, J. L., Vendramini, P. H., Reis-de-Oliveira, G., Bispo dos Santos, K., Toledo-Teixeira, D. A., Parise, P. L., Martini, M. C., Marques, R. E., Carmo, H. R., … Moraes-Vieira, P. M. (2020). Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis. *Cell Metabolism*, *32*(3), 437-446.e5. https://doi.org/10.1016/J.CMET.2020.07.007

Daniloski, Z., Jordan, T. X., Wessels, H. H., Hoagland, D. A., Kasela, S., Legut, M., Maniatis, S., Mimitou, E. P., Lu, L., Geller, E., Danziger, O., Rosenberg, B. R., Phatnani, H., Smibert, P., Lappalainen, T., tenOever, B. R., & Sanjana, N. E. (2021). Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells. *Cell*, *184*(1), 92-105.e16. https://doi.org/10.1016/J.CELL.2020.10.030

DeGrace, M. M., Ghedin, E., Frieman, M. B., Krammer, F., Grifoni, A., Alisoltani, A., Alter, G., Amara, R. R., Baric, R. S., Barouch, D. H., Bloom, J. D., Bloyet, L. M., Bonenfant, G., Boon, A. C. M., Boritz, E. A., Bratt, D. L., Bricker, T. L., Brown, L., Buchser, W. J., … Suthar, M. S. (2022). Defining the risk of SARS-CoV-2 variants on immune protection. *Nature 2022 605:7911*, *605*(7911), 640–652. https://doi.org/10.1038/s41586-022-04690-5

Dhamija, S., & Menon, M. B. (2021). Measuring lncRNA Expression by Real-Time PCR. *Methods in Molecular Biology*, *2348*, 93–111. https://doi.org/10.1007/978-1-0716-1581-2_6/FIGURES/2

Dieterle, M. E., Haslwanter, D., Bortz, R. H., Wirchnianski, A. S., Lasso, G., Vergnolle, O., Abbasi, S. A., Fels, J. M., Laudermilch, E., Florez, C., Mengotto, A., Kimmel, D., Malonis, R. J., Georgiev, G., Quiroz, J., Barnhill, J., Pirofski, L. anne, Daily, J. P., Dye, J. M., … Jangra, R. K. (2020). A Replication-Competent Vesicular Stomatitis Virus for Studies of SARS-CoV-2 Spike-Mediated Cell Entry and Its Inhibition. *Cell Host and Microbe*, *28*(3), 486-496.e6. https://doi.org/10.1016/j.chom.2020.06.020

Evans, J. R., Feng, F. Y., & Chinnaiyan, A. M. (2016). The bright side of dark matter: lncRNAs in cancer. *The Journal of Clinical Investigation*, *126*(8), 2775–2782. https://doi.org/10.1172/JCI84421

Farzani, T. A., Chov, A., & Herschhorn, A. (2020). A protocol for displaying viral envelope glycoproteins on the surface of vesicular stomatitis viruses. *STAR Protocols*, *1*(3), 100209. https://doi.org/10.1016/j.xpro.2020.100209

Feng, Z., Wang, Y., & Qi, W. (2020). *The Small Intestine, an Underestimated Site of SARS-CoV-2 Infection: From Red Queen Effect to Probiotics*. https://doi.org/10.20944/PREPRINTS202003.0161.V1

Flynn, R. A., Belk, J. A., Qi, Y., Yasumoto, Y., Wei, J., Alfajaro, M. M., Shi, Q., Mumbach, M. R., Limaye, A., DeWeirdt, P. C., Schmitz, C. O., Parker, K. R., Woo, E., Chang, H. Y., Horvath, T. L., Carette, J. E., Bertozzi, C. R., Wilen, C. B., & Satpathy, A. T. (2021). Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions. *Cell*, *184*(9), 2394-2411.e16. https://doi.org/10.1016/J.CELL.2021.03.012

Fournier, C., Hoffmann, T. W., Morel, V., Descamps, V., Dubuisson, J., Brochot, E., Francois, C., Duverlie, G., Castelain, S., & Helle, F. (2018). Claudin-1, miR-122 and apolipoprotein E transductions improve the permissivity of SNU-182, SNU-398 and SNU-449 hepatoma cells to hepatitis C virus. *Journal of Viral Hepatitis*, *25*(1), 63–71. https://doi.org/10.1111/JVH.12767

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman,

O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, *154*(2), 442–451. https://doi.org/10.1016/J.CELL.2013.06.044

