

Multiplot Studio v1.1

DESCRIPTION: Interactive, scatter-plot visualization tool for analyzing differential gene expression.

AUTHOR: Scott P. Davis (Tempero Pharmaceuticals), in collaboration with Christophe Benoist (Harvard Medical School).



CONTACT: sdavis.bioinformatics@gmail.com

SUMMARY: Multiplot Studio is a complete redesign of the original Multiplot which is intended to correct a number of its shortcomings, add increased functionality and streamline the overall analysis process. It is based on entirely new source code and provides a complete revamped user interface.

PARAMETERS:	NAME	OPTIONAL	DESCRIPTION
	<code>dataFile</code>	No	GenePattern dataset (*.GCT), Multiplot Studio Workspace file (*.MSDATA) or *.CSV (see below for requirements).
	<code>classFile</code>	Yes	Class file (*.CLS). Though not required, the lack of a class file will result in fewer calculation options. Ignored if dataFile is a Workspace (.MSDATA).

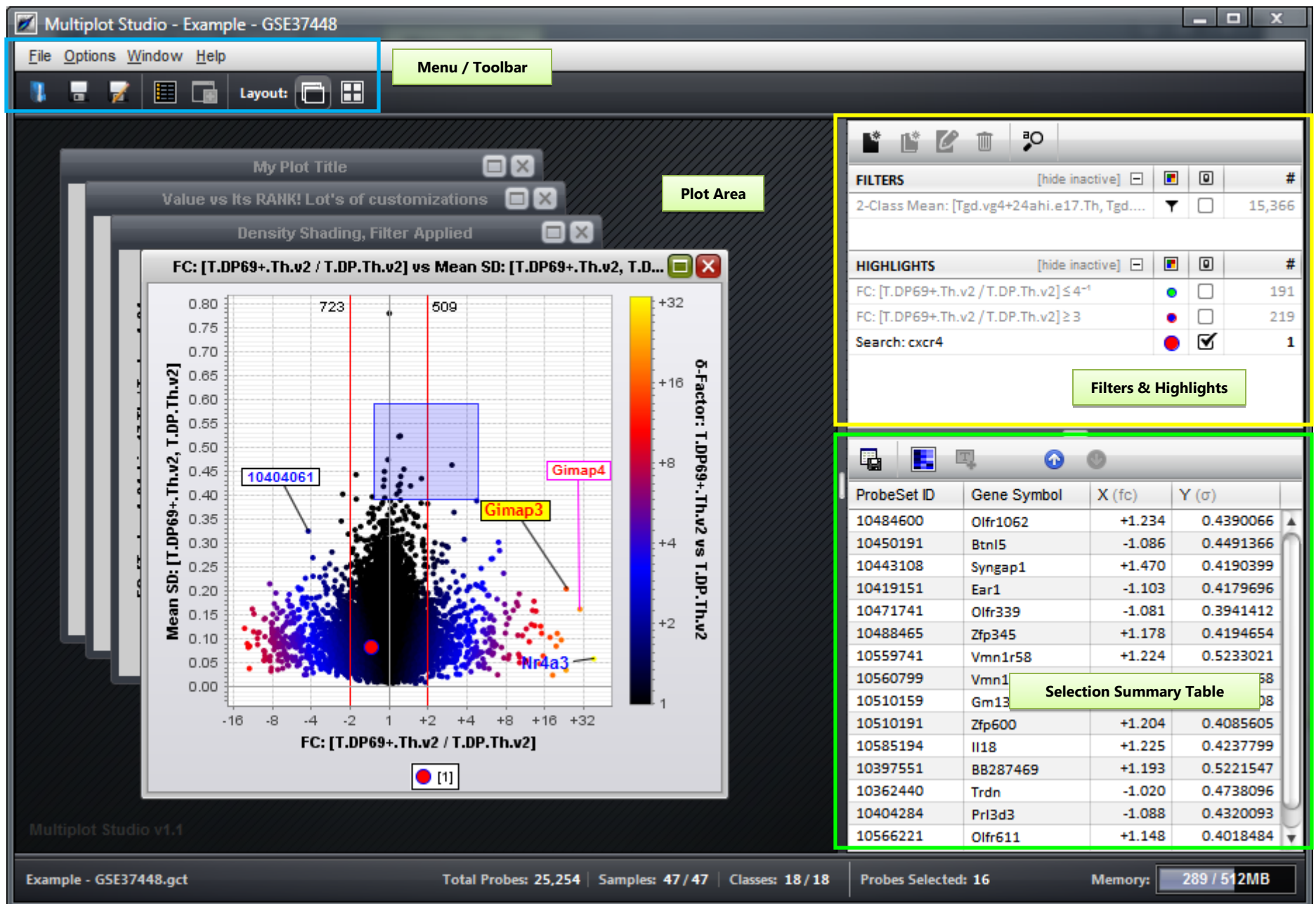
PLATFORM: Task Type: Visualizer
CPU Type: Any
OS: Any
Language: Java (1.7+)

TABLE OF CONTENTS

TABLE OF CONTENTS	2
PRIMARY INTERFACE OVERVIEW	3
IMPORTING DATA	4
INPUT/OUTPUT DATA FORMATS	4
SELECTING SAMPLES	5
COLUMN DATA TYPES	5
SAMPLE SELECTION DIALOG	6
EXPORTING DATA	7
SAVING A WORKSPACE	7
SAVING PLOT IMAGES	7
EXPORTING SUBSET DATA	7
Primary/Secondary ID Lists	7
Plotted Values	7
Original Source Data	7
CREATING PLOTS	8
WORKING WITH MULTIPLE PLOTS	8
PLOT LAYOUTS	8
PLOT TOOLBAR BUTTONS	8
DEFINING PLOT DATA	9
CALCULATIONS	9
CALCULATION SCOPES	10
ADDITIONAL OPTIONS	10
Plot Type	10
Use Rank	10
Advanced Shading	10
MOUSE MODE/DATA POINT SELECTION	11
PLOT APPEARANCE CUSTOMIZATION	12
LOG2 SCALE DISPLAY FORMAT	12
Negative/Down Fold Change Values	12

ANNOTATIONS	13
TEXT ANNOTATIONS	13
FOLD CHANGE MARKERS	14
FILTERS & HIGHLIGHTS	15
FILTER/HIGHLIGHTS TABLE	15
CREATING	16
TYPES OF CRITERIA	17
Calculation-based	17
Identifier-based	17
ADDITIONAL FILTER/HIGHLIGHT FUNCTIONS	18
Batch Create Multiple ID-Based Highlights	18
Populate Summary Table from Highlight(s)	18
Create Highlight from Summary Table	18
DATA TABLE/HEATMAP	19
Data Columns	19
Text vs. Heatmap	19
Color Scale	19
Scale Normalization	19
Column Visibility	19
Column Sorting	19

