

# BMI 702: Biomedical Artificial Intelligence

Foundations of Biomedical Informatics II, Spring 2024

Lecture 14: Design of chemical and genetic perturbations, drug repurposing, protein design, emerging uses of generative AI



Marinka Zitnik  
[marinka@hms.harvard.edu](mailto:marinka@hms.harvard.edu)

# Outline for today's class

- High-throughput genetic and chemical perturbations
- Drug repurposing, indication and contra-indication prediction
- Generative protein design
- Generative AI agents



Words and genes share a correspondence:  
their **meanings** arise from their **context**.

Gene perturbation measurements across diverse cell contexts  
induce **semantics for genes**

(under the right approach)

“apple” is a **polysemic** word...



grow an apple

buy an apple|

... whose **particular meaning** is resolved via **sentence context**.



grow an apple

- grow an apple tree
- grow an apple tree from seed
- grow an apple tree in a pot
- grow an apple tree indoors

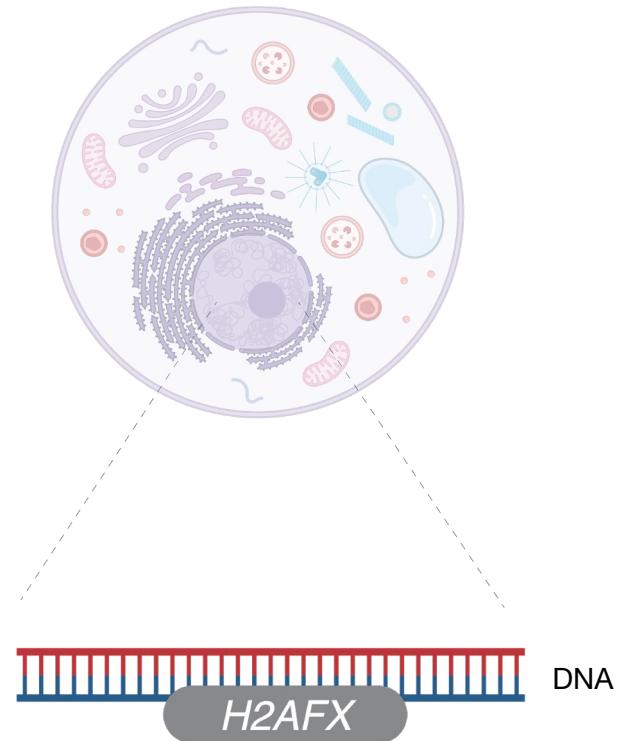


buy an apple|

- buy an apple watch
- buy an apple gift card
- buy an apple tv

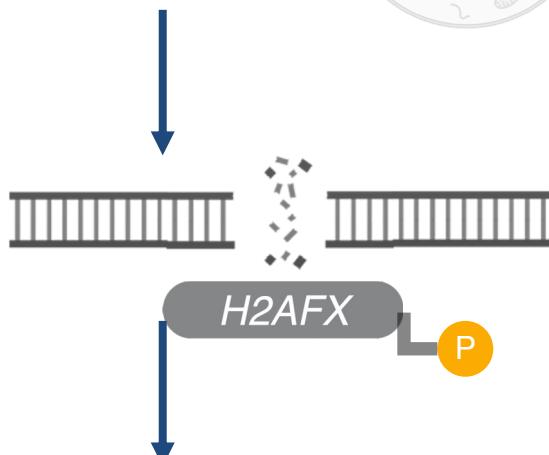
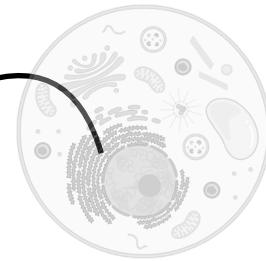
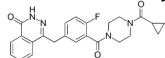


H2AFX is a **pleiotropic** gene...



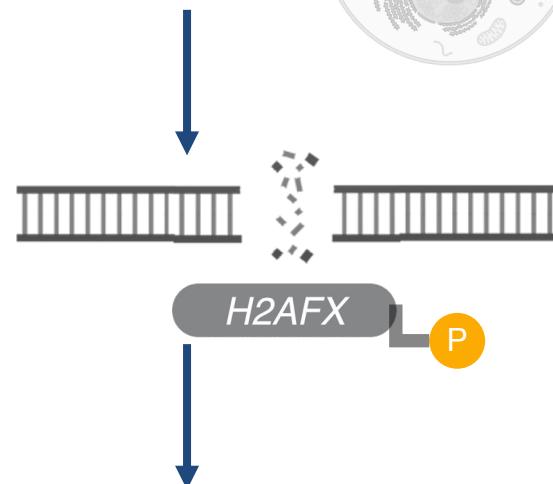
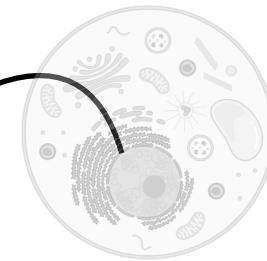
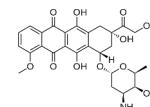
... whose **particular function** is resolved via **cell context**.

Olaparib



Homologous  
Recombination

Doxorubicin



End Joining

While unsupervised learning of word polysemy is **common**...

**Data:** corpus  
of sentence contexts

**Approach:** word embeddings  
w/ linear semantics

$$king - man + woman \approx queen$$

unsupervised learning of gene pleiotropy is **unsolved**

**Data:** ?

**Approach:** ?

$$geneA - func1 + func2 \approx geneB$$

# Our goal for today

Unsupervised learning of gene pleiotropy with applications to therapeutic science

**Data:**

?

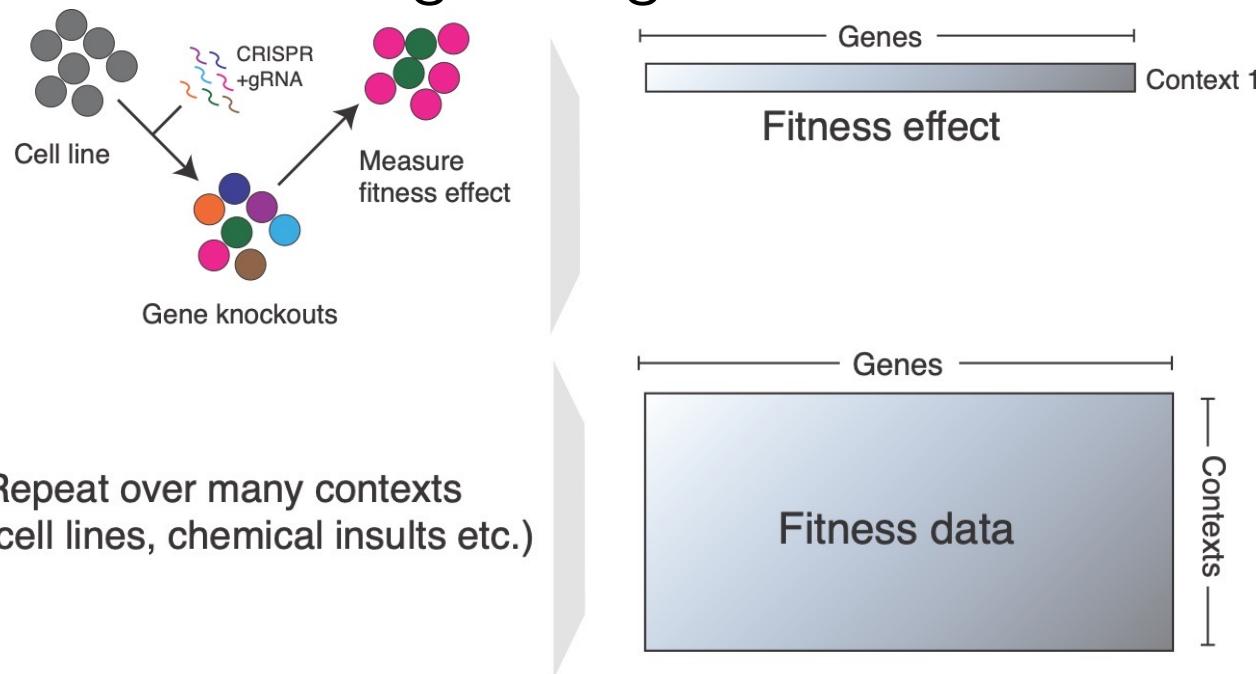
**Approach:**

?

$$geneA - func1 + func2 \approx geneB$$

# Data

Use gene perturbation effect measurements for inferring biological functions

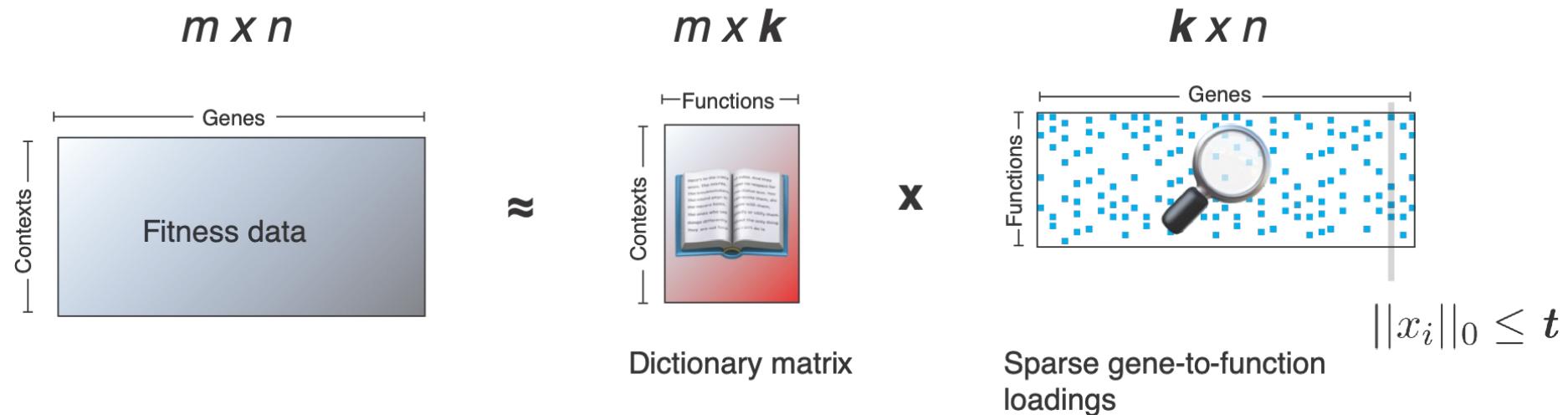


Why perturbation datasets? Alternative data types:

- **Transcriptomics:** gene co-expression is necessary but not sufficient for co-function
- **Protein-protein interactions:** direct interactions are not necessary for co-function

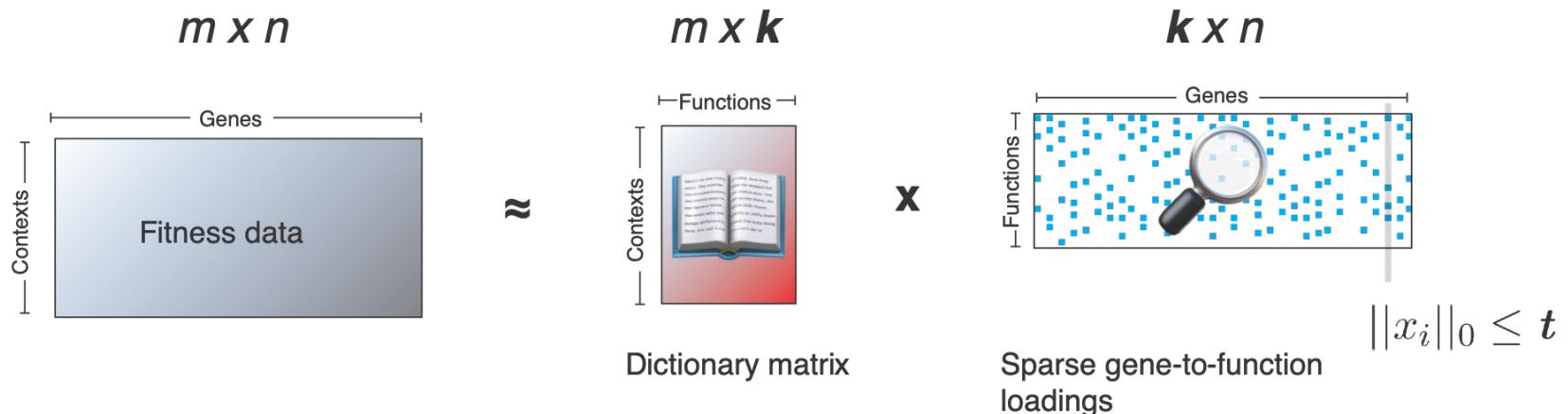
# Approach: Webster

- Low-dimensional vector embeddings that satisfy three criteria:
  - Sparse
  - Latents are biologically meaningful
  - Account for redundancy between cell contexts

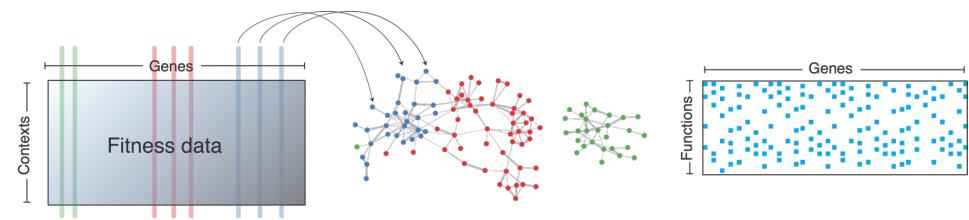


# Approach: Webster

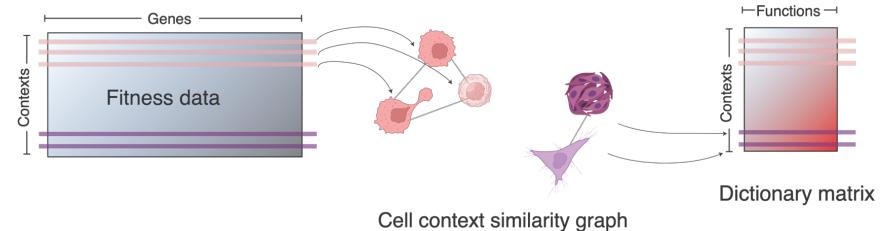
Webster learns a dictionary matrix that sparsely approximates gene effects...



1 ... while preserving interpretable relationships between genes



2 ... and accounting for redundancies between cell contexts

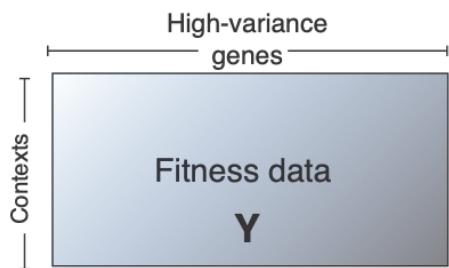


# Overview of Webster

## Preprocessing

Raw fitness data

Standardize cell lines  
Center gene effects  
Filter genes by variance



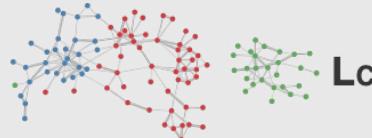
## Graph-regularized dictionary learning *Objectives*

Reduce dimensionality

$$\mathbf{Y} \approx \mathbf{D} \times \mathbf{X}$$

$$\|Y - DX\|_F^2$$

Preserve gene similarity



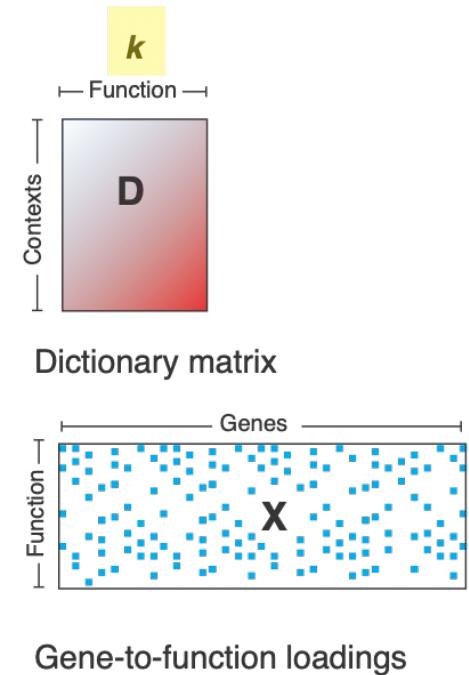
$$Tr(XL_cX^T)$$

Preserve cell context similarity



$$Tr(D^T LD)$$

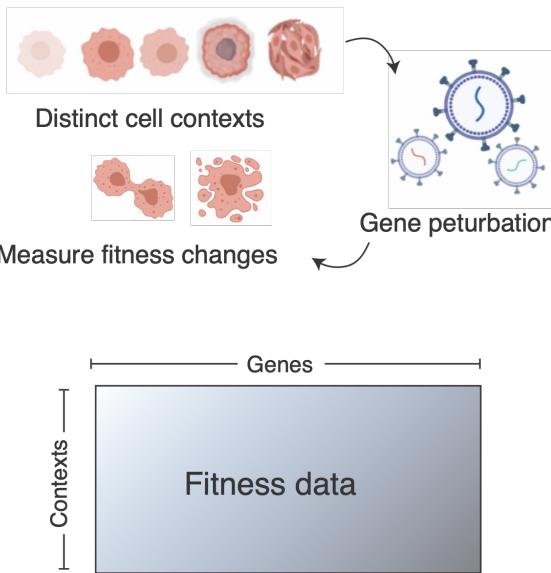
## Output



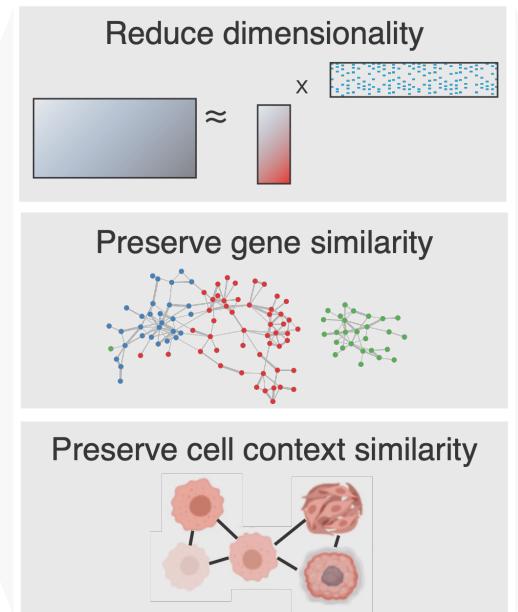
= key hyperparameters

# Its key parameters are dictionary size ( $K$ ) and sparsity on loadings ( $T$ )

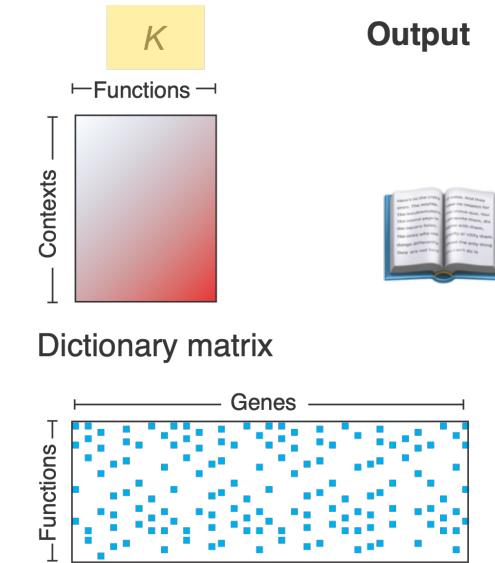
Input



Graph regularized dictionary learning



$K$

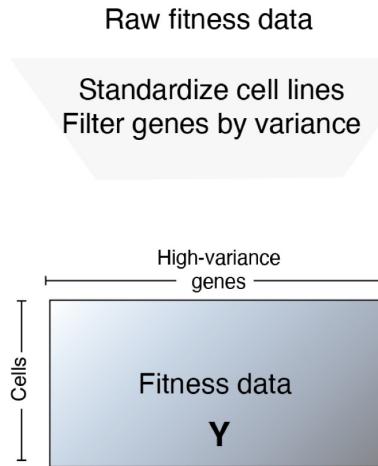


Output

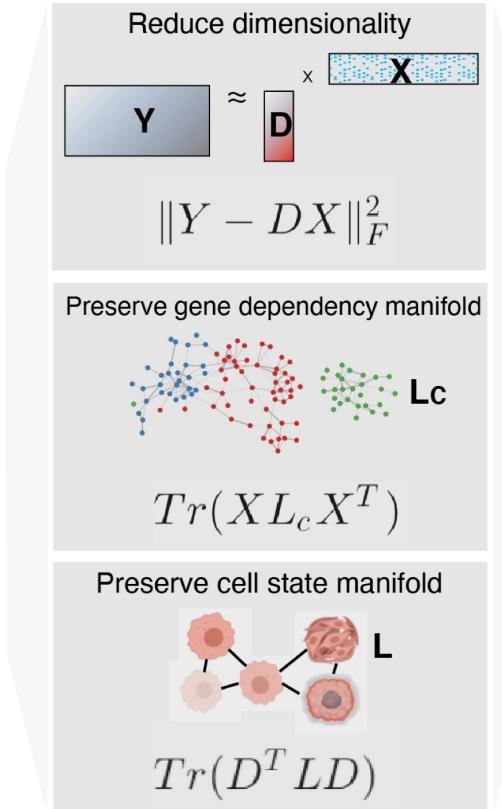


# Model optimization

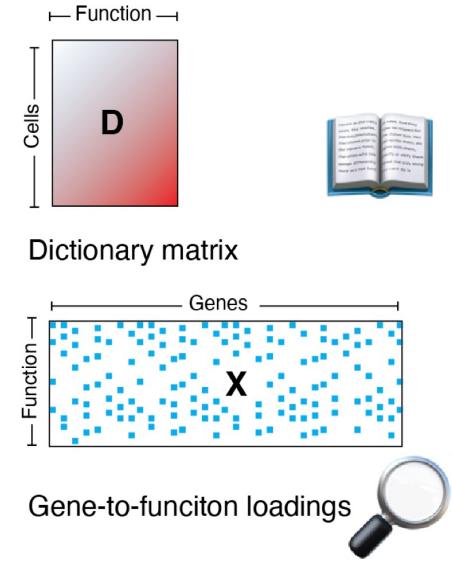
## Preprocessing



## Graph-regularized dictionary learning Objectives



## Output



## Parameters

$$\begin{aligned} \arg \min_{D, X} & \|Y - DX\|_F^2 + \alpha Tr(D^T L D) \\ & + \beta Tr(X L_c X^T) \quad \text{s.t.} \quad \|x_i\|_0 \leq T \quad \forall i. \end{aligned}$$

