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], Yfat Yahalom-Ronen[3], Nir Paran[3], Theodor Chitlaru[1], Ofir Israeli[1], Inbar Cohen-Gihon[1], Moshe Aftalion[1], Reut Falach[1], Uri Elia[1], Ital Nemet[4], Limor Kliker[4], Michal Mandelboim[4], Adi Beth-Din[1], Tomer Israely[3], Ofer Cohen[1], 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 led 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 phosphatidylglycerol biosynthesis processes appeared to have major anti-viral functions. Comparative analysis of the different VOCs revealed the host factors KREMEN2 and SETDB1 as potential unique candidates 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 revealed that GATA6 directly regulates ACE2 transcription and accordingly, is 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 resulted 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, the cause of the Coronavirus Disease 2019 (COVID-19), responsible for 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 evade 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 engagement of 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. In the host cell cytoplasm, the viral RNA 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 could not provide direct information pertaining to 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 globally disrupt individual genes of the entire host genome and screen for those whose disruption resulted in resistance to viral infection and consequently to host-cell survival. Several reports documented genome-wide CRISPR-Cas9-mediated loss-of-function screens and identification of 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 exhibited as high as 95% efficacy in preventing clinical cases and 100% efficacy in preventing severe disease with the original virus [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, R_0, 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 the genomic approach was expanded to the original WT SARS-CoV-2 and two additional variants, Alpha and Beta. 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 a 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 viral infection and/or 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 sgRNAs 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. Genomic DNA was harvested from surviving cells 7-9 days post infection, and sgRNAs abundance was determined by sequencing (Fig.1A). In the initial analysis, to identify cellular functions important for infection of Calu-3 cells with SARS-CoV-2, irrespective of the variant, the sgRNA distributions measured in the 3 screens were collectively compared to the control non-infected cells. The analysis 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) (Table 1). The sgRNA score- distribution of the full library, non-targeting control sgRNAs and that of top enriched and depleted- sgRNAs hits demonstrated high technical quality of the screens (Fig 1B and Table 2).

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 contribute to immune regulation through NF-kB and various other pathways. Interestingly, the transcription level of GATA6 was shown to be elevated in the lung transcriptome of COVID-19 patients strongly implying that GATA6 takes part of the viral-host cross-talk [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 regarded as a host -defense mechanism for 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 a 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 suggested to be important for SARS-CoV-2 infectivity in Calu-3 cells but not in other cell lines [21]. 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.

In addition to sgRNAs that were enriched following infection, the screen enabled identification of sgRNAs that were depleted in virus-infected cultures suggesting the genes targeted by these sgRNAs may have an anti-viral function. However, gene disruption by itself may affect cell viability resulting in non-specific susceptibility to infection. To further evaluate whether the significant depletion of genes was due to a defect in cell growth or to their anti-viral properties, an additional screen was performed in which Calu-3 cells presented with the CRISPR library were allowed to grow for 7 days, at the end of which those containing sgRNAs influencing viability were depleted (Table 3) [57]. Comparison between sgRNAs depleted in these cells which were not subjected to infection to those depleted from the infected cells, identified several bona-fide anti-viral candidates such as MMP23B, LARS2, HUS1, STX4 and RER1 (Table 2). Interestingly, the MMP23B metallopeptidase (MMPs) was reported to be downregulated in the lungs of COVID-19 patients. The reported downregulation of MMP23B in the presence of the virus may represent a host-pathogen cross-talk mechanism which includes a counter-reaction of the virus for avoiding cellular antiviral responses [58].

Validation of the genetic screens

To validate the candidate genes distinguished by the SARS-CoV-2 screens, individual Calu-3 KO cell-lines were generated by expressing a subset of sgRNAs targeting specific genes which exhibited high levels of enrichment or depletion in the SARS-CoV-2 screens (enrichment score -log10>4) (Table 1). Correct disruption of targeted genes was confirmed by DNA sequencing of their respective chromosomal loci or by RT-PCR analysis (Supplementary Table 1). The sensitivity of the mutated cells to SARS-CoV-2 infection was then addressed by quantification of the viral titer. Results show that disruption of many pro-viral candidates significantly impaired viral infection, while mutation of anti-viral genes markedly increased SARS-CoV-2 RNA levels (Fig. 1E). In addition, ablation of pro-viral candidates protected cells from SARS-CoV-2-induced cell-death while mutations in anti-viral candidates resulted in increased cell death compared to infected WT cells (Fig. 1F).

Since several top pro-viral candidate genes identified in the SARS-CoV-2 screens (e.g. GATA6, IRF6 and CUL5) did not emerge as essential for SARS-CoV-2 infection in previously published CRISPR screens carried in other cell types, we addressed the possibility that the observed phenomenon is related to their expression in different cell types. Indeed, inspection of the relative transcription levels of GATA6, IRF6 and CUL5 in published mRNA-seq datasets of Calu-3 [59], A549 [60] and Vero-E6 [61] cell lines, revealed that GATA6 and IRF6 are expressed at much higher levels in Calu-3 cells (IRF6 expression is below threshold in Vero-E6 and A549) (Fig 1G). CUL5 expression appeared to exhibit similar levels in various cell-lines, therefore it is conceivable that its detection in the current screen, may be attributed to a Calu-3 cell-specific modulation in the course of infection. 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 the understanding of SARS-CoV-2 entry. Finally, inspection of data from two recently documented Calu-3 infection screens [21, 62] revealed that multiple top pro-viral

candidates identified in our screens were enriched at commensurate extent also in these independent screens (Fig 1H).

Overall, corroboration of previously reported host-encoded viral-entry factors reported here, together with novel pro-viral host genes as well as the implied functional pathways defined by these factors, demonstrates the high potential of CRISPR-knockout screens to distinguish targets essential for viral pathogenicity which may serve for designing host dependent anti-viral therapy.

Functional pathways and interaction networks of SARS-CoV-2 host factors

Gene set enrichment analysis (GSEA) confirmed that anti-viral genes (depleted from the library) are strongly enriched for mitochondrial function, specifically mitochondrial translation and respiratory electron transport chain, and translation functions, including translation elongation and amino acid activation (Fig. 2A). Another large group of depleted genes were found to participate in the phosphatidylglycerol biosynthetic process. To examine whether these depleted host-pathways affect the virus specifically or exert an indirect effect on cell viability, GSEA was performed also on the data of the control viability screen conducted with uninfected cells. The analysis revealed that most of the pathways suspected as anti-viral are actually important for cell survival regardless of infection. The only pathway that was confirmed as specific anti-viral was the phosphatidylglycerol biosynthetic process (normalized enrichment score (NES) =-1.94, FDR =0.007 in the infection screen, NES =-1.35, FDR =0.526 in the cell viability screen) (Fig. 2A). Interestingly, phosphatidylglycerol was reported to regulate the innate immune response in lungs [63] and consequently was suggested as a treatment against SARS-CoV-2 infection by restoring the lung tissue [64]. No functional pathways were found to be significantly represented by analysis of enriched genes found as essential (pro-viral) for SARS-CoV-2 infection.

For further inspection of these 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 progression of the disease caused by different variants of SARS-CoV-2 indicate that these emerging variants may require both common and distinct host factors. Identification of 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 the design of future therapies. Accordingly, a comparative analysis of the datasets of the screens involving either one of the SARS-CoV-2 variants: 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 that the main mechanism for entry into the target cells is similar for the three SARS-CoV-2 variants and reiterating the robustness of the screens. Another host gene that was consistently highly enriched in the screens of cells infected by the three variants was the transcription factor GATA6. As of today, GATA6 role in SARS-CoV-2 infection was not reported and is 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.

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. 3B). In spite of the inherent variability of the experimental system, which may

complicate distinction of differential dependencies between screens in comprehensive manner, the top hits, representing genes encoding for functions strongly affected by the infection, exhibited high reproducibility. This is demonstrated by the correlation analysis of gene enrichment between biological replicates (Supplemental Fig. S1). Thus, since only genes strongly enriched or depleted are considered, the present report provides only a partial landscape of the host factors involved in the different VOCs infection. The pro-viral genes, ACE2, and GATA6 were consistently highly enriched in the screens of the cells infected with all variants, while the anti-viral MMP23B LARS2 HUS1 STX4 and RER1 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 (Fig 3B). Two such genes were KREMEN2 and SETDB1 that in the evaluation of the top five enriched genes of the Alpha variant screen (which exhibited the best correlation between replicates) revealed an opposite pattern in the screens involving the Beta variant and the WT SARS-CoV-2, suggesting that the product of these genes fulfill a distinct role during the life cycle of the Alpha variant. (Fig. 3C). Notably, the proteins KREMEN1 and KREMEN2 have been identified as co-receptors for Dickkopf (Dkk) proteins, hallmark secreted antagonists of the canonical Wnt signaling pathway [65, 66]. In addition, KREMEN1 serves as host-cell entry receptor for a major group of Enteroviruses [67]. Remarkably, in a recent screen in search of additional SARS-CoV-2 receptors, KREMEN1 was identified as an alternative viral-entry receptor [68]. Thus, the observation documented here suggests that KREMEN2 may play a similar role specifically during the life cycle of the Alpha variant, consequently contributing to its enhanced infectivity compared to other SARS-CoV-2 variants.

