PharmacoVis: A personal cancer treatment recommendation tool

Abstract—Precision oncology, an upcoming field assigning personalized treatments based on specific genomic mutations or molecular alterations in patient tumors, has enjoyed an increasing interest in the research field of oncology. Machine learning techniques can be used to predict the effectiveness of drug therapies based on the omic profiles describing the genomic, transcriptomic or proteomic expression in cancer cells. However, leveraging the increasingly large availability of pharmacogenomic datasets and drug response prediction models has proved challenging. To address this issue, PharmacoVis, a visualization tool combining an integrated drug sensitivity database and a predictive model based on transcriptomic gene expression data, is developed. PharmacoVis facilitates drug recommendations based on patient specific tumor gene expression profiles and analysis of drugs in a large drug screening database.

Index Terms—Precision oncology, drug sensitivity visualization, cancer drug screening, drug response prediction, personalized cancer treatment

1 Introduction

In contemporary medicine, assigning appropriate treatments to cancer patients remains an important challenge. Many risks are associated with common current treatments, including damage to healthy tissue by radiation or cytotoxic agents [5]. These risks are results of cancer treatments not exclusively affecting tumor cells. Cytotoxic chemotherapy, for instance, administers drugs that kill excessively proliferating cells [8] [7] [4]. Not surprisingly, tissues with higher proliferation rates are vulnerable to such therapeutic approaches.

Cancer cells evolve from healthy cells through an accumulation of genomic alterations [6]. In order to make drugs which are more specific to cancerous cells then cytotoxic agents that target excessive cell replication, molecularly targeted agents are being developed which inhibit the proteins that are abnormally activated as a result of somatic genetic alterations [10].

The introduction of these novel targeted agents, combined with the increasing availability of patient profiling resources, allows for a new approach to matching cancer patients with drug treatments. Next generation sequencing methods have led to the development of genome-wide genomic profiling methods [9] [13]. The genomic information on cancer patients available through these methods and clinical data about (targeted) treatment outcomes and cancer progression provide new therapeutic options. *Precision oncology* seeks to find molecularly-targeted treatments for cancer patient sub-types and individual patients [1].

To bridge the gap between physicians looking to find the best treatment for a (cancer) patient and drug screening studies and drug sensitivity prediction models, PharmacoVis is developed. The visualization tool provides an interface to recommend effective drugs and find cancer cell lines with similar genomic expression profiles for a patient specific profile. Furthermore, it allows visual access to the large consensus pharmacogenetic database PharmacoDB. This database contains drug sensitivity information on a large number of cancer cell lines over which PharmacoVis facilitates visual exploration [12]. The exploration tool extends the predictive functionality of PharmacoVis with visual analytics to validate drug recommendations.

Note that the data sources and models used to develop PharmacoVis are by no means complete. They can (and should) be extended with new omic-profiles, better prediction models and more extensive databases for PharmacoVis to become a mature clinical tool. PharmacoVis serves as a proof of concept and as an extensible framework to be used in the development of future drug recommendation tools.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2 Problem Description and Task Analysis

The development of innovative precision oncology models, predicting drug sensitivity for cancer cell lines based on genomic, transcriptomic and proteomic profiles, enables scientists to quantitatively determine which treatment is expected to be most effective for new cancer profiles [2]. This provides a method for recommending drugs based on patient specific tumor cell profiles. Despite the potential of precision oncology applications, a concrete instrument to predict and simultaneously analyse drug responses for unique patient tumor profiles is not available.

2.1 Problem description

PharmacoVis is developed to address this issue and unites two oncology research fields in one intuitive visual tool: Drug response prediction and cell line drug sensitivity analysis. Hence, the problem tackled by the tool is also twofold.

2.1.1 Drug recommendation

In order to recommend a drug to a particular patient, PharmacoVis can be used to predict tumor cell drug sensitivity based on its transcriptomic gene expression profile. This is done by training a machine learning model to predict the drug response of a cell line based on the transcriptomic profile corresponding to that cell line. The machine learning model employed in the PharmacoVis tool is learned from the transcriptomic data published for a prediction challenge in 2014. This challenge, organized by the Dialogue for Reverse Engineering Assessment and Methods (DREAM) and National Cancer Institute (NCI), had the goal of identifying and benchmarking effective drug sensitivity prediction models [3]. Note that the model used for prediction in PharmacoVis is not the advanced model which won the challenge, but a simple model designed as a proof of concept.

PharmacoVis visualizes the drug response prediction of the drugs in the NCI challenge data and provides a treatment recommendation. The recommended drugs can than be used as a starting point for further study in the tool.

2.1.2 Cell line and drug sensitivity analysis

After a drug recommendation is made, PharmacoVis can provide insight in the recommended treatment and cell lines which are similar to the one for which the recommendation was made. The interactive visualisation interface allows the user to explore cancer cell lines, drugs and their associated combined drug sensitivity experiment results.

