



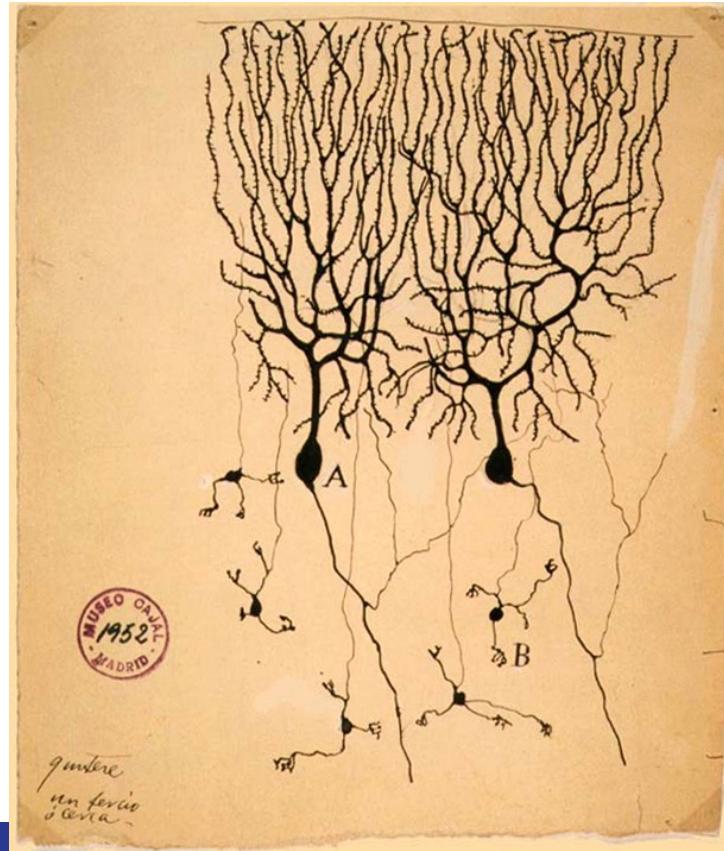
Getting the most of your microscopy data with high-content image analysis

Beth Cimini, PhD

Senior Group Leader

Imaging Platform, Broad Institute, Cambridge MA USA

Image analysis used to be much less technical



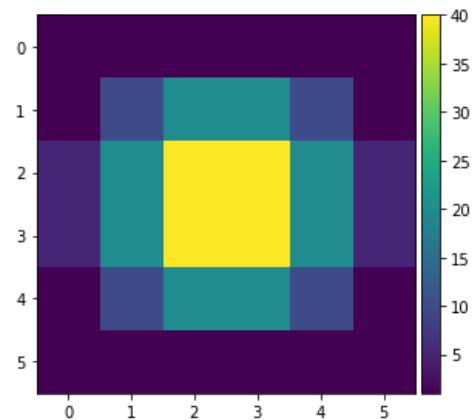
Drawing of Purkinje cells (A) and granule cells (B) from pigeon cerebellum by Santiago Ramón y Cajal, 1899. Instituto Santiago Ramón y Cajal, Madrid, Spain

Digital images are ultimately just arrays

```
skimage.io.imshow(a)
```

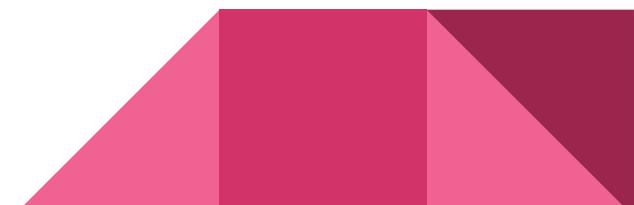
```
/usr/local/lib/python3.8/site-packages/skimage/_show.py:100: UserWarning: Image has negative values; displaying image with stretched contrast.  
    lo, hi, cmap = _get_display_range(image)
```

```
<matplotlib.image.AxesImage at 0x125202070>
```



a

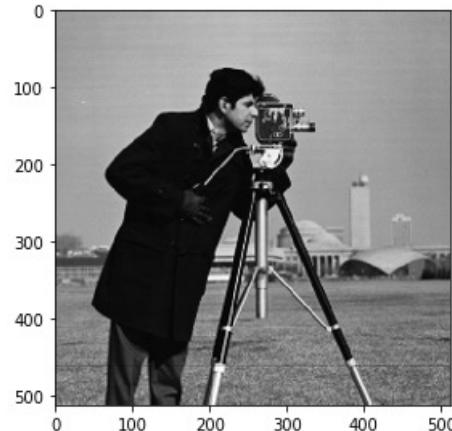
```
array([[ 1,  1,  1,  1,  1,  1],  
       [ 1, 10, 20, 20, 10,  1],  
       [ 5, 20, 40, 40, 20,  5],  
       [ 5, 20, 40, 40, 20,  5],  
       [ 1, 10, 20, 20, 10,  1],  
       [ 1,  1,  1,  1,  1,  1]])
```



Digital images are ultimately just arrays

```
skimage.io.imshow(b)
```

```
<matplotlib.image.AxesImage at 0x12552cca0>
```



```
b
```

```
array([[156, 157, 160, ..., 152, 152, 152],  
       [156, 157, 159, ..., 152, 152, 152],  
       [158, 157, 156, ..., 152, 152, 152],  
       ...,  
       [121, 123, 126, ..., 121, 113, 111],  
       [121, 123, 126, ..., 121, 113, 111],  
       [121, 123, 126, ..., 121, 113, 111]], dtype=uint8)
```

Human brains are not built for quantification



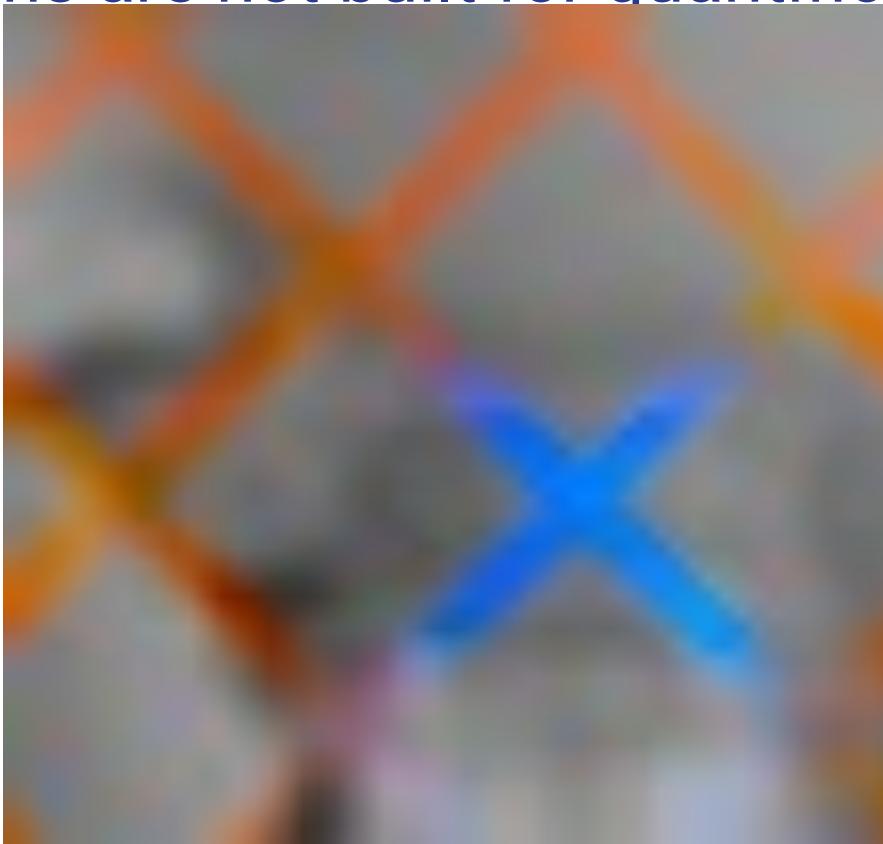
Photo from LGM by Manuel Schmalsteig CC-BY-2.0, Illusory Color Remix by Øyvind Kolås - <https://pippin.gimp.org/>

Human brains are not built for quantification



Øyvind Kolås

Human brains are not built for quantification



Øyvind Kolås

Different ways of thinking about images

1 Measure/score known phenotypes

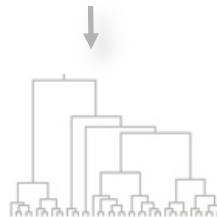
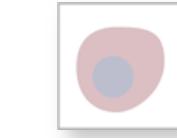


Jake Wintermute
@SynBio1

A brief history of molecular biology in the 90s



2 Profile to characterize samples





CellProfiler™

cell image analysis software

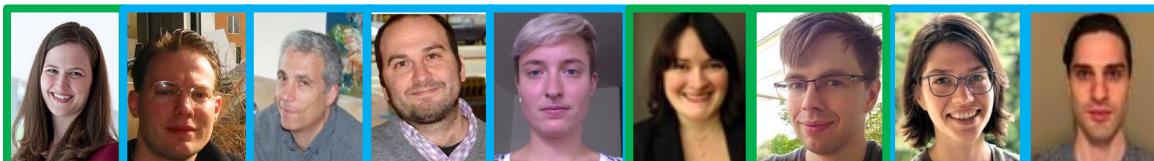
Free and open-source; Windows, Mac, Linux

Cited in 1,500+ papers per year

Used in 7/10 top pharma companies

In the Top 10 most popular papers in Genome Biology

Ranked most flexible and usable in independent analysis (Wiesmann et al.)



Anne
Carpenter

Ray
Jones

Lee
Kamentsky

Allen
Goodman

Claire
McQuin

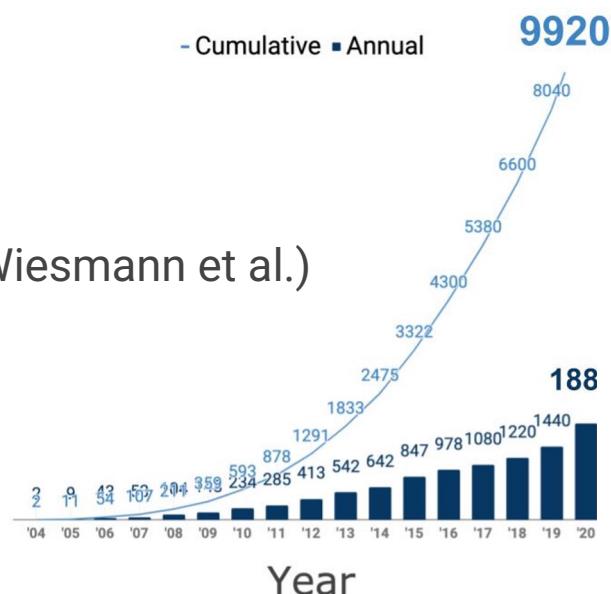
Beth
Cimini

David
Stirling

Alice
Lucas

Nodar
Gogoberidze

Publications citing CellProfiler



Software overview



CellProfilerTM
cell image analysis software

Image analysis &
quantification



CellProfiler AnalystTM
data exploration software

Image-centric
data analysis &
machine learning

Software overview



CellProfilerTM
cell image analysis software

**Measure
everything**

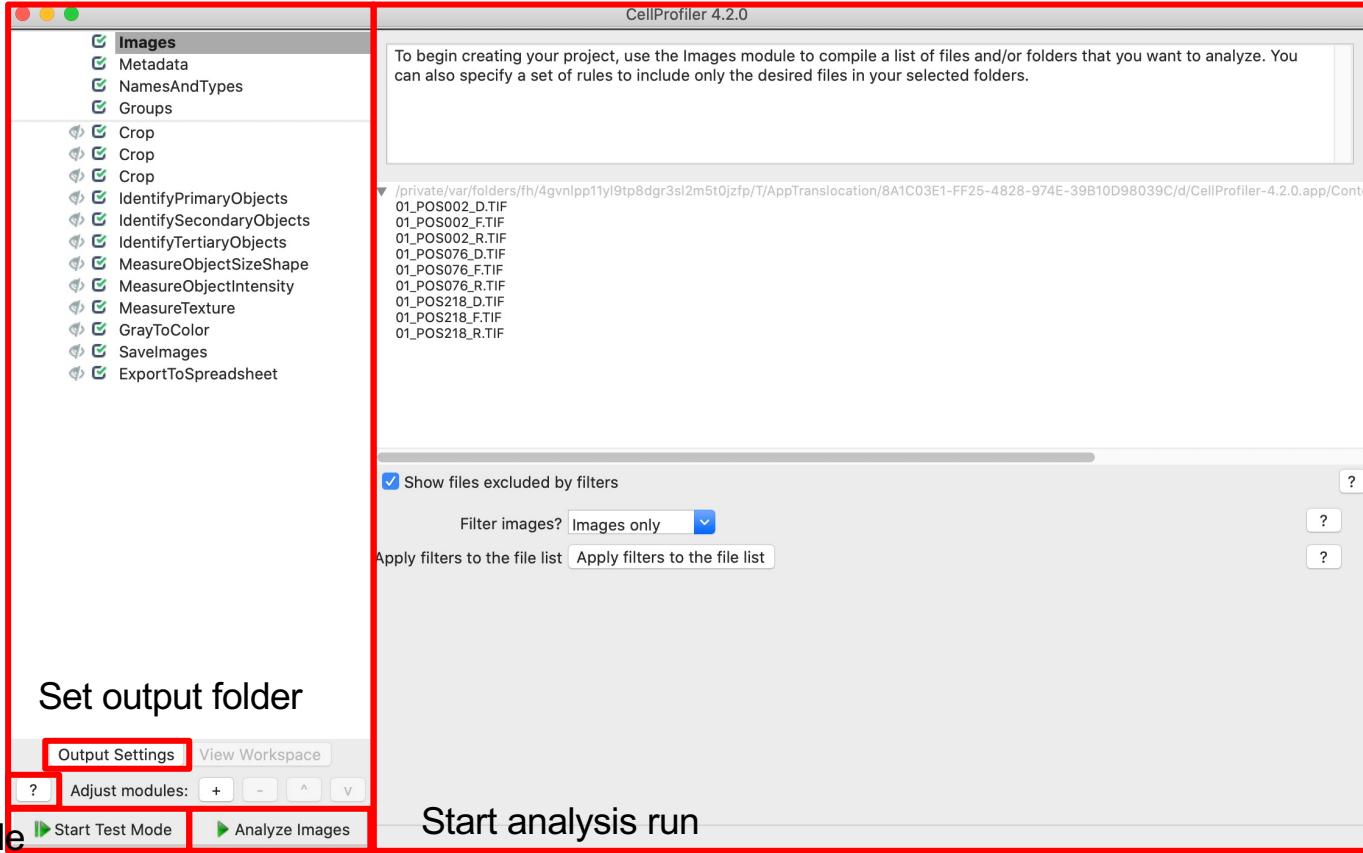


CellProfiler AnalystTM
data exploration software

**Ask question
later**

The CellProfiler interface

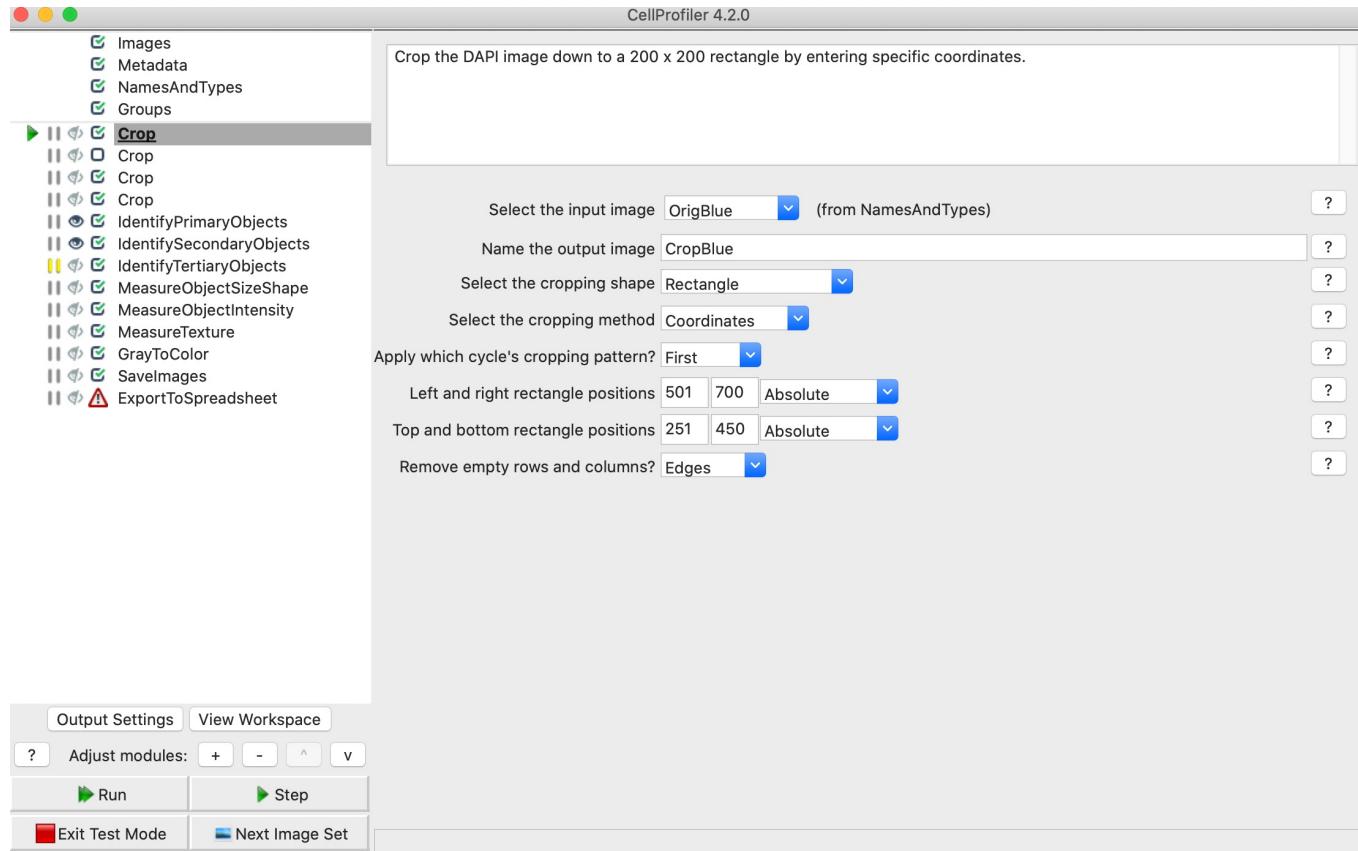
Pipeline panel



Settings panel

Module help
Start test mode

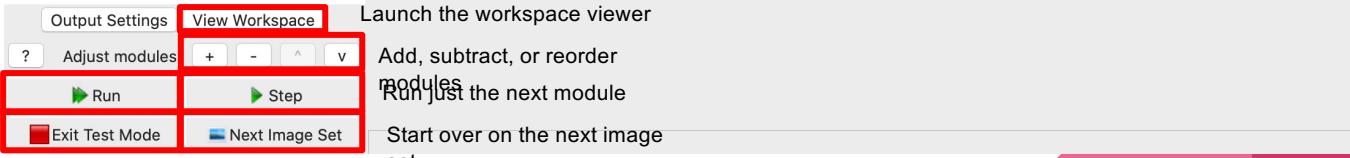
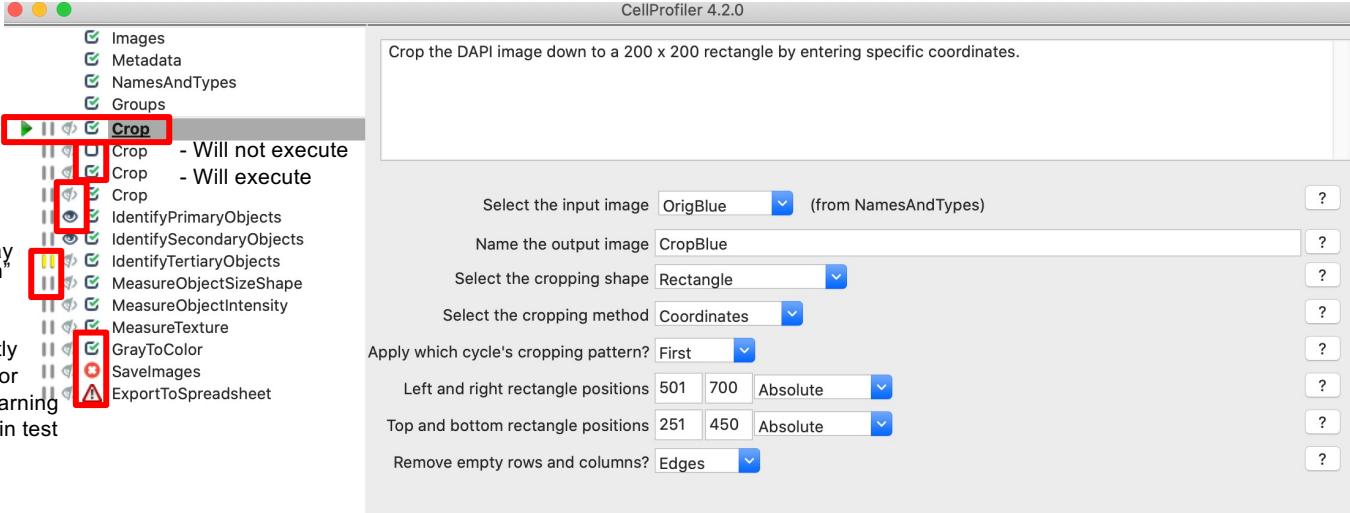
The CellProfiler interface



The CellProfiler interface

The next module to run

- Won't show display
- Will show display
- Will pause during "Run"
- Won't pause during "Run"
- Module set correctly
- Module has an error
- Module giving a warning (such as "won't run in test mode")



The CellProfiler interface

Set what feeds into and out from every module

CellProfiler 4.2.0

Identify the nuclei from the DAPI image. Three-class thresholding performs better than the default two-class thresholding in this case.

