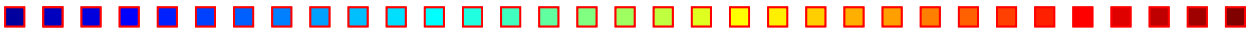
ScatterSlice
v 1.0
User Guide



Revised: February 2013

Introduction

This guide is intended as a thorough explanation of practical usage of ScatterSlice, a program for performing automated cell-to-cell variability analysis (CCVA). Follow the instructions on the next pages to install ScatterSlice, then see the Tutorial for a brief example.

An extensive example of the application of CCVA be found in Cotari & Voisinne et al. Science Signaling 2013.

ScatterSlice is a program for adding dimensionality to flow cytometric data. The program is built on an R framework, using TclTk as a means of producing a graphical user interface. Usage of ScatterSlice, however, requires only that the user have basic familiarity with flow cytometry data.

Installation: OS X

Step 1: Install necessary software

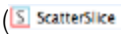
Visit http://cran.r-project.org/ to download R. Install with default options.

Visit http://cran.r-project.org/bin/macosx/tools/tcltk-8.5.5-x11.dmg to download tcltk. Install.

Visit http://xquartz.macosforge.org/landing/ to download X11. Install.

Step 2: Install ScatterSlice

Automatic Installation:

- Download ScS_Package.zip
- Unzip ScS_Package.zip archive. Move the ScS_Package folder to a documents folder.
- Open the ScS_Package folder.
- Move ScatterSlice application (

) and the "ScatterSlice" folder to your ***Applications*** folder.
- Double-click "install.packages.command". This will open Terminal and run a script to install the R packages required to run ScatterSlice.
- After the script completes, you can quit the Terminal application.
- Drag the ScatterSlice file icon from Applications to the Dock for easy launching of the application.

Occasionally, the permissions on the Applications folder prevent installation. To change permissions:

- Open *Macintosh HD* in the finder and right (control+) click the *Applications* folder and select "Get Info".
- Click the small lock icon in the bottom-right corner of the Info window. Enter your password if prompted.
- In the "Sharing & Permissions" area at the bottom, change the permissions to allow yourself to "Read & Write" to the Applications folder.

Step 3: Start ScatterSlice

Either double-click the icon in the Applications folder, or single-click the icon in the dock.

Installation: Windows

Step 1: Install necessary software

Visit http://cran.r-project.org/ to download R. Install with default options.

Visit http://www.activestate.com/activetcl/downloads to download tcltk. Install.

Visit http://sourceforge.net/projects/xming/files/latest/download to download X11 for windows.

Step 2: Install ScatterSlice

Download ScS_Package.zip

Unzip ScS_Package.zip archive.

Copy the "ScatterSlice" folder into your Program Files folder.

Open R and install required packages by pasting the following command into the R window:

```
install.packages(c("tkrplot", "foreach", "doMC",  "gplots") repos="http://cran.r-project.org")
```

Step 3: Start ScatterSlice

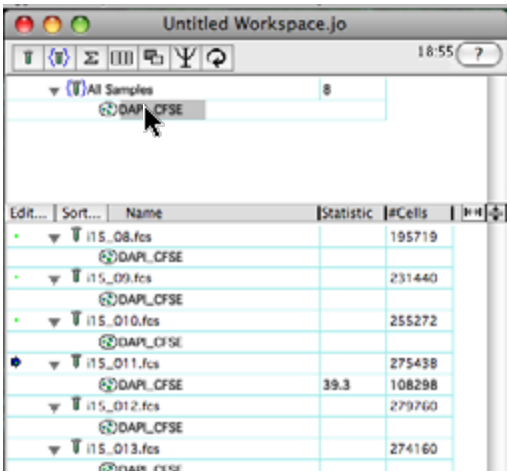
Paste the following commands into the R window, changing the drive letter if your Program Files directory is not on the C drive.

```
setwd("C:/Program Files"); source("Main.R")
```

