

BOSS Tools Instructions and Notes

Updated: 2/16/16

General Notes

- Biomarkers, Outcomes and Stats Software (BOSS) is brought to you by Michael McNamara and the Laboratory of Cancer Immunology, led by Dr. William Redmond, Earle A Chiles Research Institute, Providence Cancer Center (Portland, OR) Providence Health and Services. It is licensed under a Creative Commons Attribution 4.0 International License.
- Additional contributors: Ian Hilgart, Diego Barragan.
- The BOSS tools are designed to run in the Google Chrome web browser. They may not function properly in other browsers.
- These instructions are also available in video format @ <http://www.mycancerproject.org/boss/>
- If you would like to contribute to the development of this software, visit the GitHub page @ <https://github.com/theoregonconnection/BOSS/>
- If you would like to report any bugs, have suggestions, or need additional information feel free to send me an email. Michael.McNamara (at) providence.org.

Formatting the Data Table

Helpful Hints

- Use the “BOSS Demo Data File” as a template and replace the existing data with your data. That file can be downloaded @ <http://www.mycancerproject.org/boss/>
- Data tables can be edited in Excel, or similar programs.
- The BOSS tools only read files in CSV format. Use the “Save As” option to save an Excel worksheet in CSV (comma delimited) format.
- Do not use these symbols anywhere in your data table: (), “ ‘ ;
- These symbols should only be used in column headers, where specified: :
- These symbols are OK: { } | * - + . ? & ^ \$ # @ !
- Everything is case-sensitive

Data Types

There are 7 types of data that are recognized by the BOSS programs. The program uses the header of each column to determine the type of data that it will find in that column.

Data Type	Header Text <i>bold = required format Italics = customizable</i>	Customizable Header	Data Must be Numeric?
Patient/Specimen ID	PatientID	No	No
Data Timepoint	Timepoint	No	Yes
Type of Cancer/Disease	CancerType	No	No
Type of Treatment	Treatment	No	No
Group Subset Info	Group - <i>Anything</i>	Yes	No
Outcomes Data	Outcomes - <i>Anything</i>	Yes	Yes
Parameter Data	<i>Category</i> : <i>Parameter</i>	Yes	Yes

Data Types

Required/Standardized

- PatientID
- Timepoint
- Treatment
- CancerType

Optional/Customizable

- Groups
- Outcomes
- Parameters

PatientID	Timepoint	Treatment	CancerType	Group - Mouse Strain	Outcomes - Average Rate of Tumor Growth (m	NK Cells: Percent CD44+
14015 (CT26) P8 P1 PT1_A1_A	8	PD1/CTLA4	CT26	BALBc		3.85
14015 (CT26) P8 P1 PT1_A2_A	8	PD1/CTLA4	CT26	BALBc		40.5
14015 (CT26) P8 P1 PT1_A3_A	8	PD1/CTLA4	CT26	BALBc		57.3
14015 (CT26) P8 P1 PT1_A4_A	8	PD1/CTLA4	CT26	BALBc		25.3
14015 (CT26) P8 P1 PT1_A5_A	8	PD1/CTLA4	CT26	BALBc		37
14015 (CT26) P8 P1 PT1_A6_A	8	PD1/CTLA4	CT26	BALBc		28.9
14015 (CT26) P8 P1 PT1_A7_A	8	PD1/CTLA4	CT26	BALBc		21.6
14015 (CT26) P8 P1 PT1_A8_A	8	PD1/CTLA4	CT26	BALBc		34
14015 (CT26) P8 P1 PT1_A9_A	8	PD1/CTLA4	CT26	BALBc		30.6
14015 (CT26) P8 P1 PT1_A10_A	8	PD1/CTLA4	CT26	BALBc		39.2
14015 (CT26) P8 P1 PT1_A10_A	8	PD1/CTLA4	CT26	BALBc		30.7

Explanation of Data Types

PatientID

The **PatientID** value is a unique identifier for each patient (or animal on study). The PatientID value should be the same for a given patient at all timepoints. The column header is case sensitive and must be “PatientID”.

Timepoint

The **Timepoint** column identifies the timepoint for all of the data in that row. **Every PatientID-Timepoint combination should have its own row, all of the data associated with that specific patient for that timepoint will be in that row.** Timepoints must be in numeric form (eg. do not write “**day 8**”, put in just the number **8**). The timepoint structure is arbitrary – but we generally set the pre-treatment baseline as Timepoint 0. If you use a timepoint of 0, the software will assume that is a baseline value, and will attempt to calculate the change in each parameter value between the baseline and each experimental timepoint. The column header is case sensitive and must be “Timepoint”.

CancerType

The **CancerType** column identifies the type of cancer (or other disease) for the data in that row. The values in this column are case sensitive: **CT26** and **ct26** would be treated as separate types of cancers. The column header is also case sensitive and must be “CancerType”.

*Note: The combination of the CancerType value and Treatment value is used to define a basic cohort for that patient (eg. Cohort = **CT26 – PD1/CTLA4**). These cohorts are used to populate the Cohort Selection options in the user interface.*

Treatment

The **Treatment** column identifies what type of treatment that patient received in this study. This value is designed for a high-level summary of the treatment, rather than an exhaustive description. The values in this column are case sensitive (**PD1/CTLA4** and **pd1/ctla4**) would be treated as separate treatments. The column header is also case sensitive and must be “Treatment”. *Note: Treatment information can be replaced with other types of information (eg. cancer subtype) to create different cohorts (eg. Cohort = **Melanoma – Cutaneous**). These cohorts are used to populate the Cohort Selection options in the user interface.*

Groups

Group columns can be used to identify specific populations of patients (eg. Group – Race). Group columns all begin with “Group - “. The software looks for any headers that begin with “Group - “ and defines whatever comes after the dash as the group category (eg. Group –

Race). The software then reads down the data table and defines all of the options within that group category (eg. Caucasian, Asian, Native American, etc). These values are used to populate the Group Selection options in the user interface. There is no limit to the number of group categories, or the number of subgroups within each category. Group categories should be unique – avoid duplication.

Outcomes

Outcomes columns can be used to identify many types of outcome (eg. Outcomes – Overall Survival Days). Outcomes columns all begin with “Outcomes - “. The software looks for any headers that begin with “Outcomes - “ and defines whatever comes after the dash as a type of outcome (eg. Outcomes – **Overall Survival Days**). Outcomes data must be in numeric form. These values are used to populate the Outcomes Selection options in the user interface. There is no limit to the number of outcome types. Outcome types should be unique – avoid duplication. The data in the Outcomes section is used to populate the x coordinate of an x,y coordinate chart.

Hint: Outcomes data is specific to the timepoint

Parameter Data

Parameter columns can be used to identify any type of type of parameter data (eg. CD4 T Cells: Percent FoxP3+). Parameter columns are identified by the presence of a colon (:) in the column header. Whatever comes before the colon is defined as the parent parameter (eg. CD4 T Cells). Whatever comes after the colon is defined as the specific (child) parameter for that parent (eg. Percent FoxP3+). **The headers of parameter columns are the only place in the data set where the use of colons is allowed.** The software looks for any headers that contain a colon and identifies the parent parameter and child parameter. Parameter data must be in numeric form. The parameter header values are used to populate the Parameter Selection options in the user interface. There is no limit to the number of parameters. Parent-Child parameter pairs should be unique – avoid duplication.


Hint 1: You can add extra sub-sets for both the child and parent by using a dash (eg. CD4 T Cells – Effector Memory: Percent FoxP3+). You can also use different types of data, such as the sample source, as the parent category (eg. Peripheral Blood: CD4 T Cells – Percent FoxP3+). Meaning, you can structure parameter data however you want, as long as you use a single colon.

Hint 2: You can analyze outcomes data as a parameter by duplicating the outcomes column and changing the header on the duplicated column from “Outcomes - “ to “Outcomes: “.

Formatting Rows

Each PatientID-Timepoint pair should have a unique row. All of the data for that patient at that timepoint should be in the same row. **Do not create multiple rows for a patient at the same timepoint.**

Unique Rows for PatientID-Timepoint Pairs



PatientID	Timepoint	Treatment	CancerType
Patient C1	0	Control {Untreated}	Melanoma
Patient C1	21	Control {Untreated}	Melanoma
Patient C1	85	Control {Untreated}	Melanoma
Patient C2	0	Control {Untreated}	Melanoma
Patient C2	21	Control {Untreated}	Melanoma
Patient C2	85	Control {Untreated}	Melanoma

Exporting Data Table from Excel for Analysis

The BOSS tools accept CSV (normal comma delimited) files. They will not read a file in Excel format (eg. .xls files). Use the “Save As” option in Excel to save your data table in CSV format.

