bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genome-scale metabolic modeling reveals SARS-CoV-2-induced host metabolic reprogramming and identifies metabolic antiviral targets

Kuoyuan Cheng[1,2], Laura Riva[3,*], Sanju Sinha[1,2], Lipika Ray Pal[1], Nishanth Ulhas Nair[1], Laura Martin-Sancho[3], Sumit K. Chanda[3], Eytan Ruppin[1,4,#]

1. Cancer Data Science Laboratory (CDSL), National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD, USA.

2. Biological Sciences Graduate Program (BISI), University of Maryland, College Park, MD, USA.

3. Immunity and Pathogenesis Program, Infectious and Inflammatory Disease Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA.

4. Department of Computer Science, University of Maryland, College Park, MD, USA.

* Current address: Calibr, a division of The Scripps Research Institute, 11119 North Torrey Pines Road, La Jolla, CA 92037, USA

corresponding author, email: eytan.ruppin@nih.gov

Abstract

Tremendous progress has been made to control the COVID-19 pandemic, including the development and approval of vaccines as well as the drug remdesivir, which inhibits the SARS-CoV-2 virus that causes COVID-19. However, remdesivir confers only mild benefits to a subset of patients, and additional effective therapeutic options are needed. Drug repurposing and drug combinations may represent practical strategies to address these urgent unmet medical needs. Viruses, including coronaviruses, are known to hijack the host metabolism to facilitate their own proliferation, making targeting host metabolism a promising antiviral approach. Here, we describe an integrated analysis of 12 published *in vitro* and human patient gene expression datasets on SARS-CoV-2 infection using genome-scale metabolic modeling (GEM). We find that SARS-CoV-2 infection can induce recurrent and complicated metabolic reprogramming spanning a wide range of metabolic pathways. We next applied the GEM-based metabolic transformation algorithm (MTA) to predict anti-SARS-CoV-2 targets that counteract the virus-induced metabolic changes. These predictions are enriched for validated targets from various published experimental drug and genetic screens. Further analyzing the RNA-sequencing data of remdesivir-treated Vero E6 cell samples that we generated, we predicted metabolic targets that act in combination with remdesivir. These predictions are enriched for previously reported synergistic drugs with remdesivir. Since our predictions are based in part on human patient data, they are likely to be clinically relevant. We provide our top high-confidence candidate targets for their evaluation in further studies, demonstrating host metabolism-targeting as a promising antiviral strategy.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The coronavirus disease 2019 (COVID-19), a serious respiratory disease caused by the coronavirus SARS-CoV-2 has evolved into a major pandemic incurring millions of deaths worldwide (all dates as of December 2020; WHO Coronavirus Disease Dashboard, 2020). Despite unprecedented global efforts in response to this serious health threat including abundant studies on the disease biology (e.g. Zhou et al. 2020, Hoffmann et al. 2020, etc., reviewed in Tay et al. 2020, etc.), preclinical antiviral drug/target screens or predictions (e.g. Riva et al. 2020, Wei et al. 2020, Danoliski et al. 2020, etc., with compiled resources like Kuleshov et al. 2020, etc.), and thousands of registered clinical trials on COVID-19 (International Clinical Trials Registry Platform, 2020), therapeutic options remain scarce. Remdesivir, a viral RNA-dependent RNA polymerase inhibitor represents the only drug approved by the drug regulatory authorities of several countries, including the U.S. Food and Drug Administration (FDA) (Beigel et al. 2020), and confers only mild clinical benefits to a subset of COVID-19 patients (WHO Solidarity Trial Consortium et al. 2020). The Janus kinase (JAK) inhibitor baricitinib (in combination with remdesivir), and virus-neutralizing antibodies bamlanivimab, and casirivimab plus imdevimab have obtained Emergency Use Authorization (EUA) from the FDA (U.S. Food and Drug Administration, 2020). Dexamethasone and other corticosteroids have been recommended by the U.S. National Institutes of Health (NIH) for hospitalized patients requiring supplemental oxygen (RECOVERY Collaborative Group et al. 2020; National Institutes of Health, 2020). Besides, several SARS-CoV-2 vaccines have been approved or authorized for emergency use in different countries (Dong et al. 2020; U.S. Food and Drug Administration, 2020). Nevertheless, there is still an urgent unmet medical need for the fast identification and development of highly effective anti-COVID-19 therapies.

Viruses are known to “hijack” host cell metabolism to complete their own intracellular life cycle (Mayer et al. 2019), modulating diverse pathways including carbohydrate, lipid, amino acid and nucleotide metabolism (Mayer et al. 2019; Sanchez et al. 2015). Coronaviruses including MERS-CoV rearrange cellular lipid profiles upon infection (Yan et al. 2019; Yuan et al. 2019). Recent studies have reported that SARS-CoV-2 also induces changes in numerous metabolic pathways including TCA cycle, oxidative phosphorylation and lipid metabolism among others in human patient samples (Gardinassi et al. 2020; Ehrlich et al. 2020). Notably, counteracting the metabolic demands of viruses including MERS-CoV have been shown to abolish their ability to infect the host cells (Mayer et al. 2019; Yuan et al. 2019), and the PPARα-agonist fenofibrate can reverse some of the SARS-CoV-2-induced metabolic changes and reduce the viral load (Ehrlich et al. 2020). Therefore, targeting the virus-induced metabolic changes can be a promising novel antiviral strategy (Mayer et al. 2019), and can be especially valuable in anti-SARS-CoV-2 drug repurposing to address the current urgent COVID-19 crisis considering that many existing drugs are metabolism-targeting.

Genome-scale metabolic models (GEMs) are *in silico* constraint-based models that comprehensively encompass the cellular network of metabolic reactions, metabolic proteins, and metabolites (Baart et al. 2012). GEM analysis has been repeatedly shown to generate accurate predictions and informative hypotheses for metabolism research (Gu et al. 2019). Notably, we have previously developed numerous GEM-based algorithms including iMAT

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Shlomi et al. 2008), which computes genome-wide metabolic fluxes from gene expression profiles, and the metabolic transformation algorithm (MTA; Yizhak et al. 2013), which predicts metabolic targets whose inhibition facilitates transformation between specified cellular metabolic states (e.g. from diseased to healthy states). More recently, Valcárcel et al. has described a variant of MTA named rMTA with improved performance (Valcárcel et al. 2019). Incorporating such high-performance GEM methods in the analysis of data on SARS-CoV-2 infection provides us with a unique opportunity to understand the metabolic demands of SARS-CoV-2 and to systematically predict anti-SARS-CoV-2 targets that counteract the virus-induced metabolic alterations.

Here we apply GEM algorithms in a comprehensive analysis of 12 published bulk/single-cell RNA-sequencing (RNA-seq/scRNA-seq) and mass spectrometry (MS)-based proteomics datasets on SARS-CoV-2 infection, involving both *in vitro* and human patient samples. We find that metabolic reprogramming represents one of the most consistent molecular changes in SARS-CoV-2 infection besides immune responses, and characterize the complex patterns of metabolic flux alterations. Using MTA, we predicted anti-SARS-CoV-2 targets that reverse the virus-induced metabolic changes, either as single targets or in combination with remdesivir (the latter using our new RNA-seq data on remdesivir treatment). The predictions are highly enriched for reported anti-SARS-CoV-2 targets identified from various experimental screens, and we provide our top high-confidence candidate targets for the evaluation in further studies.

Results

Integrated analysis of multiple gene expression datasets identifies coherent immune and metabolic changes in SARS-CoV-2 infection

Multiple studies have characterized the gene expression changes during SARS-CoV-2 infection in different *in vitro* and *in vivo* settings. We collected a total of 12 published relevant datasets spanning a wide range of sample types (various cell lines, primary bronchial epithelial cells, nasopharyngeal swab and bronchoalveolar lavage fluid, i.e. BALF samples from patients) and assay platforms (bulk RNA-seq, scRNA-seq, and MS-based proteomics). These datasets are summarized in **Table 1**. With each of the datasets, we performed differential expression (DE) analysis comparing the SARS-CoV-2-infected or positive samples to the non-infected control or negative samples (Methods; Table S1). For the single-cell datasets, we focused on the airway epithelial cell which is known as the major virus-infected cell type. Comparing the datasets with a principle component analysis (PCA) plot based on the inverse normal-transformed DE log fold-change values (**Fig. 1A**; Methods) suggests that the cell lines tend to have distinct DE profiles from the patient samples, although different patient datasets exhibit considerable variation depending on sample type and sequencing platform. This pattern is confirmed by the comparison of the top significant DE genes from each pair of datasets (using Fisher's exact tests, top n=400 DE genes from each dataset were used, all with at least FDR<0.1; Methods; **Fig. 1B**). Moreover, examining only the top DE genes appears to mitigate the technical variation across datasets, with reasonable coherence demonstrated by odds ratio median value 2.18 and maximum 13.60 (adjusted P median 5.09e-5, minimum<2.22e-16; **Fig. 1B**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Summary of the published gene expression datasets on SARS-CoV-2 infection analyzed in this study.

Dataset Name*	Sample Type	Sample Size†	Platform	Reference
Vero	Vero E6 cell line	6	bulk RNA-seq	Riva et al. 2020
NHBE	Primary normal human bronchial epithelial cell	6	bulk RNA-seq	Blanco-Melo et al. 2020
A549	A549 human lung adenocarcinoma cell line with exogenous ACE2 expression	6	bulk RNA-seq	Blanco-Melo et al. 2020
Calu-3	Calu-3 human lung adenocarcinoma cell line	6	bulk RNA-seq	Blanco-Melo et al. 2020
293T	HEK293T human embryonic kidney cell line	12	bulk RNA-seq	Weingarten-Gabby et al. 2020
Caco-2	Caco-2 human colorectal adenocarcinoma cell line	6	MS-based proteomics	Bojkova et al. 2020
Swab.Butler	NP swab samples from human individuals	580	bulk RNA-seq	Butler et al. 2020
Swab.Lieberman	NP swab samples from human individuals	484	bulk RNA-seq	Lieberman et al. 2020
BALF	BALF from human individuals	6	bulk RNA-seq	Xiong et al. 2020
SC.Liao	BALF from human individuals (epithelial cells were used in analysis)	13	scRNA-seq	Liao et al. 2020
SC.Chua.Basal	NP and bronchial samples from human individuals (basal cells were used in analysis)	24	scRNA-seq	Chua et al. 2020
SC.Chua.Ciliated	NP and bronchial samples from human individuals (ciliated cells were used in analysis)	24	scRNA-seq	Chua et al. 2020

* These are the names used in figure labels throughout the text.
† The total number of replicates (virus-infected and control combined) used for analysis in in vitro datasets, or the total number of human individuals (patients and controls combined) used for analysis in in vivo datasets. In some datasets, only a subset of all the available samples were analyzed.
Abbreviations: NP, nasopharyngeal; BALF, bronchial alveolar lavage fluid; RNA-seq, RNA-sequencing; MS, mass spectrometry; scRNA-seq, single-cell RNA-sequencing.

