

‘TPWshiny’

User Manual

July, 2021

Version 1.0

Title: An interactive RShiny app to visualize and explore the genome-wide characterization of NCI-60 cell lines in response to cancer agents.

Authors: *Biometric Research Program, Division of Cancer Treatment and Diagnosis (DCTD), National Cancer Institute, Rockville, MD, USA*

Contact: TPW Support (ncitpwsupport@mail.nih.gov)

TPWshiny

TPWshiny is a powerful standalone visualization tool for the genome-wide characterization of NCI-60 human cancer cell lines and response to 15 different anticancer drugs at different time points.

The underlining data comes from the “NCI Transcriptional Pharmacodynamics Workbench” (NCI TPW), an extensive compilation of directly measured transcriptional responses to anticancer agents. The NCI TPW data is publicly available through a web interface that allows limited user interaction with the data (<https://tpwb.nci.nih.gov/>). We developed “TPWshiny” as a standalone, easy to install, R Shiny application to facilitate more interactive data exploration.

This package provides an intuitive interface to facilitate better understanding of tumor cell characteristics of nine different tissue types and help researchers further their investigations in this field. The data is presented in interactive scatter plots, heat maps, time series and Venn diagrams. All information can be queried by drug concentration, time point, gene and tissue type. Researchers can download the data for further analysis.

Table 1 and Table 2 outline the data that is available in this resource. For more details on the methods of data collection and background information, please refer to the original publication: Monks, A, et al. "The NCI Transcriptional Pharmacodynamics Workbench: a tool to examine dynamic expression profiling of therapeutic response in the NCI-60 cell line panel." *Cancer Research* (2018): canres-0989. <https://www.ncbi.nlm.nih.gov/pubmed/30355619>

Table 1. Dataset of TPWshiny.

Data Type	Number of data points
Anticancer Agent	Azacytidine, Bortezomib, Cisplatin, Dasatinib, Doxorubicin, Erlotinib, Geldanamycin, Gemcitabin, Lapatinib, Paclitaxel, Sirolimus, Sorafenib, Sunitinib, Topotecan
Concentration	Untreated, low concentration, high concentration
Time point	2hr, 6hr, 24hr
Number of gene expression arrays	7623
Number of genes per chip	12704
Number of GI50 measurements	866
Number of pathways	65
Number of transcription factors	121
Number of group of receptors	55
Expression profiling array	Affymetrix HTA array plates U133A

Table 2. Dosage used at high and low concentration for each drug.

Anticancer Agent	High concentration (nM)	Low concentration (nM)
Azacytidine	5000	1000
Bortezomib	100	10
Cisplatin	15000	3000
Dasatinib	2000	100
Doxorubicin	1000	100
Erlotinib	10000	1000

Geldanamycin	1000	100
Gemcitabin	2000	200
Lapatinib	10000	1000
Paclitaxel	100	10
Sirolimus	100	10
Sorafenib	10000	5000
Sunitinib	1000	200
Topotecan	1000	10
Vorinostat	5000	1000

User guide

Installation

This user-friendly visualization tool is freely available and easy to install. Users can download the ready to use, self-extracting package for Windows or macOS, and R source code from the project website (<https://brb.nci.nih.gov/TPWshiny/>).

Download options and installation instructions:

- WINDOWS operating system (includes R portable)
 - download the standalone TPWshiny for Windows;
 - run the self-extracting exe file to create a "TPWshiny" directory;
 - from the extracted TPWshiny folder, run the "TPWshiny.bat" file to start the application.
- MAC operating system (Intel)
 - download the standalone TPWshiny for Mac;
 - extract the zip file on your local computer. This will create an application directory called "TPWshiny";
 - in the newly created folder, double click the "TPWshiny.command" file. This should bring up the TPWshiny graphical interface.
 - For the first execution, the user may see a security pop-up. To allow TPWshiny to run, access the "Security & Privacy" menu under the "System preferences" and select the "General" tab to enable the option "Anywhere" in the section "Allow apps downloaded from".
 - If the "Anywhere" option is not visible, the user needs to disable the Gatekeeper by opening a command line terminal and run the command 'sudo spctl --master-disable'.
 - If the TPWshiny window appears empty/white, please close the app and run it a second time.
- R code (that can be launched directly within an R console or RStudio)
 - download the R code TPWshiny.R (requires R 3.6 or above)
 - install the 'shiny' package in R
 - to launch: set the working directory as the directory where you saved TPWshiny.R (using the command "setwd") and run the app with the command `shiny::runApp("TPWshiny.R")`
 - alternatively, in RStudio, you can simply click the "> Run App" button located at the top right corner of the RStudio interface
- TPWshiny has also been deployed on a shinyapp server.

