CRISPR screens for host factors critical for infection by SARS-CoV-2 variants of concern identify GATA6 as a central modulator of ACE2.

Ma'ayan Israeli[#1], Yaara Finkel[#2], Ofir Israeli[1], Adi Beth-Din[1], Inbar Cohen-Gihon[1], Yfat Yahalom-Rone[3], Nir Paran[3], Tomer Israely[3], Theodor Chitlaru[1], Uri Elia[1], Ofer Cohen[1], Ital Nemet[4], Limor Kliker[4], Michal Mandelboim[4], Noam Stern-Ginossar[2], and Adi Bercovich-Kinori[1]

[1] Department of Biochemistry and Molecular Genetics, Israel Institute for Biological Research, Ness Ziona, Israel

[2] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

[3] Department of Infectious Diseases, Israel Institute for Biological Research, Ness Ziona, Israel

[4] Central Virology Laboratory, Public Health Services, Ministry of Health and Sheba Medical Center, Tel Hashomer, Israel

[#] Contributed equally

Abstract

The global spread of SARS-CoV-2 lead to the most challenging pandemic in this century, posing major economic and health challenges worldwide. Revealing host genes essential for infection by multiple variants of SASR-CoV-2 can provide insights into the virus pathogenesis, and facilitates the development of novel broad-spectrum host-directed therapeutics. Here, employing genome-scale CRISPR screens, we provide a comprehensive data-set of cellular factors that are exploited by WT-SARS-CoV-2 as well as two additional recently emerged variants of concerns (VOCs), Alpha and Beta. These screens identified known and novel host factors critical for SARS-CoV-2 infection, including various components belonging to the Clathrin-dependent transport pathway, ubiquitination and Heparan sulfate biogenesis. In addition, the host metabolic pathways pertaining to mitochondrial activity as well as phosphatidylglycerol biosynthesis processes appeared to have major anti-viral functions. Comparative analysis between the different VOCs revealed the receptor KREMEN2 as unique gene required only to the Alpha variant, providing a possible explanation for the increased infectivity of this variant. Furthermore, the analysis identified GATA6, a zinc finger transcription factor, as an essential pro-viral gene for all variants inspected. We reveal that GATA6 directly regulates ACE2 transcription and is accordingly, critical for SARS-CoV-2 cell entry. Analysis of clinical samples collected from SARS-CoV-2 infected individuals showed an elevated level of GATA6, indicating the important role GATA6 may be playing in COVID-19 pathogenesis. Finally, pharmacological

inhibition of GATA6 results in down-modulation of ACE2 and consequently to inhibition of the viral infectivity. Overall, we show GATA6 represents a target for the development of anti-SARS-CoV-2 therapeutic strategies and reaffirm the value of the CRISPR loss-of-function screens in providing a list of potential new targets for therapeutic interventions.

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a recently emerged type of respiratory syndrome-related coronaviruses, responsible for the Coronavirus Disease 2019 (COVID-19), the cause of the most challenging pandemic in this century. Despite an unprecedented worldwide research effort that resulted in the rapid development of a variety of vaccines against SARS-CoV-2, the pandemic remains uncontrolled in many countries and continues to take a devastating toll on both human health and global economic activity. The emergence of SARS-CoV-2 variants with enhanced transmissibility and pathogenesis, or which may escape the efficacy of pre and post exposure countermeasures, presents further challenges in the continuous struggle against the virus and the tragical consequences of the pandemic.

Coronaviruses are enveloped, positive-sense single-stranded RNA viruses with a genome of approximately 30 kb in length [1]. The initial steps of SARS-CoV-2 infection involve the specific binding of the coronavirus spike protein (S) to the cellular entry receptor angiotensin-converting enzyme 2 (ACE2) [2-5]. The binding to the receptor and the subsequent membrane fusion commences the process of viral entry into the host cell by endocytosis allowing the virion to release the genetic material into the cytoplasm where it is translated into continuous polypeptides that are cleaved into 16 nonstructural proteins, which then facilitate the transcription of sub-genomic RNA that is translated into structural and accessory proteins. These viral proteins rearrange host membranes to form the endoplasmic reticulum-localized viral replication complex (RC) in which the viral genomic RNA is replicated by the viral RNA-dependent RNA polymerase (RdRP) complex. The

assembled components further undergo maturation in the Golgi apparatus to form the mature virion that is released outside the host cell by exocytosis [1, 6-8]. Every step of the viral life cycle, from entry to budding, is orchestrated through interactions with host cellular proteins. For example, cellular proteases, such as TMPRSS2, Cathepsin L, and Furin, are exploited for the cleavage of the viral spike protein of SARS-CoV-2, mediating efficient membrane fusion with the infected cell [9-13]. Identification of additional key host proteins involved in the various steps of the infection is essential for the development of countermeasures and represents the objective of intense efforts. Several proteomics studies addressed the SARS-CoV-2-host interactions by utilizing immunoprecipitation mass spectrometry (IP-MS) of affinity-tagged viral proteins and proximity labeling of the virus replicase complex [14-17]. The proteomics approach revealed comprehensive interactomes and physical contacts between many viral and cellular proteins, and highlighted potential targets for drug repurposing. Yet, this approach cannot provide direct information about the essentiality of these host components for enabling the SARS-CoV-2 life cycle. Another strategy to explore SARS-CoV-2- host interactions is to disrupt individual genes of the entire host genome and screen for genes whose disruption results in resistance to viral infection and consequently to host-cell survival. Several reports documented genome-wide CRISPR-Cas9-mediated loss-of-function screens and identify host factors that are functionally required for SARS-CoV-2 infection in various cell types [10, 18-25]. These studies identified genes, metabolic and signaling networks that previously have not been considered as potential therapeutic targets for SARS-CoV-2, such as the cholesterol synthesis pathways [19].

As of today, following a year-long period of combating Covid-19 essentially by medical care and social distancing, the contribution of many effective vaccines started to materialize [26, 27]. Covid-19 vaccines showed as high as 95% efficacy in preventing clinical cases and 100% efficacy in preventing severe disease with original viral variants [28, 29]. The average mutation rate of SARS-CoV-2 remains low and steady. However, the continuing global spread of the SARS-CoV-2, together with selective pressure for immune escape, led to the emergence of new

SARS-CoV-2 variants raising concerns regarding their relative ability to escape from natural and vaccine-induced immunity. Four novel variants that generated different patterns in the pandemic expansion, are of considerable public-health concern, as concluded by the WHO organization following judicious monitoring and classification of the new lineages (hence eloquently coined VOC, acronym for variants of concern): (i) Alpha (also known as B.1.1.7, VOC-202012/01- Variant of Concern, year 2020, month 12, variant 01) emerged in the UK. (ii) Beta variant (also known as B.1.351 variant, 501Y.V2), emerged in South Africa. (iii) Gamma, (also known as the P.1 lineage, 501Y.V3 variant), first identified in Brazil (iv) Delta (also known as B.1.617.2. variant) originally surfaced in India [30-33]. The Alpha lineage has a total of 17 nonsynonymous mutations relative to the original Wuhan strain, of which 7 replacements and 2 deletions reside in the spike protein. The Alpha variant was shown to be significantly more infectious than other lineages (increasing the effective reproduction number, Rt, by a factor of 1.35) which resulted in its rapid global expansion. As of May 2021, the variant has been detected in over 100 countries, with a worldwide daily prevalence of over 75% [34]. The Beta variant emerged independently of the Alpha variant but shares with it mutations at several loci. The significant anxiety caused by this variant, owe to the co-occurrence of the N501Y mutation in the spike protein receptor binding domain (RBD, also found in the Alpha variant) with additional mutations K417N/T and E484K. Viral variants with the triple combination of mutations exhibit reduced susceptibility to vaccine-induced and convalescent sera [35-41]. Fortunately, the Beta variant is less abundant with a reported worldwide daily prevalence of 1% [42, 43]. In addition to the assessment of full immunological and clinical implications of the SARS-CoV-2 new variants, there is an urgent need for a more profound understanding of their biology, and for integrating this information towards the development of therapeutic approaches targeting host elements essential to all variants.

In this study we have performed a genome-wide CRISPR loss-of-function screen in a lung derived human cell line infected with SARS-CoV-2. In an attempt to probe shared and differential host factors that may be essential for the virus in the course

of infection we expanded our genomic approach to the WT SARS-CoV-2 and two additional VOCs. This strategy led to the discovery of known and novel SARS-CoV-2-host interactions, enabled the identification of both common factors as well as those specific for particular variant, and revealed a pivotal role of the host pleiotropic regulator GATA6 which may represent a promising target for therapy.

Results

Genome wide CRISPR screens for the identification of host factors essential for SARS-CoV-2 infection

To identify host factors essential for cell survival in response to human SARS-CoV-2, a CRISPR-based genome-wide gene-knockout screen was performed using the human lung epithelial cell line Calu-3, which is highly permissive to SARS-CoV-2 infection due to elevated levels of endogenous ACE2 expression. The library employed for the screen was the Brunello CRISPR library composed of 76,441 targeting single guide RNAs (sgRNAs) with an average of four gRNAs per gene and 1,000 non-targeting control sgRNAs [44]. For the identification of host factors exhibiting either broad or restricted essentiality for the different variants of SARS-Cov-2, Calu-3 cells transduced with the CRISPR library, were infected separately with each of the SARS-CoV-2 variants of concern (VOCs): SARS-CoV-2 B.1.1.7(Alpha), SARS-CoV-2 B.1.351 (Beta), and the original wild-type (WT) SARS-CoV-2 (GISAID accession EPI_ISL_406862). Genomic DNA was harvested from surviving cells 7-9 days post infection, and sgRNAs abundance was determined by sequencing (Fig.1A). Integration of the datasets generated by screening cells infected with these various SARS-CoV-2 variants identified numerous genes which upon disruption, conferred resistance to death by viral infection (referred to as pro-viral) or sensitization to viral infection (referred to as anti-viral). The sgRNA score- distribution of the full library and that of sgRNA targeting the top hits for pro-and anti-viral genes demonstrated high technical quality of the screens (Fig 1B). As expected, the pro-viral gene, exhibiting the highest enrichment score was ACE2, the receptor serving as the portal for SARS-CoV-2 cell-entry (log fold change (LFC)=1.534, false discovery rate (FDR)= 0.007).

