Recursion

Mapping and navigating biology and chemistry with genome-scale imaging

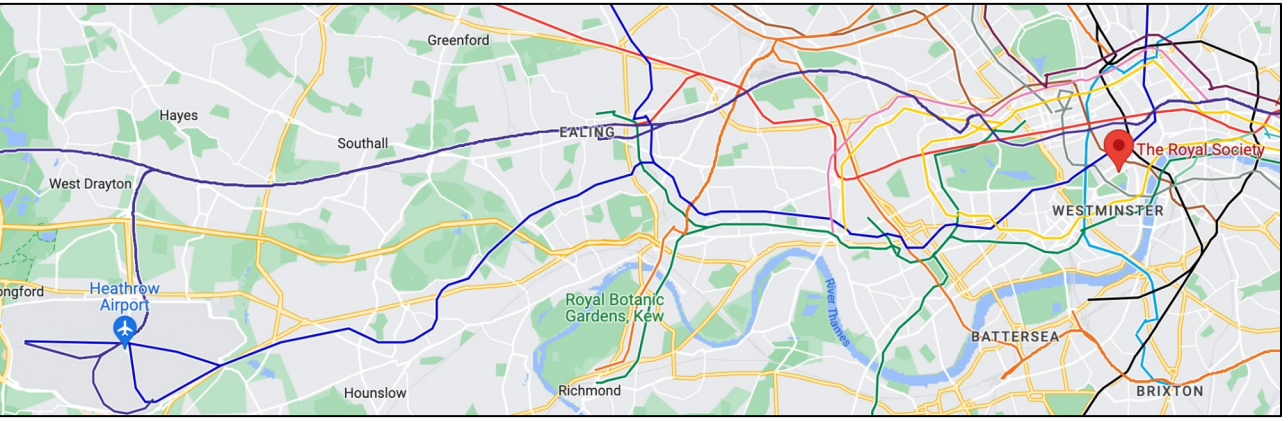
Imran S. Haque, PhD

Twitter: *@ImranSHaque*
Mastodon: *@ihaque@genomic.social*

The Royal Society: Machine learning and AI in biological science, drug discovery and medicine

2023-03-01

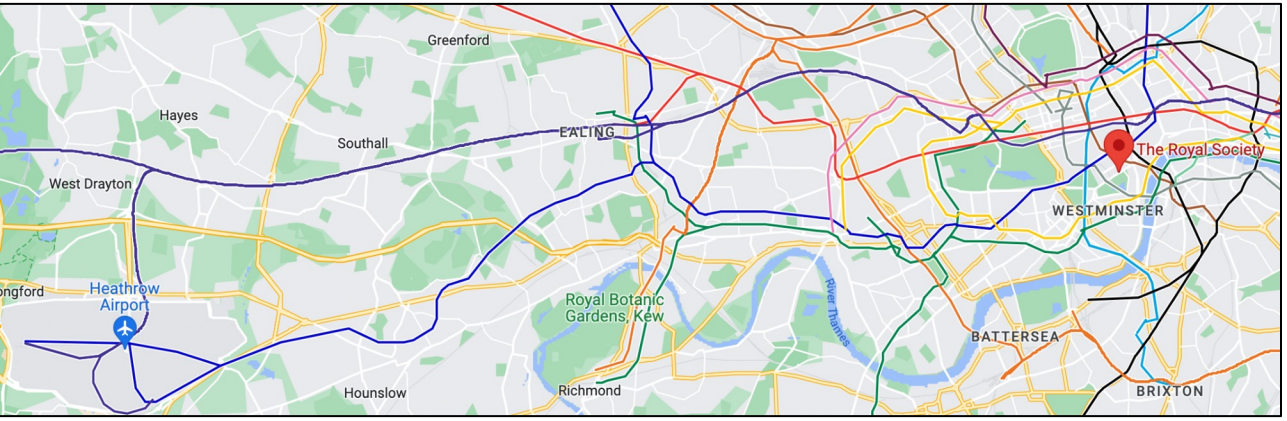
Imagine if one tool could...

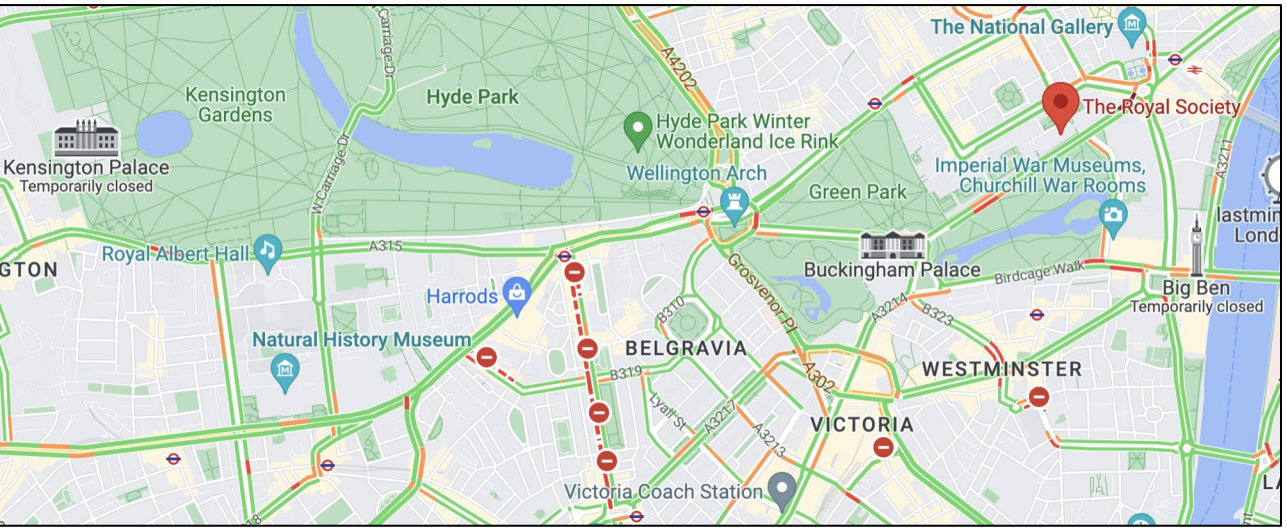


Imagine if one tool could...

Show the entire landscape of routes to the same destination.

