Henry Ford Health
Henry Ford Health Scholarly Commons

Behavioral Health Articles | Behavioral Health Services / Psychiatry

9-15-2023

# MAP3K19 regulatory variation in populations with African ancestry may increase COVID-19 severity

Zhongshan Cheng

Yi Cai

Ke Zhang

Jingxuan Zhang

Hongsheng Gui

*See next page for additional authors*

Follow this and additional works at: https://scholarlycommons.henryford.com/behavioralhealth_articles

## Authors

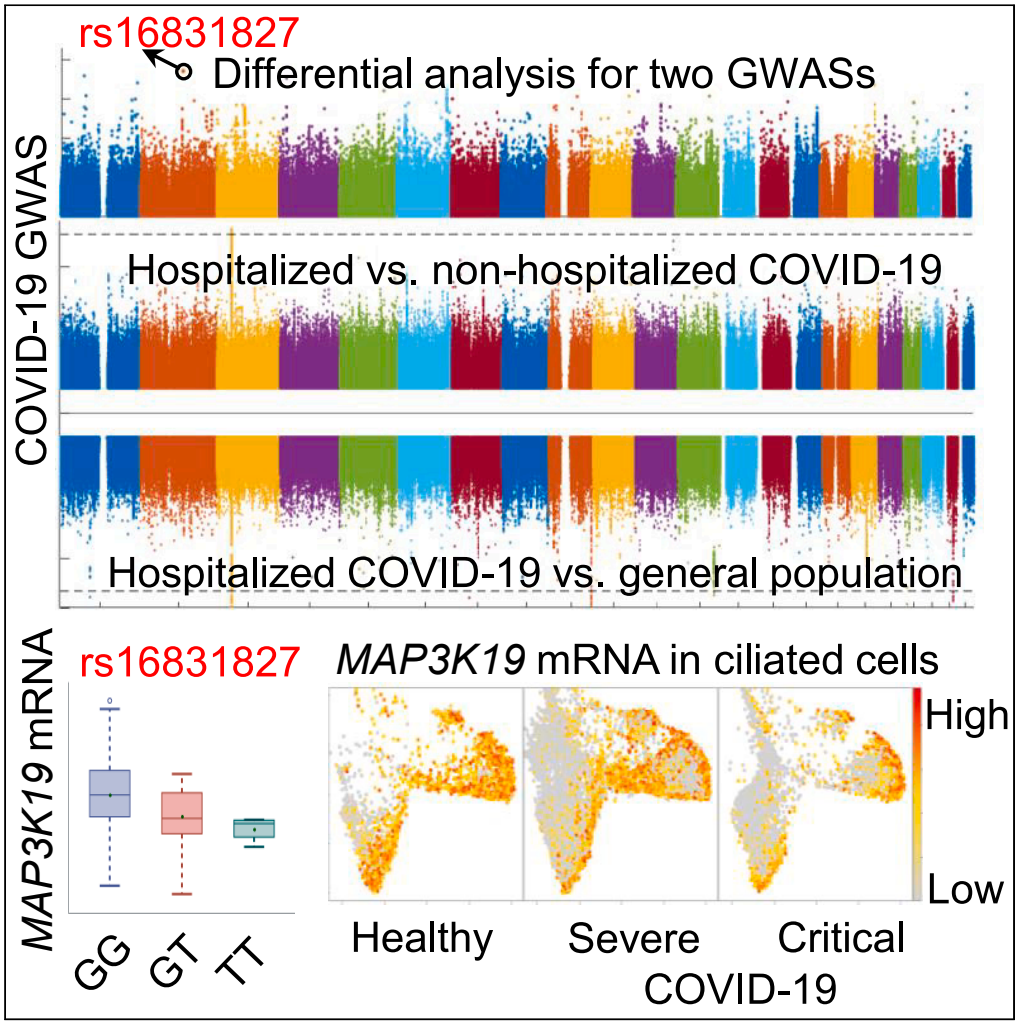
Zhongshan Cheng, Yi Cai, Ke Zhang, Jingxuan Zhang, Hongsheng Gui, Yu-Si Luo, Jie Zhou, and Brian DeVeale

iScience

CellPress
OPEN ACCESS

Article

# *MAP3K19* regulatory variation in populations with African ancestry may increase COVID-19 severity




Zhongshan Cheng, Yi Cai, Ke Zhang, ..., Yu-Si Luo, Jie Zhou, Brian DeVeale

cheng.zhong.shan@gmail.com (Z.C.)
brian.deveale@uwindsor.ca (B.D.)

## Highlights

Identified ancestry-linked genetic risk variants for COVID-19 hospitalization

COVID-19 risk SNP rs16831827*T is prevalent in populations with African ancestry

rs16831827*T is associated with reduced *MAP3K19* expression

Lower *MAP3K19* expression in ciliated cells is associated with severe COVID-19

Cheng et al., iScience 26, 107555
September 15, 2023 © 2023 The Author(s).
https://doi.org/10.1016/j.isci.2023.107555

iScience

CellPress
OPEN ACCESS

Article

# *MAP3K19* regulatory variation in populations with African ancestry may increase COVID-19 severity

Zhongshan Cheng,[1,9,*] Yi Cai,[2] Ke Zhang,[3] Jingxuan Zhang,[4] Hongsheng Gui,[5] Yu-Si Luo,[6] Jie Zhou,[7] and Brian DeVeale[8,*]

## SUMMARY

**To identify ancestry-linked genetic risk variants associated with COVID-19 hospitalization, we performed an integrative analysis of two genome-wide association studies and resolved four single nucleotide polymorphisms more frequent in COVID-19-hospitalized patients with non-European ancestry. Among them, the COVID-19 risk SNP rs16831827 shows the largest difference in minor allele frequency (MAF) between populations with African and European ancestry and also shows higher MAF in hospitalized COVID-19 patients among cohorts of mixed ancestry (odds ratio [OR] = 1.20, 95% CI: 1.10–1.30) and entirely African ancestry (OR = 1.30, 95% CI: 1.02–1.67). rs16831827 is an expression quantitative trait locus of *MAP3K19*. *MAP3K19* expression is induced during ciliogenesis and most abundant in ciliated tissues including lungs. Single-cell RNA sequencing analyses revealed that *MAP3K19* is highly expressed in multiple ciliated cell types. As rs16831827*T is associated with reduced *MAP3K19* expression, it may increase the risk of severe COVID-19 by reducing *MAP3K19* expression.**

## INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection causes coronavirus disease 2019 (COVID-19).[1] SARS-CoV-2 is transmitted by respiratory droplets and aerosols between individuals.[2] SARS-CoV-2 infects cells via viral spike (S) protein binding to ACE2 on the surface of target cells, such as respiratory epithelia cells where ACE2 is abundant.[3] Subsequent cleavage of the S protein by the serine protease TMPRSS2 or cathepsin L/B primes the infection by allowing fusion of the viral and lysosomal membranes.[4] COVID-19 symptoms vary across individuals, with acute respiratory distress and excessive inflammation being hallmarks of severe COVID-19.[5,6]

COVID-19 severity is influenced by multiple risk factors, including sex, age, body mass index, comorbidities, and ethnicity.[7] Genetic factors influencing the severity of COVID-19 have been reported by large-scale genome-wide association studies (GWASs). These include GWASs comparing cohorts with severe COVID-19, defined by hospitalization, to the general population.[8–11] They revealed several single nucleotide polymorphisms (SNPs) or indels associated with the risk of severe COVID-19, including rs11385942, an intronic indel of *LZTFL1* that tags a gene cluster harboring *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*, rs10735079 that maps to a gene cluster comprised of *OAS1*, *OAS2*, *and OAS3*, and three other SNPs (rs74956615, rs2109069, and rs2236757) that map to *TYK2*, *DPP9*, and *IFNAR2*, respectively.[10] Among these genes, *OAS1-3*, *IFNAR2*, *CCR9*, and *CXCR6* are involved in innate-immune responses to SARS-CoV-2 infection.[10,12] Whereas *LZTFL1* is an airway cilia regulator,[13] and cilia may be a site of SARS-CoV-2 entry and influence pathogenesis.[14] However, these COVID-19 risk SNPs or genes were all found in populations of European ancestry and few COVID-19 genetic risk factors have been identified in other populations.