We then performed gene set enrichment analysis (GSEA) (Subramanian et al. 2005) on the DE results from each dataset (Table S2), and further compared the datasets on the pathway level by the significantly enriched pathways (FDR<0.1; Methods). Reassuringly, the level of coherence across datasets on the pathway level is even stronger, with a median odds ratio of

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.25 (maximum is infinity followed by 35.49) across pairs of datasets (adjusted P median 2.01e-5, minimum<2.22e-16; **Fig. 1C**). Examining the most consistently enriched pathways across the datasets while giving higher importance to the various *in vivo* patient datasets (**Fig. 1D**; Table S3; Methods), we see many up-regulated pathways involved in innate immune response to viral infection, e.g. interferon signaling and cytosolic DNA sensing. Among the pathways involving coherently down-regulated genes upon SARS-CoV-2 infection, we find antigen presentation, as well as numerous pathways spanning many major categories of cellular metabolism, e.g. oxidative phosphorylation (OXPHOS; energy metabolism), sphingolipid metabolism (lipid metabolism), valine, leucine and isoleucine degradation (amino acid metabolism), and N-glycan biosynthesis. These may reflect the specific metabolic requirements of SARS-CoV-2 or underlie its pathogenic effects (see Discussion). Visualizing a more complete landscape of metabolic pathway alterations across the datasets reveals further consistent, although weaker changes (based on GSEA normalized enrichment score, i.e. NES; **Fig. 1E**; Table S2; Methods). These results suggest that besides immune response, metabolic reprogramming represents one of the most robust changes induced by SARS-CoV-2 infection across various systems, consistent with the key roles of metabolism in viral infection. We next focused on characterizing the SARS-CoV-2-induced metabolic changes in the infected host cells on the metabolic flux level.

Genome-scale metabolic modeling (GEM) identifies SARS-CoV-2-induced patterns of metabolic flux changes

Since gene expression does not necessarily correlate with protein level or enzyme activity and thus may not truthfully reflect metabolic activity (Maier et al. 2009), we applied GEM to infer the metabolic fluxes (i.e. rates of all metabolic reactions) across the datasets. Specifically, for each dataset, the iMAT algorithm (Shlomi et al. 2008) was applied to the median expression profiles of the control and virus-infected samples to compute the refined metabolic models representative of the two respective groups, and the flux distribution of each metabolic reaction was obtained by sampling (Methods). The flux distributions of the control and infected groups were compared and reactions with differential fluxes (DF) were identified (Methods; Table S4). We again examined the consistency across the datasets, here on the flux level, by checking the overlap of the top DF reactions between each pair of datasets. Like on the gene expression level, we are assured by the overall high level of coherence of the DF reactions (odds ratio median 2.40, maximum 28.13; adjusted P value median and minimum both <2.22e-16; **Fig. 2A** shows the result for the positive DF reactions, the result is similar for negative DF reactions; We note that the sign of DF represents the direction of flux change with regard to the positive direction of a reaction, which can be reversible, and not the increase or decrease of the absolute flux). Although no reaction shows fully consistent changes across all 12 datasets, we identified a set of most consistently changed reactions across datasets while giving higher importance to the *in vivo* patient datasets (Table S5A; Methods), and examined the metabolic pathways they are enriched in with Fisher's exact tests (significant pathways with FDR<0.1 shown in **Fig. 2B**; Table S5B). We see that consistent flux changes are found in various noteworthy pathways including metabolite transport (mitochondrial and extracellular), pentose phosphate pathway, hyaluronan metabolism, pyrimidine biosynthesis, glycine, serine, and threonine metabolism, TCA cycle, inositol phosphate metabolism, fatty acid elongation, among

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

others. Many of these pathways have been implicated in the infection and life cycle of different viruses including SARS-CoV-2 (Mayer et al. 2020, Ou et al. 2020, Gardinassi et al. 2020, Li et al. 2020, Thomas et al. 2020, Ehrlich et al. 2020, Bojkova et al. 2020; see Discussion). Below we closely inspect the flux alteration patterns within these pathways with specialized methods.

First, we focused on the cellular transport pathway, seeking to identify the specific transport pattern summarized for each metabolite species (Methods; Table S5C). Some of the metabolites showing the most consistent patterns across datasets are summarized in **Fig. 2C**. In addition to biological variation, technical noise from the modeling can also contribute to the lack of full consistency in these results. Nevertheless, several glucogenic amino acids, including alanine, asparagine, and glutamate are seen to have increased cellular import from the extracellular space to the cytoplasm (or decreased export) in the virus-infected vs control group across more than half of the datasets. On the other hand, nicotinamide and tetrahydrofolate show relatively consistent patterns of decreased cellular import (or increased export; **Fig. 2C**). Next, we inspected the other pathways not related to metabolite transport by visualizing their alteration patterns overlaid on the metabolic network, for virus-infected vs the control group. For example, the pyrimidine (*de novo*) biosynthesis pathway contains consistently increased fluxes towards the synthesis of UMP (the precursor of pyrimidines; **Fig. 2D**), consistent with the nucleic acid synthesis needs of the virus. In the TCA cycle, we mostly see increased flux driving the normal cycle, together with the interconversion and mitochondrial transport of the cytoplasmic counterparts of TCA cycle intermediates (**Fig. 2E**). This seems contradictory to the consistent gene expression-level decrease of the “TCA cycle and respiratory electron transport" pathway (Fig. 1D). However, upon closer inspection, we find that the gene-level enrichment of this pathway is mostly driven by the decrease in electron transport chain genes (Table S2), and consistently OXPHOS reactions show decreased fluxes (Table S5A). As examples of pathways with more complex flux change patterns, in the inositol phosphate metabolism pathway, we see increased fluxes converging to phosphatidylinositol 4,5-bisphosphate (pail45p_hs[c]) and inositol (inost[c]), but decreased fluxes to inositol 1-phosphate (mi1p_DASH_D[c]; **Fig. 2F**); in the glycine, serine, and threonine metabolism pathway, we see decreased conversion of serine to glycine but increased serine degradation to pyruvate (**Fig. 2G**); in the fatty acid elongation pathway, we see that the synthesis and interconversion of different fatty acids show distinct flux changes (**Fig. 2H**). These highly intricate metabolic programs revealed by the GEM analysis are consistent with many previous reports and possibly reflect the specific metabolic demands of SARS-CoV-2 during its life cycle (see Discussion), which also demonstrates the value of the modeling approach over gene expression-level analyses.

Prediction of anti-SARS-CoV-2 targets that act via counteracting the virus-induced metabolic changes

We have demonstrated that SARS-CoV-2 can induce recurrent and complex alterations in host cell metabolism. As was proposed previously, targeting the virus-induced metabolic changes can be an effective antiviral strategy (Mayer et al. 2019), which we adopted here to predict anti-SARS-CoV-2 targets. Specifically, we applied the GEM-based rMTA algorithm (Valcárcel et al. 2019) to each of our collected datasets to predict metabolic reactions whose

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

knockout (KO) can transform the cellular metabolism from the SARS-CoV-2-infected state to the non-infected normal state (Methods; Table S6). MTA computes a score for each of the metabolic reactions in the cell, and usually the 10-20% reactions with the highest MTA score contain promising candidate targets (Yizhak et al. 2013). We first compared the top 10% MTA-predicted reactions across datasets and found that they have reasonable overlap (odds ratio median 3.42, maximum 16.47, Fisher's exact test adjusted P median and minimum both<2.2e-16 across all pairs of datasets; **Fig. 3A**). Interestingly, some strong overlaps are seen between certain cell line and patient datasets, consistent with the recurrent metabolic changes across these datasets as seen above.

To validate these predictions, we collected multiple validation sets of reported anti-SARS-CoV-2 gene targets or drugs identified from large-scale chemical or genetic screens. These include CRISPR-Cas9 genetic screens in Vero E6 cells (Wei et al. 2020) and in cells with exogenous *ACE2* expression (A549^{ACE2}, Daniloski et al. 2020), and additional lists of experimentally validated drugs reported in different *in vitro* studies compiled by Kuleshov et al. 2020 (Methods). We first tested for significant overlap between our top 10% MTA-predicted targets from each of the datasets and the validation sets described above with Fisher's exact tests (after mapping all validated target genes or drugs to the metabolic reactions; Methods). Strongly significant overlaps were found between our predictions from 8 out of the 12 datasets with the antiviral hits (i.e. those whose KO inhibits SARS-CoV-2 infection) identified in the CRISPR-Cas9 screens (FDR<3.95e-4; **Fig. 3B**; Table S6B), these significant datasets include the Vero (Riva et al. 2020) and A549^{ACE2} data (Blanco-Melo et al. 2020) from the same cell types as those used in the CRISPR-Cas9 screens, but encouragingly also include three *in vivo* patient datasets (Lieberman et al. 2020, Liao et al. 2020, Chua et al. 2020). Further examining the experimentally validated anti-SARS-CoV-2 drug sets from previous studies (compiled by Kuleshov et al. 2020), we also found a few cases of significant overlap (FDR<0.1; **Fig. 3C**; Table S6C), although these drug sets are relatively small, and when we pooled all validated drugs compiled by Kuleshov *et al.*, their targets are only enriched in the predictions from the SC.Chua.Basal dataset (**Fig. 3C**). Some of our top predicted reactions are also mapped to host proteins identified to interact with SARS-CoV-2 proteins from Gordon et al. 2020 and Stukalov et al. 2020, although we only see weak statistically significant overlaps with the Stukalov et al. data but not the Gordon et al. data (FDR<0.1; **Fig. 3D**; Table S6D), suggesting that most of our predicted targets are not directly involved in the SARS-CoV-2 life cycle. Overall, our MTA-based top predictions obtained strong validation from the published CRISPR-Cas9 screens, with further but weaker support from the drug screens and host-virus protein-protein interaction (PPI) data.

We further take advantage of the genome-wide CRISPR-Cas9 screens to more closely evaluate the performance of our MTA predictions. Unlike in many of the drug-screen datasets where the screens are low-throughput or complete screen results were not available, we were able to confidently define positive and negative sets (i.e. genes whose KO inhibits or promotes the viral infection, respectively) from the CRISPR-Cas9 screen data. The positive and negative sets were defined in a balanced way (Methods), with which we performed ROC curve analysis of our MTA predictions from each of the dataset (Methods). Although the MTA prediction is only based on the transformation of cellular metabolic states and does not consider the possible

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

effect of other anti-/pro-viral mechanisms, we see that the predictions based on 6 of the datasets achieved area under ROC curve (AUROC) values above 0.6 and as high as 0.81, although two of the other datasets apparently yielded AUROC significantly lesser than 0.5 (**Fig. 3E**; see Discussion). As examples, ROC curves from the two best-performing datasets are shown in **Fig. 3F**, we note that the steep proportions of the curves showing close to 1 specificity and decent sensitivity indeed correspond to about top 10-20% of the MTA predictions, consistent with our previous experience with MTA. These results testify that our metabolism-targeting strategy using the MTA algorithm is able to achieve reasonable prediction performances.

Next, we seek to integrate our predictions from the 12 datasets into a final consensus list of high-confidence candidate targets for further extensive experimental validation and investigation. We applied a procedure to pick highly recurrent top predictions across both the *in vitro* and *in vivo* datasets (Methods), resulting in a final list of 66 candidate target metabolic reactions of which 59 are mapped to genes and 15 are targeted by known drugs (Table S7A). This final list of candidates are also strongly enriched for the positive targets identified in the two anti-SARS-CoV-2 CRISPR-Cas9 screens described above (Wei et al. 2020 and Daniloski et al. 2020; odds ratio=26.56, P<2.2e-16). These candidates are enriched for metabolic pathways including cellular transport, inositol phosphate and aminosugar metabolism, and within the cellular transport pathway, transporters for various amino acids, succinate, citrate and adenosine are enriched (FDR<0.1; **Fig. 3G**; Table S7B; Methods). These are consistent with the known biology of SARS-CoV-2, e.g. phosphoinositides are known to be critical for SARS-CoV-2 cell-entry by endocytosis, and inhibiting phosphatidylinositol-3,5-bisphosphate with the drug apilimod has been shown to suppress SARS-CoV-2 entry (Ou et al. 2020).

