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

Gargi Tawarmalani

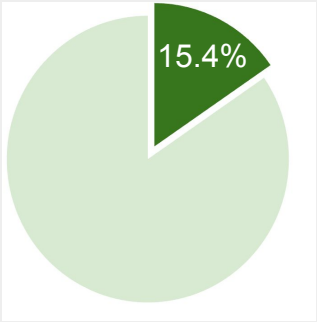
Unless otherwise noted, all table, graphs, and charts were created by the author

Introduction & Motivation for Project

Prostate Cancer (PCa) Statistics:

- 1 in 8 men get PCa in lifetime
- 375,000 deaths worldwide (2020)

Prostate Cancer as Percentage of New Cancers in US (2025)



Current limitations:

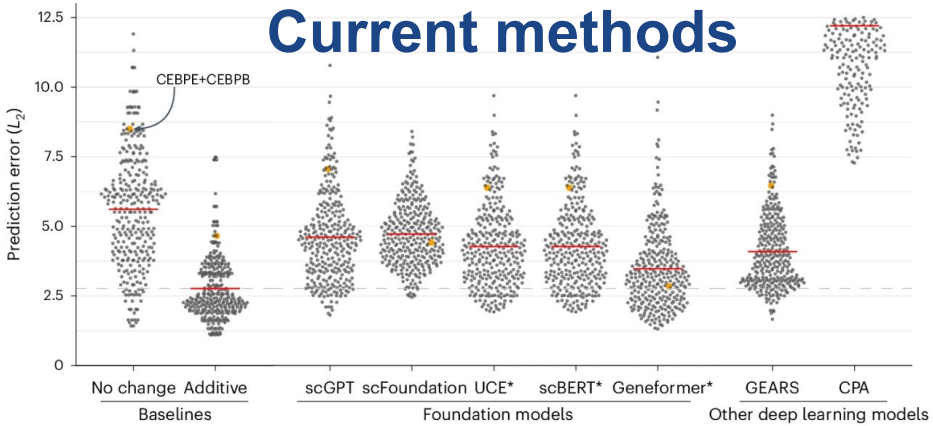
- Single drug leads to drug resistance
- Combination therapies combat resistance but hard to test in labs

Individual drug genetic data exists. PROC-DOT explores its use.

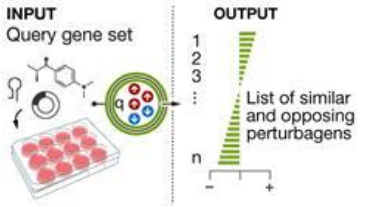
Hypothesis: Additive model predicts genetic expression for drug combinations in cancer

Introduction: Current Methods and Project Goals

Current methods



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



Source: https://pmc.ncbi.nlm.nih.gov/articles/PMC5990023/figure/F2/

Limitations:

1. Need lots of data
2. Not interpretable (Neural-net)
3. Do not use expression magnitude

Project Goals:

- Predict effective drug combinations for PCa
- Develop interpretable models that leverage limited available data on drug combinations
- Repurpose already approved drugs from all disease areas

Methodology: Process Flow

Step 1

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

Step 2

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

Step 3

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

Step 4

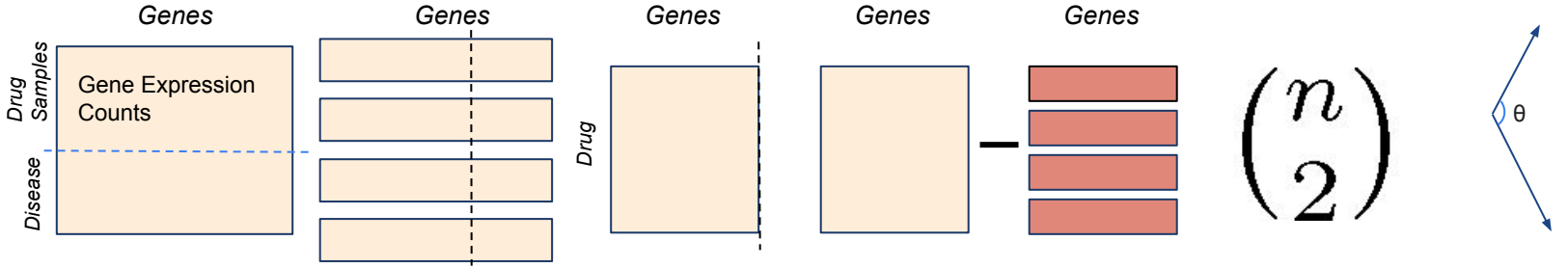
Remove disease effect from drug effects and perform Principal Component Analysis (**PCA)** on the drug effects