Using BOSS Tools to Visualize and Analyze a Data Table

Overview

There are currently two publicly available BOSS tools – **BOSS Relationships** and **BOSS Comparisons**. The **BOSS Relationships** tool analyzes the correlations between biomarkers and outcomes, and characterizes their predictive utility. The **BOSS Relationships** tool generates an interactive Scatter Chart. The **BOSS Comparisons** tool analyzes how parameters differ between cohorts of patients. The **BOSS Comparisons** tool generates an interactive Column Chart. Use of these visualization does not require any special software, just a compatible web browser (Google Chrome), and a compatible Data Table (CSV format). Visualizations are rendered with the Google Charts library (<https://developers.google.com/chart/>).

Accessing the BOSS Tools

Both of the BOSS tools, along with a demo data set and instructions are publicly accessible @ <http://www.mycancerproject.org/boss/>. Source code and development notes can also be accessed on GitHub @ <https://github.com/theoregonconnection/BOSS/>. If you have any thoughts or ideas that you would like to share, or to report bugs, feel free to post your comments on the GitHub page, or email Michael.McNamara@providence.org.

Getting Started

The easiest place to start is by using the **BOSS Demo Data File** (<http://www.mycancerproject.org/boss/>).

Step 1: the **BOSS Demo Data File** to your desktop.

Step 2: Open the **BOSS Relationships** tool by clicking on the link. ***Note: Use Google Chrome with all BOSS tools.***

Webpage @ <http://www.mycancerproject.org/boss/>

Helpful Hint

All visualizations are designed to run in Google Chrome. Other browsers are not supported and may not render properly.

Files

[BOSS Relationships](#)

[BOSS Comparisons](#)

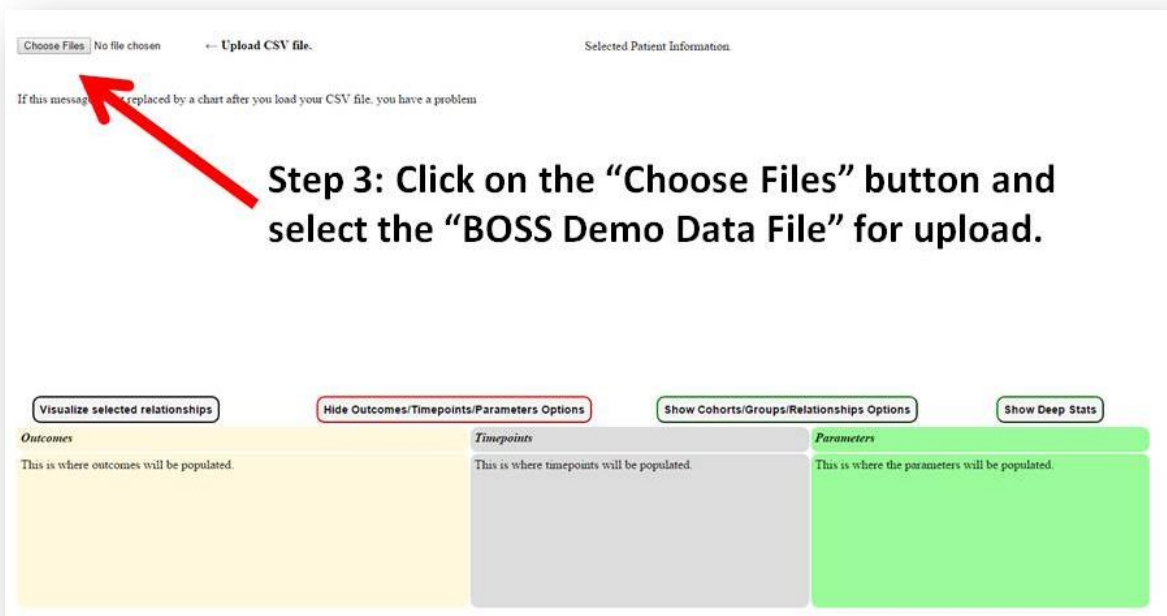
[BOSS Demo Data File](#)

[Instructions](#)

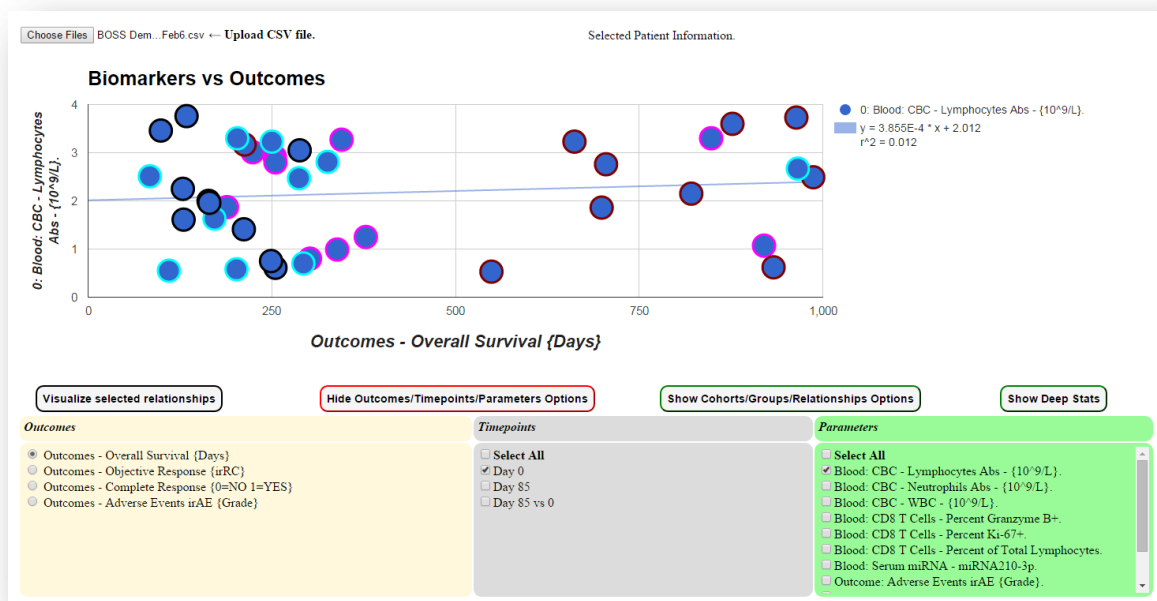
← Step 2: Click on link to open BOSS Relationships.

← Step 1: Download Demo Data File to desktop.

Step 3: Upload the **Boss Demo Data File** by clicking the “**Choose Files**” button. The button should be in the top-left corner of the screen. Select the **Boss Demo Data File** from your desktop.



Step 4: The **BOSS Relationships** tool should populate itself with the uploaded data and look like this:

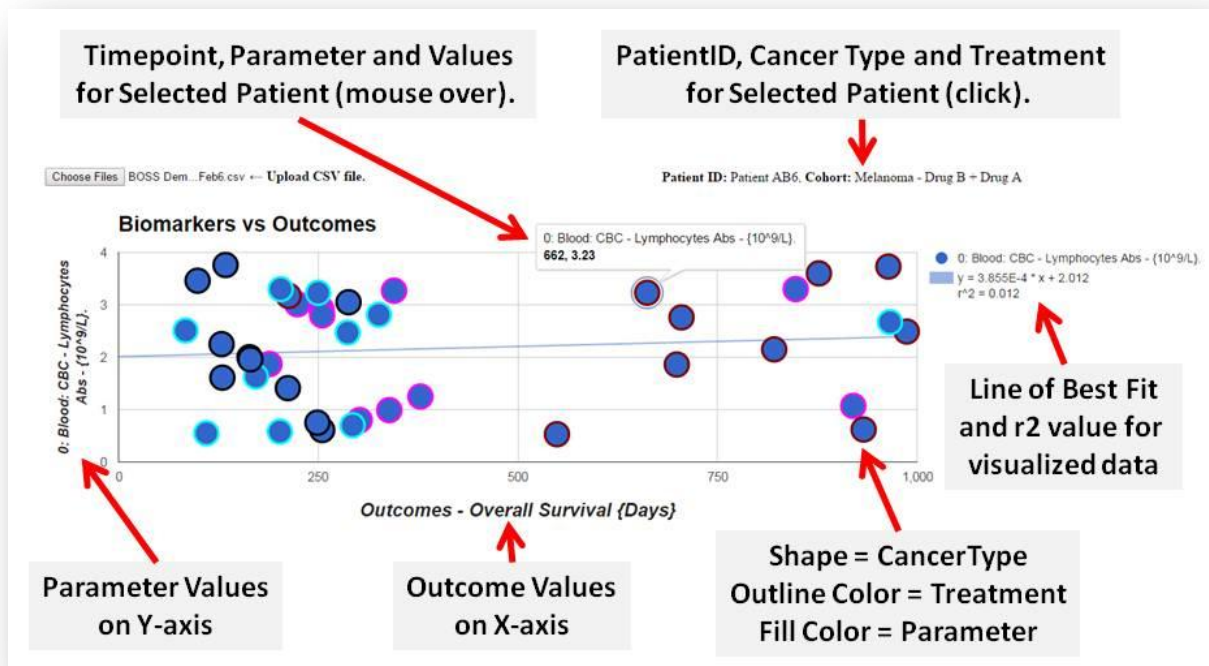


Now you are ready to use the **BOSS Relationships** tool. Uploading data tables to **BOSS Comparisons** works exactly the same. Both tools read the same data tables. For specific information on how to use each tool, see below.

BOSS Relationships

Understanding the Visual Display

- Outcomes values are plotted on the X axis, and parameter values are on the Y-axis.
- Each individual shape represents one patient. The shape corresponds with the **CancerType** value for that patient (eg. Circle = Melanoma). The outline color on the shape corresponds with the **Treatment** value for that patient (eg. Pink = Treatment A).
- Mousing over a shape will open a bubble displaying the **Timepoint**, **Parameter** and values for the outcome and parameter. Clicking on a shape will display the **PatientID** and Cohort (**CancerType – Treatment**) values in the upper-right corner of the screen.
- The fill color on the shapes corresponds to the **Parameter (at a given Timepoint)** being visualized.
- The Line of Best Fit (calculated by linear regression) and r^2 value for each Outcome-Parameter relationship being visualized is displayed on the right side of the screen.



Understanding the Selection Options

All of the selection options are dynamically generated from the uploaded data table. The goal of the selection system is to maximize the number of relationships that the user can visualize and analyze. The selection options are in tabs that can be toggled into and out of view by clicking on the buttons immediately below the chart. Click the “Visualize Selected Relationships” button to update the chart with your selections.

Updates chart to show selected data

Visualize selected relationships

Toggles Selectors for:
Outcomes
Timepoints
Parameters

Hide Outcomes/Timepoints/Parameters Options

Toggles Selectors for:
Cohorts
Groups
Basic (Total) Stats

Show Cohorts/Groups/Relationships Options

Toggles Stats For:
Cohorts
Groups
Cohort-Groups

Show Deep Stats

Outcomes

- Outcomes - Overall Survival (Days)
- Outcomes - Objective Response (uRC)
- Outcomes - Complete Response (0=NO 1=YES)
- Outcomes - Adverse Events uAE (Grade)

Timepoints

- Select All
- Day 0
- Day 85
- Day 85 vs 0

Parameters

- Select All
- Blood: CBC - Lymphocytes Abs - (10⁹/L)
- Blood: CBC - Neutrophils Abs - (10⁹/L)
- Blood: CBC - WBC - (10⁹/L)
- Blood: CD8 T Cells - Percent Granzyme B+
- Blood: CD8 T Cells - Percent Ki-67+
- Blood: CD8 T Cells - Percent of Total Lymphocytes
- Blood: Serum miRNA - miRNA210-3p
- Outcome: Adverse Events uAE (Grade)

Outcomes

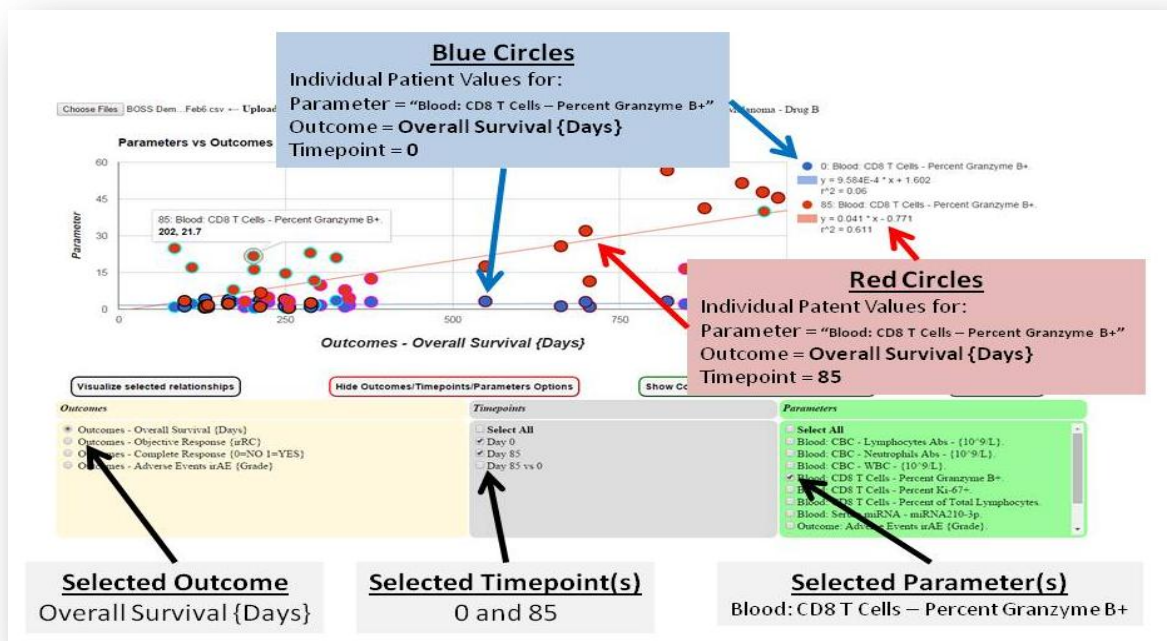
Outcomes selection boxes are generated from the “Outcomes – ” column headers in the uploaded data table. Only one outcome can be selected at a time.

Timepoints

The timepoint selection menu is populated with the timepoints from the uploaded data table. Multiple timepoints can be selected at the same time. The fill color of the shapes reflects the visualized Parameter and Timepoint. The same parameter at different timepoints will be represented by different colors. If there is a baseline timepoint (Timepoint = 0) in the data set and additional experimental timepoints, the software will attempt to evaluate the difference between each parameter at each experimental timepoint and the baseline value, for each patient. The software creates a new timepoint that represents those differentials (eg. **Day 85 vs 0**) and populates that into the timepoint selection menu. In this example, selecting this option will show you the change in the selected parameter between Timepoint **85** and Timepoint **0**. These differential values are calculated by subtracting the parameter value at the baseline from the parameter value at the experimental timepoint for each patient, where both values exist.