You can access it directly at <https://ncibrp.shinyapps.io/TPWshiny/>

IMPORTANT NOTE: Some operating systems may not allow by default the installation of "unofficial apps" downloaded from the internet. This issue is not specific to the TPWshiny executable file, but it depends on the OS security requirements. If the TPWshiny executable file doesn't seem to run, please inspect the operating system security warnings and allow the TPWshiny application to be downloaded and executed in your system. This issue should only occur once, as OS security settings are saved on the user machine.

Data exploration features available in TPWshiny

This tool contains four different tabs to explore the data:

- **Single Gene:** to query the expression profile of a specific gene in the different cell lines treated with selected anti-cancer drug at a specific concentration and time point.
- **Geneset Patterns:** to query the expression profiles of group of genes (e.g., pathways, receptors) in the different cell lines treated with selected anti-cancer drug at a specific concentration and time point.
- **Time profiles (by drug):** to query the expression profiles a specific gene at the three different time point, across different drugs of interest selected by the user.
- **Common genes for sensitivity/resistance:** to query the expression of the 100 genes most correlated with GI50 with selected drugs, and understand how the lists of genes overlap between different drugs, in order to understand negatively/positively correlated genes that may identify common mechanisms of sensitivity/resistance.

The main interface is displayed in Figure 1.



Figure 1. TPWshiny main page. 1) The major data exploration tabs are accessible at the top. Functionalities within each tab are explained in the rest of this document. 2) the Left panel displays the main control options of the selected tab. Depending on the type of analysis, auto-filling dropdown boxes and check boxes are available for intuitive user interaction. 3) The Center panel displays the plots of the selected data sets. Multiple views of the same data are displayed in sub-tabs (e.g., with the “Single Gene” tab, different visualization options are Scatter plot, Bar plot, Timeplot and table data). The plots offer interactive capabilities (including zoom, tooltip, reference lines and save as image). Please refer to other parts of this document for more information. 4) The Right panel displays statistical information, sorting options and download capabilities depending on the type of plot. See below for additional details.

“Single Gene” tab

We designed different options to accommodate different needs of the users in exploring the NCI TPW data. The first query a researcher may want to explore is how the expression of a single gene of interest treated with specific anti-cancer drug changes in comparison to the baseline across different tissue types.

The “Single Gene” tab contains four different options to display this information: scatterplot, barplot, timeplot and the numerical data. Checkboxes, dropdowns and other buttons in the left and right areas allow the user to select specific parts of the dataset to be displayed in the dynamic plot.

All the functionalities of the different tabs are controlled by the selection of the gene and the drug in the left panel. By typing a gene symbol, (e.g., LIF) and selecting a drug from drop-down menu (e.g., Doxorubicin), the different tabs described below will display a series of graphs that show the expression of the selected gene LIF in different cell lines and different cancer categories.

In the first sub-tab (Scatterplot), the plot displays the log₂ fold change expression of the selected gene (on the vertical axis) vs. log₁₀GI50 of the selected drug (on horizontal axis). Cell lines from different tissues are displayed according to the selected checkboxes: each point represents an individual cell line and each color represents specific tissue (Table 3).

Table 3. Nine cancer categories of NCI 60 cell lines.

