Applying AI to Accelerate Assay Development to Pandemic Speed

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

Recursion

23 Jun 2020
SLAS Transformed

Disclosures

Employee of Recursion, which has committed to free non-discriminatory licensing for any of its intellectual property around discoveries related to the treatment of COVID-19.

This presentation discusses (results of) screening for off-label use of therapeutics against SARS-CoV-2/COVID-19.

©2020 Recursion

About Recursion

We are building a vertically-integrated biotech, leveraging massive empirical datasets at each step to accelerate drug discovery

Target-agnostic hit ID in human disease model

Distal, high-dimensional datasets “fit for AI”

Target insights

Relational database of biological stimuli and chemistry

Confirm activity in orthogonal functional assay

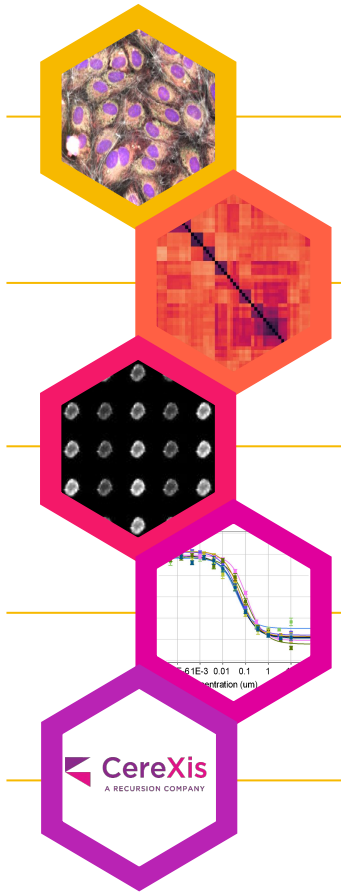
Human-derived and clinically-relevant assays

Optimize leads

NextGen predictive ADMET and comp. chemistry tools

Confirm activity in clinical setting

Innovative portfolio company structure



©2020 Recursion

Outline

The Goal: Use AI to enable a standard assay and standard analysis to generalize across disease contexts to rapidly discover new treatments…fast!

And apply this platform to discover repurposable treatments for COVID-19.

I. **AI: building phenomics**

II. **Building a high-dimensional SARS-CoV-2 assay**

III. **Results of drug screens against SARS-CoV-2**

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

Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2

Katie Heiser, Peter F. McLean, Chadwick T. Davis, Ben Fogelson, Hannah B. Gordon, Pamela Jacobson, Brett Hurst, Ben Miller, Ronald W. Alfa, Berton A. Earnshaw, Mason L. Victors, Yolanda T. Chong, Imran S. Haque, Adeline S. Low, Christopher C. Gibson

doi: https://doi.org/10.1101/2020.04.21.054387

©2020 Recursion

I. AI: building “phenomics”

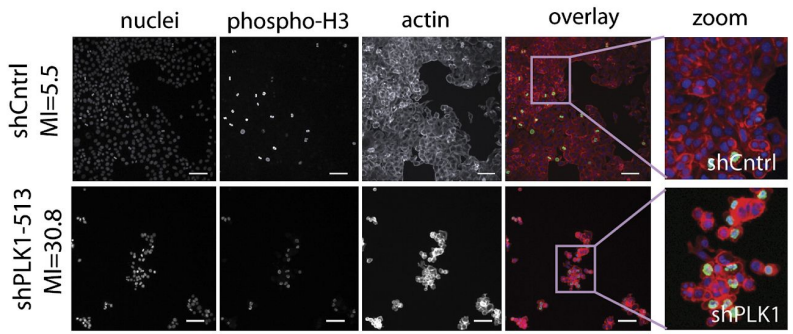
©2020 Recursion

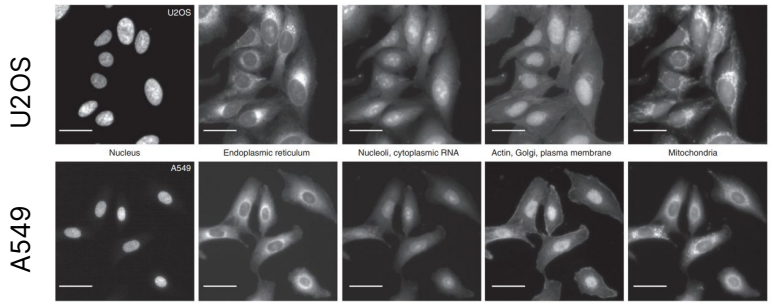
From high-content imaging to morphological profiling

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

High-content imaging: let's use 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. Standardize experiment (fast!), but do you lose information?