$k$  = latent dimension size

$\alpha$  = weight of cell Laplacian

$L$  = cell Laplacian (num neighbors, metric)

$\beta$  = weight of gene Laplacian

$L_c$  = gene Laplacian (num neighbors, metric)

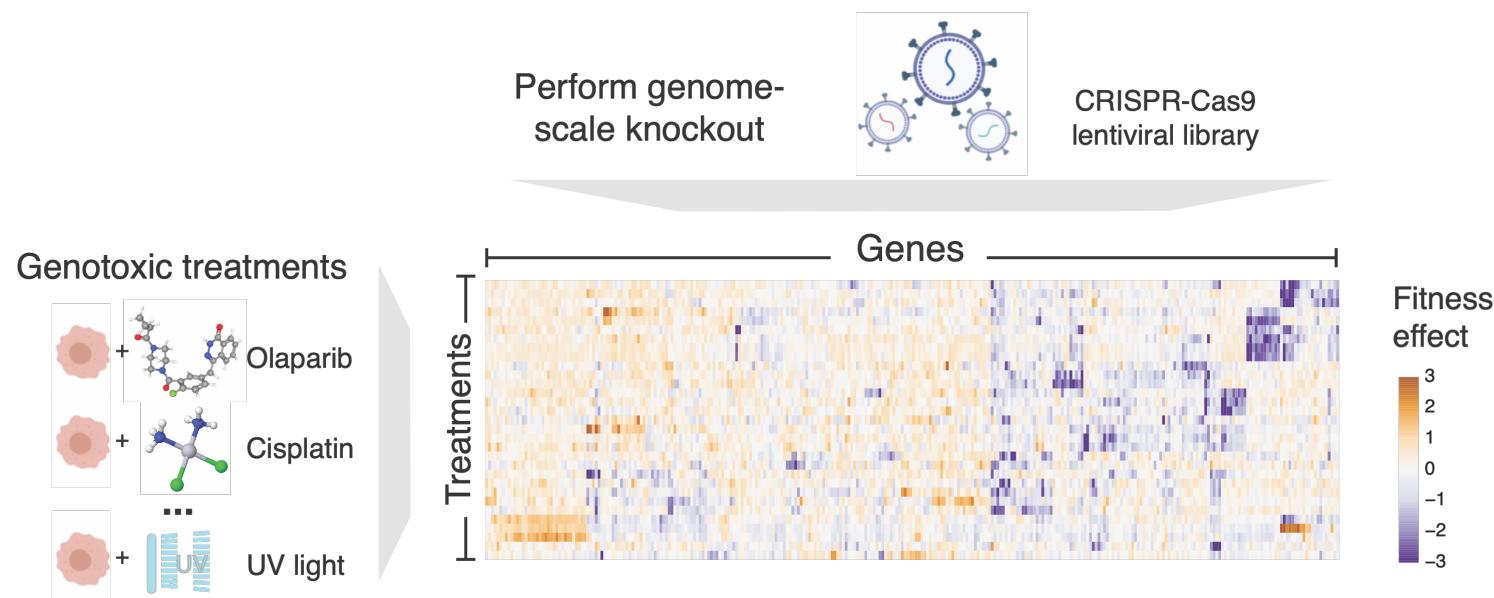
$T$  = sparsity

# Applications to three screens of gene perturbation effects

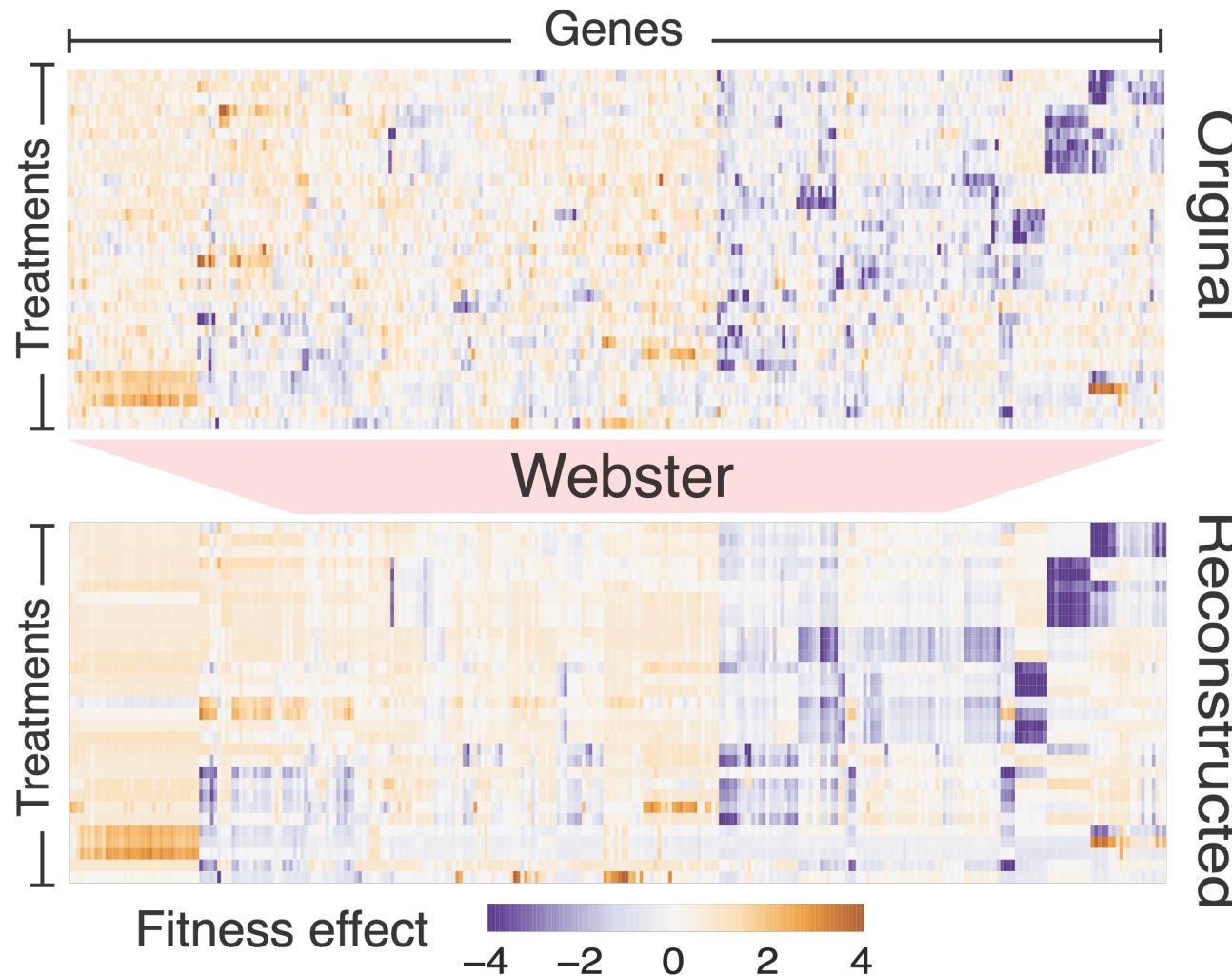
- 1) Genotoxic screens
- 2) Cancer fitness screens
- 3) Compound sensitivity screens

# Part 1: Genotoxic screens

Olivieri et al. 2020: fitness effect of gene knockout in presence of genotoxins

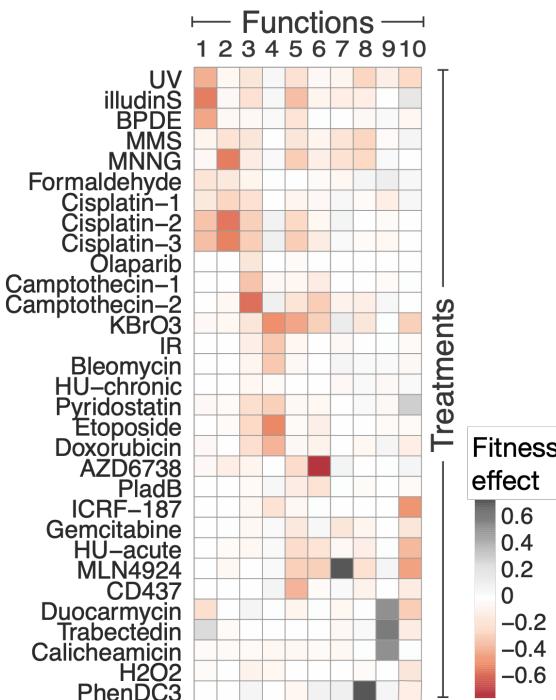


# Webster approximates the input data matrix...

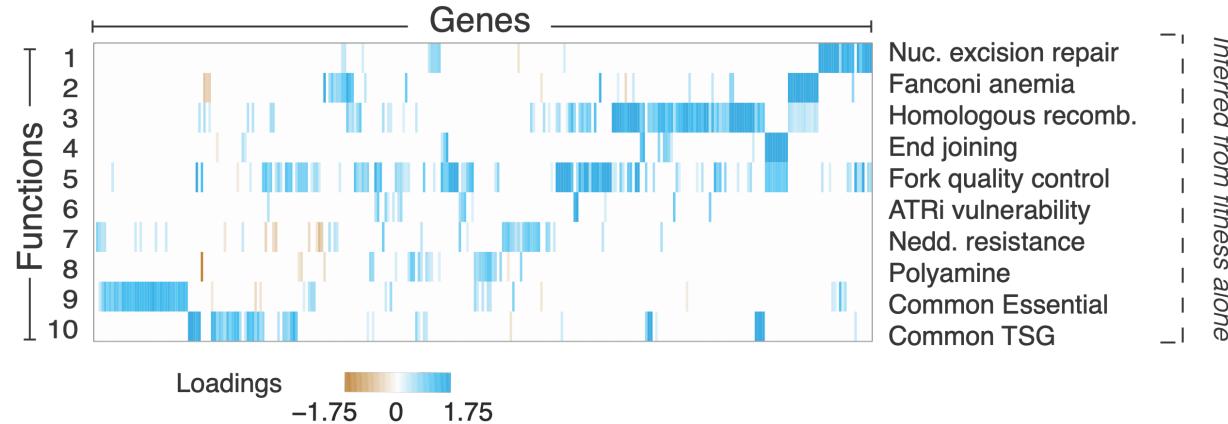


# ... as a product between a dictionary matrix and a loadings matrix

Dictionary matrix



Gene-to-function loadings

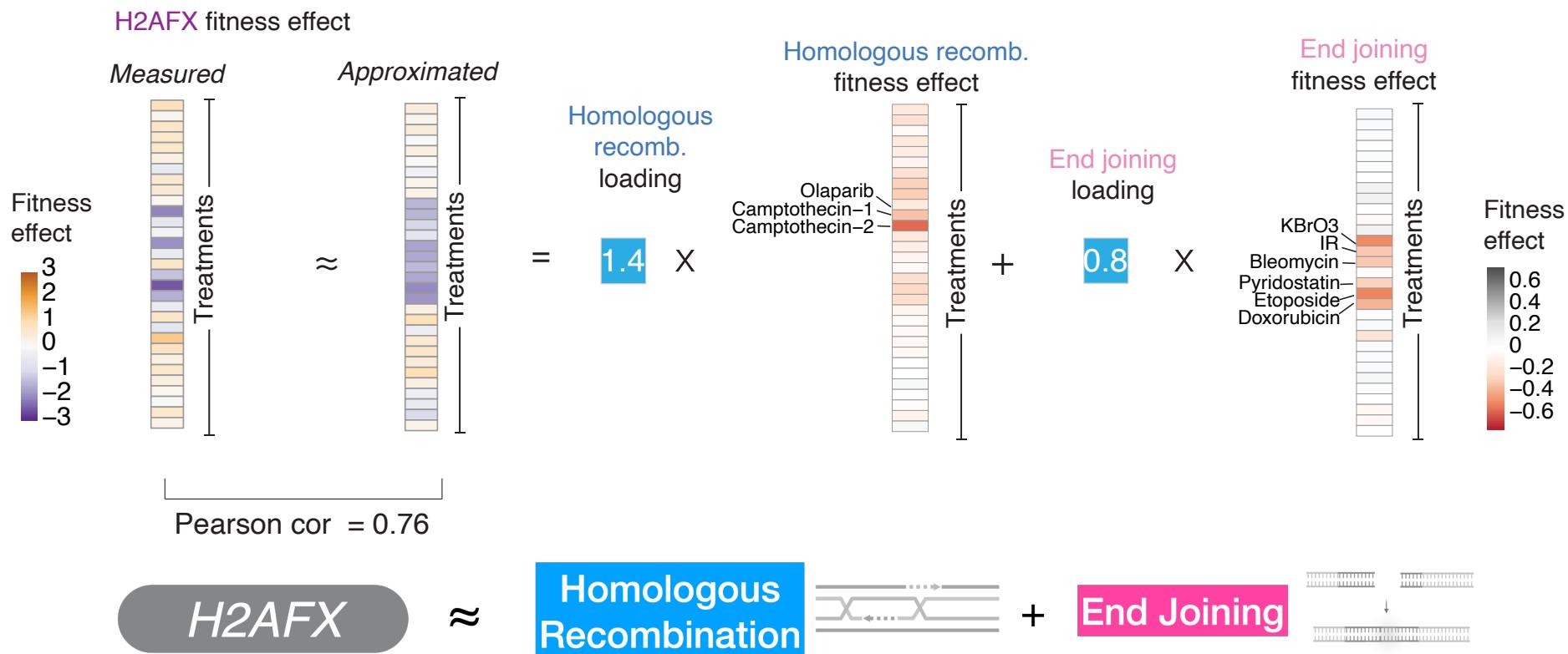


Literature annotations



Learned gene-to-function loadings recover  
biological genesets hidden during model training

# Latents inferred by the model recapitulate pleiotropy *without prior knowledge*

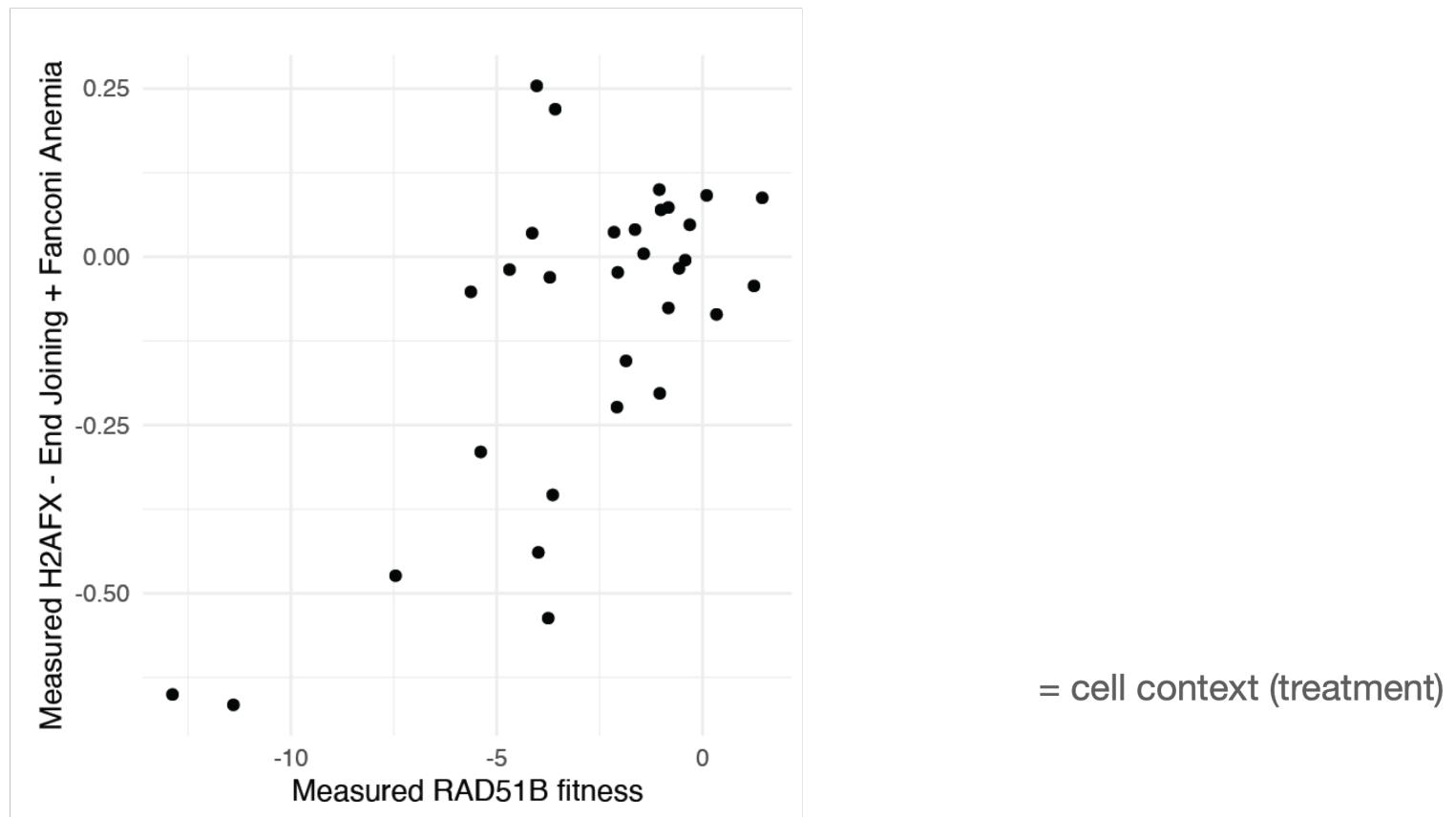


(hidden during model training!)

# Latents are biologically meaningful

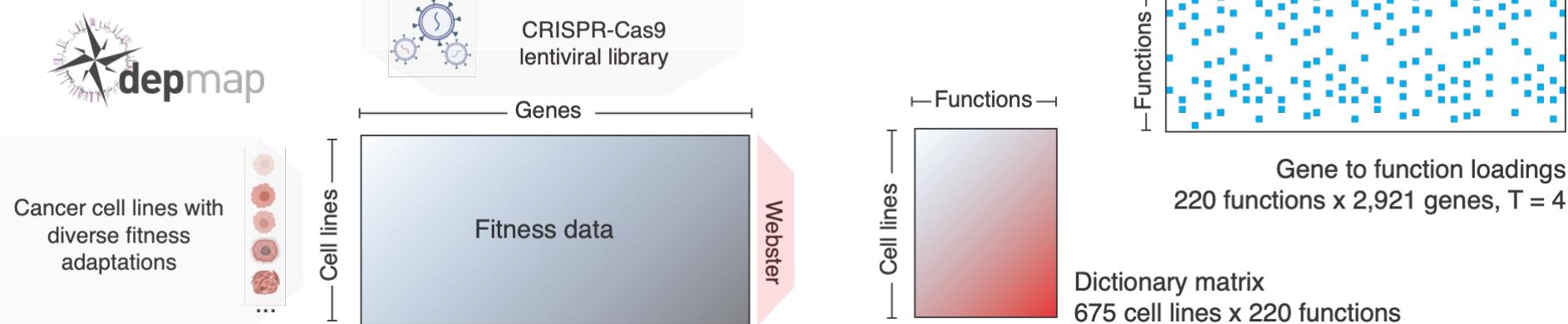
$$geneA - func1 + func2 \approx geneB$$

**H2AFX - End Joining + Fanconi Anemia  $\approx$  RAD51B**



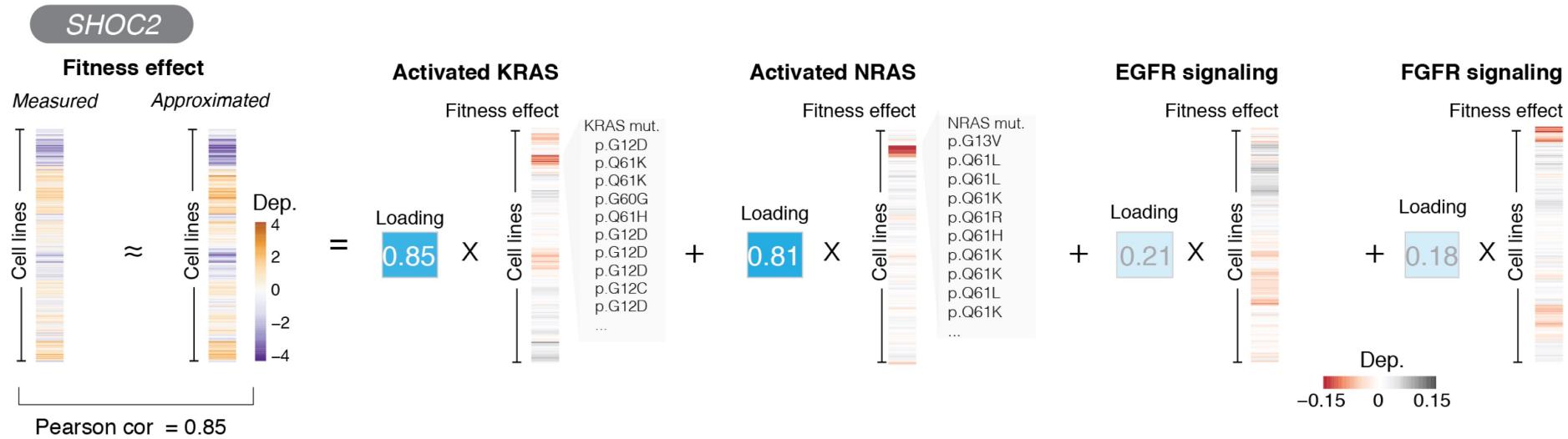
= cell context (treatment)