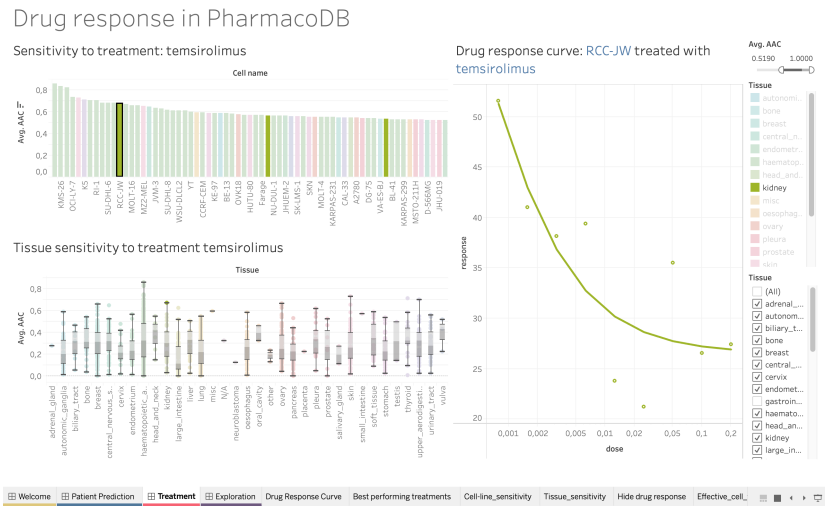


Fig. 1: The treatment exploration dashboard in pharmacoVis. This instance displays the drug response sensitivity to treatment *temsirolimus* with the specific drug response curve for cell line RCC-JW.

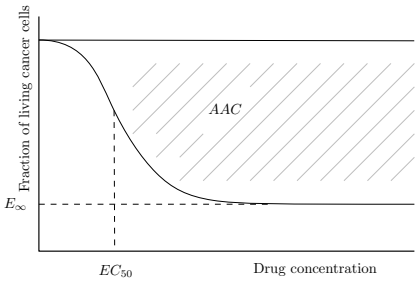


Fig. 2: The hill-slope curve modelling drug response

The analysis of similarity is performed by looking for similar transcriptomic profiles which are similar to the profile of the patient tumor cells in the combined NCI challenge and PharmacoDB data. Analysis of individual cell lines and drugs is done using data in the curated PharmacoDB database. Furthermore, the visualizations of the interaction between a cell line and drug originate from the drug sensitivity data in the pharmacoDB database.

2.2 Pharmacogenomic data description

Integrating both the transcriptomic profile data from the NCI challenge and the drug sensitivity data from PharmacoDB allows PharmacoVis to simultaneously provide visual analytics for drug recommendation and drug validation. The transcriptomic data making up transcriptomic profiles used for the NCI challenge is expressed as the *Transcripts per million* format and is an estimation of the relative molar concentration (the fraction of the sum of all mRNA concentrations corresponding to one gene):

$$TPM = \frac{r_g r l 10^6}{f l_g T}. \tag{1}$$

Here, r_g is the number of reads of a particular gene region g, r_l is the average number of nucleotides mapped per read, fl is the feature length (the mappable region of gene g) and T is the total number of transcripts sampled in a sequencing run for all genes G:

$$T = \sum_{g \in G} \frac{r_g r l}{f l_g}. \tag{2}$$

Drug sensitivity for a cell line is expressed in the pharmacoDB dataset using the *Area Above the Curve* metric (AAC). The drug response curve is used to compute this metric and is calculated by fitting a Hill Slope curve to the observed measurements for each dataset. The model fitting the fraction $y(x)$ of susceptible cancer cells killed by a concentration x of the drug describing the drug response curve is given by the following equation:

$$y(x) = E_\infty + \frac{1 - E_\infty}{1 + (\frac{x}{EC_{50}})^{HS}}. \tag{3}$$

The drug concentration at which half of the cells are alive is EC_{50}, such that $y(EC_{50}) = \frac{1}{2}$. HS denotes the parameter describing the cooperativity of binding and E_∞ is the fraction of cancer cells not at all susceptible to the drug. In figure 2 the metric AAC and the relation to the drug response curve is depicted.

2.3 The dataset

The pharmacoDB dataset is freely available for academic use here and consists of various tables. The tables that we used are shown below, together with the attributes in that table that we used.

- Experiments: Every row in this table represent an experiment that is done on a specific cell with a specific drug. This is the table that connects all the used tables together, every row in the experiments has an id of the cells, drugs, dataset and tissue. the table doesn't have any other attributes. The results of the experiments available in this table make up the drug response/susceptibility measures used in PharmacoVis. The results of 650.894 experiments are stored in the dataset.

- Cells: This table has every cell-line on which an experiment is done in it. Every cell is linked to one tissue via a tissue-id attribute. The other used attribute in this table is the cell_name attribute, which obviously holds the cell_ name of that particular cell-line. PharmacoDB contains drug response measurements on 1691 cancer cell lines.

- Drugs: This table is used as a mapping from the drug_id to the drug name. In total, susceptibility to 759 treatments is stored.

- Datasets: This table is used as a mapping from dataset_id to the dataset name. This dataset is the source dataset for which an experiment is conducted. PharmacoDB integrates 7 datasets.

- Dose_repsone: This table has a row for every experiment which contains the dose of the drug that is used and the response on the cell-line.

- profiles: has a row for every experiment containing a lot of information about the drug-response curve for that cell-line, drug pair that is tested in that experiment. The used values are E_∞, EC_{50} and HS. These profiles are modelled to fit the experiment results available in the *experiments* table.

The total pharmacoDB database is about 2.7GB. However, we do not use all the tables, restricting our analysis to tables relevant to drug response analysis with a total size of approximately 400MB. The large unused part of the pharmacoDB database contains genomic information on the cancer cell lines available in the database.

The NCI challenge data, available here, contains two csv files for 10 tissue types. For each tissue one file with the drug susceptibility measures for the cell lines belonging to this tissue and one file containing the transcriptomic for these cell lines. In developing the drug sensitivity prediction tool, transcriptomic profiles for 594 cell lines and treatment susceptibility measures for 192 drugs from the NCI challenge data are used, divided over the 10 tissue types. These transcriptomic profiles and drug response files together take up approximately 110 MB.

