

# BMI702 Final Project Report

*Niklas Rindtorff*

*4/15/2019*

## Final Project

### Introduction

Cancer is an extremely heterogeneous disease. While the availability of genomic testing of tumors is continuously increasing, linking the correct drug to the right patient remains a challenge. Part of this challenge is the difficulty to predict drug responses to a large number of drugs based on a tumors genetic profile.

In order to increase our understanding of how genetic markers predict drug response, laboratory tumor models, such as 2D cancer cell lines, can be treated with a large number of therapeutics in-vitro. The drug response of these tumor models, in combination with their genetic data, can then be associated.

### Part 1: The CancerRxGene Dataset

In this project I present the **CancerRxGene Dataset** by the Genomics in Drug Sensitivity in Cancer Project. It contains genomic information for >1000 cancer cell lines together with in-vitro drug response information. The data can be used to find genomic biomarkers of drug response in cancer cell lines. In the future, this dataset will be maintained by the DepMap project, a joint effort by the Wellcome Trust Sanger Institute and the Broad Institute. The data is publicly available. To the best of my knowledge, it is one of the greatest non-commercial drug sensitivity datasets in cancer models.

**Who publishes the dataset?** The data is published by the Genomics in Drug Sensitivity in Cancer Project. The project is a joint effort by the Cancer Genome Project, the Wellcome Trust Sanger institute and the Center for Molecular Therapeutics at MGH.

**Why do they publish it?** The data is funded by public entities, such as the Wellcome Trust and is published in scientific publications. To increase the access of the project's data, the data is freely available online.

**How frequently is it updated?** The data is updated in an irregular schedule, when a batch of cancer cell line vulnerability data has been collected and pre-processed. The last update of the dataset was published in March 2018.

**What variables does it contain?** The dataset contains multiple tables: \* meta-data about every cancer cell line, such as culture medium and tissue of origin. \* copy-number variants for every cell line, in the form of a binary matrix \* gene expression data for every cell line, in the form of a numeric matrix \* drug response data for every cell line. The data contains *IC50* values for >60 drugs.

**How is it delivered?** The data is available on the project's website and is split across multiple Excel files.

**Are there any restrictions on its use or availability?** Yes, the data is subject to restrictions: *"Users have a non-exclusive, non-transferable right to use data files for internal, non-commercial research and educational purposes Please note: The data files are experimental and academic in nature and are not licensed or certified by any regulatory body. Genome Research Limited provides access to data files on an "as is" basis and excludes all warranties of any kind (express or implied)."*

**Does it make use of any standard (or proprietary) ontologies or terminologies?** The data uses *COSMIC IDs* to identify cell lines. The gene expression data is annotated using the Ensemble nomenclature.

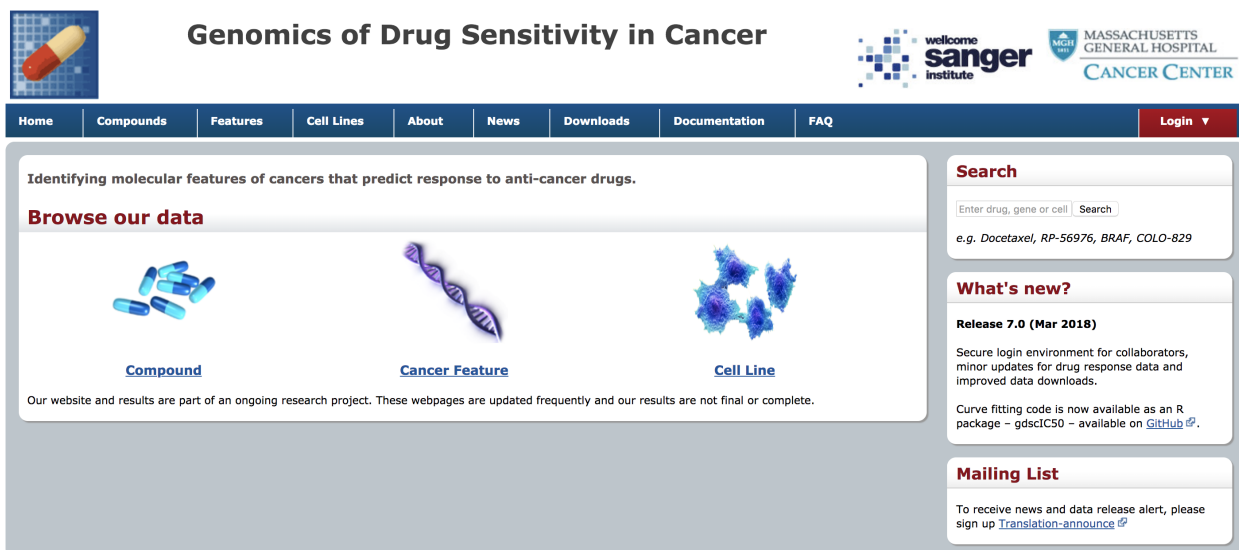


Figure 1: Landing page of the CancerRxGene Portal

The gene mutation dataset is annotated using the *HUGO* terminology. The drugs tested are not linked to a standardized ontology.

**What are some questions that have been asked previously using the data?** Previously, this data has been used to find predictive models that link genetic markers with drug response Iorio et al., 2016. In addition visualization projects have created applications to explore this dataset.

