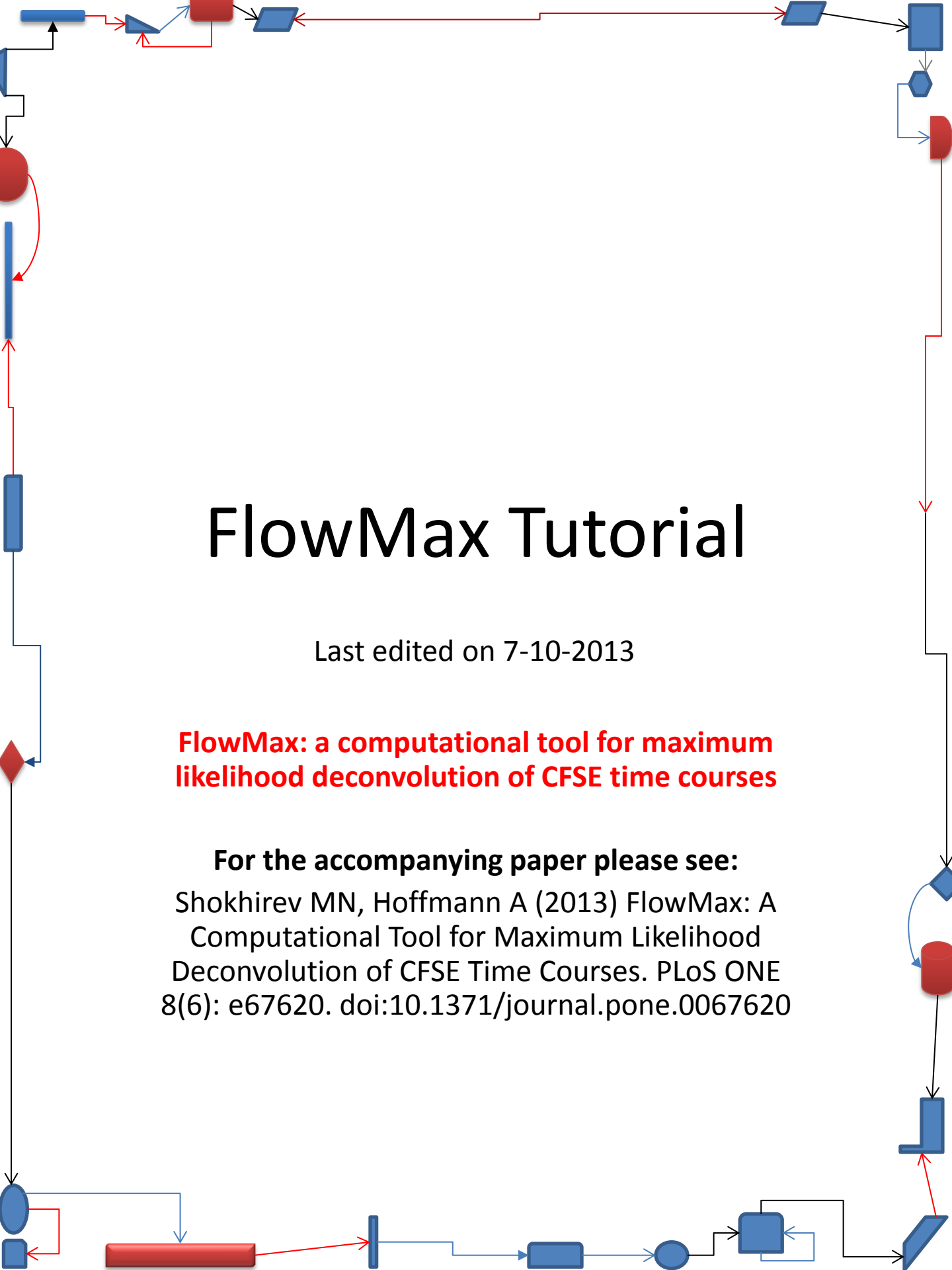


FlowMax Tutorial

Last edited on 7-10-2013

FlowMax: a computational tool for maximum likelihood deconvolution of CFSE time courses

For the accompanying paper please see:

Shokhirev MN, Hoffmann A (2013) FlowMax: A Computational Tool for Maximum Likelihood Deconvolution of CFSE Time Courses. PLoS ONE 8(6): e67620. doi:10.1371/journal.pone.0067620

Contents

Running FlowMax

FlowMax was created using the cross-platform Java 1.6 programming language and packaged as a standalone "jar" file.

Running FlowMax (operating system independent):

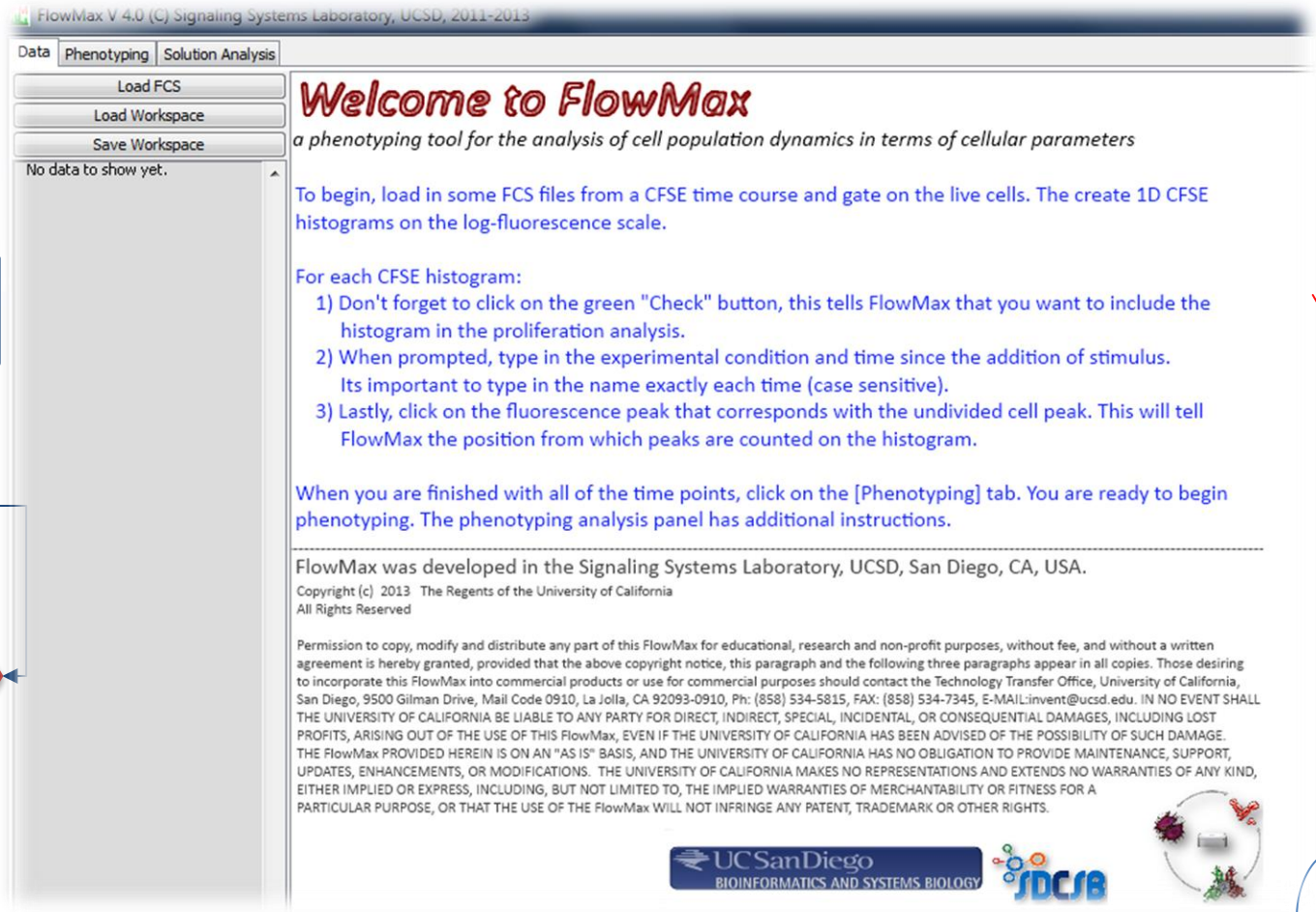
1. Ensure that the latest java virtual machine has been installed on your computer. In the command prompt or console typing: "java" should bring up usage instructions. If java is not recognized, you may need to add its location to your $PATH variable.
2. FlowMax may require extra memory for proper execution and should not be running from a restricted directory (such as "Program Files" in Windows) as it will need to create and edit files during operation.