Step 5

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

Step 6

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



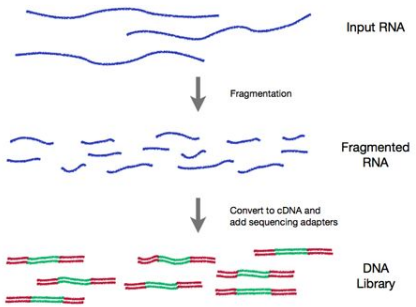
Methodology: Design & Validation Plan

Data Summary:

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

Validation

RNA-seq Data



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

Variables

Control: Cell Line

Independent: Gene Expression of each Drug

Dependent: Combination Drug Gene Expression

Software: EdgeR, Python

Use GSE206471 dataset with drug combinations

Convert scRNA-seq data to psuedobulk data by aggregating gene counts

edgeR analysis for differential gene expression of drugs using glmQLFit

PCA on drug gene expression data, lifted and converted differential expression into counts data

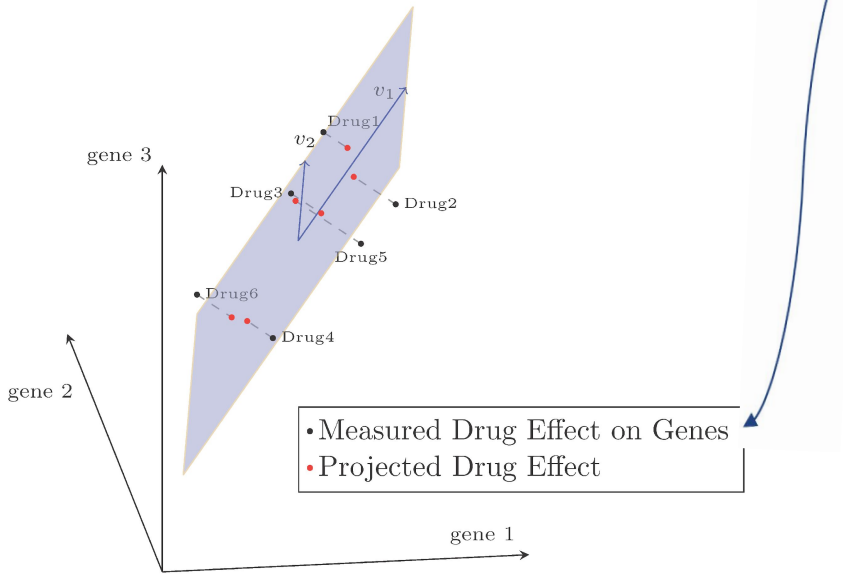
Compared actual and predicted gene expressions for combination drugs. Predictions were in principal component space

Methodology: Backend Models

Proposed Model (Fit by Negative Binomial) :

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

Depiction of PCA projection:



Dimension reduction using PCA:

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

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

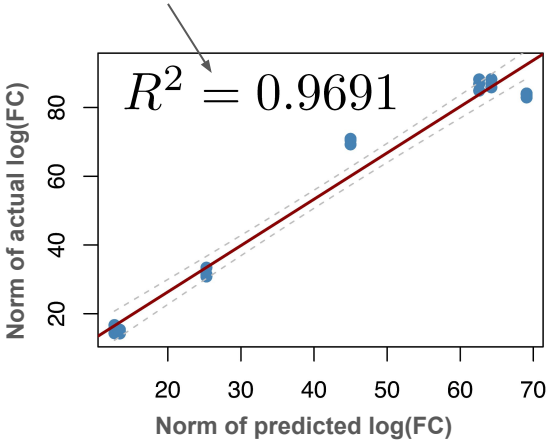
$$\beta^{\text{center}} \approx \beta^{\text{approx}} = U S_k V^{\top} \quad (\text{use } k \text{ singular values})$$

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

Results: Validation from scRNA-seq (Pseudo-bulk)

Actual Gene Expression of Combination Drugs Compared to That Predicted By Model

High R^2 value shows high correlation



	Actual (Averaged across Wells)			Predicted		
Drug Combination	Component 1	Component 2	Component 3	Component 1	Component 2	Component 3
Dacinostat_Dasatinib	33.04	-0.92	-0.55	25.15	0.36	-3.98
Givinostat_Dasatinib	4.44	-1.11	-0.74	7.79	0.10	-1.86
Givinostat_SRT2104	4.74	-0.89	-1.19	3.93	-0.51	-1.01
Panobinostat_Alvespimycin	29.13	-2.10	8.22	29.66	2.89	2.69
Panobinostat_Dasatinib	35.94	-2.65	4.93	28.78	-1.30	2.08
Panobinostat_SRT2104	37.98	-2.09	3.00	24.91	-1.91	2.94
SRT2104_Alvespimycin	6.42	6.00	-0.23	4.89	4.93	-0.39

