Genentech
A Member of the Roche Group

NVIDIA

Mila

Using AI to discover new antibiotics

Tommaso Biancalani
Senior Director
March 20th 2024

This is the BRAID team!

Heads
Tommaso Biancalani
Hector Corrada-Bravo

DELTA
Gabriele Scalia
Nate Diamant
Ziqing Lu
Alex Tseng
Ehsan Hajiramezanali

ReLU
Gokcen Eraslan
Avantika Lal
Laura Gunsalus

Post-docs
Hejin Huang
Sepideh Maleki
Masatoshi Uehara

Admin
Vilma Bermudez

SCimilarity
Graham Heimberg
Jenna Collier
Max Gold
Tony Kuo

MAGIC
Aicha BenTaieb
Max Gold
Alma Andersson
Shreya Gaddam
Kam Hon Hoi

Perturbations
Dave Richmond
Jan-Christian Huetter
Jacob Levine
Rahul Mohan
Alex Wu
Heming Yao
Phil Hanslovsky
Burkhard Hoeckendorf
Lin Qin

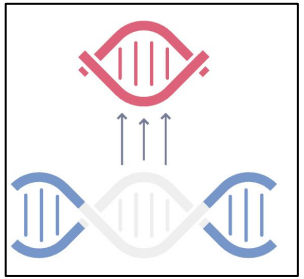
CTGi
Bo Li
Yiming Yang
Joshua Gould
Jose SL Lonzano

Corrada-Bravo lab
Alsu Missarova
Changlin Wan

Rough sketch of a drug discovery pipeline

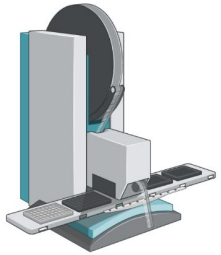
This work focuses on finding more "hits" (molecules exhibiting properties of interest)

Target Identification



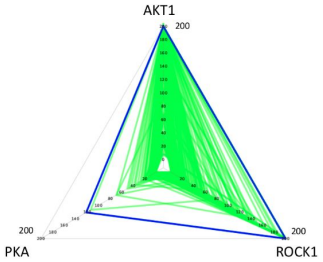
Understand disease through biology

Hit Finding



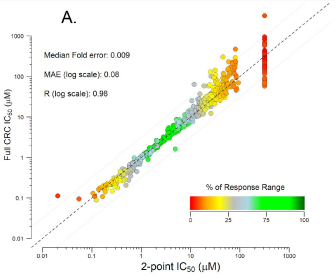
Identify candidate molecules

Lead Optimization



Optimize the molecules into actual drugs

Safety Assessment



Assess toxicity and large scale production

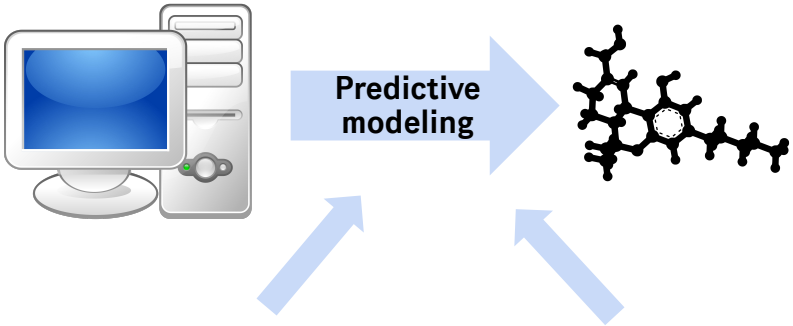
...to the clinic

Our goal: To use AI to find **more** candidate molecules

There are too many (possible) small molecules…

Pharma companies typically screen millions of compounds, but possible ones are estimated >> 10^{23}

- **Solutions:** Using computer algorithms to predict what molecules do

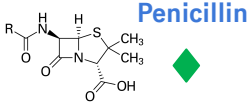


- **However…**

*We've been working on **virtual screening** for decades, with a level of success that can be characterized as quite variable but (to be honest) often underwhelming.*

-- Derek Lowe, Science (2020)

We are searching for drugs that are *different* than those we know



Penicillin

Known chemical space (small)

Unexplored chemical space (very vast)

Compound legend

- Natural products
- Internal library compound
- Synthesizable compound
- Newly-discovered active

AI strategy for virtual screens

Train on , (known activity)

Predict activity on

Test activity on to discover

How can AI discover *different* molecules?

Let's reformulate the problem using a simplified analogy

Universe of known dogs
What we train our models with

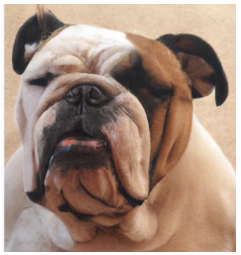
Shepards

Terriers

AI/ML

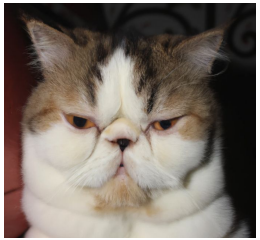
Unexplored animal space
Where discoveries are made

Bulldogs



Not-a-dog

Cat
(but looks like a bulldog)



How can AI discover *different* molecules?