Hamming, I., Timens, W., Bulthuis, M. L. C., Lely, A. T., Navis, G. J., & van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *The Journal of Pathology*, *203*(2), 631–637. https://doi.org/10.1002/path.1570

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, *181*(2), 271. https://doi.org/10.1016/J.CELL.2020.02.052

Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M., & Weissman, J. S. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *ELife*, *5*(September2016). https://doi.org/10.7554/eLife.19760

Icard, P., Lincet, H., Wu, Z., Coquerel, A., Forgez, P., Alifano, M., & Fournel, L. (2021). The key role of Warburg effect in SARS-CoV-2 replication and associated inflammatory response. *Biochimie*, *180*, 169–177. https://doi.org/10.1016/J.BIOCHI.2020.11.010

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2021). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology 2021 23:1*, *23*(1), 3–20. https://doi.org/10.1038/s41580-021-00418-x

Jiang, F. (2017). CRISPR-Cas9 Structures and Mechanisms. *Article in Annual Review of Biophysics*. https://doi.org/10.1146/annurev-biophys-062215-010822

Jiang, M.-C., Ni, J.-J., Cui, W.-Y., Wang, B.-Y., & Zhuo, W. (2019). Emerging roles of lncRNA in cancer and therapeutic opportunities. *American Journal of Cancer Research*, *9*(7), 1354–1366. http://www.ncbi.nlm.nih.gov/pubmed/31392074%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6682721

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols 2017 12:4*, *12*(4), 828–863. https://doi.org/10.1038/nprot.2017.016

Koch, J., Uckeley, Z. M., Doldan, P., Stanifer, M., Boulant, S., & Lozach, P.-Y. (2021). TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells. *The EMBO Journal*, *40*(16), e107821. https://doi.org/10.15252/EMBJ.2021107821

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, *55*(3), 105924. https://doi.org/10.1016/J.IJANTIMICAG.2020.105924

Liao, K., Xu, J., Yang, W., You, X., Zhong, Q., & Wang, X. (2018). The research progress of LncRNA involved in the regulation of inflammatory diseases. *Molecular Immunology*, *101*, 182–188. https://doi.org/10.1016/J.MOLIMM.2018.05.030

Liu, S. J., Horlbeck, M. A., Cho, S. W., Birk, H. S., Malatesta, M., He, D., Attenello, F. J., Villalta, J. E., Cho, M. Y., Chen, Y., Mandegar, M. A., Olvera, M. P., Gilbert, L. A., Conklin, B. R., Chang, H. Y., Weissman, J. S., & Lim, D. A. (2017). CRISPRi-based genome-scale identification of functional

long noncoding RNA loci in human cells. *Science*, *355*(6320). https://doi.org/10.1126/science.aah7111

Liu, W., & Ding, C. (2017). Roles of LncRNAs in viral infections. *Frontiers in Cellular and Infection Microbiology*, *7*(MAY). https://doi.org/10.3389/fcimb.2017.00205

Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biology*, *10*(6), 924–933. https://doi.org/10.4161/rna.24604

Mac Kain, A., Maarifi, G., Aicher, S.-M., Arhel, N., Baidaliuk, A., Munier, S., Donati, F., Vallet, T., Tran, Q. D., Hardy, A., Chazal, M., Porrot, F., OhAinle, M., Carlson-Stevermer, J., Oki, J., Holden, K., Zimmer, G., Simon-Lorière, E., Bruel, T., … Roesch, F. (2022). Identification of DAXX as a restriction factor of SARS-CoV-2 through a CRISPR/Cas9 screen. *Nature Communications 2022 13:1*, *13*(1), 1–13. https://doi.org/10.1038/s41467-022-30134-9

Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., Nagata, N., Sekizuka, T., Katoh, H., Kato, F., Sakata, M., Tahara, M., Kutsuna, S., Ohmagari, N., Kuroda, M., Suzuki, T., Kageyama, T., & Takeda, M. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2- expressing cells. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(13), 7001–7003. https://doi.org/10.1073/PNAS.2002589117/ASSET/75C1AE47-6B19-43A5-98A3-12B416409EFE/ASSETS/GRAPHIC/PNAS.2002589117FIG01.JPEG

Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T. S., Imai, K., Miyoshi, K., Kishigami, S., Ueno, T., Iwatani, Y., Suzuki, T., & Tokunaga, K. (2021). SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-21118-2

Pal, M., Berhanu, G., Desalegn, C., & Kandi, V. (2020). Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. *Cureus*, *12*(3). https://doi.org/10.7759/CUREUS.7423