PRIMARY INTERFACE OVERVIEW




IMPORTING DATA


INPUT/OUTPUT DATA FORMATS

EXTENSION	FILE TYPE	USE	FILE FORMAT DESCRIPTION				
*.GCT	GenePattern Dataset	I/O	See documentation.				
*.CLS	GenePattern Class File	I	This file defines which samples/replicates within a dataset belong to the same class/population. See documentation.				
*.CSV	Comma-Separated Value	I/O	<p>Can contain both numerical and text data. Contains 1 header ROW and 2 header COLUMNS. The first row and column must contain unique identifiers (this is essentially a comma-separated GCT without its first two rows and with the ability to contain text data as well). The order of columns appearing after the first two header columns does not matter, however upon reading the file Multiplot Studio will rearrange them such that all numerical columns are packed together and appear before any text columns.</p> <p><i>Example...</i></p> <table><tr><th>A CSV HAVING THIS STRUCTURE...</th><th>IS IMPORTED AS...</th></tr><tr><td>ID1, ID2, DATA1, TEXT1, DATA2, TEXT2, DATA3, TEXT3, TEXT4</td><td>ID1, ID2, DATA1 ... DATA3, TEXT1 ... TEXT4</td></tr></table> <p><i>ID1 = Primary, unique identifier</i> <i>ID2 = Secondary, optional non-unique identifier</i></p> <p>❶ A CLS file can be loaded with a CSV file, however it is important to note that the class assignments defined within a CLS only pertain to numerical data columns, and text columns are not considered. Therefore, it is highly recommended that, for the sake of consistency between a CSV and its corresponding CLS file, users prearrange their CSV files such that all data columns are already packed together and appear immediately after the first two header columns.</p>	A CSV HAVING THIS STRUCTURE...	IS IMPORTED AS...	ID1, ID2, DATA1, TEXT1, DATA2, TEXT2, DATA3, TEXT3, TEXT4	ID1, ID2, DATA1 ... DATA3, TEXT1 ... TEXT4
A CSV HAVING THIS STRUCTURE...	IS IMPORTED AS...						
ID1, ID2, DATA1, TEXT1, DATA2, TEXT2, DATA3, TEXT3, TEXT4	ID1, ID2, DATA1 ... DATA3, TEXT1 ... TEXT4						
*.MSDATA	Multiplot Studio Workspace	I/O	This is a self-contained binary file containing all information needed to restore a previously saved workspace (source data file, class file, highlights/filters, plots, settings, etc.). This file is generated by Multiplot Studio and cannot be externally modified by the user.				
*.TXT	Plain Text	I/O	Used for probe ID/gene symbol lists. As an input, can be used to create highlights & filters. As an output, a list of the corresponding identifiers for a user-defined subset.				
*.PNG	Portable Network Graphics	O	Image file depicting either a plot or heatmap. For plots, users can specify the exact output resolution.				
*.SVG	Scalable Vector Graphics	O	Alternative vector-based graphics format used to create hi-resolution, scalable images for use in publications.				

Automatic Loading of Class Files (CLS): When a GCT file is loaded, Multiplot will look to see if a CLS file with the same base name exists. If it finds such a file in the same folder as the GCT, it will automatically load it. If a CLS file cannot be found, the user will be notified and asked if they wish to manually search for one or proceed without class definitions. Failure to define classes (populations) will result in fewer calculations being available to the user.

SELECTING SAMPLES

 It is often the case that a user may wish to work with only a subset of samples/populations within a dataset. In the case of populations, it may be that only some of those populations are of interest to them. In the case of samples, it may be that one or more samples within a population have been determined to be below the user's QC standards and therefore wish to omit them when calculating the corresponding class means. In either case, a user may select or unselect samples as they see fit. Samples or populations that are not currently selected do not appear as available operands in the dropdown lists for calculations, nor do they appear as available columns in the heatmap view.

 **Dependencies:** Multiplot Studio will keep track of any samples/populations that are currently being used in plots, highlights or filters. While in use, those items cannot be unselected, as indicated by the presence of a lock icon in place of the checkbox. In the case of a population for which there is currently some dependency, Multiplot Studio will prevent the user from unselecting that population as a whole, however it may still allow them to unselect specific samples within that population. In most cases, the rule is that, for a population that is currently in use, at least one of its samples must be selected. However, in some cases, if a population is being used in a calculations which has restrictions on the minimum group size (e.g. ANOVA, T-test P-value, etc.), then it will only allow the user to unselect samples if the resulting group size meets said requirement.

COLUMN DATA TYPES

In order for Multiplot Studio to be able to accurately perform calculations on a range of data types and to be able to display those results on the appropriate axis scale, it must know something about how the data is distributed. Accordingly, the sample/population selection window also contains a dropdown box that allows the user to provide this information. Usually, it is not necessary to change the default values, as Multiplot Studio does a good job of automatically determining the data type upon reading a data file. However, there may be certain situations in which the predicted default is not accurate. In such cases, the user can override this default value and explicitly set the appropriate data type(s). A sample/population can only be assigned data types that are consistent with the observed data range of the column in question (e.g. one cannot assign **P-value** to a column that has values outside the range [0,1]). Moreover, if a class file has been loaded, the user cannot change the data type of individual samples, only of the population as a whole.

DATA TYPE	RANGE	DESCRIPTION
Signal	$[-\infty, \infty]$	Any value: a signal, score, measurement, etc. which can possibly be \leq to 0, and is generally not exponentially increasing. For data of this type, the standard arithmetic mean is most appropriate.
Signal (exponential)	$(0, \infty]$	Default. This is accurate for most microarray gene expression datasets in which expression values typically increase exponentially. For data of this type, the geometric mean is the appropriate mean to calculate.
Ratio	$(0, \infty]$	Pre-calculated fold change or percentage.
Log2(Ratio)	$[-\infty, \infty]$	Pre-logged fold change. Computationally, treated the same as "Signal".
P-value	$[0, 1]$	A pre-calculated probability. Computationally, treated the same as "Signal".

Rules for Automatic Data Type Assignment:

- GCT** With respect to the data types in a standard GCT, Multiplot assumes homogeneity. If the global range of the data is within $(0, \infty]$, Multiplot will set all data columns to "Signal (exponential)", since this is accurate for the vast majority of microarrays. If, however, a GCT contains any values \leq to 0, all columns are set to "Signal".
- CSV** Datasets of this type are *not* assumed to contain homogeneous data types, so Multiplot Studio will attempt to assign the most appropriate data on a per-population basis, based on the observed range of the data and the presence or absence of certain key words in the sample's name (such as "fold change", "pvalue", "log2", etc.). Because CSV files are expected to be less consistent in the types of data they contain, Multiplot Studio will always prompt the user to view & verify the default values that it has assigned by showing the sample selection dialog when the file is first read.

SAMPLE SELECTION DIALOG

Sample/Population Selection

Although the population as a whole cannot be unselected because it is in use, one or more individual samples within the population can be.

Data Type

Notice that, for this population, "Pvalue" is not available as an option. This is because the population contains data outside of the range [0,1]

The screenshot shows the 'Data Selection' dialog box with a table of samples and populations. The table has three columns: Name, Range, and Type. The 'Name' column shows hierarchical data with expand/collapse icons and checkboxes. The 'Range' column shows numerical ranges. The 'Type' column shows the data type for each item, with a dropdown menu open for one of the items showing options: Signal (exponential), Signal, Signal (exponential), Ratio, and Log2(Ratio).