SETDB1 (SET domain bifurcated histone lysine methyltransferase 1), is a histone H3K9 methyltransferase that contributes together with TRIM28 to heterochromatin formation [69]. Recent studies have drawn attention to the involvement of SETDB1 in epigenetic control of the immune response, including antiviral response and IFN production [70-72]. Interestingly, the transcriptional levels of IFN-I, IFN-II, TRIM28, and SETDB1 were found to be elevated in SARS-CoV-2 infected children with mild

symptoms, while their levels decreased in children with severe clinical pictures or MIS-C suggesting that they may play important roles in conditioning the evolution of the infection [73]. SETDB1 was also identified in a HCoV-NL63 genetic screen as a possible host chromatin regulator that promote successful infection [22].

To validate KREMEN2 and SETDB1 specificity for the Alpha variant sgRNAs targeting either KREMEN2 or SETDB1 were expressed in Calu-3 cells that were subsequently infected by WT-SARS-CoV-2 or by either one of the VOCs. In both KREMEN2 and SETDB1 KO cell lines, a significantly impaired viral infection of the Alpha VOC was measured. However, disruption of either one of the genes did not result in a considerate change in viral titer by infection with the WT-SARS-CoV-2 or the Beta variant (Fig. 3D). While the data presented in this report clearly suggests that the KREMEN2 and SETDB1 genes encode functions specifically assisting the Alpha variant infection, additional studies will be required to understand their mechanism of function and strain specificity, and most importantly to determine if their role relates to possible enhanced viral infectivity or pathogenicity.

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 SARS-CoV-2 infection

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 [74, 75]. Both GATA6 and SFTPA genes were upregulated in SARS-CoV-2-infected lungs, 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 manipulations indeed resulted in depletion of GATA6 and ACE2 expression as determined both by Real-Time PCR and Western blot analysis (Fig. 5B, 5C). As expected, while WT Calu-3 cells display a cytopathic effect following WT-SARS-CoV-2 infection, ACE2 gene-disruption protected cells from the infection (Fig. 4A). In agreement with the results of the CRISPR screens, GATA6-disrupted cells were almost fully protected from cell death following WT-SARS-CoV-2 infection (Fig. 4A). 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 WT-SARS-CoV-2 viral load in infected GATA6-disrupted cells 24 and 48 hours post infection, demonstrated a significant reduction of viral replication subsequent to reduction in GATA6 content (Fig. 4B). Further inspection of the effect of GATA6 disruption on the infection by the Alpha and Beta VOCs resulted with an increased resilience to cell death of GATA6 KO cells as seen by microscopy imaging and crystal violet staining (Fig 4C, 4D), and impaired viral infection (Fig 4E). The effect of GATA6 gene disruption was also interrogated on the infectivity of the novel Delta (B.1.617.2) VOC (which emerged in the course of preparing the current report). The data in Fig. 4D and 4E establish that the Delta variant exhibited impaired infectivity in the GATA6 KO cells confirming the strain-independent essentiality of GATA6 for SARS-CoV-2 infection.

As mentioned, previously published mRNA-seq transcriptomic datasets of Calu-3 [59], A549 [60] and Vero-E6 [61] cell lines, revealed that GATA6 is expressed at much higher levels in Calu-3 cells (Figure 1G), suggesting that GATA6 was not identified in screens carried out in other cell-lines due to cell-type specificity. To further substantiate this hypothesis, GATA6 was disrupted in Vero-E6 cells (Supplemental Fig. S2), and KO cells were subjected to infection of WT-SARS-CoV-2. Results demonstrated similar levels of infection between WT Vero-E6 cells

and GATA6 disrupted cells (Fig. 4F). Cell viability following infection was also similar between the WT and GATA6 KO Vero-E6 cells (Fig 4G). It is therefore conceivable that the function fulfilled by GATA6 in Calu-3 cells differs from that in Vero-E6, in line with the fact that GATA6 is a regulatory transcriptional factor which exerts its function by modulating the transcriptional level of various target genes which may not be necessarily the same 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 4H). Of note, the analysis did not reveal statistically relevant differences in ACE2 expression between COVID-19 and control healthy individuals (Fig 4H). 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 different from that of ACE2 receptor. Upregulation of GATA6 was also observed in a recent study in which the host transcriptome from nasopharyngeal-swab samples of four COVID-19 patients was profiled [76]. In line with the above *in-vivo* results, the relative levels of GATA6 transcription were found also to be elevated during SARA-CoV-2 infection of Calu-3 cells (Fig 4I).

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

To further elucidate which viral life-cycle step is facilitated by GATA6 expression, SARS-CoV-2 infection was visualized by immunofluorescent labeling of the virus in infected GATA6-disrupted cells in comparison to WT cells (Fig. 5A). The analysis enabled detection of SARS-CoV-2 in the cytoplasm of WT cells as soon as 5 hours post infection with increased viral load at 24 hours post infection. In contrast, a significant reduction of the infection of SARS-CoV-2 in the GATA6-disrupted cells was observed suggesting that GATA6 facilitates an early step in the viral life cycle possibly related to the entry of the virus into the cells (Fig. 5A).

Since SARS-CoV-2 entry into cells is primarily mediated by ACE2, the expression of ACE2 in ACE2 disrupted, WT and GATA6 disrupted cells was determined by Western-blot analysis. Notably, the decrease in GATA6 was followed by a significant reduction in ACE2 expression compared to control cells (Fig. 5B). Real-Time PCR analysis confirmed the Western-blot results indicating that this effect involves a regulatory mechanism exerted at the transcription-level, suggesting that GATA6 acts as a transcriptional factor affecting ACE2 expression (Fig. 5C). The data were validated by trans-complementation experiments which established that extrachromosomal cDNA-mediated expression of GATA6 (N-terminally triple FLAG-tagged and C-terminally triple AU1-tagged full length human wt GATA6) in GATA6 KO cells restored elevated ACE2 expression, as measured by Western Blot analysis (Fig. 5B). Furthermore, complementation of GATA6 expression by exogenous GATA6 cDNA restored susceptibility of the cells to SARS-CoV-2 infection as measured by the virus titers (Fig. 5D) and crystal violet staining (Fig. 5E). Based on the data it is possible therefore to assume that the low ACE2 levels associated with abrogation of GATA6 expression are a manifestation of the role of GATA6 in viral entry. However, since GATA6 is a known pleiotropic transcription factor, its abrogation could theoretically affect transcription of additional genes that may have a role in SARS-CoV-2 infection. In an attempt to further confirm that GATA6 is related to viral entry to the cell through regulation of ACE2 expression, GATA6 KO cells over-expressing ACE2 were infected with SARS-CoV-2. The data demonstrated that the expression of ACE2 in GATA6 KO cells restored SARS-CoV-2 titers suggesting that the main mechanism through which GATA6 influences SARS-CoV-2 infection is through modulation of ACE2 expression levels (Fig. 5D). In an attempt to further determine the 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 showed that ACE2 expression decreases as early as 12 hours post GATA6-siRNA transfection (Fig. 5F).

As a transcription factor GATA6 recognizes a well -defined DNA-binding specific consensus site (T/AGATAA/G or reverse complement sequences C/TTATCT/A)

[77]. 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 (upstream of the initiation ATG codon) (Figure 5G) [78]. To investigate the possible binding of GATA6 to ACE2 promoter an Electrophoretic Mobility Shift Assays (EMSA) was implemented. Incubation of lysates from HEK293T cells over expressing GATA6 with labeled oligonucleotides derived from the promoter region of ACE2 resulted in the formation of electrophoretic slow-migrating protein-DNA complexes, which were not generated in the presence of competing unlabeled oligonucleotides (Fig. 5H). This observation supports the notion that GATA6 may induce the expression of ACE2 via direct binding at a sequence-specific GATA6 consensus site.

The connection between GATA6 and ACE2 expression, together with the observations that GATA6 transcription levels were up-regulated in SARS-CoV-2 infected cells (Fig.4I) and SARS-CoV-2 positive clinical samples (Fig. 4H) raised the possibility that over-expression of GATA6 might increase cell susceptibility to viral-infection. To address this possibility, Calu-3 cells over-expressing GATA6 were subjected to SARS-CoV-2 infection. Western blot analysis demonstrated that the elevated levels of GATA6 expression were accompanied by increased levels of ACE2 (Supplemental Fig. S3A). Yet, the over-expression of GATA6 did not affect the RNA levels of SARS-CoV-2 measured 48 hours post infection (Supplemental Fig. S3B).

