

## ARTICLE OPEN



# A translational genomics approach identifies *IL10RB* as the top candidate gene target for COVID-19 susceptibility

Georgios Voloudakis<sup>1,2,3,4,5,6,7</sup>✉, James M. Vicari<sup>1,2,3,4,5,6,36</sup>, Sanan Venkatesh<sup>1,2,3,4,5,6,7,36</sup>, Gabriel E. Hoffman<sup>1,2,3,4,5,6,36</sup>, Kristina Dobrindt<sup>1,3,4,5,6,8,9,36</sup>, Wen Zhang<sup>1,3,4,5,6</sup>, Noam D. Beckmann<sup>5,6,10</sup>, Christina A. Higgins<sup>1,11,12,13</sup>, Stathis Argyriou<sup>1,2,3,4,5,6</sup>, Shan Jiang<sup>1,3,4,5,6</sup>, Daisy Hoagland<sup>14,15,16</sup>, Lina Gao<sup>17,18</sup>, André Corvelo<sup>19</sup>, Kelly Cho<sup>20,21</sup>, Kyung Min Lee<sup>22</sup>, Jiantao Bian<sup>22,23</sup>, Jennifer S. Lee<sup>24,25</sup>, Sudha K. Iyengar<sup>26,27,28</sup>, Shih-Wen Luoh<sup>18,29</sup>, Schahram Akbarian<sup>1,4,5,8</sup>, Robert Striker<sup>30,31</sup>, Themistocles L. Assimes<sup>12,25</sup>, Eric E. Schadt<sup>5,6,32</sup>, Julie A. Lynch<sup>12,23</sup>, Miriam Merad<sup>33,34,35</sup>, Benjamin R. tenOever<sup>11,12,14,15,16</sup>, Alexander W. Charney<sup>1,3,5,6,8,10</sup>, Mount Sinai COVID-19 Biobank, VA Million Veteran Program COVID-19 Science Initiative, Kristen J. Brennand<sup>1,3,4,5,6,8,9</sup>, John F. Fullard<sup>1,2,3,4,5,6</sup> and Panos Roussos<sup>1,2,3,4,5,6,7</sup>✉

Recent efforts have identified genetic loci that are associated with coronavirus disease 2019 (COVID-19) infection rates and disease outcome severity. Translating these genetic findings into druggable genes that reduce COVID-19 host susceptibility is a critical next step. Using a translational genomics approach that integrates COVID-19 genetic susceptibility variants, multi-tissue genetically regulated gene expression (GRex), and perturbation signatures, we identified *IL10RB* as the top candidate gene target for COVID-19 host susceptibility. In a series of validation steps, we show that predicted GRex upregulation of *IL10RB* and higher *IL10RB* expression in COVID-19 patient blood is associated with worse COVID-19 outcomes and that in vitro *IL10RB* overexpression is associated with increased viral load and activation of disease-relevant molecular pathways.

npj Genomic Medicine (2022)7:52; <https://doi.org/10.1038/s41525-022-00324-x>

## INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), is the latest of the betacoronaviruses to pose a global health threat. Of the recent respiratory virus pandemics, SARS-CoV-2 demonstrates the highest transmissibility<sup>1</sup>. Despite the fact that the overwhelming majority of affected individuals have mild symptoms, infection-fatality risk in an urban area of a developed country (e.g., New York City) is still high, ranging from 1.4% for young adults (25–44 years old) to 19.1% for more susceptible older individuals (aged 75 years and older)<sup>2</sup>. Unexplained heterogeneity in susceptibility to the disease and severity of illness exists, even when accounting for known risk factors such as age<sup>3</sup>.

The COVID-19 Host Genetics Initiative (HGI)<sup>4,5</sup> coordinates a global effort to elucidate the genetic basis of COVID-19

susceptibility. Ongoing efforts have uncovered multiple risk loci for COVID-19 susceptibility; however, these risk variants only partly explain interindividual variability and, as many of the variants reside within noncoding regions of the genome, the formulation of testable hypotheses to elucidate their potential effects is challenging. To translate these genetic findings into novel therapeutics for COVID-19, we sought to prioritize druggable gene targets by developing a multidisciplinary translational genomics framework that integrates genetic studies of COVID-19 susceptibility, genotype-tissue expression datasets, and perturbation signature libraries. We provide evidence from in vitro, in vivo, and retrospective epidemiological studies that validate the association of the top candidate gene, interleukin 10 receptor subunit beta (*IL10RB*), with COVID-19 outcome severity. Overall, our study identified gene targets with direct translational value to

<sup>1</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>2</sup>Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>3</sup>Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>4</sup>Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>5</sup>Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>6</sup>Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>7</sup>Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA. <sup>8</sup>Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>9</sup>Department of Psychiatry, Yale University, New Haven, CT, USA. <sup>10</sup>Mount Sinai Clinical Intelligence Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>11</sup>Department of Microbiology, Grossman School of Medicine, New York University, New York, NY, USA. <sup>12</sup>Department of Medicine, Grossman School of Medicine, New York University, New York, NY, USA. <sup>13</sup>Vilcek Institute of Graduate Biomedical Sciences, New York University, New York, NY, USA. <sup>14</sup>Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>15</sup>Virus Engineering Center for Therapeutics and Research, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>16</sup>Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>17</sup>Biostatistics Shared Resources, Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. <sup>18</sup>VA Portland Health Care System, Portland, OR, USA. <sup>19</sup>New York Genome Center, New York, NY, USA. <sup>20</sup>VA Boston Healthcare System, Boston, MA, USA. <sup>21</sup>Division of Aging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>22</sup>VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT, USA. <sup>23</sup>Division of Epidemiology, University of Utah, Salt Lake City, UT, USA. <sup>24</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. <sup>25</sup>VA Palo Alto Health Care System, Palo Alto, CA, USA. <sup>26</sup>Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA. <sup>27</sup>Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA. <sup>28</sup>VA Northeast Ohio Healthcare System, Cleveland VA Medical Center, Cleveland, OH, USA. <sup>29</sup>Department of Medicine, Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. <sup>30</sup>Division of Infectious Diseases, Department of Medicine, University of Wisconsin, Madison, WI, USA. <sup>31</sup>William S. Middleton Memorial Veterans Hospital, Madison, WI, USA. <sup>32</sup>Sema4, Stamford, CT, USA. <sup>33</sup>Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>34</sup>Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>35</sup>Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>36</sup>These authors contributed equally: James M. Vicari, Sanan Venkatesh, Gabriel E. Hoffman, Kristina Dobrindt. ✉Lists of members and their affiliations appear in the Supplementary Information. ✉email: georgios.voloudakis@mssm.edu; panagiotis.roussos@mssm.edu

modulate host physiology and immune response, and increase resilience to SARS-CoV-2 infection.

## RESULTS

### Overview of the multidisciplinary translational genomics framework

We developed a translational genomics framework that integrates three major sources of data (GWAS, genotype-tissue expression datasets, and perturbagen signature libraries) to identify and validate susceptibility genes for targeted therapeutics (Fig. 1). We first integrated GWASs for COVID-19 phenotypes with multi-tissue transcriptomic imputation models to predict genetically regulated gene expression (GReX) changes associated with COVID-19 susceptibility (Fig. 1a; Output 1).

We developed and applied a gene prioritization approach (Fig. 1b; Output 2) that integrates GReX with shRNA signature libraries<sup>6</sup> to identify key genes whose expression: (a) is predicted to be dysregulated in COVID-19 susceptible individuals, and (b) can be targeted to reverse the transcriptome-wide gene expression dysregulation that is associated with COVID-19 susceptibility. From this prioritization step, we identified *IL10RB* as the top gene target candidate, which we then subjected to three validation steps. The first validation step was to increase phenotypic specificity by examining whether *IL10RB* GReX is associated with COVID-19 severity in patients that tested positive for SARS-CoV-2 (Fig. 1b: EHR validation). In addition, we performed a GReX-based phenome-wide association study (PheWAS) to further understand the effect of *IL10RB* on pre-existing relevant phenotypes (before the emergence of COVID-19). The second validation step sought to associate *IL10RB* gene expression in peripheral blood with COVID-19 severity in a patient cohort (Fig. 1b: in vivo validation). The third, and final, validation step involved isogenic manipulation of *IL10RB* gene expression in vitro to study its effect on SARS-CoV-2 viral load and transcriptional dysregulation (Fig. 1b: in vitro validation).

### COVID-19 transcriptome-wide association study

We performed a transcriptome-wide association study (TWAS) leveraging 2 transformed cell lines and 40 peripheral tissue models from two cohorts (GTEx: Genotype-Tissue Expression v8<sup>7</sup> and STARNET: Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task<sup>8</sup>;  $n = 16,738$  reliably imputed genes; Supplementary Table 1) by using GWAS summary statistics for COVID-19 phenotypes<sup>4</sup> (Supplementary Table 2). Overall, for COVID-19 phenotypes, we observed a high correlation among the imputed transcriptomes of different tissues (Supplementary Fig. 1), even though the imputed transcriptomes of the COVID-19 phenotypes are quite diverse (Supplementary Fig. 2; a more detailed description of the different phenotypes can be found in the Supplementary Information). 17 genes were significantly associated with COVID-19 infection and outcomes (FDR-adjusted  $p \leq 0.05$ ) when considering all COVID-19 phenotypes and 42 tissues: *CCR1*, *CCR2*, *CCR3*, *CCR5*, *CXCR6*, *DNPH1*, *DPP9*, *IFNAR2*, *IL10RB*, *IL10RB-AS1*, *KCNN3*, *KIF15*, *OAS1*, *OAS2*, *OAS3*, *PDE4A*, and *TMEM241*. Some of these genes were identified in more than one COVID-19 phenotype (e.g., *IL10RB* and *IFNAR2*), whereas others (e.g., *OAS2*) were only associated with one (Supplementary Fig. 3); the GWAS contributing the highest number of gene-trait associations is “hospitalized COVID vs. population” (B2 phenotype as per COVID-19 HGI; 13 out of the 17 genes are captured; Supplementary Fig. 3). We also observed that most gene-trait associations are detected in Blood (STARNET), and Mammary artery (STARNET), identifying 7 and 6 gene-trait associations, respectively (Supplementary Fig. 4). For the remainder of our analysis we thus focused on the B2 phenotype which we will refer to as “COVID-19 associated hospitalization” as: (1) it is the highest

powered GWAS, and (2) conceptually, it captures genetic determinants protecting individuals both from infection (since the control group could have been SARS-CoV-2 negative or nonhospitalized positive) and from severe outcomes of COVID-19. Unsurprisingly, B2 exhibits moderate GReX correlation with both these phenotypes (Supplementary Fig. 2).