# Part 2: Cancer fitness screens

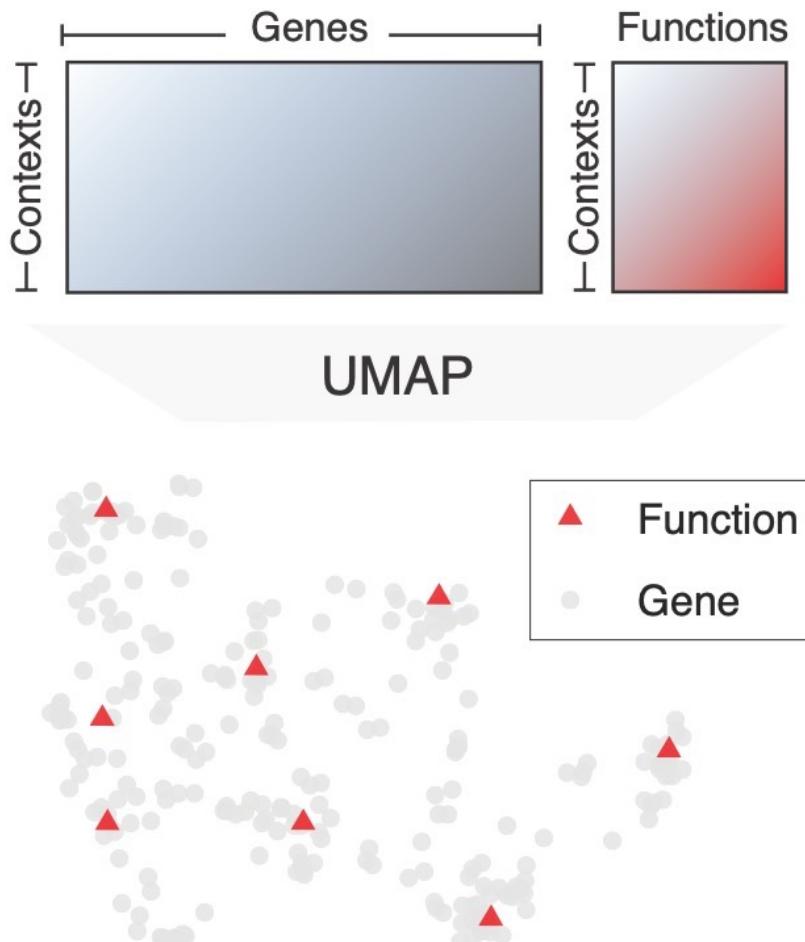


# Pleiotropic genes obey linear semantics in the latent space

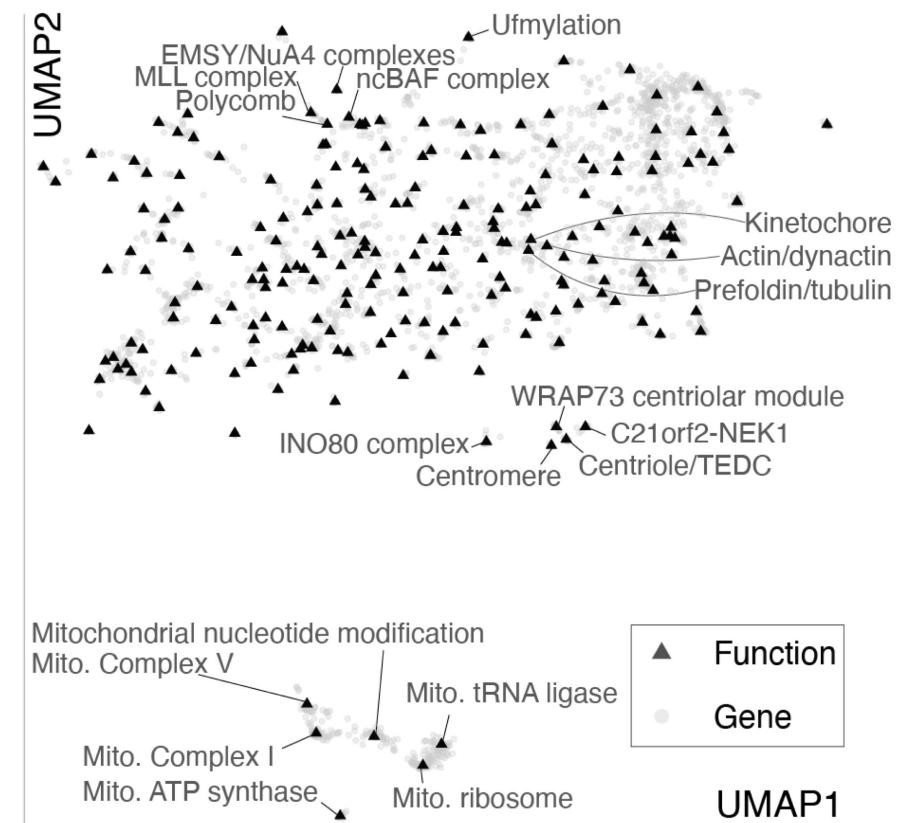
**$SHOC2 \approx \text{Activated KRAS} + \text{Activated NRAS} + \text{EGFR Signaling} + \text{FGFR Signaling}$**



# Joint embedding space of genes and functions

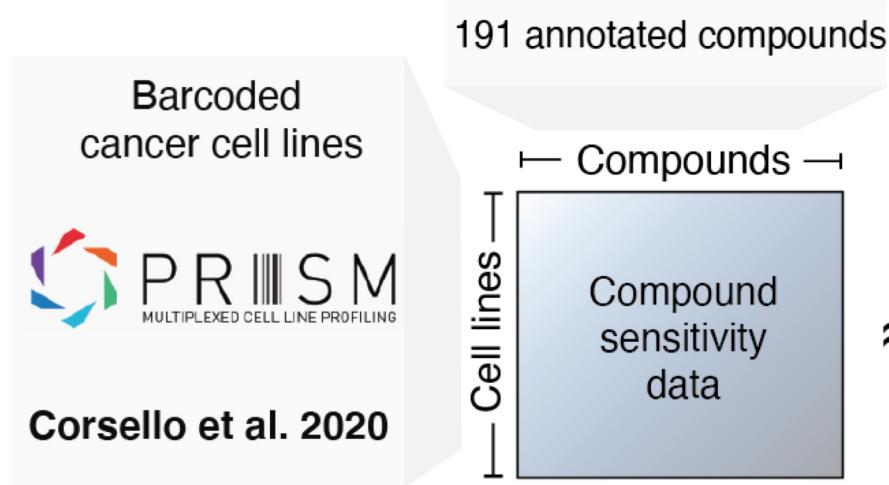


It captures interpretable processes in cancer

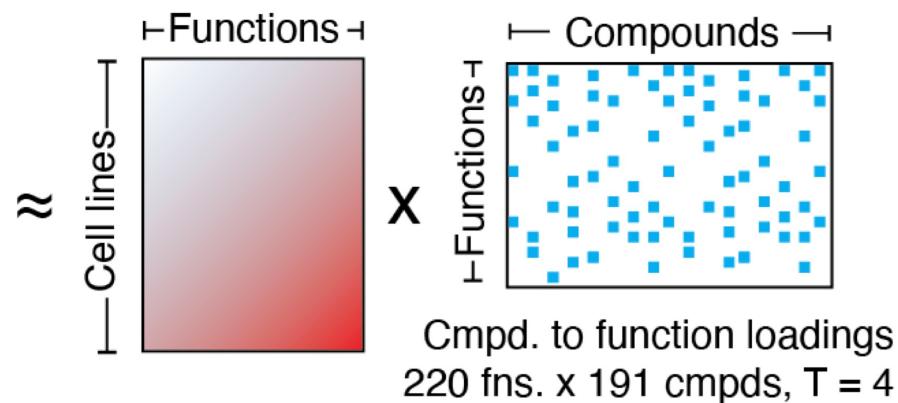


# Part 3: Compound sensitivity screens

**Query:** Drug Repurposing dataset



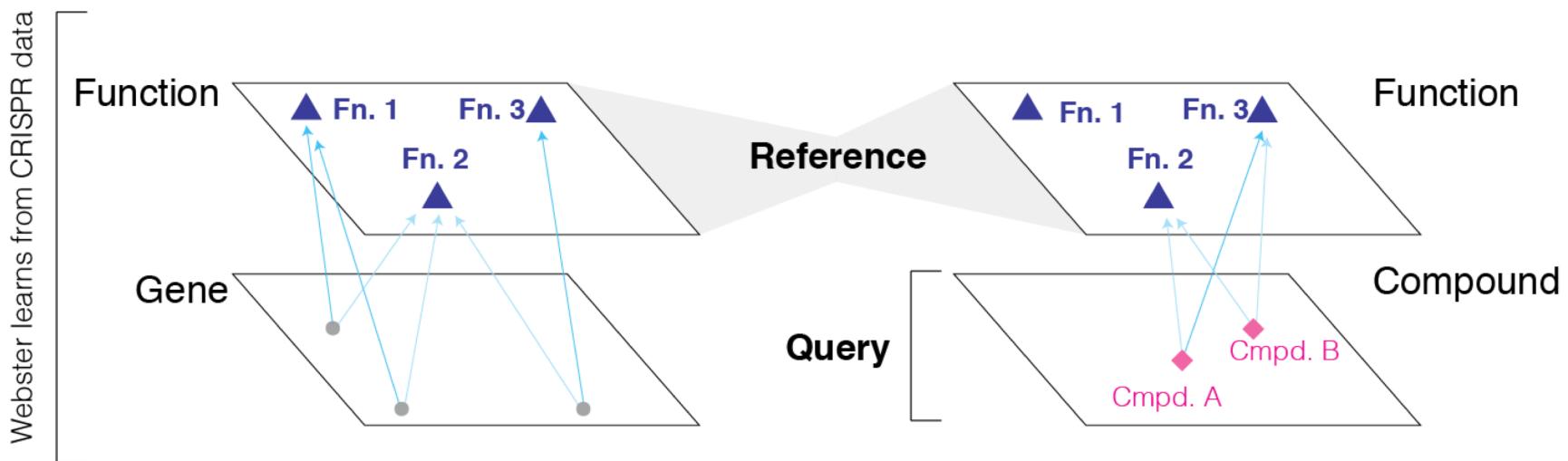
**Reference:**  
CRISPR dictionary



Modeling compound sensitivity profiles as mixtures of functions learned from CRISPR

# Modeling compounds as mixtures of latent functions

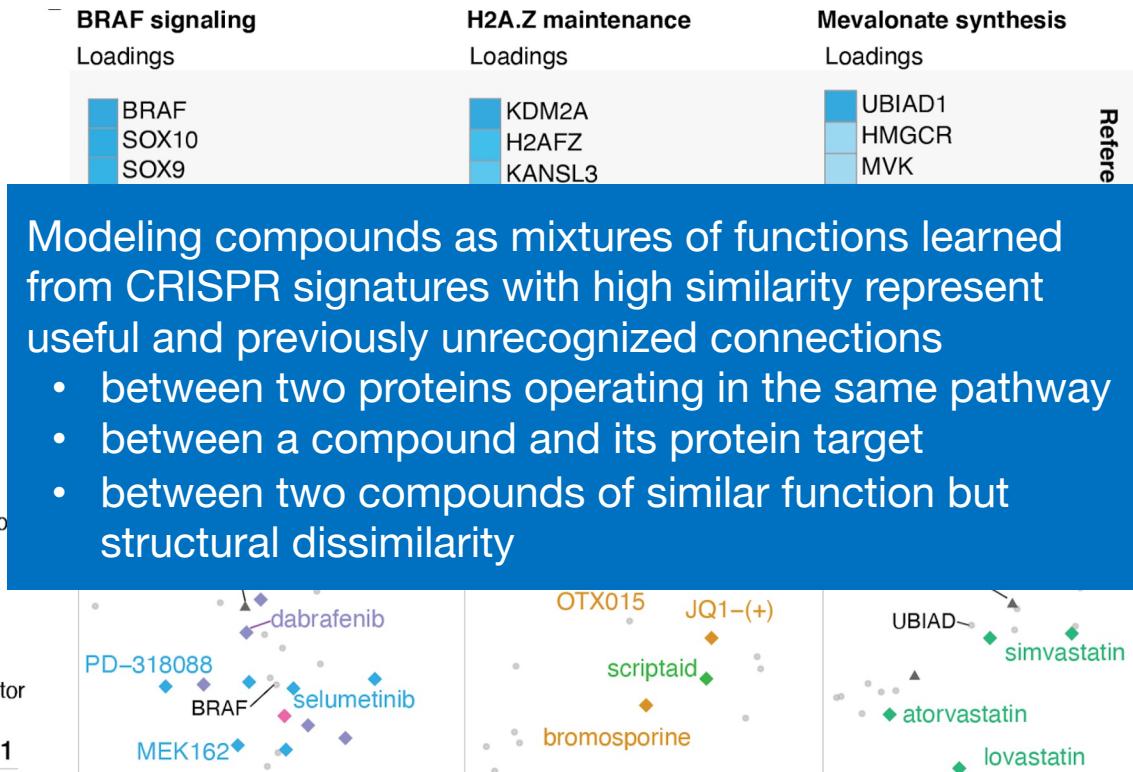
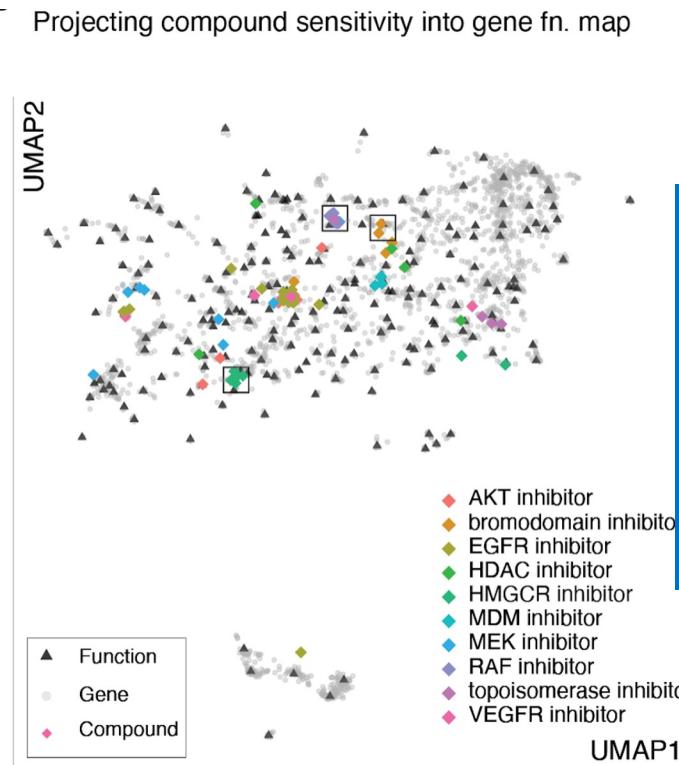
Reference-query projection



- Modeling compounds as mixtures of functions learned from CRISPR signatures with high similarity represent useful and previously unrecognized connections
  - between two proteins operating in the same pathway
  - between a small-molecule and its protein target
  - between two small-molecules of similar function but structural dissimilarity
- Such a catalog of connections can serve as a functional look-up table of compounds to predict sensitivity and genotoxic profiles and to inform therapeutic use

# Compounds' mechanisms of action

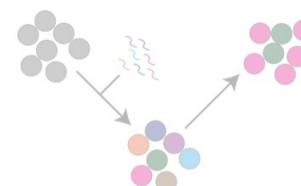
Compounds are embedded nearby gene functions, reflecting their mechanism of action



# Key takeaways

- Analogously to word semantics, genes can be modeled as **distributions over latent bio functions**
  - **Sparse learning** is an effective strategy for learning bio functions from high-dimensional chemical and genetic perturbations
  - New perturbations can be **projected** into learned space

**Data:** high-dimensional gene perturbation measurements



**Approach:** sparse approximation embeddings

$$\begin{matrix} \text{matrix } A \\ \times \\ \text{matrix } B \end{matrix}$$

$$geneA - func1 + func2 \approx geneB$$

# <https://depmap.org/webster>

Webster

Published Paper at Cell Systems    Code for paper    Dictionary learning code    Figshare data  
Design write-up

Explore relationships between genes and biological functions learned from CRISPR fitness screens using Webster.

Read The Paper: "Sparse Dictionary Learning Recovers Pleiotropy From Human Cell Fitness Screens" [For More Details.](#)

+ About this tool

Genotoxic

Select function group

ATRi vulnerability (V3)

Nedd. resistance (V5)

Polyamine (V1)

Search to select a gene or function

Selected function:  
**ATRi vulnerability (V3)**

Pan UMAP w/  
A S D  
OR  
↔ ↑ ↓  
⟳ ⟲

2d   3d  

● Functions   ● Genes   ● Gene positive association   ● Gene negative association

Native mouse controls: <=>= pan right left   ^~^= zoom

highlighted in plot  
**Gene DHX35**  
(ex. ### loading, function name)  
**1.08**  
ATRi vulnerability (V3)  
1.00  
Fork quality control (V9)  
**Approximation quality (Pearson)**  
0.74

# Outline for today's class

- High-throughput genetic and chemical perturbations
- Drug repurposing, indication and contra-indication prediction
- Generative protein design
- Generative AI agents



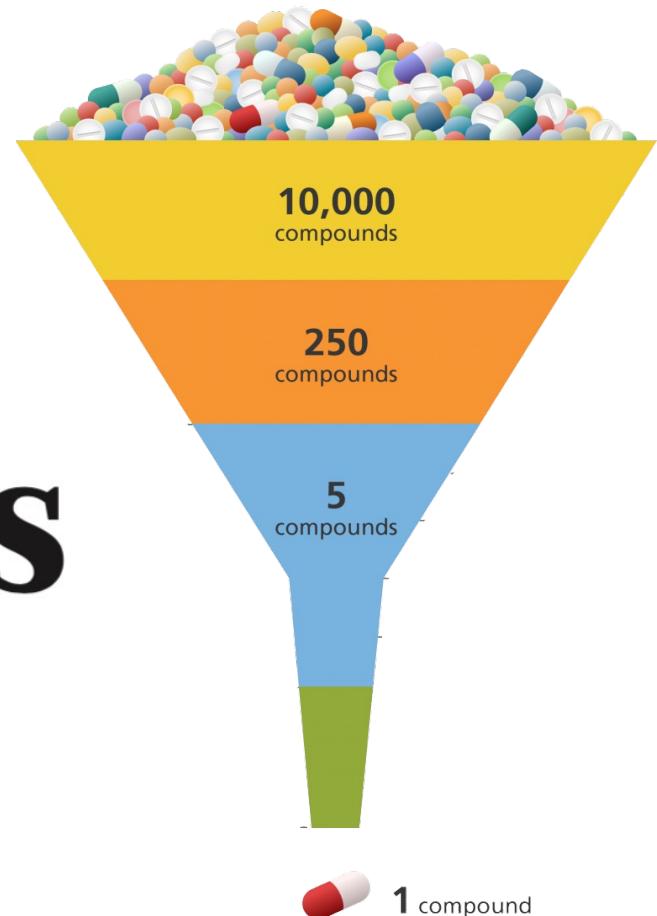
# New tricks for old drugs

*Faced with skyrocketing costs for developing new drugs, researchers are looking at ways to repurpose older ones — and even some that failed in initial trials.*



## A SHORTER TIMESCALE

Because most repositioned drugs have already passed the early phases of development and clinical testing, they can potentially win approval in less than half the time and at one-quarter of the cost.

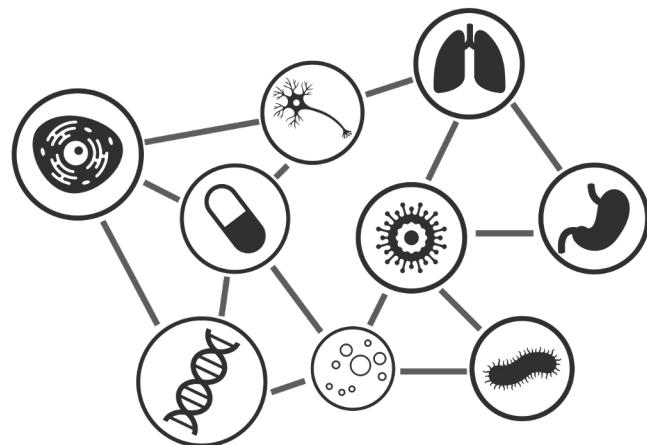


## Drug repositioning

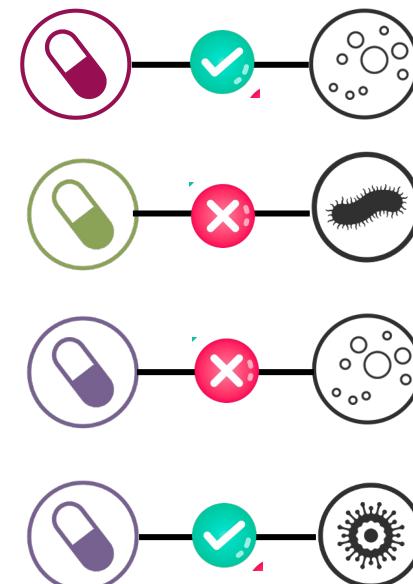
**~6 years, ~\$300 million**

# Therapeutic use prediction

Comprehensive knowledge graph  
of 17,080 clinically-recognized diseases



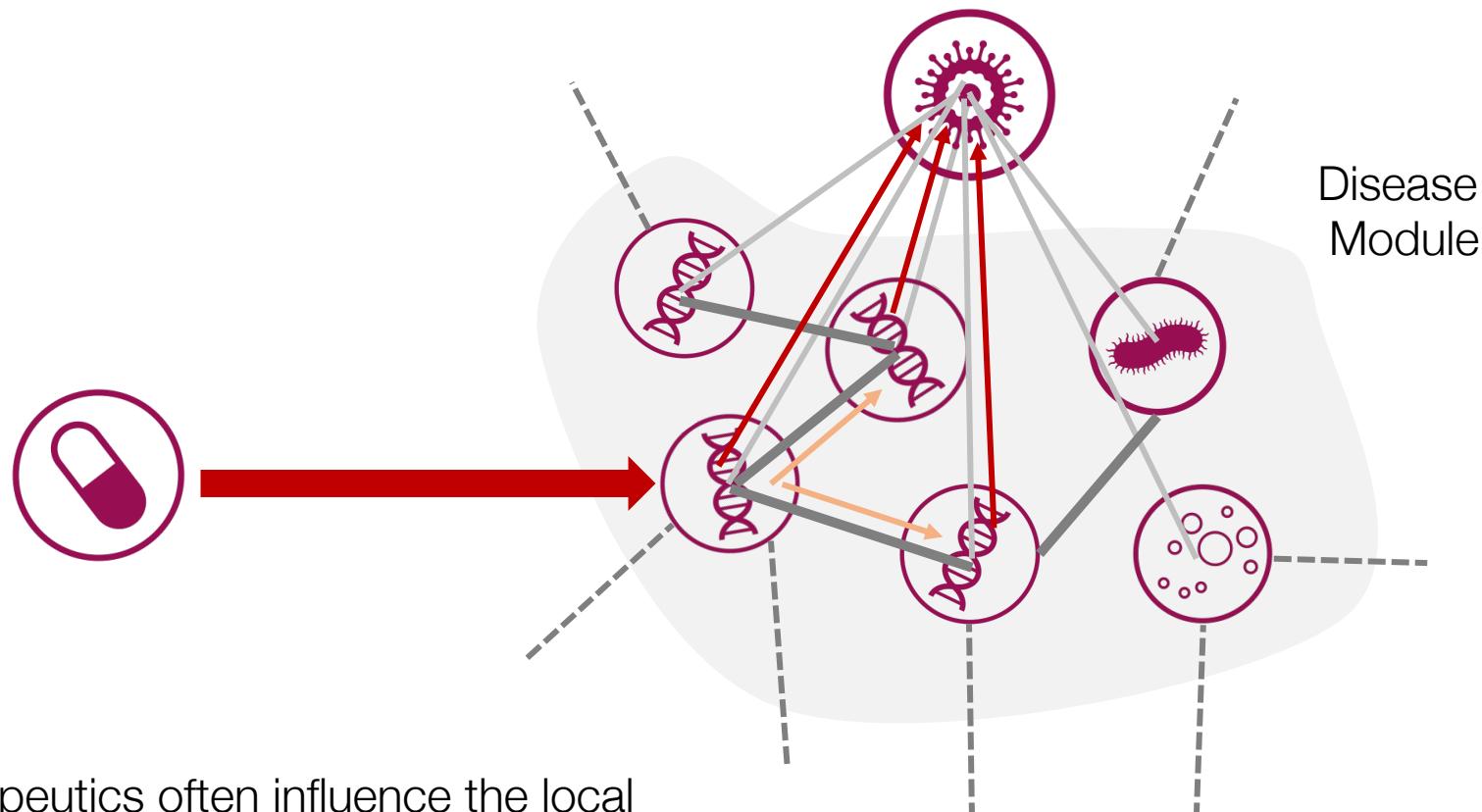
Process various therapeutic tasks, such as indication and contraindication prediction, in a unified formulation