We leverage the diversity of the training set to “learn how to learn” (meta-learning)

Universe of known dogs
(training set)

Shepards

AI/ML

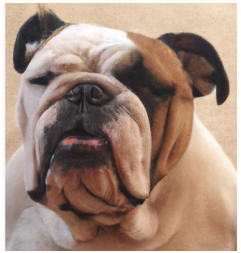
meta-learning

Terriers

AI/ML

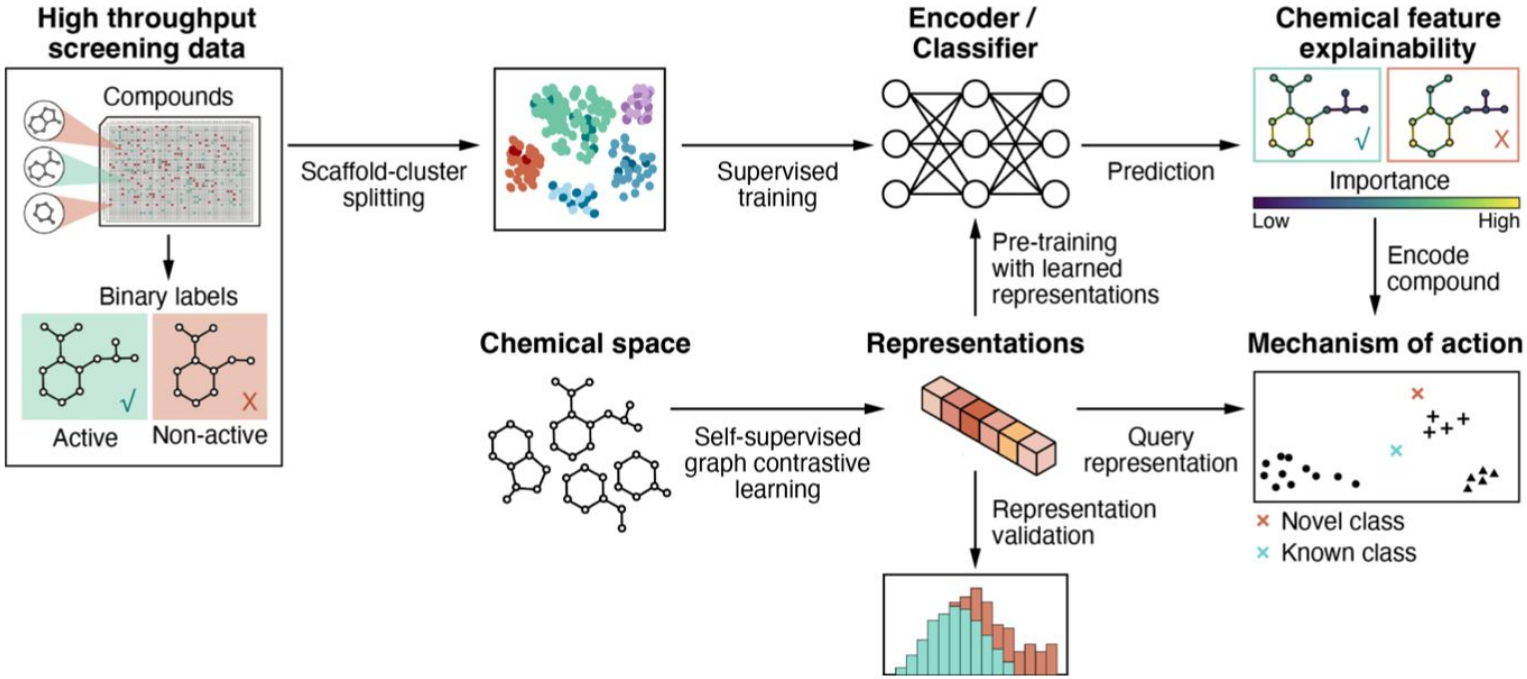
Unexplored dog space
(test set)

Bulldogs



GNEprop: our computational strategy for virtual screens

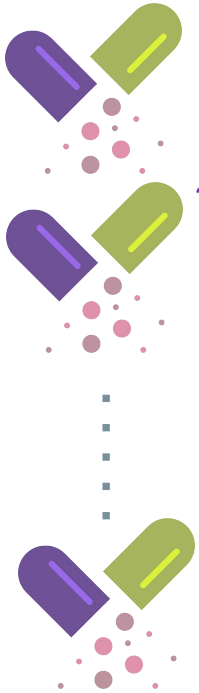
GNEprop stands for Graph Neural Encoder of chemical properties



PyTorch geometric

Our goal is antibiotic discovery in Gram-negative bacteria

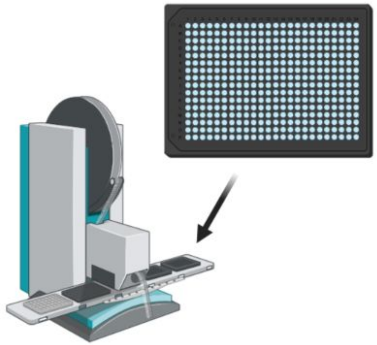
We screened 2M molecules to identify those that kill the bacterium *E. coli* (2017)

