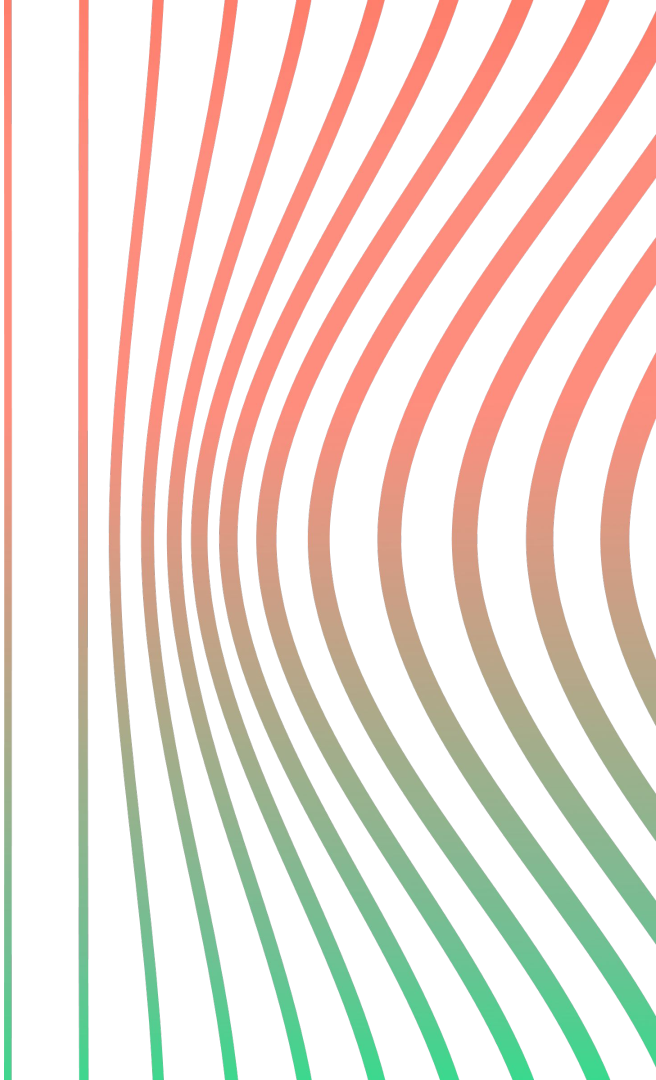
ICML
International Conference
On Machine Learning

Active Learning for Efficient Discovery of Optimal Combinatorial Perturbations

Jason Qin[1], Hans-Hermann Wessels[1], Carlos Fernandez-Granada[2], Yuhan Hao[1]