Prediction of metabolic targets for anti-SARS-CoV-2 in combination with remdesivir

Given that our MTA-based prediction of single anti-SARS-CoV-2 metabolic targets has yielded promising results, we proceed to extend the same strategy for the prediction of targets that can be combined with remdesivir to achieve higher antiviral efficacy. To this aim, we cultured Vero E6 cells infected by SARS-CoV-2, with or without remdesivir treatment. A control group (no viral infection or remdesivir treatment) a remdesivir-only group (no viral infection) were also included (Methods). Bulk RNA-seq was performed to obtain the gene expression profiles of these samples (Methods). Visualizing the gene expression data with a PCA plot, we see that remdesivir can indeed effectively reverse the virus-associated expression changes (mostly along the first PC axis), but also results in additional orthogonal changes along the second PC axis (**Fig. 4A**). Performing a GSEA analysis comparing the virus+remdesivir group to the normal control group, we see that many pathways show significant differences in their expression, including some metabolic pathways, e.g. cholesterol and steroid biosynthesis (**Fig. 4B**; Table S8; Methods). Some of these differences can be attributed to the incomplete reversion of virus-induced expression changes by remdesivir, while others may arise from remdesivir-specific effects (**Fig. 4B**). Further computing the metabolic flux profiles representative of each group of samples with iMAT (Shlomi et al. 2008) then inspecting the flux-level PCA plot (**Fig. 4C**; Methods), we observe a similar pattern from that seen on the gene

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression level. The differential fluxes between the virus+remdesivir and the control group are enriched for various metabolic pathways (FDR<0.1; **Fig. 4D**; Table S9; Methods), many also have differential fluxes comparing virus-infected samples to control (see Fig. 2B), suggesting that these metabolic changes are not fully reversed to normal by remdesivir. We hypothesize that further reversing the cellular state in the virus+remdesivir group towards the healthy control state may be an effective combinatory targeting strategy to improve the antiviral efficacy of remdesivir.

As before, we focused on the domain of cell metabolism and applied MTA on our data from Vero E6 cells to predict targets for reversing the metabolic flux profile in the virus+remdesivir group towards normal (Methods; Table S10A). Trying to validate these predictions, we obtained a list of 20 experimentally tested drugs showing synergistic anti-SARS-CoV-2 effects with remdesivir in the Calu-3 cell line (Nguyenla et al. 2020). Despite the cell type difference, we observed that the targets of these drugs are significantly enriched by our top 20% MTA predictions (Fisher's exact test P=0.011, odds ratio 4.83; there is also a trend of enrichment by the top 10% MTA predictions with odds ratio 2.01, but it failed to achieve statistical significance at P=0.30). The top 20% predictions recovered 6 of the 11 metabolic reaction targets from Nguyenla et al., corresponding to drugs including cilostazol, ezetimibe, ivosidenib and valdecoxib. Some of the top predicted targets overlap with our predicted single anti-SARS-CoV-2 targets as described above, e.g. various inositol phosphate metabolism reactions. This is consistent with the observation that the virus-induced metabolic changes in these pathways are not effectively reversed to normal by remdesivir (as seen from Fig. 2B and 4D). Performing a pathway enrichment analysis, we see that these top predictions are enriched in bile acid, IMP, and heme biosynthesis pathways, among others (**Fig. 4E**; Table S10B); cholesterol and steroid metabolism pathways, which were seen to be different on the gene expression level between the virus+remdesivir and control samples (see Fig. 4B), are also enriched for the top predicted targets (**Fig. 4E**). We provide the top 20% predictions ranked by the MTA score and known drugs for these targets in Table S10A. These predictions represent candidate targets that can potentially improve the antiviral efficacy of remdesivir in combination and warrant further testing in future studies.

Discussion

In this study, we provide a comprehensive GEM analysis integrating 12 published gene expression datasets on SARS-CoV-2 infection, spanning multiple *in vitro* and *in vivo* sample types and expression profiling platforms. We revealed the complexity of host metabolic reprogramming by SARS-CoV2, and further predicted anti-SARS-CoV-2 single or combinatory (with remdesivir) targets that act via counteracting the virus-induced metabolic changes. Our GEM-based prediction algorithm showed good performance based on validation with published targets from *in vitro* screens, and the predicted targets represent highly promising candidates for further experimental testing.

To date, a large number of studies have been published that contributed to our fast understanding of the host molecular changes associated with SARS-CoV-2 infection. These studies involve a variety of different experimental models and/or sample types, making it

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

necessary to perform a systematic analysis across datasets and evaluate the robustness and clinical relevance of the findings in human patients. Although we do not aim to (and cannot) include all relevant published data, we tried to cover datasets on both popular *in vitro* models of SARS-CoV-2 infection as well as human patients (nasopharyngeal swab and BALF samples). While many other studies have shed light on the systemic and immune cell-specific response characteristic of SARS-CoV-2 infection (e.g. Zheng et al. 2020 and many of the patient studies we collected in Table 1), our focus is specifically on the virus-infected host cells, i.e. primarily the airway epithelial cells *in vivo*. Therefore, for human samples, in addition to bulk RNA-seq, we analyzed scRNA-seq data to separate the distinctive changes within the epithelial cells from, e.g. various types of immune cells. In terms of methodology, we have avoided a more formal effect size-based meta-analysis due to the challenge arising from the wide range of expression profiling platforms (bulk RNA-seq, scRNA-seq, MS-based proteomics) and the technical variation involved. Rather, we relied more on P values, and made subjective decisions that give higher importance to the various patient datasets when defining consistent findings, aiming to obtain results of higher clinical relevance. By integrated analysis of all these data, we found that metabolism is one of the cellular domains that exhibit the most coherent changes across datasets in SARS-CoV-2 infection (besides immune responses; Fig. 1D). This finding is consistent with our prior knowledge on the need of a wide spectrum of viruses to manipulate host metabolism for viral proliferation (Mayer et al. 2019), and sets a solid basis for our GEM-based metabolic flux analysis and antiviral target prediction.

The application of GEM in complement to gene expression-level analysis is a central part of our study. It is known that gene expression does not always perfectly correlate with protein level or enzyme activity (Maier et al. 2009). Besides, many metabolic reactions are reversible, while the directions of reactions are important biologically, such information is missing on the gene level. By taking advantage of the additional information in the topological constraints of the metabolic network, GEM allows us to infer the actual metabolic fluxes, thus revealing extra complexity in SARS-CoV-2-induced metabolic reprogramming, as is evident from Fig. 2D-H. Many of these inferred metabolic changes are consistent with what's known about SARS-CoV-2 and other related viruses. For example, the highly coherent increase in pyrimidine biosynthesis (Fig. 2D) corresponds to the increased need of viral genome replication and gene expression (Bojkova et al. 2020), and pyrimidine *de novo* synthesis inhibitors have been shown to have anti-SARS-CoV-2 effects (Xiong et al. 2020). Inositol phosphate metabolism (Fig. 2F) is important for the life cycle of many viruses due to the structural or signaling roles of different phosphoinositides (Beziau et al. 2020), with the inhibition of certain phosphoinositides disrupting endocytosis and blocking SARS-CoV-2 cell-entry (Ou et al. 2020). Fatty acid synthesis was reported to increase in SARS-CoV-2 infection (Ehrlich et al. 2020), whereas our results suggest a more complex pattern for different fatty acid species (Fig. 2H), which echoes the results of several metabolomics studies (Shen et al. 2020, Barberis et al. 2020, Thomas et al. 2020). Both TCA cycle and OXPHOS have been shown to decrease based on gene expression during the virus infection (Ehrlich et al. 2020, Gardinassi et al. 2020), and have been implicated in the systemic syndromes of the virus (Li et al. 2020). While we did recapitulate such gene expression changes, our GEM analysis only confirmed the decrease of OXPHOS (Table S5A) but revealed a mostly normal-functioning TCA cycle on the flux level,

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which appears to be driven by 2-oxoglutarate transported from the cytosol (Fig. 2E). Despite that GEM can help to suggest such intricate flux-level patterns, these computed fluxes should be verified with isotope labeling experiments, and their biological significance in the virus infection needs to be further investigated.

Given the importance of metabolism during virus infection, targeting host metabolism has already been proposed as a promising novel antiviral strategy (Mayer et al. 2019). For this, the MTA algorithm we previously developed (Yizhak et al. 2013), again a method under the GEM framework, can be particularly valuable for the metabolic target discovery. MTA has been successfully applied to predict lifespan extending interventions in yeast (Yizhak et al. 2013), a metabolic cancer driver gene (Auslander et al. 2017), and a novel therapeutic target for intractable epilepsy (Styr et al. 2019). A recent variant of MTA named rMTA was shown to deliver better performance (Valcárcel et al. 2019), and here we used an optimized implementation of rMTA in our study. MTA/rMTA are not based on supervised machine learning technique, and do not use any of the validation datasets for target prediction. Yet, we were able to achieve decent performance during the validation (Fig. 3B-F). It is particularly encouraging to see that in several cases, the validation data (which mostly originated from *in vitro* experiments) correlated well with the predictions based on *in vivo* human data. In some datasets, however, our predictions were not successfully validated by the genetic screen data (Fig. 2B,E). One reason could be that MTA can only consider the metabolism-related effects and ignores other potential mechanisms that determine the antiviral efficacy of a target. Nevertheless, it could also be due to biological differences between the datasets used for prediction and those used for validation. To avoid overdependence on the limited validation sets available in defining the final consensus candidate target list, we did not explicitly exclude any dataset used for prediction but enforced the inclusion of the human patient data to achieve higher clinical relevance. The prediction for combinatory targets with remdesivir also showed promising preliminary results, although our validation is more limited in this case due to lack of data. Follow-up studies are warranted to solidly test and validate these predicted targets for potential further antiviral therapy development.

In summary, we identified prevalent and intricate metabolic reprogramming in the host cell as a feature of SARS-CoV-2 infection, and further predicted single and combinatory antiviral targets with promising performance seen in preliminary validations. These targets should be rigorously validated experimentally. Since our predictions are in part based on human patient data, they are likely to have high clinical relevance and may ultimately help to achieve better efficacy in COVID-19 treatment. Our study demonstrates the targeting of host metabolism as a promising antiviral strategy and highlights the power of GEM analysis to advance the understanding of cell metabolism during viral infection and antiviral target prediction.

Materials and Methods

Differential gene expression analysis

We obtained each of the gene expression datasets on SARS-CoV-2 infection from the sources listed in Table 1. For the bulk RNA-sequencing (RNA-seq) datasets whose read count data is

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

available at the time of analysis, we performed differential expression (DE) analysis comparing the SARS-CoV-2-infected or positive samples to the non-infected control or negative samples with DESeq2 (Love et al. 2014). For Weingarten-Gabby et al. 2020, a nested design for DE was needed and DESeq2 failed to run properly, and limma-voom (Law et al. 2014) was used. For Butler et al. 2020 and Xiong et al. 2020, we obtained the DE results provided from the supplementary materials of the respective publication, as we were not able to obtain their gene-level expression data at the time of analysis. For Butler et al. 2020, among their multiple versions of DE results we used the one from limma-voom with sva correction "Voom:Positive_vs_Negative:10M_samples:sva_correction_2sv". We also took the DE results of the proteomic data from Bojkova et al. 2020 as provided by the authors, and used the 24 hours post-infection data which is the latest time point available with the largest number of DE proteins). For the single-cell RNA-sequencing (scRNA-seq) datasets, The "FindMarkers" function in the R package Seurat (Stuart et al. 2019) was used to call the MAST method (Finak et al. 2015) for DE analysis in each annotated cell type, with "logfc.threshold" set to 0 to obtain full results across genes. We focused on the airway epithelial cells since our major aim in this study is to investigate the changes in the cell types infected by the SARS-CoV-2 virus, these include the "Epithelial" cell type from Liao et al. 2020 and the "Ciliated" and "Basal" cell types from Chua et al. 2020 (other epithelial subtypes from these datasets yielded no significant DE genes). All DE results are given in Table S1.