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 exhibiting specific roles in the viral pathogenicity may serve as appropriate candidates for antiviral therapy with minimal host side effects. 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 [79], 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 binding to ACE2 as shown by Electrophoretic Mobility Shift Assays (Fig 6A, note also the increase in the amount of the free unbound DNA probe upon competition with unlabeled sequences, which confirms the sequence specificity of the binding). Notably, Calu-3 cells retained their viability even at high doses of Pyrrothiogatain (Fig.6B). Inhibition of GATA6 resulted in lower ACE2 expression (Fig. 6C) supporting the causal connection between GATA6 and ACE2. A decrease in the amount of GATA6 protein was also observed (Fig. 6C) suggesting that Pyrrothiogatain may induce a feed-back circuit lowering GATA6 and/or a conformational change affecting its expression. Most importantly, treatment of infected Calu-3 cells with Pyrrothiogatain resulted in a significant decrease in the viral load of the culture (Fig. 6D). Immunofluorescence analysis further evidenced cytoplasmic SARS-CoV-2 staining in untreated cells while significantly reduced infection was observed in Pyrrothiogatain-treated cells (Fig. 5E). The data cannot rule out the possibility that other members of the GATA family are affected by Pyrrothiogatain, yet other previous transcriptomic surveys did not evidence the upregulation of other GATAs neither did our screen distinguish the enrichment of other GATAs upon infection (Table 1). Furthermore, with the exception of GATA6, none of the other 5 members of the GATA family revealed significant levels of transcription in the Vero-E6, A549 and Calu-3 cell lines (Fig. 6F).

The data strongly supports the possibility that novel therapeutic strategies to countermeasure viral infection via down-modulation of ACE2 may employ specific GATA6 targeting. GATA6 targeting may seem a rather unlikely therapeutic approach considering the fact that it is a pleiotropic transcription factor affecting many loci. Yet, recently, the involvement of transcription factors in many pathologies prompted preclinical and clinical assessment of their therapeutic value, including studies addressing proteins belonging to the GATA family [80-82]. The possibility to target transcription factors for selective therapeutic intervention, is facilitated by high combinatorial interplay and compensations of the transcription factors which lowers the side effects associated with targeting one particular factor.

Since transcription factors (including GATA6) may play different roles in different cell types (for example, GATA6 acts a tumor suppressor in astrocytoma while it is overexpressed in human colon cancer and pancreatic carcinoma), some of these studies, including those focusing on GATA family members, mitigate the undesired effects by combining specificities in gene and tissue targeting.

In summary, the screens conducted in human lung epithelial cells, provided 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, the comparative analysis highlights commonalities and differences between SARS-CoV-2 variants and enables the identification of the receptor KREMEN2 and SETDB1 as possible unique genes 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.

Finally, the study evidenced the 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, a 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. These observations together with the loss-of-function screen data reported here contribute to the better understanding of SARS-CoV-2 pathogenesis and may represent an important basis for the 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

The original 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) and B.1.617.2 (Delta) variants of concern (VOCs) were kindly provided by Michal Mandelboim, MOH Central Virology Lab, Tel Hashomer, Israel. Original, B.1.351 and B.1.617.2 VOC stocks were propagated (2-3 passages) on Vero-E6 cells while B.1.1.7 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. For a control reference for sgRNA enrichment analysis cells were harvested 48 hours after seeding. For viability screen, mock cells were harvested 7 days after seeding and served as a control for the depletion of genes that are essential for cell survival. 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 [83]. 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 S2. 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 [84] 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. All genes and their enrichment scores can be found in Table 1.

Pathway and network analysis

Directed scores calculated from MAGeCK RRA scores of all screens combined (treating all screens as one experiment) were used as input for gene set enrichment analysis (GSEA version 4.1) with GO biological process (c5.bp) from MSigDB (version 7.4) [85, 86]. 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 [87] using default parameters. The STRING network was imported into Cytoscape [88]. 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 Vero-E6 and Calu-3 KO cell lines.

DNA oligos (Integrated DNA Technologies, Inc.) containing sgRNA sequences (see Supplementary Table 2) 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 or Vero-E6 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, Real-Time PCR or sequencing. For GATA6 complementation experiment, GATA6 knockout 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. Seventy-two hours post transfection, the expression of GATA6 was analyzed by western blot.

Infection of Vero-E6 and Calu-3 KO cell lines

Cells were seeded 48 hours prior to infection at a density of $4*10^5$ in 12 well plates and infected as described above at a MOI of 0.002. Plates were incubated for 1 hour at 37 °C to allow viral adsorption. The medium was then aspirated and replaced with fresh medium. At that point (t=0 hour), 24 hpi and 48 hpi cell medium was taken for viral titration.

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.

Imaging and Immunofluorescence assay

Calu-3 knockout or WT cells were seeded at a density of $4*10^5$ in 12 well plates, infected as described above or left uninfected. Imaging was performed on a Nikon eclipse TS100 microscope using a WD 7.0 10X/0.25 Ph1 DL objective and DS-Ri1 camera. For Immunofluorescence experiments, Calu-3 knockout or WT cells were seeded at a density of $5*10^4$ cells per well in 8-well-µ-slides (Ibidi), infected as described above or left uninfected. 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.

Cell viability assays

The effect of Pyrrothiogatain on Calu-3 cell proliferation was determined using the XTT assay (Cell Proliferation Kit, 20-300-1000, Biological Industries, Bet Haemek, Israel) as per the manufacturer's instructions. Briefly, 10^5 Calu-3 WT cells were seeded in 96-well plate. Twenty-four hours later, the cells were incubated with Pyrrothiogatain at different concentrations (150, 300, 500 µM) for additional 48 hours. Then, XTT reagent was added for 15 minutes before reading the change in absorbance at 650 and 475 nm using the SpectraMax 250 plate reader (Molecular devices). Specific absorbance measurements were given as the mean ± SD absorbance calculated from 2 repeat wells/sample. The mean specific absorbance was normalized at each time point to that of the non-treated control.

For crystal violet staining, following 48 hours of SARS-CoV-2 infection the medium was removed from the cells that were then fixed and stained with 0.1% crystal violet solution (Biological Industries, Israel) for 5 minutes. Then stain was aspirated and plates were rinsed once with tap water and air dried.

Analysis of GATA6, CUL5 and IRF6 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 the published RPKM values 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. No information is available on the

level of symptoms manifested by each tested positive individual. 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 S2) 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 S2) 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 are listed in Supplementary Table 2.

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, ab22600), 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 S2, 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 $MgCl_2$. In competition assays, unlabeled competitor probes were included in the binding reactions in 200-fold excess relative to labeled probes. For inhibition of the binding of GATA6 to the promotor region of ACE2 Pyrrothiogatain was included in the binding reaction in 50,100 or 500 uM. 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.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Acknowledge

We thank Emanuelle Mamroud and Shahar Rotem for technical advice and helpful discussion; Shay Weiss for biosafety guidance and Tamar Aminov and Inbar Chomsky for technical assistance.

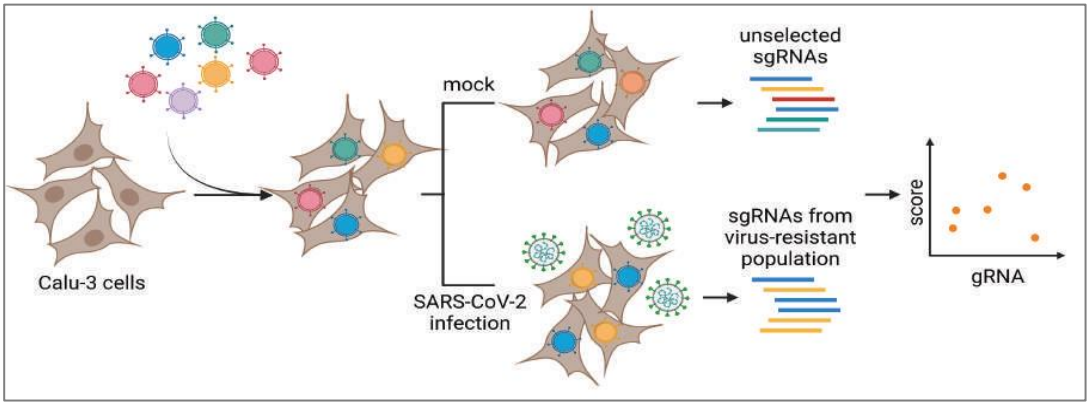
References

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.

