Zero to potential COVID-19 treatments in <4 weeks

with deep-learning-driven drug screens

Imran S. Haque, PhD
VP Data Science, Recursion
@ImranSHaque

Recursion

Apr 12-16 2021
GTC 2021

Disclaimers

Speaker is an employee of Recursion, and the views and opinions in this presentation are his own, and not necessarily those of Recursion Pharmaceuticals; their accuracy and reliability are not guaranteed. The material presented is general background information in summary form and may not be current or complete. This presentation is intended for educational purposes only.

About Recursion

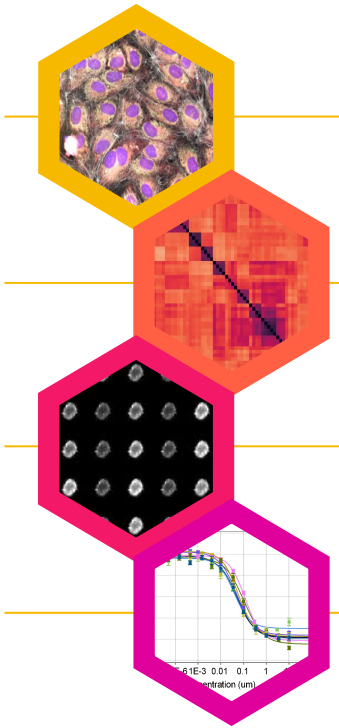
We are a clinical-stage biotechnology company decoding biology by integrating technological innovations across biology, chemistry, automation, data science and engineering in our mission to radically improve the lives of patients and industrialize drug discovery.

Target-agnostic hit ID in human disease model

Target insights

Confirm activity in orthogonal functional assay

Optimize leads



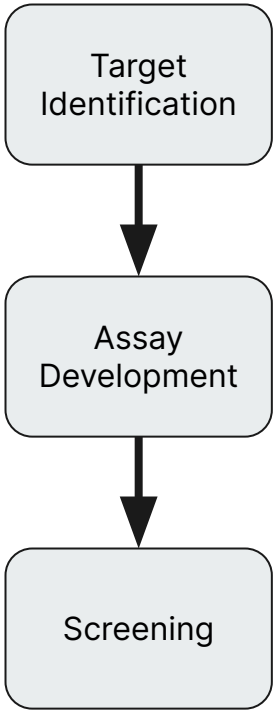
Distal, high-dimensional datasets “fit for AI”

Relational database of biological stimuli and chemistry

Human-derived and clinically-relevant assays

NextGen predictive ADMET and comp. chemistry tools

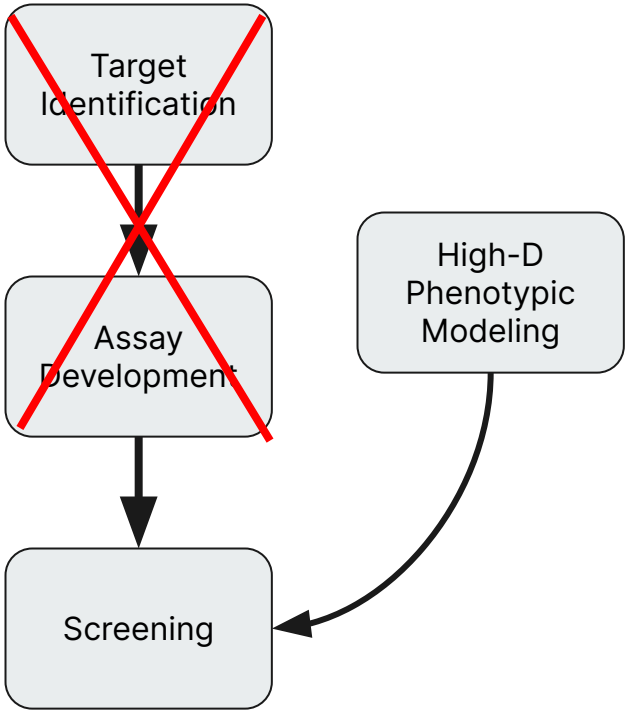
Accelerating the Front-End of Drug Discovery



Traditional drug discovery approaches start with two steps that are not only slow, but can limit how well screening results translate into real human trials.

