A unified foundation-model framework for drug-response prediction in human cell types

Amar Yadav*, Joy Makwana*, Aditya Goel*, Advika Lakshman*, Nischal Gupta, Aaryan Gupta, Lavanya Nemani, Ramesh Hariharan, Badri Padhukasahasram
Strand Life Sciences, Bangalore, India *Contributed equally

VISIT OUR WEBSITE

strand

Contact
Badri Padhukasahasram
Vice President - Data Science
91-9591670961 badri.p@strandls.com

Introduction

- Traditional cell line screens fail to capture tissue-specific and patient-specific heterogeneity in drug response (e.g. stromal, epithelial and immune cell interactions)
- Single-context predictions have poor clinical translation—e.g. KRAS/NRAS inhibitors show variable efficacy across patient cohorts and cell types

Why Context-Aware Prediction is Essential:

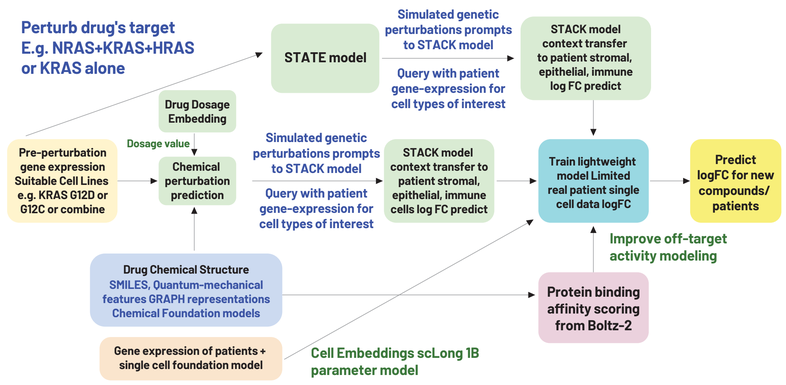
- One drug, many contexts: Mutation status + tumor microenvironment + immune infiltration = vastly different response trajectories.
- Dosage & chemical space matter: Drug dosing embeddings + chemical perturbation predictions capture dose-response nonlinearities missed by static models.
- Off-target liability: Single-target predictions (e.g., KRAS alone) ignore critical off-target effects on unintended proteins, a major cause of clinical failures.

Solution:

Sequence of AI Foundation models enables genetic and chemical perturbation simulation, followed by grounding of predictions in human cell type context with off-target activity awareness guided by protein-ligand interactions scoring.

From cell-lines to real patients

A unified framework for more realistic predictions: Cell lines to real patients

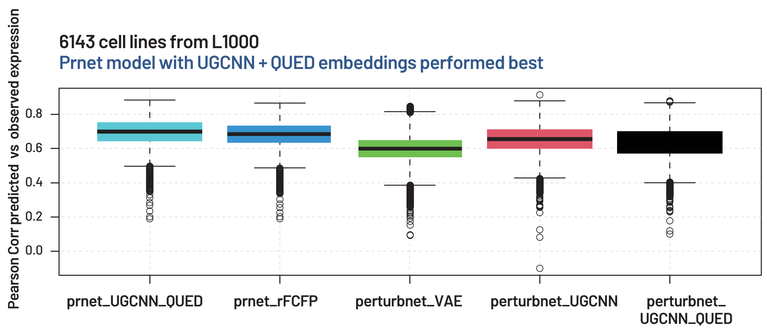


Optimize Protein-ligand binding

Drug	Fraction of times target protein binding score is > non-target protein DeepPurpose	Fraction of times target protein binding score is > non-target protein Boltz2
Upadacitinib	84.492% (83.316%-84.463%*)	98.880% (98.858%-99.040%*)
Deucravacitinib	94.540% (94.392%-94.644%*)	100.00% (99.742%-100.00%*)
Zasocitinib	32.614% (32.440%-32.806%*)	98.880% (98.586%-99.040%*)

*Exact 95% binomial confidence intervals
Boltz2 better than DeepPurpose for separating non-target and target proteins

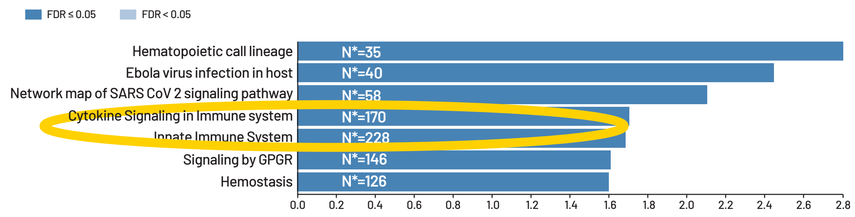
Optimizing small molecule representations