TxGNN is a model for identifying therapeutic opportunities for diseases with limited treatment options and molecular understanding. It is a graph neural network pre-trained on a comprehensive knowledge graph of 17,080 clinically-recognized diseases and 7,957 therapeutic candidates

**Applications:**  
Drug repurposing/virtual screening  
Understanding disease mechanisms  
Understanding treatment effects

# TxGNN: Mechanistic view of therapeutic effects

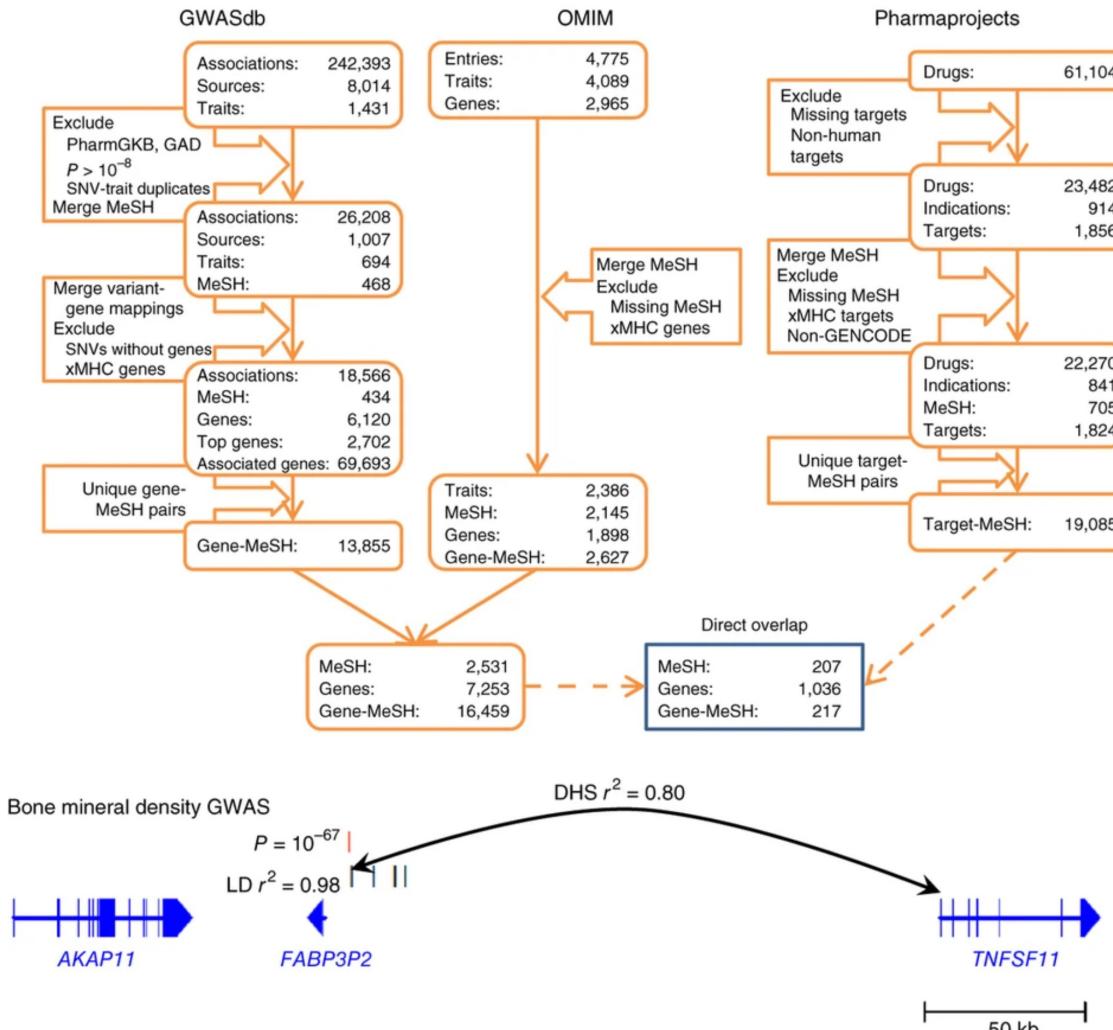


Therapeutics often influence the local biological system of disease-associated agents to create therapeutic effects

# TxGNN: Mechanistic view of therapeutic effects

- Growing insight into genes that influence human disease may affect how drug targets and indications are selected
- Questions:
  - How well the current archive of genetic evidence predicts drug mechanisms?
  - Can using the growing wealth of human genetic data to select the best targets and indications have a measurable impact on the successful development of new drugs?

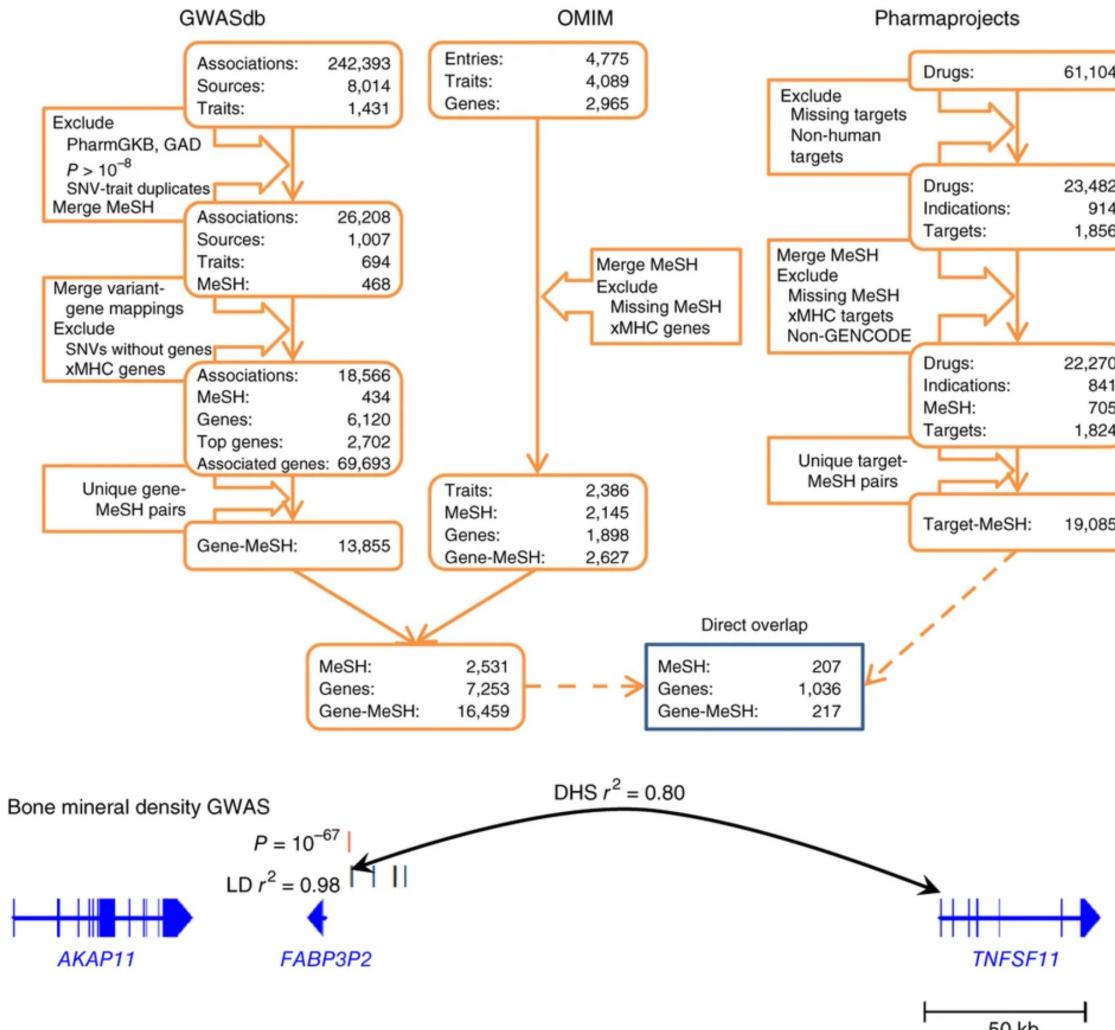
# TxGNN: Mechanistic view of therapeutic effects



Summary of each data resource and the key filtering and processing steps applied to create the final set of gene-trait and drug target-indication combinations investigated in this study. GWASdb sources correspond to unique PubMed IDs or other unique data sources given for each association. GAD, Genetic Association Database

Approach to mapping genetically associated variants to genes. Example illustrated with the bone mineral density GWAS association with rs9533090 (depicted in red). Of five SNPs in strong LD with rs9533090 ( $r^2 \geq 0.8$ ), one falls within a DNase I-hypersensitive site (DHS) that was found to have a signal correlated with the DHS of the TNFSF11 gene transcription start site

# TxGNN: Mechanistic view of therapeutic effects

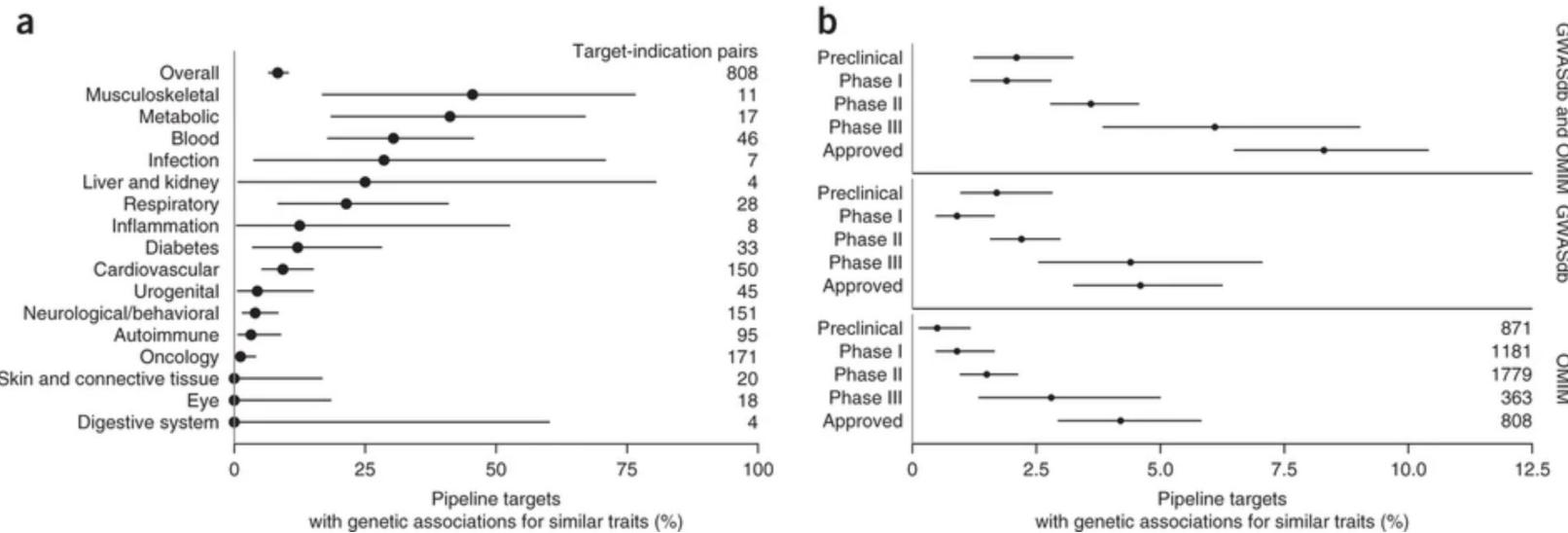


Summary of each data resource and the key filtering and processing steps applied to create the final set of gene-trait and drug target-indication combinations investigated in this study. GWASdb sources correspond to unique PubMed IDs or other unique data sources given for each association. GAD, Genetic Association Database

Approach to mapping genetically associated variants to genes. Example illustrated with the bone mineral density GWAS association with rs9533090 (depicted in red). Of five SNPs in strong LD with rs9533090 ( $r^2 \geq 0.8$ ), one falls within a DNase I-hypersensitive site (DHS) that was found to have a signal correlated with the DHS of the TNFSF11 gene transcription start site

# TxGNN: Mechanistic view of therapeutic effects

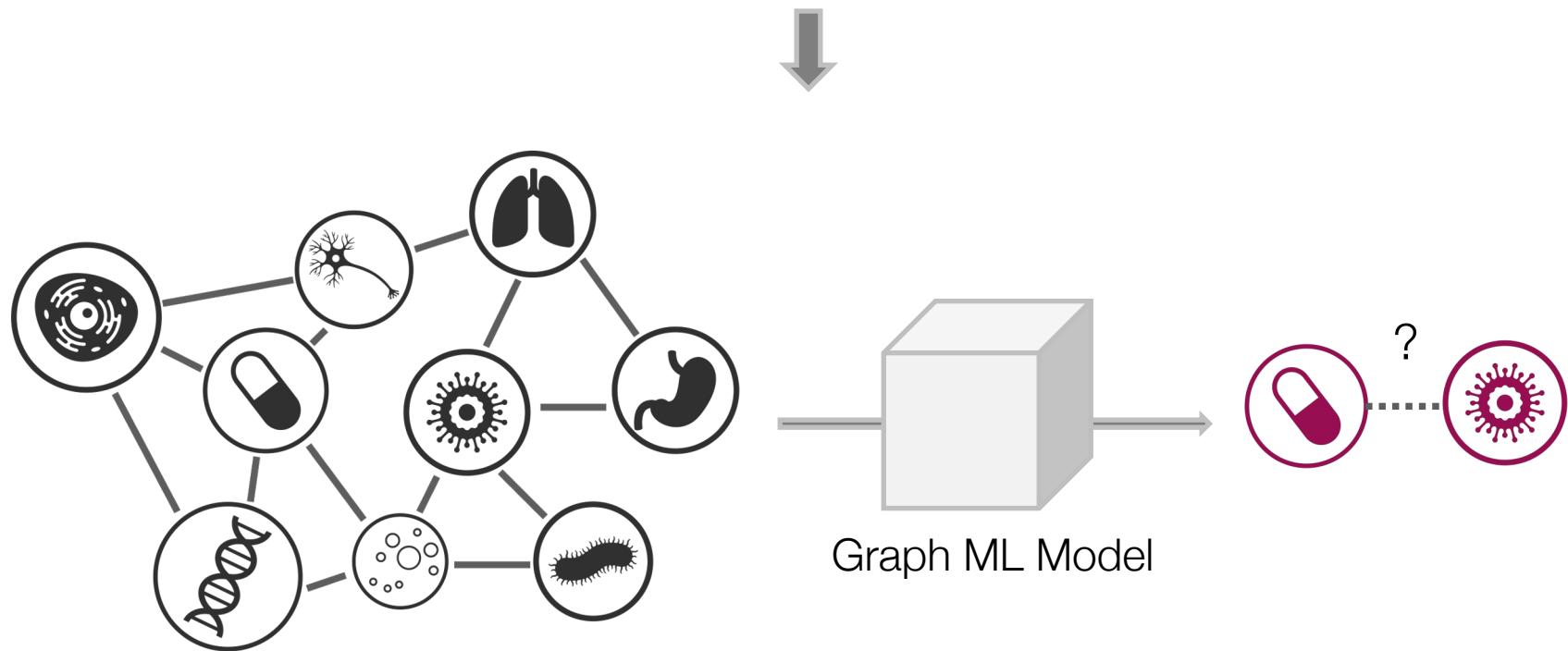
Percentage of target-indication pairs for drugs approved in the United States or the European Union overlapping with gene-trait combinations



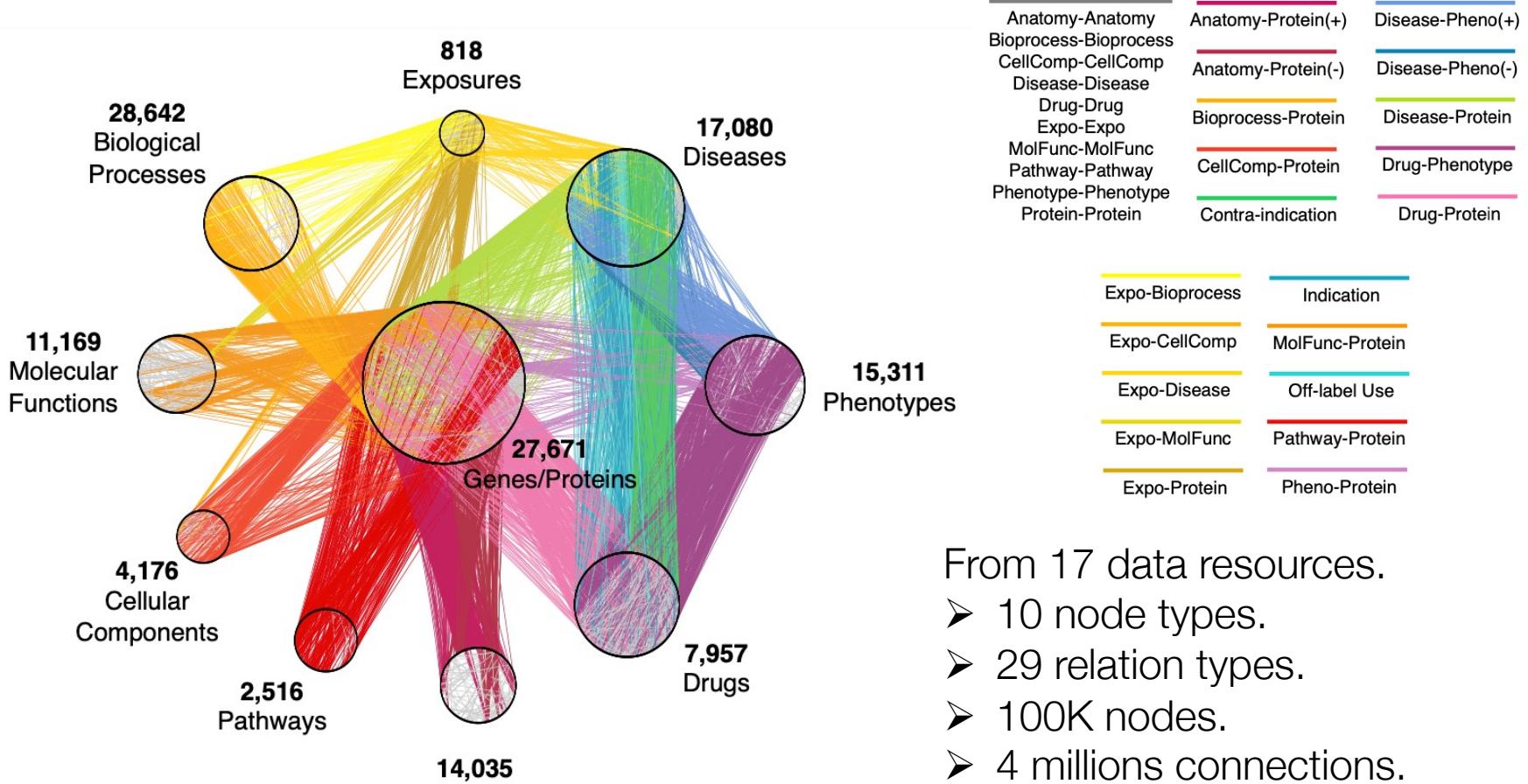
- Among well-studied indications, the proportion of drug mechanisms with direct genetic support **increases significantly across the drug development pipeline, from 2.0% at the preclinical stage to 8.2% among mechanisms for approved drugs**, and varies dramatically among disease areas.
- Selecting genetically supported targets could double the success rate in clinical development

# TxGNN

To model this mechanistic view, we need to ground the model in known mechanisms of diseases and treatment effect