Tissue Type	Cell Line
Breast	BT-549, HS-578T, MCF7, MDA-MB-231, MDA-MB-468, T-47D
CNS	SF-268, SF-295, SF-539, SNB-19, SNB-75, U251
Colon	COLO-205, HCC-2998, HCT-116, HCT-15, HT29, KM12, SW-620
Leukemia	CCRF-CEM, HL-60, K-562, MOLT-4, RPMI-8226, SR
Lung	A549, EK VX, HOP-62, HOP-92, NCI-H226, NCI-H23, NCI-H322M, NCI-H460, NCI-H522
Melanoma	LOX, M14, MALME-3M, MDA-MB-435, SK-MEL-2, SK-MEL-28, SK-MEL-5, UACC-257, UACC-62
Ovarian	IGR-OV1, NCI-ADR-RES, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8, SK-OV-3
Prostate	DU-145, PC-3
Renal	786-0, A498, ACHN, CAKI-1, RXF-393, SN12C, TK-10, UO-31

In the scatter plot, points above the $y=0$ line denote increased gene expression while the points below the zero line denote reduced gene expression (relative to the same cell line untreated by the drug). Three correlations coefficients (Pearson, Spearman, and Kendall) are automatically calculated and displayed on the right of the selected dataset (Figure 2A).

From the left panel, the user can change the drug, its concentration, and time point to display.

Another parameter that can be visualized is the ‘Resistance Measure’ displayed in the scatterplot: the three available options are “GI50”, “Doubling time” or “Multiple Resistance”.

Finally, four options for the types of regression line to display are available in the right panel, in the “Regression type” drop-down menu (None, Linear, Polynomial and Loess). The confidence interval level can be customized using a slider with values ranging from 0 to 0.99.

In the second sub-tab (Barplot) the \log_2 gene expression fold change of the selected gene, drug, tissue, and drug concentration are displayed as bars. In this view, the bars extending to the right indicate increased gene expression, and the bars extending to the left reduced gene expression relative to cell lines untreated by the drug. Each bar represents a cell line, color coded by the cancer category (tissue type) (Figure 2B). Users have the options to select their drug of interest and gene. In order to identify common patterns, researchers can sort the expression based on tissue type, gene expression, $\log(\text{GI50})$, doubling time and multiple resistance (in increasing or decreasing order). Cancer categories of interest can be toggled on/off using the checkboxes on the left panel.

The third sub-tab (Timeplot) generates a plot showing the \log_2 fold change of expression of the selected gene and drug for all the three available time plots (Figure 2C). Each line represents a cell line, color coded by the cancer category. The X-axis shows the time point and the Y-axis shows the gene expression of the particular gene. Points above the $y=0$ line denote increased gene expression while the points below the zero line denote decreased gene expression at that particular time point relative to the respective untreated cell line. Using selection options in the right panel, users have the option to display only the expression of genes falling in specific quartiles.

Under the fourth sub-tab (Data), a table with all the numerical data points is displayed. The table has sorting options and it can be easily downloaded for further local analysis using the “Download” button. Users can select if to download the complete set of data for the selected gene/drug

combination, or only download the subset of tissue types selected in the previous graphical tabs (Figure 2D).

