

User Manual of “Biomarker Explorer” R-Shiny Web App

Author: Yuanyuan Xiao (Yuanyuan.Xiao@gilead.com)

Developer: Feiyang Niu (Feiyang.Niu@gilead.com); Qinghua Song (Qinghua.Song@gilead.com)

1. Description:

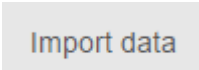
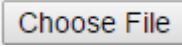
This R-shiny web app was developed at Gilead Sciences, Inc. for biomarker statisticians and scientists to generate graphs and tables for biomarker exploratory analysis including time profiling and association, in an interactive platform. The app can be accessed via http://rshiny.gilead.com/dev/biomarker_explorer/.

The app consists of multiple navigation-bar pages with each fulfilling distinct purposes as described below

- **Import data:** the user uploads the data file in this page.
- **Time profiling:** the user explores the pattern of certain variable of interest over time through some most commonly used types of graphs.
- **Association:** the user explores the distribution of one variable or association relationship between two variables in this page.
- **Survival:** the user performs survival analysis including Kaplan Meier curve and log-rank test.
- **Output results:** the user prepares the parameter specifications of output graph in a foot and title (TNF) file of Excel format and obtains all the desired graphs in a zipped file.

2. User Direction:

Step 1. Data File Upload

- Click  on the navigation bar on the top of the platform to import data.
- Click the  button to select and upload user's data file onto the web-based platform with file extension being one of the following:
 - RData, R saved workspace data file. Note that only one data object is allowed in the saved workspace for the platform to correctly identify the data to use
 - CSV/TSV, delimiter-separated files
 - XLS/XLSX, Excel file
 - TXT, plain text file
 - sas7bdat, SAS data file

To ensure that the application runs correctly, the data file the user uploads must satisfy the following requirements (You may click the link “Need input data template?” in


Choose file to import (Need input data template?)

to download an input data template):

- ❖ In data frame format
- ❖ The following columns are required
 - STUDYID (study id)
 - USUBJID (subject id)
 - PARAM (biomarker names)
 - AVAL (biomarker measurements)
 - BASE (biomarker baseline measurements)
 - CHG (biomarker change from baseline measurements)
 - PCHG (biomarker percentage change from baseline measurements)
 - PRCHG (biomarker proportion of baseline measurements)
 - XTICKER (numeric visit time, e.g. 6)
 - XTICKERL (nominal visits containing time unit, e.g. ‘Week 6’)
- ❖ The following columns are optional
 - UNIT (biomarker measurement unit)
 - PFSTIME (progress free survival time)
 - PFSEVENT (progress free survival event, 1 when event happens and 0 otherwise)
 - OSTIME (Overall survival time)
 - OSEVENT (Overall survival event, 1 when event happens and 0 otherwise)

In order to perform the survival analysis, at least one of the following pairs must be present

- PFSTIME and PFSEVENT
 - OSTIME and OSEVENT
-
- ❖ Missing entries should either be left empty or marked ‘NA’

 View top 10 rows

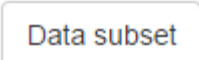
Once the uploading is complete, a button will appear on top of the main panel. Click it to view top 10 rows or more. An error message will appear on top of the main panel if the uploaded data don’t meet the data requirement. An example is given below

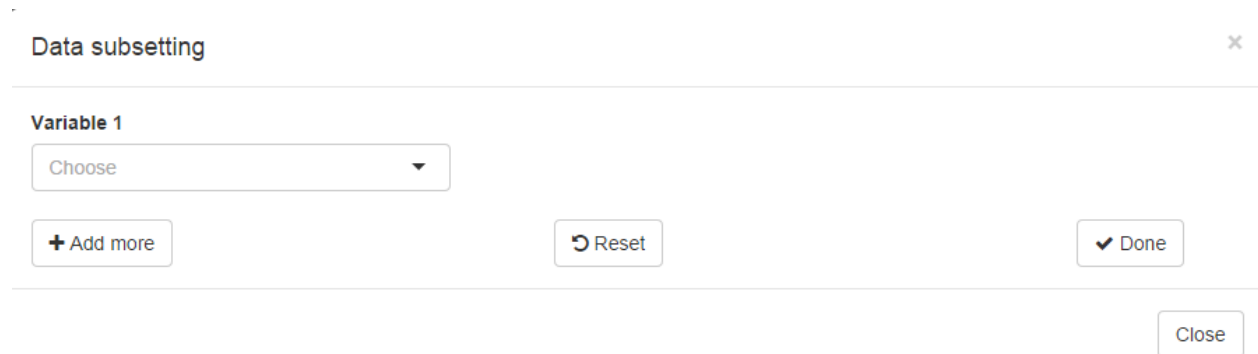
Error: Data importation failed!

