SpotAssayQuant Tutorial

Sarah Hawbaker

Contents

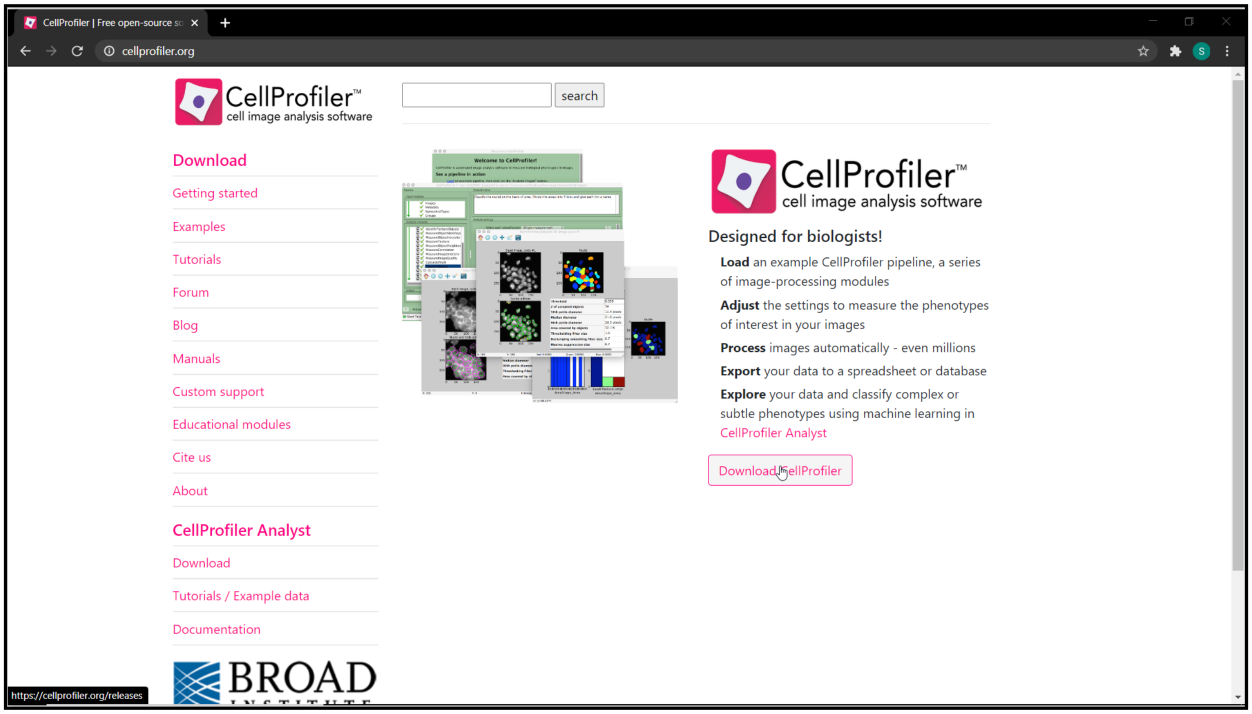
Overview

This pipeline batch processes spot assay image sets and outputs all result data to .csv files. In addition, it outputs a series of supplementary data and result images for spot checking. You may find that most of this data is unecessary; it was made with a particularly problematic and inconsistent dataset in mind. I have chosen to keep the auxillary data in for robust testing and verification. See the Input and Output section for more information.

Installation and Setup

Installing Cellprofiler

CellProfiler is available here: https://cellprofiler.org/. Click **Downloads » CellProfiler**. Make sure to download the basic version and not CellProfiler Analyst.



Designed for biologists!

Load an example CellProfiler pipeline, a series of image-processing modules

Adjust the settings to measure the phenotypes of interest in your images

Process images automatically - even millions

Export your data to a spreadsheet or database

Explore your data and classify complex or subtle phenotypes using machine learning in CellProfiler Analyst

