

Drug Repurposing for COVID-19 through Weakly Supervised Learning

Nahal Mirzaie, Forbes Burkowski, Vahid Salimi, Ali Sharifi Zarchi, Mohammad Hossein Rohban

I Like many other viral infections, COVID-19 is caused by a virus that changes the morphology of host cells. These changes can be captured with fluorescent microscopes and used in **high-throughput screening (HTS)** experiments.

II Recursion Pharmaceuticals conducted HTS experiments to investigate the prophylactic effect of **1672 drugs** against COVID-19. Their results were published in a dataset called *RxRx19a*.

III We propose an innovative learning algorithm that is trained using only the morphologies of non-drugged healthy and infected cells and then can be used to predict the effectiveness of any unseen drug against the pathogen.

We show that our algorithm scores are **IV robust**, **V stable**, and **VI sensitive to drug toxicity and cell death**.

VII **Four out of five** top drugs based on our algorithm are in clinical studies for COVID-19 treatment. This includes **Remdesivir**.

INTRODUCTION

I During the last few decades, image-based High Throughput Screening (HTS) has proven its worth in facilitating drug discovery. This methodology relies on the fact that chemical components change the morphology of cells. These changes can be tracked and analyzed to select from drug candidates the hit that is most capable of reversing the cell morphologies that are impacted by the disease.

II The advent of a new Betacoronavirus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), led to the COVID-19 pandemic. During this critical time when discovery of a treatment for this unknown disease was urgently needed, the Recursion pharmaceutical company published the *RxRx19a* data set [Heiser et al., 2020]. This dataset was derived from HTS experiments to explore the therapeutic potential of 1672 drugs for COVID-19 treatment.

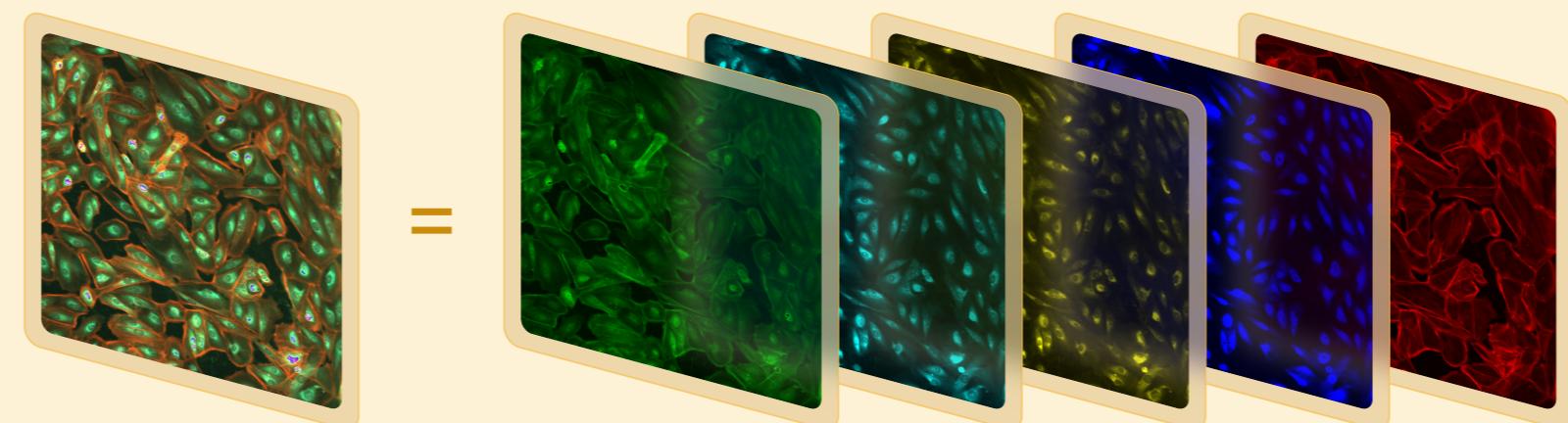


Figure 1

Site-level image of *RxRx19a* healthy non-drugged cells sample (HRCE-2, plate 23, well Q08, site 1). Samples are dyed with 5 fluorescent colors to indicate 5 different components in cells (Phalloidin, Hoechst, Concanavalin A, SYTO14, Wheat Germ Agglutinin).

Samples in this study were categorized into 3 classes, non-drugged healthy (mock and UV-irradiated SARS-CoV-2), infected (active SARS-CoV-2), and infected drugged samples (treatments). For preparing treatment samples, a drug with a specific dosage was added post-seeding, and then samples were contaminated with the active SARS-CoV-2 virus.