The following required columns are not found in the data:

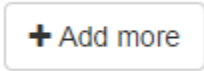


STUDYID
USUBJID
PARAM
AVAL
BASE
CHG
PCHG
PRCHG
XTICKER
XTICKERL

Step 1.1 Data filtering

Once the data uploading is complete, a button  will appear on top of the main panel. Click it to open a modal window to perform data filtering as follows.

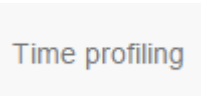


The modal window titled "Data subsetting" has a close button (X) in the top right corner. It contains a section labeled "Variable 1" with a dropdown menu showing "Choose". Below this are three buttons: "+ Add more", "Reset", and "✓ Done". A "Close" button is located at the bottom right of the modal.

- Choose a column name from the drop-down list of **Variable 1**
- Corresponding, a numeric slider input or a drop-down list will appear depending on the continuity of the selected column in **Variable 1**. Rows will be filtered according to the interval determined by the numeric slider input or the categories by the drop-down input list.
- Click  to apply more filtering criteria or  to finish the data filtering process.
- Click  to clear all applied data filtering criteria.

Step 2 Visualization

Step 2. 1 Time profiling

- Click  on the navigation bar on the top of the platform.
- Click the “Study” to select a specific Study or Studies of interest. If the user wants to compare time-profiling patterns among multiple studies, he/she can leave this blank and choose “STUDYID” in the “Group” select input.
- Click the “Cohort” to select a specific cohort of interest. Leave it blank if the user wants to perform analysis across all cohorts.
- Click the “Biomarker” to select the Biomarker of interest. One Biomarker can be selected at each.

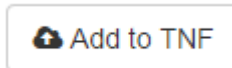
- Click the “Y variable” to select the one of the biomarker-dependent Y variables in the time profiling graph:
 - AVAL
 - BASE
 - CHG
 - PCHG
 - PRCHG
- Click “Output Type” to select one of the graph types:
 - Spaghetti plot
 - Mean + SE
 - Mean + SD
 - Median + IQR
 - Boxplot
 - Summary table

Once the “Study”, “Biomarker”, “Y Variable” and “Graph Type” are specified, a graph will be displayed correspondingly on the main panel. Please note that data of all studies will be displayed if the “Study” input is left blank.

- Click “Group” to select a group variable (such as “SEX”, “DISEASE TYPE”, etc) over which the plots can be plotted. This is optional.

Once the user is satisfied with the shown time-profiling graph and would like to save it to the

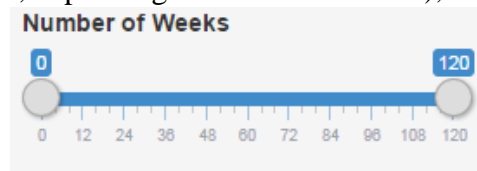
final outputs, the user can click



button. Consequently, the current specification of the graph will be added as a new row to the TNF file, which can be easily previewed or downloaded in “Output results” page.

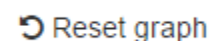
There are several ways that the user can interact with the static plot generated by specifying the below input widgets.

- **Zooming:** typically instead of the entire scope of the graph, one is interested in certain sub-region. This can be easily accomplished in one of the following two manners.
 - **Slider bar:** Move the end point in the slider bar below, “Number of XX” (XX can be Weeks, Days, and etc., depending on the units of time), to change the range of X axis

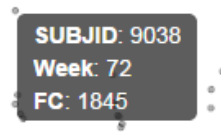


- **Brush:** Move the mouse on the plot, click, hold and drag, to zoom in to the region to

be focused. To go back, double click on the graph or click beneath the graph.

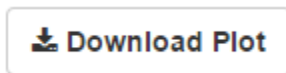


- **Tooltip:** When the mouse hovers over a data point, a small semi-transparent panel will appear at the lower right of the mouse displaying crucial information of the data point.



