UPTEC X 22029
Examensarbete 30 hp
Juli 2022

UPPSALA UNIVERSITET

ProTargetMiner one step further

Deep comparative proteomics of Dying vs. Surviving cancer cells treated with anticancer compounds

Albin Lundin

Civilingenjörsprogrammet i molekylär bioteknik

UPPSALA UNIVERSITET

ProTargetMiner one step further – Deep comparative proteomics of Dying vs. Surviving cancer cells treated with anticancer compounds

Albin Lundin

Abstract

Cancer is a leading cause of mortality worldwide, responsible for nearly one in six deaths. Thus, there is a need for a greater understanding of cancer for the development of novel therapeutics. This master thesis project aims to compare the proteome signatures between dying and surviving cancer cells treated with diverse anticancer drugs.

The first aim is to investigate if drug targets behave similarly and have the same sign (up- or down-regulation) in dying versus surviving cells. The second aim is to validate that combining the dying cancer cell's proteome with the surviving cell's can help improve drug target rankings for anticancer treatments. The third aim is to identify proteins and pathways involved in life and death decisions by comparing dying and surviving states in response to the anticancer drugs in different cell lines.

First, we demonstrate that drug target behaviour in dying versus surviving cells is almost identical for nine diverse anticancer compounds with a correlation of 0.93. To identify drug targets, orthogonal partial least squares-discriminant analysis (OPLS-DA) modelling was performed to contrast the proteome signature of one anticancer drug against all other drugs and rank the proteins based on the magnitude of the model's predictive component. There were occasions when the dying cells gave better rankings than the surviving ones. In some cases, the best target rankings were obtained when combining the data from both surviving and dying cells.

To identify proteins and pathways involved in life and death decisions, OPLS-DA modelling contrasting the two states was performed, and heatmaps and scatterplots of dying and surviving $\log_2$ fold changes were made. As a result, several pathways involved in cell survival and cell death were identified. In addition, at least six proteins consistently differentially regulated between the surviving and dying cells were identified. Such proteins can be considered as putative survival (resistance) or sensitivity biomarkers and serve as potential drug targets for the development of novel anticancer agents.

Teknisk-naturvetenskapliga fakulteten

Uppsala universitet, Utgivningsort Uppsala/Visby

Handledare: Dr. Amir Ata Saei and Prof. Roman Zubarev Ämnesgranskare: Prof. Jonas Bergquist

Examinator: Prof. Siv Andersson

Döende cancerceller – nyckeln till att överleva cancer?

Populärvetenskaplig sammanfattning

Albin Lundin

Minst en av tre personer i Sverige kommer få en cancerdiagnos någon gång under sin livstid (Bergman & Johansson 2018). I många fall går cancer att bota om den upptäcks tidigt och om medicineringen är effektiv, men trots detta är cancer en av de ledande dödsorsakerna världen över och orsaker ungefär ett av sex dödsfall (WHO 2022). Även om forskningen har gjort stora framsteg indikerar de höga dödstalen ett behov av nya och effektivare cancerläkemedel.

För att nya cancerläkemedel, och läkemedel generellt sett, ska bli godkända för användning behövs i många fall kunskap om vilket deras målprotein är. Att hitta detta målprotein är däremot lättare sagt än gjort eftersom varje människa har minst 20,000 olika proteiner kodade i sitt genom (Pray 2008). För att effektivare cancerläkemedel ska utvecklas behöver man också rikta in sig på nya målproteiner, det vill säga proteiner som spelar en avgörande roll för cancerns utveckling.

För att underlätta processen att hitta målproteiner för cancerläkemedel har metoder som utnyttjar masspektrometri utvecklats de senaste åren, vilka analyserar proteiner i överlevande cancerceller som utsatts för olika läkemedel. Genom att jämföra hur proteinnivåerna förändras som respons på läkemedel kan målproteinerna identifieras. Det finns däremot utrymme för förbättring av dessa metoder och ett sådant sätt kan vara att inkludera döende cancerceller i analysen.

Detta examensarbete syftar till att testa om målproteiner för olika cancerläkemedel är enklare att identifiera om både överlevande och döende cancerceller utsätts för denna typ av analys. Eftersom både överlevande och döende cancerceller analyseras, kan dessa jämföras för att hitta proteiner som skiljer dem åt. Dessa proteiner skulle vara involverade i celldöd och överlevnad och kan av den anledningen demonstrera hur cancern motstår behandling och undviker död. Därmed skulle dessa vara intressanta målproteiner för framtida cancerbehandlingar.

Resultatet av detta arbete visade att döende cancerceller kan vara av nytta vid identifiering av målproteiner för olika cancerläkemedel. I vissa fall kunde de döende cancercellerna identifiera målmolekylerna på ett bättre sätt än de överlevande, och i andra fall gav en kombinerad analys upphov till bättre identifiering. Utöver detta kan åtminståne sex proteiner med stora nivåskillnader mellan överlevande och döende cancerceller identifieras, vilka kan visa sig vara intressanta målproteiner för framtida cancerläkemedel.

Table of contents

Abbreviations

CV	Coefficient of Variation
GO	Gene Ontology
IRS	Internal Reference Scaling
LC-MS/MS	Liquid chromatography coupled tandem mass spectrometry
MOA	Mechanism of Action
MS	Mass Spectrometry/Mass Spectrometer
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
TMT	Tandem Mass Tag

1 Introduction

A critical part of drug discovery and development is the identification of drug targets and mechanisms of action, which can be very challenging processes. It is estimated that 7-18% of FDA-approved drugs lack a defined drug target showing that it is not an absolute necessity to characterize them. However, the identification of specific targets for a compound seems to increase approval chances (Moffat *et al.* 2017). Two common approaches used in drug target discovery are target-based discovery and phenotypic drug discovery (Moffat *et al.* 2017). The former relies on screening an already known drug target against a drug library to see which drug interacts with the target. In the latter approach, cells or a model system are exposed to various drugs to see which cell shows the desired phenotype (Lansdowne 2018). The downside of target-based discovery is that prior knowledge of the drug target is needed, while phenotypic drug discovery faces challenges in target deconvolution (Moffat *et al.* 2017).

In recent years, methods involving mass spectrometry have been developed to identify drug targets and mechanisms of action. These methods are called Functional Identification of Target by Expression Proteomics (FITExP) (Chernobrovkin *et al.* 2015) and ProTargetMiner (Saei *et al.* 2019), and these can identify drug targets in datasets comprising thousands of proteins. Both methods analyze the protein regulation in surviving cells from multiple cell lines treated with anticancer compounds. Using multiple cell lines and assuming that the drug target is regulated similarly across all cell lines, the drug target can be more easily identified since unrelated proteins regulated in a cell-specific way can be filtered out (Chernobrovkin *et al.* 2015). Furthermore, when using multiple anticancer compounds, the proteome responses can be contrasted against each other to filter out proteins that are always up or down-regulated regardless of treatment (e.g. proteins involved in detoxification or those generally related to cell death), which allows for more accurate identification of the drug target and mechanism of action (Chernobrovkin *et al.* 2015). OPLS-DA modelling is used in both methods to visualize the drug targets (Chernobrovkin *et al.* 2015, Saei *et al.* 2019).

When cell lines are exposed to anticancer treatments, the matrix attached or surviving cells transition to cell death by first detaching from the matrix (Saei *et al.* 2018). Since the detached cells further progress towards death, they have the most significant response to the anticancer compounds. Therefore they might reflect the drug target and mechanism of action better than the surviving, still attached cells. A study comparing the proteome response between dying and surviving cells for three anticancer compounds (Saei *et al.* 2018) found that drug targets and mechanisms were highly similar for the dying and the surviving cells. Furthermore, it was found that combining proteomics data from dying and surviving cells can improve the drug target identification in most cases. Also, studying dying and surviving cells made it possible to identify proteins differentially regulated between the dying and surviving states regardless of the treatment and cell line used, hinting on cell death and survival pathways.

One of the downsides of the study comparing the proteome response between dying and surviving cells (Saei *et al.* 2018) was the few anticancer drugs used. In the ProTargetMiner

paper (Saei *et al.* 2019), proteomic data from 9 drugs and the three major cancer cell lines A549, MCF-7, and RKO were obtained from surviving cells, and now corresponding data have been collected for dying cells.

1.1 Goals

This Master Thesis project aims to analyze the ProTargetMiner data complemented with data from dying cells to validate the findings made in the previous project comparing the proteome response in dying and surviving cells (Saei *et al.* 2018). This project has the following four goals:

1. To validate that dying has a more significant impact on the cellular proteome than the treatment and cell line used.
2. To test the hypothesis that the regulation of drug targets is similar between the dying and surviving cells across the three cell lines and nine anticancer compounds used in the study.
3. To validate that merging proteomics data from dying and surviving cells can improve drug target rankings extracted from OPLS-DA models.
4. To compare the proteomes of dying and surviving cells to discover pathways and proteins contributing to cell death and survival, regardless of the cell line or anticancer drug used. While proteins found to be differentially regulated in the dying state would be hypothetically involved in cell death, proteins regulated in the surviving cells would be involved in cell survival and drug resistance. Cancer treatment aims to kill the cancerous cells, so an increased understanding of how they resist drug treatment could offer new opportunities for treating cancer.

1.2 General introduction to Mass Spectrometry

A mass spectrometer is an analytical tool used for measuring ionized molecules' mass-to-charge (m/z) ratio. It consists of an ion source that ionizes molecules, a mass analyzer that measures the m/z ratio, and a detector that identifies the number of ions at each m/z value. Electrospray ionization, or ESI, is a common ion source used to ionize peptides in a solution, which can be coupled with a liquid chromatography unit. The ion source is coupled to a liquid chromatography unit, where the complex molecular solution can be separated over time by liquid chromatography and analyzed individually (Aebersold & Mann 2003).

When using Mass Spectrometry in proteomics, it is often peptides and not whole proteins that are analyzed in the mass spectrometer. Whole proteins can be challenging to handle and are not always soluble under the same conditions. Also, it is difficult to predict which protein might give rise to a measured protein m/z peak due to all modifications possibly present on a protein. When working with peptides obtained by exposing whole proteins to a protease during sample preparation, it is much easier to obtain sequence information from the mass spectrometer, which is needed for identifying and quantifying proteins (Steen & Mann 2004).

Two mass spectrometers can be coupled together and used for tandem MS. In this case, the mass to charge ratio of the ionized peptide is measured in the first MS, MS1. Before entering the second mass spectrometer, MS2, peptides are isolated and fragmented by an inert gas. The mass spectra from MS2 can then be used to identify the individual amino acids present in the precursor peptide necessary for its identification (Steen & Mann 2004).

For multiplexing MS analyses, Tandem Mass Tag (TMT) labelling can be performed. TMT labels consist of a reporter group with a different weight for each label, a mass normalizer that makes the overall weight of different TMT labels the same, and an amine-reactive group used to attach the label to the N-terminus or lysine residue of the peptide. Different samples can be labelled with different TMT labels before being pooled and analyzed together by Tandem MS. Since the labels have the same overall weight and chemical properties, the same peptides from all samples will elute simultaneously from the liquid chromatography unit, giving rise to a single peak in the MS1 spectra. In the MS2 spectra, the intensities of the different reporter ions are used to quantitate the peptide amount in each sample, while the peptide fragments peaks are used to identify the peptide (Thermo Fisher Scientific).

1.3 Internal reference scaling (IRS) normalization

For LC-MS/MS, more analytes are eluting from the chromatography unit than the mass spectrometer can handle. As a result, the mass spectrometer samples a relatively small number of analytes meaning that proteins not detected by the mass spectrometer may still be present in the sample. Furthermore, the mass spectrometer can sample the analyte anywhere between the baseline and the analyte elution peak. Therefore, the same analyte could be sampled at different intensities between mass spectrometry runs, meaning that technical replicates analyzed in different runs would not have the same abundance even if the same amount of protein were present in the replicates. Furthermore, while the samples are prepared in parallel, the high-pH reverse phase fractionation process can also induce slight variations between multiplexed samples. Overall, these phenomena induce a batch effect that must be removed before data analysis. One way of removing this type of batch effect is by a normalization technique called Internal Reference Scaling, or IRS (Plubell *et al.* 2017).

The principle behind IRS normalization is relatively simple. By analyzing an identical pool of proteins in each TMT experiment, scaling factors can be calculated to adjust each protein of the pooled standard to the same intensity across all TMT experiments. The scaling factor of each TMT experiment is then applied to all ion channels putting all samples on the same intensity scale. Thus, apart from the samples being normalized, any present batch effects between the TMT experiments are also removed (Plubell *et al.* 2017).

1.4 Orthogonal Partial Least Squares Discriminant Analysis Modelling

Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) is a multivariate data analysis tool that can classify data with one or more classes. It is well suited to work with noisy

data, of which biological data is a typical example. OPLS is an extension of Partial Least Squares (PLS) modelling, a supervised classification approach used to assess relationships between a descriptor matrix X and a response matrix Y. The descriptor matrix contains the sample data, while the response matrix contains the class of each sample. PLS works both in terms of quantitative and discriminant analysis, the latter case being called Partial Least Squares Discriminant Analysis (PLS-DA). The difference between OPLS-DA and PLS-DA lies in how they discriminate between the classes. While the loading vectors of a PLS-DA model will contain mixtures of discriminatory and non-discriminatory properties, OPLS-DA modelling can separate the discriminatory loading vector (predictive component) from the non-discriminatory vectors (orthogonal components). This property makes it easier to interpret which features have the highest predictive power or are most specific for the classes. In addition, when only discriminating between two classes, the models are straightforward to interpret since there is only one predictive component. The orthogonal component of an OPLS-DA model represents the within-class variation (Bylesjö *et al.* 2006).

When it comes to validating OPLS-DA performance, there are several measures. The R2 value is a measure that represents the percentage of the predictive variance explained by the full model, while the Q2 value represents the model's predictive performance obtained through cross-validation. Both measures take values between 0 and 1, where values close to 1 indicate better model performance (Thévenot *et al.* 2015).

1.5 Experimental procedures

Human lung adenocarcinoma A549 cells, breast adenocarcinoma MCF7 cells, and colon carcinoma RKO cells were grown in flasks for 24 hours. Samples were treated with anticancer compounds in triplicates for 48 hours, causing 50% cytotoxicity (LC50). The matrix attached cells were collected and trypsinized, washed, and lyzed. The detached cells were collected by centrifuging the cell media and were washed. 50 µg of protein were kept for each sample. Proteins were digested with Trypsin for 6 hours. TMT reagents were added and left to incubate for 2 hours, whereafter the samples were combined. The peptides were separated into 96 fractions using a Dionex Ultimate 3000 2DLC (Liquid chromatography) system over a 48-minute gradient. The fractions of detached and attached cells for MCF7 and RKO and the detached cells of A549 were concatenated into 24 samples. The attached cells of A549 were concatenated into 16 samples. The protein samples were analyzed in randomized order by LC-MS/MS. The anticancer compounds used can be seen in Table 1 and were selected to have a high diversity of targets and MOA, based on ProTargetMiner (Saei *et al.* 2019).