Recursion





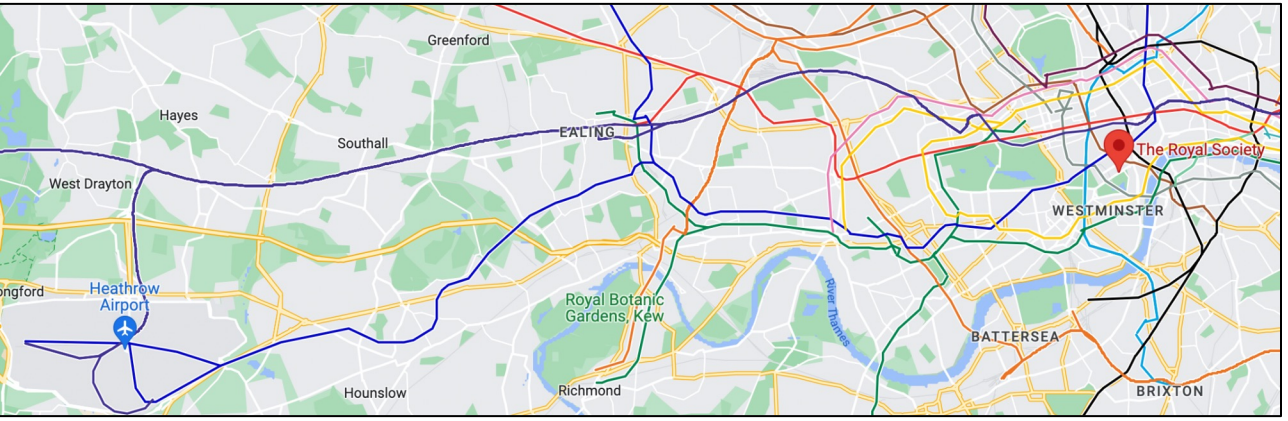
Imagine if one tool could...

Show the entire landscape of routes to the same destination.

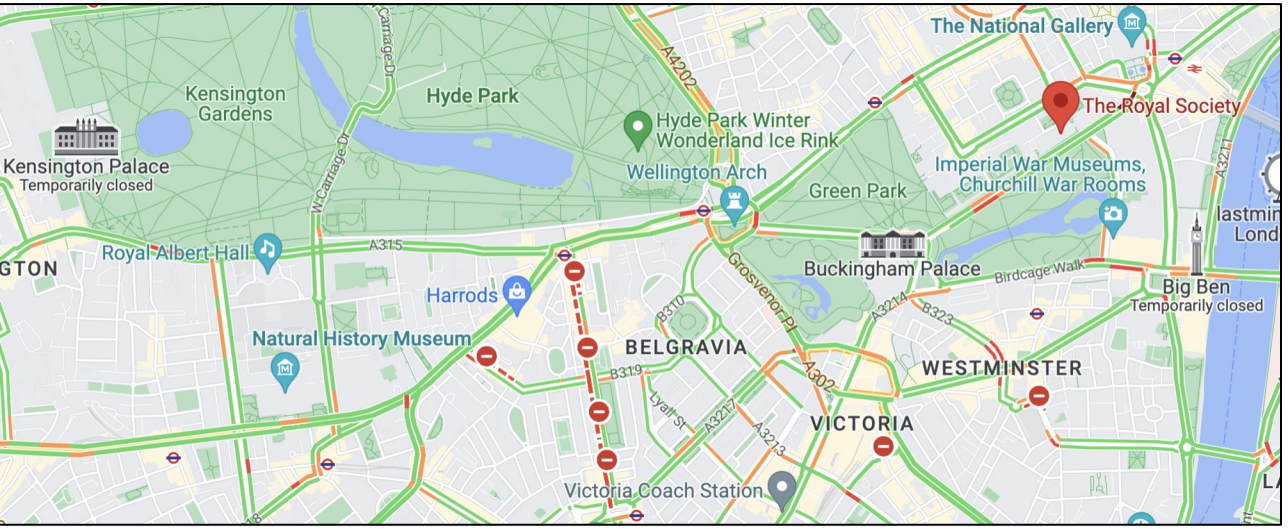
Show you ways to get around roadblocks.

Recursion

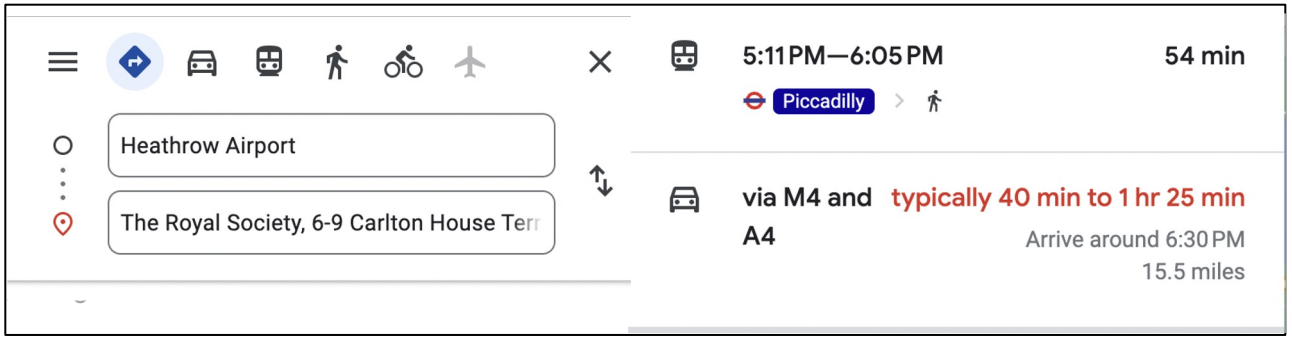
Imagine if one tool could...



Show the entire landscape of routes to the same destination.