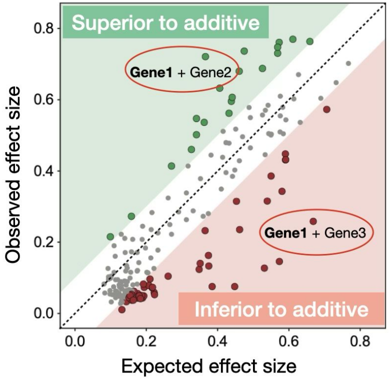
Neptune Bio

NYU

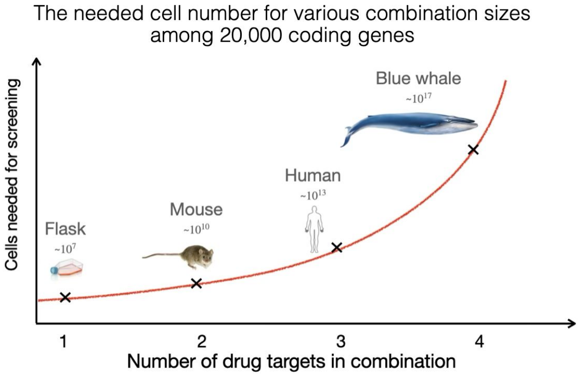


Motivation - Understanding Combinatorial Biology

Goal: Understand gene combinations for novel biology, drug discovery



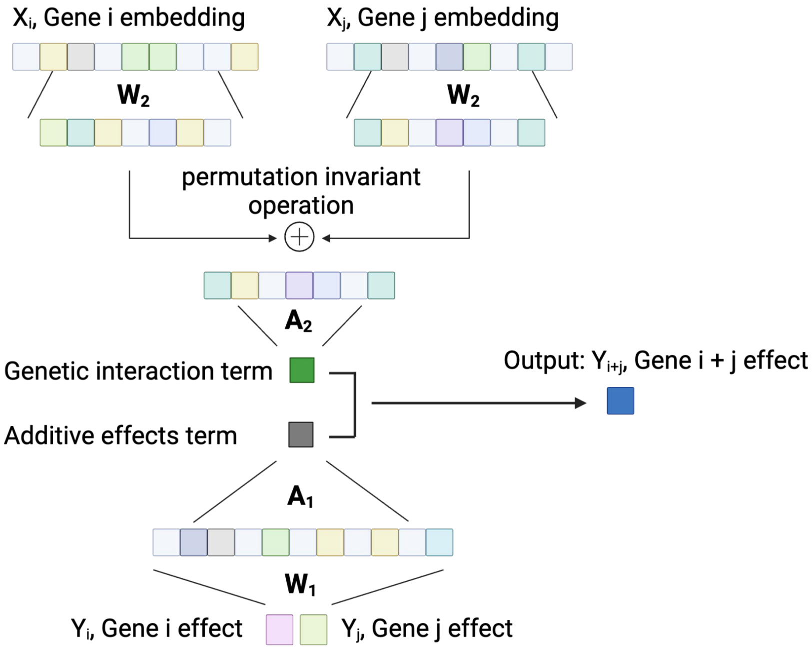
Limitation: Exhaustive exploration of combinatorial space is infeasible


The needed cell number for various combination sizes among 20,000 coding genes

Our Contribution: NAIAD

1) Accurately model non-linear interactions
2) Active learning framework to identify strongest combinations with limited testing

Modeling Non-Linear Interactions



X_i, Gene i embedding

X_j, Gene j embedding

W_2

W_2

permutation invariant operation

A_2

Genetic interaction term

Additive effects term

Output: Y_{i+j}, Gene i + j effect

A_1

W_1

Y_i, Gene i effect

Y_j, Gene j effect

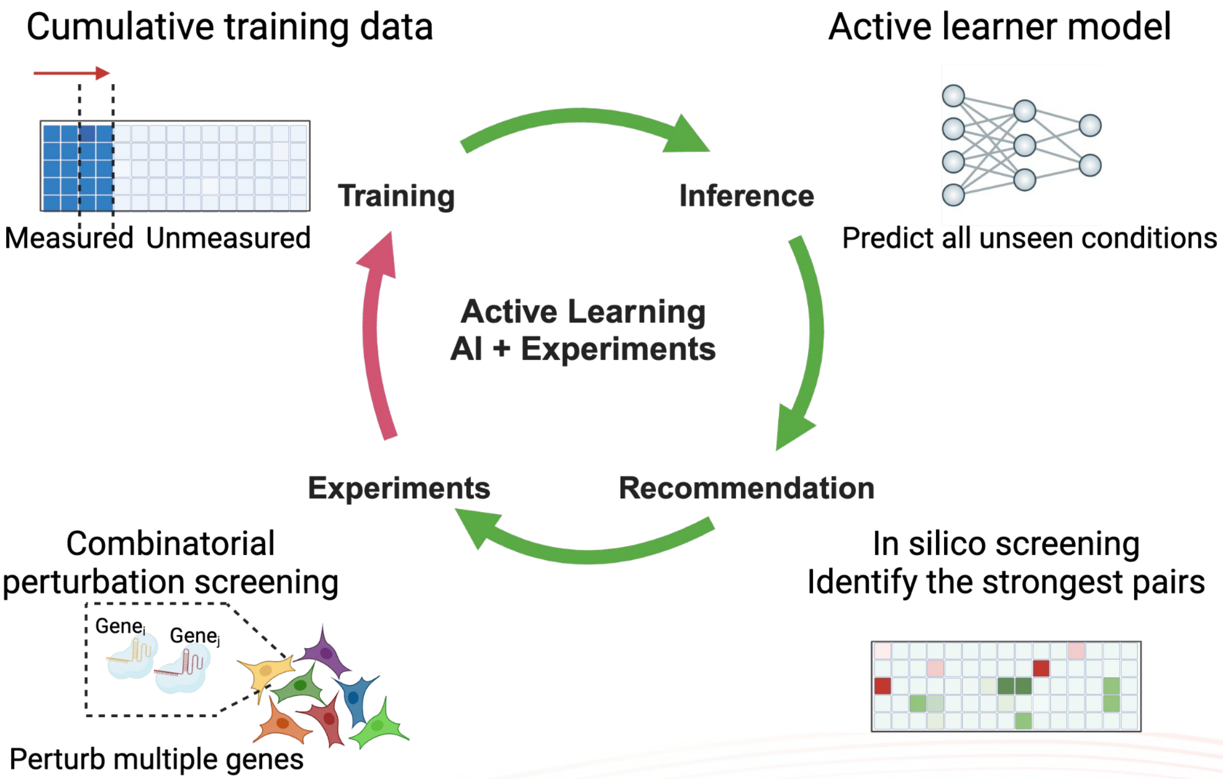
$$Y_{i+j} = \phi([Y_i, Y_j]W_1)A_1^T + f(\phi(W_2X_i^{\text{gene}}), \phi(W_2X_j^{\text{gene}}))A_2^T$$