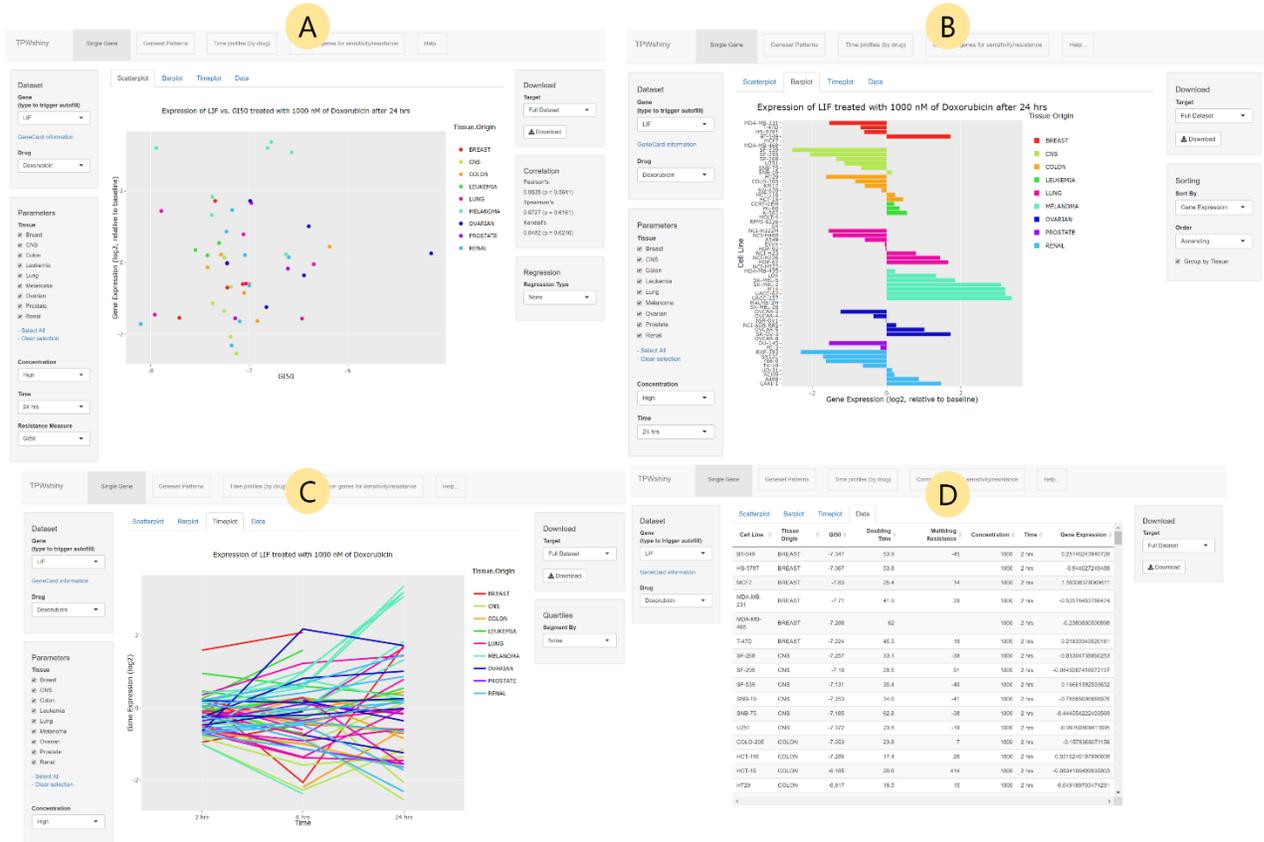


Figure 2. Different plots for the query gene LIF treated with the low concentration (1000 nM) of Doxorubicin at 24hr in 9 different cancer categories. Figure A shows the scatter plot of relative expression of a gene versus log(GI50). Figure B shows the bar plot of gene expression of LIF in different cell lines. Figure C displays the time plot and Figure D provides the dataset in a table format. The users have the options to download either the full dataset, or only the dataset displayed in the graphs of previous tabs (if only a subset of tissue types was selected).

“Geneset Patterns” tab

The analysis available under the “Geneset Patterns” tab allows the users to study how biologically relevant patterns are affected by the 15 drug treatments. In this tab, users have four different options of geneset lists for analysis (Pathway, Receptors, Transcription Factors, and User defined).

In the “Pathway” analysis, user can visualize the effect of a drug on different biological pathways in the form of colored heatmaps. The different cell lines are listed on the X-axis, while the Y-axis lists the genes involved in the selected Pathway (Figure 3). The genes and the cell lines are both clustered by default, but users have the option to change this selection using the dropdown menu in the right panel, next to the heatmap. The main color scheme used to display the gene expression values can be selected via a dropdown menu that offers several color options. Heatmaps have dynamic visualization features that allow users to move the mouse cursor over a data point and see its exact value. Zooming, toggle of spike lines and saving the current image are all features available from the dynamic menu displayed in the top-right corner (Figure 3). Each dynamic plot has a variety of buttons to interact with the displayed data (Table 4).

