Predict the response of chemotherapeutics in clinical trials

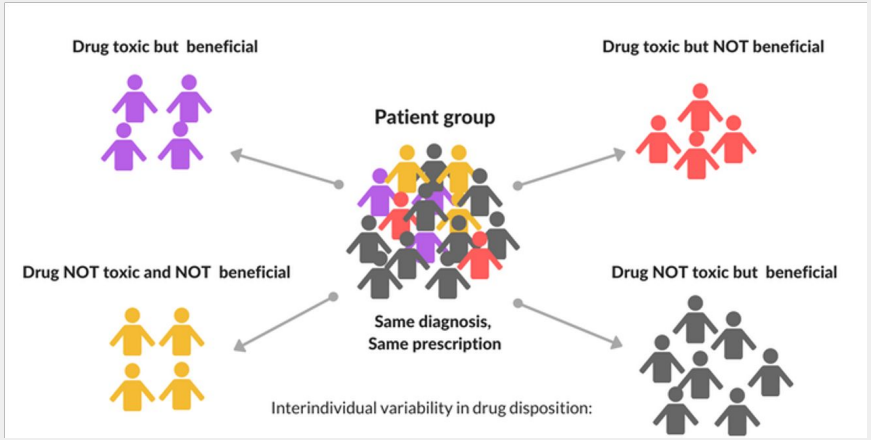
Yuting Shan
Eunice Songthangtham
Zhaoliang Zhou

Outline

- Introduction/Motivation
- About the data
- Data manipulation
- Methods
- Results
- Discussion/Future directions

Introduction

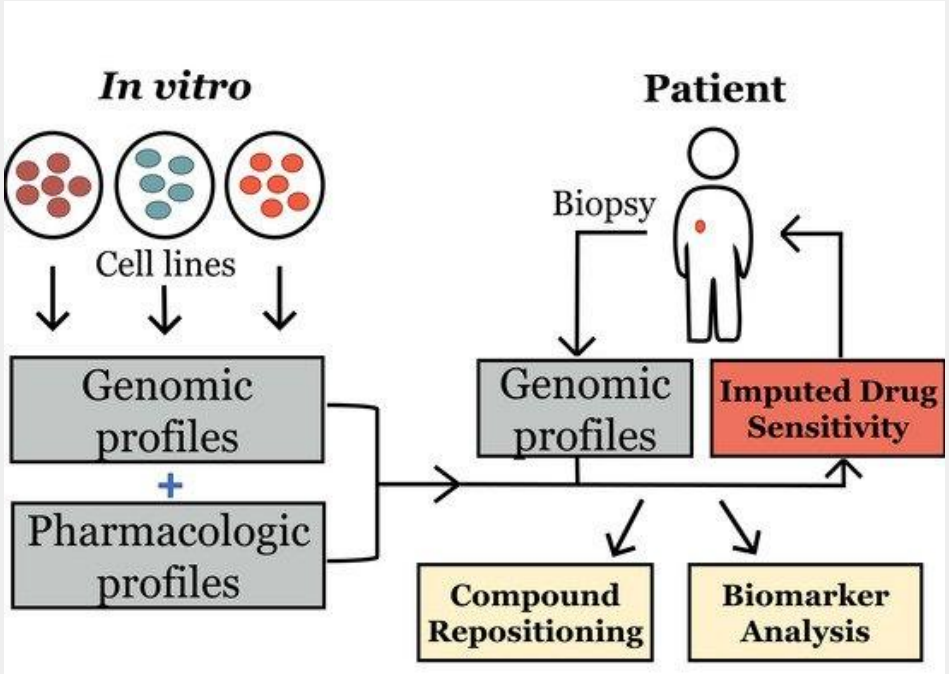
- Patients' responses to chemotherapeutic agents are highly variable. The side effects of those agents are potentially lethal (eg. febrile neutropenia)[1].
- It is very important to identify and apply molecular biomarkers to **predict** response to chemotherapeutics before patients physically receive them.
- Drug response is a very **complex trait**. Considering the cumulative effect of many biomarkers may better predict these complex phenotypes.



Introduction

- Based on paper: *Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines* by Paul Geeleher et al [2].