Palacios-Rápalo, S. N., De Jesús-González, L. A., Cordero-Rivera, C. D., Farfan-Morales, C. N., Osuna-Ramos, J. F., Martínez-Mier, G., Quistián-Galván, J., Muñoz-Pérez, A., Bernal-Dolores, V., del Ángel, R. M., & Reyes-Ruiz, J. M. (2021). Cholesterol-Rich Lipid Rafts as Platforms for SARS-CoV-2 Entry. *Frontiers in Immunology*, *12*, 5384. https://doi.org/10.3389/FIMMU.2021.796855/XML/NLM

Pandya, G., Kirtonia, A., Sethi, G., Pandey, A. K., & Garg, M. (2020). The implication of long non-coding RNAs in the diagnosis, pathogenesis and drug resistance of pancreatic ductal adenocarcinoma and their possible therapeutic potential. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1874*(2), 188423. https://doi.org/10.1016/J.BBCAN.2020.188423

Park, B. K., Kim, D., Park, S., Maharjan, S., Kim, J., Choi, J. K., Akauliya, M., Lee, Y., & Kwon, H. J. (2021). Differential Signaling and Virus Production in Calu-3 Cells and Vero Cells upon SARS-CoV-2 Infection. *Biomolecules & Therapeutics*, *29*(3), 273. https://doi.org/10.4062/BIOMOLTHER.2020.226

Payne, D. J., Dalal, S., Leach, R., Parker, R., Griffin, S., McKimmie, C. S., Cook, G. P., Richards, S. J., Hillmen, P., Munir, T., Arnold, L., Riley, K., McKinley, C., Place, S., Baretto, R. L., & Newton, D. J. (2021). The CXCR6/CXCL16 axis links inflamm-aging to disease severity in COVID-19 patients. *BioRxiv*, 2021.01.25.428125. https://doi.org/10.1101/2021.01.25.428125

Petiot, E., Cuperlovic-Culf, M., Shen, C. F., & Kamen, A. (2015). Influence of HEK293 metabolism on the production of viral vectors and vaccine. *Vaccine*, *33*(44), 5974–5981. https://doi.org/10.1016/J.VACCINE.2015.05.097

Prydz, K., & Saraste, J. (2022). The life cycle and enigmatic egress of coronaviruses. *Molecular*

Microbiology, *117*(6), 1308–1316. https://doi.org/10.1111/MMI.14907

Pulido-Quetglas, C., & Johnson, R. (2021). Designing libraries for pooled CRISPR functional screens of long noncoding RNAs. *Mammalian Genome 2021*, *1*, 1–16. https://doi.org/10.1007/S00335-021-09918-9

Rebendenne, A., Chaves Valadão, A. L., Tauziet, M., Maarifi, G., Bonaventure, B., McKellar, J., Planès, R., Nisole, S., Arnaud-Arnould, M., Moncorgé, O., & Goujon, C. (2021). SARS-CoV-2 Triggers an MDA-5-Dependent Interferon Response Which Is Unable To Control Replication in Lung Epithelial Cells. *Journal of Virology*, *95*(8). https://doi.org/10.1128/JVI.02415-20/ASSET/6D37141D-124F-4749-A48E-1968D7E81198/ASSETS/IMAGES/LARGE/JVI.02415-20-F0009.JPG

Rebendenne, A., Roy, P., Bonaventure, B., Chaves Valadão, A. L., Desmarets, L., Arnaud-Arnould, M., Rouillé, Y., Tauziet, M., Giovannini, D., Touhami, J., Lee, Y., DeWeirdt, P., Hegde, M., Urbach, S., Koulali, K. El, de Gracia, F. G., McKellar, J., Dubuisson, J., Wencker, M., … Goujon, C. (2022). Bidirectional genome-wide CRISPR screens reveal host factors regulating SARS-CoV-2, MERS-CoV and seasonal HCoVs. *Nature Genetics 2022*, 1–13. https://doi.org/10.1038/s41588-022-01110-2

Renganathan, A., & Felley-Bosco, E. (2017). Long noncoding RNAs in cancer and therapeutic potential. *Advances in Experimental Medicine and Biology*, *1008*, 199–222. https://doi.org/10.1007/978-981-10-5203-3_7/FIGURES/3