Use advanced settings? Yes No

the input image: CropBlue (from Crop #05)

to be identified: Nuclei

units (Min,Max): 10 40

diameter range? Yes No

Discard objects touching the border of the image? Yes No

Threshold strategy: Global

Thresholding method: Minimum Cross-Entropy

Threshold smoothing scale: 1.3488

Threshold correction factor: 1

Lower and upper bounds on threshold: 0 1

Log transform before thresholding? Yes No

Method to distinguish clumped objects: Shape

Method to draw dividing lines between clumped objects: Shape

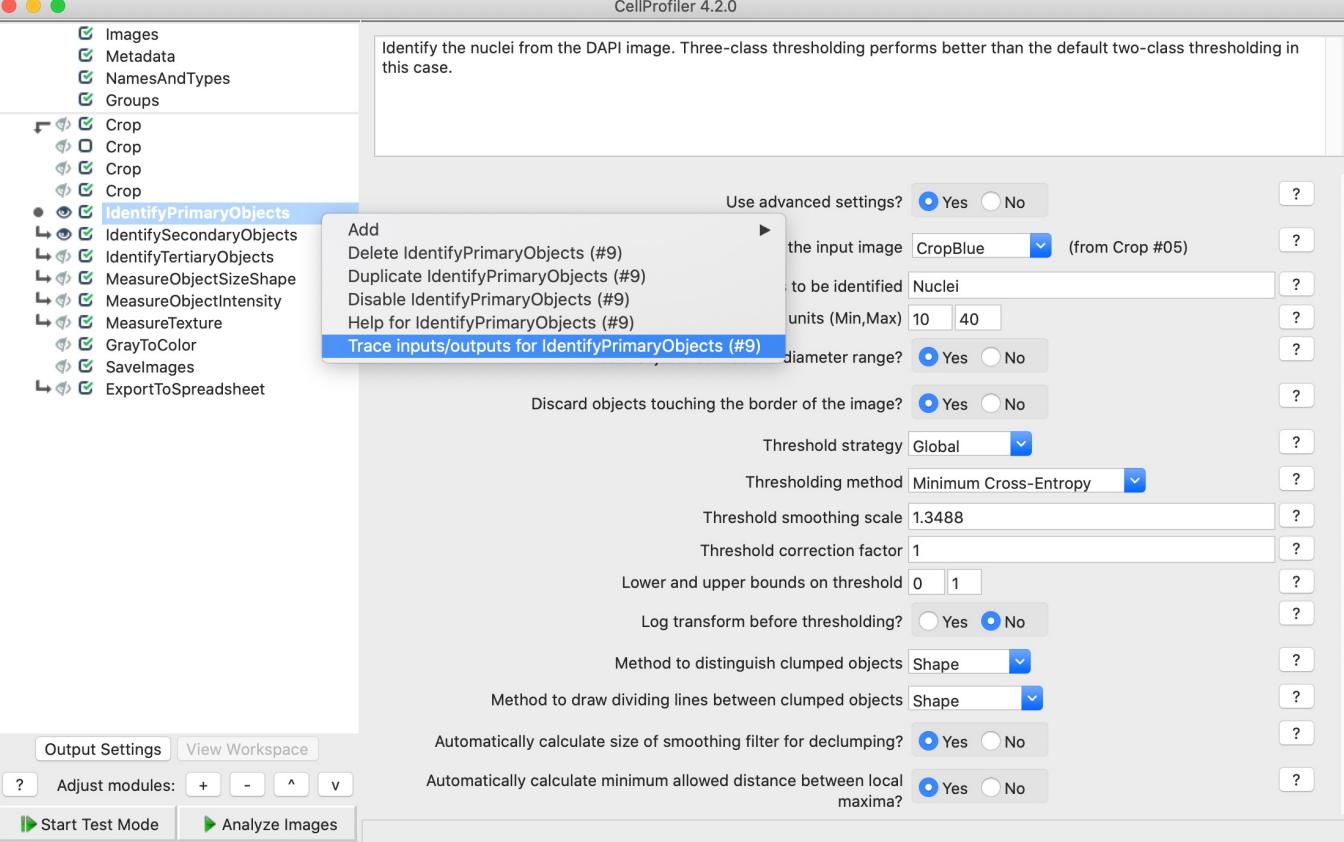
Automatically calculate size of smoothing filter for declumping? Yes No

Automatically calculate minimum allowed distance between local maxima? Yes No

Output Settings View Workspace

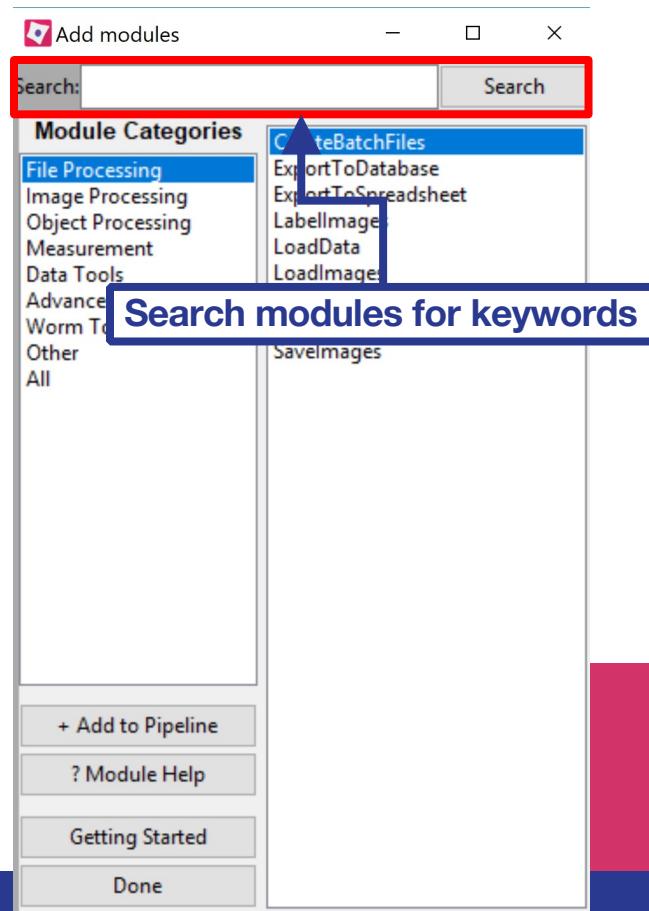
Adjust modules: + - ^ v

Start Test Mode Analyze Images



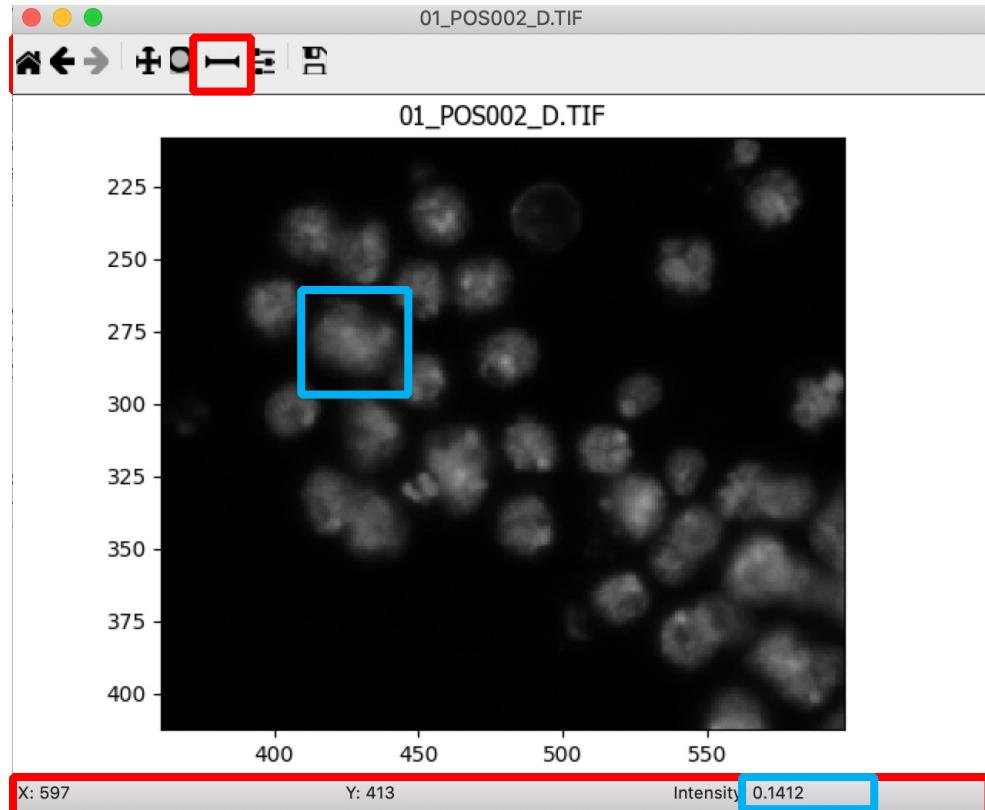
Module categories

- . **File processing:** Image input, file output
- . **Image processing:** Often used for pre-processing prior to object identification
- . **Object processing:** Identification, modification of objects of interest
- . **Measurement:** Collection of measurements from objects of interest
- . **Data Tools:** Measurement exploration, measurement output
- . **Advanced:** Typically modules for 3D analyses
- . **Worm Toolbox:** *C. elegans*-specific operations



CellProfiler figure windows

- . The figure window has additional menu options
- . Toolbar menu: Home, pan, zoom in/out
- . CellProfiler Image Tools
 - Show pixel data (location, intensity)
 - Measure length between any two points just by clicking and dragging



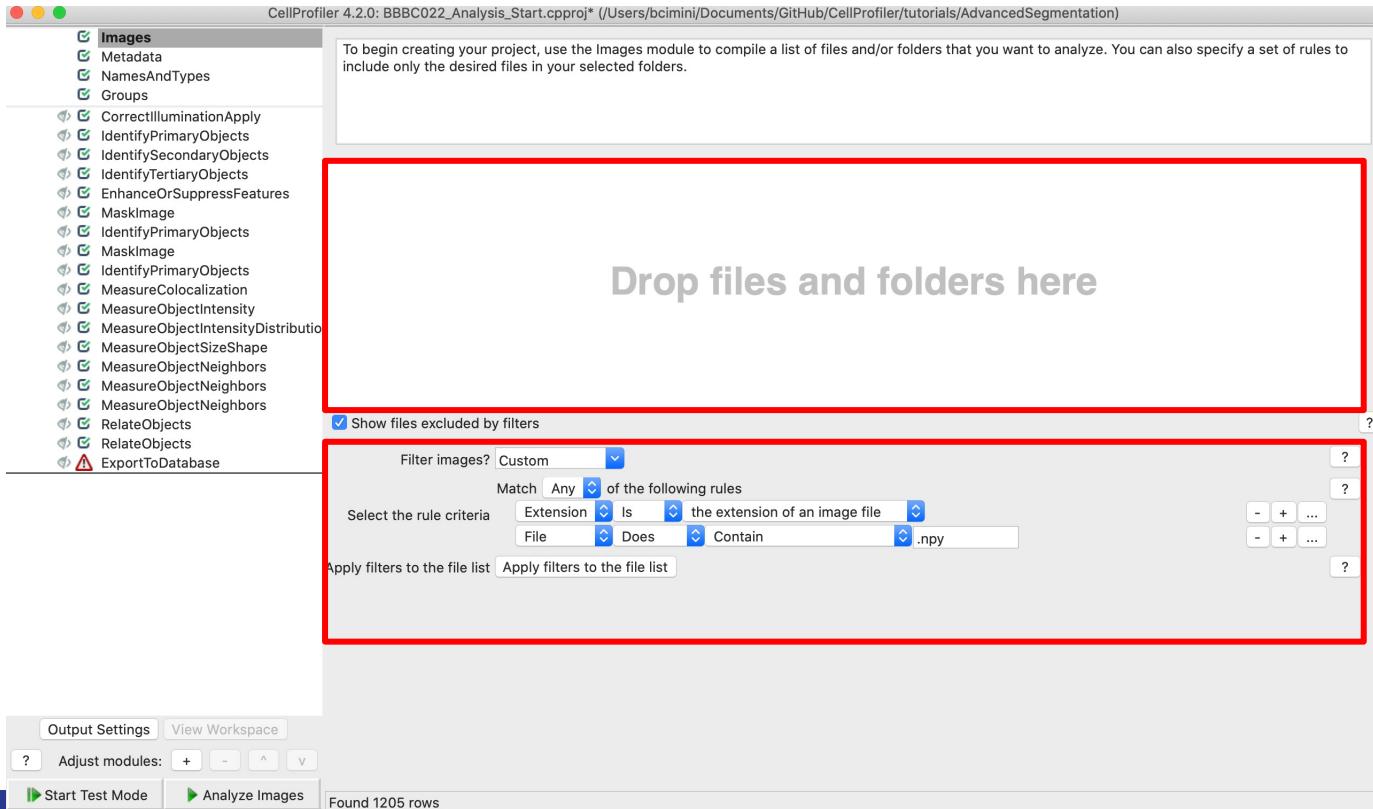
Input Modules

- 4 modules in total, handle all the “bookkeeping” of what your experimental setup is
 - Images (mandatory) – tell CellProfiler which images you want to analyze
 - Metadata (optional if one field of view per file) – give CellProfiler metadata from the file header OR file name
 - NamesAndTypes (mandatory) – tell CellProfiler if 2D vs 3D, how to break down channels, any other bookkeeping
 - Groups (optional outside of tracking, Z projection, or whole-plate correction pipelines) – tell CellProfiler if it is important to keep any image sets together during processing
- Unlike the rest of CellProfiler, these 4 modules are NOT linear, they work together in concert- so if Groups is on and is set incorrectly, NamesAndTypes will break
- See a great blog post about this, with links to a video tutorial, at broad.io/CellProfilerInput

Input Modules - Images

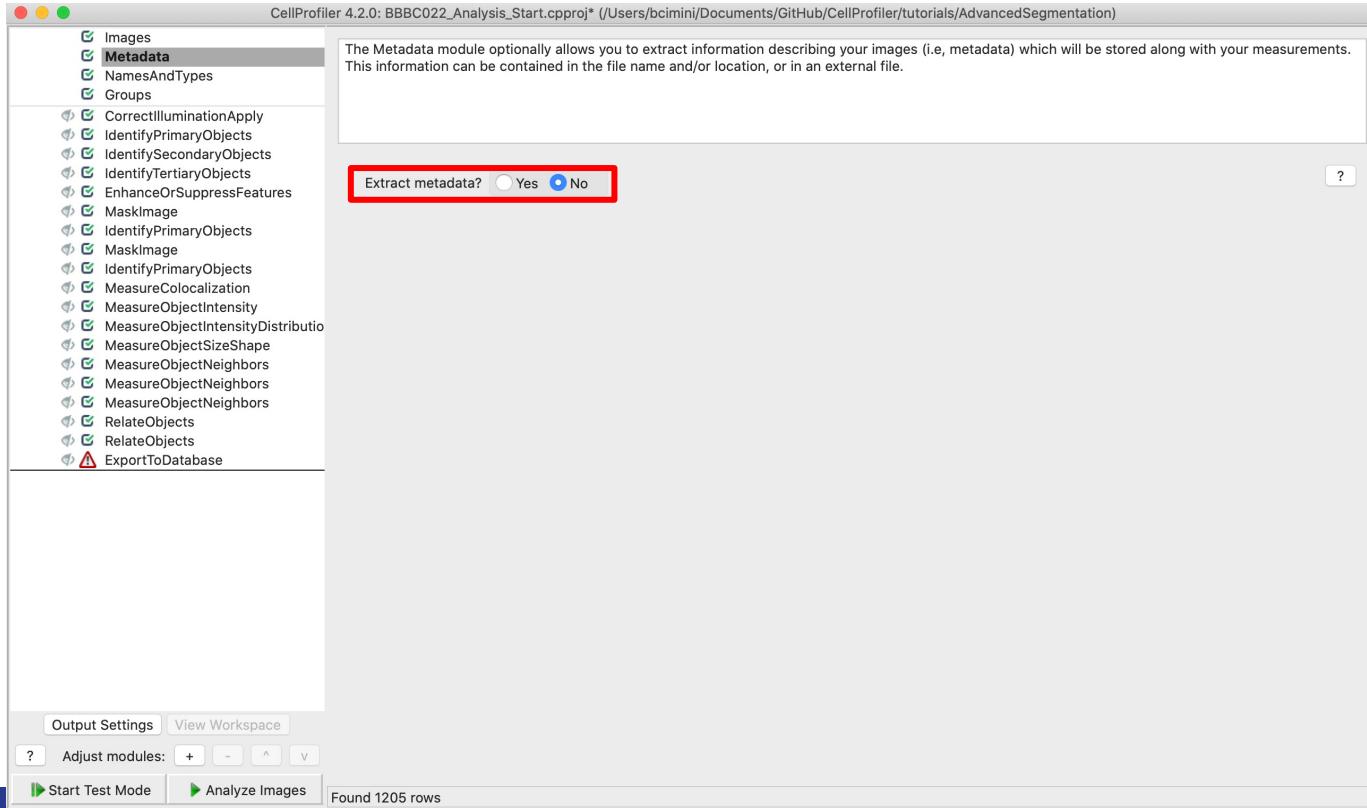
Drop
images
here (or
right click to
search)

Optionally
filter by text
or file type



Input Modules - Metadata

Keep this off if not using



Input Modules - Metadata

Metadata
can come
from any or
all of 3
places

Screenshot of CellProfiler 4.2.0 showing the Metadata module configuration.

The sidebar on the left lists available modules:

- Images
- Metadata
- NamesAndTypes
- Groups
- CorrectIlluminationApply
- IdentifyPrimaryObjects
- IdentifySecondaryObjects
- IdentifyTertiaryObjects
- EnhanceOrSuppressFeatures
- MaskImage
- IdentifyPrimaryObjects
- MaskImage
- IdentifyPrimaryObjects
- MeasureColocalization
- MeasureObjectIntensity
- MeasureObjectIntensityDistribution
- MeasureObjectSizeShape
- MeasureObjectNeighbors
- MeasureObjectNeighbors
- MeasureObjectNeighbors
- RelateObjects
- RelateObjects
- ExportToDatabase

The main panel contains the following configuration for the Metadata module:

- Extract metadata?**: Yes (radio button selected)
- Metadata extraction method**: Extract from file/folder names (selected)
- Metadata source**: Extract from file/folder names (highlighted with a red box)
- Regular expression to extract from file name**: (?P<Plate>[0-9]{5})_w(?P<ChannelNumber>[0-9])
- Extract metadata from**: Images matching a rule
- Select the filtering criteria**: Match All of the following rules
- File**: Does not Contain .npy

Below this, another section for Metadata extraction method is shown:

- Metadata extraction method**: Extract from file/folder names
- Metadata source**: Folder name
- Regular expression to extract from folder name**: (?P<Plate>[0-9]{5})

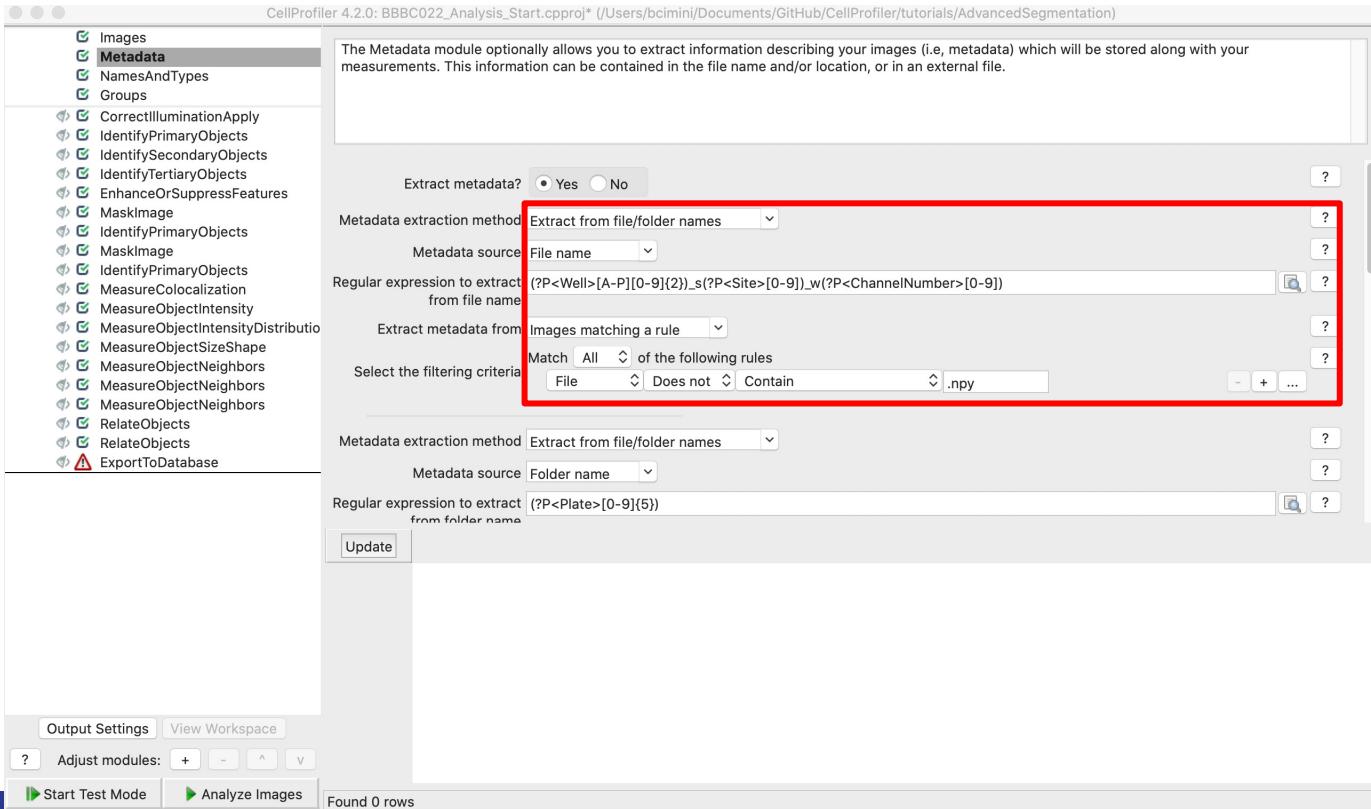
Update	Path / URL	Series	Frame	ChannelNumber	FileLocation	Plate	Site	Well
1	/Users/bcimin...5_IllumER.npy	0	0		file:/Users/b...5_IllumER.npy	20585		
2	/Users/bcimin...umHoechst.npy	0	0		file:/Users/b...umHoechst.npy	20585		
3	/Users/bcimin...IllumMito.npy	0	0		file:/Users/b...IllumMito.npy	20585		
4	/Users/bcimin...mPh_golgi.npy	0	0		file:/Users/b...mPh_golgi.npy	20585		
5	/Users/bcimin...IllumSyto.npy	0	0		file:/Users/b...IllumSyto.npy	20585		
6	/Users/bcimin...0160D4FB2.tif	0	0	1	file:/Users/b...0160D4FB2.tif	20585	1	A01
7	/Users/bcimin...04D6F7FF2.tif	0	0	2	file:/Users/b...04D6F7FF2.tif	20585	1	A01
8	/Users/bcimin...8709F6E33.tif	0	0	3	file:/Users/b...8709F6E33.tif	20585	1	A01
9	/Users/bcimin...64B9F0088.tif	0	0	4	file:/Users/b...64B9F0088.tif	20585	1	A01
10	/Users/bcimin...E12ADC245.tif	0	0	5	file:/Users/b...E12ADC245.tif	20585	1	A01
11	/Users/bcimin...E5C741B75.tif	0	0	1	file:/Users/b...E5C741B75.tif	20585	2	A01

At the bottom of the main panel, it says "Found 1205 rows".

Buttons at the bottom include: Output Settings, View Workspace, Adjust modules, Start Test Mode, Analyze Images.