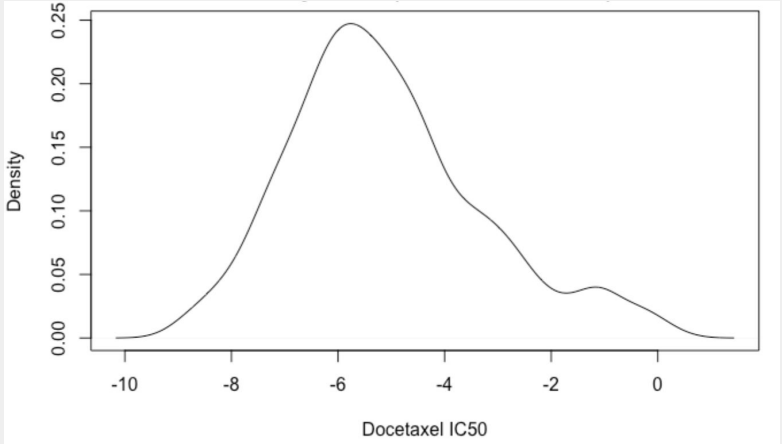
Training data

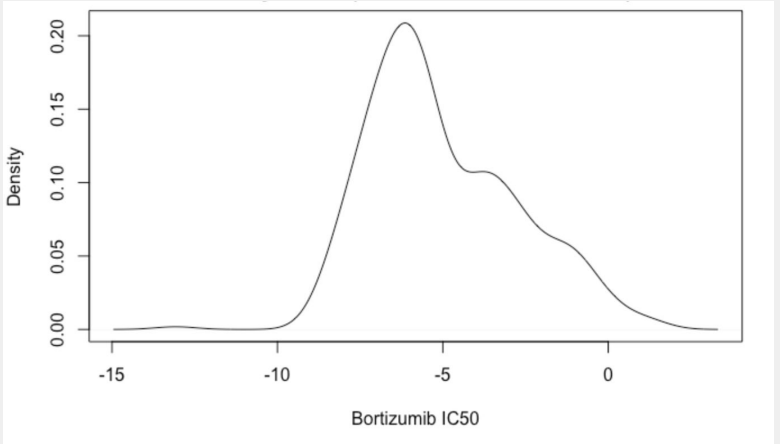


Test data

About the data - training data

- In vitro data: Cancer Genome Project (CGP):
  - Consists of almost 700 cancer cell lines.
  - Baseline gene expression microarray data: over 12,000 genes were sequenced.
  - Drug sensitivity data: a total of 138 drugs were screened over all the cell lines.
    - IC50: the concentration of the drug needed to inhibit cell viability by 50%.





About the data - test data

- In vivo data: clinical trial datasets with **tumor gene expression** before drug treatment and **drug response phenotype** (eg. responder, non-responder).
  - Trial 1: docetaxel in breast cancer
    - 24 breast cancer biopsies
    - 10 samples are sensitive to docetaxel (<25% tumor remaining), 14 samples are resistant to docetaxel (>25% tumor remaining).
    - 8,399 genes were sequenced
  - Trial 2: bortezomib in myeloma
    - 169 bone marrow biopsies
    - 85 samples are sensitive to bortezomib, 84 samples are resistant to bortezomib.
    - 22,645 genes were sequenced

Objective

- The paper used ridge regression to fit their model. We want to try different methods including lasso, elastic net, SVM, neural networks, and PCR to compare the performance of those different models.
- The computational efficiency of different models will also be compared.
- We aim to use lasso regression to do variable select in order to identify the potential biomarkers in predicting in vivo drug response.

Data manipulation

- Same data preprocessing as the paper:
- Robust multi-array average algorithm (*rms()* function in the affy library)
- Background correction, quantile normalization, and median-polish summarization
- Map cell line data and clinical trial data to official gene symbols
- Subset genes that are present in both datasets
- Homogenized two datasets using *Combat()* from sva library
- Removed genes with lowest variability because they never add predictive value

