Predicting New Indications of Known Drugs Based On Spearman Correlation Between Gene Signatures

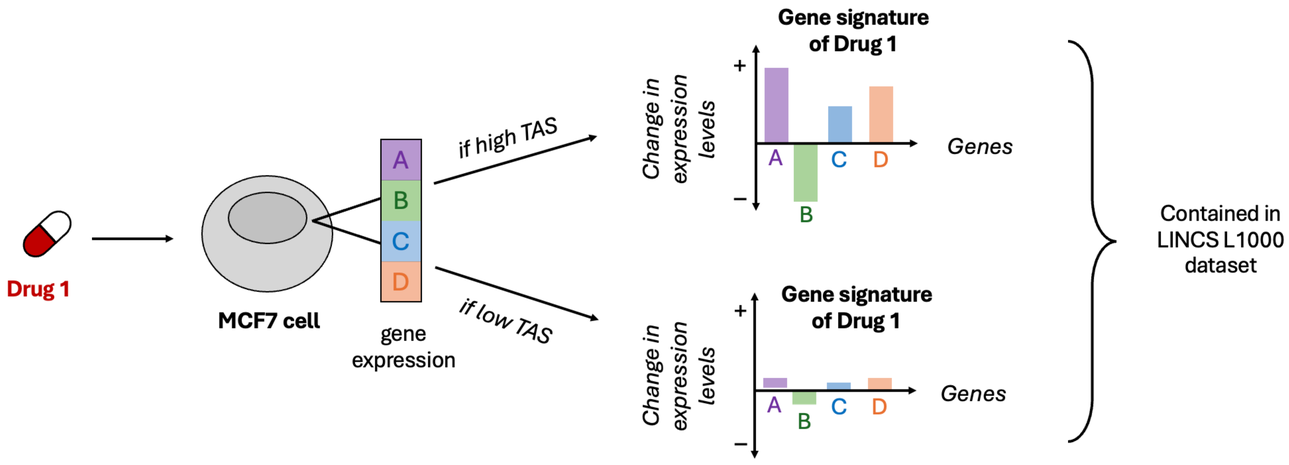
Katie Huang[1], Rachel D. Melamed[2]

[1] Department of Biomedical Engineering and Biotechnology, University of Massachusetts Lowell

[2] Department of Biological Sciences, University of Massachusetts Lowell

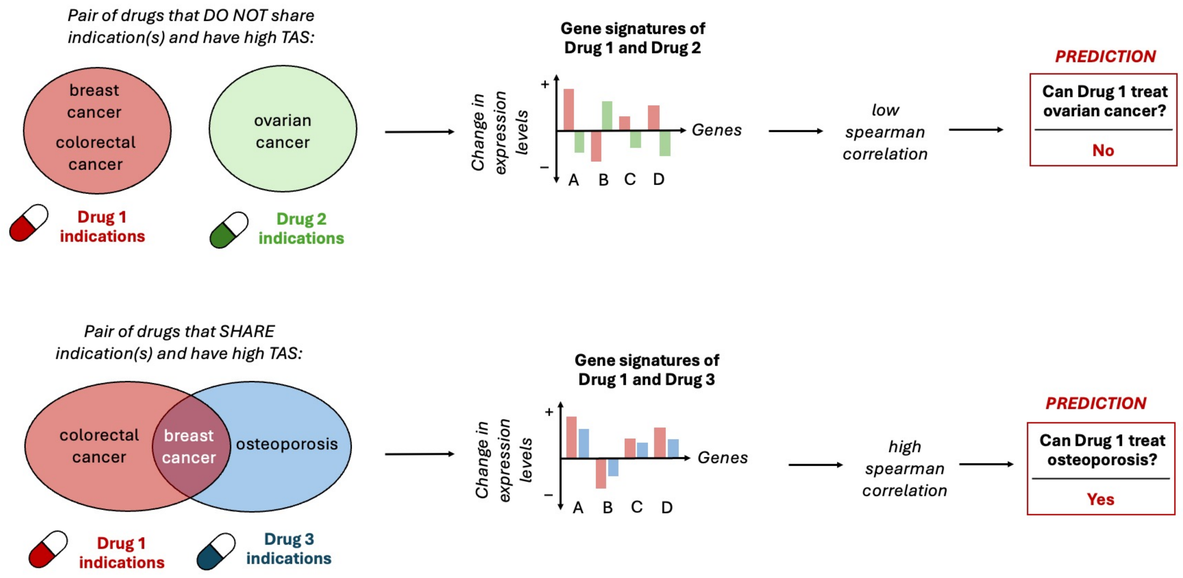
INTRODUCTION

Drug repurposing offers a quicker and potentially more cost-effective approach to finding new treatments by reusing existing drugs for alternative therapeutic indications. To accelerate the discovery of drug repurposing, the LINCS L1000 project has measured the *in vitro* effect of hundreds of drugs on gene expression across many cancer cell lines, some with high transcriptional activity scores (TAS) and others with low TAS. Low TAS in samples suggests that the drug has a minimal impact on gene expression, while high TAS indicates that the drug induces significant changes in gene expression.