Parameters

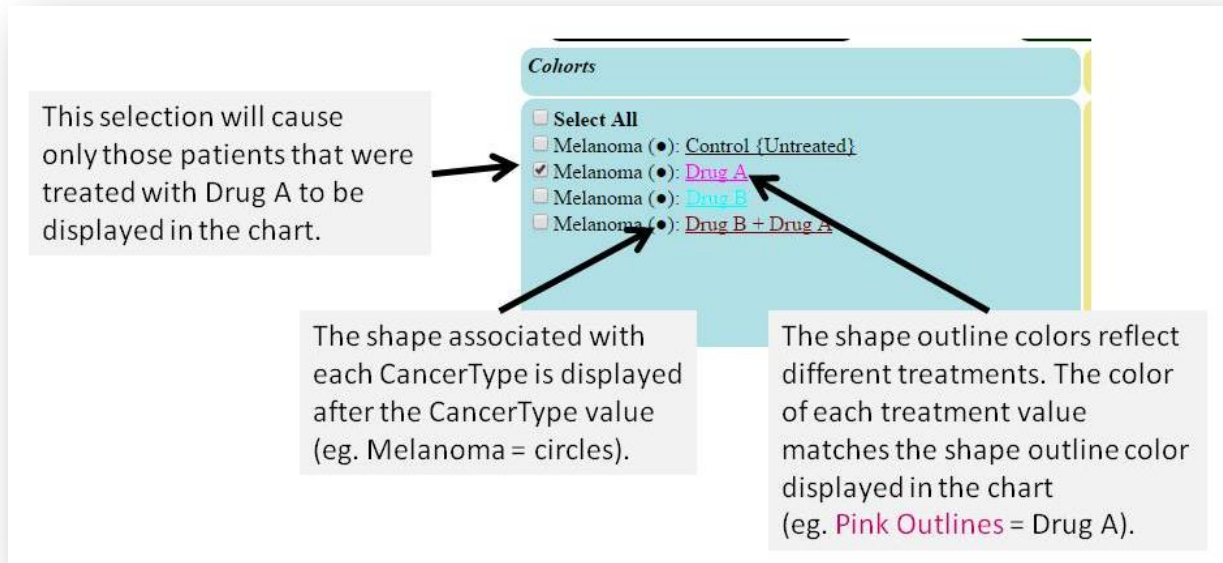
The parameter menu populated with the parameters defined in the column headers of the uploaded data table. Parameter options will automatically sort into alphabetical order. Multiple parameters can be selected at the same time.



Cohort Selection Options

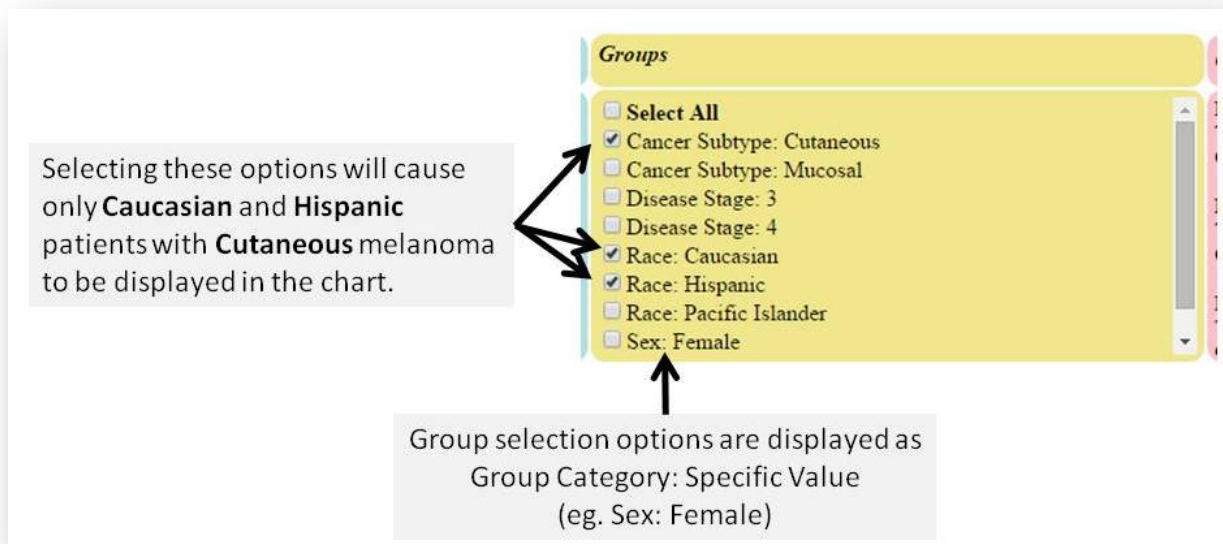
The **Cohorts Selection Options** are created by dividing the total patient population into cohorts based on their CancerType and Treatment values (eg. Melanoma – Drug A). Multiple cohorts can be selected at the same time. The shape of the data points and their outline color are defined by the patient's CancerType and Treatment, respectively. *Note: There are a limited number of shapes and colors available, so only the first 7 cancer types will receive unique shapes, the*

remainder will all be circles. Only the first 10 treatments will receive unique outline colors, the remainder will have no outline. Checking the “Select All” option will show all of the cohorts, regardless of any individual selections. When the chart is updated by clicking the “Visualize selected relationships” button, only the selected cohorts will be visualized. The line of best fit (LOBF) equation and r^2 value will automatically update to reflect the subset of patients being visualized. When **Cohorts** and **Groups** selection options are both used, only those patients that meet all of the selected criteria will be shown.



Group Selection Options

The **Groups Selection Options** are created from the **Groups** categories in the data table **and** the options contained in those columns. For example, a Groups data column may have the header “Groups – Sex”, and the possible values could be “Male”, “Female” and “Unknown”. In this case, “Sex” is the category. When the software identifies a **Groups** column, it will automatically read down that column and make a list of all of the unique values it contains. This data is then populated into the **Groups Selection Options** in the format of *Group Category: Specific Value* (eg. Sex: Male). Selecting just the “Sex: Male” option will cause the chart to show only data from male patients. Multiple selections within the same category are additive (eg. Selecting both “Sex: Male” **and** “Sex: Female” will cause both male and female patients to be displayed). Selections from different categories are subtractive (eg. Selecting both “Sex: Male” **and** “Age: 30-39 years” will cause only patients who are male **and** 30-39 years of age to be displayed). There is no limit to the number of group selections that can be made. Checking the “Select All” option will show everyone, regardless of any individual selections. When the chart is updated by clicking the “Visualize selected relationships” button, only the patients who meet all the selected criteria will be visualized. The line of best fit (LOBF) equation and r^2 value will automatically update to reflect the subset of patients being visualized. When **Cohorts** and **Groups** selection options are both used, only those patients that meet all of the selected criteria will be shown.



Statistical Analyses in the BOSS Relationships Tool

Overview

All of the statistical analyses in the BOSS relationships program use the same basic approach.

1. Parameter-Outcome relationships with less than 5 value pairs are ignored.
2. For the subset of patients being analyzed, the relationship between the selected parameter and the selected outcome is analyzed by conventional linear regression (least squares), which generates an equation for the line of best fit (LOBF) and an r^2 value for the goodness of fit.
3. A second r^2 value is calculated after excluding the value points that contain the max and min values for both the parameter and outcome, in that data set. This second r^2 value is averaged with the original r^2 value to create the " r^2 {adj}" . The point of this is to make it easier to flag correlations that are driven by values at the extremes of the data set.

Note: I'm not sure this is actually useful, and it will probably be removed when we update the software. Most users should probably just ignore this value.

4. Using the LOBF determined by linear regression, the probability that the parameter predicts the outcome is estimated in the following manner:
 - a. The null hypothesis is assumed to be a line with a slope of 0 and a Y-intercept at the mean of the selected parameter values.
 - b. The highest and lowest parameter values in the selected data set are used to define the observed range for that parameter.
 - c. For each individual data point, the software estimates the probability that the data point is as close to the LOBF as it is observed to be. It does this by

calculating the vertical distance between the data point and the LOBF. Then it calculates the maximum possible vertical distance that a data point could be from the LOBF at that X position (based on the observed range of parameter values). The probability estimate for that data point's proximity to the LOBF is defined as (observed distance / max distance). The same method is used to estimate the probability that the data point is as close to the null hypothesis line as it is observed to be. This process is repeated for every data point in the selected data set. All of these values are collected into two groups: LOBF probabilities and null hypothesis probabilities.

- d. The group of LOBF probabilities is compared the group of null hypothesis probabilities by *t-test*. The p value generated by this *t-test* is an estimate of the probability that the group of LOBF probabilities are **not** significantly different than the group of null hypothesis probabilities. This p value is what is reported in the stats outputs.
5. *It is important to keep in mind that the statistical approach used in this software is not intended to be a definitive analysis of the tested relationships. The point is to make it easy to identify relationships that are **likely** to be meaningful – from large and complex data sets. Furthermore, the code for this software is open-source, and we encourage members of the research community to add/improve the statistical capabilities where they see fit.*

Individual Statistical Analyses

Overall Relationships

The software attempts to evaluate the relationship between every parameter and every outcome at every timepoint where sufficient data exists. For the **Overall Relationships** analysis the software evaluates all of the patients in the data table with the requisite parameter and outcome data, irrespective of their group or cohort. The statistical analysis is described above. The relationships are sorted by p value (lowest to highest), and the top ~5000 are populated into the **Overall Relationships** window. These results can be downloaded in CSV format by clicking on the “Export” button in the window header.