2.4 Task analysis

PharmacoVis is designed to be used by physicians to assist them in prescribing appropriate cancer treatments. Hypothetically, the physician would acquire a sample from the patient tumor tissue, sequence this sample and obtain a transcriptomic gene expression profile of the patient tumor cells. He/She could then turn to PharmacoVis to formulate a recommended pharmacological treatment for the patient specific tumor. First a recommendation based on the transcriptomic profiles would be provided, but the physician could study this and similar drugs further. By selecting a drug or cell line, more information about this object is visualized. An oncologist could use his or her domain knowledge to direct this analysis of drug responses and derive a suitable treatment.

Three tasks a physician might undertake to conclude his or her analysis are defined below. These use cases are used in designing PharmacoVis and provide a method of evaluation for this tool.

2.4.1 Tasks 1: Find a recommended drug

The starting point of an analysis will be a recommended drug, based on the transcriptomic profile provided by the physician. The tool provides an overview of possible treatments and shows the patient tumors expected sensitivity to these drugs. Here, the user can select a next topic of study, such as a recommended drug or cell lines similar to the patient profile. Questions that would be answered by this task is: Which drug is recommended for this patient?

2.4.2 Task 2: explore drug sensitivity

The physician might be interested in compare for a specific drug how sensitive it is against certain cell-lines. Or the other way around, compare how sensitive a cell-line is to different drugs. After comparing different drugs a physician might be interested in exploring what other effects a certain drug has against different cell-lines. Question that could be answered by this task could be: To what drug is this cell-line the most sensitive? or: To which alternative cell-lines is this drug effective?

2.4.3 Task 3: Validate drug recommendation

When a recommended drug has been identified, the oncologist is likely interested in knowing more about this treatment. PharmacoVis allows the user to select drugs to study which cell lines are typically treated with this drug. Furthermore, PharmacoVis can be used to show the experimental drug sensitivity of a specific cell line and drug as available in PharmacoDB. A question that could be answered by this task is: How effective will this drug be for the Patient?

3 Visualization Design

This section describes the visualization design of PharmacoVis. The visual analysis tool consists of the three dashboards and a landing screen. Section 3.1 describes the drug recommendation dashboard, Section 3.2 explains the treatment exploration dashboard and Section 3.3 describes the dashboard for exploring cell-lines. When starting his/her analysis a physician will start at a landing screen. From this starting point a patient and recommendation model can be selected to compute the recommended treatment. However, when the user is not interested in a particular patient, he/she can select a drug or cell-line and start with the treatment sensitivity or cell-line sensitivity dashboard respectively. Figure 3 shows the landing dashboard.

3.1 recommendation

To aid oncologists in finding suitable treatments PharmacoVis employs a drug response prediction model and based on the results of this prediction a recommendation is made. This section describes how PharmacoDB implements this prediction model and how it subsequently uses the model to visualize the recommended treatments. Figure 4 depicts the dashboard visualizing the treatment recommendation.

3.1.1 Drug recommendation implementation

When the user has selected a target patient and drug prediction model, PharmacoVis will load the recommended treatments based on the prediction model in the recommendation dashboard. PharmacoVis connects to a remote python server which is used to load both the patient transcriptomic profile and treatment response prediction model. Inside this python environment the expected drug response values for the patient are computed and passed back to the recommendation dashboard.

In the current version of PharmacoVis 5 test patient profiles and one model are available as a proof of concept. The 5 test profiles are randomly selected from the NCI challenge data and not used to train the model used to predict drug responses. The prediction model currently available in PharmacoVis is a simple linear regression model. However, the python script retrieving the predicted drug response values does so for an arbitrary patient and model identifier. This makes it easy for any user of PharmacoVis to develop the their own recommendation model and use this tool to visualize the drug response predicted by their model. Hence, as long as the drugs for which susceptibility prediction is trained are available in PharmacoDB, the drug recommendation functionality is easily extend with more advanced prediction models. Furthermore, this allows the user to deploy these recommendations on new (unseen) patient profiles.

3.1.2 Visual Encoding drug recommendation

The design of the visual encoding for the drug recommendation dashboard is limited to the nature of the attributes obtained from the python script performing the drug response prediction for the drugs in pharmacoDB. Since simpler models, such as the linear regression model used in this proof of concept, predict a single numerical AAC value for the drugs it was trained on, the visual design encodes this numerical value together with the categorical drug name attribute.

A physician using PharmacoVis is interested in the most effective treatment for their patient, he/she is interested in the hierarchical ordering of the predicted AAC values. To encode this a bar chart plotting the predicted AAC measure of drug response against the categorical drug name is used. This visualisation is sorted in descending order, making it easy for users to identify which drugs are most likely to be effective. Furthermore the bar chart clearly shows the expected AAC for these drugs. In Figure 4 this drug sensitivity bar chart is shown on the left of the recommendation dashboard.

Before investigating the expected most effective drug further with the drug sensitivity data in PharmacoDB, the user is likely interested in some general information on this treatment. In the recommendation dashboard PharmacoVis is able to search PubChem, a comprehensive database containing chemical information on a