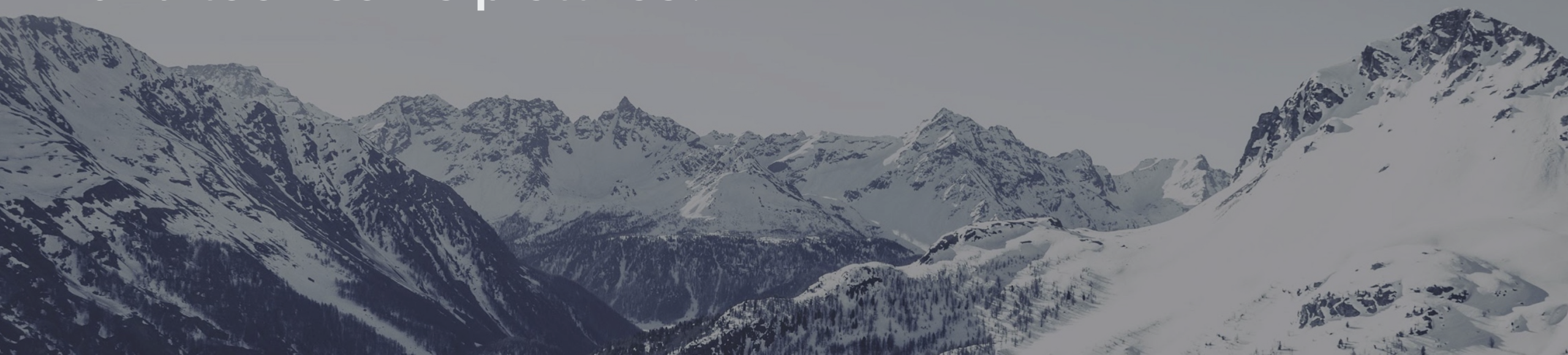
Show you ways to get around roadblocks.



Show the relationships among all those options.

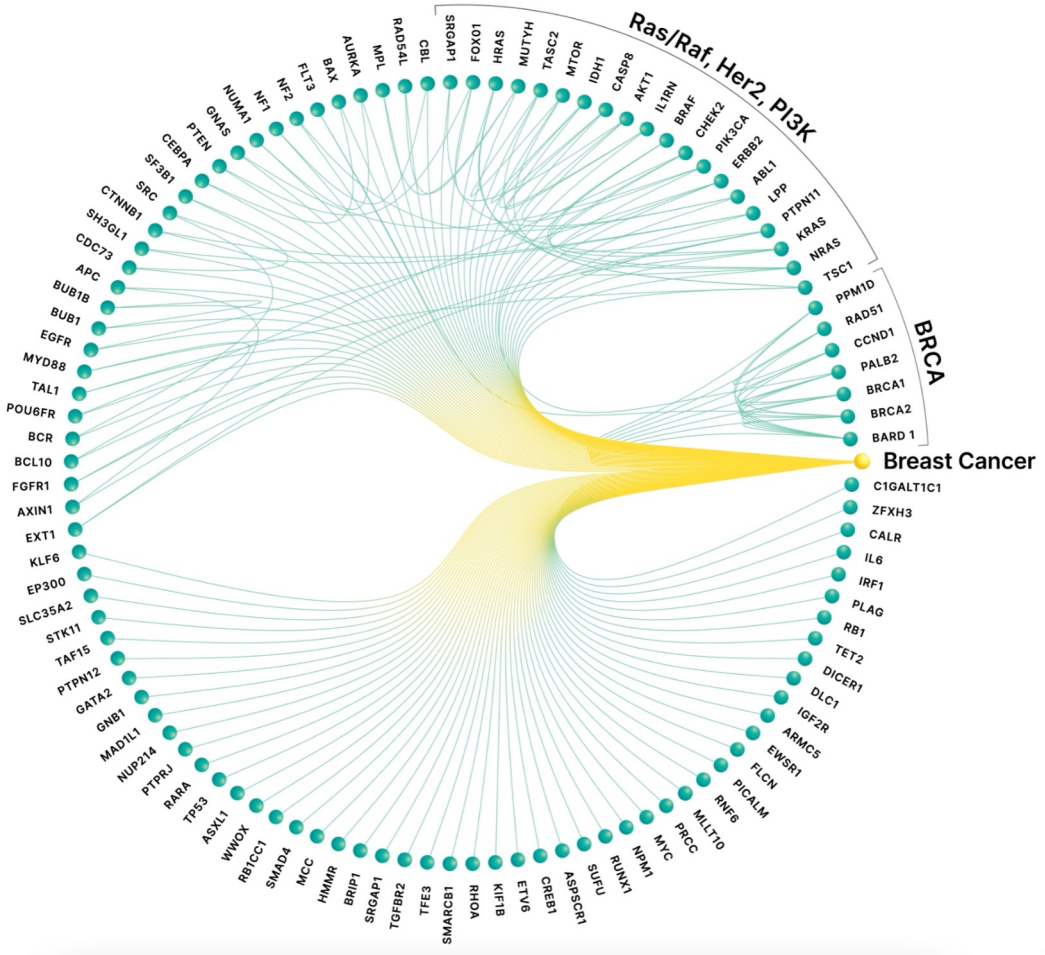
Recursion

What would happen if you took cells and knocked out each gene in the genome or individually dosed 100,000s of molecules at multiple concentrations, and took some pictures?



Maps of Biology: ***de novo*** **pathway reconstruction**

- Arrayed CRISPR knockouts of 17K genes in our maps
- Starting from query on “genes associated with breast cancer”, with no other prior data, Recursion map automatically reconstructs decades of known relationships – BRCA complex, Ras/Raf pathway, etc.

