

# Rapid Identification of Druggable Targets and the Power of the PHENotype SIMulator for Effective Drug Repurposing in COVID-19

Naomi MARIA (✉ [dr.nimaria@gmail.com](mailto:dr.nimaria@gmail.com))

The Feinstein Institutes for Medical Research <https://orcid.org/0000-0002-3452-2515>

Rosaria Valentina Rapicavoli

University of Catania

Salvatore Alaimo

University of Catania

Evelyne Bischof

School of Clinical Medicine, Shanghai University of Medicine and Health Sciences

Alessia Stasuzzo

University of Catania

Jantine Broek

Courant Institute NYU

Alfredo Pulvirenti

University of Catania <https://orcid.org/0000-0002-9764-0295>

Bud Mishra

New York University

Ashley Duits

Red Cross Blood Bank Foundation, Curaçao <https://orcid.org/0000-0002-6368-8266>

Alfredo Ferro

University of Catania

---

## Article

**Keywords:** COVID-19, drug repurposing, PHENotype SIMulator, druggable targets

**Posted Date:** April 14th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-287183/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Rapid Identification of Druggable Targets and the Power of the PHENotype SIMulator for Effective Drug Repurposing in COVID-19

**Naomi I. Maria**<sup>1,2,#\*</sup>, **Rosaria Valentina Rapicavoli**<sup>3,4,#</sup>, **Salvatore Alaimo**<sup>4,#</sup>, **Evelyne Bischof**<sup>5,6,7</sup>, **Alessia Stasuzzo**<sup>8</sup>, **Jantine A.C. Broek**<sup>9</sup>, **Alfredo Pulvirenti**<sup>4</sup>, **Bud Mishra**<sup>9,10,#\*</sup>, **Ashley J. Duits**<sup>2,11,#</sup>, **Alfredo Ferro**<sup>4,#</sup>

<sup>1</sup>Institute of Molecular Medicine, The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA, <sup>2</sup>Red Cross Blood Bank Foundation Curaçao, Willemstad, Curaçao. <sup>3</sup>Department of Physics and Astronomy, University of Catania, <sup>4</sup>Bioinformatics Unit, Department of Clinical and Experimental Medicine, University of Catania, Italy. <sup>5</sup>Department of Advanced Biomedical Sciences, University of Naples Federico II, Via Pansini, Naples, Italy, <sup>6</sup> School of Clinical Medicine, Shanghai University of Medicine and Health Sciences, Pudong, Shanghai, China, <sup>7</sup> Insilico Medicine, Hong Kong Special Administrative Region, China. <sup>8</sup>Department of Chemical Sciences, University of Catania, Italy. <sup>9</sup>Department of Computer Science, Mathematics, Engineering and Cell Biology, Courant Institute, Tandon and School of Medicine, New York University, New York, USA. <sup>10</sup>Simon Center for Quantitative Biology, Cold Spring Harbor Lab, Long Island, USA. <sup>11</sup>Curaçao Biomedical Health Research Institute, Willemstad, Curaçao

#These Authors contributed equally

\*Corresponding authors:

**Bud Mishra:** [mishra@nyu.edu](mailto:mishra@nyu.edu)

Courant Institute of Mathematical Sciences  
Room 402, 251 Mercer Street, New York, NY 10012  
516-717-9432(mobile), 212-998-3464 (office),

**Naomi I. Maria:** [dr.nimaria@gmail.com](mailto:dr.nimaria@gmail.com); [nimaria@northwell.edu](mailto:nimaria@northwell.edu)

## ABSTRACT

The current, rapidly diversifying pandemic has accelerated the need for efficient and effective identification of potential drug candidates for COVID-19. Knowledge on host-immune response to SARS-CoV-2 infection, however, remains limited with very few drugs approved to date.

Viable strategies and tools are rapidly arising to address this, especially with repurposing of existing drugs offering significant promise. **Here we introduce a systems biology tool, the**

**PHENotype SIMulator, which – by leveraging available transcriptomic and proteomic**

**databases – allows modeling of SARS-CoV-2 infection in host cells in silico to *i*) determine**

with high sensitivity and specificity (both>96%) the viral effects on cellular host-immune

response, resulting in a specific cellular SARS-CoV-2 signature and *ii*) utilize this specific

signature to narrow down promising repurposable therapeutic strategies. Powered by this tool,

coupled with domain expertise, we have identified several potential COVID-19 drugs including

methylprednisolone and metformin, and further discern key cellular SARS-CoV-2-affected

pathways as potential new drugable targets in COVID-19 pathogenesis.

## INTRODUCTION

The rapid emergence and spread of the virulent novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has hijacked and largely disrupted human civilization as we know it, bringing about countless global challenges but also the urgent need for innovative vaccine and drug discovery<sup>1</sup>. Soon after its emergence in Wuhan (Eastern China) in late 2019, coronavirus disease 2019 (COVID-19) was declared a pandemic by the World Health Organization<sup>2</sup>, as it continuously spreads and holds the world hostage. As of February 2021, over 100 million COVID-19 cases and 2 million deaths have been reported worldwide. It is apparent that our civilized, well-organized and hitherto functioning societies were not adequately prepared nor equipped to deal with the high infectivity, transmissibility, mortality and global impact of the COVID-19 pandemic. While vaccine development and deployment is well underway, widespread distribution remains challenging, and at present only an antiviral (Remdesivir) and glucocorticoids (Dexamethasone/ Methylprednisolone) have been approved for treatment of severe COVID-19. Recently, the virus-neutralizing antibody cocktail (Casirivimab and Imdevimab, termed REGN-COV2) also received emergency use authorization for treatment of mild to moderate COVID-19 in high-risk patients<sup>3</sup>. Otherwise, no established drug is available to prevent or adequately treat COVID-19 and in the absence of a clear etiological understanding, treatment has remained largely supportive and symptomatic<sup>4,5</sup>. Making matters worse, the mutating virus is now posing additional challenges.

Therefore, next to *in vitro* studies, *in silico* studies are of great value for rapid and effective drug discovery. Indeed, computational structure-based drug design and immunoinformatics have recently resulted in identification of potential SARS-CoV-2 target proteins and drugs that are being selected for further testing<sup>4,6,7</sup>. Another promising avenue for obtaining effective and readily available therapeutic strategies is the repurposing of drugs already approved for other indications. Drug repurposing strategies provide an attractive and effective

approach based on available drug characteristics – drug-related pharmacology and toxicology – for rapid therapeutic selection<sup>8</sup>. If we could, with higher probability, identify and pre-select the most promising hypothesis-based candidates using *in silico* systems biology tools, prior to costly and laborious *in vitro* and *in vivo* experiments and ensuing clinical trials, we could significantly improve disease-specific drug development<sup>9</sup>.

Several *in silico* techniques have been developed, mainly making use of molecular modeling of key viral proteins for virtual screening of drug candidates simulating receptor-drug molecular dynamics<sup>5,6</sup>. In order to increase the effectivity of identifying candidate drugs for combating COVID-19, it is crucial to build on a more in-depth knowledge of the molecular basis of the immune signaling pathways regarding host-virus interaction and SARS-CoV-2-induced immunopathology. Only if we better understand how this particular virus affects host cells in detail, on a transcriptomic, proteomic level and beyond<sup>4,5,10</sup>, will we be able to effectively treat COVID-19 patients. It is becoming evident that treatment should not only focus on direct anti-viral effects in mild cases but should also encompass potential (cytokine storm induced) aberrant host-response in severe cases<sup>4,11,12</sup>. Taken together, this points towards the importance of a more detailed and targeted approach for COVID-19, where antivirals or steroids alone might not suffice and specifically targeting the (aberrant) host-response is imperative<sup>4,7,8</sup>.

Recently in literature, tools and algorithms devised to perform simulation on biological networks have been described<sup>13,14</sup>. Here we aim to utilize our systems biology tool, the *PHENotype SIMulator* (PHENSIM), to leverage the power of pathway analysis by simulating tissue-specific infection of host cells of SARS-CoV-2 and subsequently perform *in silico* drug selection for potential repurposing. PHENSIM is a web-based user-friendly computational tool that allows phenotype prediction on selected cells, cell-lines and tissues, using a probabilistic algorithm<sup>15</sup> via “message passing”<sup>16</sup> across a network of meta-pathways<sup>17</sup>. These meta-pathways are obtained by joining all validated biological pathways, enriched with gene

regulatory elements<sup>18</sup>. The algorithm thus computes, under-specified biological contexts, by iteratively propagating the effects and alterations of one or more biomolecules (differentially expressed genes (DEGs), proteins, microRNAs, or metabolites), thus making use of published virus-human interaction data<sup>7</sup>. Here we compare our results with available data from recently published *in vitro* studies based on transcriptomics and proteomics in different model systems<sup>5,10</sup>. Relevant and significantly affected pathways are further detailed on a protein interaction level. Finally, we show the potential of the PHENSIM in selecting promising hypothesis-driven COVID-19 drug candidates, which has applicability to other diseases and broader aspects of clinical practice, thereby outlining the potential power of PHENSIM in drug repurposing in COVID-19 and beyond.

## RESULTS

### ***PHENSIM model: from in vitro to in silico***

Innovative approaches to rapidly elucidate a pathogens' mechanism of action have proven crucial for containing the global burden of communicable diseases. The PHENSIM approach, described here, is based on the definition of a newly introduced protocol for *in silico* simulation of novel emerging pathogens, such as SARS-CoV-2, and it aims at elucidating distinct host-responses and molecular mechanisms triggered by that particular pathogen, all while defining possible candidate drugs for indication repositioning.

For our strategy to be viable, even when only limited direct knowledge is available on the host-pathogen interaction, we need direct infection (*in vitro*) data that can be exploited to predict such interactions. To acquire this knowledge, we therefore employ transcriptomic and proteomic experiments of *in vitro* infected vs. normal, pathogen-free cell lines. When available, we leverage Differentially Expressed Genes (DEGs) as a means to simulate the direct and indirect effect of the virus on a host without a priori knowledge regarding the mechanism of infection. Using DEGs as input for our cell *PHENotype SIMulator* PHENSIM [<sup>15,19</sup>] we define a signature of pathogen predicted effects on human pathways (*pathogen alterations profile*; here termed the “*viral signature*”; see Fig. 1). To build the viral signature, we use pathway endpoints; an endpoint is a biological element in a pathway whose alteration, based on current knowledge, affects the phenotype in a specific way <sup>20</sup>.

By leveraging PHENSIM we aimed to determine the impact of such viral infection induced alterations on an array of human cell lines *in silico*. Simulation results are used to define a “viral signature”, that can then be employed to identify candidate drugs. Once a cell-specific SARS-CoV-2 viral signature is defined, potential repositioning drugs can be identified by building a “drug signature” database queried by means of a similarity measure using pathway endpoints (Fig. 1). Given a candidate drug identified through a database (i.e. Drugbank or

Pubchem) and literature (Pubmed) search, we define all known targets and alterations (up/down-regulations caused by the drug). Alterations are then provided as input to PHENSIM, together with the corresponding cell-specific *viral signature*. Next, distinct endpoint pathways<sup>20</sup> are identified and resulting *drug signatures* relating to a specific candidate can subsequently be compared with acquired *viral signatures* to evaluate the inhibitory potential of that candidate drug. Both viral and drug signatures are collected in a database, where a similarity search is performed using a Pearson correlation  $\rho(x,y)$  since the propagation algorithm is linear in time complexity<sup>21</sup>; see methods section *equation (1)*. All drugs whose correlation with the virus is negative (green) are considered possible repositioning candidates, since they predict inhibition of the viral signature, whereas a positive correlation (red) suggests exacerbation of the viral signature when introducing the candidate drug.

### ***Validation of PHENSIM transcriptomic strategy in SARS-CoV-2-infected host cells***

To validate our PHENSIM model on a transcriptomic level in the context of SARS-CoV2, we sought to replicate the *in vitro* experiments using publicly available data presented by Blanco-Melo *et al*<sup>10</sup>. The in-depth transcriptomic analysis of SARS-CoV-2 elicited host-response by Blanco-Melo *et al.* recently revealed an inappropriate inflammatory response driven by reduced innate antiviral defenses, with low or delayed type I and type III interferon (IFN) and exaggerated inflammatory cytokine response, with elevated chemokines and IL-6<sup>10</sup>.

As SARS-CoV-2 largely affects the lungs and respiratory tract, and because of its apparent affinity for lung tissue, the authors make use of several respiratory epithelial cell lines to assess the transcriptomic host-response. Here we use PHENSIM to reproduce transcriptomic effects *in silico*, as described *in vitro* for the following cell lines, namely undifferentiated normal human bronchial epithelial (NHBE) cells, cultured human airway epithelial cells (Calu-3) cells and A549 lung alveolar cells. The comparison of these results is depicted in Fig. 2. A549 cells



are described to be relatively non-permissive to SARS-CoV-2 replication in comparison to Calu-3 cells, which is attributed to low expression of the viral entry receptor angiotensin-converting enzyme (ACE)2<sup>10,22</sup>. Thus, A549 cells were transduced with human ACE2 (A549-ACE2), which enabled apparent SARS-CoV-2 replication at low-MOI (multiplicity of infection of 0.2). Furthermore, to induce significant IFN-I and -III expression, a high MOI of approximately 2-5 was necessary.

Here we leveraged the data published by *Blanco-Melo et al.* to run our PHENSIM simulation pipeline. In Fig. 2A we show representative genes, namely anti-viral, IFN stimulated genes (ISGs) and inflammatory cytokines and chemokines, considered important for the course of SARS-CoV-2 infection. The heatmap shows perturbed expression, either up- or down-regulated, based on results obtained by *in vitro* (left column for each depicted cell-line) experiments for the different cells assessed in comparison to *in silico* PHENSIM predictions (right column; Fig. 2A). An unbiased approach of this predictive comparison is shown in Fig. 2B, displaying the top 10 up- and downregulated DEGs based on *in vitro* SARS-CoV-2 infection, as assessed in the different cells at low and high MOI (0.2 and 2) and with ACE2 addition in A549 lung alveolar cells. For each of the top *in vitro* acquired DEGs (left; checkered boxes), the PHENSIM predicted result is shown side-by-side (right). At first glance, PHENSIM reaches high predictive accuracy for Calu-3 human airway epithelial cells and A549-ACE2 and high MOI of 2, at least for the top DEGs (Fig. 2B). To quantify the overall predictive accuracy of PHENSIM, genome-wide transcriptomic data was assessed for all scenarios as described in Fig. 2. Overall accuracy of *in vitro* predicted transcriptomic results are shown in Table 1, ranging from 51.66--83.74% for A549-ACE2 MOI 0.2 - to NHBE cells. Sensitivity of perturbation prediction for nodes accurately predicted as perturbed, ranged from 95.83--100.00% sensitivity with 97.67--99.86% specificity for this in-depth SARS-CoV-2 transcriptomic analysis. Furthermore, the positive

predictive value (PPV) and False negative rate (FNR) are shown for each tested scenario (see Table 1).