The transcription factor GATA6 (GATA Binding Protein 6) scored as the second-strongest pro-viral hit (LFC=1.24, FDR= 0.028) (Figure 1C, 1D). GATA6 is a DNA-binding protein belonging to a family of zinc finger transcription factors that participate in the regulation of cellular differentiation during vertebrate development and contributes to immune regulation through various pathways such as NF-kB. Interestingly, the transcription level of GATA6 was shown to be elevated in the lung transcriptome of Covid-19 patients [45]. Other top pro-viral hits included CUL5, IRF6, AP1G1, EP300 and EXTL2 (Figure 1C, 1D). The CUL5 gene encodes Cullin5, a core component of multiple SCF-like ECS (Elongin-Cullin 2/5-SOCS-box protein) E3 ubiquitin-protein ligase complexes, which mediate the ubiquitination and subsequent proteasomal degradation of target proteins. While ubiquitination of viral proteins can be used by the host as a defense mechanism by destroying the incoming pathogen, viruses have adapted to exploit this cellular process to enhance various steps of the replication cycle and increase pathogenesis [46]. Indeed, in addition to highly modified ubiquitin-proteome of SARS-CoV-2-Infected host cells, ubiquitination modifications were observed also on SARS-CoV-2 proteins, and these modifications were reported to inhibit the host innate immune response [46-48]. IRF6 gene is a transcriptional activator with well-established role in the development of the epidermis. A recent study suggests a role for IRF6 in the modulation of the NFk-B-pathway that influences cell-survival by regulating host responses during viral infection, such as cytokine production [49]. AP1G1 (Adaptor Related Protein Complex 1 Subunit Gamma 1) encodes a component of clathrin-coated vesicles that were shown to mediate endocytosis and intracellular trafficking of SARS-CoV-2 [50, 51]. Of note, the clathrin-coating complex, to which AP1G1 belongs, was recently identified as important for SARS-CoV-2 infectivity in a screen conducted in Calu-3 cells but not in screens conducted in other cell lines [21]. Taken together, the data therefore suggest that some of the host entry factors necessary for SARS infection exhibit cell-line specificity illustrating the importance of using relevant cell models to maximize our understanding of SARS-CoV-2 entry. EP300 encodes a histone acetyltransferase that is recruited by SMAD complexes to activate target genes in response to TGF-β receptor signaling [52]. Interestingly,

in a recent search for functional modifiers of ACE2 surface abundance, EP300 was found to regulate ACE2 transcription and hence to influence cellular susceptibility to SARS-CoV-2 infection [53]. EXTL2 encodes a glycosyltransferase involved in the chain elongation step of heparan sulfate (HS) biosynthesis [54, 55]. Several heparan sulfate biosynthetic genes were also markedly enriched in other CRISPR knockout SARS-CoV-2 screens conducted in Huh7.5.1 cells [22, 23] consistent with a recent report that SARS-CoV-2 infection is co-dependent on heparan sulfate and ACE2 [56]. Since HS is ubiquitously expressed on the surfaces and in the extracellular matrix of all cell types, EXTL2 may play a central pro-viral role through heparan sulfate production that may determine SARS-CoV-2 tropism.

Among genes encoding for factors influencing negatively viral infection (anti-viral factors), the screens identified RPP21 (LFC=-4.08, FDR=0.022) and UQCRC1 (LFC=-2.75, FDR=0.022) as the top 2 significantly depleted genes (Figure 1C ,1D). RPP21 encodes a subunit of nuclear ribonuclease P, which processes the 5-prime leader sequence of precursor tRNAs [57]. UQCRC1 encodes a component of the ubiquinol-cytochrome c oxidoreductasel that is part of the mitochondrial electron transport chain which drives oxidative phosphorylation. Similarly to a previous survey [10], the present screen also distinguished a significant enrichment of mitochondrial ribosomal proteins (MRPS) genes (e.g., MRPL17, MRPS33, MRPL41, MRPL34, MRPL53 and MRPS12) (highlighted in red in Figure 1D). The emerging notion of the role of mitochondrial dynamics in viral infections, unequivocally attribute mitochondria with a key role for cellular metabolism and innate immunity. Overall, the corroboration of previously reported entry factors reported here, together with novel pro-viral host genes and the implied functional pathways, demonstrates the high potential of CRISPR-knockout screens to distinguish targets for host dependent anti-viral therapy.

<u>Functional pathways and interaction networks of enriched SARS-CoV-2 host factors</u>

Gene set enrichment analysis (GSEA) confirmed that anti-viral genes (depleted from the library) are strongly enriched for mitochondrial function (Fig. 2A),

specifically mitochondrial translation (normalized enrichment score (NES) =-2, FDR =0.004) and respiratory electron transport chain (NES =-1.9, FDR =0.009). Additional functional categories to which anti-viral genes belong, were related to translation, including translation elongation (NES =-1.9, FDR =0.009) and amino acid activation (NES =-1.9, FDR =0.01). This exemplifies the expected absolute dependency of the virus on host factors for translation of its genes. Another large group of anti-viral genes were found to participate in the phosphatidylglycerol biosynthetic process (NES =-1.94, FDR =0.007). Interestingly, phosphatidylglycerol was reported to regulate the innate immune response in lungs [58] and consequently was suggested as a treatment against SARS-CoV-2 infection by restoring the lung tissue [59]. The current results suggest also that phosphatidylglycerols may play a supporting role in SARS-CoV-2 infection. No functional pathways were found to be significantly represented by analysis of genes found as essential (pro-viral) for SARS-CoV-2 infection.

For further inspection of the pro-viral genes, STRING-db interaction network analysis was employed for the top 200 enriched genes (Fig. 2B). The connected network components annotated using REACTOME pathways functionally implicated several cellular pathways. A large group of these genes, including the largest cluster, is related to immune system functions, and includes the genes NFKB1, IRF6, IL2Ra and EP300. In addition, the analysis identified gene groups belonging to lipid metabolism, and genes related to vesicle mediated transport, both functions known to be important determinants of viral infection.

Differential analysis of host-factors dependencies exhibited by variants of SARS-CoV-2

Reported differences in the infectivity as well as in the evolution of the disease caused by different variants of SARS-CoV-2 indicate that these emerging variants may require both common and distinct host factors. These shared and diverged host requirements may enable a better understanding of the host-pathogen interactions characterizing SARS-CoV-2 infection, and most importantly may facilitate future therapies. Accordingly, a comparative analysis of the datasets of

the screens involving either one of the SARS-CoV-2 variants of concern: SARS-CoV-2 B.1.1.7(Alpha), SARS-CoV-2 B.1.351 (Beta), and the original WT SARS-CoV-2, was conducted. The individual variant analysis enabled distinction of the top-scoring host-genes involved in the infection of each SARS-CoV-2 lineage (Fig. 3A). The cellular entry factor, ACE2, was found to be highly enriched for all lineages, illustrating the robustness of our screens and illustrating that the main mechanism for entry into the target cells is similar for the three SARS-CoV-2 variants. Another host gene that was consistently highly enriched in the screens of cells infected by the three variants was the transcription factor GATA6. GATA6 role in SARS-CoV-2 infection was so far not reported and will be extensively addressed below. Conversely, the analysis identified different genes specifically involved in the infection by either one of the variants. Such are ARIH2 and CHMP2B that exhibited a higher enrichment score in the screen of the cells infected by the Alpha and Beta variants, respectively. Additional significantly enriched genes are highlighted in Figure 3A.

Next, we assessed the extent of host-factor sharing between cells infected by the different variants, by Venn analysis of the top 100 genes which emerged as viral infection facilitators in the genome-wide screens (Figure 3B). The analysis evidenced that only the 2 genes, ACE2 and GATA6, were scored high in each of the three datasets, indicating that their key-role in the course of SARS-CoV-2 life cycle is lineage-independent. With the exception of ACE2 and GATA6, a low overlap in the top-ranked hits was observed: only 1 gene was found to be common to cells infected by Alpha and the WT variant (HEATR5B), 3 common genes to Beta and the WT variant (TMEM127, CALB2, ADIPOR1) and 5 to the Alpha and Beta variants (OSBPL10, AMBP, KIF5B, KMT2C, SMYD1). These partially mutual pro-viral genes are involved in biological processes such as endocytosis regulation (HEATR5B), mTOR signaling (TMEM127), calcium signaling (CALB2), fatty acid oxidation and glucose uptake (ADIPOR1), lipid biogenesis (OSBPL10) and histone modification (KMT2C, SMYD1).

To further distinguish the role of these host-genes in the infection of each of the variants, a pairwise correlation of the genome-wide datasets was conducted, (Fig. 3C). The pro-viral genes, ACE2, GATA6 and VEPH1 were consistently highly enriched with all variants, while the anti-viral genes NHLRC2, GART, LARS2, RPP21, PMPCB, CIAO1 and ELP5 were depleted for all three variants. Most interestingly, some genes were enriched in the screen conducted with one mutant while depleted in the screens carried-out with other variants. One of these genes, KREMEN2 (highlighted in figure 3C) was highly enriched following infection with the Alpha variant and exhibited an opposite pattern in the screens involving the Beta variant and the WT SARS-CoV-2 suggesting that this gene may therefore fulfill a distinct role during the life cycle of the Alpha variant. Notably, the proteins KREMEN1 and KREMEN2 have been identified as co-receptors for Dickkopf (Dkk) proteins, hallmark secreted antagonists of canonical Wnt signaling [60, 61]. In addition, KREMEN1 serves as host-cell entry receptors for a major group of Enteroviruses [62]. Remarkedly, in a recent screen in search of additional SRAS-CoV-2 receptors, KREMEN1 was identified as an alternative viral-entry receptor [63]. Thus, our observation suggests that KREMEN2 may play a similar role during the life cycle of the Alpha variant specifically, consequently contributing to its enhanced infectivity compared with other SARS-CoV-2 variants. Overall, the comparative analysis shows that the survival-based CRISPR screen provides datasets of host factors common and distinct for a variety of SARS-CoV-2 lineages that should be further validated and taken under consideration for the design of broad-spectrum therapies for controlling SARS-CoV-2 infection.