If these steps could be replaced by a *uniform* design *spanning biology*, initiating programs against rare or emergent diseases could be made much faster and cheaper.

Accelerating the Front-End of Drug Discovery



Traditional drug discovery approaches start with two steps that are not only slow, but can limit how well screening results translate into real human trials.

If these steps could be replaced by a *uniform* design *spanning biology*, initiating programs against rare or emergent diseases could be made much faster and cheaper.

Let's talk about how to do that.

Outline

I. How deep learning enables a uniform phenomics assay to map biology

II. Pivoting phenomics to rapidly search for treatments against COVID-19

Data and results in our preprint and at **rxrx.ai**:

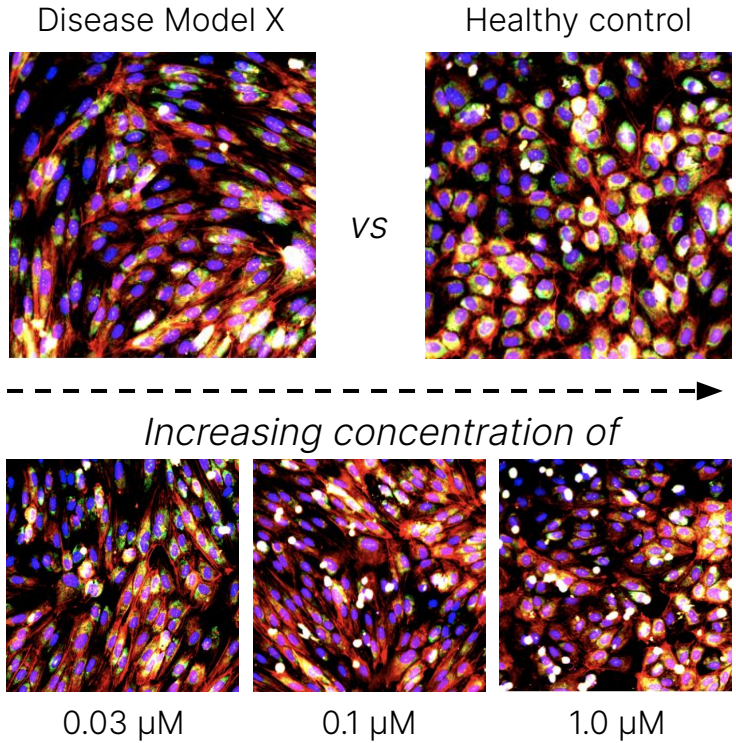
Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery

Michael F. Cuccarese, Berton A. Earnshaw, Katie Heiser, Ben Fogelson, Chadwick T. Davis, Peter F. McLean, Hannah B. Gordon, Kathleen-Rose Skelly, Fiona L. Weathersby, Vlad Rodic, Ian K. Quigley, Elissa D. Pastuzyn, Brandon M. Mendivil, Nathan H. Lazar, Carl A. Brooks, Joseph Carpenter, Brandon L. Probst, Pamela Jacobson, Seth W. Glazier, Jes Ford, James D. Jensen, Nicholas D. Campbell, Michael A. Statnick, Adeline S. Low, Kirk R. Thomas, Anne E. Carpenter, Sharath S. Hegde, Ronald W. Alfa, Mason L. Victors, Imran S. Haque, Yolanda T. Chong, Christopher C. Gibson

doi: https://doi.org/10.1101/2020.08.02.233064

I. Building phenomics: deep learning on multichannel microscopy

What is high-dimensional phenotypic modeling?