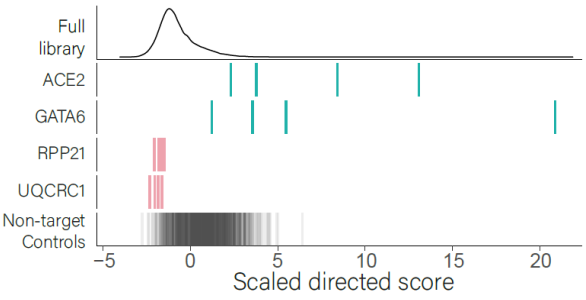
36. Huang, B., et al., *Serum sample neutralisation of BBIBP-CorV and ZF2001 vaccines to SARS-CoV-2 501Y.V2.* Lancet Microbe, 2021. **2**(7): p. e285.
37. Madhi, S.A., et al., *Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant.* N Engl J Med, 2021. **384**(20): p. 1885-1898.
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. Hart, T., et al., *High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities.* Cell, 2015. **163**(6): p. 1515-26.
58. Rosa, B.A., et al., *IFN signaling and neutrophil degranulation transcriptional signatures are induced during SARS-CoV-2 infection.* Communications Biology, 2021. **4**(1): p. 290.
59. Finkel, Y., et al., *SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis.* Nature, 2021. **594**(7862): p. 240-245.
60. Bercovich-Kinori, A., et al., *A systematic view on influenza induced host shutoff.* Elife, 2016. **5**.
61. Finkel, Y., et al., *The coding capacity of SARS-CoV-2.* Nature, 2021. **589**(7840): p. 125-130.
62. Rebendenne, A., et al., *Bidirectional genome-wide CRISPR screens reveal host factors regulating SARS-CoV-2, MERS-CoV and seasonal coronaviruses.* bioRxiv, 2021: p. 2021.05.19.444823.
63. Voelker, D.R. and M. Numata, *Phospholipid regulation of innate immunity and respiratory viral infection.* J Biol Chem, 2019. **294**(12): p. 4282-4289.
64. Bollag, W.B. and J.N. Gonzales, *Phosphatidylglycerol and surfactant: A potential treatment for COVID-19?* Med Hypotheses, 2020. **144**: p. 110277.
65. Mao, B., et al., *Kremen proteins are Dickkopf receptors that regulate Wnt/beta-catenin signalling.* Nature, 2002. **417**(6889): p. 664-7.
66. 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.
67. 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.
68. Gu, Y., et al., *Interaction network of SARS-CoV-2 with host receptome through spike protein.* bioRxiv, 2020.
69. Schultz, D.C., et al., *SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins.* (0890-9369 (Print)).
70. Jiang, Y., et al., *Epigenetic activation during T helper 17 cell differentiation is mediated by Tripartite motif containing 28.* (2041-1723 (Electronic)).
71. Kamitani, S., et al., *KAP1 regulates type I interferon/STAT1-mediated IRF-1 gene expression.* (1090-2104 (Electronic)).

72. Krischuns, T., et al., *Phosphorylation of TRIM28 Enhances the Expression of IFN-β and Proinflammatory Cytokines During HPAIV Infection of Human Lung Epithelial Cells.* (1664-3224 (Electronic)).
73. Tovo, P.A., et al., *COVID-19 in Children: Expressions of Type I/II/III Interferons, TRIM28, SETDB1, and Endogenous Retroviruses in Mild and Severe Cases. LID - 10.3390/ijms22147481 [doi] LID - 7481.* (1422-0067 (Electronic)).
74. Bruno, M.D., et al., *GATA-6 activates transcription of surfactant protein A.* J Biol Chem, 2000. **275**(2): p. 1043-9.
75. 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.
76. Islam, A., et al., *Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2.* (1479-5876 (Electronic)).
77. Chou, C.F., et al., *ACE2 orthologues in non-mammalian vertebrates (Danio, Gallus, Fugu, Tetraodon and Xenopus).* Gene, 2006. **377**: p. 46-55.
78. 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.
79. 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.
80. Bushweller, J.A.-O., *Targeting transcription factors in cancer - from undruggable to reality.* (1474-1768 (Electronic)).
81. Henley, M.A.-O. and A.A.-O. Koehler, *Advances in targeting 'undruggable' transcription factors with small molecules.* (1474-1784 (Electronic)).
82. Huilgol, D.A.-O., et al., *Transcription Factors That Govern Development and Disease: An Achilles Heel in Cancer. LID - 10.3390/genes10100794 [doi] LID - 794.* (2073-4425 (Electronic)).
83. 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.
84. 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.
85. 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.
86. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
87. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-d613.
88. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.

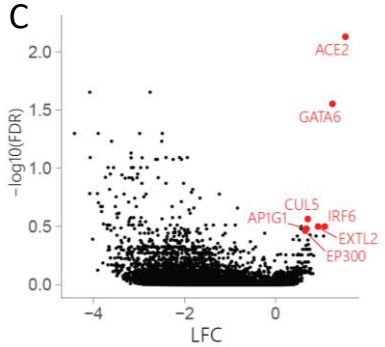
A


Calu-3 cells
mock
SARS-CoV-2 infection
unselected sgRNAs
sgRNAs from virus-resistant population
score
gRNA

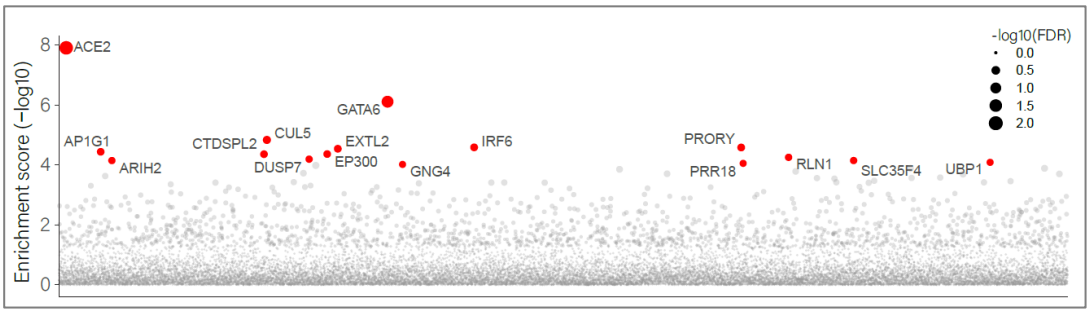
B


Full library
ACE2
GATA6
RPP21
UQCRC1
Non-target Controls
Scaled directed score

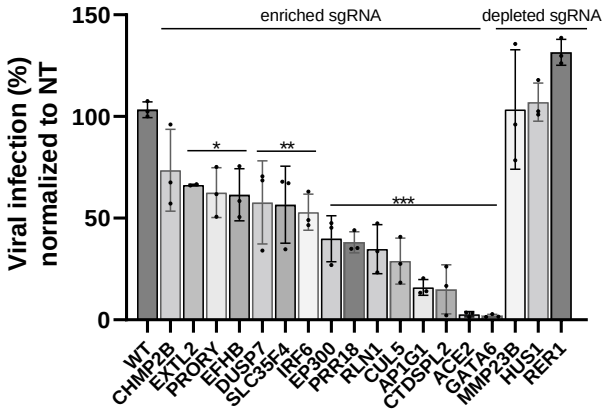
C


ACE2
GATA6
CUL5
AP1G1
IRF6
EXTL2
EP300
−log10(FDR)
LFC

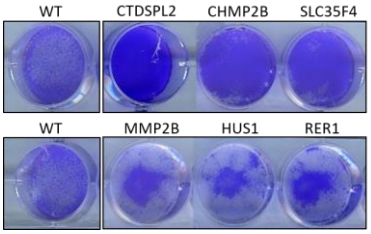
D


ACE2
GATA6
AP1G1
ARIH2
CTDSPL2
CUL5
DUSP7
EXTL2
EP300
GNG4
IRF6
PRORY
PRR18
RLN1
SLC35F4
UBP1
-log10(FDR)
Enrichment score (−log10)

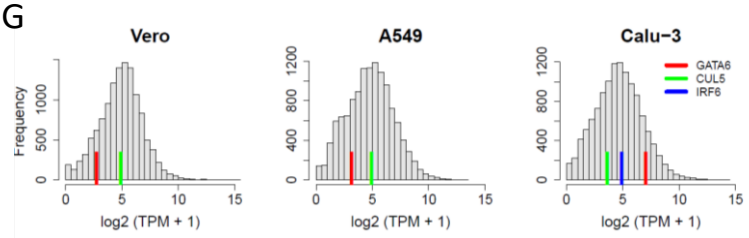
E


enriched sgRNA
depleted sgRNA
Viral infection (%) normalized to NT
WT
CHMP2B
EXTL2
PRORY
EFHB
DUSP7
SLC35F4
IRF6
EP300
PRR18
RLN1
CUL5
AP1G1
CTDSPL2
ACE2
GATA6
MMP23B
HUS1
RER1

F


WT
CTDSPL2
CHMP2B
SLC35F4
WT
MMP2B
HUS1
RER1

G


Vero
A549
Calu-3
GATA6
CUL5
IRF6
Frequency
log2 (TPM + 1)

H

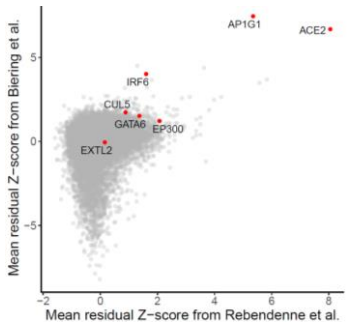

AP1G1
ACE2
IRF6
CUL5
GATA6
EP300
EXTL2
Mean residual Z−score from Biering et al.
Mean residual Z−score from Rebendenne et al.

