ToxPi™ User Manual
v2.3

Quick Start

Begin the program by double clicking the **ToxPi.jar** file or running the command
"`java -jar ToxPi.jar`" from the ToxPi directory in a terminal or command prompt.
If the interface does not appear, see the detailed installation instructions, including
Java requirements, in the following *Installation* section.

Run an example analysis by clicking the ***Add File*** button and selecting the
format_B.csv file from the data folder (can also drag and drop the file into the list to
the left of the ***Add File*** button). This file has a header that contains slice information
so the ***Recreate From File*** button can be clicked, which will automatically select all
chemicals and open the <u>Slices</u> tab using the pre-defined slice information. Continue
to the <u>Results</u> tab by clicking on the ***Calculate ToxPi Scores*** button.

At this point all tabs will be active. Refer to the sections below for more details
about each tab.

Keyboard shortcuts

Buttons that are shaded blue are selected and can be triggered by pressing *Enter*,
while those with a blue border are focused and can be triggered by pressing
Spacebar. For example, the demo above can be executed by launching **ToxPi.jar**,
pressing *Spacebar*, selecting the file, then press *Spacebar* two more times to get to
the <u>Results</u> tab.

Installation

Using ToxPi GUI requires Java. Most users will already have some version of Java installed. Java is free to download and use. If ToxPi GUI will not open, it may require an installation or update of Java.

Standard installation (i.e. For most users)

Java 8+ with JavaFX is required to run ToxPi GUI. Oracle Java versions 8-10 include JavaFX. Java can be downloaded from Oracle: https://www.oracle.com/technetwork/java/javase/downloads/index.html.

After installation/update, ToxPi GUI can be launched by double clicking the **ToxPi.jar** file included in the .zip download from www.toxpi.org. Depending on a user's local settings, it may be necessary to right-click or Control-click the **ToxPi.jar** app, then choose ***Open*** from the dialog that appears.

The ToxPi GUI can also be launched by running the command "`java -jar ToxPi.jar`" from the ToxPi directory in a terminal or command prompt.

Advanced installation
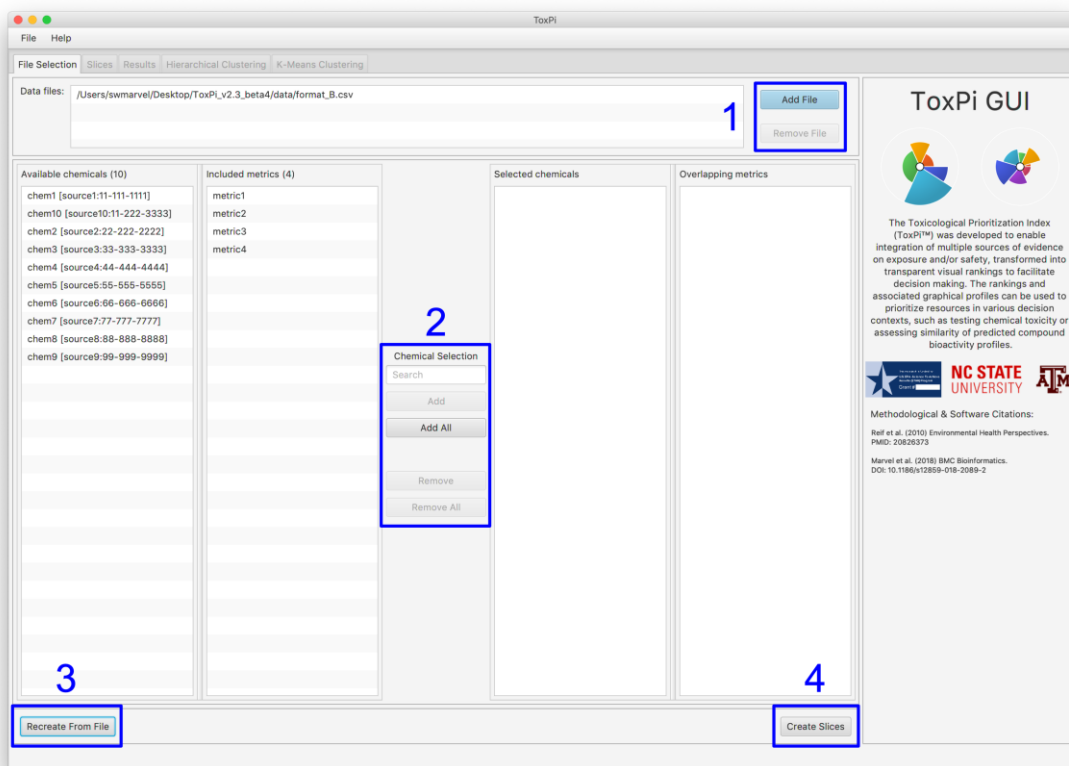
Users with Oracle Java 11+ or OpenJDK (typically users who are developing their own software in Java) will require download of the OpenJFX SDK (https://gluonhq.com/products/javafx/).