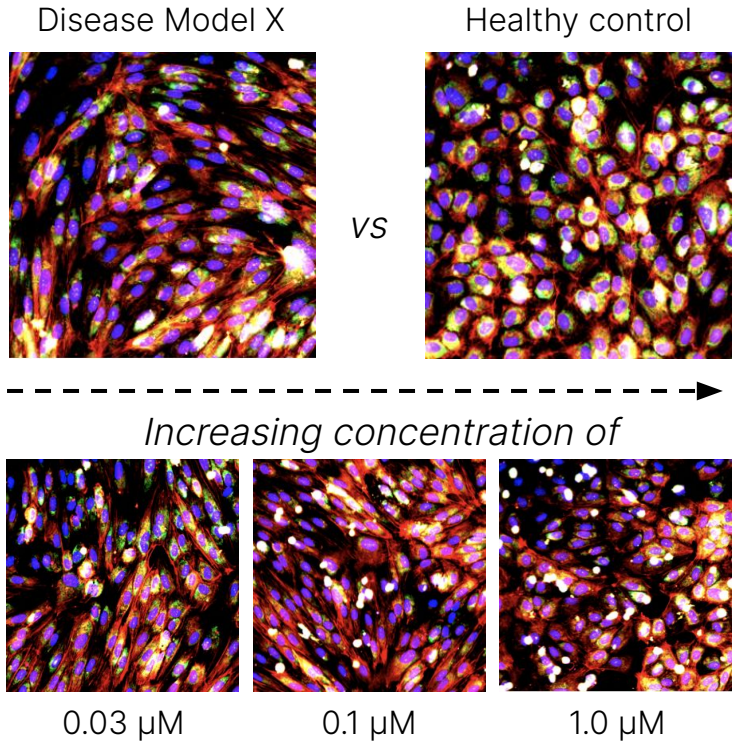


Rather than picking individual *targets*:

1. Build *causal models* of disease in human cells (genetic knockouts, soluble factors, infectious agents).
2. Find treatments that make the cells look more like “healthy”

*Note: images shown above depict a disease model with visible phenotype for illustrative purposes only; primary utility of Recursion platform is to readily distinguish non-visible phenotypes

What is high-dimensional phenotypic modeling?



Rather than picking individual *targets*:

1. Build *causal models* of disease in human cells (genetic knockouts, soluble factors, infectious agents).
2. Find treatments that make the cells look more like “healthy”

...what does “look healthy” mean?

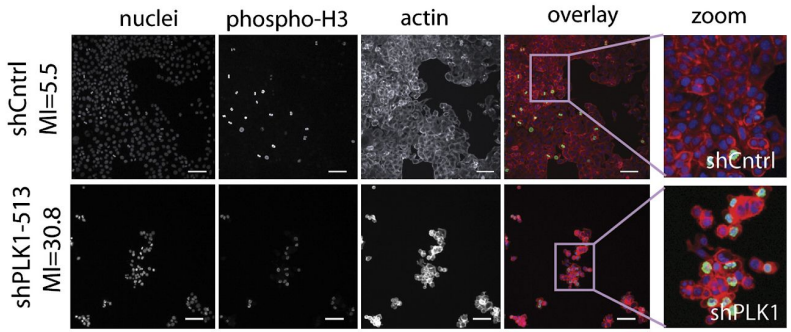
*Note: images shown above depict a disease model with visible phenotype for illustrative purposes only; primary utility of Recursion platform is to readily distinguish non-visible phenotypes

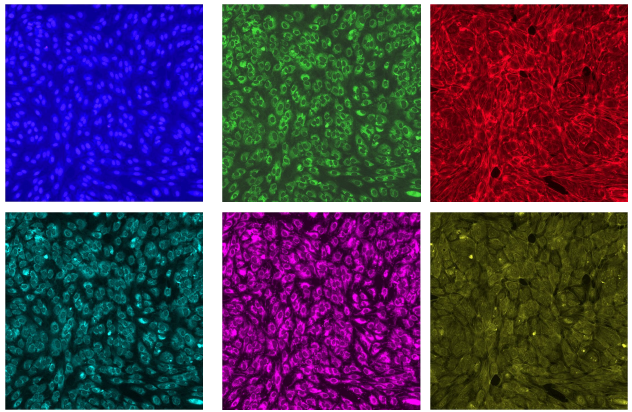
The data: 6-channel fluorescent microscopy

Microscopy is an incredibly **data-rich** technique…if you can see the right things: flexible, spatial, intrinsically single-cell.

Traditional high-content imaging: uses specific stains to highlight the specific pathway we're interested in interrogating. High-content, but also custom for every experiment!