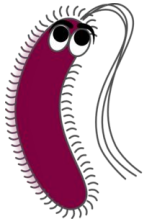


~2.1M compounds

5 µM

High-throughput screening robot

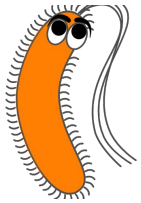




E. coli ΔtolC

Efflux-deficient and disrupted outer membrane

~6k hits



E. coli WT

~ 850 hits

Compounds are labeled as

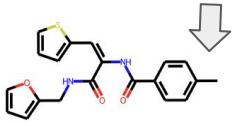
- Active (growth rate inhibition > 80%)
- Non-active (growth rate inhibition < 20%)

How to assess if the model is generalizing on novel scaffolds

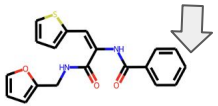
We evaluate the model by predicting activity cliffs on unseen scaffolds

TASK 1
Learn chemistry

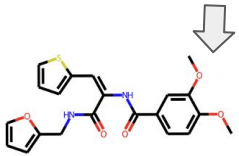
TASK 2
Characterize activity cliffs



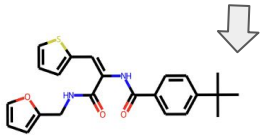
Inactive



Inactive

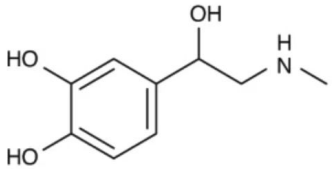


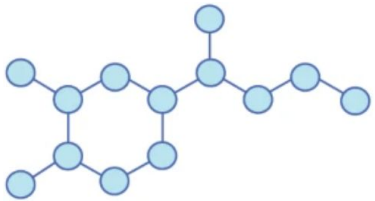