GATA6 is essential for viral entry of SARS-CoV-2

The transcription factor GATA6, emerged as the second-strongest pro-viral host factor (inferior only to the viral receptor ACE2) for all three variants tested in the screens. GATA6 is a member of a small family of zinc finger DNA-binding transcription factors that play an important role in the regulation of cellular differentiation. Of note, GATA6 was shown to promote the transcription of SFTPA gene, which is involved in immune and inflammatory responses, and lowers the

surface tension in the alveoli [64, 65]. Both GATA6 and SFTPA genes were upregulated in SARS-CoV-2-infected lung, while the GATA6 antagonist LMCD1 was downregulated [45]. To confirm the essentiality of GATA6 for SARS-CoV-2 infection the expression of GATA6 in Calu-3 cells was abrogated by CRISPR-mediated targeting of GATA6, followed by inspecting the sensitivity of cells to SARS-CoV-2 Infection. CRISPR-mediated ACE2 disruption in Calu-3 cells served as a positive control. The gene targeting manipulation indeed resulted in depletion of GATA6 and ACE2 expression as measured by Real-Time PCR analyses (Fig. S1). As expected, while WT Calu-3 cells display cytopathic effect following SARS-CoV-2 infection, ACE2 gene-disruption protected cells from the infection (Fig. 4A). In agreement with the results of our CRISPR screens, GATA6-disrupted cells were also almost fully protected from cell death following SARS-CoV-2 infection. Furthermore, the extent of resilience to viral infection was commensurate with that exhibited by cells in which ACE2 expression was abrogated (Fig. 4A). Further investigation of SARS-CoV-2 viral load in infected GATA6-disrupted cells demonstrated significant reduction of viral replication 48 hours post infection subsequent to reduction in GATA6 content (Fig. 4B).

Since GATA6 did not emerged as essential for SARS-CoV-2 infection in previously published SARS-CoV-2 CRISPR screens carried in other cell types, we addressed the possibility that the observed phenomenon is related to GATA6 expression in different cell types. Inspection of the relative expression levels of GATA6 in published datasets of Calu-3, A549 and Vero-E6 cell lines [66-68], revealed that GATA6 is indeed expressed at much higher levels in Calu-3 cells (Fig 4C). Furthermore, the relative levels of GATA6 increased along infection of Calu-3 cells (Fig 4D). It is therefore conceivable that the function fulfilled by GATA6 in Calu-3 cells is not identical to that in A549 and Vero-E6, in line with the fact that GATA6 is a regulatory transcriptional factor which exerts its function by modulating the transcriptional level of a various target genes which may not be necessarily identical in different cell lines.

To further explore the physiologically and/or clinical relevance of GATA6 to SARS-CoV-2 infection, the level of GATA6 RNA in clinical nasopharyngeal swab samples

from 20 COVID-19 patients was analyzed and compared to that in samples collected from 20 healthy people. The analysis demonstrated that COVID-19 patients exhibited significantly higher levels of GATA6 than controls ($p<0.05$) (Fig 4E). Of note, the analysis did not reveal statistically relevant differences in ACE2 expression between COVID-19 and control healthy individuals. This observation further demonstrates that GATA6 may be involved in the susceptibility to viral infection, and suggests that GATA6 by itself is upregulated in response to infection by a regulatory inductive circuit differs from that of ACE2 receptor.

To further elucidate which viral life cycle step is facilitated by GATA6 expression, SARS-CoV-2 infection was visualized by applying immunofluorescent labeling of the virus in infected WT in comparison to GATA6-disrupted cells. The analysis enabled detection of SARS-CoV-2 in the cytoplasm of WT cells. In contrast, there was significant reduction of the infection of SARS-CoV-2 in the GATA6-disrupted cells (Fig. 5A) suggesting that GATA6 facilitates an early step in the viral life cycle possibly related to the entry of the virus into the cells.

<u>GATA6 is a novel regulator of ACE2 that may serve as a potential host directed therapeutic target</u>

Since SARS-CoV-2 entry into cells is primarily mediated by ACE2, the expression of ACE2 in these cells was determined by Western-blot analysis. The analysis included WT, ACE2 disrupted, and GATA6 disrupted cells. Notably, the decrease in GATA6 was followed by a significant reduction in ACE2 expression compared to control cells (Fig. 5B). Real-Time PCR analysis validated the Western-blot results indicating that this effect involves a regulatory mechanism exerted at the transcriptome-level, suggesting that GATA6 acts as a transcriptional factor affecting ACE2 expression (Fig. 5C). In an attempt to further confirm these results and determine the possible temporal connection between GATA6 expression and that of ACE2, the expression of GATA6 was abrogated by siRNA targeting and the expression of ACE2 was interrogated at early time-points post GATA6 silencing. The analysis shows that ACE2 expression decreases as soon as 12 hours post GATA6-siRNA transfection (Fig. 5D).

GATA6 is a known transcription factor recognizing well-defined DNA-binding specific sites (T/AGATAA/G or reverse complement sequences C/TTATCT/A) [69]. Accordingly, the promoter of the ACE2 gene was screened for the GATA6 consensus DNA binding site. The conserved GATA-binding motif was found at nucleotide positions −341 and −403 (where the A of the ATG codon is +1) (Figure 5E) [70]. To further investigate the possible sequence-specific binding of GATA6 to ACE2 promoter an Electrophoretic Mobility Shift Assays (EMSA) was applied. Incubation of lysates from HEK293T cells over expressing GATA6 with labeled oligonucleotides derived from the promoter region of ACE2 results in the formation of protein-DNA complexes, which was prevented by competitive unlabeled oligonucleotides (Fig. 5F). This observation supports the notion that GATA6 may induce the expression of ACE2 directly via a sequence-specific GATA6 binding sites.

CRISPR loss-of-function surveys for host factors based on virus-induced cell death, may identify genes encoding for products essential to the virus, and yet dispensable to cell viability. Accordingly, such host factors may serve as excellent candidates for antiviral therapy. Therefore, we explored whether pharmacological inhibition of GATA6 affects ACE2 expression and consequently, coronavirus infection. Pyrrothiogatain, previously reported to inhibit the DNA-binding activity of GATA6 and other members of the GATA family [71], was tested for its ability to induce downregulation of ACE2 and subsequent inhibition of viral infection. As expected, the drug exhibited a dose-dependent inhibition of GATA6 accompanied by insignificant cytotoxicity in Calu-3 cells (Fig. S2). Inhibition of GATA6 resulted in a lower ACE2 expression (Fig. 5G, Fig.S1) confirming the causal connection between GATA6 and ACE2. Next, cells pretreated with Pyrrothiogatain for 48 hours were infected with SARS-CoV-2. Immunofluorescence analysis evidenced cytoplasmic SARS-CoV-2 staining in untreated cells while significantly reduced infection was observed in Pyrrothiogatain-treated cells (Fig. 5H). The data strongly supports the possibility that novel therapeutic strategies to countermeasure viral infection via down-modulation of ACE2 may employ specific GATA6 targeting.

In summary, our screens in human lung epithelial cells, provide a comprehensive catalog of cellular factors and functional pathways critical for the infection of WT-SARS-CoV-2 and additional VOCs. These include known and novel host-factors such as the viral receptor ACE2 and various components belonging to the Clathrin-dependent transport pathway, ubiquitination and Heparan sulfate biogenesis. In addition, our comparative analysis highlights commonalities and differences between SARS-CoV-2 variants and enables the identification of the receptor KREMEN2 as unique gene required only to the Alpha variant. The differences exhibited by the viral variants with respect to the essentiality of specific host factors for their life-cycle is of outmost importance since they may provide an explanation for the reported differences in their pathogenicity, infectivity and disease progression.

Our study found an absolute requirement of GATA6 for infection by WT-SARS-CoV-2 as well as the Alpha and Beta variants. Furthermore, analysis of clinical samples of COVID-19 infected patients showed an elevated level of GATA6, suggesting that in the course of the disease, viral-induced modulation of the level of GATA6 occurs. Further investigations revealed that GATA6 regulates ACE2 expression and is critical for the entry step of SARS-CoV-2. Notably, pharmacological inhibition of GATA6 results in down-modulation of ACE2 and consequently to inhibition of the viral infectivity suggesting that this protein may serve as a therapy target. Taken together, these findings represent a wide source of information for the understanding of SARS-CoV-2 pathogenesis and for future development of host-directed therapies.

Material &Methods

Cell lines

HEK293T (ATCC- CRL-3216), Vero-E6 (ATCC- CRL-1586) and Calu-3 cells (ATCC- HTB-55) were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% heat-inactivated fetal bovine serum (FBS), 10 mM non-essential amino acids (NEAA), 2 mM L-Glutamine, 1 mM Sodium pyruvate and 1% Penicillin/Streptomycin (all from Biological Industries, Israel) at 37°C, 5% CO_2. All cell lines tested negative for mycoplasma.

Virus stocks

SARS-CoV-2 (GISAID accession EPI_ISL_406862) was kindly provided by Bundeswehr Institute of Microbiology, Munich, Germany. SARS-CoV-2 B.1.1.7 (Alpha) and B.1.351 (Beta) variants of concern (VOCs) were kindly provided by Michal Mandelboim, MOH Central Virology Lab, Tel Hashomer, Israel. Original and B.1.351 VOC Virus stocks were propagated (2-3 passages) on Vero-E6 cells while B.1.1.7 VOC was propagated on Calu-3 cells. All viruses were tittered on Vero-E6 cells. Handling and working with SARS-CoV-2 virus were conducted in a BSL3 facility in accordance with the biosafety guidelines of the Israel Institute for Biological Research (IIBR).