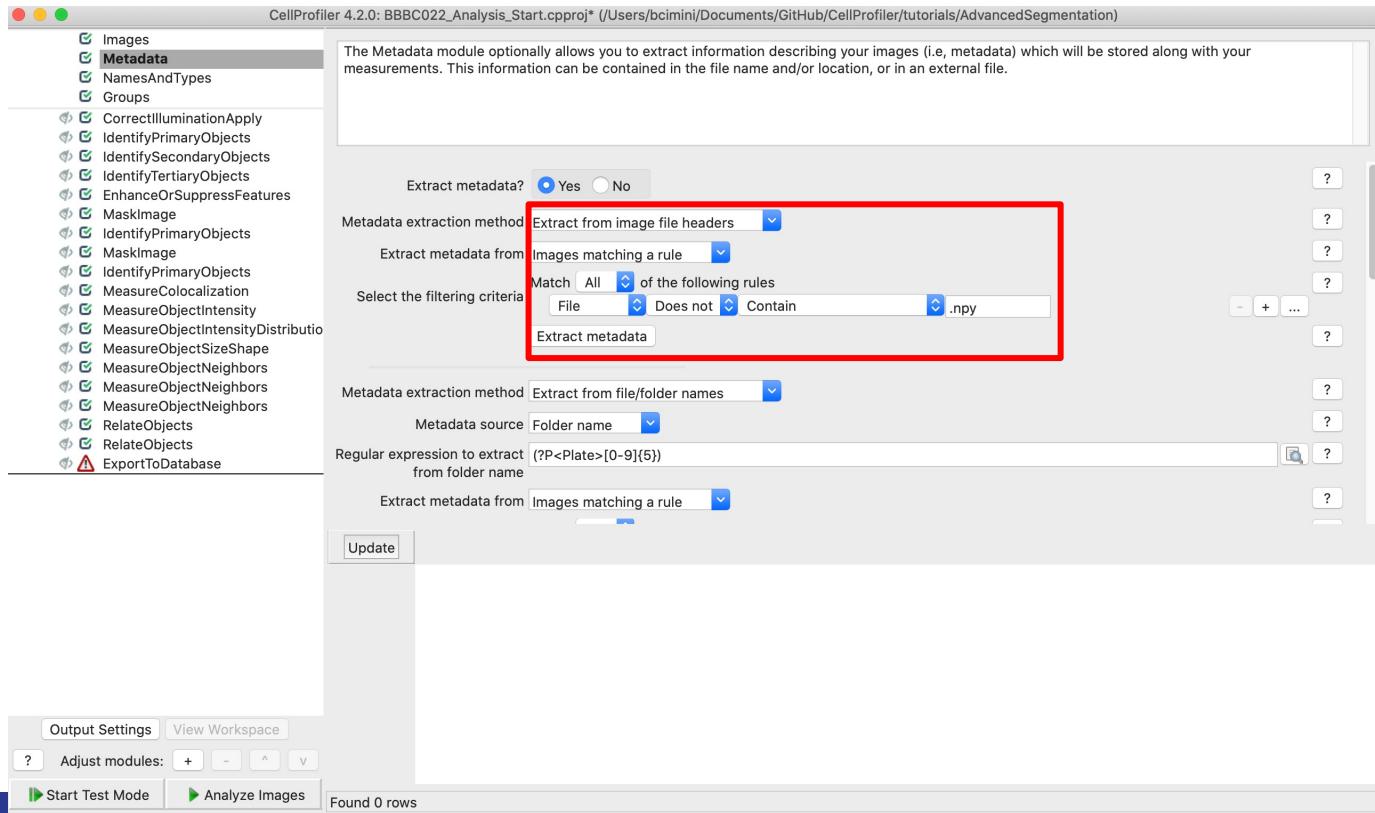
Input Modules - Metadata

1) A regular expression based on the file or folder name (see module help for resources on how to configure)



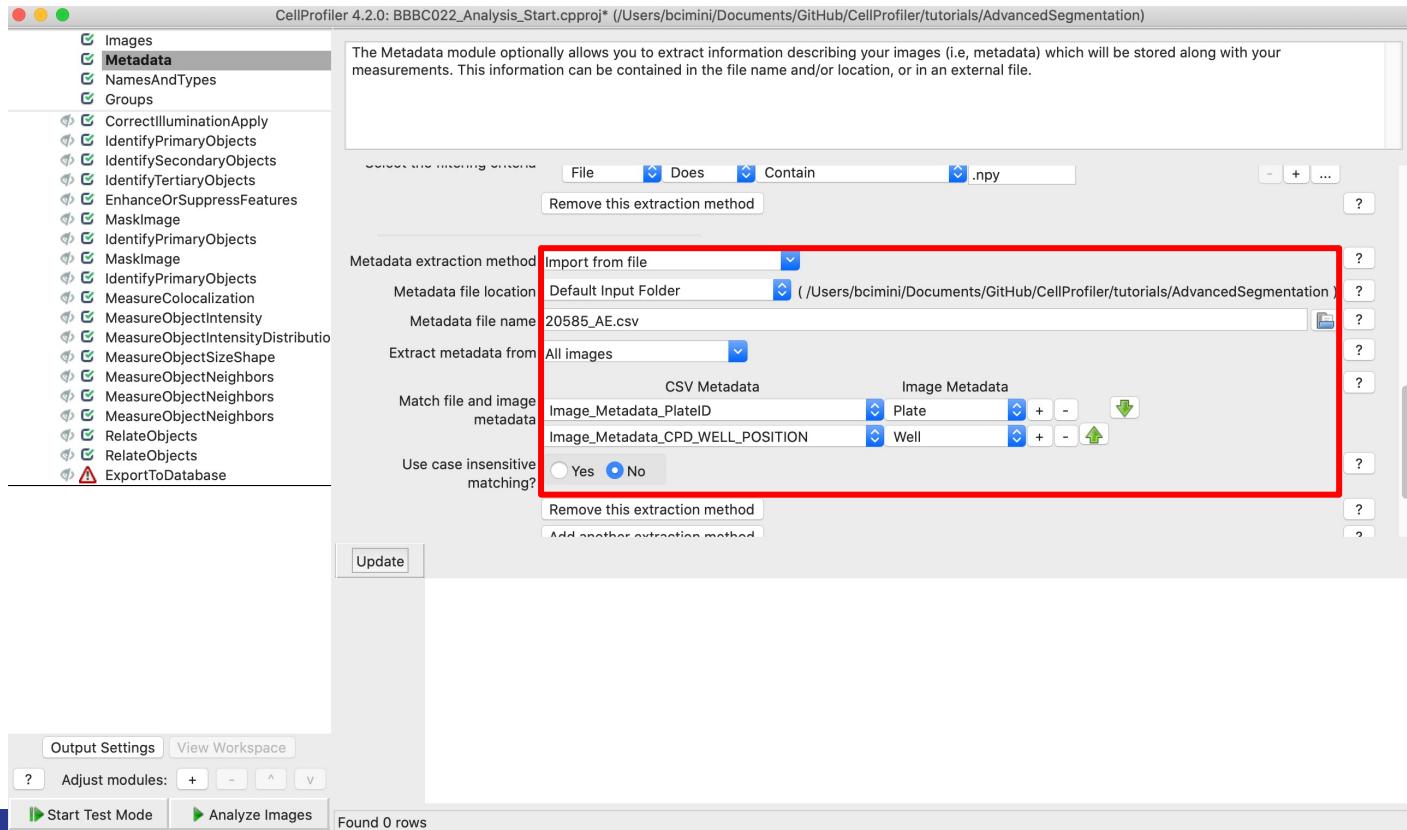
Input Modules - Metadata

2) Extracted from the file headers
(make sure to hit the “Extract Metadata” button!)



Input Modules - Metadata

3) Brought in from an internal file- can't be used on its own, can only be used in addition to 1 or 2



Input Modules - Metadata

When
you're
done, hit
“Update”,
and the
metadata
will
populate in
the box

CellProfiler 4.2.0: BBBC022_Analysis_Start.cpproj* (/Users/bcimini/Documents/GitHub/CellProfiler/tutorials/AdvancedSegmentation)

The Metadata module optionally allows you to extract information describing your images (i.e., metadata) which will be stored along with your measurements. This information can be contained in the file name and/or location, or in an external file.

Extract metadata? Yes No

Metadata extraction method Extract from file/folder names

Metadata source File name

Regular expression to extract from file name `(?P<Well>[A-P][0-9]{2})_(?P<Site>[0-9])_w(?P<ChannelNumber>[0-9])`

Extract metadata from Images matching a rule

Select the filtering criteria Match All of the following rules
File Does not Contain .npy

Metadata extraction method Extract from file/folder names

Metadata source Folder name

Regular expression to extract from folder name `(?P<Plate>[0-9]{5})`

Update	Path / URL	Series	Frame	ChannelNumber	FileLocation	ImageNumber	Image_FileName_OrigER
1	/Users/bcimin...5_IllumER.npy	0	0		file:/Users/b...5_IllumER.npy		
2	/Users/bcimin...umHoechst.npy	0	0		file:/Users/b...umHoechst.npy		
3	/Users/bcimin...IllumMito.npy	0	0		file:/Users/b...IllumMito.npy		
4	/Users/bcimin...mPh_golgi.npy	0	0		file:/Users/b...mPh_golgi.npy		
5	/Users/bcimin...IllumSyto.npy	0	0		file:/Users/b...IllumSyto.npy		
6	/Users/bcimin...0160D4FB2.tif	0	0	1	file:/Users/b...0160D4FB2.tif	9	IXMtest_A01_s...F055DD8A4.tif
7	/Users/bcimin...04D6FF7F2.tif	0	0	2	file:/Users/b...04D6FF7F2.tif	9	IXMtest_A01_s...F055DD8A4.tif
8	/Users/bcimin...8709F6E33.tif	0	0	3	file:/Users/b...8709F6E33.tif	9	IXMtest_A01_s...F055DD8A4.tif
9	/Users/bcimin...64B9F0088.tif	0	0	4	file:/Users/b...64B9F0088.tif	9	IXMtest_A01_s...F055DD8A4.tif
10	/Users/bcimin...E12ADC245.tif	0	0	5	file:/Users/b...E12ADC245.tif	9	IXMtest_A01_s...F055DD8A4.tif

Output Settings View Workspace

? Adjust modules: + - A V

Start Test Mode Analyze Images

Found 1205 rows

Input Modules - NamesAndTypes

Set if you have one vs >1 images per image set

Set 2D vs 3D

Determine how to recognize each channel (ie metadata, file name, etc)

Give each channel a descriptive name

Set the image type
(Grayscale, color, object label matrix, etc)

CellProfiler 4.2.0: BBB022_Analysis_Final.cppproj* (/Users/bcimini/Documents/GitHub/CellProfiler/tutorials/AdvancedSegmentation)

The NamesAndTypes module allows you to assign a meaningful name to each image by which other modules will refer to it.

Assign a name to **Images matching rules**

Process as 3D? Yes No

Select the rule criteria Match All of the following rules

Metadata Does Have ChannelNumber matching 1

Name to assign these OrigHoechst images

Select the image type Grayscale image

Set intensity range from Image metadata

Duplicate this image

Select the rule criteria Match All of the following rules

Metadata Does Have ChannelNumber matching 2

Update

Output Settings View Workspace

Adjust modules: + - ^

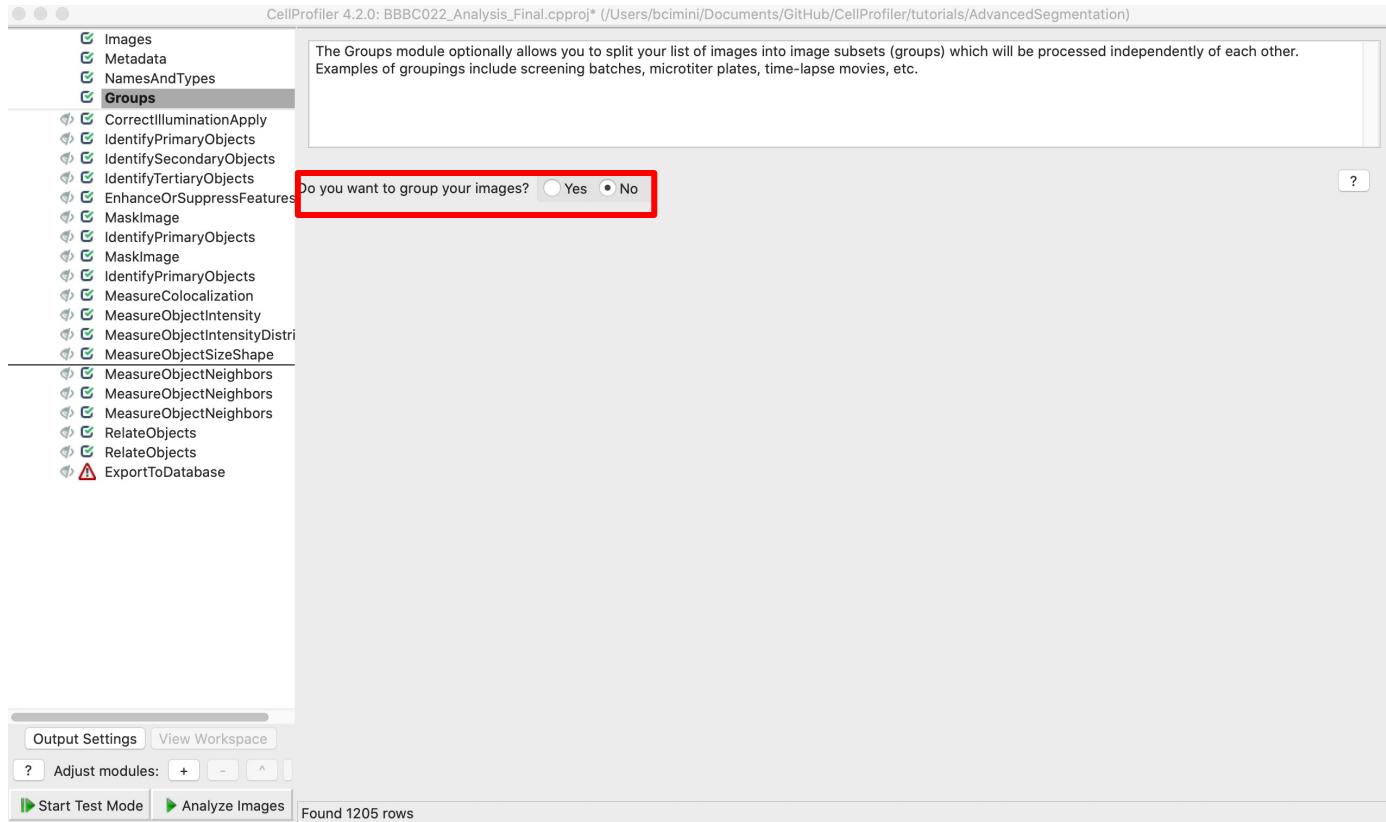
Start Test Mode Analyze Images Found 1205 rows

The screenshot shows the CellProfiler 4.2.0 software interface with the 'NamesAndTypes' module selected. The main panel displays configuration options for assigning names to images based on matching rules. A red box highlights the 'Name to assign these' field containing 'OrigHoechst images'. Another red box highlights the 'Select the image type' dropdown set to 'Grayscale image'.

Input Modules - Groups

Keep this off if not using

You need it if it matters which images are processed together



Input Modules - Groups

Set
variable(s)
to group by

CellProfiler 4.2.0: BBBC022_Analysis_Final.cpproj* (/Users/bcimini/Documents/GitHub/CellProfiler/tutorials/AdvancedSegmentation)

The Groups module optionally allows you to split your list of images into image subsets (groups) which will be processed independently of each other. Examples of groupings include screening batches, microtiter plates, time-lapse movies, etc.

Do you want to group your images? Yes No

Metadata category: Well

Add another metadata item

	Group: Well	Count
1	A01	2
2	A02	2
3	A03	2
4	A04	2

Grouping list

	Group number	Group index	Group: Well	Path: IllumER	File: IllumER	Pa
1	1	1	A01	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
2	1	2	A01	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
3	2	1	A02	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
4	2	2	A02	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
5	3	1	A03	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
6	3	2	A03	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
7	4	1	A04	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
8	4	2	A04	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
9	5	1	A05	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
10	5	2	A05	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
11	6	1	A06	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
12	6	2	A06	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
13	7	1	A07	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci
14	7	2	A07	/Users/bcimin...C022_20585_AE	20585_IllumER.npy	/Users/bci

Image sets

Output Settings View Workspace

? Adjust modules: + - ^

Start Test Mode Analyze Images Found 1205 rows

Image Processing Modules

- All classical image analysis segmentation works off of the principle- “my image should have my objects of interest bright and everywhere else dark”
- **For segmentation purposes**, basically any transformation you can think of to achieve this is “legal”
 - **Emphatically**, this does not apply to images you are actively measuring, just images to be fed into a segmentation
- To achieve this, you may need to enhance dim signal that really is there and/or mask out bright signal that truly is (but is not supposed to be segmented at the current time)
- The more transformations you do before segmentation, the more critical it is to check your segmentations against your raw images (with OverlayOutlines, OverlayObjects, or the WorkspaceViewer)

Overlaying segmentations on raw images is critical (but easy!)

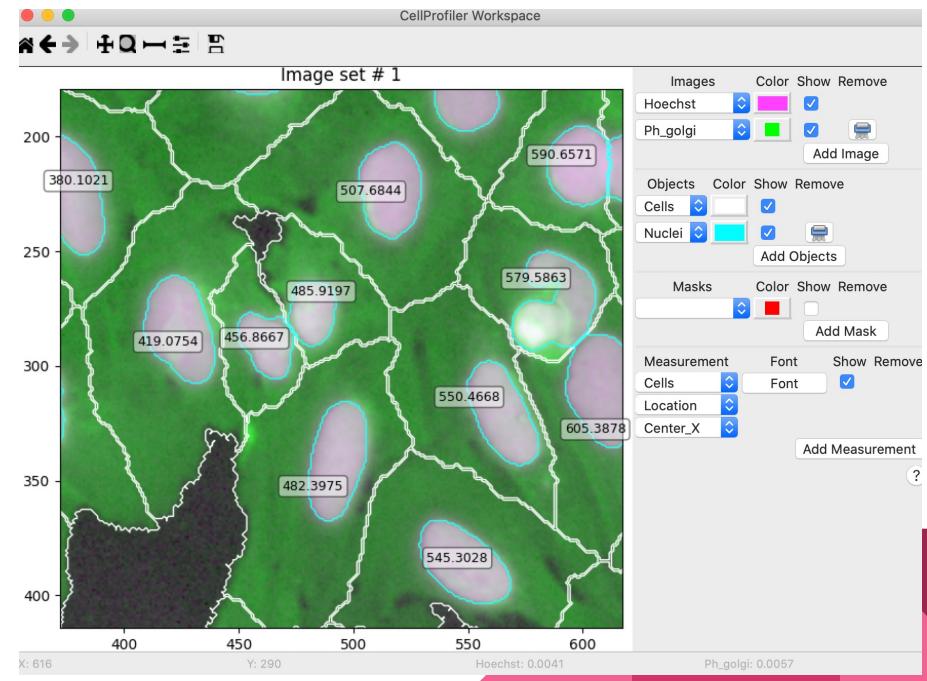
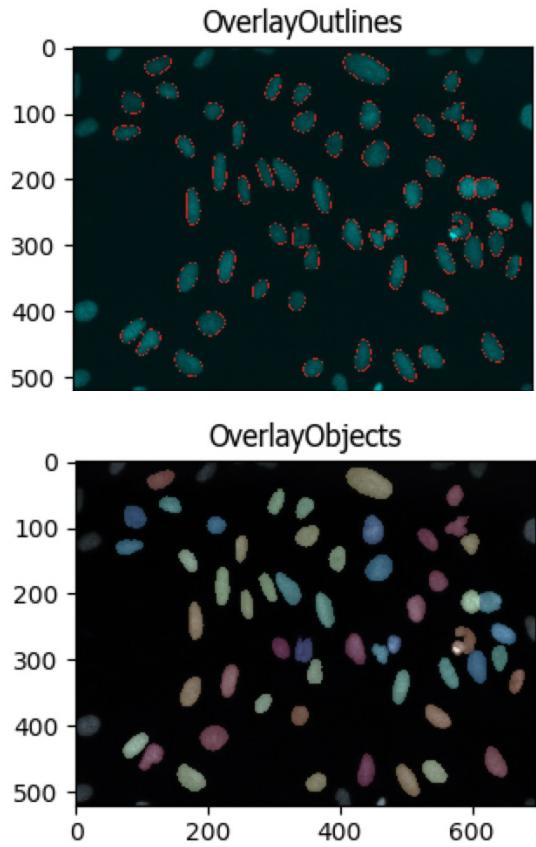


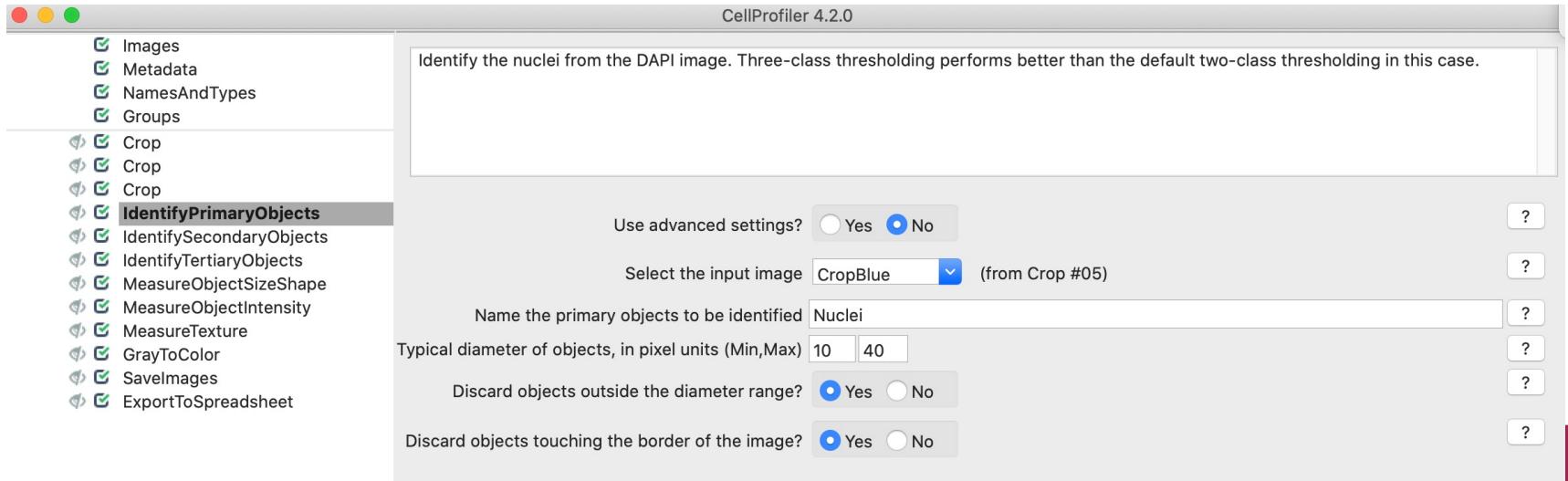
Image Processing modules in CellProfiler

(Not exhaustive, just inspirational)

- In the Image Processing category:
 - EnhanceEdges
 - EnhanceOrSuppressFeatures- works on tubes, speckles, and/or dark holes
 - Smooth
 - Crop
 - ColorToGray / GrayToColor
- In the “Advanced” category:
 - GaussianFilter
 - MedialAxis
 - RemoveHoles
 - GaussianFilter
 - MatchTemplate
 - RunImageJMacro

Primary object identification

Many options for thresholding, splitting and merging, etc.