Toggle Groups, Cohorts and Relationships Options

Click to Download Results In CSV (Excel) Format

Visualize selected relationships | Show Outcomes/Timepoints/Parameters Options | Hide Cohorts/Groups/Relationships Options | Show Deep Stats

Cohorts

- Select All
- Melanoma (•) Control (Untreated)
- Melanoma (•) Drug A
- Melanoma (•) Drug B
- Melanoma (•) Drug B + Drug A

Groups

- Select All
- Cancer Subtype: Cutaneous
- Cancer Subtype: Mucosal
- Disease Stage: 3
- Disease Stage: 4
- Race: Caucasian
- Race: Hispanic
- Race: Pacific Islander
- Sex: Female

Overall Relationships | Export

P Value: 4.51e-13, r2: 0.9992, r2 (adj): 0.9994
 Timepoint-Parameter: Day 85: Outcome: Overall Survival (Days)
 Outcome: Outcomes - Overall Survival (Days)

P Value: 4.51e-13, r2: 0.9992, r2 (adj): 0.9994
 Timepoint-Parameter: Day 0: Outcome: Overall Survival (Days)
 Outcome: Outcomes - Overall Survival (Days)

P Value: 1.58466e-11, r2: 0.9944, r2 (adj): 0.89
 Timepoint-Parameter: Day 85: Outcome: Objective Response (inRC)
 Outcome: Outcomes - Overall Survival (Days)

Top 5000 Parameter-Outcome correlations sorted by p value (includes all patients in data table, Irrespective of cohort or group)

Deep Stats

The purpose of the deep stats analysis is to identify specific subsets of patients where there may be unique Parameter-Outcome correlations that are more robust for that subset than for the patient population as a whole. *Note: In some cases, there may be an insufficient number of patients for analysis. In these cases, the relationship summary windows will be empty.*

Download Results From Each Analysis

Toggle Deep Stats Tab

Visualize selected relationships | Show Outcomes/Timepoints/Parameters Options | Show Cohorts/Groups/Relationships Options | Hide Deep Stats

Cohort Relationships | Export

P Value: 0.0032922902412568, r2: 0.7984, r2 (adj): 0.8039
 Timepoint-Parameter: Day 85: Blood: CD8 T Cells - Percent Granulysin B+
 Outcome: Outcomes - Complete Response (0=NO, 1=YES)
 Cohort: Melanoma - Drug B + Drug A

P Value: 0.0035718098947537, r2: 0.8978, r2 (adj): 0.8344
 Timepoint-Parameter: Day 85: Blood: CD8 T Cells - Percent Granulysin B+
 Outcome: Outcomes - Objective Response (inRC)
 Cohort: Melanoma - Drug B

Group Relationships | Export

P Value: 5.30349125e-8, r2: 0.8706, r2 (adj): 0.8625
 Timepoint-Parameter: Day 85: Outcome: Overall Survival (Days)
 Outcome: Outcomes - Objective Response (inRC)
 Group: Group - Cancer Subtype: Cutaneous

P Value: 8.09925807e-8, r2: 0.8784, r2 (adj): 0.8732
 Timepoint-Parameter: Day 0: Outcome: Overall Survival (Days)
 Outcome: Outcomes - Objective Response (inRC)
 Group: Group - Cancer Subtype: Cutaneous

P Value: 1.101001744e-7, r2: 0.9020, r2 (adj): 0.9416

Cohort-Group Relationships | Export

P Value: 0.0920617153010749, r2: 0.6041, r2 (adj): 0.582
 Timepoint-Parameter: Day 0: Blood: Serum miRNA - miRNA210-3p
 Outcome: Outcomes - Overall Survival (Days)
 Cohort: Melanoma - Control (Untreated)
 Group: Group - Cancer Subtype: Cutaneous

P Value: 0.4556237848875272, r2: 0.1191, r2 (adj): 0.1923
 Timepoint-Parameter: Day 0: Blood: CD8 T Cells - Percent of Total Lymphocytes
 Outcome: Outcomes - Overall Survival (Days)
 Cohort: Melanoma - Control (Untreated)
 Group: Group - Cancer Subtype: Cutaneous

Correlation Analysis Within Each Cohort

Correlation Analysis Within Each Group

Correlation Analysis Within Each Group Within Each Cohort

Cohort Relationships

The point of the **Cohort Relationships** analysis is to identify the most robust Parameter-Outcome correlations within specific cohorts (groups of patients defined by their CancerType

and Treatment). To generate **Cohort Relationships** the software runs the correlation analyses described above for each individual cohort. The result sets for each cohort are then combined, and these combined results are then sorted by p value. The top ~5000 correlations are displayed in the **Cohort Relationships** window. These results can be downloaded in CSV format by clicking on the “Export” button.

Group Relationships

The point of the **Group Relationships** analysis is to identify the most robust Parameter-Outcome correlations within specific subgroups of patients, such as male patients (Sex: Male). To generate **Group Relationships** the software runs the correlation analyses described above for each individual sub-group within each group category. The result sets for all of these analyses are then combined, and the results are sorted by p value. The top ~5000 correlations are displayed in the **Group Relationships** window. These results can be downloaded in CSV format by clicking on the “Export” button.

Cohort-Group Relationships

The point of the **Cohort-Group Relationships** analysis is to identify the most robust Parameter-Outcome correlations within specific subgroups of patients within each cohort, such as patients with melanoma who received Drug A and are also male (CancerType = Melanoma **and** Treatment = Drug A **and** Sex = Male). To generate **Cohort-Group Relationships** the software runs the correlation analyses described above for each individual sub-group within each group category, for every cohort. The result sets for all of these analyses are then combined, and the results are sorted by p value. The top ~5000 correlations are displayed in the **Cohort-Group Relationships** window. These results can be downloaded in CSV format by clicking on the “Export” button.

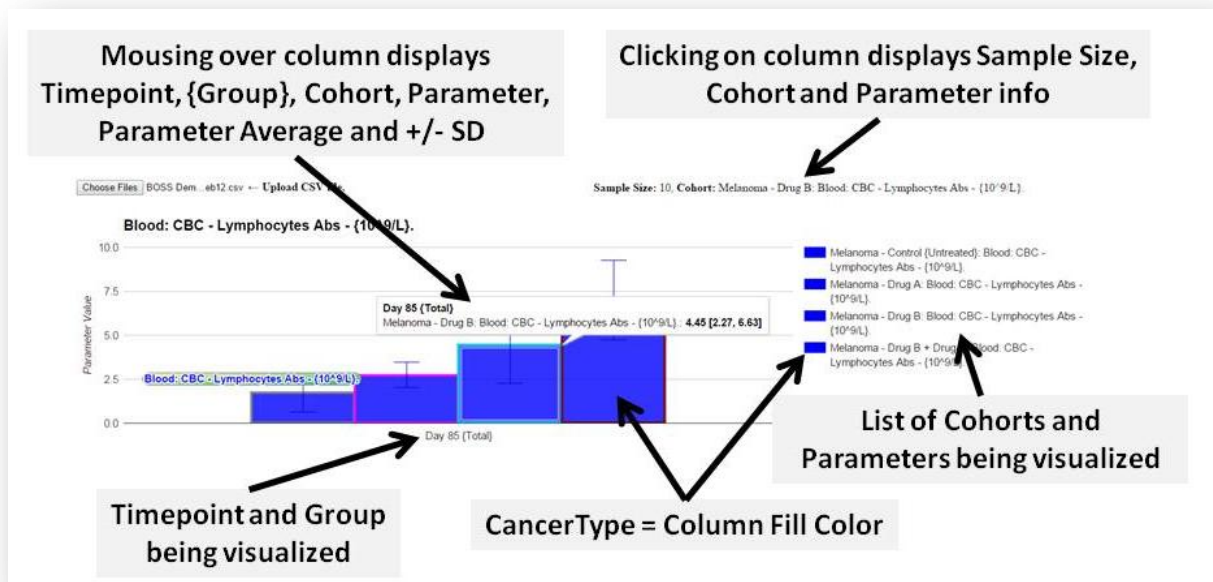
BOSS Comparisons

Overview

The purpose of the BOSS Comparisons tool is to make it easier to compare parameters between subsets of patients on the basis of CancerType, Treatment, Timepoint and any kind of Group category that you wish to define. Additionally, the goal is to facilitate the statistical comparison between subsets and to make it easier to mine complex data sets and identify where significant differences exist.