Interactions between **gene embeddings** capture non-linearity of interactions

Embedding dimension scales with dataset size

Single-gene effects condition predictions based on predicted additive effect

Selecting Strong Combinations via Active Learning



Cumulative training data

Measured Unmeasured

Active learner model

Predict all unseen conditions

Training

Inference

Active Learning
AI + Experiments

Experiments

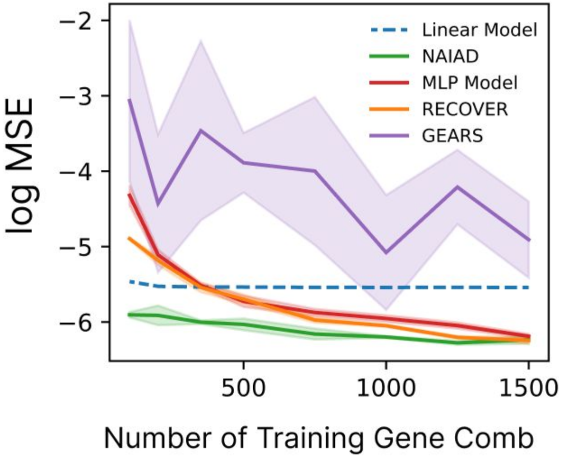
Recommendation

Combinatorial
perturbation screening

Perturb multiple genes

In silico screening
Identify the strongest pairs

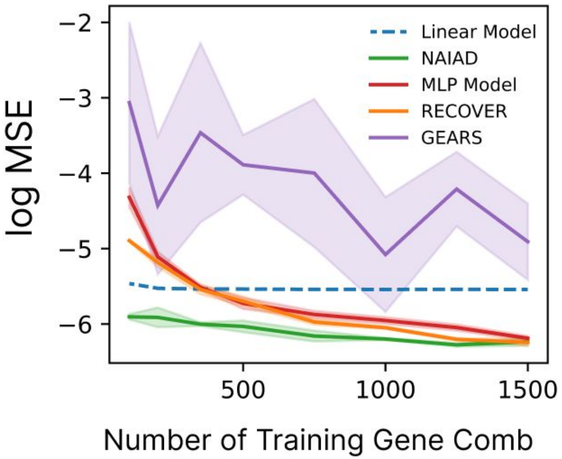
NAIAD Captures Combinatorial Phenotype Effectively