Gene set enrichment analysis of the differential expression results

Using the DE log fold-change values from each dataset, gene set enrichment analysis (GSEA) (Subramanian et al. 2005) was performed using the implementation in the R package fgsea (Korotkevich et al. 2019). The gene set/pathway annotations used were the Reactome (Jassal et al. 2020) and KEGG (Kanehisa et al. 2021) subsets from the "Canonical Pathway" category in version 7.0 MSigDB database (Liberzon et al. 2011). For metabolic pathways (in Figure 1E), those under the category "Metabolism" from KEGG (Kanehisa et al. 2021) were used. All GSEA results are given in Table S2.

Comparison of the differentially expressed genes and pathways across datasets

The DE results across datasets were compared in a descriptive manner. As a first approach, the DE log fold-change values were inverse normal-transformed across all genes within each dataset, which preserves only the order (i.e. rank) of DE effect sizes, and then PCA was applied to the transformed data. As a second approach, top significantly DE genes or enriched pathways from each pair of datasets were tested for significant overlap using Fisher's exact tests. For DE genes, the number of significant genes vary greatly across datasets based on a single FDR cutoff, so we uniformly took the top n=400 DE genes ranked by FDR from each dataset (all cases have at least FDR<0.1). For enriched pathways we simply took those with FDR<0.1 from each dataset. To identify the consistent DE changes across datasets, a formal meta-analysis of all 12 datasets is challenging given the wide range of assay platforms and DE algorithms used. So instead, we adopted a subjective criteria that give high importance to the various *in vivo* patient datasets, such that the results may be more clinically relevant: we identified pathways that are significantly (FDR<0.1) enriched in the consistent direction

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(up/down-regulation) in at least one of the bulk RNA-seq patient datasets and also at least one of the scRNA-seq datasets, while never showing significant enrichment (FDR<0.1) in the opposite direction in any of the datasets (for the results in Figure 1D; Table S3).

Computation of metabolic fluxes from gene expression data with genome-scale metabolic modeling

For each dataset, we used the genome-scale metabolic modeling (GEM) algorithm iMAT (Shlomi et al. 2008) to compute the metabolic flux profile from gene expression data. iMAT requires gene-length normalized expression values in the bulk RNA-seq datasets, for this we computed TPM values with Salmon (Patro et al. 2017) from the raw fastq files for datasets where TPM data is not provided. Then for each dataset, we took the median expression values of the control and virus-infected samples respectively as the representative expression profile for each group, and used it as input to iMAT. The human genome-scale metabolic model (GEM) Recon 1 (Duarte et al. 2007) was used as the base model to iMAT. The main reason we used the older human GEM version was due to the significantly longer computation time and resources needed if later larger models were used instead. The output of iMAT is a refined GEM for the each of the virus-infected and control groups in each dataset, with metabolic reaction bounds adjusted to achieve maximal concordance with the gene expression data while satisfying the stoichiometric constraints of the cellular metabolic network (Shlomi et al. 2008). Each output model defines a space rather than a single unique solution of the global metabolic flux profile, and artificial centering hit-and-run (ACHR) was used to sample the metabolic space and obtain the flux distribution of each metabolic reaction in each condition (control or virus-infected) and dataset. All GEM analyses were performed using our in-house R package named gembox, with the academic version of IBM ILOG CPLEX Optimization Studio 12.10 as the optimization solver.

Differential flux analysis of virus-infected vs control group in each dataset

The flux distributions of the control and infected groups were compared to identify reactions with differential fluxes (DF). Since an arbitrarily large number of sample points can be sampled from the metabolic space of each group, resulting in statistical tests with arbitrarily small P values, we adopted the following effect size-based criterion for DF reactions: absolute rank biserial correlation (an effect size measure of the difference between the two flux distributions in the control and virus-infected groups) >0.1, and absolute relative flux change (i.e. the absolute difference of the mean fluxes between the two groups over the absolute mean flux in the control group) >5%. Positive DF reactions have flux value difference in infected vs control group >0, and vice versa for negative DF reactions. Note that for non-reversible reactions, flux values are non-negative and the sign of DF can be interpreted similarly to differential gene expression; for reversible reactions, flux values can be negative, representing reactions happening in the reverse direction, thus the sign of DF needs to be interpreted differently, e.g. negative DF represents flux shift towards the reverse direction and not necessarily decrease in absolute flux. The DF results are given in Table S4.

Analysis of reactions with consistent differential fluxes across datasets and their pathway enrichment analysis

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To compare the DF results across datasets, the DF reactions from each pair of datasets were tested for significant overlap using Fisher's exact tests (separately for positive and negative DF). Since no reaction shows fully consistent DF across all 12 datasets analyzed, similarly as with the DE analysis, we identified the DF reactions with high level of consistency especially in the *in vivo* patient datasets, such that the results may be more clinically relevant: we identified DF reactions in the consistent direction (positive/negative) in at least one of the bulk RNA-seq patient datasets and also at least one of the scRNA-seq datasets, while showing DF in the opposite direction in no more than 3 datasets, resulting in 293 positive and 349 negative DF reactions (Table S5A). The metabolic pathway enrichment of these DF reactions was analyzed with Fisher's exact tests (results in Table S5B), with the "subSystems" slot in the metabolic model used as pathway annotation. Note that due to the special interpretation of the sign of DF values as explained above, the GSEA used for gene expression-level analysis is not appropriate for pathway enrichment analysis on the flux level.

Differential transport flux analysis by each metabolite species

Differential transport flux analysis on the metabolite level was first performed for each dataset. Specifically, in each group (control or virus-infected), the sampled flux values on the reaction level for all transport reactions across the cell membrane were summarized by each metabolite, taking the direction of transport (i.e. import or export) into consideration to produce the net cellular transport flux of the metabolite (positive means net import, vice versa). The metabolites with differential net transport fluxes between the control and virus-infected groups were identified as in the reaction-level DF analysis described above. The consistent DF metabolites were then identified in the same way as with the DF reactions described above. The results are given in Table S5C.

Analysis of the consistent flux alteration patterns in different metabolic pathways

For each of the significantly enriched metabolic pathways identified in the consistent DF reaction analysis described above, we defined the "consensus" direction of each reaction as represented by those shown in the virus-infected group from the majority (>6 out of 12) of the datasets, and also similarly for the "consensus" direction of DF for each reaction. The consensus directions of reactions and their DF were overlaid onto network diagrams of the pathways and visualized, where metabolites are represented by nodes, reactions are represented by directed (hyper) edges with edge direction corresponding to the consensus reaction direction and edge color corresponding to the consensus DF direction. Parts of the metabolic pathways where reactions are not consistently altered across datasets are greyed out or removed to increase the clarity. Upon visual inspection, potential futile loops in the network are also removed from the visualizations.

Prediction of anti-SARS-CoV-2 target metabolic reactions with metabolic transformation algorithm

For each of the collected datasets, the DE result of virus-infected vs control samples as well as the representative flux distribution of the virus-infected group computed with iMAT (Shlomi et al. 2008) followed by ACHR sampling were used as inputs for the GEM-based

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

metabolic transformation algorithm (MTA; Yizhak et al. 2013; a variant called rMTA was used; Valcárcel et al. 2019) to predict metabolic reactions whose knock-out can transform cellular metabolic state from that of the virus-infected to that of the control samples (full prediction results from all datasets in Table S6A). The output of rMTA is a score (rMTA score) for each metabolic reaction, with higher scores corresponding to better candidates for achieving the metabolic transformation as specified above. From our previous experience (Yizhak et al. 2013), the top 10-20% MTA predictions contain promising targets. The human Recon 1 GEM (Duarte et al. 2007) was used for the MTA analysis. The rMTA algorithm implemented in our in-house R package named gembox was used, with the academic version of IBM ILOG CPLEX Optimization Studio 12.10 as the optimization solver. To compare the MTA predictions across datasets, the top 10% predictions from each pair of datasets were tested for significant overlap using Fisher's exact tests.

Validation of the MTA-predicted anti-SARS-CoV-2 metabolic targets

Multiple datasets of reported anti-SARS-CoV-2 gene targets or drugs identified from large-scale chemical or genetic screens were collected to validate our predictions. Gene-level results of two published CRISPR-Cas9 genetic screens (Wei et al. 2020 and Daniloski et al. 2020) were obtained from the supplementary materials of the respective publication. For Wei et al., gene hits with FDR<0.1 and mean z score>0 (i.e. KO inhibits the viral infection) were taken; Daniloski et al. reported two screens with different multiplicities of infections (MOIs) and provided only single-sided FDR, so gene hits with FDR<0.1 from either screen were taken. The union set of hits from both studies were used. Lists of experimentally validated drugs reported in different studies compiled by Kuleshov et al. 2020 were downloaded from https://maayanlab.cloud/covid19/, which are then mapped to the genes they inhibit using data from DrugBank v5.1.7 (Wishart et al. 2018). Additionally, host proteins identified to interact with SARS-CoV-2 proteins were obtained from the supplementary materials of Gordon et al. 2020 and Stukalov et al. 2020. The genes from these validation datasets are mapped to metabolic reactions wherever applicable based on the human GEM Recon 1 (Duarte et al. 2007) data. Then, the significant overlap between the top 10% MTA-predicted targets from each dataset and each of the validation sets described above was tested with Fisher's exact tests on the reaction level (full results in Table S6B-D). Reaction-level test is performed because multiple reactions can be mapped to the same gene, and performing Fisher's exact test on the gene-level fails to consider such multiple mapping and is thus inappropriate.

For ROC analysis, negative sets (i.e. genes whose KO promotes SARS-CoV-2 infection) were defined based on the two CRISPR-Cas9 screens described above. For Wei et al., gene with FDR<0.1 and mean z score<0 were taken; since Daniloski et al. provided only single-sided FDR, the log fold-change threshold corresponding to the FDR<0.1 cutoff was identified, and genes with more extreme log fold-changes in the opposite direction were taken. The union of the negative sets from both studies was used. Both the positive (described in the previous paragraph) and negative sets of genes are then mapped to metabolic reactions as described above. The negative set defined as such contains a relatively balanced number of reactions compared to the positive set (131 vs 81). The rMTA score for the reactions produced by MTA was used as the predicted value for ROC analysis. The R package pROC (Robin et al. 2011)

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was used to compute the AUROC values and their 95% confidence intervals (the latter computed with bootstrapping).

Defining and analyzing the consensus set of candidate anti-SARS-CoV-2 metabolic targets across datasets

Based on top 10% MTA predictions from the 12 datasets (6 *in vitro* and 6 *in vivo*), the metabolic reaction targets that are recurrent in at least 3 of the *in vitro* datasets, and also in 3 of the *in vivo* datasets (i.e. the intersection of the two) were taken to be the final consensus candidate targets with high-confidence support across datasets. These target reactions are also mapped to genes using the human GEM Recon 1 data (Duarte et al. 2007), and further mapped to known drugs inhibiting the gene targets using data from DrugBank v5.1.7 (Wishart et al. 2018; target list given in Table S7A). The metabolic pathway enrichment of these targets was analyzed with Fisher's exact tests, with the "subSystems" slot in the metabolic model used as pathway annotation. For enrichment analysis of cellular transport reactions by each metabolite species, the reactions responsible for transporting each metabolite across the cell membrane were identified from Recon 1 and used as reaction sets for the Fisher's exact tests.