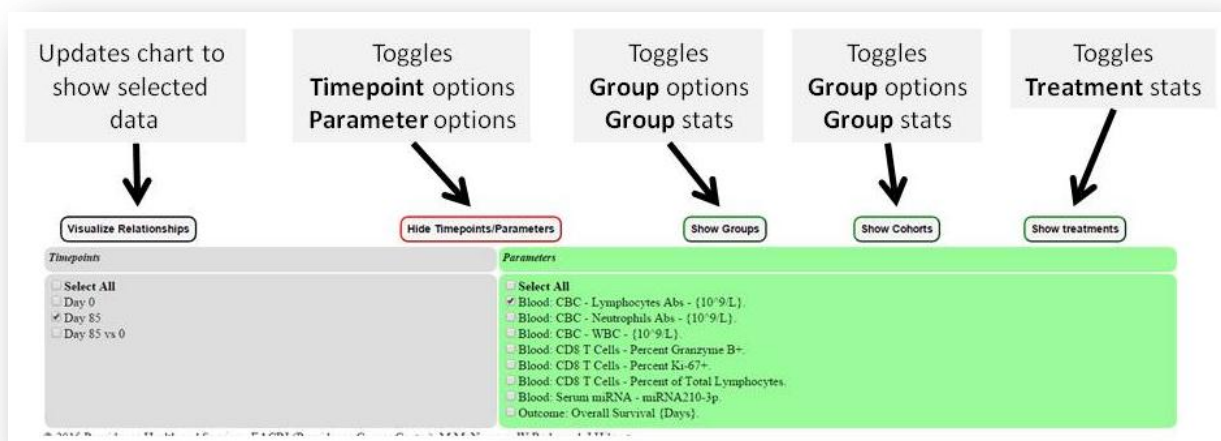
Understanding the Visual Display

- Parameter values are plotted on the Y axis.
- The X-axis displays the **Timepoint** and the **Group** being visualized (eg. Day 0 {Sex:Male}). The default group value is {Total}, which includes parameter values from all patients, irrespective of their specific group values.
- The fill color on the columns corresponds to the **CancerType** being visualized.
- The outline color corresponds to the **Treatment**.
- The height of each column reflects the average of the **Parameter** values for all of the patients in the cohort defined by the **CancerType** and **Treatment** being visualized. Patients with no data for a given parameter are excluded. The error bars reflect the standard deviation of the parameter values within the cohort.
- Black columns are composites that include parameter values from all of the patients in a given **Treatment** group, irrespective of **CancerType**.
- If there is no **Parameter** data for a selected **Cohort** there will be blank space on the chart where the column would normally be.
- Mousing over a column will open a bubble displaying the **Timepoint, Parameter, Cohort (CancerType and Treatment)**, average parameter value and standard deviation. Clicking on a column will display the sample size (n), Cohort and Parameter in the upper-right corner of the screen.



Understanding the Selection Options

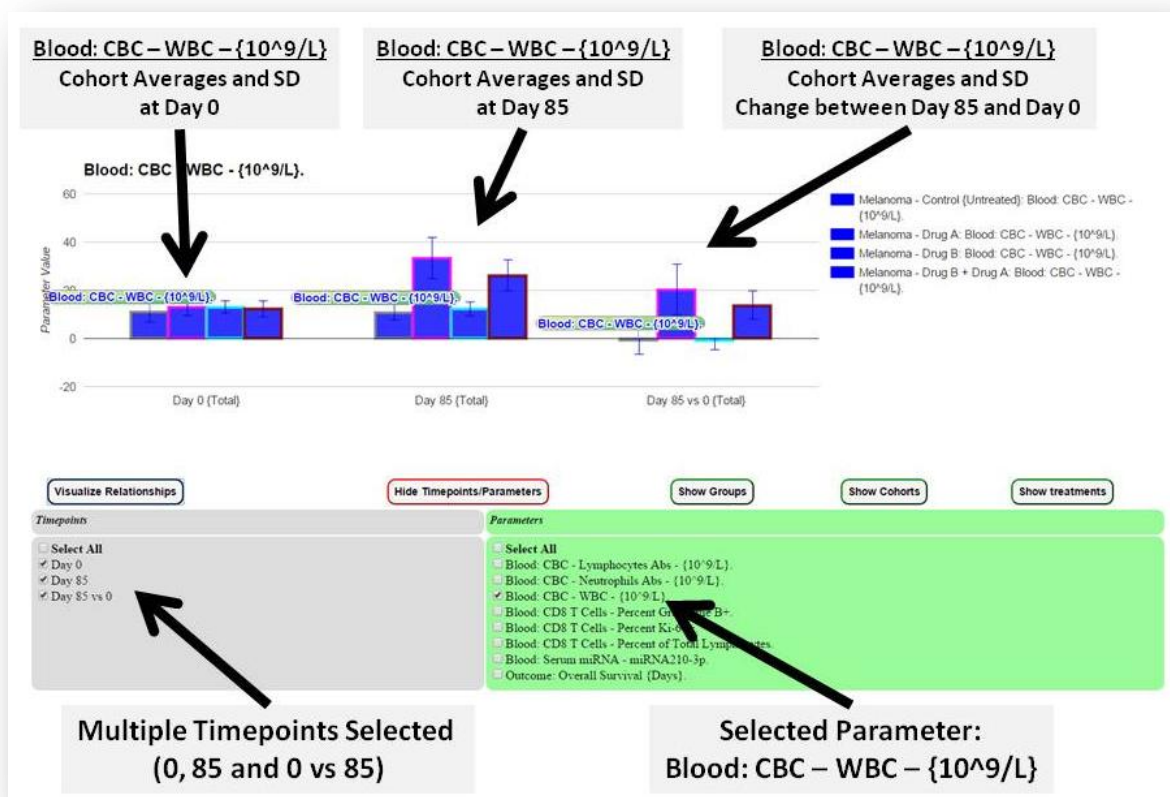
All of the selection options are dynamically generated from the uploaded data table. The goal of the selection system is to maximize the number of relationships that the user can visualize and analyze. The selection options are in tabs that can be toggled into and out of view by clicking on the buttons immediately below the chart. Click the “Visualize Relationships” button to update the chart with your selections.



Timepoints

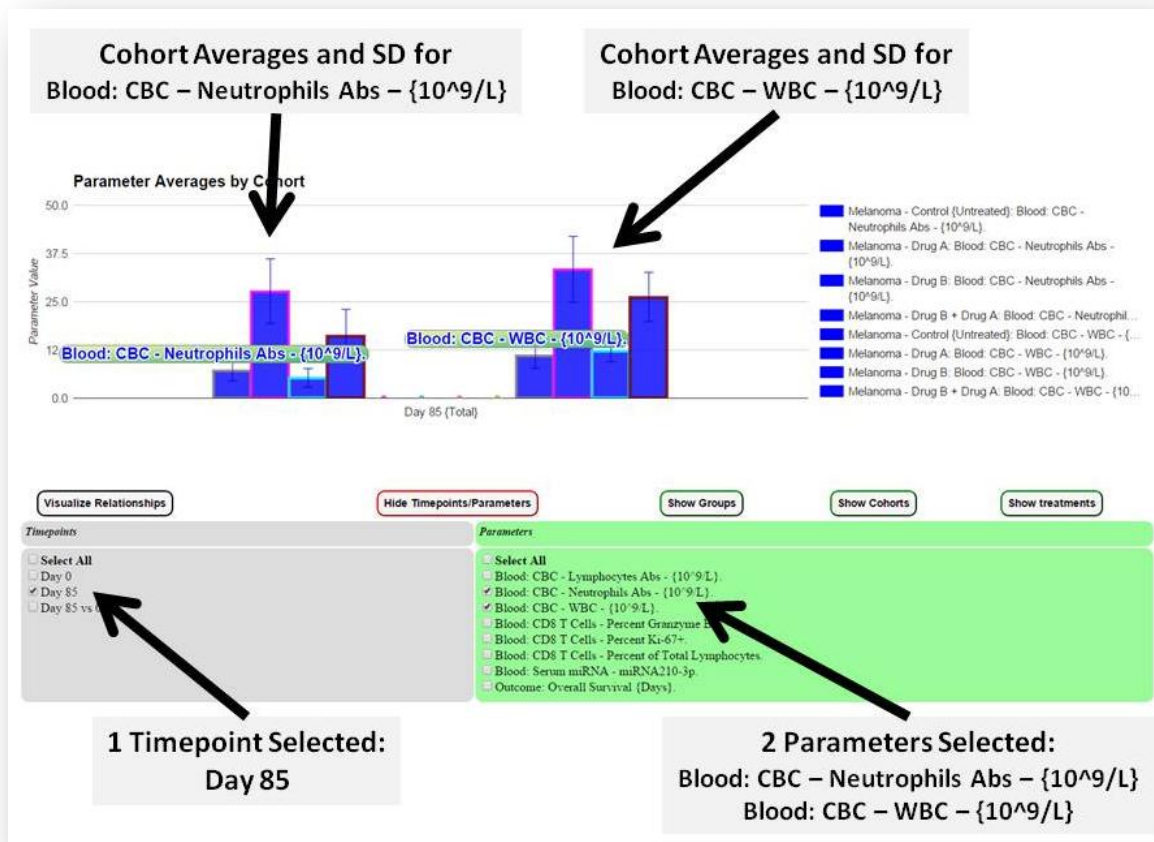
The **Timepoints** selection menu is populated with the timepoints from the uploaded data table. Multiple timepoints can be selected at the same time. When multiple timepoints are selected,

