Bericht_probe.Rmd

My Report Title

Deniz Korkmazel, Hanna Friess, Evgeniya Abadhieva, Maria Paula Fuentes

July 16, 2023

Introduction

The advancement in the creation of new drugs to combat chronic diseases and pathologies that today represent a global priority in health research, is one of the great challenges of medicine and biotechnology nowadays. This is especially evident in diseases such as cancer, due to its high incidence and mortality rate (Carcas, 2014).A large number of potential anti-tumoral drugs tend to fail in clinical trial stages, as they are not sufficiently effective or safe for human use, and simply do not reach the stage of being approved for distribution on the market. Moreover, the approval processes are remarkably long, lasting between 10 and 17 years. (Loakeim-Skoufa et al., 2023).Likewise, the most effective and novel treatments tend to be expensive, making them inaccessible for oncological patients from middle and low-income countries. This situation has generated a growing interest in the search for more affordable, less toxic, and faster-to-obtain alternatives.

An innovative strategy is "drug repurposing", which consists on re-evaluating existing and approved drugs to treat diseases different from the original application, which could have an anti-tumoral effect. Once their potential against cancer is discovered, clinical trials can be initiated in patients, reducing both the time and final cost to the public, as these drugs are already available in the market (Loakeim-Skoufa et al., 2023)

Cancer is among the deadliest illnesses, characterized by a remarkably high death rate Currently, one of the types of cancer that persists as a serious health problem is gastric cancer, despite a downward trajectory in global incidence. It is recognized as the second leading etiology of cancer-related mortality globally and registers as the fifth most prevalent cancer diagnosis (Carcas, 2014).Among the different types of gastric cancer, adenocarcinoma is the most common (Cebrián, Concha, & Fernández-Urién, 2016).In the context of early-stage tumors, the first therapeutic options are endoscopic resection, surgery, and or an integrated regimen of chemotherapy or chemoradiation. However, the treatment of advanced or metastatic gastric cancer is not very far reaching, and normally in this stage the overall survival time rate in this group is less than one year (Carcas, 2014).This landscape highlights the urgency to develop and explore new therapeutic alternatives, such as drug repurposing, to improve the survival and quality of life of patients with gastric cancer.

In this study, our main objective is to systematically explore the responses of a broad spectrum of drugs in a diverse range of cancer cell lines, with special emphasis on gastric cancer. With the help of additional information on factors such as gene expression, genetic alterations and description of target molecules, we aim to identify potential therapeutic agents for the treatment of cancer through the analysis of intermolecular relationships. More specifically, we aim to identify subgroups of these potential treatments based on their molecular properties. Ultimately, this research aims to contribute to the evolving field of drug repositioning as a viable strategy to advance gastric cancer therapeutics.

Material and Methods

2.1 Data Sets

For this project, we were provided with two datasets consisting of seven data frames in total. These data frames can be divided into three distinct categories. The first category encompasses the data frame "prism," which provides essential information regarding the effect of various treatments on the cell growth of 481 different cell lines. The second category focuses on gene expression and somatic alteration information. It includes: the data frame "prism.exp", which contains data on the levels of gene expression for 19,177 different genes in the cancer cell lines. Additionally, the "prism.cnv" data frame, which provides gene copy values indicating gene amplifications or deletions and the"prism.snv" data frame, which contains information about observed mutation types in the samples. Lastly, the "prism.achilles" data frame shows gene knockdown scores, indicating the impact of gene knock-out on cell growth. The third category is annotation, which captures additional details. The data frame "prism.treat" contains information about the treatments themselves, such as the drugs used, associated pathways, and target molecules. The data frame "prism.cl" provides information about the 481 cell lines. By organizing the data frames into these three categories, we can effectively analyze and interpret the provided datasets for our project.

2.2 Data cleanup

To reduce the dimensionality of the "prism", which originally had dimensions of 481x11168, we created two additional dataframes. The first dataframe contained only cell lines related to gastric cancer, and we removed any drugs that had missing values (NA), resulting in a reduced dimension of 17x2767. Or in the case of the fold change analysis they were replaced with the mean value. For the second dataframe, which included cell lines from five different types of cancer, we replaced the NA-values with the median value of the drug response across all examined cell lines. The same approach of replacing NA-values , but this time with mean values, was applied to the original "prism.exp" dataframe as well.Furthermore, we created for each data frame a separate data frame including only data regarding the gastric cancer cell lines.

2.3 Exploring the effect of treatments on cell growth and identifying potential targets and mechanisms of action

We conducted a comprehensive analysis of drug responses in gastric cancer cell lines and also in breast cancer cell lines as comparison for a further analysis, with a focus on reducing high-dimensional prism dataframe. Therefore, we applied Principle Component Analysis (PCA) on "prism". Furthermore, we performed Pearson correlation on each individual principal component to assess statistically the strength and the direction of the linear relationship between the drugs and the cell growth of cancer cell lines. Combining PCA and Pearson correlation provided us information about the associations between drug treatments and their effects on cell growth, helping us identify potential targets and molecular pathways(Goncalves et. al.).

2.4 Finding clusters of the cell lines based on gene expression and drug response

Principal component analysis (PCA) was performed on the "prism" and "prism.exp" data. By calculating the variance of each row in the transposed data frames, we obtained the variance of drug response and gene expression for each cell line. Plotting the variances in histograms provided a better understanding of the variability in drug response and gene expression.

The visualization of variances helped us establish a minimum variance threshold for gene expression and drug response in cell lines. Only drugs and genes with high variance were selected for PCA. To visualize the PCA results, we created a scree plot and scatter plot for both "prism" and "prism.exp"(Supplementary). We applied K-means clustering to group the cell lines based on PCA results. To determine the optimal number of clusters, we used the elbow and silhouette methods. The elbow plot shows the sum of squared distances between data points and cluster centers for different numbers of clusters, with the "elbow" indicating the optimal number of clusters. The silhouette method, considering distances within and between clusters, suggests the optimal number of clusters when analyzing data. (Supplementary).

2.5 Fold Change Analysis: Evaluating Non-Oncology Versus Oncology Treatment Responses

A fold change analysis was conducted to compare the responses of potential non-oncology treatments to oncology treatments in general or only for gastric as the control group, so all treatments matching these criteria were identified from the dataset "prism". A similar approach was used to identify the corresponding non-oncology treatments. The average response for each cell line for the various oncology and non-oncology therapy groups was then determined independently. So consequently, a fold change was then computed for each cell line, defined as the ratio of the average non-oncology treatment response to the average oncology treatment response. This provided a measure of the relative response to non-oncology versus oncology treatments for each cell line. A pseudo count was added to both the numerator and the denominator during the fold change calculation to avoid division by zero and stabilize the ratio, especially for small values.

2.6 Analyzing the correlation between the drug response and the target of the drug

Obtaining the target of each drug of which the drug response is shown in the "prism" data frame by using the "prism.treat" data frame and analyzing the correlation between the drug response and the expression as well as the gene copy number of the drugs target in the cell line. The drugs that were identified through PCA as contributing the most to the reduction of cell growth in gastric cancer cell lines were utilized for this analysis. Lists were created that contain all calculated correlations of the drug responses of a drug in gastric cancer cell lines and the gene expression, evaluated by the TPM values shown in "prism.exp" data frame, or the gene copy numbers taken from the prism.cnv data frame of the drugs targets in the cell lines. Then the drug-gene pairs were selected showing the highest absolute values

as for their correlation of response and expression or gene copy number values. These were plotted in a scatter plot to visualize the linear correlation between them.

