Bioinfo_Report_Team5.1.Rmd

Data Analysis Report - Topic 5, Group 1

Examining HDAC inhibitors in context of gastric cancer

Authors: Deniz Korkmazel, Hanna Friess, Evgeniya Abadzhieva, Maria Paula Fuentes

Data Science Project SoSe 2023

Supervisor : Carl Herrmann

Tutor: Ana Luisa Costa

Deadline: 17.07.2023

Contents

List of abbreviations

Abbreviation	Explanation
BCR	B cell receptor
CNV	Copy number values
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial–mesenchymal transition
ER	Estrogen receptor
FLT3	Fms Related Receptor Tyrosine Kinase 3
HDAC	Histone deacetylase
MTOR	Mammalian target of rapamycin
NA	Not available
PC	Principal component
PCA	Principal Component Analysis
PCR	Pearson correlation
TPM	Transcripts per million

Abstract

The development of effective treatments for diseases is a time-consuming and expensive process, leading to the growing application of drug repurposing as a method to discover new applications of already existing drugs for treating various diseases. In this report, we analyzed data on multiple drugs and their impact on cell growth in cell lines across different cancer types with a special focus on gastric cancer. In order to provide context for the analysis of drug responses across cancer cell lines, various genetic aspects were taken into account. Specifically, gene expression, CNVs, and gene knock-out scores in the cancer cell lines were observed. Fold change analysis showed that oncology drugs show a higher potential in decreasing cell growth across all cell lines. HDAC inhibitors as a subgroup of the oncological drugs turned out to be a promising group to be repurposed in gastric cancer as well as breast cancer. Searching for further similarities in gastric and breast cancer, clustering the PCA scores of gene expression across cell lines of five cancer types, led to a distinct cluster, containing breast and gastric cancer. Finally prediction models for two promising drugs to be repurposed for gastric cancer were created based on the CNVs of their target genes.

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 (Ioakeim-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 in the case of cancer 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 (Ioakeim-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, De la Concha, and 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. Ultimately, this research aims to contribute to the evolving field of drug repurposing 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 in 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.

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 relation between the drugs and the cell growth of cancer cell lines.(Gonçalves et al. (2020)).

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

Principal component analysis (PCA) was performed on the "prism" (drug responses) and "prism.exp" (gene expression) data. By calculating the variance of each row in the transposed data frames and then plotting them for visualization in histograms, it provided a better understanding of the variability in the data and therefore it 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 scatter plots for both "prism" and "prism.exp". 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.

2.5 Fold change analysis: evaluating non-oncology versus oncology treatment responses

Fold change analysis was conducted to compare the responses of non-oncology treatments to oncology treatments in a general way and then only for gastric cancer cell lines (excluding here in the oncology treatment group the gastric oncology treatments) . All treatments matching these criteria were identified

from the dataset "prism". 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 Creating prediction models based on strong correlations

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 calculating the correlation between the drugs and their targets regarding gene expression as well as CNVs. Drug-gene pairs showing the highest correlation were selected and linear regression was applied to create a prediction model for the drug responses of the selected drugs based on the CNVs of their target genes.

Results

3.1 Differential treatment efficacy in oncology and non-oncology drugs: A fold change analysis

Our study involved two-fold change analyses, both general and gastric cancer cell line -specific, to assess the efficacy of oncology and non-oncology treatments on various cell lines. A preliminary general analysis, visualized through a box plot (SM: Fig. 18), suggested that oncology treatments were more effective at inhibiting cell growth for most cell lines. This was further evidenced by a fold change bar plot (SM: Fig. 15) where the majority of cell lines exhibited a fold change value below 1, but above 0. However, a subset of cell lines presented with larger positive fold changes, suggesting a heightened effectiveness of non-oncology treatments.This hole preliminary general analysis was statistically proved through normality testing (p=0.3101 (oncology) and p= 0.06227(non- oncology)) and plotting (SM: Fig. 16). And through one side testing it was proofed that oncology treatments were generally more effective at inhibiting cell growth (p-value=2.2e-16). The focus of our study was the more detailed fold change analysis, which compared non-oncology treatments relative to oncology treatments in gastric cancer cell lines, specifically excluding gastric oncology treatments.The reason behind this particular focus was based on the previous finding that oncology treatments generally outperformed non-oncology treatments in terms of efficacy. The results revealed that most gastric cell lines also exhibited a fold change value below 1 (Fig. 1), implying superior efficacy of oncology treatments. However, three cell lines (ACH-000919, ACH-000344, and ACH-000898) showed a fold change value exceeding 1, suggesting these specific cell lines responded better to non-oncology treatments.Statistical validation of these fold change analysis was performed using Welch's Two Sample t-test, after data normality checks via the Shapiro-Wilk test. The test and additionally visualizations (SM: Fig. 17) affirmed a reasonably normal data distribution (p=0.364 (oncology) and p= 0.0798 (non- oncology)). So, with this information it was performed a two-sided t-test and it revealed no significant difference between the mean responses of the two treatment groups (p-value=0.2803). Additionally, a boxplot (SM: Fig. 18) visually supported this finding, showing comparable medians for the two treatment types in the specialized gastric cancer cell line analysis. This suggests similar impacts on cell growth from both treatment types, providing us with the freedom to focus our research on cancer drugs outside of gastric cancer-specific drugs or completely non-oncologic drugs.

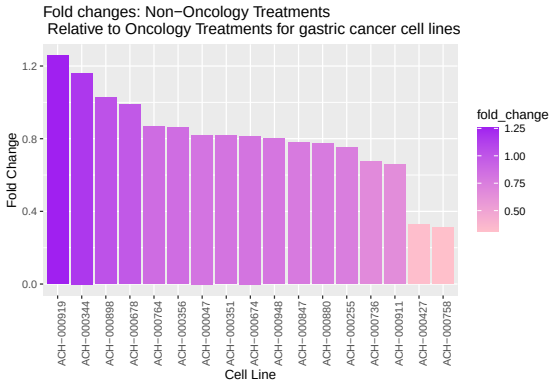

Figure 1: Representation of Fold Change Analysis: Comparing Oncology and Non-Oncology Treatment Efficiency Across the 17 gastric cancer cell lines

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

Our primary objective was to reduce the dimensionality of the "prism" dataframe by applying PCA, focusing specifically on gastric cancer cell lines. 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 (SM: Fig.10). 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 (SM: Fig. 11). 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 mechanisms of action we reduced the "prism.treat" extracting these 52 selected drugs. We created 2 barpots to visualize the most frequent targets and mechanism of action among these drugs (Fig. 2). The first barplot revealed that HDAC emerged as the most commonly targeted molecule among the 52 selected drugs. To be more concrete: *HDAC1-HDAC9.* Consequently, HDAC inhibition was identified as the dominant mechanism of action among the selected drugs.