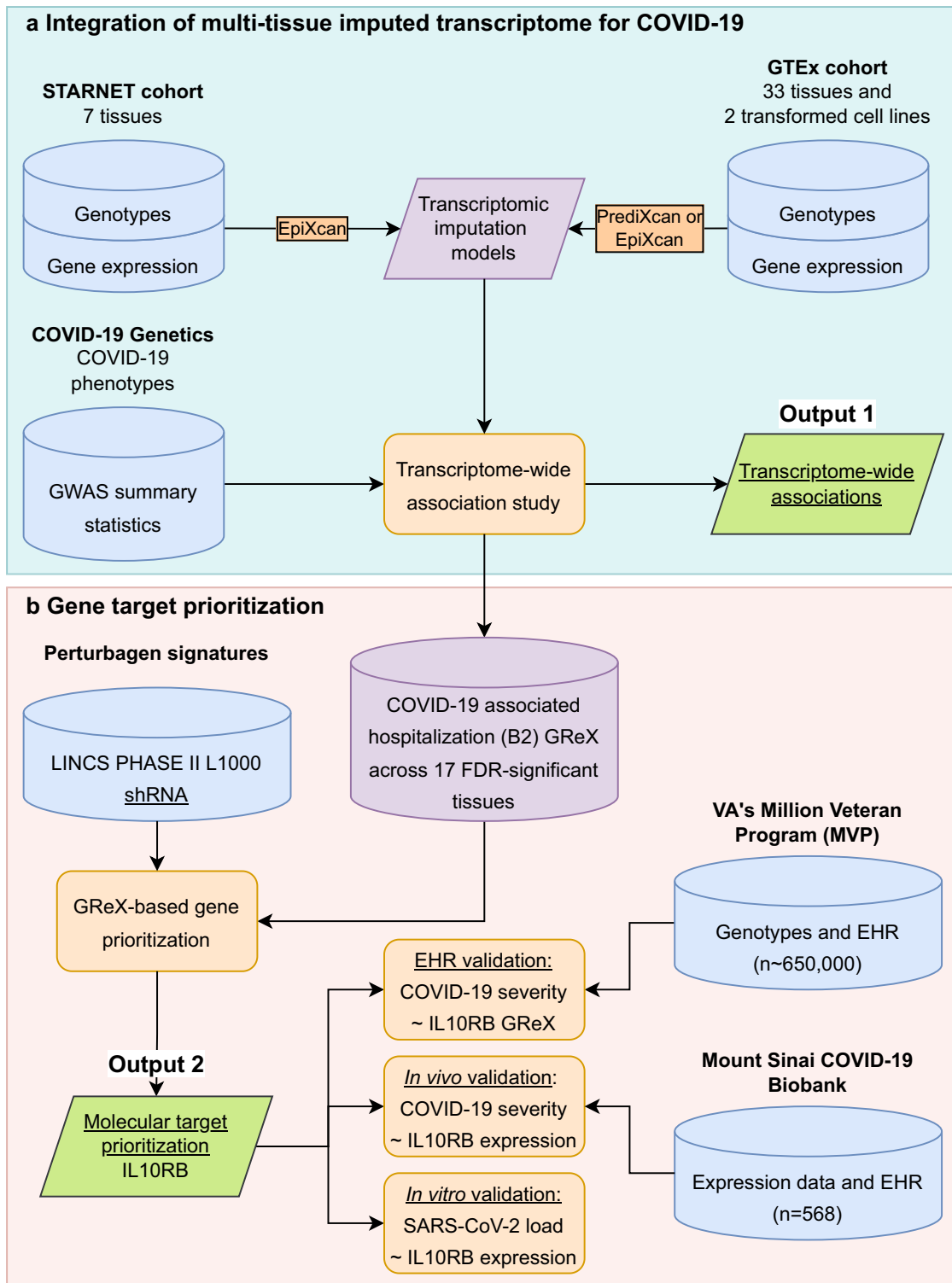
When considering only the COVID-19-associated hospitalization (B2) phenotype, we identified 88 gene-trait-tissue associations corresponding to 26 unique gene-trait associations (*AFF3*, *CCR1*, *CCR2*, *CCR3*, *CCR5*, *CDCP1*, *CRHR1*, *CXCR6*, *DPP9*, *IFNAR2*, *IL10RB*, *IL10RB-AS1*, *KANSL1-AS1*, *KCNN3*, *LINC02210*, *LRR37A2*, *LRR37A4P*, *MAPT*, *OAS1*, *OAS3*, *PDE4A*, *PIGK*, *PLEKHM1*, *PSMD2*, *THBS3*, *ZNF778*; FDR-adjusted  $p \leq 0.05$  while only considering B2; Supplementary Data 1–7 for all results split by COVID-19 phenotype) across 11 genomic regions (Fig. 2a for protein-coding genes). Testing a multitude of tissues across different cohorts allowed us to detect the consistent patterns of GReX dysregulation; for example, *IL10RB* is FDR significant in nine different tissue models (Fig. 2a). Significant genes (from at least one tissue) are enriched for pathways mainly involved in immune host response (Fig. 2b); a finding which is replicated when employing an LD-aware, competitive pathway TWAS method (JEPEGMIX2-P<sup>9</sup>; Supplementary Data 8). Overall, these results indicate that genetically-associated changes in genes involved in immune-related pathways predispose individuals to more severe COVID-19 outcomes.

### Gene target prioritization identifies *IL10RB* as a key regulator

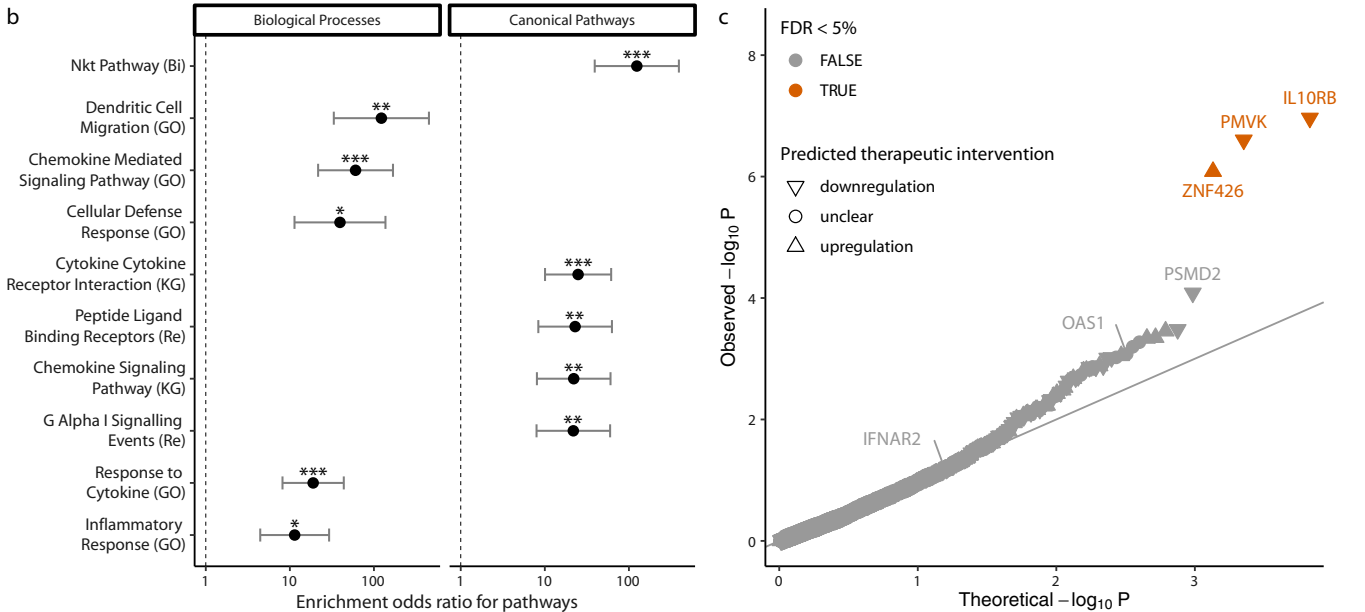
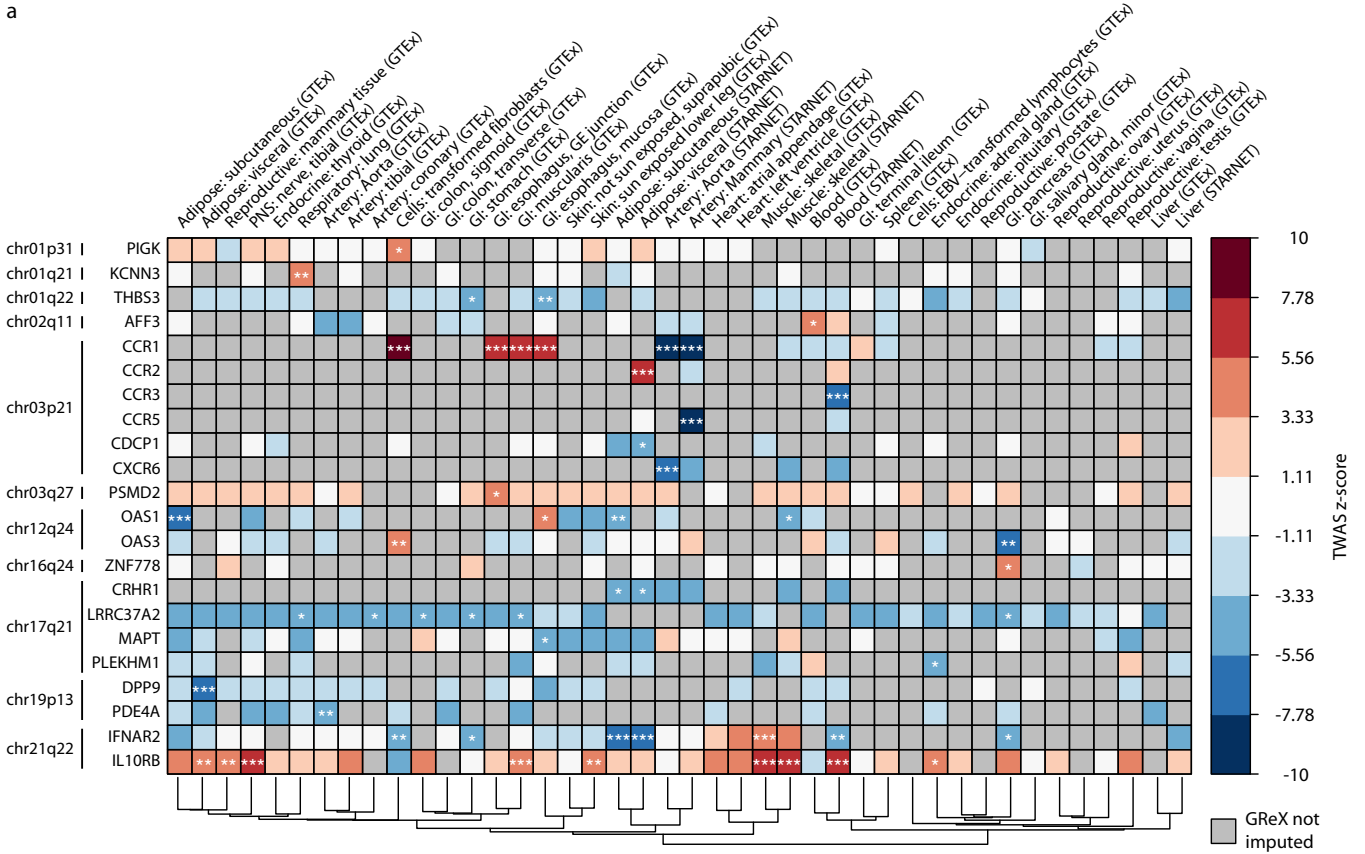
Towards prioritizing genes as putative molecular targets for intervention, we next aimed to identify genes whose perturbation is predicted to be therapeutic by antagonizing the GReX associated with COVID-19 susceptibility. Using a computational shRNA antagonism approach<sup>10,11</sup> (Supplementary Fig. 5), whose output we integrated with the TWAS findings, we identified *IL10RB* as the top most significant candidate for gene targeting (Fig. 2c). *IL10RB* was predicted to be significantly upregulated in individuals susceptible to COVID-19 hospitalization, with downregulation predicted to significantly antagonize the polygenetically driven gene expression differences associated with COVID-19 hospitalization (Supplementary Fig. 5). In mouse models, *IL10RB* overexpression was shown to increase susceptibility to lethal bacterial superinfections in the lungs via postviral increased IL-10 signaling, which dampens the immune response<sup>12</sup>, and by direct disruption of the lung epithelial barrier through increased expression of type III interferons (IFN $\lambda$ )<sup>13</sup>. *IFNAR2*, which is FDR significant in the TWAS and in the same locus as *IL10RB* (less than 2kbp from *IL10RB*), was not significant in this analysis, thereby nominating *IL10RB* as the most promising candidate in the locus while deprioritizing *IFNAR2*. Since this was a new approach, we tested both *IL10RB* and *IFNAR2* in our downstream analyses to determine how well our prioritization strategy performed in differentiating candidate genes in such close proximity to one another.

### Predicted upregulation of blood *IL10RB* is associated with COVID-19 severity and increased incidence of respiratory failure

The COVID-19-related hospitalization GWAS utilized a broadly-defined phenotype to increase cohort inclusion and sample size. To enhance granularity and phenotyping depth of *IL10RB* GReX association with COVID-19 associated hospitalization, we determined whether predicted upregulation of blood *IL10RB* was a good predictor of COVID-19 outcome severity and death in individuals who tested positive for SARS-CoV-2. We performed individual GReX imputation and association analysis in the VA’s Million Veteran Program (MVP)<sup>14</sup>, where the severity of COVID-19-related outcomes was deduced from EHR of COVID-19 positive cases ( $n = 23,226$ ; cohort characteristics in Supplementary Table 3). *IL10RB* GReX was associated with increased incidence of



**Fig. 1 Data-driven GREx (genetically regulated gene expression)-based approach for molecular target prioritization for COVID-19.** **a** Multi-tissue ( $n = 42$ ) transcriptome-wide association study using: (1) GWAS summary statistics from the COVID-19 Host Genetics Initiative for COVID-19 phenotypes, and (2) transcriptomic imputation models trained in the STARNET (Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task) and GTEx (Genotype-Tissue Expression) cohorts. **b** Gene target prioritization via integration of multi-tissue transcriptomes (17 FDR-significant tissues for COVID-19 associated hospitalization) with perturbational transcriptomic profiles from LINCS (library of integrated network-based cellular signatures) identified *IL10RB* as the top candidate. In a series of validation experiments we found that: (1) blood *IL10RB* genetically regulated gene expression (GREx) is associated with COVID-19 severity in the VA's Million Veteran Program—"EHR validation", (2) COVID-19 severity was associated with increased assayed *IL10RB* expression in patients' blood—"in vivo validation", and (3) increasing *IL10RB* expression resulted in higher SARS-CoV-2 viral load in two different model cell systems for SARS-CoV-2 infection and replication—"in vitro validation".