Having this valuable dataset, we seek to find the best drug for reverting the Cytopathic effect of SARS-CoV-2.

METHOD

III We present a pipeline that can estimate the prophylactic effect of a drug deployed against a pathogen by using an innovative learning algorithm inspired by [Caicedo et al., 2018].

- Our model uses only the morphologies of non-drugged healthy and infected cells, and does not require the morphologies of healthy (uninfected) cells that have been given the drug.
- We extracted phenotypic traits at the single-cell level using a *CellProfiler* pipeline.
- Normalization and batch effect removal applied on site-level profiles to ensure assay quality.

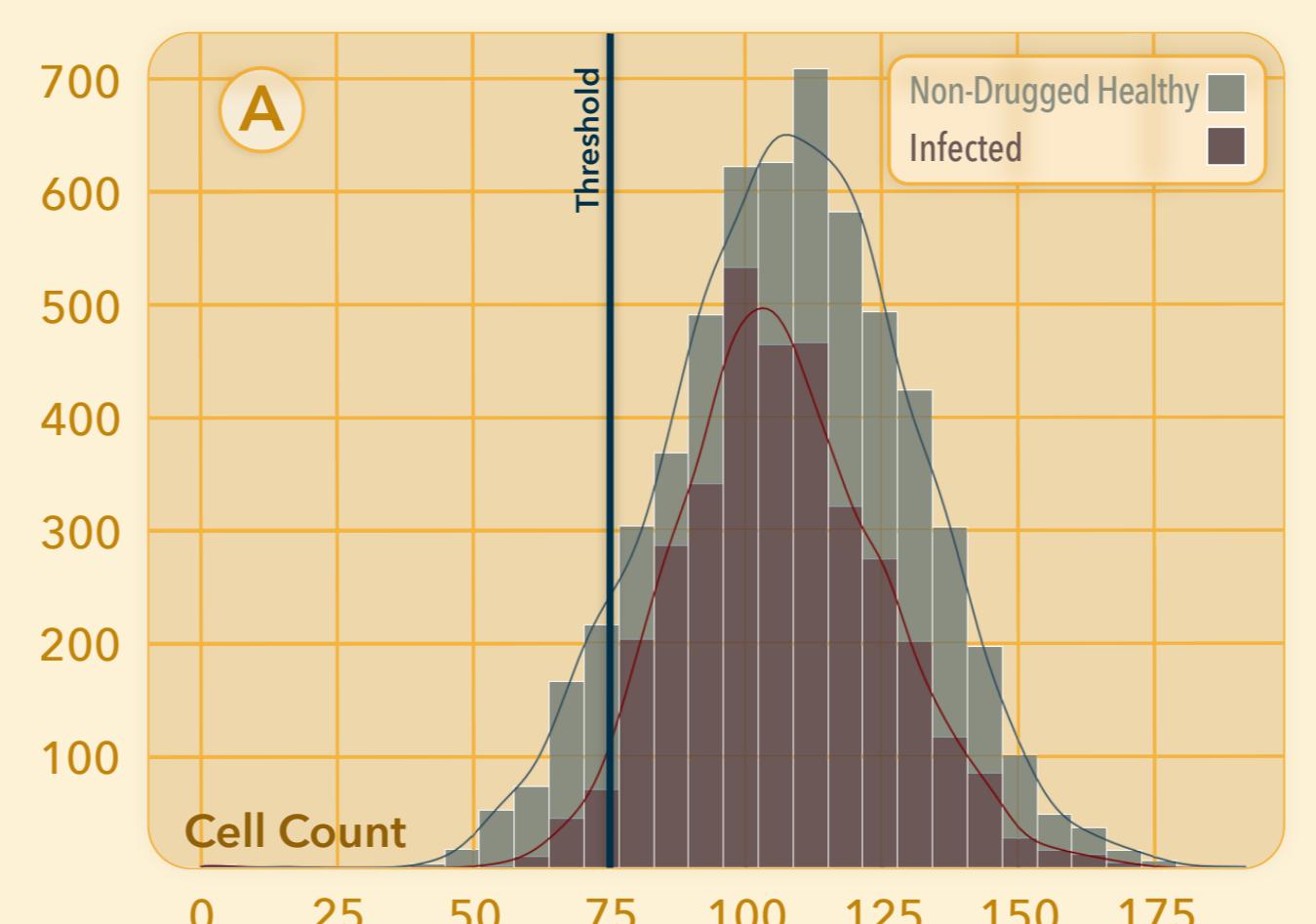


Figure 2

VI **A** Well-level cell count histogram. Wells with less than 75 cells show abnormal cell death. Because most of samples below the threshold are from healthy samples, models easily converge to learn and predict cell death and consequently drug toxicity as a healthy phenotype. We manually labelled cell death as disease before training our model. **B** Digoxin and Mitoxantrone, rank 2nd and 5th best drug proposed by [Saberian et al., 2021], Thimerosal, rank 5th proposed by [Mascolini et al., 2021] show cell death.