©2020 Recursion

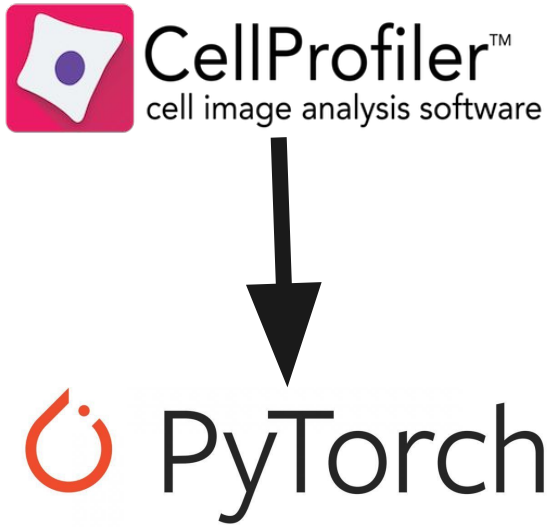
Moffat J et al. Cell 2006
Bray M-A et al. Nat Protocols 2016

Phenomics: AI-standardized Content Extraction

Fundamental challenge of morphological profiling: **is all of the information you want actually in the image? And can you extract it?**

Cell Painting as originally described relies on hand-engineered image analysis features as implemented in CellProfiler.

Deep learning has supplanted hand-engineered features in other domains of image analysis (“computer vision”): given a large enough dataset, can it do so here?



©2020 Recursion

RxRx1

Recursion released RxRx1 to the community, a large dataset of siRNA perturbations (125k images,>1100 siRNAs, 4 cell types), to answer two questions:

1. **How large a dataset is needed to train an effective deep network?**

2. **Is all the information there in Cell Painting? (How well is it possible to classify perturbations?)**

©2020 Recursion

RxRx1

Recursion released RxRx1 to the community, a large dataset of siRNA perturbations (125k images,>1100 siRNAs, 4 cell types), to answer two questions:

1. **How large a dataset is needed to train an effective deep network?**

2. **Is all the information there in Cell Painting? (How well is it possible to classify perturbations?)**

Top results in 2019 NeurIPS CellSignal competition were able to achieve >99% accuracy in classifying which siRNA was on a cell, based on this ~46GB dataset.

©2020 Recursion

II. Building a high-dimensional SARS-CoV-2 Assay

©2020 Recursion

Project Goals

Find

broadly-applicable therapeutics for

rapid translation using a

primary human cell model of

active SARS-CoV-2 infection examining

more than viral titer and cell count.

©2020 Recursion

Project Goals

Find

broadly-applicable therapeutics for

rapid translation using a

primary human cell model of

active SARS-CoV-2 infection 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.)

©2020 Recursion

Project Goals

Find

broadly-applicable therapeutics for

rapid translation using a

primary human cell model of

active SARS-CoV-2 infection examining

more than viral titer and cell count.

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

©2020 Recursion

Project Goals

Find

broadly-applicable therapeutics for

rapid translation using a

primary human cell model of

active SARS-CoV-2 infection examining

more than viral titer and cell count.

Immortal cell lines have different and unstable biology compared to normal cells; use primary cells to maximize chances of translatability.

©2020 Recursion

Project Goals

Find

broadly-applicable therapeutics for

rapid translation using a

primary human cell model of

active SARS-CoV-2 infection examining

more than viral titer and cell count.