Inactive



Active

TASK 3
Structural explainability



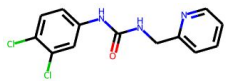


How to assess if the model is generalizing on novel scaffolds

We evaluate the model by predicting activity cliffs on unseen scaffolds

Scaffold hop

Training set

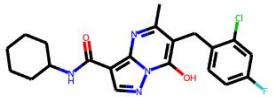


Active

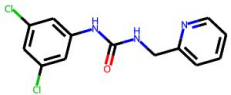
Difficult to predict

Easy to predict

Test set

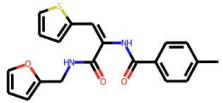


Active

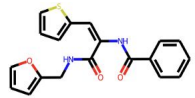


Active

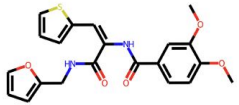
Activity cliffs



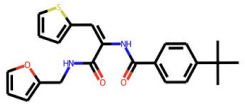
Inactive



Inactive



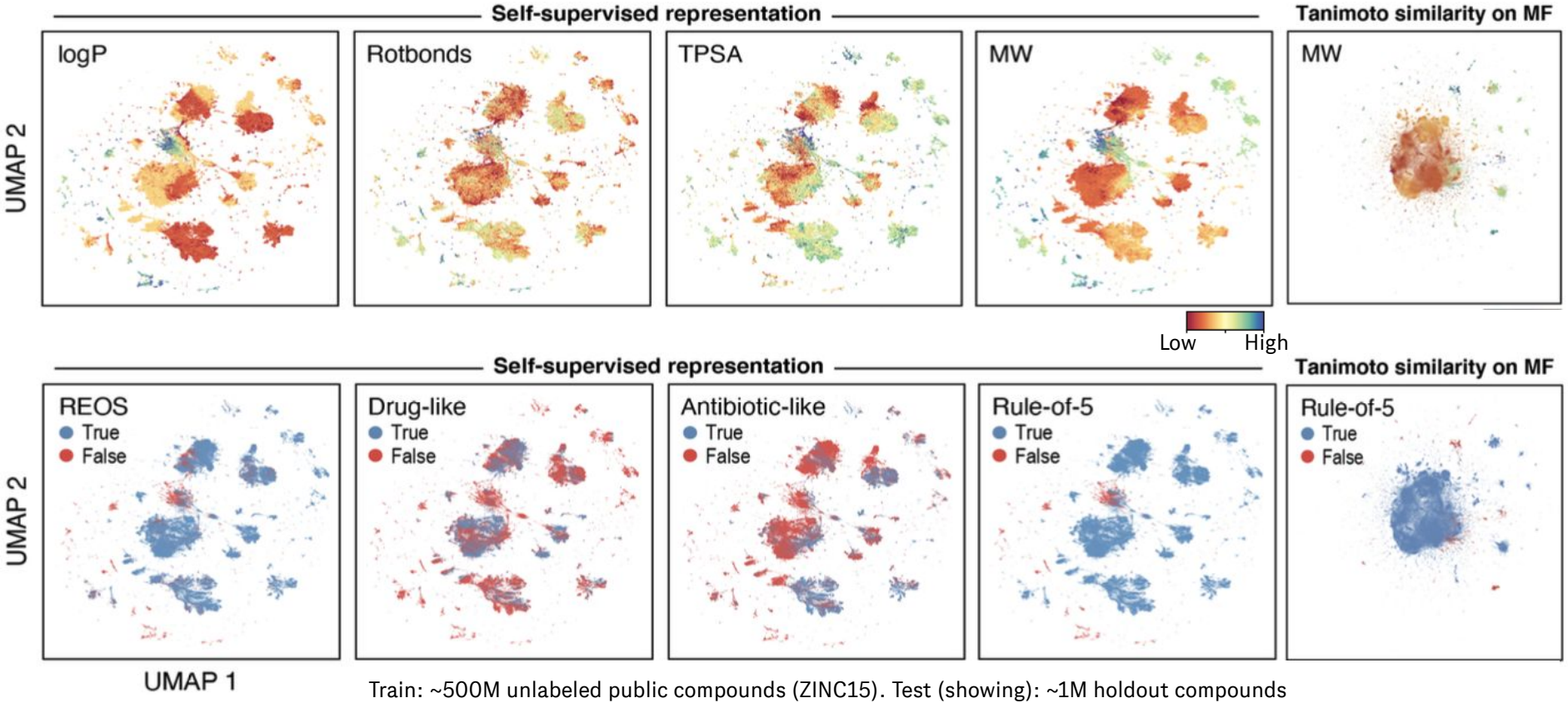
Inactive



Active

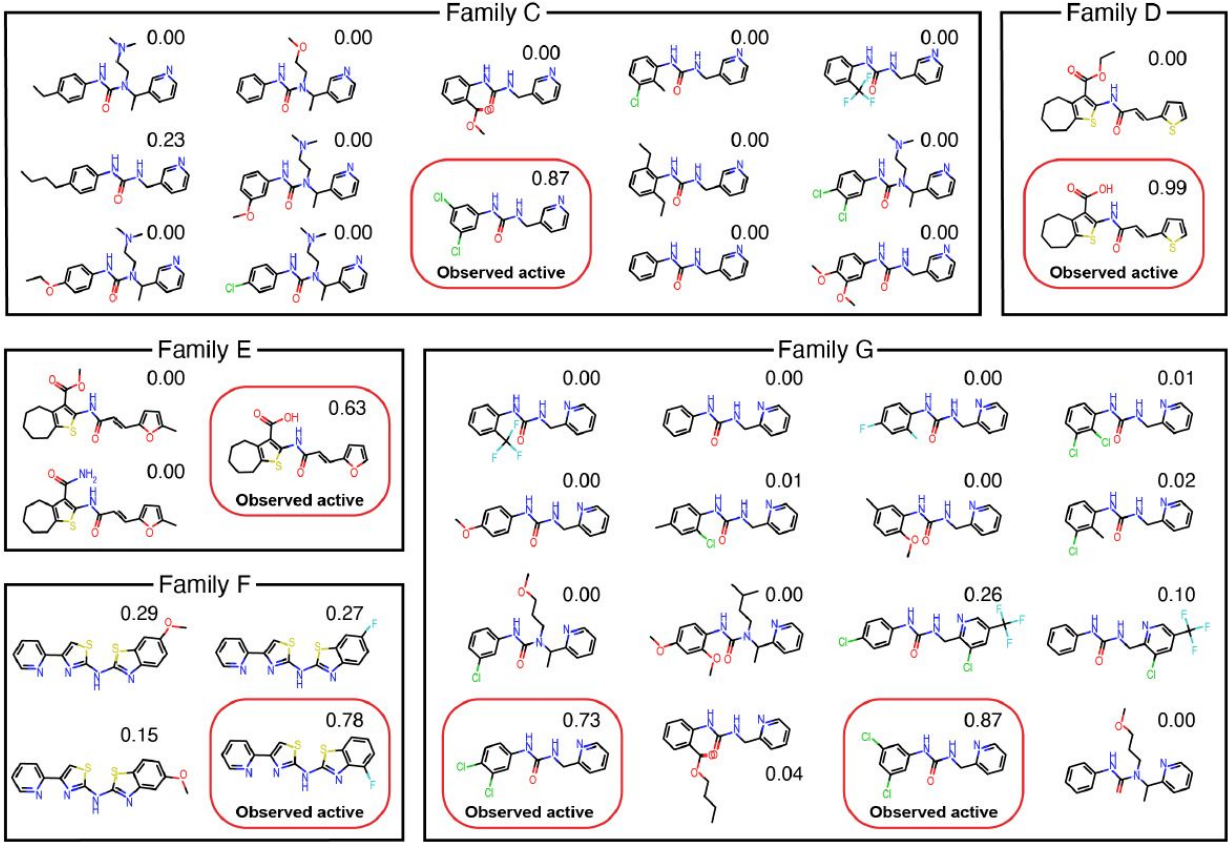
TASK 1: Learning chemistry via a self-supervised representation