Name	Range	Type
▼ B.PBMC.IL2.Gluc=1000nM.act_3d (3/3)	[0.9171 ... 7706.6906]	Signal (exponential)
<input checked="" type="checkbox"/> B.PBMC.IL2.Gluc=1000nM.act_3d#1	[1.7089 ... 2946.4093]	Signal
<input checked="" type="checkbox"/> B.PBMC.IL2.Gluc=1000nM.act_3d#2	[1.234 ... 6650.5452]	Signal (exponential)
<input checked="" type="checkbox"/> B.PBMC.IL2.Gluc=1000nM.act_3d#3	[0.9171 ... 7706.6906]	Ratio
▼ B.PBMC.IL2.PDE4i=100nM.act_3d (1/2)	[0.9211 ... 7570.487]	Log2(Ratio)
<input checked="" type="checkbox"/> B.PBMC.IL2.PDE4i=100nM.act_3d#1	[0.9211 ... 7506.2646]	Signal (exponential)
<input type="checkbox"/> B.PBMC.IL2.PDE4i=100nM.act_3d#2	[1.014 ... 7570.487]	Signal (exponential)
▶ <input checked="" type="checkbox"/> B.PBMC.IL2.PDE4i=500nM.act_3d (2/2)	[0.831 ... 7825.0261]	Signal (exponential)
▼ <input checked="" type="checkbox"/> B.PBMC.IL2.PGE1=1000nM.act_3d (4/4)	[1.0044 ... 7012.0356]	Signal (exponential)
<input checked="" type="checkbox"/> B.PBMC.IL2.PGE1=1000nM.act_3d#1	[2.3607 ... 4570.5093]	Signal (exponential)
<input checked="" type="checkbox"/> B.PBMC.IL2.PGE1=1000nM.act_3d#2	[1.3618 ... 4702.7575]	Signal (exponential)
<input checked="" type="checkbox"/> B.PBMC.IL2.PGE1=1000nM.act_3d#3	[1.2224 ... 7012.0356]	Signal (exponential)
<input checked="" type="checkbox"/> B.PBMC.IL2.PGE1=1000nM.act_3d#4	[1.0044 ... 6725.9831]	Signal (exponential)
▶ <input checked="" type="checkbox"/> B.PBMC.IL2.Sidero=1000nM.act_3d (4/4)	[0.8599 ... 7970.1968]	Signal (exponential)
▶ <input checked="" type="checkbox"/> B.PBMC.IL2.TGFB.act_3d (4/4)	[0.9229 ... 7524.1388]	Signal (exponential)
2.VD3=1000nM.act_3d (4/4)	[0.7069 ... 7691.6767]	Signal (exponential)
2.act_3d (8/8)	[1.0274 ... 7967.2785]	Signal (exponential)

The context menu is open over the list, showing options: Check selected, Uncheck selected, Check all, Uncheck all, Set selected data types to..., Set all data types to..., Expand all, and Collapse all.

Context Menu

Right-click to invoke a context menu from which values can be set for multiple items

Data Type – Sample vs. Population

When a class file has been loaded, only the data type of the population as a whole can be changed (indicated by the fact that the corresponding values of the individual samples are greyed out).

EXPORTING DATA


SAVING A WORKSPACE

One of the major features of Multiplot Studio is the ability to save one's entire workspace for later analysis. The result of this operation is a single ***.MSDATA** file, which contains all the information required to restore a workspace to its original state, including the source dataset (***.GCT** | ***.CSV**), class file (***.CLS**), and any ID lists (***.TXT**) that may have been used to create filters or highlights. In fact, because this file is in a compressed binary format, it is almost always smaller in size than the original source dataset.

SAVING PLOT IMAGES

Plot images may be **copied** to the clipboard or **saved to disc** via the plot's context menu (right-click to invoke). Multiplot Studio allows users can specify the exact resolution of the image, or they may choose to export it in its current size. Images can also be exported in a resolution-independent, scalable vector-based format for use in publications, by selecting ***.svg** as the output type. **WARNING:** SVG's can be VERY large and take a long time to export, so it is recommended that users stick with PNG unless absolutely necessary.

EXPORTING SUBSET DATA

Various data associated with the currently selected subset (i.e. those items appearing in the Selection Summary Table) can be exported by clicking the  button of the toolbar above the table. The available export options for the current subset are summarized below:


- | | |
|--|--|
| <u>Primary/Secondary ID Lists</u> | A *.TXT file containing either the primary or secondary IDs (one per line) of the current table contents. |
| <u>Plotted Values</u> | A four-column *.csv file containing all the values currently in the table (primary & secondary ID, X & Y values). To export the table values exactly as they appear (i.e. values that may have a particular formatting applied), hold SHIFT while selecting the corresponding menu item. |
| <u>Original Source Data</u> | A *.GCT which is a subset of the original dataset containing all the original columns but only for the currently selected rows. |



Copy to Clipboard – To copy the contents of the currently selected table rows to the system clipboard, press **CTRL-C**. To copy the values exactly as they appear (i.e. values that may be formatted a certain way, such as fold change values), press .

CREATING PLOTS





WORKING WITH MULTIPLE PLOTS

As many as four plots may be created by  clicking the button on the main toolbar. At any given moment, only one plot is considered to be the “current” or **active plot**. This will be indicated by the light grey color of the enclosing plot frame, while all other plots’ frames will be dark grey. The contents of the selection summary table are based on the selection of the currently active plot, and the checkboxes associated with any filters or highlights will always indicate whether or not they are enabled for the currently active plot. Accordingly, switching the focus to another plot by clicking on it will change the contents of those tables to reflect the state of the current plot.


PLOT LAYOUTS

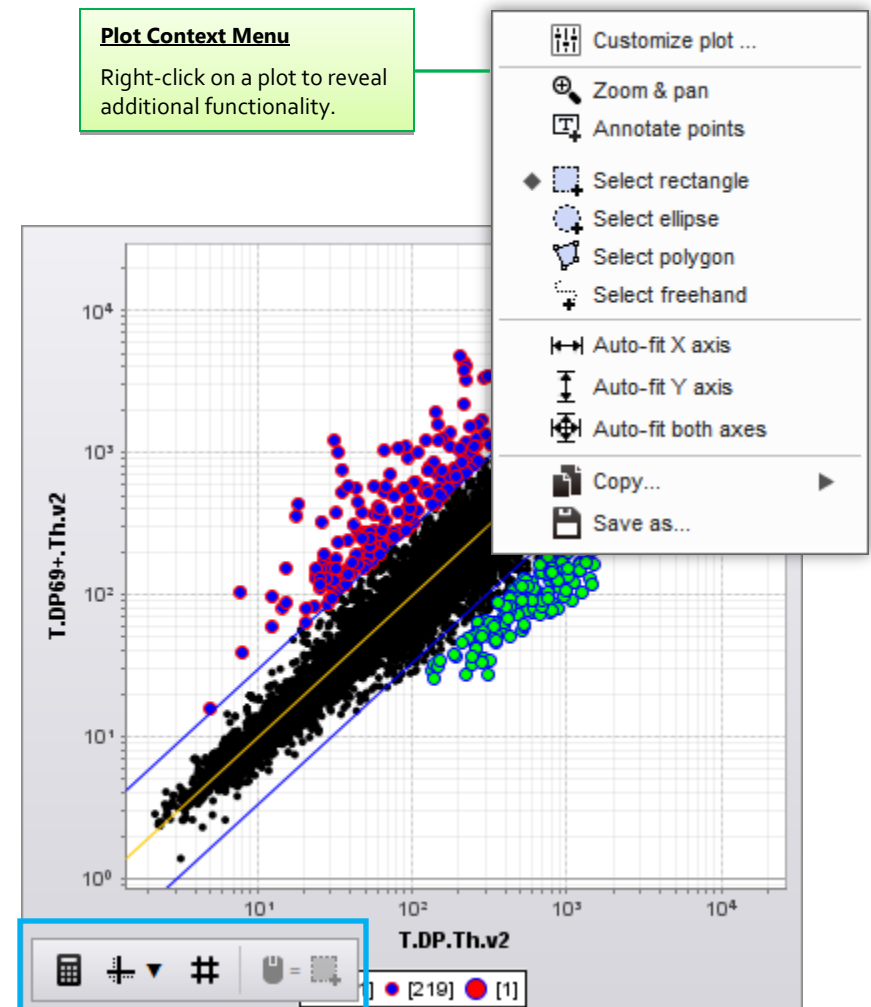
When working with more than one plot, there user can switch between two different layout modes – free layout and grid mode. In free layout (default), the user may reshape/resize individual plot windows however they choose, and place them anywhere within the desktop. In grid mode, the plots are laid out in a 2×2 snap-to grid. In this latter mode, plots may be rearranged by dragging them to another grid location, which switches the two plots in question.