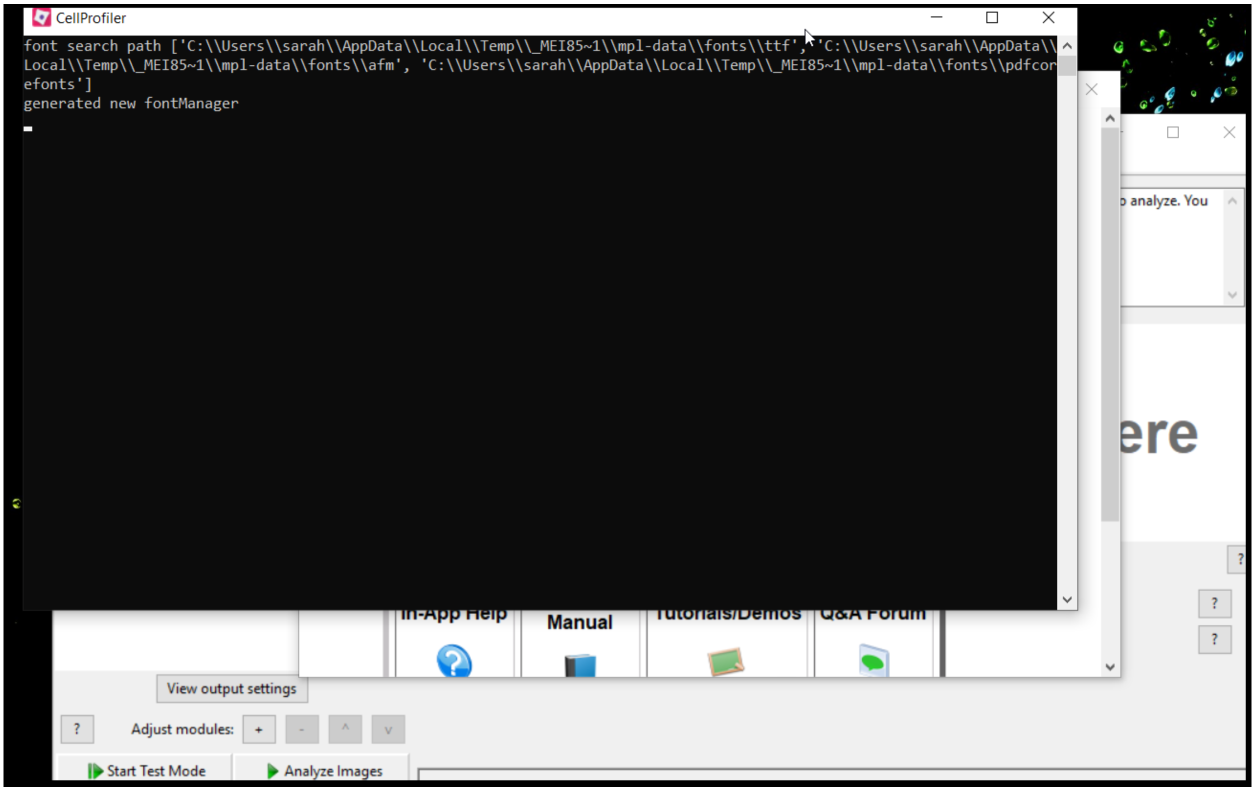
Choosing the Pipeline

Pipelines are located in the `Pipelines` folder. There are two pipelines available, depending on the format of your input images. For grayscale images, use the GrayScaleInput pipeline. However, you may find that you want more control over the thresholding values and method. In that case, there is a BinaryInput pipeline available. Note that it is not difficult to modify the `Threshold` module in the GrayScaleInput pipeline, I have just created these two options for ease-of-use.

For sample thresholding macros in ImageJ, see the `FIJI Macros` folder.

Pipeline Setup

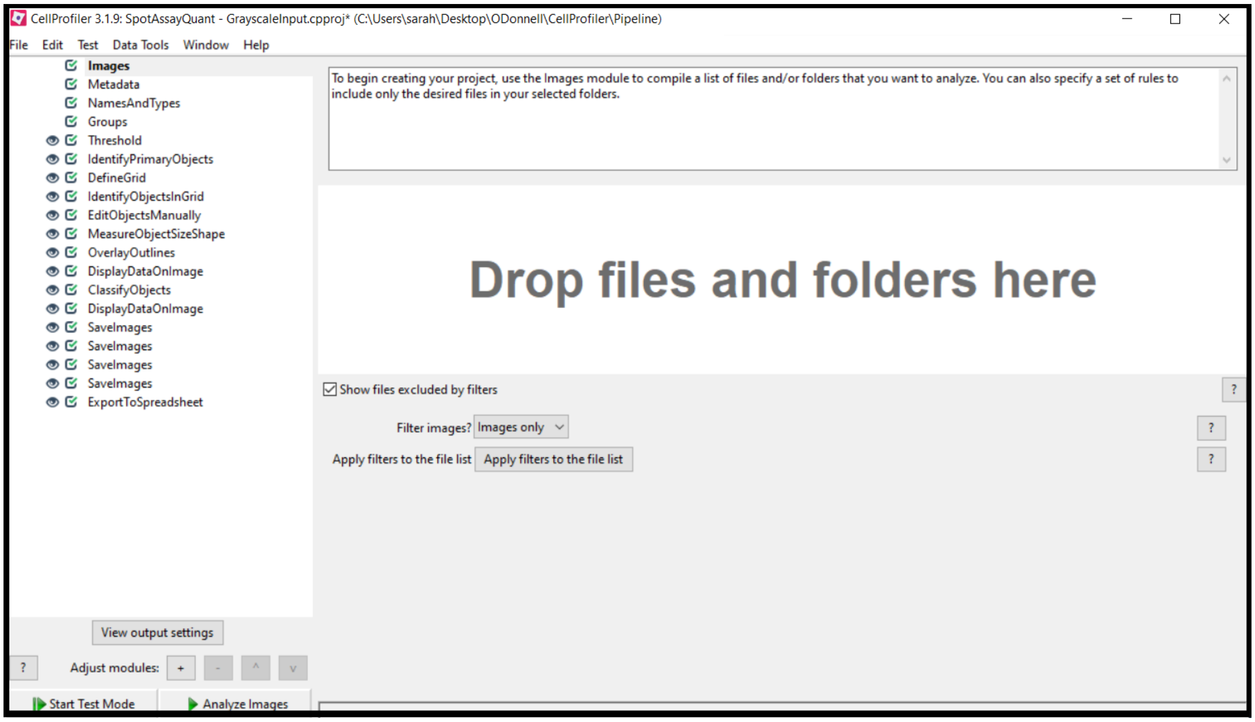
1. **Open CellProfiler**. Upon opening CellProfiler, you will see a black terminal window pop up and run a few lines of code. This is just the .exe file launching. Do not close this window while you are running the program, or it will exit. I think it's best just to minimize it to avoid accidental termination.



```
font search path ['C:\\Users\\sarah\\AppData\\Local\\Temp\\_MEI85~1\\mpl-data\\fonts\\ttf', 'C:\\Users\\sarah\\AppData\\
Local\\Temp\\_MEI85~1\\mpl-data\\fonts\\afm', 'C:\\Users\\sarah\\AppData\\Local\\Temp\\_MEI85~1\\mpl-data\\fonts\\pdfcor
efonts']
generated new fontManager
```

Close the introduction window.