Genome-wide CRISPR screens

To generate CRISPR KO libraries, a total of $4*10^8$ Calu-3 cells were transduced with lentivirus of human Brunello Human CRISPR library (Addgene #73179, gift from David Root and John Doench) in the presence of 0.5 mg/ml polybrene (TR-1003, Sigma), at a MOI of 0.3. Two days post-transduction, 5 ug/ml puromycin (ant-pr-1, InvivoGen) was added to the media and transduced cells were selected for seven days. Twenty-four hours prior to infection with SARS-CoV-2, $1.5*10^7$ Calu-3 library-cells were seeded in 75-cm2 flasks. Mock infected cells were harvested 48 hours after seeding and served as a reference for sgRNA enrichment analysis. Three screens, using 3 variants of SARS-CoV-2, were performed with $3*10^7$ cells in each screen which is sufficient for the representation of each sgRNA into ~400 unique cells. All screens were performed in duplicate. For SARS-CoV-2 infection cells were washed once with RPMI without FBS and infected with SARS-

CoV-2 virus, at a MOI of 0.02 to 0.04, in the presence of 20 µg per ml TPCK trypsin (Thermo scientific) and 2% FBS. Flasks were incubated for 1 hour at 37 °C to allow viral adsorption. Then, RPMI medium supplemented with 2% FBS, was added to each well [72]. Nine days post infection with SARS-CoV-2, surviving cells pellet was dissolved in 10% triton and boiled for 30 minutes. Genomic DNA (gDNA) was extracted using a QIAamp DNA mini kit (Qiagen). sgRNA sequences were amplified by one-step PCR using primers with illumine adapters and the gDNA as a template. A master mix consisted of 75 µl ExTaq DNA Polymerase (Clontech), 1,000 µl of 10X ExTaq buffer, 800 µl of dNTP provided with the enzyme, 50 µl of P5 stagger primer mix (stock at 100 µM concentration), and 2,075 µl water. Each PCR reaction consisted of 50 µl gDNA, 40 µl PCR master mix and 10 µl of uniquely barcoded P7 primer (stock at 5 µM concentration). PCR cycling conditions: an initial 1 min at 95°C; followed by 30 sec at 95°C, 30 sec at 53°C, 30 sec at 72°C, for 28 cycles; and a final 10 min extension at 72°C; hold at 4°C. The p5 stagger primer and uniquely barcoded P7 primers were synthesized at Integrated DNA Technologies, Inc. Primers sequences are listed in supplementary Table S1. Pooled PCR products were purified with Agencourt AMPure XP-SPRI magnetic beads according to manufacturer's instruction (Beckman Coulter, A63880). Samples were sequenced on a Illumine MiSeq platform. Reads were counted by alignment to a reference file of all possible sgRNA present in the library. The read was then assigned to a condition on the basis of the 8 nt index included in the p7 primer. The lentiviral plasmid DNA pool was sequenced as a reference.

Analysis of CRISPR-Cas9 genetic screen data.

MAGeCK v0.5.6 [73] was used to count sgRNA from FASTQ files and to analyze the selection effect of genes based on the change in sgRNA distribution, using the robust rank aggregation (RRA) algorithm with normalization to total reads. Directed score for each gene was calculated by taking the enrichment MAGeCK RRA scores for genes with a positive fold change from control and the depletion MAGeCK RRA score for genes with a negative or zero-fold change. Directed scores for sgRNAs was calculated by multiplying the MAGeCK sgRNA scores by

1 or (-1) according to the direction of change. These scores were scaled to sd=1 and centered to the mean.

Pathway and network analysis

Directed scores were used as input for gene set enrichment analysis (GSEA version 4.1) with GO biological process (c5.bp) from MSigDB (version 7.4) [74, 75]. Results were filtered to get significantly enriched pathways with false discovery rate of less than 0.01. To further analyze functional pathways enriched in positive screen hits, the 200 genes that ranked the highest based on MAGeCK enrichment score were selected and used as input for STRING protein-protein interaction network analysis (ref Szklarczyk et al., 2019) using default parameters. The STRING network was imported into Cytoscape [76]. Genes were highlighted based on association with selected REACTOME pathways. Genes that were not highlighted or connected to at least two other genes were excluded from the graph.

Generation of Calu-3 KO cell lines.

DNA oligos (Integrated DNA Technologies, Inc.) containing sgRNA sequences (see supplementary Table S1) were annealed and ligated into lentiCRISPRv2 (Addgene, #52961, gift from Feng Zhang). Lentivirus was packaged by co-transfection of constructs with the 2nd generation packaging plasmids pMD2.G and PsPax using jetPEI (Polyplus-transfection) into 6-well plates with HEK293T cells according to protocol. Sixty hours post transfection supernatants were collected, centrifuged at 1500 rpm for 5 minutes and filtered through a 0.45 µm filter. Calu-3 cells were transduced with lentiviruses in the presence of 7 ug/ml polybrene (TR-1003, Sigma) and then selected with puromycin for 7 days. Knockout was confirmed by western blot and real-time PCR. For GATA6 complementation experiment, cells were complemented by transfection of 2.5 µg pBabe 3XFLAG-wt GATA6-3XAU1 puro vector (Addgene #72607) using Lipofectamine 3000 (L3000015 ThermoFisher) according to manufacturer's instructions. Forty-eight hours post transfection, the expression of GATA6 was analyzed by western blot and real-time PCR.

Imaging and Immunofluorescence assay

Calu-3 knockout or WT cells were seeded at a density of $5*10^4$ cells per well in 8-well-µ-slides (Ibidi), or $4*10^5$ in 12 well plates, infected as described above or left uninfected. Imaging was performed on a Zeiss AxioObserver Z1 wide-field microscope using a ×10 objective and Axiocam 506 mono camera at the indicated time point. For Immunofluorescence experiments, cells were washed once with PBS, fixed with 4% paraformaldehyde (PFA) in PBS for 15 minutes and permeabilized with using 0.5% Triton X-100 (Sigma T9284) for 2 minutes. The fixed cells were blocked in PBS containing 2% FCS and stained with primary antibodies (diluted 1:200 in blocking buffer) for 1 hour at room temperature. After washing with PBS, cells were incubated with Alexa Fluor 488 conjugated anti-rabbit for 0.5 hour. Nuclei were stained with DAPI (200 µg/ml D9542 Sigma) for 5 minutes at room temperature. Imaging was performed on an LSM 710 confocal scanning microscope (Zeiss, Jena, Germany). Primary antibodies used in this study: Rabbit anti-GATA6 (Abcam ab175349), hyperimmune Rabbit serum from intervenous (i.v) SARS-CoV-2 infected Rabbits.

Viral titration

Vero-E6 cells were seeded in 12-well plates (5×10^5 cells/well) and grown overnight in growth medium. Serial dilutions of SARS-CoV-2 were prepared in infection medium (MEM containing 2% FBS with NEAA, glutamine, and Penicillin/Streptomycin), and used to infect Vero-E6 monolayers in duplicates or triplicates (200 µl/well). Plates were incubated for 1 hour at 37 °C to allow viral adsorption. Then, 2 ml/well of overlay [MEM containing 2% FBS and 0.4% tragacanth (Merck, Israel)] was added to each well and plates were incubated at 37 °C, 5% CO2 for 72 hours. The media were then aspirated, and the cells were fixed and stained with 1 ml/well of crystal violet solution (Biological Industries, Israel). The number of plaques in each well was determined, and SARS-CoV-2 titer was calculated.

Analysis of GATA6 expression in different cell lines

Published mRNA-seq data used for A549, Vero-E6 and Calu-3 cells are from Kinori et al. 2016, Finkel et al. 2021a and Finkel et al. 2021b., respectively. Gene level

values were normalized to TPM (transcripts per million) by dividing each value by the sum of each sample and multiplying by one million.

Analysis of clinical samples

Negative and positive qRT-PCR nasopharyngeal swab samples from symptomatic and asymptomatic individuals were collected as part of routine scanning of nursing homes. Ethical review and approval were waived, since the samples used for this study were leftovers of anonymized samples. Viral RNA was extracted using RNAdvance Viral XP kit (Beckman Coulter). From each sample 200 µL were added to LBF lysis buffer, and further processed on the Biomek i7 Automated Workstation (Beckman Coulter), according to the manufacturer's protocol. Each sample was eluted in 50 µL of RNase-free water. Real-time RT-PCR assays were performed using the SensiFASTTM Probe Lo-ROX one-step kit (Bioline). In each reaction the primers final concentration was 600 nM and the probe concentration was 300 nM. Thermal cycling was performed at 48°C for 20 minutes for reverse transcription, followed by 95°C for 2 min, and then 45 cycles of 94°C for 15 sec, 60°C for 35 sec. Primers and probes (listed in supplementary Table S1) were designed using the Primer Express Software (Applied Biosystems) and purchased from Integrated DNA Technologies, Inc.

siRNA inhibition assay

Calu-3 Cells were transfected with siRNA validated for GATA6 or negative control (TriFECTa Kit DsiRNA Duplex, IDT) (listed in supplementary Table S1) in the presence of Lipofectamine RNAiMAX reagent (Life Technologies), according to manufacturer's standard protocol. At the indicated time after transfection, cells were harvested. All experiments were performed in triplicate, and representative results are reported.

RT-PCR

Total RNA was extracted using RNeasy Mini Kit (Qiagen 74104), reverse transcribed using qScript cDNA Synthesis kit (95047-025, Quanta bio) according to protocols, and subjected to real-time PCR analysis using perfecta SYBR Green FastMix Low ROX (Quanta bio 95074-250). Data shown are the relative abundance of the indicated mRNA normalized to that of GAPDH. Gene-specific

primers were as follows ACE2: GGTCTTCTGTCACCCGATTT (forward) and CATCCACCTCCACTTCTCTAAC (reverse); GATA6: CACTACCTGTGCAACGCCTG (forward) and GCCGTGATGAAGGCACGC (reverse); GAPDH: TGCACCACCAACTGCTTAGC (forward) and GGCATGGACTGTGGTCATGAG (reverse).

Western blot

Cells were lysed in RIPA Lysis Buffer (Merck 20-188) in the presence of cOmplete protease inhibitor cocktail (Roche 11697498001). Lysates were nutated at 4°C for 10 minutes, then centrifuged at 20,000 × g for 15 minutes at 4°C. Equal amounts of cell lysates were denaturated in 4X Laemmli sample buffer (Bio-RAD #1610747), separated on 4-12% NuPAGE Bis-Tris gels (invitrogen), blotted onto nitrocellulose membranes and immunoblotted with primary antibodies αGATA6 (Abcam ab175349), αGAPDH (Cell Signaling 14C10), αACE2 (Sino Biological #10108-T60). Secondary antibody used was IRDye® 800CW conjugated Goat anti-Rabbit (Licor). Reactive bands were detected by Odyssey CLx infrared imaging system (Licor). Protein concentration was measured by the BCA protein assay kit (Pierce 23225). Protein quantification was performed on Licor software.

EMSA

Electrophoretic mobility assay (EMSA). HEK293T cells were transfected with 3 µg pBabe 3XFLAG-wt-GATA6-3AU1 puro plasmid (Addgene #76207) using Lipofectamine 3000 (L3000015 ThermoFisher) according to manufacturer's instructions. Forty-eight hours post transfection, cells were lysed in RIPA Lysis Buffer (Merck 20-188) in the presence of cOmplete protease inhibitor cocktail (Roche 11697498001) and nutated at 4°C for 10 minutes. The cell lysate was cleared by centrifugation, concentrated and dialyzed into PBS using Amicon Ultra-4 30K (UFC803024, Merck). The EMSA was preformed using two sets of double stranded DNA oligonucleotides of ~40 bp spanning the putative GATA6 binding sites in the promotor region of ACE2. 5'-IRDye700-labled single-stranded nucleotide probes or PCR primers, as detailed in supplementary Table S1, were purchased from Integrated DNA Technologies, Inc. Complementary single-stranded nucleotides were mixed at 100nM concentration, heated at 100ºC for 5

minutes, and then cooled down to room temperature to allow duplex formation. Binding reactions were carried out using the Odyssey® EMSA Buffer Kit (829-07910, Li-Cor) according to manufacturer's instructions, in the presence of NP-40 and MgCl2. In competition assays, unlabeled competitor probes were included in the binding reactions in 200-fold excess relative to labeled probes. The reactions were incubated for 30 minutes at room temperature in the dark. 1X Orange dye was added, and the samples were run on 8% TBE gel in 0.5% TBE buffer (LC6775, Novex) at 70 V for 90 min in the dark and imaged on an Odyssey® infrared imaging system (Li-Cor Biosiences).