Primary object identification

Which image to use, and what to call the objects

CellProfiler 4.2.0

Identify the nuclei from the DAPI image. Three-class thresholding performs better than the default two-class thresholding in this case.

Use advanced settings? Yes No

Select the input image: CropBlue (from Crop #05)

Name the primary objects to be identified: Nuclei

Typical diameter of objects, in pixel units (Min,Max): 10 40

Discard objects outside the diameter range? Yes No

Discard objects touching the border of the image? Yes No

Output Settings View Workspace

? Adjust modules: + - ^ v

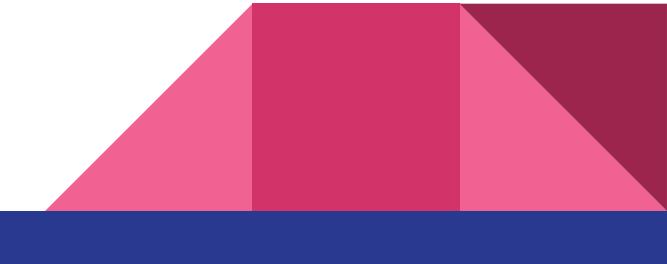
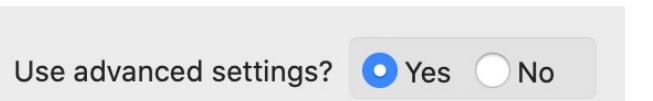
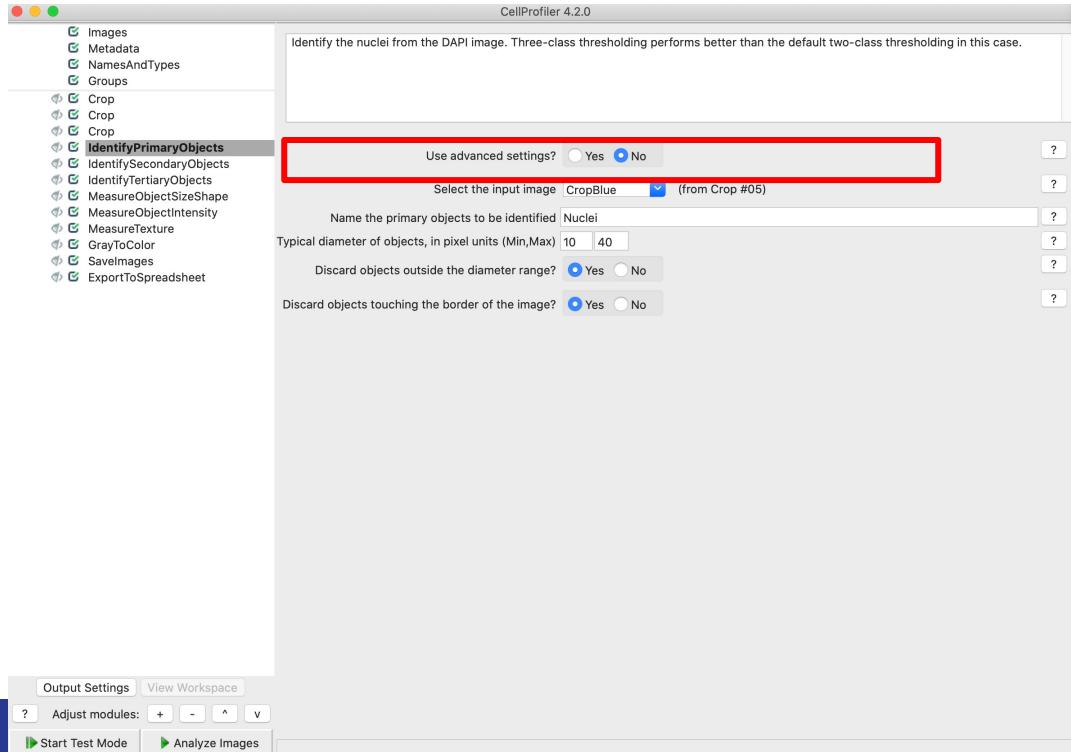
Start Test Mode Analyze Images

Select the input image CropBlue (from Crop #05)

Name the primary objects to be identified Nuclei

Primary object identification

Use defaults for thresholding, splitting and merging, etc.



Primary object identification

You have options as to which objects to keep, how to threshold, etc

CellProfiler 4.2.0

Identify the nuclei from the DAPI image. Three-class thresholding performs better than the default two-class thresholding in this case.

Use advanced settings? Yes No

Select the input image: CropBlue (from Crop #05)

Name the primary objects to be identified: Nuclei

Typical diameter of objects, in pixel units (Min,Max): 10 40

Discard objects outside the diameter range? Yes No

Discard objects touching the border of the image? Yes No

Threshold strategy: Global Local

Thresholding method: Minimum Cross-Entropy

Threshold smoothing scale: 1.3488

Threshold correction factor: 1

Lower and upper bounds on threshold: 0 1

Log transform before thresholding? Yes No

Method to distinguish clumped objects: Shape

Method to draw dividing lines between clumped objects: Shape

Automatically calculate size of smoothing filter for declumping? Yes No

Automatically calculate minimum allowed distance between local maxima? Yes No

Speed up by using lower-resolution image to find local maxima? Yes No

Display accepted local maxima? Yes No

Fill holes in identified objects? After declumping only

Handling of objects if excessive number of objects identified: Continue

Output Settings View Workspace

Adjust modules: + - ^ v

Start Test Mode Analyze Images

Discard objects outside the diameter range? Yes No

Discard objects touching the border of the image? Yes No

Handling of objects if excessive number of objects identified: Continue

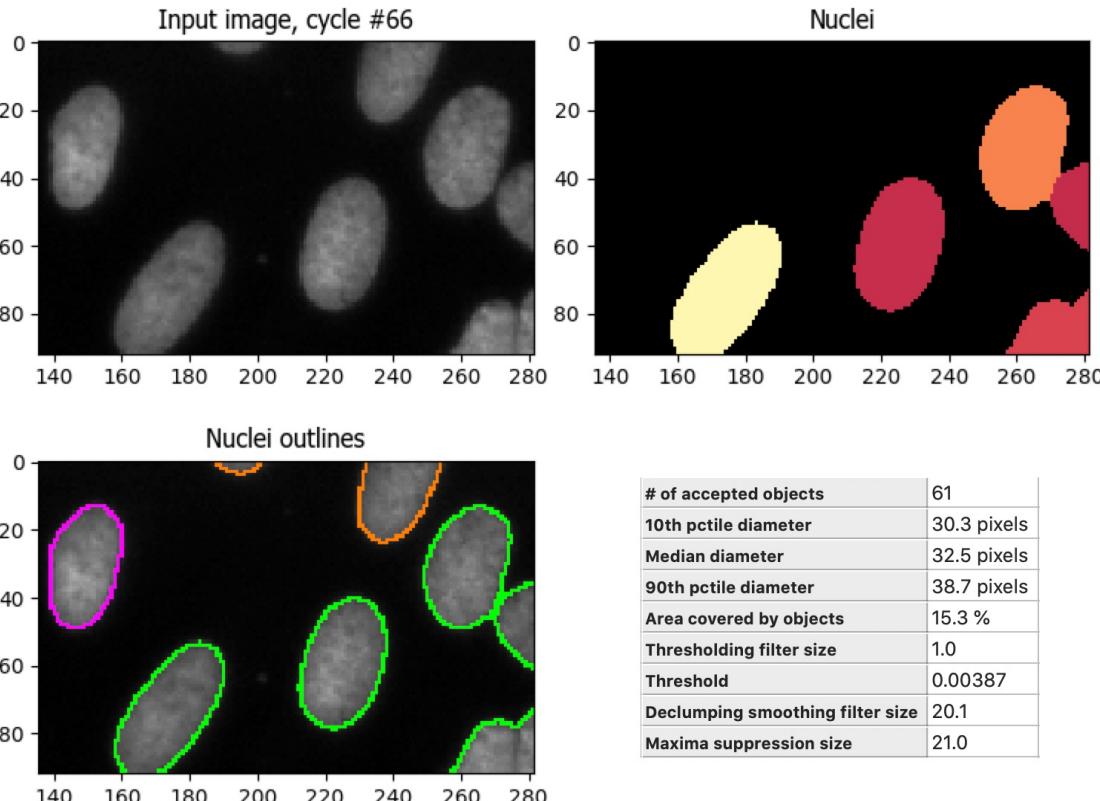
Primary object identification

Outline colors show if objects are rejected (and why)

Green = accepted
(shows in the top right)

Pink = not accepted
(outside of the size range (too large OR too small), does not show in the top right)

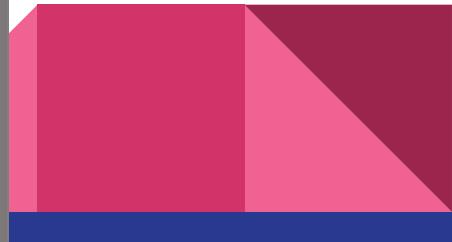
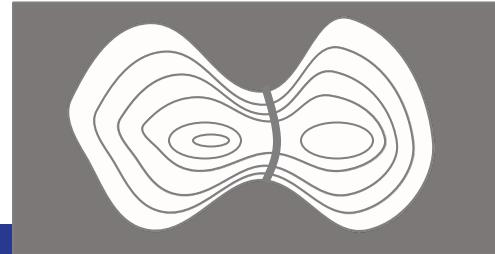
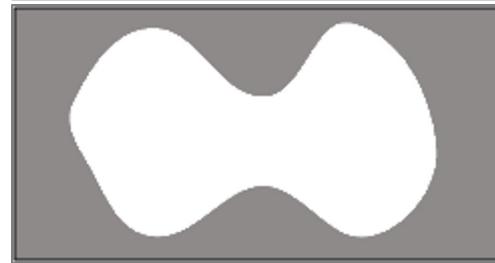
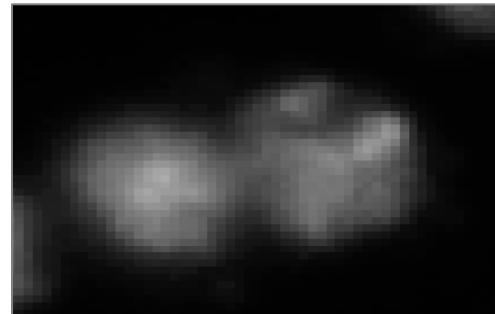
Orange = not accepted (touching the edge, does not show in the top right)



Object identification

Once the images are loaded, how do you find objects of interest?

- **Step 1:** Distinguish the foreground from the background (thresholding)
- **Step 2:** Split/merge “objects” properly



Primary object identification

How should we threshold the image?

CellProfiler 4.2.0

Identify the nuclei from the DAPI image. Three-class thresholding performs better than the default two-class thresholding in this case.

Use advanced settings? Yes No

Select the input image: CropBlue (from Crop #05)

Name the primary objects to be identified: Nuclei

Typical diameter of objects, in pixel units (Min,Max): 10 40

Discard objects outside the diameter range? Yes No

Threshold strategy: Global

Thresholding method: Minimum Cross-Entropy

Threshold smoothing scale: 1.3488

Threshold correction factor: 1

Lower and upper bounds on threshold: 0 1

Log transform before thresholding? Yes No

Method to distinguish clumped objects: Shape

Method to draw dividing lines between clumped objects: Shape

Automatically calculate size of smoothing filter for declumping? Yes No

Automatically calculate minimum allowed distance between local maxima? Yes No

Speed up by using lower-resolution image to find local maxima? Yes No

Display accepted local maxima? Yes No

Fill holes in identified objects? After declumping only

Handling of objects if excessive number of objects identified: Continue

Output Settings View Workspace

Adjust modules: + - ^ v

Start Test Mode Analyze Images

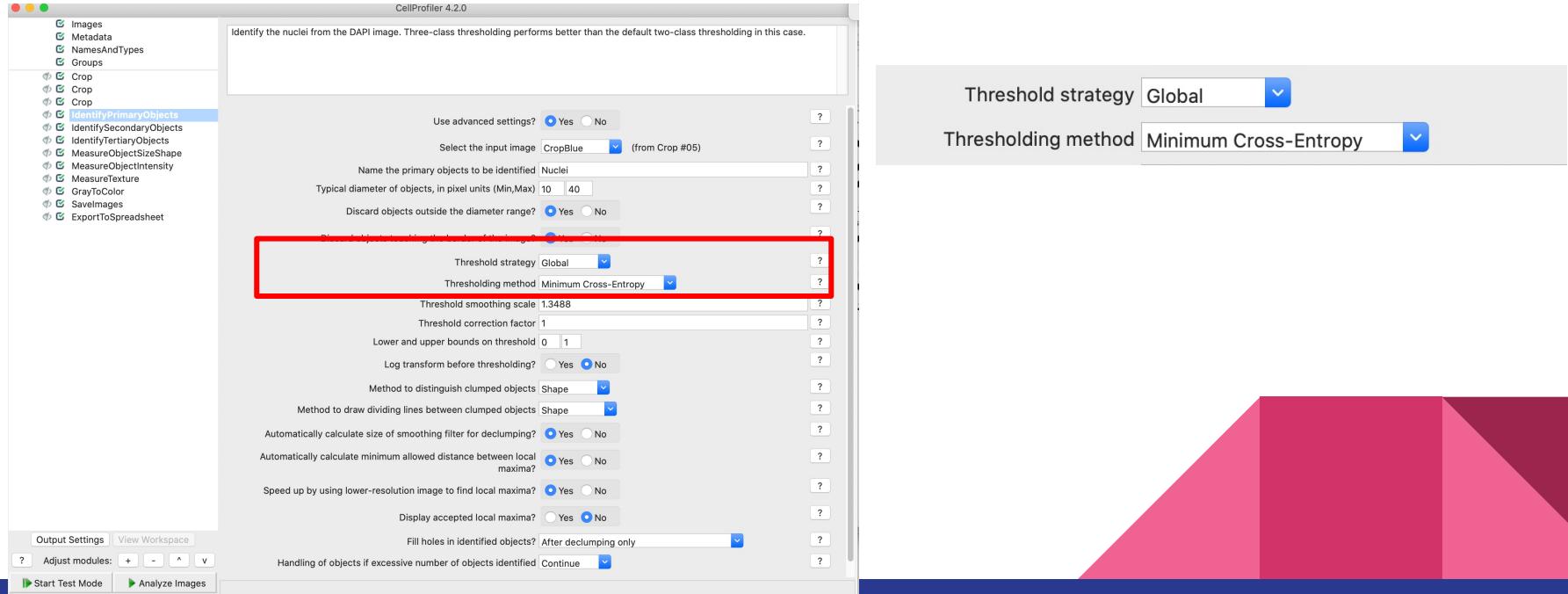
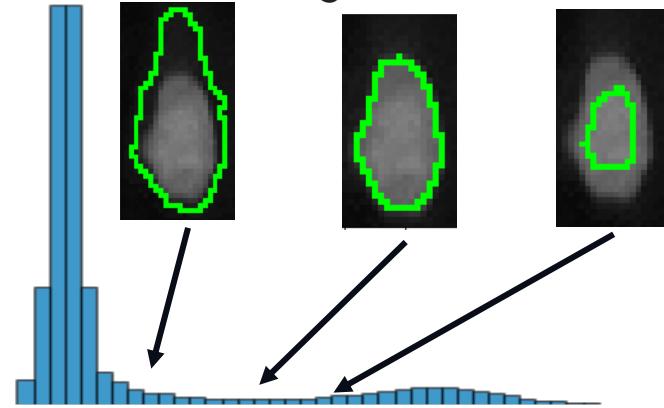


Image thresholding

Definition: Division of the image into background and foreground

What is the best threshold value for dividing the intensity into foreground and background pixels?



Methods:

- Minimum Cross Entropy: Good for readily identifiable foreground / background
- Otsu: Choose between 2- or 3-class if mid-level intensities present
- Robust Background: Good for images in which most of the image is comprised of background

Primary object identification

What adjustments and limits should be applied to the threshold?

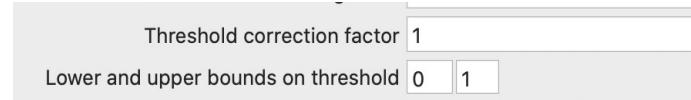
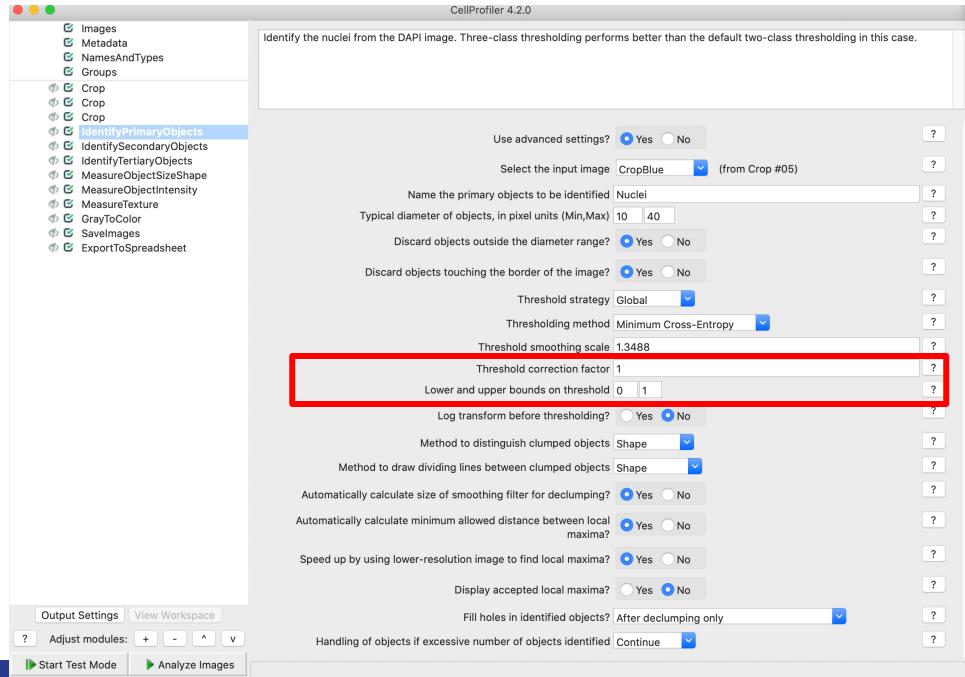


Image thresholding

Correction factors can nudge a threshold higher or lower

What is the best threshold value for dividing the intensity into foreground and background pixels?

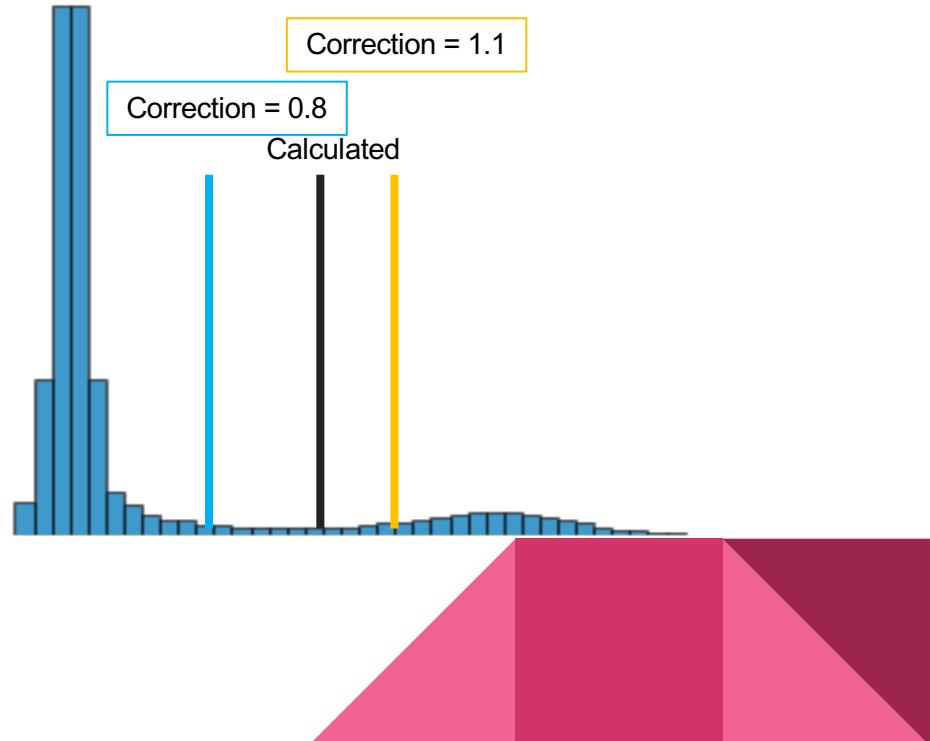


Image thresholding

Upper and lower bounds can help with odd histograms

What is the best threshold value for dividing the intensity into foreground and background pixels?