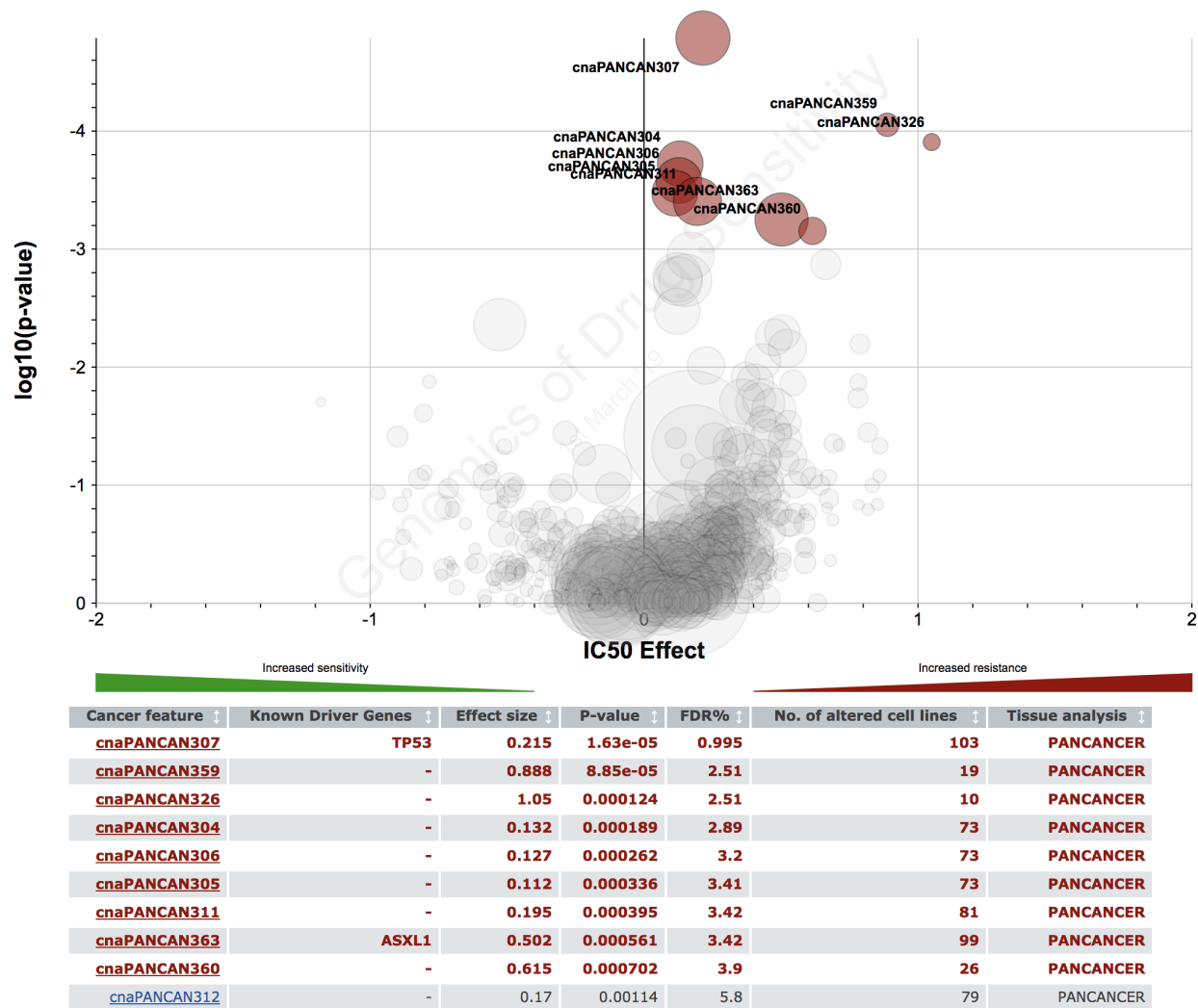


Figure 2: Exemplatory analysis in the CancerRxGene Portal: Which genomic features are associated with drug sensitivity to the chemotherapeutic 5-FU

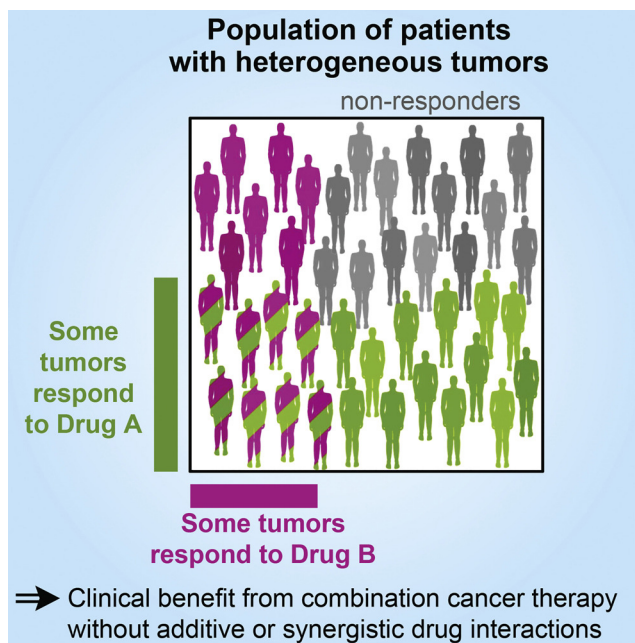


Figure 3: Visual Abstract of *Combination Cancer Therapy Can Confer Benefit via Patient-to-Patient Variability without Drug Additivity or Synergy* by Palmer et al., 2018

## Part 2: Identification of potential personalized cancer combination therapies under a non-synergy assumption

### Introduction

Recent simulation analysis of clinical trials assessing the activity of combination cancer therapies suggest, that the main driver of superiority of drug combinations over single-agent regimens is not biochemical synergy between drugs, but an increased probability of either one of the drugs causing a therapeutic effect in a patient's disease. According to this theory, combination treatments with agents that have anti-correlated response profiles are interesting candidate compounds for further clinical evaluation. In their original study the authors demonstrate that in fact the majority of currently administered combination treatments in oncology are not positively correlated.

