MMDRP:
Drug Response Prediction and Biomarker Discovery Using Multi-Modal Deep Learning

OICR Ontario Institute for Cancer Research

Molecular Genetics
UNIVERSITY OF TORONTO

Farzan Taj [†§], Lincoln D. Stein[†§]

[†] Ontario Institute for Cancer Research, Toronto, ON ON M5G 0A3, Canada | [§] Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada

Deciphering Therapy Response in Cancer Variability

Cancer treatments often face the challenge of patient-specific responses due to genetic diversity. Advancements in 'omics' technologies have paved the way for precision oncology, aiming to tailor treatments to individual patient profiles. Traditional pharmacogenomic studies have helped to understand drug sensitivity but are limited by data scarcity, inconsistent coverage across cell lines, and outdated analysis methods. Our project introduces a **M**ulti-**M**odal **D**rug **R**esponse **P**redictor (MMDRP), a sophisticated Python-based tool that utilizes deep learning to predict drug efficacy on cancer cell lines while using multiple omic data types. MMDRP incorporates novel features like:

Weight Adjustment: Prioritizing rare samples to avoid overfitting.

Multi-Modal Framework with Data Fusion: Simultaneous integration of various 'omic data for better predictions.

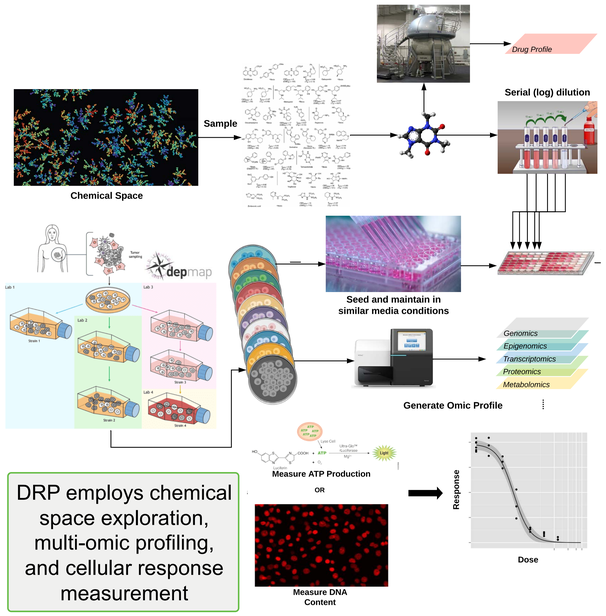
Graph Neural Networks: Advanced representation of molecular structures.

Cross-validation assessment and model interpretation demonstrates MMDRP's superior performance over traditional models, highlighting its potential in drug response prediction, drug repurposing, biomarker discovery and drug design.

The code for this project is available at:
https://github.com/LincolnSteinLab/MMDRP

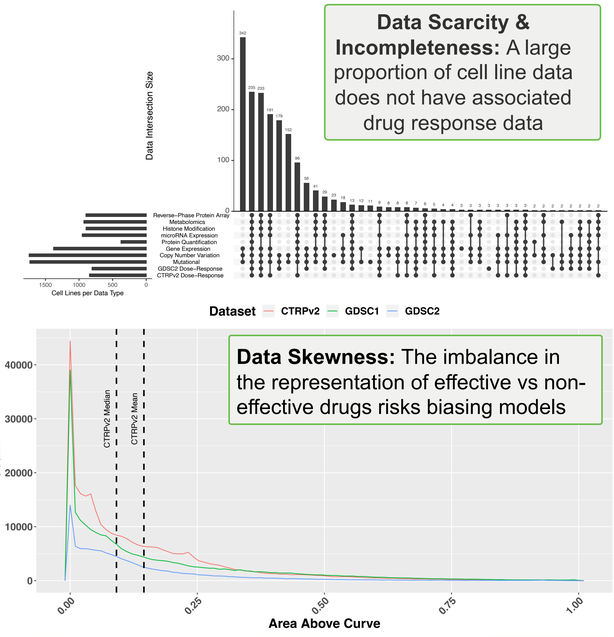
Pharmacogenomic Experiments Pave the Way for Precision Oncology

Drug sensitivity data coupled with cell line profiling data allow for better statistical modelling of cancer cells' response to drugs.



DRP employs chemical space exploration, multi-omic profiling, and cellular response measurement

Overcoming Data Limitations in Drug Response Prediction

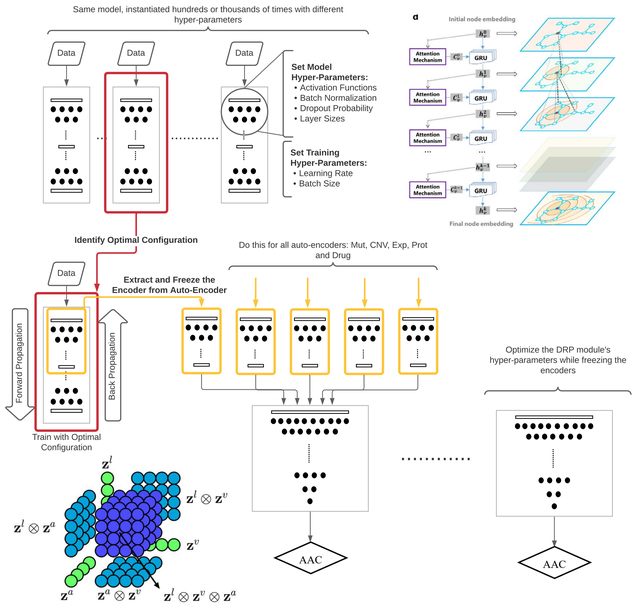


Data Scarcity & Incompleteness: A large proportion of cell line data does not have associated drug response data