Figure 1. CRISPR genome-wide screens in human Calu-3 cells identify host factors important for infection by wild-type SARS-CoV-2 and variants of concern. (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 screens of example sgRNAs. The density across the screens is plotted for the full library (top) and for the 4 sgRNAs targeting top resistance hits, top sensitization hits and the non-targeting control sgRNA. **(C)** Volcano plot showing top genes conferring resistance and sensitivity to SARS-CoV-2. Top pro-viral candidates identified in our screens are displayed. **(D)** Gene enrichment score for CRISPR screens of WT-SARS-CoV-2 and VOCs infection. Enrichment scores were determined by MaGECK. **(E)** real-time PCR quantification of WT-SARS-CoV-2 levels in wild-type and CRISPR-edited Calu-3 cells. A non-targeting (NT) sgRNA was used as control. Cells were infected using MOI=0.002 for 48 hours. Data were analyzed by one-way ANOVA with Tukey's multiple comparison test. Shown are means ± SD. *$p < 0.05$; **$p < 0.005$; ***$p < 0.0001$. **(F)** Cell viability assay of wild-type and individual CRISPR-edited Calu-3 cells infected with WT-SARS-CoV-2 at MOI of 0.002 for 48 hours and stained with crystal violet. One of three repetitions is shown. **(G)** Histograms of mRNA levels in Calu-3, Vero-E6 and A549 cells [66-68], normalized to transcripts per million (TPM). Red, green and blue lines represent the levels of GATA6, CUL5 and IRF6 respectively in each dataset. IRF6 is below threshold in A549 and Vero-E6 cells. **(H)** Comparative analysis of previously reported SARS-CoV-2 knockout screens performed in the Calu-3 cell-line [21, 62]. Top pro-viral candidates identified in our screens are displayed.

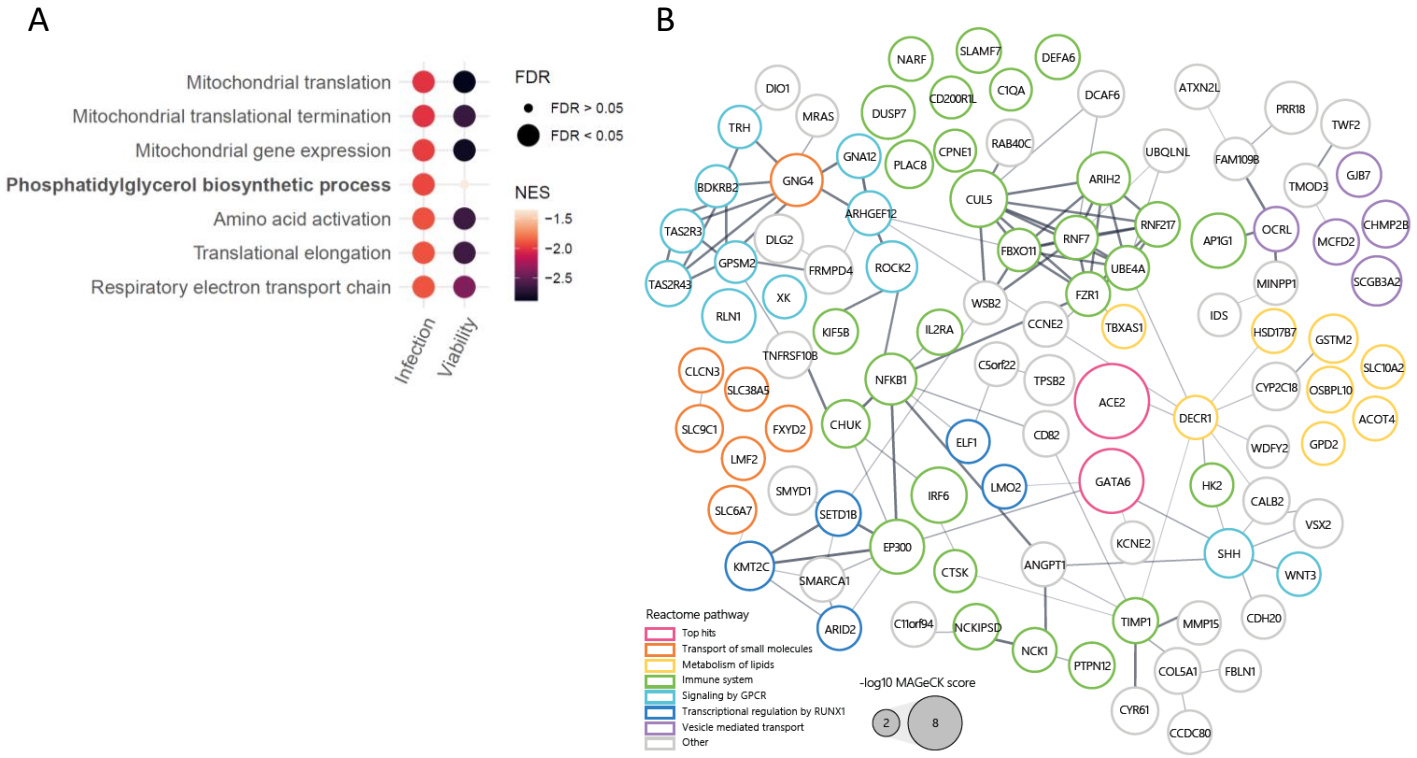


Figure 2. Gene set enrichment and interaction networks analysis identify functional pathways important for wild-type SARS-CoV-2 and variants of concern infection. (A) Gene set enrichment analysis (GSEA) was performed by analysis of the directed scores of all genes from the WT-SARS-CoV-2 and VOCs screens combined (infection), and for the uninfected cell viability screen (viability). Results from the infection screens were filtered to get significantly enriched pathways with false discovery rate (FDR) of less than 0.01. The enrichment of each of these pathways in each screen is represented as bubbles colored by Normalized enrichment score (NES) and sized by significance levels. **(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