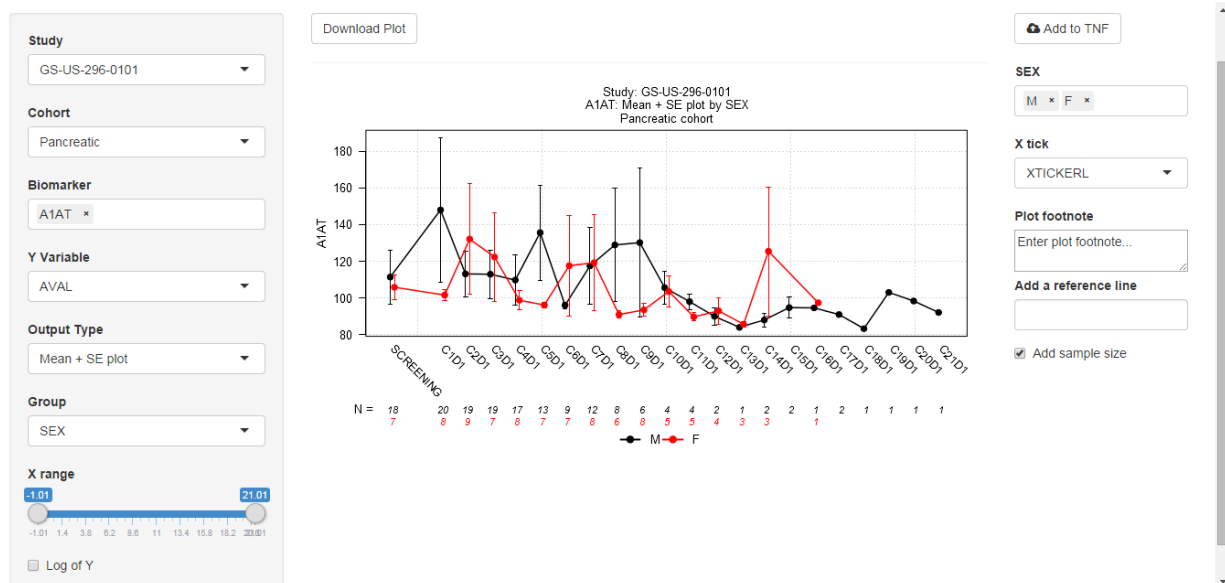
- **Line highlight:** In spaghetti plot, when the mouse hovers close enough to one line, the line will be highlighted in bold. A tooltip will appear to indicate the corresponding subject ID of the highlighted line.
- **Log scale:** click “log of Y” checkbox, then the scale of Y will be transformed into semi-logarithm.

Finally, when the user is satisfied with the graph and wants to keep a physical copy, he/she can

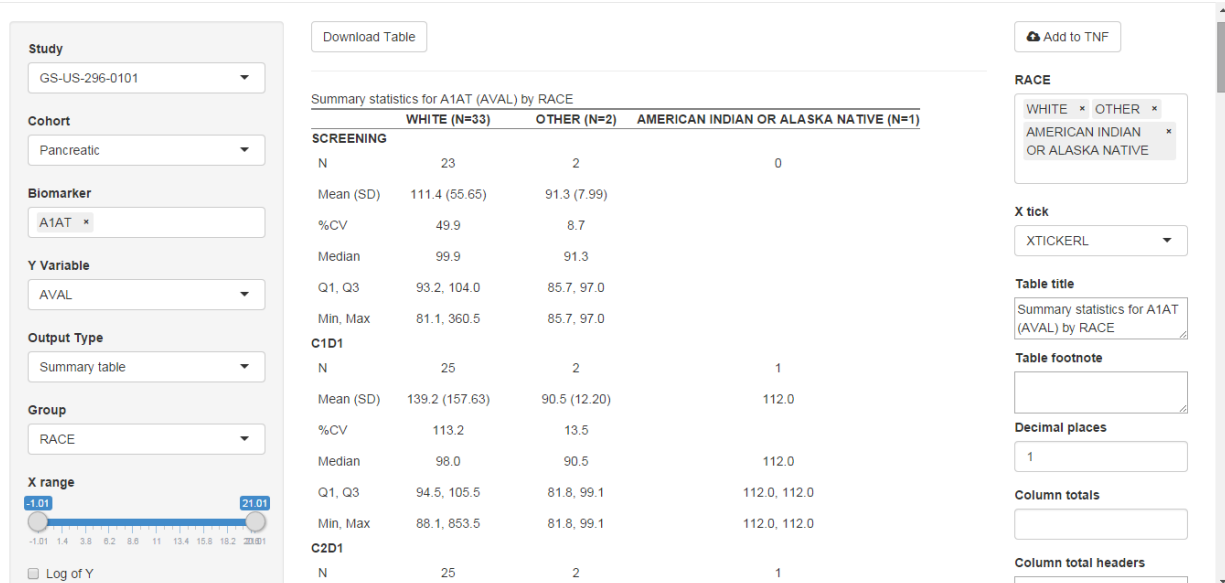


- Click the **Download Plot** on the left top of graph to download the plot into a pdf file. Note: If you’d like to download the plot in a specific area, you could use the brush tool, zoom in and download the plot before resetting the graph.

Example 1: Mean+SE plot (Output Type = Mean + SE) of the raw value (Y Variable = AVAL) of A1AT (Biomarker = A1AT), grouped by gender (Group = SEX)



Example 2: Summary table (Output Type = Summary table) of the raw value (Y Variable = AVAL) of A1AT (Biomarker = A1AT), grouped by race (Group = RACE)



Graph/Table refinement

Once the time-profiling plot is shown, there are several widgets on the right side of the plot are designed for further graph refinement.