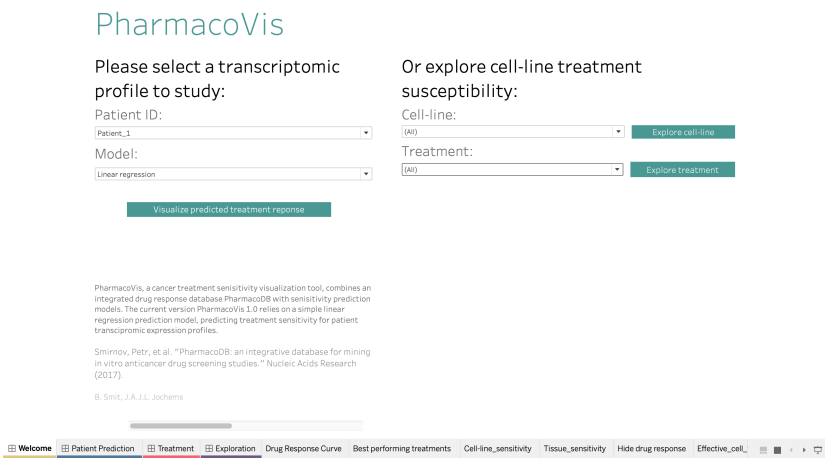


Fig. 3: The welcome dashboard of pharmacoVis. On the left, the user can load a patient transcriptomic profile and prediction model to find recommended drugs. On the right, the user can either choose a cell-line or drug to visualize its drug sensitivity information from pharmacoDB.

wide range of compounds including cancer drugs. The integrated PubChem browser, depicted on the right side of the recommendation dashboard in Figure 4, allows the user quick access to the basic characteristics of the cancer treatments likely to be effective for their patient. Futhermore, the PubChem integration provides a method of zooming in on the treatment chemical structure, synonyms, workings and detailed description.

As soon as the user navigates to the recommendation dashboard, the treatment predicted to be most effective is loaded in the pubchem browser. However, the user is free to interact with the bar chart. Hovering over a bar displays its full name and predicted drug response value and selecting a bar in the chart loads the chemical information in the PubChem browser. Finally, with a drug selected in the bar chart, the drug recommendation dashboard the user can use the "Explore treatment response" button to navigate to the treatment exploration dashboard explained in Section 3.2. This allows the user to zoom in further on the drug response data available in PharmacoDB and compare the expected treatment response for the patient with the empirical drug response on the well studied cancer cell lines.

3.2 Treatment exploration

This dashboard shows three different charts, namely: cell-line sensitivity, drug response and tissue sensitivity. Figure 1 shows this dashboard.

3.2.1 Cell-line sensitivity

The chart in the top left of the dashboard is the cell-line sensitivity graph. This chart shows the drug sensitivity of a certain drug on different cell-lines. The drug for which the sensitivity is shown is determined by which drug was selected in the prediction dashboard (Section 3.1). The shown AAC value is an average of all the experiments done on that Drug cell-line combination. To effectively display both this quantitave and categorical attribute against eachother, PharmacoVis uses a bar chart. Note that there are a lot of different cells for which PharmacoVis want to show the AAC value and a bar chart is an appropriate way to depict the value for a lot of different keys. Replacing the bars by points was also experimented with. However, in that case, every point represents a different experiment. This results in an unreadable design as there are too many cells for which the AAC value is shown. Furthermore, the dataset that we used contains only for a very small number of drug, cell-line pairs results from multiple experiments.

Selecting one cell in this chart will trigger the drug response graph to switch to show the drug response graph for that cell. It will also highlight that cell in the tissue sensitivity graph. Note that it can be the case that points for that cell are not shown in the tissue sensitivity chart. This is due to the fact that the tissue to which that cell-line belongs can be filtered out by the tissue filter on the chart.

The color of a cell in this chart represents the tissue to which it is linked in the data.

By default, a lot of cell lines are present in the visualization, it can be hard to look through the different cells in the chart. PharmacoVis allows the user to filter the chart by the AAC value, improving the readability of the chart or showing a more general overview of the drug sensitivity results. In the filter you can specify a range where the AAC value needs to be in to be showed in the chart. The large number of cells also made it hard to see which cell name belongs to which bar. This is made easier to see by hovering over a certain bar, this will show which cell is represented by that bar and the exact average AAC value.

This chart can be used to Explore and compare how sensitive all the cells for a certain drug. It can also be useful for validating that a certain drug is effective for a certain patient.

3.2.2 drug-response

The chart on the right of the treatment response dashboard is the cell line specific drug-response visualization. This chart shows the drug response for a specific drug on a particular cell-line. The drug that is shown is determined by which drug was selected on the prediction dashboard. The cell-line that is shown is determined by the cell-line that is selected on the cell-line sensitivity chart, which is on the same dashboard. The drug-response chart will only show if a cell-line is selected in the cell-line sensitivity chart. The points in the chart represent data points which contain the response of a specific drug on a specific cell for a specific dose. The points are used to determine a hill slope curve which is calculated as described in Section 2.2. The experiments in pharmacoDB, providing a dose and response combination (both numerical attributes), are trivially visualized with a scatterplot. However, only showing the points in the chart does not give enough information about the relation between the dose and response. This is partially due to the small number of datapoints per drug, cell-line pair. To aid physicians in determining what dose of the selected drug will have the desired effect on the cancerpatients, we decided to add a trendline to the chart. Modelling drug responses in pharmaceutics is typically performed using a hill-slope curve and this is implemented in pharmacoDB. However, different types of trendlines (see Figure 7) were tested.