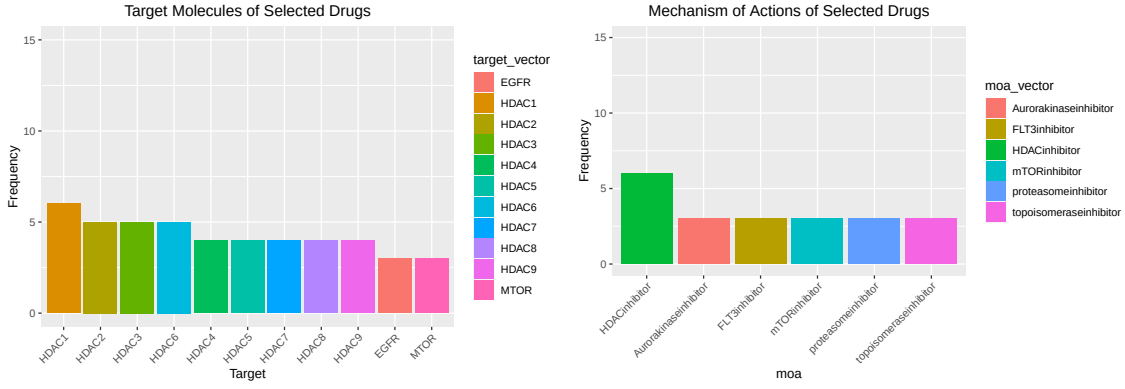

Figure 2: Barplot representation of the most frequent targets (left) and mechanism of action (right) in gastric cancer cell lines.

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 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 heatmap (Fig. 3). The heatmap 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.

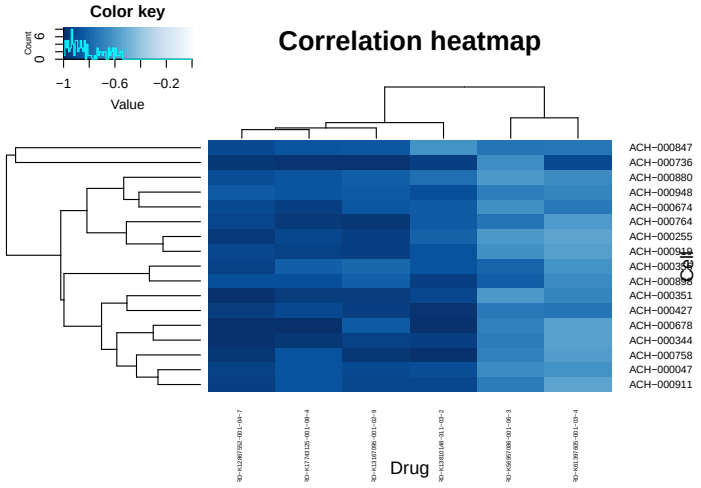

Figure 3: Heatmap of correlation values between drug concentration and drug response,regarding HDAC inhibitors and gastric cancer cell lines.

3.5 Drug responses in ACH-000948: Focus on HDAC inhibitors

In our previous result in the section fold change analysis (see.3.1), we identified that the majority of gastric cancer cell lines had a fold change below 1, indicating a greater effect with oncological treatments. Our aim was to build upon these earlier findings by conducting a more detailed investigation into these cell lines, with a particular emphasis on the potential effects of HDAC-inhibitor treatments. Given that HDAC-inhibitors are currently used as drugs for hematologic malignancies, they present potential oncology candidates for repurposing as treatments for gastric cancer.So, we selected randomly the cell line ACH-000948 for detailed analysis due to its average performance in terms of fold change (bellow 1). We identified the most efficient treatments for this cell line by setting a response threshold. Our analysis revealed that out of all treatments, only 51 demonstrated this level of efficiency on the cell line. Of these, 35 treatments aligned with the treatments identified in the first Principal Component (PC1) from the PCA conducted earlier (see.3.2). 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 were among the most efficient treatments for the selected cell line, reinforcing the notion of HDAC as a potential target for drug repurposing in gastric cancer. The subsequent boxplot comparison (Fig. 4) of the response to these four HDAC-inhibitors versus the other effective treatments revealed that the median line for the HDAC-inhibitors was lower (indicating a more negative response) compared to the rest of the treatments. This observation suggests a superior efficacy of the HDAC-inhibitor treatments compared to the other treatments, thereby reinforcing the hypothesis mentioned in section 3.2 above.

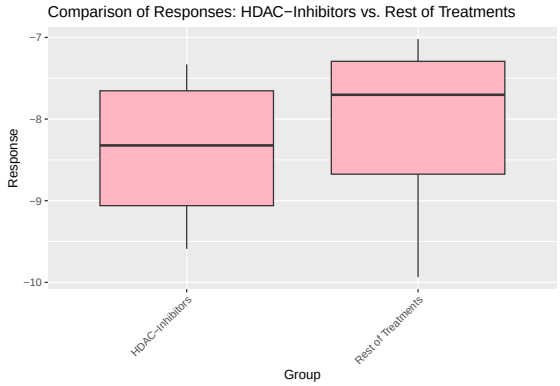

Figure 4: Boxplot Comparison of HDAC Inhibitors and Non-HDAC Inhibitors Treatments: Efficiency Analysis in Gastric Cancer Cell Line ACH-000948

3.6 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(SM: Fig. 13 and 14).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 (SM: Fig. 12). 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 (Fig. 5). 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 (SM: Fig. 14), which indicates a strong similarity of gene expression.

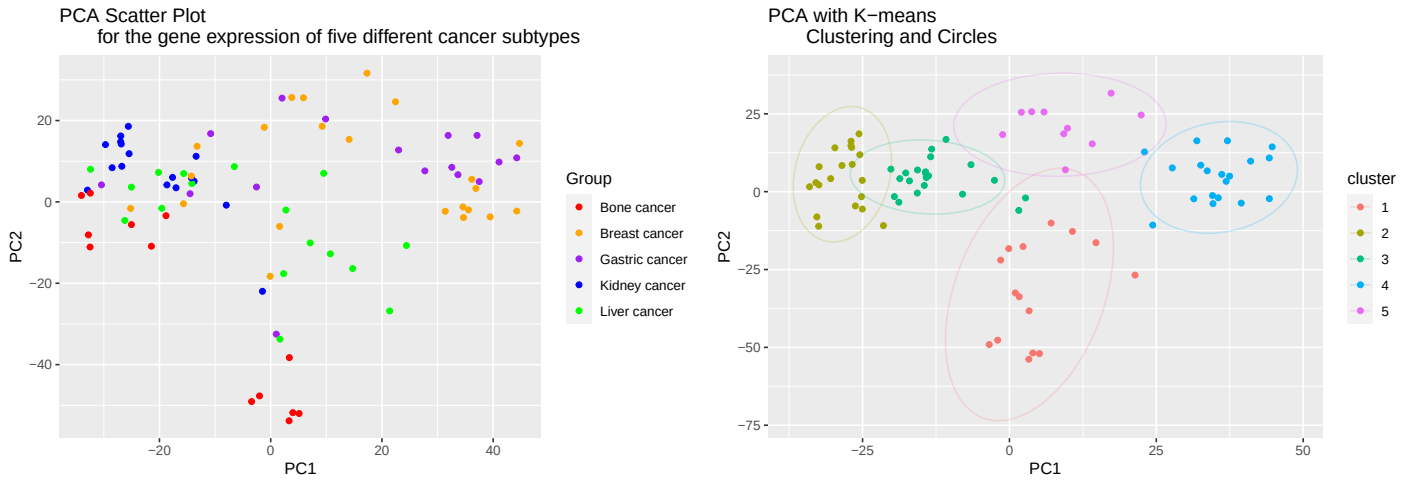

Figure 5: (Left): PCA scatter plot for gene expression of 5 different cancer subtypes. Each subtype is represented by a different color. Each dot corresponds to a cancer cell line. (Right): K - mean clustering applied for identifying different clusters

3.7 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(Fig. 5 and SM: Fig.12), 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 mechanisms of action (Fig. 6). 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.