The raw mass spectrometry data was searched in MaxQuant version 1.5.6.5 for protein quantification, and the Andromeda search engine was used to identify the proteins. In addition, the raw mass spectrometry data was IRS normalized before further data analysis.

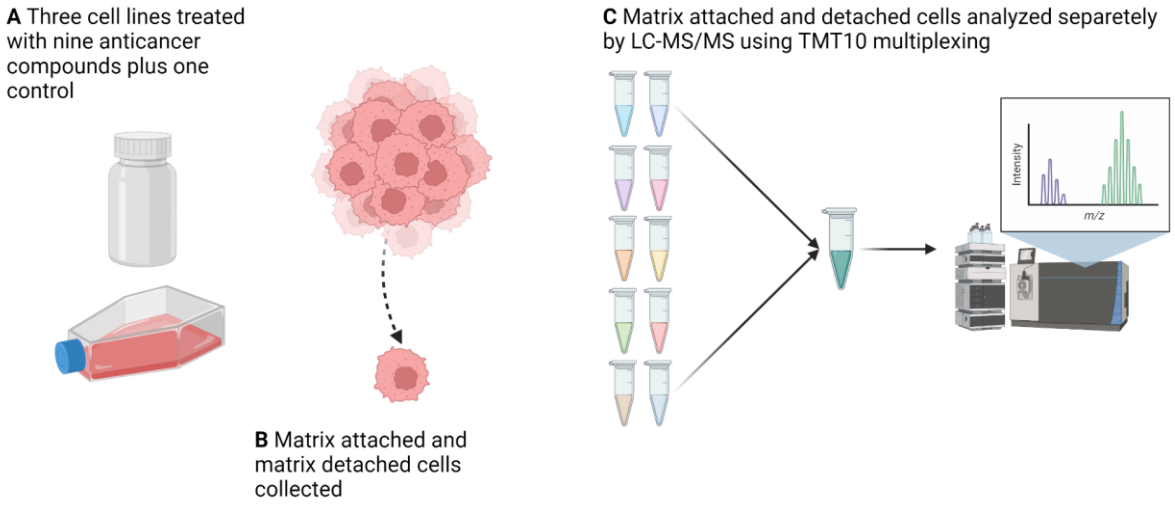


Figure 1. Overview of the experimental procedures. The three major cancer cell lines A549, MCF7, and RKO are treated with nine anticancer compounds and one control, as seen in Table 1. Matrix attached and detached cells are collected and analyzed separately by LC-MS/MS using TMT10 multiplexing.

Table 1. The drugs used in the study with drug targets from DrugBank.
*** Drug targets identified by Saei *et al.* (2019) that are not included in DrugBank**

Drug	Targets
DMSO (control)	-
8-azaguanine	PNP
Raltitrexed	FPGS, TYMS
Topotecan	TOP1MT, TOP1
Floxuridine	TYMS
Nutlin	MDM2, TP53
Dasatinib	ABL1, SRC, EPHA2, LCK, YES1, KIT, PDGFRB, STAT5B, ABL2, FYN, BTK, NR4A3, BCR, CSK, EPHA5, EPHB4, FGR, FRK, HSPA8, LYN, ZAK, MAPK14, PPAT, PARG*
Gefitinib	EGFR
Vincristine	TUBB, TUBA4A
Bortezomib	PSMB5, PSMB1, DPP3*, DPP7*

2 Methods

The dataset used in this project contains IRS normalized data for the Dying and Surviving states. Both states contain data from three carcinoma cell lines, A549, RKO, and MCF7, for which there is data for nine drugs and one control sample in triplicates. Apart from the protein intensities across the states, cell lines, and drugs, the output from MaxQuant also contains additional data, such as flags for contaminant proteins, sequence coverage, and the number of peptides used to identify the proteins. In this section, the data analysis methods used to complete

the goals presented in section 1.1 are described in detail. An overview of the workflow can be seen in Figure 2.

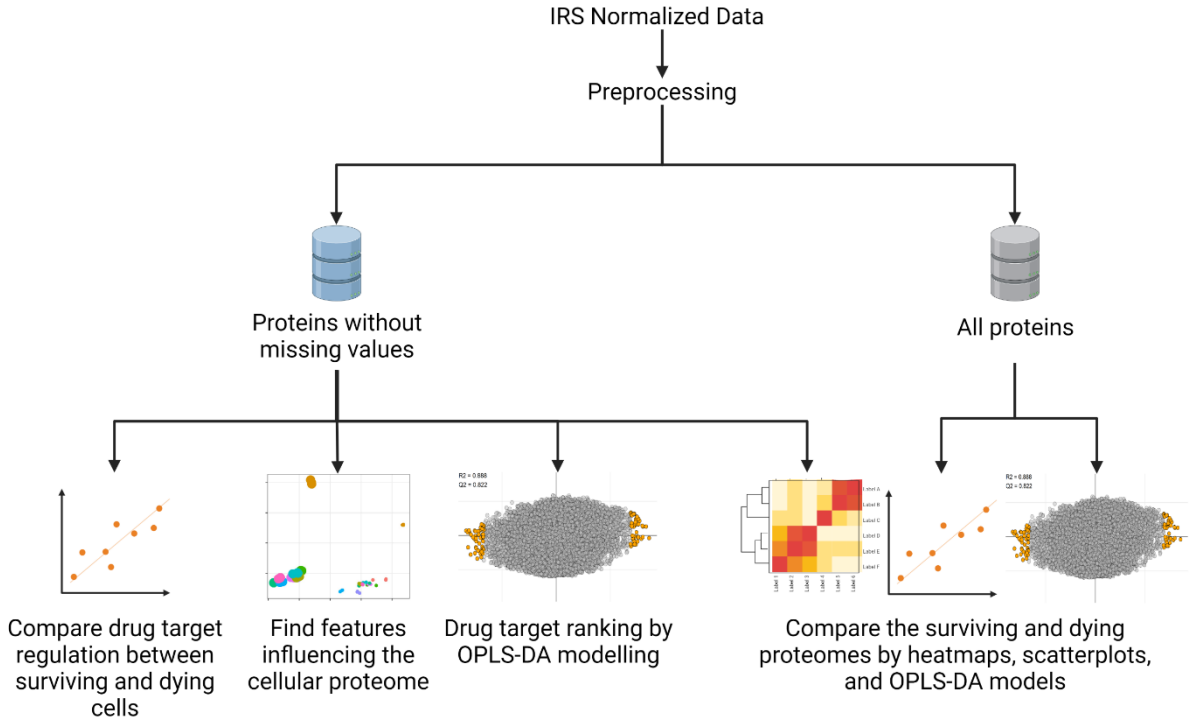


Figure 2. Overview of the workflow. First, the IRS normalized data is preprocessed, and two data sets are created. One containing only the proteins with missing values, and the other containing all proteins regardless of missing values. With the dataset containing proteins without missing values, drug target regulation is compared between surviving and dying cells and features influencing the cellular proteome are identified. Also, drug target rankings are derived through OPLS-DA modelling, and the surviving and dying states are compared through heatmaps. With the dataset containing all proteins, the dying and surviving proteomes are compared through scatterplots and OPLS-DA modelling.

2.1 Pre-processing

Proteins marked as reverse, only identified by site, contaminants, or only identified with one peptide were removed. Next, two datasets were created. One dataset contained only the proteins without missing values across all samples, used for working with drug targets. The other contained all proteins regardless of the number of missing values present across the samples, which could be of value when discriminating between surviving and dying cells. For the dataset keeping all proteins regardless of missing values, proteins that showed no variance across all samples were removed since they would not have any discriminatory power. Finally, all missing values were imputed to the minimum value of the respective cell line. The dataset allowing proteins with missing values had a final protein count of 10,483, while the dataset containing only proteins without missing values had 5,072.

For each TMT set of both datasets, $\log_2$ fold changes were calculated by $\log_2$ transforming the IRS value ratio between each drug and the control. P-values were calculated using a two-tailed t-test comparing the $\log_2$ fold changes for the replicates of each drug to the $\log_2$ fold changes of

the controls. Lastly, volcano plots were created to visualize the drug targets and mechanistic proteins.

2.2 Comparing the protein regulation between dying and surviving cells

To compare the protein log2 fold changes for the known targets of the different drugs, including those found in a previous study (Saei *et al.* 2019), mean log2 fold changes were calculated for the drug targets for each state of each cell line. The dataset containing only proteins without missing values was used in this case. The obtained mean values were visualized in a lollipop plot, highlighting the mean of each cell line and each cell state, and in a scatter plot showcasing the dying $\log_2$ fold change on the x-axis and the surviving $\log_2$ fold change on the y-axis. A linear regression was made between the dying and surviving $\log_2$ fold changes in the scatter plot, and the Pearson correlation coefficient was calculated.

2.3 Determining the feature with the highest impact on the cellular proteome

PCA plots of combined and individual cell lines were made to test what features had the most significant impact on the cellular proteome. These were made with IRS normalized values from the dataset containing proteins without missing values. The principal components were investigated to identify what characteristics contribute to the variance in the dataset. In the PCA plot combining the cell lines, a CV threshold of 15% was used, leaving 2,115 proteins.

2.4 Drug target rankings

Four different types of OPLS-DA models were built for contrasting each drug against all others, using the log2 fold changes of the dataset consisting of proteins without missing values. These models contrast each drug against all others:

1. Within the same cell line and cell state
2. Within the same cell line combining the cell states
3. Within the same cell state combining the cell lines
4. For all data combined

For all models, proteins that were more upregulated in response to the treatment than to the other drugs were specified to have positive specificity values in the predictive component, while proteins more downregulated in response to the treatment were specified to have negative specificity values.

The protein rankings were based on the specificity values obtained in the predictive component. Proteins with negative specificity were ranked separately from the proteins with positive specificity. The rankings were derived based on the order of magnitude in both cases. For instance, the protein with the highest specificity would be ranked as number one, while a protein with the 100th highest value would get ranked as number 100. The derived target rankings were put into a table to compare the exact ranking of the different targets between the different models. A boxplot was created to get a general overview of the rankings of each model.

2.5 Differentially regulated proteins and pathways between dying and surviving cells

Multiple approaches were taken to find proteins and pathways differentially regulated between the two states. These include scatterplots showcasing the surviving $\log_2$ fold changes against the dying $\log_2$ fold changes, OPLS-DA modelling contrasting the $\log_2$ fold changes of living and dying states, and making heatmaps with the IRS normalized values using k-means hierarchical clustering to cluster the proteins.

2.5.1 Scatterplots

Scatterplots were made with the dying $\log_2$ fold changes on the x-axis and the surviving $\log_2$ fold changes on the y-axis using the dataset including all proteins regardless of missing values. The scatterplots were made both on individual cell lines, comparing the cell line mean of $\log_2$ fold changes for the two states, and the combined cell line data comparing the mean of the overall $\log_2$ fold changes. For all scatterplots, four clusters were highlighted. These clusters were obtained by comparing the regulation of one state compared to the other, forming clusters of proteins showing regulation in one state while showing none or opposite regulation in the other state. For instance, one cluster highlights proteins upregulated in the dying state that shows no regulation or even downregulation in the surviving state. In all scatterplots, the $\log_2$ fold change threshold for up and downregulation was set to 0.5 and -0.5, respectively. Finally, the proteins of each cluster were submitted for pathway analysis using GOrilla (Eden *et al.* 2009).

2.5.2 OPLS-DA Modelling

Another approach used for discovering differentially regulated proteins between dying and surviving cells is OPLS-DA modelling. OPLS-DA models were built to contrast surviving versus dying states for all cell lines combined and for individual cell lines, using the dataset consisting of all proteins regardless of the number of missing values. The 50 proteins with the highest negative respective positive specificity were submitted to pathway analysis using GOrilla (Eden *et al.* 2009).

2.5.3 Barplots

To see how reproducible the results were for the found proteins in the scatterplots and OPLS-DA models across all cell lines and drug treatments, bar plots of the log2 fold changes were made showing the regulation of the dying and surviving states. Proteins showing reproducibility across the cell lines and drugs were examined further regarding their general function and involvement in cell death or survival. Barplots of the proteins found to be differentially regulated between the two states in the previous study (Saei *et al.* 2018) were also made.

2.5.4 Heatmap

As another option for finding pathways separating the two states, heatmaps with k-means hierarchical clustering of proteins were made. The heatmaps were made on $\log_{10}$ transformed IRS values (and not $\log_2$ fold changes) for the combined data of all cell lines and for individual cell lines. P-values were calculated to measure how different the living and dying states were

for each cluster. Finally, the proteins in each cluster were submitted to pathway analysis using GOrilla (Eden *et al.* 2009).

3 Results

In this section, the results are presented.

3.1 Dying has the most significant impact on the cellular proteome

PCA analysis was performed on the combined cell line data and individual cell lines to determine which feature contributes to the highest variance across the data set. As seen in Figure 3, the first two principal components are plotted against each other for the four cases. In the PCA plot of the combined cell line data, the first principal component separates the dying and surviving cells. The second component separates the cell lines.

For the PCA plots of the individual cell lines, the separation is even more apparent. In all three cases, the cellular state drives the separation of the first principal component, accounting for 80% of the total variation in the dataset for the MCF7 cell line. The second component of cell lines A549 and MCF7 separates the drug Bortezomib from the other drugs, while for the RKO cell line it separates Bortezomib and Vincristine from the other drugs. Since the first component accounts for the highest proportion of variance within the dataset, it can be concluded that dying has the most significant impact on the cellular proteome rather than the type of cell line or drug in use.