It can be the case that multiple trendlines are shown in the chart. This

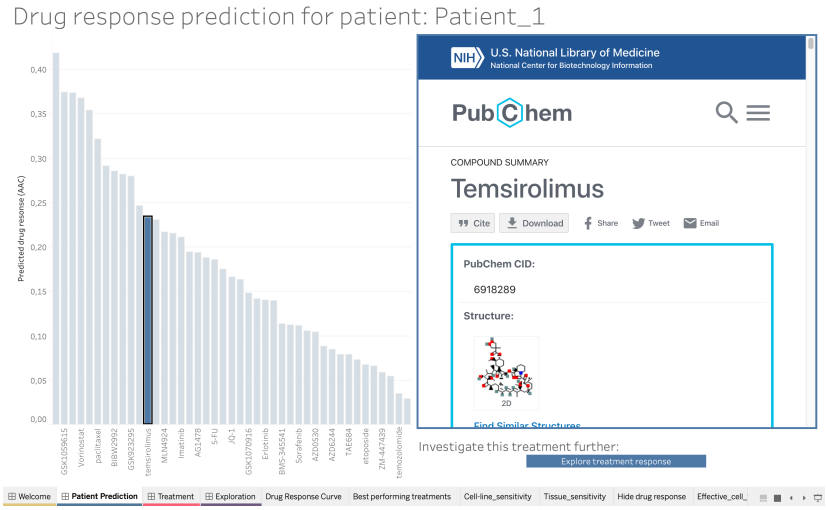
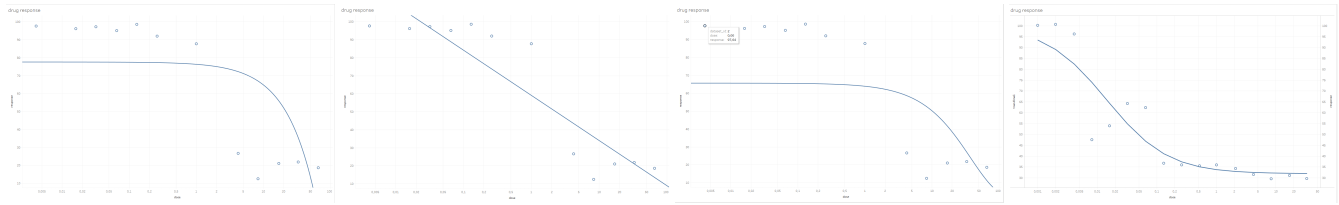


Fig. 4: The PharmacoVis drug recommendation dashboard displaying recommended treatments for test patient 1. By selecting the treatment *temsirolimus* the pubchem compound summary is displayed on the right.



(a) Linear (b) Logarithmic (c) Exponential (d) Hill slope curve

Fig. 5: Different types of trendlines for the drug response chart. Note that the scale on the x-axis is logarithmic.

is the case when there are datapoints from different datasets for this particular drug, cell-line pair. In this case the dataset to which the datapoint belongs can be determined by the color of the datapoint/trendline.

Hovering over a datapoint in the chart shows the exact dose and response value for that point. This chart can be used to how a patient could react to a the dose of a particular drug on a particular cell-line. This could then help to determine the dose of a particular drug that the patient should use. Obviously, without a selected cell-line there is no corresponding drug response curve available. Hence, so long no cell-line is selected, no drug response curve is displayed. Picking the cell with the highest sensitivity to that drug and displaying the drug-response curves for this combination was also considered, but that could lead to misinterpretation of the user as he would not know exactly what is shown in this graph.

3.2.3 Tissue sensitivity

The chart on the bottom left of the treatment sensitivity dashboard is the tissue sensitivity chart. This chart shows the sensitivity of a particular drug against different tissues. The tissue for which it is shown is determined by the drug that was selected on the prediction dashboard. As this is also a categorical attribute against a quantitative attribute a box plot should be a good way to represent this. In this case the average AAC would be shown for every tissue. However, in this case that gives not enough information as the different AAC values for a particular tissue are from a lot of differnt cell-lines. So, there is a big difference in AAC values. A way to visualize this is to show a point for every experiment done for a particular tissue. This already gives a better visualization of how the tissue responds to the drug. However, It is sometimes the case that a lot of points have a fairly similar AAC value so they will be close together in the chart. If there are too many close to each other, it can be hard to show how many are in that range. A way to visualize this is to show a box plot in the visualization. These boxplots provide an intuition about the distribution of drug response measures for treating a particular cancertype (corresponding to the tissue) with this treatment.

Hovering over a point in the chart shows the exact AAC value and which cell-line it belongs to. Selecting a cell will highlight the cell-line to which it responds in the cell-line sensitivity chart. Note that it can be the case that the cell-line is not visible in the cell-line sensitivity graph as it can be filtered out by the AAC filter.

With the number of tissues shown in the chart it can be hard to look at the box plot of a particular tissue due to it being to small. To help with this a filter in which the user can specify which tissues are shown in the chart is exposed. This will can make it easier to see the different points for every experiment for a specific tissue.

The color of the data points is determined by which tissue it belongs. In this chart it is not really necessary to have different colors per tissue because the different tissues are all on separate columns. However, to make it more consistent with the chart above this chart (the cell-line sensitivity) it is better to also give the same colors to the same tissues in this chart.

This chart can be used to summarize what the sensitivity of a particular drug is against different tissues while also showing the sensitivity against the different cells that belongs to this tissue.

3.3 Cell-line sensitivity

This dashboard consists of two charts, namely: drug sensitivity and drug-response. Figure 6 shows this dashboard

3.3.1 drug sensitivity

The chart on the left of this dashboard is the drug sensitivity chart. This chart shows the average sensitivity against a certain cell for all

