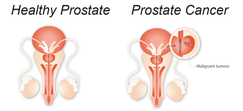
PROC-DOT: Prostate Cancer Drug-pair Optimization via Transcriptomics

Gargi Tawarmalani

INTRODUCTION

Prostate Cancer (PC) Statistics:

- 375,000 deaths worldwide (2020)
- 1 in 8 men get PC in lifetime
- $1.7 billion spent on PC prescription drugs but single drug treatment resistance often develops



Source: adapted from https://stock.adobe.com/search/images?k=prostate%28cancer&asset_id=604880030

Figure 1: Comparison of Healthy Prostate and Tumored Prostate

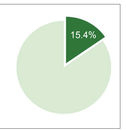
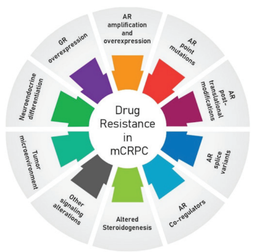


Figure 2: Prostate Cancer as percentage of new cancers in US (2025)



Source: https://pubmed.ncbi.nlm.nih.gov/36176747/#&gid=article-figures&pid=figure-1-uid-0

Figure 3: Causes of mCRPC

Resistance & Combination Therapy:

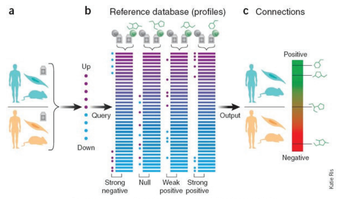
- Agressive metastatic Castrate Resistant PC (mCRPC) develops with single drug treatments
- To combat this, combination therapy is used (2+ drugs)
- Drug combination therapy has proven effective in treating cancers like lymphoma
- Hard to find effective combinations

Does the additive model predict genetic expression for drug combinations in cancer?

LITERATURE REVIEW

CMAP Query:

- Can be used to identify drugs that have certain desired genetic effects (like reversing a disease's genetic effects)
- Limitations: Does not account for gene expression magnitude only looks at genes' up and down regulation

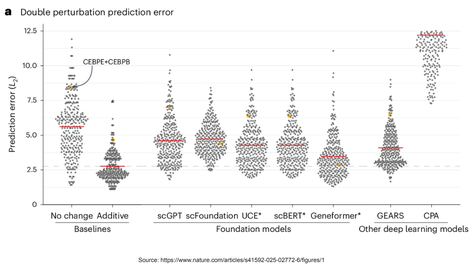


Source: https://media.springernature.com/full/springer-static/image/art%3A10.1038%2Fnchembio1206-663/MediaObjects/41589_2006_Article_BFnchembio1206663_Fig1_HTML.jpg

Figure 4: (a) Gene Expression for Disease vs Normal (b) Query takes in up and down regulated genes (c) Connect input signature to small molecule perturbations

Neural Network Models/Deep Learning:

- Predicts the combined genetic effect of gene modifications such as those from two drugs
- Limitations:
  - Not interpretable
  - Needs large datasets to run
  - Additive model outperforms (shown below)

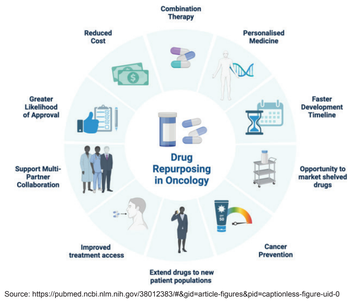


Source: https://www.nature.com/articles/s41592-025-02772-6/figures/1

Figure 5: Comparison of Current Models to Additive Model in Terms of Prediction Error

PROJECT GOALS & DESIGN

- Predict effective drug combinations for PC
- Quantify drug effects using magnitude of gene expression
- Develop an interpretable model that can leverage the limited genetic data available on drug combinations
- Repurpose FDA approved drugs from many other disease areas

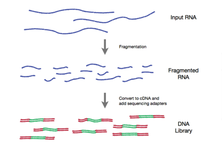


Source: https://pubmed.ncbi.nlm.nih.gov/38012383/#&gid=article-figures&pid=captionless-figure-uid-0

Figure 6: Possible benefits of drug repurposing

Datasets	Data Type	Cell Line	Samples	Drugs	Genes	Combination Drugs
GSE199800	Bulk RNA-seq	LNCaP	900	419	39376	0
GSE70466	Bulk RNA-seq	LNCaP, PReC	6	0	39376	0
GSE206741	scRNA-seq	A549	63378 cells	13	27518	21

