


# Exploring the Mechanism and Suggesting Combination Therapies for HDAC Inhibitors in Androgen Receptor-Null Prostate Cancer Using Multivariate Statistical Analysis and Data Mining Techniques

Bioinformatics and Biology Insights  
Volume 16: 1–11  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11779322221145428  


Zohaib Rana, Rhonda J Rosengren and Paul F Smith

Department of Pharmacology and Toxicology, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand.

**ABSTRACT:** Previously, we showed that novel histone deacetylase (HDAC) inhibitors, *N*1-hydroxy-*N*<sup>8</sup>-(4-(pyridine-2-carbothioamido)phenyl) octanediamide (Jazz90) and [chlorido(η<sup>5</sup>-pentamethylcyclopentadienyl)(*N*1-hydroxy-*N*<sup>8</sup>-(4-(pyridine-2-carbothioamido)-κ<sup>2</sup>*N*, *S*)phenyl)octanediamide]rhodium(III)] chloride (Jazz167), have cytostatic and anti-angiogenic effects in androgen receptor-negative prostate cancer cells and are also non-toxic in BALB/c mice. However, only univariate statistical analysis was carried out to determine the role of individual proteins. In this study, multivariate statistical analyses (MVAs) and data mining procedures were carried out with the objective of determining the molecular networks that explain the growth inhibitory potential of Jazz90 and Jazz167 in PC3 cells and to determine potential inhibitors that can be used in combination with these HDAC inhibitors. Lasso regression revealed that angiogenic factors, vascular endothelial growth factor-A (VEGF-A), and vascular endothelial growth factor receptor-2 (VEGFR-2), alongside HDAC inhibition, predicted the reduction in cell number with an adjusted *R*<sup>2</sup> value of 0.99 following Jazz90 treatment, whereas VEGFR-2, acetylation of histone-3, and HDAC inhibition predicted cell number with an adjusted *R*<sup>2</sup> value of 0.84 following Jazz167 treatment. These results were further followed up with ridge regression, hierarchical cluster analysis, random forest classification (RFC), and support vector machines. RFC and support vector machines also predicted the treatment groups with a 100% accuracy. MVAs also revealed that Jazz90 should be examined in combination with epithelial to mesenchymal transitioning inhibitors, such as simvastatin and olaparib, whereas Jazz167 should be examined with venetoclax or navitoclax. Future studies should also address the roles of VEGF-A and VEGFR-2 in cellular proliferation, whereas p27 function should be examined for its role in PC3 cell migration.

**KEYWORDS:** HDAC inhibitors, multivariate analyses, data mining, Lasso regression, random forest classification, androgen receptor-negative prostate cancer

**RECEIVED:** July 12, 2022. **ACCEPTED:** November 21, 2022.

**TYPE:** Original Research Article

**FUNDING:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Paul F Smith, Department of Pharmacology and Toxicology, School of Biomedical Sciences, University of Otago, Dunedin 9016, New Zealand. Email: paul.smith@otago.ac.nz

## Introduction

Multivariate statistical analyses (MVAs) have routinely been used in proteomic studies of cancer in humans and animals but have rarely been used for *in vitro* studies.<sup>1,2</sup> In this study, MVAs and data mining procedures were used to understand which mechanistic pathways are key to the action of novel HDAC inhibitors, *N*1-hydroxy-*N*<sup>8</sup>-(4-(pyridine-2-carbothioamido)phenyl) octanediamide (Jazz90), [chlorido(η<sup>5</sup>-pentamethylcyclopentadienyl)(*N*1-hydroxy-*N*<sup>8</sup>-(4-(pyridine-2-carbothioamido)-κ<sup>2</sup>*N*, *S*)phenyl)octanediamide]rhodium(III)] chloride (Jazz167) and suberoylanilide hydroxamic acid (SAHA), in androgen receptor-negative (AR-) prostate cancer (PC3) cells. Previously, we examined the effects of Jazz90 and Jazz167 on cell survival and migration proteins<sup>3,4</sup> (Table 1). The broader objectives of our previous studies were to identify efficacious drugs for AR-prostate cancer because this cancer variant does not have a viable drug treatment and has a poor prognosis.<sup>3</sup> However, only univariate statistical analysis was used on this data set.<sup>2</sup>

MVA and data mining analyses are specifically useful to elucidate the mechanisms of action of novel HDAC inhibitors. HDACs are a complex family of molecules that contain 18 different types of HDAC isoenzymes, out of which 11 are

zinc-dependent and 7 are NADPH-dependent.<sup>5</sup> HDACs, via deacetylation of histones and non-histone proteins, downregulate tumour suppressors in processes, including proliferation, survival, angiogenesis, and epithelial to mesenchymal transitioning<sup>5</sup> (Table 1). As Jazz90 and Jazz167 were synthesised using SAHA as a parent compound, these drugs contain a hydroxamate moiety capable of interacting with zinc ions. However, Jazz90 and Jazz167 have different cap regions in comparison to SAHA. Previous studies report that altering the cap regions on HDAC inhibitors can shift their HDAC isoenzyme inhibitory potential.<sup>6</sup> In turn, a shift in the processes regulated by HDAC inhibitors is observed. The shift in the regulated processes can be evaluated by MVA and data mining analyses, which in turn indicate the potential mechanism of action via which Jazz90 and Jazz167 function. Importantly, MVA and data mining analyses not only highlight potential drug-mediated mechanisms of action but they also elucidate the mechanism by which the drug might not act. Thus, an additional role for their use is to determine possible compound specific combination treatments to increase efficacy.

Various MVA and data mining analyses, such as hierarchical cluster, lasso and ridge regression, random forest classification



**Table 1.** Effect of Jazz90, Jazz167, and SAHA on protein levels in PC3 cells.

PROTEINS	JAZZ90	JAZZ167	SAHA
Acetylated H3	↑	↑	↑
Phosphorylated Akt	↑	NC	↑
Phosphorylated Erk	↓	NC	NC
Cyclin D	↓	↓	↓
Cyclin B	NC	NC	NC
P21	↑	↑	↑
P27	↑	↑	↑
E-cadherin	NC	NC	NC
Vimentin	NC	NC	NC
VEGF-A	↓	↓	NC
VEGFR-2	↓	↓	NC
Bcl-2	NC	NC	NC

Abbreviations: ↑ upregulation; ↓ downregulation; NC, no change in protein expression; SAHA, suberoylanilide hydroxamic acid; VEGF-A, vascular endothelial growth factor-A; VEGFR-2, vascular endothelial growth factor receptor-2.

(RFC), and support vector machine (SVM) analyses, were conducted. Cluster analysis (CA) was used to identify the associations between different proteins, whereas lasso and ridge regression were used to identify the combination of variables that could predict drug-mediated changes in cell number. RFC and SVMs were used to determine whether treated samples could be classified as drug- or control-treated while simultaneously ranking the proteins in their order of importance. The combination of proteins involved in the prediction of cell number and drug treatment not only provides an insight into their mechanism of action but may also aid in the development of biomarkers for cancers and their drug responsiveness. Although there are a multitude of different data mining procedures,<sup>2</sup> in our other studies, we have found RFC to be particularly useful,<sup>2</sup> and, in any case, the aim here was to explore the use of these kinds of methods for this data set, rather than conducting a systematic comparison of multiple data mining procedures.

## Methods

The MVAs and data mining analyses conducted here use experimental results reported previously.<sup>3,4</sup> The results obtained from standard assays to assess HDAC inhibition and expression levels of acetylation of histone-3, cell cycle (cyclin D, cyclin B, p21, and p27), apoptotic (bcl-2), epithelial to mesenchymal transitioning (E-cadherin and vimentin), angiogenic (VEGF-A and VEGFR-2), proliferative (pErk/Erk), and survival (PI3K/Akt) proteins in PC3 cells following Jazz90, Jazz167, and SAHA treatment at 0.05, 0.1, 1, 4, 6 and 8  $\mu$ M, were used in the analyses. Results used were the means obtained from experiments carried out in triplicate.