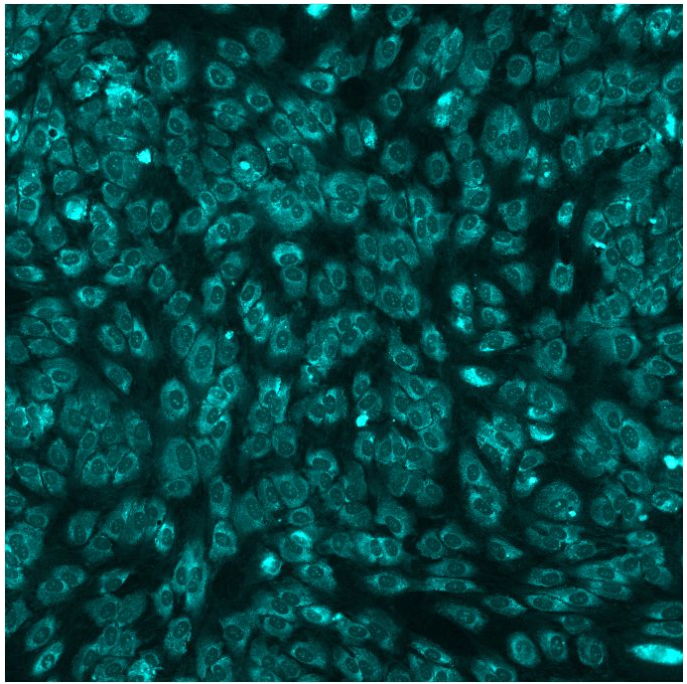
Morphological profiling: use a common set of stains across experiments. Standardizes experiment, but information is less directly encoded.

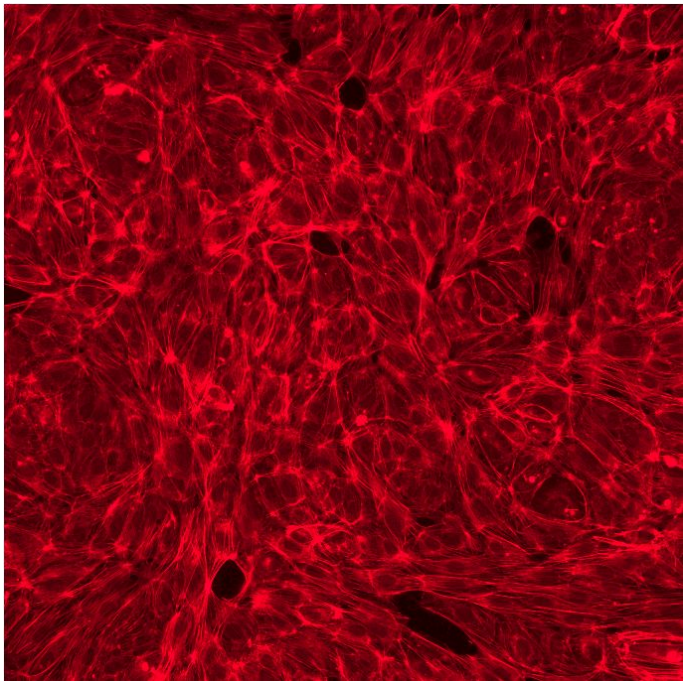




Moffat J et al. Cell 2006
Bray M-A et al. Nat Protocols 2016

The data: 6-channel fluorescent microscopy





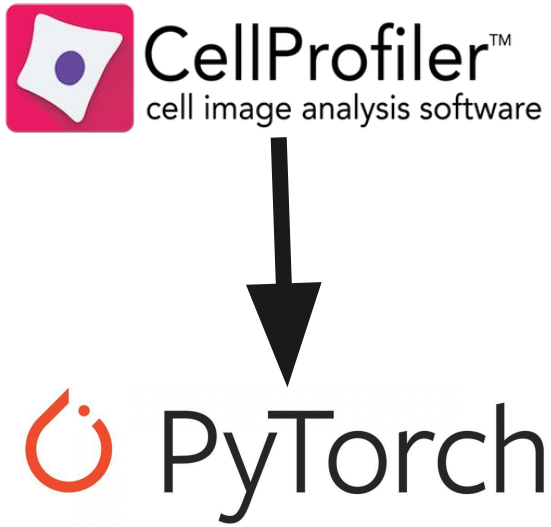
Each channel visualizes a different structure inside the cell. Great target for deep learning, but non-trivial batch-to-batch variability even between replicate experiments.

Phenomics: AI-standardized Content Extraction

Traditional approaches to morphological profiling use hand-tuned features in packages like **CellProfiler**.