2. **Load the pipeline**. Download your chosen pipeline from the `Pipelines` folder into a local directory. Load into CellProfiler via **File » Import » Pipeline from File**. Your screen should look like this:



CellProfiler 3.1.9: SpotAssayQuant - GrayscaleInput.cpproj* (C:\Users\sarah\Desktop\ODonnell\CellProfiler\Pipeline)

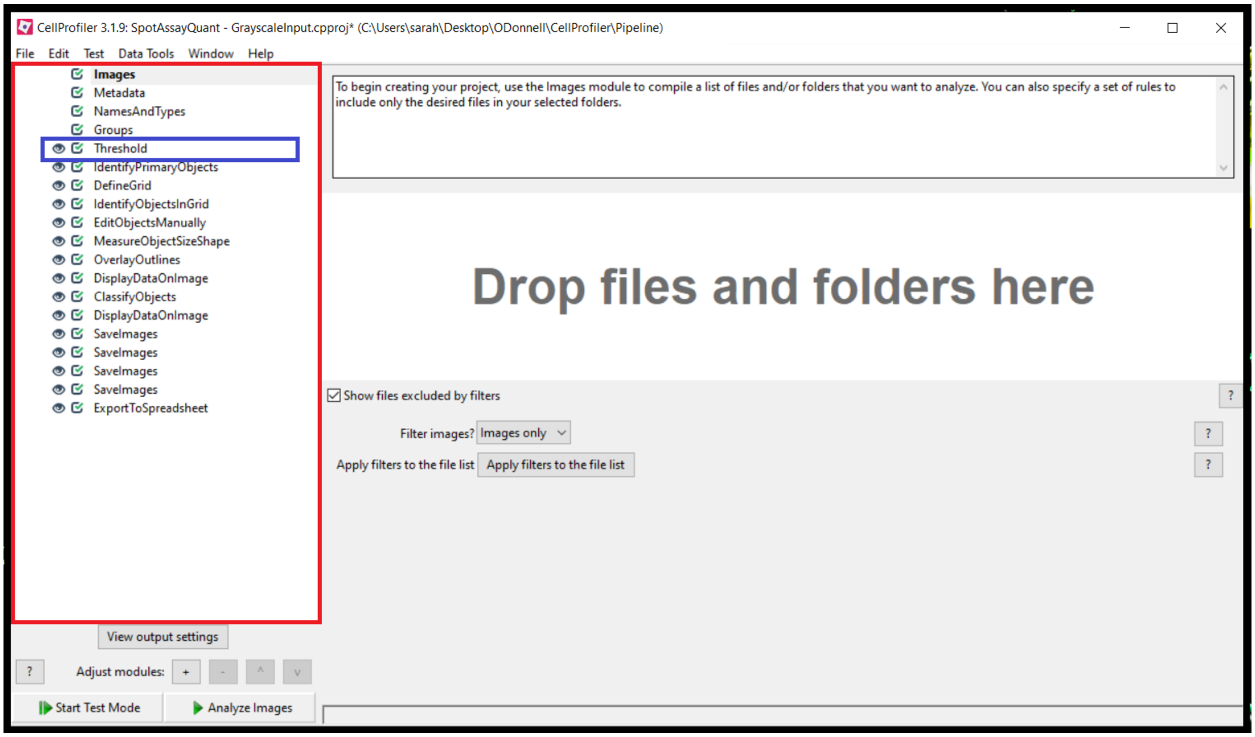
To begin creating your project, use the Images module to compile a list of files and/or folders that you want to analyze. You can also specify a set of rules to include only the desired files in your selected folders.

☑ Show files excluded by filters

Filter images? Images only

Basic Navigation

The pipeline consists of a series of modules which each apply a different function to the image. The module list is outlined in red below, and a single module in blue.



CellProfiler 3.1.9: SpotAssayQuant - GrayscaleInput.cpproj* (C:\Users\sarah\Desktop\ODonnell\CellProfiler\Pipeline)

Images
Metadata
NamesAndTypes
Groups
Threshold
IdentifyPrimaryObjects
DefineGrid
IdentifyObjectsInGrid
EditObjectsManually
MeasureObjectSizeShape
OverlayOutlines
DisplayDataOnImage
ClassifyObjects
DisplayDataOnImage
SaveImages
SaveImages
SaveImages
SaveImages
ExportToSpreadsheet

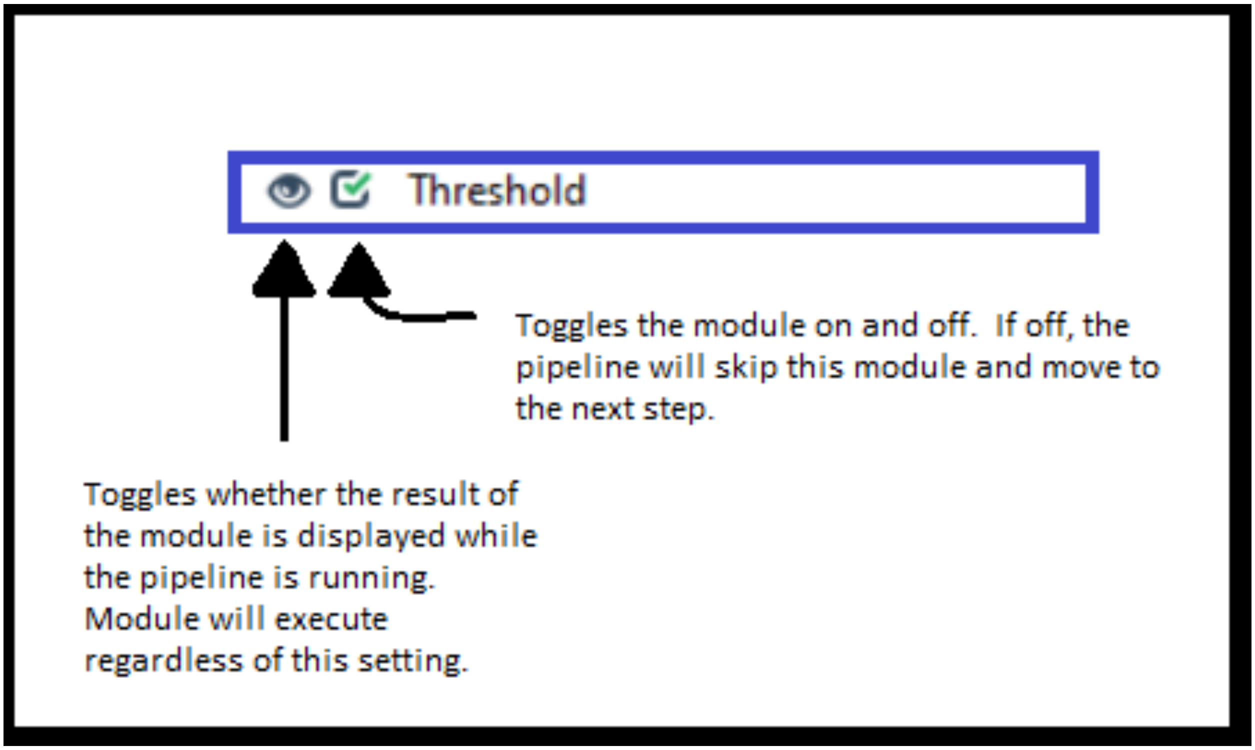
To begin creating your project, use the Images module to compile a list of files and/or folders that you want to analyze. You can also specify a set of rules to include only the desired files in your selected folders.