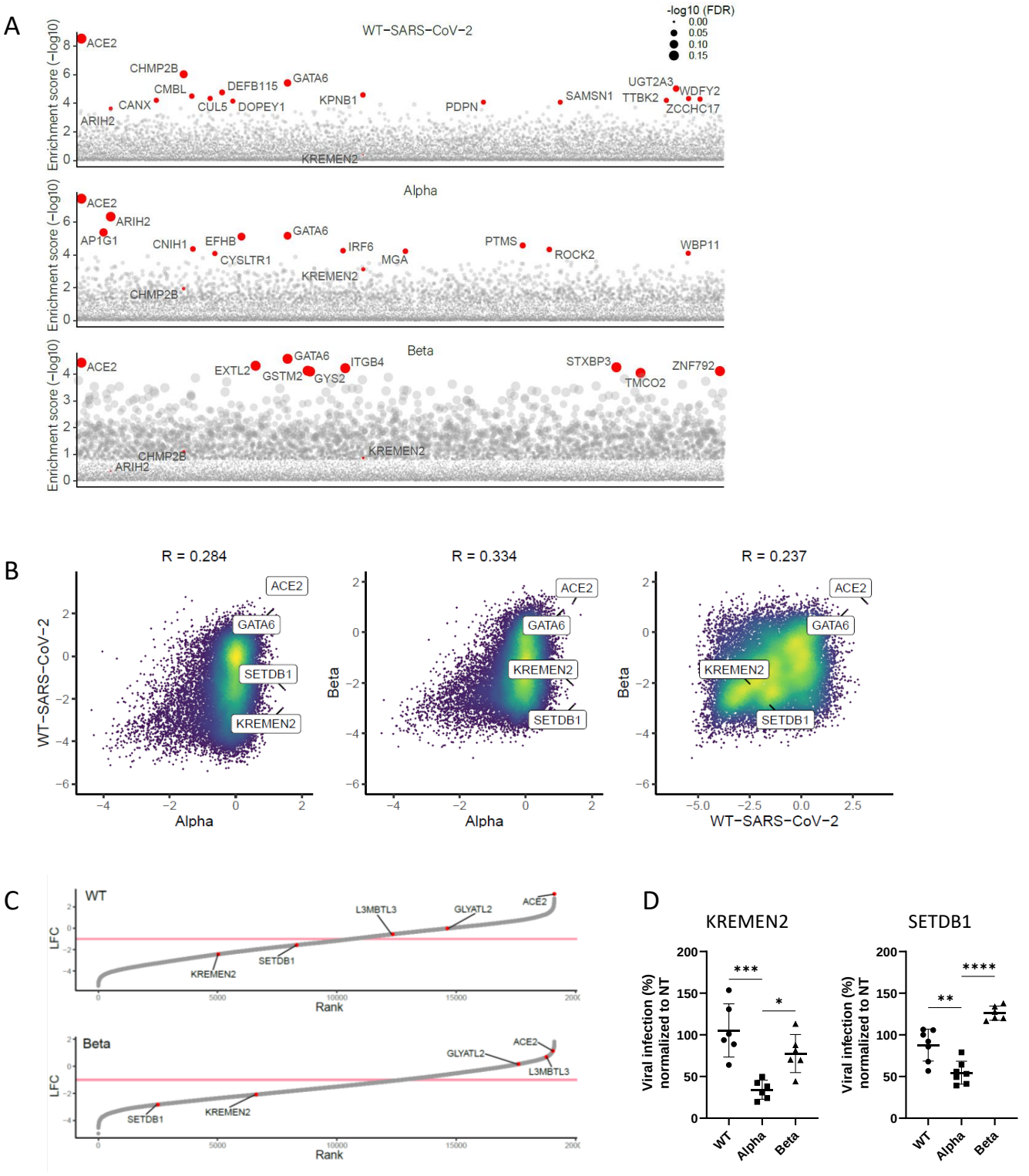


Figure 3. Differential analysis of variants of SARS-CoV-2 screens highlights shared and specific host-factors dependencies. (A) Gene enrichment score for CRISPR screens of WT-SARS-CoV-2, Alpha and Beta variants infection. Enrichment scores were determined by MaGECK. **(B)** Scatter plot depicting gene enrichment in WT-SARS-CoV-2, Alpha and Beta variants screens. R, Pearson correlation. **(C)** Rank plot showing the log fold-change (LFC) of genes in the WT-SARS-CoV-2 screen (top) or the Beta variant screen (bottom) and its rank according to their LFC (lowest LFC ranked 1). Each gene is shown as a grey point. The top five most enriched genes in the Alpha variant screen (based on fold-change) are labeled and shown in red. A pink line marks LFC of -1. **(D)** real-time PCR quantification of WT-SARS-CoV-2, Alpha and Beta VOCs levels in infected KREMEN2 and SETDB1 knockout Calu-3 cells. Cells were infected using MOI=0.002 for 48 hours. A non-targeting (NT) sgRNA was used as control. Data were analyzed by one-way ANOVA with Dunnett's multiple comparisons test, comparing each experimental condition with Alpha variant. Shown are means ± SD. *p < 0.05; **p < 0.001; ***p < 0.0005; ****p < 0.0001.

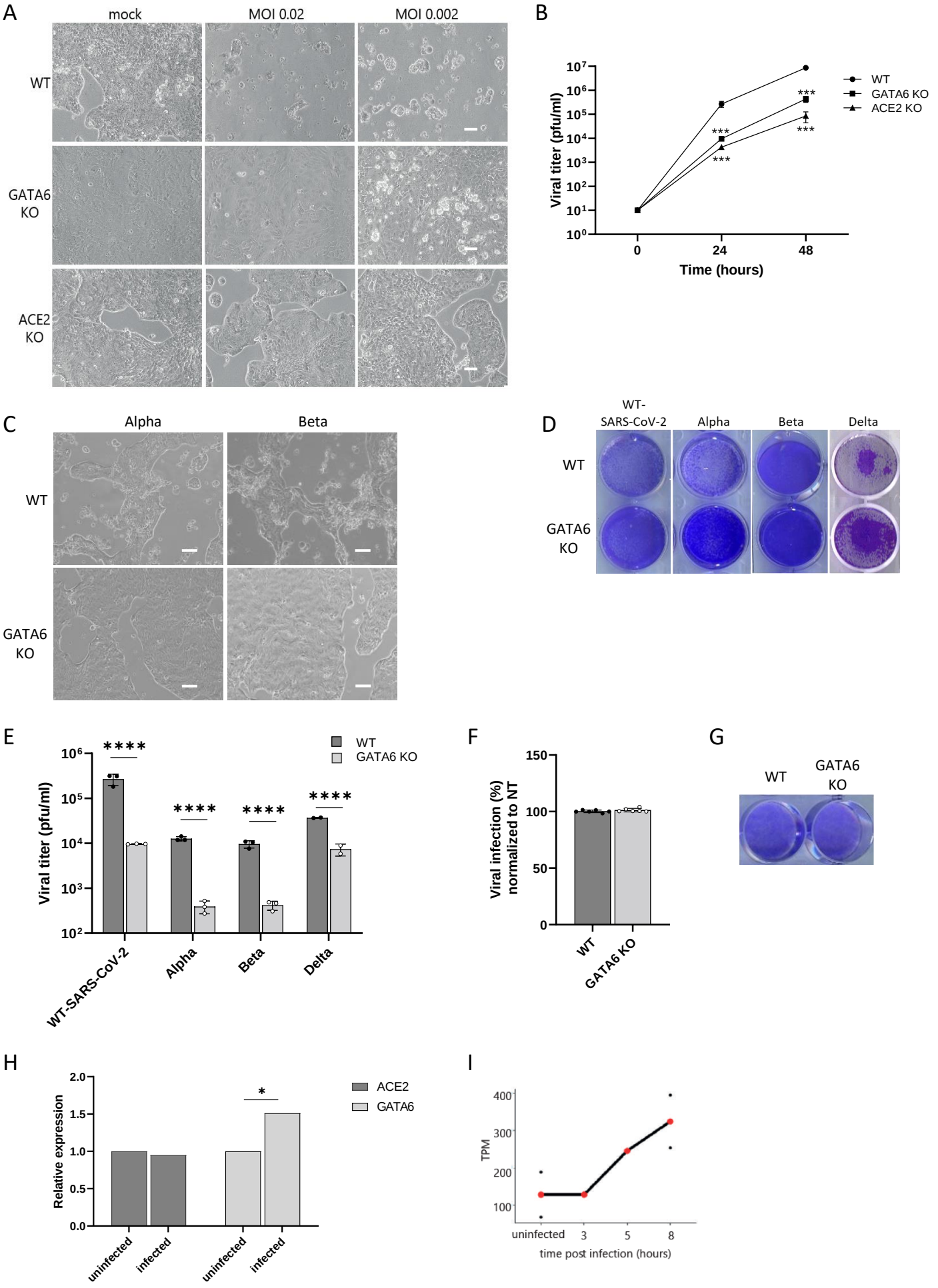
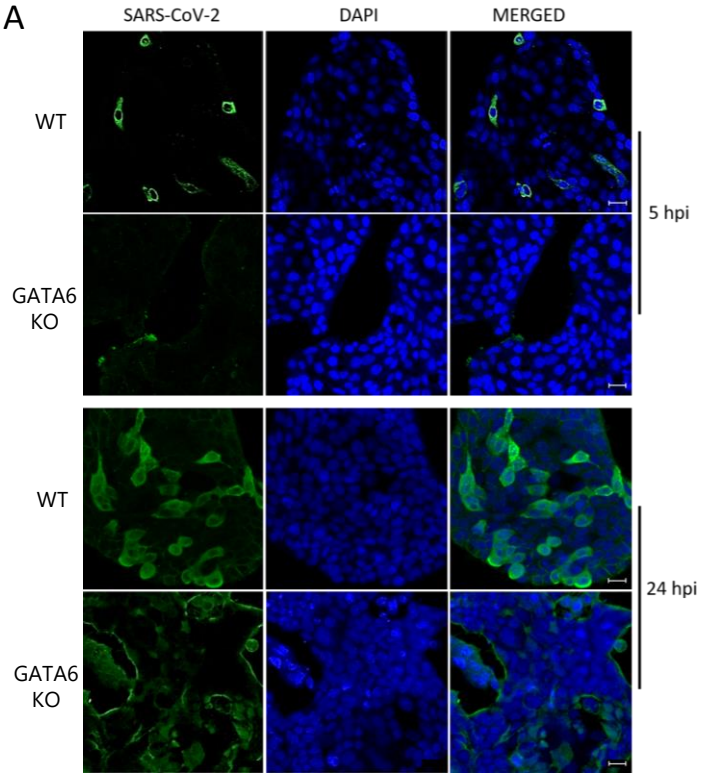
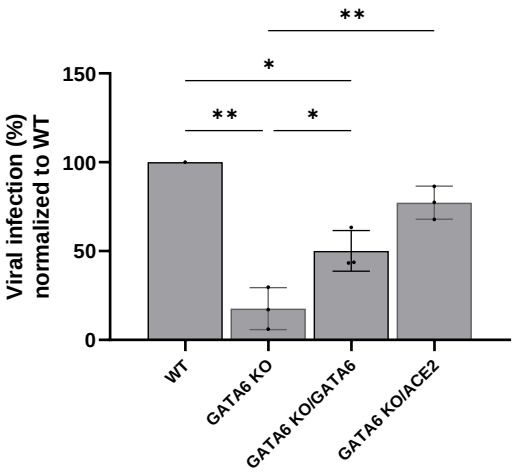

A
mock
MOI 0.02
MOI 0.002
WT
GATA6 KO
ACE2 KO
B
Viral titer (pfu/ml)
Time (hours)
WT
GATA6 KO
ACE2 KO
***
C
Alpha
Beta
WT
GATA6 KO
D
WT-SARS-CoV-2
Alpha
Beta
Delta
WT
GATA6 KO
E
Viral titer (pfu/ml)
****
WT
GATA6 KO
WT-SARS-CoV-2
Alpha
Beta
Delta
F
Viral infection (%) normalized to NT
WT
GATA6 KO
G
WT
GATA6 KO
H
Relative expression
ACE2
GATA6
*
uninfected
infected
I
TPM
uninfected
3
5
8
time post infection (hours)