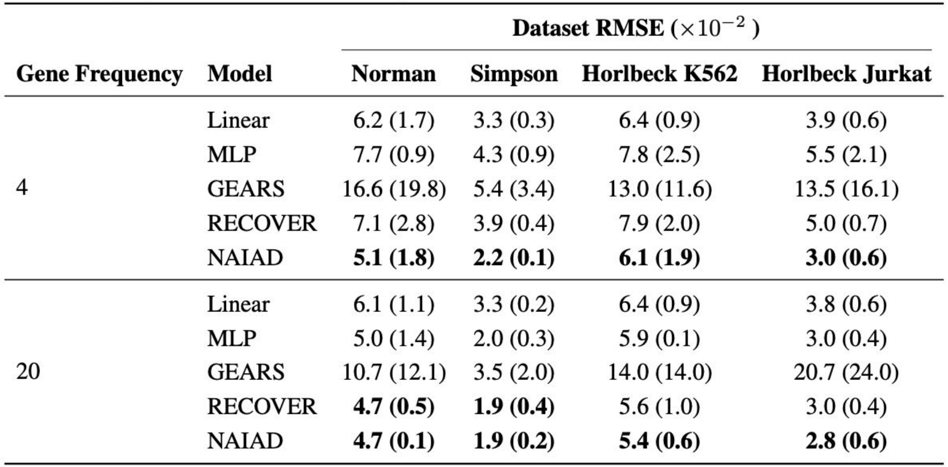


NAIAD outperforms comparable models for accurately identifying combinatorial effect

Bertin, Paul, et al. "RECOVER identifies synergistic drug combinations in vitro through sequential model optimization." Cell Reports Methods 3.10 (2023)
Roohani, Yusuf, et al.. "Predicting transcriptional outcomes of novel multigene perturbations with GEARS." Nature Biotechnology 42.6 (2024): 927-935.

NAIAD Captures Combinatorial Phenotype Effectively





		Dataset RMSE ($\times 10^{-2}$)			
Gene Frequency	**Model**	**Norman**	**Simpson**	**Horlbeck K562**	**Horlbeck Jurkat**
4	Linear	6.2 (1.7)	3.3 (0.3)	6.4 (0.9)	3.9 (0.6)
	MLP	7.7 (0.9)	4.3 (0.9)	7.8 (2.5)	5.5 (2.1)
	GEARS	16.6 (19.8)	5.4 (3.4)	13.0 (11.6)	13.5 (16.1)
	RECOVER	7.1 (2.8)	3.9 (0.4)	7.9 (2.0)	5.0 (0.7)
	NAIAD	**5.1 (1.8)**	**2.2 (0.1)**	**6.1 (1.9)**	**3.0 (0.6)**
20	Linear	6.1 (1.1)	3.3 (0.2)	6.4 (0.9)	3.8 (0.6)
	MLP	5.0 (1.4)	2.0 (0.3)	5.9 (0.1)	3.0 (0.4)
	GEARS	10.7 (12.1)	3.5 (2.0)	14.0 (14.0)	20.7 (24.0)
	RECOVER	**4.7 (0.5)**	**1.9 (0.4)**	5.6 (1.0)	3.0 (0.4)
	NAIAD	**4.7 (0.1)**	**1.9 (0.2)**	**5.4 (0.6)**	**2.8 (0.6)**

NAIAD outperforms comparable models for accurately identifying combinatorial effect

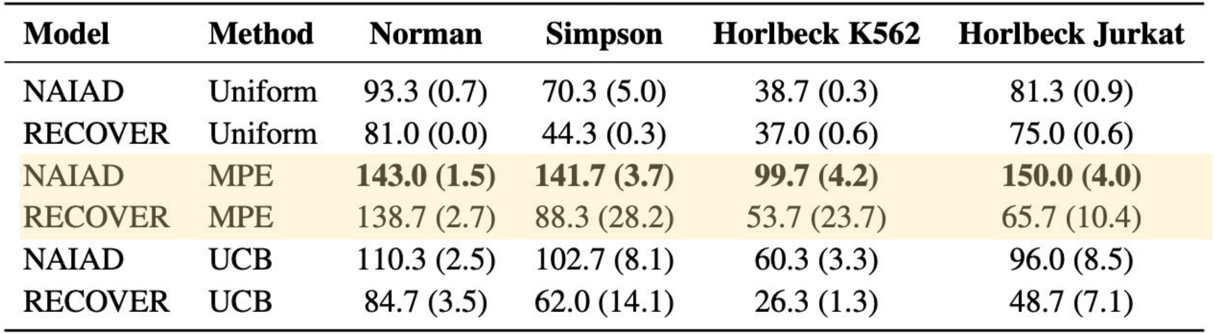
We identify a minimal “gene occurrence” frequency for training accurate embeddings