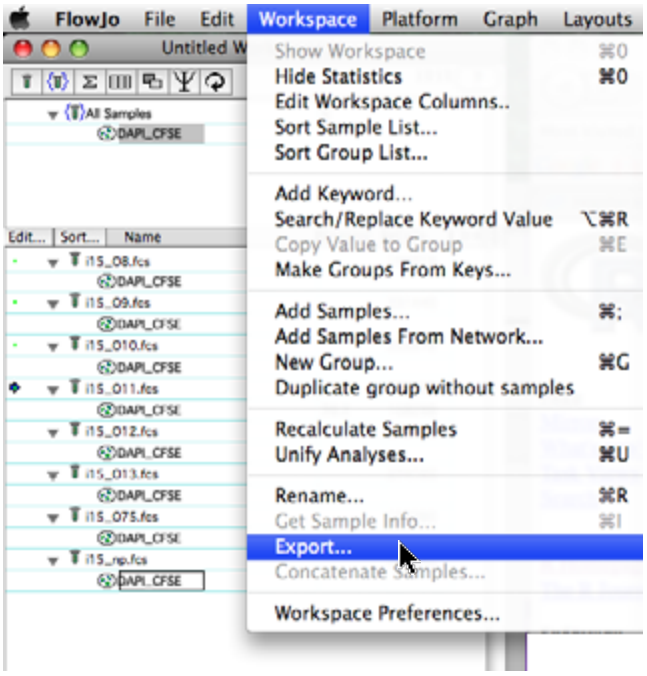
Pre-Run Data Preparation

For ScatterSlice analysis, it is necessary to first process raw flow cytometry data, performing any necessary gating and compensation before exporting your data. ScatterSlice takes data as tab-separated plain text files with a one-row header of channel names (FSC, APC, FL1 etc.). For our analysis, we typically use FlowJo (tm) to gate and export the data. The exporting process is described below.

To export from FlowJo, first select the gate you want to export.

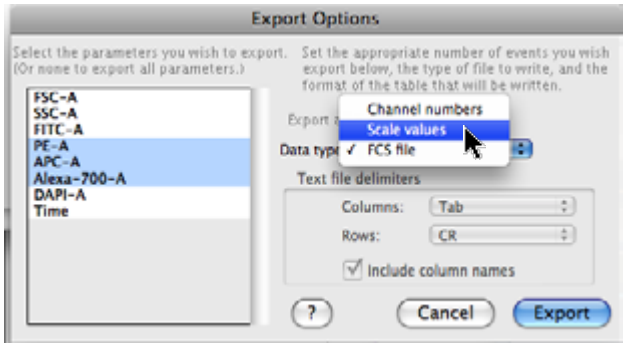


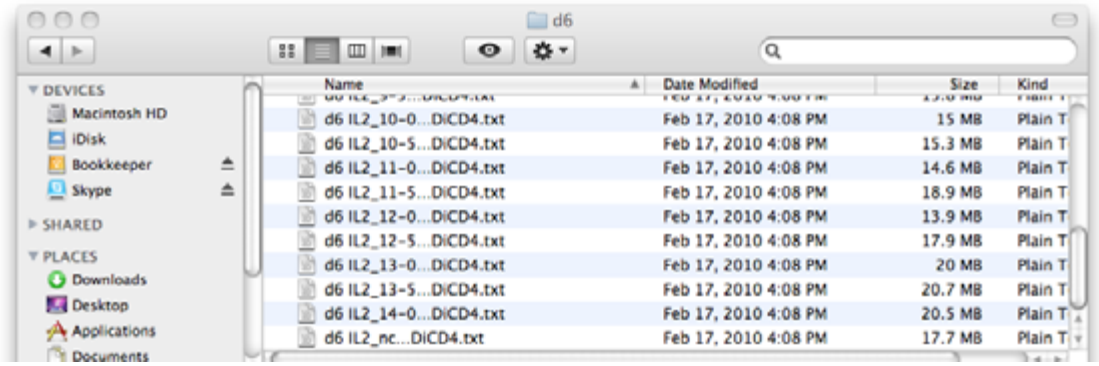
Then choose "*Export...*" under the *Workspace* menu



In the *Export Options* window, choose the channels you want to export and select the "*Scale values*" option in the Data type box. ScatterSlice currently requires three distinct channels. If you are only interested in two channels, select FSC or SSC to make up the difference. The default values in the "*Text file delimiters*" box are appropriate for our purposes (i.e. *Columns: Tab* and *Rows: CR*).

To reduce processing time, reduce the number of cells to export from 0 (all cells) to 50,000 cells.





The result of the export will be .txt files named with your tube names (up to 11 letters) followed by the subset name.

The newer version of FlowJo (X) has different exporting conventions. The files are .csv and will be named differently.

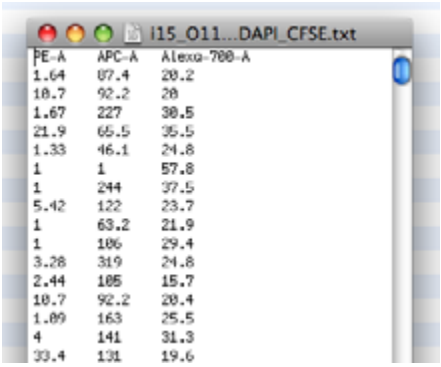
NOTES ON NAMING:

1 - FlowJo will truncate your tube names if they exceed 11 characters, so be sure that the first 11 characters will uniquely identify your tubes. If the names of the tubes are not unique within the first 11 letters, FlowJo will export, appending "-1" "-2" to all files after the first, in random order.

For example: If your tubes were named "Experiment1Tube1", "Experiment1Tube2" and you used a gate named "DAPI_CFSE", after exporting, the samples will be designated "Experiment1Tub…DAPI_CFS.txt" and "Experiment1Tub…DAPI_CFS-1". In addition, FlowJo exports files in random order, so the numbers won't help you at all.