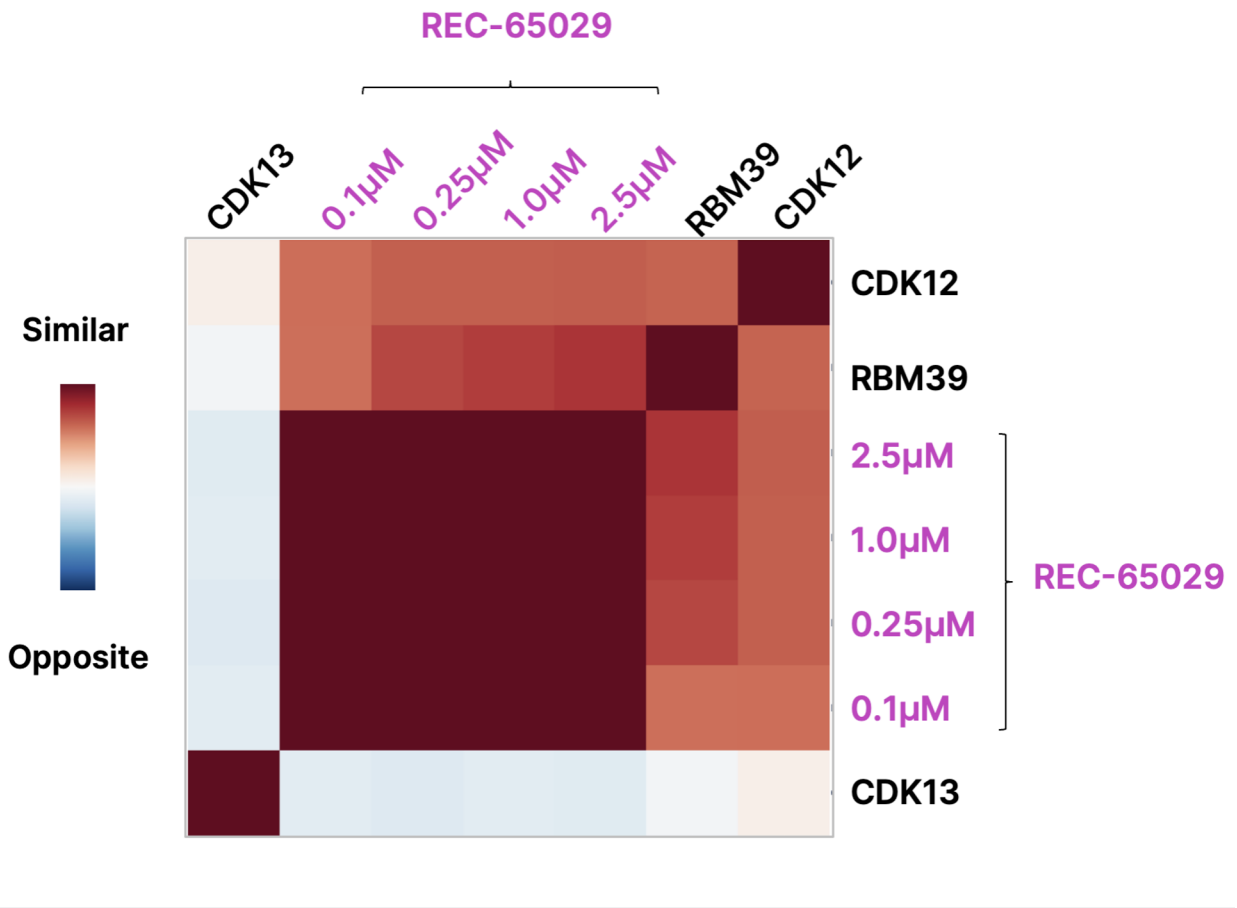


Recursion

Maps of Biology: **primary and alternative target selection**

Recursion “Target Gamma” program for HR-proficient ovarian cancer

- CDK12 has been advanced as a target to improve response in the HR-proficient setting.
- Selective inhibition of CDK12 over other CDKs, especially CDK13, is very challenging.
- Recursion maps of biology show that Inhibition of target RBM39 (e.g., with REC-65029) may mimic inhibition of CDK12 while mitigating toxicity due to CDK13 inhibition.



Recursion

Maps of Biology: **high-dimensional genome-wide screening**

One compound, biological similarity to all targets

- Recursion maps of biology allow the evaluation of concentration-response activity of each compound against *all* gene knockouts in one assay, rather than one assay *per target*.
- CRCs for bortezomib show similarity not just to proteasome subunit KOs, but also to splicing factor *SNRPD3* (known to regulate proteasomal RNAs) and masked potential targets.

(Check out MolRec™ at rxrx3.rxrx.ai!)

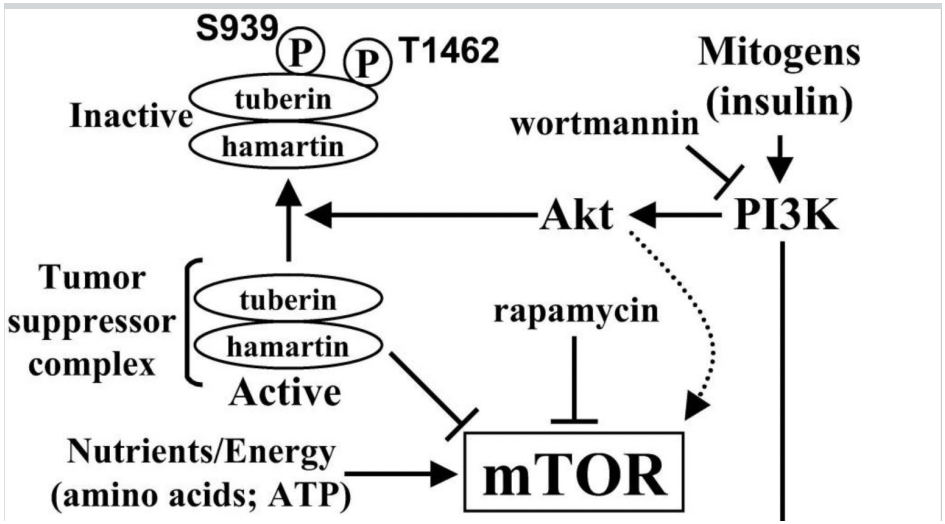
Recursion

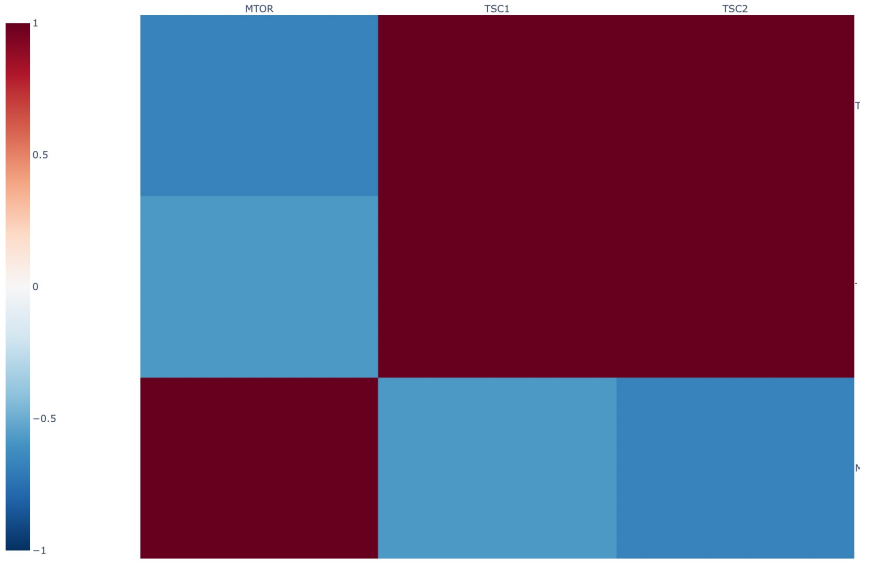
Enabling image-based maps of biology with ML/AI

Similarity is the fundamental property of maps

In order to build a map, we want to know how **similar** or **dissimilar** two biological states are.