Samelson, A. J., Tran, Q. D., Robinot, R., Carrau, L., Rezelj, V. V., Kain, A. Mac, Chen, M., Ramadoss, G. N., Guo, X., Lim, S. A., Lui, I., Nuñez, J. K., Rockwood, S. J., Wang, J., Liu, N., Carlson-Stevermer, J., Oki, J., Maures, T., Holden, K., … Kampmann, M. (2022). BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2. *Nature Cell Biology 2022 24:1*, *24*(1), 24–34. https://doi.org/10.1038/s41556-021-00821-8

Schwegmann-Weßels, C., Glende, J., Ren, X., Qu, X., Deng, H., Enjuanes, L., & Herrler, G. (2009). Comparison of vesicular stomatitis virus pseudotyped with the S proteins from a porcine and a human coronavirus. *Journal of General Virology*, *90*(7). https://doi.org/10.1099/vir.0.009704-0

Shalem, O., Sanjana, N. E., & Zhang, F. (2015). High-throughput functional genomics using CRISPR–Cas9. *Nature Reviews Genetics 2015 16:5*, *16*(5), 299–311. https://doi.org/10.1038/nrg3899

Sharon, D. M., Nesdoly, S., Yang, H. J., Gélinas, J.-F., Xia, Y., Ansorge, S., & Kamen, A. A. (2020). A pooled genome-wide screening strategy to identify and rank influenza host restriction factors in cell-based vaccine production platforms. *Scientific Reports 2020 10:1*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-68934-y

Smieszek, S. P., Polymeropoulos, V. M., Polymeropoulos, C. M., Przychodzen, B. P., Birznieks, G., & Polymeropoulos, M. H. (2021). Elevated plasma levels of CXCL16 in severe COVID-19 patients. *MedRxiv*, 2021.09.07.21263222. https://doi.org/10.1101/2021.09.07.21263222

Smieszek, S. P., Polymeropoulos, V. M., Polymeropoulos, C. M., Przychodzen, B. P., Birznieks, G., & Polymeropoulos, M. H. (2022). Elevated plasma levels of CXCL16 in severe COVID-19 patients. *Cytokine*, *152*, 155810. https://doi.org/10.1016/J.CYTO.2022.155810

Spahn, P. N., Bath, T., Weiss, R. J., Kim, J., Esko, J. D., Lewis, N. E., & Harismendy, O. (2017). PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Scientific Reports 2017 7:1*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-16193-9

Testa, D., & Banerjee, A. K. (1979). Initiation of RNA synthesis in vitro by vesicular stomatitis virus. Role of ATP. *Journal of Biological Chemistry*, *254*(6), 2053–2058.

https://doi.org/10.1016/S0021-9258(17)37764-5

Ui-Tei, K. (2013). Optimal choice of functional and off-target effect-reduced siRNAs for RNAi therapeutics. *Frontiers in Genetics*, *4*(JUN). https://doi.org/10.3389/FGENE.2013.00107

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2020). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology 2020 19:3*, *19*(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

Volders, P. J., Helsens, K., Wang, X., Menten, B., Martens, L., Gevaert, K., Vandesompele, J., & Mestdagh, P. (2013). LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Research*, *41*(D1), D246–D251. https://doi.org/10.1093/NAR/GKS915

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, *181*(2), 281. https://doi.org/10.1016/J.CELL.2020.02.058

Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C. H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., Mei, S., Cui, Y., Shi, S., Lipp, J. J., Hinterndorfer, M., Zuber, J., Brown, M., Li, W., & Liu, X. S. (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nature Protocols 2019 14:3*, *14*(3), 756–780. https://doi.org/10.1038/s41596-018-0113-7

Wang, R., Simoneau, C. R., Kulsuptrakul, J., Bouhaddou, M., Travisano, K. A., Hayashi, J. M., Carlson-Stevermer, J., Zengel, J. R., Richards, C. M., Fozouni, P., Oki, J., Rodriguez, L., Joehnk, B., Walcott, K., Holden, K., Sil, A., Carette, J. E., Krogan, N. J., Ott, M., & Puschnik, A. S. (2020). Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses. *Cell*, *184*(1), 106-119.e14. https://doi.org/10.1016/j.cell.2020.12.004

Wei, J., Alfajaro, M. M., DeWeirdt, P. C., Hanna, R. E., Lu-Culligan, W. J., Cai, W. L., Strine, M. S., Zhang, S. M., Graziano, V. R., Schmitz, C. O., Chen, J. S., Mankowski, M. C., Filler, R. B., Ravindra, N. G., Gasque, V., de Miguel, F. J., Patil, A., Chen, H., Oguntuyo, K. Y., … Wilen, C. B. (2020). Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection. *Cell*, 1–16. https://doi.org/10.1016/j.cell.2020.10.028