PLOT TOOLBAR BUTTONS

OVERLAY BUTTONS / LABELS	
	Opens dialog to define plot data.
	Modify axes properties (range/domain, log vs. linear, etc).
	Create/edit fold change marker lines.
	This is not a button, but merely a status overlay which indicates the current mouse mode.


Additional functionality such as customizing the plot’s appearance, changing the current mouse mode and copying/saving images can be accessed by right-clicking the plot area to invoke the its context menu.

 The plot toolbar only appears when the mouse is over a plot – they are not visible on any exported plot images.



Plot toolbar buttons

DEFINING PLOT DATA

In order to plot data, the user must define the values for the X and Y axes. This is accomplished by clicking the  icon in the top/left corner of the plot and specifying the following:

1. **Calculation:** List of available calculations.
2. **Scope:** Over what range of operands is the calculation performed?
3. **Operand(s):** The list of available operands.
4. **Parameters:** If applicable to the calculation, additional parameters values.

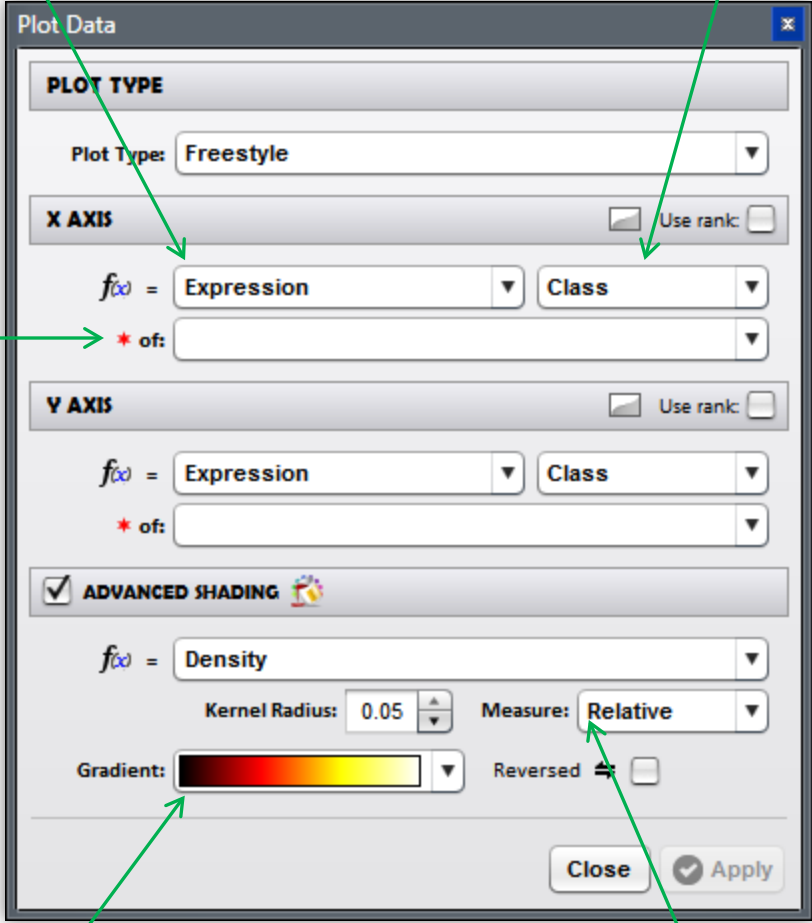
CALCULATIONS

CALCULATION	REPLICATE	CLASS	2-CLASS AVG	MULTI-CLASS	MULTI-CLASS MAX	MULTI-CLASS MIN
Expression	✓	✓	✓		✓	✓
Fold Change	✓	✓			✓	
Difference	✓	✓			✓	
Standard Deviation		✓	✓		✓	
Variance		✓	✓		✓	
Coefficient of Variation		✓	✓		✓	
T-Test P-value				✓		
One Way ANOVA				✓		
δ-Factor		✓				
Cardinality ¹		✓		✓	✓	✓
Density ²	NA	NA	NA	NA	NA	NA
Identifier Match ³	NA	NA	NA	NA	NA	NA

¹ Only available for datasets containing missing values (NaN) or having a minimum value of 0.

² Density is a special-case calculation which is only used in advanced shading. It has no operands.

³ Identifier Match is a special-case calculation which is only used in creating highlights & filters.



The image shows the 'Plot Data' dialog box with several green callout boxes pointing to specific features:

- Calculation:** Points to the 'Expression' dropdown in the X AXIS section.
- Calculation Scope:** Points to the 'Class' dropdown in the X AXIS section.
- Operand(s):** Points to the '* of:' dropdown in the X AXIS section.
- Advanced Shading:** Points to the 'Density' dropdown in the ADVANCED SHADING section.
- Parameters (visible when applicable):** Points to the 'Measure: Relative' dropdown in the ADVANCED SHADING section.

The dialog box includes sections for PLOT TYPE (Freestyle), X AXIS, Y AXIS, and ADVANCED SHADING. It also features a Gradient color bar and buttons for Close and Apply.

Plot Data Dialog

CALCULATION SCOPES

The scope of a calculation determines two things: **1)** the list of available operands, and **2)** how the calculation is applied to them. This option is only visible if a given calculation actually has more than one scope from which to choose. The meanings of the various scope values are summarized below.

① *Scopes with the word "Class" in them are only available if the user has provided a Class File (*.cls).*

SCOPE	MEANS THE RESULT IS...
Replicate	Based on applying the calculation to samples
Class	Based on applying the calculation to populations
2-Class Avg	The average of the calculation individually applied to two populations
Multi-Class	The calculation applied to multiple populations together
Multi-Class Min	The minimum of the calculation individually applied over multiple populations
Multi-Class Max	The maximum of the calculation individually applied over multiple populations
File*	Based on matching identifiers from a TXT file.
User Specified*	Based on matching identifiers given in a user-specified list.

**Only applicable to the Identifier Match calculation for highlights and filters.*

ADDITIONAL OPTIONS

Plot Type

This option allows the user to quickly generate a number of predefined plots by setting the calculation and scope for both axes.

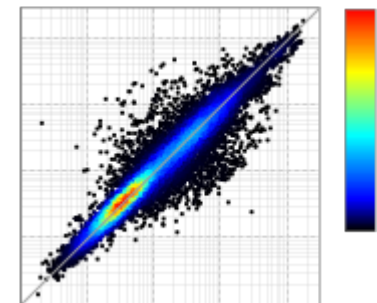
Use Rank

This option exists for each axis and allows the user to plot the rank of the calculated value in place of the value itself.

Advanced Shading

The user may also choose to paint data points according to the value of *any additional calculation*. In such a case, a third data-definition panel is shown where the user can define this criterion and indicate the color scale he/she wishes to use. This can act as an additional dimension with which to illustrate other aspects of the data than what is displayed by simply using the X and Y axes alone. However, caution should be taken when using advanced shading, as it may clash with other highlights that the user may wish to apply later on.