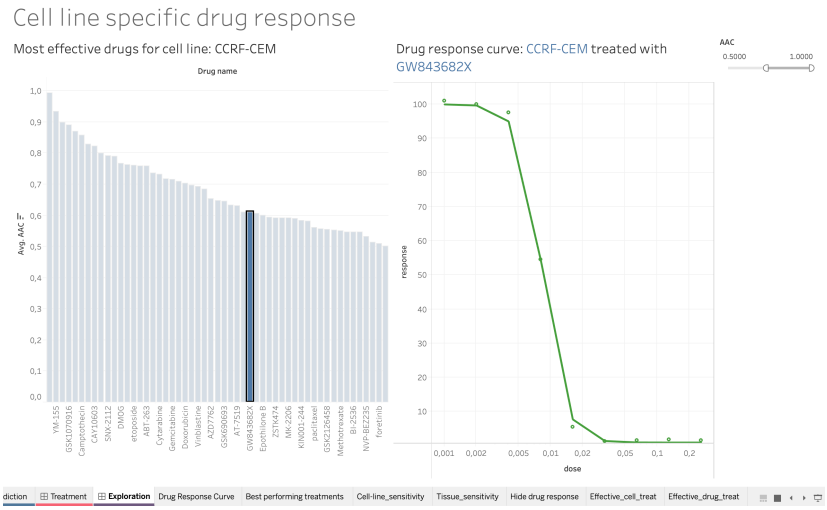


Fig. 6: The cell line sensitivity dashboard in PharmacoVis, displaying the most effective drugs for the cell line CCRF-CEM, and the drug response curve for the selected treatment GW843682X

the different drugs that are tested. This chart is similar to the cell-line sensitivity described in Section 3.2.1. The difference is that this chart is filtered to only show one cell-line and have the different drugs on the x-axis, while the cell-line sensitivity chart is filtered to show only one drug and the different cell-lines on the x-axis. The cell-line on which this chart is filtered is determined by the cell-line search box on the landing page or the selection on the previous dashboard. The reasoning behind using the bar chart as a visualization technique is the same as for the cell-line sensitivity chart. The purpose of this chart is to compare different drugs for that particular cell-line.

Selecting a drug in the chart will switch the drug-response chart on this dashboard to also show the drug-response for that drug on the cell-line selected on the landing page.

3.3.2 drug-response

This chart is exactly the same as the chart shown on the treatment exploration dashboard. The only difference is how the shown drug and cell-line are chosen. For the chart on this dashboard the cell-line is chosen by which cell-line was selected on the landing page. The drug is chosen by which drug is selected on the drug-sensitivity chart. The decision to display the drug response curve twice, is supported by the argument that physicians might want to iterate between treatments effective for their patient, cell-lines which are typically treated with this treatment and treatments which are known to be effective to these cell-lines. Having the drug response curve visible in both dashboards allows the physician to monitor the required dose for several different cell lines during this analysis.

4 Use Cases

In this section we evaluate the effectiveness of PharmacoVis to perform the tasks described in Section 2.4.

4.1 Find a recommended drug

PharmacoVis primarily is a tool to help oncologists identify effective treatments for patients based on their tumor transcriptomic profiles. Initially, a user will therefore be interested in a recommended drug. PharmacoVis intuitively ranks drugs on their predicted effectiveness in treating a patient. The user simply provides a patient identifier and selects a prediction model and is presented with the likely most effective drugs ranked on their expected drug response.

The accuracy of these predictions are however in the hands of the user. The basic linear regression model provided with PharmacoVis is not accurate enough to be used in practise. Furthermore, the recommendation only includes drugs which are used in training the prediction model. Hence for PharmacoVis to be used in a clinical setting a very powerful model needs to be developed.

In general, PharmacoVis effectively visualizes a drug recommendation and provides a treatment which is (according to the used model) most likely to be effective in treating the users patient. This treatment can be used as point for further study.

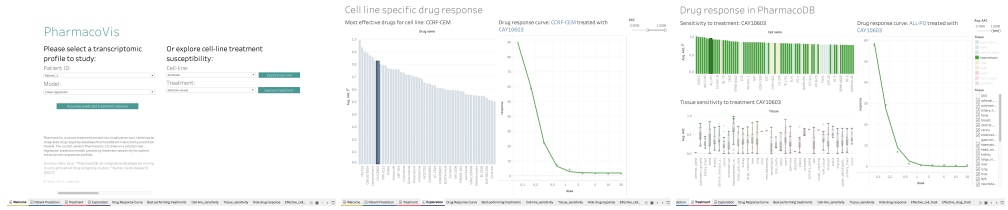
To derive this recommendation, the user specifies the target patient profile and a prediction model in the landing dashboard. This leads to the recommendation dashboard containing the expected drug response bar chart. This chart shows the drugs which can be used to treat the patient in descending order of expected susceptibility. This overview allows the user to select a specific drug study the chemical properties of this treatment. When the user is indeed interested in studying this treatment, this dashboard allows him/her to zoom in on the drug response data in PharmacoDB by navigating to the treatment exploration dashboard for the selected compound.

4.2 Validate drug recommendation

In section 4.1 the analysis a user might perform to find recommended treatment for their patient. When a drug is selected the physician is likely interested in more information on this treatment. Next to the chemical characteristics already available in the recommendation dashboard, PharmacoVis can be used to validate the drug recommendation by analysing the drug response values in PharmacoDB.

To perform this validation of the drug recommendation, the user can navigate from the recommendation dashboard to the treatment exploration dashboard, by pressing the 'Explore treatment response' button. From this dashboard the user continues his/her analysis of the selected drug. This dashboard contains a bar chart plotting the drug response for the cell lines in PharmacoDB for which the treatment is most effective, a boxplot visualizing how effective the treatment is to cancers of specific tissues and, when a cell line is selected, the drug response curve for the treatment-cell line combination.

The user navigates to this page by selecting a treatment in the recommendation dashboard. For this selected treatment, the bar chart visualizing the drug responses for known cell lines provides the user with an overview of the cancer cells for which this drug is typically effective. In designing PharmacoVis two subtasks for the validation of recommended drugs were defined. The fist task, finding cell lines which are often treated successfully with a recommended drug, is intuitively possibly through this bar chart.