Table 1: Summary of Datasets Used



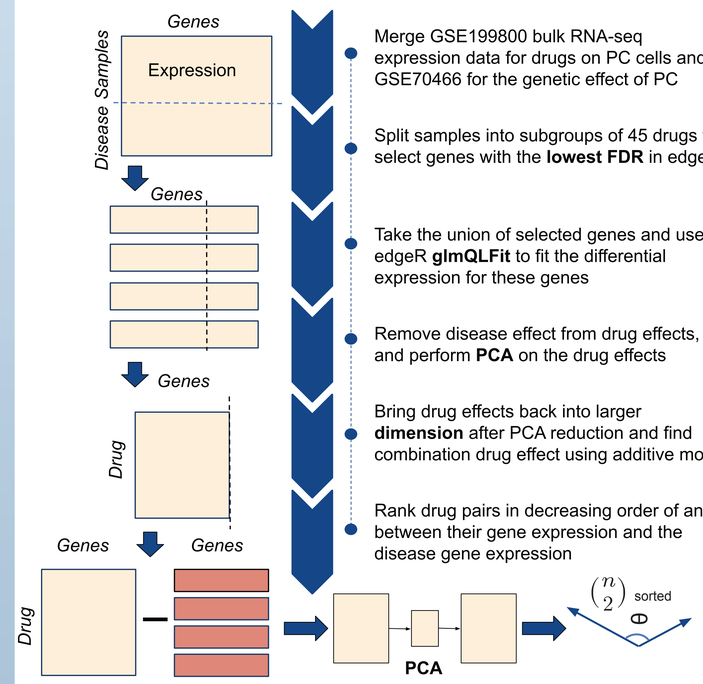
Source: https://rnaseq.uoregon.edu/img/fig-rna-seq.png

Figure 7: RNA-seq data preparation

Type	Variables
Control	Cell Lines, Genes per Study
Independent	Disease, Drug, and Plate
Dependent	Gene Expression (logFC)

Table 2: Variables Summary

METHODS



Merge GSE199800 bulk RNA-seq expression data for drugs on PC cells and GSE70466 for the genetic effect of PC

Split samples into subgroups of 45 drugs to select genes with the **lowest FDR** in edgeR

Take the union of selected genes and use edgeR **glmQLFit** to fit the differential expression for these genes

Remove disease effect from drug effects, and perform **PCA** on the drug effects

Bring drug effects back into larger **dimension** after PCA reduction and find combination drug effect using additive model

Rank drug pairs in decreasing order of angle between their gene expression and the disease gene expression

Validation

GSE206471 dataset with drug combinations

↓

scRNA-seq data to pseudo bulk data by aggregating gene counts

↓

edgeR analysis for differential gene expression of drugs using glmQLFit

↓

PCA on drug gene expression data, convert expression into counts data

↓

Compare actual and predicted gene expressions for combination drugs.

Analysis Software
Python: scanpy, anndata, pertpy,
R: edgeR, dplyr, tidyR, GEOQuery,
Illustrations: LaTeX/tikz, Illustrator, R/ggplot2, and Python/scanpy

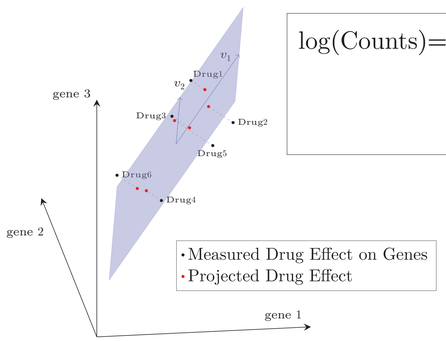


Figure 8: Depiction of PCA plot. Drug effects are projected onto principal component space as shown.

Model

$$\log(\text{Counts}) = \underbrace{\alpha}_{\text{control}} + D_{\text{disease}} \times \underbrace{\delta}_{\text{disease effect}} + D_{\text{drug}} \times \underbrace{\beta}_{\text{drug effect}} + D_{\text{plate}} \times \underbrace{\gamma}_{\text{plate effect}} + \underbrace{\log(L)}_{\text{eff. library}} + \varepsilon_{\text{error}}$$

$$\beta_{d,:}^{\text{disease}} = \beta_{d,:} - (\beta_{d,:} \cdot \delta)\delta^{\top}$$

$$\beta_{d,g}^{\text{center}} = \frac{(\beta_{d,g}^{\text{disease}} - \bar{\beta}_{:,g}^{\text{disease}})}{\sigma_{\beta_{:,g}^{\text{disease}}}}$$

$$\beta^{\text{center}} \approx \beta^{\text{approx}} = US_kV^{\top} \quad (\text{use } k \text{ singular values})$$

$$\beta_{d,g}^{\text{recon}} = \beta_{d,g}^{\text{approx}} \times \sigma_{:,g}^{\text{disease}} + \bar{\beta}_{:,g}^{\text{disease}} + \langle \beta_{d,:}, \delta \rangle \delta^{\top}$$