HYPOTHESIS

HYPOTHESIS: If a pair of drugs ***treat the same disease***, and both affect the cell (as measured by TAS), then they may have ***similar changes in gene expression***. Then, we can *predict* new indications using *similarity in gene expression.*



METHODS

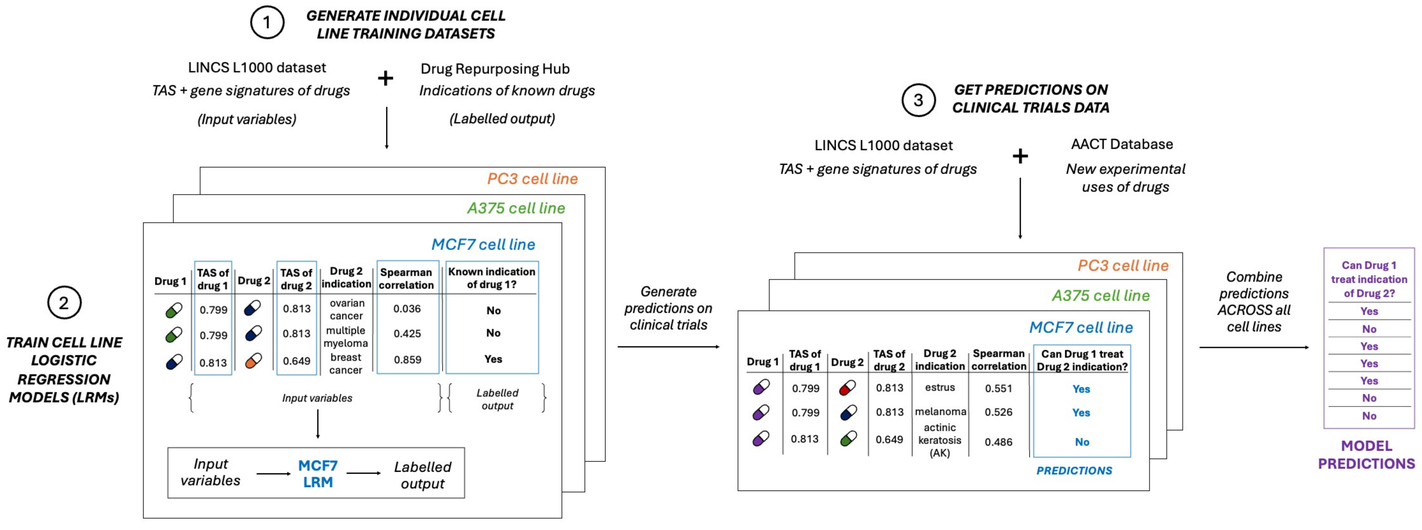


Fig 1. Overview of the process of data curation, training and assessing model's performance on clinical trials. The Level 5 data from the LINCS L1000 dataset was used to create individual datasets for three top cell lines (MCF7, A375, and PC3), filtering for drugs inducing strong expression changes as measured by transcriptional response (TAS) and known indications recorded in the Drug Repurposing Hub. The spearman correlation was then computed between drugs and indicated drugs for each indication, using the highest correlation and TAS values to predict indications with logistic regression. Logistic regression models for each cell line dataset predicted new experimental uses from the AACT database. A weighted ensemble approach was then taken to combine predictions across cell lines by calculating a weighted average of probabilities.

RESULTS

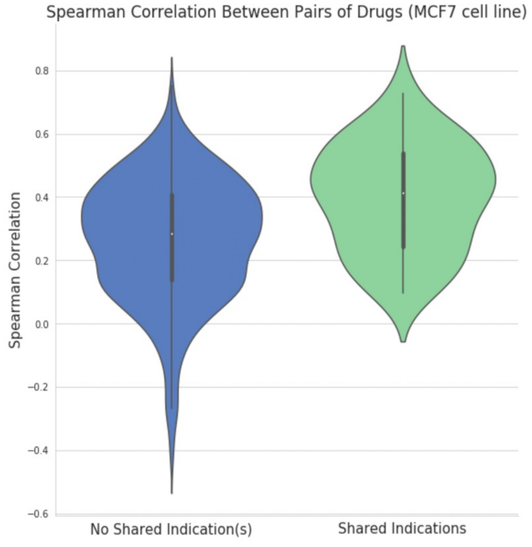


Fig 2. Spearman correlation between the changes in gene expression in pairs of drugs that had transcriptional response (TAS > 0.2) for MCF7 cell line. Pairs of drugs that shared at least one indication had more similar changes in gene expression compared to pairs that shared no indications.

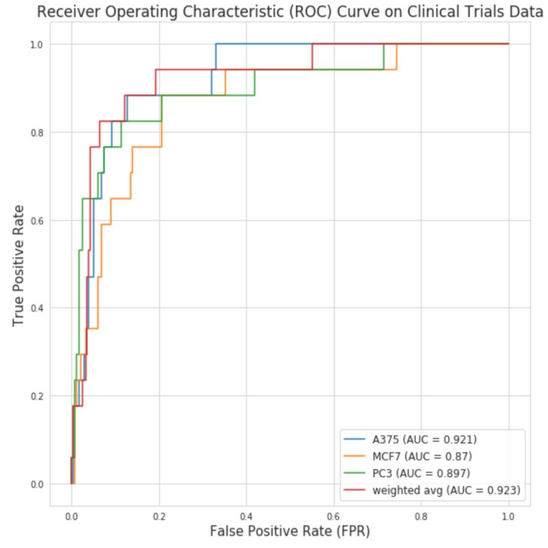


Fig 3. ROC curves with AUC scores for individual cell line logistic regression models and the weighted average ensemble model on shared drug indications from clinical trials. The ensemble model exhibited an AUC of 0.72 on all experimental uses.

SCAN ME

Contact Information

The QR code provided on the left contains additional resources and source code for this project. If you have any further questions or want to learn more about the project, feel free to contact Rachel_Melamed@uml.edu or Katie_Huang@student.uml.edu.

References

1. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).
2. Chen, B. *et al.* Reversal of Cancer Gene Expression Correlates With Drug Efficacy and Reveals Therapeutic Targets. *Nat Commun* **8**, 16022 (2017).