COVID-19 related death in individuals of European descent (EUR; logistic regression; OR = 1.13; Bonferroni-adjusted  $p = 0.01$ ;  $n = 14,262$ ) and in the trans-ethnic meta-analysis (logistic regression; OR = 1.12; Bonferroni-adjusted  $p = 0.002$ ;  $n = 23,226$ ) (Fig. 3a and Supplementary Fig. 6); *IFNAR2* GReX was not associated with COVID-19 death. However, both *IL10RB* and *IFNAR2* GReX are associated with more severe COVID-19 clinical outcomes in EUR, participants of African descent (AFR), and in the trans-ethnic meta-analyses (ordinal logistic regression; Fig. 3a and Supplementary Fig. 6; Supplementary Table 4). It is important to note the

association with severity or death controls for age, sex, Elixhauser's comorbidity score<sup>15</sup>, and ancestry-specific population structure (Fig. 3a).

To better understand the phenotypic variation linked with *IL10RB* and *IFNAR2* imputed expression, we next performed a phenome-wide association study (PheWAS) utilizing the *IL10RB* and *IFNAR2* GReX models in MVP (Fig. 3b; cohort characteristics in Supplementary Table 3; Supplementary Data 9 for a complete set of results). For *IL10RB*, among significant results, we found that COVID-19-related GReX dysregulation (higher *IL10RB* GReX in the

**Fig. 2 Transcriptome-Wide Association Study (TWAS) for COVID-19-associated hospitalization (hospitalized COVID vs. the general population) identifies associated genes, pathways, and aids in identification of druggable gene targets.** **a** FDR-significant TWAS results for COVID-19 susceptibility across all tissues. Box color indicates gene-trait-tissue association z-scores. Gray squares represent genes whose genetically regulated gene expression (GReX) could not be imputed. \*\*\*, \*\*, and \* correspond to FDR-adjusted *p*-values of association equal or less than 0.001, 0.01, and 0.05 respectively. Dendrogram on the bottom edge is shown with Ward hierarchical clustering for tissues based on all GReX (not just FDR-significant results). Displayed results are limited to protein-coding genes; cytogenetic location (at band level resolution) is also provided on the left of each gene. **b** Enrichment of COVID-19 TWAS associated genes for biological processes and canonical pathways. Odds ratio with 95% confidence interval (CI) is plotted for the significant enrichments of TWAS gene-trait associations from all tissues. Pathways are ranked based on estimated enrichment odds ratio. Analysis is limited to protein-coding genes and excludes genes residing in the major histocompatibility complex (MHC) on chromosome 6. Enrichments that are FDR significant are annotated as follows: \*, \*\*, and \*\*\* for FDR-adjusted *p* ≤ 0.05, 0.01, and 0.001 respectively; Fisher's exact test. **c** Prioritization of candidate gene targets to reverse TWAS gene-trait associations. *p*-value is estimated based on the joint statistic of two approaches ( $Z_{\text{combined}} = Z_{\text{TWAS}} + \text{pseudo } Z_{\text{GReX antagonist}}$ ) against the null. FDR-significant candidate genes are labeled orange. The direction of the predicted therapeutic intervention (upregulation or downregulation) is illustrated. *IL10RB*, *PMVK*, and *ZNF426* are FDR-significant target genes (*n* = 4302 imputed genes with reliable shRNA signatures).

blood) was positively associated with respiratory failure and tracheostomy complications and disorders of the circulatory system such as heart aneurysms and non-rheumatic mitral valve disorders, as well as cholecystitis without cholelithiasis and inflammatory conditions of the jaw. On the other hand, *IL10RB* was negatively associated with intracerebral hemorrhage, infections of the skin (e.g., lower limb cellulitis) and genitourinary system (e.g., cystitis and urethritis), type 1 diabetes, kidney disease (e.g., renal osteodystrophy), schizophrenia, functional disorders of the digestive system and bladder, and overall unspecified debility and sequela. COVID-19-related *IFNAR2* GReX dysregulation (lower *IFNAR2* GReX in the blood) shares some positive associations with *IL10RB*, including respiratory failure and heart aneurysms, but is independently associated with congestive heart failure, (chronic) renal failure and dialysis, delirium dementia, stomach cancer, and antisocial/borderline personality disorder. Furthermore, we also identified negative associations with cerebral ischemia, specific infections (cellulitis of foot, toe, and pyoderma, acute osteomyelitis), arthropathies, functional disorders of the digestive system and bladder, and overall unspecified debility and sequela. Overall, in addition to predisposing individuals to COVID-19-related hospitalization and outcome severity, increased *IL10RB* and decreased *IFNAR2* GReX are associated with respiratory failure independent of COVID-19 exposure.

### Increased *IL10RB* blood expression predicts worse COVID-19 outcome

Transcriptome imputation models can only partially explain the variance in observed *IL10RB* and *IFNAR2* gene expression ( $R^2_{\text{CV}}$  is 0.099 and 0.278, respectively). To further confirm the association of *IL10RB* with COVID-19 severity, we utilized blood gene expression profiling data from COVID-19 patients and controls at the Mount Sinai COVID-19 Biobank<sup>16</sup> (cohort characteristics in Supplementary Table 5 and Supplementary Table 6). We established a direct significant association between observed blood *IL10RB* expression and severe COVID-19 outcomes, including end-organ damage. The levels of *IL10RB* expression gradually increased with disease severity with a higher effect size in the most severe COVID-19 patient group (end-organ damage) against all other groups (Fig. 4a). Similar analysis for blood *IFNAR2* gene expression failed to demonstrate a robust association.

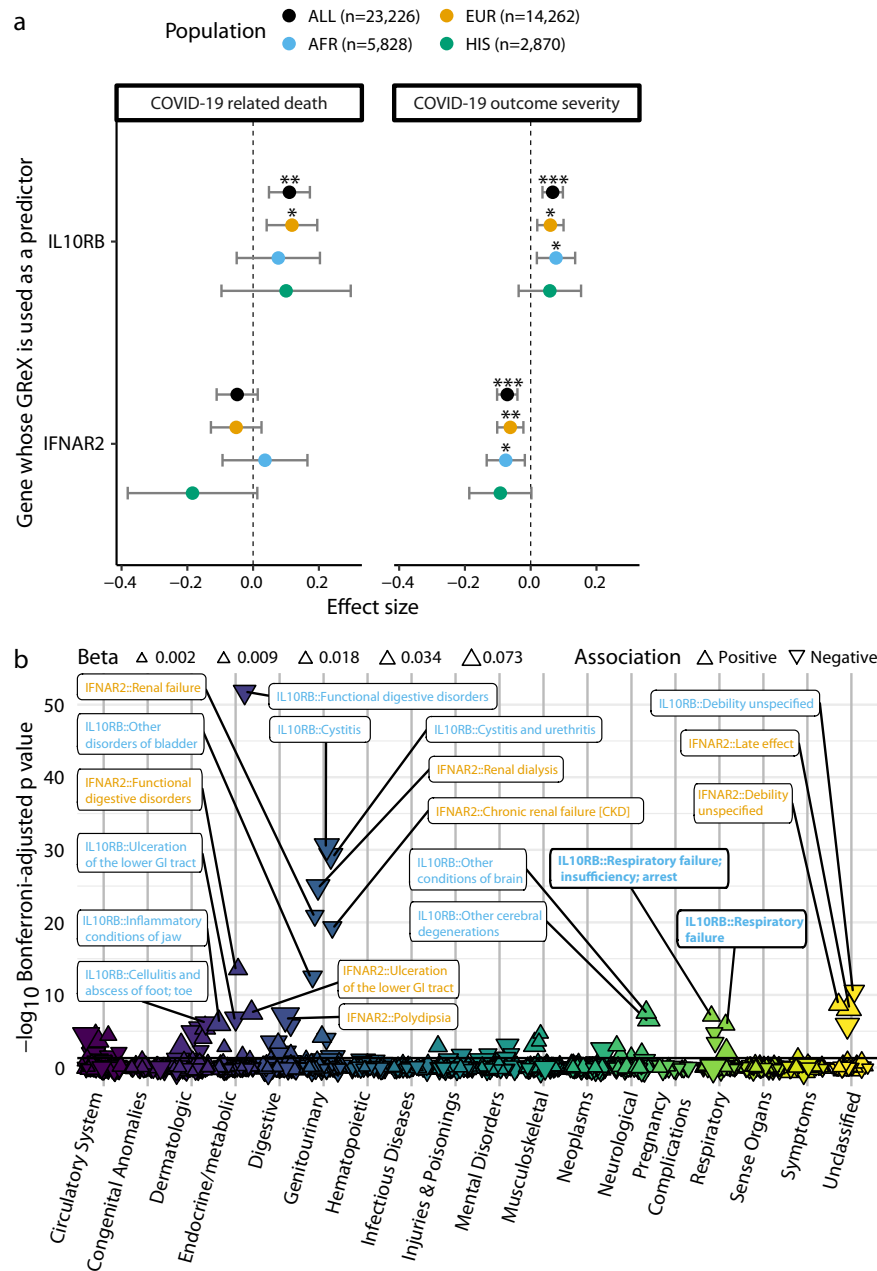
### *IL10RB* overexpression increases in vitro SARS-CoV-2 viral load

It has recently been shown that SARS-CoV-2 viral load in patients is associated with increased disease severity and mortality<sup>17</sup>. To explore the effect of *IL10RB* expression on SARS-CoV-2 viral load, we performed a series of in vitro experiments where we manipulated gene expression levels with a short hairpin (shRNA; for downregulation) and clustered regularly interspaced short palindromic repeats activation (CRISPRa; for upregulation) and quantified SARS-CoV-2 viral load (Fig. 4b). Overall, we tested (in technical triplicates) downregulation of *IL10RB* (and *IFNAR2*) by

shRNA (Supplementary Fig. 7; Supplementary Fig. 8) and upregulation of *IL10RB* by CRISPRa (by using four different guide RNAs; Fig. 4c). We performed these experiments in NGN2-glutamatergic postmitotic neurons<sup>18</sup> derived from human induced pluripotent stem cells (hiPSCs). These cells are permissive to SARS-CoV-2 and, therefore, can serve as a model cell system for SARS-CoV-2 infection<sup>19</sup>. Importantly, for the same COVID-19 phenotype (B2), *IL10RB* brain GReX has the same direction of predicted dysregulation (z-score = 5.53; dorso-lateral prefrontal cortex)<sup>20</sup> as the STARNET blood (z-score = 6.06) and GTEx lung (z-score = 3.02) models in this study. Indeed, we found that the gene expression dysregulation caused by SARS-CoV-2 infection in our model cell system mimics the transcriptional signatures corresponding to SARS-CoV-2 (and other betacoronaviruses) infection of a diverse range of cell types<sup>21</sup> (Fig. 4d; Supplementary Fig. 9).

We observed a significant increase in SARS-CoV-2 viral load (Fig. 4c; *p* = 0.0087; unpaired *t*-test) after *IL10RB* overexpression using four different guide RNAs (gRNAs) (Fig. 4c and Supplementary Table 7). Competitive pathway enrichment analysis demonstrated that overexpressing *IL10RB* in uninfected cells leads to the induction of COVID-19 relevant pathways implicated in vascular, immune system, and extracellular matrix processes (Fig. 4d), which were also activated by SARS-CoV-2 infection (Fig. 4d). Surprisingly, even in the absence of SARS-CoV-2, *IL10RB* overexpression leads to transcriptional changes reminiscent of betacoronavirus infection (Supplementary Fig. 9). In the rescue experiment, shRNA knockdown of *IL10RB* did not reduce *IL10RB* levels robustly, most likely due to low basal expression (Supplementary Fig. 7; Supplementary Table 8); it is worth noting that basal *IL10RB* expression has been shown to be low in all cell types from lung, heart, liver, and kidney<sup>22</sup>. We were also able to successfully knock down *IFNAR2* (higher basal expression; Supplementary Fig. 8; Supplementary Table 8), leading to a decrease in SARS-CoV-2 load. Interestingly, SARS-CoV-2 infection-induced expression of *IFNAR2* (Supplementary Fig. 8; Supplementary Table 9) but not *IL10RB* (Supplementary Fig. 7; Supplementary Table 9). This suggests that the increased *IFNAR2* (but not *IL10RB*) levels observed in the most severe group of COVID-19 patients (Fig. 4a) may reflect the increased likelihood of SARS-CoV-2 viremia in those patients<sup>17</sup>.