Here, we inferred SNPs with higher frequency in hospitalized COVID-19 patients with African (AFR) ancestry and sought to interrogate their impact on COVID-19 progression. To do so, we performed an integrative comparison between two COVID-19 hospitalization GWASs performed on ethnically diverse populations. This involved comparing COVID-19-hospitalized patients (n = 2,430) to non-hospitalized patients (n = 8,478) (HGI-B1) and COVID-19-hospitalized patients (n = 7,885) to the general population (n = 961,804) (HGI-B2). We revealed that rs16831827 is a COVID-19 risk SNP in individuals with African ancestry. rs16831827 is an expression quantitative trait locus (eQTL) of

[1]Center for Applied Bioinformatics, St Jude Children's Research Hospital, 262 Danny Thomas Pl, Memphis, TN 38105, USA
[2]Guangdong Key Laboratory of Regional Immunity and Diseases, Department of Pathogen Biology, Shenzhen University Medical School, Shenzhen 518000, China
[3]The Key and Characteristic Laboratory of Modern Pathogenicity Biology, School of Basic Medical Sciences, Guizhou Medical University, Guizhou, Guiyang 550025, China
[4]Hubei Key Laboratory of Embryonic Stem Cell Research, School of Basic Medical Sciences, Hubei University of Medicine, Shiyan, Hubei 442000, China
[5]Behavioral Health Services and Psychiatry Research, Henry Ford Health, Detroit, MI 48202, USA
[6]Department of Emergency, The Affiliated Hospital of Guizhou Medical University, Guizhou, Guiyang 550004, China
[7]Department of Microbiology, The University of Hong Kong, Hong Kong 999077, China
[8]The Department of Biomedical Sciences, University of Windsor, Windsor, ON N9B 3P4, Canada
[9]Lead contact
*Correspondence: cheng.zhong.shan@gmail.com (Z.C.), brian.deveale@uwindsor.ca (B.D.)
https://doi.org/10.1016/j.isci.2023.107555


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

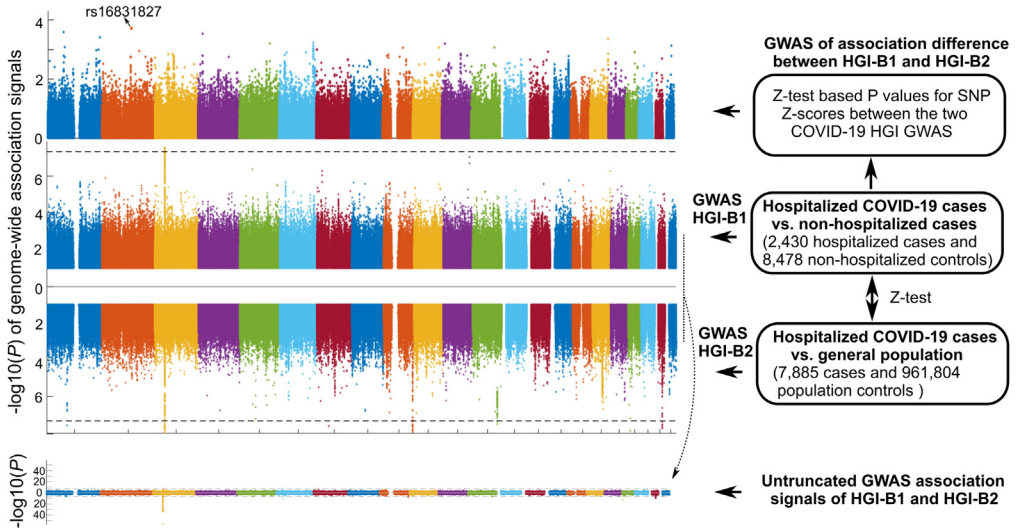



**Figure 1. Comparative analysis of two COVID-19 hospitalization GWASs**

The differential analysis between COVID-19 association effect sizes for each SNP between two GWASs. The SNP with the top independently differential association *Z* score between the two GWASs is rs16831827 (delta *Z* score [ΔZ-score] p = 1.5 × $10^{-5}$). No significant associations from the original GWASs showed differential association *Z* score ($P < 5 \times 10^{-5}$).

*MAP3K19* that is primarily expressed in ciliated cells.[15] We hypothesize that rs16831827 is a risk factor for COVID-19 hospitalization because it impairs mucociliary clearance that is exacerbated by SARS-CoV-2 infection.

## RESULTS

To identify ancestry-linked genetic risk variants for severe COVID-19, we evaluated the differential association signals of SNPs between COVID-19 hospitalized cases and either non-hospitalized COVID-19 cases or the general population (release 4, open to public on October 20, 2020). This analysis leverages two GWAS studies. The first, HGI-B1, compares hospitalized COVID-19 patients (n = 2,340) to non-hospitalized COVID-19 patients (n = 8,478) and the second, HGI-B2, compares hospitalized COVID-19 patients (n = 7,885) with the general population (n = 961,804) (see Figures 1 and S1). Both GWASs were conducted before COVID-19 vaccines were available. We initially thought that the relatively small and heterogeneous non-hospitalized cohort used in HGI-B1 may result in associations driven by ancestry as compared to the general population control cohort used in HGI-B2 (see Figure S1). To test this hypothesis, we adapted the differential Z-test statistical method to compare SNP association effect size beta (i.e., per unit increase or decrease in the outcome) from summary statistics of HGI-B1 and HGI-B2 to reveal SNPs displaying differential association signals between the two HGI GWASs.[16] ΔZ-scores were used to estimate the difference between the two effect sizes of a SNP. In detail, we selected high-quality SNPs with minor allele frequencies (MAF) > 0.01 and imputation scores > 0.6, and then evaluated the GWAS quality using Manhattan and quantile-quantile (QQ) plots (see Figure S2). Both the HGI-B1 and HGI-B2 GWAS had $\lambda$ values ~1, confirming their high quality and indicating negligible inflation of association values in each GWAS. In contrast, we found deflation in the differential effect size analysis between HGI-B1 and HGI-B2 ($\lambda$ = 0.639). Further investigation of the formula used to estimate the differential effect size of each SNP (see STAR Methods) revealed that the standard errors of two effect size beta from the two GWASs are highly correlated, mainly as a result of the overlapping samples used in HGI-B1/B2 (Figure S1). Normalizing differential Z scores genome-wide reduced the genome-wide correlation and achieved uninflated differential effect sizes between HGI-B1 and HGI-B2 ($\lambda$ = 0.991). To substantiate this approach and provide additional context for our comparison between HGI-B1 and HGI-B2, we evaluated 12 published SNPs related to COVID-19 hospitalization and 4 related to COVID-19 susceptibility.[17] Most of these genome-wide significant SNPs associated with COVID-19 severity or susceptibility are significant with ORs greater or less than 1 in HGI-B2 (Figure S3). An OR = 1 would indicate no effect on the odds of COVID-19 hospitalization, while an OR greater or less than 1 would indicate a risk or protective effect on the odds of COVID-19 hospitalization.[18] Additionally, we found that 7 of these 16 published COVID-19 risk SNPs had larger effect sizes (normalized nominal differential p < 0.05; see Table S1) in HGI-B2 than in HGI-B1, stemming from the larger sample size in HGI-B2. Conversely, none were larger in HGI-B1 than HGI-B2 (see Figure S1; Table S1). In the updated differential effect size analysis, the top independent signal, close to genome-wide significance (normalized differential p = 7.8 × $10^{-8}$), was represented by the SNP rs16831827 (*P* of ΔZ-score = 2 × $10^{-5}$; COVID-19 association p = 1.7 × $10^{-4}$, beta = 0.19, se = 0.05, and OR = 1.3 for HGI-B2; Table 1; Figure 2), which is located in the intronic region of the gene R3H domain containing 1 (*R3HDM*). Three other independent SNPs also differed in association at a threshold of $P < 5 \times 10^{-5}$ (Table 1; Figures S4A–S4D). Thus, our comparison revealed four independent SNPs with differential effect sizes between the two HGI GWASs.

**Table 1. SNPs displaying the greatest differential association signal between two GWASs that compared COVID-19 hospitalized cases to either non-hospitalized patients (HGI-B1) or the general population (HGI-B2)**

| Chr | Position (hg19) | SNP | Ref/ Alt | Gene | *P* (ΔZ-score) | *P* (*Z* score) of HGI-B1 | *P* (*Z* score) of HGI-B2 | Alt-allele frequency in HGI-B1 | Alt-allele frequency in HGI-B2 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 136323787 | rs16831827 | G/T | *R3HDM1* | $2.0 \times 10^{-5}$ (−4.3) | $7.7 \times 10^{-3}$ (−2.7) | $1.6 \times 10^{-4}$ (3.8) | 0.19 | 0.02 |
| 4 | 20980969 | rs10009407 | G/C | *KCNIP4* | $3.0 \times 10^{-5}$ (4.2) | $3.3 \times 10^{-5}$ (4.14) | $3.2 \times 10^{-1}$ (−1.0) | 0.46 | 0.39 |
| 1 | 239197542 | rs112317747 | T/C | None | $4.0 \times 10^{-5}$ (4.1) | $8.1 \times 10^{-6}$ (4.5) | $8.8 \times 10^{-1}$ (−0.15) | 0.04 | 0.03 |
| 17 | 80168281 | rs141909101 | C/T | *CCDC57* | $5.0 \times 10^{-5}$ (4.1) | $5.7 \times 10^{-5}$ (5.0) | $4.8 \times 10^{-1}$ (0.7) | 0.02 | 0.13 |

HGI: the COVID-19 host genetics initiative; Chr: chromosome; Ref: reference allele; Alt: alternative allele; ΔZ-score: delta Z-score between the effect size beta of each common SNP by comparing the two HGI GWASs. Z-score: effect size beta/se, which is not normalized.