(a) Landing dashboard (b) Exploration dashboard (c) Treatment dashboard

Fig. 7: Screenshots of different steps in the explore drug sensitivity use case.

The second subtask, finding out how a typical cell line responds to a particular treatment, is more abstract and tackled with the drug response curve. The user can study the hill slope curve fitted through the dose response measurements for the experiments in PharmacoDB to see how much of the drug is required for the desired treatment effect.

An effective analysis of drug response values in PharmacoVis relies on the domain expertise of the user. The tool is able to combine both detailed experiment specific dose response values and a general comparison of susceptible cell lines. However, in order to validate the recommended drug, the user is required to select which cell line to study and be able to interpret the drug response curve. The hill slope curve, used to model the drug response is widely used and medical professionals will therefore have no problem with this method of validation. Literature review shows us this visualisation design of the dose and drug response interaction is most appropriate [11].

The box plot showing the distribution of response values for the cell lines belonging to various tissue types is not designed with a specific task in mind. However, it can be used to develop more intuition of the effectiveness of the selected treatment. For example, when the drug treats many tissue types effectively, it is likely a cytotoxic agent not targeting any cancer specifically. These differences of effectiveness across tissue types provide valuable information when considering a drug as a potential treatment during validation. The step by step approach to retrieving and validating a drug recommendation is shown in figure 8.

4.3 Explore drug sensitivity

On the landig dashboard (see Figure 8a) start by choosing a cell-line for which the user want to explore the drug sensitivity. Clicking on the confirm button will bring the user to the cell-line sensitivity dashboard with the specified cell-line already loaded in.

In this dashboard the user can see how sensitive this cell-line is against a some drugs. The user can start by sorting the drugs by descending order by pressing the sorting icon above the y-axis header named AAC. Now the user can explore the different drugs which have a high sensitivity. However, there are a lot of different drugs shown on this chart so it is hard to see the difference and because in this case we are mostly looking at drugs with high AAC values, so the chart is filtered on a AAC value between 0.5 and 1. Now if the user presses on one of the bars of one of the drugs, for example ”CAY10618” the drug-response for that cell-line drug pair is shown in the chart on the right. Figure 8b shows the resulting dashboard. In drug-response chart the user can look at what a good dose is to give in context of that particular cell-line. The user can continue switching to drugs to look at the drug-response chart.

When the treatment exploration dashboard is shown, it is possible to look at the effect on different cell-lines for the chosen drug. The dashboard also shows a summary of how the drug responds to certain tissues. A lot of cell-lines are shown in the bar chart so the shown AAC values can again be filtered to improve readability. From here it is possible to look at different cell-lines and how they respond to the chosen drug. Selecting a cell-line will again show the drug-response curve. The box-plot will show a summary of all the AAC values for all the cell-lines that belong to a tissue. From here the user continue looking at different cell-lines in this dashboard, or switch back to the cell-line sensitivity with the same cell-line loaded in by switching to that dashboard via the tabs at the bottom of the screen.

PharmacoVis provides an intuitive interface to visually study the drug response information in pharmacoDB. In this exploration the questions raised in section 2.4 can be answered.

5 Discussion

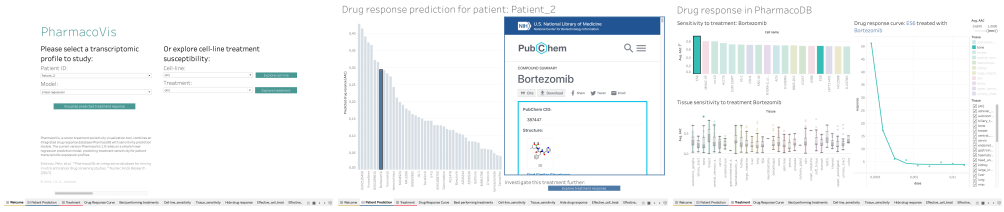
Overall, the resulting visualization is an effective visualization for the drug response data in PharmacoDB. However, in the future this visual pharmacogenomic analysis tool could be improved further. For example, the cell-line sensitivity dashboard provides a limited amount of information. Extending this dashboard could improve the tool would be step further in developing an clinical tool to predict and analyse patient treatments. To do so however, a new data source would need to be integrated. In the development of PharmacoVis cell line clustering was considered to extend the analysis from a treatment focused application. In practise however, the NCI dataset and PharmacoDB did not contain easily compatible transcriptomic profiles for cell lines making it impossible to visualize clusters of similar cell lines. So, to improve this visualization and extend the cell line dashboard in the future, a new data source with suitable expression profiles for the cell lines in pharmacoDB could be valuable.

During the development stage of PharmacoVis, a few changes to the visualization design have taken place. Initially, for each treatment (and cell line) it’s original data source was considered and used in the visual encoding. PharmacoDB, being an database containing experiments from several other data sources, records for every experiment from which source it originated. In the first visualization designs these data sources were color coding observations. However, realizing these data sources did not add relevant information to physicians, but are mostly maintained to distinguish batches of experiments, this extra visual encoding was removed from the design of PharmacoDB.

Another interesting addition to this visualization could be to include more details about the drug-response predictions. Examples of such details would be the probability of the prediction to be right or by visualizing the prediction with a confidence interval. For this to be possible, the model should be changed to a model which can also return such characteristics about the prediction. This would hence require more advanced prediction models and a more involved method of interpreting these prediction in the external python server.

For now it is only possible to predict the drug response by one linear regression model. However, it would be good if there were different and more advanced predictions models that the user could chose from. The tool is created in such a way that it is easy to add new and different models to be used in the application. Extending and refining the prediction models is an essential step, which needs to be performed before PharmacoVis can be considered a mature medical tool.