- If the “Group” variable is specified, the levels of the group variable are possible to be re-ordered by the user with an example given below

SEX

M F

- “X tick” is used to change the X tick labels in the graph
- “Footnote” is used to add any appropriate footnote to the graph
- “Add a reference line” allows the user to add one or more reference lines to the graph. When adding multiple lines, use comma to separate the line positions
- “Add sample size” checkbox toggles the sample size appearance in the graph. By default, sample sizes across possible group levels are displayed along all the visits.

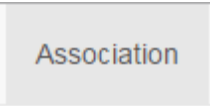
Once the time-profiling table is shown, there are several widgets on the right side of the table are designed for further table refinement.

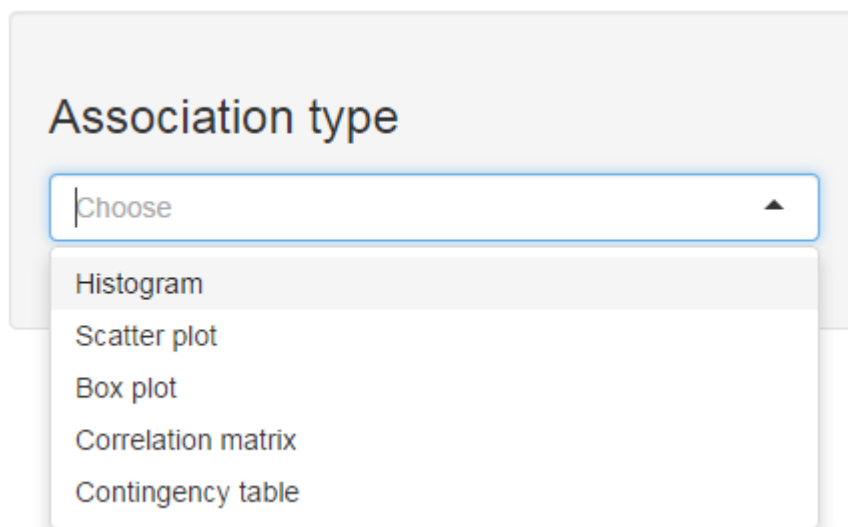
- If the “Group” variable is specified, the levels of the group variable are possible to be re-ordered by the user, same as in time-profiling graph
- “X tick” is used to change the X tick labels in the graph
- “Table title” allows the user to change the title of the table
- “Table footnote” allows the user to add any footnote to the table
- “Decimal places” control the number of decimal places of the numbers in the summary table

- “Column totals” adds first n column totals to the summary table. For example, input `3` if the user wants to add the total of the first 3 group categories. Multiple inputs are permitted, e.g. 2, 3.
- “Column total headers” specifies headers for the column totals. One row per each.
- “Add change from baseline” checkbox toggles a block of change from baseline for every visit post baseline.

Step 2. 2 Association

Association

- Click  on the navigation bar on the top of the platform.
- Click the type of Association from:



As indicated by the drop-down list, five types of association output can be produced in the “Association” page.

Clicking on each of the association type will unlock a panel of UI widgets dedicated for the selected type. Two UI widgets however are same across all the types

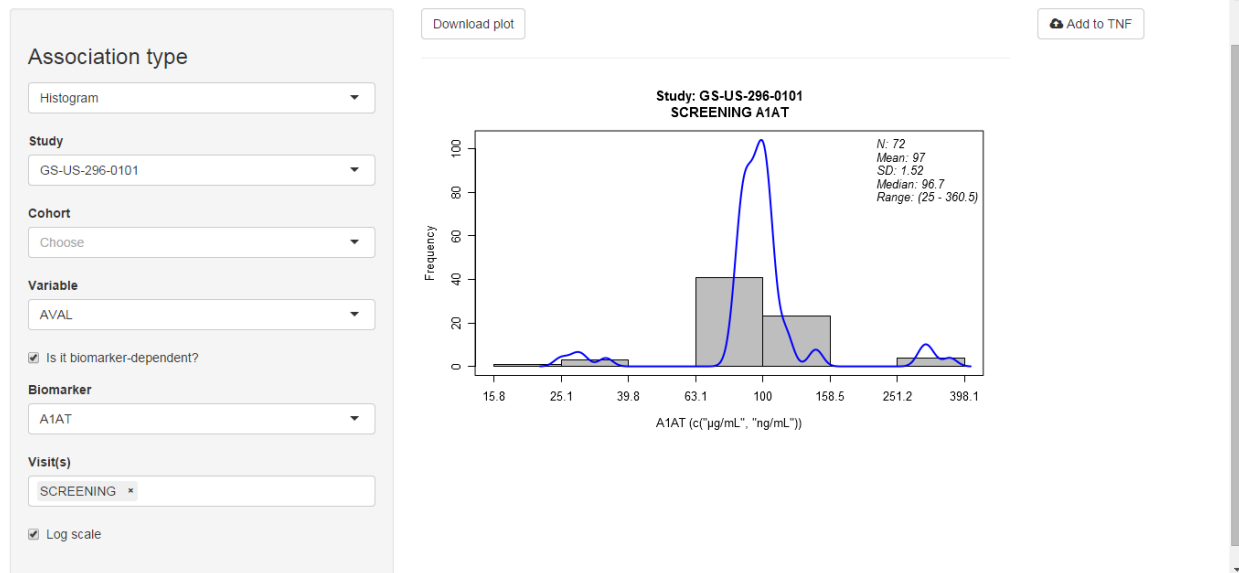
- Study: Click to open a drop-down list to choose a specific study to focus on. Leave it blank if the user wants to look at all the studies.
- Cohort: Click to open a drop-down list to choose a specific cohort to focus on. Leave it blank if the user wants to look at all the cohorts.