Results

3.1 Differential Treatment Efficacy in Oncology and Non-Oncology Drugs: A Fold Change Analysis"

We carried out two fold change analyses to compare the effectiveness of oncology and non-oncology treatments across various cell lines, focusing then especially on gastric cancer cell lines. The first analysis involved all cell lines, with 9.6 % and 90.4% of the treatments being oncology and non-oncology respectively (see Fig. in Supplementary Materials). A visual comparison of the cancer cell lines via a bar plot (Fig.) revealed that the majority had a fold change value less than 1 but above 0 . This suggests that for these cell lines, oncology treatments were generally more effective at reducing cell growth. However, a few cell lines demonstrated larger fold changes, both positive and negative. Positive fold changes (> 1) indicate a greater effectiveness of non-oncology treatments, while negative fold changes point to instances where the oncology treatment resulted in an increase in cell growth, rendering the non-oncology treatment more effective by comparison. However these latter instances can be seen as outliers or special cases.
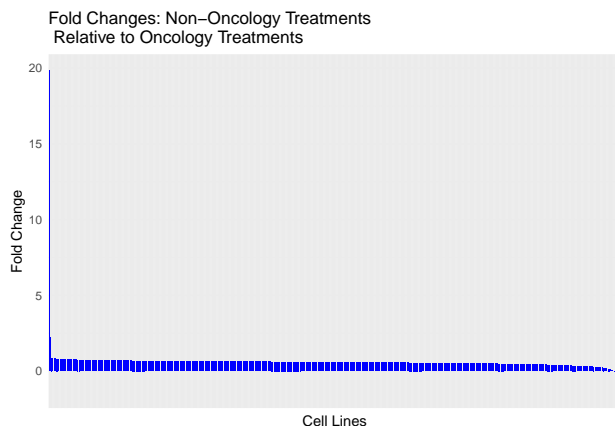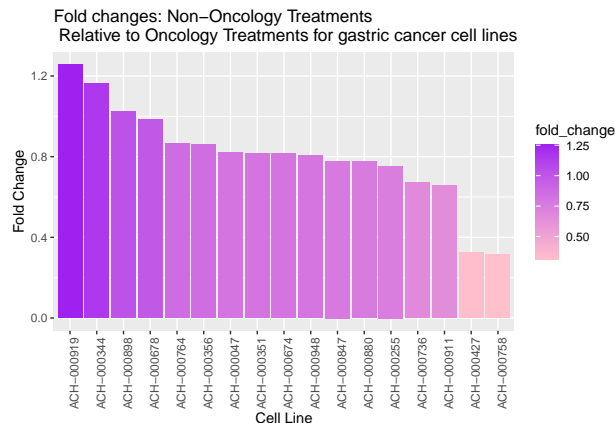

Figure 1: Figure 1(Left): Bar Plot Representation of Fold Change Analysis: Comparing Efficiency of Oncology and Non-Oncology Treatments Across Various Cell Lines.(Right):Representation of Fold Change Analysis: Comparing Oncology and Non-Oncology Treatment Efficiency Across the 17 gastric cancer cell lines

We aimed then to statistically validate our fold change analyses using Welch's Two Sample t-test, applied to both our general and gastric-specific investigations. This approach was to quantify the observed differences in effectiveness between oncology and non-oncology treatments. We first tested our data for normality using the Shapiro-Wilk test, supported by visual inspections using histograms and QQ plots. In the general analysis comparing oncology to non-oncology treatments, we obtained p-values of 0.3101 and 0.06227, respectively. For the gastric-specific analysis, the p-values were 0.364 and 0.0798, respectively. These results suggested that the null hypothesis of normality couldn't be rejected in either

Figure 2: Figure 1(Left): Bar Plot Representation of Fold Change Analysis: Comparing Efficiency of Oncology and Non-Oncology Treatments Across Various Cell Lines.(Right):Representation of Fold Change Analysis: Comparing Oncology and Non-Oncology Treatment Efficiency Across the 17 gastric cancer cell lines

analysis. Both the statistical evidence and visual representations (refer to Supplementary Materials) affirmed a reasonably normal distribution of our data, providing validity to the proceeding t-tests.

The t-tests began with a one-sided analysis, specifically aimed at determining if the mean response of the oncology treatment group was lower than that of the non-oncology treatment group. With a resulting p-value of 2.2e-16, and considering a more negative response signifies higher efficacy in reducing cell growth, the lower average response in oncology treatments suggests that these treatments are generally more effective at inhibiting cell growth.

Conversely, in the second analysis comparing oncology and non-oncology treatment efficacy, a two-sided test resulted in a p-value of 0.2803. As this value is above the threshold of 0.05, it implies insufficient evidence to conclude a significant difference between the mean responses of the two groups. Boxplots in the supplementary materials visually reinforce these findings. They show a more negative median value for oncology responses in the first analysis, while in the specialized gastric cancer cell line analysis, the medians of the two groups are nearly identical.

These last results suggest that, on average, both types of treatments may have a comparable impact on cell growth. For instance, we observed three specific cell lines (), that demonstrated a notably high fold change value. This suggests that while the overall trend may not indicate a significant difference, individual cases could still present substantial variations in response and are important for further research.

3.2 Data reduction based on drug response of gastric cancer cell lines

First of all, our primary objective was to reduce the dimensionality of the "prism" dataframe, focusing specifically on gastric cancer cell lines by applying PCA. The aim was to gain valuable insights into the relationship between gastric cancer cell lines and drugs. We performed

2 different PCA analysis and generated scatter plots to visualize the outcomes. In the first scatter plot, each dot represents a gastric cancer cell line, while in the second scatter each dot corresponds to a specific drug(Supplementary Materials).

Subsequently, we selected six from seventeen principle components, based on their contribution to the overall variance being more than 10 percent. Then, we computed the Pearson correlation of six principal components to the gastric cancer cell lines. This Pearson correlation analysis visualizes the contribution of each principal component to the cell growth of gastric cancer cell lines.
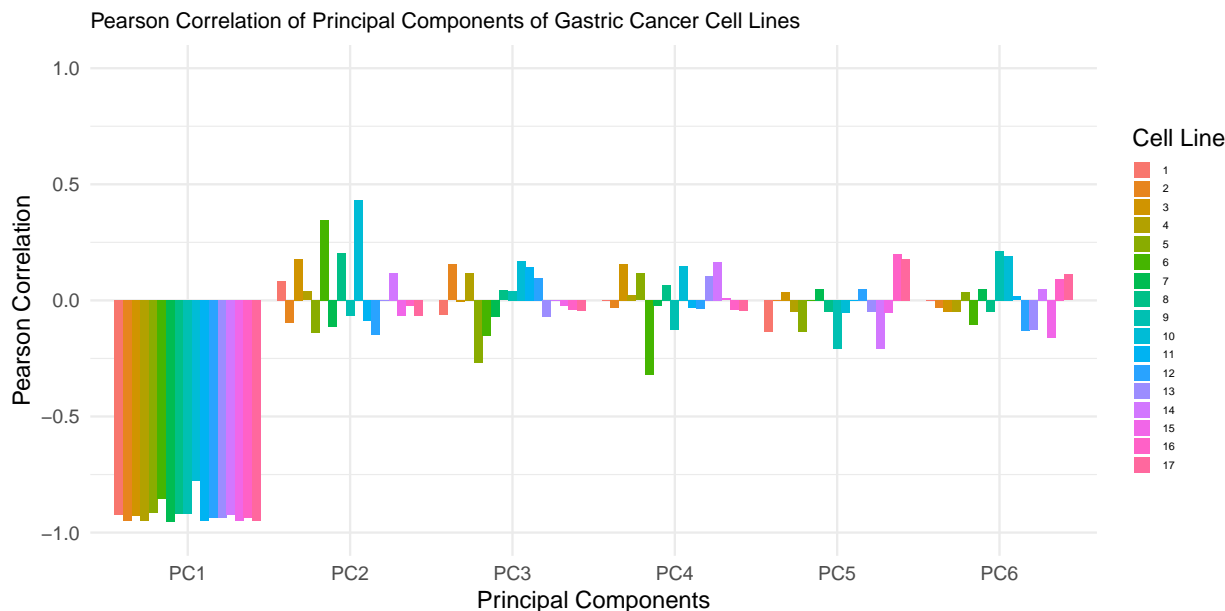
Figure 3: Figure 2: Pearson correlation of Principle components for gastric cancer cell lines. Each color represents different gastric cancer cell line.