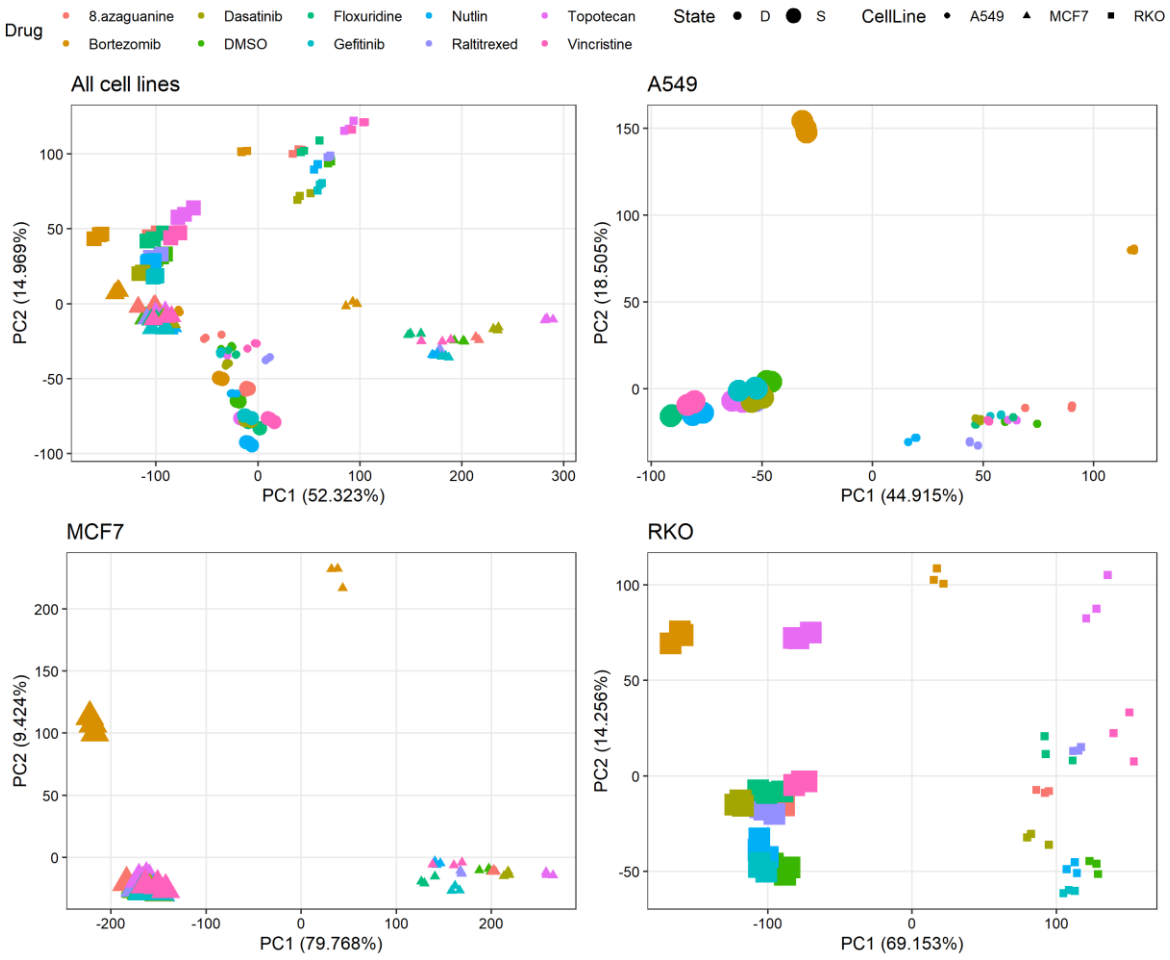


Figure 3. PCA plots for all cell lines combined and individual cell lines. The shape represents the cell line, the size represents the cell state, and the colour represents the drug treatment. There are three points for each combination of drug, cell line, and state, representing the replicates.

3.2 The regulation of drug targets is highly similar in the dying and surviving cells

The mean $\log_2$ fold changes for the drug targets were compared between the cell states for the cell lines combined and the cell lines individually. This comparison was made in a lollipop plot and scatterplots, as seen in Figure 4. The regulation of fold change of the drug targets in treated samples versus control is highly similar between the dying and surviving states in both cases. However, there are some minor differences. For instance, EGFR, the drug target for Gefitinib, has a negative $\log_2$ fold change for the MCF7 cell line in the dying state while having a slight positive $\log_2$ fold change in the surviving state. In the scatter plots, linear regressions are shown

together with the Pearson correlation coefficient between the $\log_2$ fold changes of the surviving and dying. The correlation was the largest for the combined cell line data and the RKO cell line.

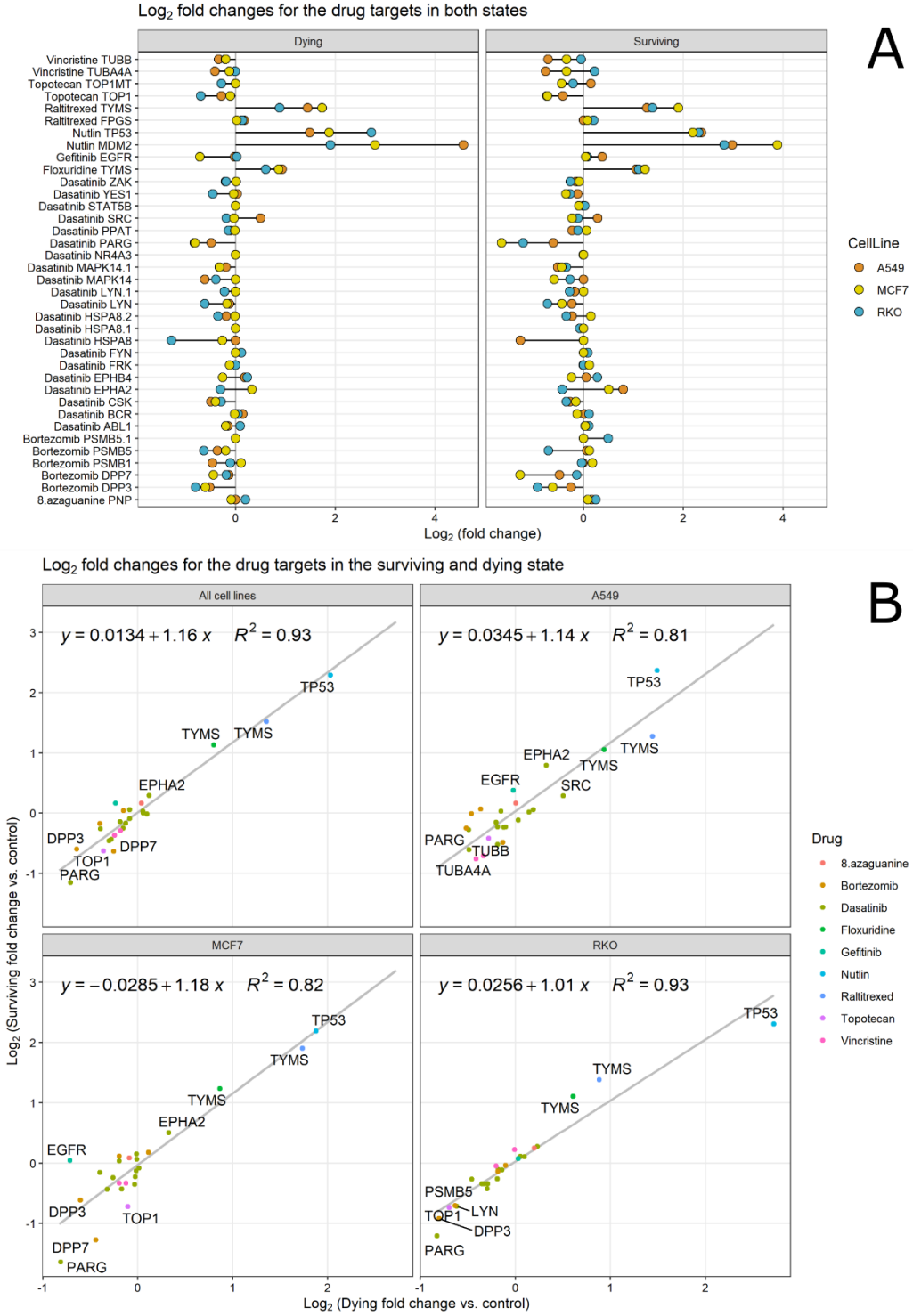


Figure 4. Log2 fold changes for the drug targets. A) Drug target $\log_2$ fold changes illustrated in a lollipop plot. On the y-axis, the drug is first specified, followed by the drug target. B) Log_2 fold changes illustrated in a scatterplot with the surviving $\log_2$ fold change on the y-axis and the dying $\log_2$ fold change on the x-axis for the combined response of all cell lines and the cell lines individually. A linear regression is made in all cases, and the Pearson correlation coefficient is displayed.

3.3 Target Rankings

The proteins without missing values were subjected to OPLS-DA analysis, contrasting the proteome of each treatment against those of all others. One of the resulting loading plots, contrasting Fluxoridine against all other drugs for the combined cell line and cell state data, is shown in Figure 5. In this figure, the drug target TYMS marked in maroon has the second-highest positive value in the predictive component (x-axis), and thus, its drug target ranking is two. All target rankings for the known drug targets can be seen in Figure 6 A. Only TYMS and PARG have high target rankings for the combined cell line and cell state data. The rest of the targets show relatively low rankings for the combined data.

The overall distribution of target rankings can be seen in Figure 6 B. For the RKO and A549 cells, the dying state has a lower median and less spread than the surviving state, meaning that dying cells generally provide better target rankings than surviving for these cell lines. For the combined cell line data, the dying cells have a higher median ranking but less spread than the surviving cells, while the dying cells of the MCF7 cell line display a higher median ranking and a similar amount of spread as the surviving cells indicating worse rankings. When it comes to the target rankings derived from the combined state data, RKO cells exhibit a general improvement with a low median ranking and the least amount of spread. For A549 and MCF7 cells, the combined state rankings display a similar median ranking to the surviving state while showing a similar spread to the dying state. Finally, for the combined cell line data, the combined state rankings have a higher mean than the surviving and dying states, but it has the lowest values for the 75th and 25th percentiles. Overall, the target rankings perform best for RKO cells when combining the cell state data.

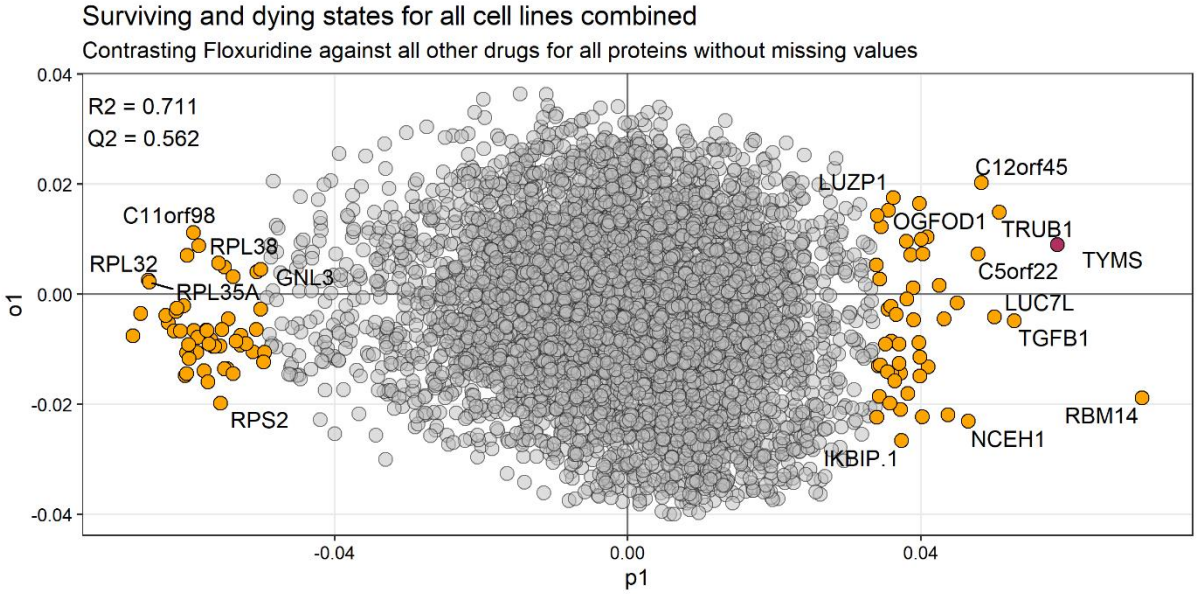


Figure 5. OPLS-DA model for Floxuridine, combining the data across all cell lines and cell states. The top proteins at each end of the predictive component (x-axis) are marked orange, and the drug target for Floxuridine, TYMS, is marked in maroon. The R2 and Q2 values of the OPLS-DA model are displayed in the top left corner.

A

		Combined States				Dying				Surviving			
		A549	Combined Cell Lines	MCF7	RKO	A549	Combined Cell Lines	MCF7	RKO	A549	Combined Cell Lines	MCF7	RKO
8.azaguanine	PNP	1804	2351	1376	1116	1149	2401	970	1409	964	2140	1301	1043
Bortezomib	DPP3	172	166	274	23	365	410	521	14	633	41	5	46
	DPP7	924	541	45	2240	1557	1045	405	2034	936	339	1	2470
	PSMB1	697	2446	1059	2422	271	1297	1333	2236	2813	1534	922	2579
	PSMB5	1296	1035	3073	118	743	500	2058	55	1683	1973	3419	222
Dasatinib	ABL1	1788	2330	1761	459	1734	1939	1738	172	2571	1765	1820	885
	BCR	1050	780	2063	104	959	850	2624	78	1744	631	2592	41
	CSK	590	215	268	12	548	175	183	29	172	231	692	25
	EPHA2	487	478	2667	211	1576	1240	2071	109	750	380	1879	413
	EPHB4	1194	743	2257	57	742	379	1080	24	2541	1282	1190	116
	HSPA8	108	123	2112	180	423	201	1164	161	503	256	2080	278
	LYN	2012	890	670	6	1725	401	187	11	1598	680	1412	12
	MAPK14	139	347	129	572	129	167	171	334	20	258	378	621
	PARG	15	16	1	22	36	33	25	79	11	8	4	7
	PPAT	692	643	2238	454	444	342	1816	800	1136	1124	2204	477
	SRC	91	934	999	277	127	529	1765	74	131	1949	158	674
	YES1	785	425	530	44	2040	642	1261	53	136	120	361	58
	ZAK	2328	1097	990	1463	2559	997	599	697	2648	1165	1367	2078
Floxuridine	TYMS	17	2	81	6	108	110	254	22	134	4	35	6
Gefitinib	EGFR	1560	1706	1180	182	520	2087	768	807	782	1182	1436	138
Nutlin	TP53	183	115	86	461	402	309	380	496	74	60	117	475
Raltitrexed	TYMS	5	5	3	26	29	21	41	53	8	1	28	13
Topotecan	TOP1	74	162	259	462	524	695	2006	535	72	42	21	595
Vincristine	TUBA4A	165	674	318	2291	348	925	425	2230	59	479	177	2426
	TUBB	91	753	740	1835	225	863	1344	792	41	322	63	2160