ToxPi GUI can be launched from the command line using "`java -jar ToxPi.jar`" if OpenJFX SDK is not required, or else by using "`java --module-path /path/to/javafx-sdk/lib --add-modules javafx.controls,javafx.fxml,javafx.swing -jar ToxPi.jar`", where `/path/to/javafx-sdk` is the location of the downloaded OpenJFX SDK folder.

Note for Linux users: double clicking will also work if OpenJDK and OpenJFX are installed using "`sudo apt install default-jre openjfx`".

File Selection

1) Add/Remove File

Input files must be comma-separated value (csv) formatted using one of the layouts shown in Figure 1 (on next page). All files must contain metric and chemical names, while additional information (i.e. metric source and type or chemical casrn and sid) is optional and certain fields must be blank. Layout A was used by ToxPi version 1 and requires slice information, which includes the slice name and weight. For layout A, if a metric is included in multiple slices, then the metric column must be replicated for each occurrence. The other layouts may contain slice information but are not required to. An example for layout D with slice information is shown below.

# Slice A!1!0xFF0000!sqrt(x)	x	x	
# Slice B!3!0x00FFFF!-log10(x)+6		x	x
	metric1	metric2	metric3
chemA	9.7493	6.3806	9.6981
chemB	6.5313	4.6236	2.7900
chemC	2.5663	2.8952	9.3871

Each row beginning with a hash tag (#) will define a slice. For each slice, the first column will contain a string beginning with the hash tag followed by the slice name, weight, color and scaling, all separated by exclamation points (!). The color and scaling can be specified as *null* for default values. Any column with a non-empty string will indicate that the corresponding metric is included in that slice, e.g. Slice A contains metric1 and metric2 in the example above.

A.

				1	1	1
				Slice A	Slice A	Slice A
				metric1_type	metric2_type	metric3_type
				metric1_source	metric2_source	metric3_source
row#	sid	casrn	name	metric1	metric2	metric3
1	chemA_SID	chemA_CAS	chemA	4.6067	4.3794	0.3531
2	chemB_SID	chemB_CAS	chemB	1.5209	2.7078	3.6432
3	chemC_SID	chemC_CAS	chemC	1.7350	1.5279	2.1769

B.

				metric1_type	metric2_type	metric3_type
				metric1_source	metric2_source	metric3_source
row#	sid	casrn	name	metric1	metric2	metric3
1	chemA_SID	chemA_CAS	chemA	6.4919	2.1528	8.4921
2	chemB_SID	chemB_CAS	chemB	1.9105	8.0640	2.8401
3	chemC_SID	chemC_CAS	chemC	1.8301	5.2784	1.0962

C.

row#	sid	casrn	name	metric1	metric2	metric3
1	chemA_SID	chemA_CAS	chemA	5.3601	2.9424	5.9277
2	chemB_SID	chemB_CAS	chemB	8.5816	0.1470	7.8053
3	chemC_SID	chemC_CAS	chemC	3.6355	4.1566	7.9331

D.

	metric1	metric2	metric3
chemA	8.2638	7.3700	1.3997
chemB	1.2432	7.6223	1.0313
chemC	4.2248	5.5914	5.5047

Figure 1.   File layouts. Formats B, C and D can optionally include slice information as described in section "1) Add/Remove File" on the previous page.

2) Add/Remove Chemical
Chemicals and their corresponding metrics can be selected or removed before proceeding to slice creation. The search field is case insensitive and will reduce the available chemicals list to those that match the input string. The ***Add All*** button will move all chemicals in the available list over to the selected list while the ***Add*** button will only move those that are selected within the available chemicals list. The two remove buttons have similar functions for removing chemicals from the selected chemicals list. Only selected chemicals are included in the analysis. Only metrics that are included in all selected chemicals can be used for slice creation.
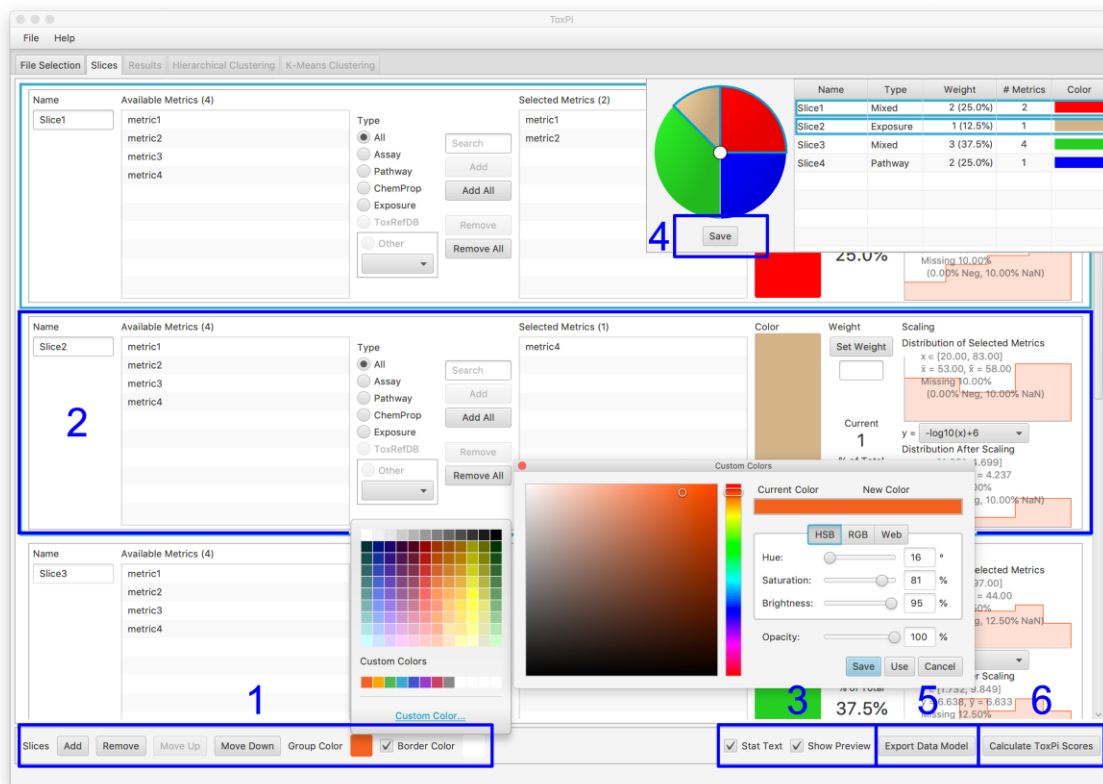
3) Recreating From File
If a single file has been added and it includes slice information, then the ***Recreate From File*** button can be used to automatically create the slices from the file header. If the selected chemicals list is empty, then all available chemicals will be automatically selected.