Drop files and folders here

☑ Show files excluded by filters

Filter images? Images only

Apply filters to the file list



Toggles the module on and off. If off, the pipeline will skip this module and move to the next step.

Toggles whether the result of the module is displayed while the pipeline is running. Module will execute regardless of this setting.

Input and Output

Before we run the pipeline, we must make sure the output all goes to the correct place. The output is currently structured as follows:

- Spreadseets: This is where all the raw data is stored.
- Outlines: This is a visual representation of the detected spots on each image.
- Object Areas: A visual respresentation of measured data drawn on the spots themselves. Currently this is set to "Area" but it can be changed in the `DisplayDataOnImage` module.
- Object Numbers: Similar to the Object Areas folder, but with the individual spot numbers displayed on each spot. This is useful to verify that the rows are being grouped correctly.
- Rows: A visualization of the row groupings.
- Graphs: A place to store any user-generated graphs post-processing. See the Post Processing section for more information.

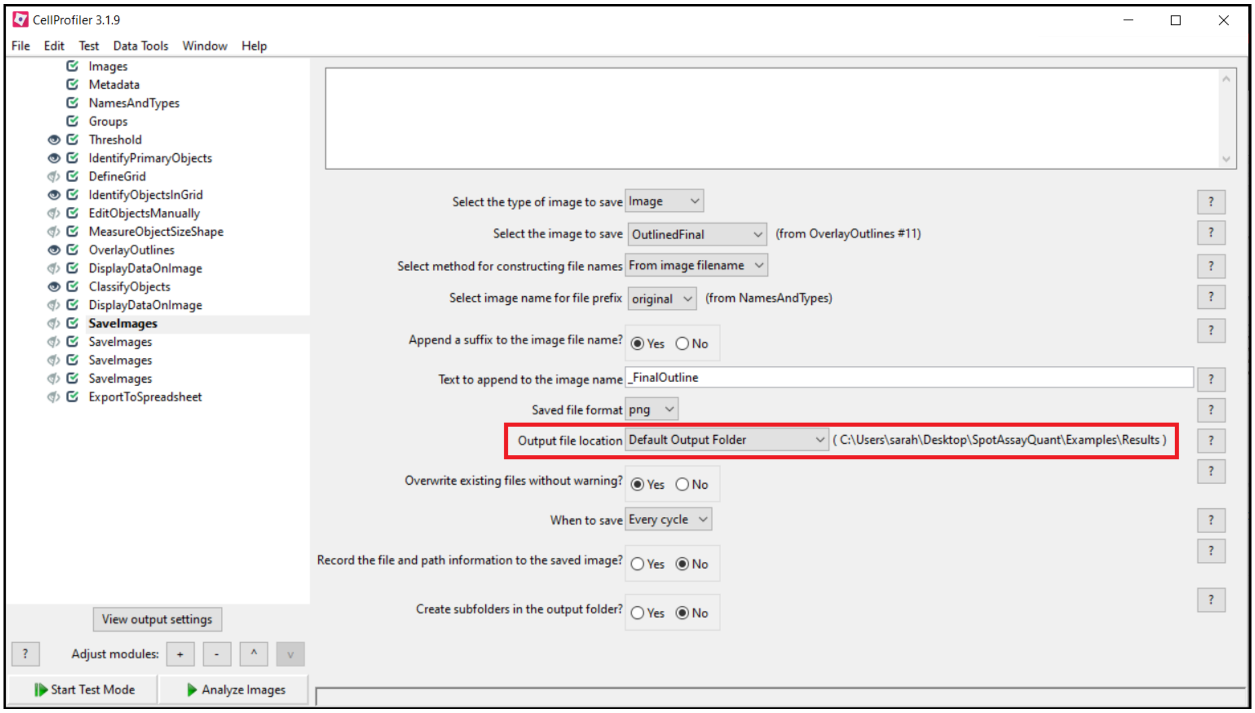
Should you not want this extra output, you can simply uncheck its associated module. You do not need to worry about setting the output folder if the module is unchecked.