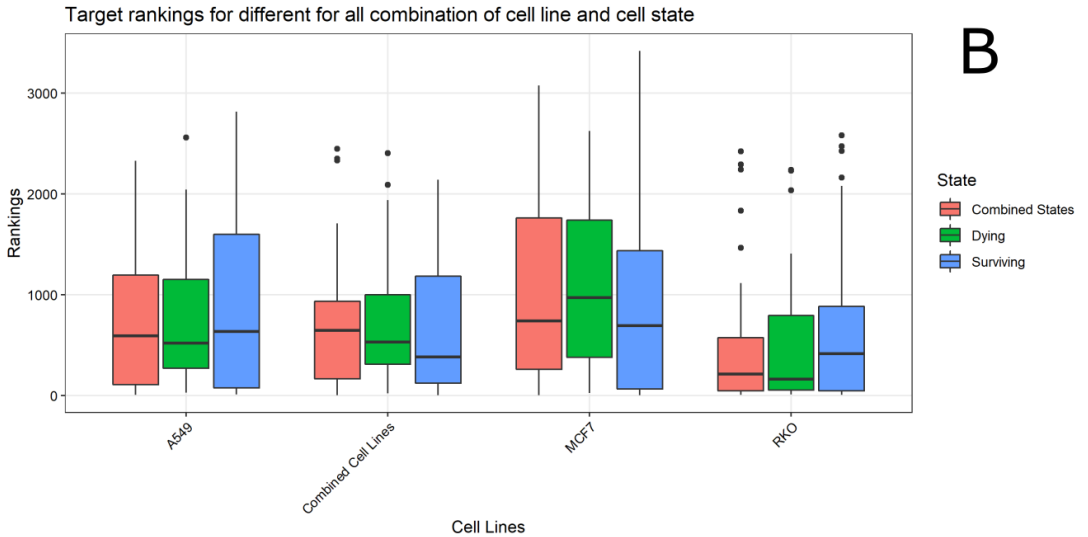


Figure 6. Target rankings. A) Exact target rankings for all combinations of cell lines and cell states. The left annotation shows the drugs and the drug targets, while the top annotation shows the cellular state and the cell lines. B) Boxplot showing the distribution of target rankings for all combinations of cell lines and cell states. The line in the middle of the box represents the median ranking. The upper and lower hinges correspond to the 75th and 25th percentile while the whiskers cover values from the hinges and extend no further than 1.5 times the Interquartile Range. Values falling outside the whiskers are marked with dots.

3.4 Comparing life and death

Several approaches were taken to compare the dying and surviving proteomes to find proteins and pathways that are always differentially regulated, especially regardless of the cell line or he anticancer agent used. In the following sections, the results from these analyses are presented.

3.4.1 Contrasting the proteome regulation in dying versus surviving cells

In Figure 7, a scatter plot comparing the $\log_2$ fold changes for the dying and surviving states is presented. In this plot, the mean $\log_2$ fold changes of all drugs in the given state is shown on the axes. Four clusters are highlighted for proteins that show up-or down-regulation in one state but none or opposite regulation in the other state. No pathways were found for the proteins upregulated in the dying state. However, *cornification* and *keratinization* pathways were associated with the proteins downregulated in the dying state. Proteins upregulated in the surviving state were associated with *extracellular structure organization* and *protein-lipid complex remodelling*, while proteins downregulated in the surviving state were associated with *positive regulation of serine-type endopeptidase activity*. Interestingly, some of these pathways are known to be involved in cell life and death processes.

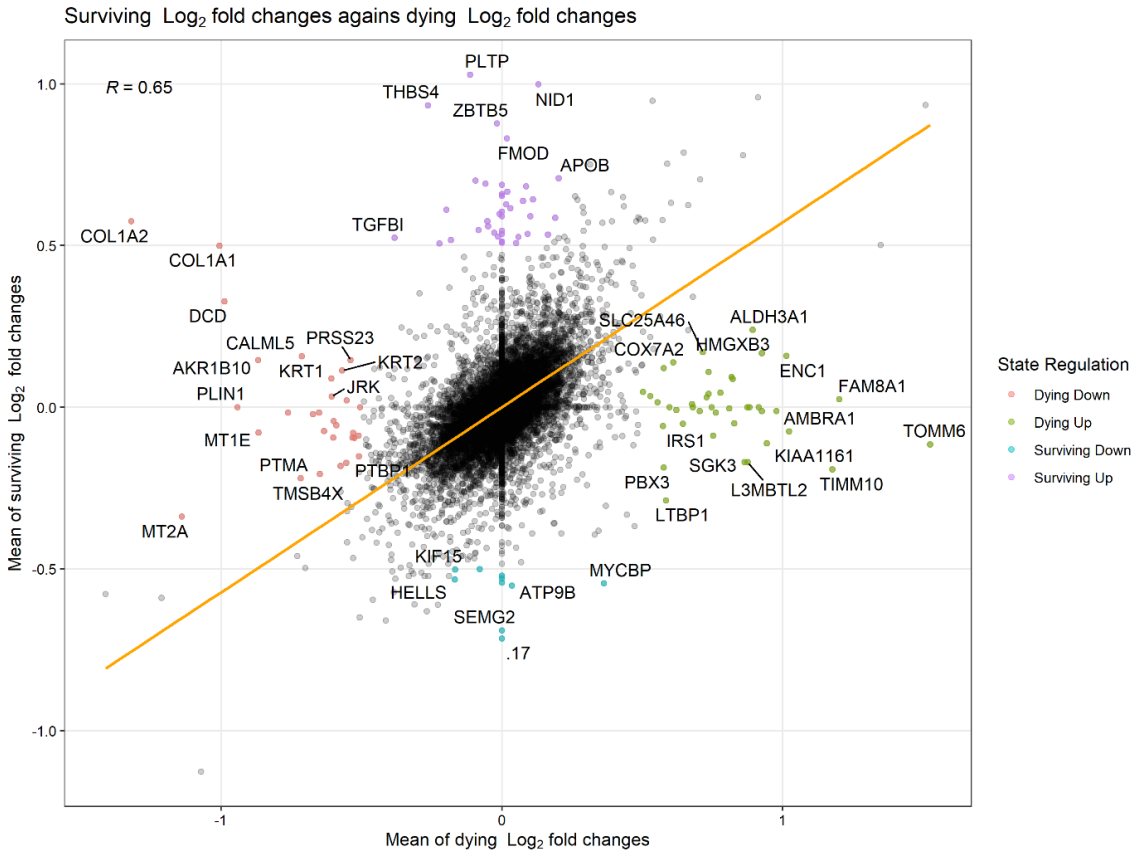


Figure 7. Scatterplot of the mean surviving $\log_2$ fold change against the mean dying $\log_2$ fold change of all the treatments and cell lines for the dataset containing all proteins regardless of the number of missing values. The clusters show up- or downregulation in one state while showing none or opposite regulation in the other. Proteins upregulated in the surviving state are marked as purple. Proteins downregulated in the surviving state are marked blue. Proteins

upregulated in the dying state are marked green. Proteins downregulated in the dying state are marked red. The log_2 fold change threshold for up- and down-regulation was set to 0.5 and -0.5, respectively. The orange line represents a linear regression between the log_2 fold changes, and the Pearson correlation is displayed in the top left corner. The regression and Pearson correlation are based on the dataset without missing values.

3.4.2 OPLS-DA modelling contrasting life and death

The loading plots obtained when contrasting surviving and dying cells through OPLS-DA modelling can be seen in Figure 8. In addition, the pathways associated with the top 50 proteins in the dying and surviving cells of each OPLS-DA model are presented in Table 2. For instance, these include *extracellular matrix organization* in the surviving state of the combined cell line data and A549 cells, *negative regulation of canonical Wnt signalling pathway* for the surviving cells of the MCF7 cell line, and *estrous cycle* for the dying cells of the RKO cell line.

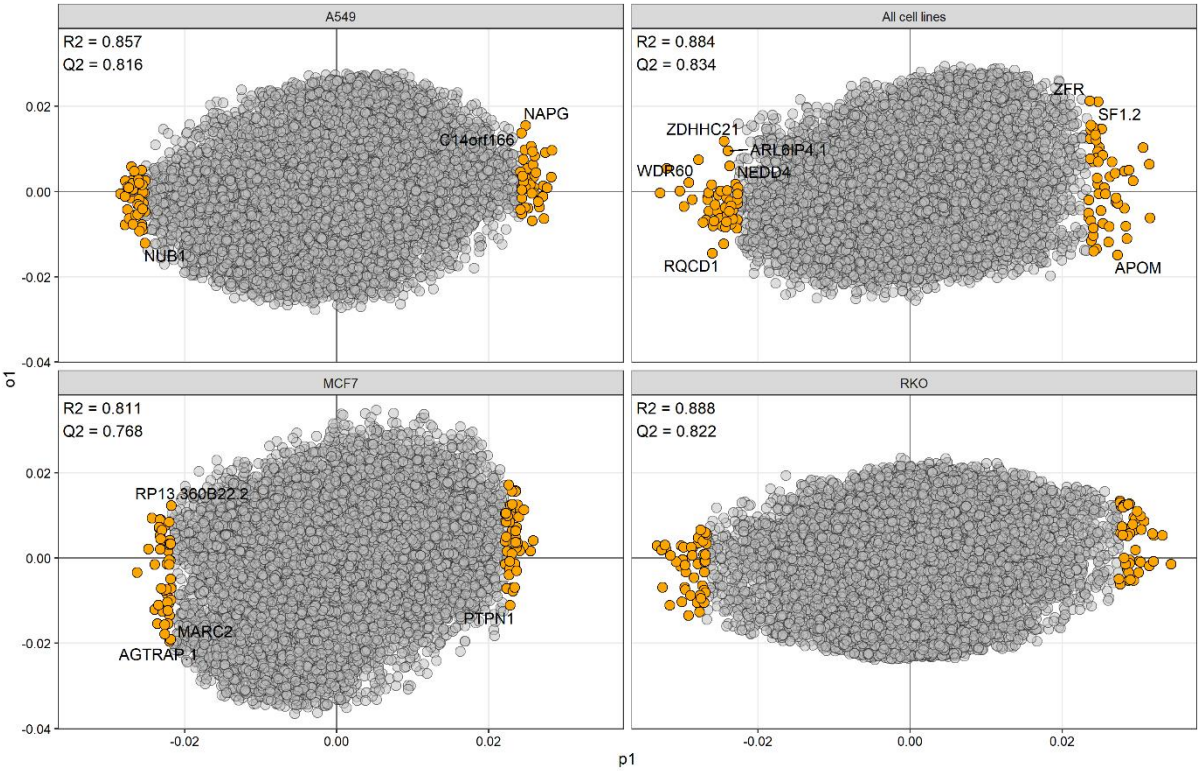


Figure 8. OPLS-DA models contrasting surviving and dying cells for individual cell lines and all cell lines combined. The circles with positive values in the predictive component (x-axis) correspond to proteins with higher regulation in the surviving state compared to the dying. Conversely, proteins with negative values in the predictive component have higher regulation in the dying state than in surviving. The top 0.5% proteins are marked in orange at each end of the predictive component, and R2 and Q2 values are displayed in the top left corner of each plot.

Table 2. Pathways for the top 50 proteins of the dying and surviving states for each OPLS-DA model presented in Figure 8.

Cell line	State	Pathways
All cell lines	Surviving	Extracellular structure and matrix organization Ossification
	Dying	Dynein heavy chain binding
A549	Surviving	Extracellular matrix organization
	Dying	Regulation of superoxide metabolic process Regulation of cellular response to stress
MCF7	Surviving	Negative regulation of canonical Wnt signalling pathway
	Dying	Protein C-linked glycosylation
RKO	Surviving	Sesquiterpenoid metabolic and catabolic processes Farnesol catabolic and metabolic processes
	Dying	Dichotomous subdivision of terminal units involved in salivary gland branching Estrous cycle

3.4.3 Comparing expression for proteins between life and death

Barplots for the proteins COL1A2, MYCBP, PRSS23, SSNA1, THBS4, and USP4 can be seen in Figure 9. These proteins stand out on the scatterplot and OPLS-DA models comparing the cellular states, and therefore, can serve as cell death or survival proteins. As displayed, COL1A2, PRSS23, and THBS4 show a consistently higher regulation in the surviving cells than in the dying ones. On the other hand, MYCBP, SSNA1, and USP4 show a consistently higher regulation in the dying cells than in surviving ones. Also, similar bar plots of the proteins that were found to be differentially regulated between the two states in the previous study (Saei *et al.* 2018) can be seen in Figure 10. Overall, the inclusion of a higher number of diverse drugs can lead to the discovery of more general cell death and survival markers than what was observed before (Saei *et al.* 2018). In both figures, the presence of missing values can be observed. In Figure 9, $\log_2$ fold changes are not present for MYCBP of dying A549 cells, and in Figure 10, most $\log_2$ fold changes are missing for RNF40.