The effect of *IL10RB* and *IFNAR2* expression levels on SARS-CoV-2 viral load was replicated in A549-ACE2 cells, which constitute a more relevant cellular context for COVID-19 infection. A549-ACE2<sup>23</sup> is a human alveolar basal epithelial carcinoma cell line (A549) that constitutively expresses ACE2 (the host receptor required for SARS-CoV-2 entry) and is able to support SARS-CoV-2 infection and replication<sup>24</sup>. Specifically, we verified that *IL10RB* expression is positively correlated with SARS-CoV-2 viral load via siRNA-mediated knockdown of *IL10RB* and exogenous *IL10RB* overexpression (Pearson's *r* = 0.88, *p* =  $8.5 \times 10^{-4}$ ; Supplementary Fig. 10; Supplementary Table 10). Consistent with the findings above in NGN2 cells, we also found that *IFNAR2* levels are positively correlated with viral load (Pearson's *r* = 0.92, *p* =  $4.7 \times 10^{-5}$ ; Supplementary Fig. 10; Supplementary Table 10).

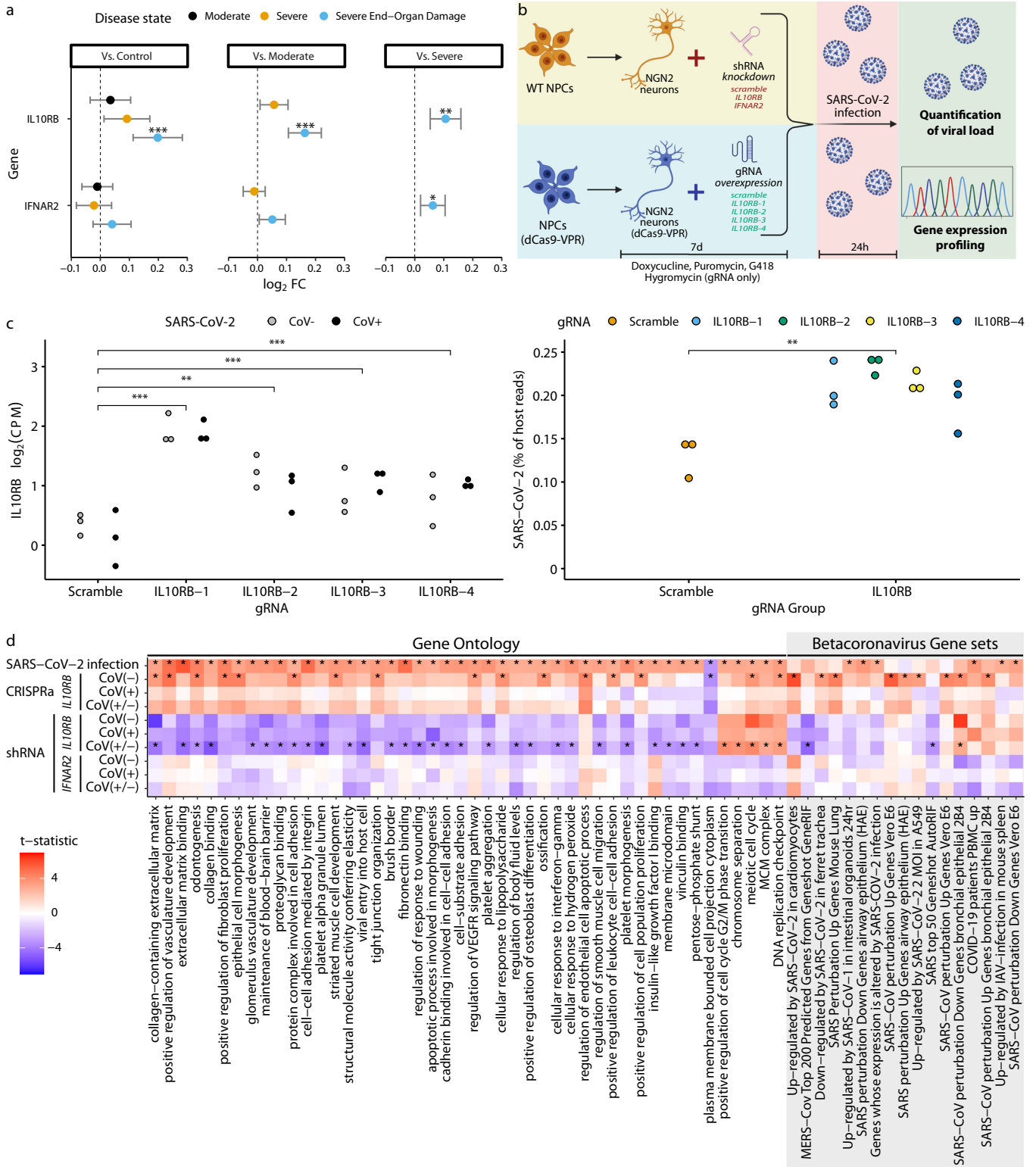


**Fig. 3 Association of blood *IL10RB* and *IFNAR2* genetically regulated gene expression (GrEx) with COVID-19-related outcomes and non-COVID-19 phenotypes.** **a** GrEx of *IL10RB* and *IFNAR2* was imputed in 23,216 individuals in the Million Veteran Program (MVP) cohort for whom COVID-19 outcome severity information was available. For COVID-19-related death (left panel) we checked the association of GrEx with the outcome of COVID-19 related death (4.8% of this cohort) under logistic regression models for *IL10RB* and *IFNAR2* GrEx, while adjusting for age, sex, Elixhauser's comorbidity score, and ancestry-specific population structure. For COVID-19 outcome severity, we applied an ordinal regression model (same predictors and covariates as above) using an outcome scale corresponding to mild (74.9% of the cohort), moderate (17%), severe (3.2%) COVID-19 related outcomes, and death (4.8%). EUR, AFR, and HIS refer to harmonized European, African and Hispanic ancestry respectively and the sample sizes are provided in the legend at the top. For both panels, a population of Asian ancestry ( $n = 266$ ) was included in the fixed effects meta-analysis (Population: "ALL" in the graph) but not plotted. \*\*\*, \*\*, and \* correspond to Bonferroni-adjusted association  $p$ -values (for  $n_{\text{genes}} \times n_{\text{outcomes}}$  for each population cohort) of equal or less than 0.001, 0.01, and 0.05 respectively. Error bars show 95% CI. **b** Phenome-wide association study (PheWAS) of *IL10RB* and *IFNAR2* blood GrEx for individuals of European descent in the MVP cohort ( $n = 296,407$ ). Phenotypes are grouped in categories shown in the x-axis, while the y-axis represents  $-\log_{10}$ (Bonferroni-adjusted  $p$ -values). Triangles represent data points for positive (pointing up) and negative (pointing down) association with GrEx; triangle size indicates the magnitude of the effect size and the color corresponds to the phenotype category. Only the top 20 associations are labeled (orange for *IFNAR2* and blue for *IL10RB*); full results are provided in Supplementary Data 9. The horizontal black line corresponds to Bonferroni-adjusted  $p = 0.05$ .

## DISCUSSION

Our multidisciplinary translational genomics framework integrates GrEx and perturbagen signature libraries to identify druggable gene targets for COVID-19 (Fig. 1). Transcriptomic imputation<sup>11,25,26</sup>

serves as the genomics backbone of this approach, and it trades off a part of SNP heritability in exchange for GrEx<sup>27</sup> which has translational potential. We first performed TWAS for all publicly available GWASs for different COVID-19 phenotypes from the



COVID-19 Host Genetics Initiative (Supplementary Table 2) and decided to base our downstream analyses on the COVID-19-associated hospitalization vs. population phenotype. We would have preferred to use a GWAS that compares severe, or hospitalized, COVID-19 cases against milder cases (infected controls), but these studies are not sufficiently powered (Supplementary Table 2); in retrospect, neither the flagship paper of the

COVID-19 Host Genetics Initiative<sup>5</sup>, nor other TWAS-based studies of COVID-19<sup>28,29</sup>, considered the GWASs based on these case/control definitions. Finally, when interpreting TWAS results, we have to take into account that there are differences in the cohort eligibility and sample collection between cohorts, which will be reflected in the TWAS results. For example, in the STARNET cohort, blood was collected from beating heart donors, whereas in GTEx

**Fig. 4** Increased *IL10RB* expression is associated with worse COVID-19 outcomes in vivo and increased SARS-CoV-2 viral load in vitro. **a** Increased *IL10RB* expression is associated with worse COVID-19 outcomes in vivo. \*, \*\*, and \*\*\* for FDR-adjusted  $p$  ( $FDR \leq 0.05, 0.01,$  and  $0.001$ , respectively). Error bars show 95% CI. **b** In vitro experimental overview. **c** CRISPRa gRNAs (*IL10RB-1*, *IL10RB-2*, *IL10RB-3*, and *IL10RB-4*) were used to overexpress *IL10RB* in hiPSC-derived NGN2-glutamatergic neurons. \*\*\*, \*\*, and \* correspond to  $p$ -values from the linear model of equal or less than 0.001, 0.01, and 0.05, respectively. For the SARS-CoV-2 viral load (right panel) we performed pairwise comparison with unpaired  $t$ -test; \*\*\*, \*\*, and \* correspond to  $p$  values equal to, or less than, 0.001, 0.01, and 0.05, respectively. **d** Competitive gene set enrichment analysis in hiPSC-derived NGN2 glutamatergic neurons. Each row represents a different experimental condition and each column a different gene set; the top row shows the effect of SARS-CoV-2 infection, while the remaining rows show the effect of gene manipulation (e.g., *IL10RB* vs. nontargeting siRNA) within a specific group (e.g., CoV(-): uninfected cells). The left side of the plot (Gene ontology gene sets; white background) indicates enrichment for canonical pathways and biological processes that are significantly enriched ( $FDR < 0.05$ ) in SARS-CoV-2 infection (top row), while the right side (Betacoronavirus Gene sets; gray background) illustrates enrichment for gene sets that correspond to betacoronavirus infections across different cell systems<sup>21</sup> (only significant results are shown;  $FDR < 0.05$ ).

the samples were obtained mostly (but not exclusively) post-mortem. Taking into account that *IL10RB* has been shown to be significantly ( $FDR = 4.55 \times 10^{-35}$ ) differentially expressed in pre- and postmortem blood samples in GTEx<sup>30</sup>, we should be cautious when, e.g., directly comparing findings for blood tissue between these two cohorts.

For our novel gene target prioritization approach, we integrated the multi-tissue TWAS results with an shRNA signature library<sup>6</sup> to identify genes whose perturbation can reverse the disease-associated GReX. This shRNA GReX antagonism approach identified *IL10RB* (21q22.11) as the most promising gene target and overcame traditional limitations of GWAS and TWAS analyses by identifying key genes within a gene cluster. Based on existing approaches, *IL10RB* would not have been the top candidate for further investigation. First, the index SNP for COVID-19 susceptibility in 21q22.11, rs13050728<sup>5</sup>, falls within an intronic region of *IFNAR2* (less than 24kbp from *IL10RB*). In addition, integrating genotype-gene expression datasets cannot identify the most likely causal gene in this locus since the index SNP is associated with gene expression changes of both *IFNAR2* and *IL10RB*<sup>5</sup>. Similarly, the genes can only be partially prioritized with targeted individual imputation (Fig. 3a; *IFNAR2* is not associated with COVID-19 death but is associated with COVID-19 severity) and cannot be prioritized with TWAS based on summary statistics (Fig. 2a) even when considering splicing<sup>28</sup> due to co-regulation<sup>31</sup> (Supplementary Fig. 11). Finally, Mendelian randomization from the VA Million Veteran Program COVID-19 Science Initiative identified both *IFNAR2* ( $p = 9.8 \times 10^{-11}$ ) and *IL10RB* ( $p = 2.3 \times 10^{-14}$ ) with good colocalization properties (PP.H4 > 0.8) and suggested a bigger role for *IFNAR2* based on PheWAS and pathway enrichment analysis for peak instrumental variants<sup>32</sup>. In our approach, *IL10RB* was prioritized primarily because its downregulation in cell lines reverses the GReX signature of COVID-19 and, secondarily, because it has a more uniform imputed transcriptional dysregulation across tissues (predominant downregulation; Fig. 2a, Supplementary Fig. 12). On the other hand, *IFNAR2* is expected to be upregulated in some tissues and downregulated in others<sup>28</sup> (Fig. 2a, Supplementary Fig. 12), e.g., there was a consistent, predicted, downregulation in adipose tissue and an opposing upregulation in muscle tissue (Supplementary Fig. 13). Unfortunately, technological innovations that would allow differential targeting of tissues are not readily available; thus, our gene target prioritization approach inherently penalizes opposing effects on GReX by integrating multiple tissues (Supplementary Fig. 5). Our study confirms that *IFNAR2* is important for COVID-19 susceptibility<sup>29,32</sup>; however, we prioritized *IL10RB* over *IFNAR2* as a suitable gene target for novel therapeutic development. To our knowledge, there is one other data-driven study that also nominated *IL10RB* from this locus by integrating information derived from single-cell/single-nucleus expression profiling of COVID-19 and healthy tissues (lung, ear, liver, and kidney)<sup>22</sup>.