Next, we asked whether the differential association of these four SNPs reflected differences in the ancestry of the HGI-B1/2 control cohorts by evaluating their MAFs across populations. The MAFs of three of these SNPs showed varied global population frequency (Figures S4A–S4D), with the top SNP rs16831827 displaying the largest frequency difference between African and other populations. In the two COVID-19 hospitalization GWASs, the MAF of rs16831827 is 0.19 in the whole HGI-B1 cohort but only 0.02 in the whole HGI-B2 cohort (Table 1). After evaluating the population frequencies of rs16831827 across 26 populations from the 1000 Genome Project, we found the SNP has higher MAF (MAF >0.45) in 7 populations with African ancestry (Figure S4A). For the 19 other populations with European or Asian ancestry, the MAFs of rs16831827 are all <0.05, indicating that the rs16831827 SNP is more prevalent in populations of African ancestry.

To validate that rs16831827 is a risk variant of severe COVID-19, we used COVID-19 hospitalization GWASs from the UK Biobank that performed GWAS on cohorts of common ancestry.[16] In this study, rs16831827 was associated with severe COVID-19 in AFR but not other populations (GWAS of hospitalized COVID-19 cases vs. non-hospitalized COVID-19 positive or COVID-19 negative of all tested AFR samples; p = 0.01, beta = 0.40, Figure S5). These UK Biobank AFR samples were not included in the two HGI COVID-19 GWASs analyzed here. To evaluate the contribution of rs16831827*T to the risk of COVID-19-related hospitalization, we calculated the average odds ratios (ORs) for the SNP in the HGI-B2 GWAS with mixed population (OR = 1.2) and the UK Biobank COVID-19-hospitalization GWAS with African ancestry (OR = 1.3). An OR = 1.25 corresponds to a 1.25x greater risk of hospitalization for COVID-19 patients. No association between rs16831827 and COVID-19 hospitalization was observed in patients with EUR ancestry in the UK Biobank GWASs, in patients from Italy with EUR ancestry in the study (p = 0.144),[9] or patients of European ancestry in the HGI-B2-EUR GWAS (Figure S5). Hence these analyses reveal rs16831827 to be a genetic risk marker of COVID-19 hospitalization specifically among patients with AFR ancestry.

To understand how rs16831827 impacts COVID-19 severity, we evaluated the relationship between rs16831827 genotype and the expression of nearby genes using the GTEx database.[19] We found that rs16831827 is an eQTL for four genes (Figure 3), including *MAP3K19* (eQTL $p = 1.5 \times 10^{-7}$ in the testis), *ZRANB3* (eQTL $p = 9.0 \times 10^{-5}$ in cultured fibroblasts), *MCM6* (eQTL $p = 1.6 \times 10^{-4}$ in the heart), and *DARS* (eQTL $p = 2.7 \times 10^{-4}$ in whole blood). Further evaluation of the correlation between rs1683817 and *CXCR4* using the GTEx eQTL dashboard revealed that higher expression of *CXCR4* is correlated with the rs16831827 risk allele in adrenal glands ($p = 2.3 \times 10^{-2}$). The most significant eQTL of the COVID-19 risk allele rs16831827*T was correlated with lower *MAP3K19* expression in the testis. Because rs16831827 is also an eQTL of *MAP3K19* in the lung (p = 0.04, Figures 3A, S6, and S7), which is central to COVID-19 pathology, we further considered whether reduced *MAP3K19* expression by rs16831827*T led to its association with severe COVID-19. We hypothesized that other regulatory SNPs of *MAP3K19*, especially rare AFR SNPs, may be associated with COVID-19 severity. Hence we searched for COVID-19 association signals among all SNPs located in the 1-Mbp window around *MAP3K19* across all Regeneron COVID-19 GWASs via its browser.[20] Among the top hits, we identified two SNPs in *MAP3K19* introns (rs186150828 and rs192473276; MAFs <0.001) that are associated with COVID-19 hospitalization among populations with African ancestry (both ORs > 100; Table S2; Figure S8). Both passed the multiple adjustment *P* threshold of $5 \times 10^{-7}$. These rare SNPs are not linked with rs16831827 among all 1000 Genome Project samples (all $R^2 < 0.01$). Hence, rs16831827 is an eQTL of *MAP3K19*, and several *MAP3K19*-linked SNPs are associated with COVID-19 severity in patients with AFR ancestry.

To explore normal *MAP3K19* expression and how it is altered in the context of SARS-CoV-2 infection, we leveraged published single-cell data[21] and found that *MAP3K19* is primarily expressed in ciliated cells, including nasal multiciliated cells[22] (see Figure S9). In respiratory tissues, *MAP3K19* is highly expressed in ciliated epithelial cells (Figures 3B, 3C, and S6). Additionally, *MAP3K19* is down-regulated in nasal multiciliated cells following SARS-CoV-2 infection in bulk RNA-seq data (Figure S10). In experimentally infected human bronchial epithelial cells, *MAP3K19* expression was acutely up-regulated after SARS-CoV-2 infection but then progressively downregulated compared to controls (Figure S11). Furthermore, the percentage of one cell type (Figures 4A and 4B), ciliated cells (Ciliated), is significantly lower in critical COVID-19 patients than in healthy controls (Figures 4A–4D), consistent with previous report that SARS-CoV-2 depletes ciliated cells.[24] No difference in the distribution of other cell types was observed across patient groups. We also observed that a lower proportion of the ciliated cells express *MAP3K19* in critical but not severe COVID-19 patients compared to healthy controls (Figure 4E). Furthermore, the expression of *MAP3K19* is consistently lower among all ciliated cells in critical COVID-19 patients compared to healthy controls (Figure 4F). Differential expression of *MAP3K19* between severe COVID-19 and healthy controls was apparent in differentiated ciliated cells (Ciliated-Diff), ciliated-viral-response cells (Ciliated-ViralResp), and secretory-ciliated cells (Secretory-Ciliated cells). In Ciliated-Diff and Ciliated-ViralResp cells only, *MAP3K19* expression is significantly lower in critical COVID-19 than in severe COVID-19 samples. As the percentage of non-ciliated cells expressing