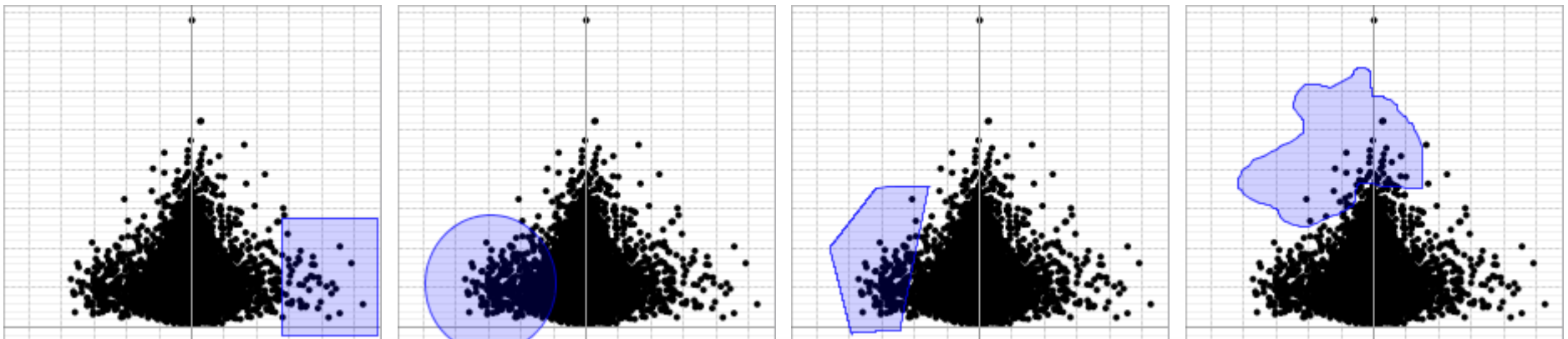
Density: A special calculation, Density, is available when using advanced shading. This calculation takes no user-provided operands (instead, the operands for this calculation are the X and Y values of the plot's axes).



MOUSE MODE/DATA POINT SELECTION

Clicking/dragging the mouse over the plot area can have a number of different effects depending on the currently selected mouse mode. Generally, the mouse mode will be in one of 3 different states:

MOUSE MODE	DESCRIPTION
Zoom & Pan	Allows the user to zoom in on a selected region of the plot. Holding CTRL while clicking allows the user to pan or “shift” the entire plot as a whole.
Annotate Points	Allows the user to add text annotations by clicking individual data points. When the user hovers over a point, a blue T icon will appear on the mouse cursor indicating that an annotation may be added there.
Selection	<p>The four “selection” modes, allow a user to define regions of the plot in which data points will be selected, causing them to appear in the selection summary table on the right. There are 4 different methods of defining a selection region...</p> <p>Rectangle: Select rectangular regions.</p> <p>Ellipse: Select elliptical regions. Hold CTRL to force circular selection.</p> <p>Polygon: Select a polygonal region by clicking multiple points to define the vertices. Right-click to finalize the defined shape.</p> <p>Freehand: Draw a freehand region.</p> <p>❗ For all selection modes except “polygon,” holding SHIFT while dragging will select any text annotations within the defined area instead of data points.</p>

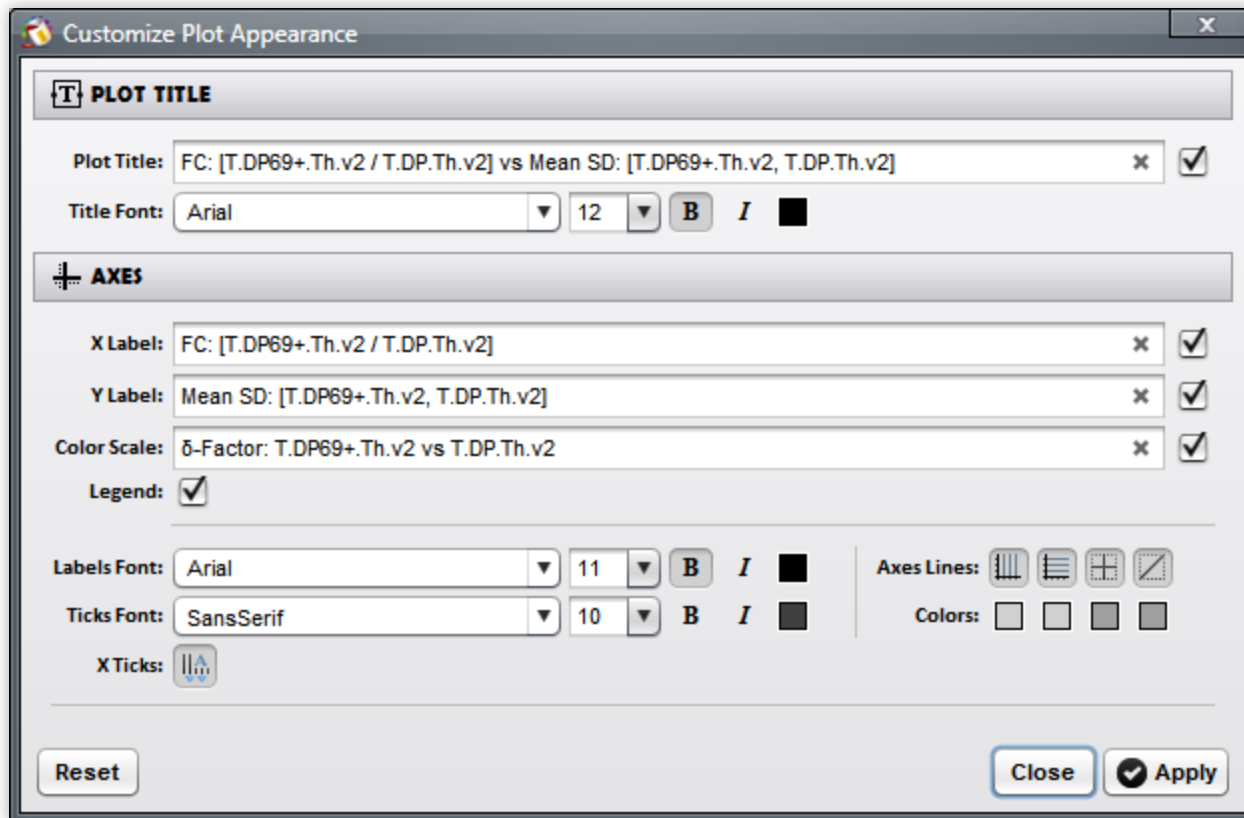


Example selection methods

PLOT APPEARANCE CUSTOMIZATION

Various aspects of a plot's appearance can be customized by right-clicking the plot and selecting the "Customize plot..." option. The list of elements that can be changed include:

Plot Element	Customizable Attributes
Title	Text Visibility Font Color
Axes Labels	Text Visibility Font Color
Legend	Visibility
Axes Ticks	Font Color Orientation
Grid Lines	Visibility



LOG₂ SCALE DISPLAY FORMAT

The way in which ratio/fold change values are displayed can be controlled via the menu: **Options | Log2 Scale Display Format**. Examples of how the various options affect the display of values are given in the table to the right.