## MVAs and data mining analyses

**Lasso and ridge regression.** Lasso and ridge regression were used to predict PC3 cell number based on the other variables studied.<sup>7</sup> Limitations of overfitting and multicollinearity were avoided using these multiple regression techniques.<sup>7</sup> Lasso regression is an example of multiple linear regression which conducts variable selection and regularisation to increase the prediction accuracy of the model. Regularisation involves adding a penalty to the coefficients of variables. In the case of lasso regression, it can shrink the coefficients to 0, and regularisation eliminates unimportant variables and makes the models sparser.<sup>7</sup> Adjusted  $R^2$  and root mean square error (RMSE) values were generated. Lasso and ridge regression were carried out, and graphs were generated using the following code in R4.0.3:

```
library(readr)
df <- read_csv("Jazz90 PC3 cell MVA (1) updated.csv")
library(caret)
set.seed(1985)
index <- createDataPartition(df$Cell_Number, p=0.5,
list=FALSE, time=1)
df <- as.data.frame(df)
train_df <- df[index,]
test_df <- df[-index,]
ctrlspecs <- trainControl(method="cv", number=2,
savePredictions="all")
lambda_vector <- 10^seq(5, -5, length=500)
set.seed(1985)
model1 <- train( Cell_Number~., data=train_df, preProcess=
c("center", "scale"), method="glmnet", tuneGrid=expand.
grid(alpha= 1, lambda=lambda_vector), trControl=
ctrlspecs, na.action=na.omit)
round(coef(model1$finalModel, model1$bestTune$lambda), 3)
varImp(model1)
ggplot(varImp(model1))
predictions1 <- predict(model1, newdata=test_df)
mod1perf <- data.frame(RMSE=RMSE(predictions1,
test_df$Cell_Number), Rsquared=R2(predictions1,
test_df$Cell_Number))
View(mod1perf)
```

As 2-fold cross-validation was used, the data were first partitioned into training (50%) and testing (50%) data sets. For lasso regression,  $\alpha = 1$ , whereas  $\alpha = 0$  for ridge regression.<sup>8</sup> Before the analysis, cell number was determined to be normally distributed using the D'Agostino–Pearson normality test. Cell number was then predicted based on the identified variables in the testing data set. Root mean squared deviation (RMSD) and adjusted  $R^2$  values were then determined.

**Random forest classification.** RFC is a supervised learning algorithm, which can be used to predict categorical variables.<sup>9</sup> Generally, it consists of 2 data sets, a training and a test data set, and therefore, the data have to be partitioned.<sup>9</sup> The training

data set is used to train the model, and the model is then applied to a test data set to predict the categorical variable.<sup>9</sup> RFC trains the model by applying a flow-like series of questions to the variables (proteins) and constructs trees. RFC can construct hundreds or thousands of trees. Each of the trees generates its own decision and predicts whether the sample fits into a certain category.<sup>9</sup> The model from the training data set can then be cross-validated using the test data set, predicting, for example, whether the cells have been drug-treated or not. This confirms whether changes in the levels of the variables can or cannot predict an outcome. As different variables are used to construct the trees, RFC also reports the mean decrease in a Gini coefficient and mean decrease in accuracy. The higher the values of the mean decrease in accuracy or the mean decrease in Gini score, the greater the importance of the variable in the model.<sup>9</sup> Of note is that RFC is largely assumption-free, and therefore, it is a useful method if the data do not fit the General Linear Model (GLM) assumptions. Furthermore, simulation studies show that RFC can provide robust results even with high-dimensional data and small sample sizes.<sup>10,11</sup>

In our study, 500 trees were generated using a 3-variable split. The reason for using 500 trees was that the error decreased at around 200 trees, and therefore, 500 trees were considered sufficient. A 3-variable split was used because previous studies suggest that  $\sqrt{\text{number of variables}}$  is optimal.<sup>11</sup> The data set was split into a training data set (70% of the samples) and a test data set (30%). All RFC analyses were performed using R and Rattle, an R package.<sup>12</sup> In R, graphs of error as a function of the number of trees, 'out-of-bag' (OOB) receiver operating characteristic (ROC) curves and a list of predictor variables of importance were plotted.

**Support vector machines.** SVMs are a supervised learning model that can be used for classification analysis, which we have used in some of our previous studies.<sup>13</sup> Here, 2 forms of SVM include linear and non-linear. Linear SVMs can be used for data that can be described using a straight line, whereas non-linear SVMs can be used for data that cannot be separated using straight lines and require kernel factors.<sup>14</sup> Linear SVM was used to classify samples based on whether they were treated or not, as the data could be separated via a straight line into treated and untreated groups. To ensure that the model was effective, cross-validation was carried out. Therefore, the data were split into training (70%) and test (30%) sets, and the SVM model generated using the training data set was tested on the test data set. An advantage of SVM is that it does not make distributional assumptions except for the independence and similar distribution of the groups of data.<sup>13</sup> Therefore, even small sample sizes can provide accurate estimates even if there are a large number of variables. All SVM analyses were performed using R and Rattle.

**Cluster analyses.** CA is an example of a non-parametric statistical method that is used to assess the associations between

independent variables without a predetermined dependent variable.<sup>15</sup> No assumptions about normality or homogeneity of the covariance matrices are made in CA. It uses measures of distance between variables, grouping them according to their degree of association. A hierarchical, Ward minimal linkage algorithm was used, along with the squared Euclidean distance,<sup>15</sup> based on our previous experience with CAs.<sup>2</sup> Thus, the clusters were formed based on the Minimisation of variance.<sup>15</sup> As the data for different variables were measured on different scales, the data were first converted to *z* scores. The results of CA can be represented on a dendrogram, in which variable similarity increases if the distance between them decreases.<sup>15</sup> SPSS 27 was used for CA.

## Results

The association between different variables was tested using a Pearson's Product Moment correlation coefficient, an example of a bivariate analysis (Figure 1). This test also determined the degree of multicollinearity between the variables. As expected, a significant positive correlation coefficient (*r*) of 0.99 ( $P < .001$ ) was observed between HDAC inhibition and cell number for all of the drugs. In the case of Jazz90, cell number also correlated with acetylation of histone-3 ( $r = -0.75$  [ $P < .01$ ]), cell cycle p21 ( $r = -0.83$  [ $P < .003$ ]), p27 ( $r = -0.83$  [ $P < .01$ ]), cyclin D ( $r = 0.85$  [ $P < .001$ ]), pErk/Erk ( $r = 0.92$  [ $P < .001$ ]) and angiogenic (VEGF-A) ( $r = 0.83$  [ $P < .01$ ]) and (VEGFR-2) ( $r = 0.92$  [ $P < .001$ ]) proteins. Jazz167 correlation analysis showed that cell number also correlated with acetylation of histone-3 ( $r = -0.8$  [ $P < .01$ ]), cell cycle p27 ( $r = -0.87$  [ $P < .001$ ]), EMT transitioning (E-cadherin) ( $r = -0.81$ , [ $P < .01$ ]), and angiogenic (VEGF-A) ( $r = 0.84$  [ $P < .01$ ]) and (VEGFR-2) ( $r = 0.91$  [ $P < .001$ ]) proteins. Results from SAHA, in comparison, showed that cell number was significantly correlated with acetylation of histone-3 ( $r = -0.95$  [ $P < .001$ ]) and cell cycle p21 ( $r = -0.95$  [ $P < .001$ ]), p27 ( $r = -0.86$  [ $P < .01$ ]), and cyclin D ( $r = 0.82$  [ $P < .01$ ]) proteins (Figure 1).