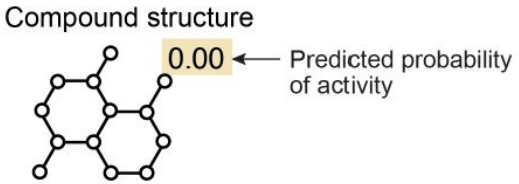
Visualization obtained via RAPIDS cuGraph

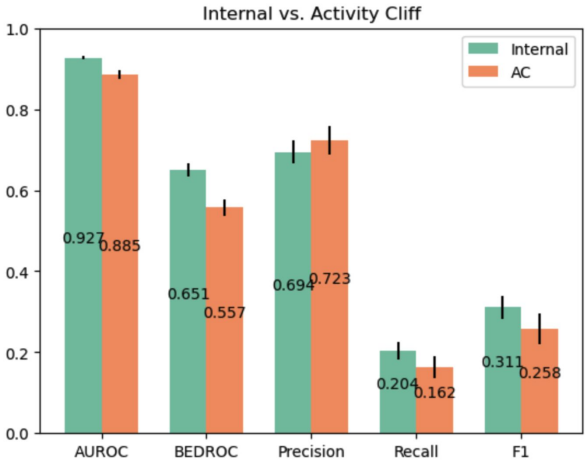


Train: ~500M unlabeled public compounds (ZINC15). Test (showing): ~1M holdout compounds

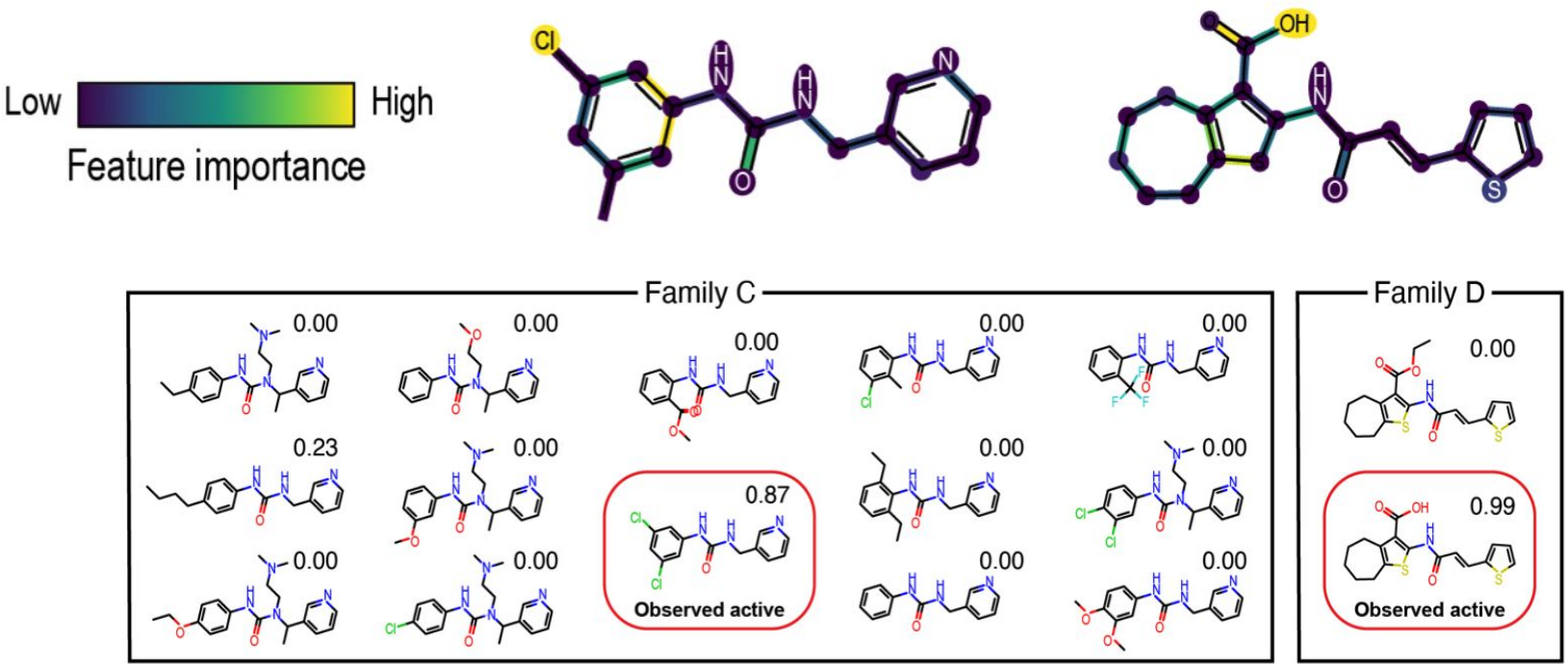
TASK 2: Predicting activity cliffs on unseen scaffolds from our annotated data