4) Create Slices
All selected chemicals and the metrics that exist for all chemicals are combined into a single data object to be used in slice creation and analysis.

Slices

1) Add/Remove/Move Slice

Add Slice will add a new slice to the end of the list. The color of the new slice will automatically be derived from the current last slice. ***Remove Slice(s)*** will remove selected slice(s). Multiple slices can be selected using the *Shift* and *Meta* (Command/Windows) keys while clicking on a slice in the main slice table, preview summary table or the ToxPi image in the preview pane. ***Group Color*** is only available if more than one slice is selected. It will apply the chosen color to all selected slices, automatically becoming slightly brighter for each additional slice. The ***Border Color*** can be toggled on/off and changed to any desired color. Changing this setting will not propagate to the <u>Results</u> or the clustering tabs since the ToxPi images are created as a step in the ***Calculate ToxPi Scores*** process. To change the border setting in those tabs the ToxPi Scores will need to be recomputed. Note for Linux users – custom color dialog popups may appear behind the main window.

2) Slice Panel

Each slice will have a variety of settings that can be modified in the slice panel

Name	Slice name. Cannot have duplicate names.
Available Metrics	The overlapping metrics from the <u>File Selection</u> tab.
Type	Limit available metrics to those with the specified type and change to the associated color.
Add/Remove Metrics	Similar function to add/remove chemicals on <u>File Selection</u> tab.

Selected Metrics	The metrics used for the slice score computation.
Color	Slice color. Note for Linux users – the custom color dialog popup may appear behind the main window.
Weight	Slice weight. The weight can be an integer or ratio of integers. To recreate the ToxPi model represented by the image above from the **format_B.csv** datafile included in the download, the weights for each of the four slices would be:
	Integer weights = {2,1,3,2} or, equivalently,
	Integer ratio weights = {1/4, 1/8, 3/8, 1/4}
Scaling	Slice scaling. **See footer note**
Summary Stats	Summary statistics for selected metrics, overlaid atop a histogram of the distribution.

3) Stat Text/Show Preview

Stat Text will toggle the summary stats text overlaid on the histograms in the slice panels. *Show Preview* will show/hide the preview pane in the upper right corner of the window. The preview pane reports basic slice information and depicts the reference pie image.

4) Save

Save the reference pie image as a pdf with slice name labels, also save the preview table as a png. Saved as "*<file>*_preview.pdf" and "*<file>*_previewTable.png". The png has a transparent background, which may cause text to blend into the background if viewed in a dark mode app.