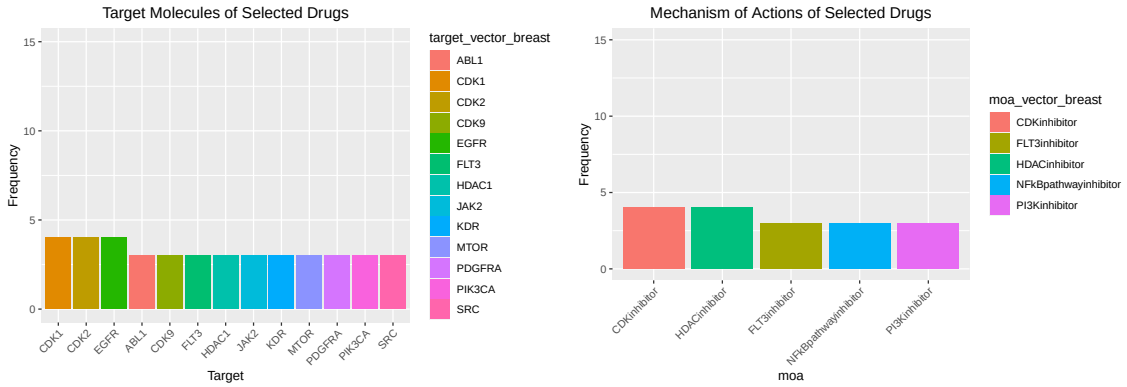

Figure 6: Barplot representation of the most frequent targets (left) and mechanism of action (right) in breast cancer cell lines.

3.8 Analyzing the correlation between drugs and their targets

Correlations between drug response and gene expression or CNVs of drug targets in gastric cancer cells identifie two drug-gene pairs for further analysis: dacinostat(BRD-K56957086-001-06-3)/*HDAC1* and GZD824 (BRD-K35329391-334-01-7)/BCR. These pairs demonstrate strong correlations between the drug response and both the gene expression and CNVs of their respective targets (SM: Fig. 20).

As before mentioned HDAC inhibitors seem to be a promising group of repurposing drugs for gastric cancer as the graphs depicting the drug response of dacinostat against *HDAC1* gene expression or CNVs in gastric cancer cell lines confirm (Fig. 7). 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 (Fig. 7). Same can be observed when looking at the scatter plots visualizing the relationship between the drug response of GZD824 and the gene expression or the CNVs of its target BCR (Fig. 8). 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 (Fig. 7). 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. 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 (Fig. 8). 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. 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.

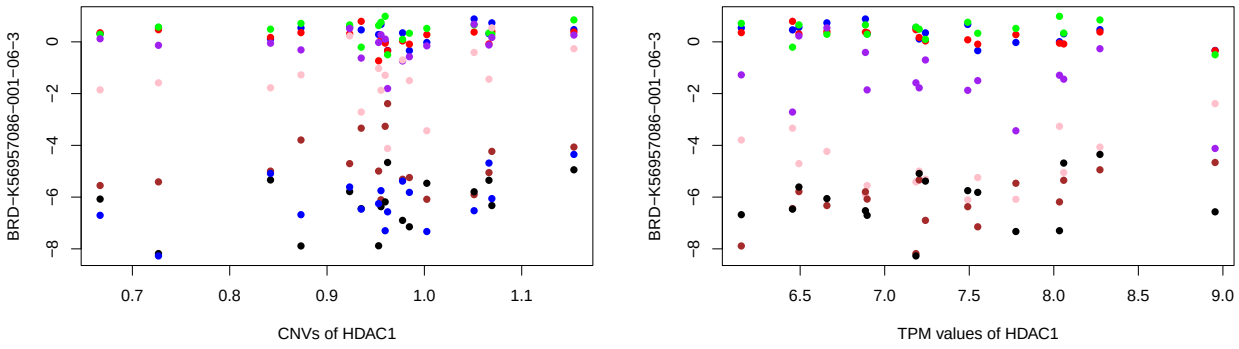

Figure 7: Plotting the drug response of dacinostat against the gene expression or CNVs of its target HDAC1

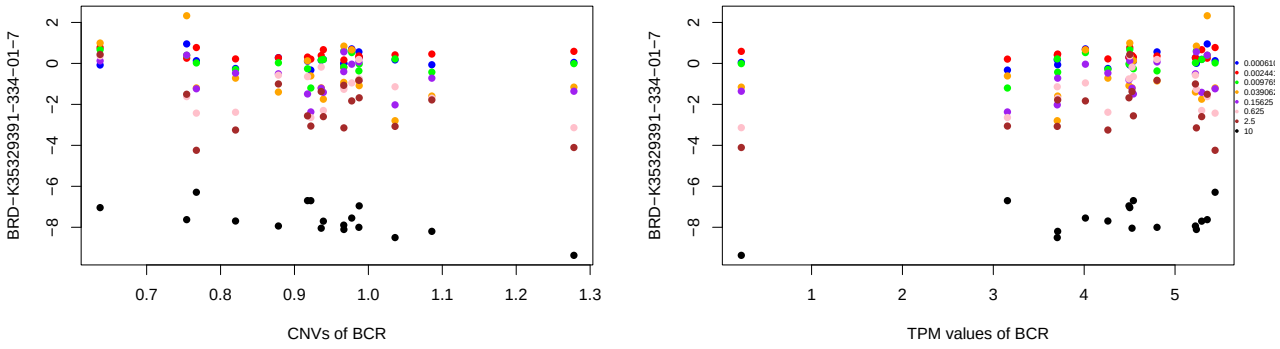

Figure 8: Plotting the drug response of GZD824 against the gene expression or CNVs of its target BCR.

3.9 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 are higher compared to the correlation between drug response and target gene expression and also show a constant increase when increasing the dosage, linear regression is 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 (SM: Fig.20). 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 models are 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 (Fig. 9). 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 (Fig.9) 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 (SM: Fig.21).