PC1 accounted for a significant portion (86.34%) of the overall variance. We identified 678 drugs as positively contributing to the PC1 and demonstrated a consistent negative effect on cell growth of all gastric cancer cell lines. Therefore, we selected a threshold as 20 for PC values of these drugs and we continued the further analysis with 52 selected drugs.

3.3 Identifying top targets and mechanism of action for drugs inducing strong negative cell growth in gastric cancer cell lines

With the aim of identifying the target molecules and the mechanism of actions we reduced the "prism.treat" dataframe by extracting these 52 selected drugs. We created 2 barpots to better visualize the most frequent targets and mechanism of action among these drugs. The first bar plot revealed that HDAC emerged as the most commonly targeted molecule among the 52 selected drugs. To be more concrete: HDAC1, HDAC2, HDAC3, HDAC4, HDAC5, HDAC6, HDAC7, HDAC8 and HDAC9. Consequently, HDAC inhibition was identified as the dominant mechanism of action among the selected drugs. The significance of these findings will be further observed in the discussion part.
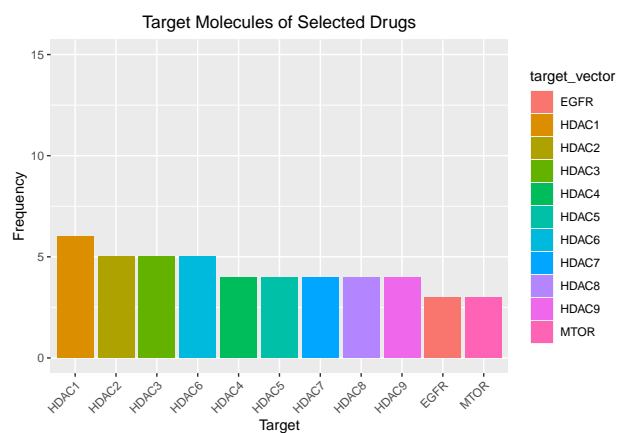
Figure 4: Figure 3: Target molecules and mechanism of actions of selected drugs regarding gastric cancer cell lines.
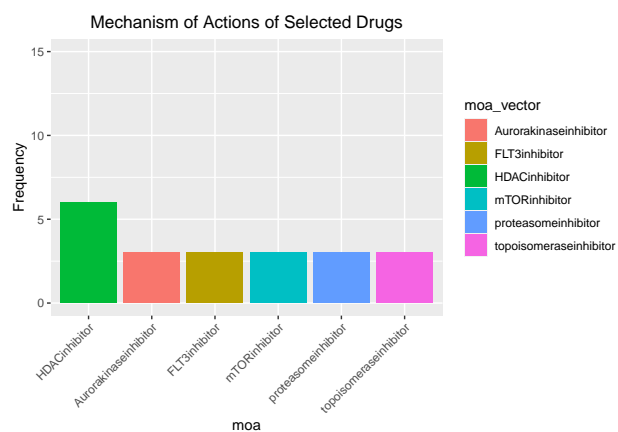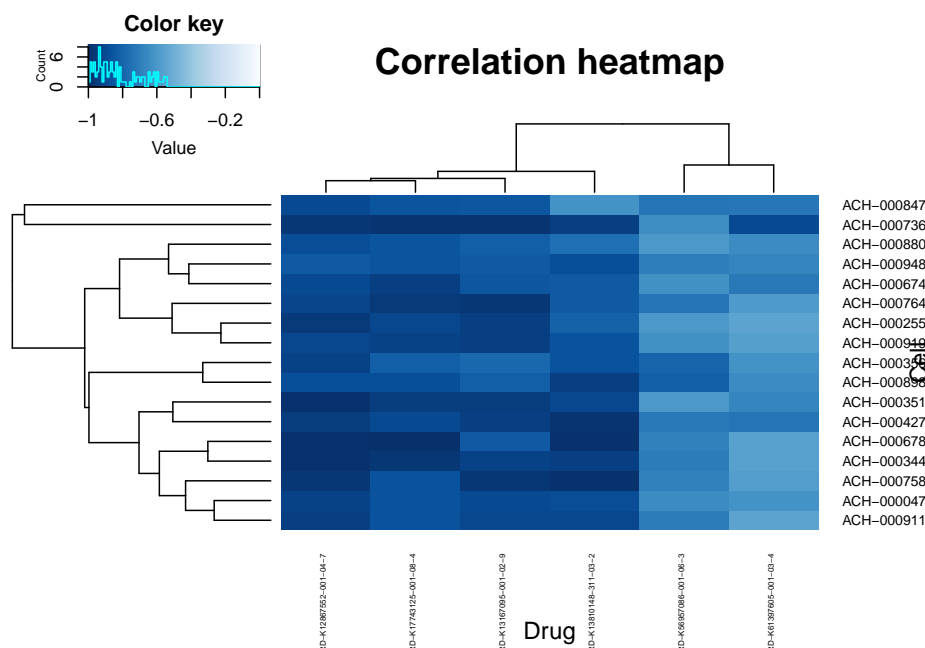

Figure 5: Figure 3: Target molecules and mechanism of actions of selected drugs regarding gastric cancer cell lines.

3.4 Exploring the Correlation Between HDAC Inhibitor Concentration and Gastric Cancer Cell Growth

Through our investigation, we identified six HDAC inhibitors that exhibited strong inhibitory effects on cell growth. To gain deeper insights into the relationship between drug concentration and drug response, we calculated the correlation between these two factors. To facilitate this analysis, we extracted the relevant information for these six treatments, including their respective concentrations and their corresponding effects on cell growth in gastric cancer cell lines. We were able to perform correlation calculations for each cell line separately. To visualize the results, we created a heat map(Figure). The heat map highlighted a robust negative correlation between the concentration of the HDAC inhibitors and cell growth. In other words, as the dosage of the inhibitors increased, the drug influence on the cell growth became stronger, leading to a more pronounced negative impact.

Figure 6: Figure 4: Heatmap of correlation values between drug concentration and drug response,regarding HDAC inhibitors and gastric cancer cell lines.

3.5 Cell lines of different cancer types form different clusters based on their drug response and gene expression

To analyze the similarities and differences regarding drug response and gene expression among cell lines, PCA and k-means clustering were performed. To find the optimal amount of clusters, silhouette and elbow method were applied(supplementary).The log fold change values from the "prism" dataset and TPM values from the "prism.exp" dataset were used, which contained data from cell lines representing five different cancer types: gastric, breast, bone, liver, and kidney.

The PCA scatter plot of drug response is showing similar drug response patterns among all of the examined cancer subtypes. Most of the cell lines formed one cohesive cluster,

which can be further divided into smaller clusters by applying k-means clustering. Notably, these similarities in drug response emerge, even when analyzing drugs with high variance. In other words, these cell lines exhibit similar responses to the drugs in terms of cell growth, independently of the particular cancer subtype.