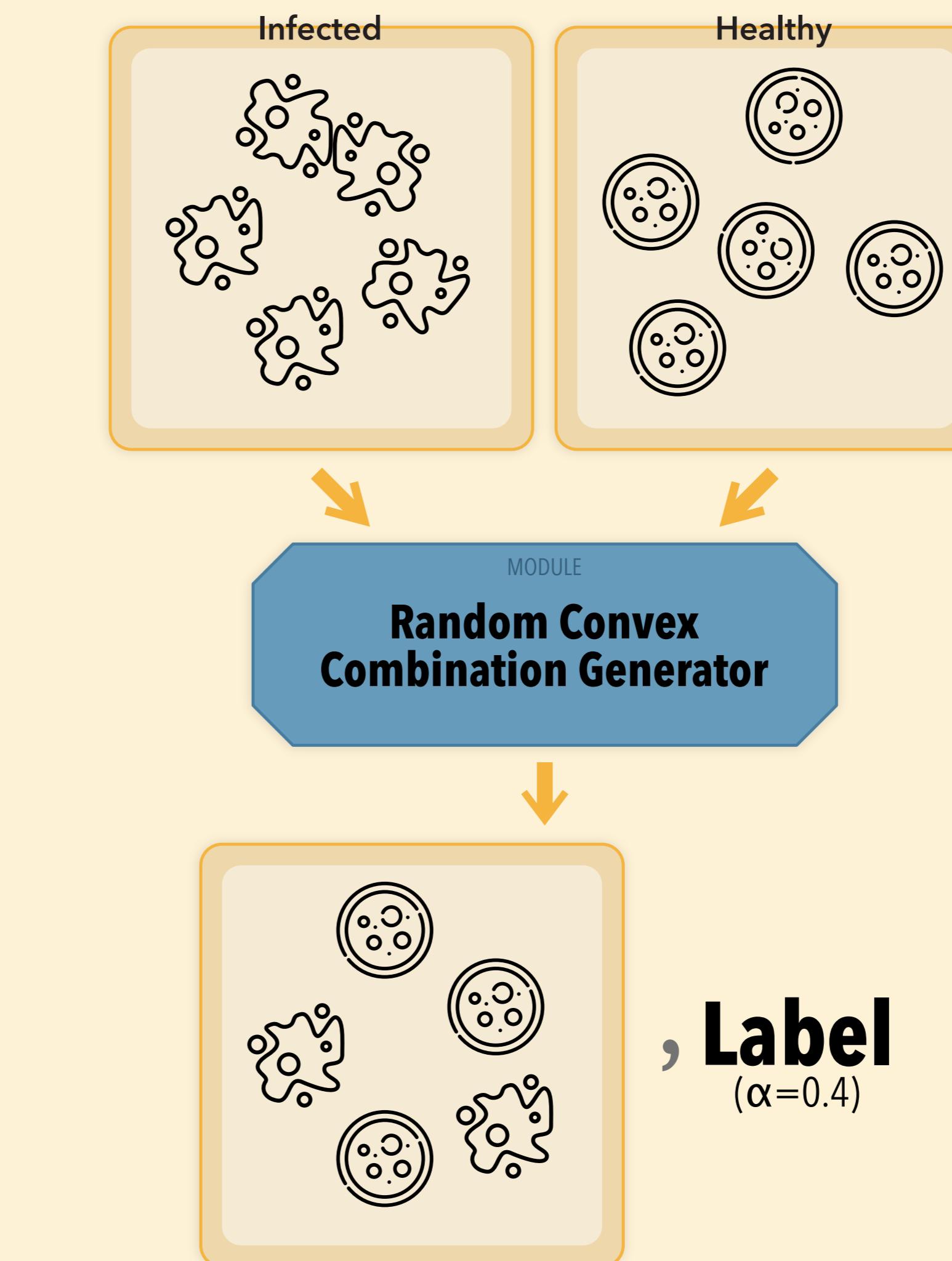


Figure 3

In the training and validation phases, our assays were augmented with new random convex combinations of infected and healthy assays that represented various levels of treatment efficacy, and the model was trained to predict the weights used in the linear combination of the combined features. Therefore, the output of such a model represented a measure of the illness of cells.