Recursion has collected over 7PB of data. Deep learning tends to perform well in such data-rich domains, but how do you:

- Bootstrap models far from natural images?
- Verify that they extract the right biological content?

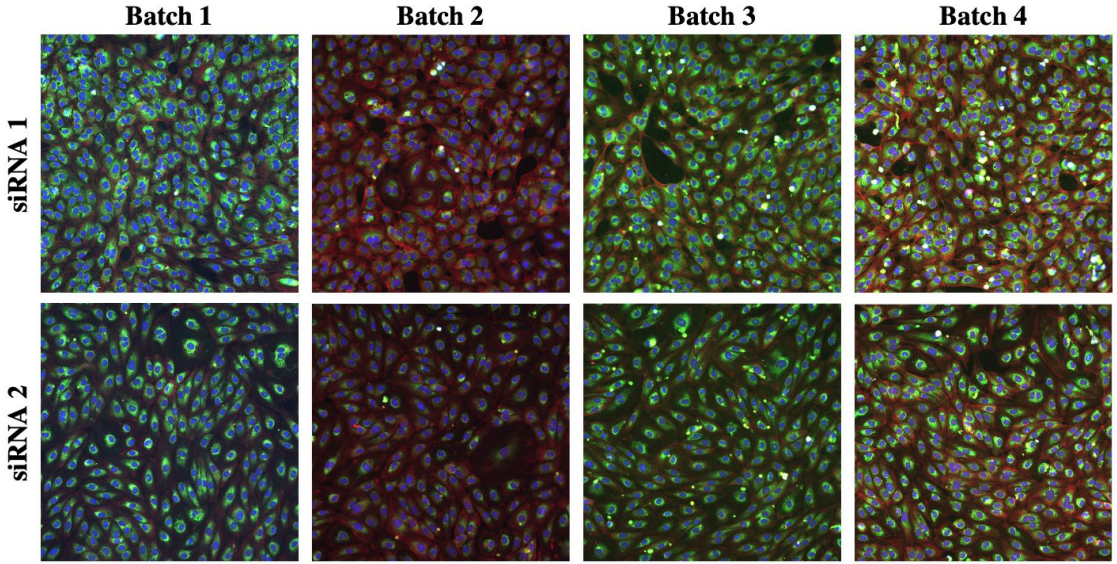


RxRx1: The ImageNet of Cells

In July 2019, Recursion released RxRx1, a large dataset of cell images tailored to answer these questions:

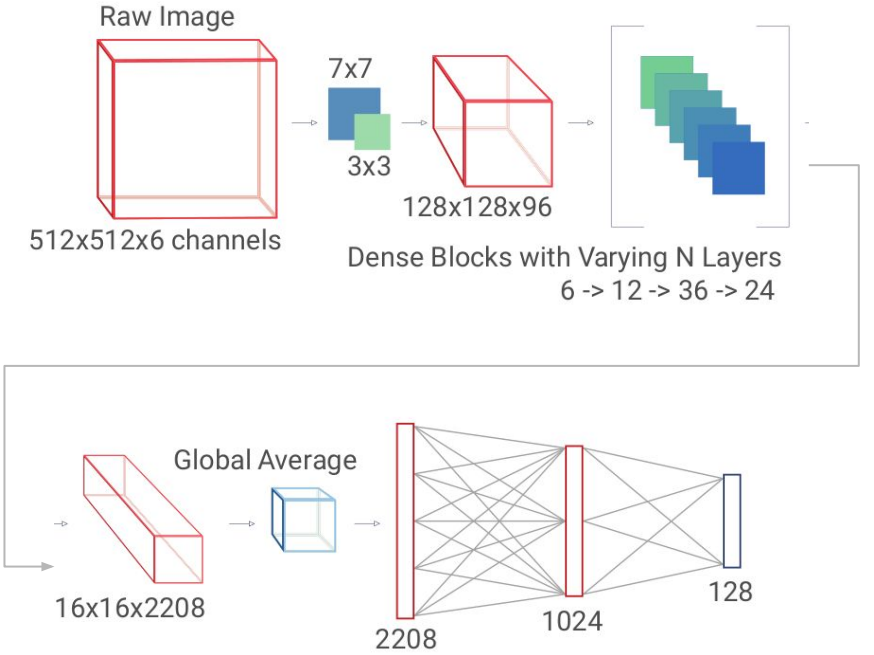
- >1100 siRNAs in 4 cell types
- 51 batches
- 125k images
- ~46GB

RxRx1 formed the backbone of the 2019 NeurIPS **CellSignal** competition, in which top competitors achieved >99% content accuracy in identifying perturbations from images alone.