In this study, my goal is to identify new potential combinations of therapeutics that are anti correlated in one of the largest in-vitro cancer drug sensitivity datasets. Because oncology today is a highly specialized field, it is not sufficient to only identify compounds that are anti-correlated from a pan-cancer perspective. Instead, I will run multiple correlation analysis on subsets of cancer cell lines defined by tissue of origin or mutation status. Because, the estimation of correlation coefficients is not stable with differing number of samples and subsetting of the dataset leads to very small subgroups, I have introduced a Bayesian sampling based method to estimate the credible intervals of correlation coefficients for small genetically defined subgroups. I have identified an anti-correlation of highly active MEK and AKT inhibitors across all available cancer types. Among many, Bayesian estimation of correlation coefficients identified NGFR receptor inhibition as a strong candidate combination for MEK inhibitors in Melanoma, thereby confirming prior reports of de-differentiation based drug resistance to MEK inhibition.

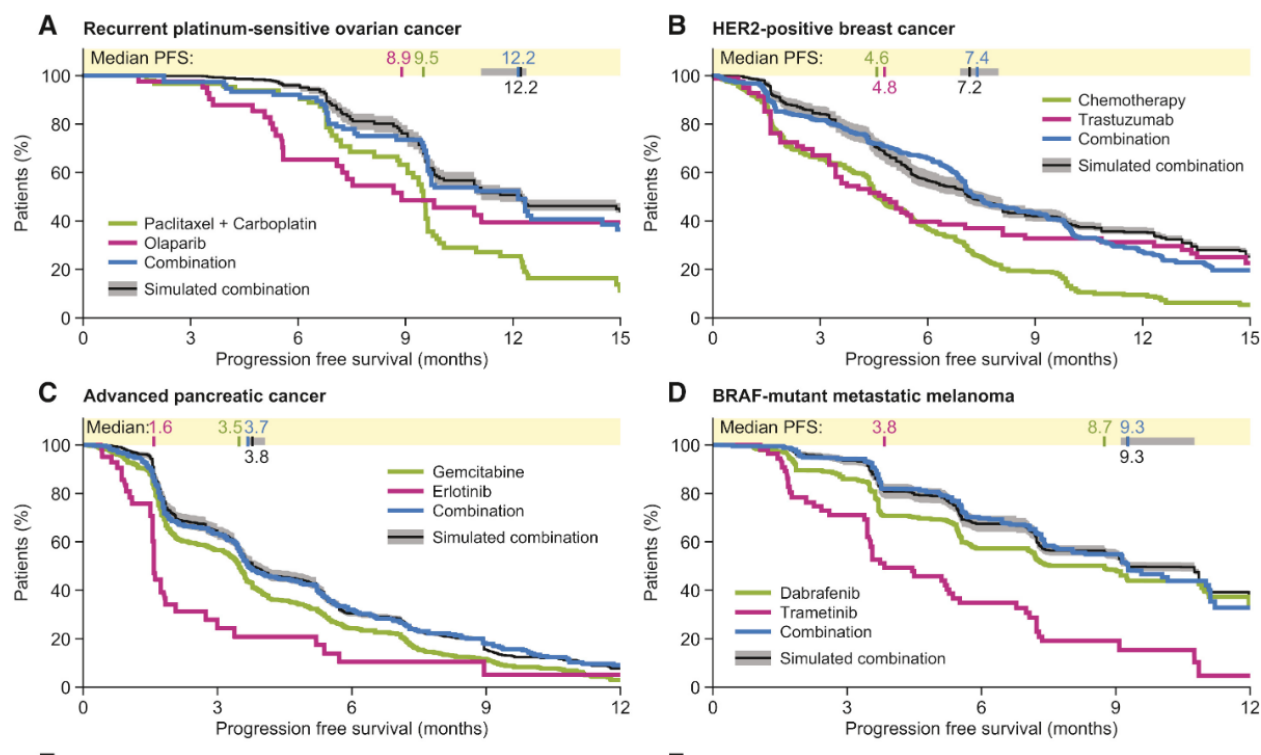


Figure 4: Simulation of combinatory drug effects by Palmer et al. suggest non-synergistic interactions of anti-cancer agents across drug- and cancer-types

### Correlation analysis of drug sensitivity profiles identifies mutually exclusive sensitivity to MEK and AKT inhibition.

We preprocessed the drug response data by log transforming and median centering IC50 values for each drug over every cell line. By transforming IC50 values this way, we are able to compare drug effects between both cell lines and drugs.

The two density plots show the distribution of IC50 values across all cell lines for each drug.

After preprocessing drug response scores, I performed Pearson correlation analyses of pairwise complete observations for each unique pair of drug sensitivity vectors.

I used available compound mechanism of action information to link every pair to the molecular targets it was inhibiting. Chi-Square testing of molecular targets in the significantly negative correlated pairs showed a strong enrichment for the combined inhibition of MEK and AKT.

In-depth analysis of the anti-correlation pattern between MEK and AKT inhibitors showed a mutually exclusive response pattern to either MEK or AKT inhibition, while the majority of tumor cells did not respond to both treatments.

Sensitivity to MEK inhibition and AKT inhibition has previously Been linked to the mutation status of biomarkers, such as *KRAS* or *PI3K*. In line with these associations, *KRAS* mutant cancer cells were more sensitive to MEK inhibition as were *PIK3CA* mutant cells to AKT inhibition.