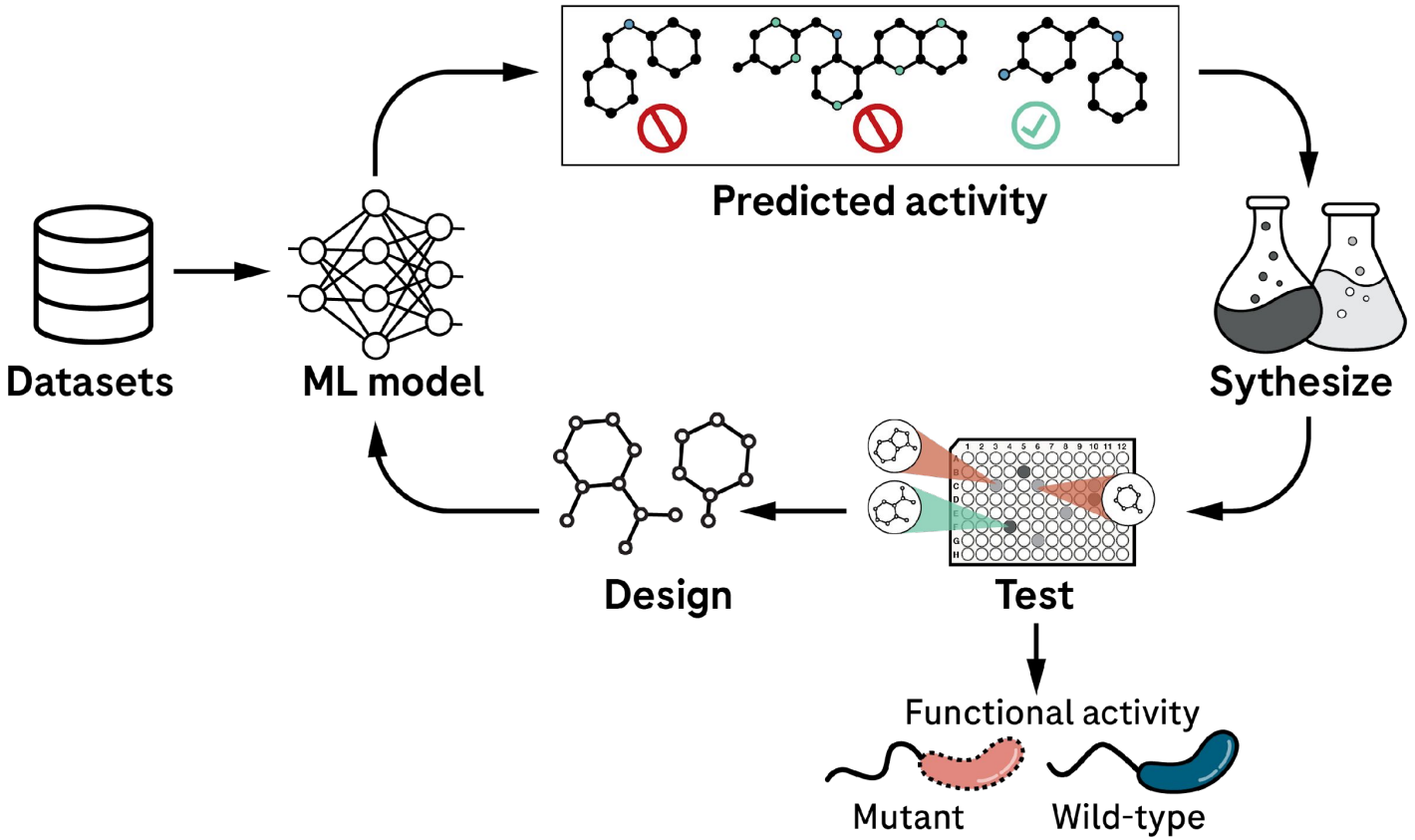




TASK 3: Explainability underlies structural parts responsible for activity cliffs



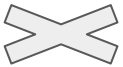
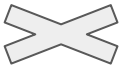
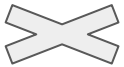
Enabling antibiotic discovery via AI-enhanced lab–in-the-loop