   To execute from command prompt or console run:

   "java –Xmx1024m –jar FlowMax.jar"

   To execute from Windows, create a batch file (new text document which you should save as "FlowMax.bat") containing the line: "java –Xmx1024m –jar FlowMax.jar" Make sure FlowMax.jar is in the same directory as the batch file and run by double-clicking "FlowMax.bat"
3. If you would like access to the FlowMax java code, you can extract it directly from the jar file.

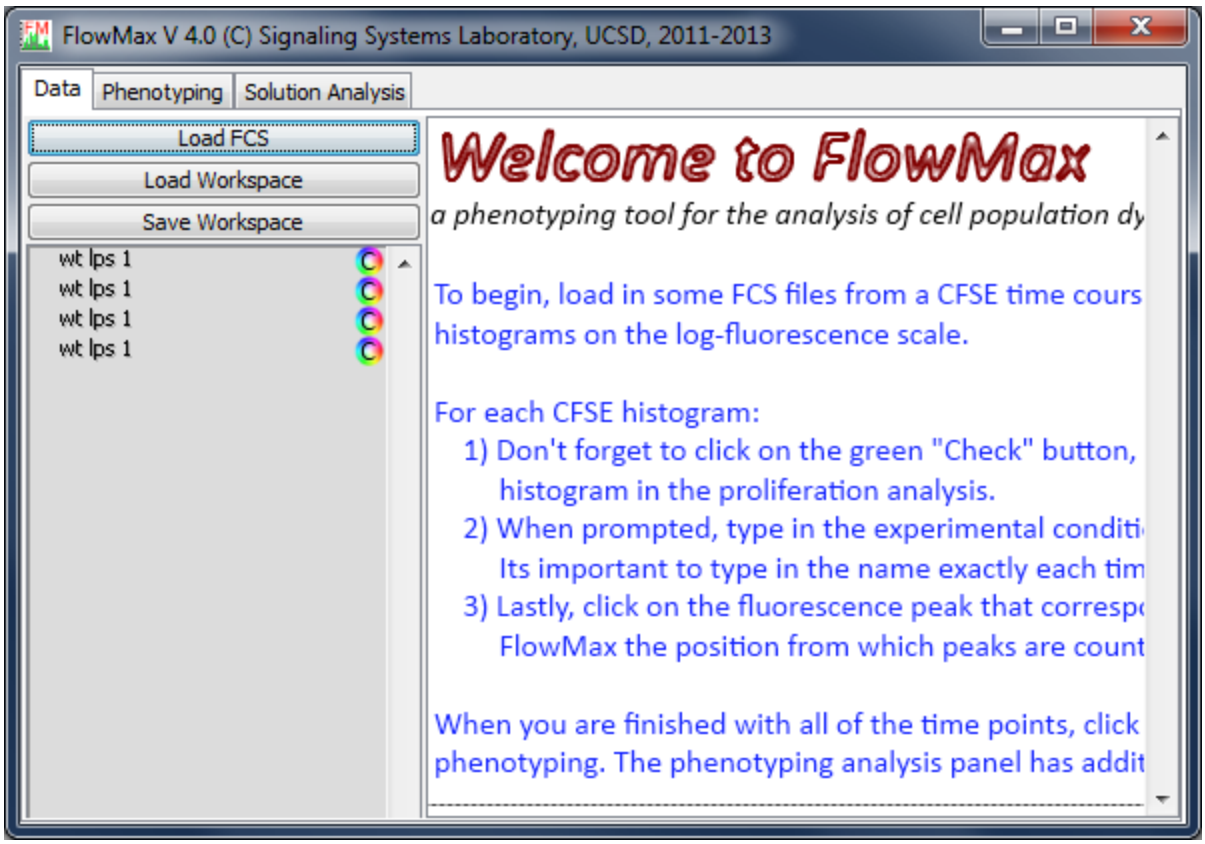
When you start up FlowMax you should see the FlowMax welcome screen:



Building log-fluorescence histograms

After successfully launching FlowMax, we will load some FCS data files representing a few time points in a typical CFSE time course:

1. Click on the “Load FCS” button, select the FCS files you would like to process, and load them into FlowMax:

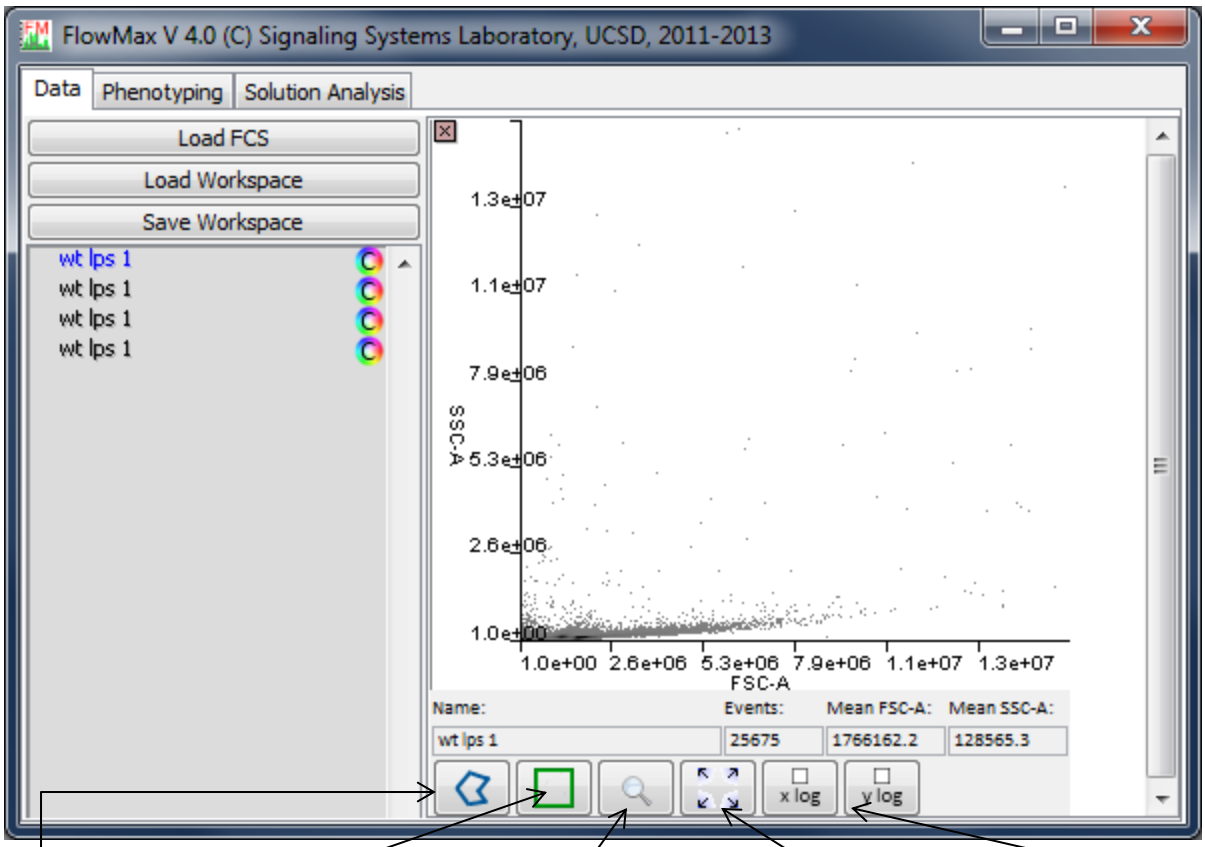


Here I have loaded four FCS files from a 145 hour CFSE time course. They should appear in separate rows on the left selection panel (I recommend using more descriptive names when collecting the data points and saving to FCS).

You can always save and load the workspace by clicking the corresponding buttons. Workspaces contain all of the gating, compensation, and colorization processing along with the relative locations of all loaded FCS files. FCS files need to be reloaded each time the workspace is loaded, so FlowMax may ask you to locate moved or missing FCS files referred to by a workspace file.

Building log-fluorescence histograms (2)

Next, I selected the first time point by double left clicking. FlowMax plots this dataset on the first two recorded axes (usually Forward and Side Scatter). The selected dataset is now blue. Notice that the name, # of events, and the mean intensities are shown at the bottom of the 2D plot. Below that is a toolbar with six buttons that are described below. You can close a plot by hitting the red [x] in the top left corner. Scroll through multiple plots with the scrollbars.