RESULTS

RxRx19a dataset publication was followed by an article that suggested a method to calculate the CPE of treatments. In our experiments, we referred to the CPE estimation algorithm of this study as *RxRx19a* on-disease score algorithm.

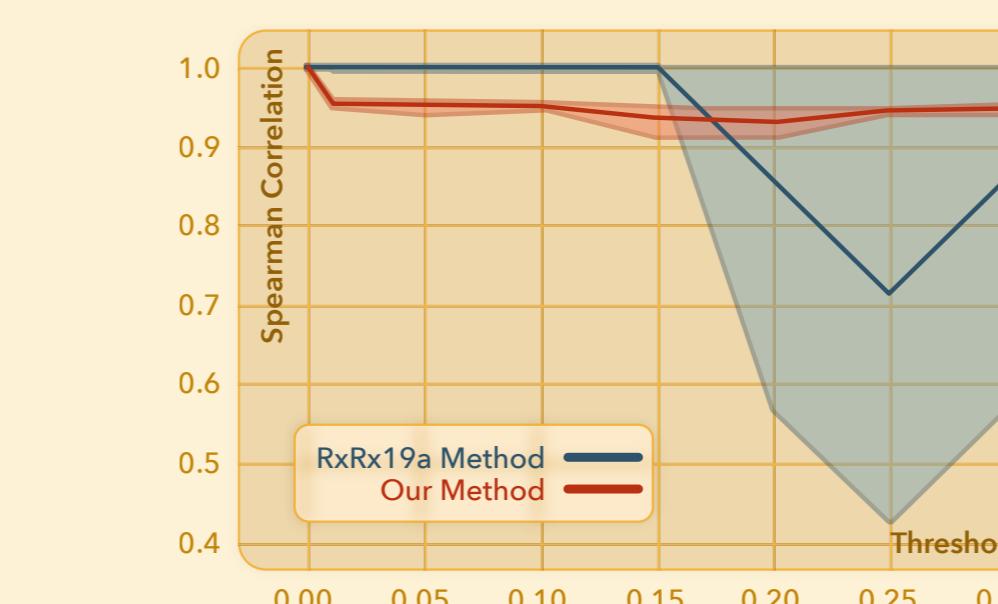


Figure 4

V In a round of five-time, a portion of training dataset was randomly omitted before a model trained on it and predicted new disease scores for treatments. The Spearman correlation of the original scores and new scores was calculated. Our model disease scores stability was enhanced due to the data augmentation process, while *RxRx19a*'s on-disease score algorithm was unstable when more than 15% of samples were missing.