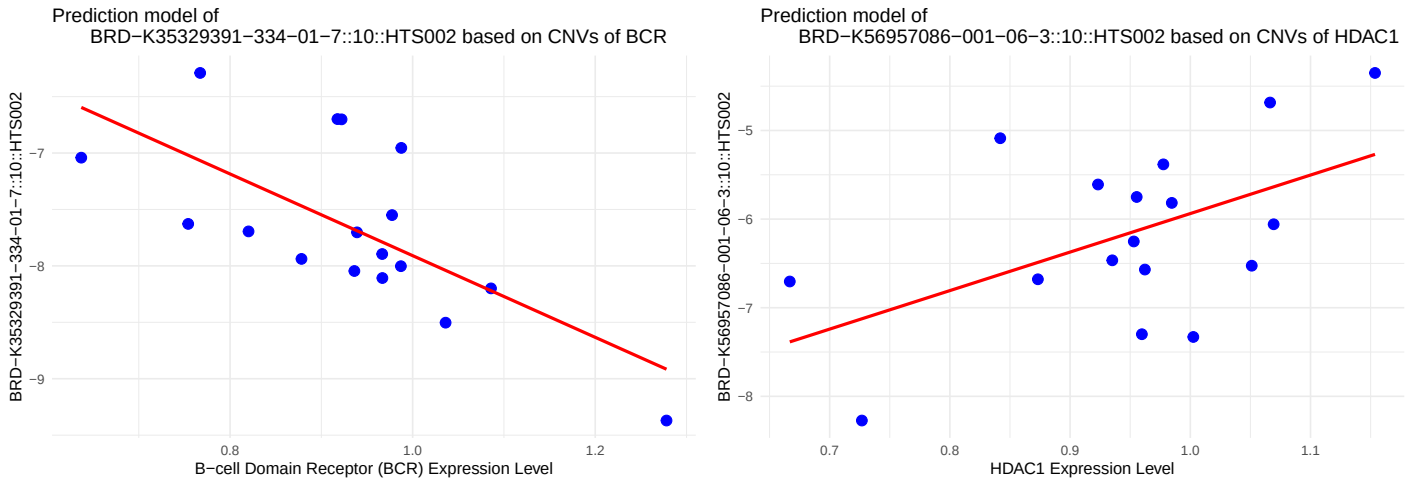

Figure 9: Prediction models for the drug response of dacinostat and GZD824 in gastric cancer cell lines.(Left)The equation explaining the results of linear regression analysis with the drug response GZD824(Y-dependent variable) and the CNVs of BCR (X- independent variable) in gastric cancer cells is:Y = -4.2900 - 3.6193 X. Both the intercept and the coefficient are statistically significant, indicated by the corresponding p-values (< 0.05). (Right) The equation which describes the created prediction model of dacinostat (Y-dependent variable) and the CNVs of HDAC1 (X- independent variable) in gastric cancer cells states as follows: Y = -10.285 + 4.347X. Both the intercept and the coefficient are statistically significant, indicated by the corresponding p-values (< 0.05).

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, fact that confirmed our theory. There were also some outlier cases that showed greater efficacy with non-oncology drugs. These results could be used for future research exploring the use of non-oncology treatments and studying the characteristics of these atypical cells. Through our focused analysis performing PCA and Pearson correlation, we selected specific drugs, which inhibit 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. 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, 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(Rajaselvi et al. (2023)). Furthermore, our analysis revealed intriguing similarities between gastric cancer and breast cancer, including shared targets (*HDAC1*) and mechanisms of action (HDAC inhibitors). 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 (Pramanik et al. (2022)). 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, we examined the the drug response and the gene expression or CNVs of its target. 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 cancer cells but no control group for this data was given, meaning no insight was given on how gene expression or CNVs differ generally from cells without cancer. 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 a benefitial approach to narrow the cancer types.

Supplementary Materials

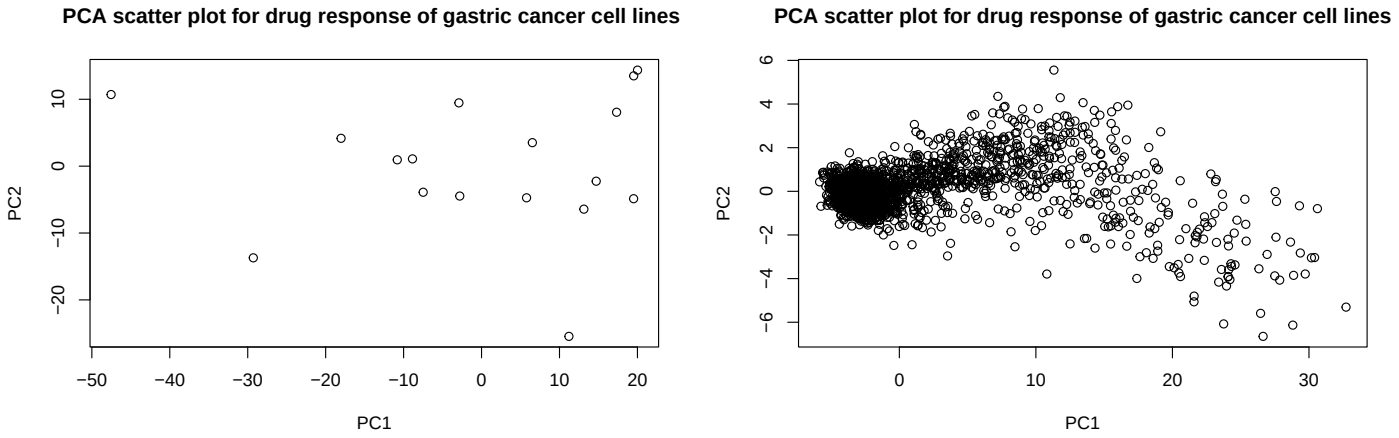

Figure 10: PCA scatter plot for drug response of gastric cancer cell lines. (Left): each dot represents a cell line. (Right): Each dot corresponds to a specific drug.

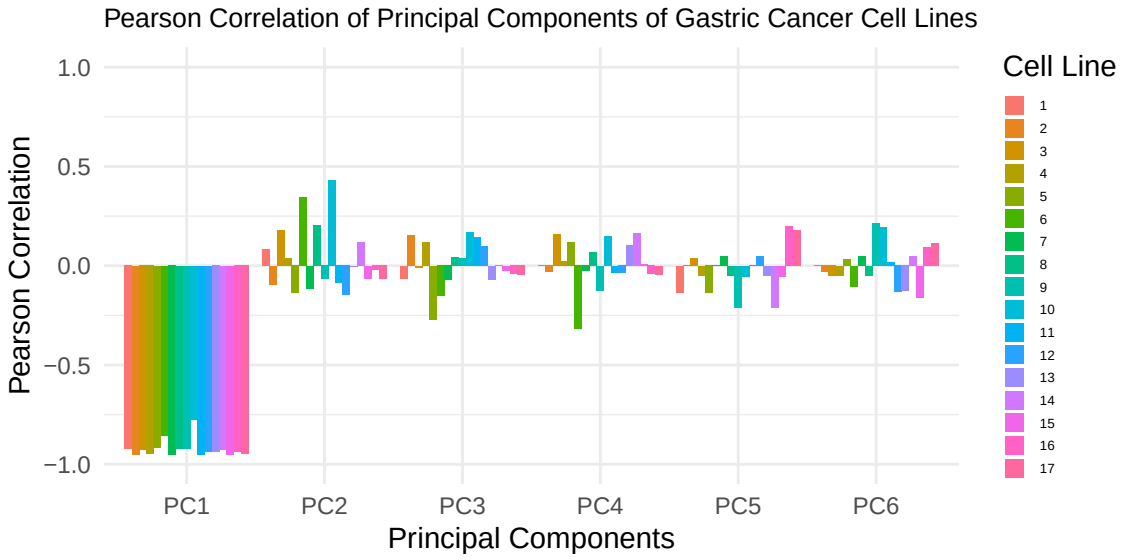

Figure 11: Pearson correlation of Principle components for gastric cancer cell lines. Each color represents different gastric cancer cell line.