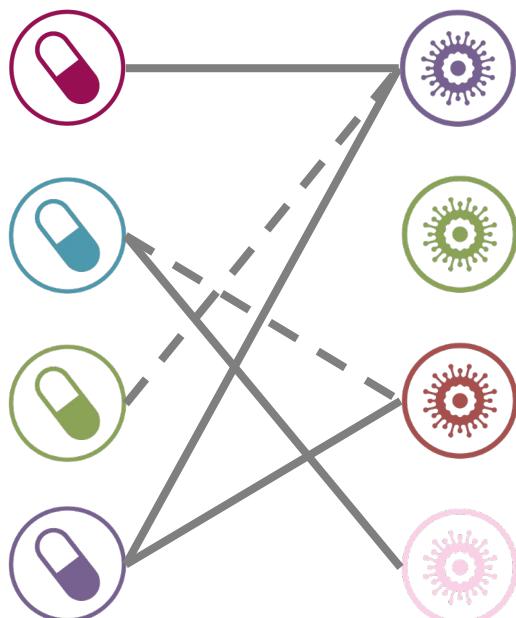


# Dataset: PrimeKG

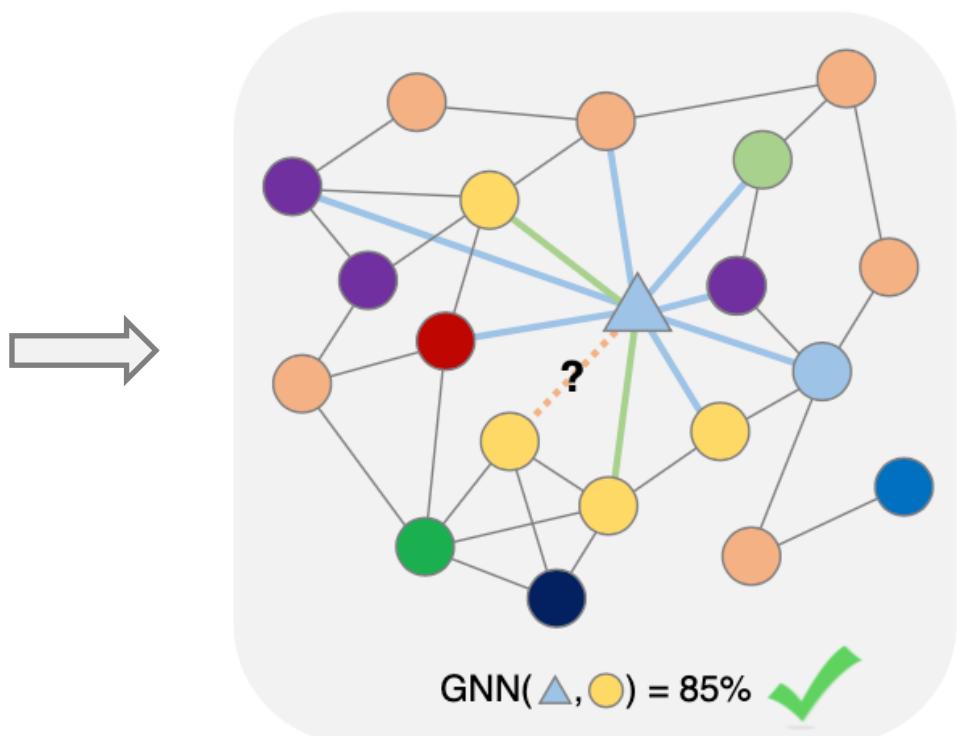
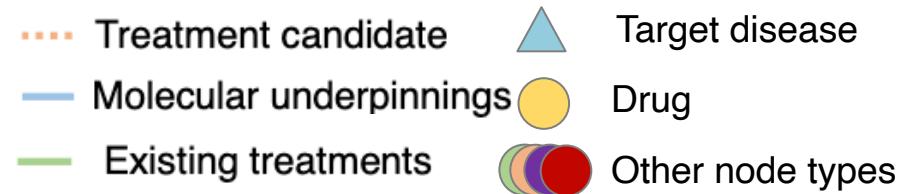


# Seeting: Baseline approach

Random split across known drug-disease pairs



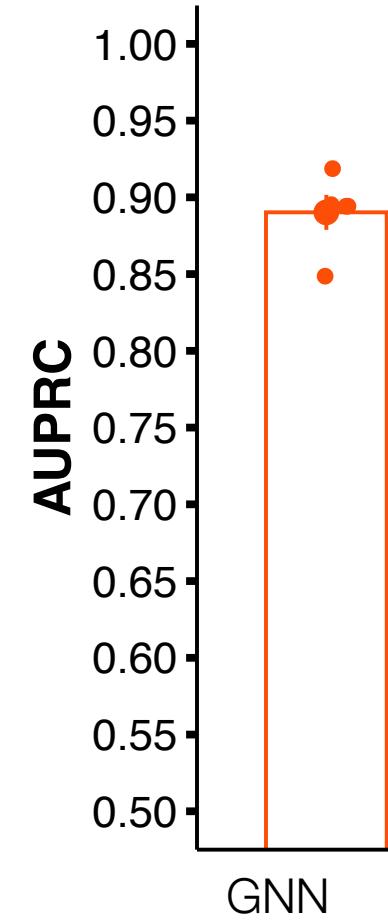
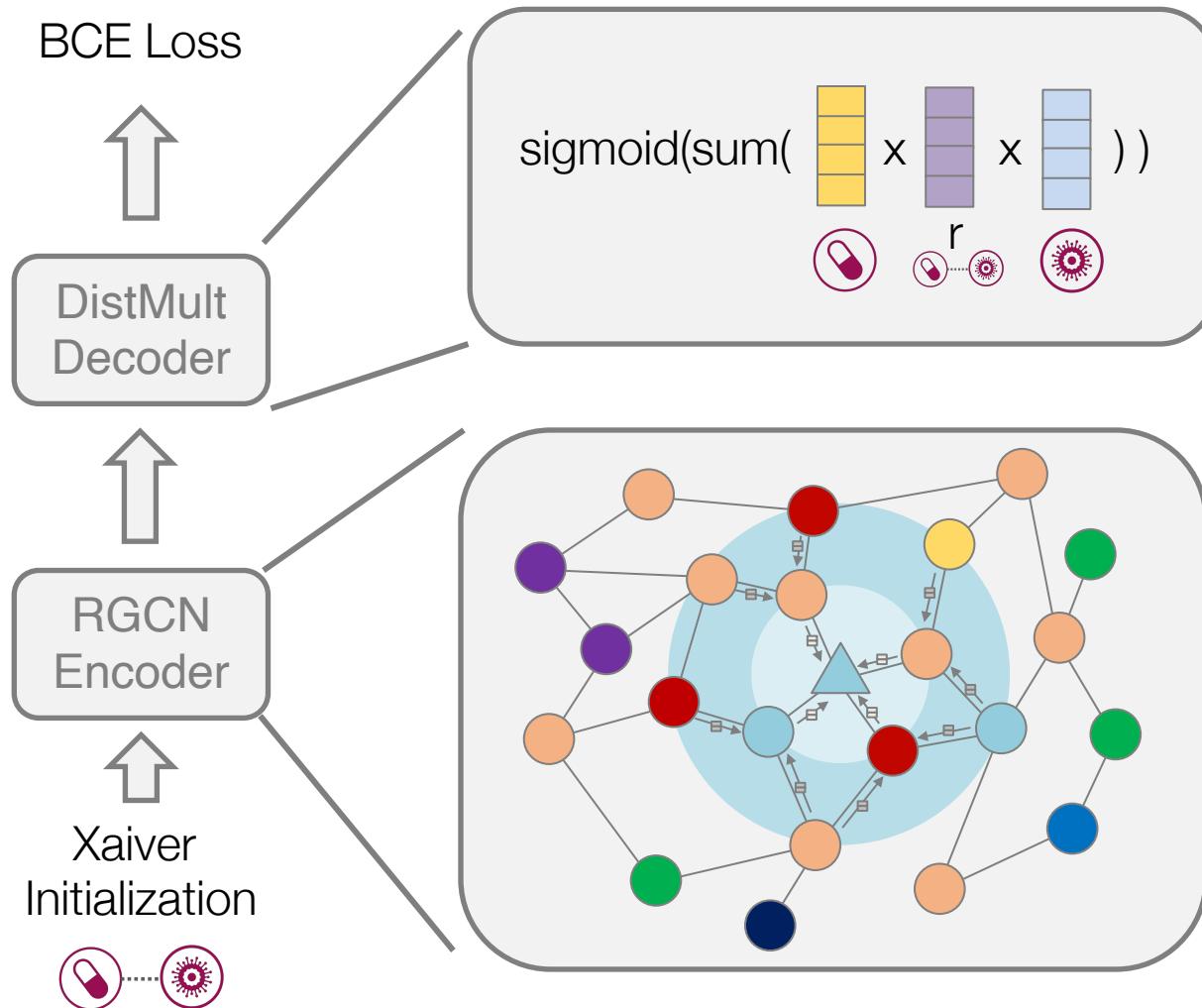
- Train Drug-Disease Pair
- - Test Drug-Disease Pair



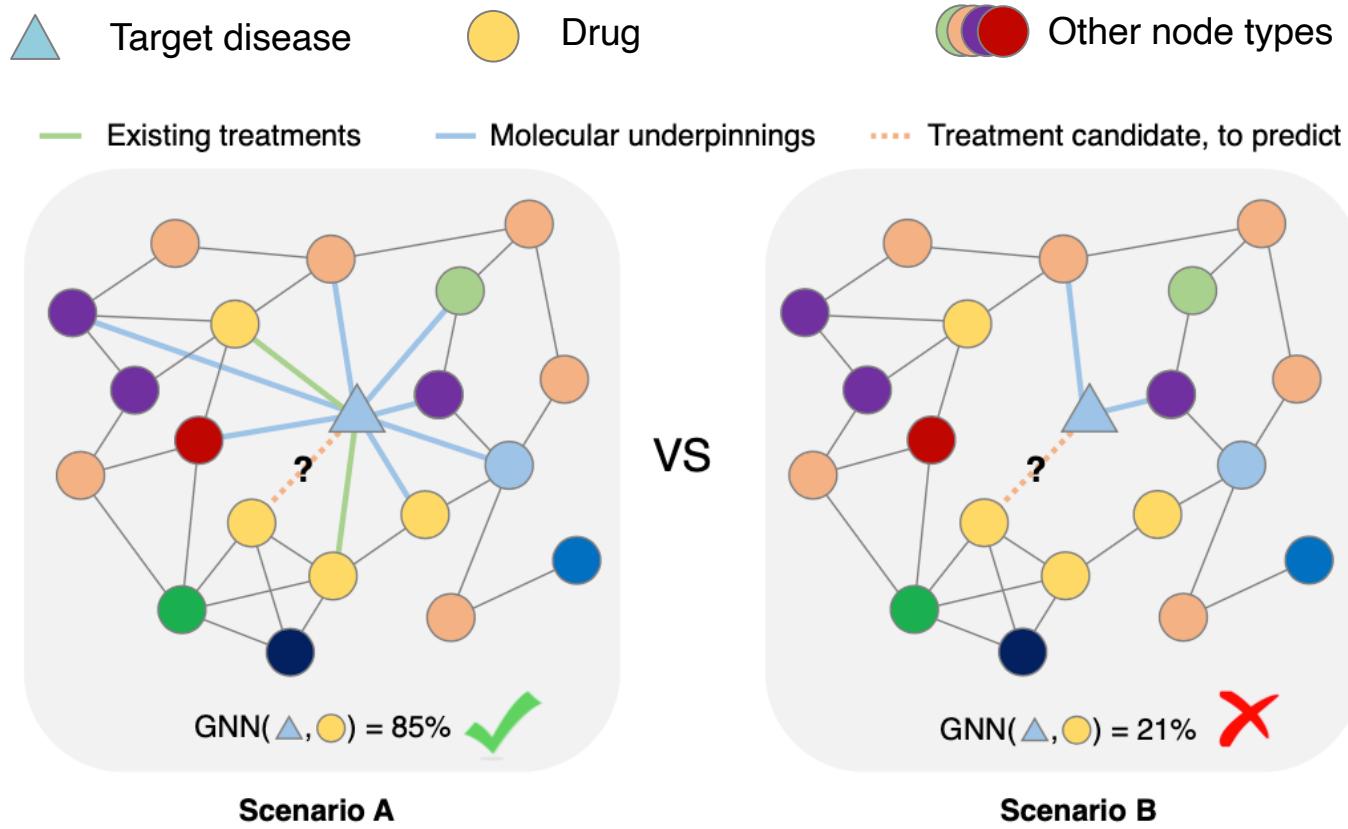
## Scenario A

- Many known treatments
- Rich molecular underpinnings

# In this setting, existing methods perform well

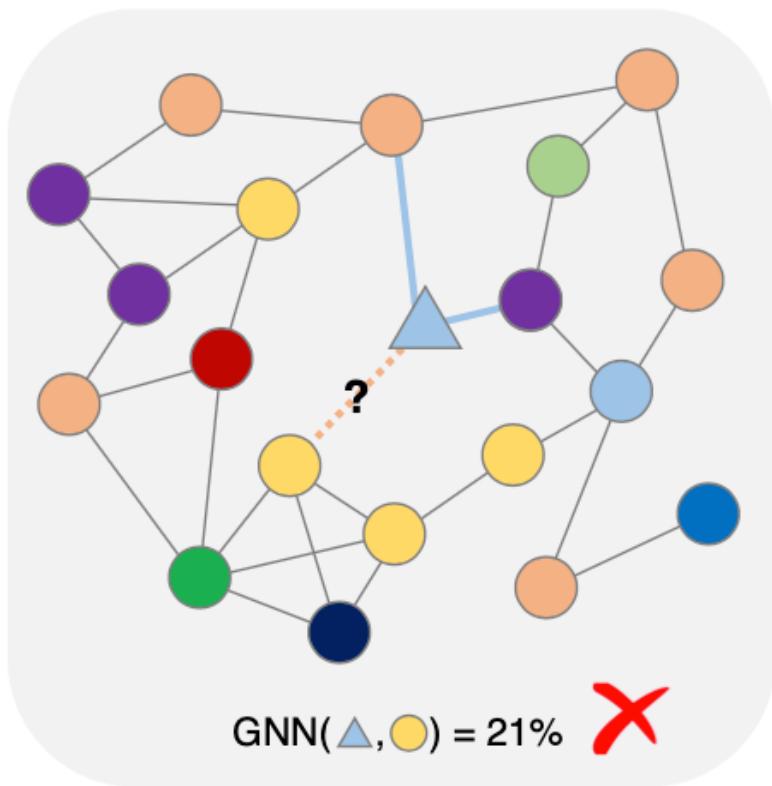


# How about other settings?



No treatments = No links between disease and any drug nodes  
Poorly characterized mechanisms = Sparse local neighborhoods

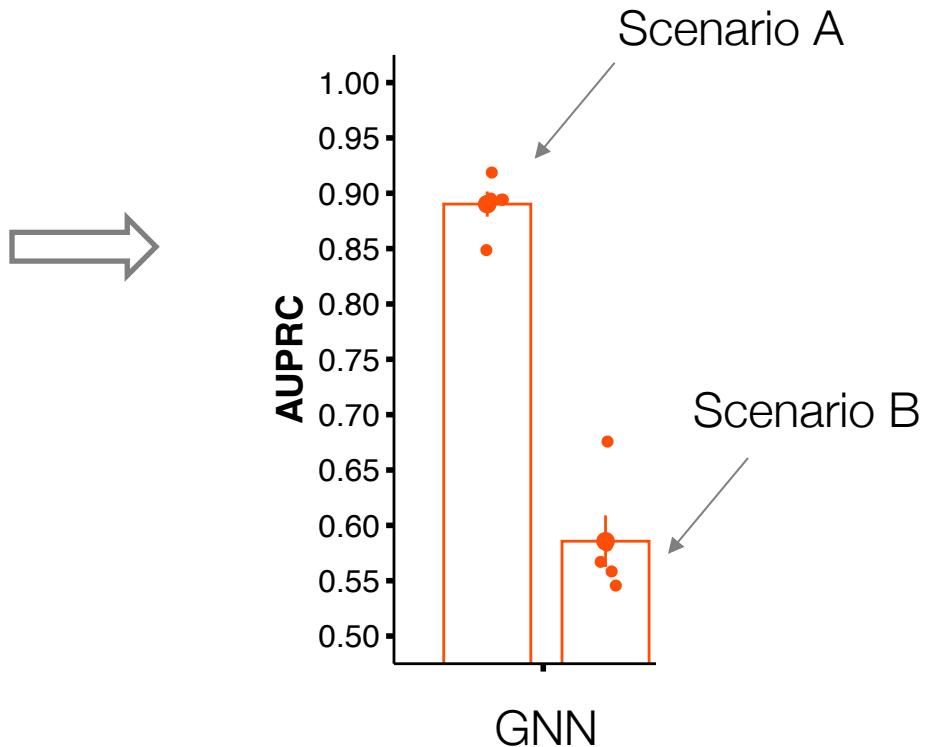
# Performance in other settings



**Scenario B**

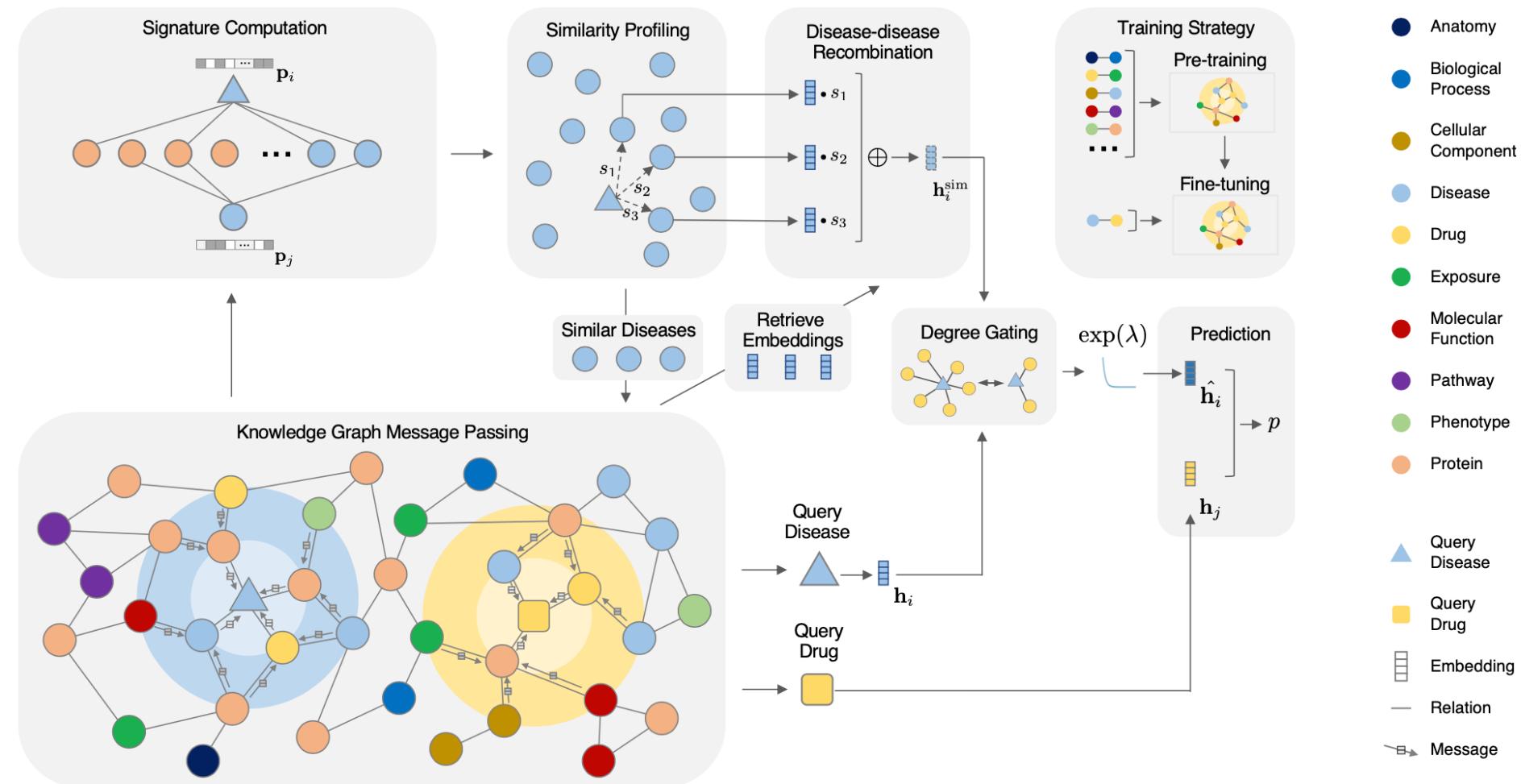
- No existing treatments
- Poorly characterized mechanisms
- Challenging to predict

Disease embeddings are less meaningful because so many relationships are unknown

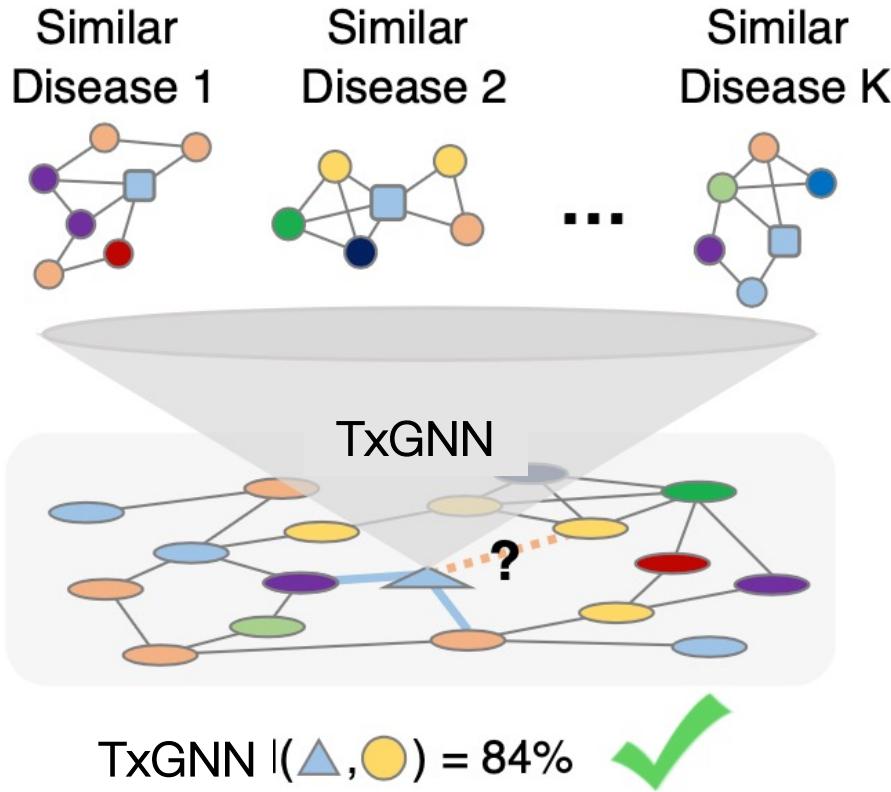


Need better disease embeddings -- Is there an inductive bias (biological rationale) that can be incorporated into the ML model?