Output	Associated Module	File Format
Spreadsheets	ExportToSpreadsheet	.csv
Outlines	SaveImages 1	.png
Object Areas	SaveImages 2	.png
Object Numbers	SaveImages 3	.png
Rows	SaveImages 4	.png
Graphs	N/A	N/A

Setting up the Output Folder:

1. Click **View output settings**, which is located beneath the module list. Set the **Default Output Folder** to your desired folder. You do not need to change the **Output file format** setting. If you want all your data to go into one folder, unsorted, you can skip the next step. I don't recommend this as with large imagesets it will be messy to sort things out afterwards.
2. For each of the associated modules, find the **Output File Location** setting. Set to your desired output folder. I like to create all empty folders ahead of time inside my main output folder. There is an example output folder in the `Examples` folder for reference.

If you structure your output this way, select **Default Output Folder sub-folder** from the drop-down menu and choose your folder from there. Do this for each **SaveImages** module and the **ExportToSpreadsheet** module.



Select the type of image to save: Image

Select the image to save: OutlinedFinal (from OverlayOutlines #11)

Select method for constructing file names: From image filename

Select image name for file prefix: original (from NamesAndTypes)

Append a suffix to the image file name? ● Yes ○ No

Text to append to the image name: _FinalOutline

Saved file format: png

Output file location: Default Output Folder (C:\Users\sarah\Desktop\SpotAssayQuant\Examples\Results)

Overwrite existing files without warning? ● Yes ○ No

When to save: Every cycle

Record the file and path information to the saved image? ○ Yes ● No

Create subfolders in the output folder? ○ Yes ● No

Adding Input Files

Click on the **Images** module. Drag and drop your files or folders into the pipeline. To clear all files from the pipeline, go to **Edit » Clear File List**.

Defining the Grid

Before you run the pipline you will need to set up the “Grid” that determines how many rows and columns of spots you want to quantify. Set this by clicking the **Define Grid** module and entering your values in the **Number of rows** and **Number of columns** fields. The default settings are 7 and 4, respectively.

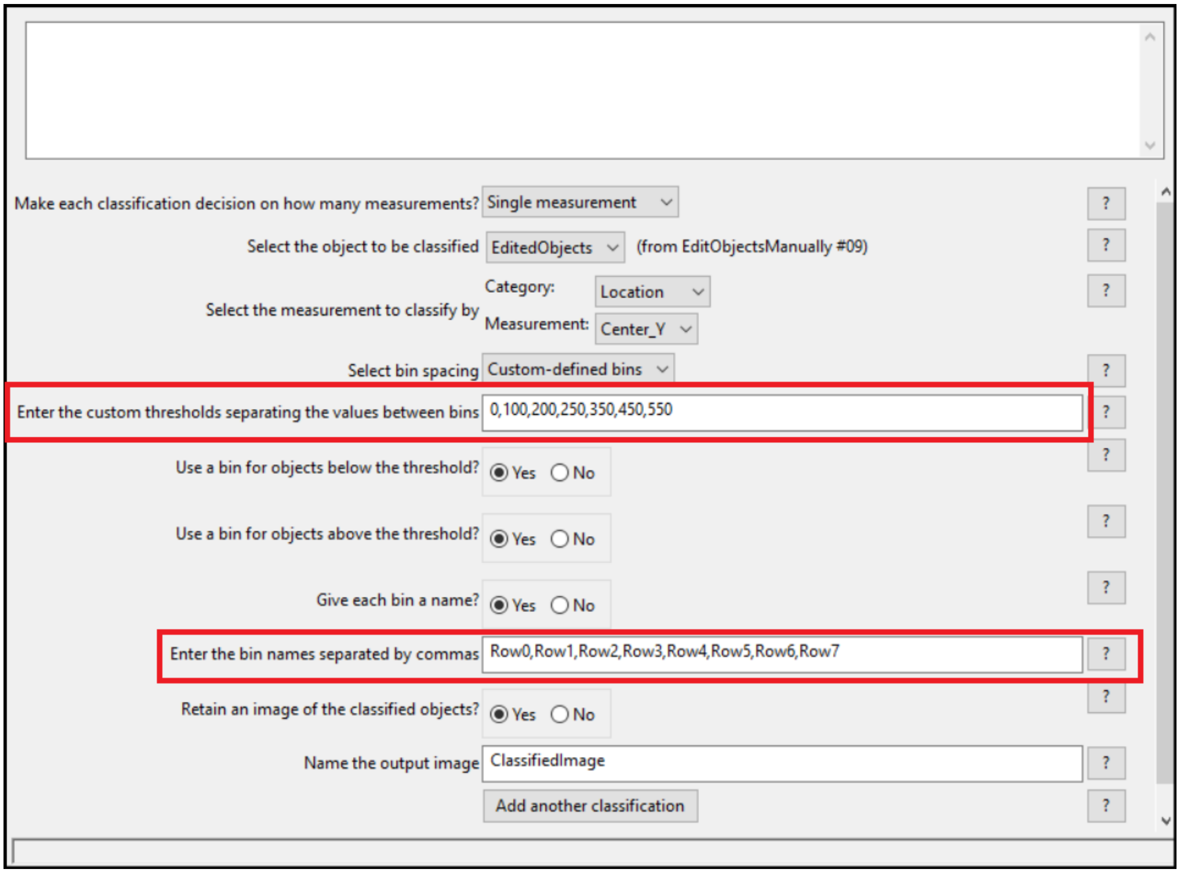
If you are sure that the image size and spot placements will be identical for each image, you can change the **Define a grid for which cycle?** setting to **Once**. Otherwise it should stay at **Each Cycle**.

Defining the Rows

CellProfiler allows you to create classifers that each spot can belong to. For our purposes, we need the image to be grouped by rows. Therefore each spot is assigned to a row depending on its y-coordinate value. This allows us to calculate the total area for each row after the pipeline has finished running.

Defining the Bins

Click on the **Classify Objects** module. Make sure all of your settings match the image below.