On the other side, the PCA scatter plot of gene expression revealed that cell lines of the same cancer type tend to cluster together, indicating similarities in their gene expression profiles. However, there were noticeable differences observed within certain cancer types, such as liver and bone cancer cell lines, resulting in a broader distribution in the scatter plot. Furthermore gastric and breast cancer cluster extremely together, forming their own clusters independently of the applied method to find the optimal amount of clusters (supplementary), which indicates a strong similarity of gene expression. This could also be another argument for their similar drug response of these two cancer subtypes.
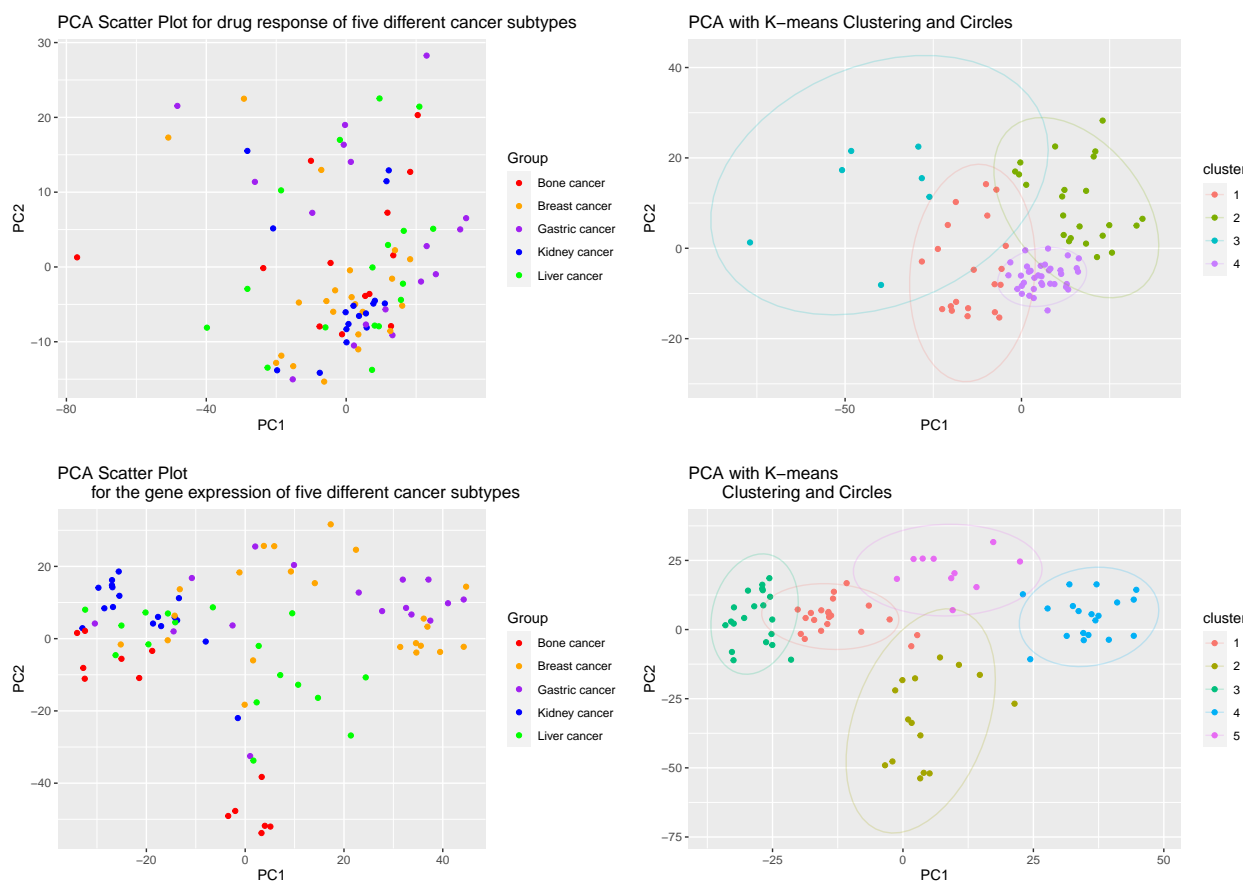
3.6 Comparative Analysis: Shared Targets and Mechanisms of Action in Gastric and Breast Cancer

In order to validate the observed clustering between gene expression patterns of gastric cancer and breast cancer and their similar drug response, we conducted further analysis to substantiate the significance of this finding. To achieve this, we performed PCA exclusively on the "prism" data set, focusing only on breast cancer cell lines, with the objective of identifying key targets and potential mechanisms of action and to compare them with gastric cancer.

The methods applied in this analysis were consistent with the approach explained in the initial sections of the report (see 2.3). Subsequently, we generated two bar plots which visualize the most commonly observed targets and mechanism of actions. It is important to note that this analysis focused exclusively on drugs that made the most substantial contributions to PC1. Remarkably, our analysis revealed the presence of three specific and noteworthy targets that are shared between gastric cancer and breast cancer subtypes: EGFR, HDAC1, and mTOR. Additionally, upon examining the second histogram, we observed a convergence of mechanisms of action. Specifically, HDAC inhibitors and FLT3 inhibitors emerged as the most frequent mechanisms common in both cancer subtypes. These findings demonstrate that breast and gastric cancer exhibit remarkable similarities in terms of shared targets, as well as common mechanisms of action. The importance of this result will be further observed in the discussion part.
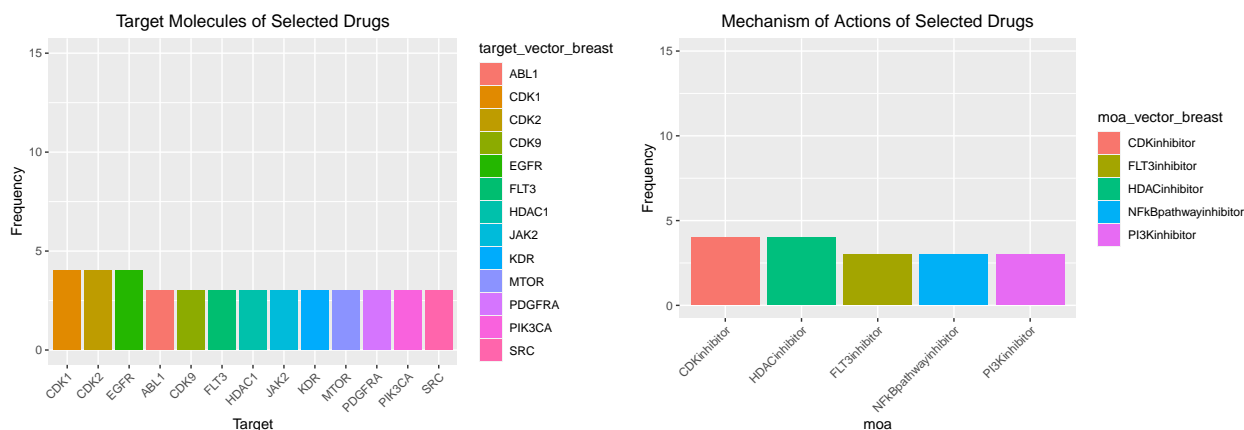

3.7 Analyzing the correlation between the response of drugs and the CNVs as well as the gene expression of the drugs target in gastric cancer cells

Correlations between drug response and gene expression or CNVs of drug targets in gastric cancer cells identified two drug-gene pairs for further analysis: dacinostat(BRD-K56957086-001-06-3)/HDAC1 and GZD824 (BRD-K35329391-334-01-7)/BCR. These pairs demonstrated strong correlations between the drug response and both the gene expression and CNVs of their respective targets.

