

I will demonstrate exploring image based data using CellProfiler Analyst in this short movie.

If you want to do the analysis yourself, you can download the example data and images that I'm using in this demo from the CellProfiler website, although keep in mind that to use CellProfiler Analyst, you're going to need to install not just CellProfiler Analyst itself, but also you'll need to install and configure a MySQL database server so that the data is accessible to CellProfiler Analyst.

When CellProfiler Analyst opens, you need to choose the properties file which tells the software where your database is located and various properties about your data.

You'll probably need the assistance of your local database administrator to get this file set up properly.

I'm going to start by displaying a scatterplot choosing two features that have been measured in the images, mean cell area vs. mean nuclear area, so all the individual cell measurements have been combined into a mean for each image.

In this data plot, each point is going to be representing a single image. I can also plot individual object data, let's say a histogram of DNA content.

That's the integrated intensity of the DNA channel within the nucleus and its customary to use the log scale for this kind of plot. It shows cell cycle distribution.

For many plot types, there are options you can configure. Here, I'm adjusting the number of bins in the histogram. In this plot, the two peaks are representative of the 2n DNA content on the left and 4n on the right. Remember this is individual cell data, so all cells in the entire experiment, no matter which image they come from, are combined in this one plot.

Now I'm plotting a parallel coordinance plot which allows plotting multiple features onto a single plot whereas scatterplots allow just two features and histograms can only display a single feature of images.

You'll see I'm choosing difference features and adding them to a list, each one of these features will become a different Y axis coordinate on the plot. Notice how the one blue point that I've selected in the original scatterplot is also going to appear blue in the parallel coordinates plot once I've created it. So I've got five coordinates here that are appearing in the parallel coordinates plot and you can see that the one blue line is representing the one blue point in the scatterplot.

I'm going to make one more scatterplot to show you how brushing in CellProfiler Analyst works. This scatterplot will again have two features plotted for image data just like the first plot we've made. We can explore relationships between features using this brushing capability.

You've seen a glimpse of it already in that the one blue point in the scatterplot appears as a blue in the parallel coordinates plot, but here you'll see it become much more powerful. I'm plotting here the ObjectCount on the Y axis and the number of neighbors that each nucleus has on the X axis.

In the parallel coordinates plot, I'm going to show the identify in the axes by choosing show info and show axes. you can see the five Y axes that are shown in the plot.

Now to brush data, I choose select and select points highlight gated up in the scatterplot. Those points now appear blue in the other plot so I can see the samples with high cell area and nuclear cell area have low object count in the parallel coordinance plot. Note also that even in the histogram which shows individual cell data, the data is linked. In blue are those cells from images that I've highlighted in the scatterplot, so the scatterplot shows whole image data and the histogram shows individual cell data.

I can further explore relationships between features by clicking the arrows in the axes to look at how other measured features look for the highlighted population and for the image population overall, and now I'm going to highlight samples with high object counts, so high cell counts, and then I can see in the other scatterplot that those images with higher cell counts tend to have smaller cell and nuclear area. that's what we would expect when cells are very crowded in the images, and to confirm that I can look at a random image in that highlighted set and it seemed to have fairly crowded samples, and I can look at an individual image with a more normal cell count and confirm that there are fewer cells and that they are larger.

Now this blue data point over here has high nuclear area and cell area. I can confirm it by looking at the raw image, and if I had processed images without lines available, I could take a look at those.

To explore a subset of data points and get an idea of what samples are there, I can highlight them and choose select points and get gene names and data which produces a short list of just the samples in that highlighted subset of data points and this experiment, I can see there are several replicates of a small number of interesting rnai hairpins that produce the data pints of interest. I can find replicates of individual data point and have them highlighted, I can also ask for web-information about a particular sample.

I can look at the raw data of those four highlighted data points, those are four replicates of one particular type of sample, and I can take a look at the numerical measurements that are made, and I can also save the data if I would like to, for just that one particular data point.

And lastly, I can make a sub-plot of object data for an individual image or set of images. Here, for example. I want to see a histogram of the DNA content for just those 4 highlighted samples, so I'm gonna highlight it on a separate plot. You'll not that this cell cycle histogram is not visible in the large histogram simply because I've only got four

images- - there're aren't many cells. If I zoomed in on that plot, I would actually see the same data.

This concludes the basics of exploring image data in CPA.