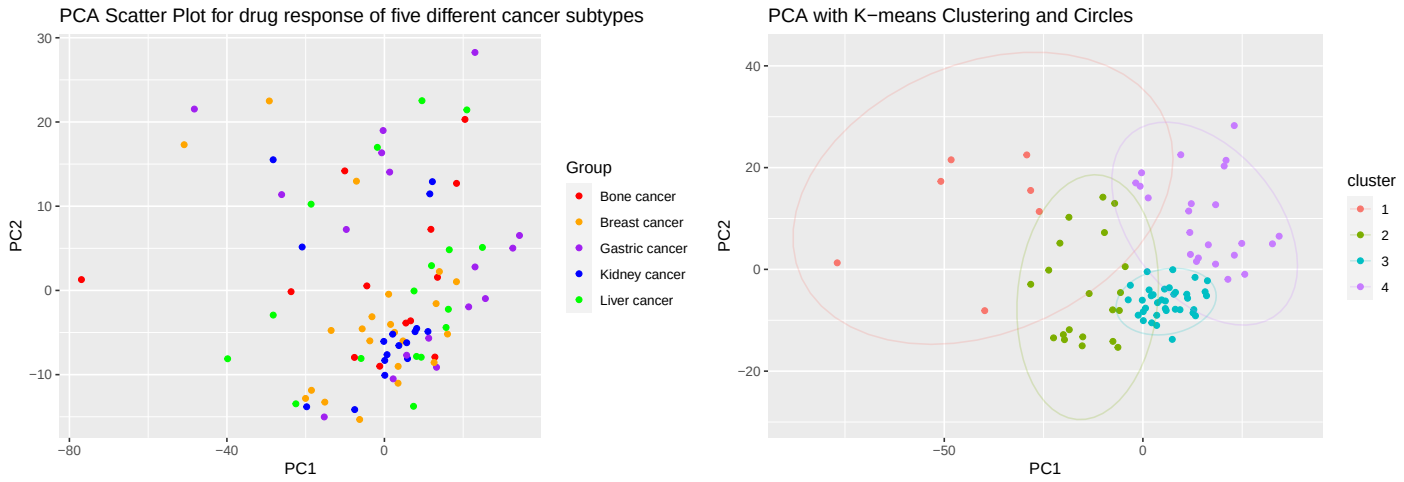

Figure 12: (Left): PCA scatter plot for drug response of 5 different cancer subtypes. Each subtype is represented by a different color. Each dot corresponds to a cancer cell line. (Right): K - mean clustering applied for identifying different clusters

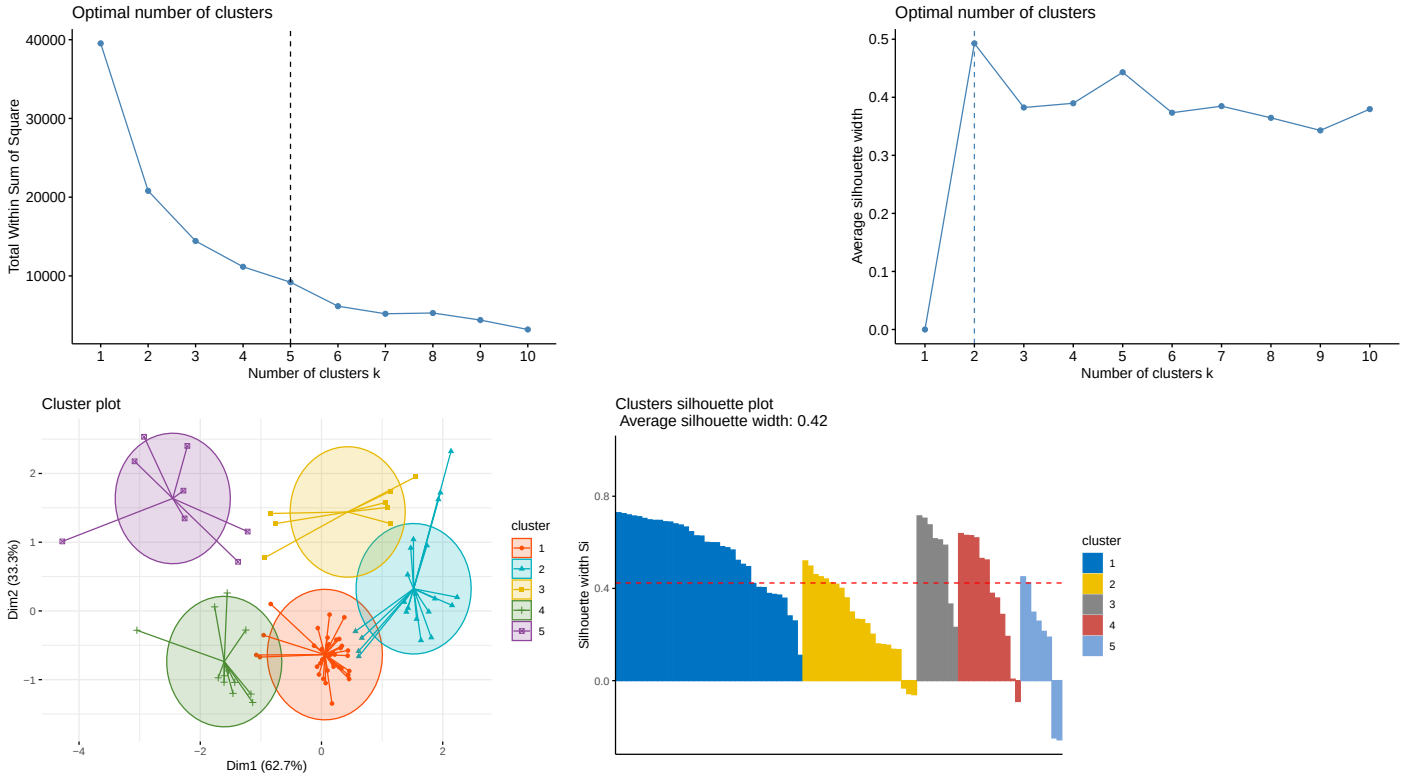

Figure 13: K-Means clustering of the PCA results for drug response from five cancer types based on the results of silhouette und elbow method