Accelerating deep learning workflows with GPUs

Even relatively small deep convolutional models on this quantity of multi-megapixel multichannel data require GPU acceleration to be tractable to train, and benefit from large amounts of memory to allow large batches and advanced normalization schemes.



RxRx1 Performance from Volta to Ampere

Iterating on model designs and hyperparameters quickly is critical as we map biology.

On an internal model training benchmark, moving from V100 hardware to the A100 **reduced training time 46%**, nearly doubling throughput.

Recursion's new **BioHive-1** supercomputer is built using NVIDIA's DGX A100 SuperPod architecture.

II. Rapid Repurposing for COVID-19

The dual nature of COVID-19

You've probably heard of COVID-19. It's caused by a virus **SARS-CoV-2** and exploded exponentially a little over a year ago.

Rapidly emerging diseases are a great opportunity for a therapeutic platform that can **rapidly** reorient itself.

But COVID isn't just one disease.

The dual nature of COVID-19

You've probably heard of COVID-19. It's caused by a virus **SARS-CoV-2** and exploded exponentially a little over a year ago.

Rapidly emerging diseases are a great opportunity for a therapeutic platform that can **rapidly** reorient itself.

But COVID isn't just one disease.

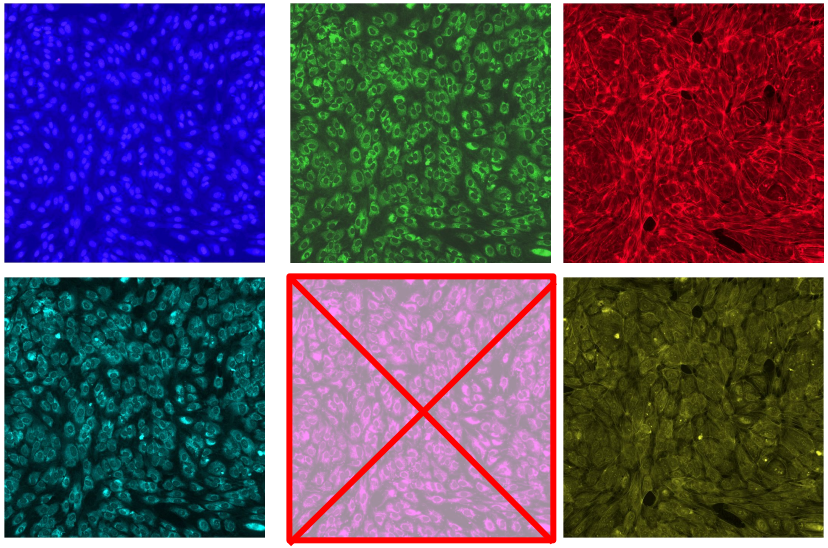
Early stage/mild disease:
viral infection by SARS-CoV-2

Late stage/severe disease:
Autoimmune over-reaction by the body's response to the virus

At Recursion, we modeled both: using active virus to model the early stage, and a patient-informed cocktail of "cytokines" (immune response proteins) for the late stage.