Toward validating *IL10RB* as a suitable molecular target, we established a direct association between increased *IL10RB* GReX (Fig. 3a) and expression (Fig. 4a) with worse COVID-19 clinical

outcomes and death. Importantly, these results provide external validity for our findings beyond individuals of European ancestry (EA)<sup>33</sup> by performing ancestry-specific analysis for GReX (Fig. 3a) and leveraging a diverse patient cohort for expression profiling (Fig. 4a). Since available multi-tissue reference training datasets comprise mostly EA subjects, there are limitations when performing non-EA GReX association analyses; e.g., differential coverage of SNP predictors across different ancestry groups (Supplementary Table 11). Overall, EA-derived imputation models perform better in similar EA target populations<sup>33</sup>, and a decrease in relative power is observed when performing trans-ethnic association studies where gene-trait associations reach significance only in EA-specific and not the trans-ethnic analysis<sup>34</sup>. However, despite these limitations, we did observe good concordance among the ancestry groups in our study (Fig. 3a). Male sex is a strong risk factor for COVID-19; while sex-specific analyses were not performed due to power considerations, both of these analyses were adjusted for sex. Finally, isogenic manipulation of *IL10RB* in a model cell system for SARS-CoV-2 using NGN2 cells<sup>19</sup> revealed that inducing *IL10RB* expression led to priming of SARS-CoV-2 pathways (Fig. 4d) and increased SARS-CoV-2 viral load upon infection (Fig. 4c) which is associated with worse COVID-19 outcomes<sup>17</sup>. A limitation of this part of the study is the model cell system that we utilized (NGN2 cells); however, this was the only cell system available to us during the challenging times of the pandemic with the capability to perform isogenic overexpression. NGN2 cells have transcriptional regulatory pathways that may differ from e.g., innate immune cells; however, we found that SARS-CoV-2 infection in this cell system mimicked the transcriptional signatures corresponding to SARS-CoV-2 infection of a diverse range of cell types<sup>21</sup> (Fig. 4d; Supplementary Fig. 9) and identified SARS-CoV-2-relevant pathway dysregulation. Finally, the effect of *IL10RB* and *IFNAR2* expression on SARS-CoV-2 viral load was replicated in a lung alveolar cell line which is a more relevant model system for COVID-19<sup>24</sup>.

It was recently shown that *IL10RB* is significantly upregulated ( $z$ -score = 4.16) in ciliated cells in COVID-19 compared with healthy lung<sup>22</sup>. *IL10RB* serves as a receptor for members of the extended IL-10 family of cytokines (*IL10RA*<sub>2</sub>*IL10RB*<sub>2</sub> heterotetramer for IL-10; *IL10RB* heterodimers with *IL22RA1*, *IL20RA*, and *IFNLR1* for IL-22, IL-26 and *IFNLR1/IFNL2/IFNL3* respectively) which emerged before the adaptive immune response and are essential in modulating host defense mechanisms, especially in epithelial cells, to limit the damage caused by viral and bacterial infections<sup>35</sup>. This family of ligands has diverse and often contradicting roles in host response with an undetermined extent of functional crosstalk between them<sup>36</sup>; thus, further molecular dissection will be required to identify the causal signaling pathway(s) of *IL10RB* in COVID-19 susceptibility. IL-10 was found to be an important mediator of enhanced susceptibility to respiratory postviral bacterial super-infections in a mouse model<sup>12</sup>. IL-22 promotes antibacterial activity<sup>37</sup> and enhances tissue regeneration and wound healing<sup>38</sup>. IL-26 is poorly understood. Finally, *IFN-λ*(IL-28/IL29) are induced by a viral infection and show antiviral activity<sup>39,40</sup>; unsurprisingly,



in a COVID-19 mouse model, administration of IFN- $\lambda$ 1a led to diminished SARS-CoV-2 replication<sup>41</sup>. However, the participation of IFN- $\lambda$ s in damaging pro-inflammatory responses remains to be evaluated since recent mouse model studies showed that: (1) viral RNA-induced IFN- $\lambda$  production causes direct disruption of the lung epithelial barrier and increases susceptibility to bacterial super-infections<sup>13</sup>, and (2) IFN- $\lambda$  signaling aggravates viral infection by impairing lung epithelial regeneration<sup>42</sup>. Clinical outcomes from pegylated IFN- $\lambda$ 1a clinical trials against COVID-19 will provide evidence about the desired modulation direction of this pathway in COVID-19 treatment (phase II clinical trials: NCT04343976, NCT04354259, NCT04534673, and NCT04344600). Possible next steps include the evaluation of readily available IL-10 neutralizing antibodies (e.g., BT063, Biotest; NCT02554019) and IL-22 neutralizing antibodies (e.g., ILV-094/095, Pfizer). While the molecular dissection of this COVID-19 susceptibility pathway is important, transiently downregulating *IL10RB* with RNA interference<sup>43</sup> may be sufficient if a lung targeting approach is developed.

## METHODS

### Transcriptome-wide association study

**Transcriptomic imputation model construction.** Transcriptomic imputation models were constructed as previously described<sup>11</sup> for peripheral tissues of the GTEx v8<sup>7</sup> (excluding brain and tissues with  $n \leq 73$ ) and STARNET<sup>8</sup> cohorts (Supplementary Table 1; Fig. 1a). The genetic datasets of the GTEx and STARNET cohorts were uniformly processed for quality control (QC) steps before genotype imputation. We restricted our analysis to samples with European ancestry as previously described<sup>11</sup>. Genotypes were imputed using the University of Michigan server<sup>44</sup> with the Haplotype Reference Consortium (HRC) reference panel<sup>45</sup>. Gene expression information was derived from RNA-seq gene-level counts, which were adjusted for known and hidden confounds, followed by quantile normalization. For GTEx, we used publicly available, quality-controlled gene expression datasets from the GTEx consortium (<http://www.gtexportal.org/>). RNA-seq data for STARNET were obtained in the form of residualized gene counts from a previously published study<sup>8</sup>. For the construction of the transcriptomic imputation models, we used elastic net-based methods; when epigenetic annotation information<sup>46</sup> was available for a given tissue, we employed the EpiXcan<sup>11</sup> method to maximize power; when not available, we used the PrediXcan<sup>25</sup> method.

**COVID-19 phenotypes GWAS summary statistics.** Summary statistics for all 7 COVID-19 phenotypes (A1, A2, B1, B2, C1, C2, and D1; Supplementary Table 2) were obtained from the COVID-19 Host Genetics initiative<sup>4</sup> (Release 4; 2020-10-20; <https://www.covid19hg.org/results/r4/>).

**Multi-tissue transcriptome-wide association study (TWAS).** We performed the gene-trait association analysis as previously described<sup>11</sup>. Briefly, we applied the S-PrediXcan method<sup>26</sup> to integrate the COVID-19 GWAS summary statistics and the transcriptomic imputation models constructed above to obtain gene-level association results.

**Gene set enrichment analysis for TWAS results.** To investigate whether the genes associated with a given trait exhibit enrichment for biological pathways, we used gene sets from MsigDB 5.1<sup>47</sup> and filtered out nonprotein coding genes, genes located at MHC, as well as genes whose expression could not be reliably imputed. Statistical significance was evaluated with one-sided Fisher's exact test and the adjusted p values obtained by the Benjamini–Hochberg method<sup>48</sup>. We also performed separately LD-aware TWAS pathway enrichment analysis with JEPeGMIX2-P<sup>9</sup> v01.1.0 with SNP and gene set annotations v0.3.0.

### Genetically regulated gene expression (GReX-) based gene targeting approach

The gene targeting approach integrates genetically regulated gene expression (GReX) information (using the TWAS gene-trait-tissue association results) with a perturbagen signature library<sup>6</sup> (Fig. 1).

**Perturbagen Library used.** We used the LINCS Phase II L1000 dataset (GSE70138) perturbagen reference library<sup>6</sup>; specifically the shRNA signatures (gene expression changes after knocking down a gene). All inferred genes (AIG;  $n = 12,328$ ) were considered. Only “gold” signatures were considered.

**Imputed transcriptomes used.** We only considered GReX from 17 EpiXcan tissue models of the B2 phenotype that had significant TWAS results (FDR adjustment was applied to all COVID-19 phenotypes and tissues; Supplementary Table 1; the steps of the method are also outlined in Supplementary Fig. 5).

**Signature antagonism of trait GReX.** Each signature from the shRNA signature library (e.g., *IL10RB* shRNA treatment for 96 h in MCF7 cells) was assessed and ranked for its ability to reverse the trait-associated imputed transcriptomes using a previously published CDR method<sup>10</sup>.

**Summarization of the effect of signatures across tissues.** Signatures were grouped by perturbagen (shRNA), and we first tested whether the signatures for a specific perturbagen were more likely to be ranked higher or lower (Mann–Whitney U test). Then, we obtained a perturbagen-specific GReX antagonism pseudo-z-score which is defined as the negative Hodges–Lehmann estimator (of the median difference between that specific shRNA vs. the other shRNAs) divided by the standard deviation of the ranks of the shRNAs as follows:  $\text{pseudo } z_{\text{GReX antagonism}} = -\frac{\text{Hodges–Lehmann estimator}_{\text{perturbagen}}}{\text{SD average ranks of all perturbagens}}$ . A positive pseudo-z-score is interpreted as a potential therapeutic candidate, whereas a negative pseudo-z-score would suggest that the shRNA is not antagonizing the imputed transcriptome and is thus likely to exacerbate the phenotype. FDR is estimated with the Benjamini–Hochberg procedure<sup>48</sup>. All shRNAs were considered.

**Gene prioritization approach.** For prioritization, we estimated, for each gene, the p-value corresponding to the joint statistic of the two approaches ( $z_{\text{combined}} = z_{\text{TWAS}} + \text{pseudo } z_{\text{GReX antagonism}}$ ).

### IL10RB and IFNAR2 GReX association with COVID-19 severity and other phenotypes in the Million Veteran Program

**Cohort.** Within the broader cohort of the Million Veteran Program<sup>14</sup>, for the COVID-19 severity analysis we used all COVID-19-positive individuals as of March 11, 2021 ( $n_{\text{EUR}} = 14,262$ ,  $n_{\text{AFR}} = 5828$ ,  $n_{\text{HIS}} = 2870$ ,  $n_{\text{ASN}} = 266$ ; EUR: European; AFR: African; HIS: Hispanic; ASN: Asian) from MVP release 4. For the phenome-wide association study (PheWAS), we used individuals of European Ancestry ( $n = 296,407$ ) from MVP release 3. Ancestries were defined by the HARE method<sup>49</sup>. Genotypes used for imputation were filtered by Minor Allele Frequency ( $>0.01$ ), Variant level Missingness ( $<0.02$ ), as well as imputation  $R^2$  ( $>0.9$ ). We considered the MVP severity cohort an independent cohort from the GWAS since less than 7% (1519) of its participants were included in the COVID-19-related hospitalization GWAS (“B2\_ALL\_eur\_leave\_23andme”; release 4), comprising less than 7% and 0.2% of the GWAS's cases and total individuals, respectively. For the individual imputation, we used the EpiXcan tissue model of blood from the STARNET cohort for the following reasons: (1) as tissue, blood is relevant (immune cells) and accessible—allowing for testing and validation, and (2) as an imputation model, the blood (STARNET) is the most powerful

model (identifying the most FDR-significant gene-trait associations; Supplementary Fig. 4), allows the concurrent study of both *IL10RB* and *IFNAR2* (Fig. 2a), and is based on collection from beating heart donors (in contrast to the GTEx model which is based mostly on postmortem blood). These analyses were conducted under the protocol “VA Million Veteran Program COVID-19 Science Initiative”, which was approved by the Veterans Affairs Central Institutional Review Board and by the Research & Development Committee at the James J. Peters VA Medical Center.

**Phenotypes.** There are four COVID-19 severity levels: mild, moderate, severe, and death (see Supplementary Table 12 for more information on the phenotypic definition and counts).

**Transcriptomic imputation.** GRex for blood *IL10RB* and *IFNAR2* was calculated with the EpiXcan<sup>11</sup> Blood (STARNET) transcriptomic imputation model (Supplementary Table 1). For TWAS, we only considered SNPs with imputation  $R^2 \geq 0.3$ . Ancestry-specific principal component analysis was performed using the EIGENSOFT<sup>50</sup> v6 software as previously described<sup>51</sup>.