Bertin, Paul, et al. "RECOVER identifies synergistic drug combinations in vitro through sequential model optimization." Cell Reports Methods 3.10 (2023)
Roohani, Yusuf, et al.. "Predicting transcriptional outcomes of novel multigene perturbations with GEARS." Nature Biotechnology 42.6 (2024): 927-935.

NAIAD Identifies Strong Perturbations Efficiently

Metric True Positive Rate (TPR) of identifying strongest perturbations

Gene modality



Model	Method	Norman	Simpson	Horlbeck K562	Horlbeck Jurkat
NAIAD	Uniform	93.3 (0.7)	70.3 (5.0)	38.7 (0.3)	81.3 (0.9)
RECOVER	Uniform	81.0 (0.0)	44.3 (0.3)	37.0 (0.6)	75.0 (0.6)
NAIAD	MPE	**143.0 (1.5)**	**141.7 (3.7)**	**99.7 (4.2)**	**150.0 (4.0)**
RECOVER	MPE	138.7 (2.7)	88.3 (28.2)	53.7 (23.7)	65.7 (10.4)
NAIAD	UCB	110.3 (2.5)	102.7 (8.1)	60.3 (3.3)	96.0 (8.5)
RECOVER	UCB	84.7 (3.5)	62.0 (14.1)	26.3 (1.3)	48.7 (7.1)

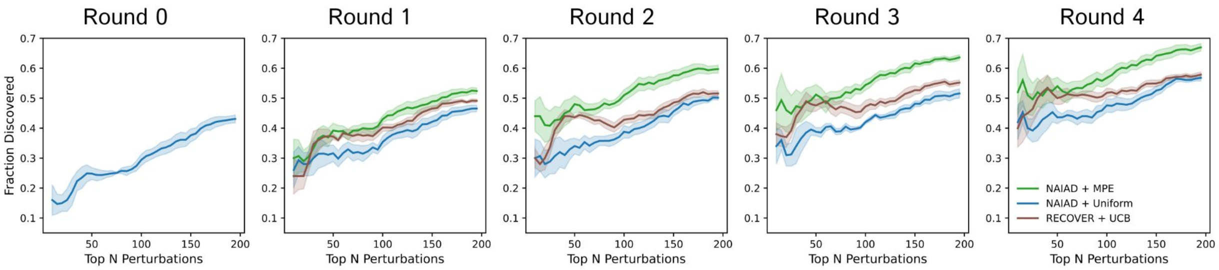
NAIAD Identifies Strong Perturbations Efficiently

Metric True Positive Rate (TPR) of identifying strongest perturbations

Gene modality

Model	**Method**	**Norman**	**Simpson**	**Horlbeck K562**	**Horlbeck Jurkat**
NAIAD	Uniform	93.3 (0.7)	70.3 (5.0)	38.7 (0.3)	81.3 (0.9)
RECOVER	Uniform	81.0 (0.0)	44.3 (0.3)	37.0 (0.6)	75.0 (0.6)
NAIAD	MPE	**143.0 (1.5)**	**141.7 (3.7)**	**99.7 (4.2)**	**150.0 (4.0)**
RECOVER	MPE	138.7 (2.7)	88.3 (28.2)	53.7 (23.7)	65.7 (10.4)
NAIAD	UCB	110.3 (2.5)	102.7 (8.1)	60.3 (3.3)	96.0 (8.5)
RECOVER	UCB	84.7 (3.5)	62.0 (14.1)	26.3 (1.3)	48.7 (7.1)

Drug modality



Thank You!

Code

https://github.com/NeptuneBio/NAIAD

Contact

jason@neptune.bio

Team

Jason Qin

Harm Wessels

Carlos Fernandez-Grenada

Yuhan Hao

NYU

Neptune Bio