Polygon Gate Tool

Rectangular Gate Tool

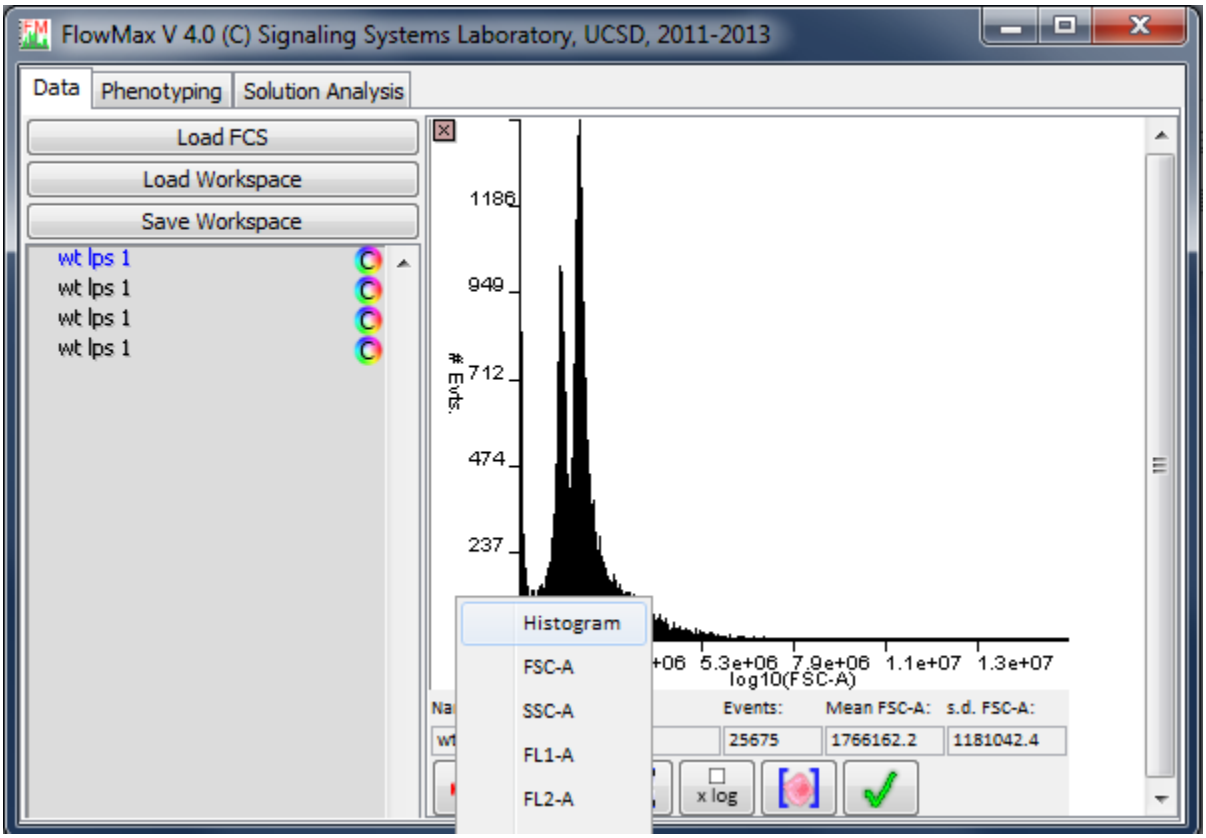
Zoom Box Tool

Zoom Out Tool

Change linear/log X and Y scales

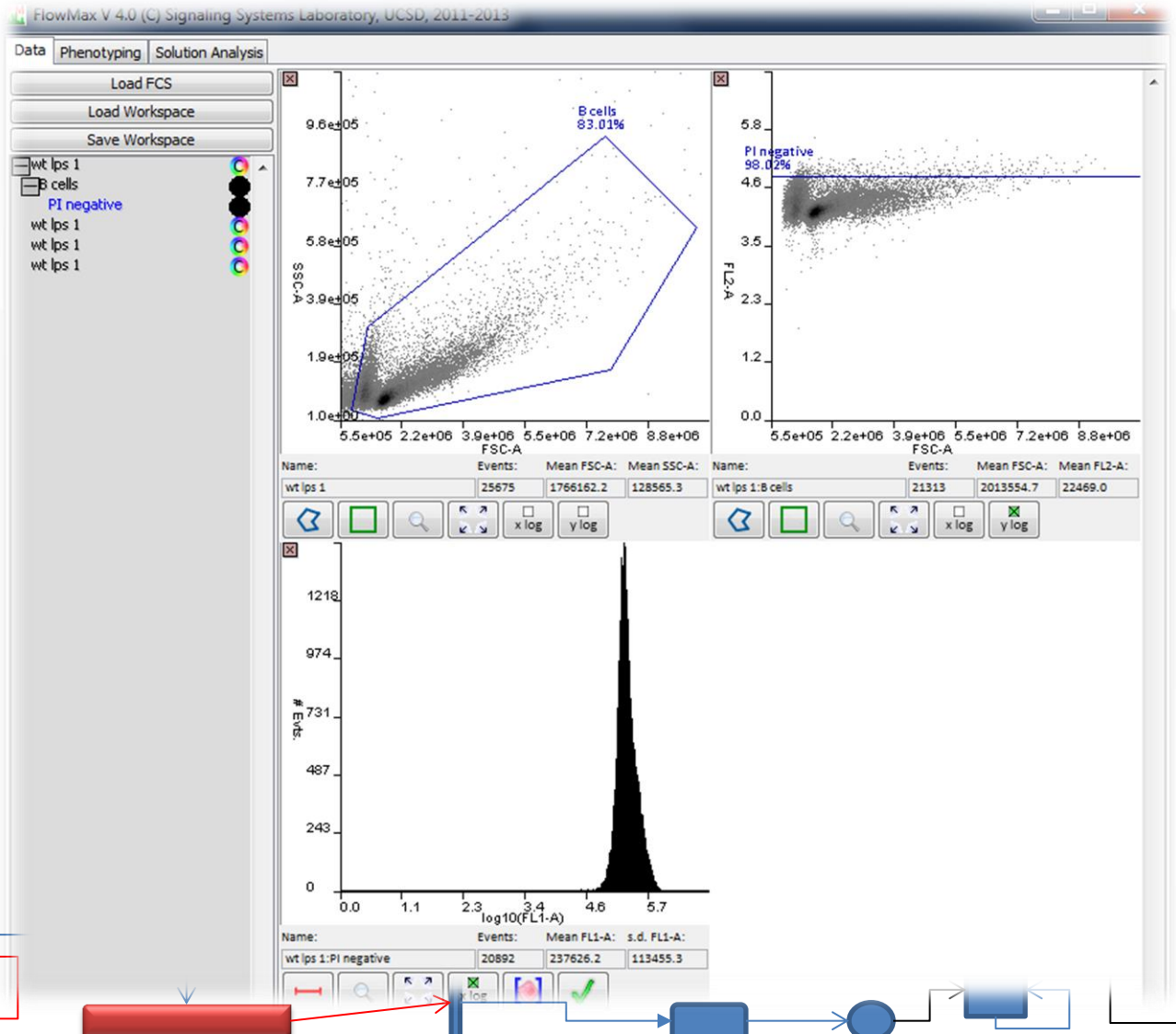
Building log-fluorescence histograms (3)

To change axes, right click on either the x or y axes labels and select a different measurement from the list. In addition, you can choose to plot a 1D histogram by selecting the histogram option when right-clicking the y axis:



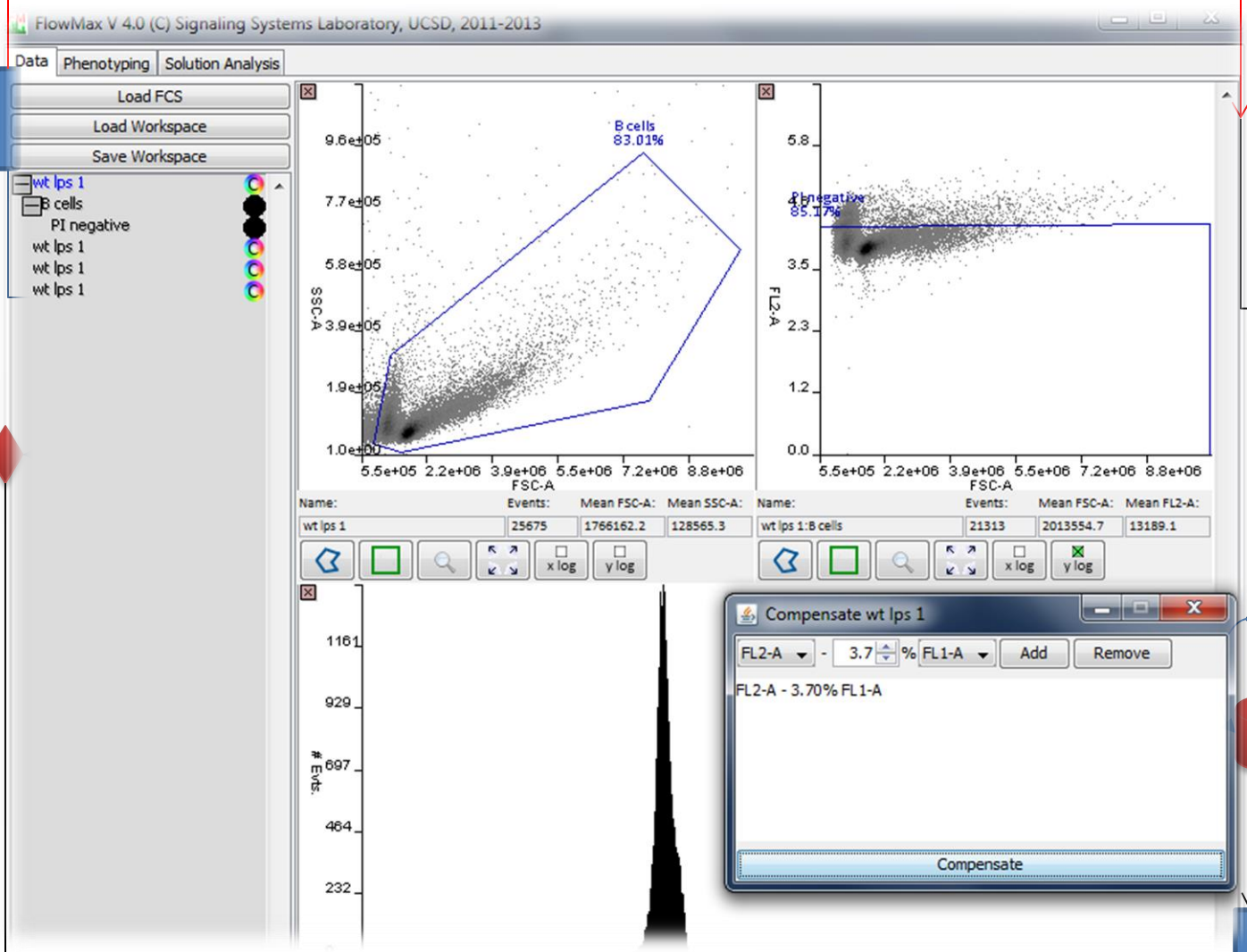
Building log-fluorescence histograms (4)

FlowMax assumes that the selected log-fluorescence histograms represent the log-fluorescence CFSE histogram of a single population of living cells. Therefore, you will need to make sure that the final 1D log-fluorescence histogram represents the CFSE fluorescence of living cells (for example 7AAD negative or PI negative). If multiple populations are present, its important to gate on a subset (e.g. naïve B cells only). Below I have highlighted my gating strategy for the sample dataset:



Building log-fluorescence histograms (5)

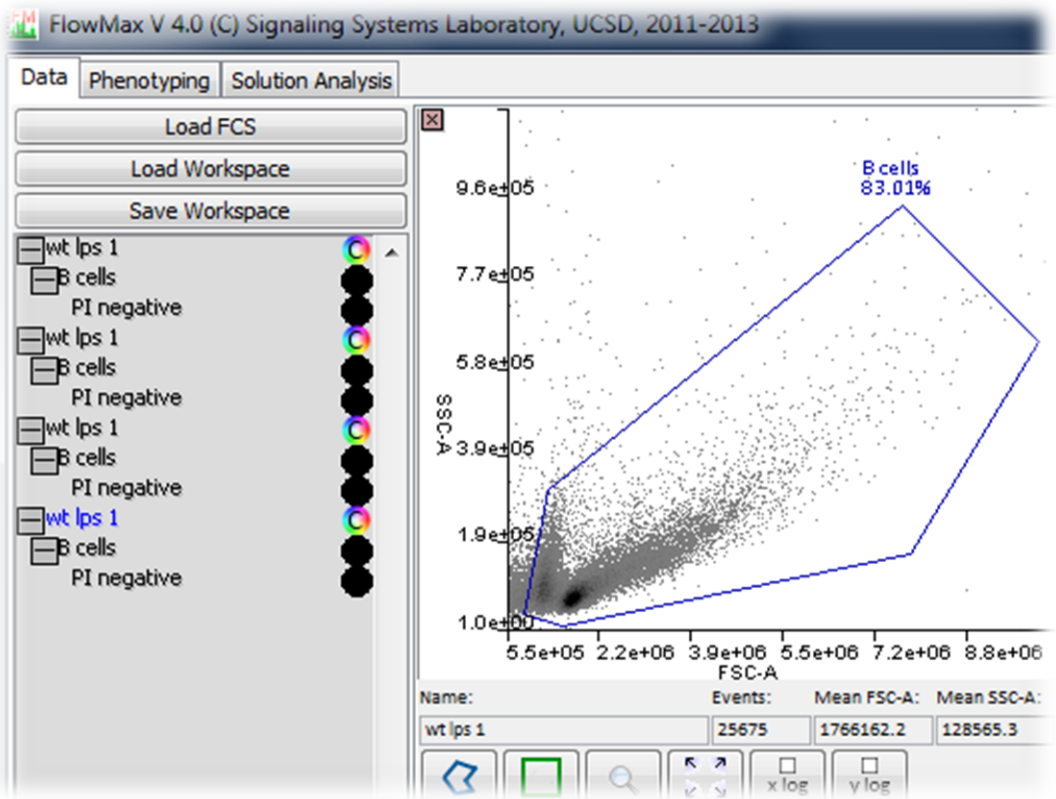
If fluorescence “spillover” is suspected, its possible to compensate mathematically, by subtracting a fraction of the fluorescence from one channel from another channel. To do this, click the multi-colored (C) icon to the right of the dataset, and add compensation rules to the list as I have done below:



It may be necessary to adjust some gates to account for the reduced values, as I have done with the “PI negative” gate above.

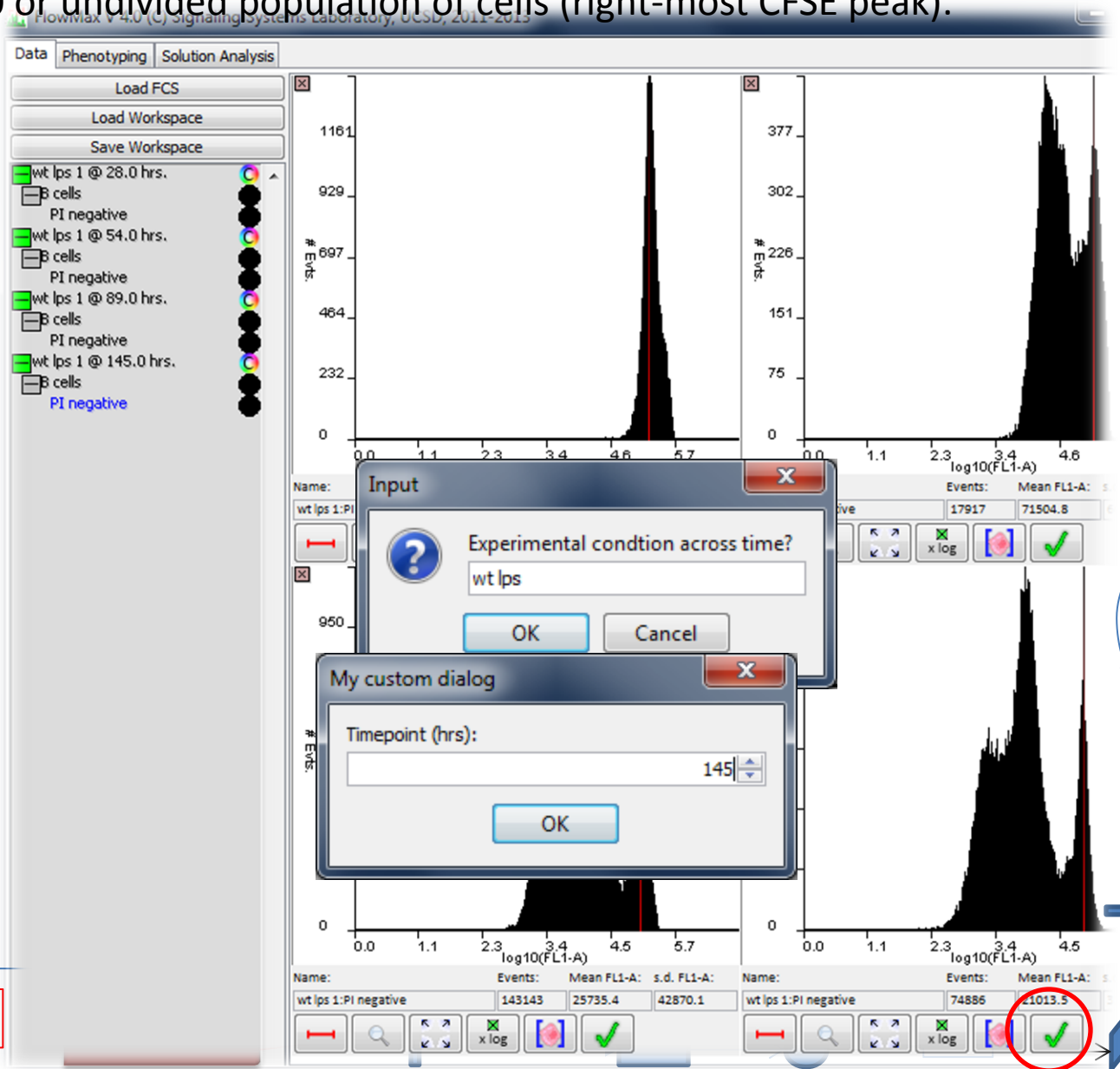
Building log-fluorescence histograms (6)