GPUs vs deadly respiratory viruses

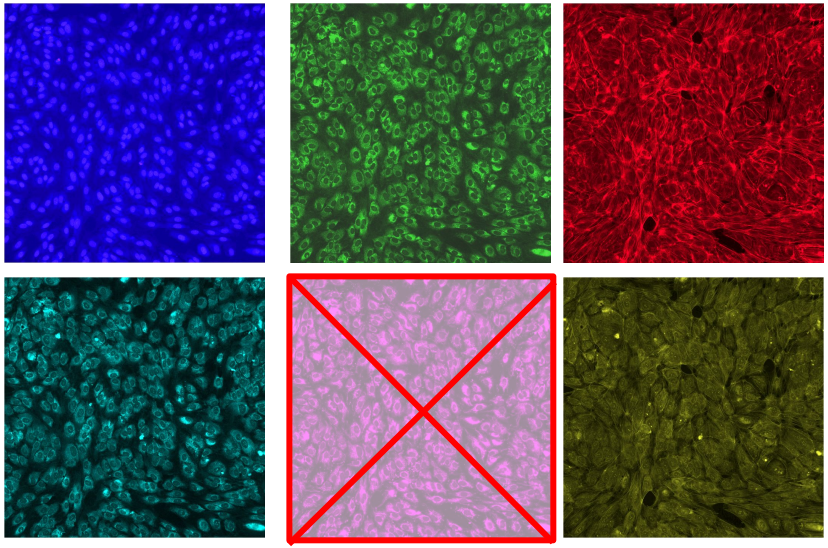
Using active SARS-CoV-2 provided a more faithful model of COVID, but forced our scientists to work in Biosafety Level 3 conditions, which prevented the use of one of our cell stains. Now what?



GPUs vs deadly respiratory viruses

Using active SARS-CoV-2 provided a more faithful model of COVID, but forced our scientists to work in Biosafety Level 3 conditions, which prevented the use of one of our cell stains. Now what?

NVIDIA GPU acceleration enabled us to retrain image embedding models without this channel in only 2 days.



Project Goals

Find

broadly-applicable therapeutics for

rapid translation using

primary human cell models of

COVID-19 examining

more than viral titer and cell count.

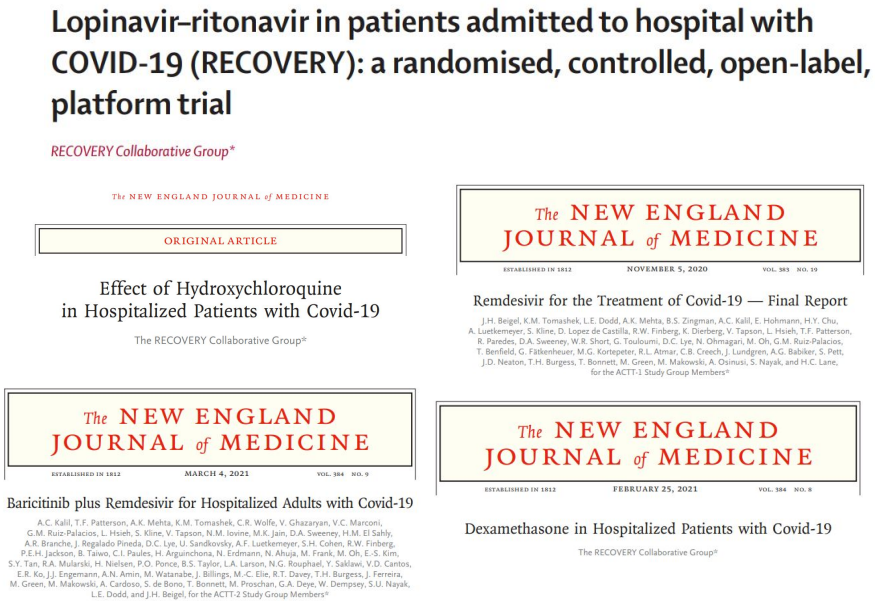
Therapeutics that are not orally available or have manufacturing difficulties will be difficult to apply broadly in a global pandemic. (This rules out biologics.)

Our objective was rapid repurposing: to expedite efficacy trials, what could we find that already has significant human safety data?

How did we do?

Drug	Prediction	Correct?
Hydroxychloroquine	✘	✔
Lopinavir	✘	✔
Ritonavir	✘	✔
Remdesivir	✔	✔
Baricitinib	✔	✔
Dexamethasone	✘	✘

Recursion platform predicted 5 of 6 subsequent randomized clinical trials correctly, in both early and late COVID, with more trials ongoing.