Whitt, M. A. (2010). Generation of VSV Pseudotypes Using Recombinant ΔG-VSV for Studies on Virus Entry, Identification of Entry Inhibitors, and Immune Responses to Vaccines. *Journal of Virological Methods*, *169*(2), 365. https://doi.org/10.1016/J.JVIROMET.2010.08.006

Yahalom-Ronen, Y., Tamir, H., Melamed, S., Politi, B., Shifman, O., Achdout, H., Vitner, E. B., Israeli, O., Milrot, E., Stein, D., Cohen-Gihon, I., Lazar, S., Gutman, H., Glinert, I., Cherry, L., Vagima, Y., Lazar, S., Weiss, S., Ben-Shmuel, A., … Israely, T. (2020). A single dose of recombinant VSV-ΔG-spike vaccine provides protection against SARS-CoV-2 challenge. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-20228-7

Yan, P., Luo, S., Lu, J. Y., & Shen, X. (2017). Cis- and trans-acting lncRNAs in pluripotency and reprogramming. *Current Opinion in Genetics & Development*, *46*, 170–178. https://doi.org/10.1016/J.GDE.2017.07.009

Yaneske, E., Zampieri, G., Bertoldi, L., Benvenuto, G., & Angione, C. (2021). Genome-scale metabolic modelling of SARS-CoV-2 in cancer cells reveals an increased shift to glycolytic energy production. *FEBS Letters*, *595*(18), 2350–2365. https://doi.org/10.1002/1873-3468.14180

Yao, R. W., Wang, Y., & Chen, L. L. (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biology*, *21*(5), 542–551. https://doi.org/10.1038/s41556-019-0311-8

Yi, K., Zhang, Y., Wang, Y., Wang, Z., Xie, M., Jin, Z., & Zhao, T. (2019). Long noncoding RNA and its

role in virus infection and pathogenesis. *Frontiers in Bioscience (Landmark Edition)*, *24*, 777–789.

Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., & Shen, F. (2019). Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *International Journal of Molecular Sciences*, *20*(22). https://doi.org/10.3390/ijms20225573

Zhang, Y., Yin, C., Zhang, T., Li, F., Yang, W., Kaminski, R., Fagan, P. R., Putatunda, R., Young, W. Bin, Khalili, K., & Hu, W. (2015). CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. *Scientific Reports 2015 5:1*, *5*(1), 1–14. https://doi.org/10.1038/srep16277

Zhu, Y., Feng, F., Hu, G., Wang, Y., Yu, Y., Zhu, Y., Xu, W., Cai, X., Sun, Z., Han, W., Ye, R., Qu, D., Ding, Q., Huang, X., Chen, H., Xu, W., Xie, Y., Cai, Q., Yuan, Z., & Zhang, R. (2021). A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry. *Nature Communications*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-21213-4

Zupin, L., Fontana, F., Clemente, L., Ruscio, M., Ricci, G., & Crovella, S. (2022). Effect of Short Time of SARS-CoV-2 Infection in Caco-2 Cells. *Viruses 2022, Vol. 14, Page 704*, *14*(4), 704. https://doi.org/10.3390/V14040704