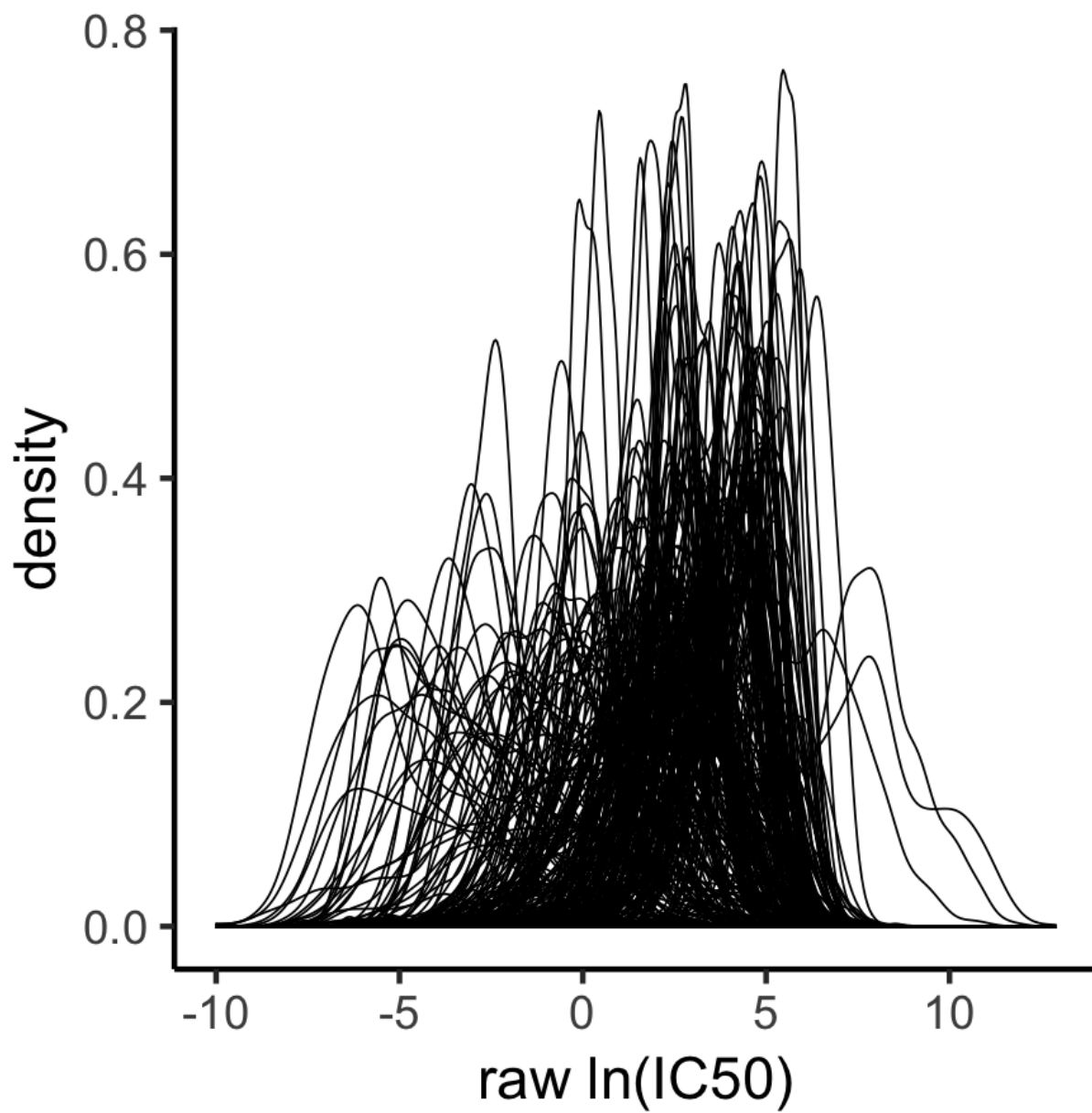


Figure 5: Distribution of  $\log(\text{IC}_{50})$  values before centering

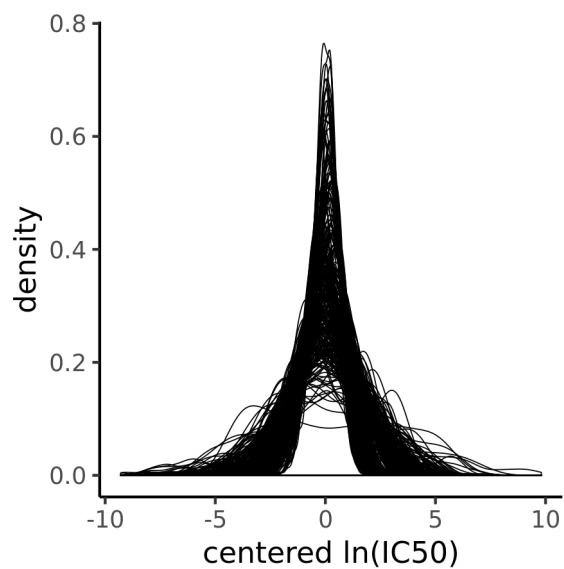


Figure 6: Distribution of  $\log(\text{IC}_{50})$  values after centering

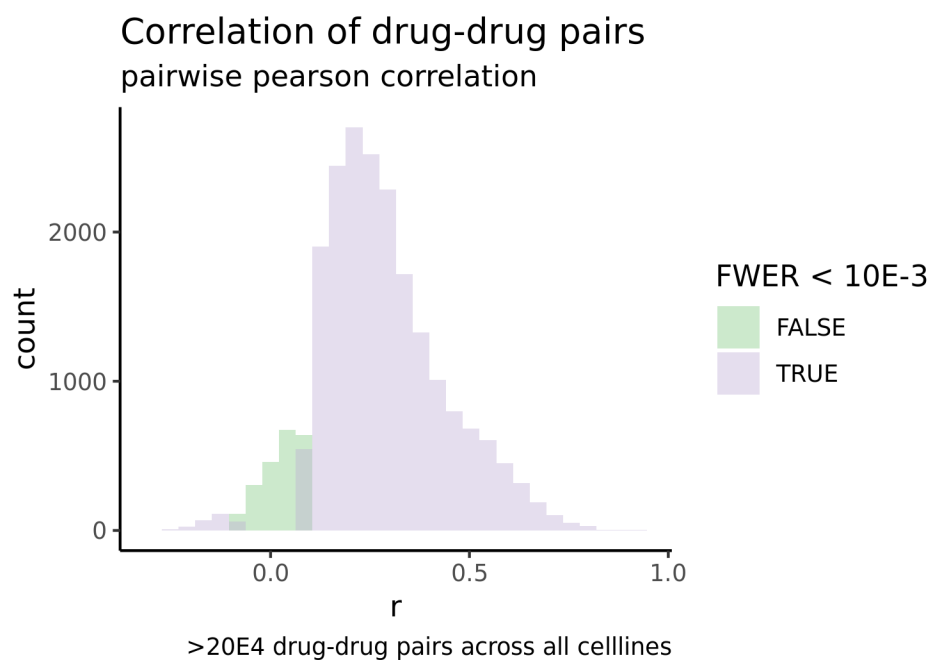


Figure 7: Correlation of drug-drug pairs across all possible combinations using all available cellline models