Lopinavir–ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial

RECOVERY Collaborative Group*

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19

The RECOVERY Collaborative Group*

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 NOVEMBER 5, 2020 VOL. 383 NO. 19

Remdesivir for the Treatment of Covid-19 — Final Report

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MARCH 4, 2021 VOL. 384 NO. 9

Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 FEBRUARY 25, 2021 VOL. 384 NO. 8

Dexamethasone in Hospitalized Patients with Covid-19

The RECOVERY Collaborative Group*

L/R: RECOVERY, Lancet 2020, 396:1345-52.
HCQ: RECOVERY, N Engl J Med 2020; 383:2030-2040
RDV: ACTT-1, N Engl J Med 2020; 383:1813-1826
Bari: ACTT-2, N Engl J Med 2021; 384:795-807
Dexa: RECOVERY, N Engl J Med 2021; 384:693-704

RxRx19

We released the images, metadata, deep learning embeddings, and a results explorer for our COVID screens to the community (licensed CC-BY):

- **RxRx19a:** Viral infection screen
- **RxRx19b:** Cytokine storm screen

https://rxrx.ai/rxrx19

	RxRx1	RxRx2	RxRx19a	RxRx19b
Release Date	June 2019	August 2020	April 2020	August 2020
Cell Types	HUVEC RPE U2OS HepG2	HUVEC	HRCE Vero	HUVEC
Stains (Channels)	Hoechst ConA Phalloidin Syto14 MitoTracker WGA	Hoechst ConA Phalloidin Syto14 MitoTracker WGA	Hoechst ConA Phalloidin Syto14 WGA	Hoechst ConA Phalloidin Syto14 MitoTracker WGA
Plate Density	384-well	1536-well	1536-well	1536-well
Imaging Sites per Well	2	4	4	1
Perturbations Evaluated	1,138 siRNAs	434 soluble factors at 6 concentrations	1,672 small molecules at 6+ concentrations Three viral conditions (active virus, irradiated, mock)	1,856 small molecules at 4-6 concentrations in three COVID-19-associated cytokine storm conditions (severe storm, healthy, and no cytokines)
Total Number of Images	125,510	131,953	305,520	70,384

Conclusions

RxRx19

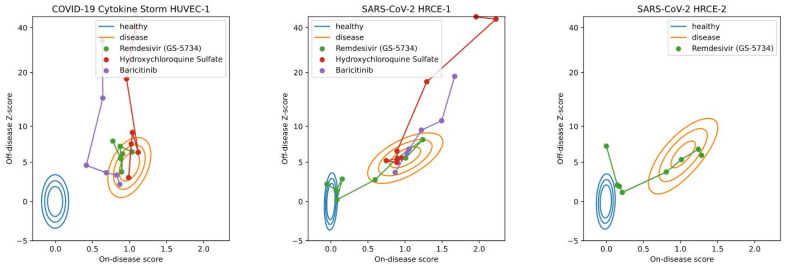
Recursion has released >300,000 morphological profiling images (~450GB), metadata, and deep learning embeddings of the images from this COVID-19 screen to the community (licensed CC-BY).

https://www.rxrx.ai/rxrx19

	RxRx1	RxRx19
Cell Types	HUVEC RPE U2OS HepG2	HRCE Vero
Stains (Channels)	Hoechst (Ch1) ConA (Ch2) Phalloidin (Ch3) Syto14 (Ch4) MitoTracker (Ch5) WGA (Ch6)	Hoechst (Ch1) ConA (Ch2) Phalloidin (Ch3) Syto14 (Ch4) WGA (Ch5)
Plate Density	384-well	1536-well
Imaging Sites per Well	2	4
Perturbations Evaluated	1,138 siRNAs	1,672 small molecules at 6+ concentrations Three viral conditions (active virus, irradiated, mock)
Total Number of Images	125,510	305,520
Image Dimension	512×512×6	1024×1024×5
Dataset Size	~46GB	~450GB

Conclusions

- Large relatable data sets and deep learning enable phenotypic screens “mapping biology”, rather than one-off bespoke assays.
- GPU acceleration not only enables giant scale models, it allows them to be retuned instantly for emergent threats: **<4 weeks from start to results.**
- Recursion’s screens successfully predicted 5 of 6 subsequent clinical trials for early and late COVID-19.
- 860+ GB of data from our screens are open to the community to accelerate the development of methods and pandemic treatments.



860 GB of datasets https://www.rxrx.ai/rxrx19

results browser https://covid19.rxrx.ai/


Recursion
Questions?
imran.haque@recursion.com
@ImranSHaque / twitter.com/ImranSHaque
rxrx.ai (RxRx1, RxRx19 datasets)