5) Export Data Model

Save a data file containing the chemicals selected in the *File Selection* tab with a header for the currently defined slices. Saved as "*<file>*_data.csv".
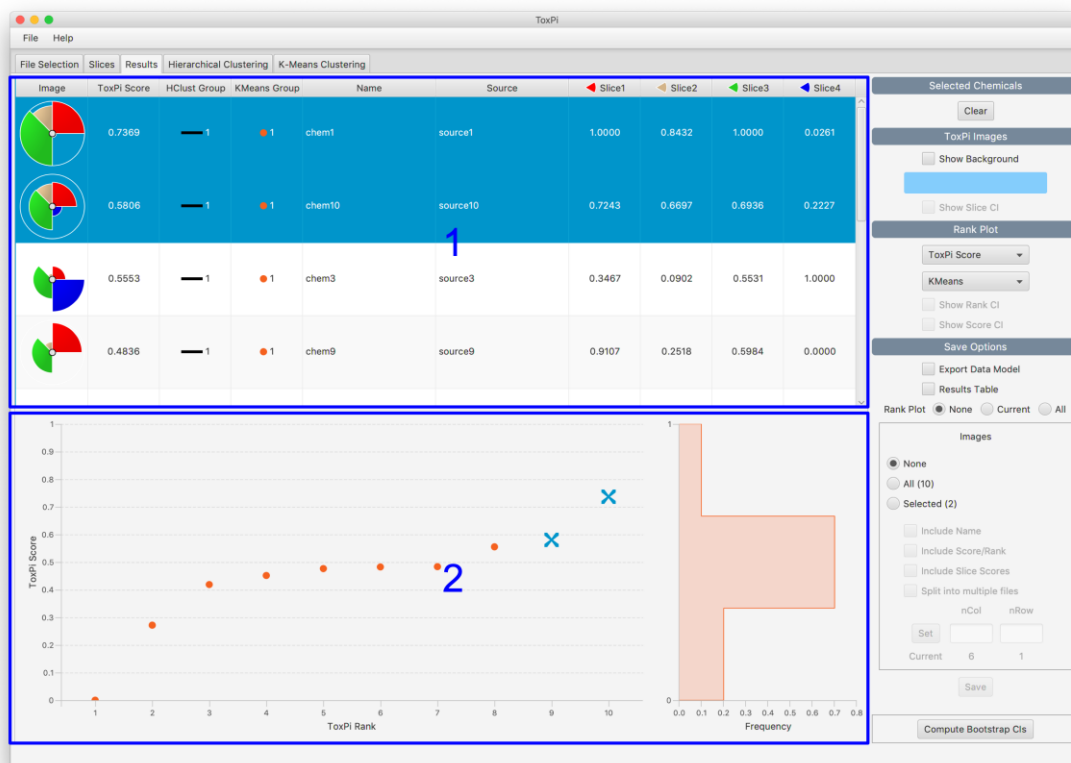
6) Calculate ToxPi Scores

Compute ToxPi scores using current slice settings.

**Scaling note:

Up to v2.2, missing data (i.e. negative values, NA/NaN) and invalid data-scaling input (e.g. 0 into "-ln(x)") were converted to 0 before summing across metrics to obtain a slice score and subsequent slice-scaling to the range [0, 1]. In v2.3, these missing/invalid data are ignored when computing slice scores and any chemical lacking valid data will be assigned a slice score of 0 after the slice-scaling step.

For example, say a set of chemicals with valid input data have slice scores in the range [10, 20] but a small set of chemicals only contain missing or invalid data. Up to v2.2, those chemicals would be assigned a slice score of 0 before slice-scaling, resulting in slice scores being in the set {0, [10, 20]}. After slice-scaling, the missing/invalid data chemicals would have a slice score of 0 while those with valid input data would be in the range [0.5, 1]. In v2.3, the missing/invalid data chemicals are ignored during slice-scaling so chemicals with valid input data would have slice scores in the range [0, 1] after slice-scaling (rather than [0.5, 1]) and then any chemical with only missing/invalid data would be assigned a score of 0.

Results

1) Results Table

Each chemical will have a ToxPi image and score along with scores for each slice. In addition, group labels for both hierarchical clustering and k-means clustering are reported. Note that while manual coloring of groups in the Hierarchical Clustering tab is allowed, those colors/groups will not be reported on the Results tab, only the groups due to changing the auto-color settings will be reported. Clicking on the image will open an enlarged ToxPi image in a tab on the right side of the results table while clicking on any other part of the row will select the chemical. Multiple chemicals can be selected by holding the *Shift* or *Meta* (Command/Windows) keys. Chemicals can be sorted by clicking on the column names and holding shift when clicking on additional columns will do hierarchical sorting.

2) Rank Plot

The left plot shows chemical scores vs. rank. Chemicals can be selected by clicking on the points. Tooltips that show the chemical name, rank and score will popup when hovering over a point. Click and drag a region to zoom in while double-clicking will reset the plot area. The right plot shows a histogram of the ToxPi scores.

Selected Chemicals - *Clear*

Remove all selected chemicals in the Results, Hierarchical Clustering, and K-Means Clustering tabs.

ToxPi Images
Show Background will add a background color to the ToxPi images. Note for Linux users – the custom color dialog popup may appear behind the main window. ***Show Slice CI*** will show/hide the confidence intervals (CIs), which first requires the CIs be computed using the ***Compute Bootstrap CIs*** button.