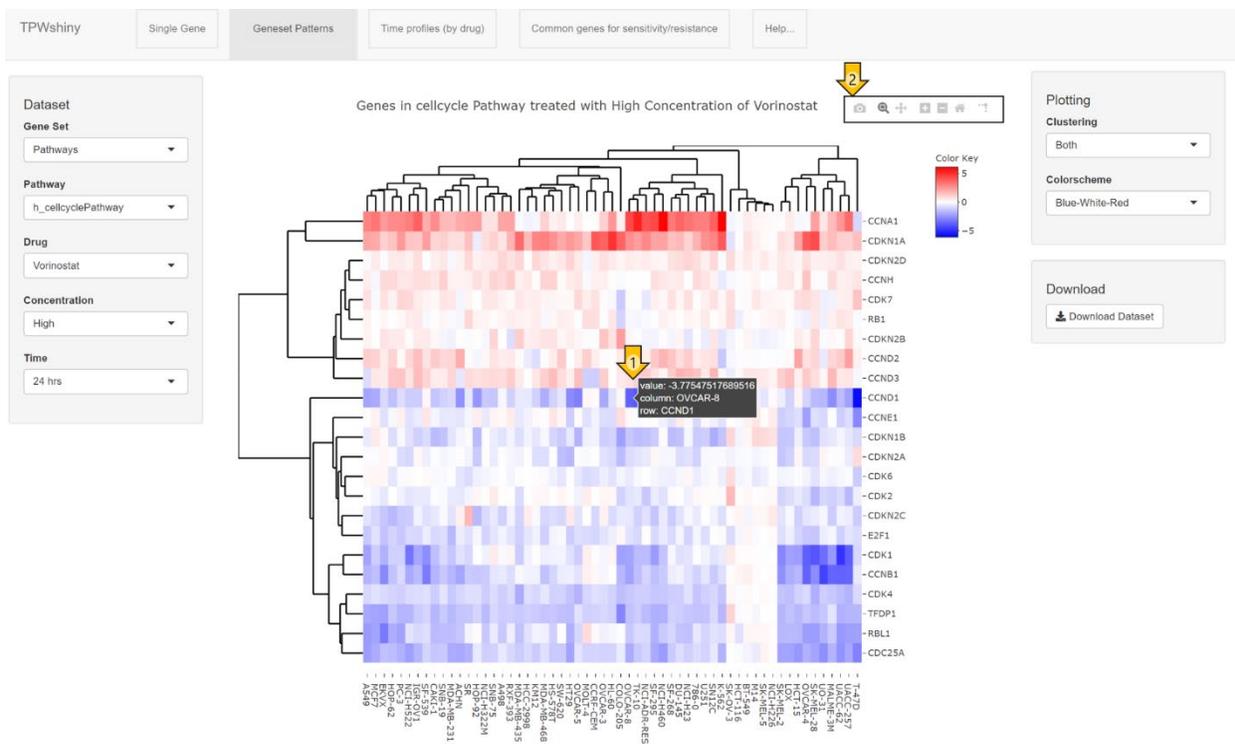


Figure 3. Heatmap of the effect of the high concentration of vorinostat on expression of the genes involved in h_cellcyclePathway at 24hr. 1) An arrow highlights how hovering over any graphical element in the heatmap brings up the tooltip with the information about that specific data point, as well as activating other control buttons (2) that allow researchers to dynamically interact with the plot by zooming, panning and saving the plot as image.

Table 4. A variety of buttons used throughout the dynamic plots to interact with the displayed graph.

The buttons appear in the right top corner of the plot area when the user moves the mouse cursor anywhere in the plot area.

Control button	Description
	Download figure in PNG format.
	Zoom in over exact data points.
	Drag the image to view the specific area. User can also pan horizontally to the left and right and pan vertically up and down while holding the mouse.
	Plot zoom in. Mouse selection will also allow zooming in.
	Plot zoom out.
	Auto scale all axes, returning them to their original formats.
	Reset all axes and return to the default visualization area.
	In some plots, show the closest data point where mouse hovers.
	In some plots, compare the closest data point where mouse hovers.
	Box selection button, used to select specific data points in a scatterplot.
	Lasso select button, used to select specific data points in a scatterplot.
	Spike lines to connect the selected data points to the axis.

All numerical data used to create these plots can be downloaded for further local analysis using the button in the right panel area.

2. Drug effects on expression of all transcription factors (Figure 5B). By selecting the drug of interest, the tool will generate heatmap of all transcription factors expressed in all cell lines.
3. Drug effects on gene expression of a specific transcription factor (Figure 5C). By selecting a specific transcription factor, the tool will generate the heatmap of expression of that transcription factor in all cell lines.