Finally, once you are happy with all of the gating and the computational compensation, its time to select the log-fluorescence CFSE histograms you are interested in modeling. One way to speed this up, is to copy and paste all of the gates from one dataset unto the other datasets by right-clicking the top-level gate in the selection panel on the left copying and pasting onto other datasets ("Ctrl-C and Ctrl-V" keyboard shortcuts should work as well):



Building log-fluorescence histograms (7)

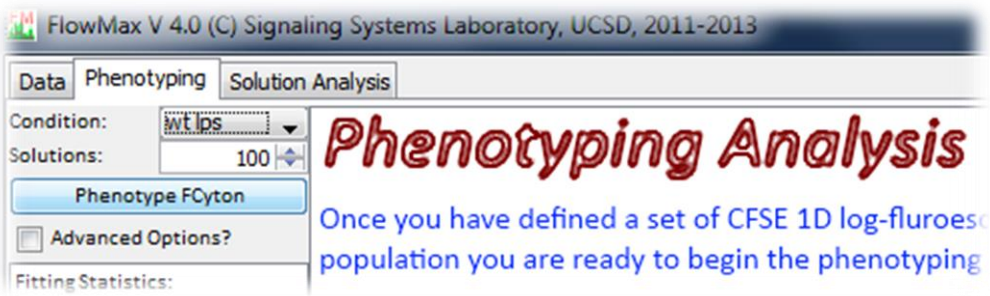
Now that all of the 1D CFSE histograms are constructed, click on each 1D histogram (PI-negative gate) for each FCS dataset and click on the green check button. A popup window will appear asking you to indicate the name of the experimental time course that this histogram belongs to and the time after stimulation with 0.5 hour resolution. Finally, you need to click on the center of the generation 0 or undivided population of cells (right-most CFSE peak):

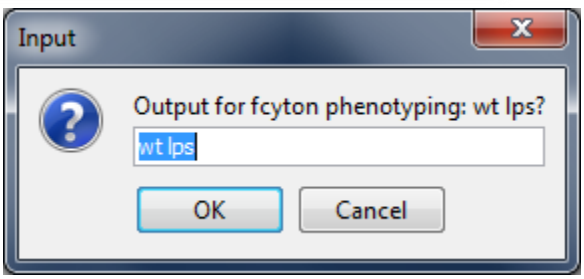


Proliferation Analysis

Now that all of the CFSE histograms have been constructed, FlowMax is ready to fit the pcyton/cyton models to the histograms using a simple cell fluorescence model as an adaptor during fitting. The methodology for how FlowMax performs the fitting is described in detail in the accompanying publication. In short, first FlowMax fits a mixed Gaussian model to each histogram assuming approximate halving of the dye fluorescence with each division, and equal coefficient of variability. Then, FlowMax uses the best-fit cell fluorescence parameters during cell population model fitting to the histograms.

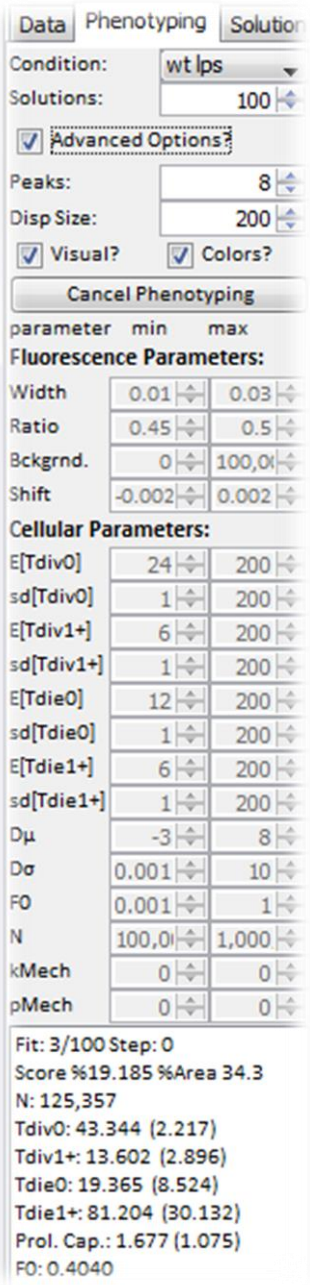
FlowMax automates this process as much as possible requiring the user to simply switch to the “Phenotyping” tab, select the time course label (e.g. wt lps in our example) , click the “phenotype pcyton” button, and indicate where to save the analysis results once they are loaded:





Proliferation Analysis (2)

If you know the ranges for some of the cell fluorescence and population model parameters, you can specify the minimum and maximum values under the “Advanced Options” panel.



The number of solutions should be high enough to allow for all possible solution clusters to be found. This will vary widely on the quality and type of data.

The number of peaks should be set to ensure that the entire histogram is modeled. Typically eight peaks is the maximum.

Display size determines the size of the histograms being drawn during fitting as well as their saved size.

Deselecting “Visual” will improve running time on some older machines.

Selecting “Colors” will draw colored semi-transparent backgrounds for all histograms indicating the quality of the current fit.

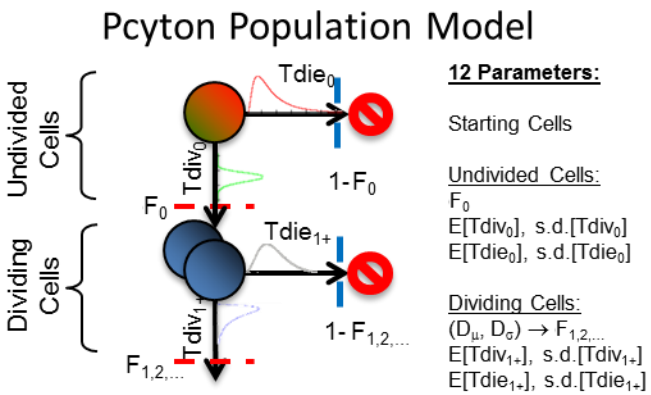
Coefficient of Variation or CV (This parameter determines the standard deviation of the gaussians)

Ratio determines the distance between peaks in a log-transformed histogram.

The shift is a linear adjustment to all peak used to compensate for poor selection of the undivided peak.