# Approach: TxGNN



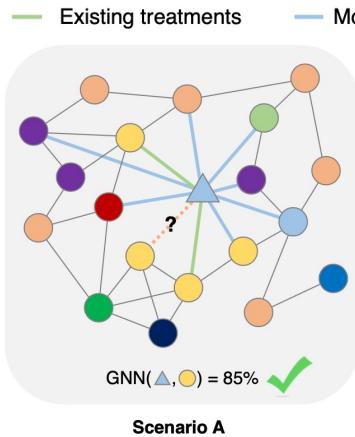
# TxGNN: Transfer learning across diseases



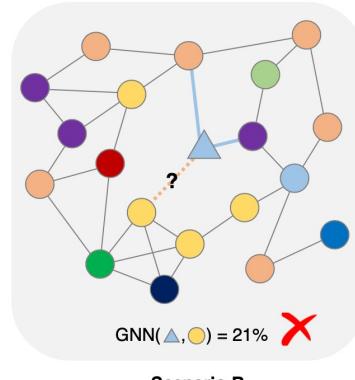
- (1) identify similar diseases
- (2) leverage disease similarities

# Results: Therapeutic use prediction

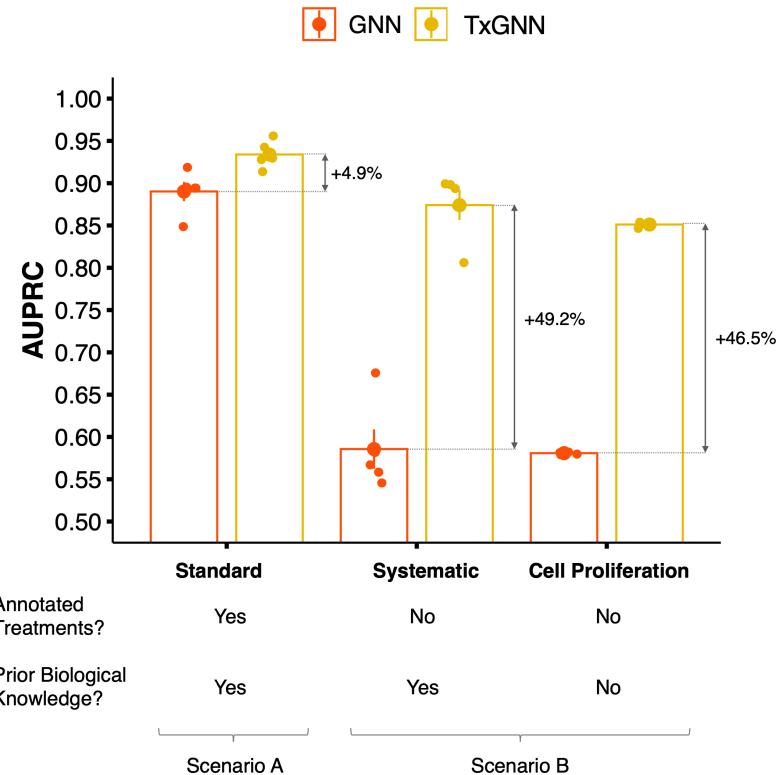
- Once trained, TXGNN can perform zero-shot inference on new diseases without additional parameters or fine-tuning on ground truth labels



- Many known treatments
- Known molecular understanding
- “Easy” to predict

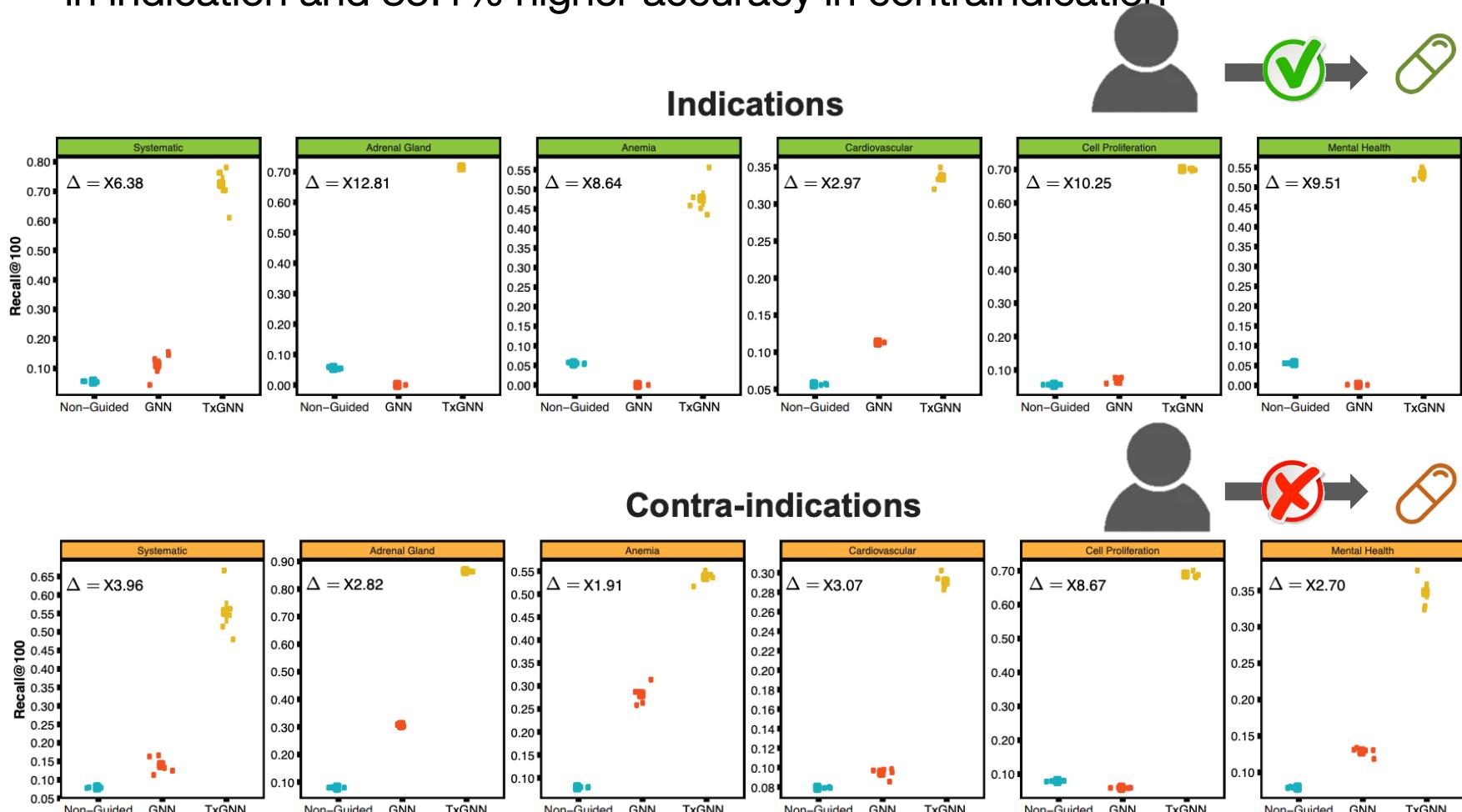


- No known treatments
- Poor molecular understanding
- “Hard” to predict



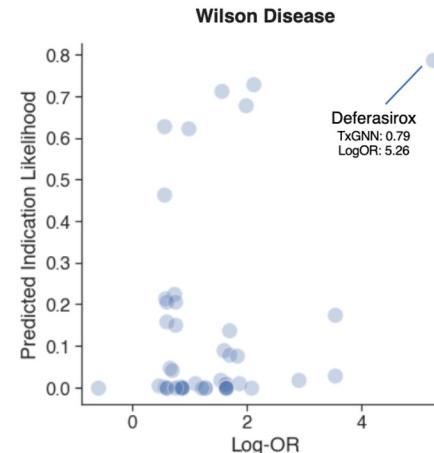
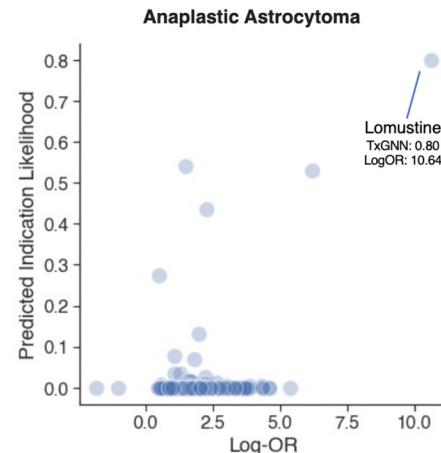
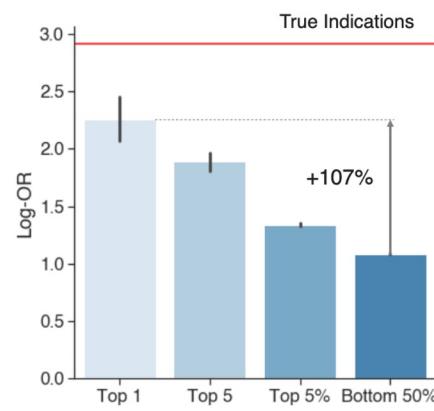
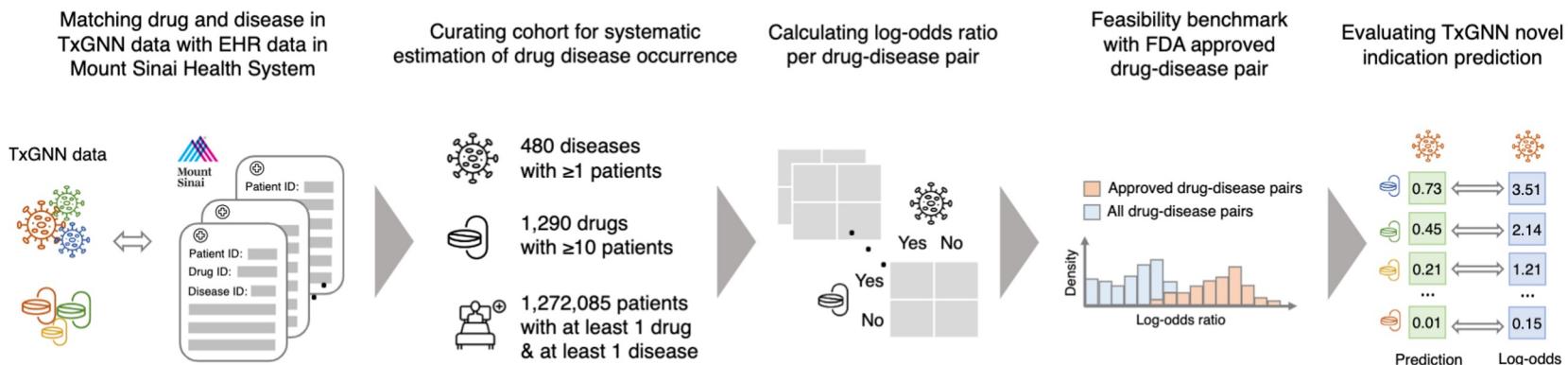
# Results: Therapeutic use prediction

- TxGNN improves over existing methods, with up to 49.2% higher accuracy in indication and 35.1% higher accuracy in contraindication



# Results: Therapeutic use prediction

- TxGNN's novel predictions are consistent with off-label prescription decisions made by clinicians in a large healthcare system



# Results: Therapeutic use prediction

- TxGNN can also predict therapeutic use for recent FDA approvals

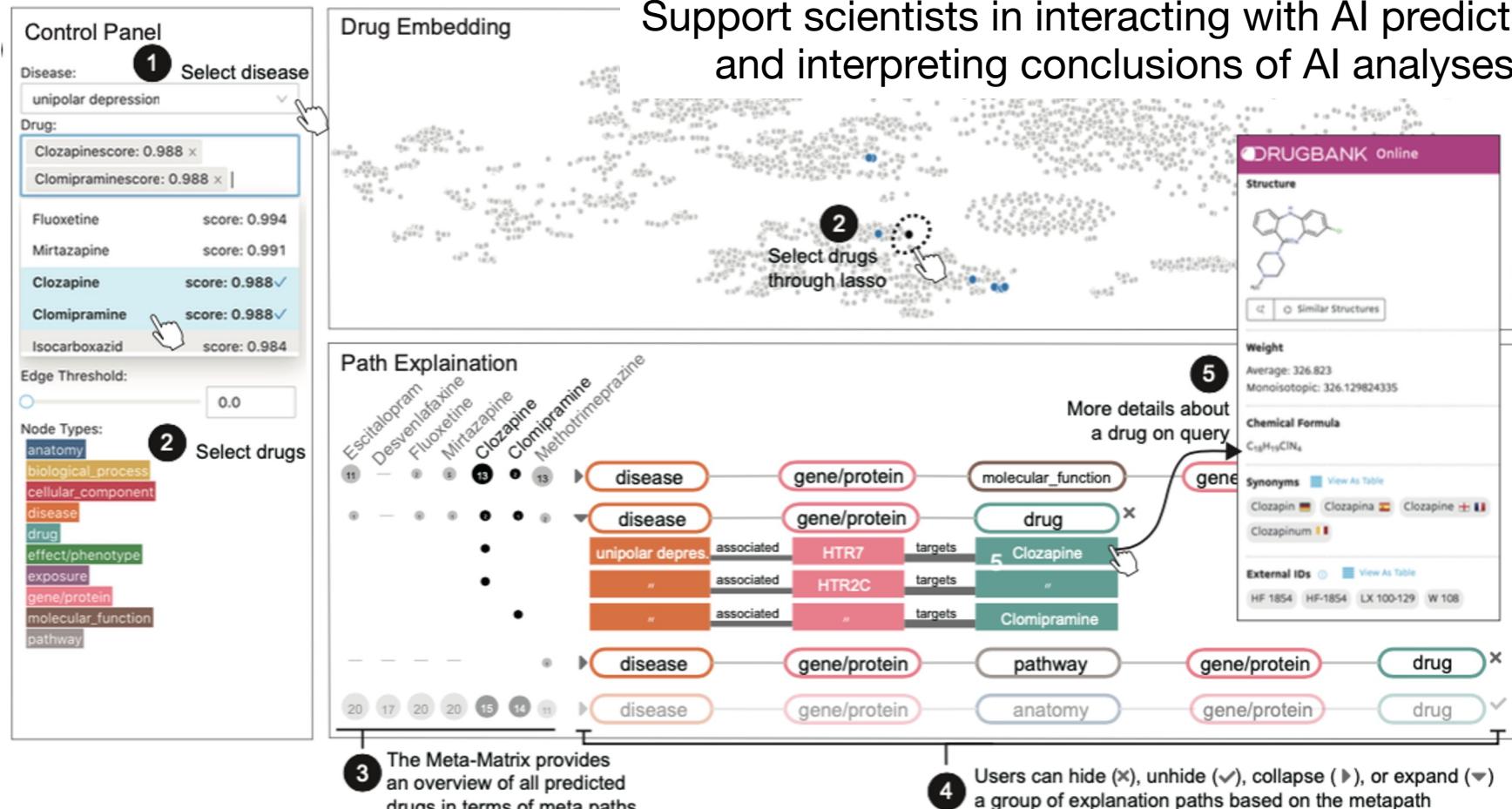
Drug name	Ingredient	Disease	Approval date	Company	FDA Number	Orphan	Prediction	Percentile
Welireg	Belzutifan	von Hippel-Lindau disease	08/13/2021	Merck	NDA215383	Yes	0.720	4.11%
Livtency	Maribavir	Cytomegalovirus infection	11/23/2021	Takeda	NDA215596	Yes	0.033	66.37%
Tezspire	Tezepelumab-Ekko	Asthma	12/17/2021	AstraZeneca	BLA761224	No	0.233	32.41%
Leqvio	Inclisiran Sodium	Familial hypercholesterolemia	12/22/2021	Novartis	NDA214012	No	0.301	19.32%
Adbry	Tralokinumab	Atopic dermatitis	12/27/2021	Leo Pharma	BLA761180	No	0.040	50.37%
Vabysmo	Faricimab-Svoa	Macular degeneration	01/28/2022	Genentech	BLA761235	No	0.938	2.25%
Vonjo	Pacritinib Citrate	Myelofibrosis	02/28/2022	Cti Biopharma	NDA208712	Yes	0.011	63.14%
Ztalmy	Ganaxolone	CDKL5 disorder	03/18/2022	Marinus	NDA215904	Yes	0.335	18.73%
Mounjaro	Tirzepatide	Type 2 diabetes mellitus	05/13/2022	Eli Lilly	NDA215866	No	0.286	12.50%
Vtama	Tapinarof	Psoriasis	05/23/2022	Dermavant	NDA215272	No	0.261	32.70%

# AI-clinician collaboration

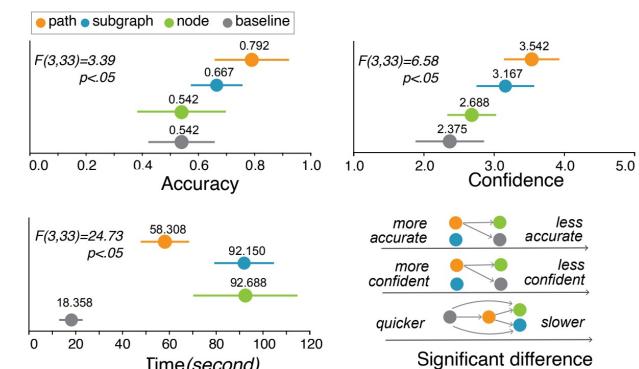
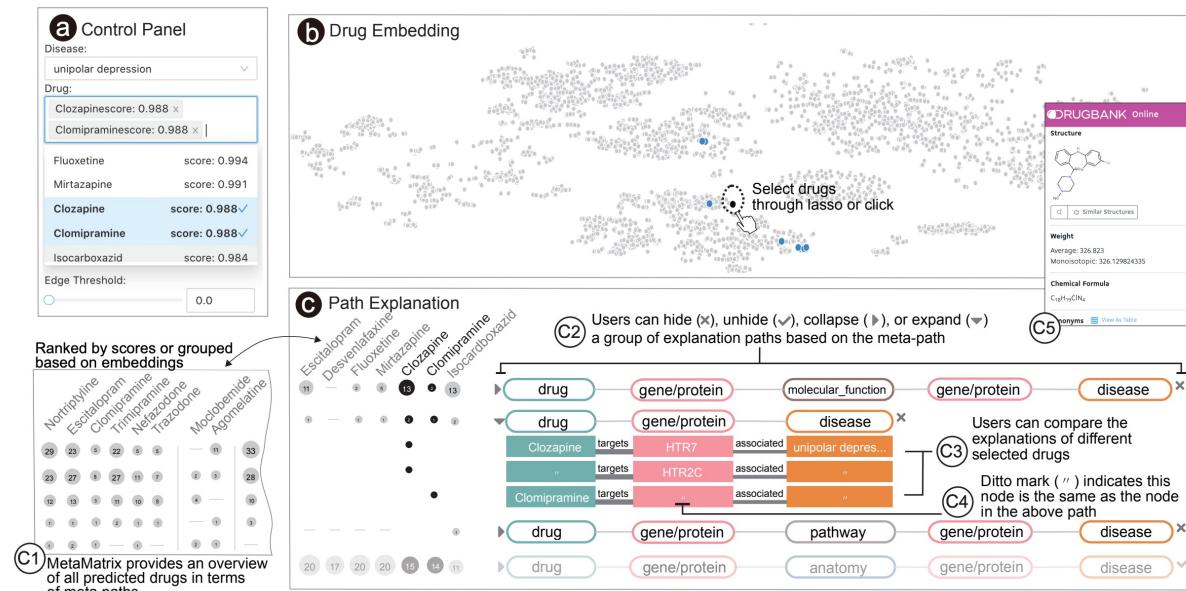
"Will clozapine treat unipolar depression? What is the disease treatment mechanism?"