In order to further verify PHENSIM's robustness in whole genome pathway analysis, we next explored PHENSIM's ability to predict significantly affected signaling pathways in SARS-CoV-2 infection. In Fig. 2C we highlight PHENSIM's predicted perturbation of a select set of affected pathways during infection, as recently identified to be of importance by Catanzaro *et al.* 2020 and Draghici *et al.* 2020 (also see Supplementary Fig. S1 and S2), such as IL-17, JAK-STAT and TNF signaling pathways, Toll-like Receptor (TLR), NOD-like receptor and RIG-I-like receptor signaling pathways as well as complement and coagulation cascades.

For further verification of our PHENSIM pathway analysis prediction *in silico*, we compared our results with those obtained using our previously described MITHrIL (Mirna enriched paTHway Impact anaLysis) tool<sup>20</sup> to analyze the Blanco-Melo *et al.* acquired *in vitro* data (Fig. 2D-E). Given DEGs, MITHrIL first computes a perturbation for each gene in the meta-pathway (as described in Methods section). The perturbation can be considered as the predicted state that the node will have given the input DEGs. Next, we sum the perturbation of all nodes for each pathway to acquire the "accumulated perturbation," or the Accumulator. The accumulator is equivalent to a pathway expression and is a sum of all perturbations computed for that particular pathway. MITHrIL pathway analysis for A549-ACE2 at low viral load (MOI 0.2) revealed Chemokine, JAK-STAT, PI3K-Akt signaling and cytokine-cytokine interaction as a few of the top upregulated pathways, according to impact (circle size), significance (color-gradient for adjusted p-value) and accumulated perturbation computed for that particular pathway (accumulator).

For A549-ACE2 at high viral load (MOI 2.0; Fig. 2E), next to similar pathways at low viral MOI, Toll-like receptor (TLR) and NOD-like receptor signaling were among the top pathways observed, corresponding to the observation that high viral MOI was needed to induce significant

type I IFN signaling <sup>10</sup>. Interestingly, both at low and high MOI various metabolic pathways were significantly affected with a negative accumulator. Overall, the MITHrIL analysis results show the most affected pathways to be similar to the PHENSIM *in silico* predicted results.

### ***Modeling proteomics in SARS-CoV-2-infected host cells leveraging PHENSIM***

Using combinatorial profiling of proteomics and translomics to study host-infection on a cellular and molecular level give opportunity to study relevant viral pathogenicity in the search of potential drug targets <sup>5</sup>. As SARS-CoV-2 has been detected in stool and can replicate in gastrointestinal cells <sup>23,24</sup>, Bojkova *et al.* use the human colon epithelial carcinoma cell line Caco-2 to study SARS-CoV-2 infection <sup>5</sup>. With their novel method, multiplexed enhanced protein dynamics (mePROD) proteomics, they determined SARS-CoV-2-specific translome and proteome changes at high temporal resolution <sup>25</sup>, and were able to quantify translational changes occurring during SARS-CoV-2 infection *in vitro* over the course of 24 hours at multiple timepoints (at 2, 4, 10 and 24h) <sup>5</sup>.

### ***PHENSIM proteomic validation***

To validate PHENSIM on a proteomic level, we used our *in silico* approach to replicate the *in vitro* SARS-CoV-2 infection of human Caco-2 cells <sup>5</sup>. As viral genome copy number in cell culture supernatant and all viral protein levels assessed reached peak levels at 24h post infection, and the proteome underwent most extensive modulation <sup>5</sup>, we focused on this particular time-point for more in-depth comparison of protein expression and functional pathway analysis (Fig. 3). The PHENSIM simulation results obtained by leveraging the proteomic data 24hrs post SARS-CoV-2 infection are shown in Fig. 3. We provide an unbiased assessment by comparing the PHENSIM obtained Average Node perturbation *in silico*, to the 30 most perturbed proteins according to Bojkova *et al.* In order to compare *in vitro* to *in silico* protein

expression levels a representative selection of relevant proteins involved in infection is depicted in the heatmaps in Fig. 3B and C. In Fig. 3B, proteomic perturbation of the top differentially expressed proteins (DEPs; n=30) as predicted by PHENSIM (right, solid) is compared side-by-side to perturbation results from Bojkova *et al.* (left, checkered). Next, in Fig. 3C the top DEPs described by Bojkova *et al.* (right) is compared to PHENSIM predicted perturbation. Based on this selection of proteins we can denote a relatively high prediction rate for PHENSIM, although not all proteins are predicted to full accuracy. When quantifying the predictive power of PHENSIM on this protein-wide analysis, PHENSIM simulated results showed a predictive accuracy of 97.9% to the described *in vitro* proteomic data at 24hrs, where significant perturbation prediction was at 97.87% sensitivity and 97.96% specificity for this particular dataset (see Table 2).

#### *PHENSIM proteomics from in vitro to in silico*

Next, to compare the Reactome-based *in vitro* functional pathway analysis to our PHENSIM *in silico* approach, a representative selection of significantly affected pathways – correctly predicted by PHENSIM – is depicted in Fig. 3D. Pathways were selected according to the cellular mechanisms highlighted by Bojkova *et al.* <sup>5</sup>. The centered heatmap shows an increasing activity score (top to bottom) as predicted by PHENSIM for each pathway. An in-depth analysis of proteomic pathway at 24hrs revealed distinct upregulation of various pathways involving cellular metabolism such as fatty acid degradation, glycolysis and glyconeogenesis, carbon metabolism, inflammatory and immune signaling pathways and also cellular senescence signaling pathways (Fig. 3D). A selection of KEGG pathways similar to the Reactome protein interaction pathways are further highlighted in more detail, depicting the main protein-protein interactions for that particular pathway, with upregulated proteins depicted in red and

downregulated proteins in blue. As authors used Reactome as their pathway knowledge-base, here we show only KEGG pathways matching their Reactome counterpart.

### *PHENSIM predicts a metabolic signature in SARS-CoV-2 infection in silico*

As a metabolic signature was identified by PHENSIM's proteomic in silico simulation of SARS-CoV-2 infection (Fig. 3D), we next assessed the degree of intersection between the perturbed genes of these metabolic pathways in order to reject the hypothesis that a common set of altered proteins is driving the significant perturbation of these closely related metabolic pathways. All metabolic pathways considered essential for SARS-CoV-2 infection according to the acquired Bojkova *et al.* data (Fig. 3) were included in the analysis (FDR-adjusted p-value < 0.05) and a PHENSIM activity score was determined (see Supplementary Table S2). The affected general metabolic pathways showed very low degree of shared sub-pathway overlap. The Venn diagrams in Supplementary Fig. S4 show all possible intersections for the following top metabolic pathways: (i) Fatty acid degradation, Amino sugar and nucleotide sugar metabolism, Glycolysis /Gluconeogenesis, Citrate cycle (TCA cycle), and Purine metabolism; (ii) Glycolysis /Gluconeogenesis, Citrate cycle (TCA cycle), Purine metabolism, Carbon metabolism, and Pyrimidine metabolism.

### ***PHENSIM Drug repurposing strategy for COVID-19***

The next step in our PHENSIM approach is the employment of our drug strategy in order to test candidate drugs for potential COVID-19 repurposing. This approach takes advantage of existing knowledge on drug-related pharmacology and toxicology for rapid therapeutic selection<sup>8</sup>. As schematically described in Fig. 1, once a cell-specific viral signature is defined, it can be exploited to search for possible repositioning candidates by leveraging our select drug signature

database. We used a Pearson correlation  $p(x,y)$  to compare the viral and drug signatures, which gives rise to a correlation score specific to that candidate drug, computed for SARS-CoV-2 infection in a particular setting. Here we set out to test a selection of hypothesis- and data-driven candidate drugs as shown in Fig. 4. One such drug which regrettably failed to live up to its anticipated potential to effectively treat COVID-19 is the antimalarial drug hydroxychloroquine (HCQ), currently approved for rheumatologic implications, although associated with cardiac toxicity<sup>26-28</sup>.

Although the efficacy of corticosteroids in viral acute respiratory distress syndrome (ARDS) remains controversial, recent evidence on drugs such as Dexamethasone and Methylprednisolone are showing promise in COVID-19<sup>29,30</sup>. Furthermore, the potential beneficial effects of blocking the mTOR pathway with use of mTOR-inhibitors such as Metformin, Everolimus or Rapamycin (the later not evaluated here) in COVID-19 patients have been hypothesized, however its effects on gene expression and distinct signaling pathways remain to be satisfyingly established. In light of targeting cell immunometabolism, 2-Deoxy-Glucose (2DG) was recently proposed as possible therapeutic in COVID-19<sup>5</sup>. Lastly, therapeutic targeting of excessive host inflammation by inhibiting Bruton tyrosine kinase (BTK) – for example the BTK-inhibitor Acalabrutinib – in severe COVID-19 was recently described<sup>31</sup>.

We evaluated a select set of candidate drugs for potential repurposing in SARS-CoV-2 infection as shown in Fig. 4 (and Supplementary Fig. S5). In Fig. 4 we show the results for A549-ACE2 cells at both a low (0.2, Fig. 4A) and high MOI (2, Fig. 4B). The other cell type scenarios are shown in Supplementary Fig. S5. The drug candidates having a positive effect on ameliorating SARS-CoV-2 infection *in silico* have a negative correlation score (green) between viral and drug signature, whereas candidate drugs worsening the disease phenotype have a positive correlation (red). Indeed, for both low and high viral load (MOI), Methylprednisolone, Metformin, Dexamethasone and Acalabrutinib positively correlated with the viral signature

(green) which points to an effective therapeutic to target SARS-CoV-2 infection in A549 cells in the presence of ACE2, however, the order of the candidate drugs differed somewhat between the two.

Using CoVariation analysis, we next looked at individual pathway contribution for each of the repositioning candidates evaluated here. The acquired Pearson correlation when comparing viral and drug-based signatures was dissected into components to show individual pathway contribution (see Fig. 4C-D and Supplementary Fig. S6). The overall effect of a candidate drug can be seen as the sum of the individually affected pathways, where anti-correlation is depicted in purple and positive correlation in orange. Here we use Methylprednisolone as an example for A549 cells expressing ACE2 receptor at low (0.2, Fig. 4C) and high MOI (2.0, Fig. 4D). Only significantly affected pathways are shown here. Pathway accumulation plots for the top 4 candidate drugs are shown in Supplementary Fig. S6 for A549-ACE2 at low and high MOI to illustrate the variation and effectiveness of the tested drug candidates and show how the drugs were ordered based on best candidate (top; most pathways are anti-correlated shown in purple), to least likely candidate of interest (bottom; mostly positively correlated pathways in orange). Some top anti-correlated pathways for Methylprednisolone, highly contributing to the final result of this drug candidate based on our PHENSIM analysis include the JAK-STAT pathway, the Toll-like receptor pathway, MAPK and PI3K-AKT signaling pathways. Next to similar pathways of importance affected for A549-ACE2 at both low and high viral MOI such as JAK-STAT, Toll-like receptor (TLR), NOD-like receptor, RIG-I-like receptor and MAP-kinase (MAPK) signaling, Focal adhesion and Neurotrophin signaling pathway were among the top pathways observed at high viral load (MOI 2.0; Fig. 4D).

*PHENSIM Methylprednisolone treatment of SARS-CoV-2 infected host cells in silico*

As a next step in our drug repurposing efforts, we simulate the simultaneous host-cell infection of SARS-CoV-2 and *in silico* treatment with Methylprednisolone (MP), hereby combining the drug action and pathogen infection on a host-cell, in order to further assess MP as top candidate. We simulate SARS-CoV-2 viral infection and simultaneous MP treatment *in silico*, in order to more closely resemble the *in vivo* situation (Fig. 5). The heatmap in Fig. 5 depicts the results of transcriptomic pathways analysis of host-cell SARS-CoV-2 infection, based on Blanco melo *et al. in vitro* (Fig. 5; left column A) and PHENSIM simulation *in silico* (Fig. 5.; middle column B), compared to MP treatment of *in silico* SARS-CoV-2 infected host cells (Fig. 5.; right column C). Here we visualize the pathways identified in Fig. 2C, and show the effects of MP treatment on these top affected pathways during SARS-CoV-2 infection in particular host-cells. All identified upregulated pathways during infection, where significantly inhibited by MP treatment, showing it's known anti-inflammatory and immunosuppressive effects.



## Discussion

The current pandemic has accelerated the need for efficient and effective identification of potential drug candidates for COVID-19 pathogenesis. Knowledge on host-immune response to SARS-CoV-2 infection, however, remains limited with very few drugs approved to date. Various viable strategies and tools are rapidly arising to address this, where repurposing of existing drugs can offer a feasible mechanism of deployment (4). Here we introduce one such strategic approach, the *PHENotype SIMulator*, which allows the modeling of SARS-CoV-2 infection *in silico* and implementation to select promising candidates for further *in vitro* and *in vivo* analysis. We show that PHENSIM can effectively be used to 1) determine the viral effects on cellular host-immune response and cellular pathways and 2) evaluate a myriad of therapeutic strategies.

As previously described, PHENSIM uses a probabilistic randomized algorithm to compute the effect of a particular biological scenario on gene regulation, protein expression, miRNA and metabolite involvement, with use of KEGG meta-pathway analysis<sup>15</sup>. Here, we simulate SARS-CoV-2 viral infection based on publicly available data to acquire a specific cellular SARS-CoV-2 signature. To verify a transcriptomic-based PHENSIM strategy, we compared the *in silico* simulation results to publicly available transcriptomic data of SARS-CoV2 infected cell lines<sup>10</sup>. We find that PHENSIM performs at a high overall accuracy with high PPV, sensitivity and specificity for all airway and lung-related cell lines evaluated. Key SARS-CoV2 infection-related signaling pathways could be discerned as such, comprising the viral signature. PHENSIM predictive performance was further validated using our previously described MITHRIL transcriptomic pathway analysis<sup>20</sup>, showing similar results. Interestingly, key signaling pathways proposed to be crucial in SARS-CoV2 infection<sup>4</sup>, were shown to be significantly perturbed in all cell lines studied *in silico* using PHENSIM, simulation offering promising potential molecular drug targets.

The PHENSIM strategy was also suitable for a proteomics/translatome-data based approach. PHENSIM simulation was compared to published SARS-CoV2 infection specific proteomic effects in host cell lines <sup>5</sup>. Comparing simulation results to proteomic data 24 hours post infection showed, for the proteins available in KEGG, high accuracy with a PPV, sensitivity and specificity well above 97%. Inhibition of several of these protein associated pathways pathways was shown to prevent viral replication in human cells <sup>5</sup>.