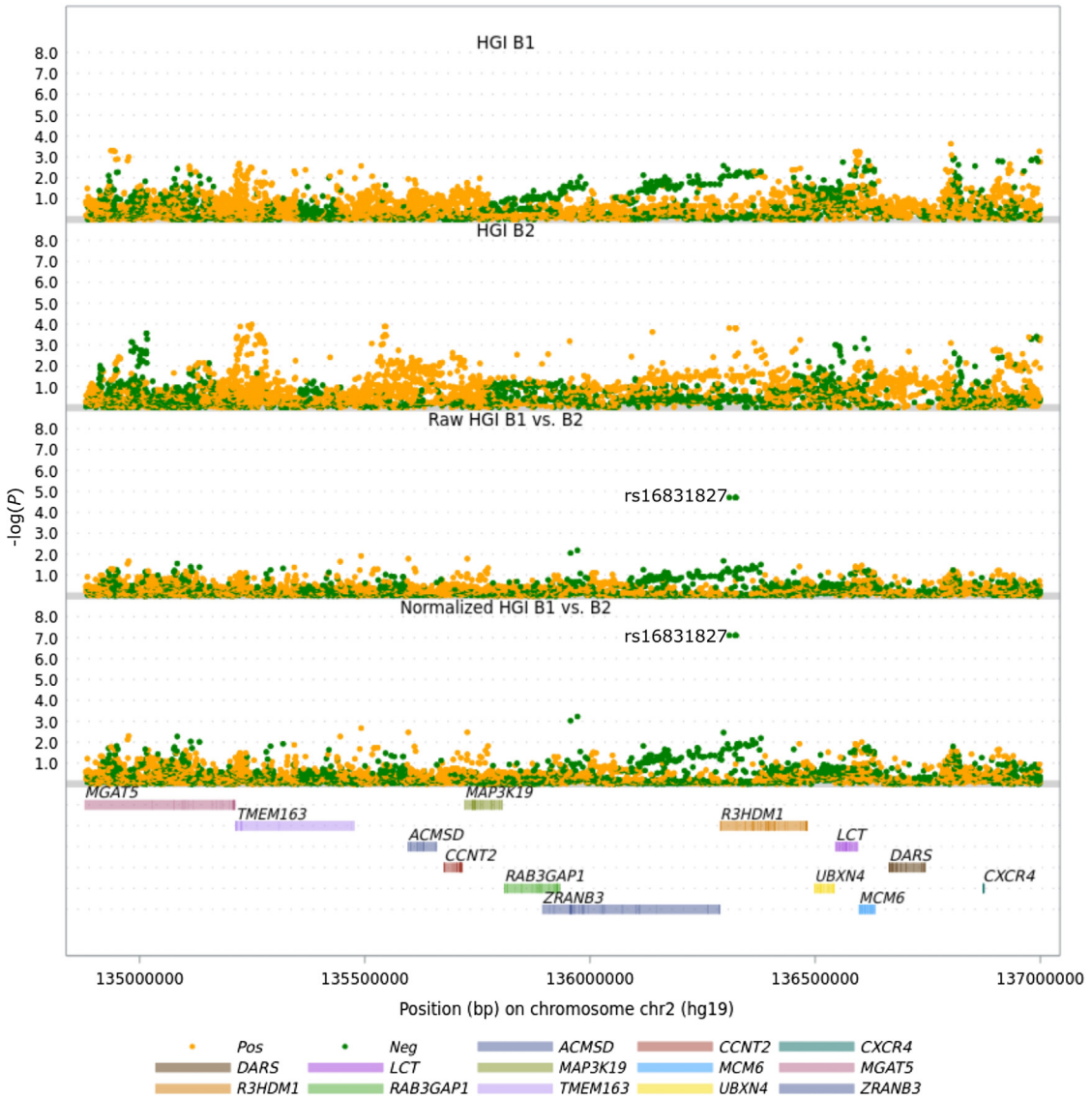



**Figure 2. Local Manhattan plots of COVID-19 association signals around *MAP3K19***

The four scatterplots illustrate the association signals among 4 GWASs: HGI-B1, HGI-B2, raw HGI-B1 vs. HGI-B2, and normalized HGI-B1 vs. HGI-B2. After adjusting for sample overlap, the final differential effect size of rs16831827 for the HGI-B1 to HGI-B2 comparison is $p = 7.8 \times 10^{-8}$. The positive and negative directions of effect sizes are colored orange and green in the scatterplots.

*MAP3K19* and *MAP3K19* expression in non-ciliated cells are extremely low, we did not conduct differential expression analyses. Taken together, *MAP3K19* is primarily expressed in ciliated cells. In critical COVID-19 patients, the expression of *MAP3K19* is downregulated in Ciliated, Ciliated-Diff cells, Ciliated-ViralResp cells, and Secretory-Ciliated cells compared to healthy controls.

## DISCUSSION

Here we identified ancestry-linked SNPs that affect COVID-19 severity. Our evaluation revealed that rs16831827 (an eQTL of *MAP3K19)* is a COVID-19 risk SNP found predominantly in individuals with African ancestry. The association of rs16831827 with COVID-19 hospitalization in individuals with AFR ancestry was also observed in an independent GWAS of entirely AFR ancestry (UK Biobank) and not observed in European populations. We also revealed two additional rare variants of *MAP3K19* associated with COVID-19 hospitalization in populations with AFR ancestry. In contrast, no rare *MAP3K19* variants associated with COVID-19 severity emerged in populations with European ancestry. Our results suggest that the *MAP3K19* SNP rs16831827 is a genetic risk marker for COVID-19 hospitalization specifically in patients with AFR ancestry.

Our analyses indicate that the contribution of rs16831827 to the risk of COVID-19 hospitalization is consistent as the ORs derived from the HGI-B2 GWAS and the UKB AFR replication COVID-19 hospitalization GWAS are similar. The OR of rs16831827 is 1.2 (95% confidence interval:

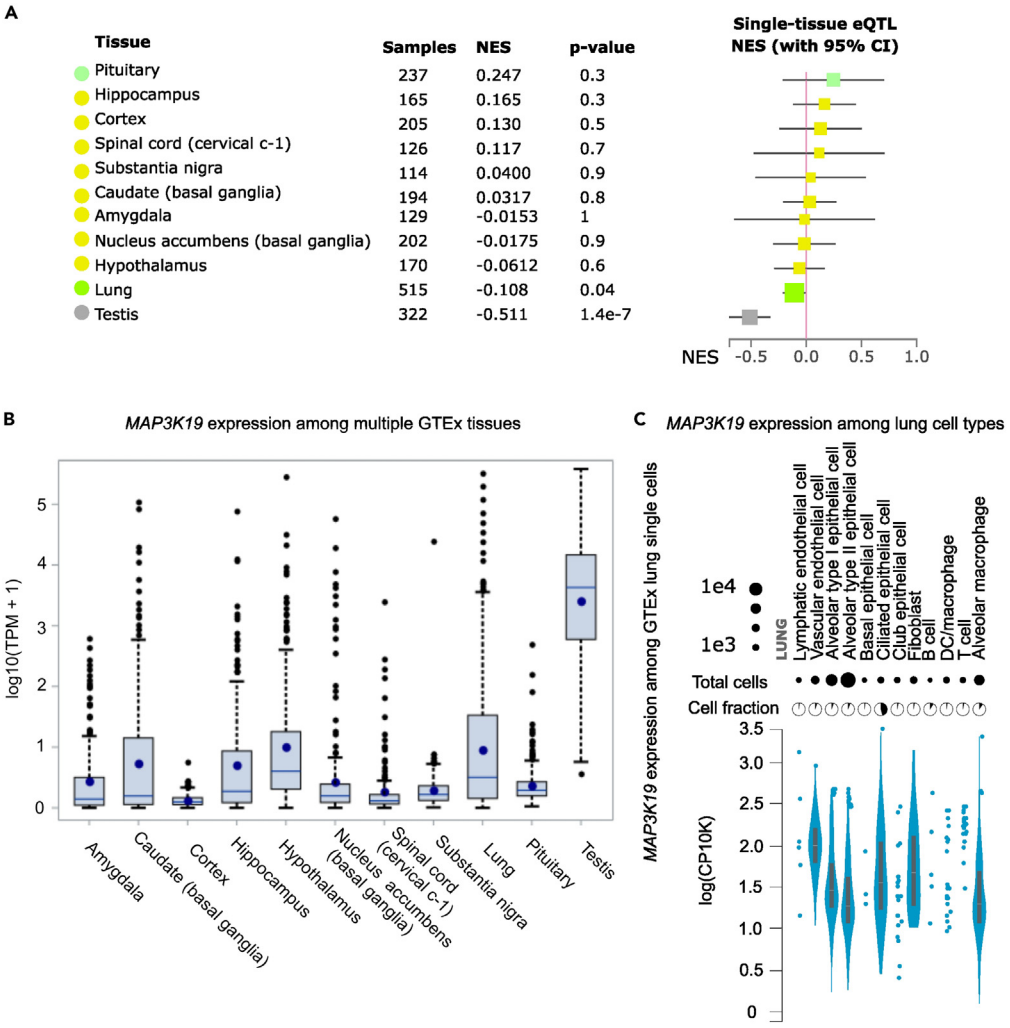



**Figure 3. Analysis of rs16831827 and *MAP3K19* expression**

(A) eQTL analysis of rs16831827 in GTEx tissues.

(B) Expression of *MAP3K19* across GTEx tissues. The error bar indicates the 95% confidence interval (CI) of odd ratio (OR).

(C) *MAP3K19* expression among lung single cell types. Only tissues with median normalized expression of Transcripts Per Kilobase Million (TPM) > 0.05 were included in panel A and B. The box-and-whisker plots display the mean (dot within box), median (line inside the box), inter-quantile interval (box), minimum (lowest value of whisker), and maximum (maximum value of whisker), with outliers represented by dots up or down the whiskers. Note: the lower or upper whisker is the line specifically goes from the minimum to the lower quartile or the line links the upper quartile to maximum. NES: normalized effect size.

1.1–1.3) in the HGI-B2 GWAS (COVID-19-hospitalized cases vs. general population controls), indicating that a COVID-19 case harboring the rs16831827 risk allele T is 20% more likely to be hospitalized than someone without it. Similarly, the rs16831827 risk allele T in the replication cohort from UKB demonstrates OR = 1.3 (95% CI: 1.02–1.6).