GATA6 inhibition assay

Pyrrothiogatain (3-(2,5-dimethyl-1H-pyrrol-1yl) thiophene-2- carboxylic acid) was purchased from Santa Cruz (cat #sc-352288A). Cells were treated with the indicated concentration of Pyrrothiogatain (50mM stock dissolved in DMSO) for 48 hours prior to infection with SARS-CoV-2. Infection and subsequent immunofluorescence imaging were carried as described above.

Acknowledge

We thank Tamar Aminov and Inbar Chomsky for technical assistance. We thank Emanuelle Mamroud for providing valuable feedback

References

1. Fung, T.S. and D.X. Liu, *Human Coronavirus: Host-Pathogen Interaction.* Annu Rev Microbiol, 2019. **73**: p. 529-557.
2. Hoffmann, M., H. Kleine-Weber, and S. Pöhlmann, *A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells.* Mol Cell, 2020. **78**(4): p. 779-784.e5.
3. Letko, M., A. Marzi, and V. Munster, *Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses.* Nat Microbiol, 2020. **5**(4): p. 562-569.
4. Li, W., et al., *Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus.* Nature, 2003. **426**(6965): p. 450-4.
5. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* Nature, 2020. **579**(7798): p. 270-273.
6. Hartenian, E., et al., *The molecular virology of coronaviruses.* J Biol Chem, 2020. **295**(37): p. 12910-12934.
7. Shereen, M.A., et al., *COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses.* J Adv Res, 2020. **24**: p. 91-98.
8. V'Kovski, P., et al., *Coronavirus biology and replication: implications for SARS-CoV-2.* Nat Rev Microbiol, 2021. **19**(3): p. 155-170.
9. Bertram, S., et al., *TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium.* J Virol, 2013. **87**(11): p. 6150-60.
10. Hoffmann, H.H., et al., *Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors.* Cell Host Microbe, 2021. **29**(2): p. 267-280.e5.
11. Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280.e8.
12. Shirato, K., M. Kawase, and S. Matsuyama, *Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2.* J Virol, 2013. **87**(23): p. 12552-61.
13. Simmons, G., et al., *Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry.* Proc Natl Acad Sci U S A, 2005. **102**(33): p. 11876-81.
14. Gordon, D.E., et al., *Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms.* Science, 2020. **370**(6521).
15. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.
16. Stukalov, A., et al., *Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV.* Nature, 2021. **594**(7862): p. 246-252.
17. V'Kovski, P., et al., *Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling.* Elife, 2019. **8**.
18. Baggen, J., et al., *Genome-wide CRISPR screening identifies TMEM106B as a proviral host factor for SARS-CoV-2.* Nat Genet, 2021. **53**(4): p. 435-444.
19. Daniloski, Z., et al., *Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells.* Cell, 2021. **184**(1): p. 92-105.e16.
20. Heaton, B.E., et al., *SRSF protein kinases 1 and 2 are essential host factors for human coronaviruses including SARS-CoV-2.* bioRxiv, 2020.
21. Biering, S.B., et al., *Genome-wide, bidirectional CRISPR screens identify mucins as critical host factors modulating SARS-CoV-2 infection.* bioRxiv, 2021(doi: https://doi.org/10.1101/2021.04.22.440848).
22. Schneider, W.M., et al., *Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks.* Cell, 2021. **184**(1): p. 120-132.e14.
23. Wang, R., et al., *Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses.* Cell, 2021. **184**(1): p. 106-119.e14.
24. Wei, J., et al., *Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection.* Cell, 2021. **184**(1): p. 76-91.e13.
25. Zhu, Y., et al., *A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry.* Nat Commun, 2021. **12**(1): p. 961.
26. Dagan, N., et al., *BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting.* N Engl J Med, 2021. **384**(15): p. 1412-1423.
27. Shilo, S., H. Rossman, and E. Segal, *Signals of hope: gauging the impact of a rapid national vaccination campaign.* Nat Rev Immunol, 2021. **21**(4): p. 198-199.
28. Baden, L.R., et al., *Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine.* N Engl J Med, 2021. **384**(5): p. 403-416.
29. Polack, F.P., et al., *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.* N Engl J Med, 2020. **383**(27): p. 2603-2615.
30. Abdool Karim, S.S. and T. de Oliveira, *New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications.* N Engl J Med, 2021. **384**(19): p. 1866-1868.
31. Faria, N.R., et al., *Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil.* Science, 2021. **372**(6544): p. 815-821.
32. Rambaut, A., et al., *A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology.* Nat Microbiol, 2020. **5**(11): p. 1403-1407.
33. Tegally, H., et al., *Detection of a SARS-CoV-2 variant of concern in South Africa.* Nature, 2021. **592**(7854): p. 438-443.
34. https://covlineages.org/global_report_B.1.1.7.html, PublicHealthEngland. 2020, 2020.

35. Andreano, E., et al., *SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma.* bioRxiv, 2020.
36. Huang, B., et al., *Serum sample neutralisation of BBIBP-CorV and ZF2001 vaccines to SARS-CoV-2 501Y.V2.* Lancet Microbe, 2021. **2**(7): p. e285.
37. Madhi, S.A., et al., *Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant.* N Engl J Med, 2021. **384**(20): p. 1885-1898.
38. Wang, Z., et al., *mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants.* Nature, 2021. **592**(7855): p. 616-622.
39. Weisblum, Y., et al., *Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants.* Elife, 2020. **9**.
40. Wu, K., et al., *mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants.* bioRxiv, 2021.
41. Health, E.P., *Investigation of novel SARS-CoV-2 variant - Variant of Concern 202012/01 - Technical briefing 3 London, United Kingdom.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950823/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3_-_England.pdf, 2020.
42. Latif, A.A., et al., *P.1 Lineage Report. outbreak.info* https://outbreak.info/situation-reports?pango=P.1&loc=BRA&loc=USA&loc=USA_US, 2021.
43. Latif, A.A., et al., *B.1.351 Lineage Report* https://outbreak.info/situation-reports?pango=B.1.351, 2021.
44. Doench, J.G., et al., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.* Nat Biotechnol, 2016. **34**(2): p. 184-191.
45. Islam, A. and M.A. Khan, *Lung transcriptome of a COVID-19 patient and systems biology predictions suggest impaired surfactant production which may be druggable by surfactant therapy.* Sci Rep, 2020. **10**(1): p. 19395.
46. Valerdi, K.M., et al., *The Role of the Host Ubiquitin System in Promoting Replication of Emergent Viruses.* Viruses, 2021. **13**(3).
47. Cao, Z., et al., *Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response.* Cell Mol Immunol, 2021. **18**(3): p. 746-748.
48. Zhang, H., et al., *Ubiquitin-Modified Proteome of SARS-CoV-2-Infected Host Cells Reveals Insights into Virus-Host Interaction and Pathogenesis.* J Proteome Res, 2021. **20**(5): p. 2224-2239.
49. Joly, S., et al., *Interferon Regulatory Factor 6 Has a Protective Role in the Host Response to Endotoxic Shock.* PLoS One, 2016. **11**(4): p. e0152385.
50. Bayati, A., et al., *SARS-CoV-2 infects cells after viral entry via clathrin-mediated endocytosis.* J Biol Chem, 2021. **296**: p. 100306.
51. Wang, H., et al., *SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway.* Cell Res, 2008. **18**(2): p. 290-301.
52. Feng, X.H., et al., *The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation.* Genes Dev, 1998. **12**(14): p. 2153-63.
53. Sherman, E.J., et al., *Identification of ACE2 modifiers by CRISPR screening.* bioRxiv, 2021.
54. Cagno, V., et al., *Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias?* Viruses, 2019. **11**(7).
55. Zhu, W., J. Li, and G. Liang, *How does cellular heparan sulfate function in viral pathogenicity?* Biomed Environ Sci, 2011. **24**(1): p. 81-7.
56. Clausen, T.M., et al., *SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2.* Cell, 2020. **183**(4): p. 1043-1057.e15.
57. Jarrous, N., et al., *Function and subnuclear distribution of Rpp21, a protein subunit of the human ribonucleoprotein ribonuclease P.* Rna, 2001. **7**(8): p. 1153-64.
58. Voelker, D.R. and M. Numata, *Phospholipid regulation of innate immunity and respiratory viral infection.* J Biol Chem, 2019. **294**(12): p. 4282-4289.
59. Bollag, W.B. and J.N. Gonzales, *Phosphatidylglycerol and surfactant: A potential treatment for COVID-19?* Med Hypotheses, 2020. **144**: p. 110277.
60. Mao, B., et al., *Kremen proteins are Dickkopf receptors that regulate Wnt/beta-catenin signalling.* Nature, 2002. **417**(6889): p. 664-7.
61. Zebisch, M., et al., *Structure of the Dual-Mode Wnt Regulator Kremen1 and Insight into Ternary Complex Formation with LRP6 and Dickkopf.* Structure, 2016. **24**(9): p. 1599-605.
62. Staring, J., et al., *KREMEN1 Is a Host Entry Receptor for a Major Group of Enteroviruses.* Cell Host Microbe, 2018. **23**(5): p. 636-643.e5.
63. Gu, Y., et al., *Interaction network of SARS-CoV-2 with host receptome through spike protein.* bioRxiv, 2020.
64. Bruno, M.D., et al., *GATA-6 activates transcription of surfactant protein A.* J Biol Chem, 2000. **275**(2): p. 1043-9.
65. Silveyra, P. and J. Floros, *Genetic complexity of the human surfactant-associated proteins SP-A1 and SP-A2.* Gene, 2013. **531**(2): p. 126-32.
66. Finkel, Y., et al., *The coding capacity of SARS-CoV-2.* Nature, 2021. **589**(7840): p. 125-130.
67. Finkel, Y., et al., *SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis.* Nature, 2021. **594**(7862): p. 240-245.
68. Bercovich-Kinori, A., et al., *A systematic view on influenza induced host shutoff.* Elife, 2016. **5**.
69. Chou, C.F., et al., *ACE2 orthologues in non-mammalian vertebrates (Danio, Gallus, Fugu, Tetraodon and Xenopus).* Gene, 2006. **377**: p. 46-55.
70. Pedersen, K.B., et al., *The transcription factor HNF1α induces expression of angiotensin-converting enzyme 2 (ACE2) in pancreatic islets from evolutionarily conserved promoter motifs.* Biochim Biophys Acta, 2013. **1829**(11): p. 1225-35.
71. Nomura, S., et al., *Pyrrothiogatain acts as an inhibitor of GATA family proteins and inhibits Th2 cell differentiation in vitro.* Sci Rep, 2019. **9**(1): p. 17335.