Standard:	0.1	0.2	0.25	0.333	0.5	1	2	3	4	5	10
±Symmetric:	-10	-5	-4	-3	-2	1	+2	+3	+4	+5	+10
Signed exponent:	10 ⁻¹	5 ⁻¹	4 ⁻¹	3 ⁻¹	2 ⁻¹	1	2	3	4	5	10
Log2(Ratio):	-3.322	-2.32	-2	-1.58	-1	0	1	1.58	2	2.32	3.322
Percentage:	10%	20%	25%	33%	50%	100%	200%	300%	400%	500%	1000%


Negative/Down Fold Change Values: Whenever the display format is set to **Standard**, **±Symmetric** or **Signed exponent**, Multiplot Studio allows negative (or “down”) fold change values to be entered by the user using the intuitive (and keyboard friendly) **±Symmetric** notation. That is, a user may specify a fold change of **0.25** as **-4**. For most people, this represents a far more intuitive interpretation of down/negative fold changes values.

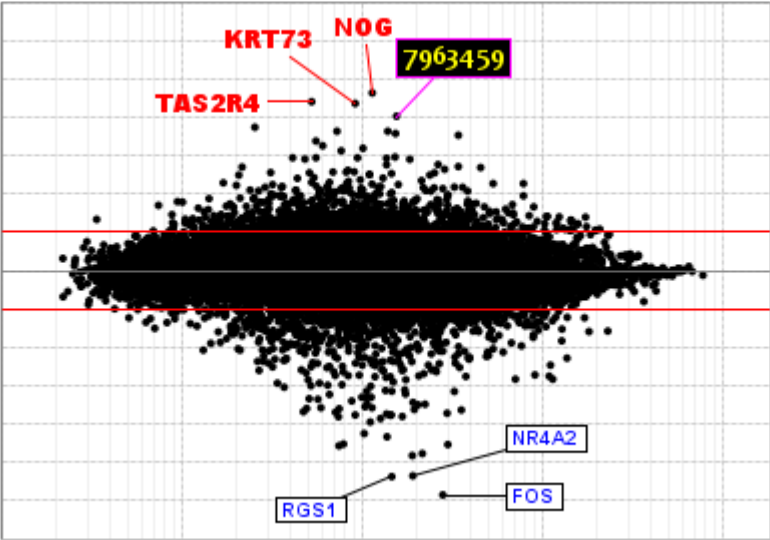
ANNOTATIONS

TEXT ANNOTATIONS

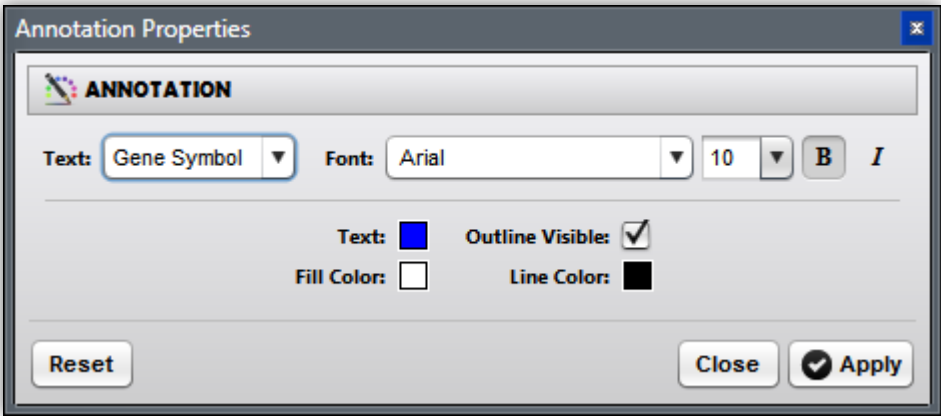
Text annotations can easily be added to a plot in order to label certain data points. Annotations will remain on the plot until the user explicitly removes them or until the associated data points are removed via filtering. This means that even if the user changes the plot's data (i.e. plots a new calculation or selects new operands), any existing annotations will remain on the plot and their locations within the plot will simply be updated to reflect its new (X,Y) coordinates. The only exception to this is if the new coordinates contains an missing/NaN value, in which case the annotation will not be displayed.

① **Batch Operations:** Several annotations can simultaneously be moved, deleted or customized in groups. See below for details.

FUNCTION	HOW TO...
Creating	There are two ways in which text annotations can be added to a plot: 1) Right-clicking a plot and selecting the "Annotate Points" mouse mode . 2) Highlighting rows in the selection summary table, then clicking the button  on the toolbar above the table.
Multiple Selection	Multiple annotations can be selected for batch operations by holding Shift while either clicking on it or selecting a region.
Deleting	Delete one or more annotations by right-clicking on it a selecting the "Delete" option.
Customizing	Customize one or more annotations by right-clicking on it a selecting the "Customize" option.
Moving	To move an annotation, simply click and drag it to any part of a plot. To move multiple annotations as a group, hold Shift while dragging any one of the annotations within a selected group.



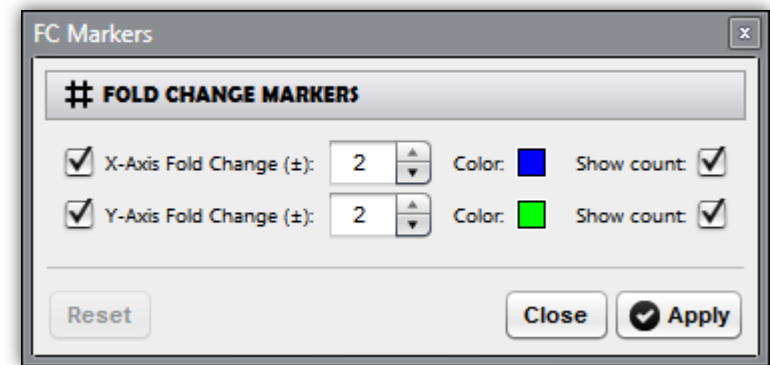
Plot with various styles of text annotations.



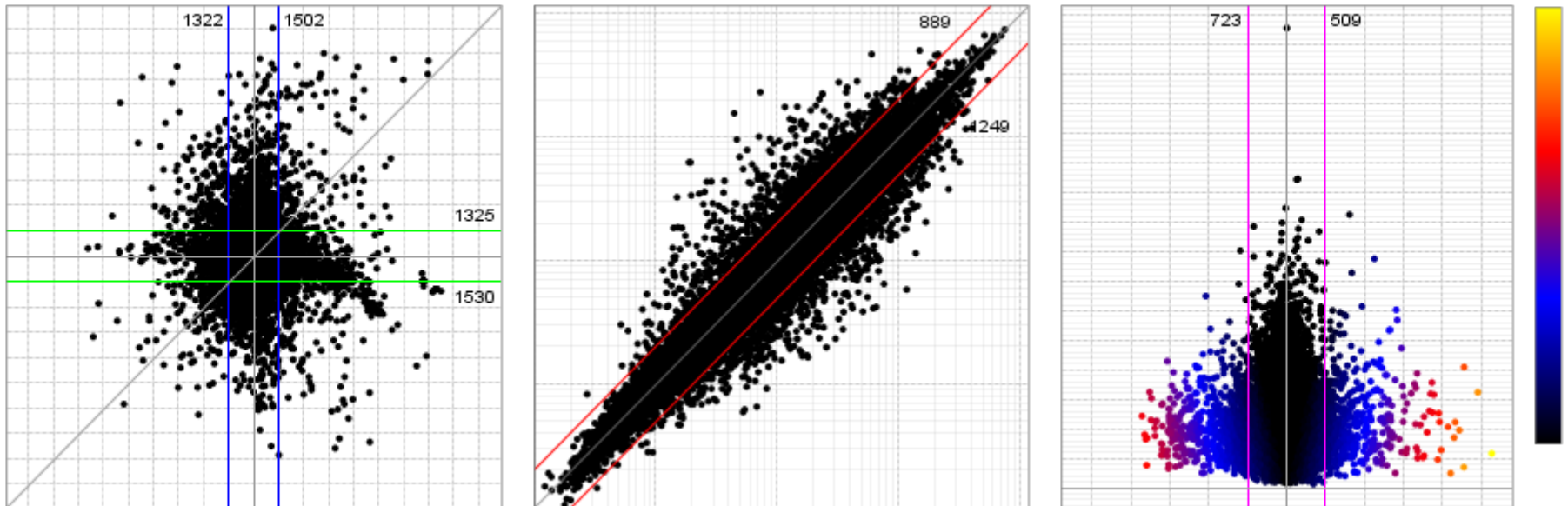
Text annotation customization dialog.