**GRex association with COVID-19 severity.** Associations of GRex and COVID-19 severity (Supplementary Table 12) were independently performed in each ancestry group. All associations were performed on scaled GRex with the following covariates: Elixhauser comorbidity index<sup>15</sup> for 2 years, sex, age, age squared (age<sup>2</sup>), and top 10 ancestry-specific principal components. To estimate the effect of GRex on COVID-19-related death, we used a logistic regression analysis (binomial distribution) where death was defined as “1”, while mild, moderate, and severe cases were defined as “0”. To estimate the effect of GRex on COVID-19-related outcome severity, we performed an ordinal logistic regression where COVID-19 severity was ordered as follows: mild, moderate, severe, and death. The ancestry-specific associations were meta-analyzed with a fixed-effect model using the inverse-variance method to estimate the effect of *IL10RB* and *IFNAR2* GRex in the total population.

**GRex PheWAS.** Phecodes<sup>52</sup> assigned to clinical encounters up to 2018 (predating the COVID-19 pandemic) were grouped into categories using Phecode Map v1.2 with manual curation for some uncategorized phecodes (as provided in Supplementary Data 10). All phecodes with at least one count in more than 25 individuals in the cohort were considered for further analysis. Association testing was performed between scaled GRex and counts of each phecode with a negative binomial distribution—this is the appropriate distribution to capture the full range of phecode counts since variance was higher than the mean phecode count in 99.95% of the phecodes evaluated (1840/1841; the only exception was “Other disorders of purine and pyrimidine metabolism”). The following covariates were used: total number of phecodes per individual, length of the record, sex, age, and top 10 ancestry-specific principal components. Phecodes with nonconvergent regression models were dropped. Significance was tested at the 0.05 false discovery rate (FDR) level.

### Gene expression profiling and EHR-based phenotyping in the Mount Sinai COVID-19 Biobank

Bio-specimens for this analysis were obtained from 568 individuals<sup>16</sup>; some of whom ( $n = 392$ ) contributed more than one bio-specimen. The complete biobank dataset and analyses will be presented in Thompson et al.<sup>53</sup>.

**RNA-seq gene expression profiling.** RNA was extracted from whole blood samples and used to prepare RNA-seq libraries which were quality controlled and sequenced, as previously described<sup>53,54</sup>. In

addition, we confirmed that no samples were mislabeled. RNA-seq reads were processed, quality controlled, and aligned to a reference genome as previously described<sup>53,54</sup>. After removing lowly expressed genes (keeping genes with counts per million  $>1$  in at least half of the number of control subjects in the cohort), we normalized the raw count data of the 21,194 remaining genes using voom<sup>55</sup> with dream<sup>56</sup> weights from the variancePartition R package<sup>57</sup>. Samples that failed to pass all quality controls were removed from further analyses. Principal component analysis was performed using the prcomp R function to explore covariate effect on gene expression variance genome-wide. Batch effect was calculated on a per gene basis using technical replicates sequenced in all batches, as previously described<sup>53</sup>. The following additional covariates were included in the model: Subject ID, number of days since first blood sample, RNA Library Prep Plate, Sex, DV200 Percent, and PICARD metrics PCT\_R2\_TRANSCRIPT\_STRAND\_READS, PCT\_INTRONIC\_BASES, WIDTH\_OF\_95\_PERCENT, MEDIAN\_5PRIME\_BIAS, MEDIAN\_3PRIME\_BIAS<sup>53</sup>. Cell type proportions were calculated with CIBERSORTx<sup>58</sup> using the LM22 reference<sup>53,59</sup>. COVID-19 severity-associated cell types are identified as having nonzero coefficients in a linear mixed model lasso procedure (R package glmmlasso) predicting COVID-19 severity<sup>53</sup>. Finally, we used dream<sup>56</sup> (a precision-weighted linear mixed model to consider repeated measures) for differential gene expression analysis while accounting for covariates identified by variancePartition<sup>57</sup>, above, as well as the proportions of COVID-19 severity associated with cell types identified above. A total of 6 DE signatures were generated (Severe end-organ damage Vs. Severe; Severe end-organ damage Vs. Moderate; Severe end-organ damage Vs. Control; Severe Vs. Moderate; Severe Vs. Control; Moderate Vs. Control). Multiple testing was controlled separately for each DE comparison accounting for the 21,114 genes tested using the false discovery rate (FDR) estimation method of Benjamini–Hochberg<sup>48</sup>.

**COVID-19 severity scale.** Phenotypic information was obtained by the EMR of the Mount Sinai Health System, which is reviewed by a screening team that includes practicing physicians. Each bio-specimen was associated, when possible—given the information in the EMR—with a COVID-19 severity measurement that corresponded to the time of collection. There are four levels of severity<sup>60</sup>: controls, moderate, severe, and severe end-organ damage, summarized in Supplementary Table 13.

**Ethics statement.** This study was approved by the Human Research Protection Program at the Icahn School of Medicine at Mount Sinai (STUDY-20-00341). All patients admitted to the Mount Sinai Health System were made aware of the research study by a notice included in their hospital intake packet. The notice outlined details of the specimen collection and planned research. Flyers announcing the study were also posted in the hospital, and a video was run on the in-room hospital video channel. Given the monumental hurdles of consenting sick and infectious patients in isolation rooms, the Human Research Protection Program allowed for sample collection, which occurred at the time of clinical collection and included at most an extra 5–10 cc of blood prior to obtaining research consent. Limited existing clinical data obtained from the medical record was collected and associated with the samples. As the research laboratory processing needed to begin proximal to sample collection, a portion of the data was generated prior to obtaining informed consent. During or after hospitalization, research participants and/or their legally authorized representative provided consent to the research study, including genetic profiling for research and data sharing on an individual level. In those circumstances where consent could not be obtained (13.8% of subjects, 0% of subjects who completed the post-discharge checklist), data already generated could continue to be used for analysis purposes only when not doing so would

have compromised the scientific integrity of the work. The data were not identifiable to the researchers doing the analyses.

### Manipulation of *IL10RB* and *IFNAR2* expression in NGN2 cells and their effect on SARS-CoV-2 viral load and transcriptional profiles

**Overview.** NGN2-glutamatergic neurons were derived from hiPSC-NPCs of donor NSB2607<sup>61</sup> as previously described<sup>62</sup>. gRNAs were designed using the CRISPR-ERA (<http://crispr-era.stanford.edu>) web tool and cloned into a lentiviral transfer vector using Gibson assembly<sup>18,62</sup>; shRNAs were ordered as glycerol stocks from Sigma. Wild type or dCas9 expressing NPCs were infected with rTA and NGN2 lentiviruses, as well as desired shRNA or CRISPRa lentiviruses, and differentiated for 7 days before SARS-CoV-2 infection with a multiplicity of infection (MOI) of 0.1 or mock infection for 24 h. After the completion of the experiment, RNA was isolated, quality metrics were obtained, and 200 ng of RNA was processed through a total RNA library prep using the KAPA RNA Hyper Prep Kit + RiboErase HMR kit (Roche, cat no: 8098140702) following the manufacturer's instructions. The resulting libraries were sequenced on the NovaSeq 6000 S4 flow cell (Illumina), obtaining 2 × 150 bp reads at 60 M reads per sample. An additional 20 ng of RNA was run on a SARS-CoV-2 targeted primer panel using AmpliSeq Library Plus and cDNA Synthesis for Illumina kits (Illumina, Cat no: 20019103 and 20022654). All samples were normalized, pooled, and run on the NovaSeq 6000 S4 in a 2 × 150 run targeting 750 K reads per sample. The STAR aligner v2.5.2a<sup>63</sup> was used to align reads to the GRCh38 genome (canonical chromosomes only) and Gencode v25 annotation. The module featureCounts<sup>64</sup> from the Subread package v1.4.3-p1<sup>65</sup> was used to quantify genes. RSeQC v2.6.1<sup>66</sup> and Picard v1.77<sup>67</sup> were used to generate QC metrics. Differential expression analysis was performed with limma<sup>68</sup> using the first two components of multidimensional scaling and RIN as covariates. Competitive gene set testing using sets from Gene Ontology<sup>69</sup> and the COVID-19 Drug and Gene Set Library<sup>21</sup> was performed with camera<sup>70</sup>. SARS-CoV-2 quantification was performed by taxonomically classifying short-read data with taxMaps<sup>71</sup>. The AmpliSeq approach confirmed the presence or absence of the virus in our samples. A more detailed version of this section can be found in the Supplementary Methods section.

**hiPSC-NPC culture and donor.** hiPSC-NPCs of line NSB2607 (male, 15 years old, European descent)<sup>61</sup> were cultured in hNPC media (DMEM/F12 (Life Technologies #10565), 1 × N-2 (Life Technologies #17502-048), 1 × B-27-RA (Life Technologies #12587-010), 20 ng/mL FGF2 (Life Technologies)) on Matrigel (Corning, #354230). hiPSC-NPCs at the full confluence (1–1.5 × 10<sup>7</sup> cells/well of a six-well plate) were dissociated with Accutase (Innovative Cell Technologies) for 5 min, spun down (5 min × 1000 × g), resuspended, and seeded onto Matrigel-coated plates at 3–5 × 10<sup>6</sup> cells/well. Media was replaced every 24 h for 4 to 7 days until the next passage.

**SARS-CoV-2 virus propagation and infections.** SARS-related coronavirus 2 (SARS-CoV-2), isolate USA-WA1/2020 (NR-52281) was deposited by the Center for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH. SARS-CoV-2 was propagated in Vero E6 cells in DMEM supplemented with 2% FBS, 4.5 g/L D-glucose, 4 mM L-glutamine, 10 mM Non-Essential Amino Acids, 1 mM Sodium Pyruvate and 10 mM HEPES. Virus stock was filtered by centrifugation using Amicon Ultra-15 Centrifugal filter unit (Sigma, Cat # UFC910096) and resuspended in viral propagation media. All infections were performed with either passage 3 or 4 SARS-CoV-2. Infectious titers of SARS-CoV-2 were determined by plaque assay in Vero E6 cells in Minimum Essential Media supplemented with 4mM L-glutamine, 0.2% BSA, 10 mM

HEPES and 0.12% NaHCO<sub>3</sub>, and 0.7% Oxoid agar (Cat #OXP0028B). All SARS-CoV-2 infections were performed in the CDC/USDA-approved BSL-3 facility of the Global Health and Emerging Pathogens Institute at the Icahn School of Medicine at Mount Sinai in accordance with institutional biosafety requirements.

**gRNA design and cloning and shRNAs.** gRNAs were designed using the CRISPR-ERA (<http://crispr-era.stanford.edu>) web tool. gRNAs were selected based on their specific locations at decreasing distances from the TSS as well as their lack of predicted off-targets and E scores (<http://crispr-era.stanford.edu>). For lentiviral cloning: synthesized oligonucleotides were phospho-annealed (37 °C for 30 min, 95 °C for 5 min, ramped-down to 25 °C at 5 °C per min), diluted 1:100, ligated into BsmB1-digested lentiGuide-Hygro-mTagBFP2 (addgene Plasmid #99374) and transformed into NEB10-beta E. coli, according to manufacturer's instructions (NEB # C3019H). shRNAs were ordered as glycerol stocks from Sigma (IL10RB # SHCLNG-NM\_000628; IFNAR2 # SHCLNG-NM\_000874). Gibson Assembly of Vectors: Unless specified, all cloning reagents were from NEB, and plasmid backbones were from Addgene (<https://www.addgene.org/>). Primers were synthesized by Thermo Fisher Scientific. All fragments were assembled using NEBuilder HiFi DNA Assembly Master Mix (NEB, no. E2621X). All assemblies were transformed into either DH5a Extreme Efficiency Competent Cells (Allele Biotechnology, no. ABP-CE-CC02050) or Stbl3 Chemically Competent E. coli (Thermo Fisher Scientific, no. C737303). Positive clones were confirmed by restriction digest and Sanger sequencing (GENEWIZ). The following vectors have been deposited at Addgene: lenti-EF1a-dCas9-VP64-Puro, lenti-EF1a-dCas9-VPR-Puro, lenti-EF1a-dCas9-KRAB-Puro, lentiGuide-Hygro-mTagBFP2, lentiGuide-Hygro-eGFP, lentiGuide-Hygro-dTomato, lentiGuide-Hygro-iRFP670, and pLV-TetO-hNGN2-Neo. lentiGuide-dTomato and lentiGuide-mTagBFP2-Hygro lentiGuide-Puro (Addgene, no. 52963) were digested with Mlu1 and BsiWI. dTomato was amplified from AAV-hSyn1-GCaMP6f-P2A-NLS-dTomato (Addgene, no.51085). HygroR sequence was amplified from lentiMS2-P65-HSF1\_Hygro (Addgene, no. 61426). mTagBFP2 was amplified from pBAD-mTagBFP2 (Addgene, no. 3463). The P2A self-cleaving peptide sequence was amplified using a reverse primer of HygroR and a forward primer of mTagBFP2. All sequences are provided in Supplementary Table 14.