Histogram

To produce a histogram, the user needs to further specify the following widgets

- Variable: Choose a variable to plot a histogram on. After a variable is specified, a checkbox “Is it biomarker-dependent” will appear. Check the box if the variable is biomarker-dependent, e.g. AVAL, CHG and etc.
- Biomarker: The name of the biomarker if “Is it biomarker-dependent” checkbox is checked.
- Visit(s): Select one or more visits.
- Log scale: Check the box to have semi-log scale.

An example of histogram is given below



Scatter plot

To produce a scatter plot, the user needs to further specify the following widgets

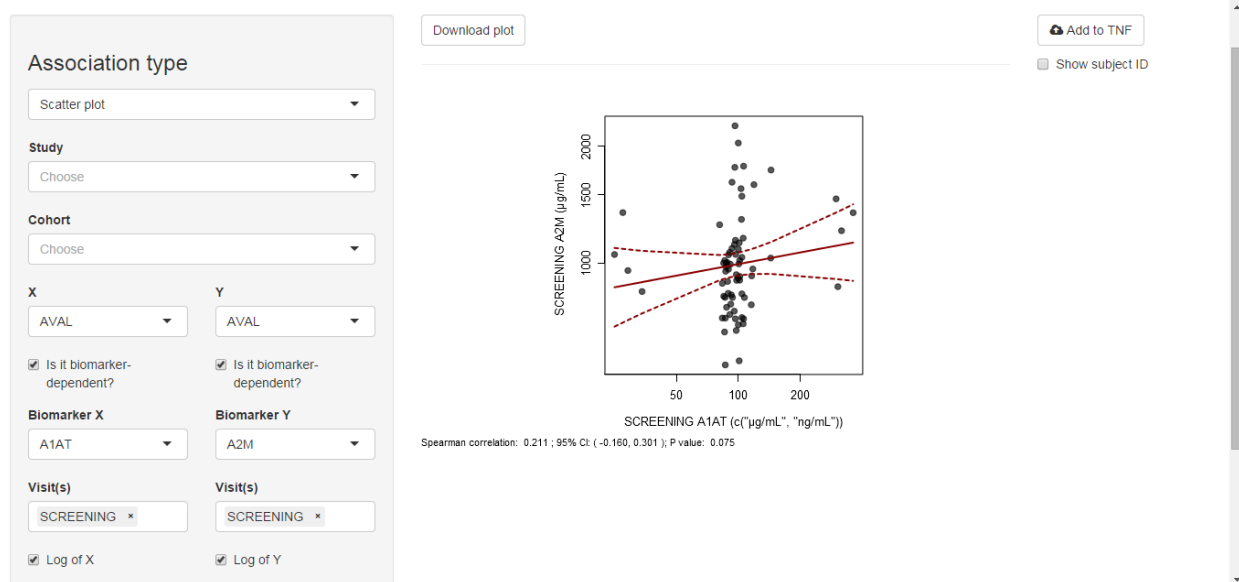
- X: Choose a variable for the X axis in the scatter plot. After a variable is specified, a checkbox “Is it biomarker-dependent” will appear. Check the box if the variable is biomarker-dependent, e.g. AVAL, CHG and etc.
- Y: Choose a variable for the X axis in the scatter plot. After a variable is specified, a checkbox “Is it biomarker-dependent” will appear. Check the box if the variable is biomarker-dependent, e.g. AVAL, CHG and etc.
- Biomarker X: The name of the biomarker for X variable if “Is it biomarker-dependent” checkbox is checked.
- Biomarker Y: The name of the biomarker for Y variable if “Is it biomarker-dependent” checkbox is checked.

- Visit(s): Select one or more visits for both X and Y variable. Note that when multiple visits are specified, visits for X and Y should match with each other.
- Log of X/Y: Check the box to have semi-log scale.