We next used the transcriptomic-based PHENSIM approach to compare the viral signatures, computed with respect to model cell lines, to *in-silico*-derived drug signatures for a selection of drugs analyzed. Our overall correlation results show several potential drug repurposing candidates negatively correlating with SARS-CoV-2, varying from corticosteroids such as MP (already approved for treatment of COVID-19 patients) to biologics such as BTK-inhibitors that are currently being studied in clinical trials <sup>31</sup> to metformin <sup>32</sup>. Individual signaling pathway contribution to the observed correlation score could be further delineated for each individual drug, providing specific targets for in depth analysis and potential for pathway-specific therapeutic targeting. As expected, the individual pathways most targeted by the *in silico* drug interventions (Fig. 4C&D) were similar to pathways found most perturbed by PHENSIM during host transcriptomic response to SARS-CoV-2 viral infection (Fig. 2C&D), emphasizing their potential therapeutic effects. HCQ, although hypothesized to be a good potential candidate to treat COVID-19, has not proven effective *in vivo*<sup>33</sup>. The exact reason why HCQ has failed in COVID-19 remains to be fully understood. Interestingly, COVID-19 is associated with a variety of hematologic complications <sup>34</sup>, and increased HCQ use during the COVID-19 pandemic has induced the emergence of methemoglobinemia, including tissue hypoxia and reduced oxygenation <sup>35,36</sup> amongst others. Evidently, evaluating the risk-benefit ratios – drug safety and efficacy – is crucial when selecting drugs to be repurposed for COVID-19 <sup>8</sup>, which particularly holds true for HCQ <sup>27,28,37</sup>.

In Fig. 4, we depict our drug repurposing PHENSIM approach that functions as a screening tool for initial drug candidate screening, based on the anti-correlation of the viral and drug signatures, and gives a particular score for each candidate; a negative correlation constitutes higher potential for that particular drug. This broader correlation approach described in Fig. 4 and Supplementary Fig. S5 and S6 can be used to screen large sets of candidate drugs. Next, as depicted in Fig. 5, a more dynamic and extensive analysis can be simulated, in order to simulate the interaction between the SARS-CoV-2 host-cell infection (column B) and subsequent *in silico* treatment with a candidate drug, here Methylprednisolone (column C). Although Methylprednisolone is a known broad-spectrum corticosteroid, with clear anti-inflammatory and immunosuppressive effects (as shown in Fig. 5), complete inhibition of these crucial immune signaling pathways might not be beneficial to COVID-19 patients at every stage of disease; which has been described in clinical practice. Other, more targeted drug candidates might be more beneficial to the overall functioning of the patient's immune system during the fight and recovery from COVID-19. Indeed, our detailed approach can be implemented for all other top candidates, for further in-depth evaluation of their potential. However, we should bear in mind that the simulation is simultaneous (both virus and drug) and not completely reflective of a sequential treatment of a drug during infection. We are currently leveraging our simultaneous approach to evaluate the use of Metformin in COVID-19 in more detail.

Drug repurposing towards COVID-19 is challenging, but also poses many new opportunities. Several innovative approaches have been used varying from structure assisted computer designed mini inhibitors of receptor binding domain (RBD) binding<sup>38</sup>, inhibitors of viral key enzymes like Mpro<sup>6,39,40</sup>, machine learning models predicting compound protein inhibiting activity<sup>41</sup> to infected cell-based assays drug screening<sup>42,43</sup>. Using computational tools, such as PHENSIM, allows for safe exploration of potential candidate drugs and uses previously acquired knowledge from biomedical databases to narrow the scope of possible viable biomarkers and

druggable targets. One of the clear advantages of PHENSIM is a more effective selection of hypothesis driven drugs, before initiating extensive, time-consuming and costly *in vitro* experiments that should eventually provide the basis for clinical studies. PHENSIM requires on average (depending on data availability) about 3 hours of simulation time. Another interesting possibility enabled by our approach is the potential capability to not only simulate the effect of a single drug, but also drug combinations. This expansion of PHENSIM is currently being developed (see Methods section; *data availability*). By making use of not just viral targets but also host proteins and structured pathways in the computation of the PHENSIM viral signature, we broaden the scope of potential drug targets with the added advantage that these are less prone to resistance development <sup>44</sup>. Here we simulated a select set of candidate drugs for repurposing in COVID-19, however, there are many candidates with high potential that can be added to this list, and further evaluated by our PHENSIM system *in silico* in the near future. We can also learn from and identify additional candidates based on the results obtained in this study. Based on the SARS-CoV-2 viral signature acquired by PHENSIM and recent data on IFN-involvement in COVID-19 [reviewed in <sup>45</sup>], targeting the JAK/STAT pathway using Baricitinib – approved for moderate to severe arthritis <sup>46</sup> – was recently shown to reduce time-to-recovery for hospitalized COVID-19 patients in combination with Remdesivir <sup>47</sup>, however, caution is warranted <sup>48</sup>.

One of the advantages of the PHENSIM algorithm is the capacity to add on information of particular genes of interest (as specific knowledge becomes available) to the original simulator, if absent in KEGG. In the case of SARS-CoV2 infection, the absence of some important genes involved in SARS-CoV-2 infection of host cells in the KEGG database, was considered a limitation. One gene in particular is Basigin (BSG), also known as CD147. The exact role of CD147 in SARS-CoV-2 infection remains elusive to date, however, its role in spike-protein (SP) binding and viral infiltration of host cells has been described <sup>49</sup>. In our current

approach, we model SARS-CoV-2 infection *in silico* with the addition of CD147 into KEGG, in order to investigate the role of this extracellular matrix metalloproteinase inducer (EMMPRIN) in COVID-19 (see Supplemental Table S1).

As discussed, the *in silico* model presented here provides an interesting framework that could be further developed and expanded, achieving a more complete cell signature with input of available data on processes such as cell-cell communication through ligand-receptor complexes<sup>50</sup> or viral immune evasion e.g.<sup>51</sup>. As most *in vitro* studies are performed on cell lines, tissue tropism characteristics of viral infection seem key to better understanding viral activity<sup>52</sup>. The same model could be adapted to study specific cells involved in viral infection like tissue-specific epithelial cells and immune cells (e.g. T cells and NK cells)<sup>53,54</sup>. Moreover, many interesting avenues can potentially be explored using PHENSIM, such as modeling immune-related effects of this pathogen and others, in distinct tissue-specific non-immune epithelial cells, stem cells, and beyond<sup>11,53-55</sup>. The system can be further adapted to include new data gathered on the viral translational landscape related to newly discovered open reading frames (ORFs) and potential novel polypeptides/proteins and infectivity potentiating cell surface structures like neuropilin<sup>56,57</sup>. Interestingly, integration of all the aforementioned schemes could potentially yield novel and effective drug targets<sup>58</sup>.

Here we show distinct candidate drugs having a variable effect depending on the multiplicity of infection (MOI) of virus infection in A459-ACE2 expressing cells in low (MOI 0.2; Fig. 4A) and high MOI (2.0; Fig. 4B). As shown, the sequence of top candidate drugs for repurposing is slightly different depending on the cell-type, viral load (MOI), and expression of the viral entry receptor ACE2 (see Fig. 4 and Supplementary Fig. S2). This speaks to the variability of how the virus might affect specific cell types and tissues, even within the same organ system such as the bronchial (NHBE), airway epithelial (Calu-3) and lung alveolar (A549) cells e.g., pointing to the difficulty in specifically targeting this viral infection therapeutically. It will

also depend on the stage of infection and disease state, as to which course of treatment or combination of treatments will be optimal.

As demonstrated by our results, we believe that the PHENSIM system provides a multitude of powerful systems biology functions and implements them easily and efficiently. PHENSIM is a simulation algorithm which follows the biological processes modeled by pathways. Therefore, PHENSIM is able to make a prediction of such processes and not only of the final effect, going beyond methods based on pathway enrichment. Furthermore, since pharmacological treatments may depend on the state of biological processes, PHENSIM may be of more appropriate use in this context. Comparison with other simulation algorithms such as BIONSI<sup>13,14</sup> has shown excellent performance by PHENSIM<sup>15</sup>. PHENSIM creates and builds on interpretable and intervenable mechanistic bio-chemical models, rather than combinatorial and statistical “black-box” models for joint stationary distribution of biological data, as in, say protein-protein interaction (PPI) networks, Graphical or Deep-net models.

PHENSIM gives rise to feasible validation and comparison of *in vitro* and *in vivo* experimental data<sup>4,5</sup>, gives insight into drug efficacy<sup>5,30,44</sup>, tracks specific host signal transduction pathways<sup>4</sup>, *in silico* testing of single drugs and drug combinations and further delineation of future targets (e.g. CD147) and identification of specific pathways of action of both pathogen and therapeutic compound in healthy and infected systems. For cost efficiency, validated predictive methods and assays for early elimination of potential drug candidates are of great value<sup>59</sup>. The overall efficiency (time, costs, safety) prompts to suggest implementing PHENSIM not only in viral acute pandemic settings<sup>60</sup>, but in additional curative and non-curative diseases, especially complex chronic disorders, where clinical trials are time-consuming or impossible to reduce to practice. Optimally leveraging the power of pathway analysis by simulating host cell and tissue-specific infection and performing *in silico* drug selection, has a tremendous potential beyond COVID-19, with applicability to high global burden

communicable diseases, translatable to pathogens of viral, bacterial and fungal origin, and potentially chronic disease such as inflamm-aging and diabetes. In conclusion, our PHENSIM approach will enable more rapidly initiated clinical trials and accelerated regulatory review of already pre-selected drugs with a high repurposable potential.

## METHODS

### *PHENotype SIMulator (PHENSIM) approach*

PHENSIM is a systems biology approach, simulating the effect of the alteration of one or more biomolecules (genes, proteins, microRNAs, or metabolites) in a specific cellular context using KEGG (Kyoto Encyclopedia of Genes and Genomes) meta-pathways<sup>15,20</sup>. The meta-pathway concept, introduced by us previously<sup>61</sup>, has been devised to account for pathway cross-talk in analysis. Essentially, all KEGG pathways are merged in a single graph through common nodes, where the meta-pathway is a graph in which the nodes represent molecular entities (genes, metabolites) and the edges are the known biological interactions present in the KEGG database. The meta-pathway is further completed by adding validated miRNA-targets downloaded from miRTarBase (release 8.0), miRecords and TF-miRNA-interactions obtained from TransmiR (release 2.0)<sup>62-64</sup>.

To start the simulation, PHENSIM requires a set of biomolecules as input, their direction of deregulation (activation/up-regulation or inhibition/down-regulation), and a set of inactive biomolecules in the cellular context (cell lines, tissue, e.g). The algorithm uses these details to compute synthetic Log Fold Changes (LogFC). Synthetic LogFCs are computed by sampling the normal distribution fitted to the actual LogFCs of a particular gene, as described previously<sup>15</sup>. Such values are then propagated within the meta-pathway, using the MITHrIL (Mirna enriched paTHway Impact anaLysis) pathway perturbation analysis<sup>20</sup>. MITHrIL determines how local change can affect the cellular environment by computing a “perturbation”. For each gene in the meta-pathway, the perturbation reflects its expected change of expression/activity (negative/positive for down-/up-regulation, respectively). Finally, these results are collected and synthesized using two values: the “Average Perturbation” and the “Activity Score” (AS). Given a node, the average perturbation is the mean for its perturbation values computed at each simulation step. It reproduces the expected change of expression for the entire process. The



function of AS is twofold: 1) the sign gives the type of predicted effect (activation(+); inhibition(-)), 2) the value is the log-likelihood that this effect will occur. Together with AS, PHENSIM also calculates a p-value which determines how biologically relevant the predicted alteration is for the phenomena being simulated.

All p-values computed by PHENSIM are corrected for multiple hypotheses using the q-value algorithm<sup>65</sup>. To determine this probability, PHENSIM randomly selects genes in the meta-pathway and runs the simulation on this random set. By repeating this procedure (n=1000 for our simulations), it is possible to empirically estimate the probability that a node has a higher activity score than the observed one. For this reason, we can employ such a value to determine which alterations are most specific for a particular infection, gaining novel hypotheses on the molecular action of the pathogen.

#### *PHENSIM pathogen alterations profile*

Our approach defines a protocol for the *in silico* simulation of emerging pathogen infection, aimed at defining candidate drugs for repositioning. First, we find a representation of the pathogen in the KEGG meta-pathway, which allows us to perform simulations. For a novel pathogen, such as SARS-CoV-2, interactions with the host genes might be unknown. Therefore, we can approximate this by employing expression data of pre-/post-infection samples. The rationale is that differentially expressed genes (DEGs) represent the downstream effects of the viral infection on the host; *i)* we compute DEGs between pre- and post-infection samples, *ii)* we extend the meta-pathway by adding a new node representing the virus, *iii)* the viral node is connected to each DEG with an activating (/inhibiting) edge if its LogFC computed between post- and pre-infection is positive (/negative), *iv)* we run a simulation by giving the upregulation of the viral node as input.

To build the *pathogen signature*, we use pathway endpoints; *An endpoint is a biological element in a pathway whose alteration, based on current knowledge, affects the phenotype in a specific way* <sup>20</sup>. Given the output of this simulation, we collect all Endpoint Activity Scores in a single signature, the 'pathogen alterations profile'. This profile can be exploited to search for possible repositioning candidates, by building a *drug signature* database queried by means of a similarity measure (Fig. 1). When building the simulation profile, we do not use any p-value. Indeed, we need to consider not only the alterations, which are the most specific for a particular infection, but also the alterations caused by any cellular response to the infection. Since the p-value represents the biological relevance for the phenomena that is being simulated, we can ignore its value to build the signature.

#### *PHENSIM drug signature database*

Given a particular drug identified through databases (i.e. Drugbank or Pubchem) and literature (Pubmed) searches, we define all known targets and their alterations (up/down-regulations caused by the drug), and these alterations are provided as input to PHENSIM together with the same cellular context specified for the viral simulation. The results are used to define a drug signature using pathway endpoints as described above, which are collected in a database used for repositioning.

#### *PHENSIM drug repurposing approach*

Our drug repurposing methodology is based on a similarity search performed on the drug signature database. Given a pathogen profile computed with PHENSIM, we use a correlation function to scan through each record in a drug profile database. This procedure yields a ranking on each drug in the range [-1;1], where negative values indicate that the virus alteration profile

is opposite to the drug and positive values indicate the reverse; drugs with a negative correlation are considered possible candidates for repositioning. In our experiments, we employ a Pearson correlation function to run the similarity search. Since PHENSIM is based on MITHrIL pathway perturbation analysis, which computes results in a log-linear space<sup>20</sup>, we can assume a Pearson correlation is sufficient to determine similarity between the viral and drug signature. A key characteristic of this approach is the capability to simulate both single and drug combinations. Furthermore, PHENSIM also provides a framework for extending pathways by adding new nodes and edges coming from results in the literature as well as other reputable sources.