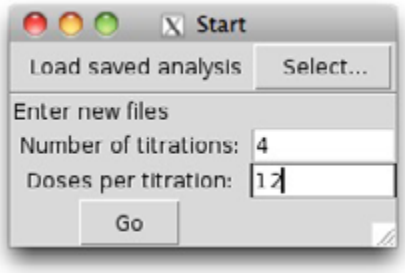
2 - For easy import into ScatterSlice, name files so that titrations are in alphabetical order, and will be grouped by titration. The alphabetizing conventions may take some getting used to, but it's worth it.

Opening one of your exported files should show columns of numbers with headers at the top of each one. Longer names may cause headers not to line up with their columns, but this will not affect ScatterSlice's ability to read the files.



Running ScatterSlice

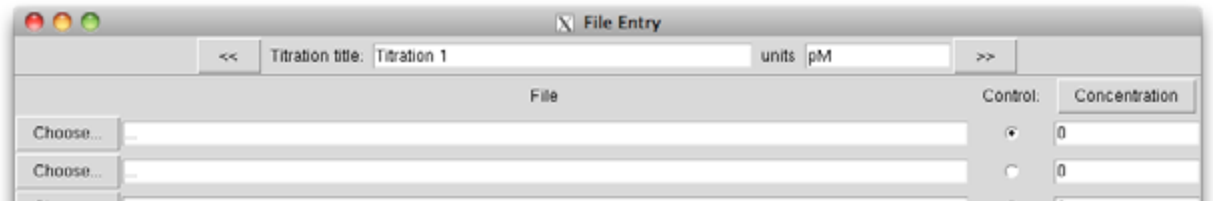
Start



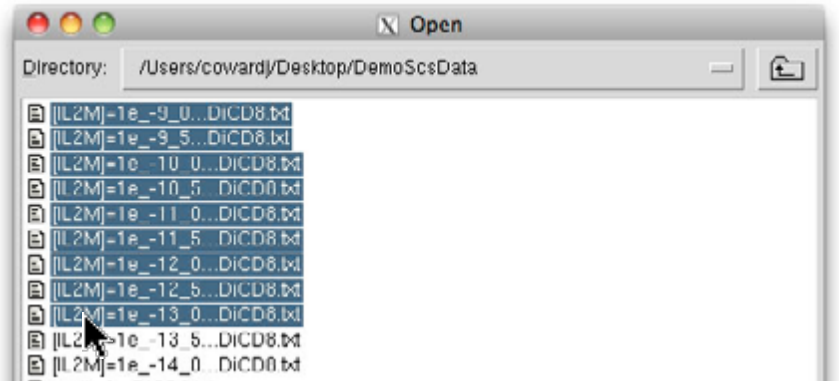
The first step in running ScatterSlice is to either load a saved analysis or input the number of files you will be entering. The program is set up to allow parallel evaluation of multiple titrations. For two to three doses per titration, ScatterSlice will estimate amplitudes from data, but not EC_{50} values. For a single dose, ScatterSlice will perform binning in two dimensions and display fluorescence intensities using a color scale.

File Entry

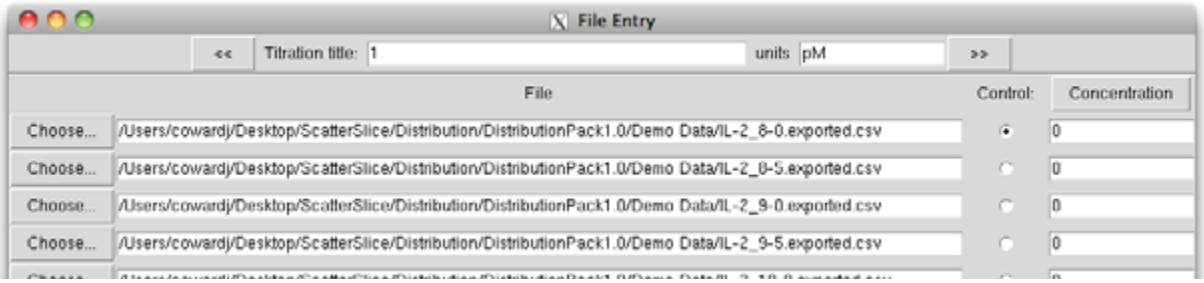
This is where you enter and name your titrations, and specify doses and units.



1. Start by using the ***Choose...*** button to open a dialog to choose your input (.txt or .csv) files. *Multiple files can be selected at once.* Click the first file, then hold shift and choose the last file. This requires having your files titled in numerical order.



After selecting files, they will populate the window as shown below.



2. Next, fill in the important information for each titration, using the "<<" and ">>" buttons to move between titration sets.

Titration Title: Specify a title for each titration. This will be used to title graphs and identify exports.

Units: Used for display and labeling.

Control: Select which file is your control file. This should be a condition with no stimulus. *Be sure to set the control for each titration* before entering a concentration series.

Concentration: Either fill in titrations manually, or click this button to input a serial dilution. Opens the window shown below.