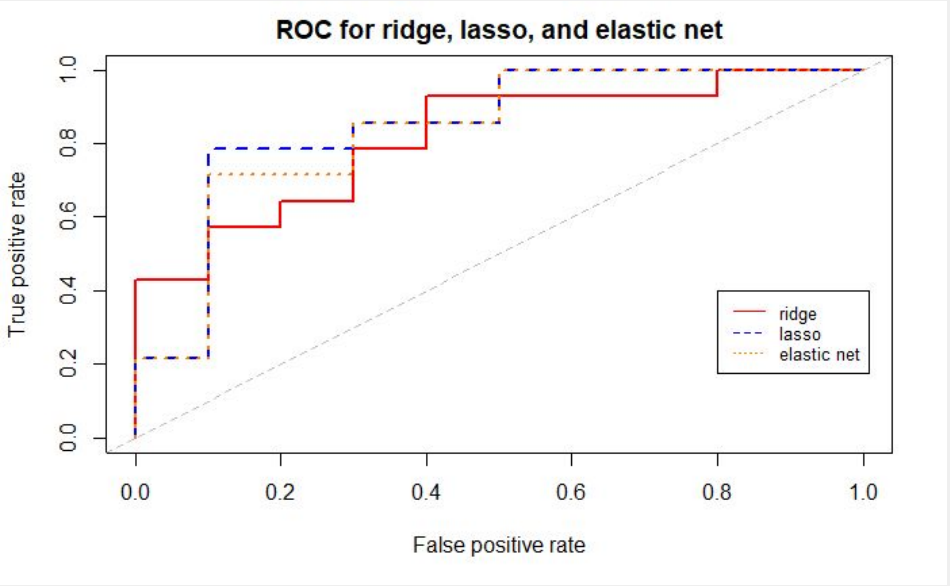
Methods

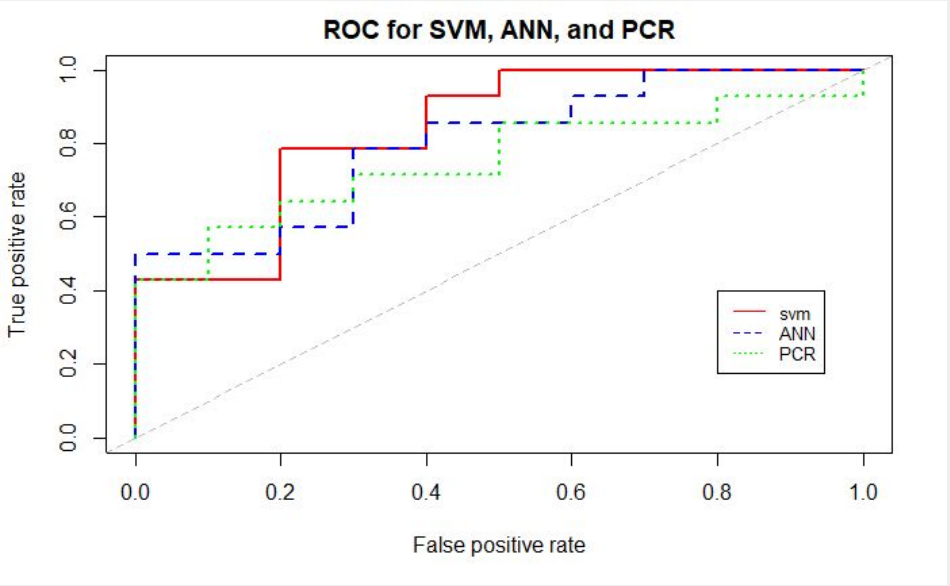
- Statistical methods implemented:
  - Ridge (linearRidge and glmnet)
  - Lasso
  - Elastic net
  - SVM
  - Artificial Neural network (ANN)
  - PCR
- Use AUC to determine performance and to compare with the performance from the publish paper
- Tuning:
  - ridge/lasso/elastic net: default 10-fold CV with *cv.glmnet()*
  - SVM/ANN: default with *tune()*
  - PCR: number of components that minimize RMSEP
- Computing environment: R version 4.1.3

Results

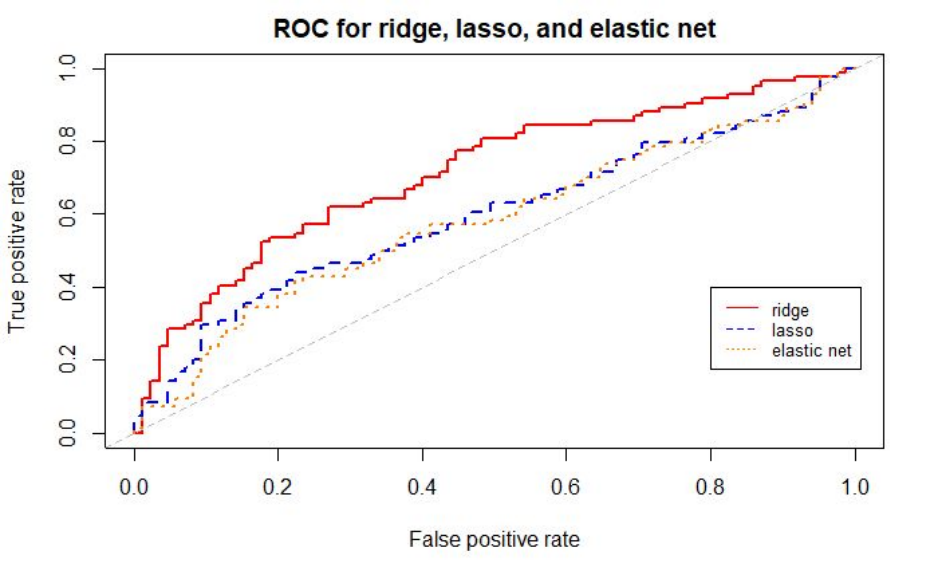
Methods/Outcome	Docetaxel	Bortezomib
Ridge (paper)	0.814	0.711
Ridge (glmnet)	0.814	0.714
Lasso	0.850	0.581
Elastic net	0.836	0.581
SVM	0.836	0.677
ANN	0.821	0.579
PCR	0.75	0.591