Data Skewness: The imbalance in the representation of effective vs non-effective drugs risks biasing models

Modular Multi-Modal Neural Network for Enhanced DRP

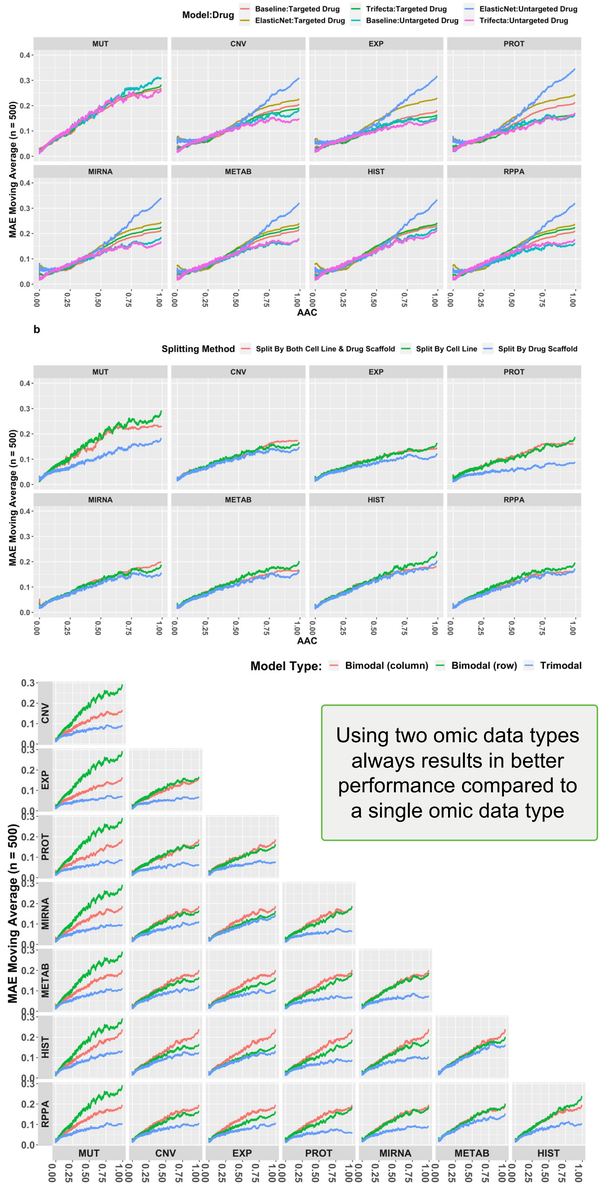
Autoencoders: Independent Learning from omic data types
AttentiveFP GNN: Better drug representation
Low-Rank Multi-Modal Fusion (LMF): better integration of omic data



Multifaceted Performance Assessment

Assessed model performance across **Drug efficacies** (AACs 0 to 1), **Drug types** (Targeted vs Untargeted), **Model configurations** (Baseline vs Refined), **Data Splitting Strategies** (Split by Cell Line, Split by Drug Scaffold, or Both)

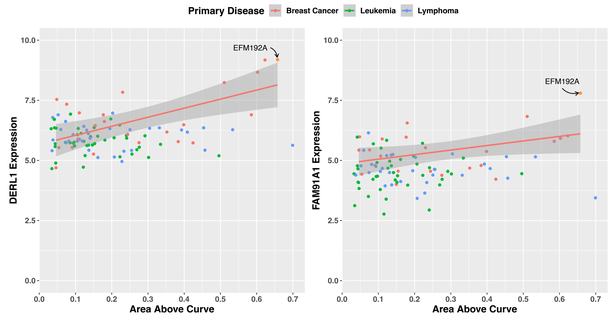
- Sample weighting via Label Distribution Smoothing (LDS), drug representation via graph neural networks (GNNs), and data integration via low-rank multi-modal fusion (LMF) all result in better predictive models.
- Predicting the effects of targeted drugs, as well as more efficacious drugs with higher AAC remains a challenge.
- MMDRP generalizes best to novel drugs than to novel cells.



Using two omic data types always results in better performance compared to a single omic data type

DRP for Drug Repurposing

Example: Ibrutinib is a small molecule that targets the BTK protein and is FDA-approved for Chronic lymphocytic leukemia and a few subtypes of lymphoma. MMDRP accurately predicts the sensitivity of breast cancer cell lines to ibrutinib, e.g., in the EFM192A breast cancer cell line. We interpreted our model using the Integrated Gradients method and identified **DERL1** and **FAM91A1** to be involved in the cell line's response to ibrutinib, both of which have been identified as potential targets in breast cancer. The expressions of DERL1 and FAM91A1 show higher correlations with AAC in breast cancer cell lines and may serve as predictive biomarkers:



Interpretation of MMDRP's GNN module highlights atoms and bonds predictive of ibrutinib's efficacy, and correctly identifies the acryloyl functional group that is responsible for binding BTK.

Ibrutinib

Summary

MMDRP can help improve pre-clinical pass rates, drug repurposing and biomarker discovery. The key findings of this work are summarized as follows:

- Current DRP datasets lack diversity, impacting predictions for underrepresented drugs, drug efficacies and cancer types.
- Other profiling data types, beyond gene expression, contribute significantly to drug response prediction
- LDS, LMF, and GNN better use existing omic and drug sensitivity data for tasks with downstream impact in precision oncology.
- The interpretation of multi-modal models can help identify novel composite biomarkers.

Future work should emphasize on prediction performance in higher AAC and targeted drugs, in addition to the deconvolution of the effects of clonal heterogeneity on drug response.

Ontario

Funding provided by the Government of Ontario.