Concentration

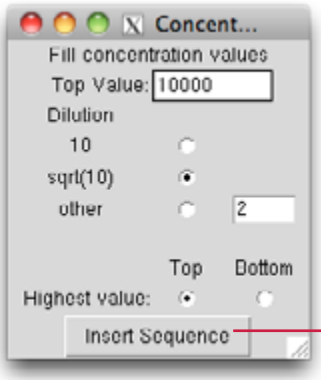
This dialog box allows you to specify the doses that will be used across all of your titrations. It is possible to fill all the titrations first, then subsequently change values for individual groups.

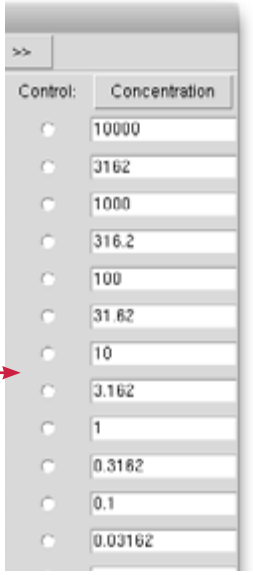
Top Value Highest dose in your titration

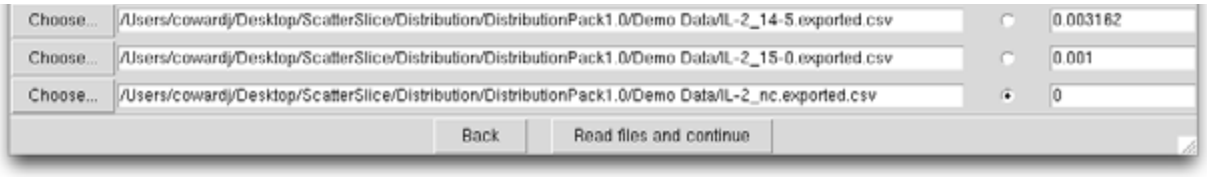
Dilution Factor by which each successive value is decreased.

Highest value Select whether the highest dose in your titration will be at the bottom or top of the list of files

Insert Sequence Clicking this button will fill the values specified, as shown at right.







After filling in all important information, use the ***Read Files and Continue*** button to read your data files into memory.

Note: During processing, there will not be an hourglass or pinwheel indicating that the computer is working. Instead, the last button you clicked (in this case Next) will remain depressed. Depending on the size of your files and the speed of your computer, this may take a few seconds. To decrease the reading time, export from FlowJo only the columns you will use in your analysis.

If you need to start over, use the ***Back*** button to go back to File Number Entry. This will keep the values you have entered.

Select Parameters

1. Labeling

The ***Channel*** column is the header of each column in the raw data, as exported by FlowJo.

Use the ***Label*** column to identify your parameters more informatively.

If all of your channels are the same across titrations it is sufficient to enter this information once; it will be applied to all titrations. Otherwise, use the **<< >>** buttons to move through your titrations.

2. Analysis Setup

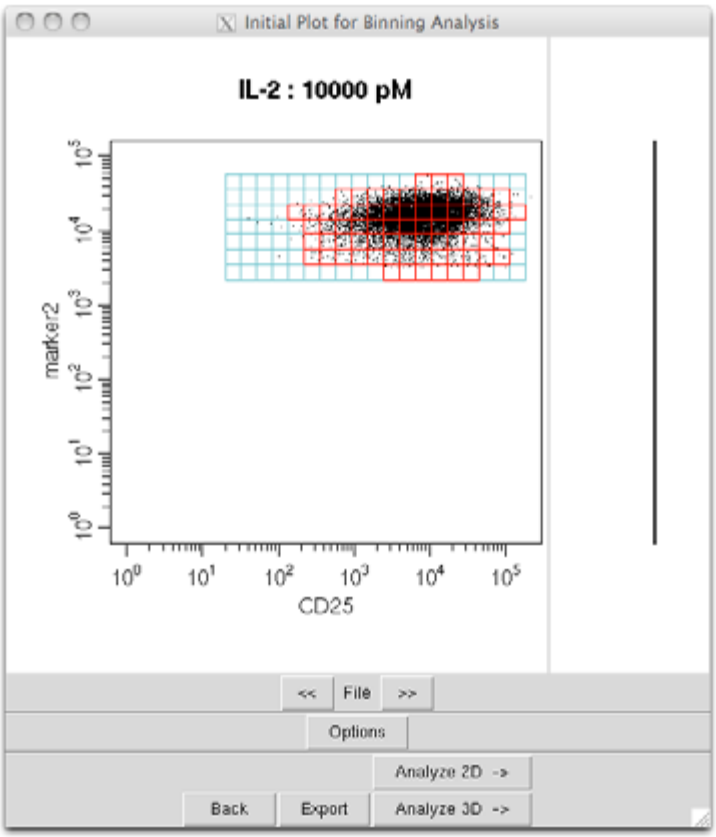
X, Y, Z: Choose which parameters will be plotted on x and y axes, and which one will be used as analysis of response. Your phospho-signal will typically be your Z parameter

Log: Choose whether data is displayed and binned on log or linear scale. Most antibody staining will require a log scale. Typically, only parameters such as Forward / Side Scatter or cell cycle DNA dyes will require linear scaling.