Rank Plot
Pull-downs: the first pull-down changes what score is used as plotting data, either the ToxPi score or that from a slice; the second pull-down changes which clustering method is used to label the symbols. ***Show Rank/Score CI*** will add confidence interval lines to the rank plot. The rank confidence interval is determined from the ToxPi score and will be unchanged if viewing slice scores, however, score confidence intervals are computed separately for both the ToxPi score and slice scores. These options first require the CIs be computed using the ***Compute Bootstrap CIs*** button.
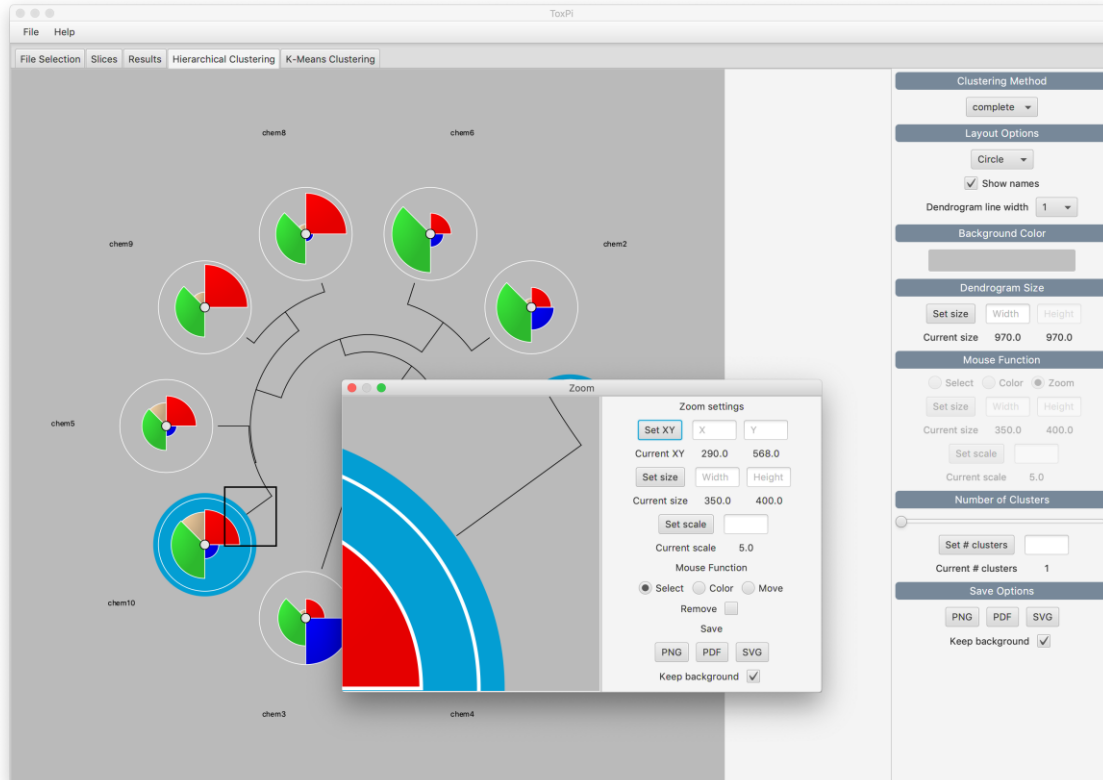
Save Options
A single save button is used to save several different figures. The base file name, *file*, will have different suffixes applied depending on the selected figures.

Export Data Model	Save a data file containing the chemicals selected in the *File Selection* tab with a header for the currently defined slices. Saves as "*<file>*_data.csv". This will create the same output as the ***Export Data Model*** button in the <u>Slices</u> tab.
Results Table	A table of ToxPi and slice scores for each chemical. Saves as "*<file>*_results.csv".
Rank Plot	A copy of the rank plot and histogram. Option to save only the currently selected score, or all scores (ToxPi score and all slice scores). Saves as "*<file>*_rank.png".
Images	A grid of ToxPi images. Can save all chemicals or the subset of selected chemicals. Names, ToxPi score/rank and slice scores can be included below the images. Finally, the number of image rows and columns per file can be specified. Picking 1 column and 1 row per file while selecting "Split into multiple files" will generate individual files for each chemical. Saves as "*<file>*_images.pdf".

Compute Bootstrap CIs
Compute confidence intervals for each slice using 1000 bootstrap samples. This can be a fairly computation-intensive step.

Hierarchical Clustering

Clustering Method

Agglomeration method used to cluster chemicals. The clustering algorithm is a Java port of the R function *hclust*. The method names correspond to the same options available in the R implementation. Clustering results can be replicated in R using the code below after saving the "Results Table" from the <u>Results</u> tab.

```
# Get results
res <- read.csv("Example_results.csv", check.names=F)
# Get slice weights
w <- sapply(sapply(names(res)[-c(1:5)],function(x) strsplit(x,
    "!")),"[",2)
w <- sapply(strsplit(w,split="/"),function(x) {y <-
    ifelse(length(x)==2,x[2],1);as.numeric(x[1])/as.numeric(y)})
# Cluster
hc <- hclust(dist(res[,-c(1:5)]*rep(w,each=nrow(res))),
    method="complete")
# Plot
plot(hc, hang=-1, labels=res$Name)
```

Layout Options

There are five choices for the layout of the cluster dendrogram ("tree"). These include 'Circle' (as shown above) and 'Bottom', 'Left', 'Top', or 'Right' for a traditional hanging-style dendrogram. The width (thickness) of the dendrogram lines is adjustable.