Make each classification decision on how many measurements? Single measurement

Select the object to be classified EditedObjects (from EditObjectsManually #09)

Select the measurement to classify by Category: Location Measurement: Center_Y

Select bin spacing Custom-defined bins

Enter the custom thresholds separating the values between bins 0,100,200,250,350,450,550

Use a bin for objects below the threshold? ● Yes ○ No

Use a bin for objects above the threshold? ● Yes ○ No

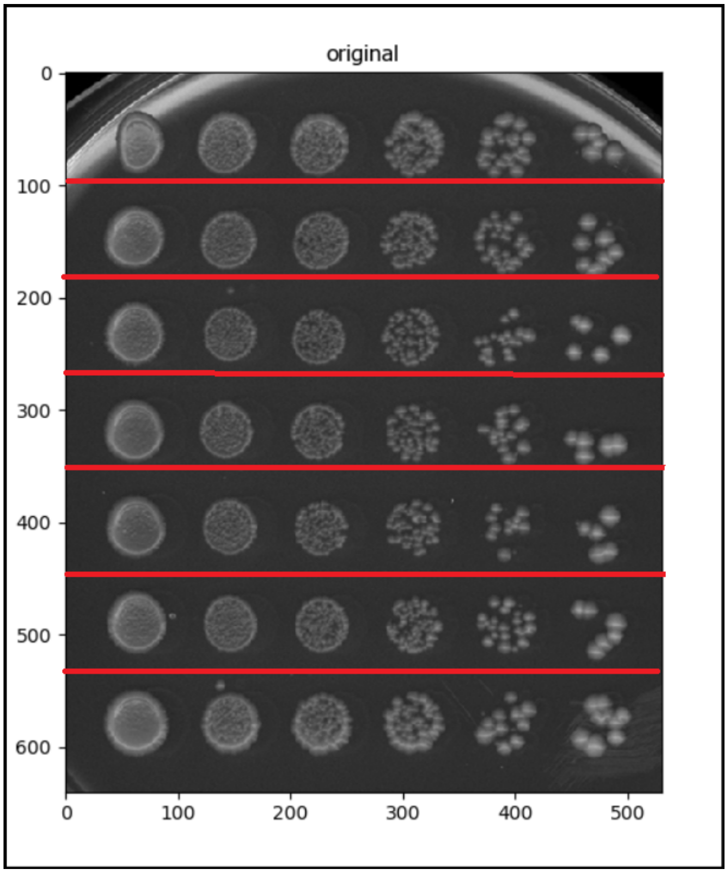
Give each bin a name? ● Yes ○ No

Enter the bin names separated by commas Row0,Row1,Row2,Row3,Row4,Row5,Row6,Row7

Retain an image of the classified objects? ● Yes ○ No

Name the output image ClassifiedImage

You will need change two fields. First is the **Custom Threshold Values** field. These are the values that define the bins. Note that these values are for the **Center_Y** position, meaning that the center of the spot, not necessarily the whole spot, needs to fall inside the bin. In the example below the values are set to `0,100,200,250,350,450,500` to define 7 rows.

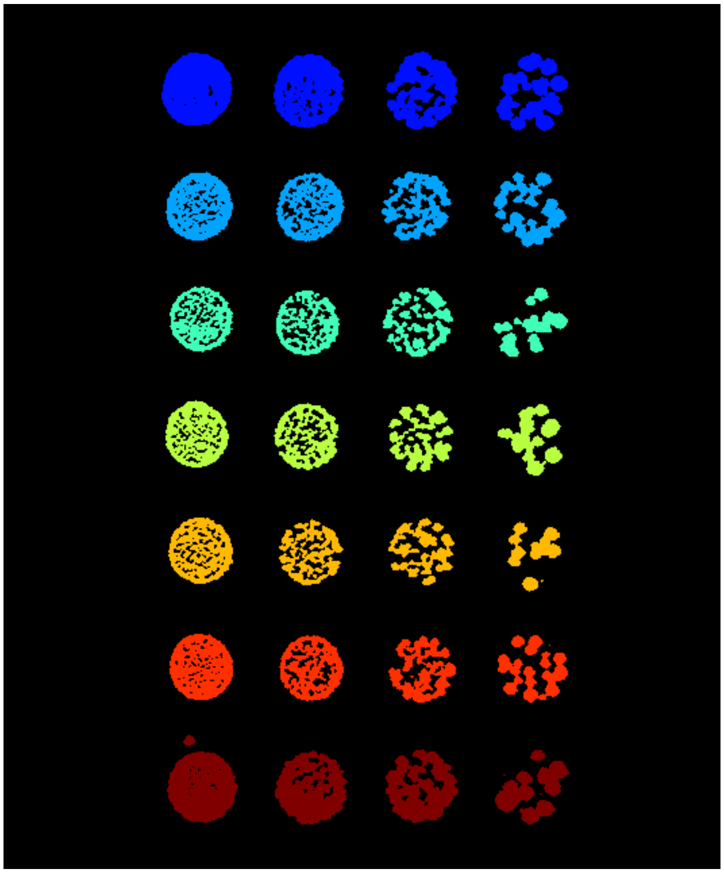


Note that you do not need to set the upper bound, but you do need to include 0.

Naming the Bins

For however many rows you have, you will need that many names + 1. I encourage you to stick to the naming convention shown in the example, where for 7 rows you have **Row1, Row2,. . . ,Row7** as well a **Row0** that CellProfiler requires to put anything it is unable to classify. These define the column headers for the .csv output file.

You will be able to see a visual representation of your row groupings in the `Rows` folder. It's a good idea to spot check a few of these to make sure your row sums will be correct.