Support scientists in interacting with AI predictions and interpreting conclusions of AI analyses



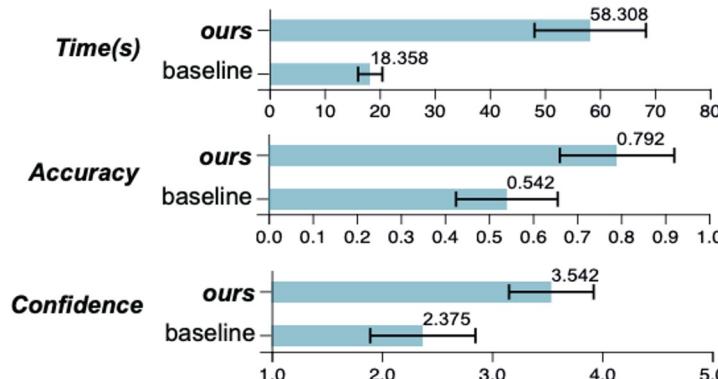
# Clinician-centered AI design



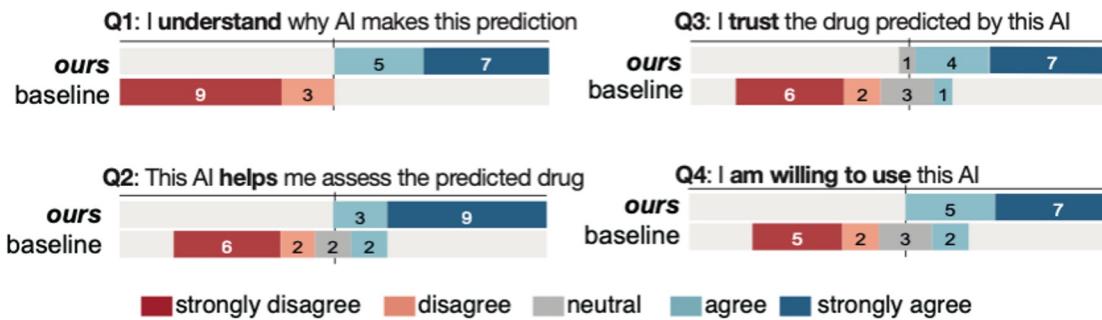
Zero-shot prediction of therapeutic use with geometric deep learning and clinician centered design, medRxiv, 2023  
 Probing GNN Explainers: A Rigorous Theoretical and Empirical Analysis of GNN Explanation Methods, AISTATS 2022  
 Extending the Nested Model for User-Centric XAI: A Design Study on GNN-based Drug Repurposing, IEEE VIS 2022 (Best Paper Award)  
 Identification of Disease Treatment Mechanisms through the Multiscale Interactome, Nature Communications 2021

# Usability study with end users

Compared to a no-explanation baseline in terms of user answer accuracy, exploration time, user confidence, and user agreement across a spectrum of usability questions



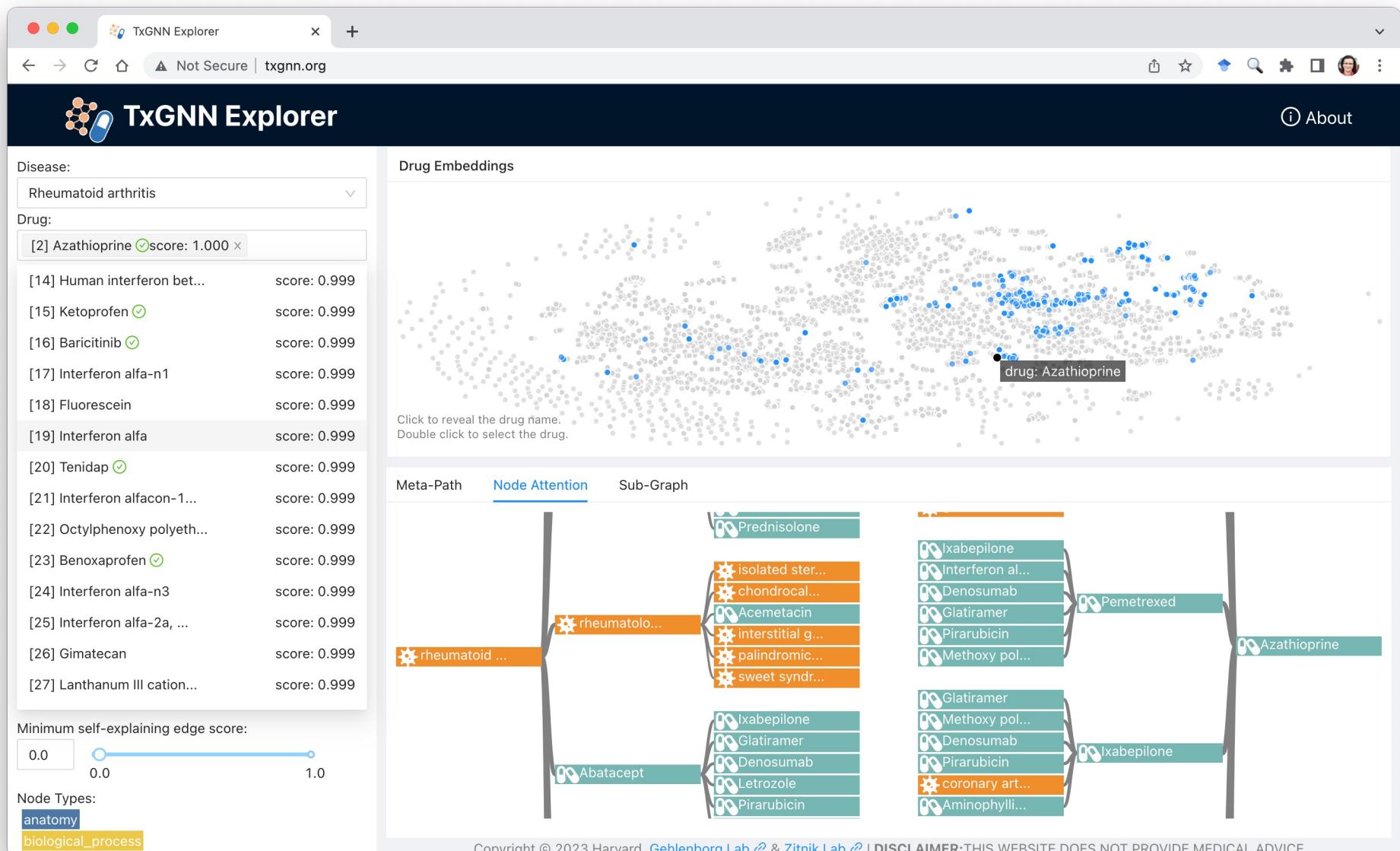
Error bars indicate the 95% confidence intervals



Agree scores are placed to the right, disagree to the left



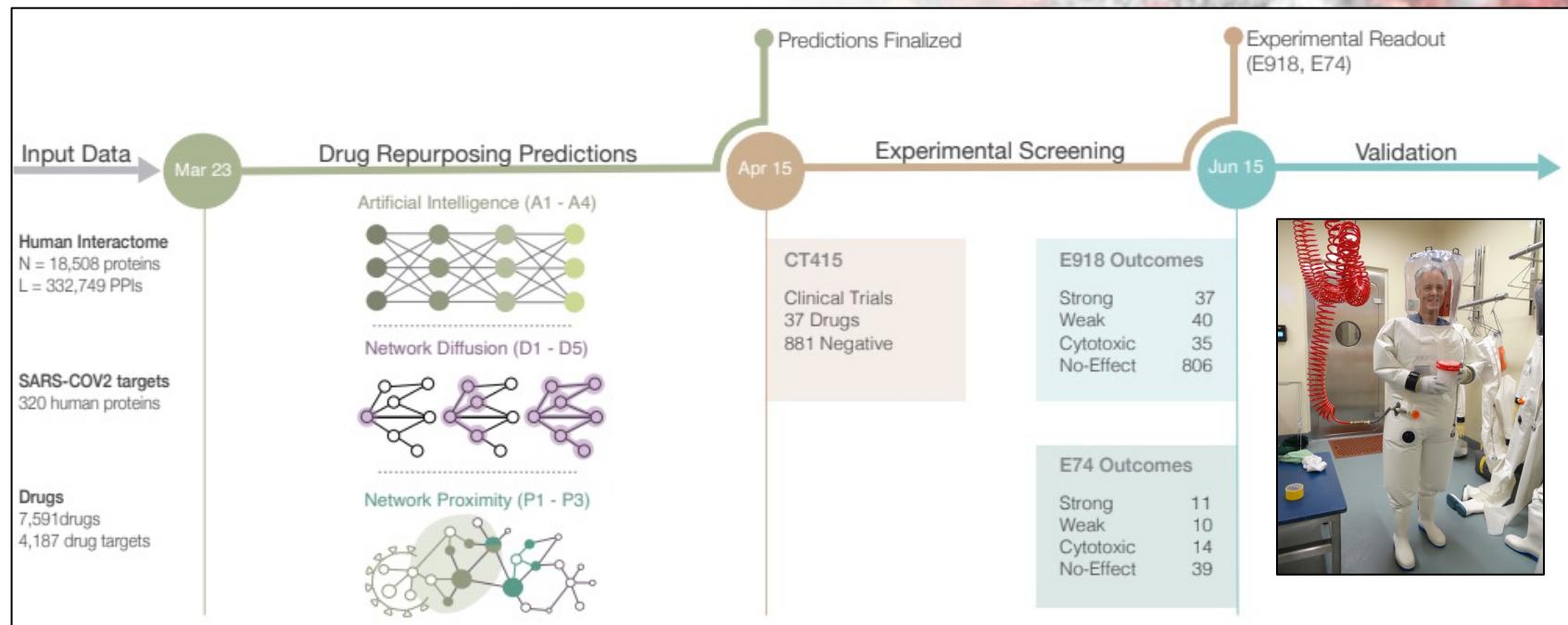
# http://txgnn.org



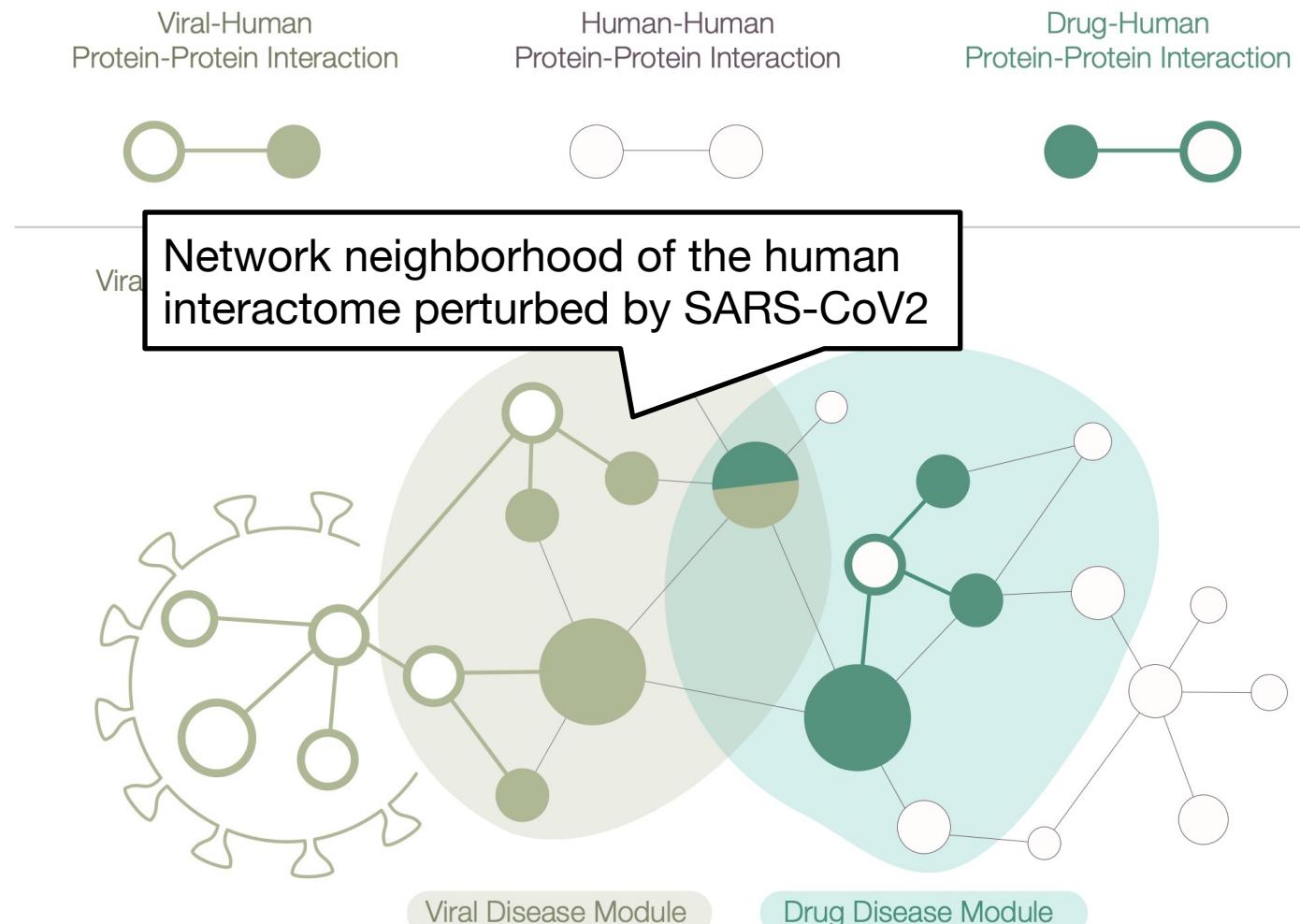
# Emerging pathogens

The traditional approach of iterative development, experimental testing, clinical validation, and approval of new drugs are not feasible

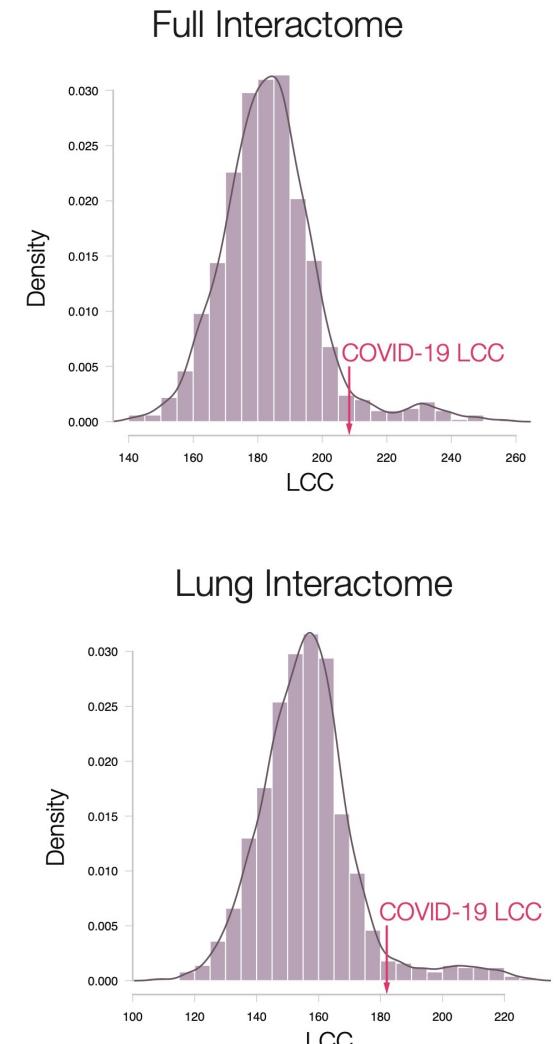
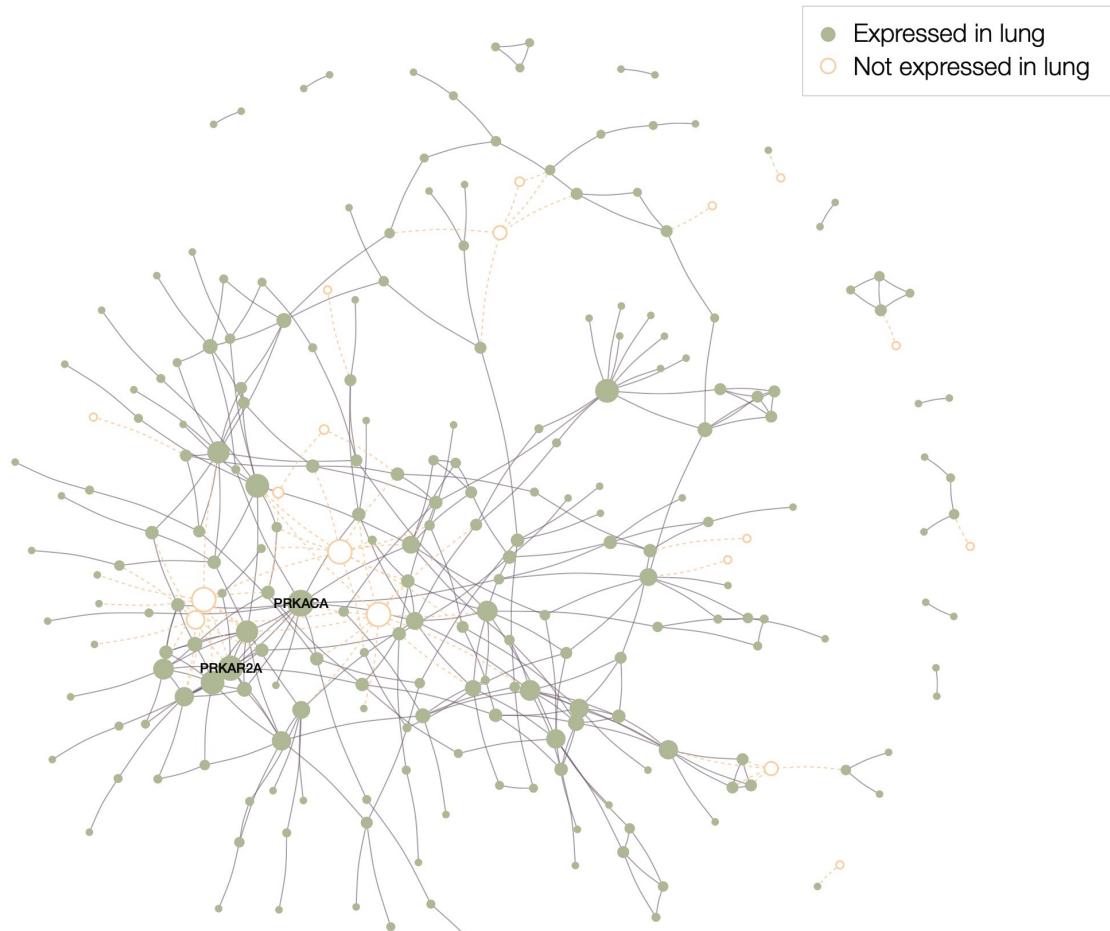
A more realistic strategy relies on drug repurposing, requiring us to identify clinically approved drugs that have a therapeutic effect in COVID-19 patients



# How to represent COVID-19? Map SARS-CoV2 targets to the human interactome



# COVID-19 disease module

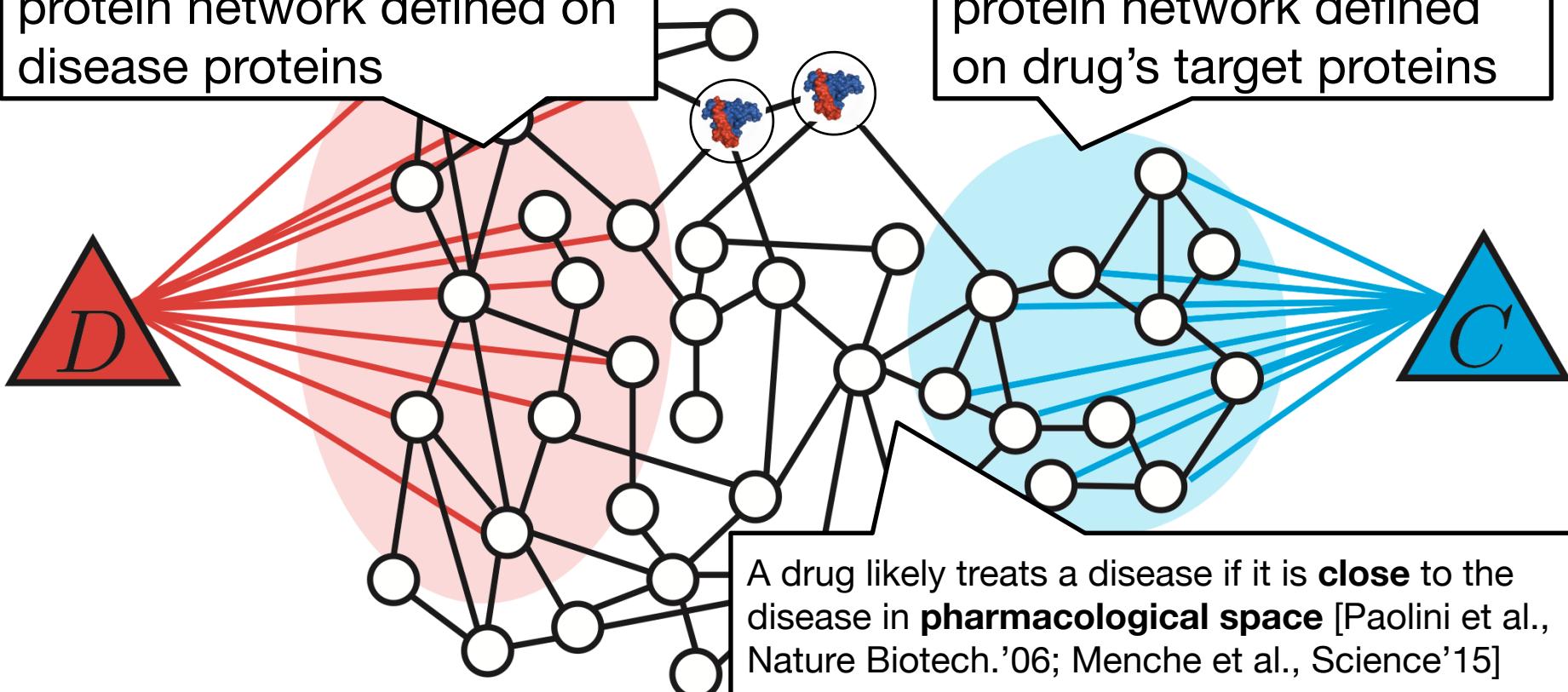


Gordon et al., Nature 2020 expressed 26 of the 29 SARS-CoV2 proteins and used AP-MS to identify 332 human proteins to which viral proteins bind

# Key Insight: subgraphs

**Disease:** Subgraph of rich protein network defined on disease proteins

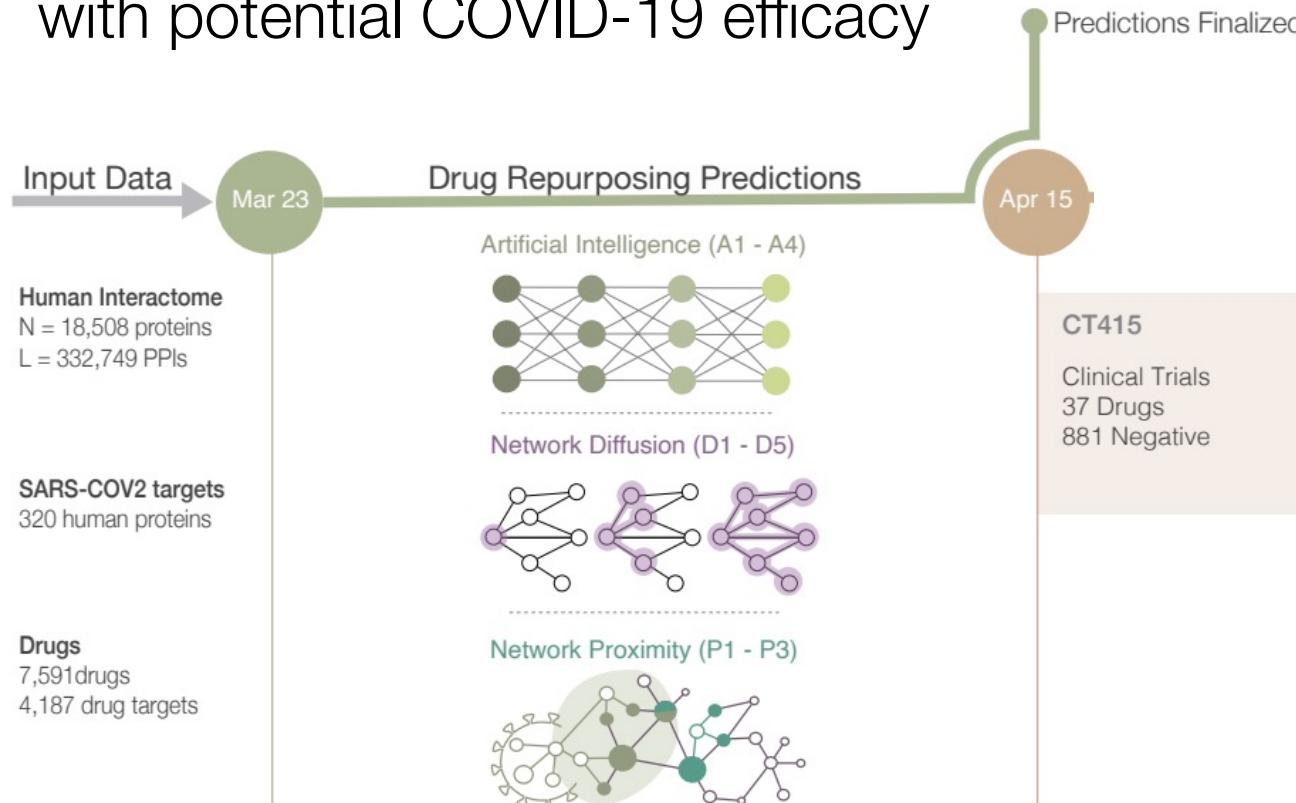
**Drug:** Subgraph of rich protein network defined on drug's target proteins