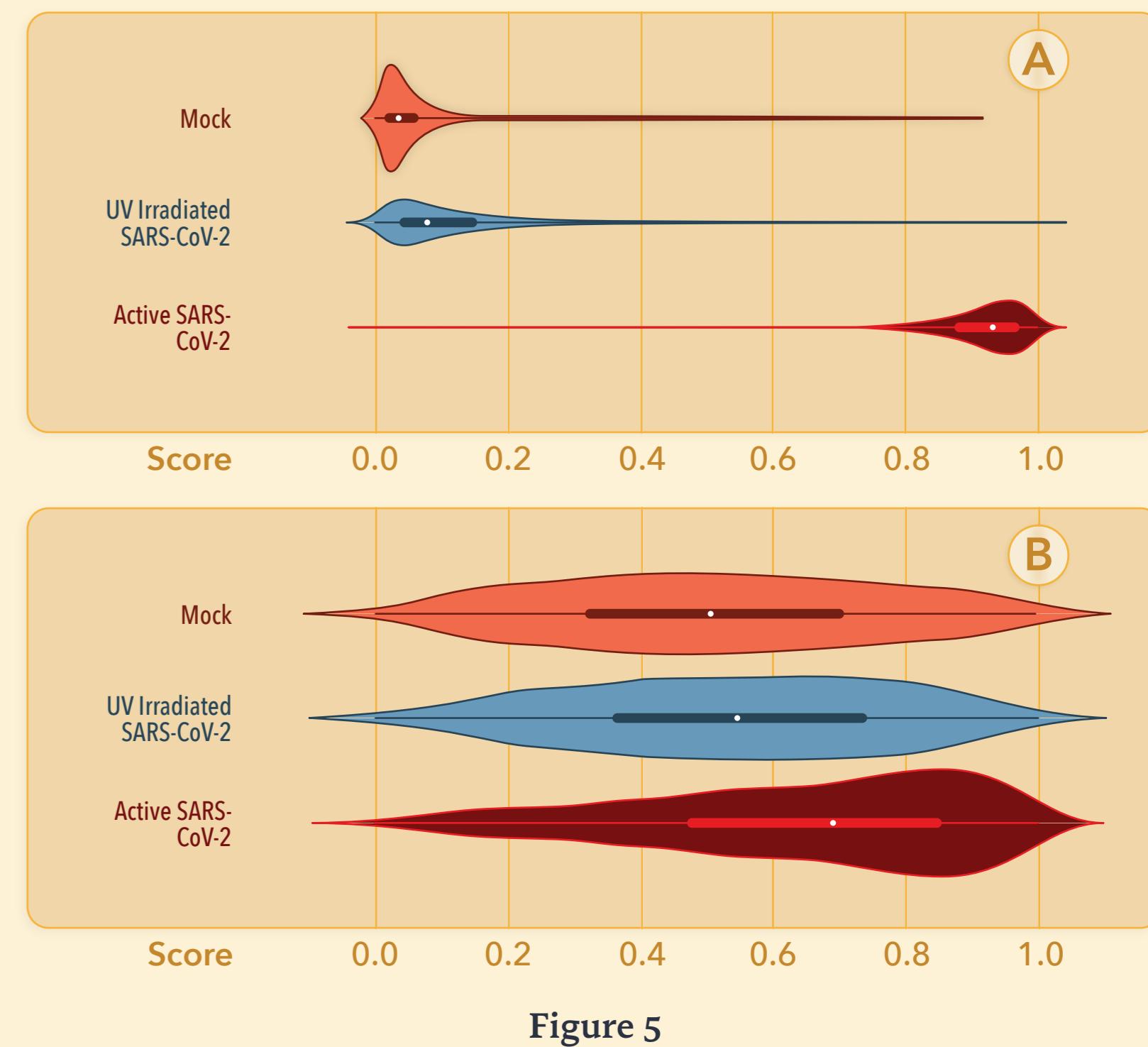


Figure 5

IV Our model is robust and learns meaningful disease-related features from samples. We trained our model on healthy mock and infected active SARS-CoV-2 samples and left out all UV-irradiated SARS-CoV-2 samples for testing. **A** Our model correctly predicted UV-irradiated samples disease scores near 0, which means healthy. **B** The *RxRx19a* algorithm, on-disease scores failed to make distinction between healthy and infected samples.

Treatment	Hit score (Our)	Hit score (RxRx19a)	Clinical Trial
Dipyridamole	0.96	0.17	NCT04410328
Eltrombopag	0.78	0.12	NCT04516837
Remdesivir	0.77	0.97	NCT05222113
Nitazoxanide	0.76	0.45	NCT04486313
FG-4592	0.75	0.44	-
GS-441524	0.72	0.99	NCT04859244
Hexachlorophene	0.70	0.52	-
Nilotinib	0.69	0.48	-
Lasalocid	0.68	0.13	-
Clofazimine	0.64	0.59	NCT04465695
Atovaquone	0.63	0.10	NCT04339426
Aloxistatin	0.59	0.79	-
Itraconazole	0.57	0.05	NCT04962022
Sunitinib	0.57	0.10	-
Buparvaquone	0.57	0.02	-

Table 1

VII Top 20 effective drugs based on our study. The higher (darker) hit scores, the better resistance against the virus. Clinical trials were selected amongst NIH clinical trial database (<https://clinicaltrials.gov>).

By perusing treatment disease scores, we statistically proved that **MEK inhibitors** and **adrenergic receptor agonists** are *unlikely* to be good prescriptions for the early stage of COVID-19 as they worsen the disease phenotype.

DISCUSSION

We developed a powerful pipeline to estimate a drug effect against a pathogen in HTS dataset. Further in-vitro and in-vivo follow-up experiments are needed to validate drug efficiency and toxicity for hits in this study.

CITATION

- Heiser et al., 2020, identification of potential treatments for covid-19 through artificial intelligence-enabled phenomic analysis of human cells infected with sars-cov-2, *bioRxiv*
 Caicedo et al., 2018, Weakly supervised learning of single-cell feature embeddings, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, doi:10.1109/CVPR.2018.00970
 Saberian et al., 2021, DEEMD: Drug Efficacy Estimation against SARS-CoV-2 based on cell Morphology with Deep multiple instance learning, 10.36227/techrxiv.1932665.v1.
 Mascolini et al., 2021, Exploiting generative self-supervised learning for the assessment of biological images with lack of annotations: A covid-19 case-study, *arXiv*