1. *We are interested in phenotypic screening, not structure-based modeling of individual proteins.*

2. *Pseudovirus constructs etc. are logistically simpler - can be run in BSL2 or less - but are not fully faithful models of SARS-CoV-2 infection.*

©2020 Recursion

Project Goals

Find

broadly-applicable therapeutics for

rapid translation using a

primary human cell model of

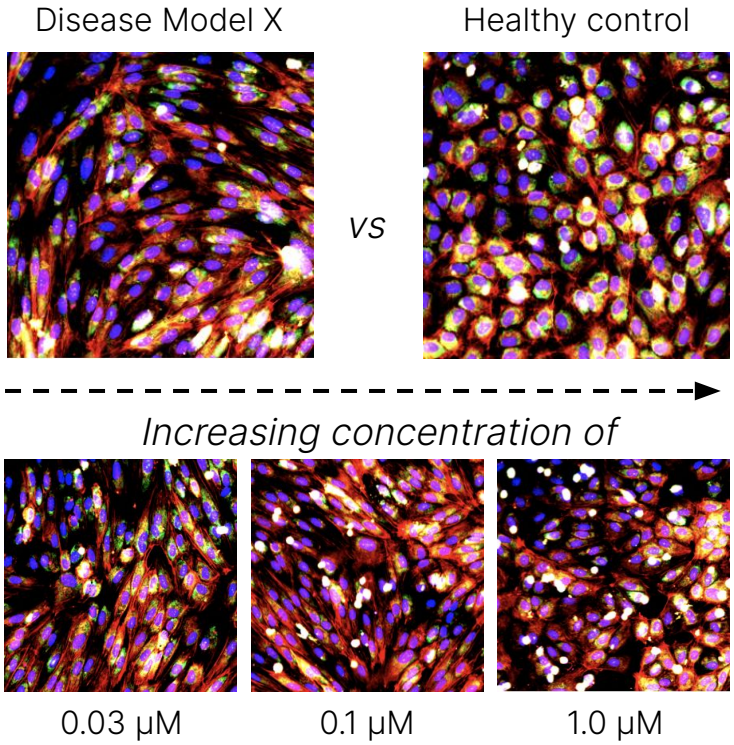
active SARS-CoV-2 infection examining

more than viral titer and cell count.

High-dimensional analysis of morphology may be able to reveal more subtle effects of viral infection and compound side-effect/toxicity.

©2020 Recursion

What is a high-dimensional assay?

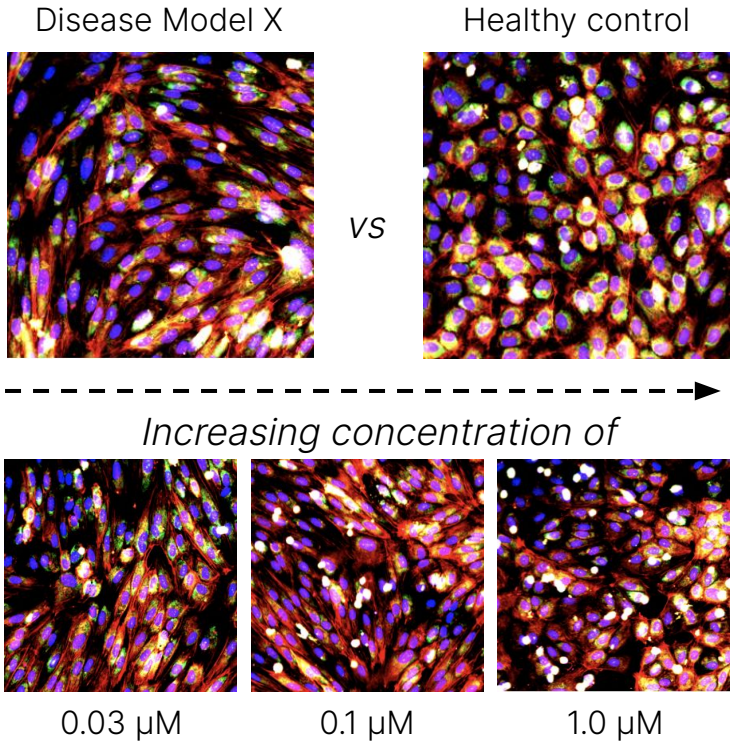


Measuring cell count is an inadequate proxy for biological effect and toxicity: these conditions have similar cell count, but *dramatically* different morphology.

©2020 Recursion

*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 a high-dimensional assay?



Measuring cell count is an inadequate proxy for biological effect and toxicity: these conditions have similar cell count, but *dramatically* different morphology.

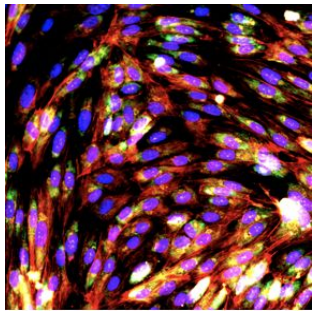
Deep learning model used in this work extracts 1,024 dimensions from each image. How can we visualize these results?