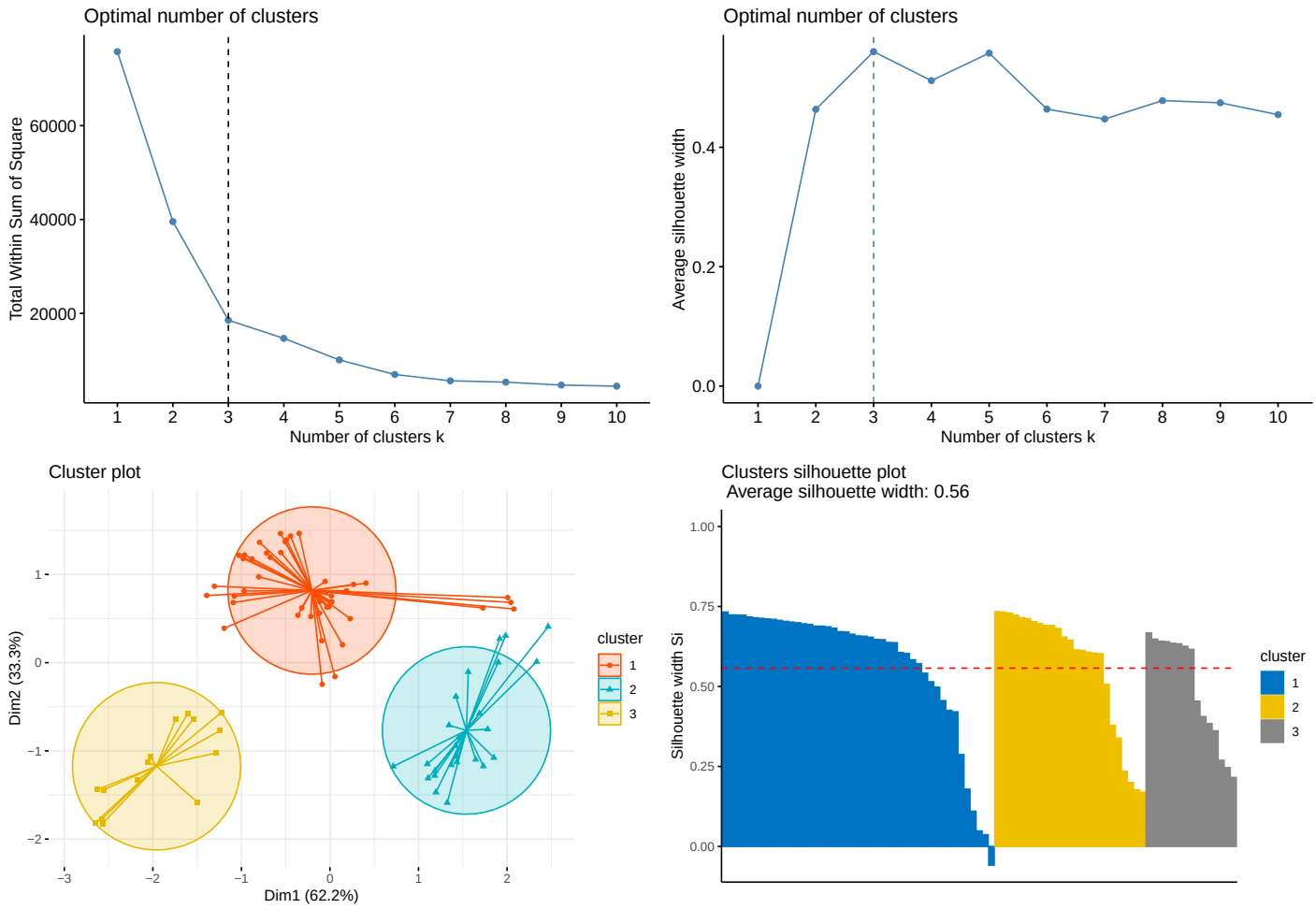

Figure 14: K-Means clustering of the PCA results for gene expression from five cancer types based on the results of silhouette und elbow method

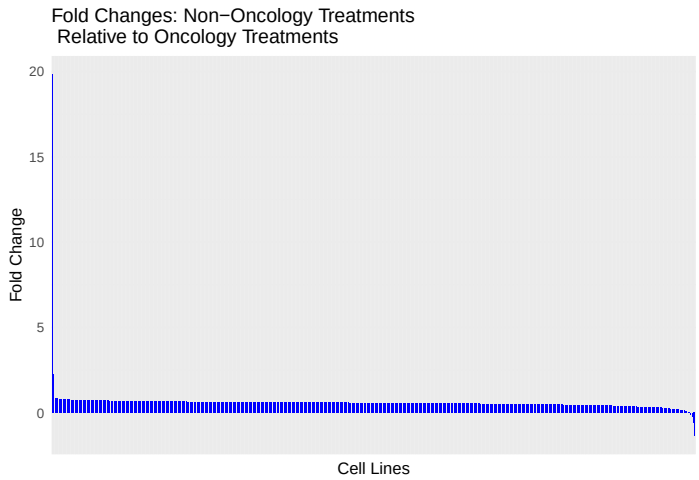

Figure 15: Bar Plot Representation of Fold Change Analysis: Comparing Efficiency of Oncology and Non-Oncology Treatments Across Various Cell Lines.