the chart spatially separates the data by timepoint. If there is a baseline timepoint (Timepoint = 0) in the data set and additional experimental timepoints, the software will attempt to evaluate the difference between each parameter at each experimental timepoint and the baseline value, for each patient. The software creates a new timepoint that represents those differentials (eg. Day 85 vs 0) and populates that into the timepoint selection menu. In this example, selecting this option will show you the change in the selected parameter between Timepoint 85 and Timepoint 0. These differential values are calculated by subtracting the parameter value at the baseline from the parameter value at the experimental timepoint for each patient, where both values exist.



Parameters

The **Parameters** menu is populated with the parameters defined in the column headers of the uploaded data table. Parameter options will automatically sort into alphabetical order. Multiple parameters can be selected at the same time. When multiple parameters are selected, they will be separated by 4 colored bumps when the chart by four colored bumps and they will be accompanied by floating bubbles with that identify the parameter. *Note: If the first selected cohort does not have data for the selected parameter, the identification bubble may not appear.*



Group Selection Options

Click on the “Show Groups” button to toggle the Group selection options and stats window.

Group Selection Options are created from the Groups categories in the data table *and* the options contained in those columns. In the selection menu both the group category and subgroup are listed in the format of **Category: Sub-group** (eg **Sex: Female**). The default selection is “Total {No Filter}”, which includes all available data, irrespective of Group information.

Selecting a specific group option will create a set of columns on the chart with data from only that selection. Group selections are not additive or subtractive – each selection will be displayed as its own subset on the chart.

Group Subset = Total {No Filter}

Parameter = Blood: CD8 T Cells – Percent Ki-67+

Timepoint = Day 85

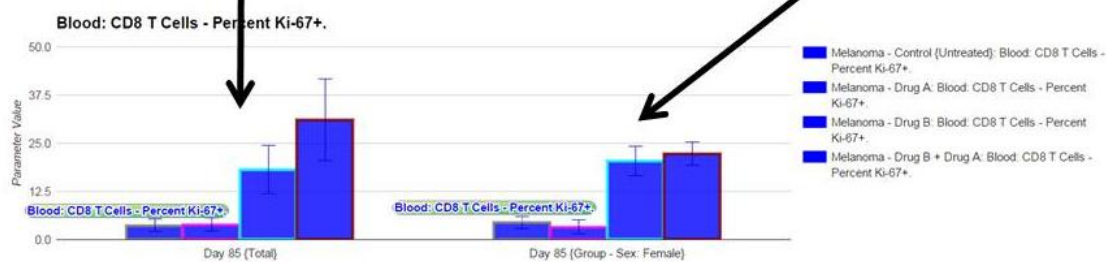
All available data, irrespective of group

Group Subset = Sex: Female

Parameter = Blood: CD8 T Cells – Percent Ki-67+

Timepoint = Day 85

Only data from Female patients



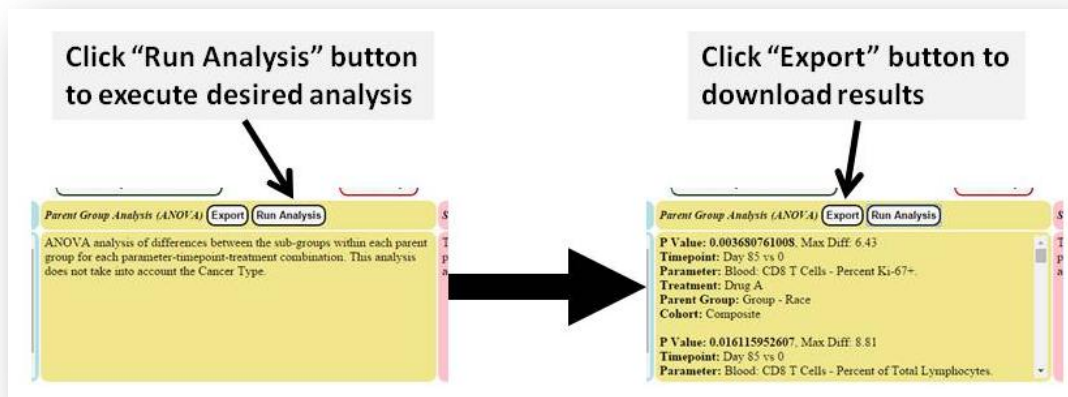
2 Group Subsets Selected:

Total {No Filter} – Shows all available data, irrespective of group
Sex: Female – Shows only data from Female patients

Notes on BOSS Comparison Statistical Analyses

The BOSS Comparison tool has 2 statistical tests to analyze how parameters differ between subsets of patients, and identify where significant differences exist. Regardless of the context, the tests use conventional ANOVA or t-test (Welch's). Because these statistical analyses can take a significant amount of time to complete (especially with larger data sets), they are not automatically executed when the chart is initiated. Instead, each analysis can be executed *a la carte* by clicking the "Run Analysis" button. The results will then populate into the frame when complete. These results can be downloaded in CSV format by clicking on the "Export" button.

Note: For large and complex data tables, statistical analyses can take more than 1 min. In these cases your browser may indicate that it is not responding. If you leave the browser alone for a few minutes, the stats processing will usually complete and the results will populate into the frame. We are working on adding a time estimate and/or progress bar and hope to incorporate that into the next software update.



Group Statistical Analyses

The BOSS Comparison tool has 2 statistical tests to analyze how parameters differ between group subsets, and identify where significant differences exist. The point of these tests is to address the question: "For every parameter at every timepoint, are there significant differences between sub-groups?" Group Statistical Analyses **only** compare sub-groups within the same group category (eg. Group Category is "Sex" and the sub-groups being compared are "Male" and "Female"). The BOSS Comparison tool does not support comparison between unrelated sub-groups (eg. "Male" vs "Age 30-39"). The ANOVA analysis is designed to highlight group categories where significant differences exist and the t-test analysis is designed to identify specific sub-groups where there are significant differences between for a given parameter (at a specific timepoint). The ANOVA analysis requires that a group category has at least 3 sub-groups, all other group categories will be ignored. The results menu is sorted on the basis of p value and includes the group, parameter, timepoint and max parameter value difference for each relationship analyzed. *Important note: The Group Statistical Analyses evaluate only the **Composite Cohort**, which includes all patients in each **Treatment** group, irrespective of their **CancerType**.*

P Value and Max Parameter Value Difference Between Compared Groups

Timepoint Being Analyzed

Parameter Being Analyzed

Comparison Group #1 (Reference)

Comparison Group #2 (Target)