©2020 Recursion

*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

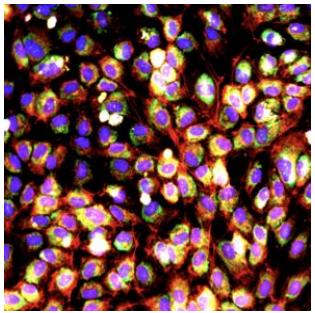
What is a high-dimensional assay?

Disease Model X

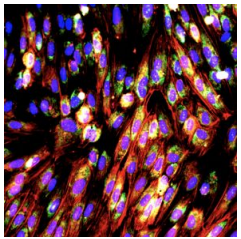


vs

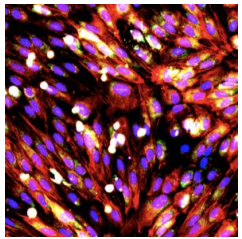
Healthy control



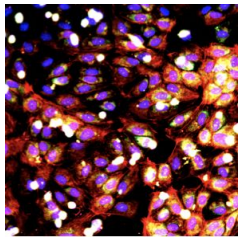
Increasing concentration of



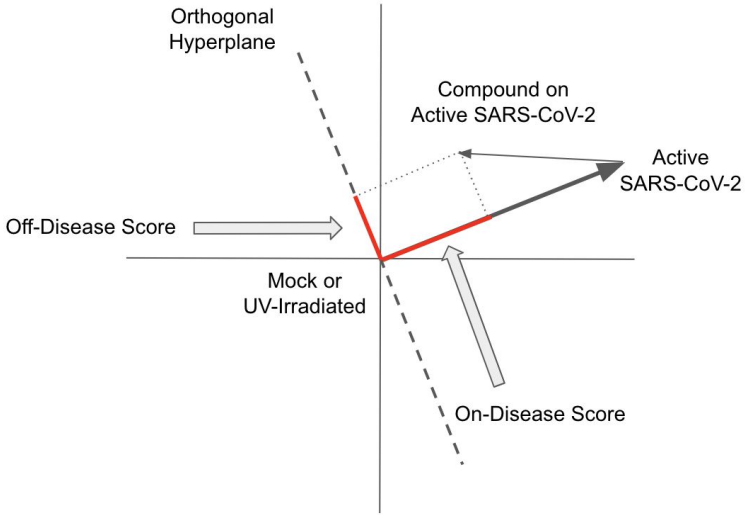
0.03 µM



0.1 µM



1.0 µM

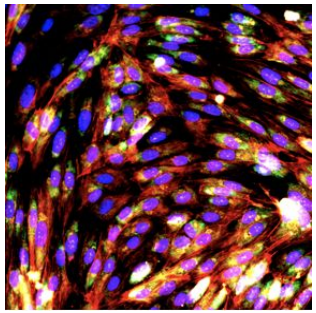


©2020 Recursion

*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

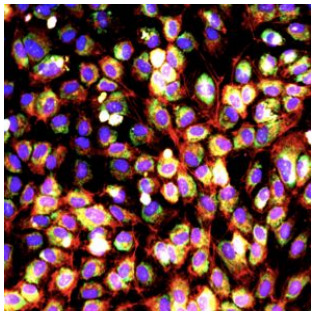
What is a high-dimensional assay?