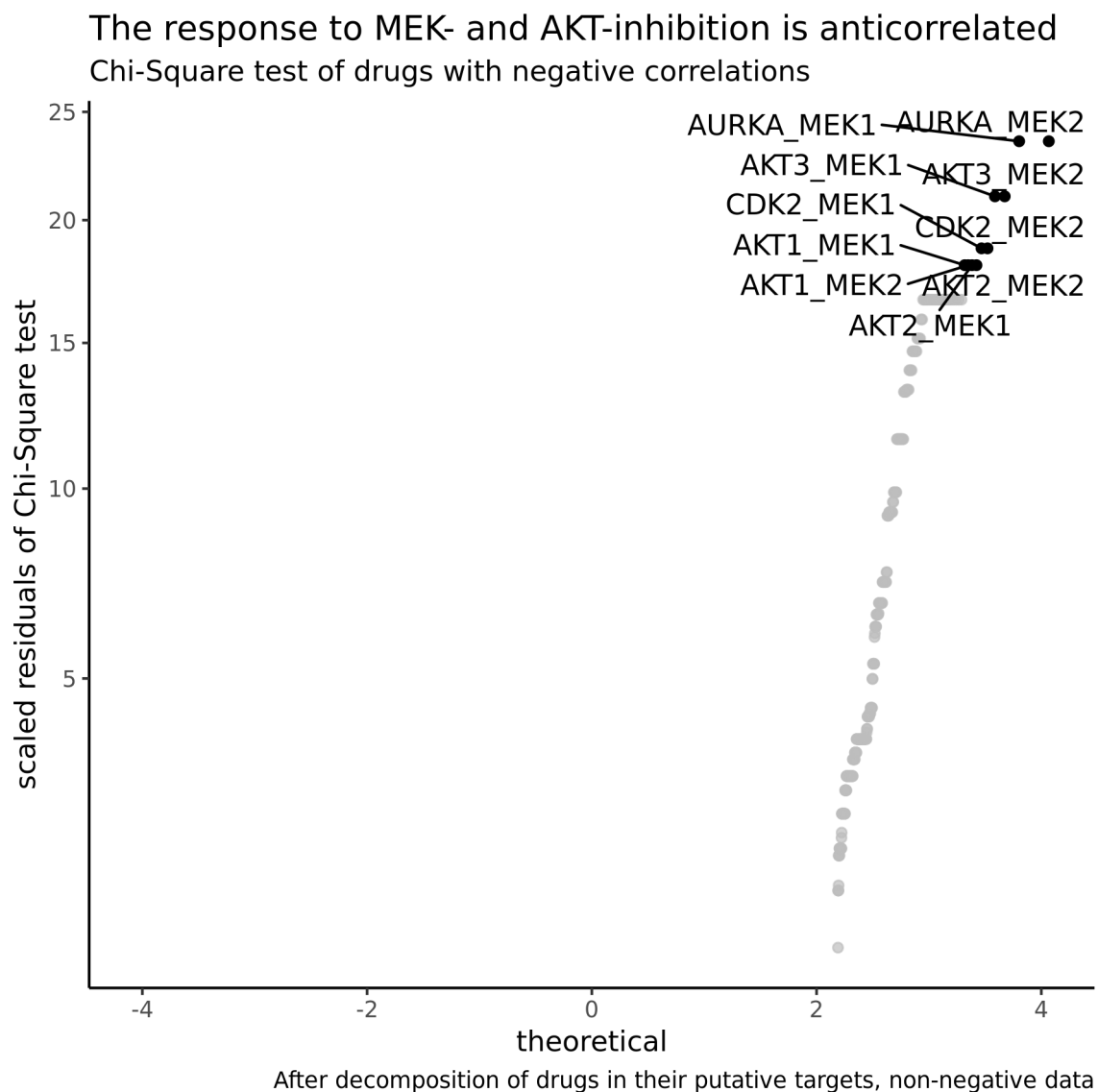


Figure 8: Enrichment of dual inhibition of MEK and AKT across all cancer cell lines. Shown are the residuals for each molecularly targeted pair after performing a Chi-Square test. Only values with a positive residual are plotted.