Background Color
This provides a color picker popup to adjust the background canvas. Note for Linux users – the custom color dialog popup may appear behind the main window.

Dendrogram Size
The size of the dendrogram can be specified. This option is useful for meeting specific graphical size requirements (e.g. adhering to journal figure requirements or poster real estate).

Mouse Function
Select
Chemicals can be selected/deselected by clicking on a ToxPi image. If the **Select** option is chosen, then groups of chemicals can be selected (or deselected if **Remove** is checked) by click and dragging over either tree branches or ToxPi images.

Color
The **Color** option can be used to manually color tree branches, but this will not change the HClust group specified in the Results tab, only setting the number of clusters will determine group labels. Note for Linux users – the custom color dialog popup may appear behind the main window.

Zoom
Clicking the **Zoom** button attaches a rectangular outline-box to the cursor that can be placed over any desired section of the dendrogram. Upon clicking, a 'Zoom' popup will appear (see inset in figure above) that shows a magnified view of the area delineated by the cursor's outline-box. Users can scroll across detail in the 'Zoom' popup by hold-dragging the mouse button over the main window dendrogram image or tracking within the 'Zoom' popup itself. The 'Zoom' popup also provides magnified versions of the **Select** and **Color** button functions. Scale and size options will adjust the level of magnification.
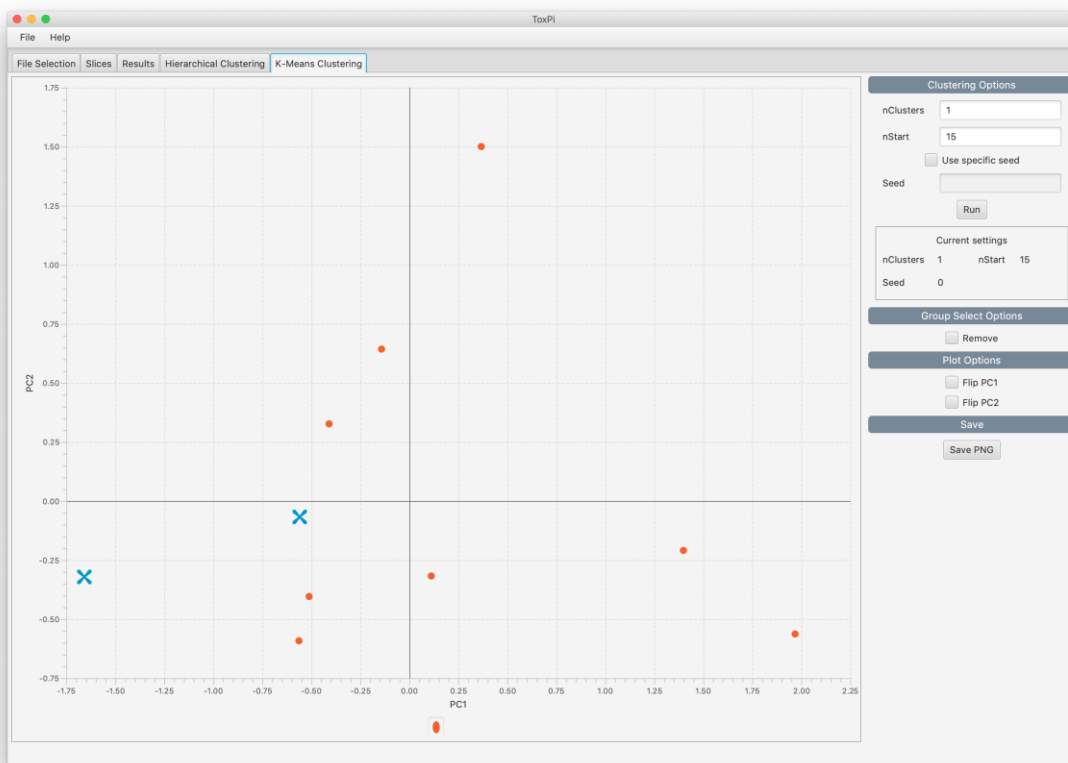
Number of Clusters
The number of auto-colored clusters can be set with a slider bar or the text box. Clusters are set as cutoffs moving from the dendrogram root downward/outward toward the leaves.

Save Options
The cluster dendrogram image can be saved to an external file with or without a transparent background.

K-Means Clustering

Clustering Options

Chemicals can be split into groups using k-means applied to the weighted slice scores. The implementation is a Java port of the R function `kmeans` using the "Hartigan-Wong" algorithm. The algorithm is run ***nStart*** times with different starting cluster locations and the best clustering result has the smallest within group sum of squares. The seed of the random number generator can be specified to replicate results. The chemicals are plotted using the first two principal components (PCs) and are colored according to the best clustering groups. The chart symbols are linked to those used in the <u>Results</u> tab Rank Plot.