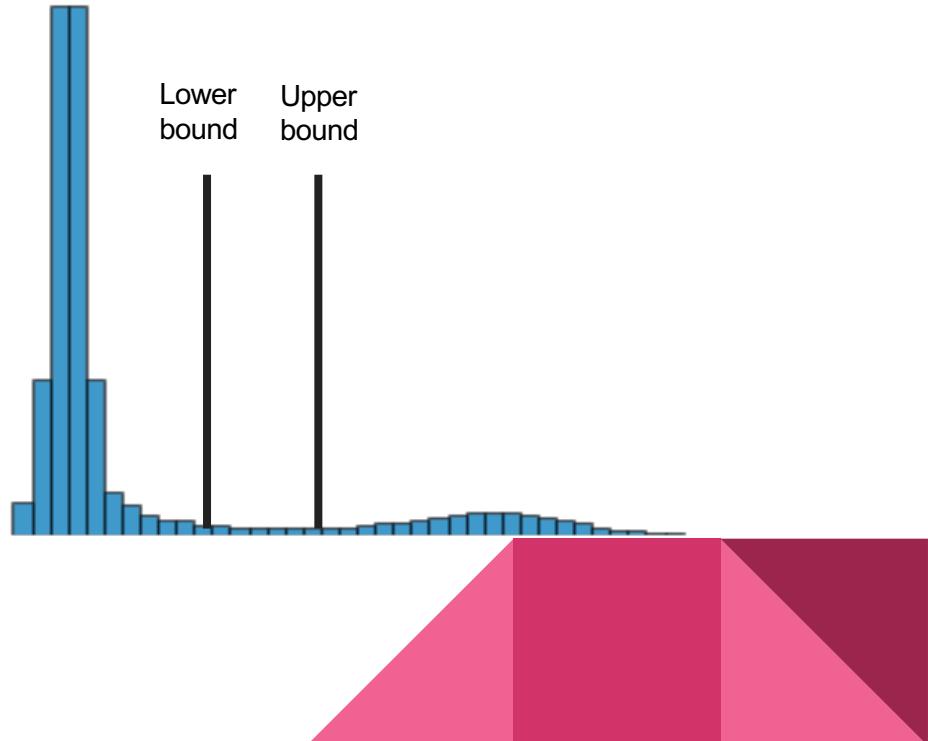


Image thresholding

Upper and lower bounds can help with odd histograms

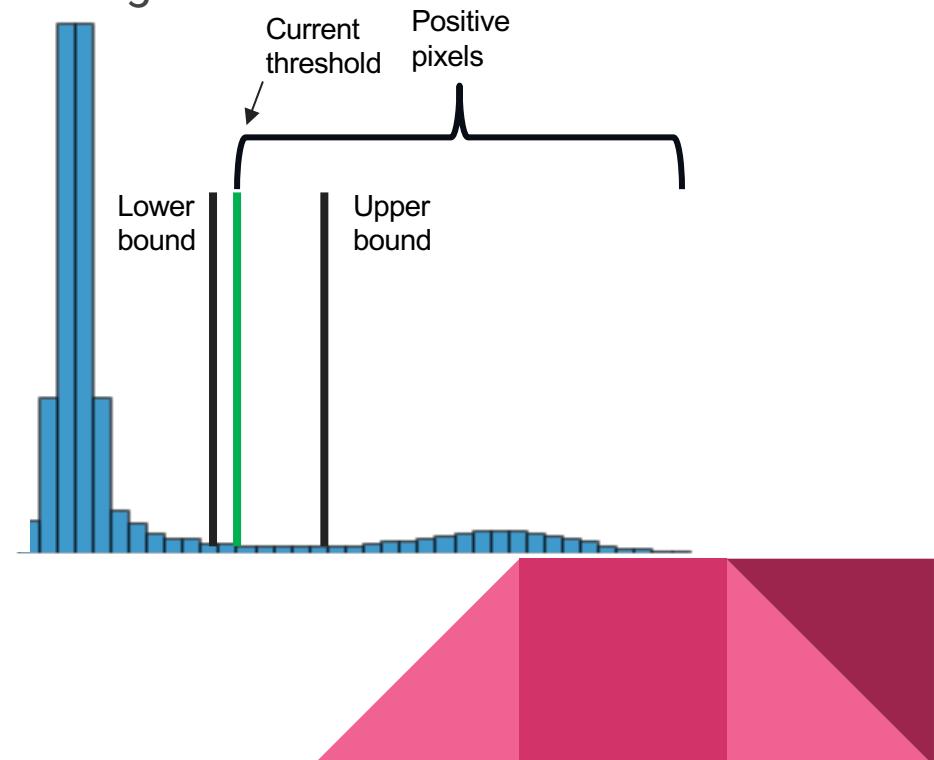
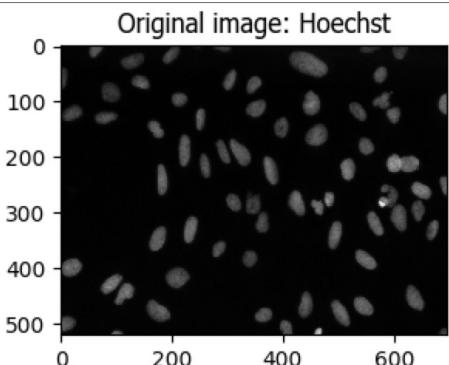


Image thresholding

Upper and lower bounds can help with odd histograms

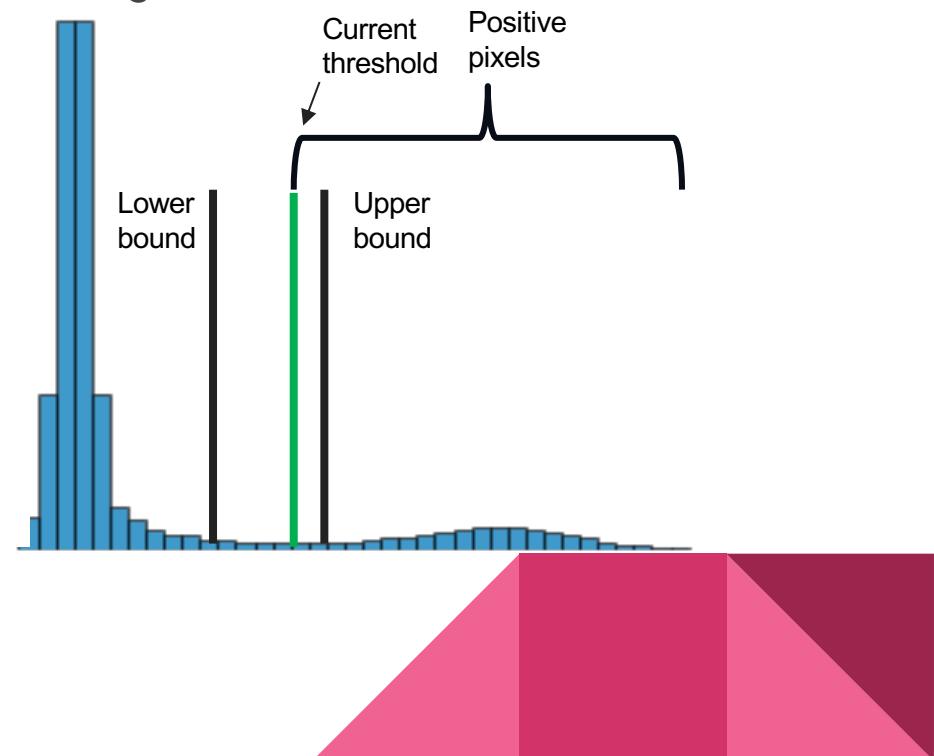
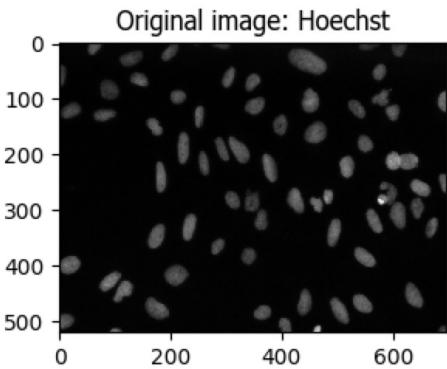


Image thresholding

Upper and lower bounds can help with odd histograms

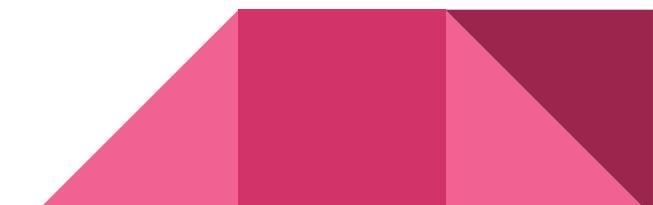
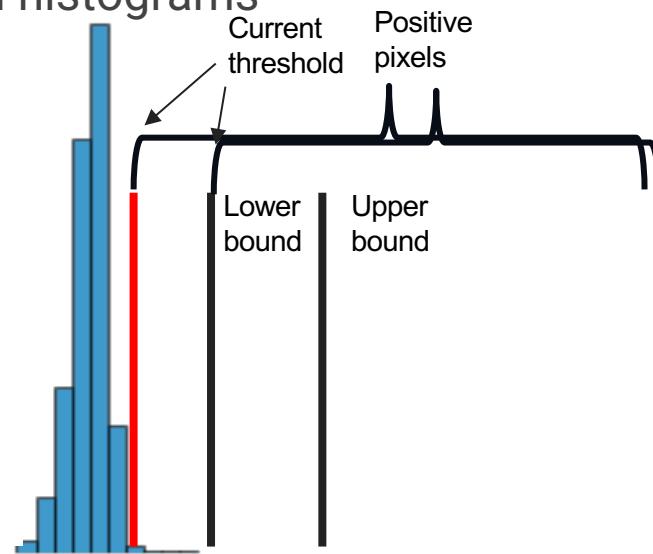
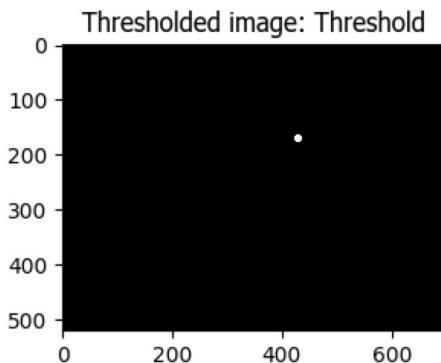
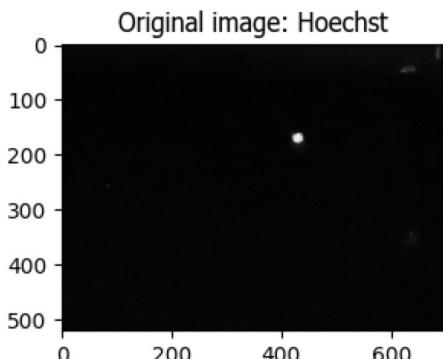
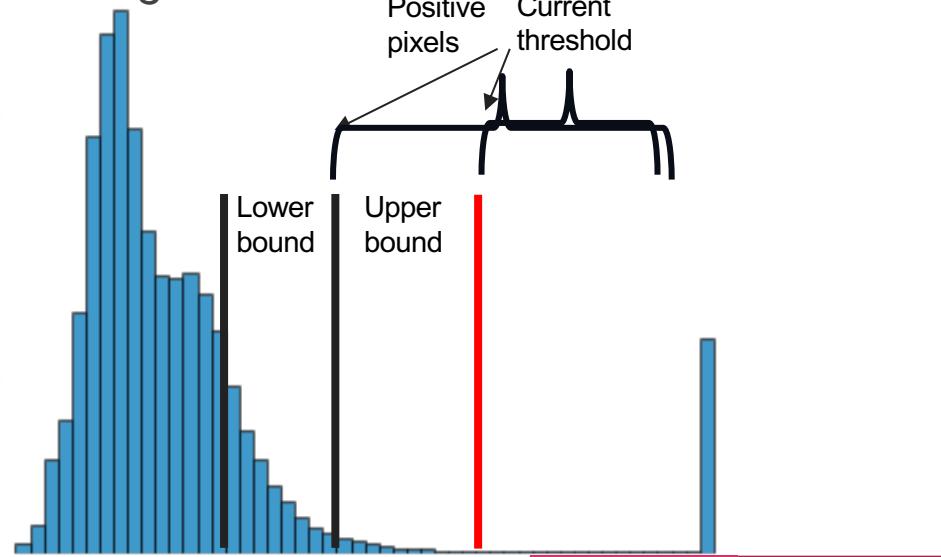
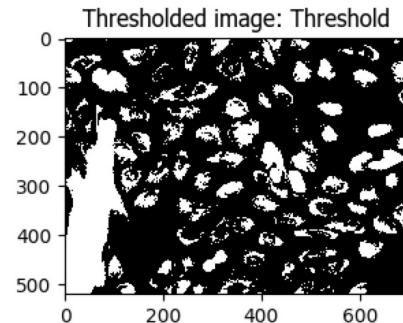
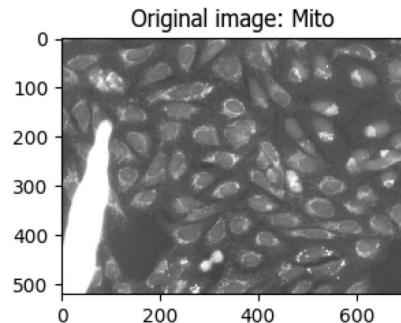


Image thresholding

Upper and lower bounds can help with odd histograms



Primary object identification

Should we transform the image before thresholding?

CellProfiler 4.2.0

Identify the nuclei from the DAPI image. Three-class thresholding performs better than the default two-class thresholding in this case.

Use advanced settings? Yes No

Select the input image: CropBlue (from Crop #05)

Name the primary objects to be identified: Nuclei

Typical diameter of objects, in pixel units (Min,Max): 10 40

Discard objects outside the diameter range? Yes No

Discard objects touching the border of the image? Yes No

Threshold strategy: Global

Thresholding method: Minimum Cross-Entropy

Threshold smoothing scale: 1.3488

Threshold correction factor: 1

Lower and upper bounds on threshold: 0 1

Log transform before thresholding? Yes No

Method to distinguish clumped objects: Shape

Method to draw dividing lines between clumped objects: Shape

Automatically calculate size of smoothing filter for declumping? Yes No

Automatically calculate minimum allowed distance between local maxima? Yes No

Speed up by using lower-resolution image to find local maxima? Yes No

Display accepted local maxima? Yes No

Fill holes in identified objects? After declumping only

Handling of objects if excessive number of objects identified: Continue

Output Settings View Workspace

? Adjust modules: + - ^ v

Start Test Mode Analyze Images

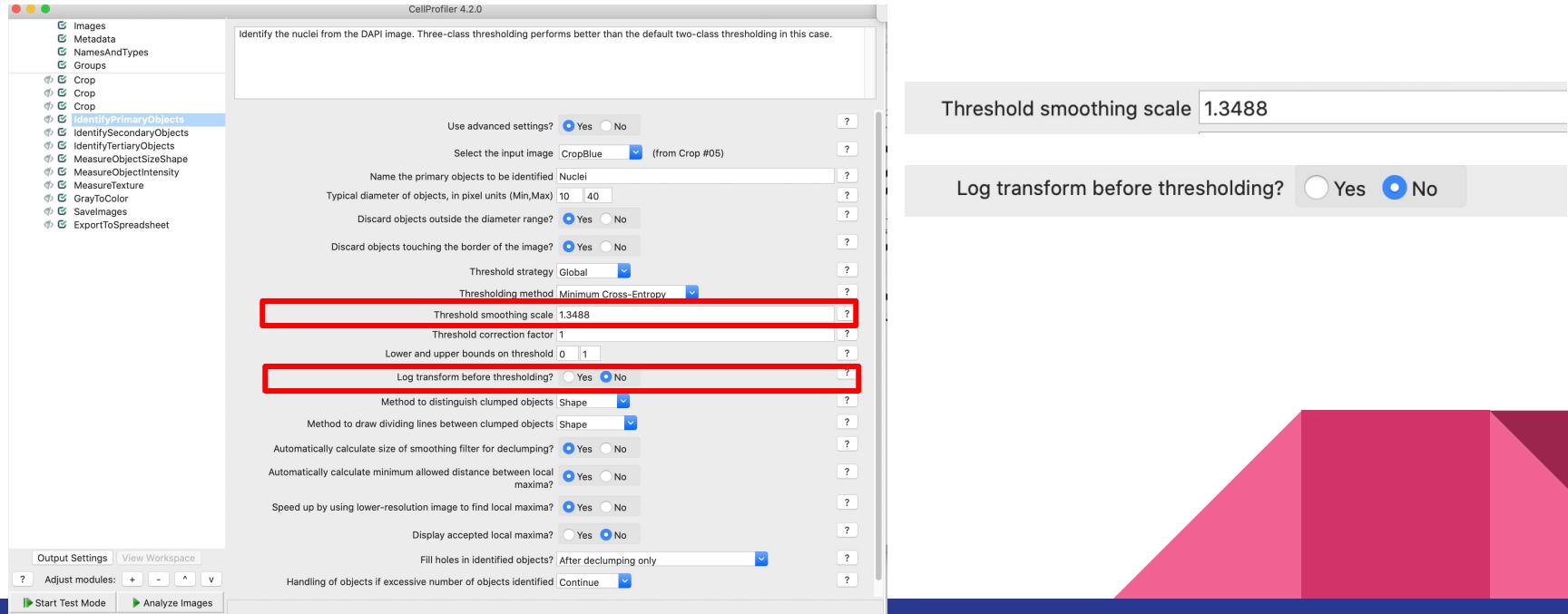


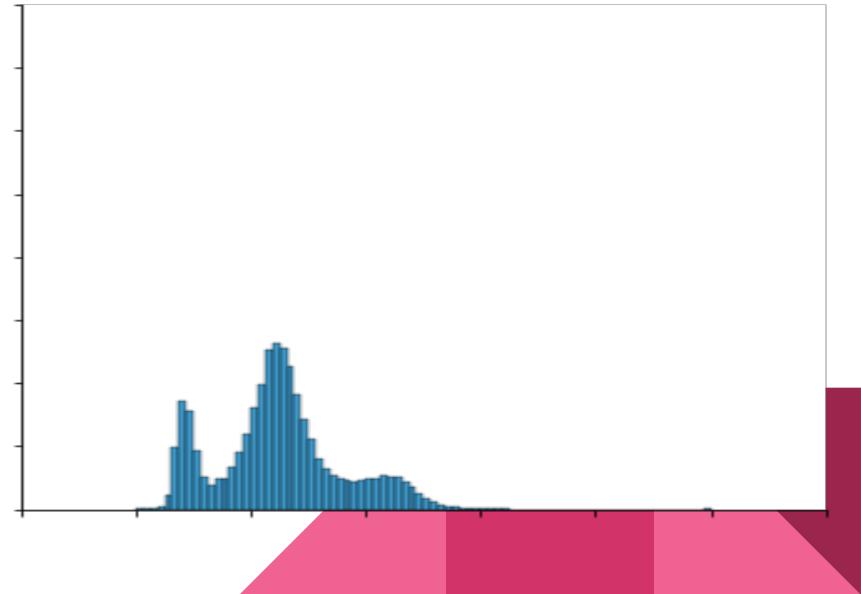
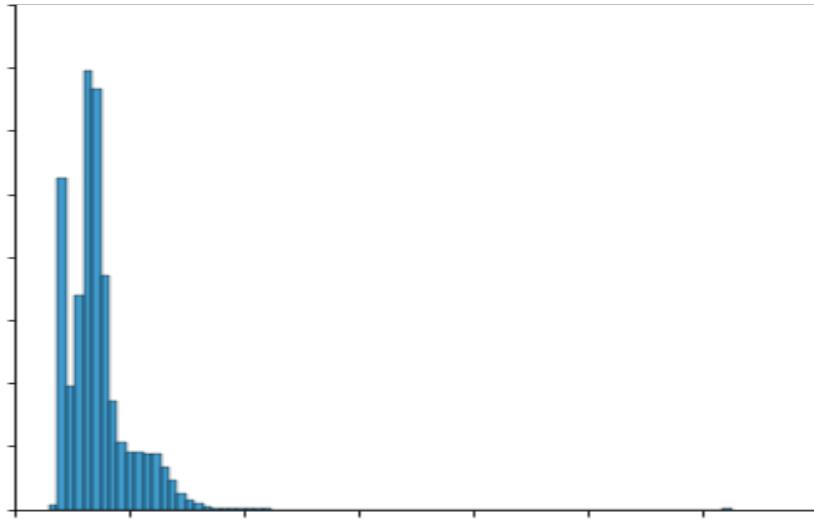
Image thresholding

Transformations to the image can make it easier to threshold

“Log transform before thresholding?” selected?

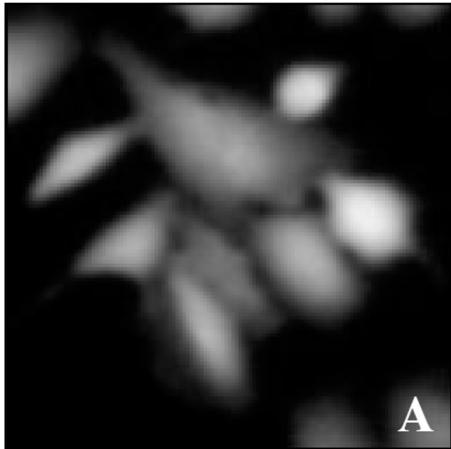
No

Yes

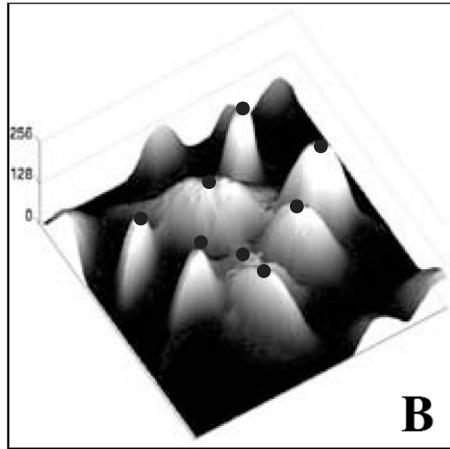


Separating touching objects

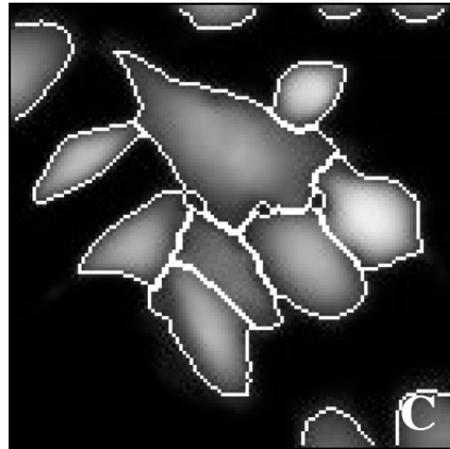
- We next need to distinguish multiple objects contained in the same “clump”



A



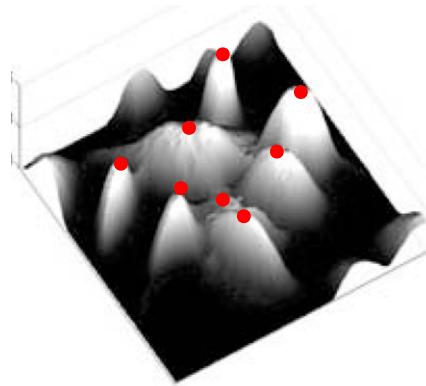
B



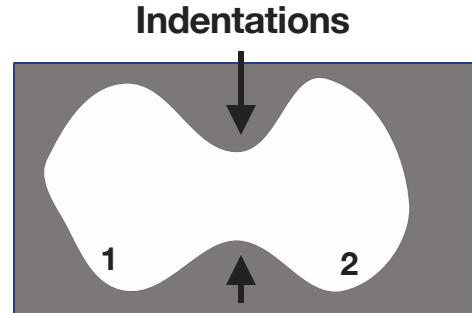
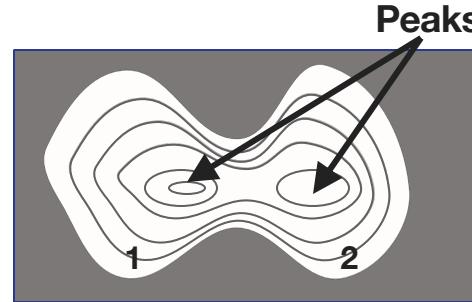
C

Separating touching objects – clump identification

- **Intensity:** Works best if objects are brighter at center, dimmer at edges



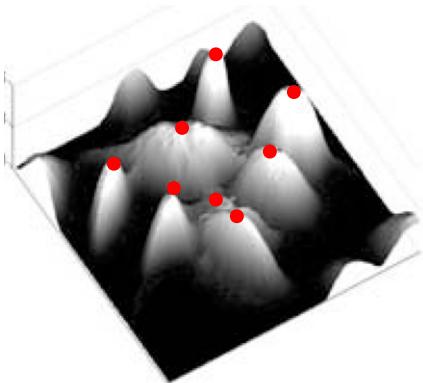
- **Shape:** Works best if objects have indentations where clumps touch (esp. if objects are round)