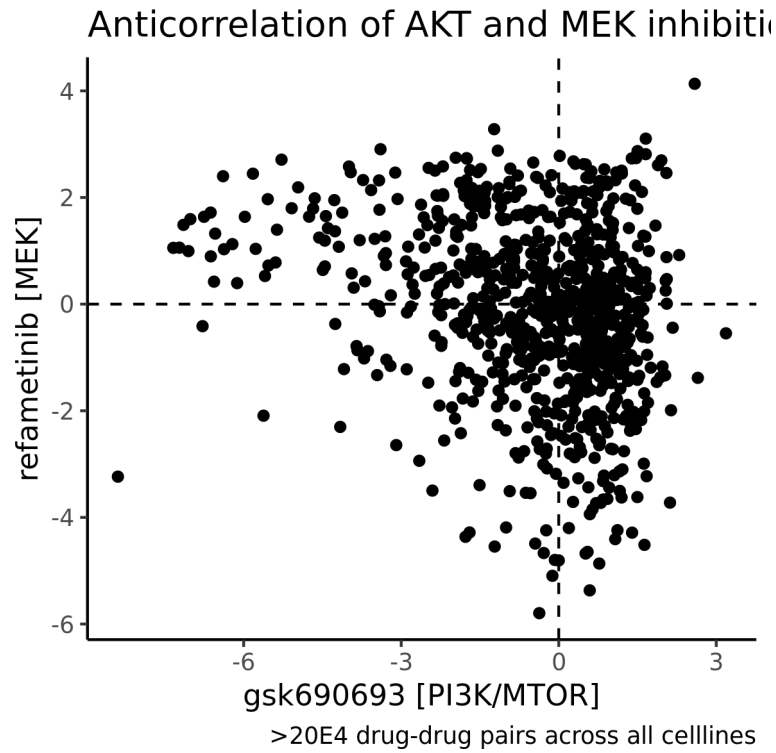


Figure 9: Anticorrelation of MEK- and AKT-inhibitor sensitivity

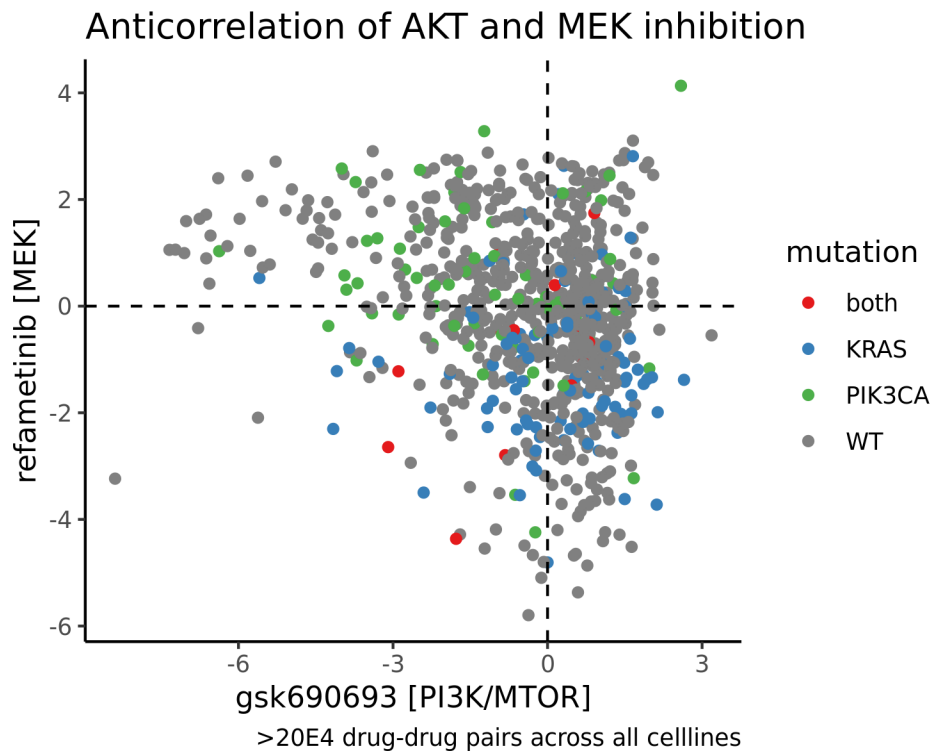


Figure 10: KRAS and PIK3CA status are linked to drug response under MEK and AKT inhibition

## Bayesian subgroup-restricted correlation analysis identifies potential tissue- and genotype-specific combination therapies.

As the previous example illustrates, known genotype- and tissue-specific predictors of drug sensitivity can underlie anti-correlation patterns, thereby reducing the clinical value of such drug combinations in a genetically-informed clinical practice. In order to identify drug combinations in more specific subgroups of cancer models, which share common genetic features of tissue types, I performed multiple correlation analyses in subgroups of cancer cell lines.

To this end, I identified more than 50 genotype or tissue related features that are at least present in more than 2% of all available cancer models. After grouping models by features and performing subgroup-specific correlation analyses I observed a strong heteroscedasticity of correlation coefficients, depending on the size of the respective subgroup.

To generate stable estimates of correlation, I generated sampling based credible intervals of correlation coefficients using a Bayesian correlation method which expects a bivariate t-distribution underlying a given dataset. Thus, it is less vulnerable to outliers, which are especially impactful in subgroups with low sample size. Because of restricted compute access (sampling of 60.000 drug pairs across subgroups took >24h on an instance with 16 CPUs) I was not able to estimate credible intervals for every drug-drug pair, but restricted the evaluation to pairs with a correlation coefficient <0 in the prior non-robust correlation analysis.

As a part of this analysis I identified a strong anti-correlation of *selumetinib* and *gw441756* in melanoma cancer models. Today, MEK inhibition is an available targeted treatment strategy of *BRAF* negative Melanoma. Recent evidence suggests that MEK inhibitor resistant melanoma cells enter a state of dedifferentiation, by over expression of NGFR, an interaction partner of NTRK1.

Another observation during the analysis of the results was an anti correlation of responses to the two pro-apoptotic compounds *r-TRAIL* and *Obatoclax* in *APC* mutant cancer cell lines. *APC* is a frequently mutated tumor suppressor gene in colorectal cancer. The identified anti-correlation can be explained by the two different mechanisms of apoptosis stimulation in human cells, the intrinsic pathway, which is modified by *Obatoclax* and the extrinsic pathway, which is influenced by *r-TRAIL*.

The timeline of this project did not allow for further in-depth analysis of potential combination therapies. I plan to pursue this project further and work towards a publication.

## Curating a database of drug response and genomic information of cancer models

In order to facilitate the use of the curated data, I created a compact SQLite database, which can be shared upon request. The database contains 4 tables:

- copy number variations
- mutations
- drug responses
- cell line covariates, such as tissue type

All tables are searchable by the *cosmic\_id*, a unique identifier for every cancer cell line in the dataset.