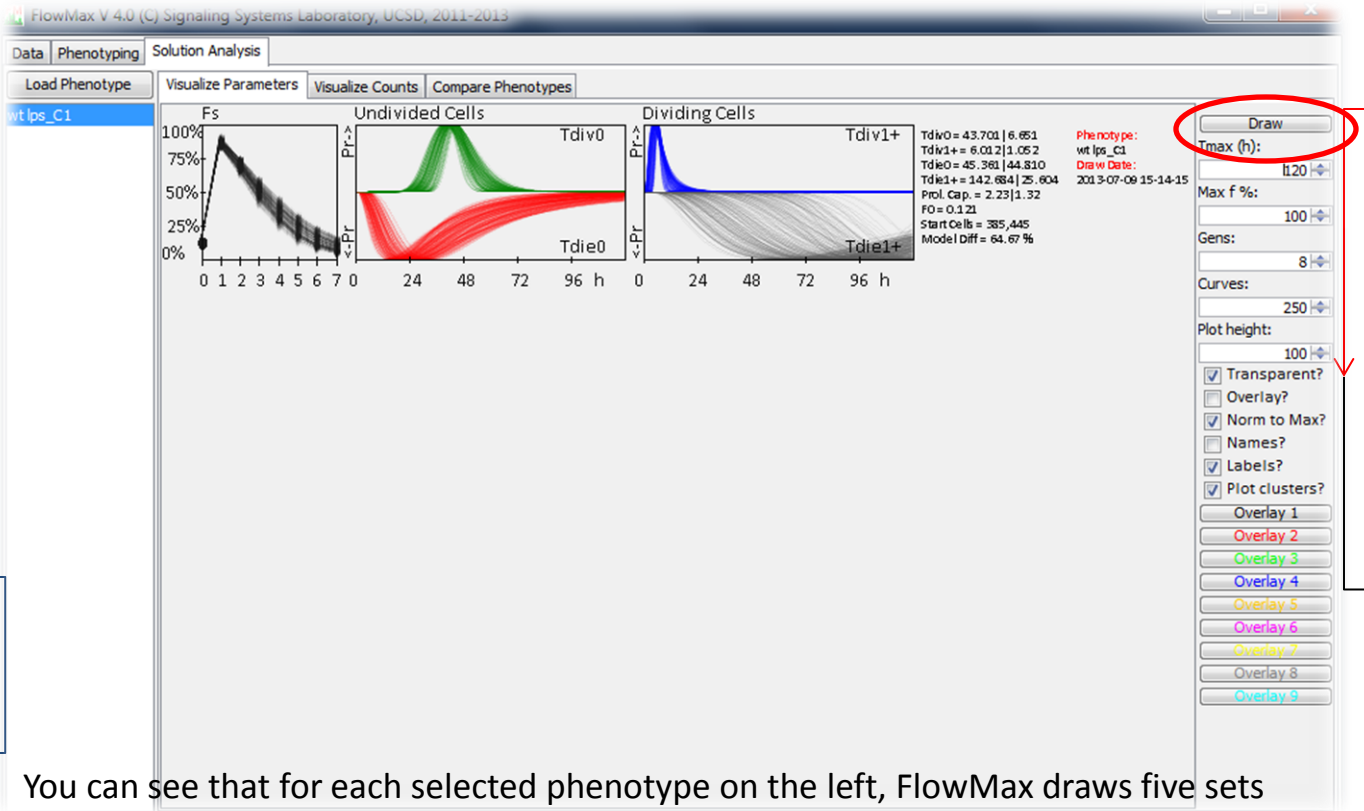
Background fluorescence determines the minimum fluorescence value (this is the cellular autofluorescence)

The time to die and time to divide parameters determine the log-normal distributions describing division and death of undivided and dividing cells, while the Dμ and Dσ parameter determine the mean and standard deviation of the number of divisions that dividing cells will undergo (proliferative capacity). The F_0 parameter determines the fraction of the generation 0 cells that respond. Responding cells are protected from death and will divide, while non-responding cells will die:



Post-processing
(visualizing population parameters)

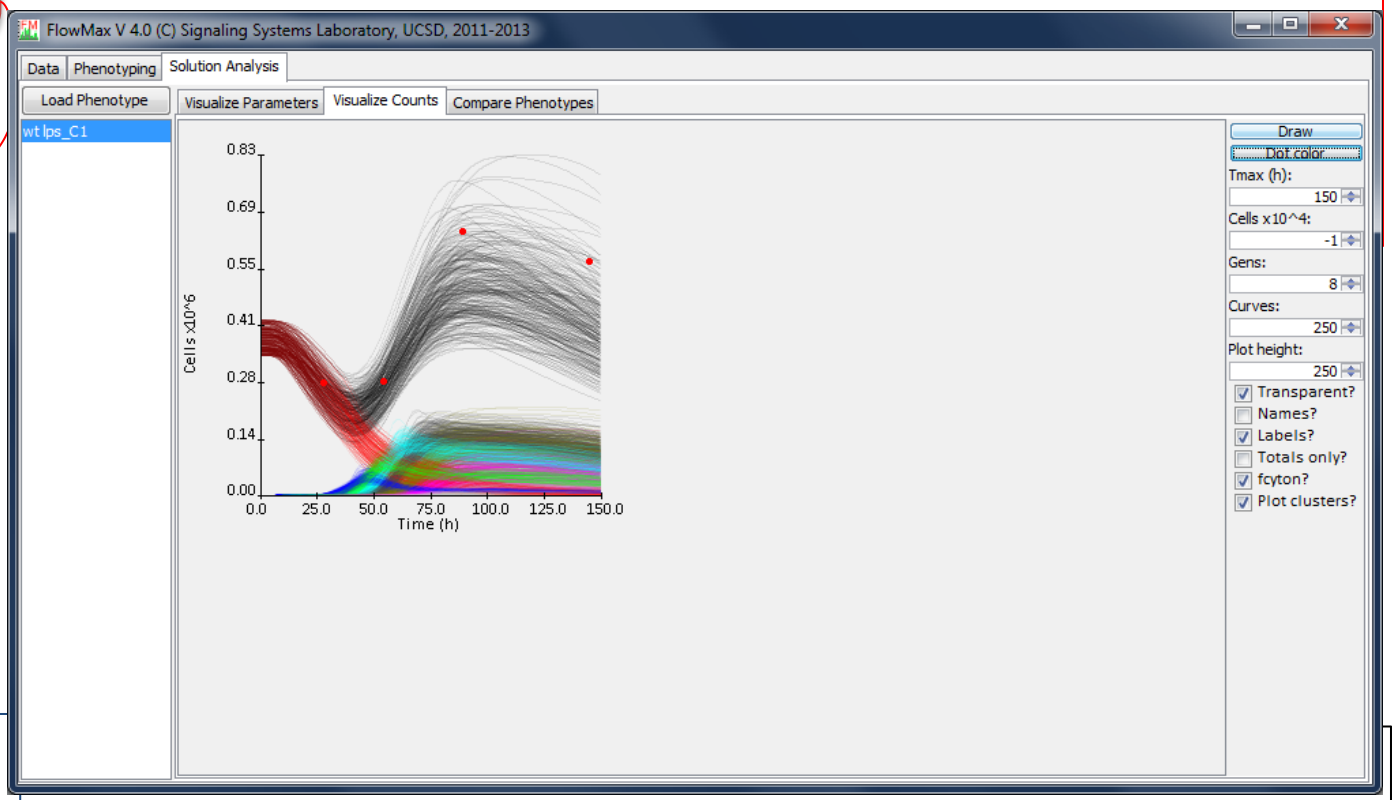
After phenotyping a CFSE time course you can visualize/analyze the result using the “Solution Analysis” tab. First lets take a look at plotting the best-fit parameters for a solution cluster:



You can see that for each selected phenotype on the left, FlowMax draws five sets of distributions representing the sampled maximum-likelihood lognormal distributions to divide and die as well as the response distribution as a function of generation number (the Fs). The cluster average population model parameters are summarized on the right.

Under the draw button you can change visualization settings such as the maximum time in hours plotted, the maximum F value, the number of generations to plot, the total number of sampled curves, the size of the graphs, as well as the transparency, overlay, normalization, and label options. When overlaying multiple phenotypes in the graph, you can change the color of each curve manually by clicking on the appropriate Overlay button and selecting the color of your choice from the color triangle (To change gamma, use the middle mouse button). Uncheck the plot clusters box to plot the individual clustered solutions instead of sampling from the maximum-likelihood cluster ranges for parameters.