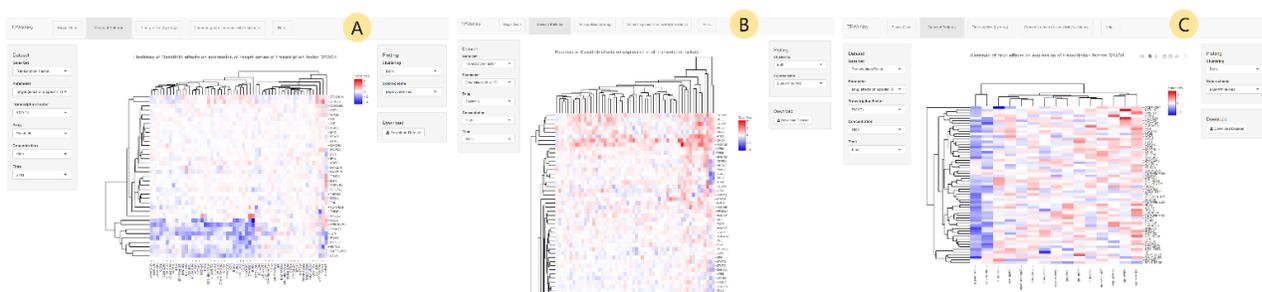


Figure 5. Effect of a drug on different transcription factors. Figure A shows the effect on the high concentration of dasatinib treatment at 6 hr on target genes of SMAD4, Figure B represents the effect of the high concentration of dasatinib at 6 hr t on expression of all transcription factors, and Figure C represents the effect of a high concentration of all the 15 drugs on the expression of the transcription factor SMAD4 at 6 hr in all cell lines.

By choosing a “User defined” list of genes, the tool allows the generation of a heatmap of any set of genes of interest. Users can either list the genes one at the time (taking advantage of the autofill capability of the text field) or can paste a space or comma separated set of gene names. When more than two gene symbols are entered, the corresponding heatmap will be displayed (Figure 6).

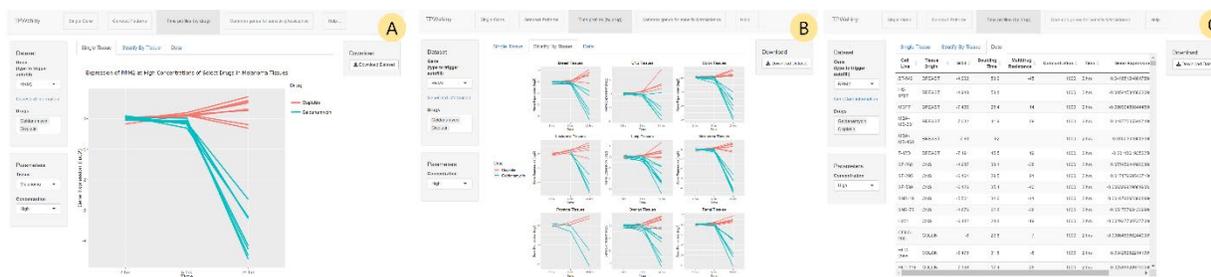


Figure 7. A) the effect of the high concentration of geldanamycin and cisplatin on the expression of the RRM2 gene in melanoma cell lines. B) the effect of the high concentration of geldanamycin and cisplatin on the expression of the RRM2 gene in cell lines across all nine tumor categories. C) the raw data used to generate the plots can be downloaded in CSV format.

“Common genes for sensitivity/resistance” tab

Using the “Common genes for sensitivity/resistance” tab, users can explore the NCI TPW data to identify potential biomarkers of drug response, following the hypothesis that centralized mechanisms of actions can be identified by looking at the correlation between gene expression changes and growth inhibition.

The distribution of the 100 genes with the largest absolute correlation between changes in gene expression and $\log(\text{GI50})$ of the specified drug can be visualized in an interactive Venn Diagram. From the left panel, the user can select the time point of interest, low or high concentration of the drug and any number of drugs between 2 and 15. Depending on the selection, the right panel will update the two graphs that identify the overlapping genes, divided between negatively correlated and positively correlated ones (tabs at the top, see Figure 8A).