Gene expression analyses of *MAP3K19* localized its expression primarily to ciliated cell types and support its involvement in ciliogenesis. *MAP3K19* is expressed in nasal multiciliated epithelial cells and airway-ciliated cells, both of which are relevant to COVID-19 pathogenesis. Nasal ciliated cells are sites of early SARS-CoV-2 replication,[25] and airway ciliated cells are a primary infection site.[26] *MAP3K19* expression is reduced at later stages of SARS-CoV-2 infection in both cell types. The most significant reduction of *MAP3K19* occurred in the Ciliated-Diff cells, which along with the other two cilial cell types, Secretory-Ciliated cells, and Ciliated-ViralResp cells, also correlated with increased COVID-19 severity. Hence these data suggest that *MAP3K19* impacts COVID-19 severity in ciliated cells that are sites of SARS-CoV-2 infection and replication.

*MAP3K19* was prioritized as the potential causal link between rs16831827 and severe COVID-19 based on its association with rs16831827 in lung tissue, which is critical to COVID-19 pathology. However, rs16831827 is located in a genomic region harboring a 4.0 Kbp deletion within the first intron of *LCT* (lactase gene) in populations of AFR but not European ancestry.[27] The SNP potentially regulates multiple genes, including *MAP3K19*, *ZRANB3* (zinc finger RANBP2-type containing 3), *MCM6* (minichromosome maintenance complex component 6), and *DARS* (aspartyl-tRNA synthetase), as well as *CXCR4* (C-X-C chemokine receptor type 4). rs16831827 shows significant associations with these

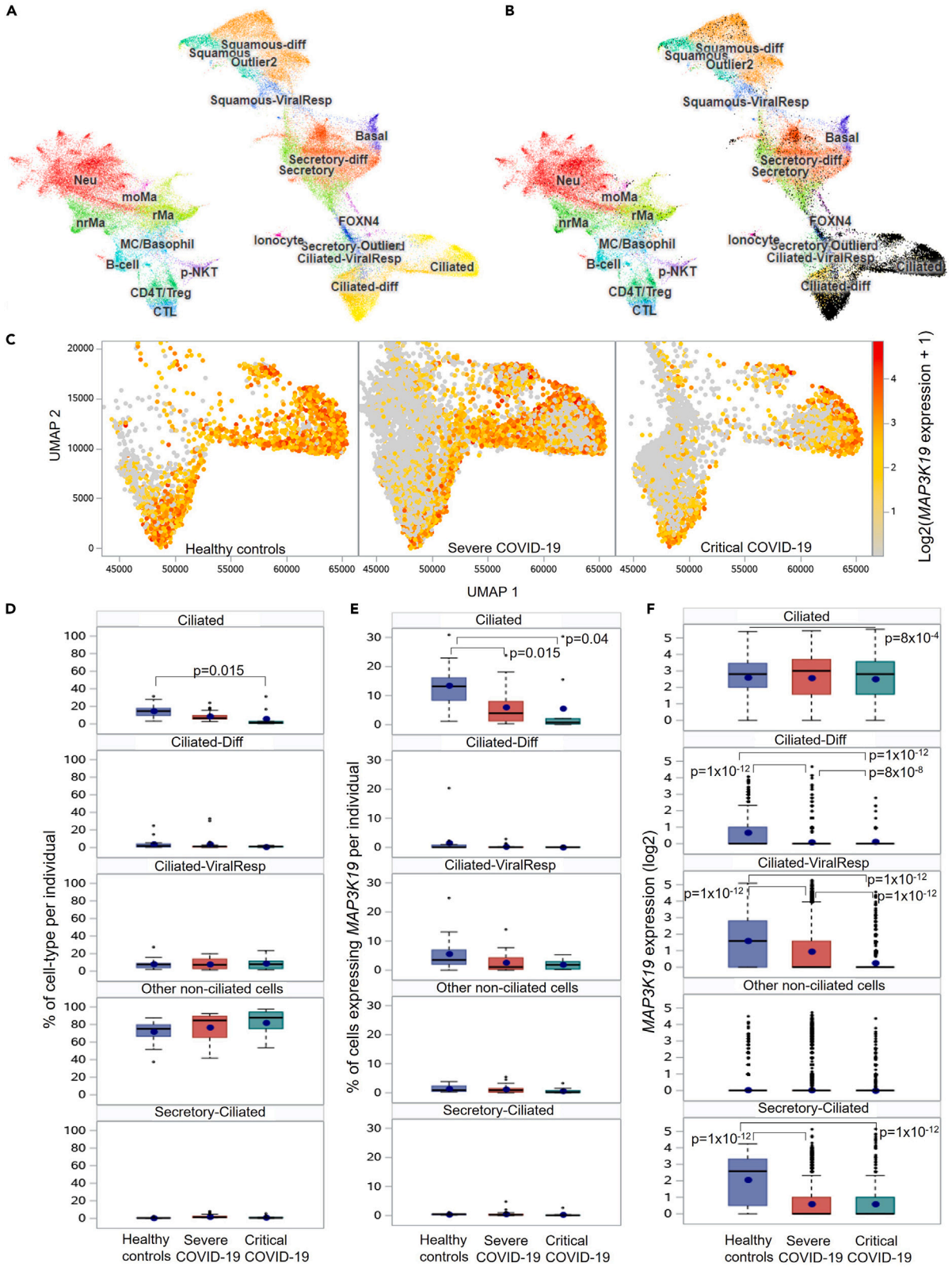

A
Squamous-diff
Squamous
Outlier2
Squamous-ViralResp
Basal
Secretory-diff
Secretory
FOXN4
Ionocyte
Secretory-Outlier
Ciliated-ViralResp
Ciliated
Ciliated-diff
Neu
moMa
rMa
nrMa
MC/Basophil
B-cell
p-NKT
CD4T/Treg
CTL
B
C
UMAP 2
UMAP 1
Healthy controls
Severe COVID-19
Critical COVID-19
Log2(MAP3K19 expression + 1)
D
% of cell-type per individual
Ciliated
Ciliated-Diff
Ciliated-ViralResp
Other non-ciliated cells
Secretory-Ciliated
p=0.015
Healthy controls
Severe COVID-19
Critical COVID-19
E
% of cells expressing MAP3K19 per individual
p=0.04
p=0.015
F
MAP3K19 expression (log2)
p=8x10-4
p=1x10-12
p=8x10-8

**Figure 4. *MAP3K19* is expressed in ciliated cells and associated with COVID-19 severity**

(A) Uniform Manifold Approximation and Projection (UMAP) of nasopharyngeal single cell expression profiles[23] from healthy controls (n = 16), severe (n = 23), and critical (n = 9) COVID-19 patients.

(B) Highlighted view of ciliated clusters expressing *MAP3K19* with black dots in the UMAP as that from panel (A). *MAP3K19* is predominantly expressed in ciliated cell types.

(C) Fewer ciliated cells show *MAP3K19* expression in individuals with COVID-19.

(D) The percentage of each cell type among all cells.

(E) The percentage of *MAP3K19*-expressing cells.

(F) *MAP3K19* expression is reduced in ciliated cells of critical COVID-19 patients. There are 4 ciliated cell types: ciliated (Ciliated), ciliated-differentiated (Ciliated-Diff), ciliated-viral-response (Ciliated-ViralResp), and secretory-ciliated (Secretory-Ciliated) cells. *MAP3K19* expression was reduced in Ciliated-Diff, Ciliated-ViralResp, and Secretory-Ciliated cells but not Ciliated cells from severe critical COVID-19 patients compared to healthy controls. *MAP3K19* expression is lower in Ciliated-Diff and Ciliated-ViralResp cells of critical relative to severe COVID-19 patients. The box-and-whisker plots display the mean (dot within box), median (line within the box), inter-quantile interval (box), minimum (lowest value of whisker), and maximum (maximum value of whisker), with outliers represented by dots above or below the whiskers. Pairwise statistical significance test was conducted with the "lsmeans" statement adjusted by the "TUKEY" method using the SAS procedure "proc GLM", with $p < 0.05$ set as the significance threshold.

first 4 genes in at least one GTEx tissue (all $P < 5 \times 10^{-4}$). *CXCR4* expression is also weakly associated with rs16831827 in adrenal glands ($p = 2.3 \times 10^{-2}$). Altered regulation of these genes, perhaps outside the lung, may impact COVID-19 severity. In particular, higher *CXCR4* expression was associated with increased COVID-19 severity by single-cell analysis of nasopharyngeal samples derived from COVID-19 patients.[23] Functional experiments would be valuable to delineate regulatory effects of each gene on COVID-19 severity. It is also possible that rs16831827 exerts its effects on severe COVID-19 through multiple genes and cell types.