In addition, the following two widgets further allow the user to add more information to the scatter plot

- Add a line: The user has the option to add one of the following three lines to the scatter plot: (1) Linear regression line, (2) Loess curve and (3) Identity curve.
- Correlation type: The user has the option to add one of the following two correlation results to the footnote of the scatter plot: (1) Pearson correlation and (2) Spearman correlation

An example of scatter plot is given below



Box plot

To produce a scatter plot, the user needs to further specify the following widgets

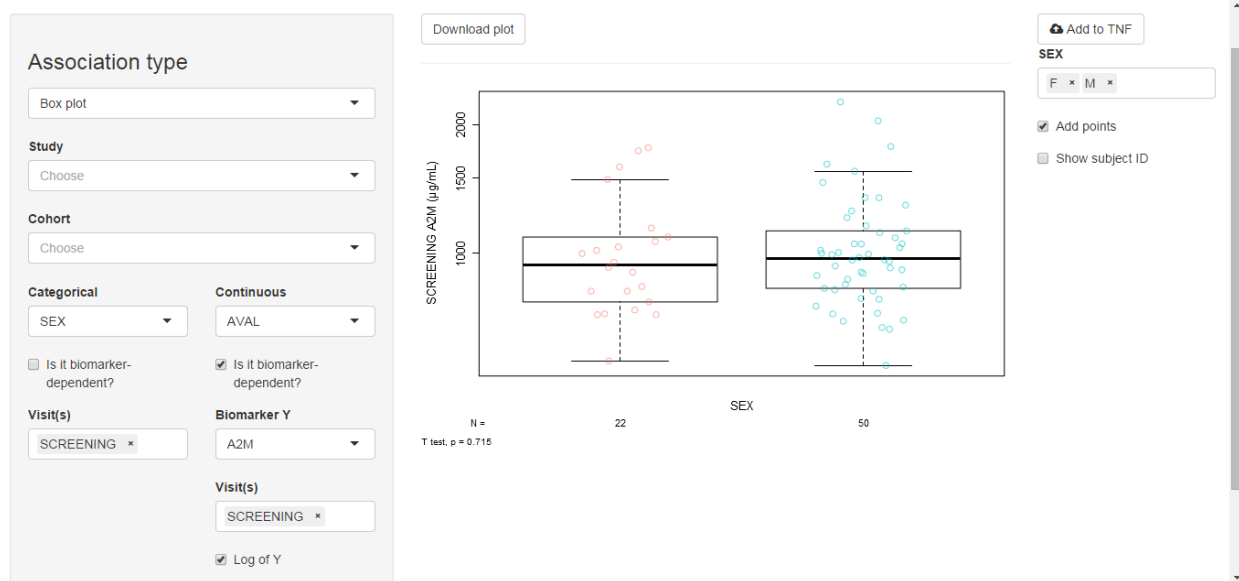
- Categorical: Choose a group variable across which the box plots will be drawn. This can be left blank if the user only wants to draw a box plot of all subjects.
- Continuous: Choose a variable with which the box plot will be drawn. After a variable is specified, a checkbox “Is it biomarker-dependent” will appear. Check the box if the variable is biomarker-dependent, e.g. AVAL, CHG and etc.
- Biomarker Y: The name of the biomarker for the continuous variable if “Is it biomarker-dependent” checkbox is checked.

- Visit(s): Select one or more visits for both group and continuous variables. Note that when multiple visits are specified, visits for X and Y should match with each other.
- Log of Y: Check the box to have semi-log scale.

In addition, the following widget further allows the user to add more information to the scatter plot

- Statistical test: Choose parametric or non-parametric method to perform test between/among all the categories of the group variable specified. The test result will be added as footnote to the box plot.