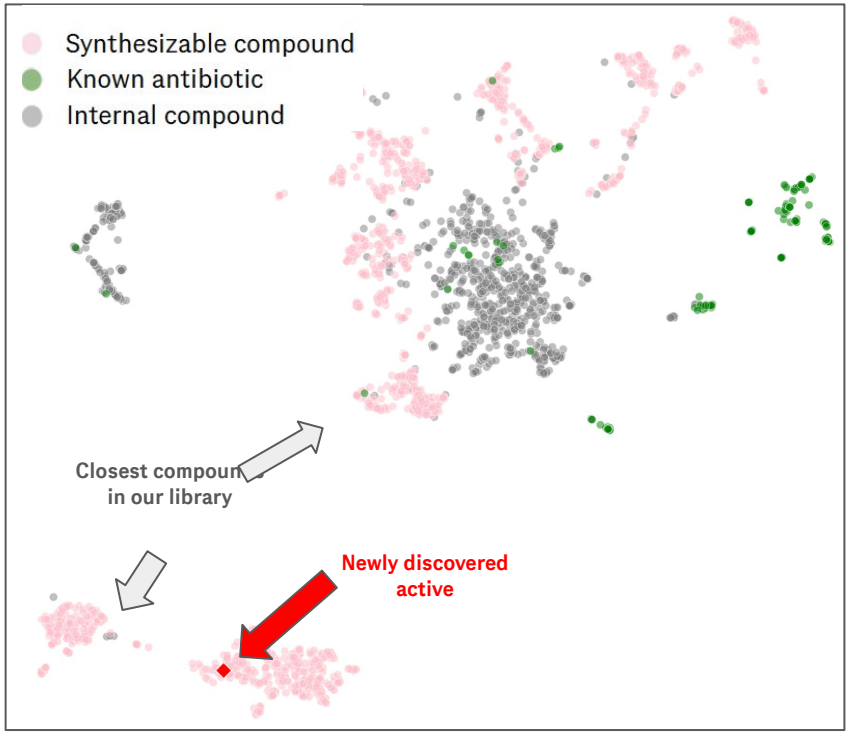
GNEprop achieves significantly increased hit rate in prospective screens

We virtually screen Enamine for activity against the *E. coli* ΔtolC mutant

Library	#
Enamine library (2020)	1.4B
GNEprop hits	44,437
GNEprop purchased	345
Confirmed hits	**82**

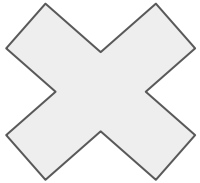
	Hit rate (2017)	GNEprop (2021)	Fold Enrichment
***E. coli* ΔtolC mutant**	**0.4%**	**24%**	**60X**
***E. coli* wild type**			

GNEprop on HTS data leads to discovery of novel molecular scaffolds active against *E. coli* Δ*tolC*



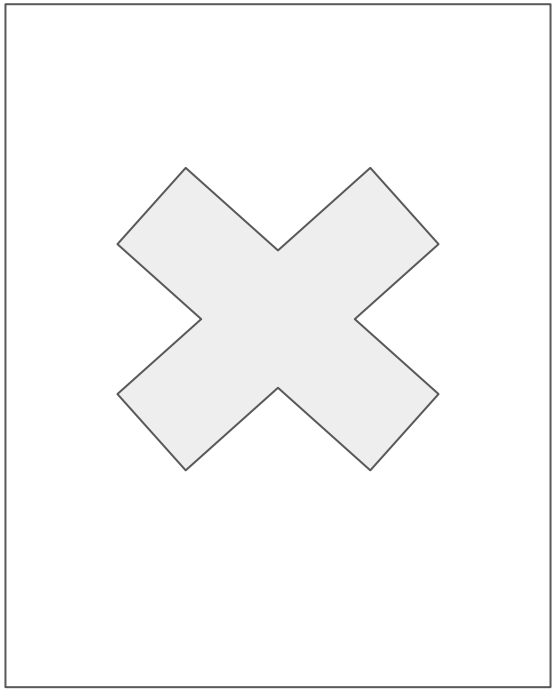
UMAP from Tanimoto distances of Morgan fingerprints of compounds

Newly discovered active

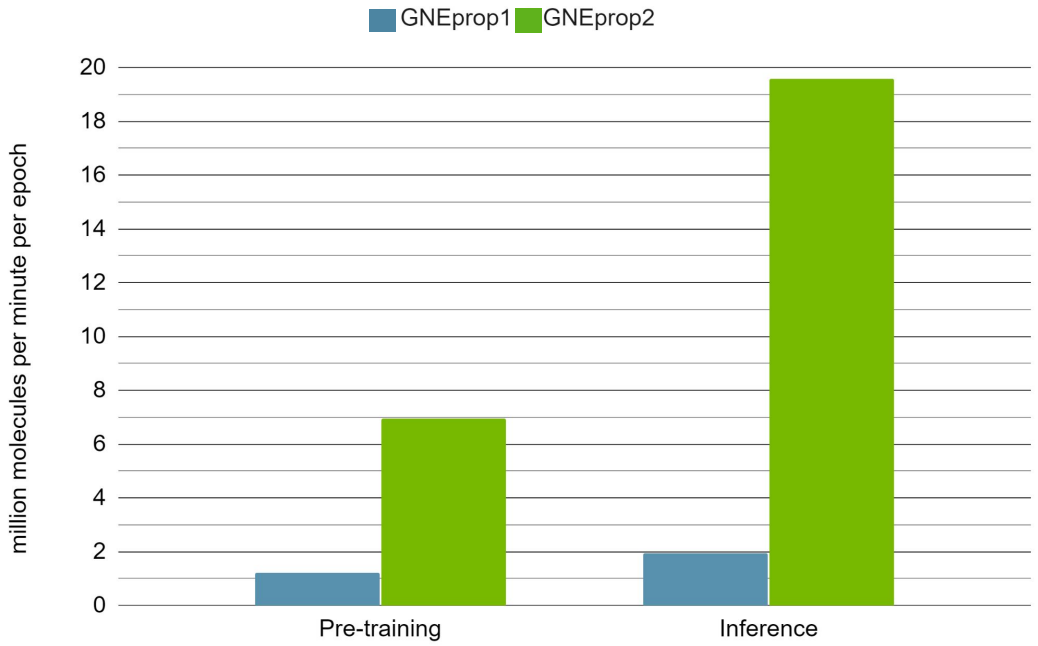


How structurally similar is the newly discovered active vs our internal library?