Disease Model X

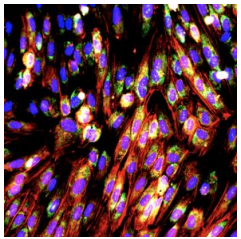


vs

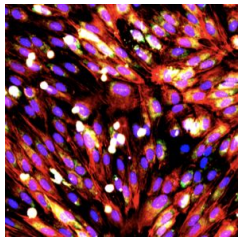
Healthy control



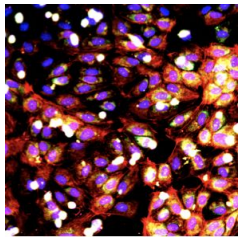
Increasing concentration of



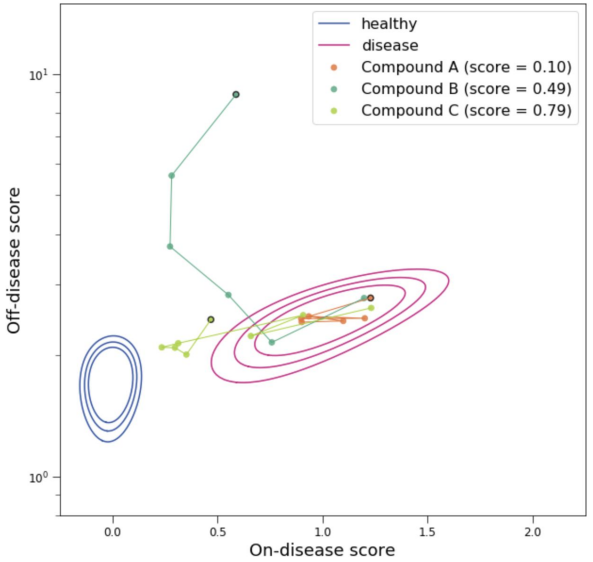
0.03 µM



0.1 µM



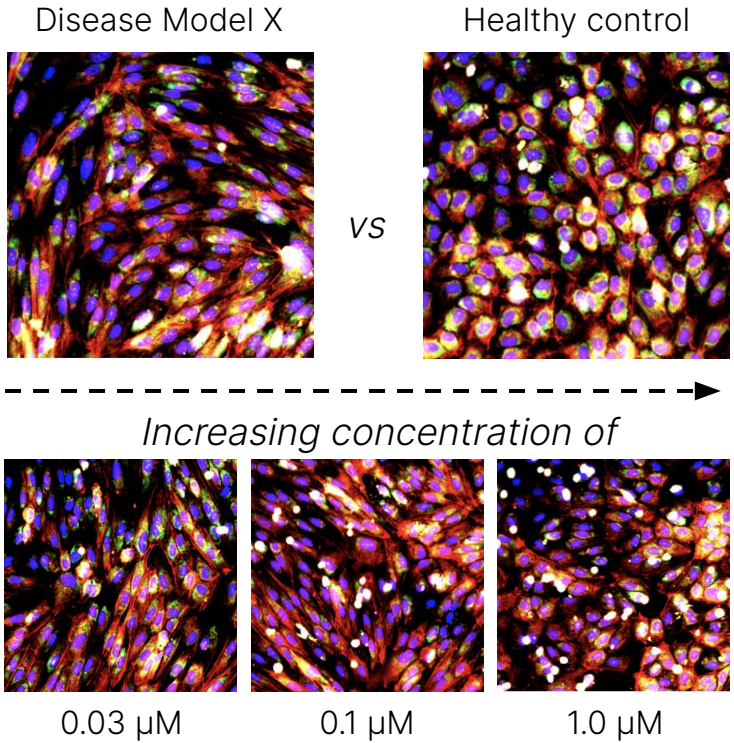
1.0 µM

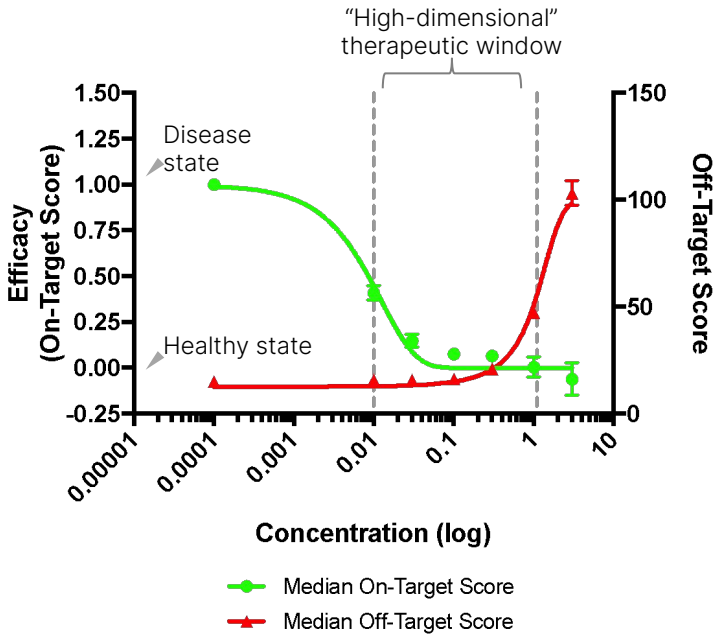


©2020 **Recursion**

*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 a high-dimensional assay?