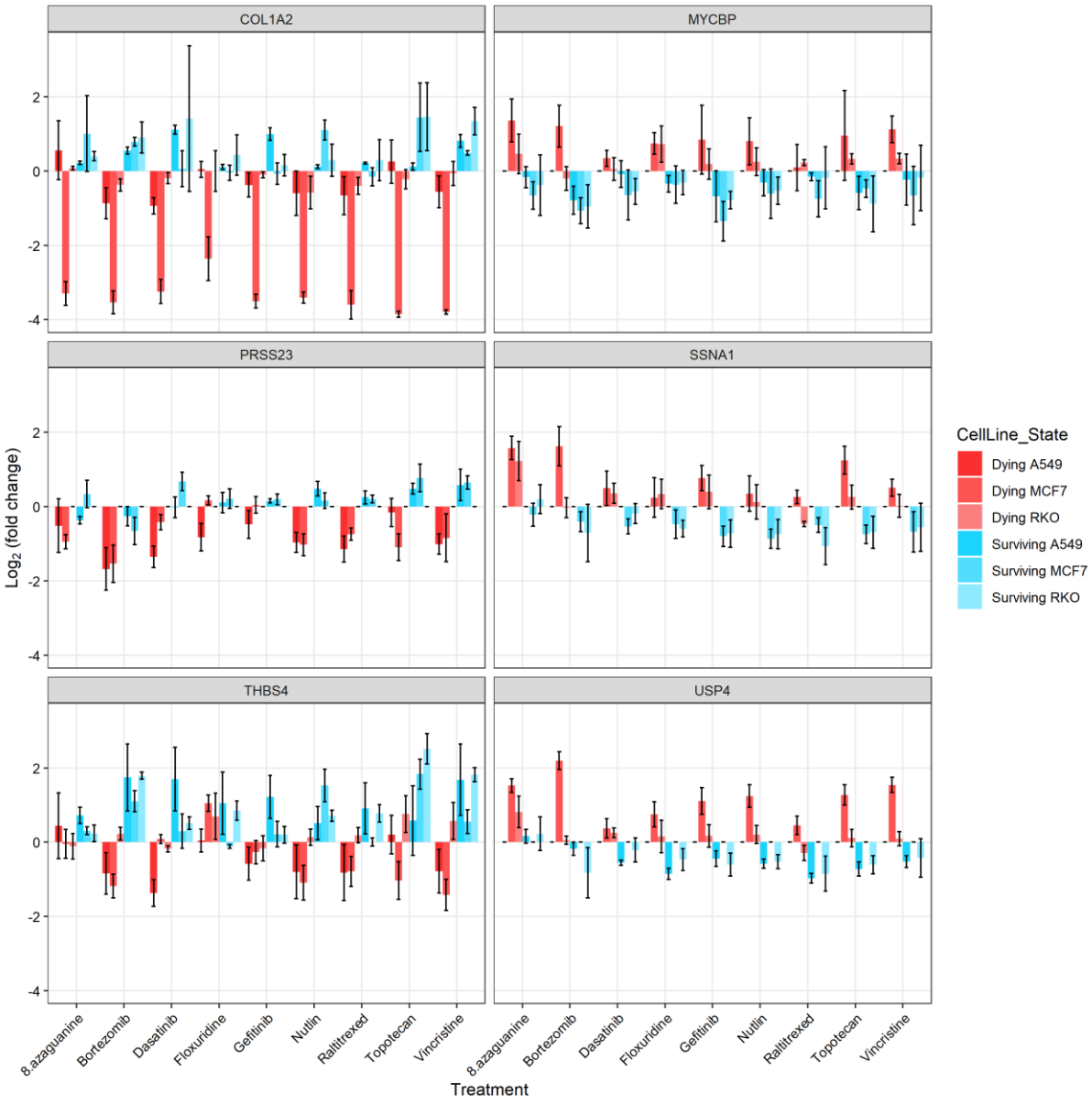


Figure 9. Log_2 fold changes for proteins standing out in the OPLS-DA models and scatter plot, showing the log_2 fold changes across all drugs and both states. The regulation in the dying state is marked in red with different shades for the different cell lines, while the regulation in the surviving state is marked in blue with different shades for the cell lines. The error bar shows the standard deviation of the replicates in each case.

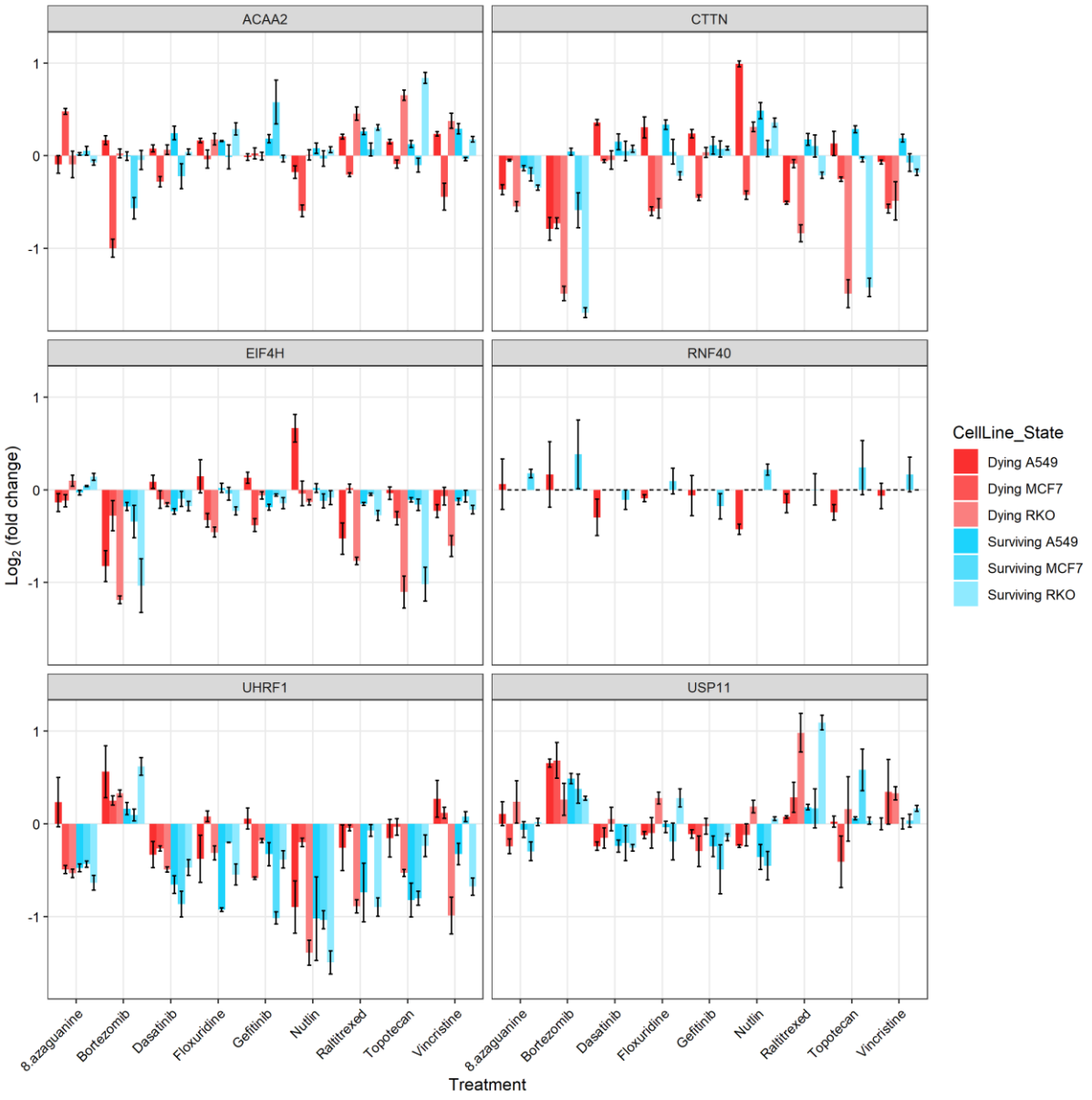


Figure 10. Log_2 fold changes for proteins found to be differentially regulated between dying and surviving cells in the previous study (Saei *et al.* 2018). Dying log_2 fold changes are shown in red and surviving log_2 fold changes are shown in blue. The different cell lines are marked in different shades of the two colours. The error bar shows the standard deviation of the replicates in each case.

3.4.4 Heatmaps

Heatmaps were made as a final approach to compare the dying and surviving cells. In Figure 11 A, a heatmap of all cell lines can be seen, in which there are some clear distinctions between the cell lines and states. The clusters in this figure are ordered based on the p-value comparing the dying and surviving states. In Figure 11 B, the top three Gene Ontology (GO) processes for each cluster can be seen along with their enrichment and false discovery rate. For example, cluster 3 showed a clear difference between the dying and surviving states of MCF7 and was found to be involved in the *regulation of the cell cycle* and the *regulation of tyrosine*

phosphorylation of STAT protein. The latter is a process necessary for establishing the JAK-STAT signalling pathway (Egger *et al.* 2003) involved in cellular processes such as cell division and cell death (Hu *et al.* 2021). Further, the proteins in cluster 5 were involved in *biological adhesion* and *mesenchyme development*, and the proteins in cluster 1 were involved in *rRNA processing*.

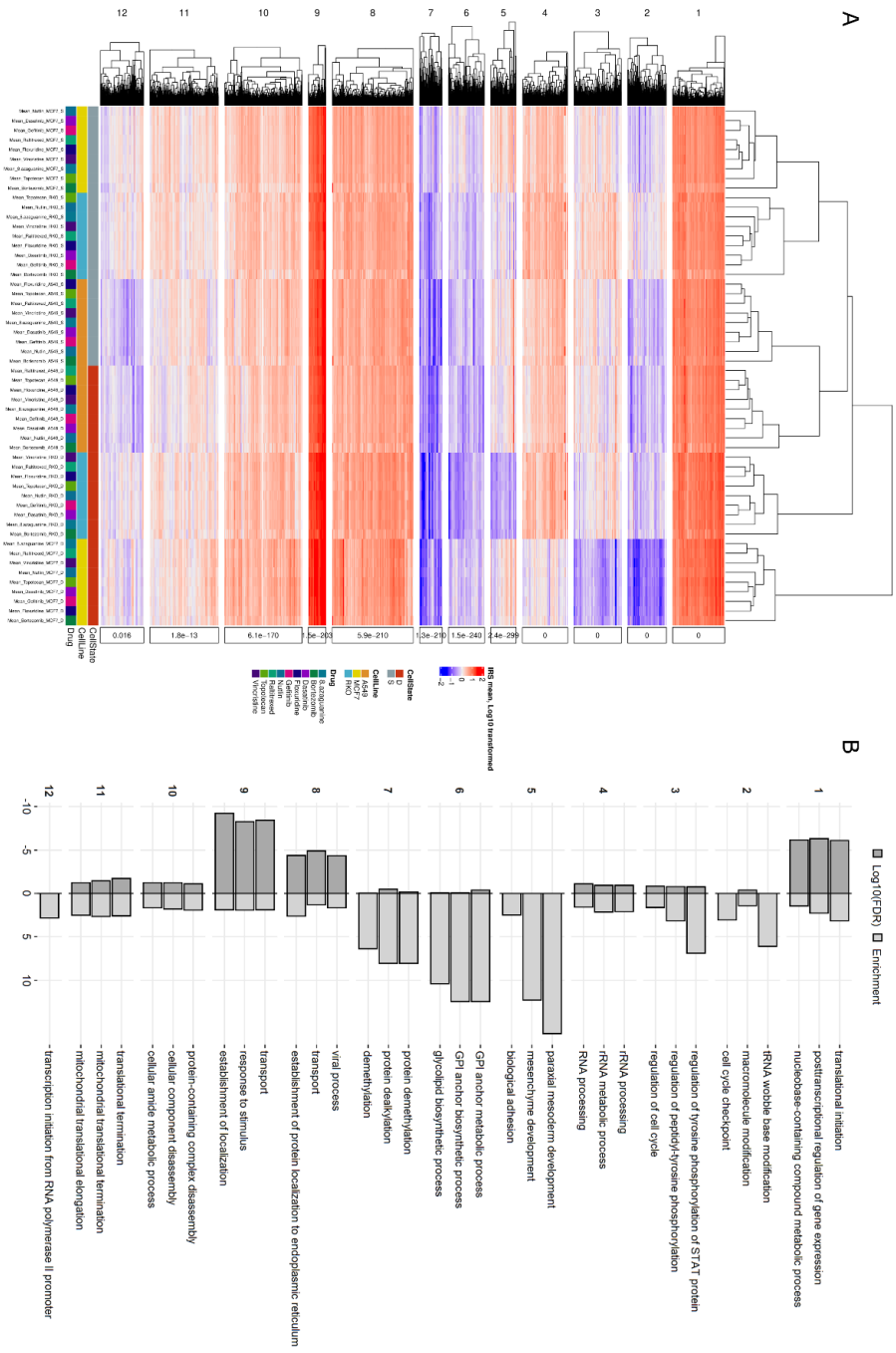


Figure 11. A) Heatmap of all cell lines using log_{10} transformed IRS normalized values. The column annotation at the bottom marks the two cellular states, the three cell lines, and the nine anticancer drugs. The clusters are ordered by p-value comparing each cluster's living and dying states. B) The top three associated GO processes obtained from Gorilla for each cluster, showing their Enrichment and False discovery rate (FDR). For cluster 12, only one GO process was found.

4 Discussion

This master thesis project compares the proteome response between dying and surviving cells treated with different anticancer compounds. The cellular states are investigated individually and together to devise the best approach in target identification. Further, the proteome response is compared between dying and surviving cells to identify proteins and pathways differentially regulated between the two states. These would contribute to cell survival, drug resistance, and cell death. The results concerning these aims are discussed in the below sections.

4.1 Dying has the most significant impact on the cellular proteome

Similar to the previous study comparing the proteome of dying and surviving cells (Saei *et al.* 2018), this study also shows that cell death has the most significant impact on the cellular proteome rather than the treatment or type of cell line used. The separation between life and death is evident in the first principal component of all PCA plots, especially for individual cell lines.

An additional way to determine what feature has the highest impact on the cellular proteome would have been to calculate distances in the multidimensional PCA space. In this case, distances between cell lines, treatments, and cell states would have been calculated to determine which feature shows the highest separation in the PCA space. However, due to the clear separation between surviving and dying cells in the first principal component of each PCA plot, this way of determining the feature with the highest impact on the cellular proteome was deemed unnecessary. The first principal component in a PCA plot shows the highest variance across the dataset, and thus it is clear that cell death has the most significant impact on the cellular proteome.

4.2 Dying and surviving cells have similar drug target regulation

The known drug targets of the anticancer compounds had a similar regulation between the dying and surviving states. There were some noticeable differences when comparing the states for individual cell lines, but when combining the cell lines, the $\log_2$ fold changes were the most consistent between the states. This can also be noticed when comparing the Pearson correlation coefficients, for which the highest correlation was found for the combined cell line data. The consistency between the states shows that dying and surviving cells have similar drug target regulations. However, it would not have been surprising if there were more differences, given that two different states with different effects on the cellular proteome are compared in this study.

Only the known drug targets that did not have any missing values across all samples were included in this comparison. If drug targets with missing values were included, missing value imputation would have been necessary, giving rise to unwanted bias. If more data was available for the drug targets with missing values and, generally, data from more cell lines and anticancer

treatments, the conclusion that the surviving and dying cells have similar regulation of drug targets would have been more general.

4.3 Combining data from dying and surviving cells can improve drug target identification

Some of the known targets for the compounds in the study showed good rankings when combining all data, e.g. TYMS as the known target for Raltitrexed and Floxuridine, and PARG for Dasatinib. These targets are among the ones with the greatest $\log_2$ fold change in response to drug treatment, as seen in Figure 4. Only TP53 have a greater $\log_2$ fold change, but at the same time, TP53 has high $\log_2$ fold changes across most drug treatments, which can be seen in Appendix A. This is due to the involvement of TP53 in various cell death processes. Thus, it seems easier to identify drug targets for drug treatments having a significant effect on the expression of the drug target using this method. It should also be noted that some of the drug targets displayed high variations in $\log_2$ fold changes within the biological replicates, which could affect the rankings.