Group Select Options

Chemicals can be selected/deselected by clicking on a chart symbols. Groups of chemicals can be selected (or deselected if ***Remove*** is checked) by click and dragging over multiple chart symbols.

Plot Options

The sign of each PC axes can be flipped. Selecting the Flip PC checkboxes will multiply the corresponding PC by -1.

Save

The entire image can be saved to an external file.

Vignettes

Load a previously-created ToxPi model

On the <u>File Selection</u> tab, click the ***Add File*** button. Select the file stored on your machine (e.g. a file in your `/Downloads` or the "**size medium.csv"** example file in the `/data` directory from the ToxPi download) and click ***Open***. Click the ***Recreate From File*** button. This will automatically select all chemicals and open the <u>Slices</u> tab using the pre-defined slice information. Click the ***Calculate ToxPi Scores*** button to see the ToxPi model in the <u>Results</u> tab. All tabs are now active.

Save a custom ToxPi model to share

Verify that all adjustments have been made by clicking the ***Show Preview*** box on the <u>Slices</u> tab. Click ***Save*** on the resulting inset window. To export the ToxPi data model, which includes the data and all ToxPi parameters needed to reload the model, click ***Export Data Model***. Choose a directory location and descriptive name for the model, which will be saved as "*<descriptive_name>*_data.csv". The exported file can now be shared with others or reloaded for further exploration/modification.

Save ToxPi profile graphics

The ***Save*** button underneath the **Save Options** section of the <u>Results</u> tab has options for saving ToxPi profile images. To save all images, click the ***All*** button. If no adjustments are made to other check-box options, the profiles will be written to a .pdf, arranged in a matrix according to the number of ***nCol*** columns. For example, if a set of 300 ToxPi images is saved with ***nCol*** = 20, the resulting .pdf will have a matrix of ToxPi images arranged into 15 rows of 20 profiles each.

Note that the images will be sorted according to the current state of the results table. The default sorting is by descending ToxPi Score. If other sorting is desired, click on the relevant column header (e.g. Sort in descending order of Slice 1 scores), and the resorted table will be used for the saved image file.

Save ToxPi rank (distributional) graphics

The ***Save*** button underneath the **Save Options** section of the <u>Results</u> tab has options for saving the Rank Plot images (the dotcharts of ToxPi scores versus rank plus associated histograms). To save only the currently-displayed Rank Plot, choose the ***Current*** button in the **Rank Plot** row. To adjust the display, scroll through options in the upper **Rank Plot** section. These options include changing the vertical axis of the Rank Plot from overall ToxPi score (default) to instead displaying a single slice score on the vertical axis. Selecting single slice scores will <u>not</u> rearrange the horizontal rank axis (though the associated histogram is redrawn), and in this

manner, the distribution of singular slices can be visualized. To save all single slice and the default overall ToxPi score Rank Plots at once, choose the **All** button in the **Rank Plot** row.

The other display option is to adjust the points ("dots") on the Rank Plot. The default is to plot each point according to its K-Means Clustering annotation. If no clustering has been chosen (i.e. **nClusters** = 1 on the K-Means Clustering tab), then the points will be plotted as orange circles. If clustering is performed, the Rank Plot characters will be adjusted accordingly. The points can be plotted according to either clustering methods by scrolling to **KMeans** or **Hclust** in the upper **Rank Plot** section.
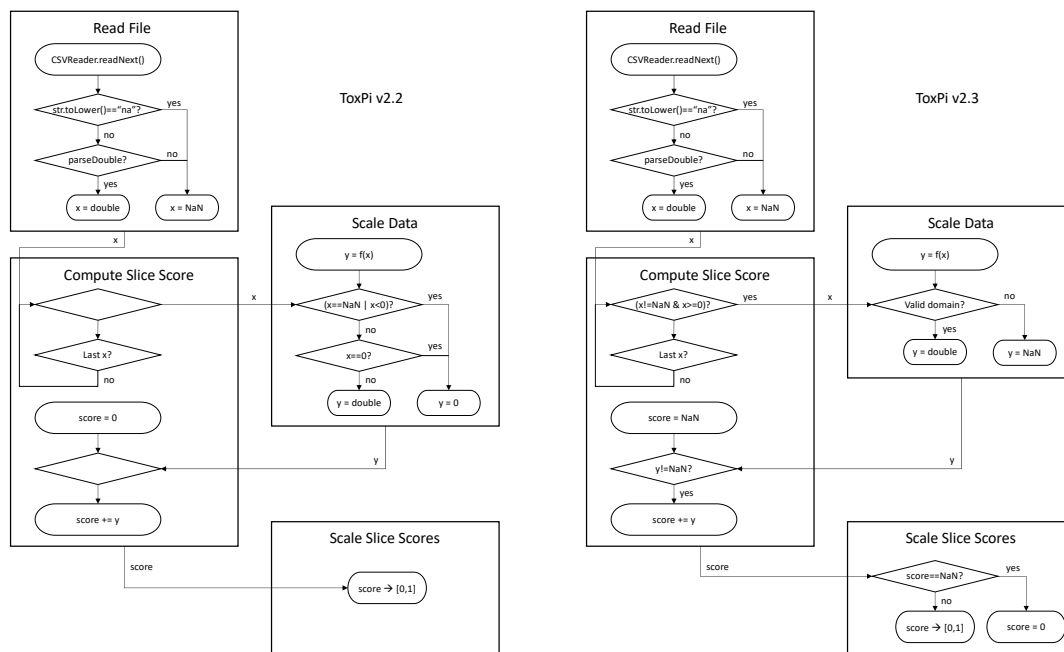
Explore scaling options