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. Scale bar 100um. 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)** Control and GATA6 disrupted Calu-3 cells were infected with the Alpha or the Beta VOCs at MOI=0.002. Cell viability was visualized by light microscopy 48 hours post infection. Scale bar 100um. Representative images of three independent replicates are shown. **(D)** Cell viability assay of Control and GATA6 disrupted Calu-3 cells infected with the indicated SARS-CoV-2 variant at MOI of 0.002 for 48 hours, and stained with crystal violet. One of three repetitions is shown. **(E)** real-time PCR quantification of WT-SARS-CoV-2, Alpha, Beta and Delta VOCs levels in infected control and GATA6 disrupted Calu-3 cells. Cells were infected using MOI=0.002 for 48 hours. Data were analyzed by one-way ANOVA with Šídák's multiple comparisons test. Shown are means ± SD. ****$p < 0.0001$. **(F)** real-time PCR quantification of control and GATA6 disrupted Vero-E6 cells infected with WT-SARS-CoV-2 at MOI of 0.002 for 48 hours. Shown are means ± SD. Data were analyzed by two-tailed student's t-test. **(G)** Cell viability assay of Control and GATA6 disrupted Vero-E6 cells infected with WT-SARS-CoV-2 at MOI of 0.002 for 48 hours, and stained with crystal violet. One of three repetitions is shown. **(H)** real-time PCR measurements of ACE2 and GATA6 mRNA levels relative to GAPDH amount in negative (uninfected) and positive (infected) nasopharyngeal swab samples from symptomatic and asymptomatic individuals Data were analyzed by two-tailed student's t-test, *$p<0.05$. **(I)** Expression levels of GATA6 at different times along SARS-CoV-2 infection in Calu-3 cells [67] normalized to transcripts per million (TPM) values. Values from single replicates shown for 03hr and 05hr. Mean values of duplicates shown for uninfected and 08hr in red, and individual replicate values are presented as black points.

A


SARS-CoV-2
DAPI
MERGED
WT
GATA6 KO
5 hpi
WT
GATA6 KO
24 hpi

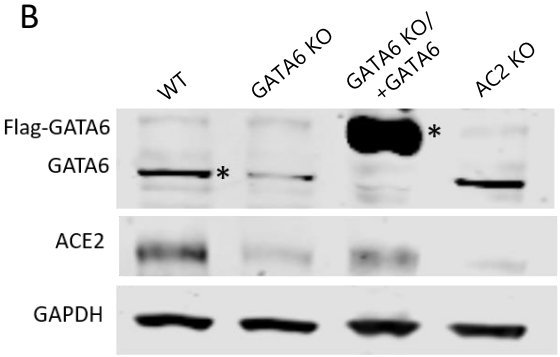
D


Viral infection (%) normalized to WT
150
100
50
0
**
*
**
*
WT
GATA6 KO
GATA6 KO/GATA6
GATA6 KO/ACE2

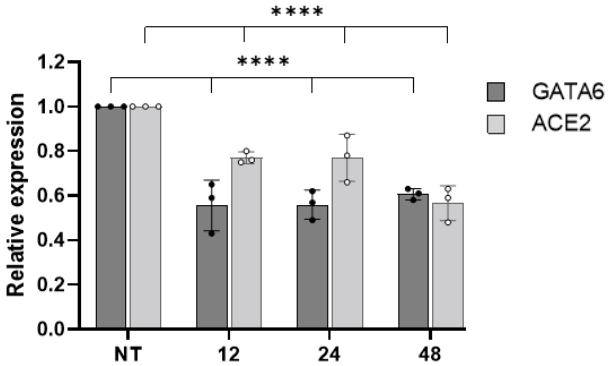
E

WT
GATA6 KO
GATA6 KO/ +GATA6
GATA6 KO/ +ACE2

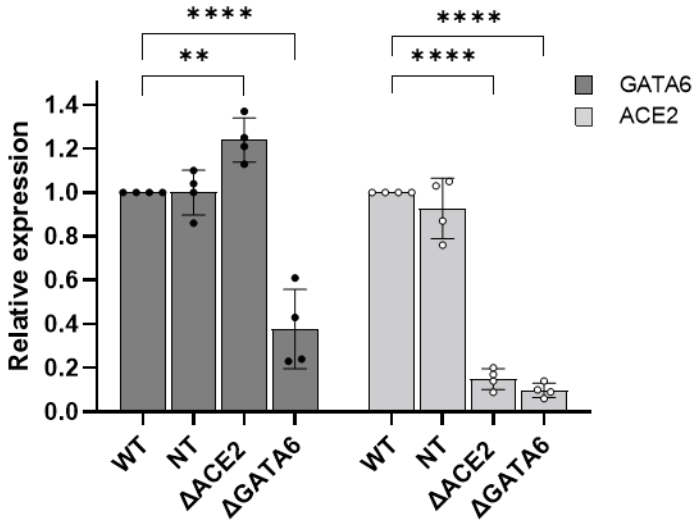
B


WT
GATA6 KO
GATA6 KO/ +GATA6
AC2 KO
Flag-GATA6
GATA6
ACE2
GAPDH
*
*

F


****
****
Relative expression
1.2
1.0
0.8
0.6
0.4
0.2
0.0
NT
12
24
48
GATA6
ACE2

C


****
**
****
****
Relative expression
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
WT
NT
ΔACE2
ΔGATA6
WT
NT
ΔACE2
ΔGATA6
GATA6
ACE2

G

GGGCAGATCACAAGGTCAGGAGATAGAGACCGTCCTGACTAACACGGTGAAACCCCGTCTCTATTAAAAA -751
TACAAAAACTTAGCTGGGCGTGGTGGTGGGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA -681
ATGGCGTGAACCTGGGAGGCAGAGCTTGCAGTGAGCCGAGATCGCGCCACTGTGCTCCAGCCTGGGCGAC -611
AGACCGAGACTCAGTCTCAAAAAAAAAAAAAAAAAAAAAACTGATCAGAAATATGAATTCCATAAAGACA -541
AGGGCAAAGTCATGTATTTGGAAGGGAAAATGTTGCCCAAGTAGAGAGTTTCTGGGAATATGATCTTGAA -471
ATAAAAATAAATGTGAGATAACCTATTAATGAAATTGTCTGAAAACCATACAAACACCAACATTATCTTC -401
ATGATCCCTAGTTCTAGACCTGTTTGGTCACTGTAAAATTATAACATTTTCCGTGTATCTTTAACAGCTT -331
TCTAGGAAAATATTAACCAAAAGTACCGGTTTTGATTTGGCCATAAAGTGACAGGAGAGGTAAGGTTCTC -261
TAGGATTAAAGAATAACGTATTCTTATTTGATTCACTTTAAAAAATTATTCTAAAATCTGTTACATATCT -191
GTCCTCTCCAGGATGAACTTTATATTGGCTCAGCAGATTGTTTACTGTGTTCTTCTTCTTTTTCTTTTTT -121
TGGTCTTTCCTGCTCAGCGCCCAACCCAAGTTCAAAGGCTGATAAGAGAGAAAATCTCATGAGGAGGTTT -51
TAGTCTAGGGAAAGTCATTCAGTGGATGTCATCTTGGCTCACAGGGGACG -1

H

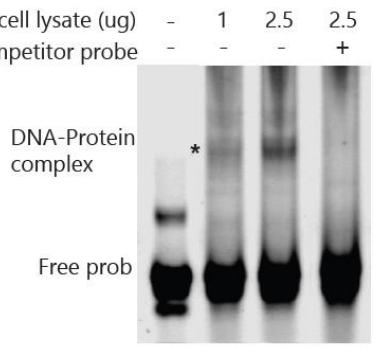

Whole cell lysate (ug) - 1 2.5 2.5
Competitor probe - - - +
DNA-Protein complex
Free prob
*

Figure 5. GATA6 is a novel regulator of ACE2. (A) WT-SARS-CoV-2 infected control and GATA6-disrupted Calu-3 cells were fixed 5 and 24 hours post infection and stained with antisera against SARS-CoV-2 (green) and DAPI (blue). Scale bar, 20um. **(B)** Western blot for the detection of GATA6 and ACE2 levels was performed in WT, ACE2 knockout, GATA6 knockout and complemented Calu-3 cells. **(C)** ACE2 and GATA6 mRNA levels relative to GAPDH amount were analyzed in WT, NT control, ACE2 and GATA6 knockout Calu-3 cells by real time PCR. **(D)** real-time PCR quantification of WT-SARS-CoV-2 in infected WT, GATA6 knockout and GATA6 knockout complemented by GATA6 or ACE2 expression Calu-3 cells. Cells were infected using MOI=0.002 for 48 hours. Data were analyzed by one-way ANOVA with Tukey's multiple comparison test Shown are means ± SD. *$p < 0.05$; **$p < 0.005$. **(E)** Cell viability of control, GATA6 knockout and GATA6 knockout complemented by GATA6 or ACE2 cells infected with WT-SARS-CoV-2 variant at MOI of 0.002 for 48 hours, and stained with crystal violet. One of three repetitions is shown. **(F)** Cells transfected with a control or siRNAs targeting GATA6 were analyzed by real-time PCR for GATA6 and ACE2 expression relative to GAPDH amount. Data were analyzed by one-way ANOVA with Tukey's multiple comparison test Shown are means ± SD. ****$p < 0.0001$. **(G)** ACE2 ATG codon upstream sequences from nucleotide -1 to -820 are shown. Sequences shown framed in bold are potential GATA binding sites. **(H)** 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.