תקציר

מגפת נגיף קורונה 2019 (COVID-19) נגרמת כתוצאה מהדבקה של הנגיף Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). נגיף הקורונה מדביק את תא המאכסן שלו על ידי קשירה של חלבון הספייק (Spike) הווירלי את רצפטור תא המטרה angiotensin-converting enzyme 2 (ACE2). וירוסים פתוגניים משתמשים בפקטורים של התא המאחסן לצורך שלבים שונים של תהליך ההדבקה. גילוי פקטורים אנושיים של התא המאכסן הנחוצים לתהליך ההדבקה של וירוס הקורונה יסייע בפיתוח טיפולים תרפויטיים חדשים ל-COVID-19, וגם ישפוך אור על מנגנון התהוות המחלה של נגיף הקורונה. בעוד שמרבית המחקרים כיום מתמקדים בזיהוי גנים מקודדי-חלבון המשפיעים על תהליך ההדבקה, תפקיד הגנום הלא-מקודד כמעט ואינו נחקר. עיקר הגנום שלנו מורכב מגנים שמשועתקים לרנ"א (חומצה ריבונוקלאית) לא-מקודד. רנ"א ארוך לא-מקודד (lncRNA) מרכיבים את הקבוצה גדולה והמגוונת ביותר של רנ"א לא-מקודד. רנ"א ארוך לא-מקודד מוגדר כתעתיקים באורך של יותר מ-200 נוקלאוטידים, שאינם כוללים מסגרת קריאה פתוחה, או שיש להם מסגרת קריאה קצרה מאד. רנ"א ארוך לא-מקודד יכול לבקר ביטוי גנים ברמת האפיגנטיקה, השעתוק, התרגום והפוסט-תרגום על ידי אינטראקציות עם דנ"א (חומצה דאוקסיריבונוקלאית), רנ"א וחלבונים. לרנ"א ארוך לא-מקודד יש תפקיד בתהליכים תאיים רבים, לכן אנו מניחים שהוא משחק גם תפקיד בתהליך ההדבקה של וירוס הקורונה. אפיון רנ"א ארוך לא-מקודד הינו מאתגר ממספר סיבות. הוא עובר העשרה בגרעין מה שמגביל את סוגי ההשתקה הגנטית שמסוגלים לפעול עליו. בנוסף, מכיוון שאין לו מסגרת קריאה פתוחה, לא ניתן להשתיק אותו על ידי חיתוך הדנ"א. לבסוף, מכיוון שיש מספר רב של תעתיקי הרנ"א ארוך לא-מקודד, יש צורך בשימוש בסקירות גנטיות נרחבות. סקירות גנטיות המבוססות על CRISPR (Clustered regularly interspaced short palindromic repeats) הינן גישה יעילה לחקר רנ"א ארוך לא-מקודד. ניתן להגביר אותן בקלות לסדר גודל כלל-גנומי, יש להן ספציפיות גבוהה והן מסוגלות להשתיק רנ"א שעובר העשרה בגרעין. באופן יותר פרטני, שיטת ה-CRISPR interference (CRISPRi) מאפשרת לבקר את ביטוי הגנים ברמת השעתוק, ללא צורך בעריכה לא הפיכה של הגנום. במחקר הזה, ביצענו סקירת CRISPR לזיהוי רנ"א ארוך לא-מקודד המשפיע על הדבקת וירוס הקורונה בתאי SNU449, שהם תאים אנושיים שמקורם מסרטן כבד. התאים עברו גם מודיפיקציה גנטית כדי לבטא ביתר את רצפטור ה-ACE2. כדי לדמות את כניסת וירוס הקורונה, השתמשנו במערכת של פסאודו-וירוס מסוג vesicular stomatitis virus (VSV) שמציג את חלבון הספייק של וירוס הקורונה על גבי הממברנה שלו. הווירוס גם הכיל רצף שמקודד לחלבון מדווח בעל פלואורסצנציה ירוקה (GFP) על מנת שנוכל לזהות ולכמת את ההדבקה. הסקירה הגנטית שלנו זיהתה רנ"א ארוכים לא-מקודדים רבים בעלי השפעה חיובית ושלילית על תהליך ההדבקה. מתוך הגנים שנמצאו, RP11-977G19.11 ו-RP11-314A20.5 עברו ולידציה כגנים החשובים להדבקת וירוס הקורונה. מצאנו ששני הגנים שעברו ולידציה מבקרים את הביטוי של גנים מקודדי-חלבון שכנים, שזהו מנגנון הפעולה הנפוץ של רנ"א ארוך לא מקודד. RP11-977G19.11 מבקר את הביטוי של Citrate synthase, שזהו גן בעל תפקיד מפתח במטבוליזם של אנרגיה. RP11-314A20.5 מבקר את הביטוי של מספר גנים שכנים, ביניהם Chemokine (C-X-C motif) ligand 16 (CXCL16), שבעל תפקיד בתגובה החיסונית. הגנים שמצאנו עשויים להוות מטרות תרפויטיות חדשות לטיפול ב-COVID-19, וגם ירחיבו את ההבנה שלנו של מנגנון ההדבקה של וירוס הקורונה. הממצאים שלנו במחקר זה רלוונטיים למגפה הנוכחית, וגם עשויים להיות חשובים לפתוגניים וירליים חדשים שמשתמשים באותם פקטורים של התא המאכסן.

המחקר נעשה בהנחיית ד"ר אסף בסטר בפקולטה לביולוגיה.

אנו מודים לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

סקירת קריספר כלל-גנומית לזיהוי רנ"א ארוך לא מקודד המשפיע על בניסת וירוס הקורונה לתא

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר מגיסטר למדעים בביולוגיה

יובל בן אליהו

הוגש לסנט הטכניון – מכון טכנולוגי לישראל

אלול תשפ"ב חיפה ספטמבר 2022