©2020 Recursion

*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

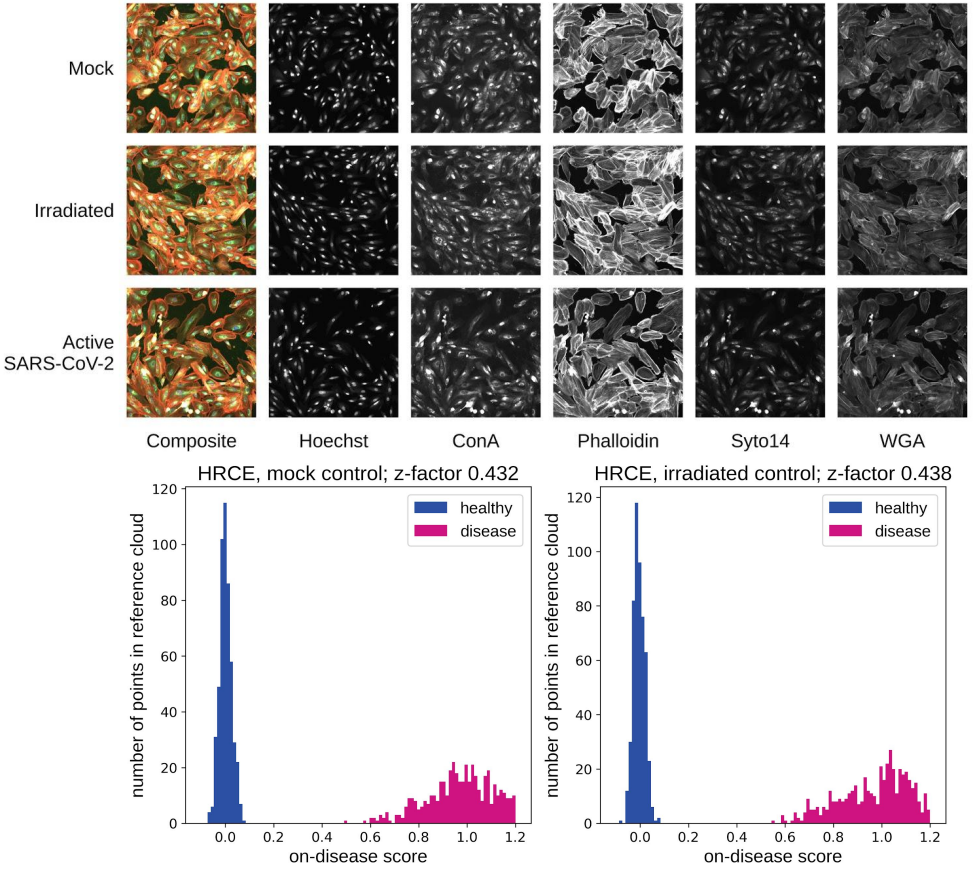
Establishing a phenotype (images)

Performed experiments in Vero (African green monkey) cells as a control, and in human cells:

- Bronchial epithelium
- Caco-2 (intestinal cell line)
- renal cortical epithelium (HRCE)

Mock and irradiated controls to capture effect of Vero supernatant.

HRCE had strongest phenotype of human primary cells and were progressed to drug screens.



©2020 Recursion

What did we screen?

Screened 1,670 approved and reference small-molecule compounds for activity in SARS-CoV-2 vs. mock assay in HRCE.

ChEMBL has imported results from our and other screens, and tables of structures and scores are available at www.ebi.ac.uk/chembl/document_report_card/CHEMBL4303122/

Molecule ChEMBL ID	Compound Key	Standard Type	Standard Relation	Standard Value	Standard Units	pChEMBL Value	Comment	Assay ChEMBL ID
CHEMBL2016757	GS-441524	Hit score	=	1.0	No Data	No Data	N=2 (0.999998436,0.999997106)	CHEMBL4303811
CHEMBL4065616	Remdesivir (GS-5734)	Hit score	=	0.9631	No Data	No Data	N=2 (0.97672611,0.949489959)	CHEMBL4303810