As multicollinearity between the variables was high, lasso regression was used instead of general multiple linear regression to identify predictors for cell number, which were normally distributed. Lasso regression indicated that HDAC inhibition was the strongest predictor of cell number for all of the drugs (Figures 2, 3, and 4A). In the case of Jazz90, HDAC inhibition, VEGF-A, and VEGFR-2 predicted cell number with an adjusted  $R^2$  value of 0.99, whereas VEGFR-2, HDAC inhibition, and acetylation of histone-3 predicted cell number with an adjusted  $R^2$  value of 0.84 for Jazz167 (Figures 2 and 3A). In comparison, HDAC inhibition was the only predictor for cell number after SAHA treatment (Figure 4A). To determine the least important variables, ridge regression was used. In the case of Jazz90, E-cadherin was the least important followed by cyclin B and bcl-2, whereas for Jazz167, cyclin D and bcl-2 were the least important variables (Figures 2 and 3B). For

SAHA, cyclin B followed by bcl-2 were the least important variables. The most important variables identified by ridge regression were similar to lasso regression (Figure 4B).

To identify the importance of variables in predicting treated and untreated groups, RFC analysis was carried out (Figure 5). For Jazz90, acetylation of histone-3, cell cycle (p27, pErk/Erk,

and p21), and angiogenic (VEGF-A) proteins were the most important variables as observed by the mean decrease in accuracy and Gini values. For Jazz167, angiogenic (VEGFR-2 and VEGF-A), and EMT (E-cadherin), proteins were identified as the most important proteins. In comparison, cell cycle proteins (p21 and p27) differentiated treated and untreated groups

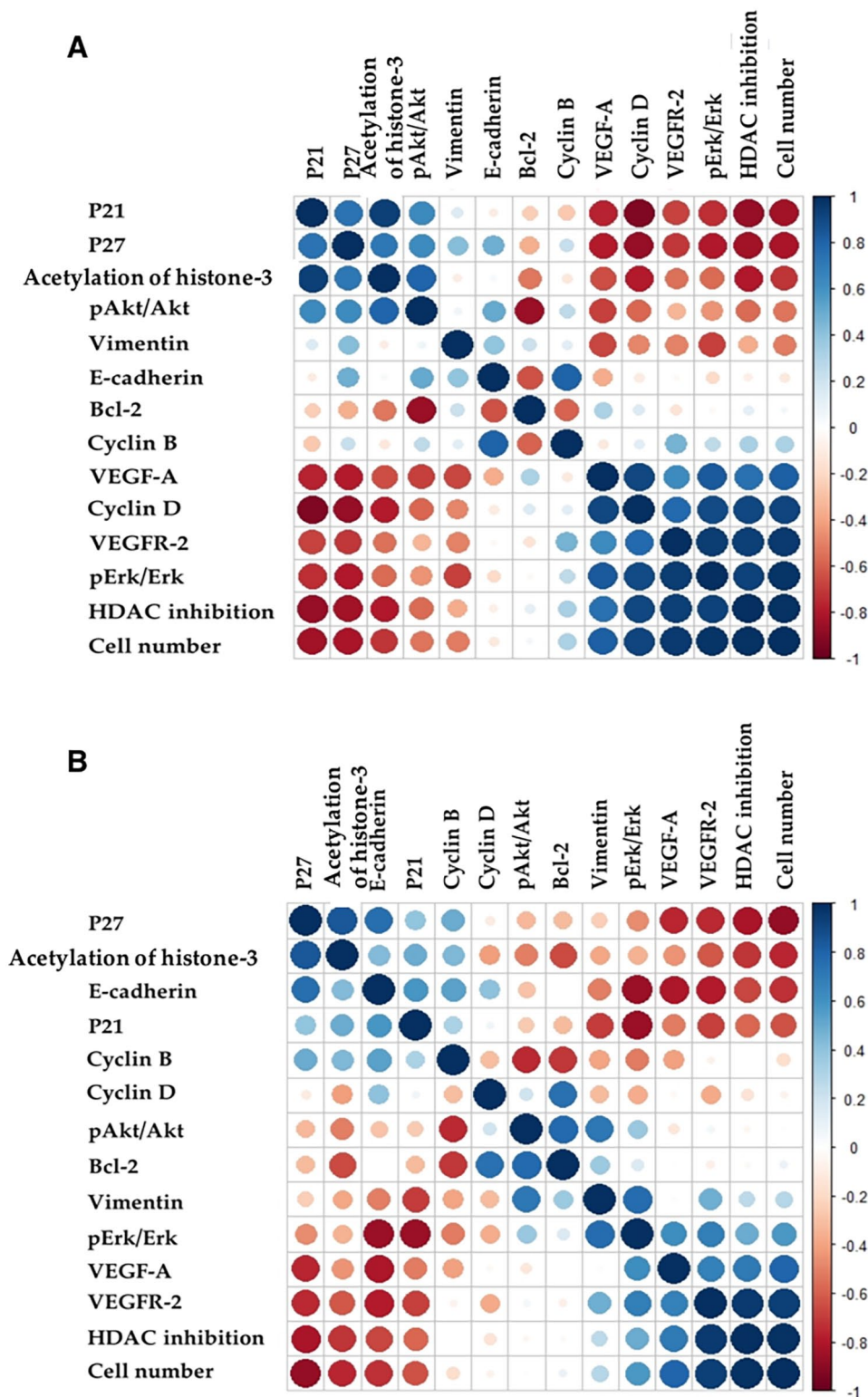
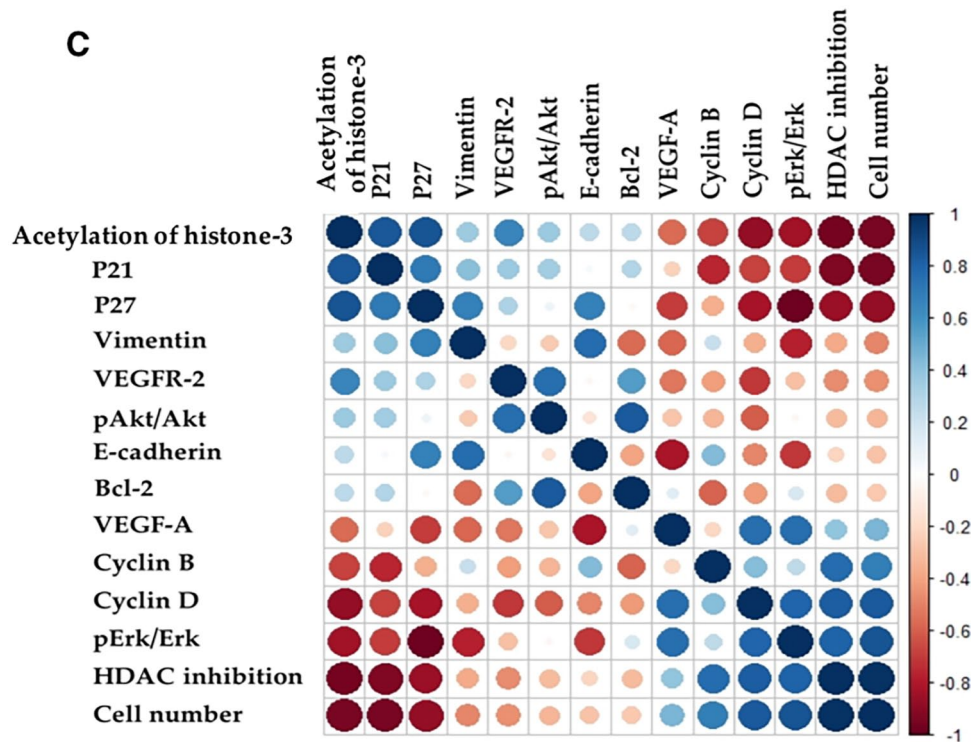


Figure 1. (Continued)



**Figure 1.** A Pearson Product Moment Correlation Coefficient matrix showing the relationship between variables relating to (A) Jazz90, (B) Jazz167, and (C) SAHA treatment. Positive correlations are shown in blue, negative correlations in red, and the size of the circle indicates the degree of the correlation.

following SAHA treatment. RFC showed that EMT proteins were the least important variables for Jazz90 and SAHA treatment. For Jazz90 and SAHA, anti-apoptotic (Bcl-2 and pAkt/Akt) proteins were also identified as less important. Cell cycle proteins, cyclin D, and cyclin B, were the least important variables for Jazz167 treatment. Further analysis of the testing data set revealed that the model differentiated between the treated and untreated groups with 100% accuracy (Figure 5G to I) for all of the drugs. The RFC results were further supported by SVM analysis, which also predicted the treated and untreated data sets with 100% accuracy.