FOLD CHANGE MARKERS

Fold change demarcation guide lines can be added to any plot containing fold change values or, as a special case, to Expression vs Expression plots in which the axes are on the log scale and the source data are nonlinear values. Whenever fold change markers are allowed to be displayed on a plot, the **#** overlay button will appear in the lower/right corner of the plot. The exact options available on the FC Markers dialog depend upon the plot's current data.



FC marker dialog



Fold change markers plotted on different types of plots

FILTERS & HIGHLIGHTS

Filters and highlights are essentially the same entity. In fact, each can easily be converted to the other in the edit dialog or via the context menu. The difference lies in their behavior: filters remove all data points that satisfy some criteria, while highlights paint those data points in a specified color. The table of available filters and highlights is located in the upper/right hand corner of the screen.

FILTER/HIGHLIGHTS TABLE

Quick Search
Create on-the-fly ID-based highlights by entering one or more id's (primary or secondary) separated by commas.

Show/hide Inactive
Toggle the visibility of filters/highlights that are not being applied to the current plot

Create | Clone | Edit | Delete
Basic filter/highlight management operations.

Context Menu
Perform operations on one or more filters/highlights.

Filter Summary
Filters can be applied as the union or intersection of all active filters. Clicking on the blue **ALL** / **ANY** text will toggle between the two modes.

On/Off
Filters and highlights can be turned on/off globally (by clicking the checkbox) or only on selected plots. To selectively turn on/off for the current plot only, right-click on a selected row to invoke the context menu (shown to the left).

Marker Appearance
Change the color, size and shape of data points for a highlight. Click to invoke popup dialog.

Display Order
Drag & drop to reorder one or more items. This is relevant for situations in which a data point may exist in more than one active highlight, in which case the order determines which highlight's marker is placed on top.

FILTERS		[hide inactive]			#
2-Class Mean: [Tgd.vg4+24ahi.e17.Th, Tgd.vg4+2...	▼	☑			15,366
MF.11c-11b+.Lu ≥ 200.00	▼	☑			7,706
Matching: ALL Σ =					7,348
HIGHLIGHTS		[hide inactive]			#
FC: [T.DP69+.Th.v2 / T.DP.Th.v2] ≥ 3		☑	●		91
FC: [T.DP69+.Th.v2 / T.DP.Th.v2] ≤ 4 ⁻¹		☑	●		23
Search: cxc4		☑	●		1

☒ Enable for all plots

☐ Disable for all plots

☒ Enable for current plot only

☐ Disable for current plot only

▼ Convert to FILTER

CREATING

All filters/highlights are defined by these four basic properties:

PROPERTY	DESCRIPTION
Type	A value indicating whether it is a Filter or a Highlight
Name	<i>Optional.</i> A name to associate with this filter/highlight. Will be automatically generated if not explicitly provided by the user.
Point Marker	If defined as a highlight, this is the appearance of points on a plot which satisfy this highlight's criteria.
Criteria	One or more conditions that must be met in order to be selected by this filter/highlight. In the case of multiple criteria, the user can specify whether they must ALL be met, or just ANY of them.

Criteria List
List of criteria which must be satisfied in order for a point to be highlighted or passed through filter.

Dynamic Distribution Plot
For calculation-based criteria, this plot shows the distribution of calculated values in order to assist the user in determining an appropriate threshold

Criterion Thresholds
Thresholds for a given calculation can be defined by sliding the control buttons OR by typing them into the text field above.

Filter/highlight creation dialog.

TYPES OF CRITERIA

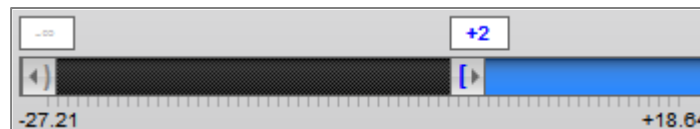
Calculation-based

Criteria of this type operate by selecting data points based on whether or not the associated values for a user-defined calculation meet some specified threshold. Calculations are defined in the exact same way they are defined when creating a plot. As soon as all of the required input fields for a given calculation are complete, Multiplot Studio will instantly perform that calculation on-the-fly and generate a rank-ordered distribution plot which is displayed on the right. The purpose of this plot is to allow the user to visualize the distribution of values in order to determine a reasonable threshold (typically near the tails of the distribution).

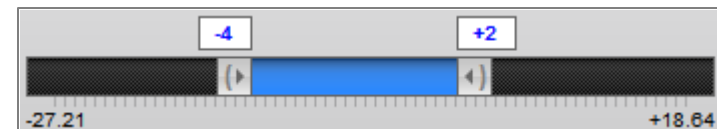
Specifying Threshold Values

Threshold values may be specified by **1**) dragging the upper/lower slider buttons below the plot, or **2**) explicitly entering the values in the text fields above the slider buttons. The color of the slider's gutter indicates whether or not a range is included (**BLUE**) or excluded (**GREY**). In addition to an arrow indicating the included range, the slider buttons also contain either a **()** or a **[]** indicating whether or not the defined [interval](#) is open (i.e. $<$, $>$) or closed (i.e. \leq , \geq). Single-clicking on a button without dragging it will toggle between an open or closed interval for that threshold value.

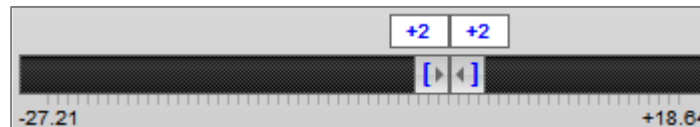
Examples:



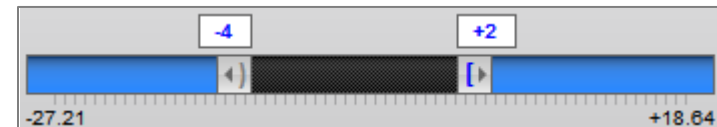
Include any x where: $x \geq +2$



Include any x where: $x > -4$ and $x < +2$



Include any x where: $x = +2$



Include any x where: $x < -4$ or $x \leq +2$

Inverting Interval

To invert the interval range of included values, simply double-click on any part of the excluded range (the dark grey part of the gutter). When hovering the mouse over any part of the excluded range, a white border will appear around that region and a tooltip will pop up indicating that this action may be performed.