In the previous study (Saei *et al.* 2018), it was found that combining the cell line data for both states could improve drug target rankings. There are cases where this is true in this study, for instance, for TYMS in Floxuridine, but it does not seem to happen to the same extent. It is difficult to say what this might depend on. It could be affected by the different drugs and cell lines used in this study compared to the previous one. If some of them have similar effects on the cellular proteome, it could be challenging to contrast them against each other to obtain target rankings. On the other hand, this study includes more drug treatments, which according to the results of the ProTargetMiner paper (Saei *et al.* 2019), should provide better target rankings. Another reasonable explanation could be the high variation in $\log_2$ fold changes for biological replicates for some of the targets. Since the OPLS-DA models take all replicates into account when discriminating between the drug treatments, a high variation between replicates would make the rankings worse.

4.4 Comparing the proteome in living and dying cells

When combining cell line data and analyzing individual cell lines, two similar pathways found for proteins with higher expression in the surviving cells than in the dying are *extracellular structure organization* and *extracellular matrix organization*. Cells only grow and differentiate when they are in the correct location, which they sense through interactions with neighbouring cells and the extracellular matrix (Gilmore 2005). When cells lose attachment to the extracellular matrix, they undergo a special form of apoptosis called anoikis (Galluzzi *et al.* 2018). For this reason, it is not surprising that the surviving cells show an increase in proteins responsible for maintaining the extracellular structures to increase their chances of survival.

Ossification, or bone formation, is a pathway found for proteins having higher regulation in surviving cells when combining the cell line data. Collagens are a primary constituent in bones,

and several are found upregulated in the surviving cells compared to the dying. Furthermore, collagens provide structural support to the extracellular space (Wu *et al.* 2022). As mentioned before, cells need to be connected to the extracellular matrix to remain alive. Therefore, if they produce collagens that support the extracellular space, they would have a greater chance of remaining attached and therefore remain alive.

Another finding in the combined response from all cell lines was that proteins involved in dynein heavy chain binding were found at higher levels in dying cells than in surviving. Dyneins are involved in several cellular processes, for instance, mitotic spindle organization and chromosome separation during mitosis (Vallee *et al.* 2004). Since dead cells do not proliferate, it does not seem surprising that dynein interactions differentiate the dying and surviving cells.

The *sesquiterpenoid and farnesol catabolic and metabolic processes* were more prevalent in surviving than dying cells of the RKO cell line. One sesquiterpenoid, Bigelovin, inhibits cell proliferation and induces apoptosis and autophagy by inhibiting the mTOR pathway (Wang *et al.* 2018). The mTOR pathway regulates cell proliferation, apoptosis, and autophagy (Zou *et al.* 2020) by regulating the production of reactive oxygen species (Wang *et al.* 2018). Like sesquiterpenoids, farnesol is an inducer of cell cycle arrest and apoptosis, which have also been found to inhibit tumorigenesis in animals (Joo & Jetten 2010). Therefore, the catabolic processes of sesquiterpenoids and farnesol in surviving cells prevent apoptosis and autophagy, keeping the cells alive.

The pathway *regulation of superoxide metabolic process* was found in dying A549 cells. Superoxide is produced in the mitochondria of apoptotic cells (Cai & Jones 1998), which is a possible explanation for why this pathway marks a difference between the dying and surviving cells.

One surprising finding is the pathway *negative regulation of canonical Wnt signalling* found in surviving MCF7 cells. Wnt signalling regulates the early and late stages of apoptosis (Pećina-Šlaus 2010). In addition, an active Wnt signalling increases cells' growth properties and has been associated with therapy resistance (Bugter *et al.* 2021). Further, suppression of Wnt signalling prevents inappropriate proliferation (Sampson *et al.* 2001). For these reasons, it is surprising that the Wnt signalling pathway is found to be negatively regulated in the surviving cells and not positively regulated. If found positively regulated, this pathway could have been linked to surviving cells' efforts to resist the drugs and continue to grow. One hypothesis for this observation could be that it is an effect induced by the anticancer compounds to kill the cells, but then that gives rise to questions about why not a similar effect is found in the dying cells.

When cancer cells detach from the cellular matrix, it is not necessarily marking the end of their life cycle. Detached cancer cells can recover and grow if recultured in fresh media (Saei *et al.* 2018), and at the beginning of metastasis cancer cells detach from the primary tumour (Fujii *et*

al. 2021). The keratinization and cornification pathways found downregulated for dying cells in the scatter plot (Figure 7) bears the hypothesis if the dying or detaching cancer cells can migrate more and colonize other tissues. Cornification is a particular type of cell death occurring in the epidermis, dependent on the high keratin levels produced through keratinization. In a study by Seltmann *et al.* (2013), it was found that the downregulation of keratins in keratinocytes directly contributes to the tumour cells' migratory and invasive behaviour by affecting cell stiffness. Therefore, as a final attempt to survive, the detached cells might downregulate their keratin expression to increase their chances of relocating elsewhere.

Proteases, mainly caspases, play an essential role in cell death by breaking down the cells (Galluzzi *et al.* 2018). In addition, serine proteases are also shown to mediate cell death in the absence of caspases by perforating the mitochondrial membrane and thus releasing cytochrome c into the cell (Egger *et al.* 2003). When cytochrome c is released in the cell, caspases are activated, leading to subsequent cell death (Jiang & Wang 2004), which explains why the pathway *positive regulation of serine-type endopeptidase activity* is found downregulated in surviving cells, as seen in Figure 7.

Three proteins with consistently higher $\log_2$ fold changes in the surviving state than the dying are COL1A2, PRSS23, and THBS4. COL1A2 has, on the one hand, been found to act as a tumour suppressor for colorectal cancer (Yu *et al.* 2018 p. 2). On the other hand, it induces cancer cell proliferation, migration, and metastasis (Xu *et al.* 2019). By the results of this study, it does not seem like COL1A2 would act as a tumour suppressor in colorectal cancer since it is upregulated in the surviving RKO cells. In this case, the more probable explanation seems to be that it induces cell proliferation, migration, and metastasis. The second protein, PRSS23, is an interesting one. Studies have shown that knockdown of this protein inhibits gastric cancer tumorigenesis, and therefore it has been suggested as a potential target for gastric cancer treatment (Han *et al.* 2019). This might also apply to the A549 and MCF7 cell lines for which PRSS23 is consistently upregulated in the surviving state while being downregulated in the dying state. If the knockdown causes the inhibition of cancer tumorigenesis, it is reasonable for this protein to be downregulated in the dying cells. For RKO cells, PRSS23 was not detected in the experiment, and unfortunately, no conclusions can be drawn for RKO cells in this case. Finally, the function of the third protein THBS4, according to UniProt (The UniProt Consortium 2021), is cell proliferation, adhesion, and attachment, which should be processes essential for the survival of cancer cells. These functions align with the protein being upregulated in surviving cells since these are apparent processes for surviving cells.

Three proteins with consistently higher $\log_2$ fold changes in the dying state than the surviving are MYCBP, SSNA1, and USP4. According to UniProt (The UniProt Consortium 2021), MYCBP may control the transcriptional regulation of the protein MYC, which activates the transcription of growth-related genes. This function potentially describes why this protein would have different expression levels in dying than surviving cells. SSNA1 is a protein that slows down the cellular growth rate, shrinkage, and catastrophe, explaining why higher levels

would be observed in dying cells than in surviving. Finally, USP4 is a deubiquitinating enzyme that can remove ubiquitin from target proteins (The UniProt Consortium 2021). It inhibits the activation of the NF-κB signalling pathway (Fan *et al.* 2011), which gives cancer cells a survival advantage by upregulating anti-apoptotic genes (Verzella *et al.* 2020). Since USP4 is downregulated in surviving cells, it means that NF-κB signalling is active and upregulates anti-apoptotic genes in the surviving cells. The contrary applies to dying cells.

The survival and death markers found in the current study with nine drugs, were stronger and showed a more consistent behaviour between dying and surviving states than those identified in the previous study with three drugs (Saei *et al.* 2018). This observation is probably because different cell lines and drug treatments were used in this study compared to the previous. Furthermore, the drugs used in the current study were selected from different principal components of a dataset with 55 drugs and thus represent diverse targets, mechanisms of action, and cell death pathways.

One concern with the comparison between Dying and Surviving cells is that the dying A549 cells only were analyzed in 16 fractions, compared to the other cells which were analyzed in 24 fractions (as mentioned in section 1.5 Experimental procedures). This might introduce some bias in the analysis since fewer proteins would be able to be identified in the dying A549 cells in general compared to the others. To avoid this, the analysis could have been performed only on the RKO and MCF7 cells, or the proteins with missing values could have been excluded. However, the proteins and pathways found in this comparison are still relevant for life and death processes.

Out of the methods used for comparing Dying and Surviving cells, OPLS-DA modelling should give the most reliable results. While scatterplots of Dying and Surviving cells can give additional information in regards to what proteins are up or down-regulated, they can be heavily influenced by the log2 fold changes of one sample since they only display mean values. If one sample has extremely high or low log2 fold change, that could have a significant impact on the mean log2 fold change. On the other hand, OPLS-DA modelling takes reproducibility into account when producing the models, both between different samples and within biological replicates. This means that the proteins identified in the OPLS-DA models would show consistent differences between the two states and therefore, this approach should produce more reliable results.

5 Conclusion

This study shows that cell death is the feature contributing most to the separation of the proteome signatures rather than the cell line or treatments. We show that the regulation of drug targets is highly similar in dying versus surviving cells. Further, both matrix attached and detached cells are helpful for drug target identification, and combining data from both states

can improve drug target rankings. However, the $\log_2$ fold change of the drug targets and the variation within the biological replicates seem to affect the final ranking. Thus, the target identification is highly dependent on the data quality. A drawback with this method of identifying drug targets is that the drug target needs to be detected by the mass spectrometer. Also, outside the scope of cancer drug target identification, identifying drug targets in dying cells is relatively limited, considering the treatment would need to exert cytotoxicity to the cells.

Most pathways found when comparing the proteome between the dying and surviving cells can be expected based on our current knowledge of cell death and survival pathways. In addition, at least six proteins showing differential regulation between the two cell states regardless of cell line and treatment were identified. Some of these proteins seem to have potential as drug targets for future anticancer treatments, and thus, these would be interesting to investigate further.

6 Acknowledgements

I would like to thank my supervisors, Dr Amir Ata Saei and Professor Roman Zubarev, for providing me with such an exciting project. Amir, thank you for your continuous support throughout this project and for always taking the time to answer my questions. I could not have done this without you, and I could not have asked for a better supervisor. Roman, thank you for your feedback and input on this project.

I would also like to thank Hassan Gharibi for sharing R wisdom and statistical expertise with me, and my subject reader, Professor Jonas Bergquist, for support and input throughout this project.

Finally, I want to thank my friend and student opponent, Linnéa Yuan Andersson. It has been truly great to have a fellow student to brainstorm idéas with and to be able to help each other out whenever necessary.

7 References

Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. Nature 422: 198–207.

Bergman O, Johansson E. 2018. Cancer i siffror 2018: popul??rvetenskapliga fakta om cancer.

Bugter JM, Fenderico N, Maurice MM. 2021. Mutations and mechanisms of WNT pathway tumour suppressors in cancer. Nature Reviews Cancer 21: 5–21.

Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. 2006. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. Journal of Chemometrics 20: 341–351.

Cai J, Jones DP. 1998. Superoxide in Apoptosis: MITOCHONDRIAL GENERATION TRIGGERED BY CYTOCHROMEc LOSS *. Journal of Biological Chemistry 273: 11401–11404.

Chernobrovkin A, Marin-Vicente C, Visa N, Zubarev RA. 2015. Functional Identification of Target by Expression Proteomics (FITExP) reveals protein targets and highlights mechanisms of action of small molecule drugs. Scientific Reports 5: 11176.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10: 48.

Egger L, Schneider J, Rhême C, Tapernoux M, Häcki J, Borner C. 2003. Serine proteases mediate apoptosis-like cell death and phagocytosis under caspase-inhibiting conditions. Cell Death & Differentiation 10: 1188–1203.

Fan Y-H, Yu Y, Mao R-F, Tan X-J, Xu G-F, Zhang H, Lu X-B, Fu S-B, Yang J. 2011. USP4 targets TAK1 to downregulate TNFα-induced NF-κB activation. Cell Death & Differentiation 18: 1547–1560.

Fujii T, Shimizu T, Katoh M, Nagamori S, Koizumi K, Fukuoka J, Tabuchi Y, Sawaguchi A, Okumura T, Shibuya K, Fujii T, Takeshima H, Sakai H. 2021. Survival of detached cancer cells is regulated by movement of intracellular Na+,K+-ATPase. iScience 24: 102412.

Galluzzi L, Vitale I, Aaronson SA, Abrams JM, Adam D, Agostinis P, Alnemri ES, Altucci L, Amelio I, Andrews DW, Annicchiarico-Petruzzelli M, Antonov AV, Arama E, Baehrecke EH, Barlev NA, Bazan NG, Bernassola F, Bertrand MJM, Bianchi K, Blagosklonny MV, Blomgren K, Borner C, Boya P, Brenner C, Campanella M, Candi E, Carmona-Gutierrez D, Cecconi F, Chan FK-M, Chandel NS, Cheng EH, Chipuk JE, Cidlowski JA, Ciechanover A, Cohen GM, Conrad M, Cubillos-Ruiz JR, Czabotar PE, D'Angiolella V, Dawson TM, Dawson VL, De Laurenzi V, De Maria R, Debatin K-M, DeBerardinis RJ, Deshmukh M, Di Daniele N, Di Virgilio F, Dixit VM, Dixon SJ, Duckett CS, Dynlacht BD, El-Deiry WS, Elrod JW, Fimia GM, Fulda S, García-Sáez AJ, Garg AD, Garrido C, Gavathiotis E, Golstein P, Gottlieb E, Green DR, Greene LA, Gronemeyer H, Gross A, Hajnoczky G, Hardwick JM, Harris IS,

Hengartner MO, Hetz C, Ichijo H, Jäättelä M, Joseph B, Jost PJ, Juin PP, Kaiser WJ, Karin M, Kaufmann T, Kepp O, Kimchi A, Kitsis RN, Klionsky DJ, Knight RA, Kumar S, Lee SW, Lemasters JJ, Levine B, Linkermann A, Lipton SA, Lockshin RA, López-Otín C, Lowe SW, Luedde T, Lugli E, MacFarlane M, Madeo F, Malewicz M, Malorni W, Manic G, Marine J-C, Martin SJ, Martinou J-C, Medema JP, Mehlen P, Meier P, Melino S, Miao EA, Molkentin JD, Moll UM, Muñoz-Pinedo C, Nagata S, Nuñez G, Oberst A, Oren M, Overholtzer M, Pagano M, Panaretakis T, Pasparakis M, Penninger JM, Pereira DM, Pervaiz S, Peter ME, Piacentini M, Pinton P, Prehn JHM, Puthalakath H, Rabinovich GA, Rehm M, Rizzuto R, Rodrigues CMP, Rubinsztein DC, Rudel T, Ryan KM, Sayan E, Scorrano L, Shao F, Shi Y, Silke J, Simon H-U, Sistigu A, Stockwell BR, Strasser A, Szabadkai G, Tait SWG, Tang D, Tavernarakis N, Thorburn A, Tsujimoto Y, Turk B, Vanden Berghe T, Vandenabeele P, Vander Heiden MG, Villunger A, Virgin HW, Vousden KH, Vucic D, Wagner EF, Walczak H, Wallach D, Wang Y, Wells JA, Wood W, Yuan J, Zakeri Z, Zhivotovsky B, Zitvogel L, Melino G, Kroemer G. 2018. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. Cell Death & Differentiation 25: 486–541.