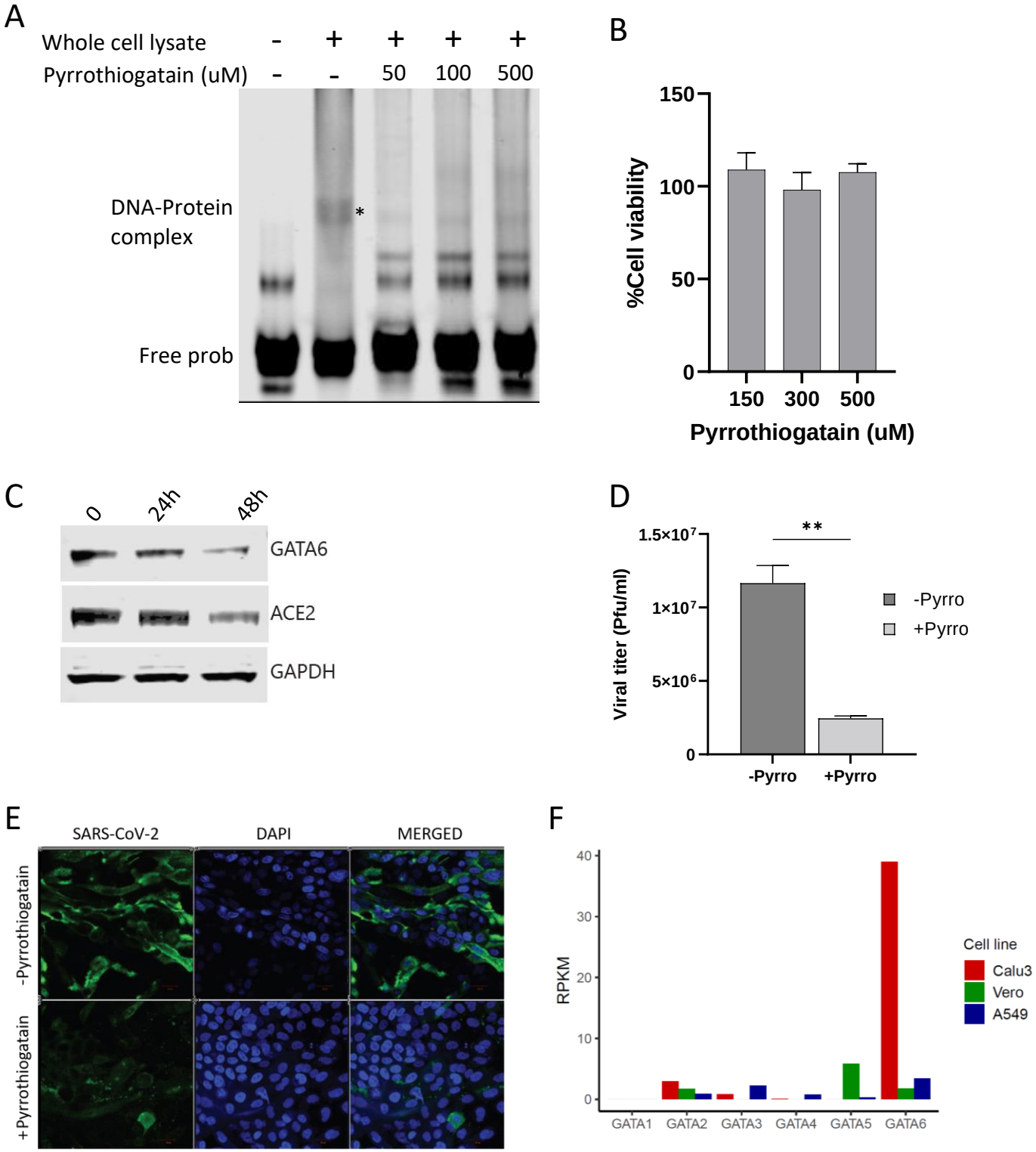
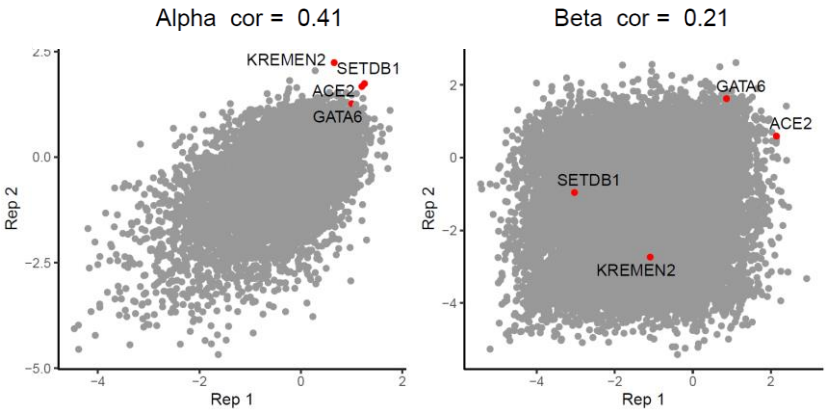


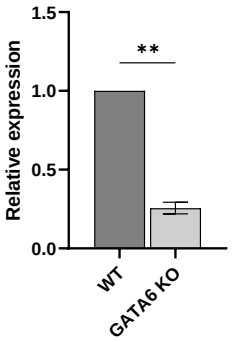
Figure 6. GATA6 may serve as a potential host directed therapeutic target. (A) Electrophoretic mobility shift assay (EMSA) of lysates from HEK293T cells over expressing GATA6 with labeled oligonucleotides derived from the promoter region of ACE2 in the presence of various concentrations of Pyrrothiogatain (0 to 500 µM). **(B)** The effect of various concentrations of Pyrrothiogatain on Calu-3 cell proliferation was determined using the XTT assay and presented as a percentage of cell viability of untreated cells. **(C)** Calu-3 cells were treated with 500uM Pyrrothiogatain and the expression of GATA6 and ACE2 were analyzed by Western blot at the indicated time. **(D)** real-time PCR quantification of WT-SARS-CoV-2 in Calu-3 cells that were pretreated with 500uM Pyrrothiogatain. Cells were infected using MOI=0.002 for 48 hours. Data were analyzed by two-tailed student's t-test. Shown are means ± SD. **$p < 0.01$. **(E)** Calu-3 cells pretreated with 500uM Pyrrothiogatain and infected with WT-SARS-CoV-2. Cells were fixed 24 hours post infection and stained with antisera against SARS-CoV-2 (green) and DAPI (blue). Scale bar, 20um. (F) mRNA read densities (reads per kilobase million, RPKM) of the six members of the GATA family in uninfected Calu-3 (red), Vero-E6 (green) and A549 cells (blue) [66-68].

Supplemental Fig. S1. LFCs in each replicate of alpha and beta



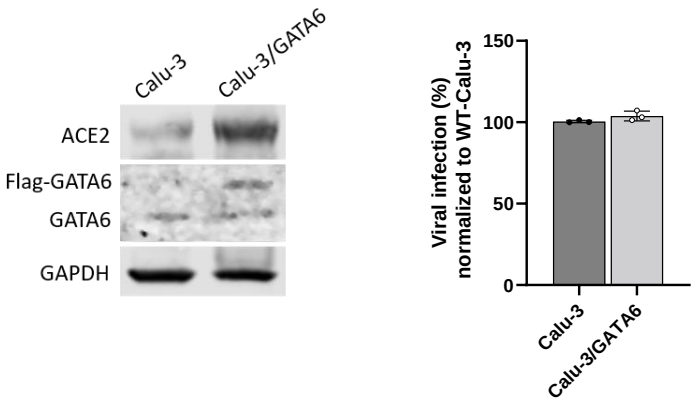
Supplemental Fig. S1. Reproducibility of the log fold change (LFC) of sgRNAs between replicates of the screens of Alpha and Beta VOCs. Top pro-viral candidates identified in our screens are displayed.

Supplemental Fig. S2. Disruption of GATA6 in Vero-E6 cells



Supplemental Fig. S2. real-time PCR measurements of GATA6 mRNA levels relative to GAPDH amount in Control and GATA6 disrupted Vero-E6 cells.

Supplemental Fig. S3. Over-expression of GATA6 in Calu-3 cells



Supplemental Fig. S3. (A) Western blot for the detection of GATA6 and ACE2 levels was performed in control and GATA6-over expressed Calu-3 cells. **(B)** real-time PCR quantification of WT-SARS-CoV-2 in control and GATA6-over expressed Calu-3 cells. Cells were infected using MOI=0.002 for 48 hours. Data were analyzed by two-tailed student's t-test. Shown are means ± SD.