Next: Performs binning and moves to initial plotting of data. This is a memory-intensive process, especially for large files. Be prepared to wait one second for every 10^5 total cells to be binned (on a computer with 2 GB RAM). Freeing memory by closing other programs will speed the process.

Initial Plot

This window allows you to set up binning and display parameters and to inspect the phospho-responses before fitting.



Binning: By default, the data is sliced into 5 bins per log decade. Bins meeting a minimum cell fill threshold are outlined in red. In the example at right, the threshold at is 2 cells (the default), so bins with 2 or more cells will have red outlines.

Display: The default display indicates only the number of cells per bin, and may display more range than necessary. Use the ***Options*** button to change these defaults (described next page).

File: Using the arrow buttons, you can page through to evaluate each of your conditions.

If you're content with the default binning (or after adjusting the options) use the analysis buttons to process the data. Both 2D and 3D methods perform Hill fits within each bin, across all concentrations. Since all bins may not be full in all conditions (especially around edges), the fitting algorithm will only attempt a hill fit on a bin if there are sufficient cells in four or more conditions.

Analyze 2D: Slices the data independently into X and Y bins (in this case, either 20 X bins or 10 Y bins, for a total of 30 perpendicular bins), then performs separate analysis on each set of slices.

Analyze 3D: First slices data into X bins, then sub-slices each of those into Y bins, (in this example there will be a total of 20 x 10 = 200 bins).

Initial Plot Options - specify binning and analysis parameters.

Plotting and ***Binning*** Options are exactly what they sound like.

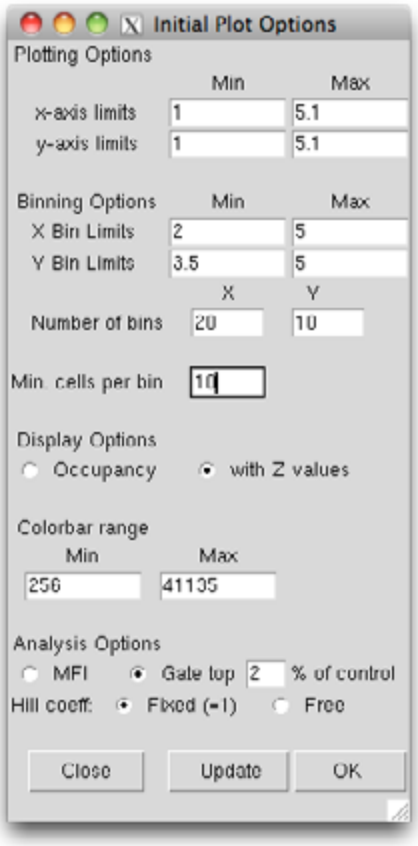
Min cells per bin: Sets the threshold for the minimum cells for bins in which ScatterSlice will attempt to fit. This helps eliminate some of the noise arising from fitting bins towards the edges.

Display Options: Offers the choice between "Occupancy," which shows the scatterplot with red outlines displaying occupancy, and "with Z values," which overlays squares on the bins, color-coded by Z value.

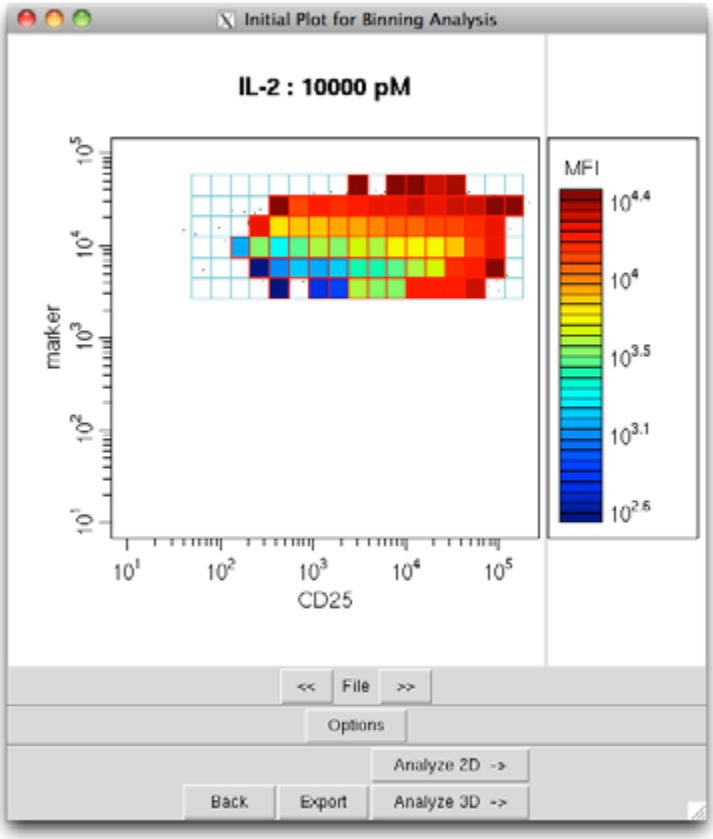
Colorbar range: Set range for bottom and top of Z value display.