Finally, to assess whether each drug candidate targets relevant infection processes, we decomposed the Pearson correlation in terms of KEGG pathways and reviewed the results. More in detail, let  $D$  and  $V$  be drug and pathogen alteration profiles, respectively. That is,  $D[e]$  is the activity score of endpoint “e” computed by PHENSIM for a drug simulation, and  $V[e]$  is the activity score for the same endpoint in the pathogen simulation. Pearson correlation  $\rho(D, V)$  can be written as *equation (1)*:

$$\rho(D, V) = \frac{\sum_e (D[e] - \bar{D})(V[e] - \bar{V})}{\sigma(D) \cdot \sigma(V)},$$

Where  $\bar{D}$  and  $\bar{V}$  are the means of  $D$  and  $V$ , respectively, and  $\sigma(\cdot)$  is the standard deviation.

Therefore, given a pathway  $P$ , we can sum the Pearson correlation components belonging to its endpoints to estimate how much it contributes to the final correlation value. More in detail, the partial correlation  $\hat{\rho}(D, V, P)$  can be computed as *equation (2)*:

$$\hat{\rho}(D, V, P) = \frac{\sum_{e \in P} (D[e] - \bar{D})(V[e] - \bar{V})}{\sigma(D) \cdot \sigma(V)}.$$

A significant feature of this partial correlation approach is that we obtain the total correlation by summing up all values for each pathway  $P$ . Therefore, we can determine which biological processes are impacted by the drug administration.

#### *PHENSIM combined drug/pathogen simulation*

To further evaluate whether the results of the correlation could be confirmed by PHENSIM, we devised a strategy to simultaneously simulate drug action and pathogen infection on a host cell line. First, we collected DEGs between pre- and post-infection samples as described in the previous section. Then, given a drug, we gather its known targets and their alterations (up/down-regulations caused by the drug) through databases (i.e. Drugbank or Pubchem) and literature (Pubmed) searches. Therefore, we extend the meta-pathway by adding two nodes, representing the virus and the drug, respectively. The virus node is connected to each DEG with an edge as described in the “*PHENSIM pathogen alterations profile*” section. Then, an activating (inhibiting) edge is added between the drug node and a target, for any up-regulated (/down-regulated) target. Finally, we can run a simulation by giving as input the simultaneous upregulation of both virus and drug nodes (results depicted in Fig. 5).

#### ***PHENSIM method validation***

To determine the efficacy of our model we used several datasets obtained in the context of SARS-CoV-2 infection. For each dataset, we computed the genome-wide Log Fold Changes (FC). As PHENSIM does not require any quantitative information, DEGs were termed upregulated if  $\text{LogFC} > 0.6$ , and downregulated if  $\text{LogFC} < -0.6$ .

#### *PHENSIM CD147 gene and pathway extension*

Prior to running all simulations, we verified if viral entry points were present in KEGG to better represent viral activity. SARS-CoV-2 is said to invade host cells via these two receptors: angiotensin-converting enzyme 2 (ACE2) and CD147 (also known as Basigin or EMMPRIN) <sup>49</sup>. In KEGG the latter gene is missing and to extend our simulation model, a new node, representing CD147 was added and connected with its known interactions and downstream nodes according to literature (Supplementary Table S1). The incoming edges to CD147 represent the possible activators/inhibitors (upstream genes), the outgoing edges represent the actions performed by CD147 towards its downstream genes. In Supplementary Table S1, we report a list of up- and downstream genes added to our simulations and their respective references containing interaction details. CD147 is a transmembrane protein of the immunoglobulin super family, expressed in many tissues and cells, acting as the main upstream stimulator of matrix metalloproteinases (MMPs) and playing a crucial role in intercellular recognition <sup>66</sup>. Over the last decade, several groups have shown that CD147 acts as a key molecule in the pathogenesis of several human diseases including infectious diseases (HIV, HBV, HCV, KSHV) <sup>66</sup>, and it has now been posed to recognize and internalize/endocytose SARS-CoV-2 in certain cell types <sup>49</sup>.

#### *PHENSIM transcriptomic reliability assessment*

To assess the reliability of the results, we focused on two fronts: *ii*) the ability of PHENSIM to predict genes altered in the expression data, and *ii*) the ability to predict the correct direction of the alteration. In detail, we define as altered all genes having an absolute LogFC>0.6. The type of the alteration is given by the sign of the LogFC (+LogFC for upregulation, -LogFC for downregulation). Predictive power of PHENSIM was assessed by means of Positive Predictive Value (PPV), Sensitivity, and Specificity. The PPV is the proportion of true positive results with respect to all positive predictions, the sensitivity is the percentage of true positives with respect

to the entire population, and the specificity is the percentage of true negatives with respect to all negative cases.

#### *PHENSIM transcriptomic approach*

For the evaluation of our strategy, we exploited transcriptomics data published in Blanco-Melo *et al.* 2020 [GSE147507] <sup>10</sup> and proteomics data coming from Bojkova *et al.* 2020 <sup>5</sup>.

The Blanco-Melo *et al.* dataset comprises RNA-seq data of infected vs. mock-treated cell-lines from human and ferret. The data were obtained by using the Illumina NextSeq 500 platform <sup>10</sup>. In our analysis, we focused solely on human cell data. In detail, 4 cell-lines were evaluated: primary human lung epithelium (NHBE), transformed lung alveolar (A549) cells, transformed lung alveolar (A549) transduced with a vector expressing human ACE2 and transformed lung-derived Calu-3 cells. For all cell-lines, sequencing data of biological triplicates was obtained from mock treated or SARS-CoV-2 infected experiments. Furthermore, for both A549 cell lines different MOIs (multiplicity of infection) were used at low 0.2 and high MOI 2.0. Following the same procedure used by Blanco-Melo *et al.*, raw counts were normalized and analyzed for differential expression using the DESeq2 pipeline. All genes with an FDR-adjusted p-value<0.05 and absolute LogFC>0.6 were considered differentially expressed. Non-expressed genes for a specific cell-line were defined as genes that showed an average read count lower than 10.

#### *PHENSIM proteomic approach*

To determine if our methodology can also exploit proteomic data, we leveraged data from Bojkova *et al.* 2020, namely proteome measurement by LC–MS/MS of control vs. SARS-CoV-2-infected human Caco-2 cell lines <sup>5</sup>. All cell lines were analyzed in triplicates (n=3) at 2, 4, 10 and

24h. Log2-ratios between infected and normal differentially expressed proteins (DEPs) ( $p$ -value $<0.05$ ) were used as input for the simulation algorithm. Non-expressed proteins for the Caco-2 cells were taken from the Human Protein Atlas (using the query “celline\_category\_rna:CACO-2;Not detected”). Since PHENSIM uses Entrez Gene Identifiers, we mapped all proteins to their gene, yielding 5809 mapped proteins. Of these 5809 mapped proteins, we could find only 1914 in the KEGG meta-pathway. We therefore combined our PHENSIM analysis with an enrichment analysis to determine if a prediction made by our methodology was based on adequate data (Supplementary Table S2).

More in detail, for each time point and each pathway, we compared the number of altered proteins predicted by PHENSIM to the expected number of altered proteins using a hypergeometric distribution. This analysis yielded an enrichment  $p$ -value combined with PHENSIM one using the Fisher’s Method <sup>67</sup> due to their independence.  $P$ -values were corrected for multiple hypotheses using the Benjamini-Hochberg correction, and all pathways with a  $p$ -value $<0.05$  were considered significant for further analyses.

As Bojkova *et al.* reported their results using Reactome Pathway analysis <sup>5</sup>, in contrast to the KEGG Pathway analysis used by PHENSIM, we selected all pathways corresponding to the SARS-CoV-2 highlighted cellular processes (translation, splicing, carbon metabolism and nucleic acid metabolism, for instance). Comparisons were performed using the Average Pathway Perturbations as reported by PHENSIM (Fig. 3D). Finally, since we found many similar metabolic pathways significantly affected *in silico* as described *in vitro* by Bojkova *et al.*, we aimed to determine if a core set of proteins was common between pathways; results are displayed in the VENN diagrams in Supplementary Fig. S4 and Table S2.

### **Performance Evaluation: PHENSIM Genome-wide and Proteome network analysis**

We compared our results with published *in vitro* experiments from Blanco-Melo *et al.*, Bojkova *et*

*al.*, and Draghici *et al.* <sup>5,10,30</sup>. First, we compared the results from Blanco-Melo *et al.* with our *in silico* predictions for NHBE, Calu-3, A549 (MOI 0.2 and 2) and A549 transduced with ACE-2 (MOI 0.2 and 2) cells. Transcriptomics data for all cell lines were collected from *GEO dataset* GSE147507 and Log2-LogFCs were computed. Next, LogFCs were compared with our predicted Activity Scores (AS) by accounting for their direction of perturbation. We compared our predictions with the genes that *Blanco-Melo et al.* reports as important in the antiviral host-response to SARS-CoV-2. Furthermore, using an unbiased approach and to verify the accuracy of PHENSIM, we assessed the top-10 upregulated and top-10 downregulated genes for each cell-line. Finally, we assessed our viral simulation with results from Draghici *et al.* 2020 and Catanzaro *et al.* 2020 <sup>4,30</sup>.

### **Pathway Analysis**

Pathway analysis was applied to the transcriptomics data to determine which biological processes were altered by the viral infection. We used 4 pathway analyses approaches to assess the most impacted pathways: 1) MITHrIL, 2) SPIA, 3) Reactome Pathways and 4) Gene Ontology Enrichment analysis. MITHrIL pathway analysis was performed as described in Alaimo *et al.*, 2016 <sup>68</sup>. We used the LogFC of DEGs for all cell lines to perform MITHrIL perturbation analysis on the KEGG meta-pathway. Therefore, all values were aggregated on a pathway basis to compute an Accumulator and a p-value. Finally, p-values were adjusted for multiple hypotheses using the Benjamini-Hochberg FDR correction. Results were filtered by an FDR-adjusted p-value of 0.05 and ranked using the Accumulator. The top-25 significant pathways were reported in Fig. 2D&E. SPIA analysis was performed as previously described by Tarca *et al.*, 2009 <sup>69</sup>; the LogFC of DEGs and ranked pathways were calculated using FDR-adjusted p-values as computed by SPIA. Pathways with a  $p < 0.05$  were considered significant.



Finally, to further expand our understanding of the biological processes affected by the infection, we performed enrichment analysis on both Reactome Pathways, using the ReactomePA package <sup>70</sup>, and Gene Ontology, using the GOfuncR package <sup>71</sup>. All results produced by the 4 pathway methodologies were collected and considered significant with an FDR-adjusted p-value<0.05 (Supplementary Fig. S3 and Supplementary material).

### **Data availability**

All input data, raw images, and source codes for PHENSIM are available at <https://github.com/alaimos/phensim-covid19> . Website: <https://phensim.atlas.dmi.unict.it><sup>15</sup>.

### **Statistical analysis**

Statistical methods for transcriptomics and proteomics were applied as described by Blanco-Mello and Bojkova *et al.* respectively <sup>5,10</sup>. For transcriptomic data, raw counts were normalized and analyzed for differential expression using the DESeq2 pipeline as previously described <sup>72</sup>. All genes with an FDR-adjusted p-value<0.05 and absolute LogFC>0.6 were considered differentially expressed. In addition, we considered all genes showing an average read count<10 as non-expressed. All p-values computed by PHENSIM are corrected for multiple hypotheses testing, using the q-value algorithm <sup>65</sup>. For proteomic data, Normalized LC-MS/MS data were downloaded and significance was tested using unpaired two-sided Student's t-tests with equal variance assumed. All values were aggregated on a pathway basis to compute an Accumulator and a p-value, and p-values were adjusted for multiple hypotheses using the Benjamini-Hochberg FDR correction. Results were filtered by an FDR-adjusted p-value of 0.05 and ranked using the Accumulator.

## REFERENCES

- 1 Kupferschmidt, K. & Cohen, J. Will novel virus go pandemic or be contained? *Science* **367**, 610-611, doi:10.1126/science.367.6478.610 (2020).
- 2 Cucinotta, D. & Vanelli, M. WHO Declares COVID-19 a Pandemic. *Acta Biomed* **91**, 157-160, doi:10.23750/abm.v91i1.9397 (2020).
- 3 Weinreich, D. M. *et al.* REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. *N Engl J Med* **384**, 238-251, doi:10.1056/NEJMoa2035002 (2021).
- 4 Catanzaro, M. *et al.* Immune response in COVID-19: addressing a pharmacological challenge by targeting pathways triggered by SARS-CoV-2. *Signal Transduct Target Ther* **5**, 84, doi:10.1038/s41392-020-0191-1 (2020).
- 5 Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature*, doi:10.1038/s41586-020-2332-7 (2020).
- 6 Panda, P. K. *et al.* Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Sci Adv* **6**, eabb8097, doi:10.1126/sciadv.abb8097 (2020).
- 7 Parks, J. M. & Smith, J. C. How to Discover Antiviral Drugs Quickly. *N Engl J Med* **382**, 2261-2264, doi:10.1056/NEJMcibr2007042 (2020).
- 8 Guy, R. K., DiPaola, R. S., Romanelli, F. & Dutch, R. E. Rapid repurposing of drugs for COVID-19. *Science* **368**, 829-830, doi:10.1126/science.abb9332 (2020).
- 9 Levin, J. M. *et al.* Artificial intelligence, drug repurposing and peer review. *Nat Biotechnol* **38**, 1127-1131, doi:10.1038/s41587-020-0686-x (2020).
- 10 Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045 e1039, doi:10.1016/j.cell.2020.04.026 (2020).
- 11 Mangalmurti, N. & Hunter, C. A. Cytokine Storms: Understanding COVID-19. *Immunity* **53**, 19-25, doi:10.1016/j.immuni.2020.06.017 (2020).
- 12 Zhong, J., Tang, J., Ye, C. & Dong, L. The immunology of COVID-19: is immune modulation an option for treatment? *Lancet Rheumatol* **2**, e428-e436, doi:10.1016/S2665-9913(20)30120-X (2020).
- 13 Rubinstein, A. *et al.* BioNSi: A Discrete Biological Network Simulator Tool. *J Proteome Res* **15**, 2871-2880, doi:10.1021/acs.jproteome.6b00278 (2016).
- 14 Yeheskel, A., Reiter, A., Pasmanik-Chor, M. & Rubinstein, A. Simulation and visualization of multiple KEGG pathways using BioNSi. *F1000Res* **6**, 2120, doi:10.12688/f1000research.13254.2 (2017).
- 15 Alaimo S., R. R. V., Marceca G.P., La Ferlita A. , Serebrennikova O.B., Tschlis P.N., Mishra B., Pulvirenti A., Ferro A. PHENSIM: Phenotype Simulator. *BioRxiv*, doi:<https://doi.org/10.1101/2020.01.20.912279> (2020).
- 16 Cantwell, G. T. & Newman, M. E. J. Message passing on networks with loops. *Proc Natl Acad Sci U S A* **116**, 23398-23403, doi:10.1073/pnas.1914893116 (2019).
- 17 Konwar, K. M., Hanson, N. W., Page, A. P. & Hallam, S. J. MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. *BMC Bioinformatics* **14**, 202, doi:10.1186/1471-2105-14-202 (2013).
- 18 Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet* **21**, 71-87, doi:10.1038/s41576-019-0173-8 (2020).
- 19 Alaimo, S. *et al.* DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. *BMC Syst Biol* **9 Suppl 3**, S4, doi:10.1186/1752-0509-9-S3-S4 (2015).