Post-processing
(visualizing population counts)

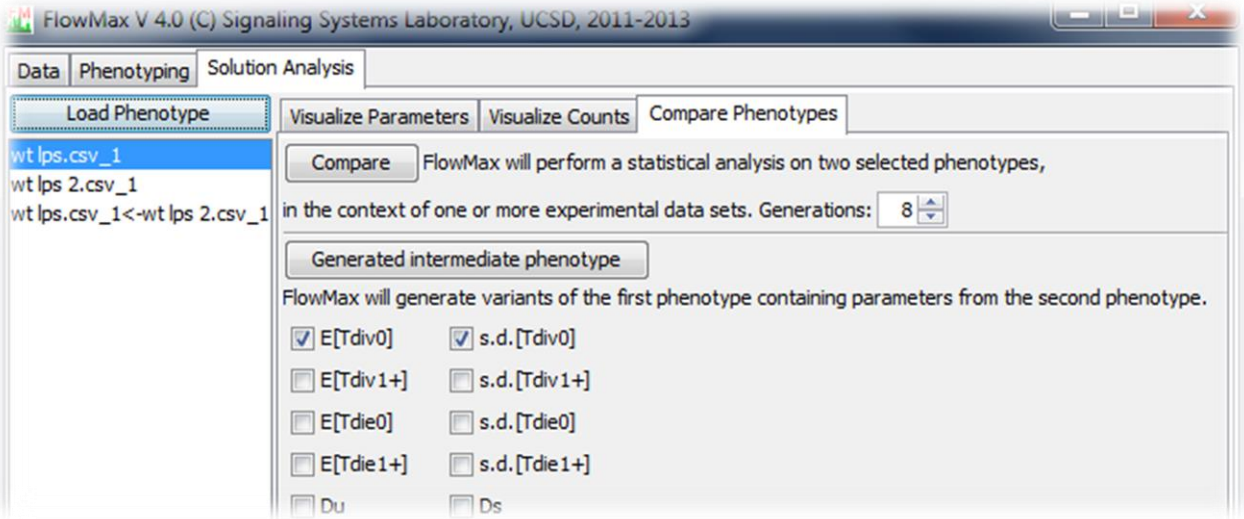


To visualize the best-fit cluster cell counts as a function of time, switch to the "Visualize Counts" sub tab, select the phenotypes from the list, and click the "Draw" button. FlowMax will sample from the maximum-likelihood parameter ranges of each selected cluster and draw the model solutions for each generation as well as the total (black) as a function of time. The experimental cell counts are plotted as dots with SD error bars (when available).

You can adjust the visualization properties on the right. Specifically you can adjust the time axis, the cell count axis, the number of generations plotted, the number of samples plotted, the size of the graph, whether plots should be transparent, whether names and labels should be plotted, whether individual generation counts should be plotted, whether the fcyton vs cyton model is to be used, and whether solutions are sampled or taken directly from the best-fit solutions in the cluster. The color of the experimental dot counts can be selected by hitting the "Dot color" button and picking the color from the color triangle (scroll mouse wheel to adjust gamma).

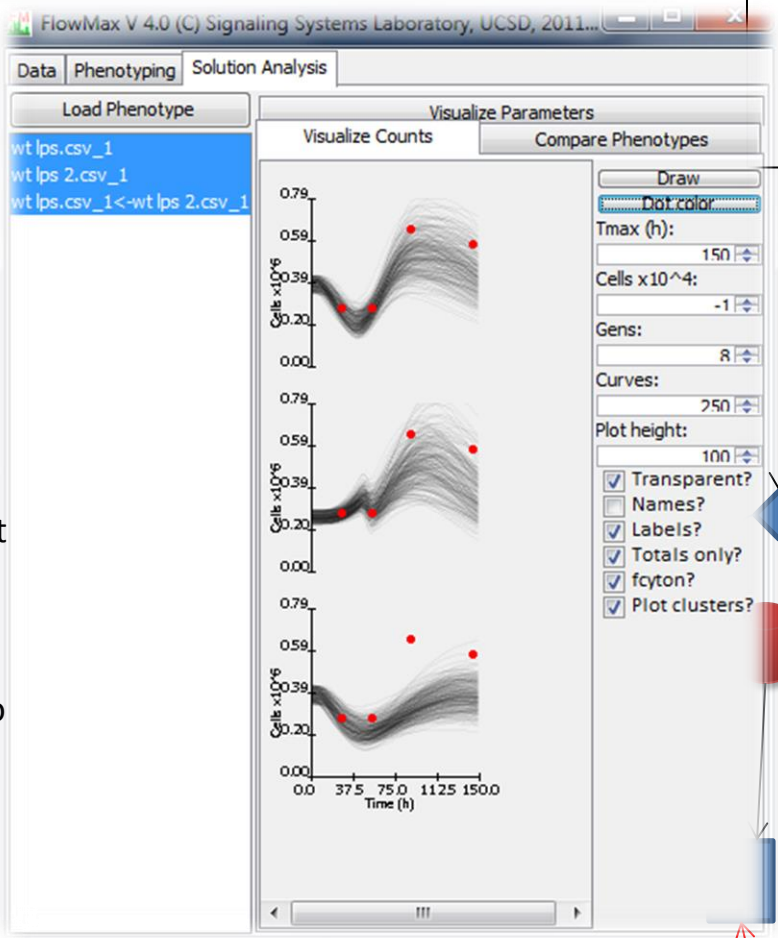
Post-processing

(generating chimeric phenotypes)



To determine which population parameters are sufficient for describing the difference between two phenotypes, FlowMax can be used to generate “chimeric” phenotypes containing parameters from both phenotypes. Check the parameters you would like to copy from the second phenotype, and hit the “Generate intermediate phenotype” button. FlowMax will create a new phenotype containing the parameters from the first phenotype with the selected parameters taken from the second phenotype.

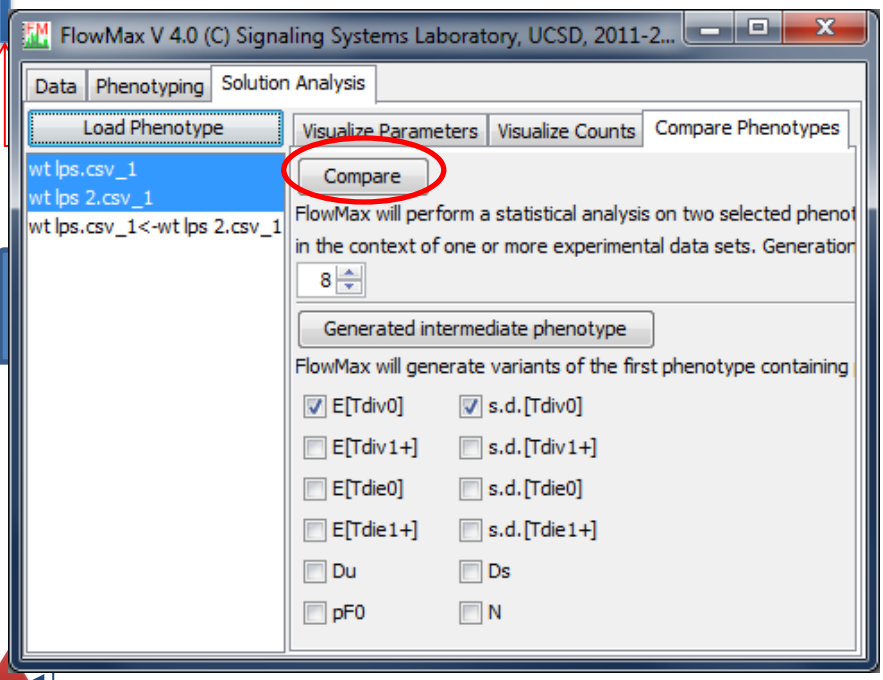
You can now plot the total cell counts to visually compare if the changes are sufficient for describing the population behavior. In the example at right, changing the $Tdiv_0$ parameters alone was insufficient for recapitulating the second phenotype.



Post-processing

(Determining if two solutions are different)

To get an upper estimate on the probability that two solutions are statistically different, a Monte Carlo sampling approach is used to estimate probabilities that any two solutions fit one or two separate experimental datasets. The methodology is described in further detail at the end of this tutorial.