The results of the hierarchical CA also showed that HDAC inhibition was closely associated with cell number for the drugs (Figure 6). However, CA also identified additional variables closely associated with cell number. For Jazz90, cell cycle proteins (cyclin D, cyclin B, and pErk/Erk) were identified in addition to angiogenic proteins, whereas angiogenic (VEGFR-2 and VEGF-A), cell cycle (pErk/Erk), and EMT (vimentin) proteins were identified for Jazz167. For SAHA, cell cycle proteins clustered closely with cell number. In the case of Jazz90 and SAHA, EMT proteins formed the most distant cluster from cell number and HDAC inhibition. In comparison, anti-apoptotic (pAkt/Akt and bcl-2) proteins formed a distant cluster from cell number for Jazz167 (Figure 6).

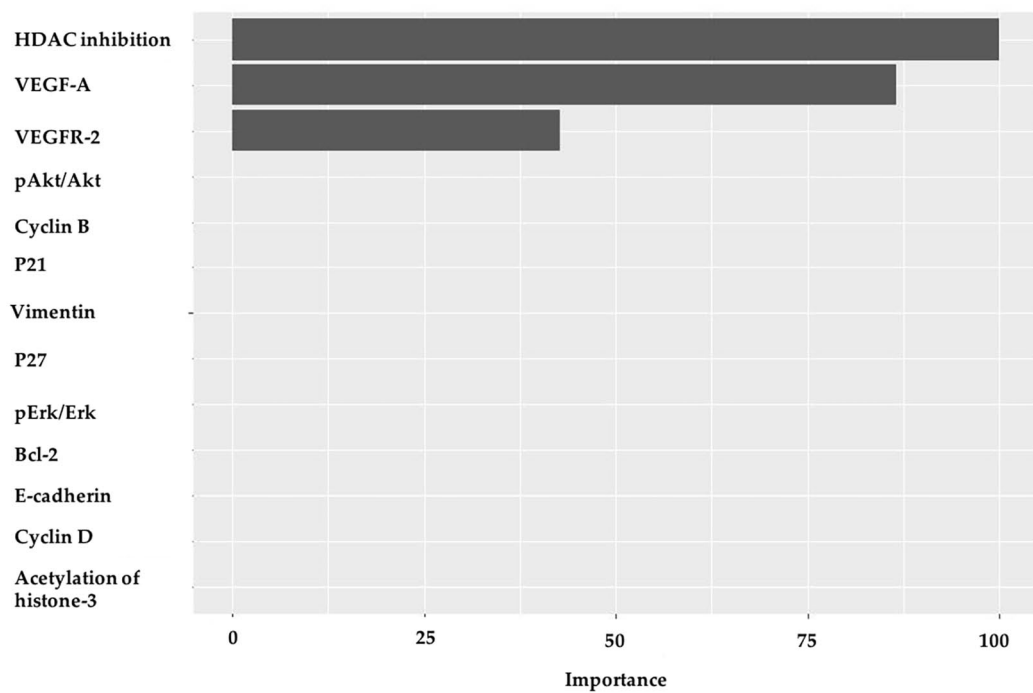
## Discussion

MVAs and data mining procedures are useful to study the interactions between systems of proteins and to develop

biomarkers in cancers. This is possible because MVAs and data mining analyses determine the importance of a variable within a network of variables and how it may change relative to the other network components. Various other studies have used MVAs to identify the protein signatures of cancers. For example, Martey et al<sup>2</sup> used a variety of multivariate statistical methods to determine a protein signature (Ki67, PKC- $\alpha$ , PP2AA- $\beta$ , and CaD1) that could be used to predict tumour growth in mice in response to a curcumin derivative encapsulated in a nanoparticle. Similarly, De Marchi et al<sup>1</sup> predicted tamoxifen response in breast cancer patients using 4 proteins (programmed cell death 4, cingulin, ovarian carcinoma immune-reactive antigen domain containing protein 1, and Ras-GTPase-activating protein-binding protein 2 proteins) identified by MVA analyses.

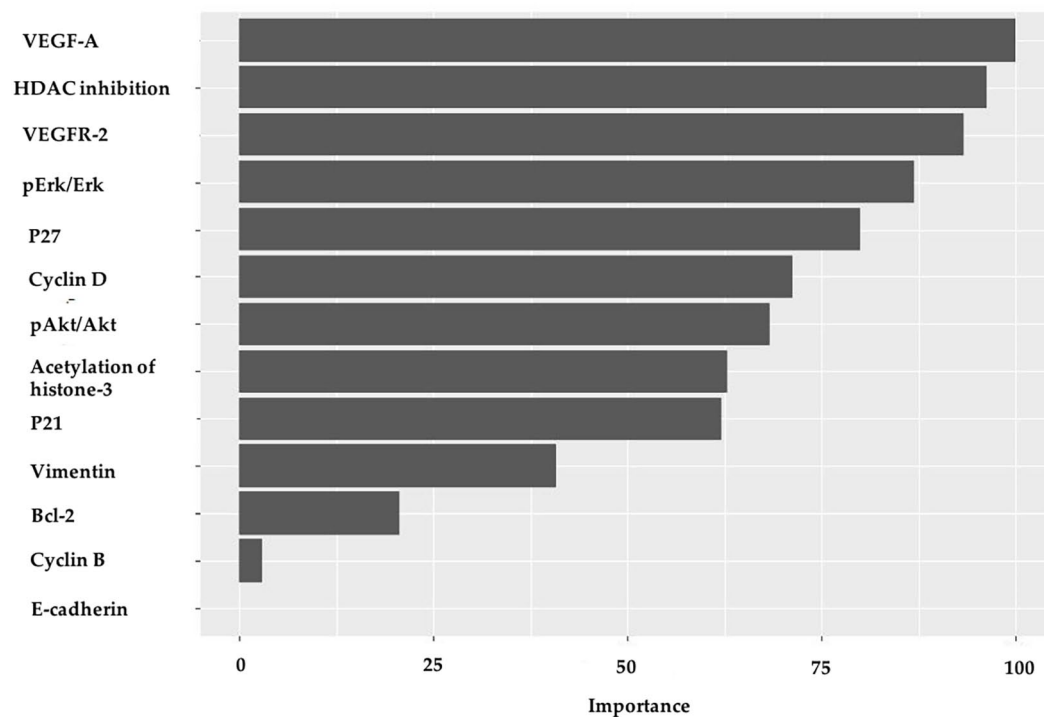
Recently, we showed that Jazz90 and Jazz167 are cytostatic HDAC inhibitors with an effect on angiogenic pathways in addition to inhibiting cell cycle arrest. MVAs were carried out to identify a protein signature that could predict cell number or whether the cells had been treated and drug efficacy. Despite inhibiting the same molecules, these drugs had different signatures. Lasso regression showed that HDAC inhibition, VEGF-A, and VEGFR-2 inhibition predicted cell number for Jazz90, whereas HDAC inhibition, VEGFR-2, and acetylation of histone-3 predicted it for Jazz167. In the case of SAHA, only HDAC inhibition was found to be predictive. RFC and CA further confirmed the importance of these angiogenic proteins in Jazz90- and Jazz167-treated prostate cancer cells. The