Top 5 structurally similar gRED compounds



GNEprop 2.0 now runs MUCH faster, due to NVIDIA and Genentech collaboration



Total pre-training time is down **from weeks to hours**
from both hardware parallelization and software optimizations.

NVIDIA

PyTorch geometric

RAPIDS

DevTech

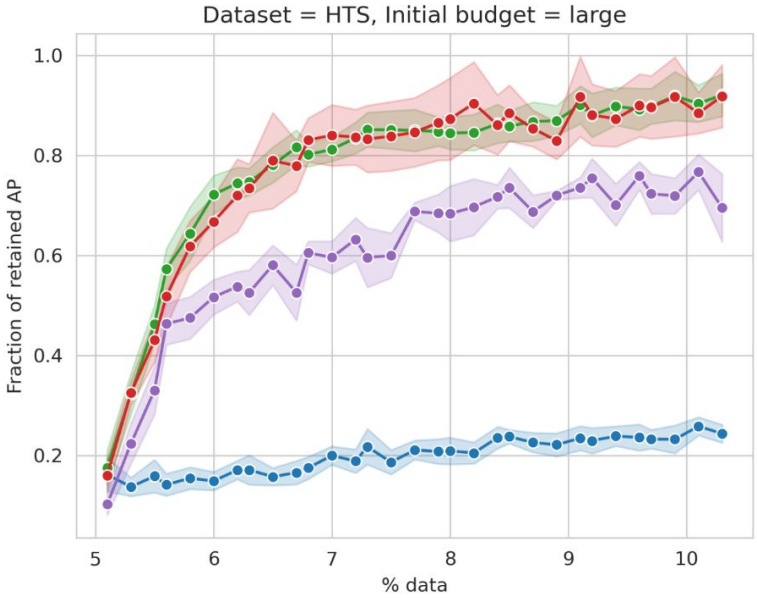
+

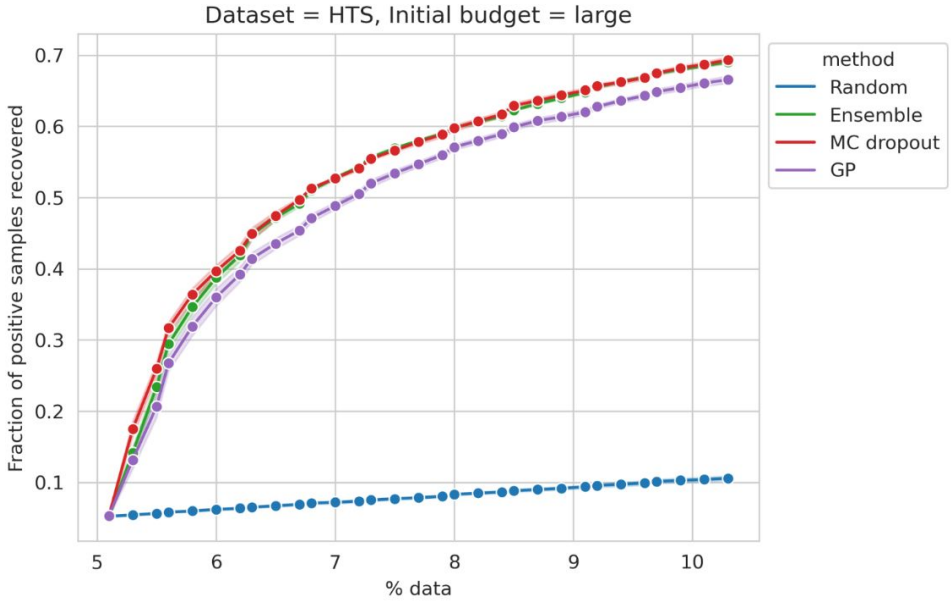
Solution Architects

Greg Zynda and Nia Dickson

Future direction: Expanding libraries combining active learning and lab-in-the-loop

Uncertainty-guided active learning allows choosing the next batch of compounds to maximize model performance (90% performance are achieved with 15% training data)





Mila

Michał Koziarski and Yoshua Bengio

Acknowledgements

Gabriele Scalia
Ziqing Lu
Jerome Luescher
Kangway Chuang

Steven Rutherford
Kerry Buchholz
Anh Miu
Man-Wah Tan

Nicholas Skelton
Leo Gendelev
Jeff Blaney

Nia Dickson
Greg Zynda
Alex Sabatier

Michał Koziarski
Yoshua Bengio