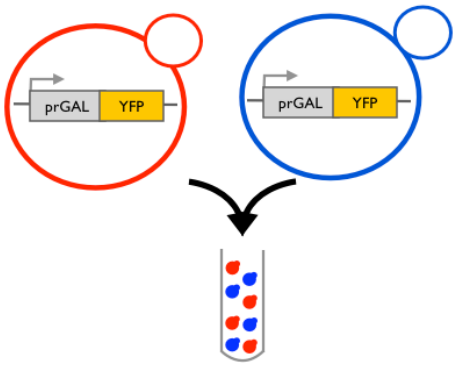
Comparing gene expression in different strains using flow cytometry

Background

We are interested in better understanding how cells respond to nutrients in their environment. One way to do this is to compare the response of related strains and, when the responses are different, to use genetics to identify the traits responsible for these differences.

In this example, we are comparing two wild type isolates of yeast from different parts of the world. We are interested in comparing how the different isolates induce a metabolic gene in response to galactose. In order to do so, we have cells that constitutively express either a red fluorescent protein (mCherry) or a blue fluorescent protein (BFP). In addition, each strain has a yellow fluorescent protein (YFP) promoter reporter for the first gene in the galactose catabolic pathway (see figure). To compare the response of each strain we inoculate at very low cell densities the mCherry and BFP strains in a flask. After 8 hours we take a sample of the culture to assay.



In Yeast Showdown I, we analyzed cells by microscopy. Microscopy is an excellent technique for measuring fluorescence quantitatively, for determining cellular localization, and for following individual cells over time; but the total number of cells that can be visualized is limited. Flow cytometry is a technique that can quickly measure the total fluorescence of a large number of cells one cell at a time. Flow cytometry has the disadvantage that the signal to noise is often lower than microscopy (think about why this is the case) and a single cell cannot be followed over time; instead one measures changes in a population of cells. Flow cytometry works by quickly passing a narrow stream of liquid containing the cells through a flow cell. Multiple lasers and filters are then able to detect a large number of fluorophores. Each distinct laser and filter combination is referred as a channel (eg GFP or mCherry). Additionally, other features such as the amount of scattered light, side scatter (SSC) and forward scatter (FSC), are detected. Both SSC and FSC give information on the size of the cells but both measurements are influenced by multiple variables besides size.

Exercises

Part 1. Load the data

The data from this experiment can be found in the associated zip file under the name “facs_data.mat”. **Try loading the data.**

The data is a matrix with raw data where every row is a cell measured in the assay and each column represents either a measured fluorescence for a specific channel (columns 2-4) or the forward scatter and side scatter (columns 5 and 6). Column 1 contains the time where the cell was measured.

Column 1	2	3	4	5	6
Time	YFP	BFP	mCherry	FSC (Forward Scatter)	SSC (Side scatter)

Before starting any of the exercises below, make a comprehensive list of all the steps that you will need

to quantitatively compare the gene expression in the two strains. Think about the types of visualizations and quantifications that you need to answer this question. **Don't just think about it write out this list!.** Now try to implement it. (Check out the hints or go on to Part 2 and Part 3 for more guidance).

Part 2. Visualizing distributions and relationships in the data

1. How would you visualize each channel (column) in the experiment? Data often needs to be mathematically transformed in order to be more clearly understood. In this case, we will log transform the data. **Don't forget to log transform the data.** Make a histogram of the log10 values of each channel in the experiment. What can you tell about the distributions of each channel? Are they unimodal, bimodal?

Keep in mind that column 1 contains time information so you can actually skip that histogram.

2. We use scatter plots to visualize relationships between variables. What can you tell between the relationships between channels? More specifically, what can you tell about the relationship between the BFP and the mCherry channel? What causes the slope in the BFP versus mCherry plot? How can you verify this?

3. How can use the mCherry and BFP values to segment between the red and blue cells? **Hint available.**

Hint: Look at our proposed work flow (under '**HINTS**' below) if you need extra help.

Part 3. Segmenting the strains and comparing their gene expression.

1. After picking the most appropriate threshold for the segmentation. Visualize your segmentation by plotting in different colors blue and red cells. How good is your segmentation? Would you include all the cells in this experiment or would you filter out some cells?

2. Now plot the YFP channel in each segmented strain. Are blue cells inducing the galactose metabolic genes (tagged in YFP) in the same manner? How similar or different are they? Is there anything special about the blue cells?

3. What are the different ways that you can use to quantify the YFP levels? Try some.

4. Try segmenting the BFP cells into on and off cells and compare all three groups (mCherry, BFP on and BFP off). Do the the on cells have different sizes compared to non inducing cells?

** Bonus**

There is another array in your zip file called "facs_data_Time1.mat" that contains the same type of data for the first time point in the experiment, that is, right after inoculation of the cells. Using the same scripts that you wrote in the previous parts, what can you learn about the gene expression levels right after inoculation?

Hints to exercises

Hint: Part 2

This is the proposed workflow for the analysis

Visualize	Compare histogram versus scatter plots of each channel.

	Observe distributions of each channel and compare relationships between channels. Make other scatter plots: fluorescence versus time. Check that the fluorescence is even throughout the run
Segment	Pick a threshold value to distinguish between strains. Compare using the mCherry or BFP channels normalized by size FSC or SSC. Use logical statements to extract rows above a certain threshold. Extract the YFP values for each strain.
Quantify	Look at the YFP distribution for each channel. Compare the shapes of the distributions. Extract statistics of the distributions: mean, medians, standard deviations and coefficients of variation. Test whether the distributions are statistically different.

Hint part 2, question 3

Plot BFP vs SSC, and BFP vs SSC. Do the same for mCherry. Can you better distinguish between the two strains? Make a histogram of BFP normalized by size, that is, BFP/SSC . Compare the histogram of BFP/ FSC.

What threshold value gives you the most accurate distinction between blue and red cells? Note that there is not a single response and multiple answers are correct.

Hint Part 3

For question 1, make scatter plots with the segmented populations using different colors.

Last updated 20120525 JW.