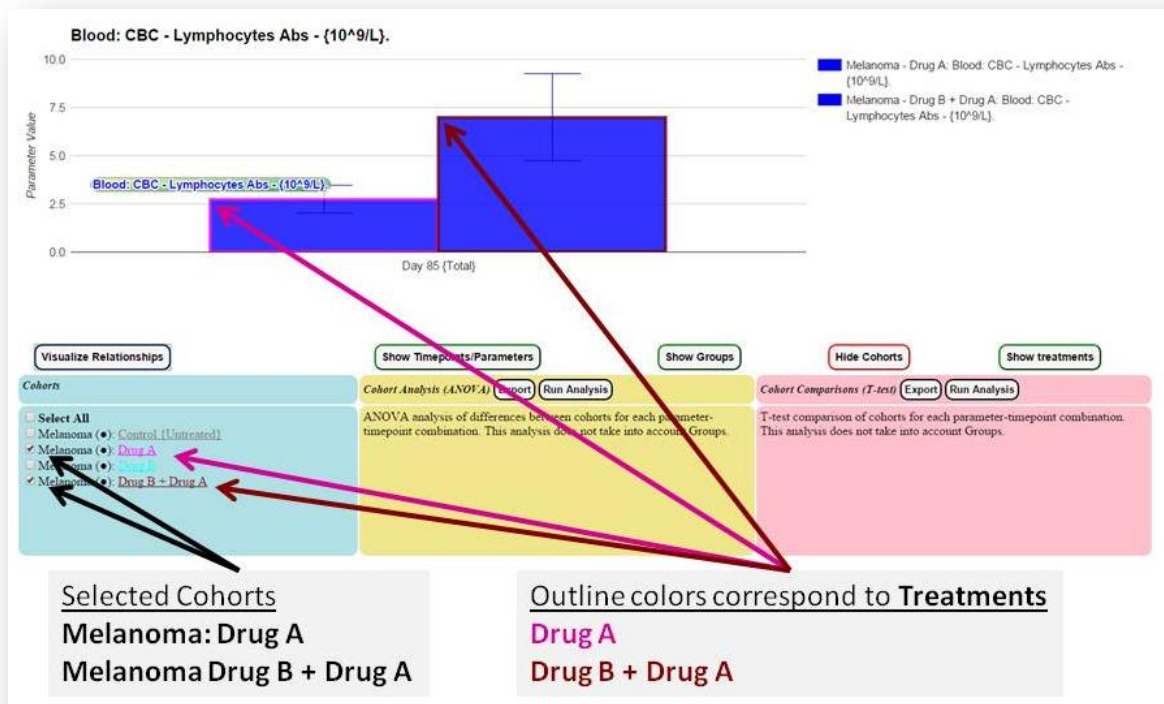
Cohort Being Analyzed
All patients are grouped by **Treatment**, irrespective of the **CancerType**. In this example, the parameter differences between groups are being analyzed for the subset of patients who were treated with "Drug A".

Sub-Group Comparisons (T-test) **Export** **Run Analysis**

P Value: 2.019506458e-7, Max Diff: 3.7
Timepoint: Day 85 vs 0
Parameter: Blood: CD8 T Cells - Percent Ki-67+.
Treatment: Drug A
Reference Group: Group - Sex: Female
Target Group: Group - Sex: Male
Cohort: Composite

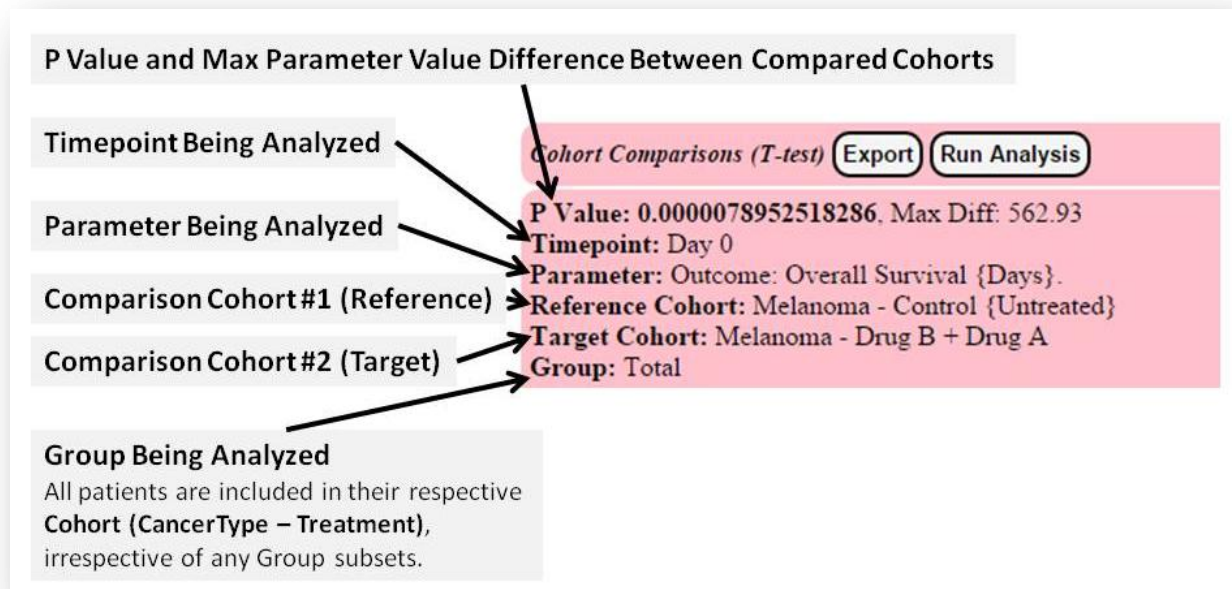
Cohorts Selection Options and Stats

To toggle the Cohorts Selection Options and Stats tab, click on the "Show Cohorts" button. The Cohorts selector allows you to choose specific patient **Cohorts**, which are defined by CancerType: Treatment (eg. Melanoma: Drug A). If there are more than 2 CancerTypes in the data table, the software will automatically attempt to create a "Composite" cohort, which includes all patients who received a given treatment, irrespective of their CancerType (eg. Composite: Drug A). All composite columns have the fill color of **Black**. The outline color on the columns reflects the **Treatment**. In the cohort selection window, the treatments are color-coded, and these colors match the column outline colors on the chart.



Cohort Statistical Analyses

The BOSS Comparison tool has 2 statistical tests to analyze how parameters differ between cohorts (CancerType: Treatment subsets), and identify where significant differences exist. The point of these tests is to address the question: “For every parameter at every timepoint, are there significant differences between cohorts?” The ANOVA analysis is designed to highlight cohorts where significant differences exist and the t-test analysis is designed to identify specific pairs of cohorts where there are significant differences between for a given parameter (at a specific timepoint). The ANOVA analysis requires that at least 3 cohorts exist in the data set. The results menu is sorted on the basis of p value and includes the cohorts, parameter, timepoint and max parameter value difference for each relationship analyzed. *Important note: Cohort Statistical Analyses do not take into account Group information. All patients with a given CancerType and Treatment are included in the respective cohort.*



Treatment Statistical Analyses

The BOSS Comparison tool has 2 statistical tests to analyze how parameters differ between patients on the basis of **Treatment** and identify where significant differences exist. The point of these tests is to address the question: “For every parameter at every timepoint, are there significant differences between treatment groups?” The **Treatment Statistical Analyses** do not take into account CancerType or Group information. They use the Composite Cohorts for comparative analysis. The ANOVA analysis is designed to highlight parameters where significant differences exist between treatment groups and the t-test analysis is designed to identify specific pairs of treatment groups where there are significant differences (at a specific timepoint). The ANOVA analysis requires that at least 3 **Treatments** exist in the data set. The results menu is sorted on the basis of p value and includes the cohorts, parameter, timepoint and max parameter value difference for each relationship analyzed. *Important note: There is currently no additional selection menu for Treatments. Treatment-specific selection options were incorporated into the Cohorts Selection Options.*

P Value and Max Parameter Value Difference Between Compared Treatments

Timepoint Being Analyzed →

Parameter Being Analyzed →

Comparison Treatment #1 (Reference) →

Comparison Treatment #2 (Target) →

Cohort Being Analyzed
All patients are included in their respective **Treatment** group, irrespective of any CancerType or Group subsets.

Treatment Comparisons (T-test) **Export** **Run Analysis**

P Value: 0.0000078952518286, **Max Diff:** 562.93
Timepoint: Day 0
Parameter: Outcome: Overall Survival {Days}.
Reference Treatment: Control {Untreated}
Target Treatment: Drug B + Drug A
Cohort: Composite

Software License Disclaimer (modified MIT)

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

The license may not give you all of the permissions necessary for your intended use. This software uses code developed by Google, HighCharts and others that is subject to additional conditions. For example, some of the components used in this software include licenses that may limit how you use this software. Users may copy, display, make derivatives of, and distribute this software to others for non-commercial purposes, but must maintain the copyright notice above and this permission notice in all copies.