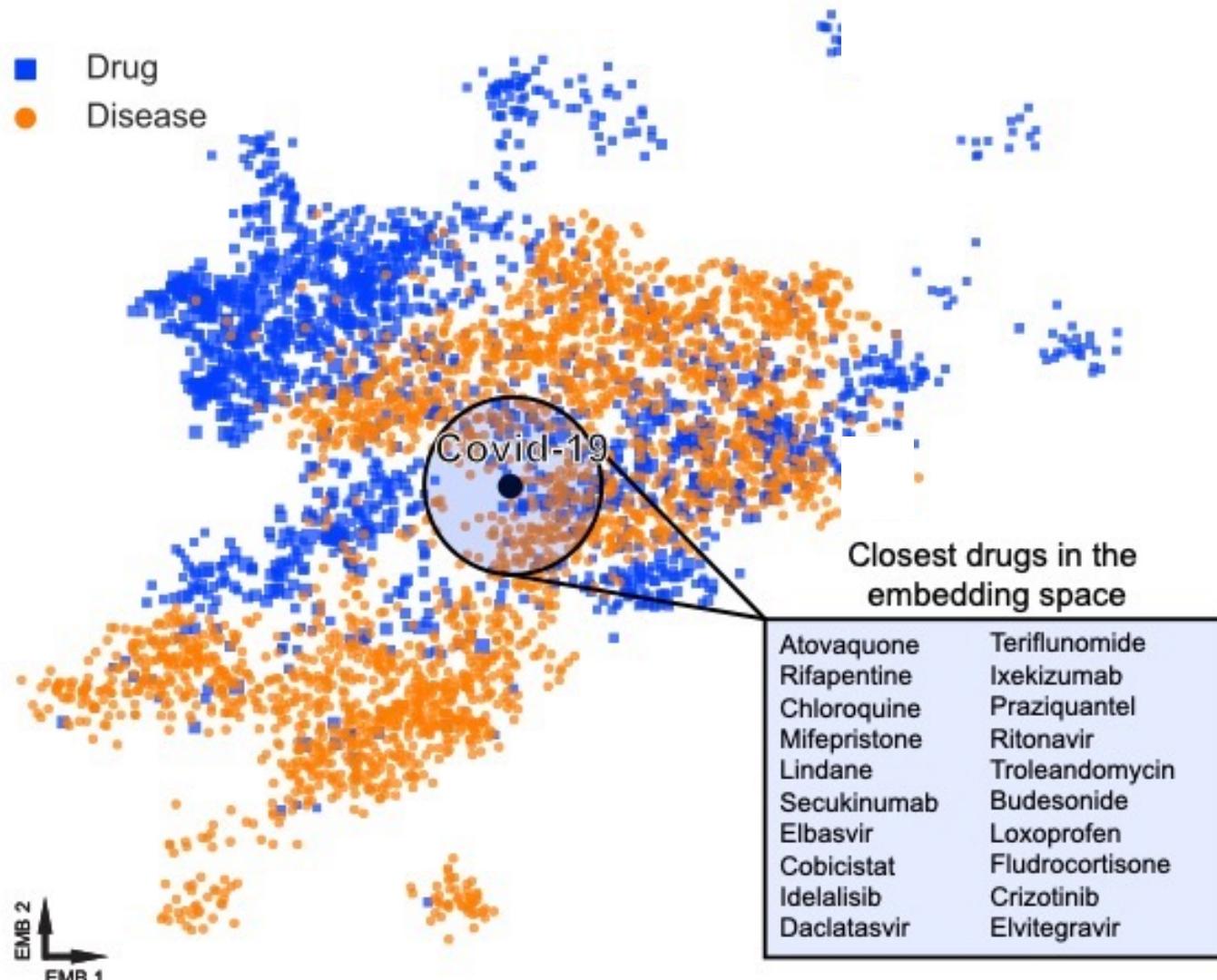
**Idea:** Use the paradigm of embeddings to operationalize the concept of closeness in pharmacological space

# Computational setup

- Proxy for ground-truth information:
  - Monitor drugs under **clinical trials**
  - Capture the **medical community's assessment** of drugs with potential COVID-19 efficacy

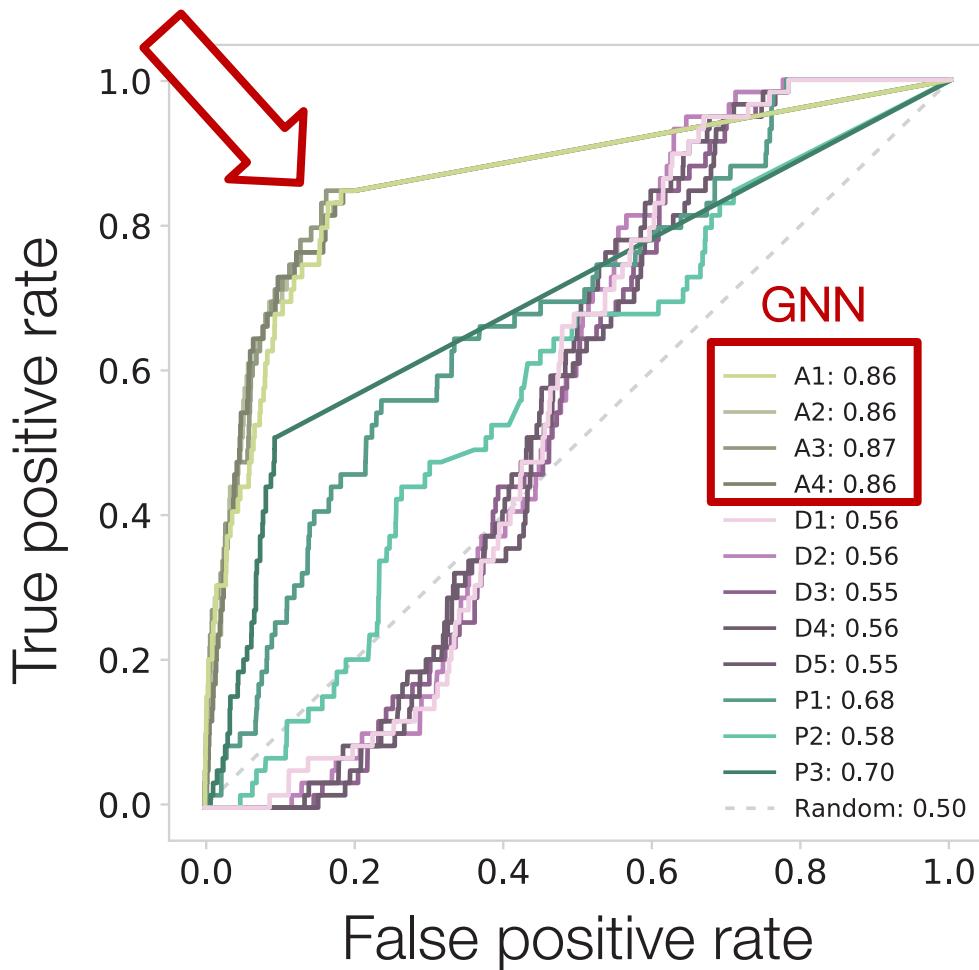


# Embedding space



# Results: COVID-19 Repurposing

## Individual ROC



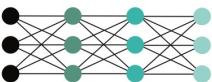
We test each pipeline's ability to recover drugs currently in clinical trials for COVID-19

The best individual ROC curves are obtained by the GNN methods

The second-best performance is provided by the proximity P3. Close behind is P1 with AUC = 0.68 and AUC = 0.58

Diffusion methods offer ROC between 0.55-0.56

# Final Prediction Model – Part #1

Input Data	Methods	Outcomes
Human Interactome N = 18,508 proteins L = 332,749 PPIs	 Network Proximity 3 pipelines	Infected Tissues/Organs
SARS-COV2 targets 320 human proteins Gordon et al, 2020	 Network Diffusion 5 pipelines	Comorbidity
Drug Targets 7,591 drugs 4,187 drug targets DrugBank	 AI Prioritization 4 pipelines	Drug Repurposing & Validation

# Final Prediction Model – Part #2

## Methods

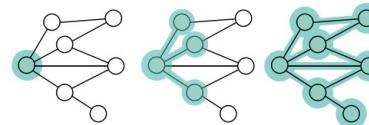
---

- A COVID-19 treatment can not be derived from the arsenal of therapies approved for specific diseases
- Repurposing strategies focus on drugs previously approved for other pathogens, or on drugs that target the human proteins to which viral proteins bind.
- Most approved drugs do not target directly disease proteins but bind to proteins in their network vicinity
- [Yildirim, Nature Biotech. 2007]
- Identify drug candidates that have the potential to perturb the network vicinity of the COVID-19 disease module.
- Implement 3 Network Repurposing Methods.



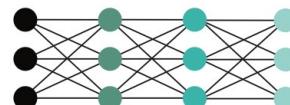
Network Proximity  
3 pipelines

---



Network Diffusion  
5 pipelines

---



AI Prioritization  
4 pipelines

# Final Prediction Model – Part #3

**Rank Aggregation Algorithm:** Maximize the number of pairwise agreements between the final ranking and each input ranking.

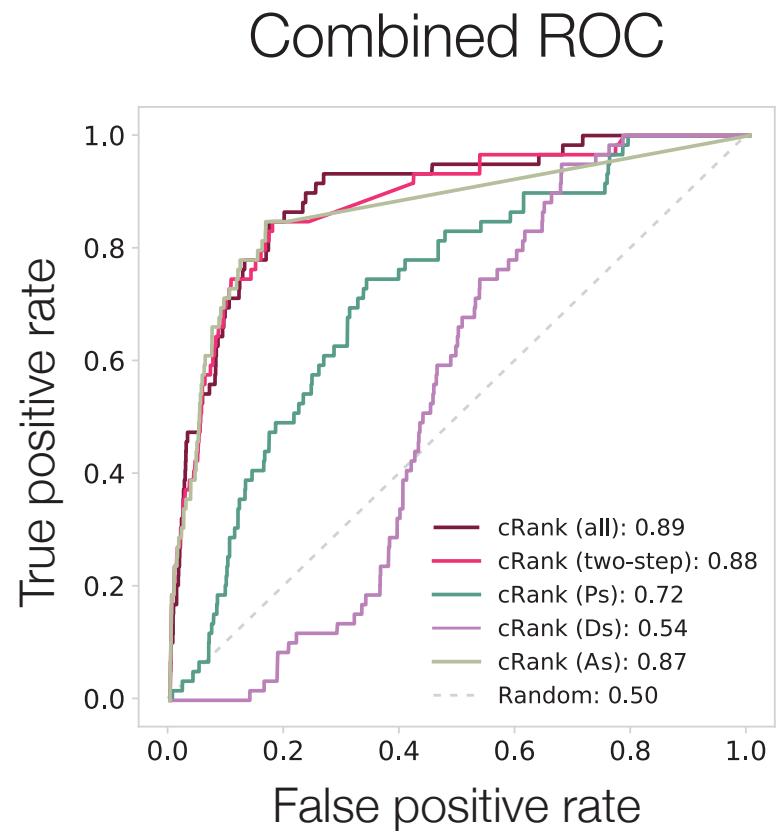
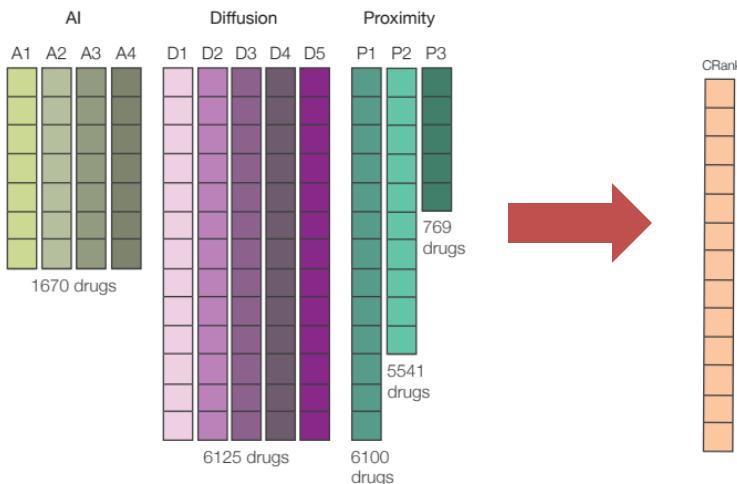
The combined performance of the AI methods is 0.87, the same as A3.

Improvement for proximity pipelines: 0.70 → 0.72.

Combined diffusion pipelines have lower performance (0.54 vs 0.56, for D1, D2, and D4).

Combining all 12 pipelines, gives AUROC=0.89, the highest of any individual or combination-based pipelines,

Individual pipelines offer complementary information harnessed by the combined ranking.



# Predicted Drug Candidates

○ # of Clinical trials from ClinicalTrials.gov

Joseph Loscalzo



86 drugs selected from the top 10% of the rank list.

Respiratory drugs (e.g., theophylline, montelukast).

Cardiovascular systems (e.g., verapamil, atorvastatin).

Antibiotics used to treat viral (e.g., ribavirin, lopinavir), parasitic (e.g., hydroxychloroquine, ivermectin, praziquantel), bacterial (e.g., rifaximin, sulfanilamide), mycotic (e.g., fluconazole), and mycobacterial (e.g., isoniazid) infections.

Immunomodulating/anti-inflammatory drugs (e.g., interferon- $\beta$ , auranofin, montelukast, colchicine)

Anti-proteasomal drugs (e.g., bortezomib, carfilzomib)

Less obvious choices: aminoglutethimide, melatonin, levothyroxine, calcitriol, selegiline, deferoxamine, mitoxantrone, metformin, nintedanib, cinacalcet, and sildenafil.

Drug	C-rank	Drug	C-rank	Drug
⑩ Ritonavir	1	Mesalazine	69	Sulfanilamide
Isoniazid	2	Pentamidine	92	Hydralazine
Troleandomycin	3	Verapamil	98	Gemfibrozil
Cilostazol	4	Melatonin	109	④ Ruxolitinib
76 Chloroquine	5	Griseofulvin	112	Propranolol
Rifabutin	6	Auranofin	118	Carbamazepine
Flutamide	7	① Atovaquone	124	Doxorubicin
② Dexamethasone	8	Montelukast	131	Levothyroxine
Rifaximin	9	Romidepsin	138	Dactinomycin
Azelastine	10	① Cobicistat	141	Tenovifir
Folic Acid	16	⑦ Lopinavir	146	Tadalafil
Rabeprazole	27	Pomalidomide	155	Doxazosin
Methotrexate	32	Sulfinpyrazone	157	Rosiglitazone
Digoxin	33	① Levamisole	161	Aminolevulinic acid
Theophylline	34	Calcitriol	164	Nitroglycerin
Fluconazole	41	① Interferon- $\beta$ -1a	173	Metformin
Aminoglutethimide	42	Praziquantel	176	① Nintedanib
67 Hydroxychloroquine	44	① Ascorbic acid	195	Allopurinol
Methimazole	47	Fluvastatin	199	Ponatinib
① Ribavirin	49	① Interferon- $\beta$ -1b	203	① Sildenafil
① Omeprazole	50	Selegiline	206	Dapagliflozin
Bortezomib	53	① Deferoxamine	227	Nitroprusside
Leflunomide	54	Ivermectin	235	Cinacalcet
Dimethylfumarate	55	① Atorvastatin	243	Mexiletine
④ Colchicine	57	Mitoxantrone	250	Sitagliptin
Quercetin	63	Glyburide	259	Carfilzomib
Mebendazole	67	② Thalidomide	262	① Azithromycin

# Experimental validation of predictions



National Emerging Infectious Diseases Laboratories (NEIDL)

CRank	Drug Name
1	Ritonavir
2	Isoniazid
3	Troleandomycin
4	Cilostazol
5	Chloroquine
6	Rifabutin
7	Flutamide
8	Dexamethasone
9	Rifaximin
10	Azelastine
11	Crizotinib

17	Celecoxib
18	Betamethasone
19	Prednisolone
20	Mifepristone
21	Budesonide
22	Prednisone
23	Oxiconazole
24	Megestrol acetate
25	Idelalisib
26	Econazole
27	Dehorserazole

Ranked lists of drugs

New algorithms:

Prioritizing Network Communities, *Nature Communications* 2018

Subgraph Neural Networks, *NeurIPS* 2020

Graph Meta Learning via Local Subgraphs, *NeurIPS* 2020

**Results:** 918 compounds screened for their efficacy against SARS-CoV-2 in VeroE6 cells:

- **37 had a strong effect** being active over a broad range of concentrations
- **40 had a weak effect** on the virus
- **An order of magnitude higher hit rate** among top 100 drugs than prior work

# Results: Network drugs

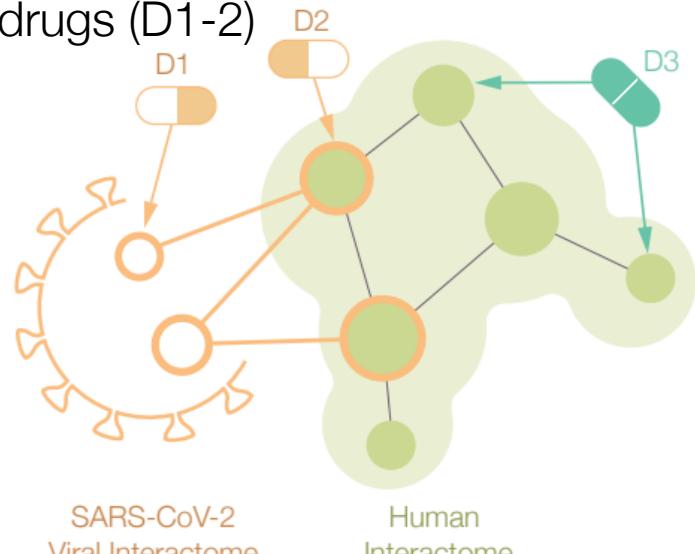
- 76/77 drugs that successfully reduced viral infection do not bind proteins targeted by SARS-CoV-2:
  - These drugs rely on **network-based actions** that cannot be identified by docking-based strategies

Strong  
Weak

CRank	Drug Name	CRank	Drug Name
5	Chloroquine	423	Pitavastatin
6	Rifabutin	431	Tenoxicam
9	Rifaximin	438	Quinidine
10	Azelastine	456	Sertraline
16	Folic acid	460	Ingenol mebutate
32	Methotrexate	463	Noregrestromin
33	Digoxin	493	Sildenafil
44	Hydroxychloroquine	499	Eliglustat
50	Omeprazole	518	Ulipristal
113	Clobetasol propionate	553	Cinacalcet
118	Auranofin	556	Perphenazine
120	Vinblastine	558	Idarubicin
199	Fluvastatin	564	Perhexiline
210	Clomifene	569	Amiodarone
233	Ibuprofen	577	Duloxetine
235	Ivermectin	585	Toremifene
243	Atorvastatin	586	Afatinib
253	Pralatrexate	601	Amitriptyline
263	Cobimetinib	626	Medazine
269	Hydralazine	635	Valsartan
297	Propranolol	651	Eletriptan
317	Osimertinib	673	Sotalol
348	Vincristine	678	Thioridazine
367	Doxazosin	695	Chlorycyclizine
397	Rosiglitazone	707	Omacetaxine mepesuccinate
398	Aminolevulinic acid	721	Candesartan

CRank	Drug Name
742	Mianserin
755	Clofazimine
767	Chlorpromazine
772	Imipramine
830	Promazine
900	L-Alanine
917	Moxifloxacin
933	Tasimelteon
995	Vandetanib
1000	Azilsartan medoxomil
1020	Frovatriptan
1034	Zolmitriptan
1035	Procarbazine
1093	Asenapine
1107	Dyclonine
1140.5	Clemastine
1194	Prochlorperazine
1222	Miglustat
1224	Prenylamine
1276	Dalfampridine
1314	Cinchocaine
1355	Methotriptazine
1396	Methylthioninium
1403	Metixene
1443	Trifluoperazine

Direct target drugs (D1-2)



58/77 drugs with positive experimental outcome are among top 750 ranked drugs

Network drugs (D3)

# L14 Quick Check

<https://forms.gle/B5PBaa2DCTLZpEqh8>

BMI 702: Biomedical Artificial Intelligence

*Foundations of Biomedical Informatics II, Spring 2024*

Quick check quiz for lecture 14: Design of chemical and genetic perturbations, drug repurposing, protein design, emerging uses of generative AI.

Course website and slides: <https://zitniklab.hms.harvard.edu/BMI702>

marinka@hms.harvard.edu [Switch accounts](#) 

Not shared

\* Indicates required question

First and last name \*

Your answer \_\_\_\_\_

Harvard email address \*

Your answer \_\_\_\_\_

Go to <http://txggn.org> and examine predictions for **rheumatoid arthritis**. Our evaluation will focus on disease-modifying antirheumatic drugs (DMARDs), which is a class of drugs indicated for the treatment of several inflammatory arthritides, including rheumatoid arthritis, as well as for the management of other connective tissue diseases and some cancers. Answer the following four questions.

1) What is the predicted rank of **sulfasalazine**, a common conventional DMARD?

2) What is the predicted rank of **methotrexate**, another common DMARD?

3) Give two examples of reasoning paths (meta-paths) used by the algorithm to relate **rheumatoid arthritis** with **sulfasalazine**. Comment the results.

4) Give two examples of reasoning paths (meta-paths) used by the algorithm to relate **rheumatoid arthritis** with **methotrexate**. Comment the results. Examine meta-paths that use this template: Disease-Drug-Gene/Protein-Drug.

Your answer \_\_\_\_\_

# Outline for today's class

- High-throughput genetic and chemical perturbations
- Drug repurposing, indication and contra-indication prediction



- Generative protein design

- Generative AI agents