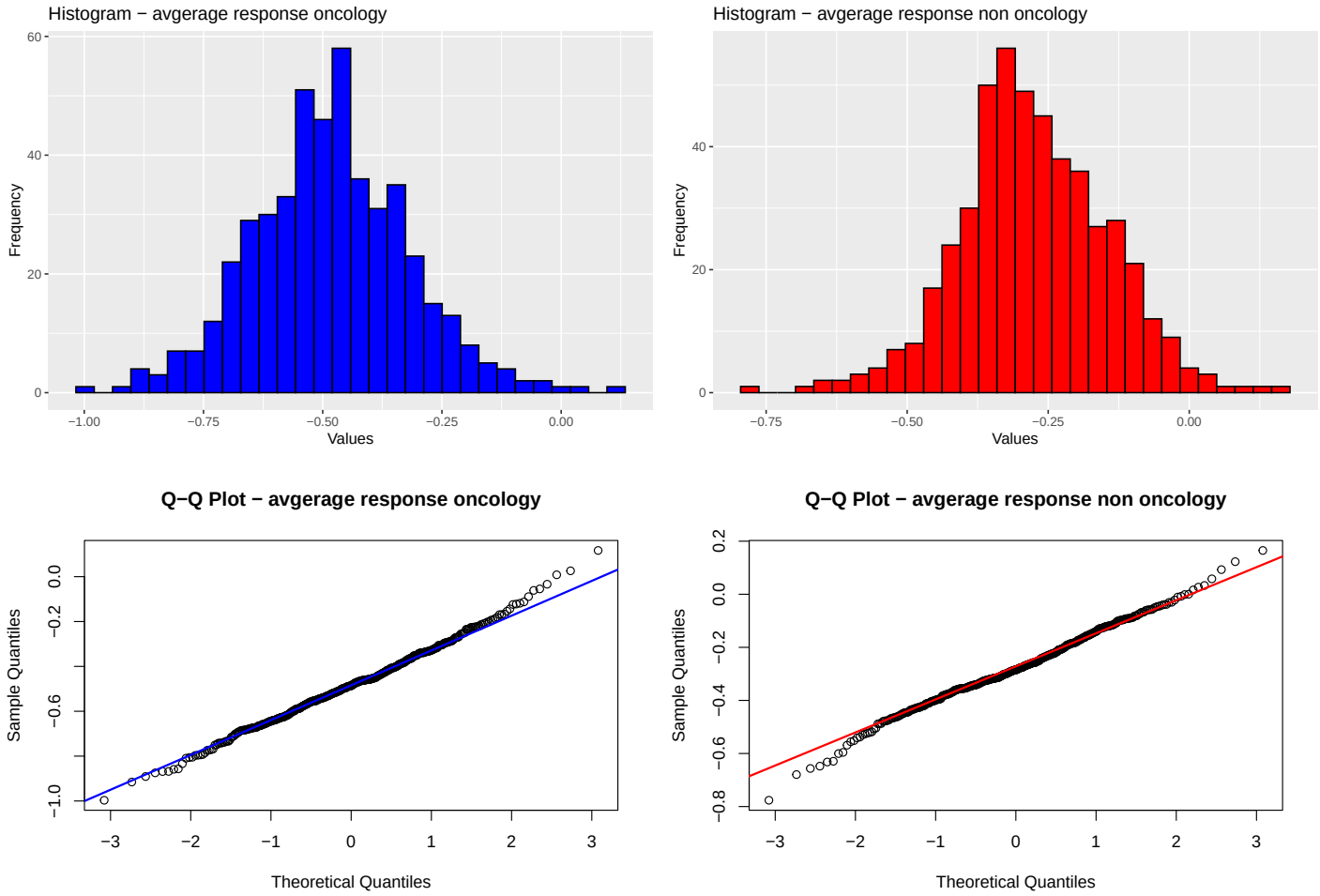

Figure 16: (Left):Histogram of Average Oncological Drug Responses for Normality Testing.(Right): Q-Q Plot of Average Oncological Drug Responses for Normality Testing

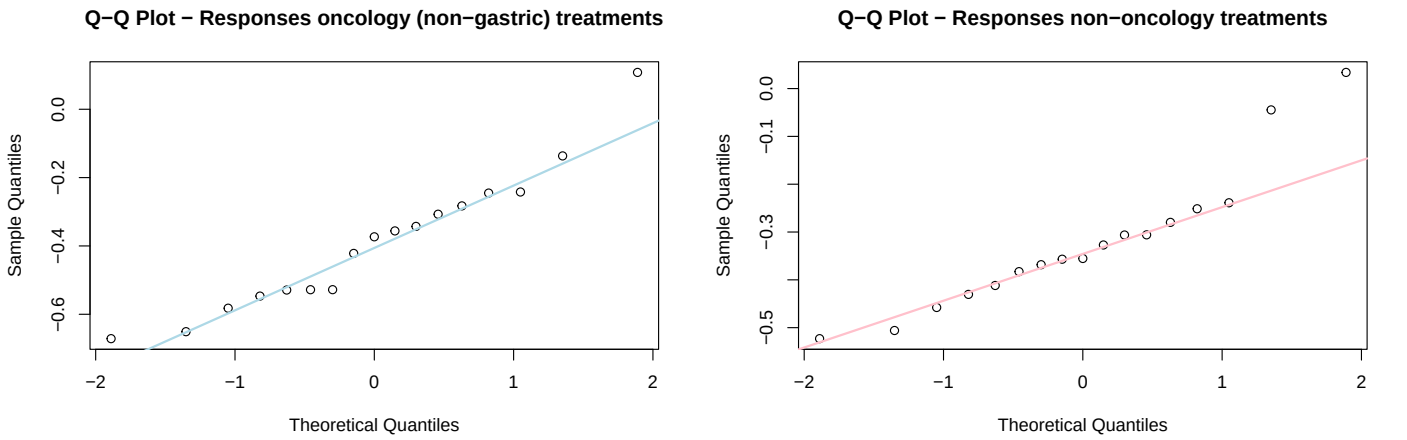

Figure 17: Q-Q Plot of Average Responses from Oncological Drugs for Normality Testing Fig X: Q-Q Plot of Average Responses from Non-Oncological Drugs for Normality Testing

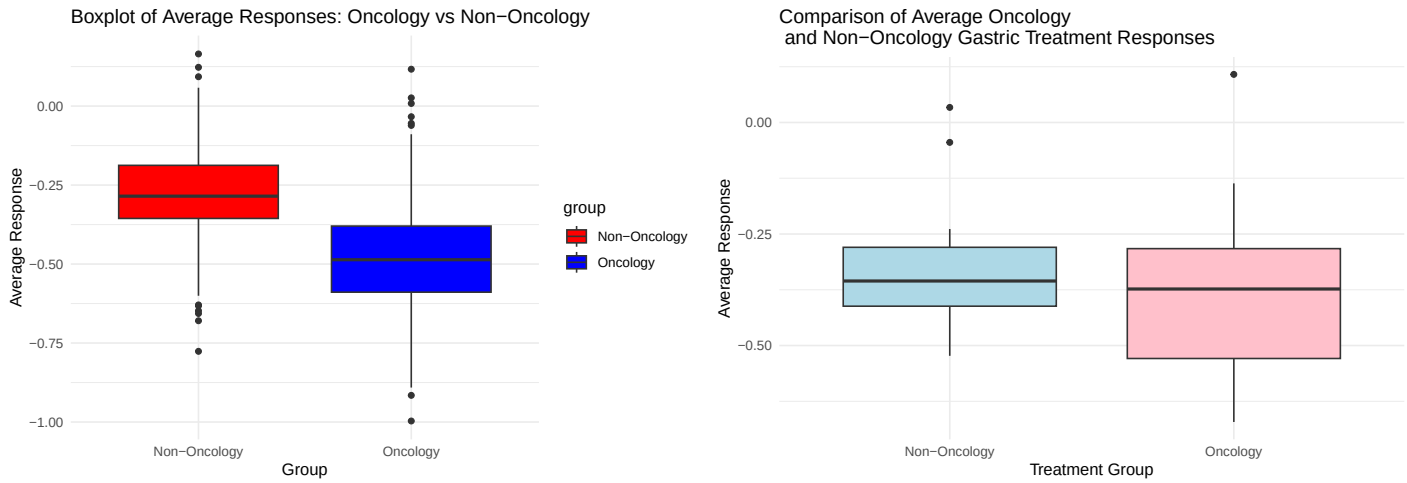

Figure 18: (Left):Boxplot of Average Responses: Comparison Between Non-Oncology and Oncology Treatments.(Right): Boxplot of Average Responses: Comparison Between Non-Oncology and Oncology Treatments in Gastric Cancer Cell Lines

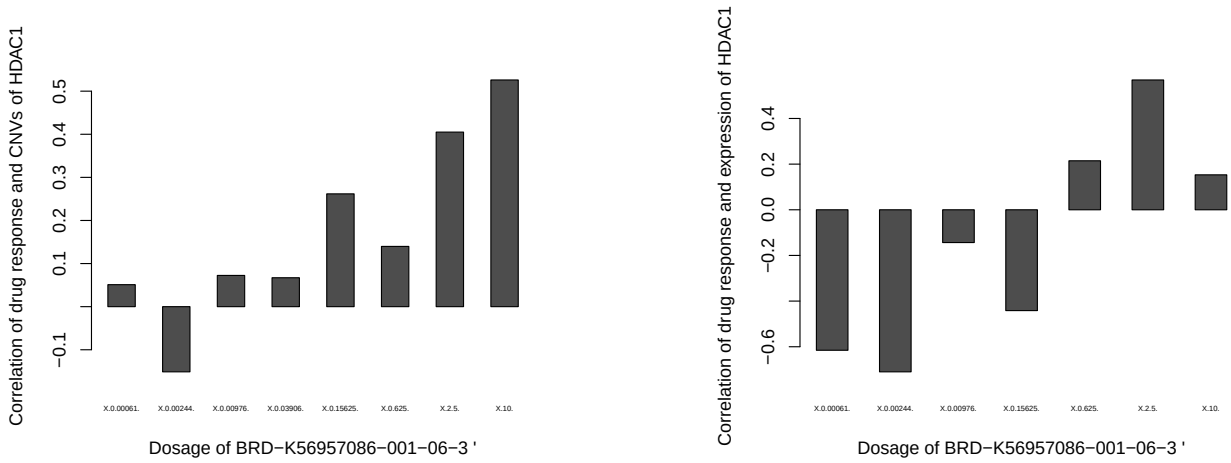

Figure 19: Barplot of the correlations between the drug responses of dacinostat and the gene expression or CNVs of its target HDAC1