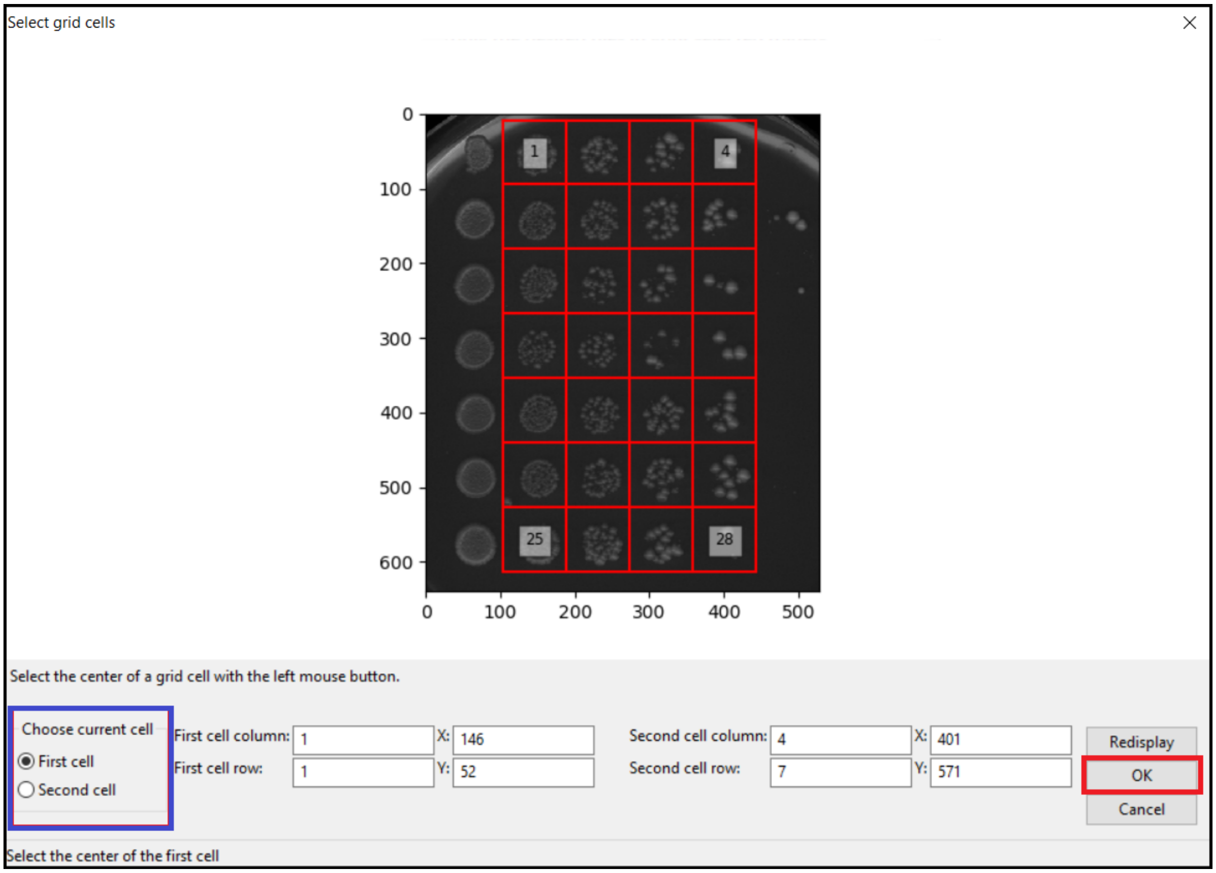
Running the Pipeline

Click **Analyze Images**. The pipeline will start. Typically it will take a moment or two to start up, but after that it will run relatively fast. As it begins to process images, you will see windows pop up depending on which modules are set to visible.

Define Grid

This is the first of two points in the pipeline where manual intervention is required. If you set the **Define a grid for which cycle?** setting to **Once** you will only need to do this step once. Otherwise it will have to be done for each image.

Click to set the coordinates for the top-left spot and the bottom-right spot. Keep setting until you feel comfortable that the grid is nearly correct. We can edit later, so it does not need to be perfect.



Select the center of a grid cell with the left mouse button.

Choose current cell

● First cell

○ Second cell

First cell column: 1 X: 146

First cell row: 1 Y: 52

Second cell column: 4 X: 401

Second cell row: 7 Y: 571

Select the center of the first cell

Edit Objects

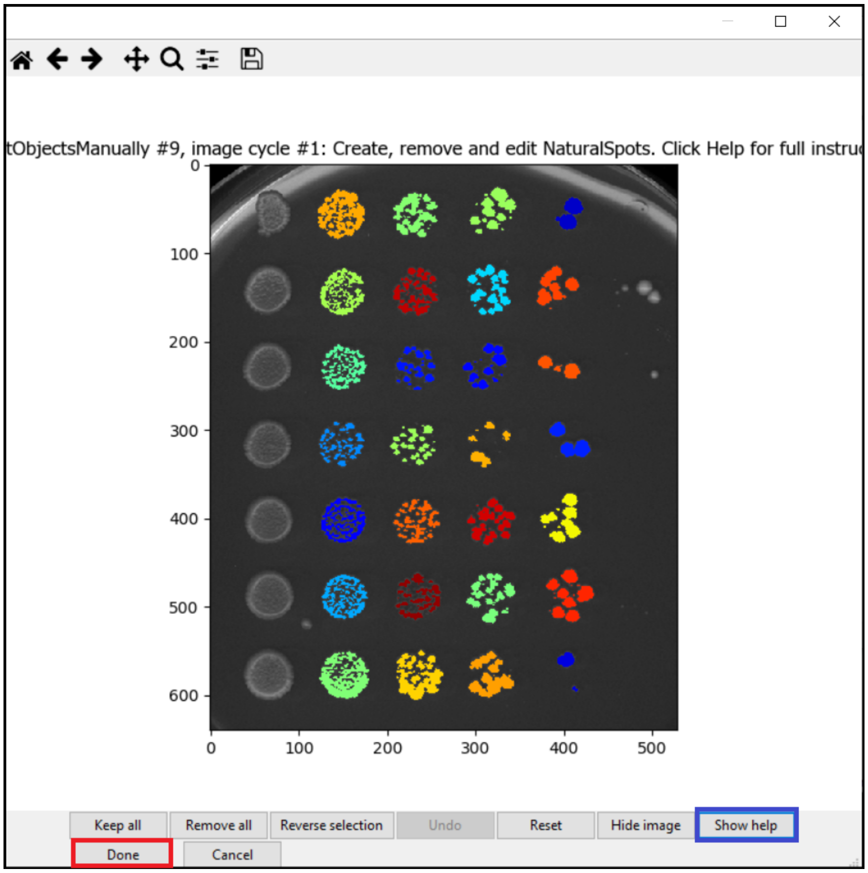
This is the second manual intervention. This module allows you to see the detected objects and edit them as needed. Most images should not need any editing at all. If you find that more than 10% need manual editing then either the settings are incorrect or this pipeline is not compatible with your dataset.