Click the comparison button after selecting two phenotypes for comparison. Ensure that the experimental datasets corresponding to those two datasets have been loaded into FlowMax (FlowMax may ask you to specify which experimental condition corresponds to each selected phenotype). FlowMax will then refit the cell fluorescence parameters to each dataset, perform the sampling, and output the results to a csv file of your choosing:

Comparing phenotype	wt lps.csv_1	and phenotyp	wt lps 2.csv_1	in the context	wt lps	and	wt lps
Param	wt lps.csv_1	wt lps 2.csv_1	low wt lps.csv	low wt lps 2.c	high wt lps.cs	high wt lps 2.csv_1	
E[Tdiv0]	43.70111893	87.72750719	2.552241082	14.33130301	3.484390055	12.27249281	
sd[Tdiv0]	6.650726641	65.18236521	3.102556795	23.92220046	4.48833675	34.81763479	
E[Tdiv1+]	6.012473623	7.842638674	0.012473623	1.842638674	2.912806075	7.132451366	
sd[Tdiv1+]	1.052189947	1.33673314	0.052189947	0.33673314	5.91625213	13.6383569	
E[Tdie0]	45.36099332	50.25922694	16.02658028	2.259226937	21.07787584	3.585065076	
sd[Tdie0]	44.81018509	1.961783754	29.04737466	0.961783754	21.62868408	4.721799415	
E[Tdie1+]	142.6840512	69.7959059	43.65037457	39.28670594	57.31594878	2.204094097	
sd[Tdie1+]	25.60443663	68.50612153	9.098823858	59.90351667	174.3955634	3.493878466	
Du	2.226702001	2.24924456	0.243535337	0.289224241	0.27452106	0.227450913	
Ds	1.323959779	1.240590708	0.154924941	0.248511017	0.278078972	0.25520914	
F0	0.120728671	0.463121327	0.020182071	0.095311186	0.016879878	0.062296952	
Starting cells	385445.2427	260597.3542	47795.26448	41211.50632	40741.13151	26540.98204	
Average %difference	18.78662995	s.d.	1.708352844				
Average %difference	21.18884104	s.d.	1.91352066				
Probability that	wt lps.csv_1	describes	wt lps	better than	wt lps 2.csv_1	does =	1
Probability that	wt lps.csv_1	describes	wt lps	is	0.495114219		
Probability that	wt lps 2.csv_1	describes	wt lps	is	0		
Conservative probabili	wt lps.csv_1	is identical to	wt lps 2.csv_1	given the data	0		

Understanding the CVS file output

After population model fitting is completed, candidate solutions are subjected to a post-processing step which involves solution filtering by score and maximum-likelihood clustering. All solutions, and clusters are saved in a comma-separated text file (CSV) along with images of best-fit cluster model overlays.

Phenotyp	w t lps	model use	fcyton Mo	Finishe	2013-07-08 20-44-05								
Generati	w t lps	performed	2013-07-0	timepoi	4	replicat	1	genera	8				
Time	Gen 0	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 6	Gen 7	Gen 0	Gen 1	Gen 2	Gen 3	Gen 4
28	2E+05	45006.3	0	1715	0	61.25	15.31	0	2E+05	45006	0	1715	0
54	59121	43433	64452.4	84736	23088	0	0	0	59121	434			
89	79152	29167	30362.5	73316	1E+05	2E+05	1E+05	13051	79152	291			
145	31501	25510.5	16233.6	2E+05	2E+05	45138	44885	18554	31501	255			
fitScore	Weight	Tdiv0 u	Tdiv0 s	Tdiv1 u	Tdiv1 s	Tdie0 u	Tdie0 s	Tdie1 u	Tdie1 s	D u	D s	F0	Start C
0.0747	4.039	43.6756	6.6201	6.012	1.052	46.19	45.81	141.4	24	2.239	1.312	0.122	4E+05
0.0747	4.04	43.7198	6.6719	6.01	1.04	44.27	44.07	143.6	25.6				
0.0747	4.056	43.6423	6.6125	6.025	1.132	42.01	41.81	142.3	24.9				
0.075	4.059	43.3714	6.2055	6.038	1.077	51.3	50.64	140.7	23.6				
0.0747	4.06	43.4754	6.3699	6.017	1.009	43.38	43.35	142.6	24.0				
0.0858	4.283	39.0762	5.43177	7.311	1.262	62.36	55.89	145.4	44.1				
0.0906	4.374	106.664	88.5375	7.12	1.189	30.89	4.517	67.84	67.0				
0.0906	4.38	104.764	86.468	7.16	1.256	27.46	2.459	65.31	64.1				
0.0906	4.383	102.707	83.0305	7.059	1.22	28.86	5.029	66.7	64.9				
0.0907	4.391	100.959	80.433	6.93	1.17	26.42	6.778	70.14	69.25	2.24	1.305	0.258	5E+05
Total solu	10	After filter	2	Numbe	1	detrmin	After fi	1	clusters remaining				
Cluster 1	Averag	64.6686											
fitScore	Weight	Tdiv0 u	Tdiv0 s	Tdiv1 u	Tdiv1 s	Tdie0 u	Tdie0 s	Tdie1 u	Tdie1 s	D u	D s	F0	Start C
0.0747	4.04	43.7198	6.6719	6.01	1.04	44.27	44.07	143.6	25.69	2.213	1.331	0.12	4E+05
0.0747	4.039	43.6756	6.6201	6.012	1.052	46.19	45.81	141.4	24	2.239	1.312	0.122	4E+05
Param	Ar. Ave	Ar. s.d.	Ar. SEM	Overla	Low er	Upper							
Tdiv0 u	43.7	0.02214	0.01566	43.7	2.552	3.484							
Tdiv0 s	6.646	0.0259	0.01831	6.651	3.103	4.488							
Tdiv1 u	6.011	9.99E-04	7.06E-04	6.012	0.012	2.913							
Tdiv1 s	1.046	0.00639	0.00452	1.052	0.052	5.916							
Tdie0 u	45.23	0.95958	0.67852	45.36	16.03	21.08							
Tdie0 s	44.94	0.86589	0.61228	44.81	29.05	21.63							
Tdie1 u	142.5	1.10801	0.78348	142.7	43.65	57.32							
Tdie1 s	24.85	0.84224	0.59555	25.6	9.099	174.4							
D u	2.226	0.01333	0.00943	2.227	0.244	0.275							
D s	1.321	0.00925	0.00654	1.324	0.155	0.278							
F0	0.121	7.90E-04	5.59E-04	0.121	0.02	0.017							
Start Cou	4E+05	6193.98	4379.81	4E+05	47795	40741							

Description of experiment

Experimental Generational Counts

All best-fit solutions: best-fit population parameters and their corresponding sensitivity analysis ranges

Filtered best-fit solutions used for maximum-likelihood clustering

Average and standard deviation of parameters in cluster

Maximum likelihood parameter ranges of the cluster

Troubleshooting/Tips

There are several potential problems that can arise during the phenotyping process that are briefly discussed here along with possible solutions. Other considerations are addressed in the accompanying publication:

1. Its important to work with a single homogenous population of cells because phenotyping a mixture of cell populations can lead to differing best-fit parameters. Its also important to minimize experimental preparation errors caused by poor/inconsistent mixing and preparation of cells (see the accompanying publication for specific details).
2. Selecting the correct position of the undivided peak during histogram construction is essential for avoiding errors. Using more frequent time points, better staining technique, and unstimulated controls can help guide peak selection.
3. Choosing time points before the first division, just after the first division occurs, when cells are dividing frequently, and after most cells have died (usually 120-140 hours after stimulation) is important for constraining the parameters during fitting (see the paper for specific examples).
4. Using multiple replicates at each time point can help eliminate biological and experimental noise by fitting to all replicates simultaneously (you can build multiple CFSE histograms per time point by indicating the same time after stimulation during construction).
5. FlowMax can interpret volume measurements recorded by some flow cytometers (such as the Accuri C6) automatically. If sample volume measurements are not available, the cellular concentration can be entered manually for each CFSE log-fluorescence histogram by clicking on the cell concentration button and entering the correct value. 
6. Selecting a sufficiently large number of candidate solutions is important to ensure that all good solutions clusters are found. In some cases, 100 solutions may be sufficient to ensure the best solution cluster(s) are found, however, in some cases 1,000 solutions may be required. To test if sufficient solutions are being sampled, repeat the phenotyping and see if the same solutions are found consistently.
7. It may help to manually adjust some parameter ranges to help constrain the fit. For example, if the first division peak doesn't appear until between 24-40 hours, the E[$Tdiv_0$] parameter may be set to between 20 and 48 hours.
8. To avoid exponential-like division and death distributions set the maximum s.d.[Tdiv/Tdie] parameters to values less than or equal to those of the corresponding E[Tdiv/Tdie] parameters (e.g. if E[$Tdiv_0$] is set to between 24-72 hours, set the s.d.[$Tdiv_0$] parameter range to 1-72).
9. Some parameters are naturally less sensitive to the data and may be poorly constrained during fitting even with perfect data (see the paper), therefore, solutions should be represented by sampling within the maximum-likelihood parameter ranges.

License