Gilmore AP. 2005. Anoikis. Cell Death & Differentiation 12: 1473–1477.

Han B, Yang Y, Chen J, He X, Lv N, Yan R. 2019. PRSS23 knockdown inhibits gastric tumorigenesis through EIF2 signaling. Pharmacological Research 142: 50–57.

Hu X, Li J, Fu M, Zhao X, Wang W. 2021. The JAK/STAT signaling pathway: from bench to clinic. Signal Transduction and Targeted Therapy 6: 1–33.

Jiang X, Wang X. 2004. Cytochrome C-mediated apoptosis. Annual Review of Biochemistry 73: 87–106.

Joo JH, Jetten AM. 2010. Molecular Mechanisms involved in Farnesol-Induced Apoptosis. Cancer letters 287: 123.

Lansdowne LE. 2018. Target Identification & Validation in Drug Discovery. WWW document 29 November 2018: http://www.technologynetworks.com/drug-discovery/articles/target-identification-validation-in-drug-discovery-312290. Accessed 15 May 2022.

Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M. 2017. Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nature Reviews Drug Discovery 16: 531–543.

Pećina-Šlaus N. 2010. Wnt signal transduction pathway and apoptosis: a review. Cancer Cell International 10: 22.

Plubell DL, Wilmarth PA, Zhao Y, Fenton AM, Minnier J, Reddy AP, Klimek J, Yang X, David LL, Pamir N. 2017. Extended Multiplexing of Tandem Mass Tags (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse Epididymal Adipose Tissue. Molecular & cellular proteomics: MCP 16: 873–890.

Pray LA. 2008. Eukaryotic Genome Complexity | Learn Science at Scitable. WWW document 2008: http://www.nature.com/scitable/topicpage/eukaryotic-genome-complexity-437. Accessed 25 May 2022.

Saei AA, Beusch CM, Chernobrovkin A, Sabatier P, Zhang B, Tokat ÜG, Stergiou E, Gaetani M, Végvári Á, Zubarev RA. 2019. ProTargetMiner as a proteome signature library of anticancer molecules for functional discovery. Nature Communications 10: 5715.

Saei AA, Sabatier P, Tokat ÜG, Chernobrovkin A, Pirmoradian M, Zubarev RA. 2018. Comparative Proteomics of Dying and Surviving Cancer Cells Improves the Identification of Drug Targets and Sheds Light on Cell Life/Death Decisions. Molecular & cellular proteomics: MCP 17: 1144–1155.

Sampson EM, Haque ZK, Ku M-C, Tevosian SG, Albanese C, Pestell RG, Paulson KE, Yee AS. 2001. Negative regulation of the Wnt–β-catenin pathway by the transcriptional repressor HBP1. The EMBO Journal 20: 4500–4511.

Seltmann K, Fritsch AW, Käs JA, Magin TM. 2013. Keratins significantly contribute to cell stiffness and impact invasive behavior. Proceedings of the National Academy of Sciences 110: 18507–18512.

Steen H, Mann M. 2004. The abc's (and xyz's) of peptide sequencing. Nature Reviews Molecular Cell Biology 5: 699–711.

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research 49: D480–D489.

Thermo Fisher Scientific. TMT Quantitation - SE. WWW document: //www.thermofisher.com/uk/en/home/industrial/mass-spectrometry/proteomics-mass-spectrometry/quantitative-proteomics-mass-spectrometry/tmt-quantitation.html. Accessed 8 May 2022.

Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. 2015. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. Journal of Proteome Research 14: 3322–3335.

Vallee RB, Williams JC, Varma D, Barnhart LE. 2004. Dynein: An ancient motor protein involved in multiple modes of transport. Journal of Neurobiology 58: 189–200.

Verzella D, Pescatore A, Capece D, Vecchiotti D, Ursini MV, Franzoso G, Alesse E, Zazzeroni F. 2020. Life, death, and autophagy in cancer: NF-κB turns up everywhere. Cell Death & Disease 11: 1–14.

Wang B, Zhou T-Y, Nie C-H, Wan D-L, Zheng S-S. 2018. Bigelovin, a sesquiterpene lactone, suppresses tumor growth through inducing apoptosis and autophagy via the inhibition of mTOR pathway regulated by ROS generation in liver cancer. Biochemical and Biophysical Research Communications 499: 156–163.

WHO. 2022. Cancer. WWW document 3 February 2022: https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed 25 May 2022.

Wu M, Cronin K, Crane JS. 2022. Biochemistry, Collagen Synthesis. StatPearls

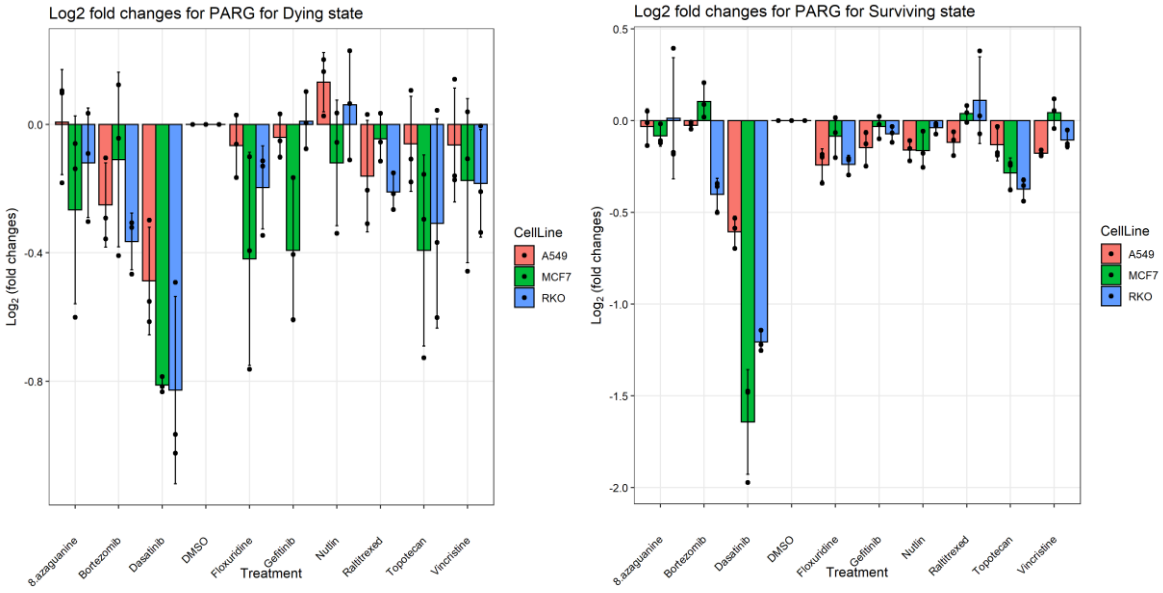
Xu S, Xu H, Wang W, Li S, Li H, Li T, Zhang W, Yu X, Liu L. 2019. The role of collagen in cancer: from bench to bedside. Journal of Translational Medicine 17: 309.

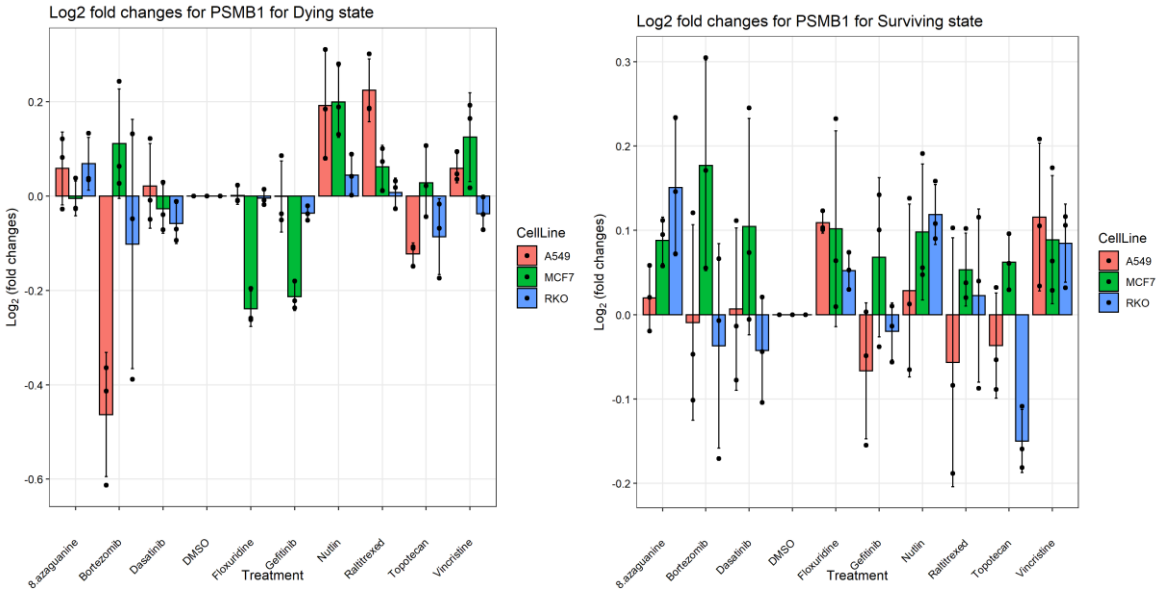
Yu Y, Liu D, Liu Z, Li S, Ge Y, Sun W, Liu B. 2018. The inhibitory effects of COL1A2 on colorectal cancer cell proliferation, migration, and invasion. Journal of Cancer 9: 2953–2962.

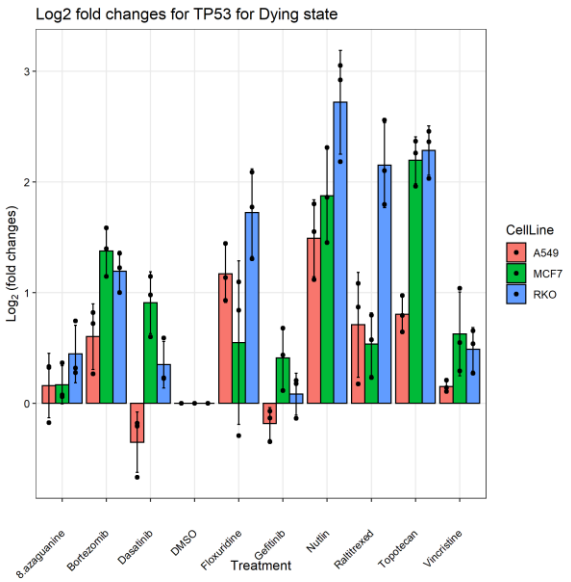
Zou Z, Tao T, Li H, Zhu X. 2020. mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. Cell & Bioscience 10: 31.

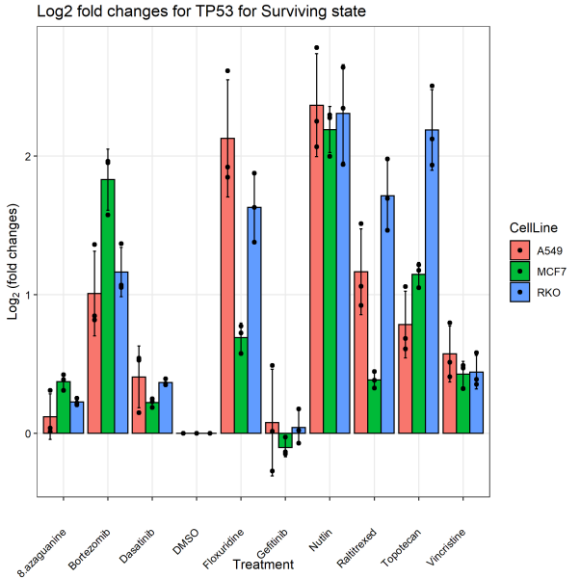
8 Appendix A – Log_2 fold changes for the drug targets across each drug

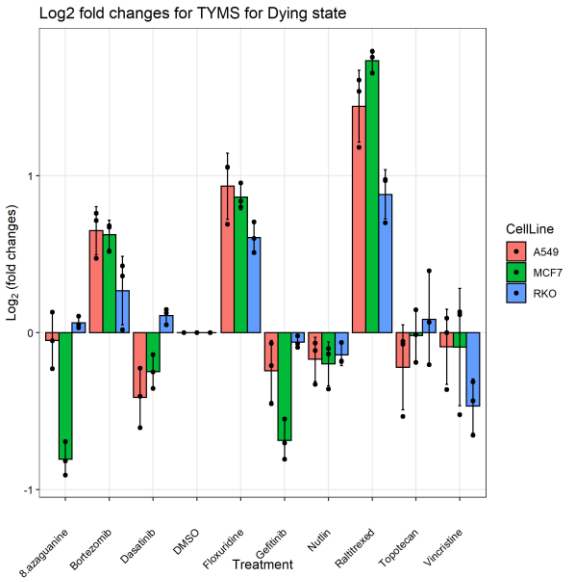
This appendix shows log2 fold changes for some drug targets in bar plots across all drug treatments and both cell states. Each bar represents the mean log_2 fold change, and each replicate is shown as a dot. The error bar displays the standard deviation.