We can intuitively think of this in a perturbative context: two perturbations are similar if they make the cell “do the same thing”, and opposite if one reverses the effect of the other.





Tee AR et al. PNAS 2002

Breadth, standardization, and computability define universal assays

We'd love to have a *universal* assay that we could apply to any perturbation rather than designing one for each gene.

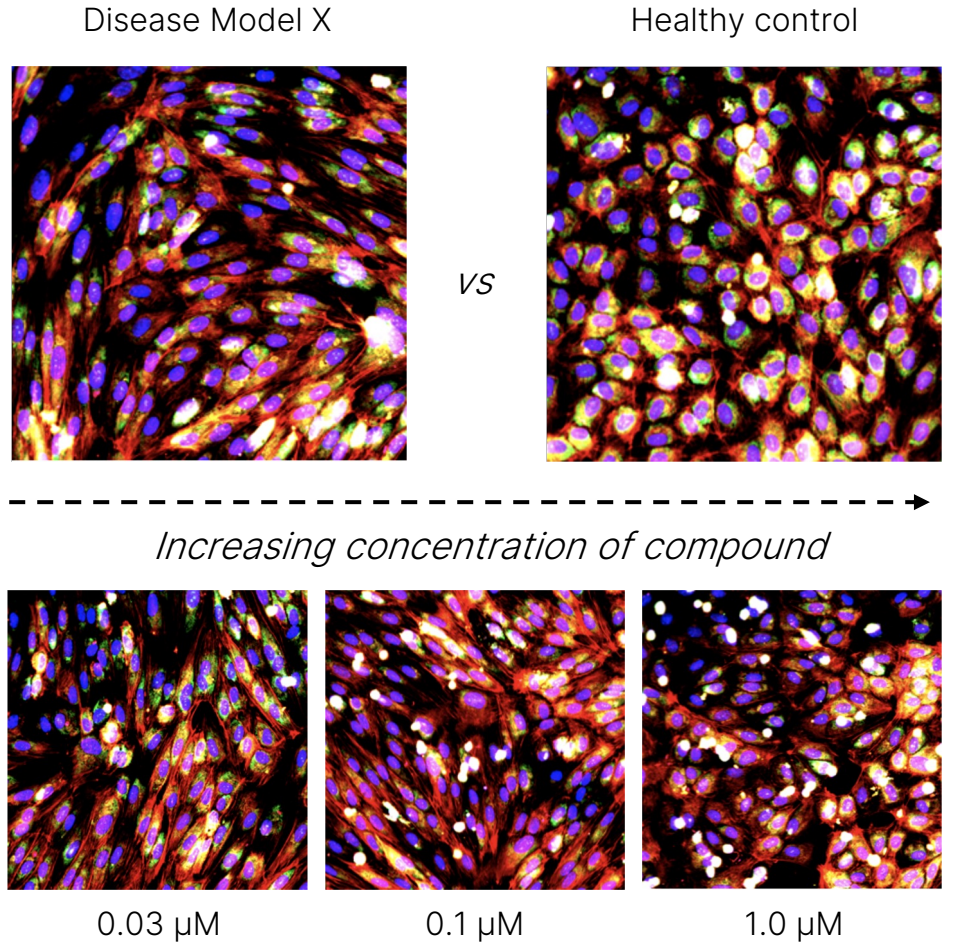
What would make such an assay useful?

- **Breadth of coverage** (which perturbations produce a response),
- **Standardization** (how much they need to be engineered for each condition),
- **Computability** (how easily we can analyze and integrate the data from each perturbation).

Molecular 'omics come close, but are $$$!

Imaging is distal, data-rich, and cheap

- Morphology is downstream of RNA and protein activity – and may observe effects molecular assays would not.
- Images can be **super cheap**.
- Recursion uses a standardized assay we call “**phenomics**”, staining six common cellular substructures.