The effects of different scaling options can be explored in the Slices tab. In the far right of the panel for each slice, histograms display data distributions before (above) versus after (below) a user's scaling choice. To aid scaling choices, a new "distributions.csv" file has been included with the distribution in the data folder. This file includes data following several commonly-encountered statistical distributions and allows users to see the effects of ToxPi scaling options. The default model for these data assigns each 'Available Metrics' data component to its own slice. Users can add additional components to a given slice to see the effects of mixing source data distributions.

Note that scaling choices in ToxPi will depend on analysis goals. For example, users wanting to highlight extremes (e.g. chemicals especially "active" on a given slice) may not want a normal or uniform distribution of scores. In general, the most effective ToxPi models will consider both the distribution and scaling of component data in building slices that include more than a single 'Available Metrics' data component.

Changelog

ToxPi 2.3

This update focuses on scaling. There are new scaling options and an implementation fix to one of the existing options. The main change is in how missing data is treated, as shown below.

File Selection
There are 2 new files in the data folder: "distributions.csv" and "missing.csv". The "distributions" file will show users what happens to their data when choosing a scaling option in the slices tab, the "missing" file will show how negatives and NA/NaN are treated when computing slice values.

Slices
Several new scaling options have been added and the implementation of "-log10(x)+log10(max(x))" has been corrected by changing log(max(x)) → log10(max(x)). This error (using natural log rather than log10) has existed since ToxPi v1. The "Scaling" panel has been redone. It now includes stats and histograms for before and after scaling. The stat text can be toggled on/off. Any color can be used for slice borders now rather than white or none.

Results
Rank plot pull-downs added to select which score is used as data and which clustering method is used to label the symbols. Option to save all rank plots now

that several data sources are available. Improved communication between results and hclust tabs when toggling confidence intervals.

General
Saving an SVG used to be gzip to SVGZ, however, this has been changed to saving as an uncompressed SVG.

ToxPi 2.2

This update adds flexibility for setting weights, updates information included in the saved **Results Table** object, and fixes the following two issues identified by users:

Files beginning with a Byte Order Mark (BOM) resulted in an error when processing file content. Files that begin with the BOM "\uFEFF" are now correctly processed.

Saving a large number of images would sometimes result in an error. There were two causes for this: 1) hardware acceleration limited the size of the JavaFX Node that could be constructed or 2) a memory leak would cause a heap overflow when outputting multiple files. The image saving process has been changed. Prior to v2.2, the image was created using a snapshot of a Scene where image shapes and text were added to a containing parent node. Now, a custom function is used to layout the image information as an SVG string that is then transcoded to a PDF.

Results
An additional change for the Results tab is the addition of slice weights and colors to the column labels for the **Results Table** object saved to file. These fields are appended to the slice names, separated by exclamation points. This change will allow reconstruction of ToxPi images using a single file, where prior versions would require information from both the results table and exported data model files.

Slices
The slice weights slider has been replaced by an explicit text box and responsive **Set Weight** button. This allows slice weights to be set as integers relative to the whole model or as fractions. This update permits finer control over weights and easier implementation of weights derived from coefficient-based models (e.g. linear regression).

ToxPi 2.1

The primary reason for this update is to prepare the software for handling larger environmental datasets. Previously, ToxPi has mainly been used to analyze chemical datasets, which typically consist of hundreds of chemicals. Now there is a need for ToxPi to work with datasets consisting of thousands of locations where environmental assay data is collected. The majority of this update is a complete rework of the Hierarchical Clustering tab and improving functions for creating SVG

images of both the clustering results and pie images. These SVG functions will be leveraged during implementation of the web-based ToxPi software development.

File Selection
Improved performance when using "Recreate from File" with large datasets – the copy data step is now skipped.

Slices
Data export option available on this tab now instead of <u>Results</u> tab. Preview table can now be saved.

Results
Chart symbol updated to match new K-Means options. Changed symbol of selected chemicals to stand out. Pie background color is customizable.

Hierarchical Clustering
New implementation greatly improves performance by drawing on a Canvas instead of adding Nodes to a Pane. New zoom functionality for better user control. Ability to select multiple chemicals added.

K-Means Clustering
Increased maximum number of clusters from 8 to 64, which required new chart symbols. Ability to select multiple chemicals added.

ToxPi 2.0

This version is the implementation described in the new GUI publication:
Marvel SW, To K, Grimm FA, Wright FA, Rusyn I, Reif DM. ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. BMC Bioinformatics. 2018 Mar 5;19(1):80.