Preparation of Vero E6 cell samples with SARS-CoV-2 infection and remdesivir treatment, RNA-sequencing, and gene expression data analysis

Vero E6 cells (ATCC® CRL-1586™) were maintained in Dulbecco's modified eagle medium (DMEM, Gibco) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco), 50 U ml−1 penicillin, 50 µg ml−1 streptomycin, 1 mM sodium pyruvate (Gibco), 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Gibco), and 1× MEM non-essential amino acids solution (Gibco). The SARS-CoV-2 USA-WA1/2020 strain was obtained from BEI Resources (NR-52281). The virus was inoculated on Vero E6 cells and the cell supernatant was collected at 72 h post-inoculation (hpi), when extensive cytopathic effects were observed. The supernatant, after clarification by centrifugation 15 min at 4 °C at 5,000 xg, was aliquoted and stored at −80□°C until use. 500,000 Vero E6 cells were seeded in 6-well plates. The following day, the cell medium was replaced with fresh medium supplemented with either DMSO or 1 µM remdesivir (Adooq Bioscience), and cells were either mock-infected or infected with SARS-CoV-2 USA-WA1/2020 (MOI=0.3). Twenty-four hours after infection, cells were collected, and total intracellular RNA was extracted using the Qiagen® RNeasy® Plus Mini Kit. Three replicates were performed for each group, resulting in a total of six samples. The quality of the extracted RNA was assessed with Agilent® 2100 Bioanalyzer. Libraries were prepared on total RNA following ribosome RNA depletion with standard protocol according to Illumina®. Total RNA sequencing was then performed on the Illumina® NextSeq system, 150bp paired-end runs were performed and 100 million raw reads per sample were generated. STAR (Dobin et al. 2013) was used to align the reads to reference genome of the African green monkey (Chlorocebus sabaeus, https://useast.ensembl.org/Chlorocebus_sabaeus/Info/Annotation), with the SARS-CoV-2 genome (https://www.ncbi.nlm.nih.gov/nuccore/NC_045512) added to the reference genome. DESeq2 (Love et al. 2014) was used for DE analysis between pairs of experimental groups (including virus+remdesivir vs control and virus vs control; DE results in Table S8A). GSEA (Subramanian et al. 2005) was performed using the implementation in the R package

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fgsea (Korotkevich et al. 2019), results are provided in Table S8B. The gene set/pathway annotations used were the Reactome (Jassal et al. 2020) and KEGG (Kanehisa et al. 2021) subsets from the "Canonical Pathway" category in version 7.0 MSigDB database (Liberzon et al. 2011).

Genome-scale metabolic modeling of the remdesivir-treated Vero E6 cell samples and prediction of anti-SARS-CoV-2 metabolic targets in combination with remdesivir

As with the metabolic modeling of the other datasets on SARS-CoV-2 infection, iMAT (Shlomi et al. 2008) together with ACHR was used to compute the metabolic flux distribution for each of the experimental groups, using the median expression TPM values of each group as the input to iMAT. Reactions with differential fluxes (DF) between groups (including virus+remdesivir vs control and virus vs control) were identified as described above, and their significant metabolic pathway enrichment was tested with Fisher's exact tests, with pathways defined by the "subSystems" from the Recon 1 model (Duarte et al. 2007; results in Table S9). The DE result of virus+remdesivir vs control group and the mean flux distribution of the virus+remdesivir group computed with iMAT were used as inputs for MTA to predict metabolic reactions whose knock-out can further transform the virus+remdesivir metabolic state to the normal control state (results in Table S10A). The top 10% and 20% MTA-predicted targets were tested for significant enrichment for the targets of a list of experimentally validated synergistic drugs with remdesivir (Nguyenla et al. 2020) using Fisher's exact test (performed on the metabolic reaction level as described above). Metabolic pathway enrichment analysis of the top MTA-predicted targets was performed as described above (results in Table S10B).

Notes on statistical analysis and visualization

R version 3.6.3 was used for all statistical tests. P values lesser than 2.22e-16 may not be computed accurately and are reported as "P<2.22e-16" throughout the text. The Benjamini-Hochberg (BH) method was used for P value adjustment throughout the text. The R packages ggplot2 (Wickham 2016), ComplexHeatmap (Gu et al. 2016) and visNetwork (https://cran.r-project.org/web/packages/visNetwork/index.html) were used to create the visualizations.

Data and code availability

The gene expression data analyzed in this study are from published studies, with detailed information given in Table 1. The bulk RNA-seq data for SARS-CoV-2 infection in Vero E6 cells with remdesivir treatment will be deposited to the GEO database (dataset ID pending). The code used for the analyses can be found in the GitHub repository: https://github.com/ruppinlab/covid_metabolism.

Acknowledgement

This research was supported in part by the Intramural Research Program of the National Institutes of Health, NCI, and used the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We acknowledge and thank the National Cancer Institute for providing financial and infrastructural support. K.C. and S.S is supported by the NCI-UMD Partnership for Integrative Cancer Research Program.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conflict of interest

The authors declare no competing interests.

Figure Legends

Figure 1. Analysis of SARS-CoV-2-induced gene expression changes with 12 published datasets. (A) PCA plot using the rank-based inverse normal-transformed differential expression (DE) log fold-change values (virus-infected compared to control samples) across all the datasets analyzed. **(B)** Visualization of the overlap of the top significant DE genes (n=400, all at least have FDR<0.1) between each pair of datasets analyzed using Fisher's exact tests (Methods). The dot size corresponds to the effect size of the overlap as measured by odds ratio, and the color corresponds to the negative log10 adjusted one-sided P value (grey means below 0.05). **(C)** Visualization of the overlap of the top significantly enriched pathways (FDR<0.1) from the gene set enrichment analysis (GSEA) between each pair of datasets analyzed using Fisher's exact tests (Methods). The meanings of dot size and color are the same as (B), dots with black border correspond to infinity odds ratio. **(D)** A summary visualization of the GSEA result for the top consistently altered pathways during SARS-CoV-2 infection across the datasets, with more importance given to the various *in vivo* patient datasets (Methods). The dot color corresponds to the negative log10 adjusted P values from GSEA, with two sets of colors (red-orange and blue-purple) distinguishing up-regulation from down-regulation (positive or negative normalized enrichment scores, i.e. NES); dot size corresponds to the absolute value of NES measuring the strength of enrichment. The left and right-hand side blocks represent the pathways that tend to be consistently up-regulated and down-regulated in infected vs control samples, respectively; within each block, the pathways are ordered by negative sum of log P values across datasets (i.e. Fisher's method). **(E)** Heatmap summarizing the landscape of metabolic pathway alterations (based on gene expression) during SARS-CoV-2 across datasets. The heatmap color corresponds to the GSEA NES values (explained above) for KEGG metabolic pathways grouped into major categories. Only the metabolic pathways with FDR<0.2 enrichment in at least one dataset are included in the heatmap. The dataset labels used in this figure correspond to those given in Table 1.

Figure 2. Genome-scale metabolic modeling (GEM)-based analysis of SARS-CoV-2-induced metabolic alterations across datasets. GEM was used to compute the metabolic fluxes from the gene expression profiles, and reactions with differential fluxes (DF) between the SARS-CoV-2-infected and control groups were identified for each dataset (Methods). **(A)** Visualization of the overlap of the top DF reactions between each pair of datasets analyzed using Fisher's exact tests (Methods). The dot size corresponds to the effect size of the overlap

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as measured by odds ratio, and the color corresponds to the negative log10 adjusted one-sided P value (grey means below 0.05). **(B)** A summary visualization of the metabolic pathway enrichment result for the top consistent DF reactions across the datasets, with more importance given to the various *in vivo* patient datasets (Methods). Y-axis represents the odds ratio of enrichment, the dot color corresponds to the adjusted P value from Fisher's exact tests, and dot size corresponds to the number of enriched reactions within each pathway. Half-dots plotted on the top border line correspond to infinity odds ratio values. The pathways on the X-axis are ordered by P value and only those with FDR<0.1 are shown. **(C)** A summary visualization of the top metabolites with consistently altered cellular transport patterns during SARS-CoV-2 infection across the datasets, with more importance given to the various *in vivo* patient datasets (Methods). The dot color corresponds to the rank biserial correlation (r_{rb}) as one measure of the DF effect size, with positive value (red color) corresponding to higher net cellular import from the extracellular space or lower net export in infected vs control samples, vice versa for negative value (blue color); dot size corresponds to another measure of DF effect size, which is the absolute percentage of flux value change relative to the flux in the control group (%dflux; Methods). The left-hand side block contains metabolites that tend to show overall higher net cellular import (or lower export) in infected vs control samples across many datasets (despite the lack of full coherence), vice versa for the right-hand side block. **(D-H)** Visualization of the relatively consistent DF patterns in selected enriched pathways. The metabolic network is based on the human GEM Recon 1 (Duarte et al. 2007), and pathways are taken from the "subSystems" annotation in the GEM. Metabolites are represented by nodes, reactions are represented by directed (hyper) edges, with edge direction corresponding to the consensus reaction direction and edge color corresponding to the consensus DF direction across datasets (Methods). Red and blue colors correspond to increased and decreased fluxes, respectively; grey color corresponds to reactions not showing consistent DF changes across datasets, some of such reactions are not shown to increase clarity. **(D)** Pyrimidine biosynthesis. **(E)** TCA cycle (left-hand side); the right-hand side contains the cytoplasmic counterparts of the TCA cycle intermediate metabolites, their interconversion and cytosol-mitochondrial transport. **(F)** Inositol phosphate metabolism. **(G)** Glycine, serine, and threonine metabolism. **(H)** Fatty acid elongation. Metabolites are labeled by their names in (D,E,G) or IDs from Recon 1 in (F,H), with suffixes denoting their cellular compartments: [c] cytosol; [m] mitochondria. The mapping between the IDs and metabolite names in (F,H) is given in Table S5D.

Figure 3. Genome-scale metabolic modeling (GEM)-based prediction of anti-SARS-CoV-2 targets that act via reversing the virus-induced metabolic alterations. The robust metabolic transformation algorithm (rMTA, Valcárcel et al. 2019) was used to predict metabolic reactions whose knock-out can reverse the SARS-CoV-2-induced metabolic changes using each of the collected datasets (Methods). **(A)** Visualization of the overlap of the top 10% MTA-predicted target reactions between each pair of datasets analyzed using Fisher's exact tests (Methods). The dot size corresponds to the effect size of the overlap as measured by odds ratio, and the color corresponds to the negative log10 adjusted one-sided P value (grey means below 0.05). **(B)** A summary visualization of the enrichment of the top 10% MTA-predicted targets from each dataset for the antiviral hits (i.e. those whose KO inhibits SARS-CoV-2 infection) identified in the two published CRISPR-Cas9 screens (Wei et al. 2020 and Daniloski et al. 2020; Methods). Y-

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

axis represents the odds ratio of enrichment, the dot color corresponds to the negative log10 adjusted one-sided P value from Fisher's exact tests, and dot size corresponds to the number of enriched target reactions. The datasets are ordered by P values, the first 8 datasets have FDR<0.1. **(C)** cases of significant enrichment (FDR<0.1) of top 10% MTA-predicted targets from each dataset for the experimentally validated anti-SARS-CoV-2 drug sets from previous studies (compiled by Kuleshov et al. 2020). Axes and the meanings of dot color and size are similar to (B) but the axes are flipped and the adjusted P values are not log-transformed. Half-dots plotted on the top border line correspond to infinity odds ratio values. **(D)** significant cases of significant enrichment (FDR<0.1) of top 10% MTA-predicted targets from each dataset for the host proteins involved in host-SARS-CoV-2 protein-protein interactions (from Stukalov et al. 2020; data from Gordon et al. 2020 was also analyzed but no significant enrichment was found). Axes and the meanings of dot color and size are similar to (C) but raw P values instead of adjusted P values are plotted (all the displayed cases have adjusted P=0.08). **(E)** A summary of the area under ROC curve (AUROC) value of the MTA prediction using each dataset, based on positive and negative sets (i.e. genes whose KO inhibits or promotes the viral infection, respectively) from the two published CRISPR-Cas9 screen data (Wei et al. 2020 and Daniloski et al. 2020; Methods). The error bars (vertical lines through the dots) represent 95% confidence intervals. **(F)** Example ROC curves from two of the best-performing datasets. The curves are colored by color gradients corresponding to the threshold of top MTA predictions. **(G)** Summary of the metabolic pathways significantly enriched (FDR<0.1) by the final list of consensus candidate targets identified across datasets (Methods). Three significant pathways including cellular transport were identified (the first three items on the X axis). The rest of the items on the X axis are for significant enrichment results (FDR<0.1) by each metabolite species within the cellular transport pathway. Y axis and the meanings of dot color and size are similar to (B).