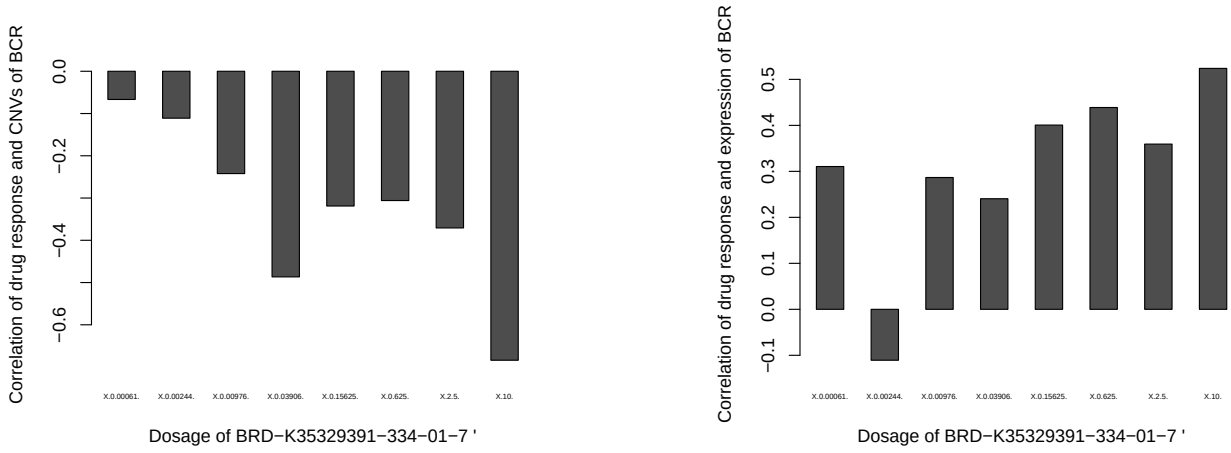

Figure 20: Barplot of the correlations between the drug responses of GZD824 and the gene expression or CNVs of its target BCR

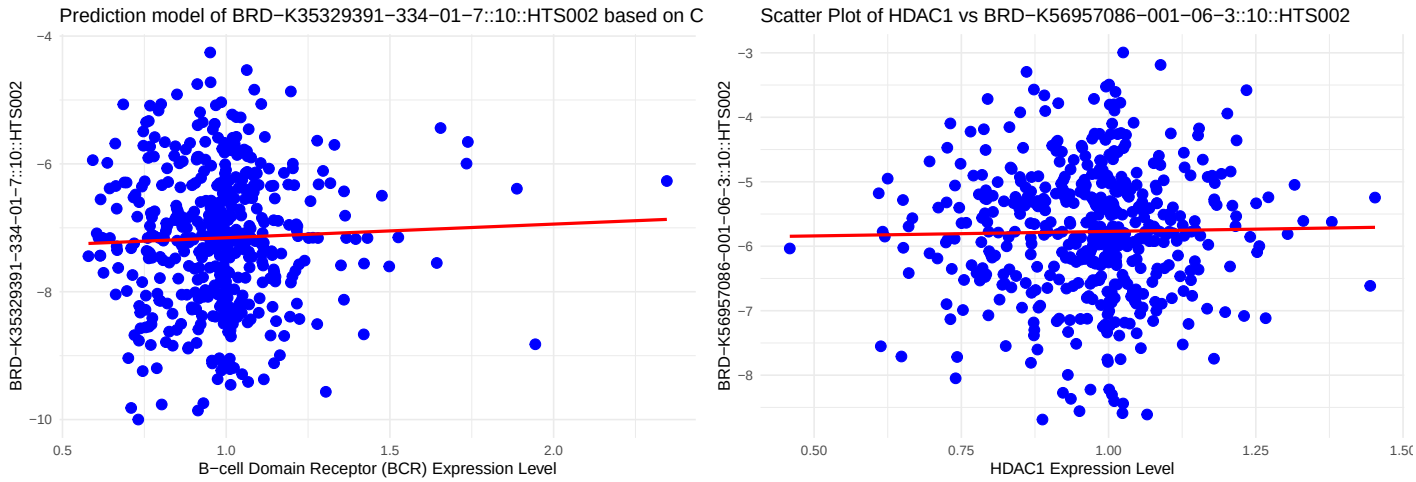

Figure 21: Prediction models for the drug response of dacinostat and GZD824 based on the CNVs of their target genes HDAC1 and BCR across all cell lines

References

Carcas, Lauren Peirce. 2014. "Gastric Cancer Review." *Journal of Carcinogenesis* 13.

Cebrián, A, G De la Concha, and I Fernández-Urién. 2016. "Cáncer gástrico." *Medicine-Programa de Formación Médica Continuada Acreditado* 12 (3): 118–27.

Gonçalves, Emanuel, Aldo Segura-Cabrera, Clare Pacini, Gabriele Picco, Fiona M Behan, Patricia Jaaks, Elizabeth A Coker, et al. 2020. "Drug Mechanism-of-Action Discovery Through the Integration of Pharmacological and CRISPR Screens." *Molecular Systems Biology* 16 (7): e9405. https://doi.org/https://doi.org/10.15252/msb.20199405.

Ioakeim-Skoufa, Ignatios, Natalia Tobajas-Ramos, Enrica Menditto, Mercedes Aza-Pascual-Salcedo, Antonio Gimeno-Miguel, Valentina Orlando, Francisca González-Rubio, et al. 2023. "Drug Repurposing in Oncology: A Systematic Review of Randomized Controlled Clinical Trials." *Cancers* 15 (11): 2972.

Pramanik, Siddhartha Das, Amit Kumar Halder, Ushmita Mukherjee, Dharmendra Kumar, Yadu Nandan Dey, et al. 2022. "Potential of Histone Deacetylase Inhibitors in the Control and Regulation of Prostate, Breast and Ovarian Cancer." *Frontiers in Chemistry* 10: 948217.

Rajaselvi, N Divya, MD Jida, KK Ajeeshkumar, Suresh N Nair, Preethy John, Zarina Aziz, and AR Nisha. 2023. "Antineoplastic Activity of Plant-Derived Compounds Mediated Through Inhibition of Histone Deacetylase: A Review." *Amino Acids*, 1–15.