As before mentioned HDAC inhibitors seem to be a promising group of repurposing drugs for gastric cancer as the graphs (Figures) depicting the drug response of dacinostat against HDAC1 gene expression or CNVs in gastric cancer cell lines confirm. The reduction of cell growth is dose-dependent of dacinostat since the dots of lower doses are clustered around zero, meaning no effect, while increasing the dose lead to a reduction of cell growth in all gastric cancer cell lines (Figures). Same can be observed when looking at the scatter plots (Figures) visualizing the relationship between the drug response of GZD824 and the gene expression or the CNVs of its target BCR. Compared to HDAC inhibitor a much higher concentration is needed so the effect of GZD824 is guaranteed since in some cell lines the second highest dose of 2.5 mM has nearly no effect on cell growth.

The correlations between drug response of dacinostat and HDAC1 gene expression vary when increasing the drug dosage (Figure). In fact, there is no consistent pattern or trend in the

correlations as the dose of dacinostat increases. Contrarily, as the drug dose of dacinostat increases, the correlations with HDAC1 CNVs consistently increase (Figure). This finding supports the hypothesis that the drug's effect in gastric cancer cells is indeed linked to the CNVs of its target gene, HDAC1. For very low dosages, the overall effect on the cell lines is minimal, likely due to the fact that the effective dose of the drug is not reached, thereby hindering a clear expression of the correlation between drug response and HDAC1 CNVs. Similar observations can be made when examining the change of the correlation between the drug response of GZD824 and the gene expression or the CNVs of its target BCR when increasing the dosage of the drug(Figures). However the drug response of GZD824 and the CNVs of BCR in gastric cancer cells show a negative correlation for all dosages of the drug (Figure). Lower CNVs suggest a weaker reduction in cell growth, while higher CNVs indicate an increased effectiveness of the drug in reducing cell growth in gastric cancer cells.
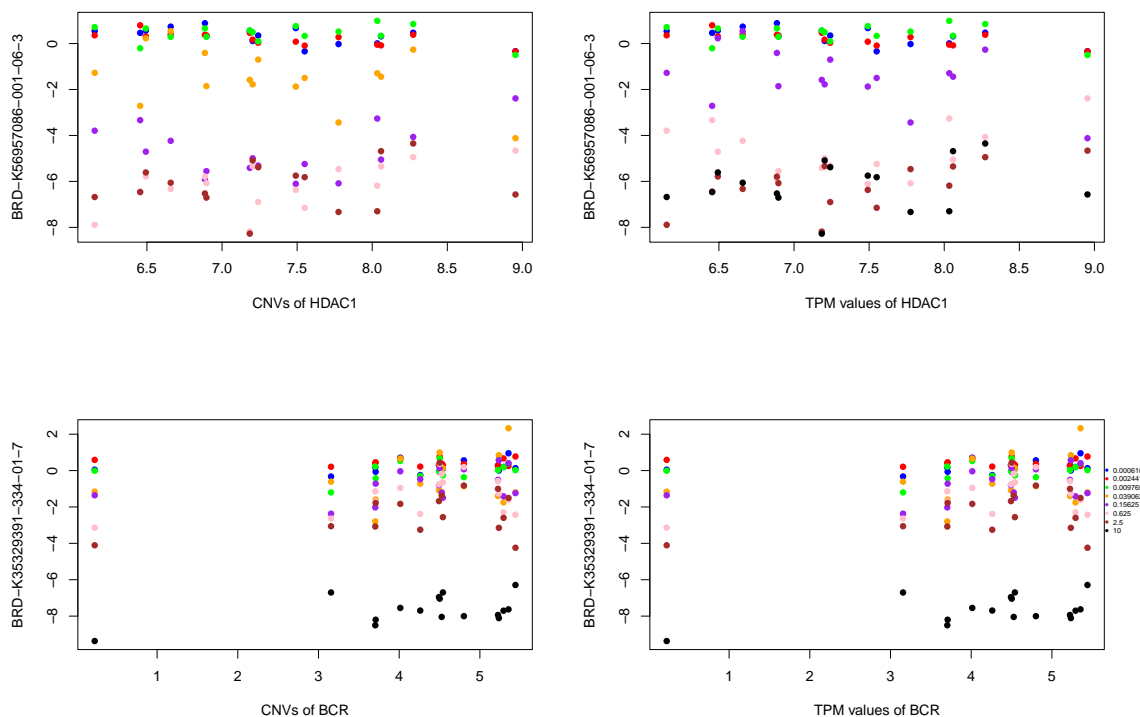
3.8 Creating prediction models for the drug response of two specific drugs based on the CNVs of their target gene

Since the correlations of both drugs with their target CNVs were higher compared to the correlation between drug response and target gene expression and also showed a constant increase when increasing the dosage, linear regression was applied to further analyze the relationship between the drug response of dacinostat and GZD824 and the CNVs of HDAC1 and BCR in gastric cancer cells. The objective is to develop a prediction model for estimating the drug response of a specific drug, using the copy number variations (CNVs) of the drug's target gene as the independent variable and perform linear regression. The model was created only for the highest dosage of the drugs namely 10 mM.

The model of dacinostat has a multiple R-squared of 0.2766, indicating that approximately 27.66% of the variance of the drug response is explained by the model. The F-statistic (5.736) is significant, indicating that the model as a whole is statistically significant. The residual standard error is 0.8758, which represents the average deviation of the observed values from the predicted values.

The model of GZD824 explains approximately 46.77% of the variance of the drug response, as indicated by the multiple R-squared value of 0.4677. The F-statistic is 13.18, with a corresponding p-value of 0.002468, suggesting that the model as a whole is statistically significant. The residual standard error is 0.5696, representing the average deviation of the observed values from the predicted values.

The R-squared values of both prediction models are not close to one, indicating that the CNVs do not have a very strong predictive power since the linear regression models explain just a moderate amount of the variance of the drug responses.

When involving other cancer types in the linear regression analysis, the result states for both drug-genes pairs that the relationship between the drug response and CNVs of the drugs target is not significant, indicated by high p-values of the F-statistics (Graphs in Appendix).
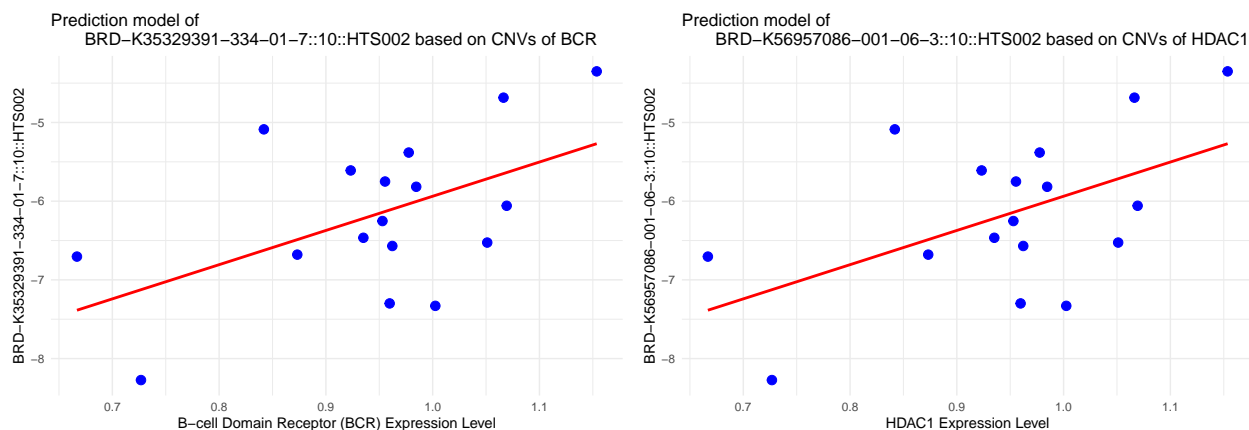
3.9 Drug responses in ACH-000948 (Fold change cell line): Focus on HDAC inhibitors)