Currently, the drugs which are shown in the visualization are restricted to the drugs in the pharmacoDB database. It would be nice to obtain a visualisation for all drugs. However for this to happen a dataset which contain other drugs on which the same experiments are performed would have to be included. Of course, adding drugs and



(a) Landing dashboard (b) Recommendation (c) Treatment dashboard

Fig. 8: Steps a user can undertake in pharmacoVis to retrieve and validate a treatment. In this case the user used the linear regression model to find a recommended treatment for test patient 2, selected the *bortezomib* treatment and validated this with the drug response information of cell-line *ES6*.

drug response information to the database will always improve this tool.

In this version of PharmacoVis the only predictions that are made are for five randomly selected test patients. For the visualization to be useful it is needed that the physician uploads a transcriptomic gene expression data of a patient for which a treatment should be found. These patients profiles are locally stored as simple csv files, and automatically included in PharmacoVis. In the future PharmacoVis could also be used to extend the software used to sequence tumor cells. This would allow oncologists to automatically retrieve relevant drug response visualizations when sequencing a tumor sample.

Finally the effectiveness of this visualisation tool relies heavily on the domain knowledge of the user. The visualisation design assumes the user is capable of directing the analysis by investigating the right treatments and drugs. He or she is required to be familiar with treatment names and be able to interpret the chemical information provided by the PubChem treatment summaries. For example, simply selecting the drug which is expected to be most effective for the patient is not necessarily smart. The side effects of cancer treatments vary greatly and in deciding which treatment is best for a patient additional effect will always need to be considered. Specifically in the context of cytotoxic agents, caution is advised. PharmacoDB contains not only molecularly targeted agents, but also provided drug response values for cytotoxic agent. These cytotoxic agents (used in chemotherapy) are often effective for many patients but impact patient wellbeing significantly.

6 Conclusion

As the availability of drug prediction models rises, medical professionals do not yet have an appropriate method of using these advanced predictions in theirs in their diagnoses. PharmacoVis bridges the gap between the potential of data mining techniques used to predict the best treatment for cancer patients and their physicians who are still unable to efficiently interpret these novel models.

PharmacoVis contains several different dashboards containing different types of visualization techniques such as a bar chart, box-plot and a scatterplot with a fitted hill slope model. The different charts and dashboards are connected using interactions which allow the selecting different elements, resulting in either highlighted or filtered elements in different charts. While all visualizations contribute to the visual analysis of the drug response data in pharmacoDB and the pharmacogenomic prediction, the interactions are used to guide the user through this analysis.

With pharmacoVis it is possible to compare different treatments for patients based on specific tumor gene expression profiles. When a suitable candidate treatment has been determined it is possible to explore the other effects of that drug. PharmacoVis also gives a summary of what a good dose of drug would be for that patient and provides a summary of the treatment, its chemical structure and workings. Furthermore, pharmacovis can be used to compare the sensitivity of different treatments on a cell-lines from the pharmacoDB database. This aids the exploration of different treatment options on specific cell-lines.

With pharmacoVis users can answer can answer questions when treating a specific patient. It can be used to identify a suitable treatment. It provides insight in the (unwanted) effects of this treatment and suggest alternative treatments. For a particular drug pharmacoVis helps to answer for which tissue this drug is most effective. For a specific cell-line it is possible to show how susceptible it is for different drugs.

While pharmacoVis is successful in visualizing the drug response data in pharmacoDB and integrates prediction models to provide visual analytics used to provide personal treatment recommendations, it is far from a mature clinical application. Extending the drug response database and introducing more advanced and accurate prediction models is necessary to develop pharmacoVis further. The field of precision oncology is quickly developing and this study is hopeful visual analysis applications like PharmacoVis will help oncologists prescribe suitable treatments to cancerpatients in the future.

References

[1] M. Ali and T. Aittokallio. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophysical Reviews*, 2018.

[2] E. I. Andersson, S. Pützer, B. Yadav, O. Dufva, and S. Khan. Discovery of novel drug sensitivities in T-PLL by high-throughput ex vivo drug testing and mutation profiling. *Leukemia*, 2018.

[3] J. C. Costello, L. M. Heiser, E. Georgii, M. Gönen, and E. Al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nature Biotechnology*, 2014.

[4] C. Davies, H. Pan, J. Godwin, R. Gray, and E. Al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *The Lancet*, 2013.

[5] V. T. DeVita and E. Chu. A history of cancer chemotherapy, 2008.

[6] M. Gerlinger, A. J. Rowan, S. Horswell, J. Larkin, and E. Al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*, 2012.

[7] P. M. Hanno, S. Bruce Malkowicz, and A. J. Wein. *Penn clinical manual of urology*. 2007.

[8] A. Heeney, F. Hand, J. Bates, O. Mc Cormack, and K. Mealy. Surgical mortality - an analysis of all deaths within a general surgical department. *Surgeon*, 2014.

[9] T. J. Husdon. International Cancer Genome Consortium. *Cancer*, 2011.

[10] C. Le Tourneau, J. P. Delord, A. Gonçalves, C. Gavoille, C. Dubot, and E. Al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): A multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *The Lancet Oncology*, 2015.

[11] H. Prinz. Hill coefficients, dose–response curves and allosteric mechanisms. *J Chem Biology*, 2009.

[12] P. Smirnov, V. Kofia, A. Maru, M. Freeman, C. Ho, N. El-Hachem, G. A. Adam, W. Ba-Alawi, Z. Safikhani, and B. Haibe-Kains. PharmacoDB: An integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Research*, 2018.

[13] J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, and E. Al. The cancer genome atlas pan-cancer analysis project, 2013.