In relation to targeting *MAP3K19* for COVID-19 treatment, higher *MAP3K19* expression protects against COVID-19 hospitalization. We are unaware of drugs that directly stimulate *MAP3K19* expression. However, treatments that improve cilium functions among COVID-19 patients[28–30] could be a feasible approach to ameliorate severe COVID-19. Alternatively, steric-blocking oligonucleotides could be developed to directly elevate *MAP3K19* abundance.[31]

Overall, our integrative GWAS, eQTL, and expression analysis indicate that rs16831827 is a potential risk factor for severe COVID-19 in patients of AFR ancestry putatively by lowering *MAP3K19* expression and disrupting cilium function.

## Limitations of the study

A limitation of our analysis is that it is based on two HGI GWASs, which were generated without adjusting for factors predisposing to COVID-19 hospitalization, such as sex, age, BMI, and diabetes. Some of these factors are more prevalent in populations with AFR ancestry. Furthermore, the GTEx and single-cell RNA-seq data analyzed here are mostly from European individuals. According to Gay et al.,[32] 103/838 GTEx individuals have AFR ancestry and most others (n = 715) are European American. Our GTEx eQTL analysis is still informative and supportive of the potential regulatory roles played by rs16831827, since in the 1000 Genome Project the MAF of rs16831827*T is 2% and 58% in the European and AFR populations, respectively. The single-cell RNA-seq analysis is complementary, correlating expression of *MAP3K19* and COVID-19 severity as a contrast based on rs16831827 genotypes is not currently feasible with available data. To confirm our observation, future investigation with *cis*-eQTL and single-cell RNA-seq data exclusively from individuals with AFR ancestry are warranted.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability statement
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Pairwise comparison of association effect sizes
  - Normalization of differential effect sizes
  - Identification of ancestry-linked COVID-19 SNPs
  - Replication of rs16831827 in other COVID-19 GWASs
  - *MAP3K19* expression upon SARS-CoV-2 infection
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107555.

## ACKNOWLEDGMENTS

We are thankful to all researchers who participated and contributed to the HGI and GRASP projects; their dissemination of the two COVID-19 hospitalization GWASs and other COVID-19 related GWASs were essential for this study. We also think other researchers who freely shared the single cell and bulk RNA-seq data related to COVID-19. We also appreciate 3 reviewers who have provided critical suggestions that have improved the manuscript dramatically.

## AUTHOR CONTRIBUTIONS

Z.C. and B.D. conceived the study. Z.C, B.D., Y.C., K.Z., J.Z, H.G., Y.L., and J.Z. analyzed the data and wrote the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## INCLUSION AND DIVERSITY

We worked to ensure diversity in experimental samples through the selection of the genomic datasets.

Received: November 29, 2022
Revised: May 20, 2023
Accepted: August 3, 2023
Published: August 7, 2023

## REFERENCES

1. Hu, B., Guo, H., Zhou, P., and Shi, Z.L. (2021). Characteristics of SARS-CoV-2 and COVID-19. Nat. Rev. Microbiol. *19*, 141–154. https://doi.org/10.1038/s41579-020-00459-7.
2. Parasher, A. (2021). COVID-19: Current understanding of its pathophysiology, clinical presentation and treatment. Postgrad. Med. J. *97*, 312–320. https://doi.org/10.1136/postgradmedj-2020-138577.
3. Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature *426*, 450–454. https://doi.org/10.1038/nature02145.
4. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell *181*, 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052.
5. Wong, C.K., Lam, C.W.K., Wu, A.K.L., Ip, W.K., Lee, N.L.S., Chan, I.H.S., Lit, L.C.W., Hui, D.S.C., Chan, M.H.M., Chung, S.S.C., and Sung, J.J.Y. (2004). Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome. Clin. Exp. Immunol. *136*, 95–103. https://doi.org/10.1111/j.1365-2249.2004.02415.x.
6. Zhang, B., Zhou, X., Qiu, Y., Song, Y., Feng, F., Feng, J., Song, Q., Jia, Q., and Wang, J. (2020). Clinical characteristics of 82 cases of death from COVID-19. PLoS One *15*, e0235458. https://doi.org/10.1371/journal.pone.0235458.
7. Velavan, T.P., Pallerla, S.R., Rüter, J., Augustin, Y., Kremsner, P.G., Krishna, S., and Meyer, C.G. (2021). Host genetic factors determining COVID-19 susceptibility and severity. EBioMedicine *72*, 103629. https://doi.org/10.1016/j.ebiom.2021.103629.
8. Gandhi, R.T., Lynch, J.B., and Del Rio, C. (2020). Mild or moderate Covid-19. N. Engl. J. Med. *383*, 1757–1766. https://doi.org/10.1056/NEJMcp2009249.
9. Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., et al.; Severe Covid-19 GWAS Group (2020). Genomewide association study of severe Covid-19 with respiratory failure. N. Engl. J. Med. *383*, 1522–1534. https://doi.org/10.1056/NEJMoa2020283.
10. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. Nature *591*, 92–98. https://doi.org/10.1038/s41586-020-03065-y.
11. Anastassopoulou, C., Gkizarioti, Z., Patrinos, G.P., and Tsakris, A. (2020). Human genetic factors associated with susceptibility to SARS-CoV-2 infection and COVID-19 disease severity. Hum. Genom. *14*, 40. https://doi.org/10.1186/s40246-020-00290-4.
12. Zhou, S., Butler-Laporte, G., Nakanishi, T., Morrison, D.R., Afilalo, J., Afilalo, M., Laurent, L., Pietzner, M., Kerrison, N., Zhao, K., et al. (2021). A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity. Nat. Med. *27*, 659–667. https://doi.org/10.1038/s41591-021-01281-1.
13. Fink-Baldauf, I.M., Stuart, W.D., Brewington, J.J., Guo, M., and Maeda, Y. (2022). CRISPRi links COVID-19 GWAS loci to LZTFL1 and RAVER1. EBioMedicine *75*, 103806. https://doi.org/10.1016/j.ebiom.2021.103806.
14. Buqaileh, R., Saternos, H., Ley, S., Aranda, A., Forero, K., and AbouAlaiwi, W.A. (2021). Can cilia provide an entry gateway for SARS-CoV-2 to human ciliated cells? Physiol. Genom. *53*, 249–258. https://doi.org/10.1152/physiolgenomics.00015.2021.
15. van der Vaart, J., Böttinger, L., Geurts, M.H., van de Wetering, W.J., Knoops, K., Sachs, N., Begthel, H., Korving, J., Lopez-Iglesias, C., Peters, P.J., et al. (2021). Modelling of primary ciliary dyskinesia using patient-derived airway organoids. EMBO Rep. *22*, e52058. https://doi.org/10.15252/embr.202052058.
16. Thibord, F., Chan, M.V., Chen, M.H., and Johnson, A.D. (2022). A year of Covid-19 GWAS results from the GRASP portal reveals potential genetic risk factors. HGG Adv. *3*, 100095. https://doi.org/10.1016/j.xhgg.2022.100095.
17. Niemi, M.E.K., Karjalainen, J., Liao, R.G., Neale, B.M., Daly, M., Ganna, A., Pathak, G.A., Andrews, S.J., Kanai, M., Veerapen, K., et al. (2021). Mapping the human genetic architecture of COVID-19. Nature *600*, 472–477. https://doi.org/10.1038/s41586-021-03767-x.
18. Szumilas, M. (2010). Explaining odds ratios. J. Can. Acad. Child Adolesc. Psychiatr. *19*, 227–229.
19. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
20. Horowitz, J.E., Kosmicki, J.A., Damask, A., Sharma, D., Roberts, G.H.L., Justice, A.E., Banerjee, N., Coignet, M.V., Yadav, A., Leader, J.B., et al. (2022). Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. Nat. Genet. *54*, 382–392. https://doi.org/10.1038/s41588-021-01006-7.
21. Speir, M.L., Bhaduri, A., Markov, N.S., Moreno, P., Nowakowski, T.J., Papatheodorou, I., Pollen, A.A., Raney, B.J., Seninge, L., Kent, W.J., and Haeussler, M. (2021). UCSC Cell Browser: visualize your single-cell data. Bioinformatics *37*, 4578–4580. https://doi.org/10.1093/bioinformatics/btab503.
22. Winkley, K., Banerjee, D., Bradley, T., Koseva, B., Cheung, W.A., Selvarangan, R., Pastinen, T., and Grundberg, E. (2021). Immune cell residency in the nasal mucosa may partially explain respiratory disease severity across the age range. Sci. Rep. *11*, 15927. https://doi.org/10.1038/s41598-021-95532-3.