An example of a box plot is given below

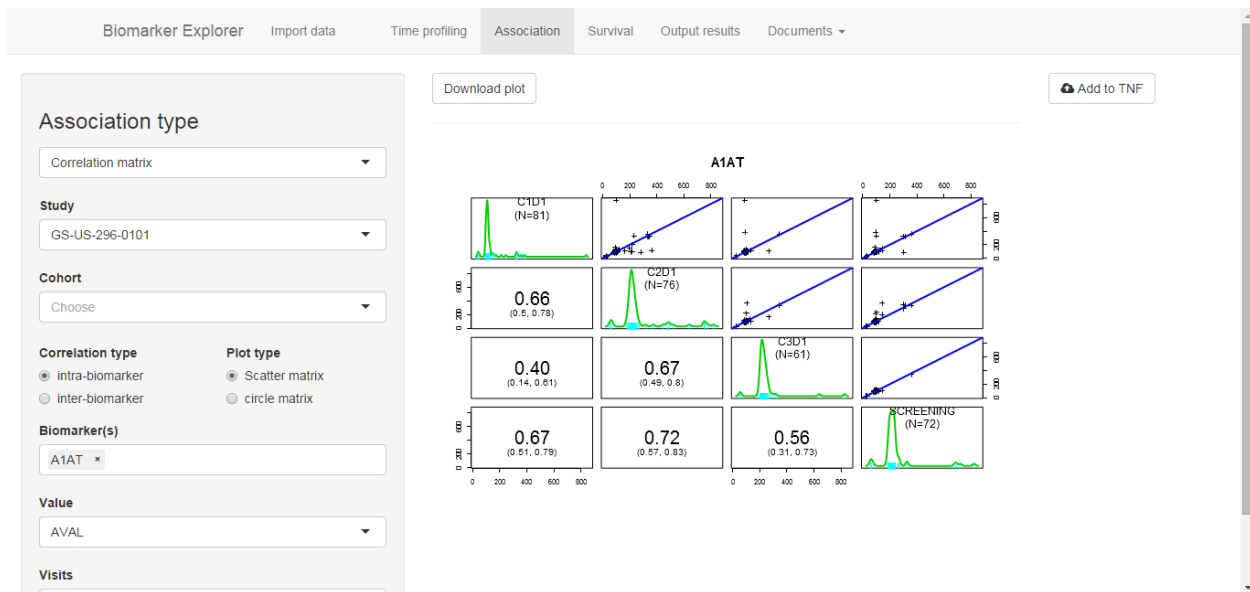


Correlation matrix

To produce a scatter plot, the user needs to further specify the following widgets

- Correlation type: The user needs to decide whether to produce a correlation matrix plot across visits for one biomarker (intra-biomarker) or across biomarkers at one visit (inter-biomarker).
- Plot type: The user needs to specify whether the correlation matrix plot is scatter plot matrix or circle plot matrix.
- Biomarker(s): Specify one or more biomarkers
- Value: Specify the type of the biomarker value, i.e. raw value, baseline value, change from baseline and etc.
- Visit(s): Specify one or more visits

An example of a correlation matrix (intra-biomarker) is given below



Contingency table

To produce a contingency table, the user needs to further specify the following widgets

- **Categorical X/Y:** The two categorical variables in two-way contingency table.
- **Visit(s):** Select one or more visits for both group and continuous variables. Note that when multiple visits are specified, visits for X and Y should match with each other.

In addition, the following widget further allows the user to add more information to the contingency table

- **Add %:** Add an additional column/row for row/column percentages.
- **Statistical test:** The user can choose one of the following contingency table test results to be added as footnote to the table: (1) Chi-square test and (2) Fish exact test

An example of a contingency table is given below

Association type

Contingency table

Study

GS-US-296-0101

Cohort

Choose

Categorical X

RACE

Categorical Y

SEX

Is it biomarker-dependent?

☐

Is it biomarker-dependent?

☐

Visit(s)

SCREENING

Visit(s)

SCREENING

Add %

☒ None ☐ Row ☐ Column

Statistical test

Chi-square

Download table

Add to TNF

Study: GS-US-296-0101

	SEX†		Total
	F	M	
RACE†			
AMERICAN INDIAN OR ALASKA NATIVE	0	2	2
ASIAN	0	1	1
BLACK	3	2	5
NOT PERMITTED	0	4	4
OTHER	0	3	3
WHITE	25	52	77
Total	28	64	92

† RACE (SCREENING) vs SEX (SCREENING)
Chi-square test, p = 0.253

Once the graph or table is shown, one of the following two action buttons will appear to allow the user to download the graph or table

- Download plot
- Download table

Once the user is satisfied with the shown association graph or table and would like to save it to

the final outputs, the user can click

Add to TNF

 button. Consequently, the current specification of the graph or table will be added as a new row to the TNF file, which can be easily previewed or downloaded in “Output results” page.