Model predicts direction of expression and magnitudes

Results: Prostate Cancer

Top combinations from all drugs

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

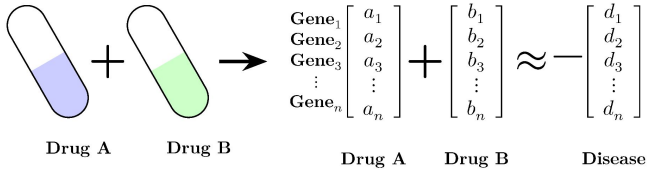
Top combinations of PCa drugs

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

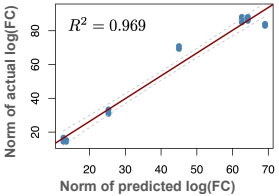
Literature Confirmation (combinations show promise):

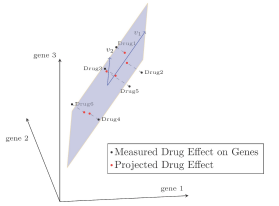
Example (ARN509 and Baricitinib): Baricitinib inhibits JAK pathways helpful in combating ADT-resistant prostate cancer due to ARN509. Combination shows promise (Deng, 2022).

PROC-DOT: Conclusions and Future Work



PROC-DOT predicts effect of drug combinations from individual drug data





Drug1	Drug2	Angle	Norm
ARN509	Baricitinib	140.81	20.39
ARN509	Pixantrone	140.06	21.38
ARN509	Pilaralisib	140.00	20.99
Gossypol	ARN509	139.59	21.17
ARN509	Propranolol	138.98	21.56
ARN509	SAR 245409.00	137.82	22.28
ARN509	Pyrrolidinedithiocarbamate ammonium	137.64	22.07
AUY922	ARN509	137.63	22.02
Carmustine	ARN509	137.46	22.10
ARN509	MK0752	137.33	22.11

Key Contributions

1. Validated
2. Interpretable
3. Predicts combinations

Limitation:

Validation with lab experiments

Extensions:

Account for dosage scaling of drugs

Impact:

1. **Reduces cost** for testing in labs
2. **Reduces time to market**

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

Deng, S., Wang, C., Wang, Y., Xu, Y., Li, X., Johnson, N. A., Mukherji, A., Lo, U., Xu, L., Gonzalez, J., Metang, L. A., Ye, J., Carla Rodriguez Tirado, Rodarte, K. E., Zhou, Y., Xie, Z., Arana, C., Annamalai, V., Liu, X., & Donald Vander Griend. (2022). Ectopic JAK–STAT activation enables the transition to a stem-like and multilineage state conferring AR-targeted therapy resistance. *Nature Cancer*, *3*(9), 1071–1087. https://doi.org/10.1038/s43018-022-00431-9

He, Y., Xu, W., Xiao, Y.-T., Huang, H., Gu, D., & Ren, S. (2022). Targeting signaling pathways in prostate cancer: mechanisms and clinical trials. *Signal Transduction and Targeted Therapy*, *7*(1). https://doi.org/10.1038/s41392-022-01042-7

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

Leslie, S. W., Soon-Sutton, T. L., Sajjad, H., & Siref, L. E. (2023, November 13). *Prostate Cancer*. National Library of Medicine; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK470550/

Murthy, M. (2021, June 11). *A Beginner's Guide to Singular Value Decomposition (SVD)*. Medium. https://mukundh-murthy.medium.com/a-beginners-guide-to-singular-value-decomposition-svd-97581e080c11

Rathkopf, D. E., Antonarakis, E. S., Shore, N. D., Tutrone, R. F., Alumkal, J. J., Ryan, C. J., Saleh, M., Hauke, R. J., Bandekar, R., Maneval, E. C., de Boer, C. J., Yu, M. K., & Scher, H. I. (2017). Safety and Antitumor Activity of Apalutamide (ARN-509) in Metastatic Castration-Resistant Prostate Cancer with and without Prior Abiraterone Acetate and Prednisone. *Clinical Cancer Research*, *23*(14), 3544–3551. https://doi.org/10.1158/1078-0432.ccr-16-2509

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

Tortorella, E., Giantulli, S., Sciarra, A., & Silvestri, I. (2023). AR and PI3K/AKT in Prostate Cancer: A Tale of Two Interconnected Pathways. *International Journal of Molecular Sciences*, *24*(3), 2046. https://doi.org/10.3390/ijms24032046