23. Trump, S., Lukassen, S., Anker, M.S., Chua, R.L., Liebig, J., Thürmann, L., Corman, V.M., Binder, M., Loske, J., Klasa, C., et al. (2021). Hypertension delays viral clearance and exacerbates airway hyperinflammation in patients with COVID-19. Nat. Biotechnol. *39*, 705–716. https://doi.org/10.1038/s41587-020-00796-1.
24. Chiu, M.C., Li, C., Liu, X., Song, W., Wan, Z., Yu, Y., Huang, J., Xiao, D., Chu, H., Cai, J.P., et al. (2022). Human nasal organoids model SARS-CoV-2 upper respiratory infection and recapitulate the differential infectivity of emerging variants. mBio *13*, e0194422. https://doi.org/10.1128/mbio.01944-22.
25. Ahn, J.H., Kim, J., Hong, S.P., Choi, S.Y., Yang, M.J., Ju, Y.S., Kim, Y.T., Kim, H.M., Rahman, M.D.T., Chung, M.K., et al. (2021). Nasal ciliated cells are primary targets for SARS-CoV-2 replication in the early stage of COVID-19. J. Clin. Invest. *131*, e148517. https://doi.org/10.1172/JCI148517.
26. Hou, Y.J., Okuda, K., Edwards, C.E., Martinez, D.R., Asakura, T., Dinnon, K.H., 3rd, Kato, T., Lee, R.E., Yount, B.L., Mascenik, T.M., et al. (2020). SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. Cell *182*, 429–446.e14. https://doi.org/10.1016/j.cell.2020.05.042.
27. Ebert, P., Audano, P.A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M.J., Sulovari, A., Ebler, J., Zhou, W., Serra Mari, R., et al. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science *372*, eabf7117. https://doi.org/10.1126/science.abf7117.
28. Campos-Gómez, J., Fernandez Petty, C., Mazur, M., Tang, L., Solomon, G.M., Joseph, R., Li, Q., Peabody Lever, J.E., Hussain, S.S., Harrod, K.S., et al. (2023). Mucociliary clearance augmenting drugs block SARS-CoV-2 replication in human airway epithelial cells. Am. J. Physiol. Lung Cell Mol. Physiol. *324*, L493–L506. https://doi.org/10.1152/ajplung.00285.2022.
29. Joskova, M., Mokry, J., and Franova, S. (2020). Respiratory cilia as a therapeutic target of phosphodiesterase inhibitors. Front. Pharmacol. *11*, 609. https://doi.org/10.3389/fphar.2020.00609.
30. Jiao, J., and Zhang, L. (2019). Influence of intranasal drugs on human nasal mucociliary clearance and ciliary beat frequency. Allergy Asthma Immunol. Res. *11*, 306–319. https://doi.org/10.4168/aair.2019.11.3.306.
31. Scharner, J., and Aznarez, I. (2021). Clinical applications of single-stranded oligonucleotides: current landscape of approved and in-development therapeutics. Mol. Ther. *29*, 540–554. https://doi.org/10.1016/j.ymthe.2020.12.022.
32. Gay, N.R., Gloudemans, M., Antonio, M.L., Abell, N.S., Balliu, B., Park, Y., Martin, A.R., Musharoff, S., Rao, A.S., Aguet, F., et al. (2020). Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. Genome Biol. *21*, 233. https://doi.org/10.1186/s13059-020-02113-0.
33. Ravindra, N.G., Alfajaro, M.M., Gasque, V., Huston, N.C., Wan, H., Szigeti-Buck, K., Yasumoto, Y., Greaney, A.M., Habet, V., Chow, R.D., et al. (2021). Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes. PLoS Biol. *19*, e3001143. https://doi.org/10.1371/journal.pbio.3001143.
34. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. *41*, D991–D995. https://doi.org/10.1093/nar/gks1193.
35. Broadbent, L., Bamford, C.G.G., Lopez Campos, G., Manzoor, S., Courtney, D., Ali, A., Touzelet, O., McCaughey, C., Mills, K., and Power, U.F. (2022). An endogenously activated antiviral state restricts SARS-CoV-2 infection in differentiated primary airway epithelial cells. PLoS One *17*, e0266412. https://doi.org/10.1371/journal.pone.0266412.
36. Marcus, J.H., and Novembre, J. (2017). Visualizing the geography of genetic variants. Bioinformatics *33*, 594–595. https://doi.org/10.1093/bioinformatics/btw643.
37. Wang, M.C., and Bushman, B.J. (1999). Integrating results through meta-analytic review using SAS software (SAS Institute).
38. Cooper, H., Hedges, L.V., and Valentine, J.C. (2019). The handbook of research synthesis and meta-analysis (Russell Sage Foundation).

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| The COVID-19 Host Genetics Initiative (HGI) hospitalization GWAS B1 (hospitalized vs. non-hospitalized COVID-19; release 4) | GRASP[16] and HGI[17] | https://grasp.nhlbi.nih.gov/downloads/COVID19GWAS/10202020/COVID19_HGI_B1_ALL_20201020.b37.txt.gz |
| The COVID-19 HGI hospitalization GWAS B2 (hospitalized COVID-19 vs. general population; release 4) | GRASP[16] and HGI[17] | https://grasp.nhlbi.nih.gov/downloads/COVID19GWAS/10202020/COVID19_HGI_B2_ALL_leave_23andme_20201020.b37.txt.gz |
| Summary statistics of differential GWAS between HGI-B1 and HGI-B2 | This study | cheng, zhongshan (2023), "Differential association analysis between two COVID-19 hospitalization GWAS from HGI", Mendeley Data, V1, https://doi.org/10.17632/7bgym75bjx.1 |
| Airway epithelium upon SARS-CoV-2 infection | Ravindra et al.[36] | https://cells.ucsc.edu/covid19-bronch-epi/exprMatrix.tsv.gz; https://cells.ucsc.edu/covid19-bronch-epi/meta.tsv; https://cells.ucsc.edu/covid19-bronch-epi/UMAP.coords.tsv.gz |
| Nasal lifespan single cell atlas | Winkley et al.[22] | https://cells.ucsc.edu/lifespan-nasal-atlas/lifespan-nasal/exprMatrix.tsv.gz; https://cells.ucsc.edu/lifespan-nasal-atlas/lifespan-nasal/meta.tsv; https://cells.ucsc.edu/lifespan-nasal-atlas/lifespan-nasal/Integrated_UMAP.coords.tsv.gz |
| COVID-19 nasopharyngeal single cell RNA-seq | Trump et al.[27] | https://cells.ucsc.edu/covid-hypertension/exprMatrix.tsv.gz; https://cells.ucsc.edu/covid-hypertension/meta.tsv; https://cells.ucsc.edu/covid-hypertension/Seurat_umap.coords.tsv.gz |
| Bulk RNA-seq of nasal multiciliated epithelial cells infected with SARS-CoV-2 | GEO[34] | GSE199743 [35] |
| Software and algorithms | | |
| SAS OnDemand for Academics | SAS Institute Inc. | https://www.sas.com/en_us/software/on-demand-for-academics.html |
| UCSC Cell Browser | Speir ML et al.[21] | https://cells.ucsc.edu/ |
| Regeneron curated COVID-19 GWASs | Regeneron[20] | https://rgc-covid19.regeneron.com/results |
| GTEx Portal (version 8) | The GTEx Consortium[19] | https://gtexportal.org/home/ |
| Genetic Variants Browser | Marcus and Novembre[36] | https://popgen.uchicago.edu/ggv/ |
| COVID-19 GWAS Analyzer | This study | https://github.com/chengzhongshan/COVID19_GWAS_Analyzer |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for data and analysis codes will be addressed by the lead contact, Zhongshan Cheng (cheng.zhong.shan@gmail.com).

### Materials availability statement

No new unique reagents were generated in this study.

### Data and code availability

- The downloading links for COVID-19 GWAS summary statistics, all single-cell or bulk RNA-seq data that are published by other researchers are listed in the key resources table. The GWAS summary statistics are the COVID-19 Host Genetics Initiative (HGI) hospitalization GWAS B1 (hospitalized vs. non-hospitalized COVID-19; release 4) and the COVID-19 HGI hospitalization GWAS B2 (hospitalized COVID-19 vs. general population; release 4). The following single-cell or bulk RNA-seq data are used in the current study: airway epithelium upon SARS-CoV-2 infection, nasal lifespan single cell atlas, COVID-19 nasopharyngeal single cell RNA-seq, and the bulk RNA-seq of nasal multiciliated epithelial cells infected with SARS-CoV-2. The final differential GWAS summary statistics reported in this paper are provided in the key resources table with its downloadable link at Mendeley data. The above summary statistics will also be shared by the lead contact upon request.
- All original code has been deposited at github and is publicly available as of the date of publication. Downloadable links are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Our study did not involve experiments on patients or animals and our analyses were only conducted on other freely and de-identified GWAS summary statistics from HGI and GRASP databases. We followed the ethical guidelines of research at St. Jude Children's Research Hospital as well as in our collaborators' Universities or institutes. The eQTL analysis and single cell expression data were shared by other researchers and we cited their publications accordingly. When the sex information was included in the meta-data belonging to these data sets used in our study, we adjusted the analysis based on the sex assigned at birth to eliminate potential sex biases. The age or developmental stages of subjects were unavailable among these summary statistics and single cell expression data sets, thus we were unable to evaluate its impact on our analyses.