Step 2. 3 Survival

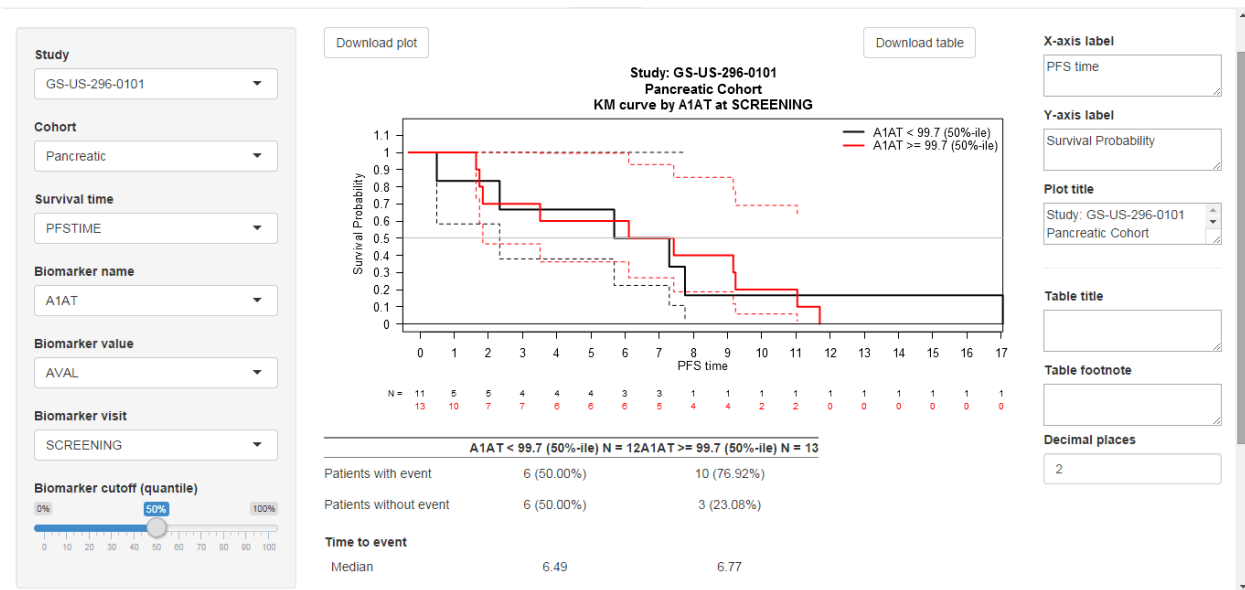
The survival page will be unlocked if one of the following two pairs of columns are present in the data

- PFSTIME and PFSEVENT
- OSTIME and OSEVENT

This page allows the user to plot a KM-curve and a table summarizing the results from a log-rank test. In order to obtain the results, the following widgets need to be specified

- Study: Click to open a drop-down list to choose a specific study to focus on. Leave it blank if the user wants to look at all the studies.
- Cohort: Click to open a drop-down list to choose a specific cohort to focus on. Leave it blank if the user wants to look at all the cohorts.
- Survival time: Specify one of the following two columns that represent the survival time
 - PFSTIME: progress free survival time
 - OSTIME: overall survival time
 If “PFSTIME” is select, “PFSEVENT” will be used as the event column in the survival analysis. Otherwise, “OSEVENT” will be used as the event column.
- Biomarker name: The name of the biomarker
- Biomarker value: Specify the type of the biomarker value, i.e. raw value, baseline value, change from baseline and etc.
- Biomarker visit: Specify the at which visit the biomarker value is to be used in the analysis
- Biomarker cutoff (quantile): Specify a quantile of the biomarker value, at which the biomarker is dichotomized into two levels: high (\geq the cutoff value) and low ($<$ the cutoff value)

An example of KM-curve and log-rank test table is given below



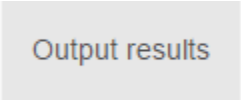
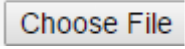

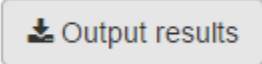
After the graph and table are presented, the user can have the following options to refine the graph and the table

- X-axis label: Change the X-axis label of the KM-curve
- Y-axis label: Change the Y-axis label of the KM-curve

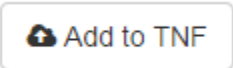
- Plot title: Change the title of the KM-curve
- Table title: Change the title of the table
- Table footnote: Add footnote to the table
- Decimal places: Control the number of decimal places of the numbers in the summary table

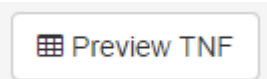
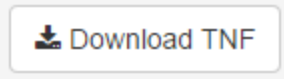
Step 3. Output Results

Step 3.1 Upload TNF file

- Click  on the navigation bar on the top of the platform.
- Click  button to select and upload a TNF file. The standardized format of TNF files for Biomarker Explorer Rshiny tool can be found by clicking the link “Need input data template?” in  to download a template.
- Once the TNF file is uploaded, a button  will appear on the top left of main panel. Click it to save the graphs and/or tables in a zipped file to the directed folder.

Step 3.2 Preview / Download TNF file

Once the user clicked  button in either time-profiling or association page, a TNF file in Excel format will be created and constantly updated at the back end that saves the specification of all the graphs or table the user would like to output. The app provides an easy way to preview or download this TNF file which can be directed uploaded in Step 3.1

- To preview the TNF file, click  button. A tab panel will be automatically created. A “Time Profiling” tab will be created if the user has saved any time-profiling graph specifications. An “Association” tab will be created if the user has save any association graph or table specifications.
- To download the TNF file, click  button. An Excel file containing all the saved graph and/or table specifications will be downloaded to the directory specified by the user.