Analysis options: Choose how Z values are calculated for display and further analysis. "MFI" - mean fluorescence intensity of the Z parameter. "Gate top X% of control" - Z value will be calculated by creating a gate that only includes the highest X% of points in the control file (selected in the "File Entry" dialog).

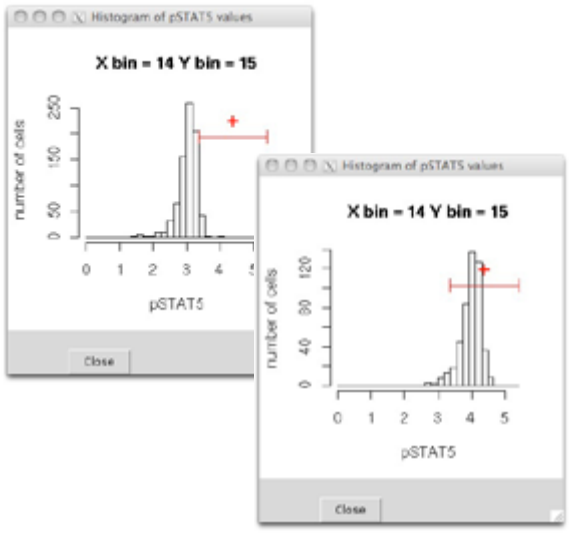
Hill coeff: option allows you to constrain the Hill coefficient ("Fixed") for dose-response fits, aiding in fitting noisy data (for more information on Hill fitting, see Appendix).



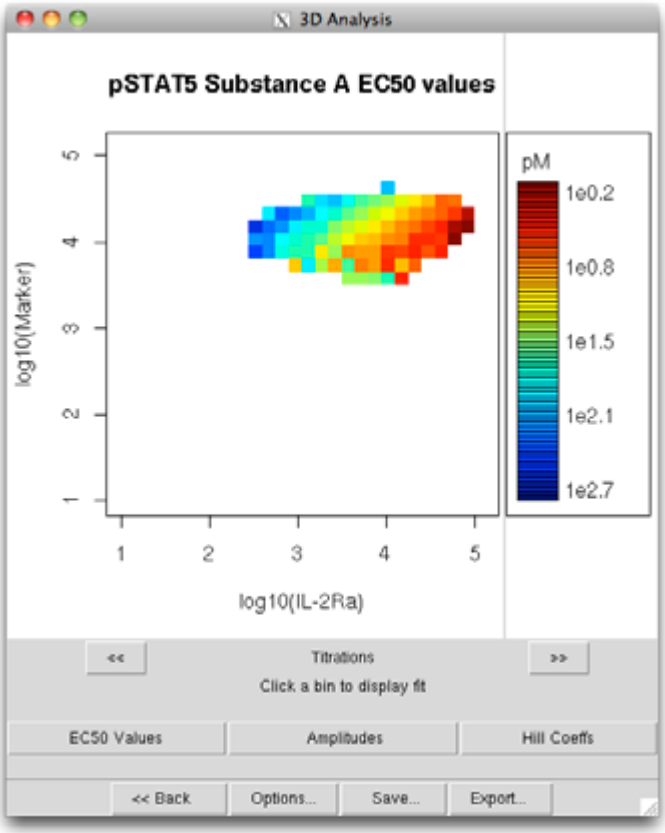
After setting the options above, the Initial Plot window looks as below. If you're interested in the plot for each dose in your titration use the ***Export*** button. This opens up a dialog to export a PDF showing binning and Z levels across all concentrations.



Clicking on a bin will pop up a histogram of the Z parameter in a new window. The gate labeled "+" is drawn if the "percent positive" is chosen in the options, and calculated using the cutoff specified in the options (in this case, top 2%).



3D Analysis



Selecting 3D analysis (and potentially waiting patiently for a minute or two, if you have many cells) displays the output: a graph of ***EC50 values***. Select ***Amplitudes*** or ***Hill Coeffs***: to display the respective graphs.

(Note: unless a "free" hill parameter was chosen in the Initial Plot options, there will be no Hill Coefficients to display.)

Color Bar: Red EC_{50} values indicate ***lower*** EC_{50}, whereas red amplitude values indicate ***higher*** amplitude. This color scheme was chosen such that increased responsiveness will always correlate with redder colors.

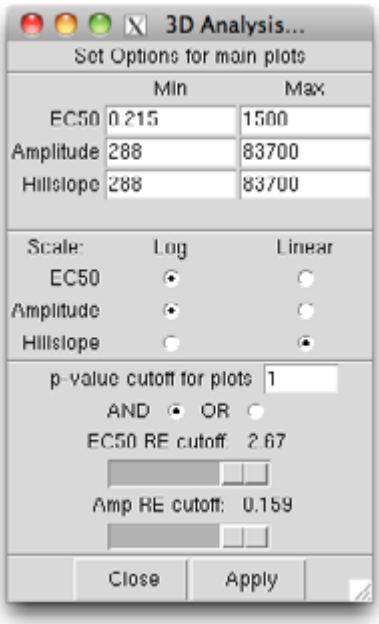
Options: Opens display options, described below

Save: Stores all data and settings. This will allow you to return to exactly this point by reloading this analysis in the first window.