©2020 Recursion

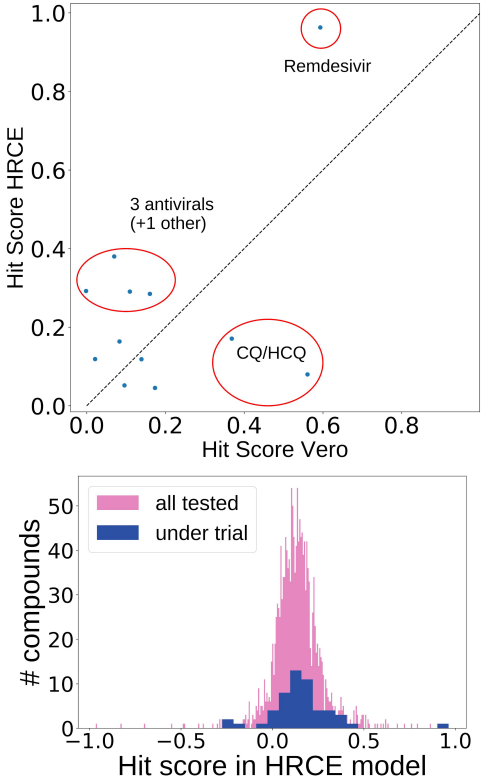
III. Screening Drugs for COVID-19

©2020 Recursion

Reference and Trial Compounds

Top: A limited set of compounds was run in both the Vero and HRCE models, with large species-specific differences in response.

Bottom: Hit scores for all compounds listed as undergoing current COVID-19 clinical trial at the time of publication (blue) versus all tested compounds (pink). With exception of remdesivir, no significant difference between trial compounds and all tested compounds.



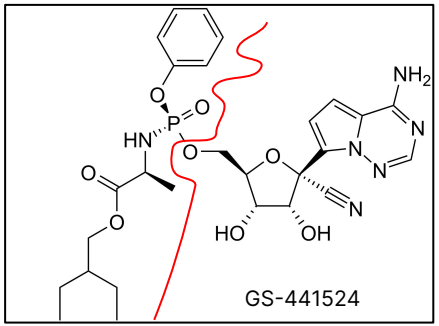
©2020 Recursion

https://www.biocentury.com/clinical-vaccines-and-therapies

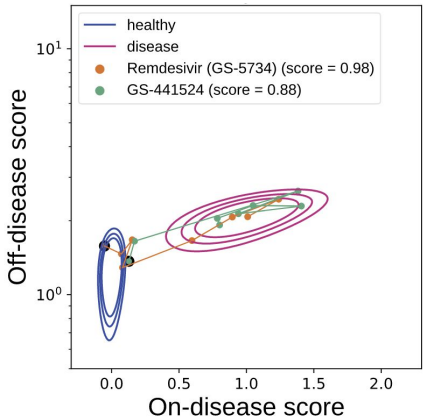
Remdesivir

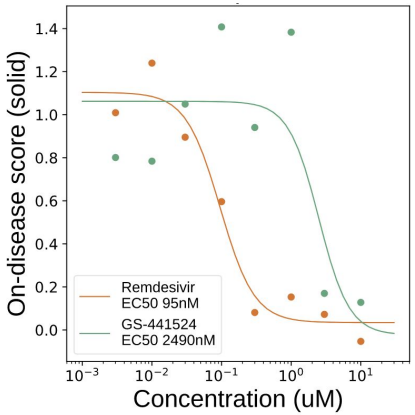
Remdesivir and its daughter nucleoside GS-441524 both specifically inhibit SARS-CoV-2 infection (low disease score, no increase in off-disease score).

As expected, remdesivir shows a signficant cellular assay potency boost over GS-441524, as a consequence of increased permeability from prodrug modifications.