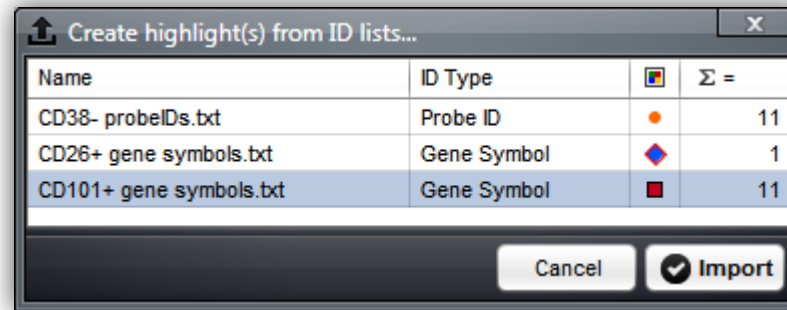
Identifier-based

A criterion based on the "Identifier Match" calculation (only an option for filters and highlights), in which data points are selected based on whether or not their primary (usually "probe ID") or secondary (usually "gene symbol") identifier matches any of those specified by the user. The source of the specified identifier(s) can either be a TXT file (1 id per line) or a user-entered, comma-separated list, and is specified by the user via the "Scope" field. The type of identifier is specified by the user as the first "operand" to this calculation.

ADDITIONAL FILTER/HIGHLIGHT FUNCTIONS


Batch Create Multiple ID-Based Highlights

Multiple ID-based highlights can be created simultaneously by dragging one or more *.TXT files containing a list of primary or secondary IDs. Each TXT file must contain only one type of identifier (i.e. all primary or all secondary), and must be in the standard ID-list format of one ID per line.




Batch ID list highlight creation dialog

Populate Summary Table from Highlight(s)

The Selection Summary Table can be populated with the members of one or more highlights (even if they are not currently active) by selecting the desired highlight(s) in the highlight table, then clicking the  button on the toolbar above the Summary Table.




Create Highlight from Summary Table

The subset of probes currently listed in the Selection Summary Table can be converted into an ID-based highlight by clicking the  button on the Summary Table toolbar.

DATA TABLE/HEATMAP

While the Selection Summary Table shows the plotted X/Y values for a given subset of data points, the Selection Data Table/Heatmap shows the corresponding original source data (including any text columns) for the same subset of points. The table can be sorted on any of the columns currently being displayed. The options controlling the display of data in this table are described below:

- Data Columns** If a class file has been provided, the user will have the option of displaying either individual replicate values or the class means.
- Text vs. Heatmap** User may toggle between displaying numerical values as text or as a heatmap.
- Color Scale** In heatmap display, the color scale can be any of a number of predefined color schemes.
- Scale Normalization** In heatmap display, the intensity of each cells values can be normalized in 3 different ways, depending on whether or not the columns being displayed contain homogeneous (uniform) or mixed data types.

NORMALIZATION	UNIFORM	MIXED	COLOR-CODED RELATIVE TO...
 Global	✓		min max defined over all rows and columns in the current table.
 Relative (Row)	✓		min max defined for each row.
 Relative (Column)	✓	✓	min max defined for each column.

- Column Visibility** The user may selectively show/hide individual columns by clicking the column-control button [☐] located in the upper/right-hand corner of the table. Note that any time a numerical column is hidden or shown, Multiplot Studio will re-check the uniformity of the data types for the columns currently being displayed in order to update the normalization options available to the user.
- Column Sorting** Columns can be sorted by **1)** column name, **2)** the value of the selected row, or **3)** the cardinality of each column (# or % of non-missing values). That latter is only available if the dataset contains NaNs (i.e. is ragged) or has a minimum value of 0 (is possibly sparse).

Data Columns

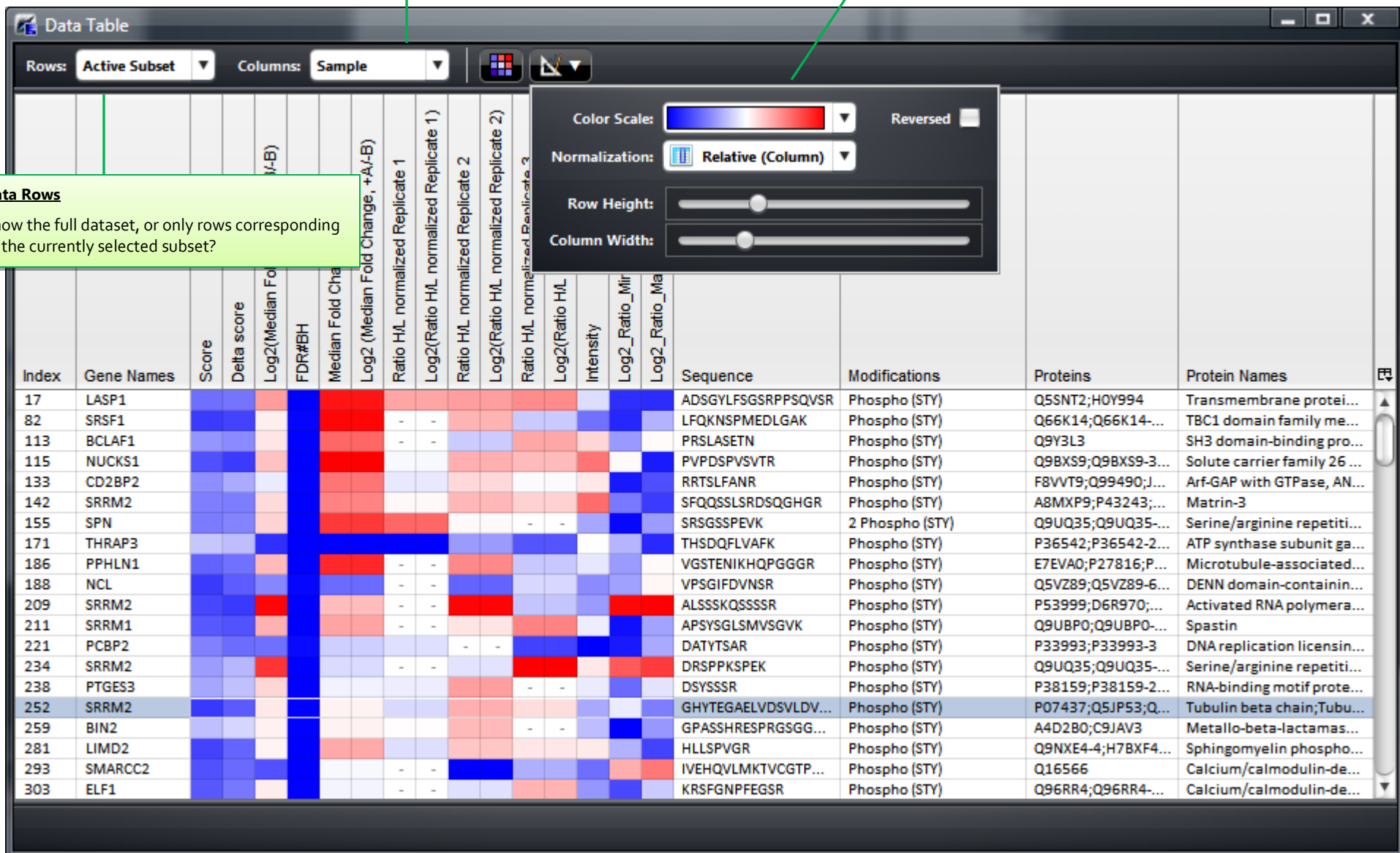
Determine whether the values shown are those of the individual replicates or of the current class means (only applicable when Class File is loaded).

Heatmap Settings

Buttons to control the use, color-scale, and normalization of values for the heatmap.

Data Rows

Show the full dataset, or only rows corresponding to the currently selected subset?



Selection Data Table/Heatmap