Export: Opens a dialog to export fit values as text files and images as PDF files.

Clicking within any bin opens a plot of the concentration and Z values used for fitting, overlaid with the fit curve.

3D Analysis Options - Settings to adjust how fits are displayed



EC50, Amplitudes, Hillslope Min/Max: sets minimum and maximum display values. Any values above the maximum will be displayed the same as the maximum (i.e. dark red), likewise for the minimum.

Scale: log or linear color scaling

p-value cutoff*: allows the user to remove some bins from the graph that were not well fit by the dose-response curve.

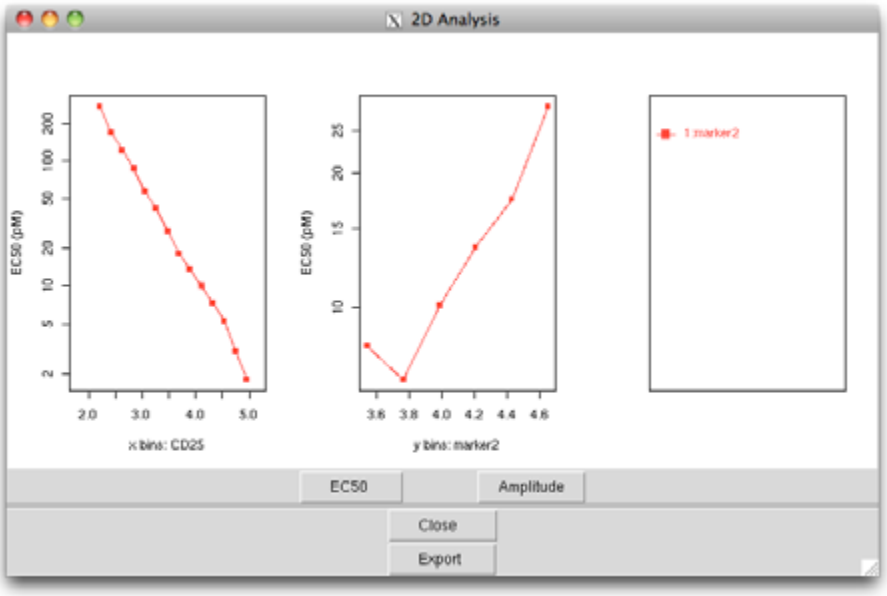
EC50 RE cutoff: allows the user to remove some bins from the graph that were not well fit by the dose-response curve, based on relative error. Relative error will vary in absolute amount from experiment to experiment. Reducing this value increases the certainty of the fits, but decreases the number of bins displayed.

The "3D Analysis" graph shown at the top of the page has been adjusted appropriately, such that the color of the bins spans the entire range of the color spectrum, without saturation at either the upper or lower bound.

***p-value cutoff**: Setting the p-value cutoff to 0.05 is not recommended. This is a statistically valid measure of how well the curve fits the data, given the uncertainty (noise) in the data. However, this means that a low p-value is easier to achieve in cases where data is noisy, but more difficult in cases where there is small variation. For this reason, this metric should only be applied as a subjective measure of goodness of fit.

2D Analysis

The 2D display within ScatterSlice is noticeably less sophisticated than the 3D display. This is because other programs exist that already display line graphs very well. For this reason, the primary function of the 2D analysis window is to quickly check your results before exporting, using the ***Export*** button.



The exported file will be named by the titration names, followed by "Analysis in 2D.txt". Each file contains a number of data tables, and is easily readable by Microsoft Excel (or R). The headers of the columns for each table are (using the data at right as an example):

x or y: The centers of the bins in the x or y direction, respectively

Substance A:IL-2Ra: Titration of substance A, binned by IL-2Ra. This is to indicate which titration and which binning was used.

The tables included in this output are:

ec: There are two tables of EC_{50} values. One for x bins and one for y bins.

ecRE: The relative error in EC_{50}, calculated from the hessian of the fit. This is a relative estimate in the size of the error given the size of the estimated value.