### A Lasso regression



RMSE	R squared
1953.39	0.99

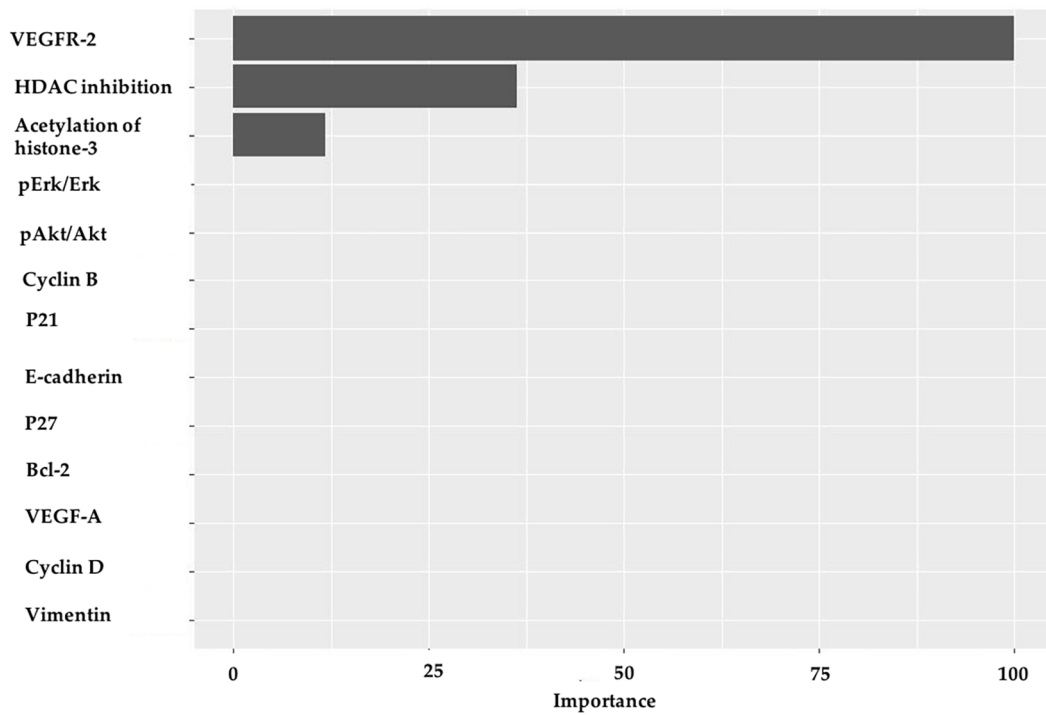
### B Ridge regression



RMSE	R squared
6272.46	0.86

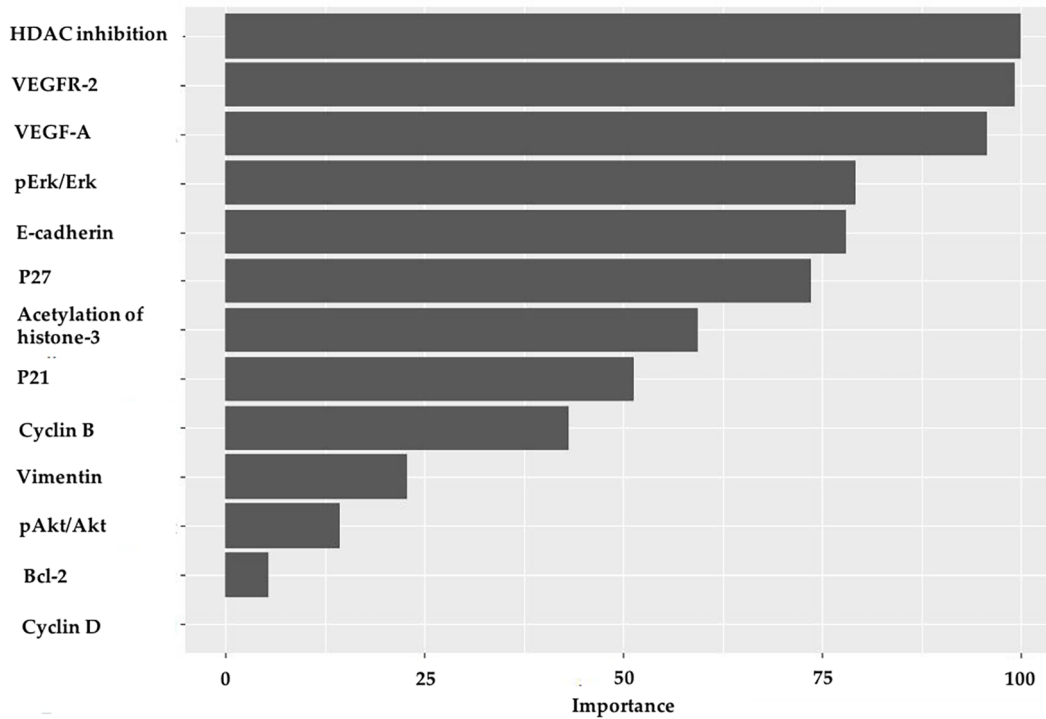
**Figure 2.** Lasso and ridge regression with 2-fold cross-validation for Jazz90. Plots depicting the importance of variables in predicting cell number were plotted for Jazz90. RMSE and adjusted  $R^2$  values were generated from cross-validation used to test the model on a test data set.