RESULTS

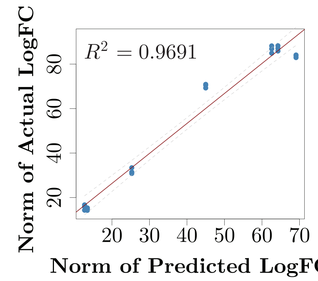


Figure 9: Additive model predicts genetic expression for drug combinations in lung cancer with high accuracy

Model predicts not only up and down regulation but magnitudes well

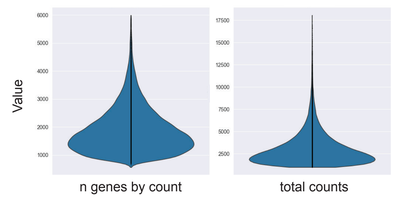


Figure 10: Data Visualization for scRNA Data

Angles between predicted and actual vectors were all less than 10 degrees

Drug1	Drug2	Angle	Norm
ARN509	Baricitinib	140.81	20.39
ARN509	Pixantrone	140.06	21.38
ARN509	Pilaralisib	140	20.99
Gossypol	ARN509	139.59	21.17
ARN509	Propranolol	138.98	21.56

Table 3: Top Predicted Combinations Across 419 Drugs in Sample Dataset

Drug1	Drug2	Angle	Norm	Rank
ARN509	Rucaparib	133.99	23.6	1
Prednisone	ARN509	131.05	24.5	2
ARN509	Talazoparib	129.06	25.12	3
Docetaxel	Rucaparib	102.7	31.74	9
Prednisone	Rucaparib	102.33	31.5	10

Table 4: Selected Drug Combinations From Top Ranked Prostate Cancer Drugs

ARN509 and Baricitinib:
Baricitinib inhibits JAK pathways that combat ADT-resistant prostate cancer. Since ARN509 is an androgen blocker, this combination shows promise

ARN509 and Pilaralisib:
Pilaralisib inhibits the PI3K/AKT pathway, while ARN509 inhibits ARs. Literature states that AR and AKT inhibition combination therapy might be advantageous

ARN509 and Rucaparib:
Rucaparib is a PARP inhibitor, while ARN509 is an AR-receptor targeted based therapy. PARP inhibitors and AR based therapies show promise in combination therapy

ARN509 and Prednisone:
These drugs have been tested in combination with Abiraterone as a drug combination

CONCLUSIONS & FUTURE WORK

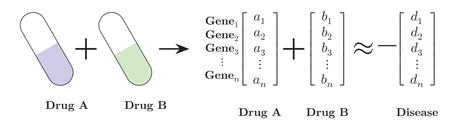


Figure 11: Summarized Steps of Methods

Benefits of PROC-DOT:

- Model is **interpretable**
- Has been **validated**
- Predicts **new combinations** from repurposed and PC drugs
- Utilizes genetic expression **magnitude and direction**

Future Work:

- Study the **effect of dosage** on combination drugs
- Generalize to predict **combinations for other diseases**

Impact:

- **Reduces cost** and testing in wet labs
- Allows **drug repurposing** of approved drugs
- Reduces time to market
- **Helpful in combating agressive PC**

KEY REFERENCES

Constantin Ahlmann-Eltze, Huber, W., & Anders, S. (2025). Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. Nature Methods, 22(8), 1657–1661. https://doi.org/10.1038/s41592-025-02772-6

Holmes, S., & Huber, W. (2026). 7 Multivariate Analysis – Modern Statistics for Modern Biology. Modern Statistics for Modern Biology. https://www.huber.embl.de/msmb/07-chap.html

Koudijs, K. K. M., Terwisscha van Scheltinga, A. G. T., Böhringer, S., Schimmel, K. J. M., & Guchelaar, H.-J. (2019). Transcriptome Signature Reversion as a Method to Reposition Drugs Against Cancer for Precision Oncology. The Cancer Journal, 25(2), 116–120. https://doi.org/10.1097/ppo.0000000000000370

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D., & Liberzon, A. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell, 171(6), 1437-1452.e17. https://doi.org/10.1016/j.cell.2017.10.049