Remdesivir



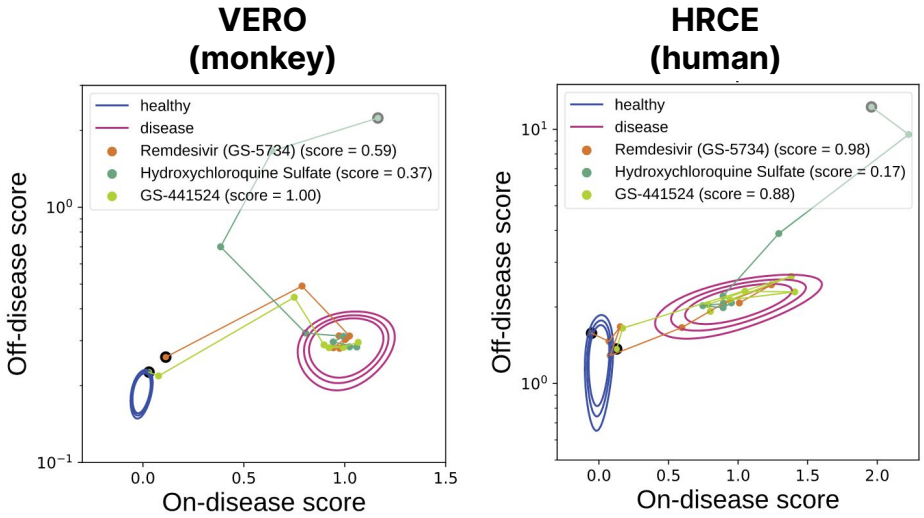


©2020 Recursion

Hydroxychloroquine

Hydroxychloroquine shows some efficacy at reducing viral load in Vero cells (replicating past in vitro work) -- but at a large cost in off-disease effect (beyond cell killing).

In human cells, no efficacy of hydroxychloroquine is detected.

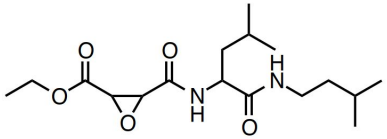


©2020 Recursion

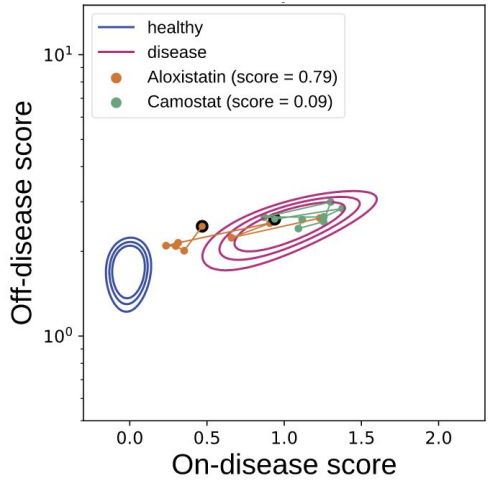
Aloxistatin (work in progress)

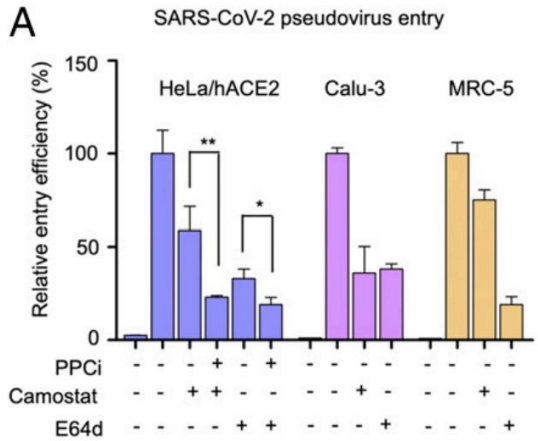
Top:
Aloxistatin (aka E64d) is a covalent cathepsin inhibitor originally in development for muscular dystrophy. It demonstrates consistent rescue of viral phenotype in Recursion HRCE platform screen with negligible off-disease effect.

Bottom:
Recursion HRCE model is likely *TMPRSS2*-negative (null effect from camostat). Data from Shang *et al* pseudovirus entry assay suggests strong activity of aloxistatin at preventing viral entry even in *TMPRSS2*+ cells.



Aloxistatin (E64d)





©2020 Recursion

Shang et al, *PNAS* 2020

Conclusions

©2020 Recursion

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

©2020 Recursion

Conclusions

- Application of morphological profiling + machine learning on large datasets enables rapid assay development without per-disease customization required: <4 weeks from zero to results (including logistics: time to contract BSL3 facility, etc.!).

- High dimensional analysis of cellular assays allows detection of potential toxic or off-target effects more subtle than what is possible with cell titer assays.

- Recursion's screen resulted in compounds with antiviral activity potentially repurposable for COVID-19.

- 450+ GB of data from this screen has been opened to the community to accelerate the development of treatments for the pandemic.

©2020 Recursion

Recursion

Questions?

imran.haque@recursionpharma.com

@ImranSHaque / twitter.com/ImranSHaque

rxrx.ai (RxRx1, RxRx19 datasets)