Negative correlation is an association between those genes generally up-regulated after a selected drug treatment in the cell lines with low $\log(\text{GI50})$ (referred to as ‘sensitive cells’), but down-regulated or unchanged in cell lines with high $\log(\text{GI50})$ (referred to as ‘resistant cells’). Positive correlation corresponds to the reverse trend.

The Upset plot visualizes the counts of overlapping genes in a concise bar graph, where each vertical bar represents a different combination of the selected drugs (identified by the corresponding connected dots), while each horizontal bar next to the drug name represents the number of genes negatively or positively correlated with that drug.

The Venn diagram visualizes the overlapping gene counts in a more traditional way, where each section represents a specific drug and transparencies helps understand the overlapping areas. The Venn diagram can display up to 7 drugs.

The list of the gene symbols in each of the common groups is listed in a data table (Figure 8B) to allow researchers to export this information and perform further gene set analysis which may guide the discovery of common biological mechanisms of resistance or sensitivity to the specific drugs.

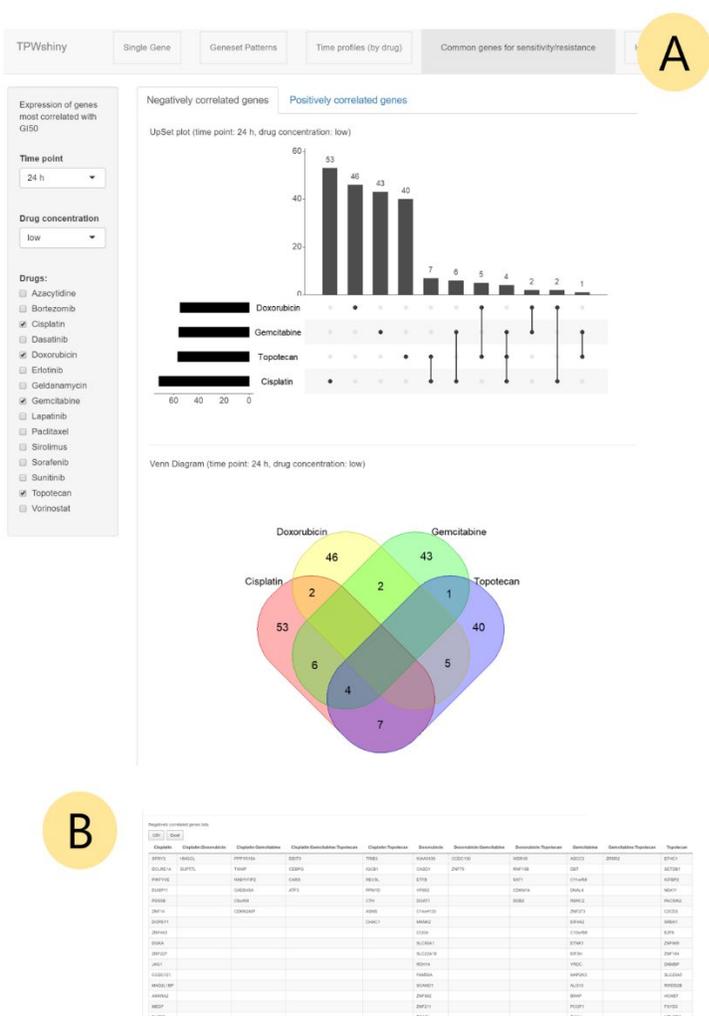


Figure 8. A) Distribution of the 100 genes with the largest absolute correlation between changes in gene expression and log(GI50) of cisplatin, doxorubicin, gemcitabine, and topotecan treated with a low concentration for 24 hr. The genes are displayed in an Upset plot (top) and Venn Diagram (bottom), to identify genes of common sensitivity and resistance. B) The list of gene names in each group is displayed below the graphs in TPWshiny.

“Help” tab

This tab contains links to the user manual, tool webpage and information about the versions of the R packages dependencies that have been tested by our Team.

To facilitate debugging any issue that may be caused by the use of R packages different than the ones used and tested by our Team, we automatically include a list of the package versions that we used in our testing as well as the versions currently used in the user session.