72. Yahalom-Ronen, Y., et al., *A single dose of recombinant VSV-ΔG-spike vaccine provides protection against SARS-CoV-2 challenge.* Nat Commun, 2020. **11**(1): p. 6402.
73. Li, W., et al., *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.* Genome Biol, 2014. **15**(12): p. 554.
74. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
75. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
76. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.

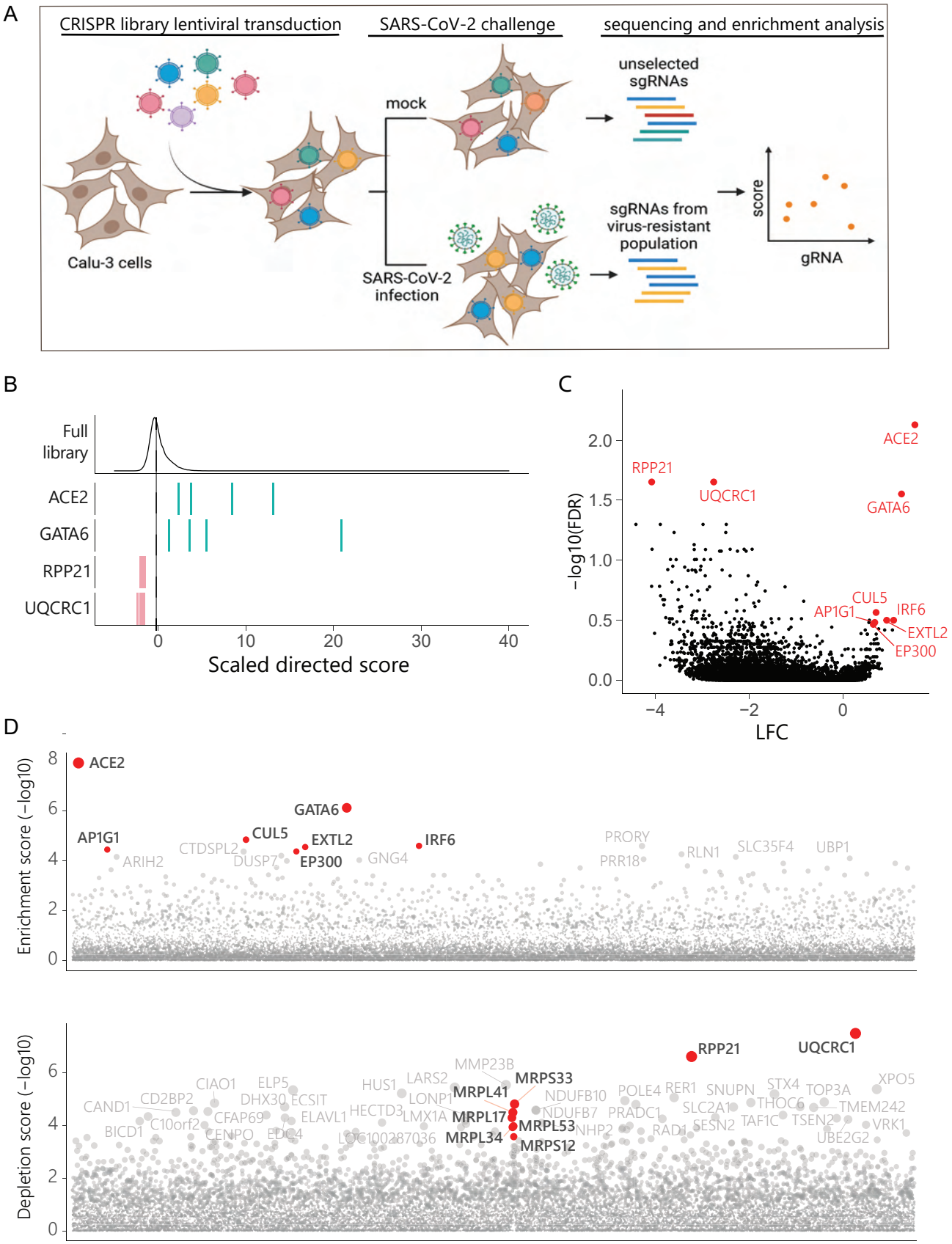


Figure 1. CRISPR genome-wide screens in human Calu-3 cells identify host factors important for infection by SARS-CoV-2. (A) Schematic of Genome-wide CRISPR screening workflow for the identification of SARS-CoV-2 host factors. Calu-3 cells were transduced with the Brunello CRISPR library, selected with puromycin and infected with WT-SARS-CoV-2, Alpha or Beta VOCs. Surviving cells were harvested 7-9 days post infection (dpi). The abundance of each sgRNA in the mock controls and selected population was determined by high-throughput sequencing and a gene enrichment analysis was performed. The figure was created with BioRender.com **(B)** Performance in the screen of example sgRNAs. The density across the screens is plotted for the full library (top) and for the 4 sgRNAs targeting top resistance hits (in blue) and top sensitization hits (in red). The dashed line indicates a residual of zero. **(C)** Volcano plot showing top genes conferring resistance and sensitivity to SARS-CoV-2. (D) Gene enrichment and depletion score for CRISPR screens of SARS-CoV-2 infection. Enrichment scores were determined by MaGECK.

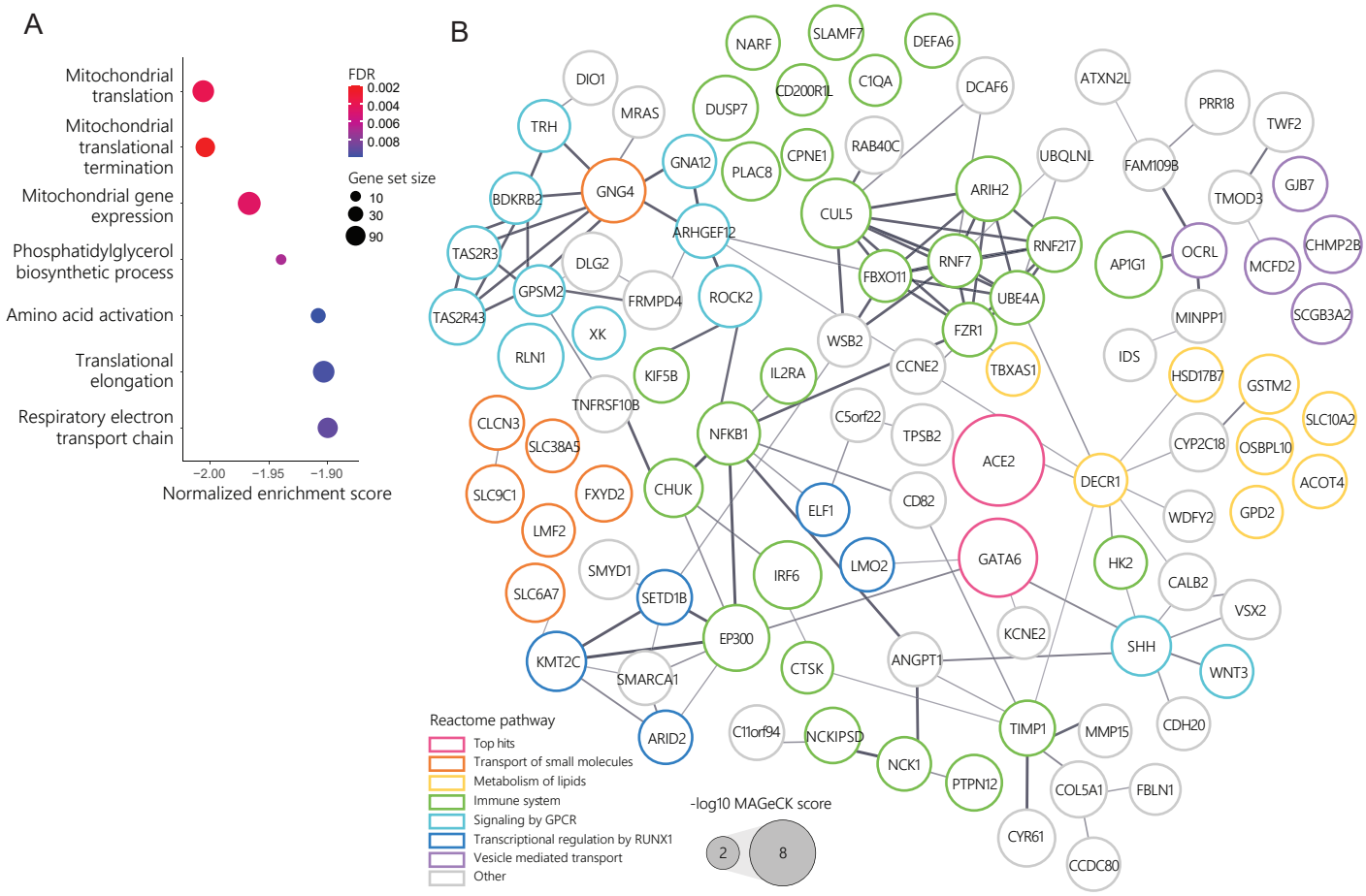


Figure 2. Gene set enrichment and interaction networks analysis identify functional pathways important for SARS-CoV-2 infection. (A) Gene set enrichment analysis (GSEA) was performed on hits from all strain screens combined. Results were filtered to get significantly enriched pathways with false discovery rate of less than 0.01. **(B)** Protein-protein interaction network for top 200 enriched hits from all strain screens combined based on STRING analysis. Edge thickness represents interaction score. Nodes are colored by selected REACTOME pathways and sized according to log10 transformed screen enrichment score.