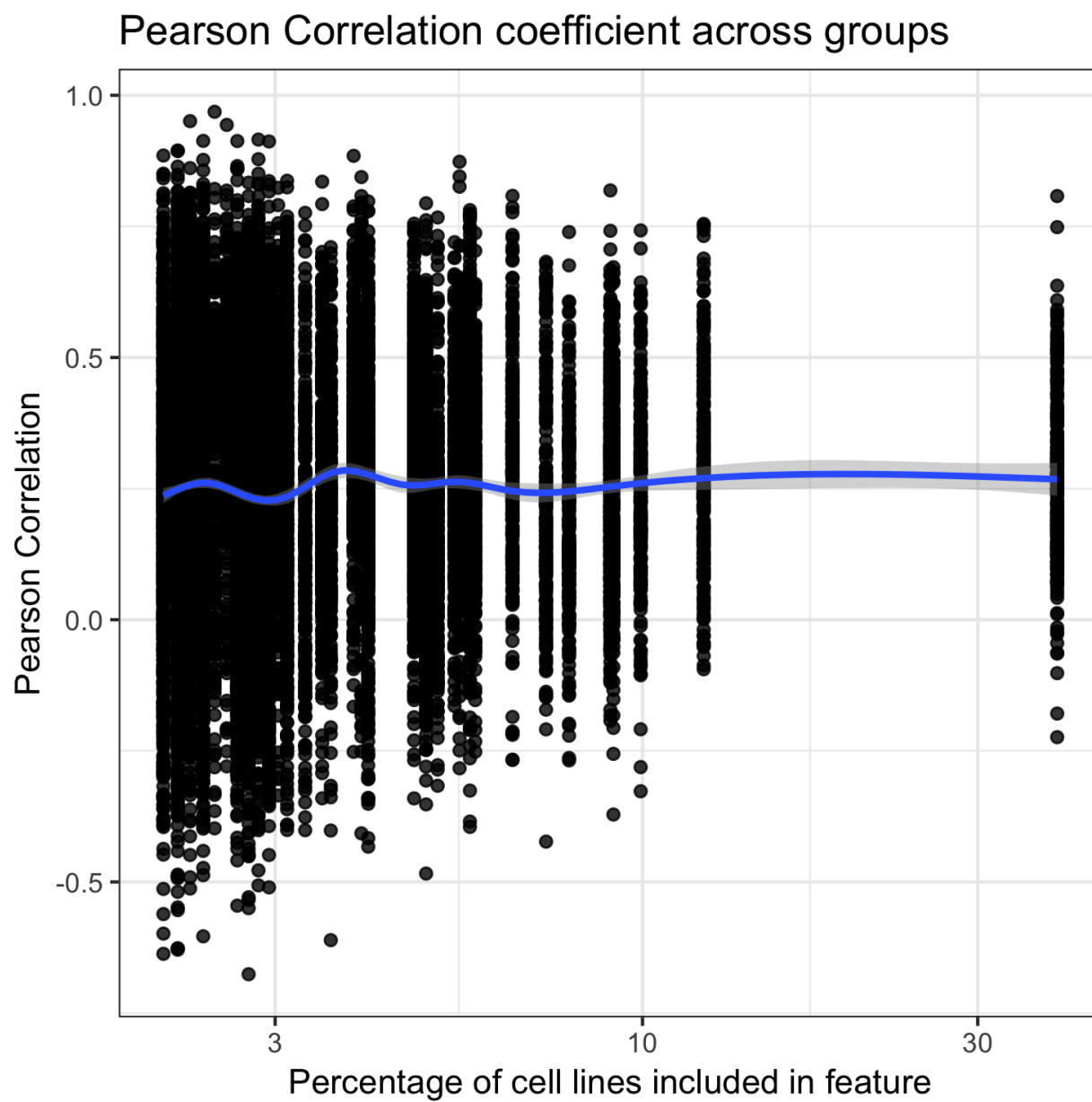


Figure 11: Increased variance of correlation coefficients as subgroup size decreases