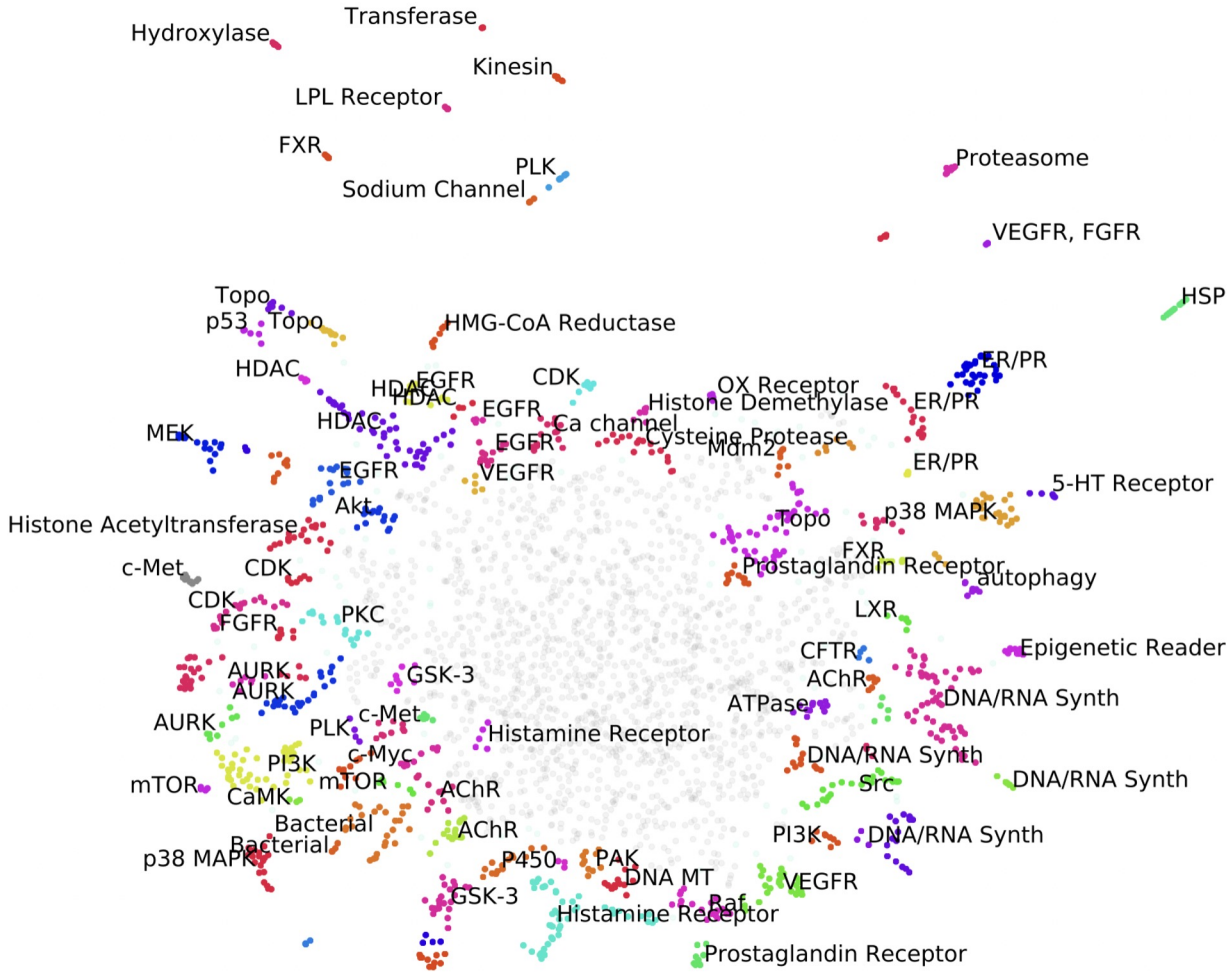


Recursion

*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

Standardized imaging assays capture broad swaths of biology

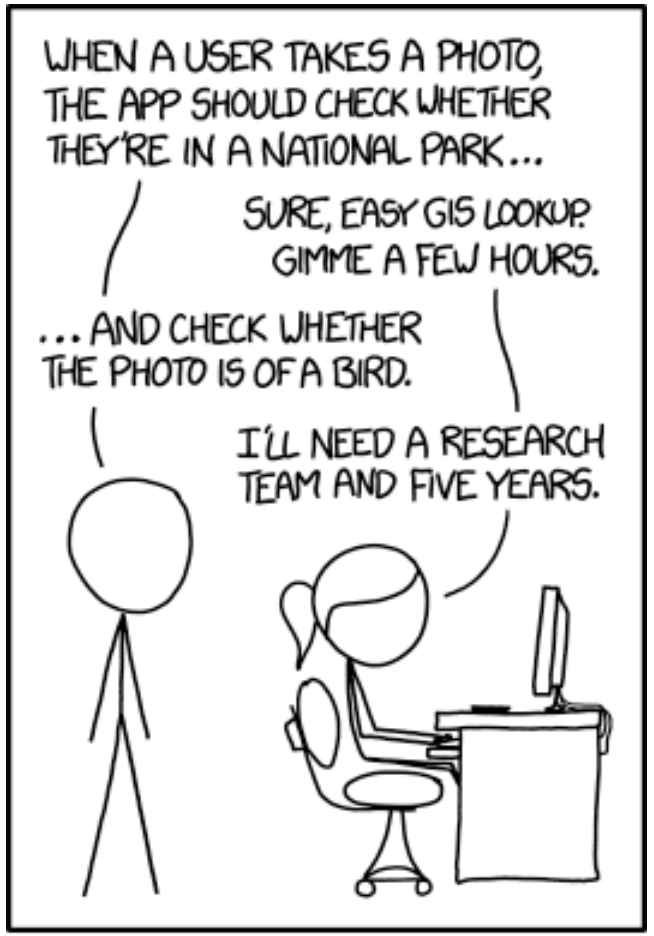
- Phenomics is an unexpectedly powerful standard assay capable of sensitive detection and quantification across 100s-1000s of mechanisms.



Recursion

Computability is a challenge with imaging data

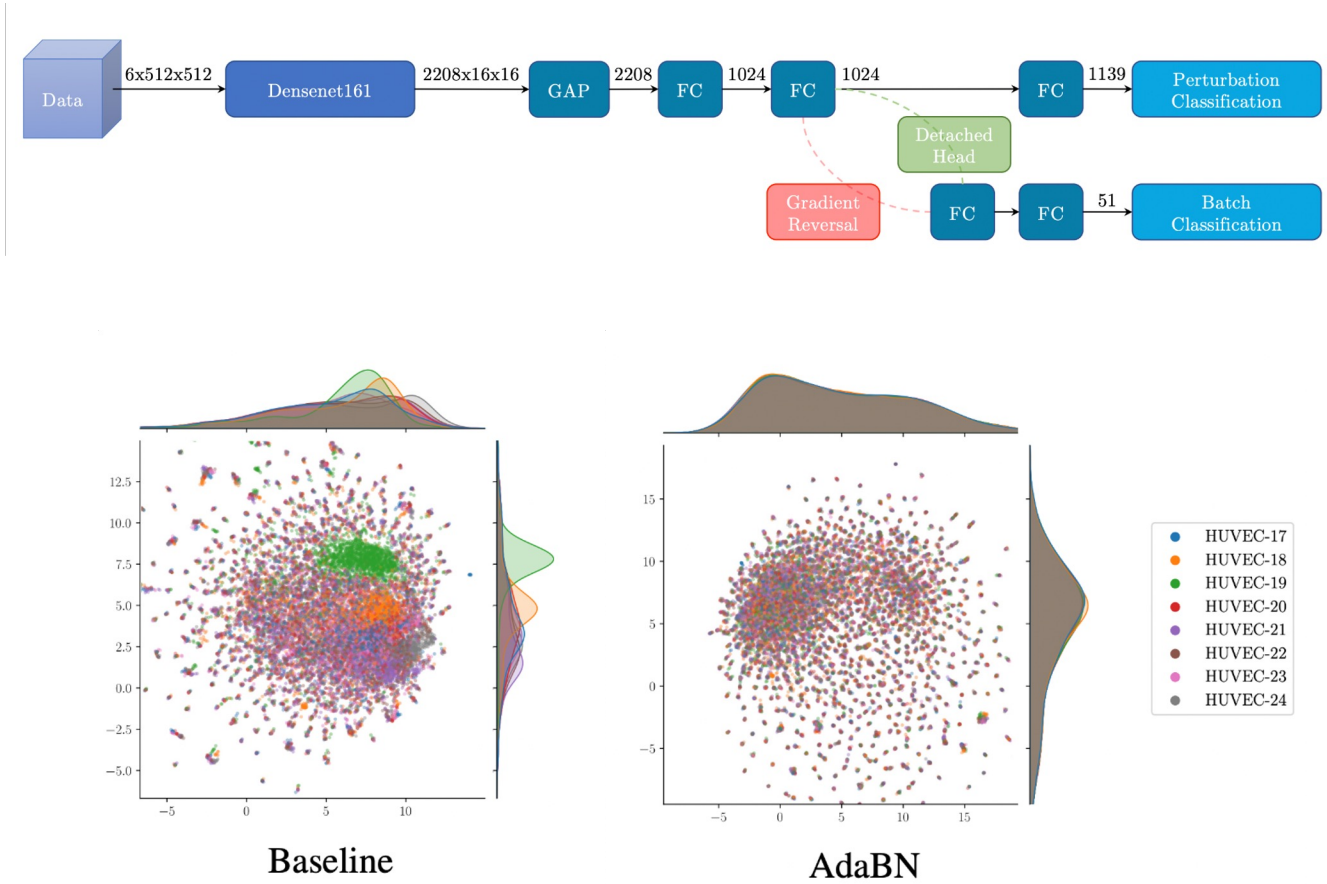
- Image data is not naturally tabular or computable – how can we relate meaningful phenotypes rather than pixels?