**Lentiviral dCas9 effectors.** To engineer a lentiviral transfer vector that expresses dCas9: VP64-T2A-Puro (EF1a-NLS-dCas9(N863)-VP64-T2A-Puro-WPRE), dCas9:VP64-T2A-Blast (EF1a-NLS-dCas9(N863)-VP64-T2A-Blast-WPRE) (Addgene, no. 61,425) was digested with BsrGI and EcoRI. T2A-PuroR was amplified from pLV-TetO-hNGN2-P2A-eGFP-T2A-Puro (Addgene, no. 79823). Fragments were then assembled using NEBuilder HiFi DNA Assembly Master Mix (NEB, no. E2621). To engineer a lentiviral transfer vector that expresses dCas9:KRAB-Puro (EF1a-NLS-dCas9(N863)-KRAB-T2A-Puro-WPRE), dCas9:VP64-T2A-Blast (EF1a-NLS-dCas9(N863)-VP64-T2A-Blast-WPRE) (Addgene, no.61425) was first digested with BamHI and BsrGI. KRAB was then amplified from pHAGE-TRE-dCas9:KRAB (Addgene, no. 50917). Fragments were assembled using NEBuilder HiFi DNA Assembly Master Mix. dCas9:KRAB-Blast was digested with BsrGI and EcoRI, and T2A-PuroR was amplified from pLV-TetO-hNGN2-P2A-eGFP-T2A-Puro (Addgene, no. 79823). Fragments were then assembled using NEBuilder HiFi DNA Assembly Master Mix. To engineer a lentiviral transfer vector that expresses dCas9:VPR-Puro (EF1a-NLS-dCas9(D10A, D839A, H840A, and N863A)-VPR-T2A-Puro-WPRE), dCas9:VPR was first amplified from SP-dCas9-VPR (Addgene, no. 63798), and T2A-PuroR was amplified from pLV-TetO-hNGN2-P2A-eGFP-T2A-Puro (Addgene, no. 79823). dCas9:KRAB-T2A-Puro was

digested with BsiWI and EcoRI. Fragments were then assembled using NEBuilder HiFi DNA Assembly Master Mix.

**NGN2-glutamatergic neuron induction of shRNA and CRISPRa treated neurons**<sup>18,62</sup>. On day -1 NPCs were dissociated with Accutase Cell Detachment Solution for 5 min at 37 °C, counted, and seeded at a density of  $5 \times 10^5$  cells/well on Matrigel-coated 24-well plates in hNPC media (DMEM/F12 (Life Technologies #10565),  $1 \times N-2$  (Life Technologies #17502-048),  $1 \times B-27-RA$ , 20 ng/mL FGF2 (Life Technologies)) on Matrigel (Corning, #354230). On day 0, cells were transduced with rtTA and NGN2 lentiviruses as well as desired shRNA or CRISPRa viruses in NPC media containing 10  $\mu$ M Thiazovivin and spinfected (centrifuged for 1 h at 1000  $\times g$ ). On day 1, media was replaced, and doxycycline was added with 1  $\mu$ g/mL working concentration. On day 2, transduced hNPCs were treated with corresponding antibiotics to the lentiviruses (1  $\mu$ g/mL puromycin for shRNA, 1 mg/mL G-418 for NGN2-Neo). On day 4, the medium was switched to Brainphys neuron medium containing 1  $\mu$ g/mL dox. The medium was replaced every second day until SARS-CoV-2 (MOI of 0.1) or mock infection on day 7. The samples were harvested in Trizol (Invitrogen, Cat #15596026) 24 h later. RNA was isolated by phenol/chloroform extraction prior to purification using the RNeasy Mini Kit (Qiagen, Cat # 74106).

**Sequencing platform.** RNA samples were submitted to the New York Genome Center and, following an initial quality check, were normalized onto two different 96-well plates for a total RNA with RiboErase assay and a SARS-CoV-2 targeted assay. For the total RNA assay, 200 ng of RNA were normalized into a plate to be run through the KAPA RNA Hyper Prep Kit + RiboErase HMR (Roche, cat no: 8098140702). This total RNA prep followed the manufacturer's protocol with minor adjustments for automation on the PerkinElmer sciclone. Briefly, the RNA first goes through an oligo hybridization and rRNA depletion, followed by 1st and 2nd strand synthesis. The cDNA was then adenylated, and unique dual indexed adapters ligated onto the ends. Finally, samples were cleaned up, enriched, and purified. The final library was quantified by picogreen and ran on a fragment analyzer to determine the final library size. Samples were normalized, pooled and run on a NovaSeq 6000 S4 in a  $2 \times 150$  bp run format, targeting 60 M reads per sample. For the SARS-CoV-2 targeted assay, we used the AmpliSeq Library Plus and cDNA Synthesis for Illumina kits (Illumina, Cat no: 20019103 & 20022654). Briefly, 20 ng of RNA were reverse transcribed, the cDNA targets were then amplified with the Illumina SARS-CoV-2 research panel (Illumina, 20020496). The amplicons were partially digested, and AmpliSeq CD Indexes were ligated onto the amplicons. The library was then cleaned up and amplified. After amplification, there was a final clean-up, and the libraries were quantified, pooled, and run on a NovaSeq 6000 S4, obtaining  $2 \times 150$  bp reads.

**SARS-CoV-2 quantification.** Short-read data were taxonomically classified using taxMaps<sup>71</sup>. As part of the taxMaps pipeline, reads were processed prior to mapping. Adapter sequences and low-quality ( $Q < 20$ ) bases were trimmed out, and low complexity reads discarded. The remaining reads were then concurrently mapped against (1) the phiX174 reference genome (NC\_001422.1); (2) the SARS-CoV-2 reference genome (NC\_045512.2); and (3) a combined index encompassing the entire NCBI's nt database, RefSeq archaeal, bacterial, fungal, protozoan and viral genomes, as well as a selection of RefSeq model organism genomes, including the human GRCh38 reference<sup>72</sup>, to produce the final classification. Given that some human sequences of ancestral origin (that constitute variation between individuals) are absent from the GRCh38 reference, a small percentage of human reads usually maps to other primate genomes and, consequently, was classified as such. To obtain

more accurate estimates of the human content in these samples, all reads classified as "primate" were considered of human origin and reclassified accordingly. SARS-CoV-2 viral load was determined as the number of SARS-CoV-2 reads over the host (human) reads.

**Competitive gene set testing.** Competitive gene set testing using sets from Gene Ontology<sup>69</sup> and the COVID-19 Drug and Gene Set Library<sup>21</sup> was performed with camera<sup>70</sup>. First, we performed differential expression analysis with limma<sup>68</sup> using the first two components of multidimensional scaling and RIN as covariates to identify the signature of SARS-CoV-2 infection in our cells while adjusting for other treatments. We then performed competitive gene set enrichment analysis for all gene ontology and betacoronavirus gene sets ( $n = 18,553$ ). For gene ontology datasets, we kept all significantly enriched gene sets ( $FDR < 0.05$ ) and kept those with a Jaccard index less than 0.2. For the betacoronavirus gene sets, we kept all the gene sets and filtered them based on a Jaccard index of 0.2. The combined SARS-CoV-2 gene set collection with the two datasets above (significantly pruned gene ontology and all pruned betacoronavirus) was used for all the following competitive gene set testing except as otherwise indicated ( $n = 296$ ). Thus, in Fig. 4d, enrichment analysis is run across the whole exploratory dataset ( $n = 18,553$ ) for SARS-CoV-2 infection (first row), whereas for all other conditions, we are only exploring the combined SARS-CoV-2 gene set collection ( $n = 296$ ).

#### Manipulation of *IL10RB* and *IFNAR2* expression in A549-ACE2 alveolar cells and their effect on SARS-CoV-2 viral load

**Overview.** ACE2-expressing A549 cells (A549-ACE2), a gift from Brad Rosenberg<sup>23</sup>, were either transfected with siRNA, or transduced with the TetOne inducible system prior to infection with SARS-CoV-2. For knockdown, A549-ACE2 cells were transfected with pooled siRNAs (Dharmacon) 48 h prior to SARS-CoV-2 infection. For overexpression, A549-ACE2 cells were transduced using the TetOne Inducible Expression System (Takara Bio) with lentivirus-containing TetOne-eGFP, -*IL10RB*, or -*IFNAR2* for 48 h followed by puromycin selection. To induce gene expression, A549-ACE2 TetOne-eGFP, -*IL10RB*, or -*IFNAR2* cells were treated with either 0 ng/mL or 100 ng/mL doxycycline for 24 h before SARS-CoV-2 infection. Cells were either mock-infected or infected with media containing SARS-CoV-2 for a multiplicity of infection (MOI) of 0.02. The cells were incubated for 48 h, and then the cells were harvested, and the virus was inactivated by Trizol (Invitrogen) or RIPA lysis buffer for safe removal from the BSL-3 facility. Lysates were stored at  $-80$  °C until further analysis. RNA was isolated from Trizol, cleaned using the Qiagen RNeasy Mini Kit (Cat # 74106), and quantified by QuBit. A starting input of 500 ng was used to prepare cDNA via the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Cat # 4368813). Real-time qPCR was performed using TaqMan probes (Thermo Fisher Scientific) for the SARS-CoV-2 S (vi07918636\_s1), *IL10RB* (hs00175123\_m1), *IFNAR2* (hs01022059\_m1) and control gene, *GAPDH* (hs02786624\_g1). qPCR reactions were performed on the Applied Biosystems QuantStudio 5 and analyzed using the  $-\Delta\Delta C_t$  method for fold change expression to validate genetic manipulation and quantify SARS-CoV-2 infection. A more detailed version of this section can be found in the Supplementary Methods section.

**Plasmids.** The pLVX.TetOne-2xstrept-eGFP plasmid was a gift from Nevan Krogan<sup>73</sup>. pLVX.TetOne-*IL10RB* and pLVX.TetOne-*IFNAR2* were cloned using the parental pLVX-TetOne vector (Takara Bio, Inc.) and cDNA encoding human *IL10RB* or human *IFNAR2* (IDT) with the InFusion HD Cloning Kit (Takara Bio, Inc.). Clones were sequence-verified by sanger sequencing (Psomagen). Sequences are provided in Supplementary Table 14.

**Cells.** HEK 293 T cells, a human kidney epithelial cell line (HEK 293 T/17, ATCC®, CRL-11268); and Vero E6 cells (Vero 76, clone E6, Vero E6, ATCC® CRL-1586; for SARS-CoV-2 propagation), an African Green Monkey kidney epithelial cell line, were authenticated by ATCC. A monoclonal ACE2-expressing A549 cell line (A549-ACE2) was a gift from Brad Rosenberg<sup>23</sup>. All cell lines were cultured under humidified, 5% CO<sub>2</sub>, 37 °C conditions in complete DMEM (10% v/v fetal bovine serum (FBS, Thermo Fisher Scientific) and 100 I.U. penicillin and 100 µg/mL streptomycin (Pen/Strept, Corning) in Dulbecco's Modified Eagle Medium (DMEM, Corning)). Cells were confirmed negative for mycobacteria monthly (Lonza).

**Viruses.** SARS-related coronavirus 2 (SARS-CoV-2) isolate USA-WA1/2020 (NR-52281) was obtained from BEI Resources, NIAID, NIH. Virus stocks were grown, processed, and titrated as previously described<sup>74</sup>. The supernatant was collected 30 h postinfection and concentrated through a 100 kDa centrifugal filter unit (Amicon). All work with live SARS-CoV-2 was performed in the CDC/USDA-approved biosafety level 3 (BSL-3) facility of the NYU Grossman School of Medicine in accordance with institutional guidelines.