**A Lasso regression**



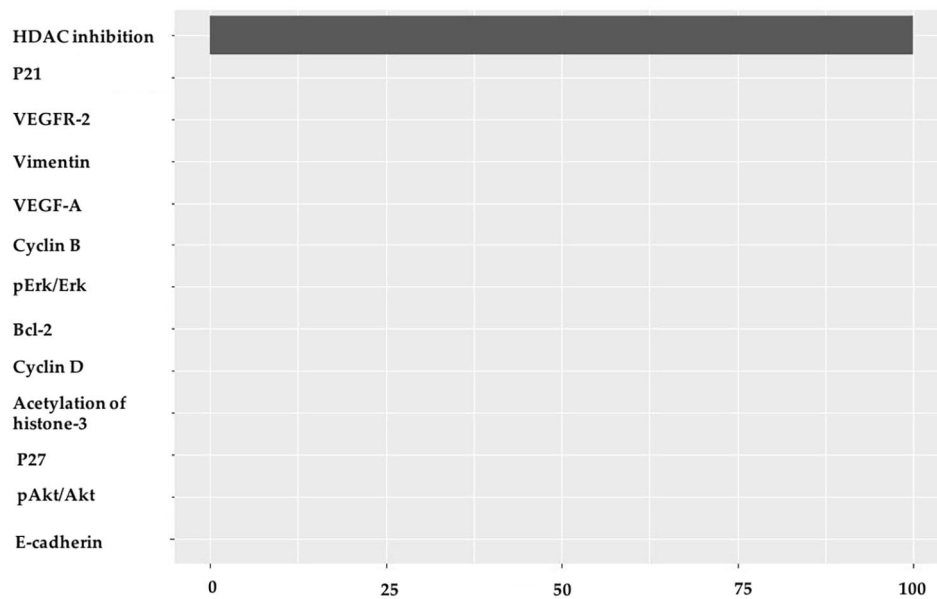
RMSE	R squared
2184.76	0.84

**B Ridge regression**

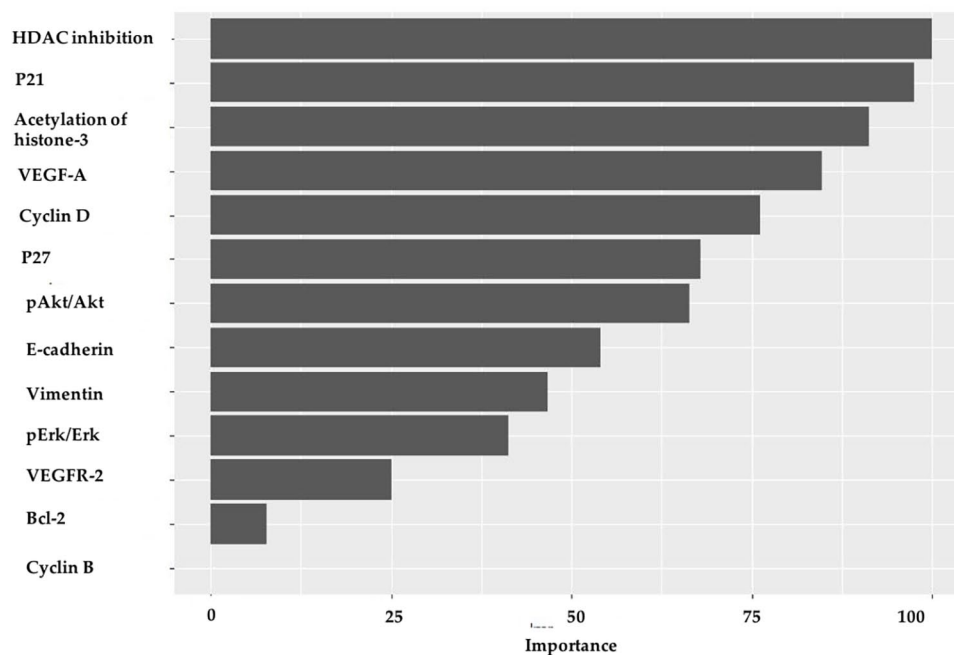


RMSE	R squared
3983.24	0.85

**Figure 3.** Lasso and ridge regression with 2-fold cross-validation for Jazz167. Plots depicting the importance of variables in predicting cell number were plotted for Jazz167. RMSE and adjusted  $R^2$  values were generated from cross-validation used to test the model on a test data set.

**A Lasso regression**

Importance	
RMSE	R squared
6242.217	0.99

**B Ridge regression**

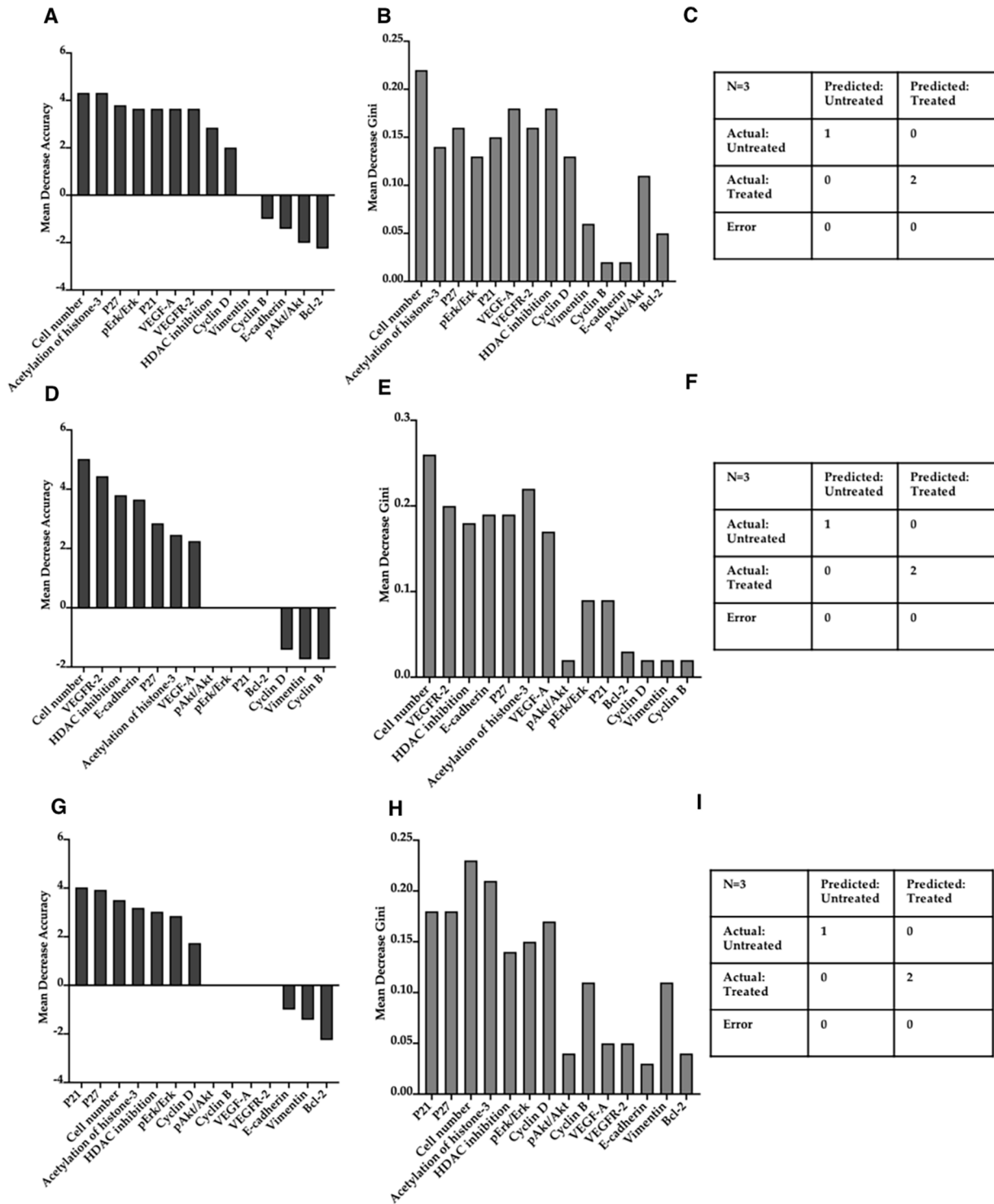
Importance	
RMSE	R squared
6272.46	0.86

**Figure 4.** Lasso and ridge regression with 2-fold cross-validation for SAHA. Plots depicting the importance of variables in predicting cell number were plotted for SAHA. RMSE and adjusted  $R^2$  values were generated from cross-validation used to test the model on a test data set.

anti-angiogenic effect of Jazz90 and Jazz167 may be attributed to their selective inhibition of HDAC6.<sup>16</sup> HDAC6 does not function via a histone-mediated pathway, instead it acts directly on non-histone proteins, such as VEGFR-2. On HDAC6 inhibition, acetylation of VEGFR-2 increases. This in turn

degrades VEGFR-2 by a clathrin-dependent internalisation process.<sup>17</sup> Jazz90 and Jazz167 are 4-fold more potent at inhibiting HDAC6 in comparison to SAHA.<sup>16</sup> High expression of VEGF-A and VEGFR-2 is also associated with a poor prognosis in prostate cancer; therefore, inhibition of these