Comparison of different approaches for summarizing chemical structure of drugs as numerical vectors i) prnet model with Uniform graph Convolution neural network (UGCNN) embeddings and quantum mechanical features (QUED) that capture electronic property of molecules ii) prnet model with rFCP (SMILES based string converted to rescaled functional class finger prints) iii) Perturbnet model with a variational autoencoder to represent drug chemical structure iv) Perturbnet model with UGCNN v) Perturbnet model with UGCNN + QUED

Computational TNF inhibition experiments

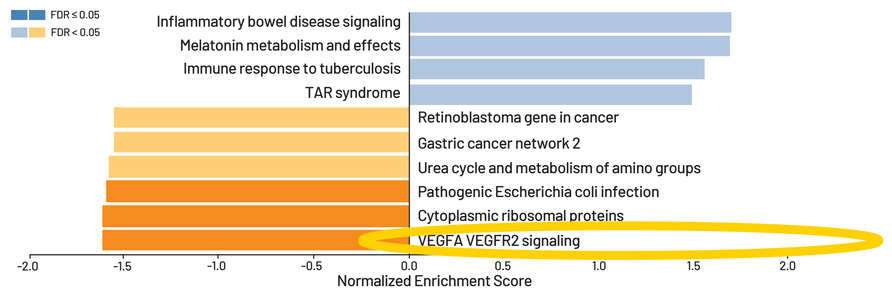
Significant pathways from overrepresentation analysis with FDR < 0.05



*N is number of differentially expressed genes overlapping the pathway

Overrepresentation analysis tests whether a pathway contains more differentially expressed genes than expected by chance
Highlights large effect transcriptional changes in pathways
Identified immune related pathways among top hits from TNF gene perturbations

Significant pathways from Gene Set Enrichment Analysis with FDR < 0.05



GSEA considers not just differential expressed genes but all genes and their ranking. Complements overrepresentation analysis. Captures coordinated but modest shifts in pathway activity such as for VEGFA-VEGFR2 signaling

Small-molecule perturbation modeling

PRnet **Perturbation-Response Network**

A deep learning model that predicts cell-type- specific transcriptomic responses to chemical or genetic perturbations. It learns a mapping from perturbation features (e.g., compound structure, dose) and cell context to differential gene expression profiles. Perturbation identity + cell-type context.

Input	Perturbation identity + cell-type context.
Output	Predicted differential gene expression.
Strength	Generalizes to unseen cell types and perturbations.
Use Case	Virtual screening of drug effects across tissues.

PerturbNet **Perturbation-Response Network**

A deep learning model that predicts cell-type- specific transcriptomic responses to chemical or genetic perturbations. It learns a mapping from perturbation features (e.g., compound structure, dose) and cell context to differential gene expression profiles. Perturbation identity + cell-type context.

Input	Perturbation descriptor + cellular state.
Output	Predicted single-cell expression distribution.
Strength	Captures heterogeneity in cellular response.
Use Case	Modeling variable drug responses at single-cell resolution.

Shared Conceptual Pipeline

Perturbation Features (compound / gene / dose) → Cell Context Encoding
Latent Space Mapping → Predicted Transcriptomic Response

Summary and Achievements

- A foundation model based pipeline aiming for realistic drug response prediction for human cell types.
- Optimal drug representations identified through extensive analysis and provided 5% improvement in pearson correlation and 10% reduction in mean square error for LINCS1000 cell lines.
- Sequence of foundation models enables context transfer for generating more accurate predictions in cell types of interest.

Conclusions

- An integrated predictive framework designed for realistic drug response prediction in diverse contexts.
- Optimization of different individual components ongoing including i) Drug representations ii) Genetic perturbation modeling iii) Protein-ligand activity iv) Foundation models for context transfer
- Genetic perturbation modeling as well as chemical perturbation modeling tools demonstrated meaningful results

- **Complementary Approaches**
  PerturbNet models full single-cell distributions.
- **Drug Discovery**
  Enable in silico screening of perturbation effects without exhaustive wet-lab experiments.
- **Foundation for Virtual Cells**
  Core building blocks toward predicting cellular behavior under arbitrary interventions.