## METHOD DETAILS

### Pairwise comparison of association effect sizes

Summary statistics of two multi-ancestry COVID-19 hospitalization GWASs comprising samples from European, African, Asian, and other populations were obtained from GRASP[16] (https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx). These data were generated by the COVID-19 Host Genetics Initiative (release 4, generated on October 20, 2020).[17] The first HGI GWAS (HGI-B1) was designed by comparing 2,430 hospitalized COVID-19 cases with 8,478 non-hospitalized COVID-19 cases. While the second COVID-19 hospitalization GWAS (HGI-B2) was performed between hospitalized COVID-19 cases (n = 8,638) and general population controls (n = 1,736,547) (Figure S1). We used a differential Z-test to compare the effect sizes of each common SNP (minor allele frequency [MAF] > 0.01, imputation score > 0.6) between the two GWASs[16] with or without normalizing the effect of overlapped samples of hospitalized cases between HGI-B1 and HGI-B2 on the estimated variance, i.e., $gwas1.se^2 + gwas2.se^2$. The quality controls for the two GWAS summary statistics and the rationale for the Z-score normalization are explained in the following section.

### Normalization of differential effect sizes

The HGI consortia generated GWAS summary statistics by analyzing diverse populations. These analyses included quality control such as evaluation of genomic inflation factor and Hardy-Weinberg Equilibrium (HWE) filtering (www.covid19hg.org/results/r4/). Our analysis began from these summary statistics by focusing on high quality SNPs with minor allele frequencies > 0.01 and imputation scores > 0.6. We performed meta-analysis comparisons between the two GWASs that revealed SNPs which differed in the two GWASs and are specific to individuals with African ancestry. In detail, we evaluated the GWAS quality using Manhattan and QQ plots (see Figure S2) and normalized differential Z-scores to reduce the genome-wide correlation that deflated comparisons of the two HGI GWASs. The rationale for the normalization is demonstrated in the following two formulas that are suitable for differential effect size analysis where there are overlapped samples among the GWASs being compared. In the current analysis, overlap in hospitalized COVID-19 cases between the two HGI GWASs lead to correlated standard errors (se) of the effect size beta of each SNP. The se values for each SNP were estimated based on the number of effect alleles in the cases and controls in each GWAS,[37] i.e., se=sqrt(1/a + 1/b + 1/c + 1/d), where a, b, c, and d denote: (a) the number of COVID-19 cases carrying the risk allele; (b) the number of controls harboring the risk allele; (c) the number of COVID-19 cases not carrying the risk allele; and (d) the number of controls not harboring the risk alleles. Notably, the estimated variance, $gwas1.se^2 + gwas2.se^2$, in the formula 1 for each SNP is overestimated because gwas1.se and gwas1.se have the common components from COVID-19 cases and thus are correlated due to use of overlapped COVID-19 hospitalized cases in the two GWASs.

Formula 1 for calculating differential effect size of each SNP from HGI-B1 and HGI-B2:

$$\Delta Z - \text{score} = \frac{\text{gwas1}.\beta - \text{gwas2}.\beta}{\sqrt{(\text{gwas1}.\text{se}^2+\text{gwas2}.\text{se}^2)}}$$

$$\text{diff.zscore} = \Delta Z - \text{score}$$

$$P = \text{pnorm}(-|\text{diff.zscore}|) * 2$$

Formula 2 for determining normalized differential effect size of each SNP between HGI-B1 and HGI-B2. The updated formula considers the correlation (due to sample overlap) between the standard errors of each SNP from the two GWASs by normalizing differential Z-scores, i.e., setting mean at 0 and adjusting it by standard deviation genome-wide on differential Z-score among all SNPs. Without this adjustment, the sqrt(gwas1.se$^2$ + gwas2.se$^2$) is overestimated,[38] resulting in smaller differential z-score for each SNP and consequently leading to less significant *P*:

$$\Delta Z - \text{score} = \frac{\text{gwas1}.\beta - \text{gwas2}.\beta}{\sqrt{(\text{gwas1}.\text{se}^2+\text{gwas2}.\text{se}^2)}}$$

$$\text{diff.zscore} = \text{Normalization}(\Delta Z - \text{score})$$

$$P = \text{pnorm}(-|\text{diff.zscore}|) * 2$$

## Identification of ancestry-linked COVID-19 SNPs

The two HGI COVID-19 hospitalization GWASs included summary statistics of (HGI-B1) 9,384,545 and 12,844,271 SNPs (HGI-B2) with MAF > 0.01, and indels were excluded in the analysis. There were 9,138,337 SNPs overlapped between these two datasets. The genome-wide significance threshold for the ΔZ-score *P* without normalization was set at $P < 5\text{x}10^{-5}$ to select SNPs demonstrating suggestive differential COVID-19 association between the two HGI COVID-19 hospitalization GWASs. As many samples overlap between the two HGI GWASs (Figure S1), most SNPs show similar effect sizes in the two GWASs. Hence the ΔZ-score *P* threshold ($< 5\text{x}10^{-5}$) was only used to select candidate SNPs for downstream analysis. Next we evaluated these SNPs based on normalized ΔZ-score *P*. To evaluate the MAFs of the top SNPs among all global populations we used Genetic Variants Browser.[36] SNPs with independent COVID-19 association signals and absolute delta MAFs > 0.2 between ancestrally diverse populations, such as AFR vs. EUR, were selected for manual investigation. We then determined whether these ancestry-specific SNPs are eQTLs in GTEx Portal.[19] Ultimately, rs16831827 (MAF=0.54 in AFR population but MAFs < 0.05 in EUR, ASN, and AMR), representing an independent association signal and also being an eQTL of *MAP3K19*, *ZRANB3*, *MCM6*, and *DARS* based on the GTEx Portal,[19] was prioritized for further investigation. Because rs16831827 shows the most significant association to *MAP3K19* in testis tissue and is nominally associated with *MAP3K19* but not other genes in COVID-19-relevant lung tissue, we prioritized *MAP3K19* as the gene underlying the association of rs16831827 with COVID-19 hospitalization.

## Replication of rs16831827 in other COVID-19 GWASs

To determine whether the association of rs16831827 with COVID-19 hospitalization was reproducible and to evaluate whether other *MAP3K19* SNPs are associated with different COVID-19 phenotypes, we queried the SNP of *MAP3K19* in publicly available COVID-19 GWASs, including GRASP (https://grasp.nhlbi.nih.gov/COVID-19GWASResults.aspx),[16] the European population specific severe COVID-19 GWAS,[9] and the Regeneron curated COVID-19 GWASs for *MAP3K19* (https://rgc-covid19.regeneron.com/).[20]

## *MAP3K19* expression upon SARS-CoV-2 infection

We evaluated *MAP3K19* expression using multiple single cell RNA-seq datasets in the UCSC Cell Browser.[21] We focused on three datasets including a nasal lifespan atlas,[22] airway epithelial upon SARS-CoV-2 infection,[33] and COVID-19 nasopharyngeal cells.[23] The last dataset included healthy controls (n = 16), as well as severe (n = 23) and critical (n = 9) COVID-19 patients. We performed differential gene expression for *MAP3K19* across these three cohorts for each cell type with SAS OnDemand for Academics (https://www.sas.com/en_us/software/on-demand-for-academics.html). The proc GLM procedure was used to evaluate a general linear model of *MAP3K19* expression across each cell type and patient groups, adjusted by age, sex, and medication, as well as for coronary artery disease (CAD), cardiovascular disease (CVD), and hypertension. The lsmeans option was used to compare *MAP3K19* expression across different sample groups with Tukey's multiple adjustment and a significance threshold set at $P \leq 0.05$. We also used a bulk RNA-seq timecourse of nasal multiciliated epithelial cells infected with SARS-CoV-2 (GEO: GSE199743).[35] In this analysis, normalized read counts of *MAP3K19* were compared in a pairwise manner between treated samples and matched mock controls using paired t-test. For visualization, GTEx[19] data across 49 tissues was plotted using the UCSC Cell Browser to compare *MAP3K19* expression across tissues.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All pairwise differential comparisons conducted in the study were performed using the "lsmeans" statements with "TUKEY" adjustment method under the procedure "proc GLM" from SAS OnDemand for Academics. Other quantification analyses or visualizations if not exclusively mentioned were carried out using SAS OnDemand for Academics.