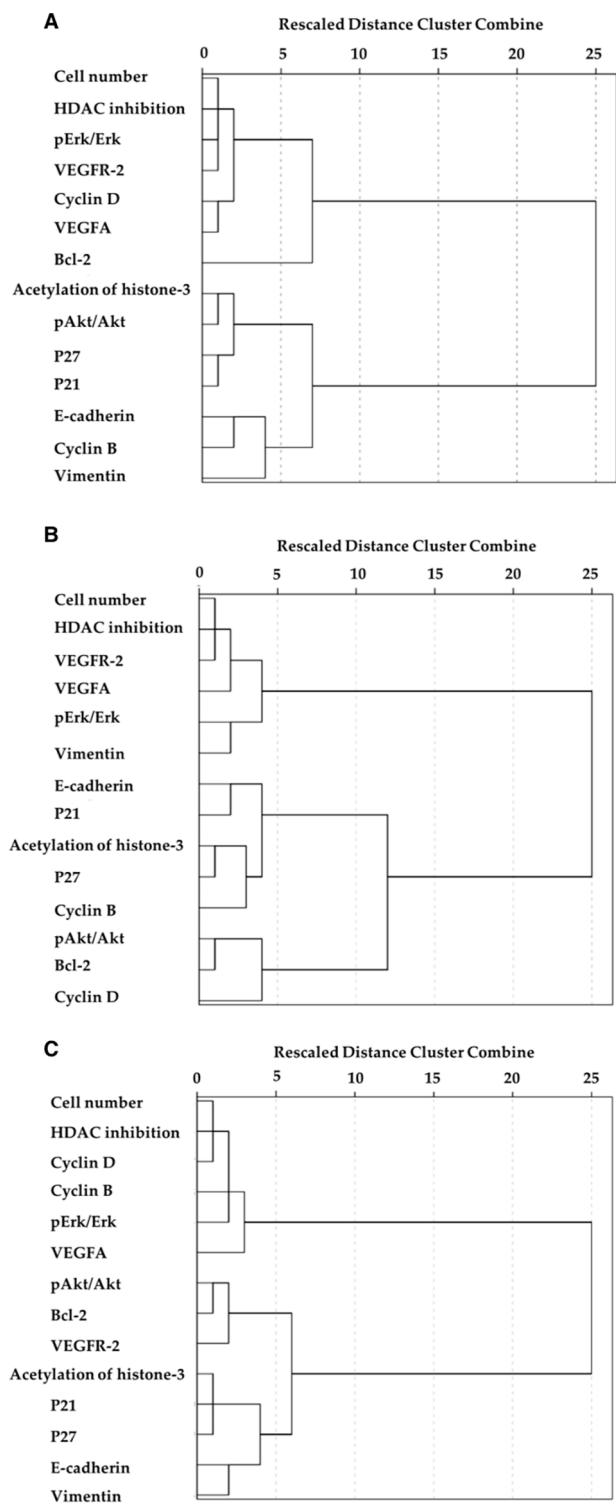


**Figure 5.** Rank order of the specific proteins critical to splitting data sets into treated and untreated groups. RFC of protein expression was conducted using R. Plots for mean decrease accuracy, which is the average measure of the decrease in accuracy in classifying groups if the variable is removed from the model, were plotted for (A) Jazz90, (D) Jazz167, and (G) SAHA. Plots for mean decrease Gini, a measure of the difference in split nodes over trees in the model for an individual variable, were computed for (B) Jazz90, (E) Jazz167, and (H) SAHA. OOB classification matrix for (C) Jazz90, (F) Jazz167, and (I) SAHA is shown.

molecules may be associated with a better outcome for patients.<sup>18</sup> Of note, no VEGF-A and VEGFR-2 inhibitors are available for prostate cancer treatment.<sup>18</sup> Despite undergoing clinical trials, VEGF-A and VEGFR-2 inhibitors, such as bevacizumab and sunitinib, have exhibited toxicity.<sup>18</sup> Since

Jazz90 and Jazz167 were not toxic in BALB/c mice, they may be useful candidates for further efficacy testing for angiogenic pathways *in vivo*.

Acetylation of histone-3 was identified as the strongest predictor of cell number in response to Jazz167 treatment using



**Figure 6.** A dendrogram showing the clustering of variables following (A) Jaz90, (B) Jaz167, and (C) SAHA treatment. The hierarchical clustering of variables method was determined using the agglomerative Ward's method and squared Euclidean distance.

lasso regression. Previously, Hanif et al<sup>16</sup> also showed that Jaz167 was 10-fold more potent at inhibiting a nuclear HDAC (HDAC1) in comparison to SAHA. Therefore, a genome-wide chromatin immunoprecipitation analysis needs to be conducted to understand the interactions between

different acetylated regions of histone-3 and promoters of tumour suppressor genes, such as *p21* and *p27*.<sup>19</sup> This is critical because prostate cancer is marked by hypoacetylation at sites, lysine 9 (H3K9), lysine 14 (H3K14), lysine 18 (H3K18), and lysine 23 (H3K23), on histone-3.<sup>20</sup> In turn, the inter-relatedness of the different genes identified can then be evaluated using the Ingenuity Pathway Analysis (IPA) software tool. Thus, these experiments will yield additional targets.

MVAs and data mining analyses are also useful to identify interactions between different signalling pathways. Cancers, in particular, involve interactions between a complex network of signalling pathways. For example, pathways involved in angiogenesis, EMT, proliferation, and survival have been linked together. The complexity is further increased by proteins assuming dual roles.<sup>21-23</sup> For this reason, the correlation of VEGF-A and VEGFR-2 with cell number is a surprising finding because these angiogenic markers are generally associated with endothelial tube formation and migration.<sup>24</sup> However, their role in proliferation and survival has been assessed in other cancers, such as ovarian and colorectal cancer, but they have not been tested in PC3 cells.<sup>23</sup> Therefore, the results from MVAs indicate that knockdown studies of VEGF-A and VEGFR-2 to elucidate their role of proliferation in PC3 cells should be conducted. Previous reports have also suggested that p21 and p27 are involved in migration in addition to regulating the cell cycle.<sup>21,22</sup> Although p21, p27, acetylation of histone-3, and cyclin B clustered closely with E-cadherin in Jaz167-treated cells, the Pearson's correlation analysis and RFC showed that p27 and E-cadherin were closely linked. Therefore, the interaction between p27 and E-cadherin needs to be investigated further in Jaz167-treated cells, especially because p27 has been linked to migration in the past.<sup>22,25</sup>

Another benefit of using MVA is the potential to identify viable drug combinations. Combination treatments are critical in cancer because proliferating cancer cells can acquire resistance to single target agents.<sup>26</sup> Previously, we showed that Jaz90 and Jaz167 are cytostatic drugs and were non-toxic in BALB/c mice at doses as high as 75 mg/kg<sup>3</sup> and, therefore, are good candidates for use in combination with other agents. Since all the proteins analysed in the study significantly predict the prognosis in prostate cancer patients and function in different pathways,<sup>18,27,28</sup> the least important proteins in our analyses should be targeted in combination with Jaz90 and Jaz167. For Jaz90, EMT transitioning proteins did not associate with cell number or treatment. Various FDA approved drugs, such as simvastatin and olaparib, have been shown to inhibit the EMT transitioning process.<sup>29</sup> Thus, further investigation should be carried out for these drugs in combination with Jaz90. In contrast, anti-apoptotic (bcl-2 and pAkt/Akt) and cell cycle proteins did not associate with cell number or treatment for Jaz167. Therefore, Jaz167 should be tested in combination with venetoclax or navitoclax, inhibitors of anti-apoptotic proteins, in PC3 cells.<sup>30</sup>

One limitation of this study is that it was restricted to only 1 cell line, PC3 cells. However, we do not believe that this

limitation undermines the demonstration of the utility of using MVAs and data mining analyses in an effort to understand the interaction of different cell signalling pathways in the development of cancer. First, very recently, others have shown that some of the same or similar methods that we used, such as CA, multiple regression, and linear discriminant analysis, can be useful in studying hepatocellular carcinoma cell lines,<sup>31</sup> head and neck squamous cell carcinoma (HNSCC) cell lines,<sup>32</sup> and nasopharyngeal carcinoma cell lines.<sup>33</sup> Therefore, based on these studies, and the general principles of MVAs and data mining procedures, it is very unlikely that these methods are applicable only to PC3 cell lines. Second, the aim of this study was not to investigate the utility of these analytic methods across a multitude of cell lines, but to illustrate how they might be used to discover how cell signalling pathways interact and form a protein signature in response to certain drug treatments.