In our previous discussion, we identified a group of cell lines with a fold change below 1 in the analysis comparing oncology and non-oncology gastric treatments. Our intention was to investigate these cell lines further, focusing on potential effects of HDAC-inhibitor treatments. The main reason being, HDAC-inhibitors have largely been used in hematologic malignancies, and we hypothesized their potential in solid tumors such as gastric cancer. We selected cell line ACH-000948 for detailed analysis due to its average performance in terms of fold change. We identified the most efficient treatments for this cell line by setting a response threshold of -7. This means we considered treatments with values of -7 and lower (more negative) as highly efficient. Our analysis revealed that out of all treatments, only 51 demonstrated this level of efficiency on the ACH-000948 cell line. Of these, 35 treatments aligned with the treatments identified in the first Principal Component (PC1) from the PCA conducted

earlier. This suggests a significant overlap between highly efficient treatments and the PC1 treatments.Furthermore, as mention before we sought to determine if any of the selected treatments coincided with the six HDAC inhibitors treatments, previously identified as potential target drugs. Remarkably, four out of the six HDAC treatments ("BRD-K13167095-001-02-9::10::HTS002", "BRD-K13810148-311-03-2::10::HTS002", BRD-K56957086-001-06-3::10::HTS002", "BRD-K61397605-001-03-4::10::HTS002") were among the most efficient treatments for the cell line, reinforcing the notion of HDAC as a potential target for drug repurposing in gastric cancer. A subsequent boxplot comparing () the responses of these four treatments against the HDAC target molecule efficient treatments illustrated that the median line for the HDAC inhibitors was lower(more negative response) than for the resto of the treatments. This suggested a superior efficacy for the for HDAC – Inhibitors treatments compared to the rest of the treatments, which again proves the hypotesis mentioned in section xxx above.
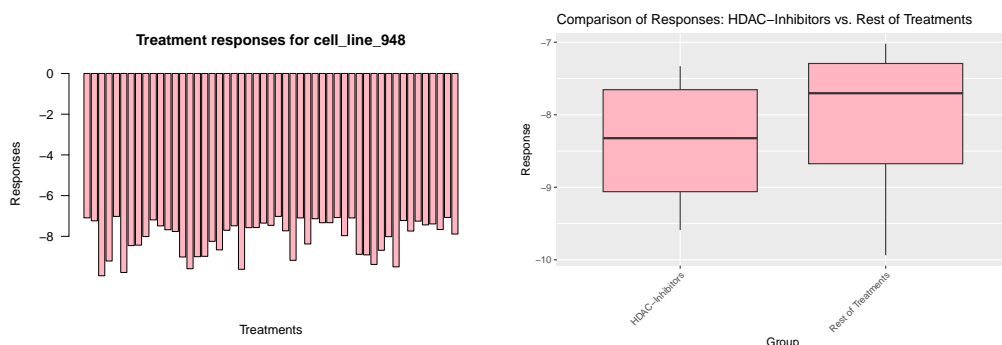

Fig X:(Left):Representation of Highly Efficient Treatments for ACH-000948 (Right):Boxplot Comparison of HDAC Inhibitors and Non-HDAC Inhibitors Treatments: Efficiency Analysis in Gastric Cancer Cell Line ACH-000948 # **Discussion**