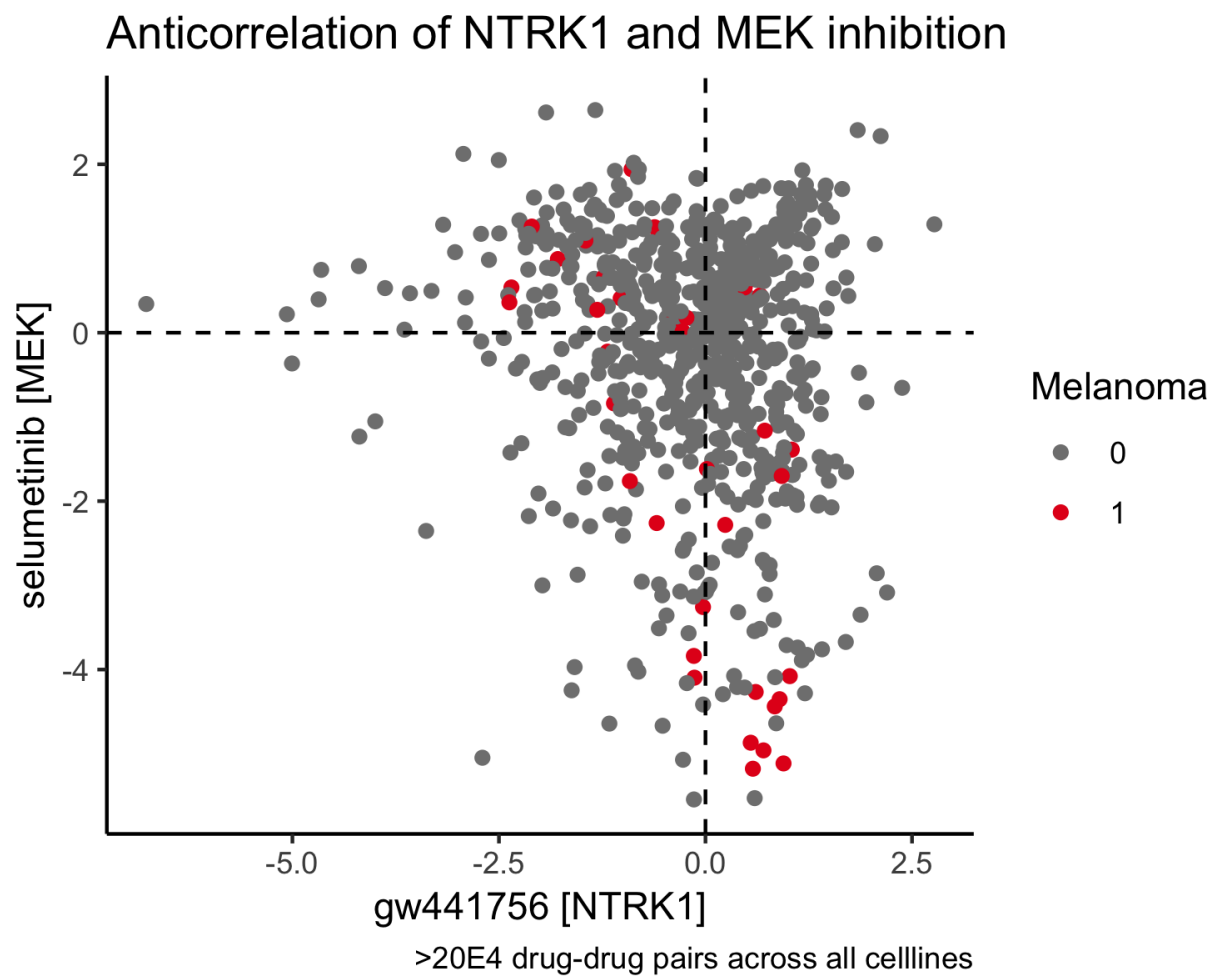


Figure 12: NTRK1 inhibitors are active in MEK inhibitor resistant Melanoma cell lines

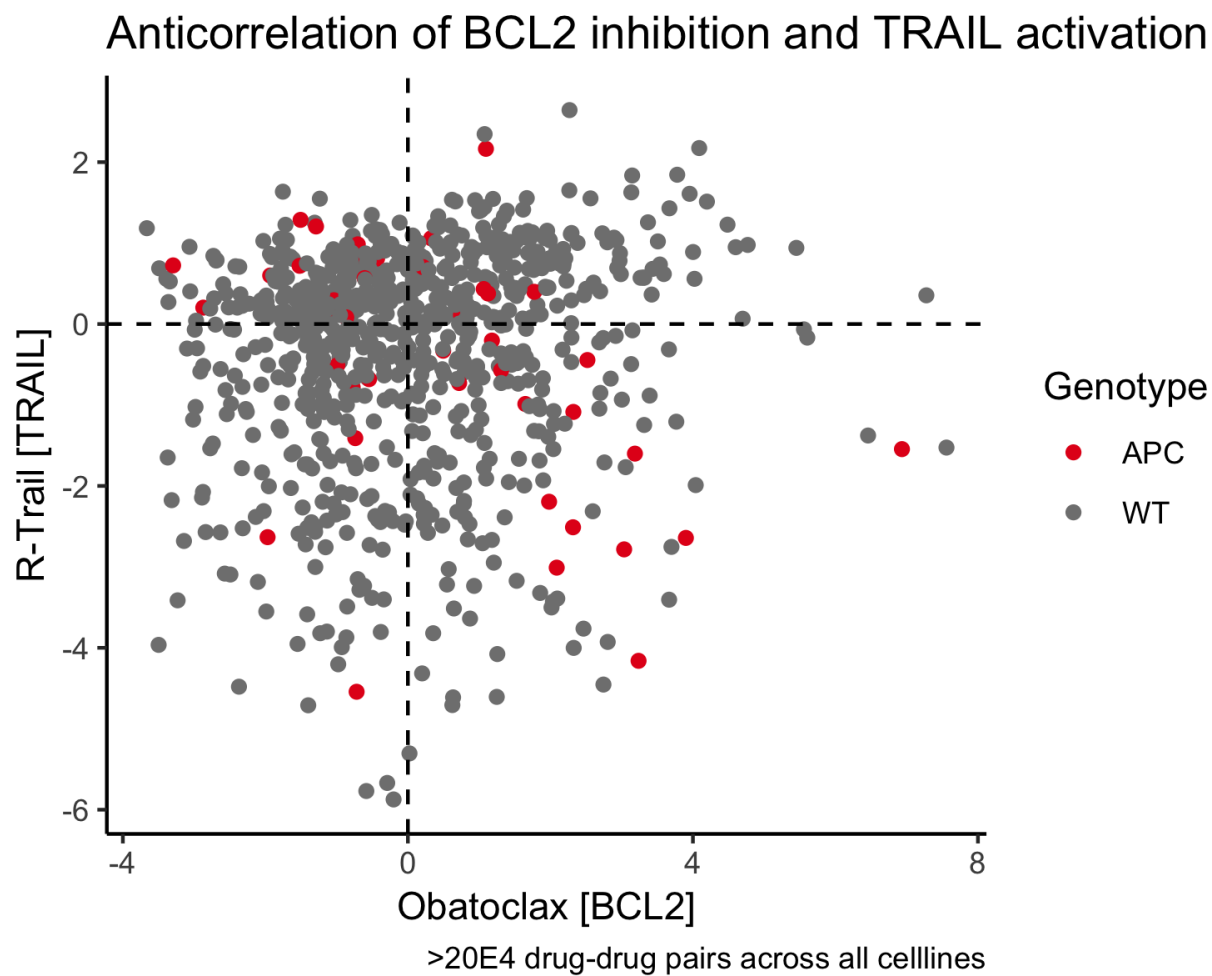


Figure 13: extrinsic stimulation of apoptosis is active in Obatoclax resistant APC mutant cancer cells