## Conclusions

Overall, our study shows the usefulness of MVAs and data mining analyses in highlighting the key mechanisms via which drugs act, in determining combination treatments and predicting the dual roles via which proteins act in different signalling pathways in in vitro cancer studies. MVAs and data mining analyses identified a protein signature and suggested combination treatments for Jazz90 and Jazz167 in PC3 cells. Jazz90 and Jazz167 were found to act via an angiogenic marker (VEGFR-2), which predicted cell number with adjusted  $R^2$  values of 0.99 and 0.84, respectively. Based on the results from MVA, further studies in PC3 cells can be used to understand the role of VEGF-A and VEGFR-2 in the proliferation of PC3 cells, as the role of these markers in proliferation remains unclear. MVAs also suggested that the combination of Jazz90 with EMT inhibitors, such as olaparib and simvastatin, may be useful, whereas Jazz167 can be tested in combination with bcl-2 inhibitors, such as venetoclax or navitoclax.

## Author Contributions

Project conception: RR; Collection of original data: ZR; Formal statistical analyses: ZR and PFS; Writing of manuscript: ZR and PFS; Manuscript editing: ZR, PFS and RR; Approval of final manuscript: all authors.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author on request.

## REFERENCES

- De Marchi T, Liu NQ, Stingl C, et al. 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Mol Oncol*. 2016;10:24-39.
- Martey ONK, Greish K, Smith PF, Rosengren RJ. A multivariate statistical analysis of the effects of styrene maleic acid encapsulated RL71 in a xenograft model of triple negative breast cancer. *J Biol Methods*. 2019;6:e121.
- Rana Z, Tyndall JDA, Hanif M, Hartinger CG, Rosengren RJ. Cytostatic action of novel histone deacetylase inhibitors in androgen receptor-null prostate cancer cells. *Pharmaceuticals*. 2021;14:103.
- Rana Z, Diermier SP, Walsh F, Hanif MG, Hartinger CJ, Rosengren R. Anti-proliferative, anti-angiogenic and safety profiles of novel HDAC inhibitors for the treatment of metastatic castration resistant prostate cancer. *Pharmaceuticals*. 2021;14:1020.
- Seto E, Yoshida M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol*. 2014;6:a018713.
- Bouchain G, Delorme D. Novel hydroxamate and anilide derivatives as potent histone deacetylase inhibitors: synthesis and antiproliferative evaluation. *Curr Med Chem*. 2003;10:2359-2372.
- Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med*. 2016;35:1159-1177.
- Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. Evaluation of the lasso and the elastic net in genome-wide association studies. *Front Genet*. 2013;4:270. <http://journal.frontiersin.org/article/10.3389/fgene.2013.00270/abstract>
- Wang H, Yang F, Luo Z. An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics*. 2016;17:60.
- Breiman L. Random Forests. *Mach Learn*. 2001;45:5-32.
- Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer; 2009.
- Williams G. *Data Mining with Rattle and R*. Springer; 2011. <http://link.springer.com/10.1007/978-1-4419-9890-3>
- Smith PF, Zheng Y. Applications of multivariate statistical and data mining analyses to the search for biomarkers of sensorineural hearing loss, tinnitus, and vestibular dysfunction. *Front Neurol*. 2021;12:627294.
- Wang L, Ma C, Wipf P, Xie XQ. Linear and nonlinear support vector machine for the classification of human 5-HT1A ligand functionality. *Mol Inform*. 2012;31:85-95.
- Manly BFJ. *Multivariate Statistical Methods: A Primer*. Chapman and Hall; 1986.
- Hanif M, Arshad J, Astin JW, et al. A multitargeted approach: organorhodium anticancer agent based on vorinostat as a potent histone deacetylase inhibitor. *Angew Chem Int Ed* 2020;59:14609-14614.
- Hrgovic I, Doll M, Pinter A, Kaufmann R, Kippenberger S, Meissner M. Histone deacetylase inhibitors interfere with angiogenesis by decreasing endothelial VEGFR-2 protein half-life in part via a VE-cadherin-dependent mechanism. *Exp Dermatol*. 2017;26:194-201.
- Melegh Z, Oltean S. Targeting angiogenesis in prostate cancer. *Int J Mol Sci*. 2019;20:2676.
- Abbas A, Gupta S. The role of histone deacetylases in prostate cancer. *Epi-genetics*. 2008;3:300-309.
- Cang S, Feng J, Konno S, et al. Deficient histone acetylation and excessive deacetylase activity as epigenomic marks of prostate cancer cells. *Int J Oncol*. 2009;35:1417-1422.
- Fizazi K, Martinez LA, Sikes CR, et al. The association of p21(WAF-1/CIP1) with progression to androgen-independent prostate cancer. *Clin Cancer Res*. 2002;8:775-781.
- Besson A. p27Kip1 modulates cell migration through the regulation of RhoA activation. *Genes Dev*. 2004;18:862-876.
- Nagano H, Tomida C, Yamagishi N, Teshima-Kondo S. VEGFR-1 regulates EGF-R to promote proliferation in colon cancer cells. *Int J Mol Sci*. 2019;20:5608.
- Peach C, Mignone V, Arruda M, et al. Molecular pharmacology of VEGF-A isoforms: binding and signalling at VEGFR2. *Int J Mol Sci*. 2018;19:1264.
- Sharma SS, Pledger WJ. The non-canonical functions of p27 Kip1 in normal and tumor biology. *Cell Cycle*. 2016;15:1189-1201.
- Bozic I, Reiter JG, Allen B, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife*. 2013;2:e00747.
- Zhang X, Zhang Z, Chen S, et al. Prognostic significance of E-cadherin expression in prostatic carcinoma: a protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. 2020;99:e19707.
- Taftachi R, Ayhan A, Ekici S, Ergen A, Ozen H. Proliferating-cell nuclear antigen (PCNA) as an independent prognostic marker in patients after prostatectomy: a comparison of PCNA and Ki-67. *BJU Int*. 2005;95:650-654.
- Ramesh V, Brabletz T, Ceppi P. Targeting EMT in cancer with repurposed metabolic inhibitors. *Trends Cancer*. 2020;6:942-950.
- Perini GF, Ribeiro GN, Pinto Neto JV, Campos LT, Hamerschlag N. BCL-2 as therapeutic target for hematological malignancies. *J Hematol Oncol Hematol Oncol*. 2018;11:65.
- Asakura N, Nakamura N, Muroi A, et al. Expression of cancer stem cell markers EpCAM and CD90 is correlated with anti- and pro-oncogenic EphA2 signaling in hepatocellular carcinoma. *Int J Mol Sci*. 2021;22:8652.
- Tao W, Tian G, Xu S, Li J, Zhang Z, Li J. NAT10 as a potential prognostic biomarker and therapeutic target for HNSCC. *Cancer Cell Int*. 2021;21:413.
- Qiu S, Weng Y, Li Y, et al. Raman profile alterations of irradiated human nasopharyngeal cancer cells detected with laser tweezer Raman spectroscopy. *RSC Adv*. 2020;10:14368-14373.