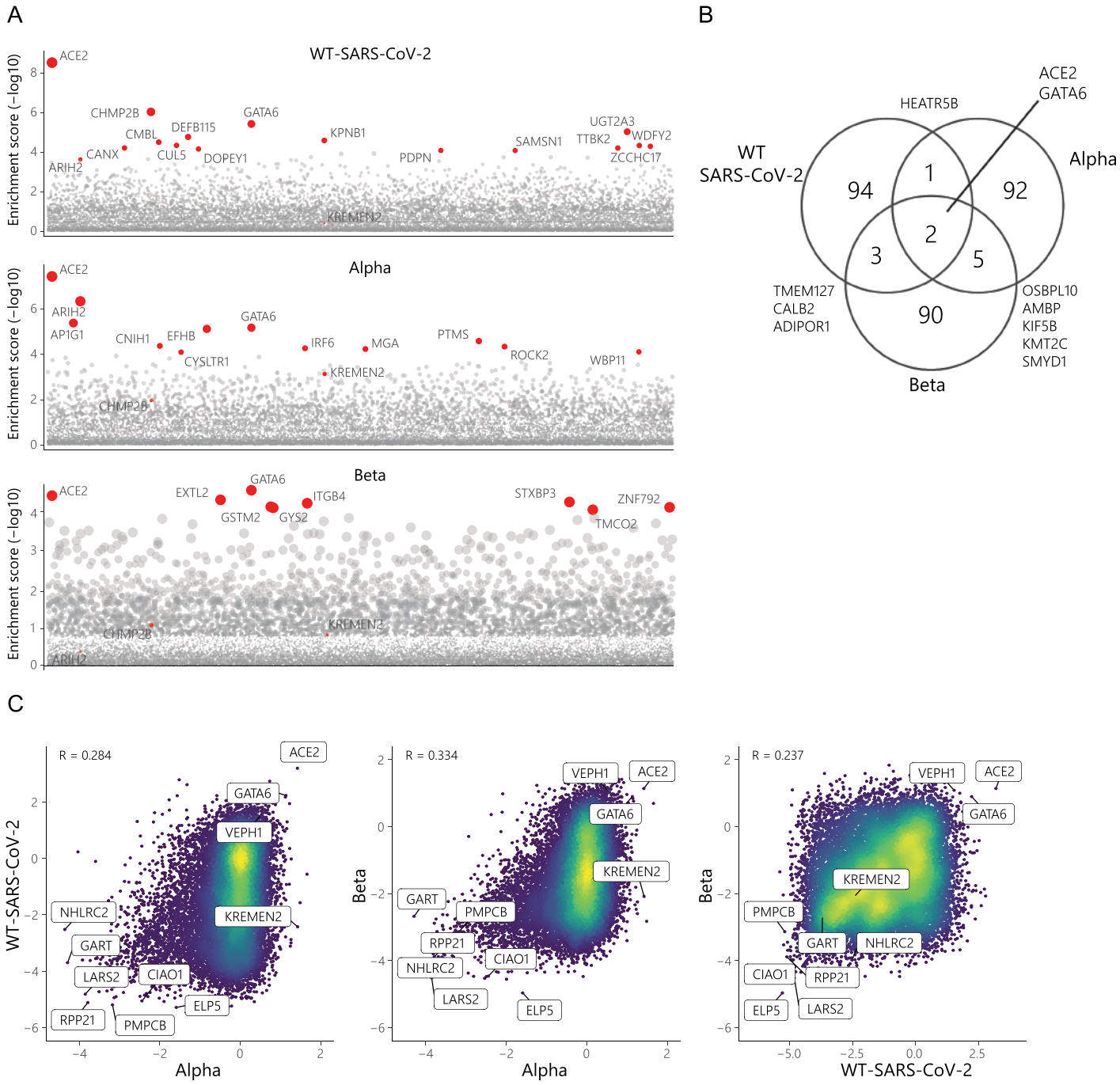


Figure 3. Differential analysis of variants of SARS-CoV-2 screens highlights shared and specific host-factors dependencies. (A) Gene enrichment score for CRISPR screens of WT-SARS-CoV-2, Alpha and Beta variants infection. Enrichment scores were determined by MaGECK. **(B)** Venn diagrams depicting overlap of the top 100 pro-viral genes enriched in the screens of WT-SARS-CoV-2, Alpha and Beta variants infection. **(C)** Scatter plot depicting gene enrichment in WT-SARS-CoV-2, Alpha and Beta variants screens. R, Pearson correlation.

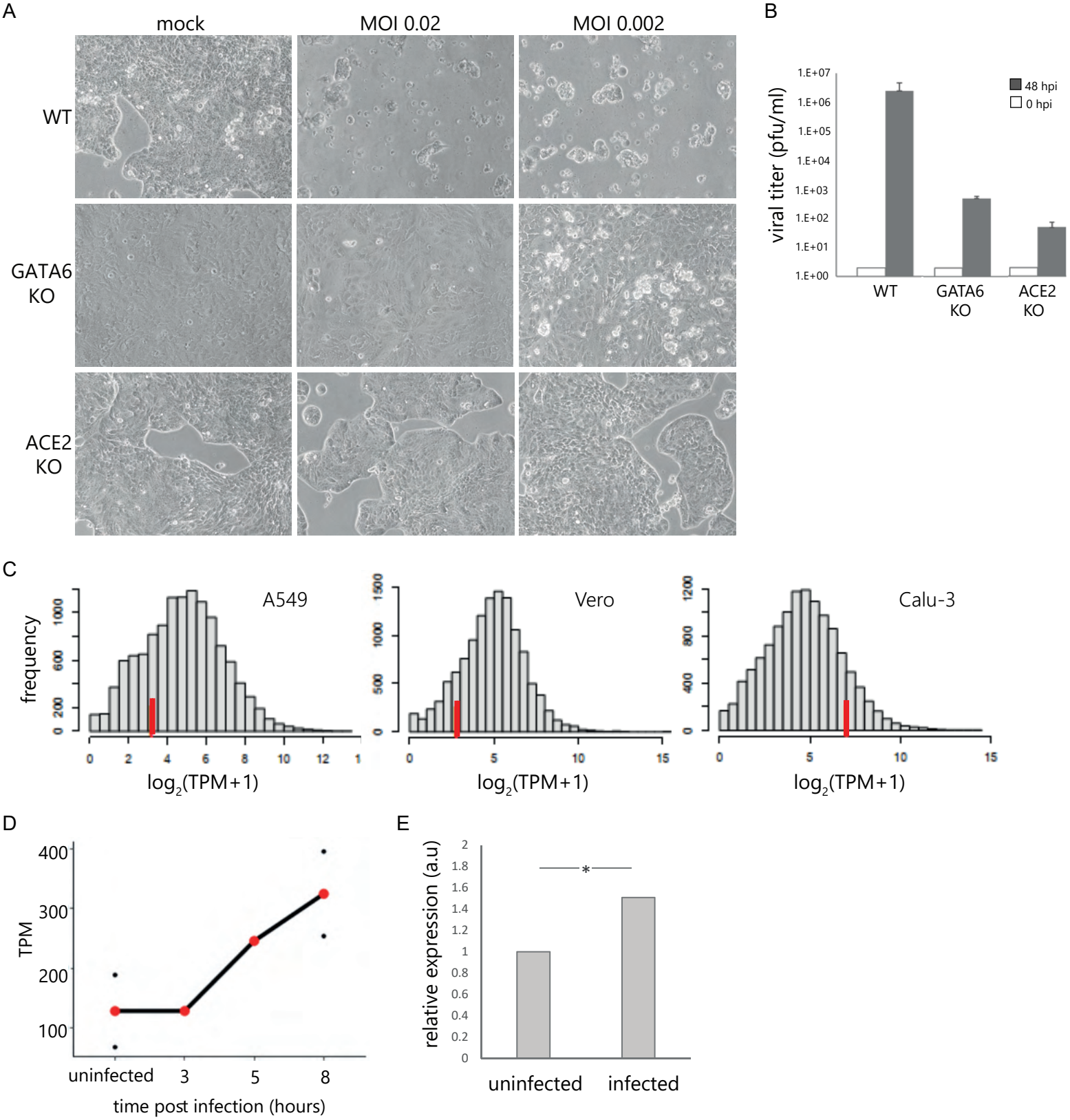
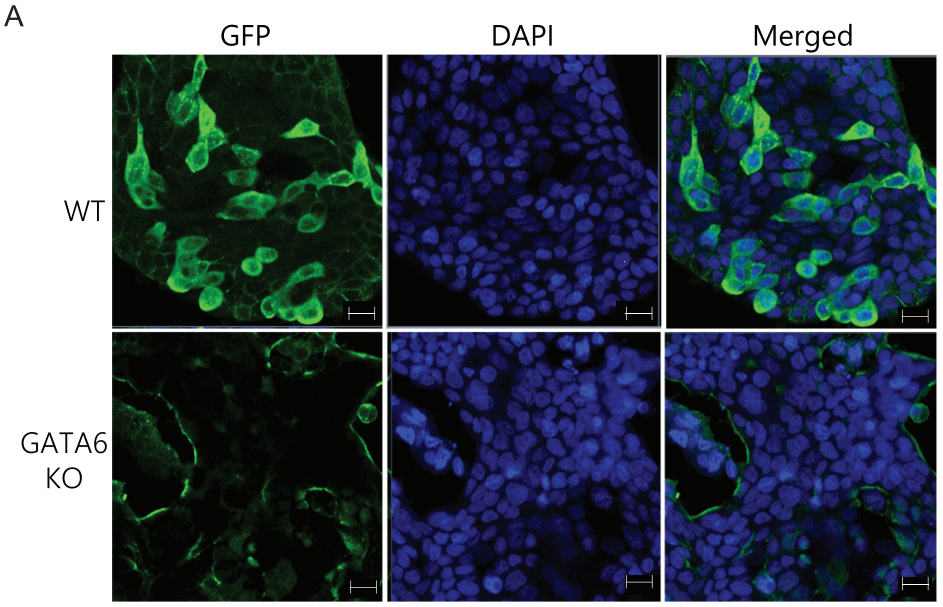
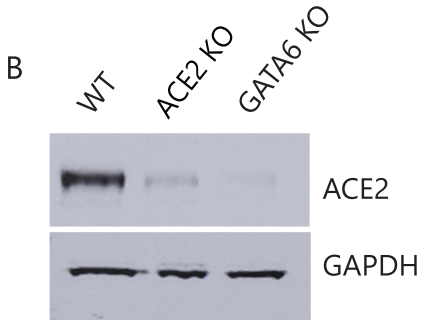
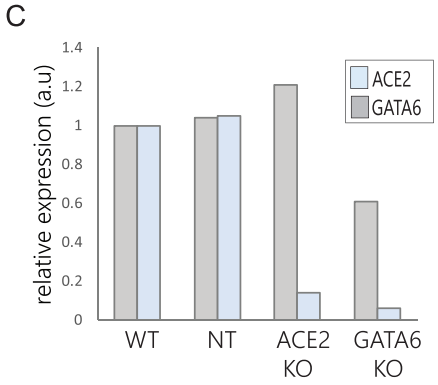
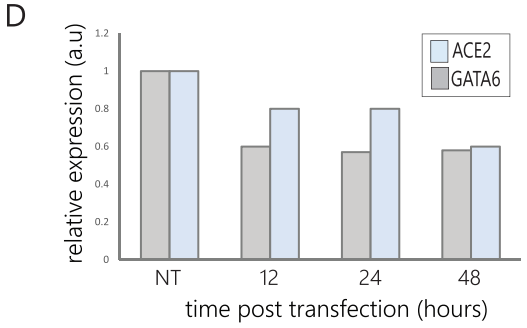