- 20 Alaimo, S. *et al.* Post-transcriptional knowledge in pathway analysis increases the accuracy of phenotypes classification. *Oncotarget* **7**, 54572-54582, doi:10.18632/oncotarget.9788 (2016).
- 21 Sirota, M. *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* **3**, 96ra77, doi:10.1126/scitranslmed.3001318 (2011).
- 22 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278, doi:10.1016/j.cell.2020.02.052 (2020).
- 23 Young, B. E. *et al.* Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. *JAMA*, doi:10.1001/jama.2020.3204 (2020).
- 24 Xiao, F. *et al.* Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology* **158**, 1831-1833 e1833, doi:10.1053/j.gastro.2020.02.055 (2020).
- 25 Klann, K., Tascher, G. & Munch, C. Functional Translatome Proteomics Reveal Converging and Dose-Dependent Regulation by mTORC1 and eIF2alpha. *Mol Cell* **77**, 913-925 e914, doi:10.1016/j.molcel.2019.11.010 (2020).
- 26 Schrezenmeier, E. & Dorner, T. Mechanisms of action of hydroxychloroquine and chloroquine: implications for rheumatology. *Nat Rev Rheumatol* **16**, 155-166, doi:10.1038/s41584-020-0372-x (2020).
- 27 Pers, Y. M. & Padern, G. Revisiting the cardiovascular risk of hydroxychloroquine in RA. *Nat Rev Rheumatol* **16**, 671-672, doi:10.1038/s41584-020-00521-x (2020).
- 28 Lane, J. C. E. *et al.* Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. *Lancet Rheumatol* **2**, e698-e711, doi:10.1016/S2665-9913(20)30276-9 (2020).
- 29 Group, R. C. *et al.* Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*, doi:10.1056/NEJMoa2021436 (2020).
- 30 Draghici S., N. T., Sonna L.A., Ziraldo C., Vanciu R., Fadel R., Morrison A., Ramesh M., I Mor G. COVID-19: disease pathways and gene expression changes predict methylprednisolone can improve outcome in severe cases. *MedRxiv*, doi:<https://doi.org/10.1101/2020.05.06.20076687> (2019).
- 31 Roschewski, M. *et al.* Inhibition of Bruton tyrosine kinase in patients with severe COVID-19. *Sci Immunol* **5**, doi:10.1126/sciimmunol.abd0110 (2020).
- 32 Bramante, C. T. *et al.* Metformin and risk of mortality in patients hospitalised with COVID-19: a retrospective cohort analysis. *Lancet Healthy Longev* **2**, e34-e41, doi:10.1016/S2666-7568(20)30033-7 (2021).
- 33 Peiffer-Smadja, N., Rebeaud, M. E., Guihur, A., Mahamat-Saleh, Y. & Fiolet, T. Hydroxychloroquine and COVID-19: a tale of populism and obscurantism. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30866-5 (2020).
- 34 Terpos, E. *et al.* Hematological findings and complications of COVID-19. *Am J Hematol* **95**, 834-847, doi:10.1002/ajh.25829 (2020).
- 35 Naymagon, L. *et al.* The emergence of methemoglobinemia amidst the COVID-19 pandemic. *Am J Hematol* **95**, E196-E197, doi:10.1002/ajh.25868 (2020).
- 36 Faisal, H., Bloom, A. & Gaber, A. O. Unexplained Methemoglobinemia in Coronavirus Disease 2019: A Case Report. *A A Pract* **14**, e01287, doi:10.1213/XAA.0000000000001287 (2020).
- 37 Fiolet, T. *et al.* Effect of hydroxychloroquine with or without azithromycin on the mortality of coronavirus disease 2019 (COVID-19) patients: a systematic review and meta-analysis. *Clin Microbiol Infect* **27**, 19-27, doi:10.1016/j.cmi.2020.08.022 (2021).
- 38 Cao, L. *et al.* De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426-431, doi:10.1126/science.abd9909 (2020).

- 39 Menendez, C. A., Bylehn, F., Perez-Lemus, G. R., Alvarado, W. & de Pablo, J. J. Molecular characterization of ebbsen binding activity to SARS-CoV-2 main protease. *Sci Adv* **6**, doi:10.1126/sciadv.abd0345 (2020).
- 40 Jin, Z. *et al.* Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289-293, doi:10.1038/s41586-020-2223-y (2020).
- 41 Kowalewski, J. & Ray, A. Predicting novel drugs for SARS-CoV-2 using machine learning from a >10 million chemical space. *Heliyon* **6**, e04639, doi:10.1016/j.heliyon.2020.e04639 (2020).
- 42 Riva, L. *et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **586**, 113-119, doi:10.1038/s41586-020-2577-1 (2020).
- 43 Touret, F. *et al.* In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *Sci Rep* **10**, 13093, doi:10.1038/s41598-020-70143-6 (2020).
- 44 Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, doi:10.1038/s41586-020-2286-9 (2020).
- 45 Calabrese L.H., W. K., Strand V., Yazdany J., Walter J.E. Type I interferon, anti-interferon antibodies, and COVID-19. *The Lancet Rheumatology*, doi:[https://doi.org/10.1016/S2665-9913\(21\)00034-5](https://doi.org/10.1016/S2665-9913(21)00034-5) (2021).
- 46 Bronte, V. *et al.* Baricitinib restrains the immune dysregulation in patients with severe COVID-19. *J Clin Invest* **130**, 6409-6416, doi:10.1172/JCI141772 (2020).
- 47 Kalil, A. C. *et al.* Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. *N Engl J Med*, doi:10.1056/NEJMoa2031994 (2020).
- 48 Calabrese, L. H. & Calabrese, C. Baricitinib and dexamethasone for hospitalized patients with COVID-19. *Cleve Clin J Med*, doi:10.3949/ccjm.88a.ccc073 (2021).
- 49 Wang, K., Chen W., Zhou Y., Lian J., Zhang Z., Du P., ..., Chen Z. SARS-CoV-2 invades host cells via a novel route: CD147-spike protein *BioRxiv*, doi:<https://doi.org/10.1101/2020.03.14.988345> (2019).
- 50 Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506, doi:10.1038/s41596-020-0292-x (2020).
- 51 Thoms, M. *et al.* Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **369**, 1249-1255, doi:10.1126/science.abc8665 (2020).
- 52 Yang, L. *et al.* A Human Pluripotent Stem Cell-based Platform to Study SARS-CoV-2 Tropism and Model Virus Infection in Human Cells and Organoids. *Cell Stem Cell* **27**, 125-136 e127, doi:10.1016/j.stem.2020.06.015 (2020).
- 53 Altmann, D. M. & Boyton, R. J. SARS-CoV-2 T cell immunity: Specificity, function, durability, and role in protection. *Sci Immunol* **5**, doi:10.1126/sciimmunol.abd6160 (2020).
- 54 Krausgruber, T. *et al.* Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296-302, doi:10.1038/s41586-020-2424-4 (2020).
- 55 Hou, Y. J. *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* **182**, 429-446 e414, doi:10.1016/j.cell.2020.05.042 (2020).
- 56 Cantuti-Castelvetri, L. *et al.* Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856-860, doi:10.1126/science.abd2985 (2020).
- 57 Daly, J. L. *et al.* Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **370**, 861-865, doi:10.1126/science.abd3072 (2020).
- 58 Finkel, Y. *et al.* The coding capacity of SARS-CoV-2. *Nature*, doi:10.1038/s41586-020-2739-1 (2020).
- 59 Gupta, R. Advancing new tools for infectious diseases. *Science* **370**, 913-914, doi:10.1126/science.abe0773 (2020).

- 60 Abrams, R. P. M. *et al.* Therapeutic candidates for the Zika virus identified by a high-throughput  
screen for Zika protease inhibitors. *Proc Natl Acad Sci U S A* **117**, 31365-31375,  
doi:10.1073/pnas.2005463117 (2020).
- 61 Alaimo, S., Marceca, G. P., Ferro, A. & Pulvirenti, A. Detecting Disease Specific Pathway  
Substructures through an Integrated Systems Biology Approach. *Noncoding RNA* **3**,  
doi:10.3390/ncrna3020020 (2017).
- 62 Huang, H. Y. *et al.* miRTarBase 2020: updates to the experimentally validated microRNA-target  
interaction database. *Nucleic Acids Res* **48**, D148-D154, doi:10.1093/nar/gkz896 (2020).
- 63 Xiao, F. *et al.* miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids  
Res* **37**, D105-110, doi:10.1093/nar/gkn851 (2009).
- 64 Tong, Z., Cui, Q., Wang, J. & Zhou, Y. TransmiR v2.0: an updated transcription factor-microRNA  
regulation database. *Nucleic Acids Res* **47**, D253-D258, doi:10.1093/nar/gky1023 (2019).
- 65 Storey J.D., B. A. J., Dabney A. and Robinson D. qvalue: Q-value estimation for false discovery  
rate control. R package version 2.20.0. <http://github.com/jdstorey/qvalue>. *Github* (2020).
- 66 Xiong, L., Edwards, C. K., 3rd & Zhou, L. The biological function and clinical utilization of CD147 in  
human diseases: a review of the current scientific literature. *Int J Mol Sci* **15**, 17411-17441,  
doi:10.3390/ijms151017411 (2014).
- 67 Fisher, R. A. 224A: Answer to Question 14 on Combining independent tests of significance. *The  
American Statistician*, **2**: 30 (1948).
- 68 Alaimo, S., Giugno, R. & Pulvirenti, A. Recommendation Techniques for Drug-Target Interaction  
Prediction and Drug Repositioning. *Methods Mol Biol* **1415**, 441-462, doi:10.1007/978-1-4939-  
3572-7\_23 (2016).
- 69 Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* **25**, 75-82,  
doi:10.1093/bioinformatics/btn577 (2009).
- 70 Yu, G. & He, Q. Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and  
visualization. *Mol Biosyst* **12**, 477-479, doi:10.1039/c5mb00663e (2016).
- 71 S., G. GOfuncR: Gene ontology enrichment using FUNC. R package version 1.10.0. *Bioconductor  
version: Release (1.8.0)* (2020).
- 72 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for  
RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

## **ACKNOWLEDGEMENTS AND FUNDING**

N.I.M. was funded in part by a fellowship award from the Netherlands-Caribbean Foundation for Clinical Higher Education (NASKHO). S.A., A.F. and A.P. have been partially supported by the MIUR PON research project BILIGeCT “Liquid Biopsies for Cancer Clinical Management”. National Cancer Institute Physical Sciences-Oncology Center Grant U54 CA193313-01 (to B.M.), and US Army grant W911NF1810427 (to B.M.).

## **AUTHOR CONTRIBUTIONS**

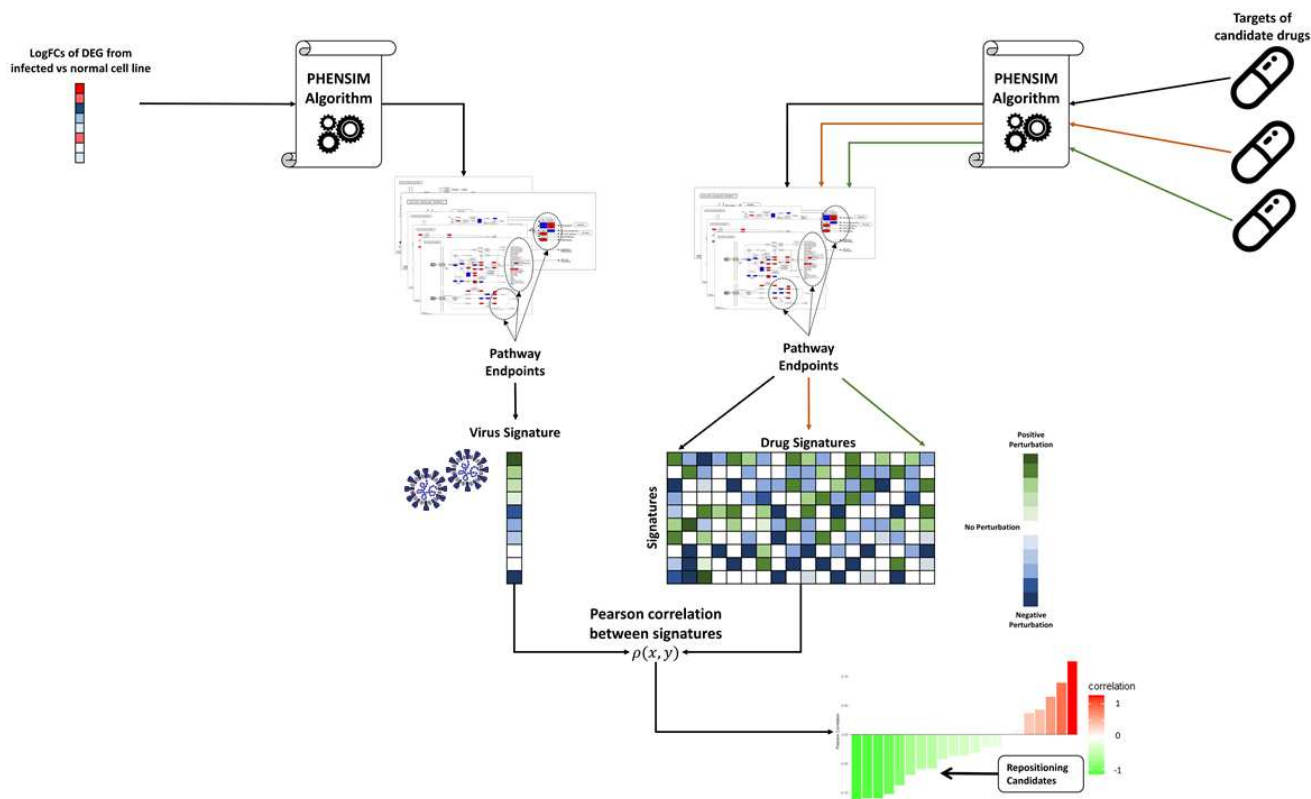
*Concept, design, drafting and revision of the manuscript:* NIM, RVR, SA, BM, AJD, AF