xkcd.com/1425

AI/ML turns unstructured images into computable data

- Computability is the fundamental challenge of mapping with images.
- DL algorithms can extract *biologically meaningful* representations of images and automatically correct issues like batch effect.
- These models have shown the power to accelerate development across cell types, and get better with more data.



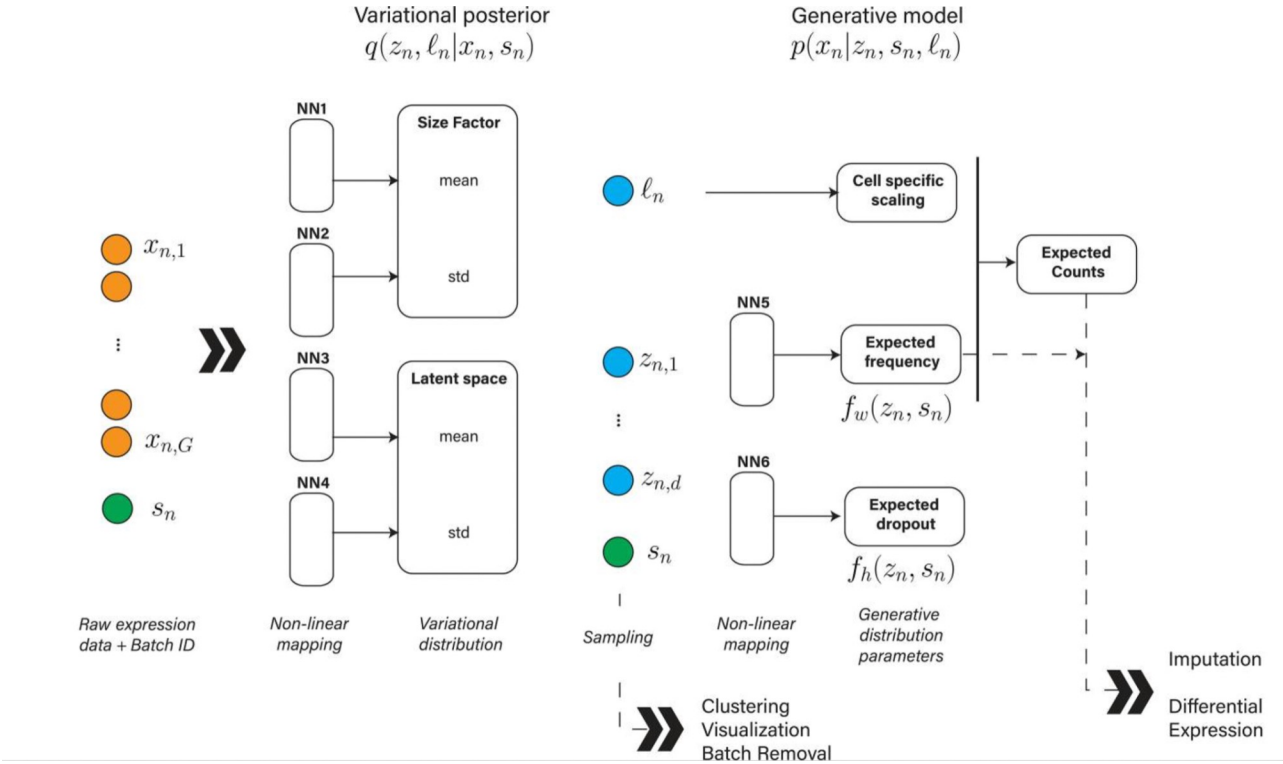
Recursion

Sypetkowski et al. https://arxiv.org/abs/2301.05768

Opaque features suffice for mapping

Deep learning features are typically uninterpretable – and that's fine!

Even 'interpretable' methods like transcriptomics are converging on the use of DL-driven features as well.



Recursion

Lopez, R., et al. Nat Methods 15, 1053–1058 (2018)

RxRx3: Enabling ML research in phenomics

RxRx3: Leading the field in open science

rxrx.ai/rxrx3

RxRx3:

- Images, metadata, and DL embeddings of knockouts of ~17K genes and multiple concentrations of ~1700 compounds.
- The largest publicly-released data set of perturbative cellular imaging, all generated at a single site with a consistent protocol.