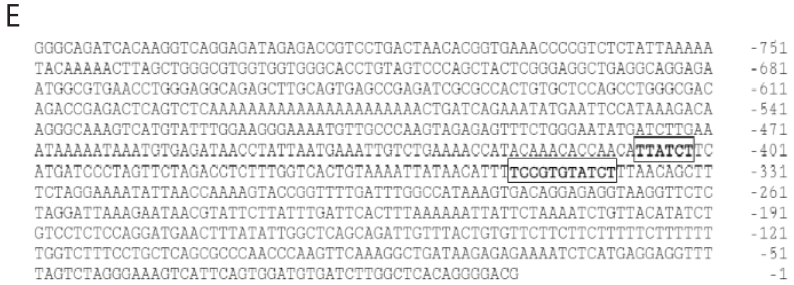
Figure 4. GATA6 is a pro-viral host factor that is upregulated in response to infection. (A) Control, GATA6 and ACE2 disrupted Calu-3 cells were infected with WT-SARS-CoV-2 at the indicated MOI. Cell viability was visualized by light microscopy 48 hours post infection. Representative images of three independent replicates are shown. **(B)** Supernatants were collected from WT and GATA6 knockout cells infected with SARS-CoV-2 at the indicated time points. Virus production, as measured by plaque-forming units (PFU) per milliliter, was determined by plaque assay. Representative results of two independent experiments are shown. **(C)** Histograms of normalized mRNA expression levels in Calu-3, Vero-E6 and A549 cells. Red lines represent the level of GATA6 in each dataset. **(D)** Expression levels of GATA6 at different times along SARS-CoV-2 infection in Calu-3 cells. Individual values shown for 3 hours and 5 hours. Mean values of duplicates shown for uninfected and 8 hours, and individual values are presented as black points. **(E)** real-time PCR measurements of GATA6 mRNA evels in negative (uninfected) and positive (infected) nasopharyngeal swab samples from symptomatic and asymptomatic individuals. p-values are derived from a Student's t-test, *p<0.05.

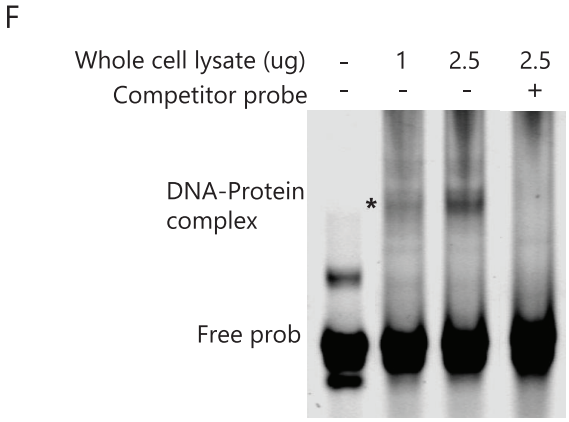

A
GFP
DAPI
Merged
WT
GATA6 KO

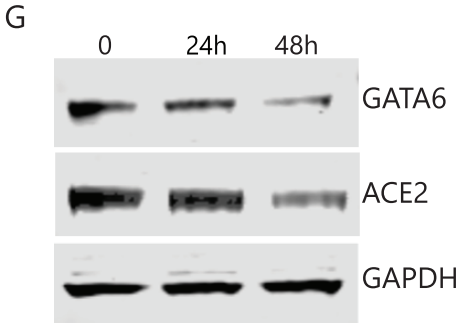

B
WT
ACE2 KO
GATA6 KO
ACE2
GAPDH


C
relative expression (a.u)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
ACE2
GATA6
WT
NT
ACE2 KO
GATA6 KO


D
relative expression (a.u)
1.2
1
0.8
0.6
0.4
0.2
0
ACE2
GATA6
NT
12
24
48
time post transfection (hours)


E
GGGCAGATCACAAGGTCAGGAGATAGAGACCGTCCTGACTAACACGGTGAAACCCCGTCTCTATTAAAAA -751
TACAAAAACTTAGCTGGGCGTGGTGGTGGGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA -681
ATGGCGTGAACCTGGGAGGCAGAGCTTGCAGTGAGCCGAGATCGCGCCACTGTGCTCCAGCCTGGGCGAC -611
AGACCGAGACTCAGTCTCAAAAAAAAAAAAAAAAAAAAAAACTGATCAGAAATATGAATTCCATAAAGACA -541
AGGGCAAAGTCATGTATTTGGAAGGGAAAATGTTGCCCAAGTAGAGAGTTTCTGGGAATATGATCTTGAA -471
ATAAAAATAAATGTGAGATAACCTATTAATGAAATTGTCTGAAAACCATACAAACACCAACATTATCTTC -401
ATGATCCCTAGTTCTAGACCTCTTTGGTCACTGTAAAATTATAACATTTTCCGTGTATCTTTAACAGCTT -331
TCTAGGAAAATATTAACCAAAAGTACCGGTTTTGATTTGGCCATAAAGTGACAGGAGAGGTAAGGTTCTC -261
TAGGATTAAAGAATAACGTATTCTTATTTGATTCACTTTAAAAAATTATTCTAAAATCTGTTACATATCT -191
GTCCTCTCCAGGATGAACTTTATATTGGCTCAGCAGATTGTTTACTGTGTTCTTCTTCTTTTTCTTTTTT -121
TGGTCTTTCCTGCTCAGCGCCCAACCCAAGTTCAAAGGCTGATAAGAGAGAAAATCTCATGAGGAGGTTT -51
TAGTCTAGGGAAAGTCATTCAGTGGATGTGATCTTGGCTCACAGGGGACG -1


F
Whole cell lysate (ug) - 1 2.5 2.5
Competitor probe - - - +
DNA-Protein complex
*
Free prob


G
0
24h
48h
GATA6
ACE2
GAPDH

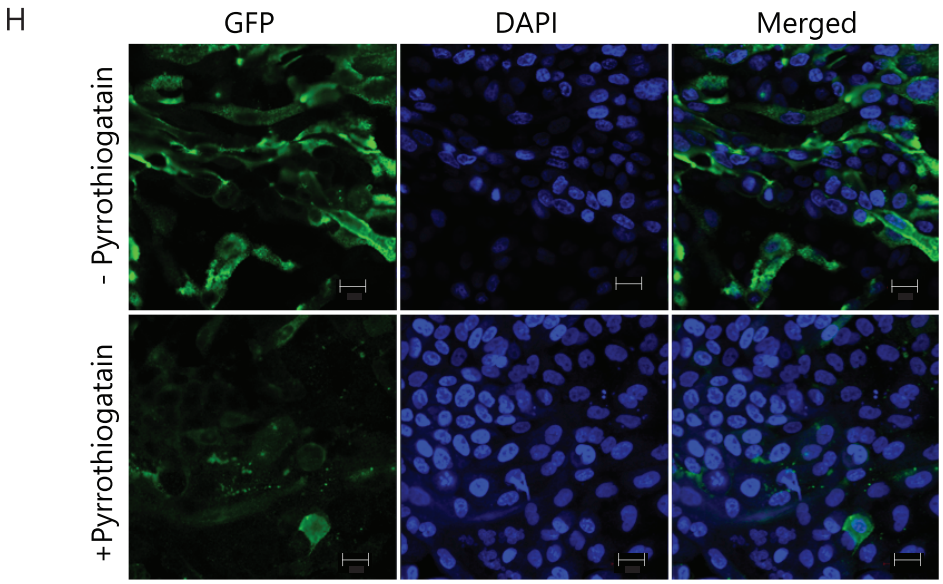

H
GFP
DAPI
Merged
- Pyrrothiogatain
+Pyrrothiogatain

Figure 5. GATA6 is a novel regulator of ACE2 that may serve as a potential host directed therapeutic target. (A) SARS-CoV-2 infected control and GATA6-disrupted Calu-3 cells were fixed 48 hours post infection and stained with antisera against SARS-CoV-2 (green) and DAPI (blue). Scale bar, 20mm. **(B)** ACE2 and GATA6 expression levels were analyzed in control, GATA6 and ACE2 disrupted Calu-3 cells by Western blot and **(C)** by real time PCR. **(D)** Cells transfected with a control or siRNAs targeting GATA6 were analyzed by real-time PCR for GATA6 and ACE2 expression. Representative results of three independent experiments are shown. **(E)** ACE2 ATG codon upstream sequences from nucleotide -1 to -820 are shown. Sequences shown framed in bold are potential GATA binding sites. **(F)** Electrophoretic Mobility Shift Assays (EMSA) of lysates from HEK293T cells over expressing GATA6 with labeled oligonucleotides derived from the promoter region of ACE2 results in the formation of protein-DNA complexes, which was prevented by competitive unlabeled oligonucleotides. **(G)** Calu-3 cells were treated with 500uM Pyrrothiogatain and the expression of GATA6 and ACE2 were analyzed by Western blot at the indicated time. **(H)** Calu-3 cells were pretreated with 500uM Pyrrothiogatain and then infected with SARS-CoV-2. Cells were fixed 24 hours post infection and stained with antisera against SARS-CoV-2 (green) and DAPI (blue). Scale bar, 20mm.