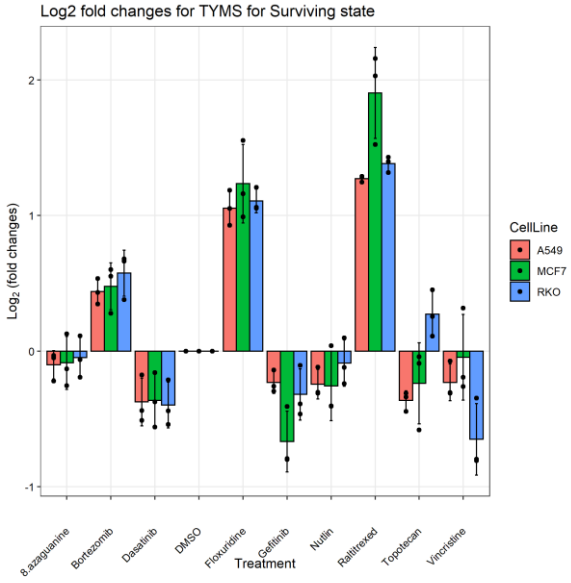





Log2 fold changes for TP53 for Dying state
Log2 (fold changes)
Treatment
CellLine
A549
MCF7
RKO
8.azaguanine
Bortezomib
Dasatinib
DMSO
Floxuridine
Gefitinib
Nutlin
Raltitrexed
Topotecan
Vincristine

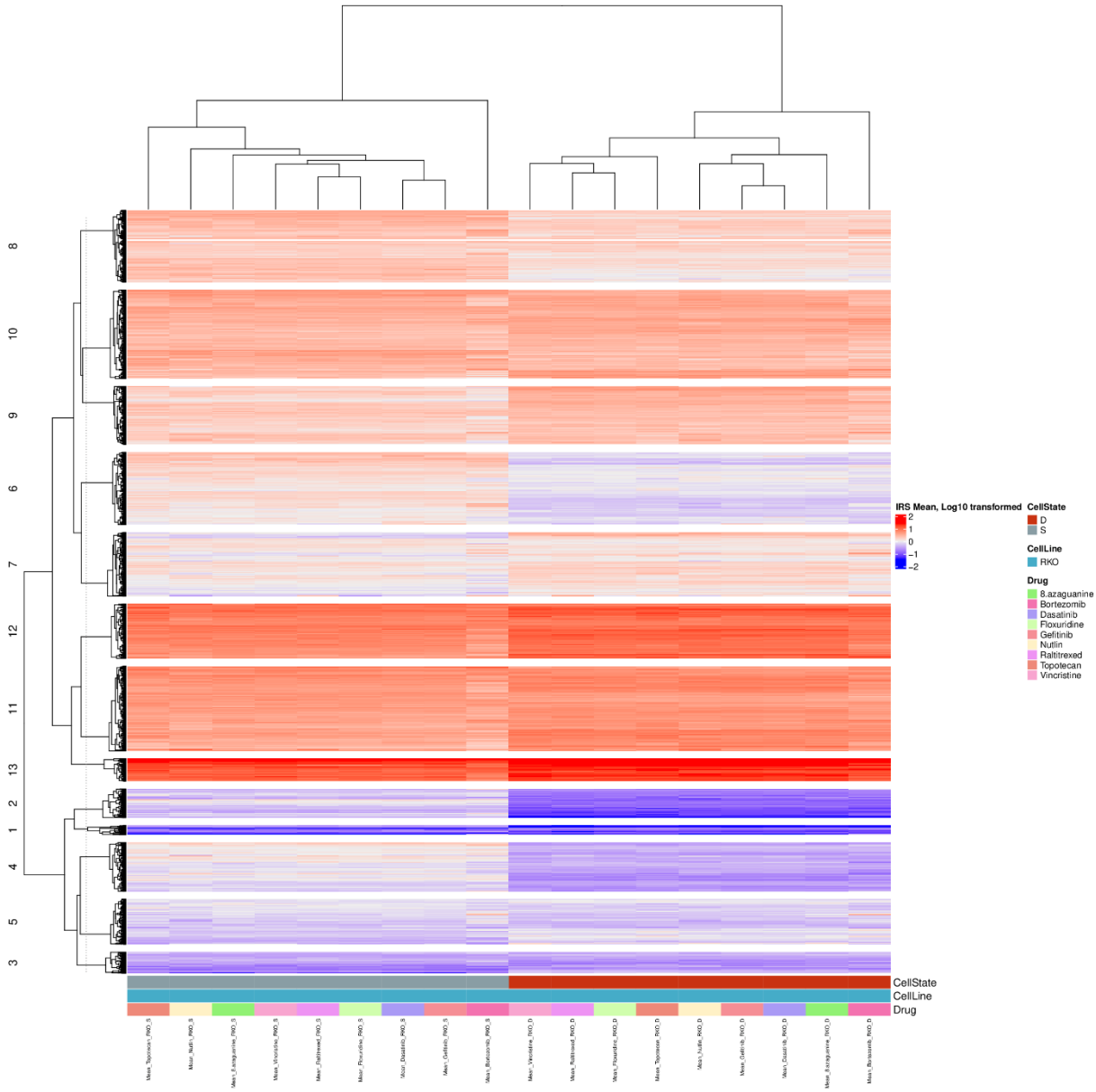

Log2 fold changes for TP53 for Surviving state
Log2 (fold changes)
Treatment
CellLine
A549
MCF7
RKO
8.azaguanine
Bortezomib
Dasatinib
DMSO
Floxuridine
Gefitinib
Nutlin
Raltitrexed
Topotecan
Vincristine

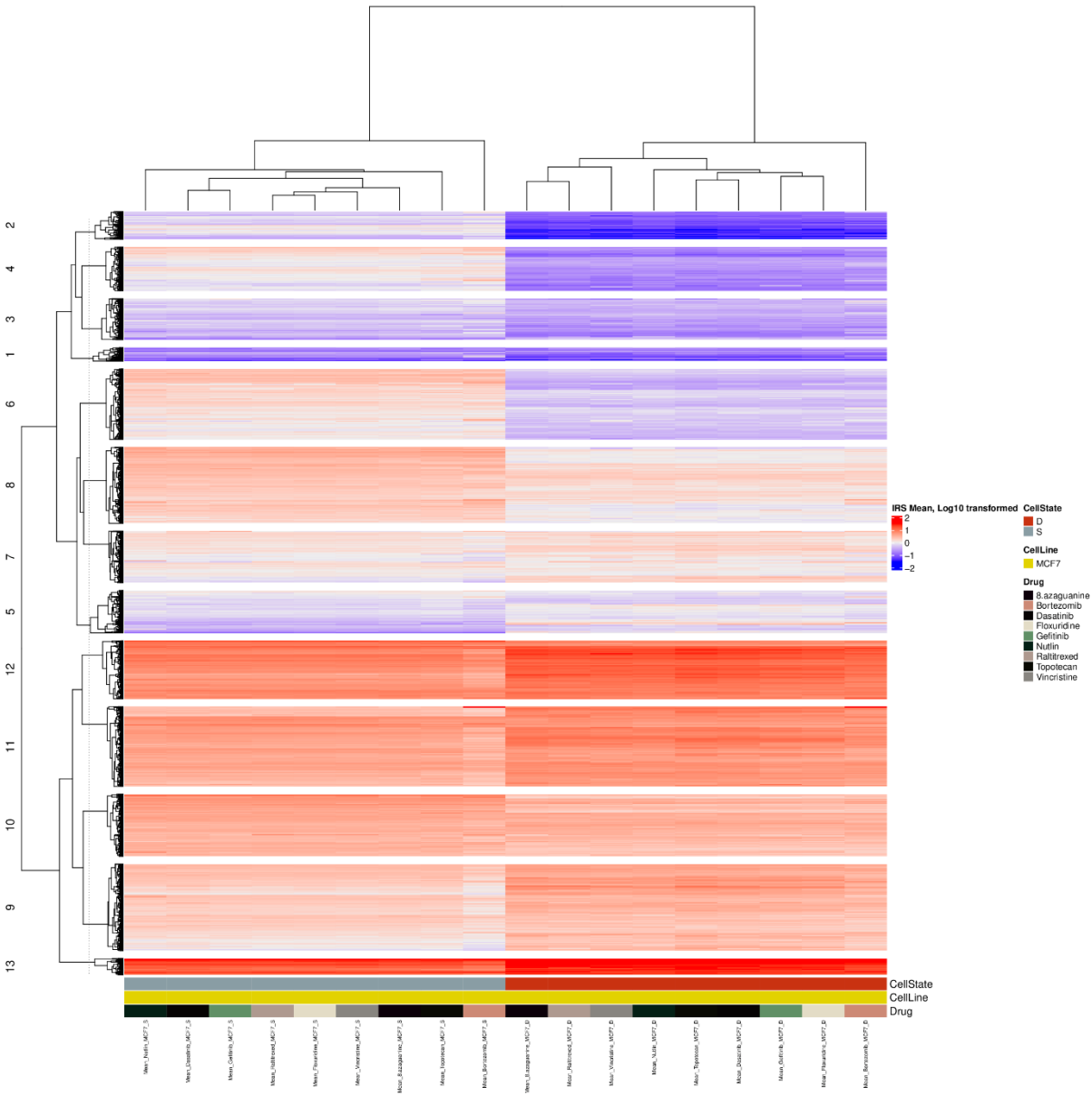

Log2 fold changes for TYMS for Dying state
Log2 (fold changes)
Treatment
CellLine
A549
MCF7
RKO
8.azaguanine
Bortezomib
Dasatinib
DMSO
Floxuridine
Gefitinib
Nutlin
Raltitrexed
Topotecan
Vincristine

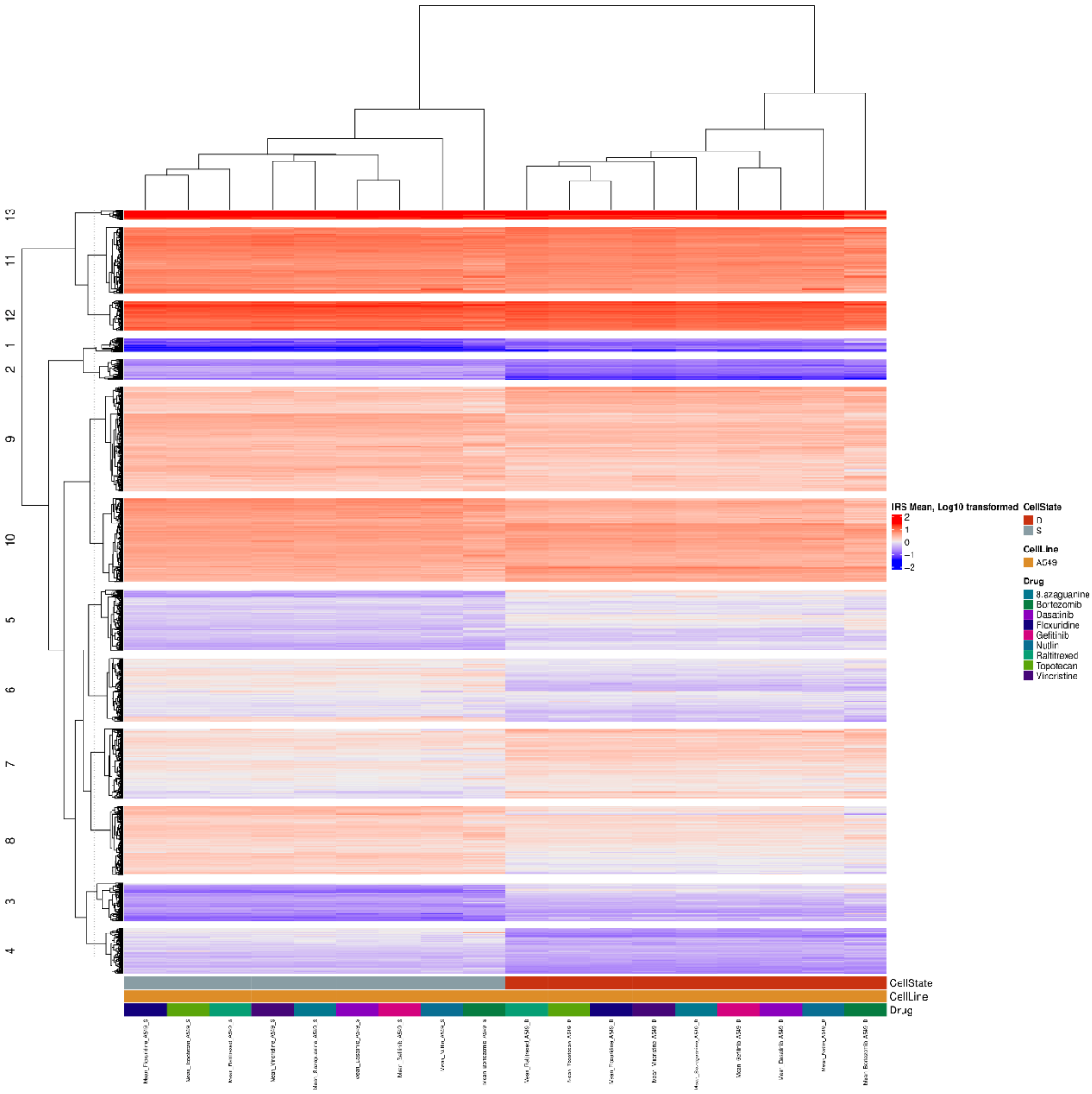

Log2 fold changes for TYMS for Surviving state
Log2 (fold changes)
Treatment
CellLine
A549
MCF7
RKO
8.azaguanine
Bortezomib
Dasatinib
DMSO
Floxuridine
Gefitinib
Nutlin
Raltitrexed
Topotecan
Vincristine

9 Appendix B - Heatmaps of individual cell lines

In this appendix, heatmaps of individual cell lines are shown. The first one is made on the RKO cell line, the second one on the MCF7 cell line, and the third on the A549 cell line.




IRS Mean, Log10 transformed
2
1
0
−1
−2
CellState
D
S
CellLine
MCF7
Drug
8.azaguanine
Bortezomib
Dasatinib
Floxuridine
Gefitinib
Nutlin
Raltitrexed
Topotecan
Vincristine
CellState
CellLine
Drug


IRS Mean, Log10 transformed
2
1
0
−1
−2
CellState
D
S
CellLine
A549
Drug
8 azaguanine
Bortezomib
Dasatinib
Floxuridine
Gefitinib
Nutlin
Raltitrexed
Topotecan
Vincristine
CellState
CellLine
Drug