The edit module has a "Show help" button which has detailed instructions for how to use it. I will give an abridged version here, but I recommend to use the help module at first.

Note that you cannot edit if the `pan` or `zoom` button is selected. Always make sure to deselect after using.

The main commands you will need are:

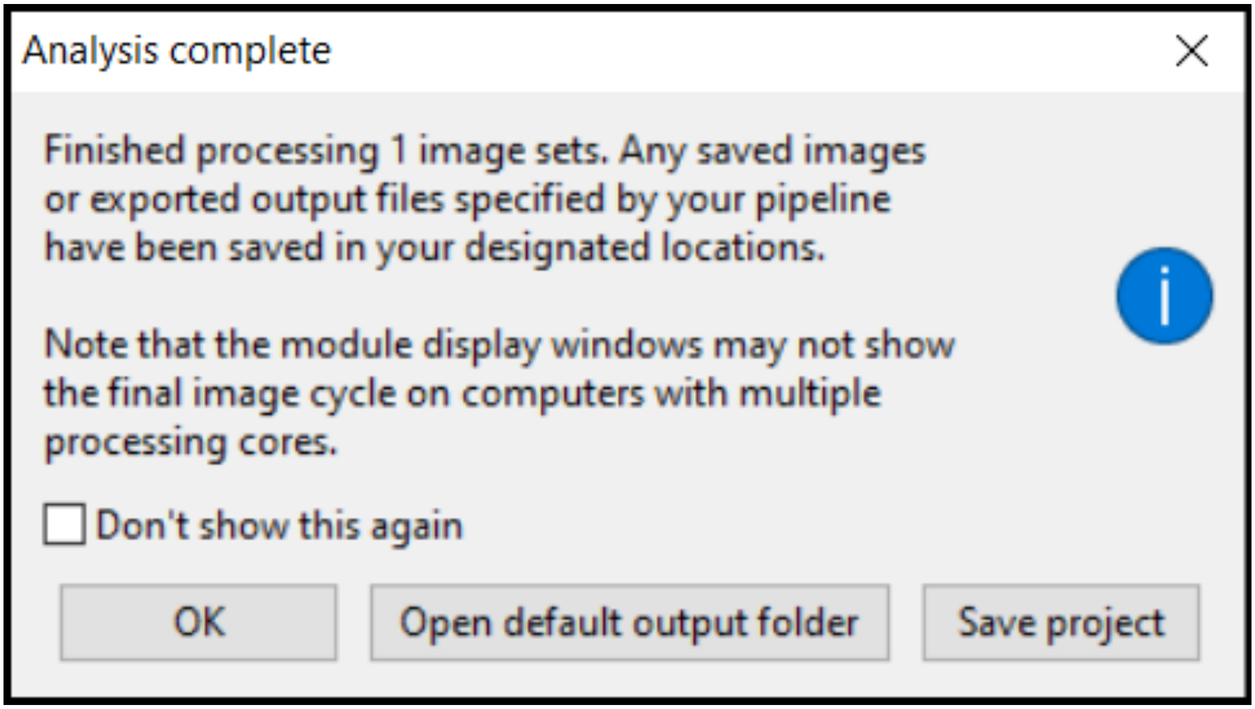
- Toggle. Press **T** to toggle the display between full colors and outlines. This will make the objects easier to see and edit.
- Edit mode: **Right click** an object to go into edit mode. This will transform the object outline into a series of control points that you can then edit with the commands below.
- Delete: Press **X** to delete a series of points by selection, **D** to delete a single point after clicking it.
- Add: Press **A** to add a control point at the location of your mouse after a single click.
- Join: Press **J** to join two areas of control points. Even if not contiguous the areas will be measured and counted as one object.
- Split: Press **S** to split an object into two. After pressing **S** you will click the two points that will form the dividing line.



tObjectsManually #9, image cycle #1: Create, remove and edit NaturalSpots. Click Help for full instru

Post Processing

When the pipeline is completed you will see this pop-up:



Analysis complete

Finished processing 1 image sets. Any saved images or exported output files specified by your pipeline have been saved in your designated locations.

Note that the module display windows may not show the final image cycle on computers with multiple processing cores.

□ Don't show this again

Go to the **Default Output Folder** to view your results.

The data is summarized and compiled using a Python script. If you are familiar with Python the script is available as a Jupyter Notebook and can be found in the `Scripts` folder. Otherwise, I am working on hosting the tool on the lab website for easy use. Until then, email me your “Spreadsheets” folder and I’ll run the script for you.