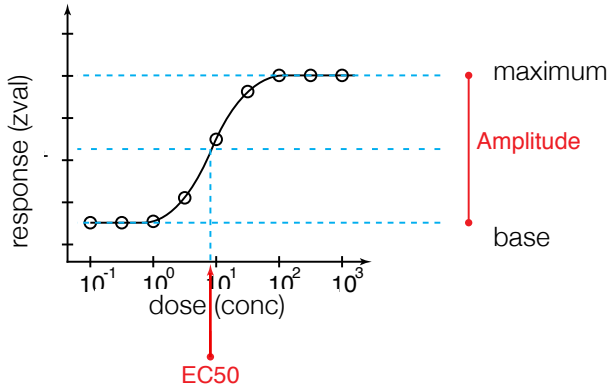
ampRE: Same as the ecRE, but for amplitude

	A	B	C
1	Values exported from ScatterSlice		
2	Tue Feb 26 17:14:54 2013		
3			
4			
5	MFI values represent bin centers		
6			
7			
8	ec		
9	x	1:CD25	
10	1.79529412	NA	
11	2.00588235	NA	
12	2.21647059	268.973602	
13	2.42705882	168.550488	
14	2.63764706	120.963842	
15	2.84823529	86.0063509	
16	3.05882353	56.8958081	
17	3.26941176	41.3633459	
18	3.48	27.2491999	
19	3.69058824	18.0316035	
20	3.90117647	13.5821318	
21	4.11176471	9.93083818	

Appendix A: Curve Fitting

Curve fitting

Our analysis fits a curve that expresses the cellular phospho-response as a function of dose of stimulus, amplitude, and EC_{50}, illustrated at right.



Dose response curves are fit to a four parameter log logistic equation:

$$\textbf{\textit{zval}} = base + \frac{maximum - base}{1 + \exp(hill(\log(\textbf{\textit{conc}}) - \log(ec50))}$$

Within each bin, two parameters are derived from the data:

zval - mean of z values within the bin, at each dose.

conc - concentration of stimulus

Four parameters are then fit, using the "drc" package:

base - lower end of response

maximum - upper end of response

hill - hillslope, a measure of curve steepness

ec50 - EC_{50} is defined as the concentration at which half of the distance between base and maximum has been traversed

For presentation purposes, the linear difference between maximum and baseline is calculated and expressed as amplitude

Appendix B: Analysis of ScatterSlice output with R

For users who wish to analyze or re-plot ScatterSlice output within R, we have included a number of variables that are saved to the global workspace as lists to allow direct access. The $ after a variable indicates that it contains further variables. Variables without a $ contain numerical data in either atomic, vector or matrix form.

ScsAnswers[[*titration*]]$ - List with length = the number of titrations. Stores the final results from fitting for each titration. Indexed numerically (e.g. ScsAnswers[[1]] is the first titration.)

Key sub-variables:

bas, amp, ec and ***hill*** - matrices of baseline, amplitude, EC50 and hillslope, respectively.

Values are stored as matrices of answers for each fit parameter. For example, *ScsAnswers[[1]]$amp[5,6]* access the amplitude for the bin 5 *x* and 6 *y* bins from the bottom-leftmost bin, in the first titration.

pvals - associated p-values.

ecRE, ampRE - relative errors from fits, calculated from the fitting hessian.

fitdata$ - list containing data used for fitting each bin's dose response, if fitting was successful. Contains ***concs*** and ***means***, vectors specifying concentration of stimulus and mean of Z values, respectively

models$ - list containing fitted parameters for each bin

Both *fitdata* and *models* contain lists named by bin number. *ScsAnswers$models[["12 , 14"]]* contains the coefficients to the hill equation fitting *ScsAnswers[[1]]$fitdata[["12 , 14"]]$means* as a function of *ScsAnswers$fitdata[["12 , 14"]]$concs*

ScsVars$ - Information used for fitting. Some of these variables apply across all titrations, whereas others are set on a per-titration basis. Those that are indexed by titration are denoted with "[[titr]]".

Key sub-variables:

NumFiles - number of input files

Control.File[[titr]] - index of control file

Concs [[titr]] - vector of concentrations entered in File Entry window

Datas[[titr]][[dose]]$ - list of ***x***, ***y***, and ***z*** data selected in Data Prep window. *x* and *y* are log transformed if this was specified in Data Prep. Numbered by order of entry.

xyzLabels[[titr]]$ - list of ***x***, ***y***, and ***z*** column labels

units[[titr]] - string specifying units, entered in File Entry

islog$ - list of whether *x, y,* and *z* were identified as log in Data Prep

ScsBins$ - Binning parameters and information

nx, ny - number of *x* and *y* bins

seqx, seqy - vector specifying sequence of *x* and *y* bin edges, including lower and upper limits. (There are *nx*+1 elements in *seqx*.)

centersx, centersy - vector of bin centers for *x* and *y* bins. Useful for plotting data.

sizes[[titr]]$, means[[titr]]$, pospct[[titr]]$ - lists holding one matrix per input file of the number of elements, mean *Z* value, and calculated percent positive cells in each bin. E.g.: *ScsBins$sizes[[1]][12,14]* returns the number of cells in the 12th *x* and 14th *y* bin of the first entered file.

pos.cut.mat[[titr]] - matrix of fluorescence intensities defining positive cutoffs, used to calculate *pospct*

binsmax, binsmin - maximum and minimum mean z value for all bins, across all files

***Full$*** - list of full set of z values in every bin, indexed by X, Y bins and file number. E.g. ScsBins$Full[["X12.Y14.Z1"]] returns a vector of values reflecting all of the cells binned into the 12th x and 14th y bin, in the first file.