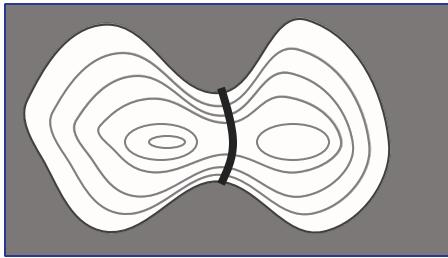
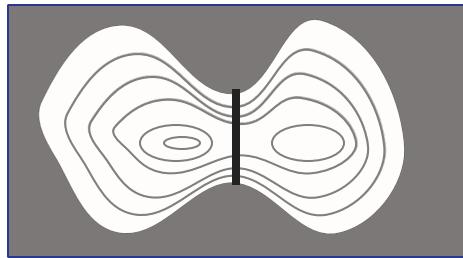
Separating touching objects

- Drawing boundaries: Two options

- **Shape:** Draws boundary lines at the indentations

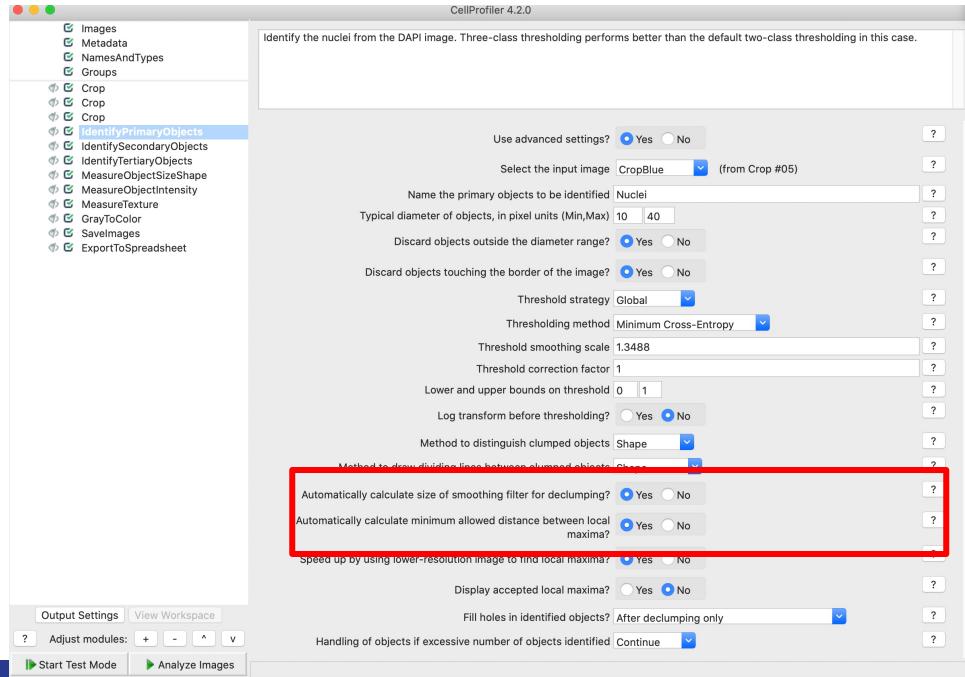


- **Intensity:** Draws boundary lines at dimmest line between objects



Primary object identification

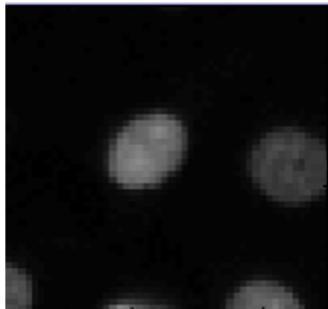
Adjusting the sizes of our expected objects



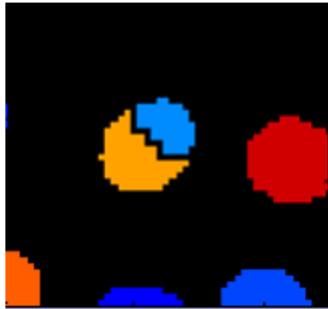
Automatically calculate size of smoothing filter for declumping? Yes No

Automatically calculate minimum allowed distance between local maxima? Yes No

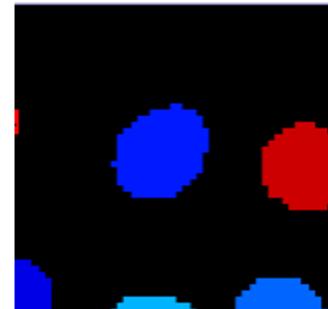
Separating touching objects



Original image



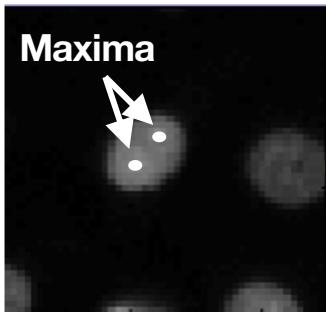
Smoothing filter
size = 4



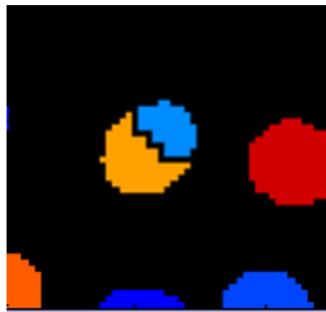
Smoothing filter
size = 8

- . Smoothing: Increase to reduce intensity irregularities which produce over-segmentation of objects

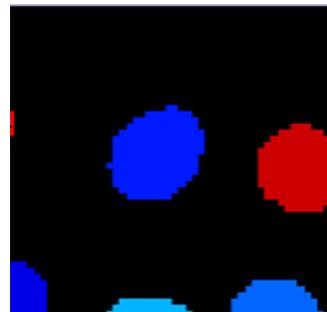
Separating touching objects



Original image



Maxima
distance = 4

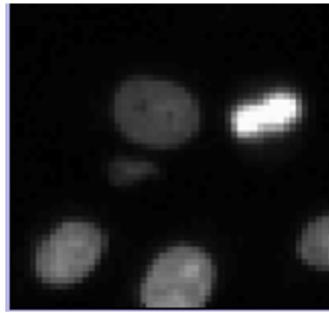


Maxima
distance = 8

- Suppress Local Maxima
- Smallest distance allowed between object intensity peaks to be considered one object rather than a clump
- Decrease to reduce improper merging of objects in clumps

Separating touching objects

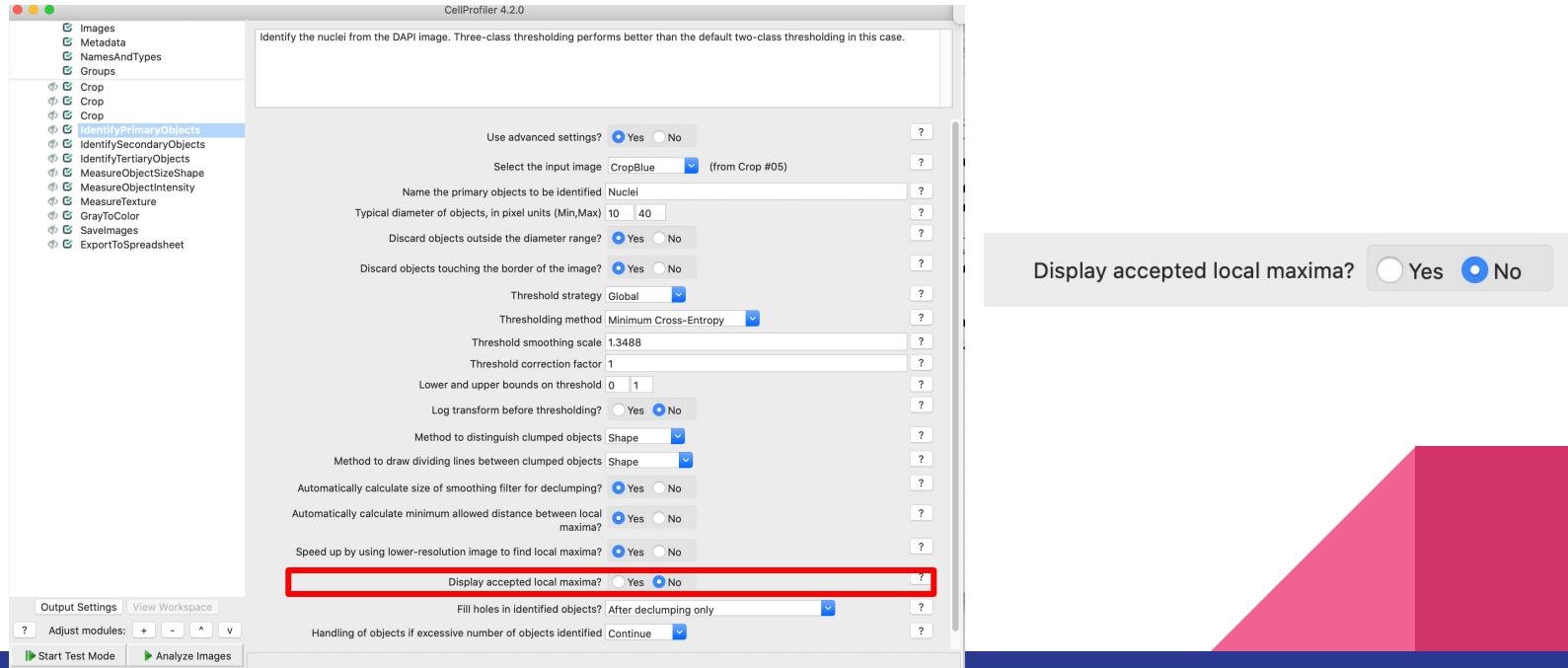
However....



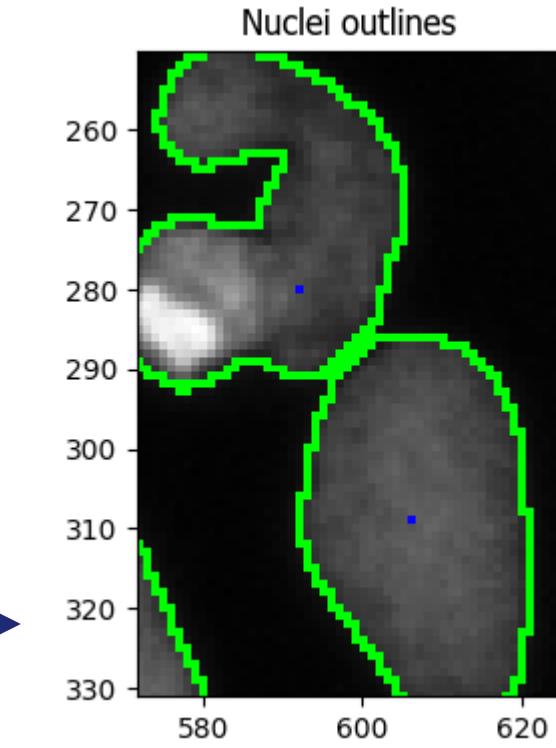
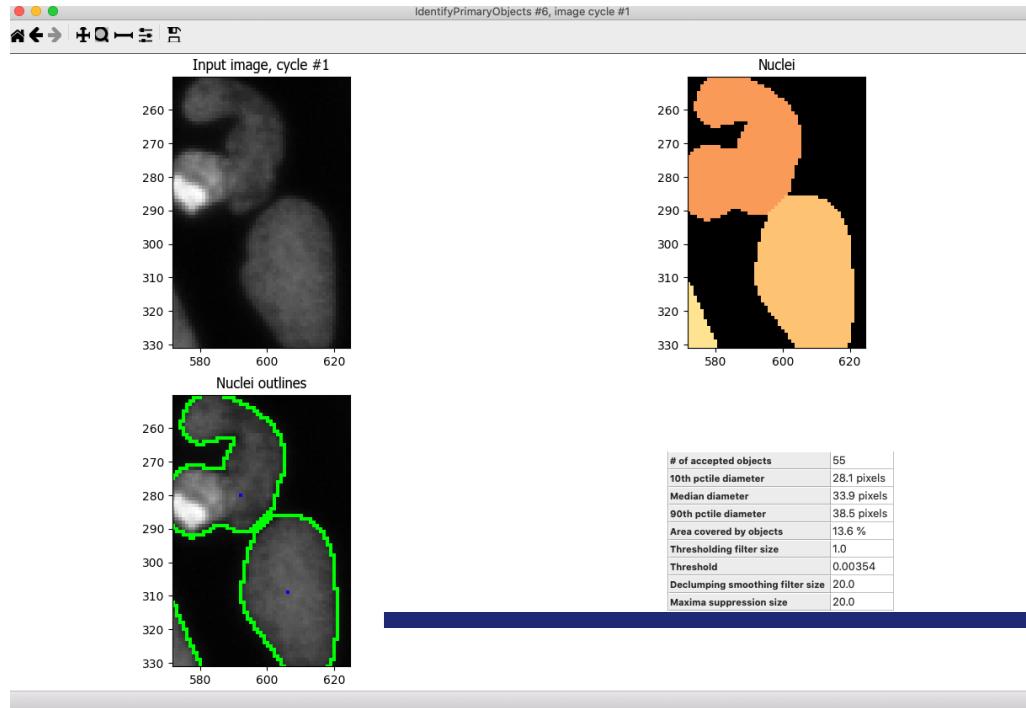
- . Over-adjusting can produce more improper segmentation than it solves
- . The proper settings are usually a matter of trial and error, and trying to globally maximize good segmentations rather than tuning to one image

Primary object identification

Maxima can be displayed to assist in creating objects

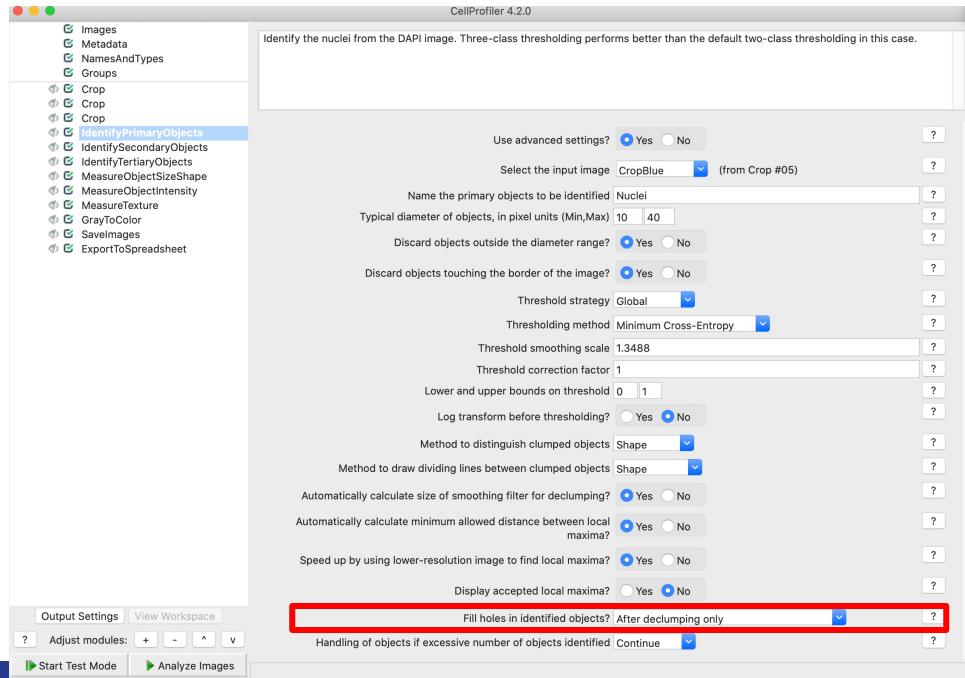


Primary object identification



Primary object identification

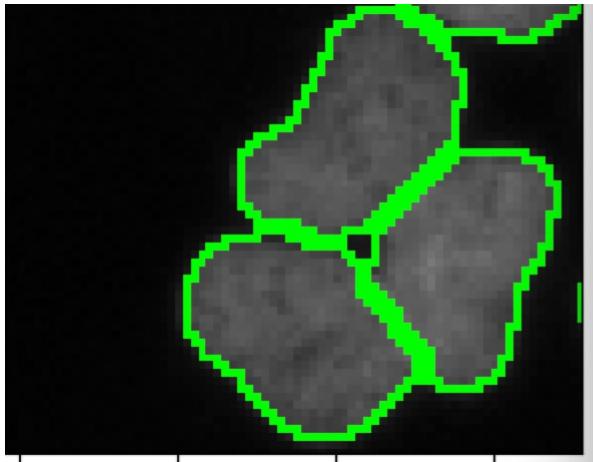
When if at all should you fill holes in the objects?



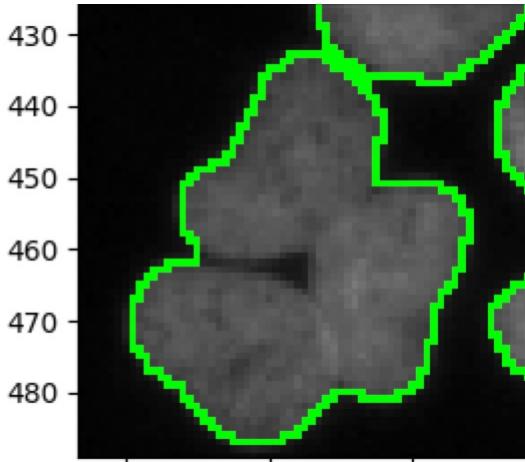
Fill holes in identified objects? After declumping only

Primary object identification

When if at all should you fill holes in the objects?



Hole filling “After declumping only”



Hole filling “After both
thresholding and declumping”

Secondary object identification

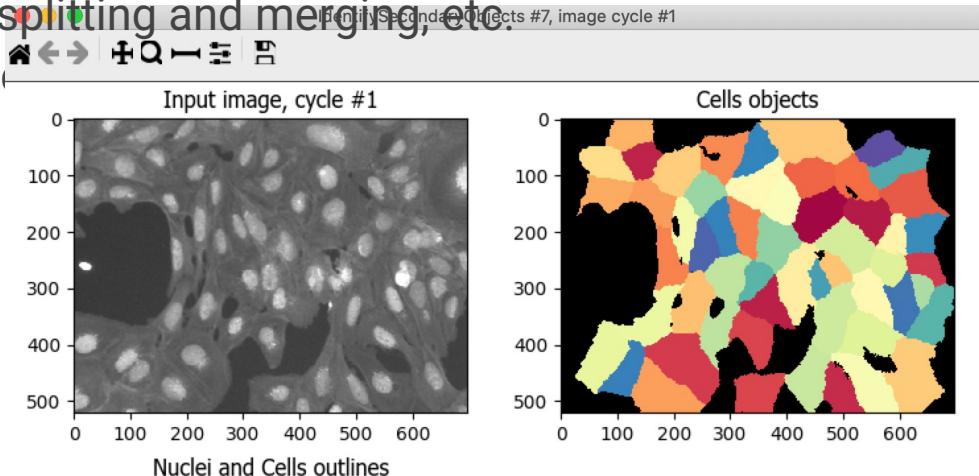
Many options for thresholding, splitting and merging, etc.

- Make a second object base

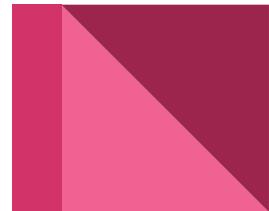
Our colors mean different things here!

Green = the input primary object

Pink = the secondary object

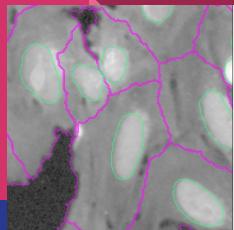
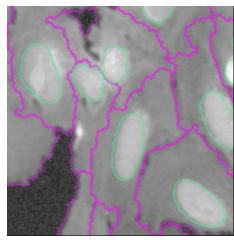
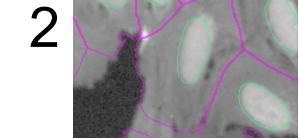
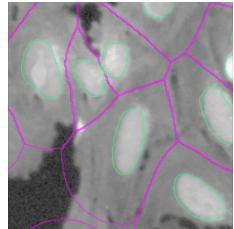


Threshold	0.003
10th pctlle diameter	58.6 pixels
Median diameter	73.6 pixels
90th pctlle diameter	91.9 pixels
Thresholding filter size	0.0
Area covered by objects	72.3 %



Secondary object identification

- A few ways to do it
- Non-image-based:
 - Distance-N – draw a circle a fixed width around the primary object (1)
- Image-based- all start by thresholding the image, then:
 - Distance-B – draw a circled a fixed width around the primary object, but only into areas that are threshold positive (2)
 - Watershed-Image and Watershed-Gradient – fill the thresholded area in a watershed, either based on the original pixel data or a transform of it (3)
 - Propagation- do watershed image, but balance those lines against trying to make each cell the same size (4)

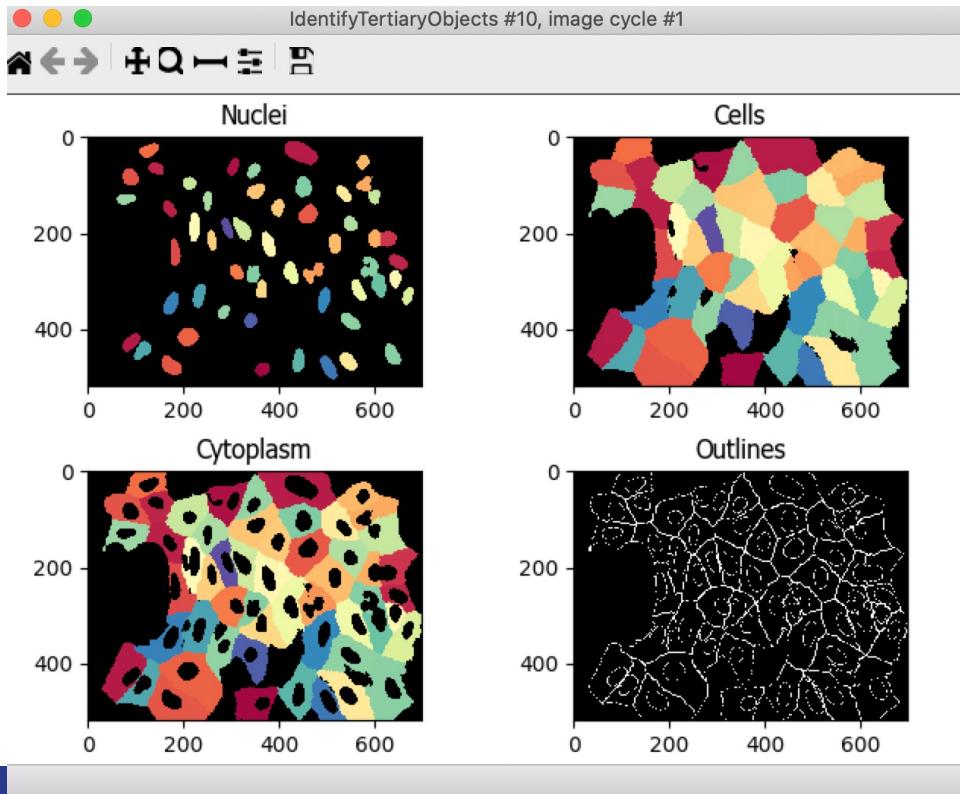


Tertiary object identification

Make a third object out of two others

The third object should be the differences of the first two

The second object should therefore also be derived from the first object (such as by IdentifySecondaryObjects or ExpandOrShrinkObjects)



Object processing modules in CellProfiler

(Not exhaustive, just inspirational)

- In the Object Processing category:
 - ExpandOrShrinkObjects
 - FilterObjects
 - MaskObjects
 - RelateObjects
 - EditObjectsManually
 - CombineObjects
 - ShrinkToObjectCenters
- In the “Advanced” category:
 - DilateObjects
 - ErodeObjects

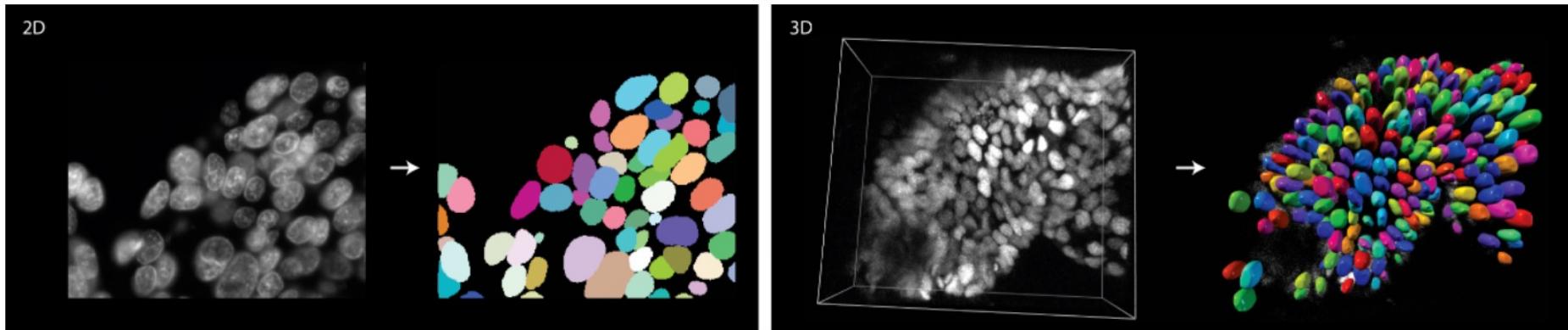


How can we push this
further?