*Concept, design, and revisions of the manuscript:* EB, AS, JACB, AP

## **COMPETING INTERESTS**

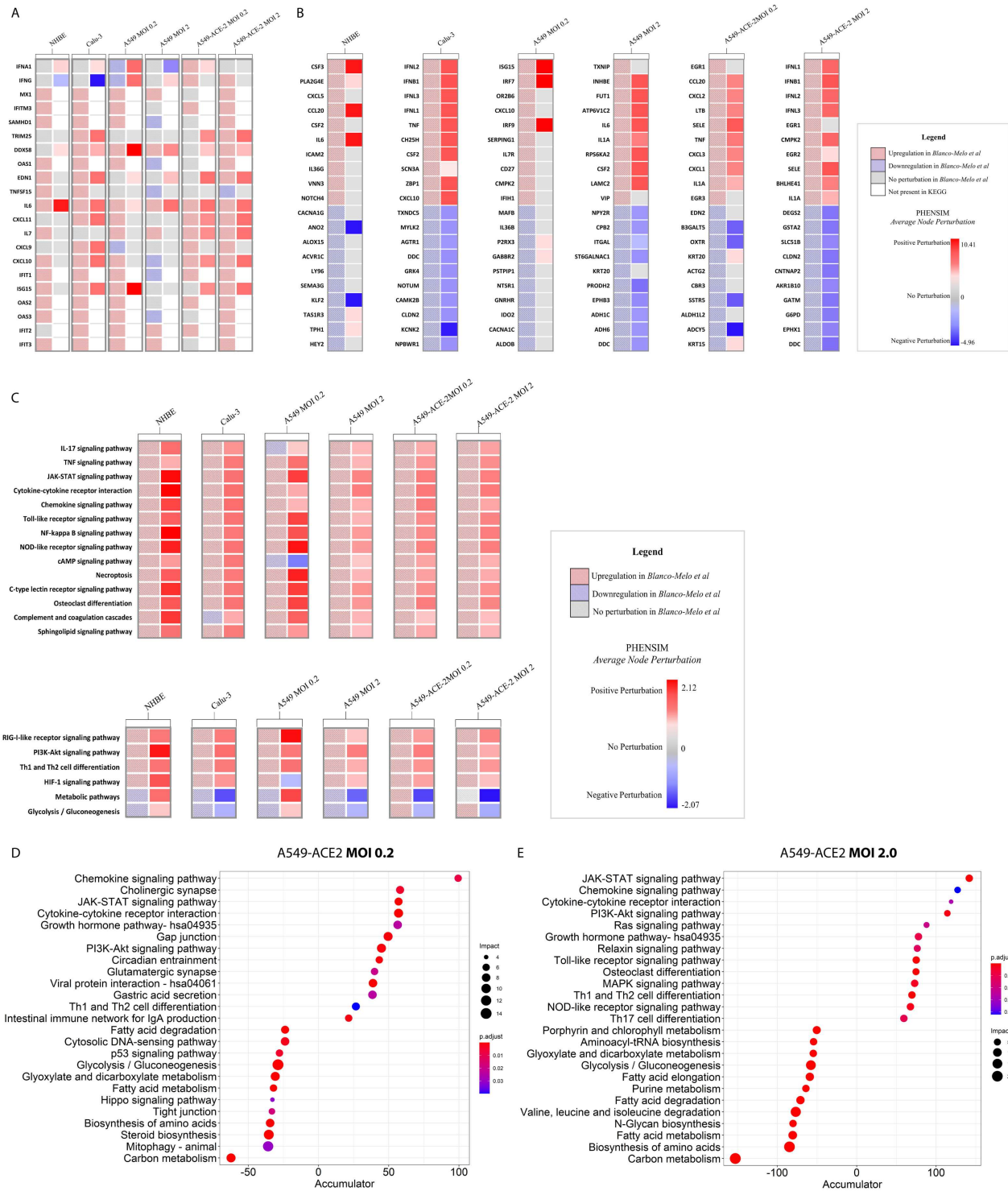
All authors declare no competing interests.

## FIGURES



**Figure 1. Schematic representation of the PHENSIM Drug repurposing Strategy.** Outline for our approach to acquire a cell-specific viral signature *in silico* using a *Transcriptomic strategy*: logFold Changes (logFCs) of Differentially Expressed Genes (DEGs) arising from transcriptomic genome wide expression analysis of SARS-CoV-2 infected vs. baseline uninfected cells, cell-lines and tissues are the main input for the *PHENotype SIMulator*. Once a cell-specific *viral signature* is defined based on gene and signaling pathway endpoints using KEGG meta-pathway analysis, PHENSIM can be exploited to search for possible repositioning candidates by building a *drug signature* database using the *Drug repurposing strategy*: multiple targets of drug candidates are used as input for PHENSIM to define drug signatures based on pathway endpoints. A Pearson correlation between the acquired virus and drug signatures  $\rho(x,y)$  gives rise to a correlation scoring system to evaluate drug repositioning candidates in a certain infected cell or tissue. Negative correlation (green) predicts promising targets that inhibit the viral signature and positive correlation (red) suggests exacerbation of the viral signature when introducing the drug.

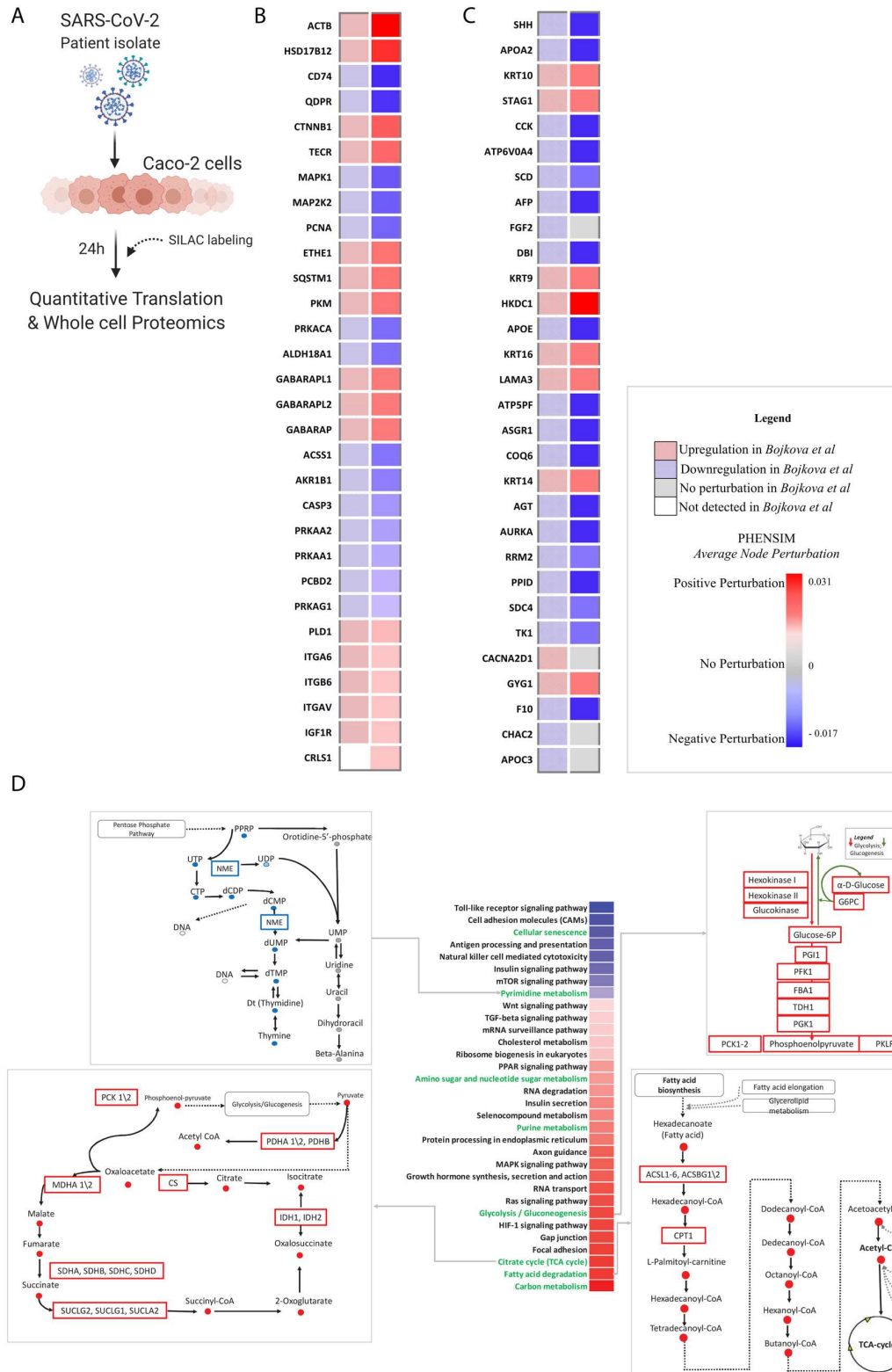




**Figure 2. *In silico* PHENSIM prediction of host transcriptional response to SARS-Cov-2. *In vitro* results from Blanco-Melo *et al.* (left column; checkered boxes) are compared to *in silico* PHENSIM predictions (right; solid) for all evaluated respiratory related cells assessed; NHBE, Calu-3, A549 cells at**



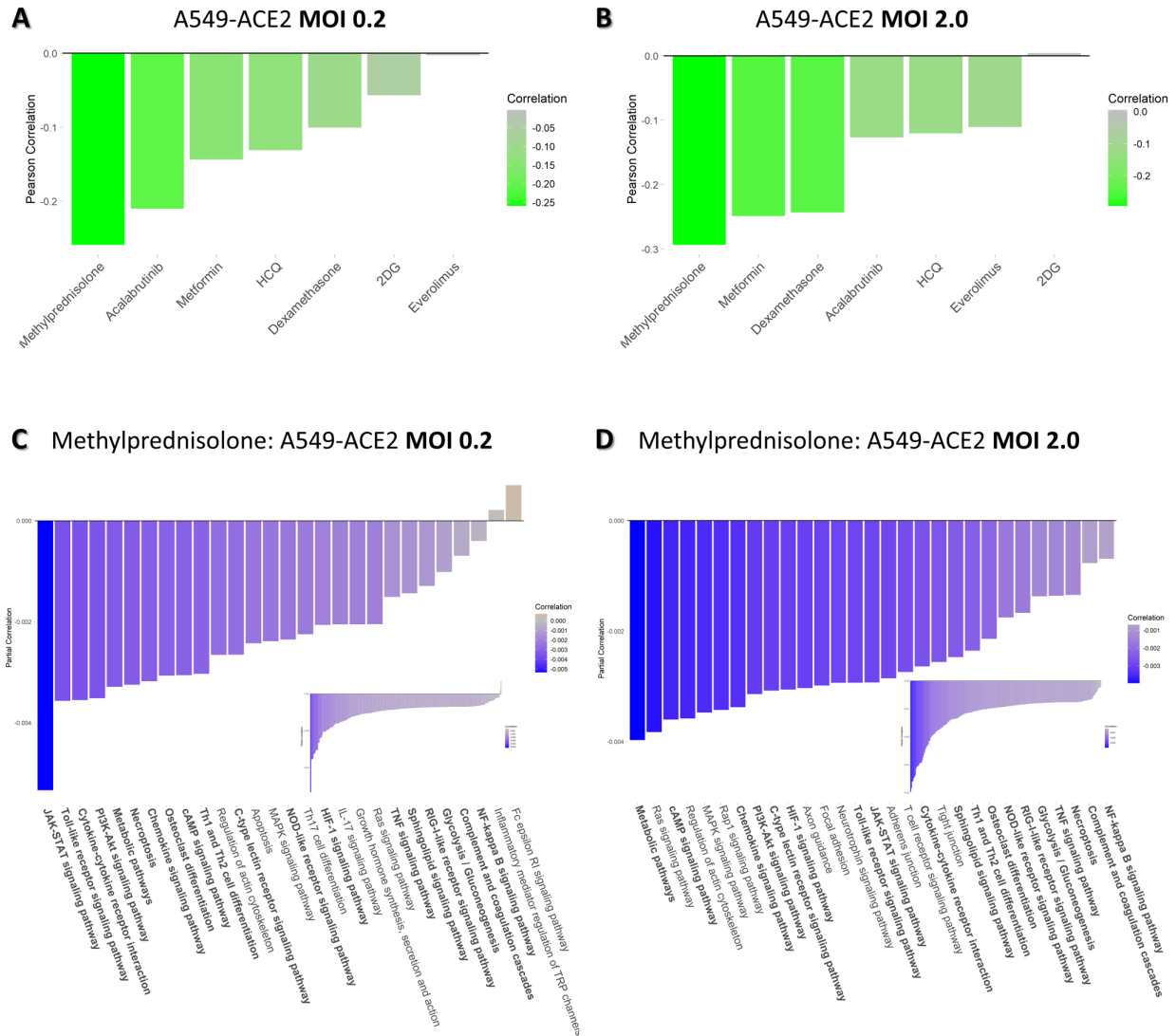
low (0.2) and high (2.0) MOI,  $\pm$  ACE2 transduction respectively. **A)** Heatmap depicting the perturbation of a select subset of anti-viral, ISGs and inflammatory genes. **B)** Heatmaps depicting unbiased analysis of the top-10 upregulated (red) and top-10 downregulated (blue) DEGs from *Blanco-Melo et al.* (left) with side-by-side PHENSIM predictions (right). For A&B, legend shows denoted perturbations for PHENSIM prediction and Blanco-Melo *et al.* See legend box for DEG annotation. **C)** Heatmap depicts whole genome pathway analysis as predicted by PHENSIM for a select set of signaling pathways of interest in all assessed cell types. Pathway selection was based on highlighted pathways affected by SARS-CoV-2 infection. Color gradient depicts the average pathway perturbation as predicted in our PHENSIM in silico experiments. **D&E)** MITHrIL pathway analysis was used to assess top meta-pathways for **D)** A549-ACE2 MOI 0.2 (low viral load) and **E)** A549-ACE2 MOI 2.0 (high viral load), according to impact (circle size) and significance (color-gradient for adjusted p-value) for the top 12 up- (+accumulator) and down-regulated pathways. The accumulator is the accumulation/sum of all perturbations computed for that particular pathway. NHBE; Normal Human Bronchial Epithelial cells, Calu-3; Cultured human airway epithelial cells, A549; Transformed lung alveolar cells, ACE2; angiotensin-converting enzyme, MOI; multiplicity of infection. DEGs; Differentially expressed genes, ISGs; IFN-stimulated genes.