Figure 4. Analysis of the gene expression and metabolic flux profile of remdesivir treatment and prediction of metabolic targets for anti-SARS-CoV-2 in combination with remdesivir. (A) PCA plot of the gene expression profiles for the Vero E6 samples from all experimental groups: control (no virus or remdesivir treatment), virus (SARS-CoV-2-infected), virus+remdesivir (SARS-CoV-2-infected treated by remdesivir), remdesivir (remdesivir treatment alone without virus). There are 3 replicates in each group. **(B)** A visualization of selected differentially expressed pathways comparing the virus+remdesivir group to the control group using gene set enrichment analysis (GSEA). Y-axis represents normalized enrichment score (NES), positive value means higher expression in the virus+remdesivir group compared to control, vice versa. Pathways on the X-axis are ordered by their NES values, all pathways displayed have adjusted P<0.05; pathways with names in red are those that are not significantly different (adjusted P>0.2) when comparing the virus group to the control group using GSEA, i.e. changes in these pathways may arise from remdesivir-specific effects. The dot color corresponds to the negative log10 adjusted GSEA P value, and dot size corresponds to the number of enriched genes (i.e. "leading edge" genes in GSEA). **(C)** PCA plot of the average metabolic flux profile computed using the iMAT algorithm (Shlomi et al. 2008; Methods) representative of each of the experimental groups, the labels are the same as (A). **(D)** A visualization of metabolic pathway enrichment result of the differential metabolic fluxes in the virus+remdesivir group vs the control group, using Fisher's exact tests (Methods). Y-axis

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

represents the odds ratio of enrichment, the horizontal dashed line corresponds to odds ratio of 1. The dot color corresponds to the negative log10 adjusted one-sided P value from Fisher's exact tests, and dot size corresponds to the number of enriched target reactions. The datasets are ordered by P values, all pathways displayed have FDR<0.1. **(E)** The robust metabolic transformation algorithm (rMTA, Valcárcel et al. 2019) was used to predict metabolic reactions whose knock-out can further transform the metabolic state of the remdesivir-treated SARS-CoV-2-infected cells back to the normal control state, using the Vero E6 cell samples (Methods). The significant metabolic pathways (FDR<0.1) enriched by the top 20% MTA-predicted targets are shown. Axes and the meanings of dot color and size are similar to (D). The pathways with names in bold text are also significantly enriched (FDR<0.1) by the top 10% MTA-predicted targets.

References

Auslander N, Cunningham CE, Toosi BM, et al. An integrated computational and experimental study uncovers FUT9 as a metabolic driver of colorectal cancer. Mol Syst Biol. 2017;13(12):956. Published 2017 Dec 1. doi:10.15252/msb.20177739

Baart GJ, Martens DE. Genome-scale metabolic models: reconstruction and analysis. Methods Mol Biol. 2012;799:107-126. doi:10.1007/978-1-61779-346-2_7

Barberis E, Timo S, Amede E, et al. Large-Scale Plasma Analysis Revealed New Mechanisms and Molecules Associated with the Host Response to SARS-CoV-2. Int J Mol Sci. 2020;21(22):8623. Published 2020 Nov 16. doi:10.3390/ijms21228623

Beigel JH, Tomashek KM, Dodd LE, et al. ACTT-1 Study Group Members. Remdesivir for the treatment of Covid-19—Final report. N Engl J Med. 2020 Nov 5;383(19):1813-1826. doi: 10.1056/NEJMoa2007764.

Beziau A, Brand D, Piver E. The Role of Phosphatidylinositol Phosphate Kinases during Viral Infection. Viruses. 2020;12(10):1124. Published 2020 Oct 3. doi:10.3390/v12101124

Blanco-Melo D, Nilsson-Payant BE, Liu WC, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell. 2020;181(5):1036-1045.e9. doi:10.1016/j.cell.2020.04.026

Bojkova D, Klann K, Koch B, et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature. 2020;583(7816):469-472. doi:10.1038/s41586-020-2332-7

Bojkova D, Costa R, Bechtel M, Ciesek S, Michaelis M, Cinatl J Jr. Targeting pentose phosphate pathway for SARS-CoV-2 therapy. Preprint. bioRxiv. 2020;2020.08.19.257022. Published 2020 August 21. doi: https://doi.org/10.1101/2020.08.19.257022

Butler DJ, Mozsary C, Meydan C, et al. Shotgun Transcriptome and Isothermal Profiling of SARS-CoV-2 Infection Reveals Unique Host Responses, Viral Diversification, and Drug Interactions. Preprint. bioRxiv. 2020;2020.04.20.048066. Published 2020 May 1. doi:10.1101/2020.04.20.048066

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, et al. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. Nat Biotechnol. 2020 Aug;38(8):970-979. doi: 10.1038/s41587-020-0602-4.

Daniloski Z, Jordan TX, Wessels HH, et al. Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells [published online ahead of print, 2020 Oct 24]. Cell. 2020;S0092-8674(20)31394-5. doi:10.1016/j.cell.2020.10.030

Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

Dong Y, Dai T, Wei Y, Zhang L, Zheng M, Zhou F. A systematic review of SARS-CoV-2 vaccine candidates. Signal Transduct Target Ther. 2020;5(1):237. Published 2020 Oct 13. doi:10.1038/s41392-020-00352-y

Duarte NC, Becker SA, Jamshidi N, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A. 2007;104(6):1777-1782. doi:10.1073/pnas.0610772104

Ehrlich A, Uhl S, Ioannidis K, Hofree M, tenOever BR and Nahmias Y. The SARS-CoV-2 Transcriptional Metabolic Signature in Lung Epithelium. 2020. Available at SSRN: https://ssrn.com/abstract=3650499 or http://dx.doi.org/10.2139/ssrn.3650499

Finak G, McDavid A, Yajima M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015;16:278. Published 2015 Dec 10. doi:10.1186/s13059-015-0844-5

Gardinassi LG, Souza COS, Sales-Campos H, Fonseca SG. Immune and Metabolic Signatures of COVID-19 Revealed by Transcriptomics Data Reuse. Front Immunol. 2020;11:1636. Published 2020 Jun 26. doi:10.3389/fimmu.2020.01636

Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583(7816):459-468. doi:10.1038/s41586-020-2286-9

Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. Genome Biol. 2019;20(1):121. Published 2019 Jun 13. doi:10.1186/s13059-019-1730-3

Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-2849. doi:10.1093/bioinformatics/btw313

Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;181(2):271-280.e8. doi:10.1016/j.cell.2020.02.052

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

International Clinical Trials Registry Platform (ICTRP). COVID-19 trials [updated on 4 December 2020]. Available online: https://www.who.int/clinical-trials-registry-platform (accessed on December 10, 2020); data from: https://www.who.int/ictrp/COVID19-web.csv

Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2020;48(D1):D498-D503. doi:10.1093/nar/gkz1031

Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545-D551. doi:10.1093/nar/gkaa970

Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis. 2019. Preprint. bioRxiv. doi:10. 1101/060012.

Kuleshov MV, Stein DJ, Clarke DJB, et al. The COVID-19 Drug and Gene Set Library. Patterns (N Y). 2020;1(6):100090. doi:10.1016/j.patter.2020.100090

Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29. Published 2014 Feb 3. doi:10.1186/gb-2014-15-2-r29

Li S, Ma F, Yokota T, et al. Metabolic reprogramming and epigenetic changes of vital organs in SARS-CoV-2 induced systemic toxicity [published online ahead of print, 2020 Dec 7]. JCI Insight. 2020;145027. doi:10.1172/jci.insight.145027

Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F, Liu L, Amit I, Zhang S, Zhang Z. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat Med. 2020 Jun;26(6):842-844. doi: 10.1038/s41591-020-0901-9.

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP, Molecular signatures database (MSigDB) 3.0, Bioinformatics, 2011;27(12):1739–1740, https://doi.org/10.1093/bioinformatics/btr260

Lieberman NAP, Peddu V, Xie H, et al. In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. PLoS Biol. 2020;18(9):e3000849. Published 2020 Sep 8. doi:10.1371/journal.pbio.3000849

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett. 2009;583(24):3966-3973. doi:10.1016/j.febslet.2009.10.036

Mayer KA, Stöckl J, Zlabinger GJ, Gualdoni GA. Hijacking the Supplies: Metabolism as a Novel Facet of Virus-Host Interaction. Front Immunol. 2019;10:1533. Published 2019 Jul 3. doi:10.3389/fimmu.2019.01533

National Institutes of Health. COVID-19 Treatment Guidelines. 2020. Available online: https://www.covid19treatmentguidelines.nih.gov/immune-based-

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

therapy/immunomodulators/corticosteroids (updated on November 3, 2020; accessed on December 10, 2020)

Nguyenla X, Wehri E, Dis EV, Biering SB, Yamashiro LH, Stroumza J, Dugast-Darzacq C, Graham T, Stanley S, Schaletzky J. Discovery of SARS-CoV-2 antiviral synergy between remdesivir and approved drugs in human lung cells. Preprint. bioRxiv. 2020;2020.09.18.302398; doi: https://doi.org/10.1101/2020.09.18.302398

Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2020;11(1):1620. Published 2020 Mar 27. doi:10.1038/s41467-020-15562-9

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417-419. doi:10.1038/nmeth.4197

RECOVERY Collaborative Group, Horby P, Lim WS, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report [published online ahead of print, 2020 Jul 17]. N Engl J Med. 2020;NEJMoa2021436. doi:10.1056/NEJMoa2021436

Riva L, Yuan S, Yin X, et al. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. Nature. 2020;586(7827):113-119. doi:10.1038/s41586-020-2577-1

Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77. Published 2011 Mar 17. doi:10.1186/1471-2105-12-77

Sanchez EL, Lagunoff M. Viral activation of cellular metabolism. Virology. 2015;479-480:609-618. doi:10.1016/j.virol.2015.02.038

Shen B, Yi X, Sun Y, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell. 2020;182(1):59-72.e15. doi:10.1016/j.cell.2020.05.032

Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. Network-based prediction of human tissue-specific metabolism. Nat Biotechnol. 2008;26(9):1003-1010. doi:10.1038/nbt.1487

Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-1902.e21. doi:10.1016/j.cell.2019.05.031

Stukalov A, Girault V, Grass V, et al. Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. Preprint. bioRxiv. 2020;2020.06.17.156455. Published 2020 Jun 17. doi:10.1101/2020.06.17.156455

Styr B, Gonen N, Zarhin D, et al. Mitochondrial Regulation of the Hippocampal Firing Rate Set Point and Seizure Susceptibility. Neuron. 2019;102(5):1009-1024.e8. doi:10.1016/j.neuron.2019.03.045

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Subramanian B, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS. Oct 2005;102(43):15545-15550. doi: 10.1073/pnas.0506580102

Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol. 2020;20(6):363-374. doi:10.1038/s41577-020-0311-8

Thomas T, Stefanoni D, Reisz JA, et al. COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. JCI Insight. 2020;5(14):e140327. Published 2020 Jul 23. doi:10.1172/jci.insight.140327

U.S. Food and Drug Administration. Coronavirus (COVID-19) Update: FDA Authorizes Monoclonal Antibodies for Treatment of COVID-19. 2020. Available online: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-monoclonal-antibodies-treatment-covid-19 (released on November 21, 2020; accessed on December 10, 2020)

U.S. Food and Drug Administration. COVID-19 Vaccines. 2020. Available online: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines (accessed on December 28, 2020)

Valcárcel LV, Torrano V, Tobalina L, Carracedo A, Planes FJ. rMTA: robust metabolic transformation analysis. Bioinformatics. 2019;35(21):4350-4355. doi:10.1093/bioinformatics/btz231

Wei J, Alfajaro MM, DeWeirdt PC, et al. Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection [published online ahead of print, 2020 Oct 20]. Cell. 2020;S0092-8674(20)31392-1. doi:10.1016/j.cell.2020.10.028

Weingarten-Gabbay S, Klaeger S, Sarkizova S, et al. SARS-CoV-2 infected cells present HLA-I peptides from canonical and out-of-frame ORFs. Preprint. bioRxiv. 2020;2020.10.02.324145. Published 2020 Oct 2. doi:10.1101/2020.10.02.324145

WHO Solidarity Trial Consortium, Pan H, Peto R, et al. Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results [published online ahead of print, 2020 Dec 2]. N Engl J Med. 2020;NEJMoa2023184. doi:10.1056/NEJMoa2023184

Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. Springer-Verlag New York. ISBN 978-3-319-24277-4.

Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-D1082. doi:10.1093/nar/gkx1037

Xiong Y, Liu Y, Cao L, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microbes Infect. 2020;9(1):761-770. doi:10.1080/22221751.2020.1747363

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Xiong R, Zhang L, Li S, et al. Novel and potent inhibitors targeting DHODH are broad-spectrum antivirals against RNA viruses including newly-emerged coronavirus SARS-CoV-2 [published correction appears in Protein Cell. 2020 Oct 8;:] [published correction appears in Protein Cell. 2020 Nov 9;:]. Protein Cell. 2020;11(10):723-739. doi:10.1007/s13238-020-00768-w

Yan B, Chu H, Yang D, et al. Characterization of the Lipidomic Profile of Human Coronavirus-Infected Cells: Implications for Lipid Metabolism Remodeling upon Coronavirus Replication. Viruses. 2019;11(1):73. Published 2019 Jan 16. doi:10.3390/v11010073

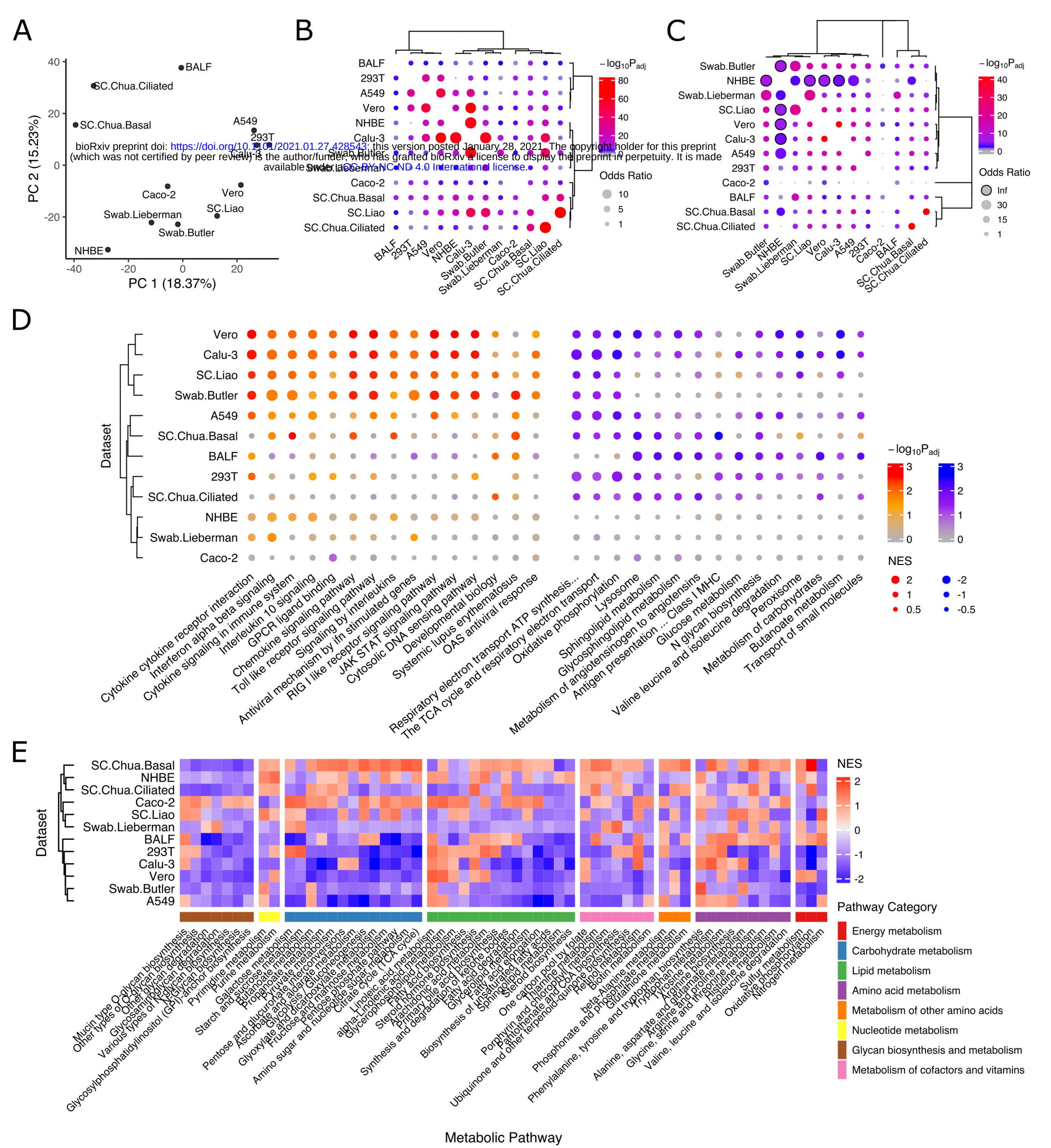
Yizhak K, Gabay O, Cohen H, Ruppin E. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. Nat Commun. 2013;4:2632. doi:10.1038/ncomms3632

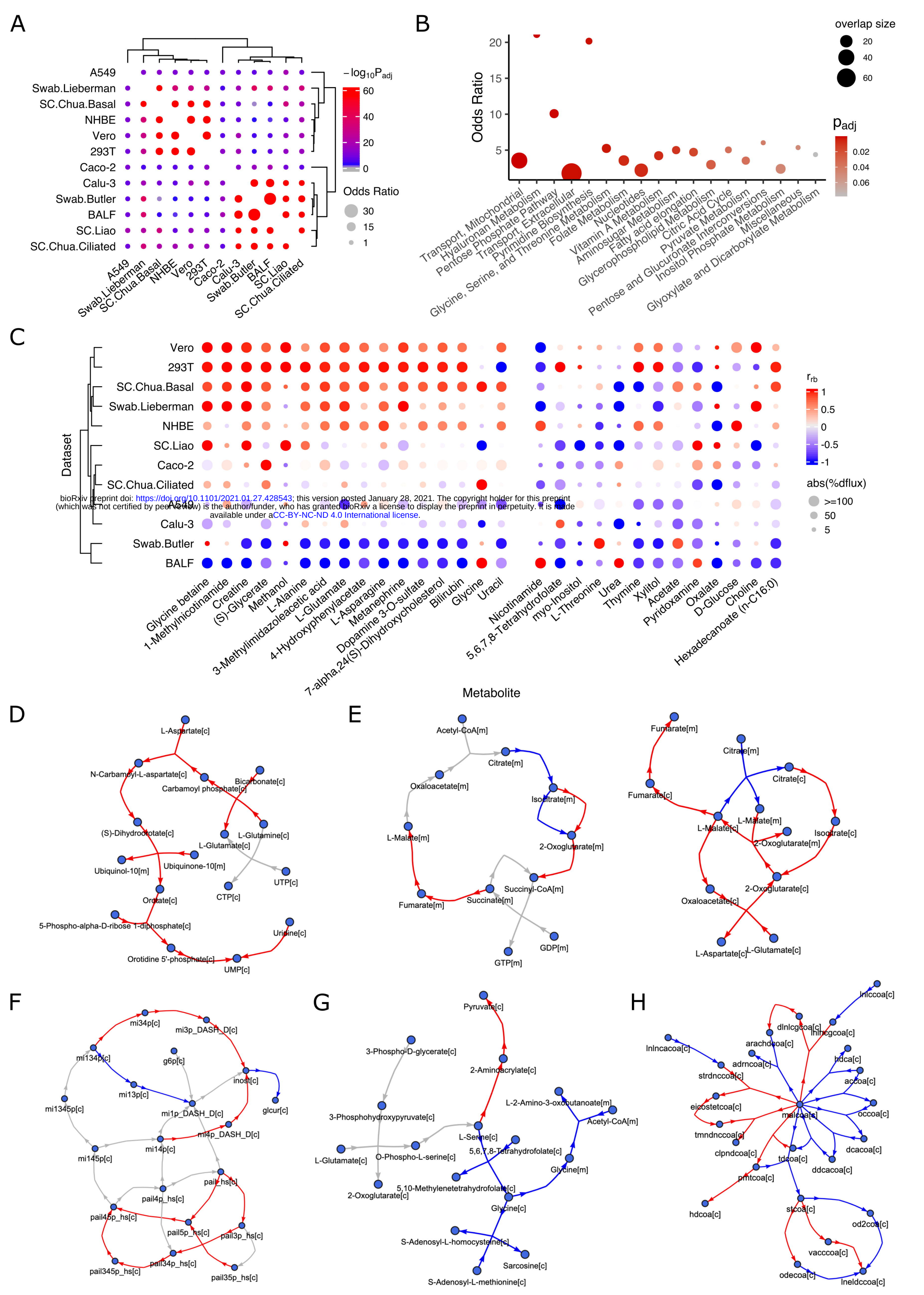
Yuan S, Chu H, Chan JF, et al. SREBP-dependent lipidomic reprogramming as a broad-spectrum antiviral target. Nat Commun. 2019;10(1):120. Published 2019 Jan 10. doi:10.1038/s41467-018-08015-x

Zheng HY, Xu M, Yang CX, et al. Longitudinal transcriptome analyses show robust T cell immunity during recovery from COVID-19. Signal Transduct Target Ther. 2020;5(1):294. Published 2020 Dec 24. doi:10.1038/s41392-020-00457-4