**Generation and induction of A549-ACE2 TetOne cell lines.** In a 6-well format, for each well,  $1.5 \times 10^6$  HEK 293 T cells were transfected with a mixture of 0.75 µg pLVX.TetOne-strept-eGFP, pLVX.TetOne-IL10RB, or pLVX.TetOne-IFNAR2; 0.75 µg gag/pol lentivirus packaging plasmid (Takara Bio Inc.), and 0.12 µg vesicular stomatitis virus G plasmid ((Takara Bio Inc.) in Opti-MEM (Corning) and Lipofectamine 2000 transfection reagent (Invitrogen) at a µg of DNA: µL of reagent ratio of 1:3. Twenty-four hours post-transfection, media was replaced with complete DMEM. Forty-eight hours post-transfection, cell supernatants were filtered through a 0.22 µm syringe filter.  $4 \times 10^6$  A549-ACE2 cells in a six-well format were then transduced per well, with 0.7 mL filtered lentivirus-containing supernatant diluted to a total of 2 mL with complete DMEM supplemented with 10 µg/mL polybrene (Corning). Forty-eight hours post-transduction, the supernatant was removed, and fresh complete DMEM containing 2.5 µg/mL puromycin was added to the transduced cells. Forty-eight hours post-puromycin treatment, cells were expanded for use in experiments. Cells were then maintained in 2.5 µg/mL puromycin when in culture. To induce gene expression prior to infection,  $2 \times 10^5$  A549-ACE2 TetOne-eGFP, -IL10RB, or -IFNAR2 cells were plated in a 24-well format in complete DMEM containing 0 ng/mL or 100 ng/mL doxycycline. Twenty-four hours post-induction, cells were infected as described below.

**siRNA transfections.**  $2 \times 10^5$  A549-ACE2 cells in a 24-well format were transfected per well with 0.5 µL 10 µM pooled siRNAs (Dharmacon) in 50 µL Opti-MEM with 1.5 µL RNAiMAX (Invitrogen). The transfection mix was incubated at room temperature for 15 min prior to addition to cells. Forty-eight hours post-transfection, cells were infected as described below. Sequences are provided in Supplementary Table 14.

**SARS-CoV-2 infections.** Media was removed from cells and mock-infected with 0.5 mL infection media (DMEM, 2% FBS, P/S) or infected with 0.5 mL infection media containing enough SARS-CoV-2 for a multiplicity of infection (MOI) of 0.02. The cells were then incubated under humidified, 5% CO<sub>2</sub>, 37 °C conditions. Forty-eight hours postinfection, media was removed from cells, and 0.5 mL Trizol (Invitrogen) or 0.15 mL RIPA lysis buffer (50 mM Tris HCL, 150 mM NaCl, 0.5% v/v NP-40, 1% v/v Triton X-100, 0.1% v/v SDS, Pierce protease inhibitors) was added to each well. Cells were then incubated at 4 °C for 15 min. Lysates were then transferred to tubes and removed from the BSL-3 facility. Lysates were stored at -80 °C until they were analyzed.

**SARS-CoV-2 quantification.** RNA was isolated from Trizol (Invitrogen) using the manufacturer-provided protocol. Isolated RNA was further cleaned using the Qiagen RNeasy Mini Kit (Cat # 74106) and supplemented with RNase inhibitor (Takara, Cat # 2313 A) 5% by volume. RNA was quantified by QuBit, and a starting input of 500 ng was used to prepare cDNA via the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Cat # 4368813). TaqMan probes for the SARS-CoV-2 S protein (vi07918636\_s1), IL10RB (hs00175123\_m1), IFNAR2 (hs01022059\_m1), and control gene, GAPDH (hs02786624\_g1), were acquired through Thermo Fisher (Cat # 4331182). Real-time PCR was performed in triplicate using 2 ng of cDNA per reaction and the Applied Biosystems TaqMan Gene Expression Mix (Cat # 4369016). Reactions were run using the Applied Biosystems QuantStudio 5, and SARS-CoV-2 was quantified using the delta-delta Ct method.

**Statistical analysis.** For pairwise comparisons, unpaired t-test was used, as indicated in Supplementary Table 10. For correlation analysis of *IL10RB* and *IFNAR2* levels with SARS-CoV-2 viral load, Pearson correlation analysis was used.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

All data associated with this study are present in the main text or the Supplementary Data files. Sequencing data from the in vitro experiments (NGN2 cells) have been uploaded to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo>) database under accession number GSE180622. Our study only does targeted replication in the Mount Sinai COVID-19 Biobank; the complete biobank dataset and analyses will be presented in Thompson et al.<sup>53</sup>.

Received: 7 August 2021; Accepted: 19 August 2022;

Published online: 05 September 2022

### REFERENCES

- Petersen, E. et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **20**, e238–e244 (2020).
- Yang, W. et al. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *Lancet Infect. Dis.* **21**, 203–212 (2021).
- Pereira, N. L. et al. COVID-19: understanding inter-individual variability and implications for precision medicine. *Mayo Clin. Proc.* **96**, 446–463 (2021).
- COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).
- COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).
- Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17 (2017).
- GTEX Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
- Franzén, O. et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
- Chatzinakos, C. et al. TWAS pathway method greatly enhances the number of leads for uncovering the molecular underpinnings of psychiatric disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **183**, 454–463 (2020).
- So, H.-C. et al. Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat. Neurosci.* **20**, 1342–1349 (2017).
- Zhang, W. et al. Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat. Commun.* **10**, 3834 (2019).
- van der Sluijs, K. F. et al. IL-10 is an important mediator of the enhanced susceptibility to pneumococcal pneumonia after influenza infection. *J. Immunol.* **172**, 7603–7609 (2004).
- Broggi, A. et al. Type III interferons disrupt the lung epithelial barrier upon viral recognition. *Science* **369**, 706–712 (2020).

14. Gaziano, J. M. et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
15. van Walraven, C., Austin, P. C., Jennings, A., Quan, H. & Forster, A. J. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med. Care* **47**, 626–633 (2009).
16. Charney, A. W. et al. Sampling the host response to SARS-CoV-2 in hospitals under siege. *Nat. Med.* **26**, 1157–1158 (2020).
17. Fajnzylber, J. et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat. Commun.* **11**, 5493 (2020).
18. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
19. Dobrindt, K. et al. Common genetic variation in humans impacts in vitro susceptibility to SARS-CoV-2 infection. *Stem Cell Rep.* **16**, 505–518 (2021).
20. Fullard, J. F. et al. Single-nucleus transcriptome analysis of human brain immune response in patients with severe COVID-19. *Genome Med.* **13**, 118 (2021).
21. Kuleshov, M. V. et al. The COVID-19 drug and gene set library. *Patterns (N. Y.)* **1**, 100090 (2020).
22. Delorey, T. M. et al. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113 (2021).
23. Daniloski, Z. et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105.e16 (2021).
24. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e9 (2020).
25. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
26. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
27. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).
28. Pathak, G. A. et al. Integrative genomic analyses identify susceptibility genes underlying COVID-19 hospitalization. *Nat. Commun.* **12**, 4569 (2021).
29. Ma, Y. et al. Integrative genomics analysis reveals a 21q22.11 locus contributing risk to COVID-19. *Hum. Mol. Genet.* **30**, 1247–1258 (2021).
30. Ferreira, P. G. et al. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nat. Commun.* **9**, 490 (2018).
31. Wainberg, M. et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019).
32. Gaziano, L. et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat. Med.* **27**, 668–676 (2021).
33. Mikhaylova, A. V. & Thornton, T. A. Accuracy of gene expression prediction from genotype data with predixcan varies across and within continental populations. *Front. Genet.* **10**, 261 (2019).
34. Huckins, L. M. et al. Analysis of genetically regulated gene expression identifies a prefrontal PTSD gene, SNRNP35, specific to military cohorts. *Cell Rep.* **31**, 107716 (2020).
35. Ouyang, W., Rutz, S., Crellin, N. K., Valdez, P. A. & Hymowitz, S. G. Regulation and functions of the IL-10 family of cytokines in inflammation and disease. *Annu. Rev. Immunol.* **29**, 71–109 (2011).
36. de Weerd, N. A. & Nguyen, T. The interferons and their receptors—distribution and regulation. *Immunol. Cell Biol.* **90**, 483–491 (2012).
37. Forbester, J. L. et al. Interleukin-22 promotes phagolysosomal fusion to induce protection against *Salmonella enterica* Typhimurium in human epithelial cells. *Proc. Natl Acad. Sci. USA* **115**, 10118–10123 (2018).
38. Ouyang, W. & O'Garra, A. IL-10 family cytokines IL-10 and IL-22: from basic science to clinical translation. *Immunity* **50**, 871–891 (2019).
39. Kotenko, S. V. et al. IFN-lambdas mediate antiviral protection through a distinct class II cytokine receptor complex. *Nat. Immunol.* **4**, 69–77 (2003).
40. Sheppard, P. et al. IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nat. Immunol.* **4**, 63–68 (2003).
41. Dinno, K. H. et al. A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures. *Nature* **586**, 560–566 (2020).
42. Major, J. et al. Type I and III interferons disrupt lung epithelial repair during recovery from viral infection. *Science* **369**, 712–717 (2020).
43. Coelho, T. et al. Safety and efficacy of RNAi therapy for transthyretin amyloidosis. *N. Engl. J. Med.* **369**, 819–829 (2013).
44. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
45. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
46. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
47. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
48. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat. Med.* **9**, 811–818 (1990).
49. Fang, H. et al. Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *Am. J. Hum. Genet.* **105**, 763–772 (2019).
50. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
51. Klarin, D. et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat. Genet.* **51**, 1574–1579 (2019).
52. Denny, J. C. et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1110 (2013).
53. Thompson, R. C. et al. Acute COVID-19 gene-expression profiles show multiple etiologies of long-term sequelae. *medRxiv* <https://doi.org/10.1101/2021.10.04.21264434> (2021).
54. Beckmann, N. D. et al. Cytotoxic lymphocytes are dysregulated in multisystem inflammatory syndrome in children. *medRxiv* <https://doi.org/10.1101/2020.08.29.20182899> (2020).
55. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
56. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).
57. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinform.* **17**, 483 (2016).
58. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
59. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* **1711**, 243–259 (2018).
60. Del Valle, D. M. et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* **26**, 1636–1643 (2020).
61. Hoffman, G. E. et al. Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nat. Commun.* **8**, 2225 (2017).
62. Ho, S.-M. et al. Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods* **101**, 113–124 (2016).
63. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
65. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
66. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
67. Broad Institute. *GitHub repository. Picard Tools*. (Broad Institute, 2021).
68. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
69. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
70. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133 (2012).
71. Corvelo, A., Clarke, W. E., Robine, N. & Zody, M. C. taxMaps: comprehensive and highly accurate taxonomic classification of short-read data in reasonable time. *Genome Res.* **28**, 751–758 (2018).
72. NCBI Resource Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **45**, D12–D17 (2017).
73. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
74. Nilsson-Payant, B. E. et al. The NF- $\kappa$ B Transcriptional Footprint Is Essential for SARS-CoV-2 Replication. *J. Virol.* **95**, e0125721 (2021).

## ACKNOWLEDGEMENTS

We would like to thank Chris Chatzinakos and Nikolaos P. Daskalakis for providing valuable feedback on their JEPeGMIX2-P method and software. This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by award #MVP035. This publication does not represent the views of the Department of Veteran Affairs or the United States Government. This study was also supported by the National Institutes of Health (NIH), Bethesda, MD under award numbers K08MH122911 (Georgios Voloudakis), R01AG050986, R01AG065582, R01AG067025, U01MH116442, R01MH109677 (Panos

Roussos), and by the Veterans Affairs Merit grants BX002395 and BX004189 (Panos Roussos). This study has also been funded in part by the Brain & Behavior Research Foundation via the 2020 NARSAD Young Investigator Grant #29350 (Georgios Voloudakis) and the Icahn School of Medicine at Mount Sinai via a COVID-19 Pilot Grant (Schahram Akbarian, Benjamin R. tenOever, Kristen J. Brennan).

## AUTHOR CONTRIBUTIONS

G.V. and P.R. contributed to overall study concept and design. G.V., G.E.H., N.D.B., A.C., K.C., B.R.tO., A.W.C., K.J.B., J.F.F., and P.R. provided domain-specific design advice. G.V., J.M.V., S.V., G.E.H., K.D., N.D.B., C.A.H., D.H., J.F.F., and P.R. contributed to writing the manuscript with input from all authors (G.V., J.M.V., S.V., G.E.H., K.D., W.Z., N.D.B., C.A.H., S.A., S.J., D.H., L.G., A.C., K.C., K.M.L., J.B., J.S.L., S.K.I., S.-W.L., S.A., R.S., T.L.A., E.E.S., J.A.L., M.M., B.R.tO., A.W.C., K.J.B., J.F.F., and P.R.). G.V., J.M.V., S.V., G.E.H., K.D., W.Z., N.D.B., C.A.H., S.A., S.J., D.H., and A.C. performed experiments and/or data analyses. All authors interpreted the data, provided conceptual comments, and approved the final manuscript. J.M.V., S.V., G.E.H., and K.D. contributed equally to this work. K.J.B. and J.F.F. had equal supervisory roles.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41525-022-00324-x>.

**Correspondence** and requests for materials should be addressed to Georgios Voloudakis or Panos Roussos.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022