**Figure 3. PHENSIM proteomic pathway analysis in SARS-CoV-2-infected human host cells.**

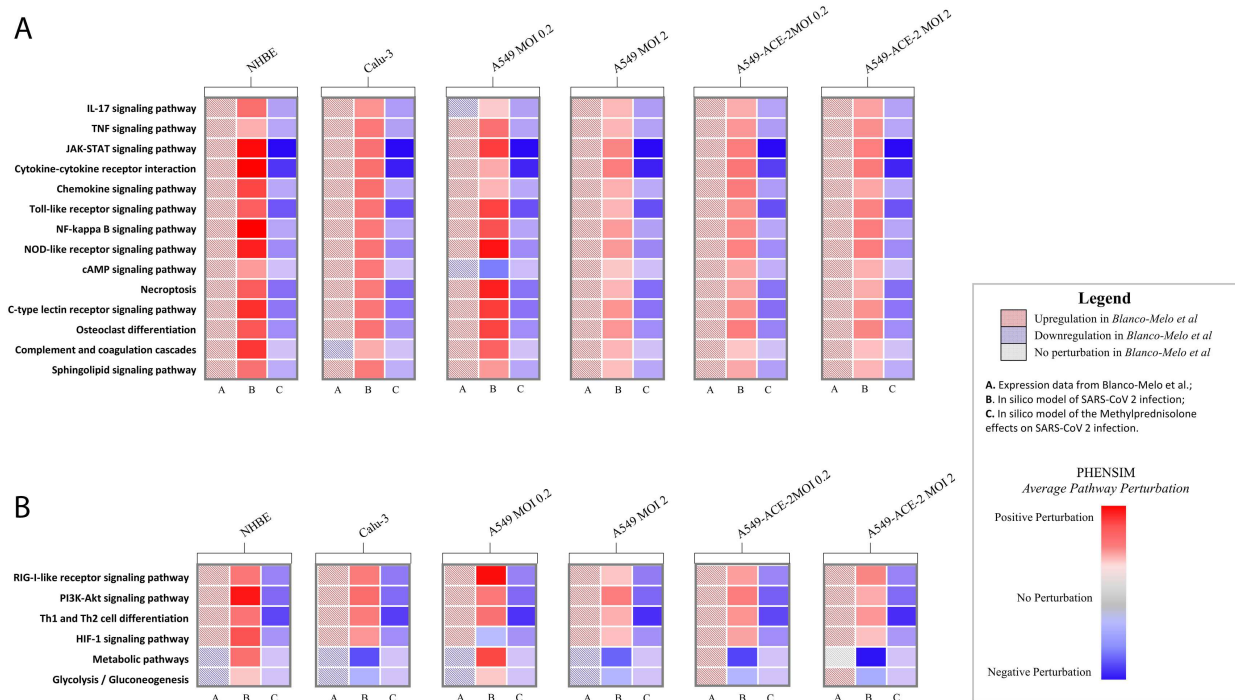
PHENSIM pathway analysis of the Caco-2 cell experiment was simulated *in silico* to reproduce *in vitro* results presented by Bojkova *et al.* at the 24hour time-point post SARS-CoV-2 infection **A**) Schematic

representation depicting the experimental design as described by Bojkova *et al. in vitro*: the human colon epithelial carcinoma cell line, Caco-2 cells, were infected and monitored for 24hrs post SARS-CoV-2 infection. Naturally occurring heavy isotype SILAC labelling was used to quantify translational changes, as this method does not affect cellular behavior allowing for unbiased pathway analysis. Quantitative translation and whole cell proteomics by LC-MS/MS was performed <sup>5</sup>. **B&C)** Heatmaps depicting a representative subset of the 30 top differentially expressed proteins (FDR<0.05) involved in viral infection after 24hr SARS-CoV-2 infection **B)** as predicted by PHENSIM *in silico* (right column, solid squares), compared to expression results as determined by Bojkova *et al.* (left column, checkered squares) and **C)** as described by Bojkova *et al.* (left column, checkered) with side-by-side PHENSIM expression prediction for that protein (right column, solid). **D)** Heatmap depicts PHENSIM simulated results *in silico* for the top 30 signaling pathways significantly affected at 24h post infection; Up- (red) and Down-regulated (blue). Pathways depicted in *green* text are signaling pathways described as significant by Bojkova *et al.* in their analysis. A select simplified KEGG-based pathways are highlighted on protein interaction level, displaying upregulated proteins in red and downregulated proteins in blue. Color gradient reflects PHENSIM activity; the value of the activity score attributed to each pathway from blue (downregulation) to red (maximum upregulation). Caco-2; the human colon epithelial carcinoma cell line, SILAC; Stable Isotype Labeling by Amino Acids in Cell culture, LC-MS/MS; Liquid chromatography mass spectrometry, DEPs; Differentially expressed proteins, Max; maximum.



**Figure 4. Drug repositioning candidates for COVID-19.** We leverage our PHENSIM drug strategy approach to test candidate drugs for potential repurposing for COVID-19 treatment. Once a cell-specific viral signature is defined, it can be exploited to search for possible repositioning candidates by building a drug signature database. A Pearson correlation  $\rho(x,y)$  between the viral and drug signatures gives rise to a correlation score. Drug candidates having a positive effect on ameliorating SARS-CoV-2 infection have a negative correlation score (green) between viral and drug signature, whereas candidate drugs worsening disease correlate positively (red). Here we show distinct candidate drugs having a variable effect depending on the multiplicity of infection (MOI) of virus infection in A459-ACE2 expressing cells in **A**) low MOI 0.2 and **B**) high MOI 2.0. This analysis shows the modeling viral load dynamics and discerning what candidate could work best in low vs higher viral load. Resulted top pathways significantly

affected by Methylprednisolone treatment are depicted for **C**) low MOI 0.2 and **D**) high MOI 2.0. Drug candidates represented here: Methylprednisolone, Metformin (mTOR-inhibitor), (Hydroxy)chloroquine (HCQ-CQ), Acalabrutinib (BTK-inhibitor), Dexamethasone, 2-Deoxy-Glucose (2DG) and Everolimus (mTOR-inhibitor). ACE2; angiotensin-converting enzyme, MOI; multiplicity of infection.



**Figure 5. Methylprednisolone inhibits key inflammatory and viral signaling pathways in host lung and airway cells after SARS-CoV-2 infection.** Heatmap depicts the effects of Methylprednisolone *in silico* in SARS-CoV-2 infection on select signaling pathways of interest (similar pathways to Fig. 2C). From left to right, **column A** shows pathway analysis results of SARS-CoV-2 infection *in vitro* as performed using the MITHrIL algorithm; **column B** shows PHENSIM results of SARS-CoV-2 infection *in silico*; **column C** shows PHENSIM simulation results of Methylprednisolone on SARS-CoV-2 infected cells *in silico*. Color gradient depicts the average pathway perturbation as predicted in our PHENSIM *in silico* experiments for **column B&C**. NHBE; Normal Human Bronchial Epithelial cells, Calu-3; Cultured human airway epithelial cells, A549; Transformed lung alveolar cells, ACE2; angiotensin-converting enzyme, MOI; multiplicity of infection.

## TABLES

**Table 1. PHENSIM transcriptomic predicted values from Blanco-Melo et al. 2020.**

	Overall Accuracy	Nodes Predicted as perturbed			Nodes predicted as non-perturbed	
		PPV	Sensitivity	Specificity	PPV	FNR
<b>A549-ACE2 MOI 0.2</b>	51.66%	93.50%	96.72%	97.67%	58.90%	41.10%
<b>A549-ACE2 MOI 2</b>	71.72%	96.88%	99.24%	99.13%	60.04%	39.96%
<b>A549 MOI 0.2</b>	83.74%	68.75%	100.00%	99.86%	85.64%	14.36%
<b>A549 MOI 2</b>	78.20%	97.41%	97.20%	99.58%	77.47%	22.53%
<b>Calu-3</b>	77.17%	96.93%	99.34%	99.30%	76.55%	23.45%
<b>NHBE</b>	82.43%	67.65%	95.83%	99.69%	86.48%	13.52%

PPV; Positive predictive value, FNR; False negative rate, NHBE; Normal Human Bronchial Epithelial cells

**Table 2. PHENSIM proteomic predicted values from Bojkova et al. 2020.**

Time (hours)	All Proteins	Proteins in Meta-pathway	Predicted Percentage	Accuracy	PPV	Sensitivity	Specificity
<b>2</b>	5809	1914	6.95%	93.98%	95.45%	92.65%	95.38%
<b>6</b>	5809	1914	11.70%	93.75%	94.35%	97.66%	81.13%
<b>10</b>	5809	1914	10.45%	94.50%	95.27%	98.17%	77.78%
<b>24</b>	5809	1914	34.95%	97.91%	98.39%	97.87%	97.96%

*All Proteins*; number (N) of proteins quantified for each timepoint.

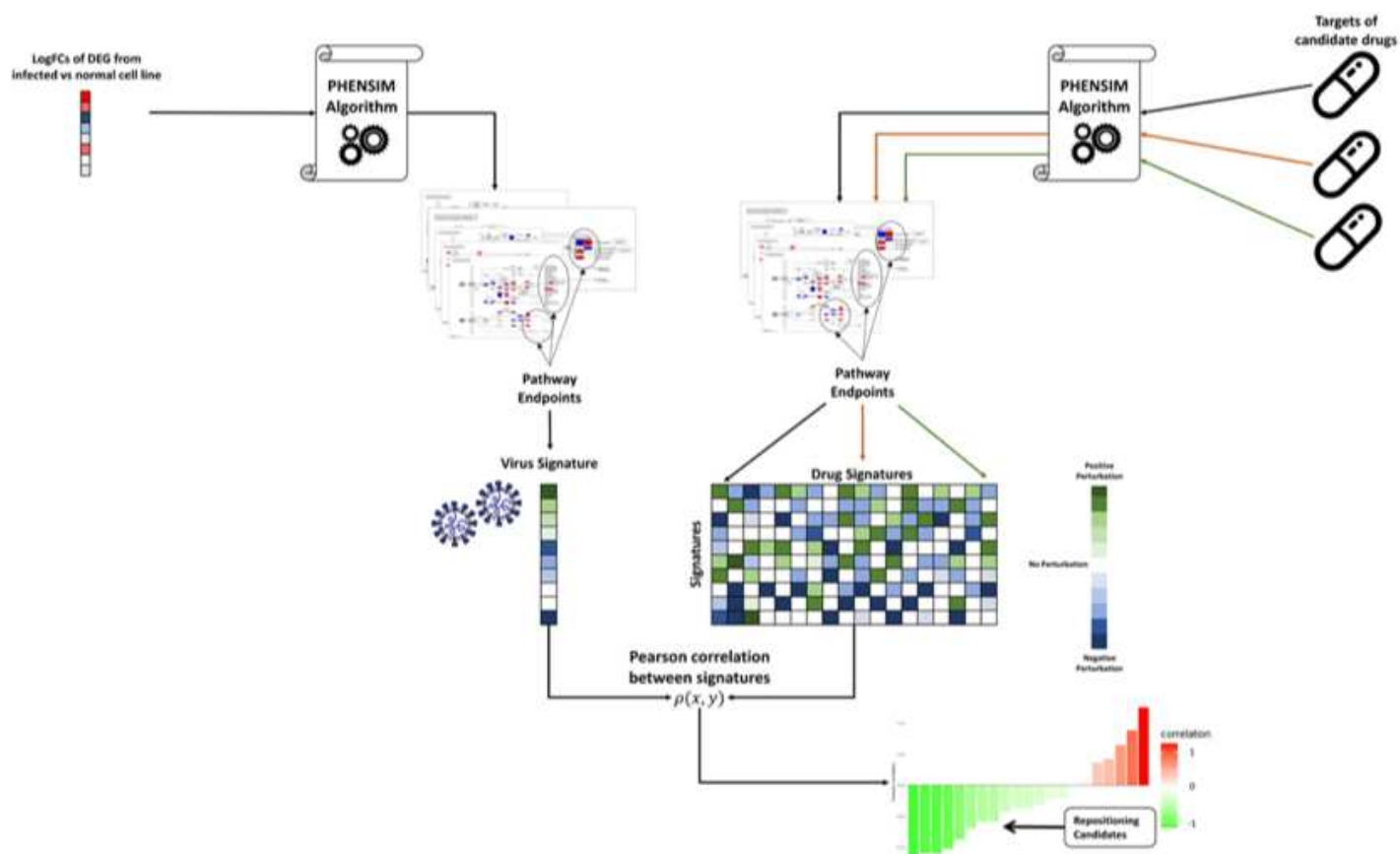
*Proteins in Meta-pathway*; number (N) of proteins present in KEGG, and therefore in the meta-pathway.

*Predicted Percentage*; Percent (%) proteins for which PHENSIM could produce a prediction.

*Accuracy, PPV, Sensitivity, and Specificity*; the metrics used to compare our model with the actual proteomics data.

*PPV*; Positive predictive value.

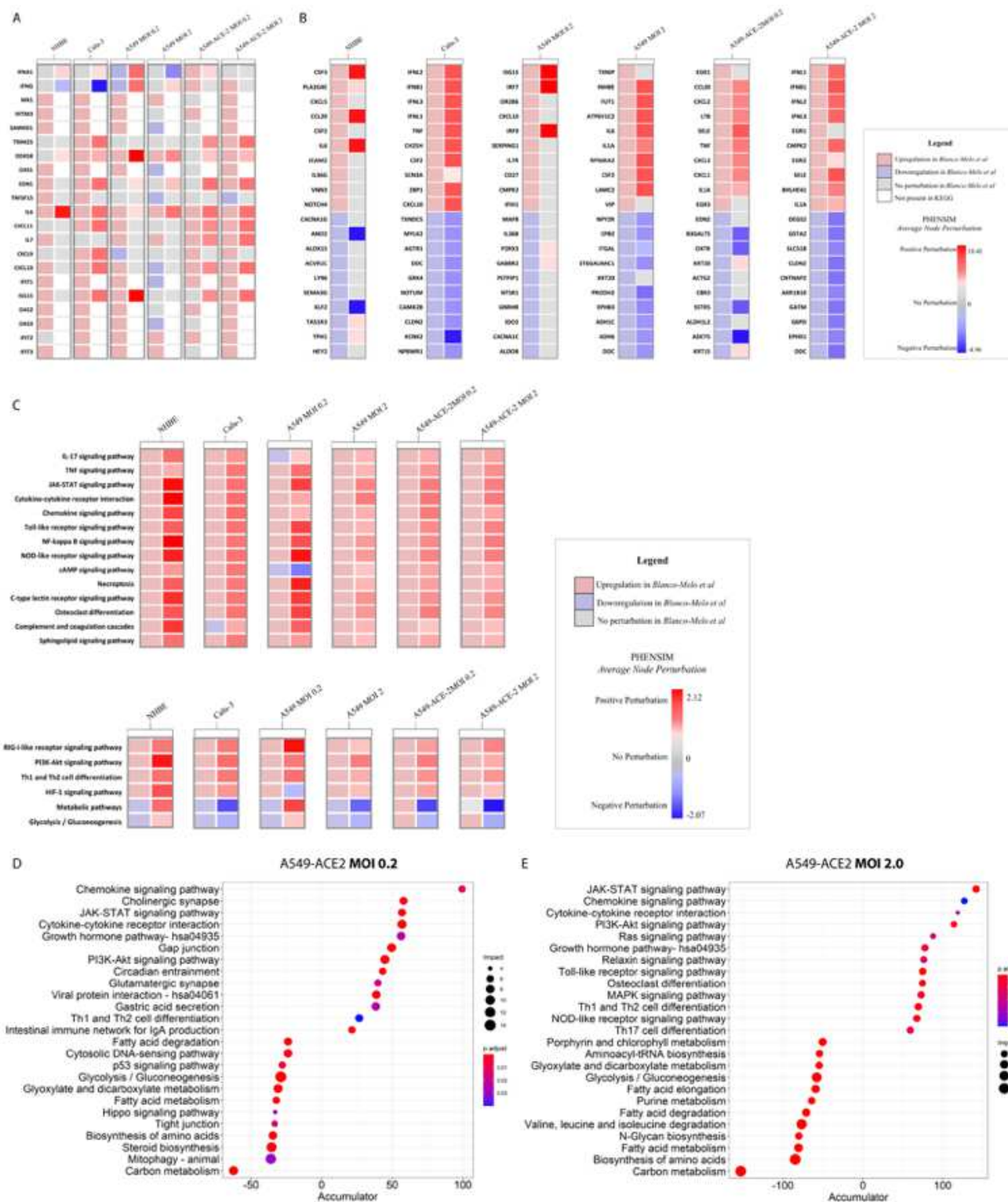
# Figures



**Figure 1**

Schematic representation of the PHENSIM Drug repurposing Strategy. Outline for our approach to acquire a cell-specific viral signature in silico using a Transcriptomic strategy: logFold Changes (logFCs) of Differentially Expressed Genes (DEGs) arising from transcriptomic genome wide expression analysis of SARS-CoV-2 infected vs. baseline uninfected cells, cell-lines and tissues are the main input for the PHENotype SIMulator. Once a cell-specific viral signature is defined based on gene and signaling pathway endpoints using KEGG meta-pathway analysis, PHENSIM can be exploited to search for possible repositioning candidates by building a drug signature database using the Drug repurposing strategy: multiple targets of drug candidates are used as input for PHENSIM to define drug signatures based on pathway endpoints. A Pearson correlation between the acquired virus and drug signatures  $\rho(x,y)$  gives rise to a correlation scoring system to evaluate drug repositioning candidates in a certain infected cell or tissue. Negative correlation (green) predicts promising targets that inhibit the viral signature and positive correlation (red) suggests exacerbation of the viral signature when introducing the drug.





