Ruprecht-Karls University of Heidelberg
Faculty of Engineering Sciences
BSc Molecular Biotechnology

Drug viability screens for oncological and non-oncological treatments for breast cancer

Data Science Project Summer Semester 2023

Topic 5 Team 4

Submission date: 17.07.2023

Luis Herfurth, Aaron Eidenmüller, Sharujan Suthakaran, Simon Westermann

Abstract

Hier muss das Abstract eingefügt werden

Contents

Abbreviations

1 Introduction

2 Materials and Methods

2.1 Data

2.1.1 Prism Datasets

Prism: effect of the treatment (columns) on cell growth of the cell lines (rows); includes drug, dosage and assay used

Prism.treat: for each treatment (rows) further information on the treatment and drug

Prism.cl: contains information about the different celllines
if we search after "breast" in the column lineage we get our 22 celllines

2.1.2 Cellline Datasets

Prism.exp: contains levels of gene expression. Celllines (rows) and genes (columns)

Prism.cnv: contains copy number levels of genes. Normal is CN = 2. Gene names (rows) and celllines (columns)

Prism.snv: marks mutation in the different celllines als functional or nonfunctional to the cancer.

Prism.achilles: has information on how important a gene is for cell survival. Was generated using knockdown celllines. Gene names (rows) and celllines (colums)

2.2 Data clean up/Filtering

Show distributions after cleanup Abbildung für prism vorher nachher. Für andere maybe im Clean up

2.3 imension reduction

UMAP -> PCA wird rausgelassen; Plot den Luis gemacht hat. Ist nicht geclustert, aber man kann erkennen, dass die MOA sich in den gleichen areas aufhalten

2.4 Clustering

kmeans -> Simon hat welche gemacht, aber die sehen kacke aus => Anna fragen

2.5 Statistical test

Wilcoxson test 0.05 alpha p adjust needed wegen sehr viele berechnungen und hoher fdr deswegen Shapiro Wilk

2.6 Lineare Regression

Prediction tool

3 Results

First include positiv results; if space is left include negativ results: UMAP, K means clustering, promoting drugs describe goal, describe process, describe outcome

3.1 Gene search engine

Goal: Arbeitsvereinfachung; Outcome: Overview over data Für Präsentation als Visualisierungstool pitchen Maybe Website

3.2 List of inhibitory drugs

Results von Data clean up und filtering. Goal: List of Inhibitory Drugs; Outcome: List of Inhibitory Drugs Bilder vergleich liste vergleich ohne threshold und mit threshold Maybe oncological drugs rein screenen

3.3 Gene analysis

3.3.1 Correlation analysis

treatment response / gene expression; Goal: finding relevant genes; Outcome: giant data frame -> used for further work

copy number / gene expression Goal: looking if hypothesis correct; Outcome: Histogram of correlations

3.3.2 Dataframe for targets involving genes

mean of data frames. Threshold for what genes are relevant. Used indings from correlation tests Goal: finding interesting genes; Outcome: Data frame with many genes -> 48 genes data set with filtering after gene knockout score

3.3.3 Statistical testing of important genes

Test wich of the found genes are for breast cancer of interest Goal: find out which one are negativ, which ones are lower than other lineages; Outcome: 2 genes

3.4 Linear regression

Perform drug by drug to avoid weird plot; For every drug one linear regression, R^2 Value and with those showing, that many of them are very good. Prediction model for concentration and drug name. Plot um das gnaze zu veranschaulichen; Goal: Regression/Prediction model; Outcome: Regression/Prediction model

4 Discussion

Search for papers mentioning certain genes found in the targets of the inhibitory drugs or

5 References

6 Appendix