Understanding and analyzing the drug response of existing drugs in cancer cell lines especially in gastric cancer cell lines to identify possible targets and metabolic pathways is important for drug-repurposing and to identify effective treatments for cancer. In our preliminary analysis using fold changes and t-tests, we found that oncological treatments were generally more effective in inhibiting cell growth in all cancer cell lines. This observation highlighted the potential for reusability of oncological drugs in different cancer types and subtypes and different by which they were created. This conclusion led us to the following fold-change analysis, in which we analyzed non-oncology treatments in relation to oncology-treatments, specifically excluding gastric oncological treatments among only gastric cancer cells.
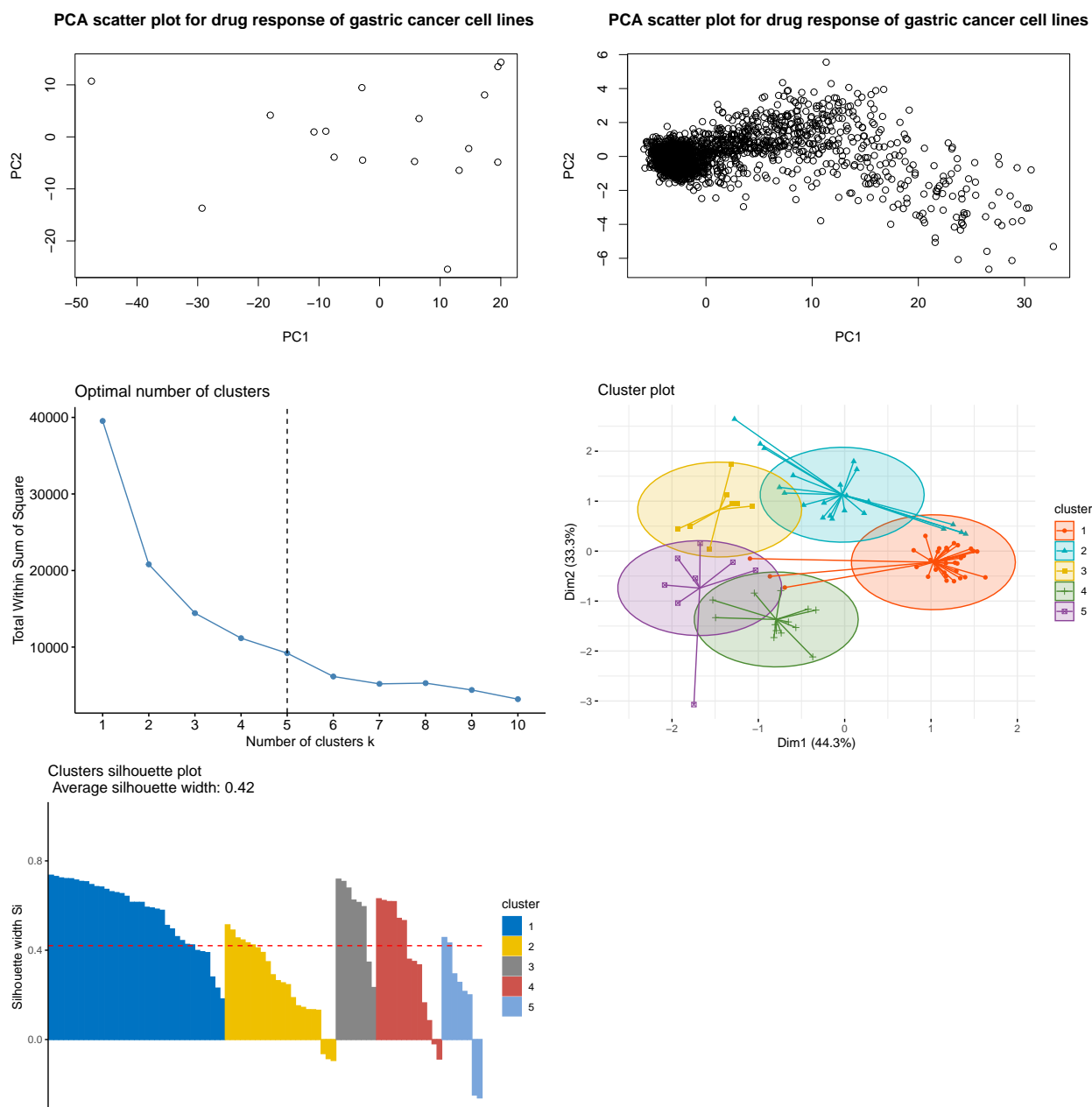
Most gastric cancer cell lines responded more effectively to oncological treatments, even those not specifically designed for gastric cancer, which confirmed our theory.
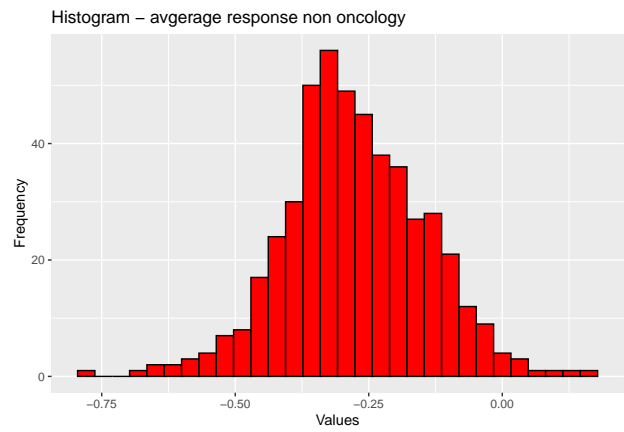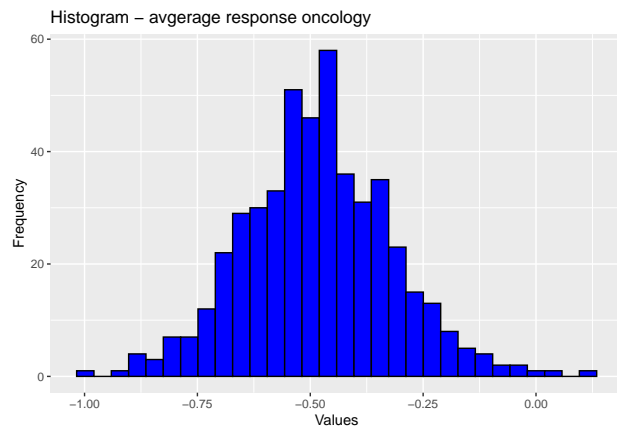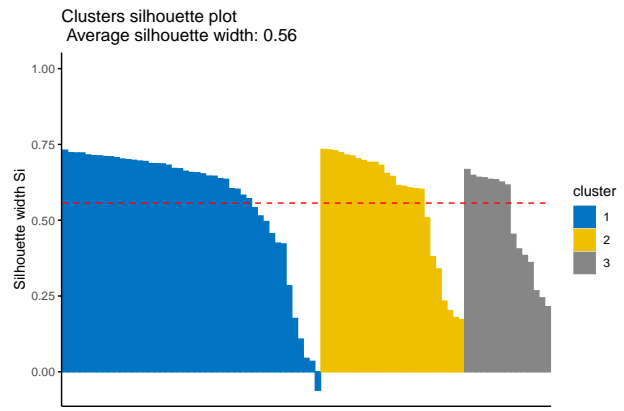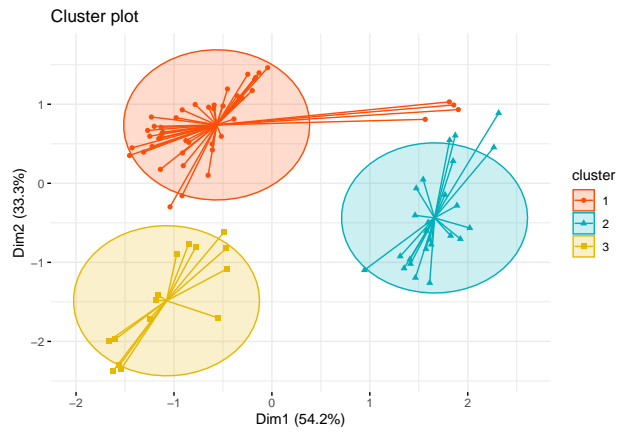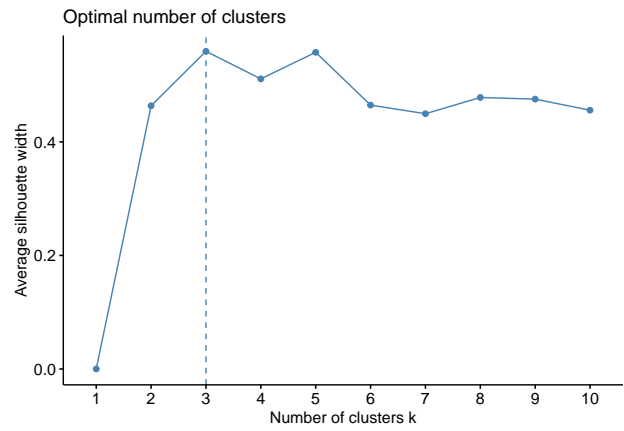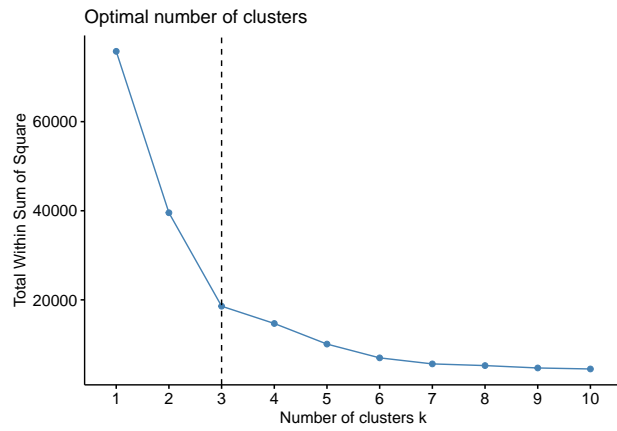
There were also some outlier cases that showed greater efficacy with non-oncology drugs. Although our study focused primarily on drug repurposing within oncology, these results point to promising avenues for future research exploring the use of non-oncology treatments and studying the unique characteristics of these atypical cell lines. Through our focused analysis performing PCA and Pearson correlation, we analyzed the relationships between drugs and cell growth in gastric cancer cell lines. By performing the analysis we selected specific drugs, which inhibit the cell growth and gained insights into their potential targets and molecular pathways. According to our results, the majority of the targets of the selected drugs were