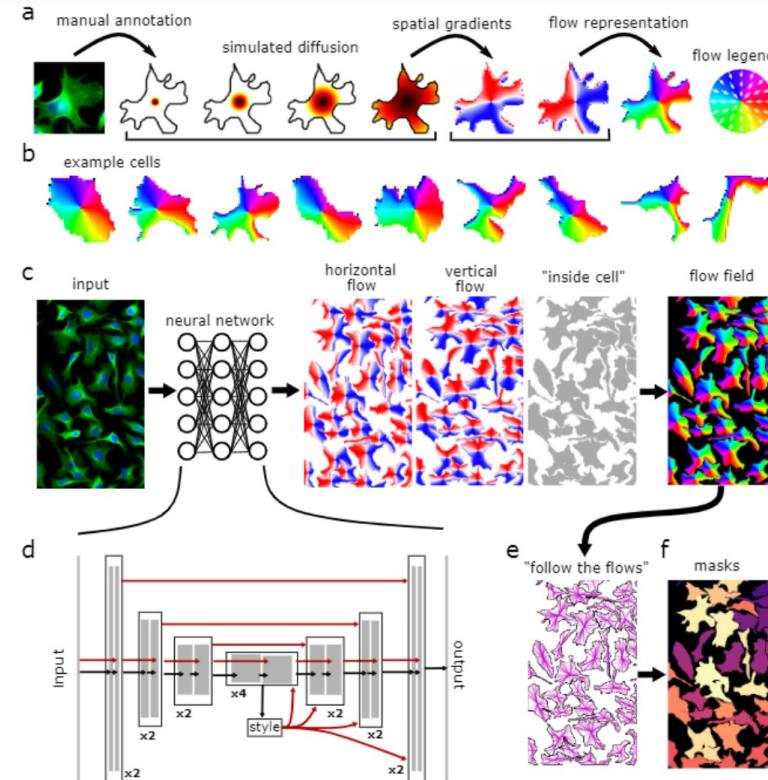
Are there ways we can
better find objects?

Stardist

StarDist - Object Detection with Star-convex Shapes

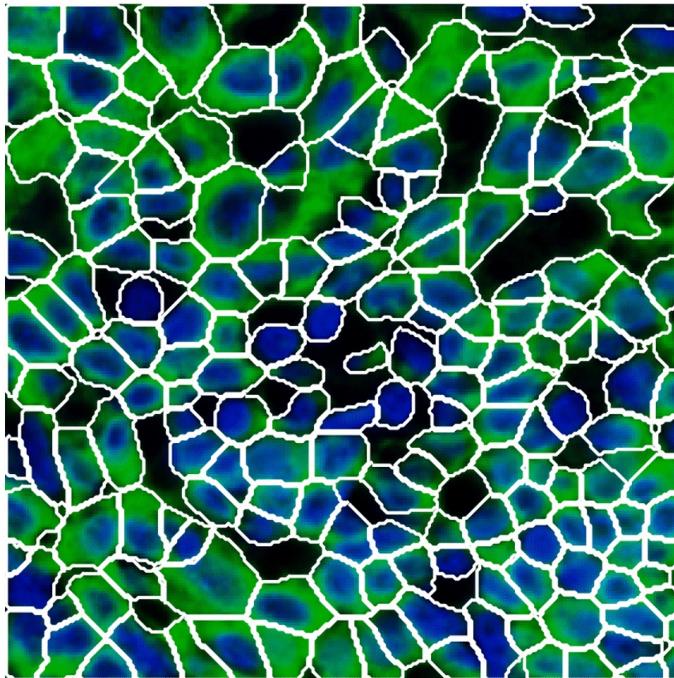


Cellpose

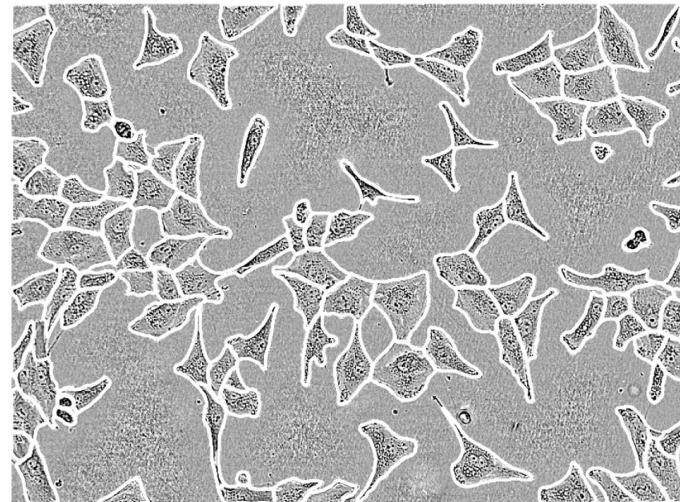


Cellpose 2

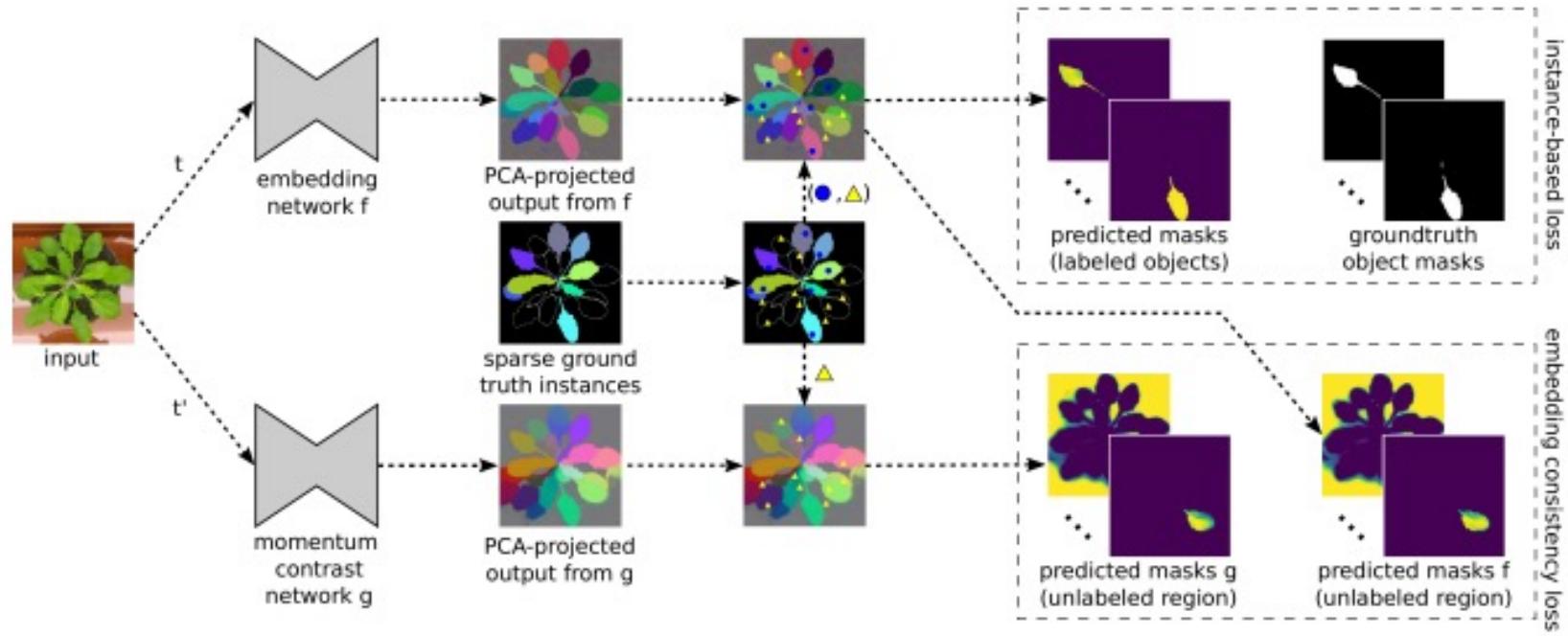
Cellpose 2.0 (finetuned)



Cellpose 2.0 (finetuned)



Future models may not need so much ground truth - SPOCO



Tips for creating a good high content analysis workflow

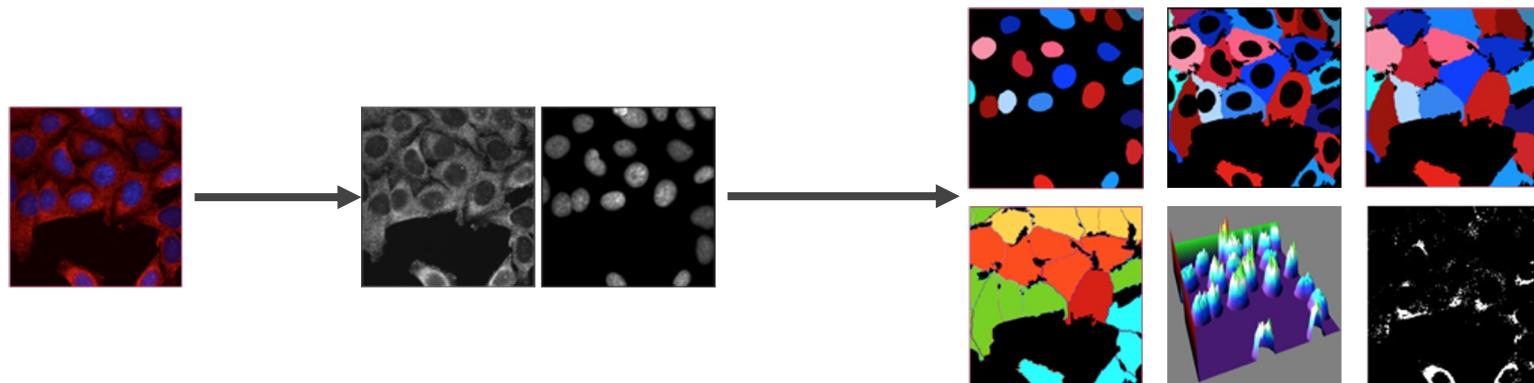
- When finding the objects that you care about, ask yourself for your whole experiment:
 - Do I generally agree with most of the object segmentations from my analysis workflow?
 - Do I have an approximately equal number of regions/images where the threshold chosen by the algorithm for this image is a bit too low vs a bit too high?
 - Do I have an approximately equal number of oversegmentations/splits and undersegmentations/merges?
 - Very important: Do both the second and third bullet points hold true for both my negative control images and my positive control (or most extreme expected phenotype(s) sample) images?

We've found our objects- now what?



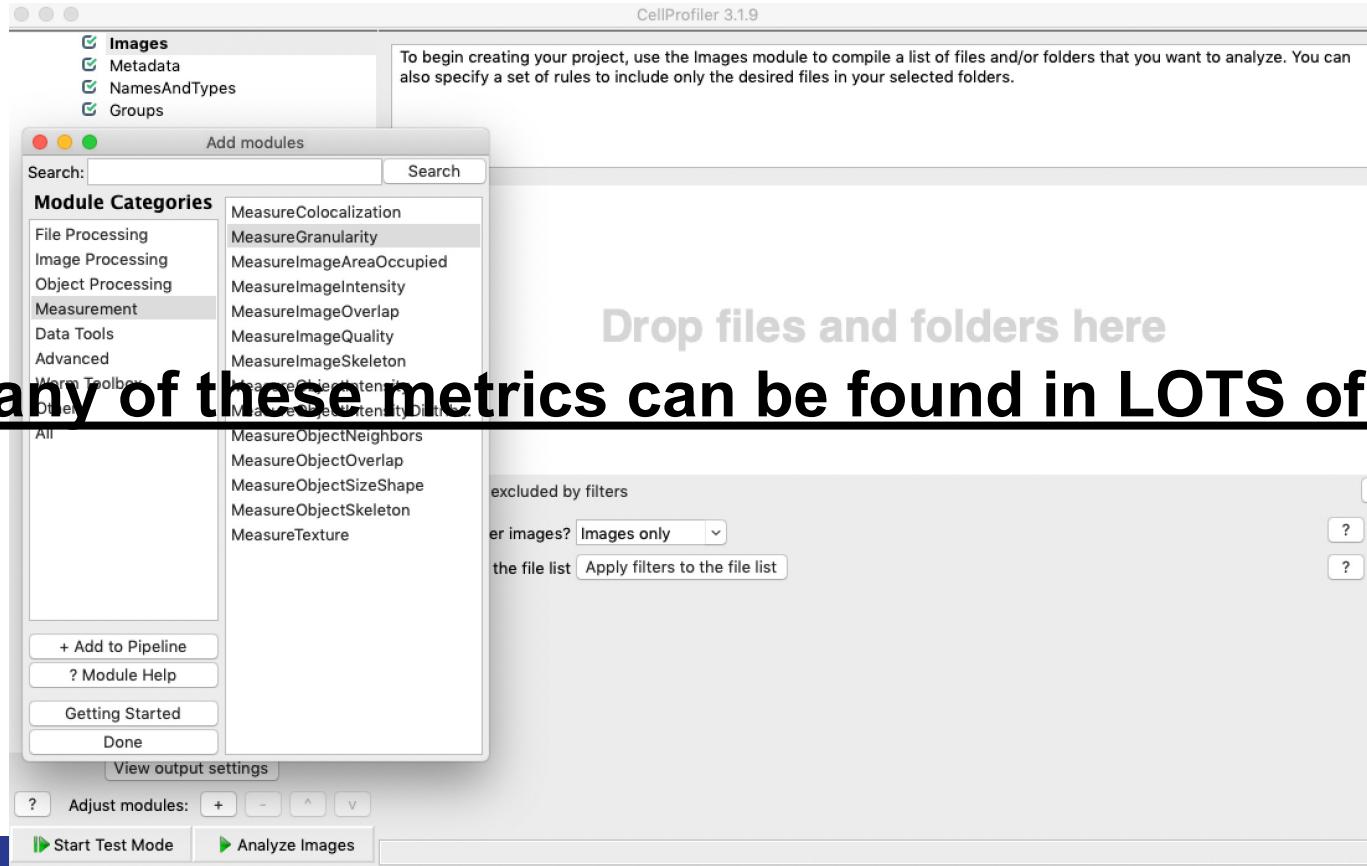
Feature extraction

- How do you get hundreds of measurements per cell?



Counts, Shapes, Sizes, Intensities, Textures, Correlations, Relationships

Software can help us build up big measurement suites for our images



Many of these metrics can be found in LOTS of tools

Objects can be measured for their... intensity

- Integrated intensity: Sum of the pixel intensities within an object
- Mean, median, standard deviation intensities
- Maximal and minimal pixel intensities
- Lower/Upper quartile of the intensity
- Object intensities may be measured from any channel, not just the channel used to identify the object
 - Example: GFP intensity may be measured using nuclei objects identified with DAPI

Objects can be measured for their... intensity distribution

Calculate intensity Magnitudes and phase

Maximum zernike moment 9

Select an image to measure image_input (from NamesAndTypes)
Add another image

Select objects to measure Cell (from IdentifySecondaryObjects)
Object to use as center? These objects
Add another object

Scale the bins? Yes No

Number of bins 4

Scale the bins? Yes No

Number of bins 4

Maximum radius 35

Magnitude: Amount of intensity within each ring

Phase (Zernike moment): Distribution of intensity within each ring

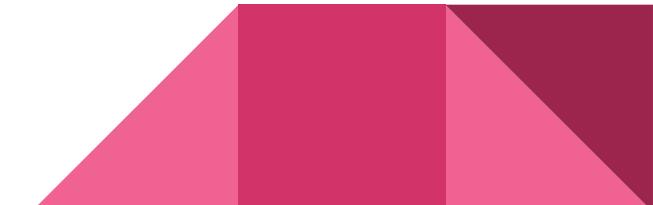
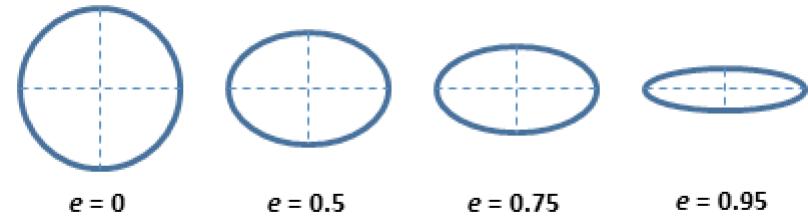
E.g. one side of the cell is brighter

Within each fraction/ring:

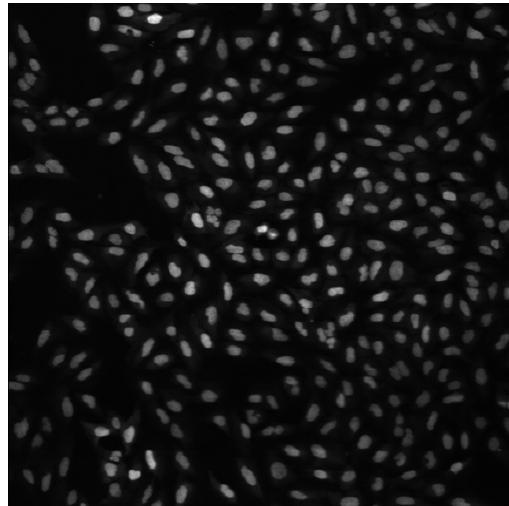
- *FracAtD*: total intensity
- *MeanFrac*: mean intensity
- *RadioCV*: divide the ring into 8 slices, measure the coefficient of variation

Objects can be measured for their... size and shape

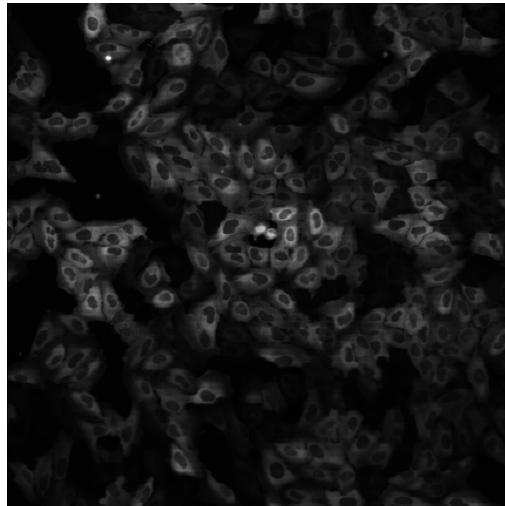
- Area
- Perimeter
- Eccentricity (circle = 1, line = 0)
- MajorAxisLength
- MinorAxisLength
- Orientation
- Solidity



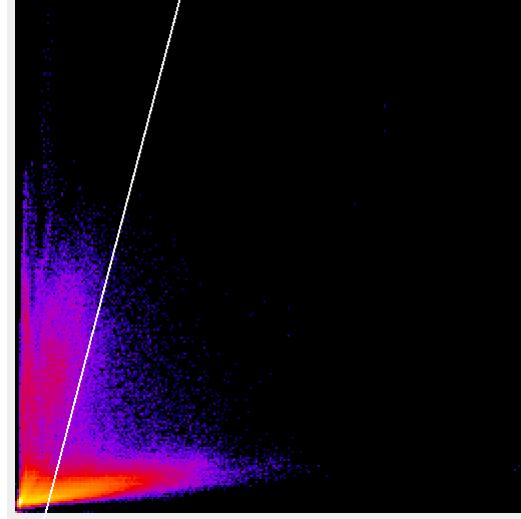
Objects can be measured for their... colocalization