**Figure 2**

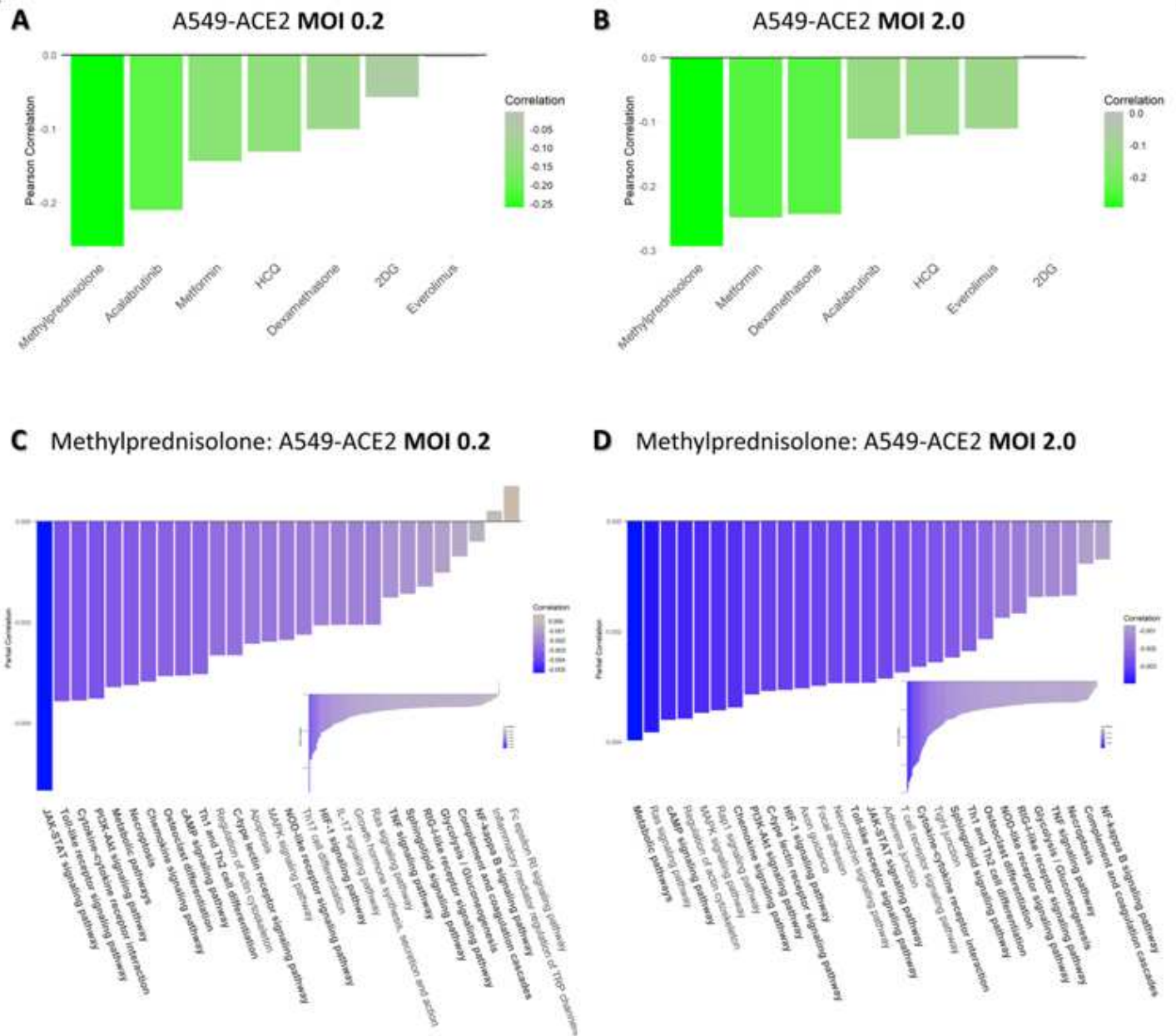
In silico PHENSIM prediction of host transcriptional response to SARS-Cov-2. In vitro results from Blanco-Melo et al. (left column; checked boxes) are compared to in silico PHENSIM predictions (right; solid) for all evaluated respiratory related cells assessed; NHBE, Calu-3, A549 cells at low (0.2) and high (2.0) MOI,  $\pm$  ACE2 transduction respectively. A) Heatmap depicting the perturbation of a select subset of anti-viral, ISGs and inflammatory genes. B) Heatmaps depicting unbiased analysis of the top-10 upregulated (red)



and top-10 downregulated (blue) DEGs from Blanco-Melo et al. (left) with side-by-side PHENSIM predictions (right). For A&B, legend shows denoted perturbations for PHENSIM prediction and Blanco-Melo et al. See legend box for DEG annotation. C) Heatmap depicts whole genome pathway analysis as predicted by PHENSIM for a select set of signaling pathways of interest in all assessed cell types. Pathway selection was based on highlighted pathways affected by SARS-CoV-2 infection. Color gradient depicts the average pathway perturbation as predicted in our PHENSIM in silico experiments. D&E) MITHrIL pathway analysis was used to assess top meta-pathways for D) A549-ACE2 MOI 0.2 (low viral load) and E) A549-ACE2 MOI 2.0 (high viral load), according to impact (circle size) and significance (color-gradient for adjusted p-value) for the top 12 up- (+accumulator) and down-regulated pathways. The accumulator is the accumulation/sum of all perturbations computed for that particular pathway. NHBE; Normal Human Bronchial Epithelial cells, Calu-3; Cultured human airway epithelial cells, A549; Transformed lung alveolar cells, ACE2; angiotensin-converting enzyme, MOI; multiplicity of infection. DEGs; Differentially expressed genes, ISGs; IFN-stimulated genes.



heavy isotype SILAC labelling was used to quantify translational changes, as this method does not affect cellular behavior allowing for unbiased pathway analysis. Quantitative translation and whole cell proteomics by LC-MS/MS was performed 5. B&C) Heatmaps depicting a representative subset of the 30 top differentially expressed proteins (FDR<0.05) involved in viral infection after 24hr SARS-CoV-2 infection B) as predicted by PHENSIM in silico (right column, solid squares), compared to expression results as determined by Bojkova et al. (left column, checkered squares) and C) as described by Bojkova et al. (left column, checkered) with side-by-side PHENSIM expression prediction for that protein (right column, solid). D) Heatmap depicts PHENSIM simulated results in silico for the top 30 signaling pathways significantly affected at 24h post infection; Up- (red) and Down-regulated (blue). Pathways depicted in green text are signaling pathways described as significant by Bojkova et al. in their analysis. A select simplified KEGG-based pathways are highlighted on protein interaction level, displaying upregulated proteins in red and downregulated proteins in blue. Color gradient reflects PHENSIM activity; the value of the activity score attributed to each pathway from blue (downregulation) to red (maximum upregulation). Caco-2; the human colon epithelial carcinoma cell line, SILAC; Stable Isotope Labeling by Amino Acids in Cell culture, LC-MS/MS; Liquid chromatography mass spectrometry, DEPs; Differentially expressed proteins, Max; maximum.



**Figure 4**

Drug repositioning candidates for COVID-19. We leverage our PHENSIM drug strategy approach to test candidate drugs for potential repurposing for COVID-19 treatment. Once a cell-specific viral signature is defined, it can be exploited to search for possible repositioning candidates by building a drug signature database. A Pearson correlation  $p(x,y)$  between the viral and drug signatures gives rise to a correlation score. Drug candidates having a positive effect on ameliorating SARS-CoV-2 infection have a negative correlation score (green) between viral and drug signature, whereas candidate drugs worsening disease correlate positively (red). Here we show distinct candidate drugs having a variable effect depending on the multiplicity of infection (MOI) of virus infection in A549-ACE2 expressing cells in A) low MOI 0.2 and B) high MOI 2.0. This analysis shows the modeling viral load dynamics and discerning what candidate could work best in low vs higher viral load. Resulted top pathways significantly affected by Methylprednisolone treatment are depicted for C) low MOI 0.2 and D) high MOI 2.0. Drug candidates

represented here: Methylprednisolone, Metformin (mTOR-inhibitor), (Hydroxy)chloroquine (HCQ-CQ), Acalabrutinib (BTK-inhibitor), Dexamethasone, 2-Deoxy-Glucose (2DG) and Everolimus (mTOR-inhibitor). ACE2; angiotensin-converting enzyme, MOI; multiplicity of infection.

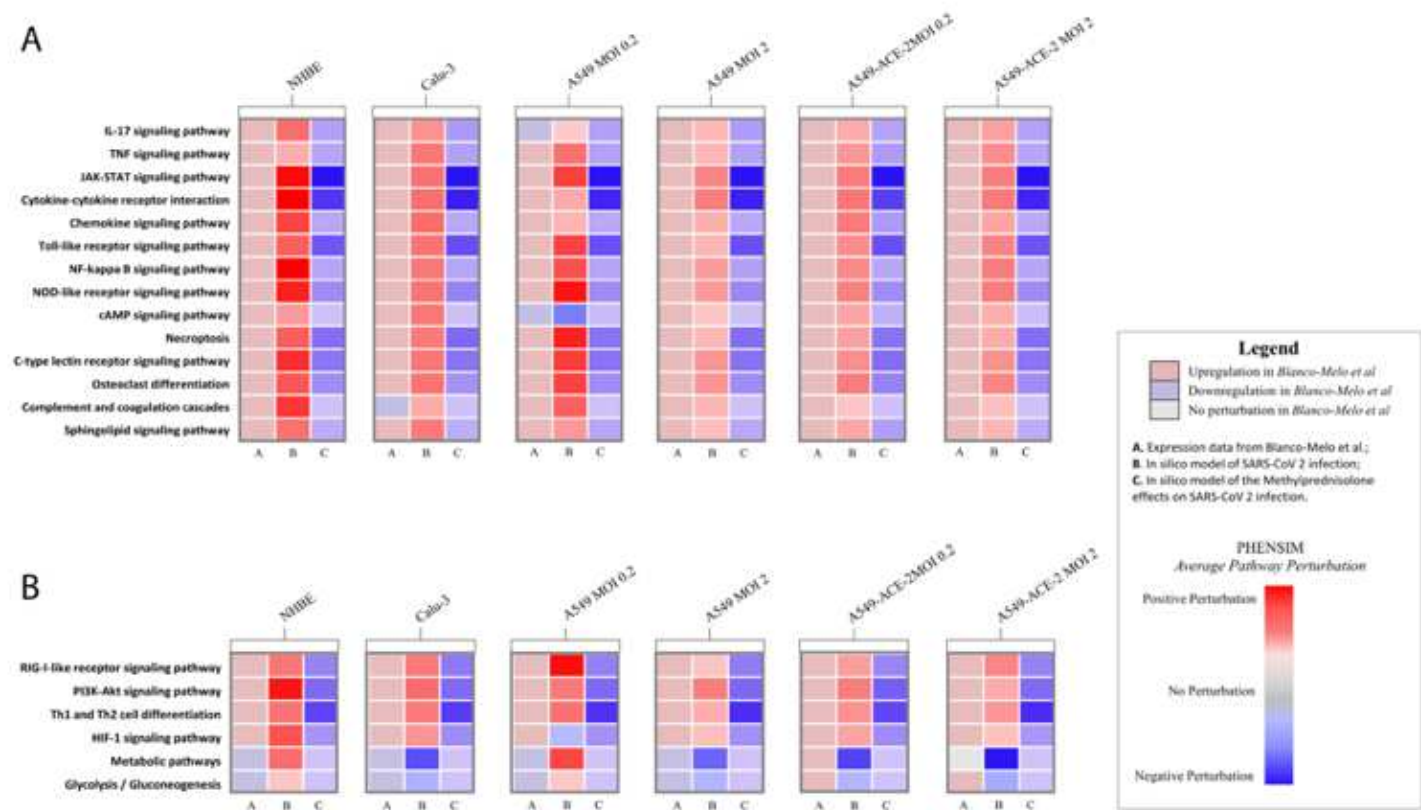


Figure 5

Methylprednisolone inhibits key inflammatory and viral signaling pathways in host lung and airway cells after SARS-CoV-2 infection. Heatmap depicts the effects of Methylprednisolone in silico in SARS-CoV-2 infection on select signaling pathways of interest (similar pathways to Fig. 2C). From left to right, column A shows pathway analysis results of SARS-CoV-2 infection in vitro as performed using the MITHrIL algorithm; column B shows PHENSIM results of SARS-CoV-2 infection in silico; column C shows PHENSIM simulation results of Methylprednisolone on SARS-CoV-2 infected cells in silico. Color gradient depicts the average pathway perturbation as predicted in our PHENSIM in silico experiments for column B&C. NHBE; Normal Human Bronchial Epithelial cells, Calu-3; Cultured human airway epithelial cells, A549; Transformed lung alveolar cells, ACE2; angiotensin-converting enzyme, MOI; multiplicity of infection.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalTableS1.CD147edges.xlsx](#)

- [SupplementalTableS2.Bojkovapvalueanalysis.xlsx](#)
- [MariaetalPHENSIMDrugRepurposingSupplementalmaterial.docx](#)
- [PathwayAnalysis1.zip](#)