subtypes of the HDAC enzyme family, followed by MTOR and EGFR. Moreover, the drugs which inhibit HDAC did have a strong negative effect on cell growth of gastric cancer cell lines. Furthermore, we wanted to investigate the correlation between drug concentration and drug response in the context of gastric cancer, specifically focusing on HDAC inhibitors. Our analysis revealed a notable negative correlation between drug concentration and cell growth, indicating that higher drug concentrations corresponded to a stronger influence on inhibiting cell growth. This analysis provides valuable insights into the potential efficacy of HDAC inhibitors in gastric cancer and underscores the importance of dosage considerations in developing more effective therapeutic approaches for treating gastric cancer. This was also verified in the further analysis using the results of the fold change, where it was corroborated that among a gastric cancer cell line that had a response in average ratio of fold change value (best effect with oncology drugs), 4 of the treatments with the best efficiency corresponded to HDAC inhibitors. Furthermore, our analysis revealed intriguing similarities between gastric cancer and breast cancer, including shared targets (EGFR, HDAC1, and mTOR) and mechanisms of action (HDAC inhibitors). Specifically, we investigated the HDAC inhibitor, which is depicted in the barplots analysis for both cancer subtypes. This observation aligns with existing scientific literature, which highlights the potential of HDAC inhibitors to reduce cell growth through their impact on critical processes such as EMT, ER reactivation, and also their ability to induce programmed cell death in breast cancer cells. This shared mechanism of action highlights the potential for repurposing HDAC inhibitors as targeted therapies that could benefit patients with both gastric and breast cancer. To investigate the relationship between a drug and its target gene in gastric cancer cells, the drug response, gene expression, and CNVs of the drug and the respective target gene were examined. This analysis identified several drug-gene pairs, which were then subjected to further analysis using linear regression. By employing linear regression, prediction models were created to estimate the drug response based on the gene expression or the CNVs of the target gene. Since the whole analysis was based on gastric cancer which represents a small group of cell lines with just approximately 17 cell lines in the data set, the limitations of the achieved results should be kept in mind. For example the correlations between dacinostat and GZD824 and the gene expression of their target genes did show inconsistency when increasing the drug doses, which could indicate that the high correlations for some dosages were more or less coincidence and are not proving a relationship between drug response and gene expression. These findings are based on gene expression and CNVs in the cancer cells but no control group for this data was given, meaning no insight was given on how the gene expression or the CNVs differ generally from cells without cancer. The limited amount of tested gastric cancer cell lines and missing information about the gene expression and the CNVs in cells without cancer make it difficult to conclude concrete facts out of this analysis. Though the idea of analyzing the relationship between the tested drug in drug repurposing and its original target gene could be an approach to identify or exclude certain cancer types based on gene expression or CNV´s for the application of the drug. The created prediction models for the drug response of dacinostat and GZD824 based on the CNVs of their target gene showed no strong predictive power when looking at the R-squared values. When four additional cancer types were included in the prediction model to increase the sample size for the linear regression analysis, the results yielded even lower explanatory power or R-squared value. Since the amount of data for gastric cancer cells is so little, it was not possible to check if the discrepancy occurred because these genes are only related to the drug response

in gastric cancer cells or if this proves that there is no significant correlation. But since all five cancer types have in common that these two drugs reduced their cell growth, it indicates that most probably analyzing the CNVs of their target gene is not a good parameter to predict the drug response. Despite the fact that the created prediction models did not prove to have a strong predictive power, the idea of analyzing the target gene or protein in the cells for which the drug is aimed to be repurposed could be an approach to narrow the cancer types that might benefit from the drug.

Supplementary Materials

16

Optimal number of clusters

Optimal number of clusters

Cluster plot

Clusters silhouette plot
Average silhouette width: 0.56

Histogram – avgerage response oncology
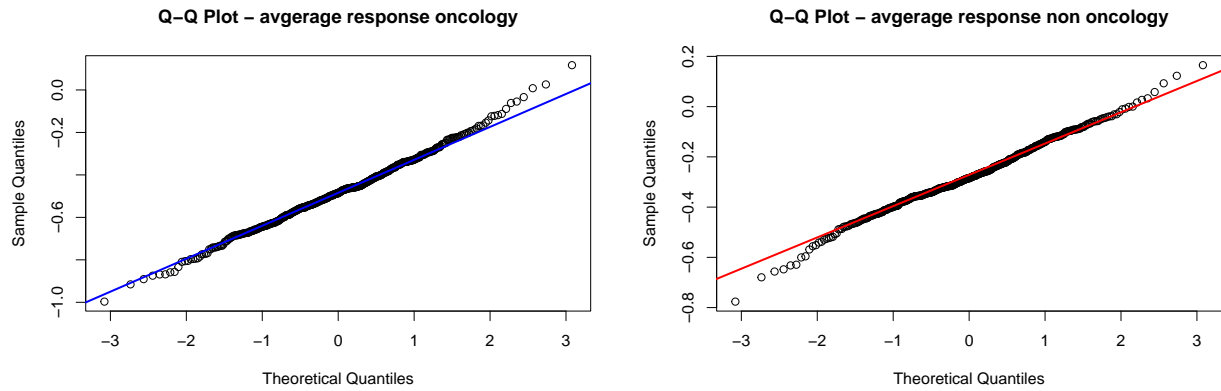
Histogram – avgerage response non oncology

Fig X(Left):Histogram of Average Oncological Drug Responses for Normality Testing.(Right): Q-Q Plot of Average Oncological Drug Responses for Normality Testing
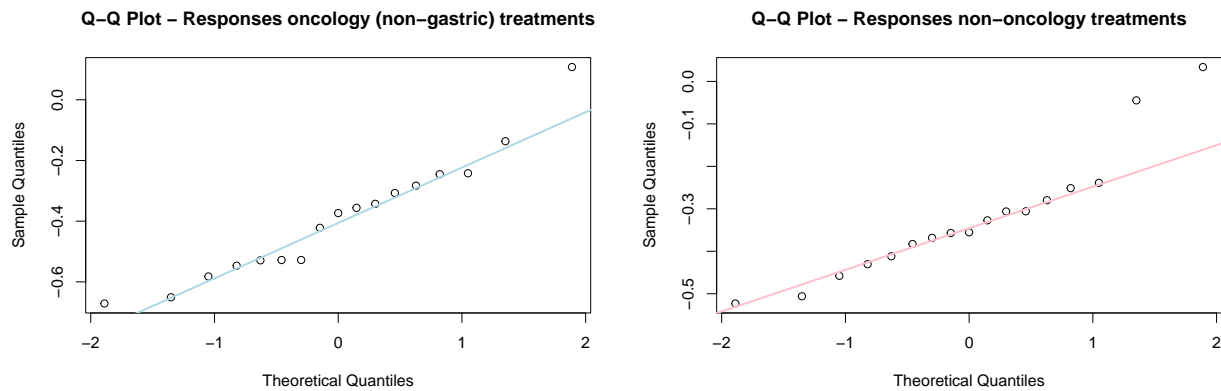

Fig X: Q-Q Plot of Average Responses from Oncological Drugs for Normality Testing Fig X: Q-Q Plot of Average Responses from Non-Oncological Drugs for Normality Testing
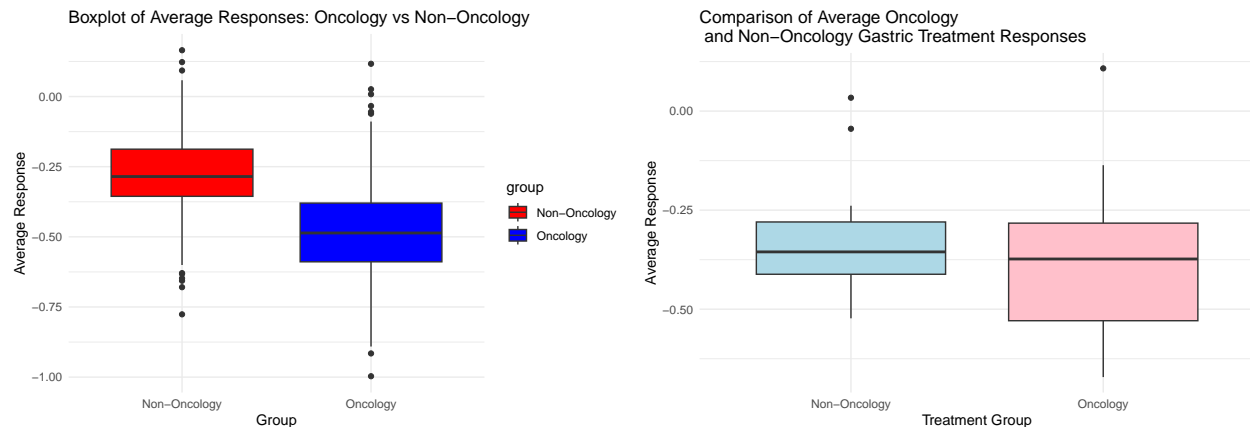

Fig X(Left):Boxplot of Average Responses: Comparison Between Non-Oncology and Oncology Treatments.(Right): Boxplot of Average Responses: Comparison Between Non-Oncology and Oncology Treatments in Gastric Cancer Cell Lines