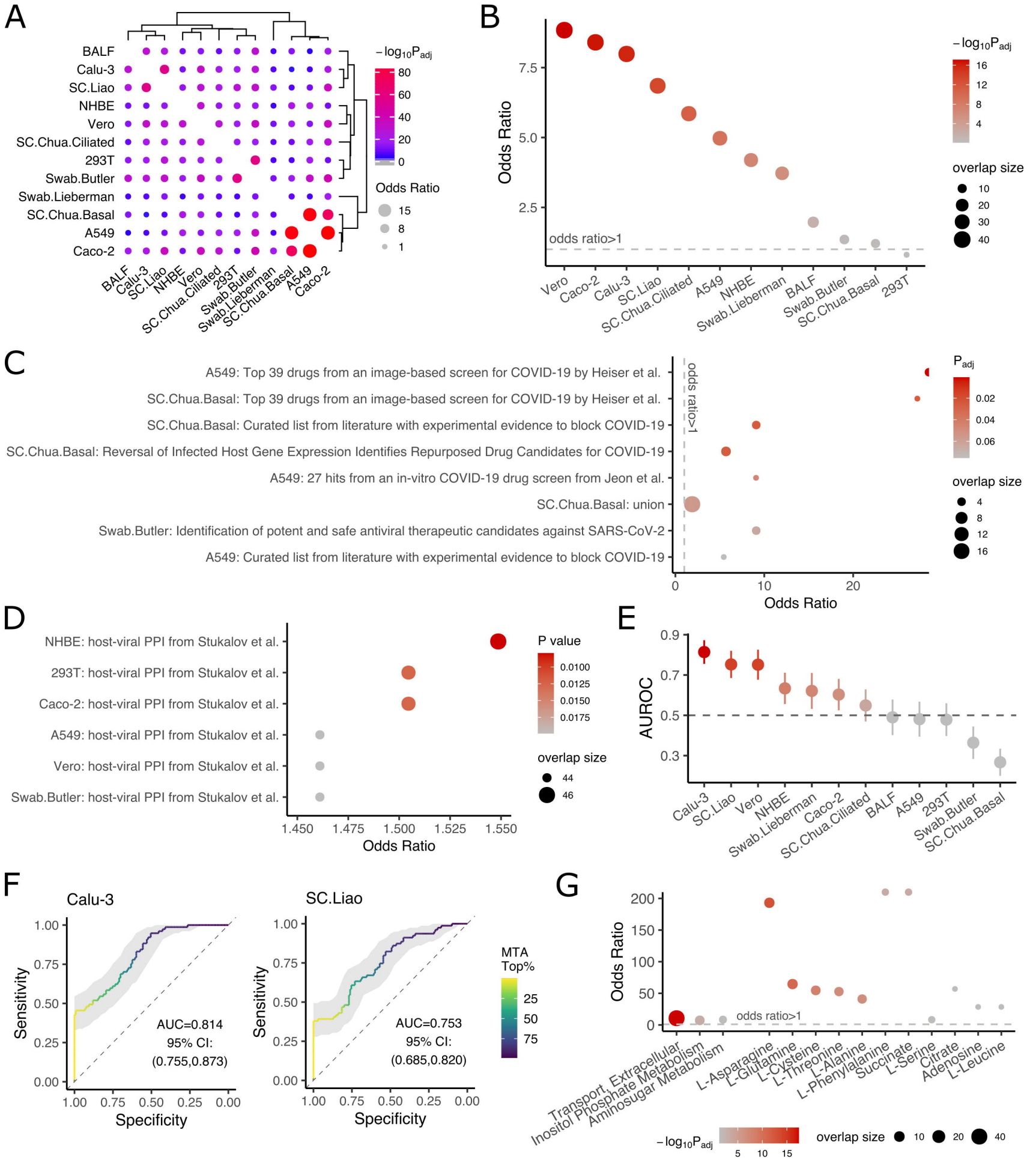
Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7

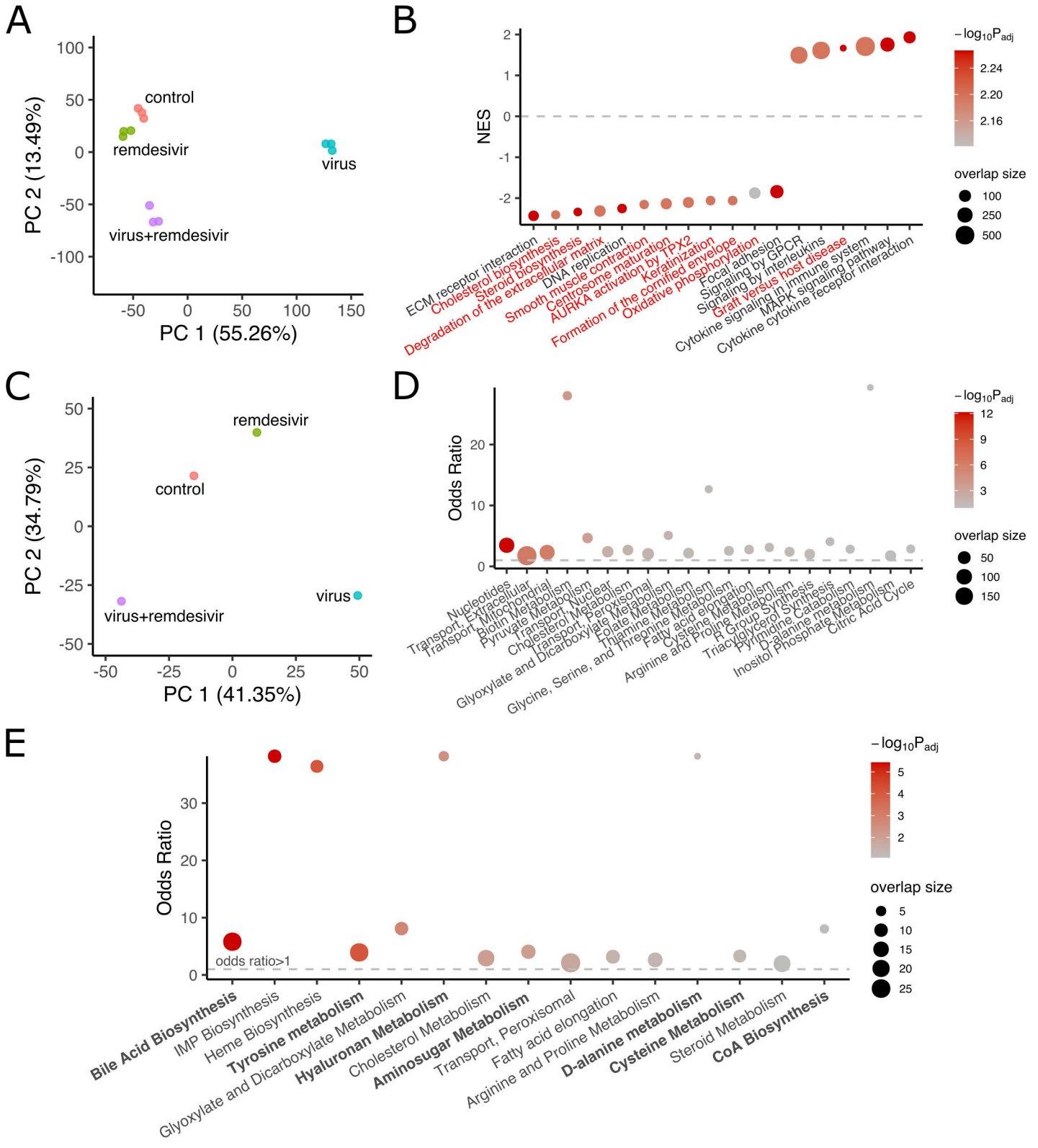
bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.





bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428543; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


A
BALF
Calu-3
SC.Liao
NHBE
Vero
SC.Chua.Ciliated
293T
Swab.Butler
Swab.Lieberman
SC.Chua.Basal
A549
Caco-2
$-\log_{10}P_{adj}$
80
60
40
20
0
Odds Ratio
15
8
1
B
Odds Ratio
7.5
5.0
2.5
odds ratio>1
Vero
Caco-2
Calu-3
SC.Liao
SC.Chua.Ciliated
A549
NHBE
Swab.Lieberman
BALF
Swab.Butler
SC.Chua.Basal
293T
$-\log_{10}P_{adj}$
16
12
8
4
overlap size
10
20
30
40
C
A549: Top 39 drugs from an image-based screen for COVID-19 by Heiser et al.
SC.Chua.Basal: Top 39 drugs from an image-based screen for COVID-19 by Heiser et al.
SC.Chua.Basal: Curated list from literature with experimental evidence to block COVID-19
SC.Chua.Basal: Reversal of Infected Host Gene Expression Identifies Repurposed Drug Candidates for COVID-19
A549: 27 hits from an in-vitro COVID-19 drug screen from Jeon et al.
SC.Chua.Basal: union
Swab.Butler: Identification of potent and safe antiviral therapeutic candidates against SARS-CoV-2
A549: Curated list from literature with experimental evidence to block COVID-19
odds ratio>1
0
10
20
Odds Ratio
P_{adj}
0.02
0.04
0.06
overlap size
4
8
12
16
D
NHBE: host-viral PPI from Stukalov et al.
293T: host-viral PPI from Stukalov et al.
Caco-2: host-viral PPI from Stukalov et al.
A549: host-viral PPI from Stukalov et al.
Vero: host-viral PPI from Stukalov et al.
Swab.Butler: host-viral PPI from Stukalov et al.
1.450
1.475
1.500
1.525
1.550
Odds Ratio
P value
0.0100
0.0125
0.0150
0.0175
overlap size
44
46
E
AUROC
0.9
0.7
0.5
0.3
Calu-3
SC.Liao
Vero
NHBE
Swab.Lieberman
Caco-2
SC.Chua.Ciliated
BALF
A549
293T
Swab.Butler
SC.Chua.Basal
F
Calu-3
SC.Liao
Sensitivity
1.00
0.75
0.50
0.25
0.00
Specificity
AUC=0.814
95% CI:
(0.755,0.873)
AUC=0.753
95% CI:
(0.685,0.820)
MTA
Top%
25
50
75
G
Odds Ratio
200
150
100
50
0
odds ratio>1
Transport, Extracellular
Inositol Phosphate Metabolism
Aminosugar Metabolism
L-Asparagine
L-Glutamine
L-Cysteine
L-Threonine
L-Alanine
L-Phenylalanine
Succinate
L-Serine
Citrate
Adenosine
L-Leucine
$-\log_{10}P_{adj}$
5
10
15
overlap size
10
20
40


A
control
remdesivir
virus
virus+remdesivir
PC 2 (13.49%)
PC 1 (55.26%)
B
NES
ECM receptor interaction
Cholesterol biosynthesis
Steroid biosynthesis
Degradation of the extracellular matrix
DNA replication
Smooth muscle contraction
Centrosome maturation
AURKA activation by TPX2
Keratinization
Formation of the cornified envelope
Oxidative phosphorylation
Focal adhesion
Graft versus host disease
Signaling by GPCR
Signaling by interleukins
Cytokine signaling in immune system
MAPK signaling pathway
Cytokine cytokine receptor interaction
$-\log_{10}P_{adj}$
overlap size
C
remdesivir
control
virus
virus+remdesivir
PC 2 (34.79%)
PC 1 (41.35%)
D
Odds Ratio
Nucleotides
Transport, Extracellular
Transport, Mitochondria
Biotin Metabolism
Pyruvate Metabolism
Transport, Nuclear
Cholesterol Metabolism
Transport, Peroxisomal
Glyoxylate and Dicarboxylate Metabolism
Folate Metabolism
Glycine, Serine, and Threonine Metabolism
Thiamine Metabolism
Fatty acid elongation
Cysteine Metabolism
Arginine and Proline Metabolism
B Group Synthesis
Triacylglycerol Synthesis
Pyrimidine Catabolism
D-alanine metabolism
Inositol Phosphate Metabolism
Citric Acid Cycle
$-\log_{10}P_{adj}$
overlap size
E
Odds Ratio
odds ratio>1
Bile Acid Biosynthesis
IMP Biosynthesis
Heme Biosynthesis
Tyrosine metabolism
Glyoxylate and Dicarboxylate Metabolism
Hyaluronan Metabolism
Cholesterol Metabolism
Aminosugar Metabolism
Transport, Peroxisomal
Fatty acid elongation
Arginine and Proline Metabolism
D-alanine metabolism
Cysteine Metabolism
Steroid Metabolism
CoA Biosynthesis
$-\log_{10}P_{adj}$
overlap size