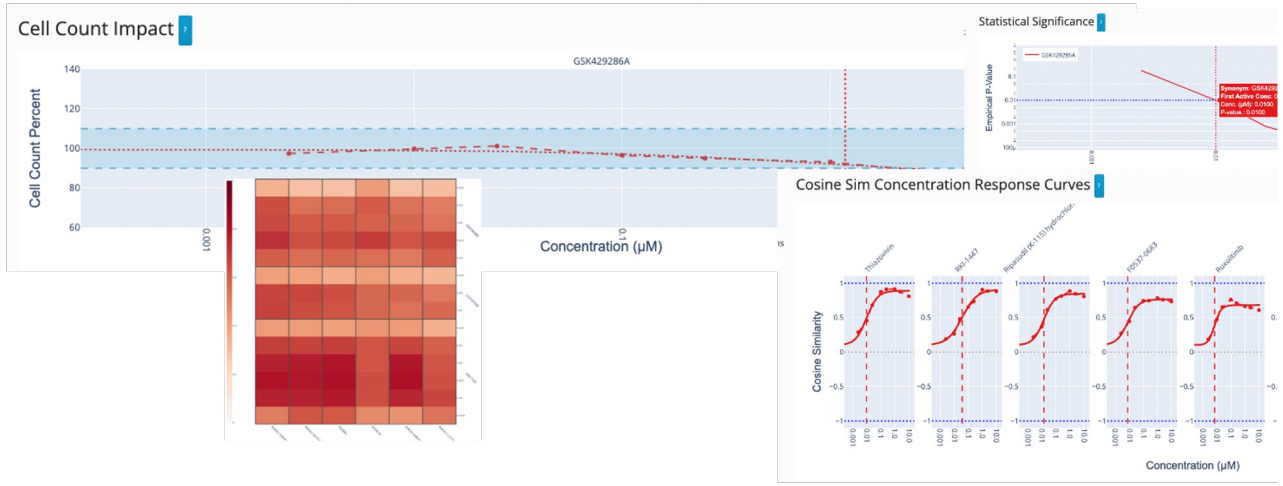
Historically, datasets have driven advances in machine learning technology.

	Dataset	Released	# of Samples
~100 TB	**Bio/Chem Phenomic Maps**		
	RxRx3	**2023**	**2.2M**
	JUMP-CP	2023	823,438
~1-5 TB	**Autonomous Driving**		
	Waymo Open Dataset	2018	~105,000
	nuScenes	2018	1000
10 GB - ~1 TB	**Image/Object recognition**		
	ImageNet (21k)	2009	14M
	COCO	2014	330,000

MolRec: A keyhole view into the Recursion Map of Biology

rxrx.ai/rxrx3

MolRec shows off the ability to relate small molecules to each other and to gene knockouts, with the suite of analyses you might expect to want in driving a discovery program – cellular toxicity, on- and off-target similarity, and compound similarity.

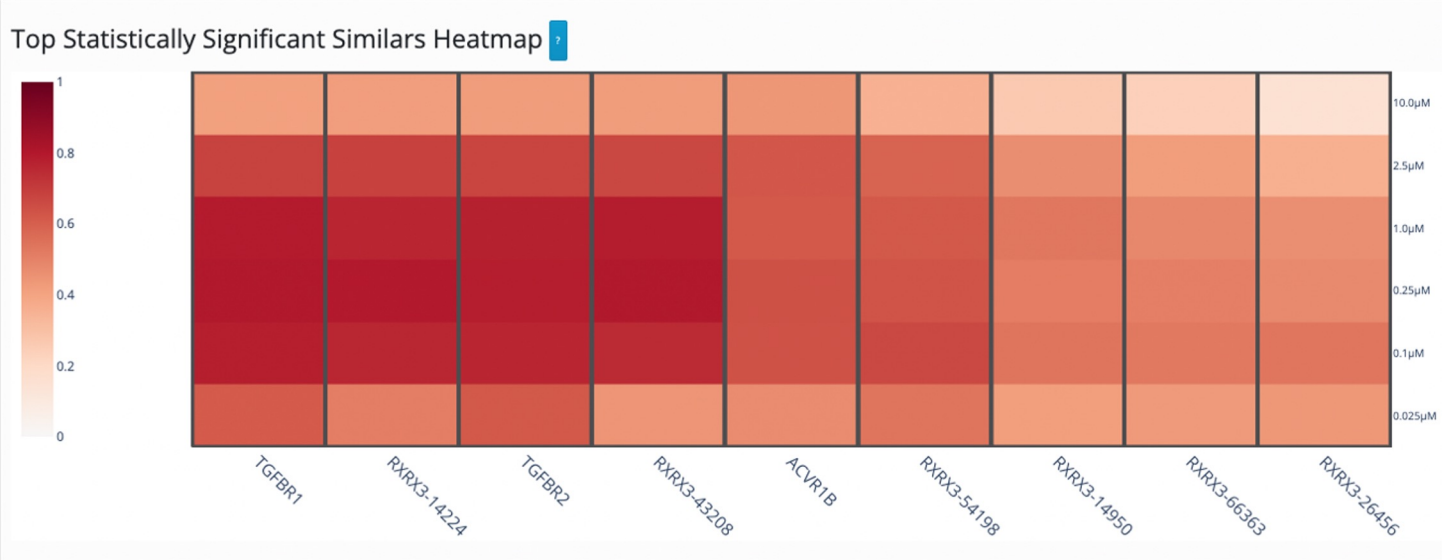


Recursion

RxRx3 and MolRec: blinded research data sets

rxrx.ai/rxrx3

RxRx3 and MolRec are partially blinded.
We expect to unblind more of these over time.



Recursion

RxRx3 and MolRec: blinded research data sets

rxrx.ai/rxrx3

RxRx3 and MolRec are partially blinded. We expect to unblind more of these over time.

But you know, if you want to pay us some money to unblind it, let's talk.

For just $0.99 more, I'll personally hand-deliver the hard drives.

Recursion

Conclusion

AI/ML Structures Unstructured Image Data for Mapping

- Standardized cellular imaging (**phenomics**) is incredibly data-rich and scalable, but produces *unstructured* data that is difficult to compute on.
- Given sufficient data, deep learning converts unstructured data into **biologically-meaningful representations** that enable *mapping* of biology.
- Recursion has released **MolRec** to provides a view into the power of *mapping and navigating*, and **RxRx3** to advance research on machine learning and cellular imaging.

Questions?

info@rxrx.ai for questions on RxRx3 and MolRec

@ImranSHaque on Twitter or @ihaque@genomic.social on Mastodon


Recursion