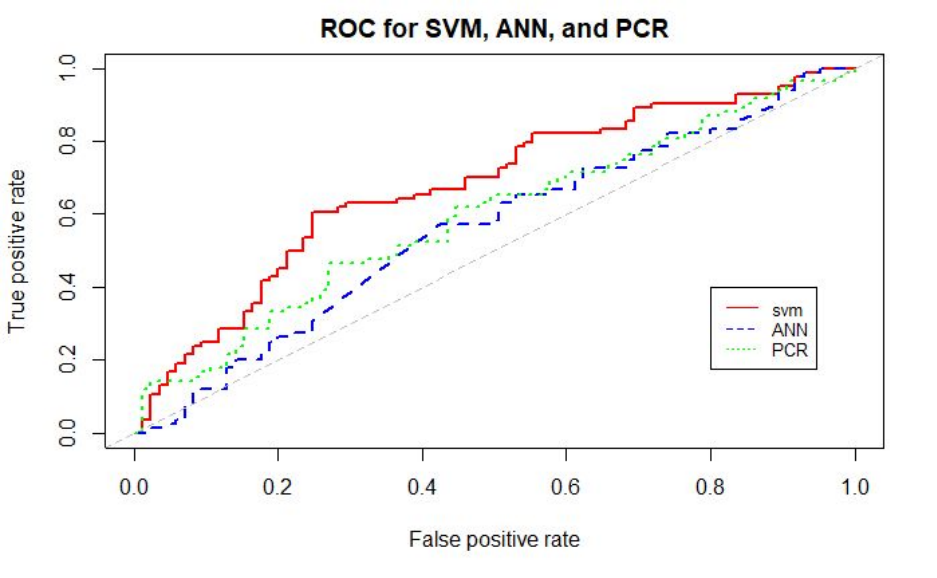
Docetaxel - ROC curves





Bortezomib - ROC curves





Discussions

- GLI3 gene common to Docetaxel and Bortezomib
- Top 2 predictors

Docetaxel	Bortezomib
ABCB1: translocate drugs across membranes BCAT2: suppresses cell death	RBM12B: tumor progression ARID5B: dysfunctions may facilitate tumorigenesis

- Limitations
  - Didn't work for Cisplatin
  - Author used microarray data
  - If the method outputs the predicted IC50 of the drug of interest, we need to come up with a cut off to separate responders and non-responders.
  - There might not be other options if the drug of interest is predicted to be ineffective.

Future Directions

- Fit the model using RNA sequencing data which is the most popular and more reliable sequencing method used currently.
- Personalized medicine
- Prevention of adverse drug reactions
- More in-depth research on particular genes and their effects on drug sensitivity

References

1. Mishra A, Verma M: Cancer biomarkers: are we ready for the prime time? Cancers (Basel) 2010, 2:190–208
2. Geeleher, Paul et al. “Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines.” Genome Biology 15 (2013): R47 - R47.
3. U.S. National Library of Medicine. (2020, August 18). Gli3 gene: Medlineplus genetics. MedlinePlus. Retrieved April 25, 2022, from https://medlineplus.gov/genetics/gene/gli3/#conditions
4. Matissek, S. J., & Elsawa, S. F. (2020). GLI3: a mediator of genetic diseases, development and cancer. Cell Communication and Signaling, 18(1), 1-20.
5. The Human Protein Atlas
6. https://arxiv.org/pdf/1205.0686.pdf

Thank you.

Any questions?