DAPI



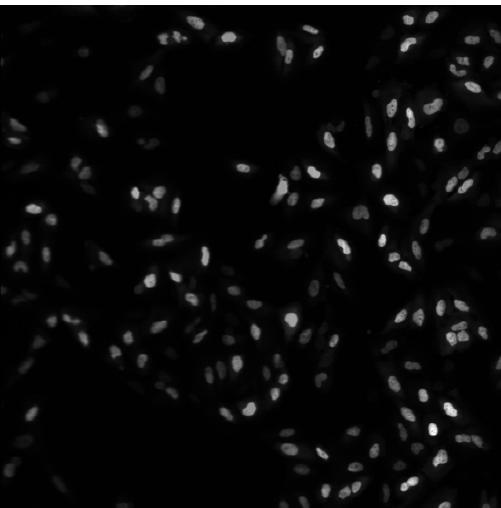
GFP



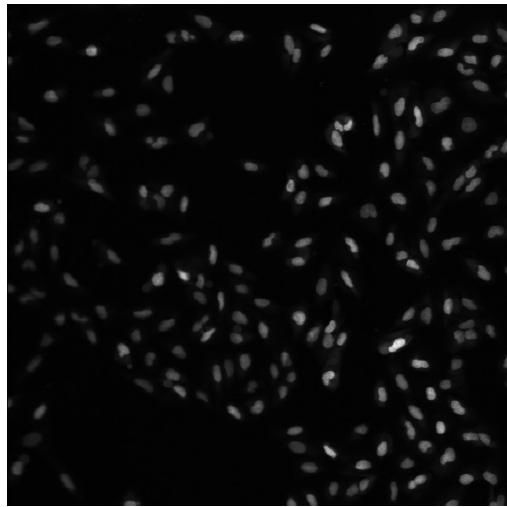
FIJI

- FIJI, Coloc2- https://imagej.net/Coloc_2

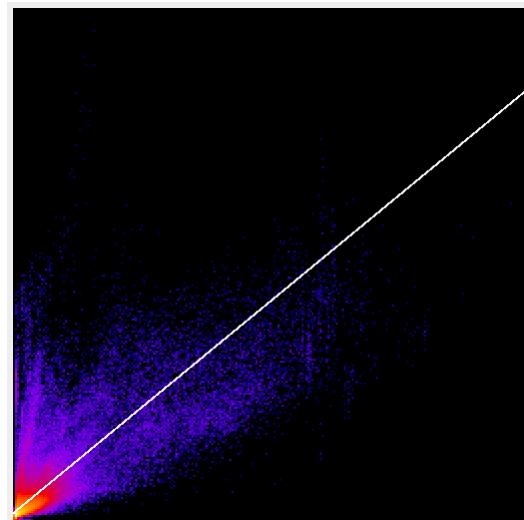
Objects can be measured for their... colocalization



DAPI



GFP



- FIJI, Coloc 2-https://imagej.net/Coloc_2

Information is hidden in images



actual phase:

G1

S

G2

Prophase

Metaphase

Anaphase

Telophase

	G1	S	G2	Prophase	Metaphase	Anaphase	Telophase
G1	0.92	0.07	0.01	0.00	0.00	0.00	0.00
S	0.23	0.66	0.11	0.00	0.00	0.00	0.00
G2	0.01	0.08	0.90	0.01	0.00	0.00	0.00
Prophase	0.00	0.04	0.38	0.58	0.01	0.00	0.00
Metaphase	0.00	0.01	0.13	0.02	0.82	0.02	0.00
Anaphase	0.00	0.00	0.11	0.01	0.03	0.83	0.02
Telophase	0.02	0.05	0.02	0.00	0.00	0.01	0.91

predicted phase



Minh
Doan

Claire
McQuin

Holger Hennig,
Rostock U

Paul Rees,
Swansea U

Huw Summers,
Swansea U

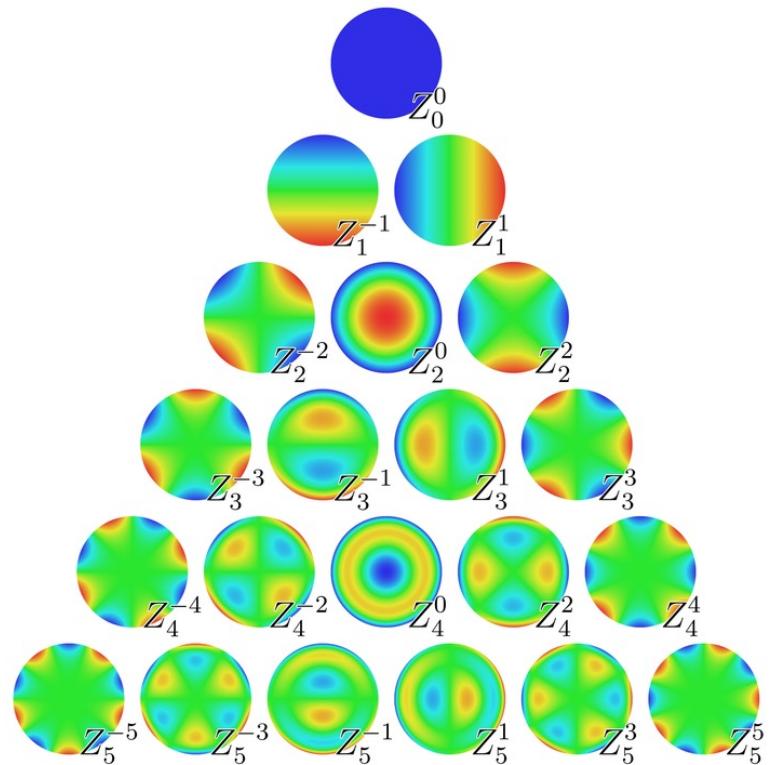
Andrew Filby,
Newcastle U

Fabian Theis,
Helmholtz Zentrum, München

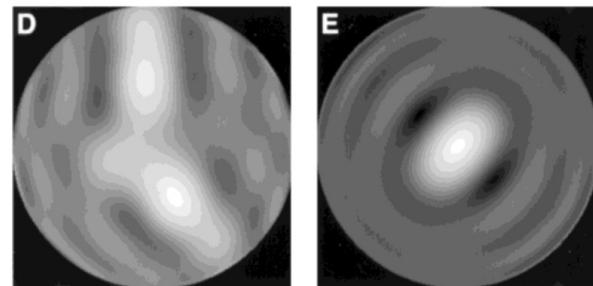
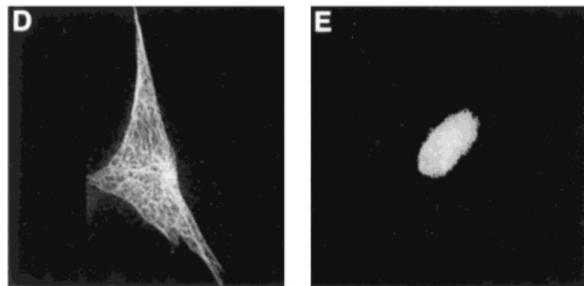
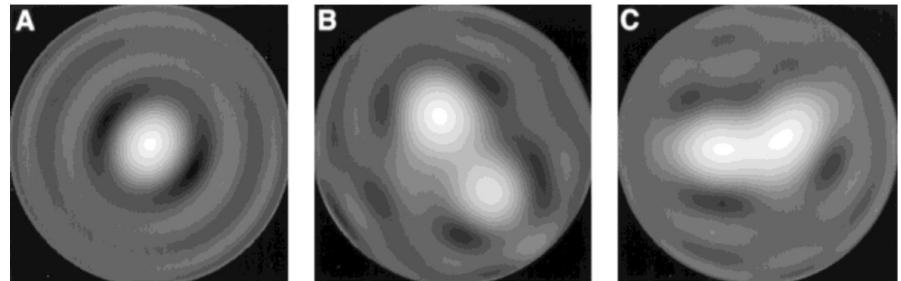
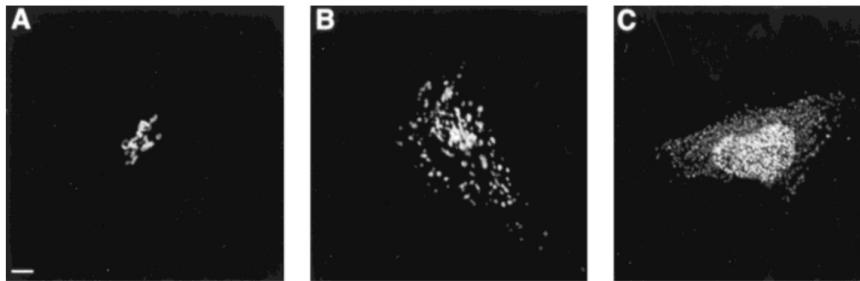
What if we could use the
information that we can't
see?

Objects can be measured for their... Zernike Polynomials

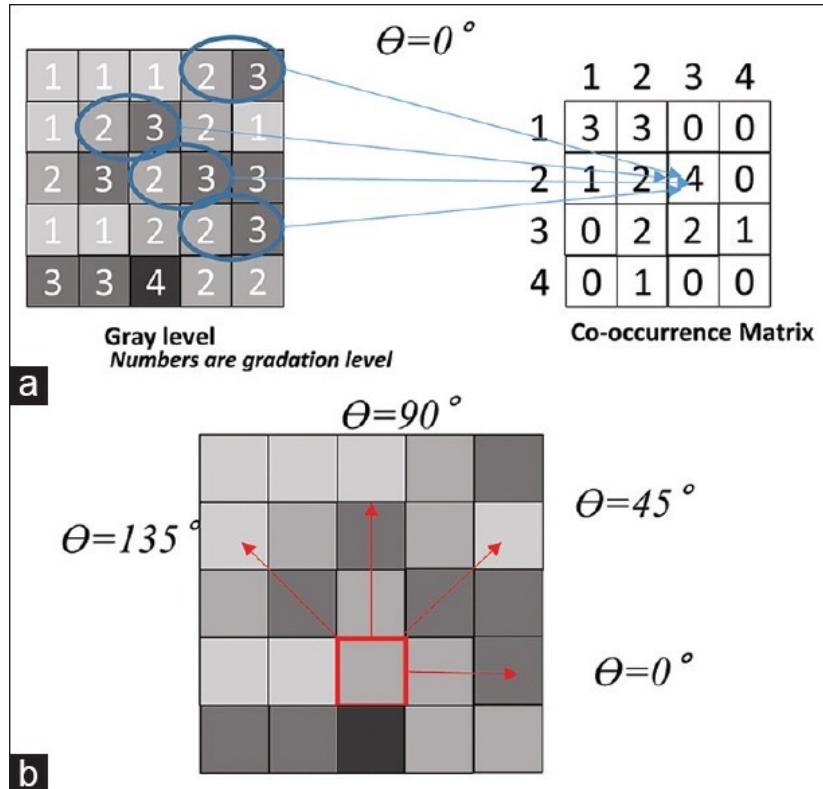
Zernike Polynomials



Zernike polynomials can give accurate descriptions of an object's shape



Objects can be measured for their... texture



File utilities in CellProfiler

■ Savelimages

- A variety of file types, bit depths, compression options

■ ExportToSpreadsheet

- Saves as CSV or TSV

- Can customize which measurements to save

■ ExportToDatabase

- Saves to SQLite or MySQL

- Can optionally make a properties file for CellProfiler Analyst (more later)

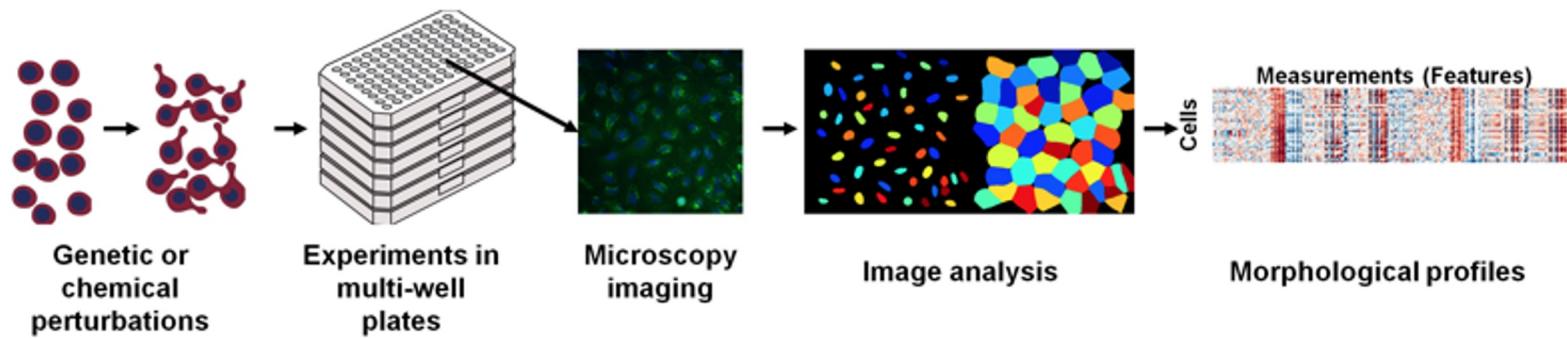
Running on large image sets on CellProfiler

- A few – a few hundred images
 - Can likely run on your local machine
 - CellProfiler will automatically multithread process up to your number of CPUs
- A few hundred – a few tens of thousands of images
 - Talk to your local sysadmin about running on a cluster (directly or with [Docker](#))
 - Check out our [instructions on getting started](#)
- A few tens of thousands – a few million images
 - Can consider cloud processing
 - Check out our [Distributed-CellProfiler package](#) for running on AWS

“The thing I want to do doesn’t exist in CellProfiler!”

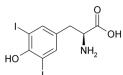
- Are you sure it doesn’t?
 - Search the help, and/or post on image.sc (more later)
- Is it an image processing utility that exists in ImageJ/Fiji?
 - Try out the RunImageJMacro module, which can point to your system ImageJ/Fiji
- Go ahead and write your own!
 - We have [templates to expand](#) and a [video on how to do it](#)

Morphological Profiling



Morphological profiling can predict small molecule activity

Treat cells
with compounds



Imaging

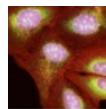


Image
analysis



Match
morphological
profiles



Vebjorn
Liosa

Rob ter
Horst

Actual class

Actin disruptors	Act
Aurora kinase inhibitors	Aur
Cholesterol-lowering	Ch
DNA damage	DD
DNA replication	DR
Eg5 inhibitors	Eg5
Epithelial	Epi
Kinase inhibitors	KI
Microtubule destabilizers	MD
Microtubule stabilizers	MS
Protein degradation	PD
Protein synthesis	PS

Class predicted by morphology

	Act	Aur	Ch	DD	DR	Eg5	Epi	KI	MD	MS	PD	PS
Act	3					1					1	
Aur		7			1					4		
Ch			6									
DD				8	1							
DR					4	3						
Eg5						6				6		
Epi				2			6					
KI								5				
MD					1		2		9	2		
MS									4	4	1	
PD											4	1
PS												8

Acc.

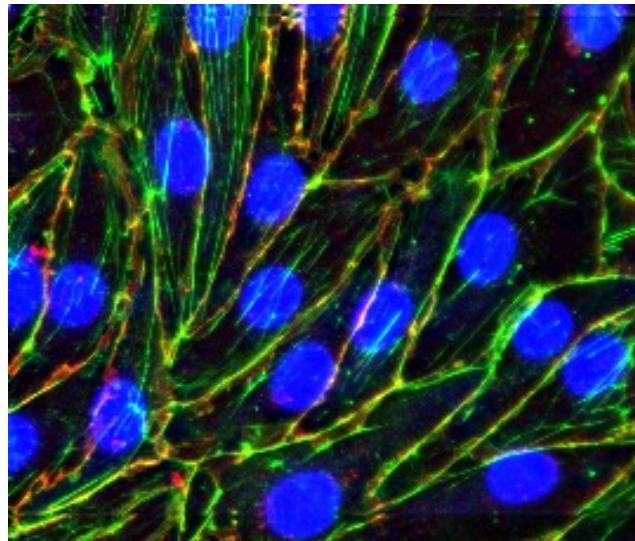
60 %
61 %
99 %
84 %
39 %
48 %
75 %
100 %
65 %
39 %
63 %
99 %

Overall accuracy: 67 %

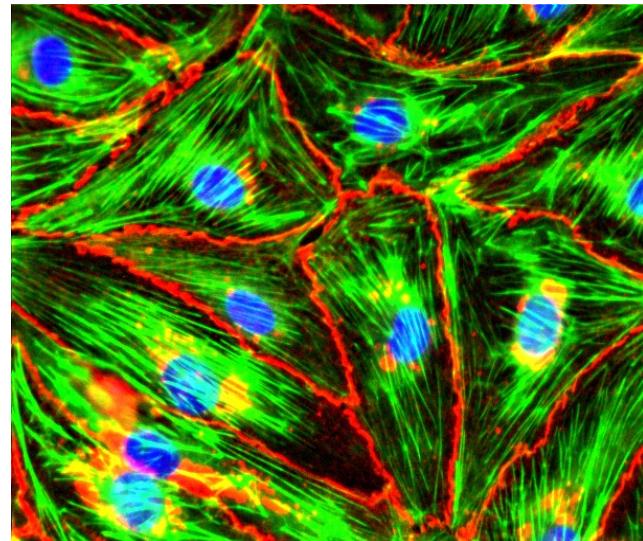
Morphological profiling can identify drugs for disease

+
drug?

Healthy

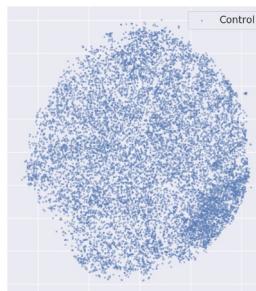


Disease (CCM knockdown)

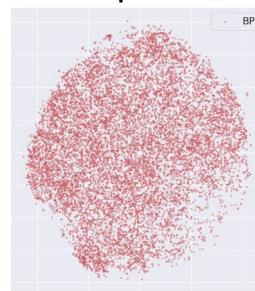


Disease states can be morphologically distinct

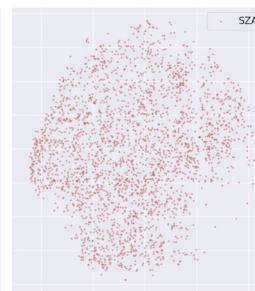
Controls



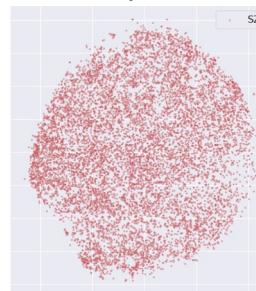
Bipolar



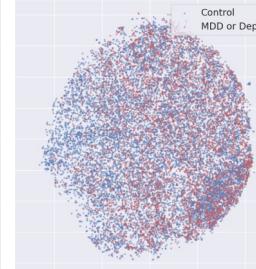
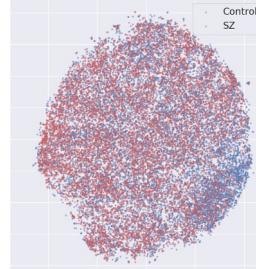
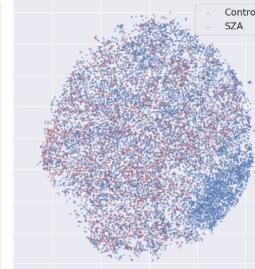
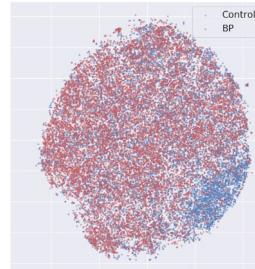
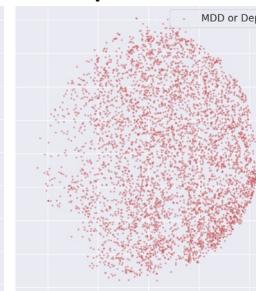
Schizoaffective



Schizophrenia



Depression



Mohammad
Rohban



Kyle
Karhofs



Marzieh
Haghghi



Bruce
Cohen,
McLean
Hospital



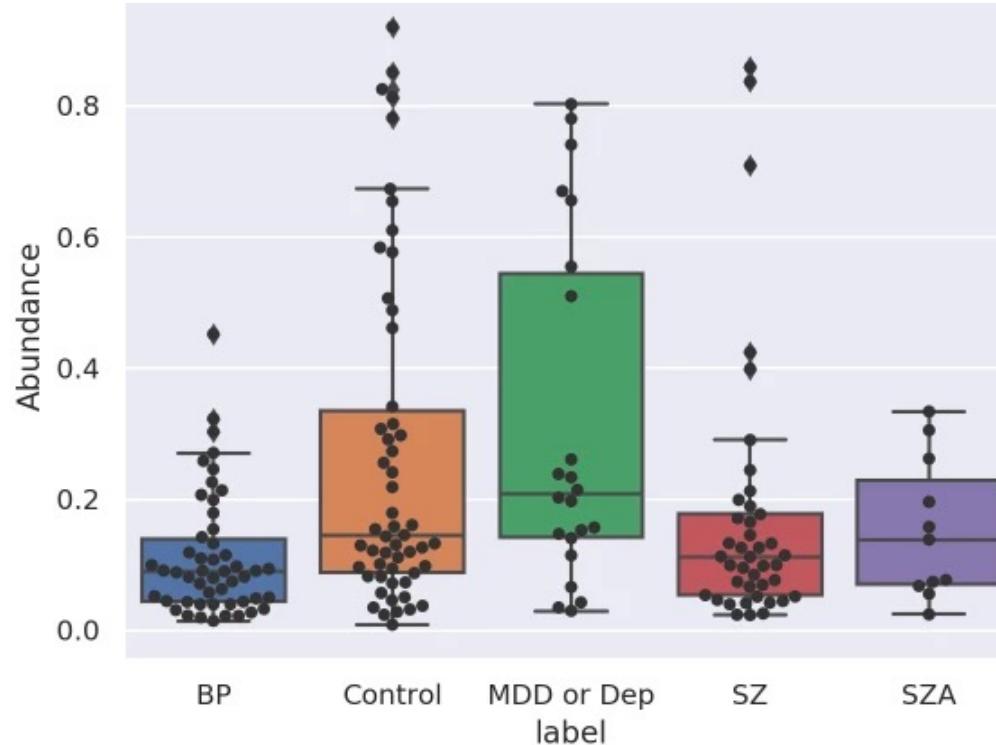
Donna
McPhie



Rakesh
Karma-
charya

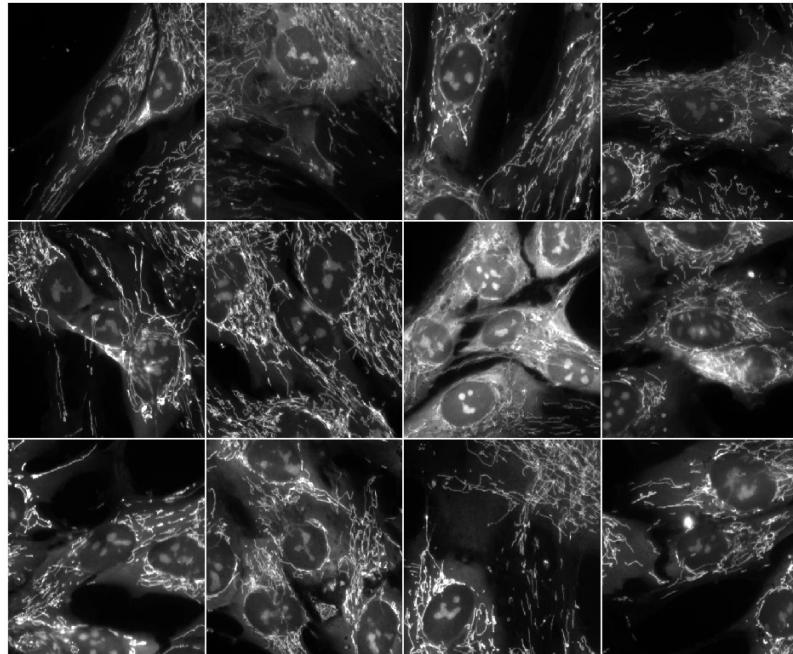
Disease states can be morphologically distinct

Even in subtle ways not obvious to the naked eye

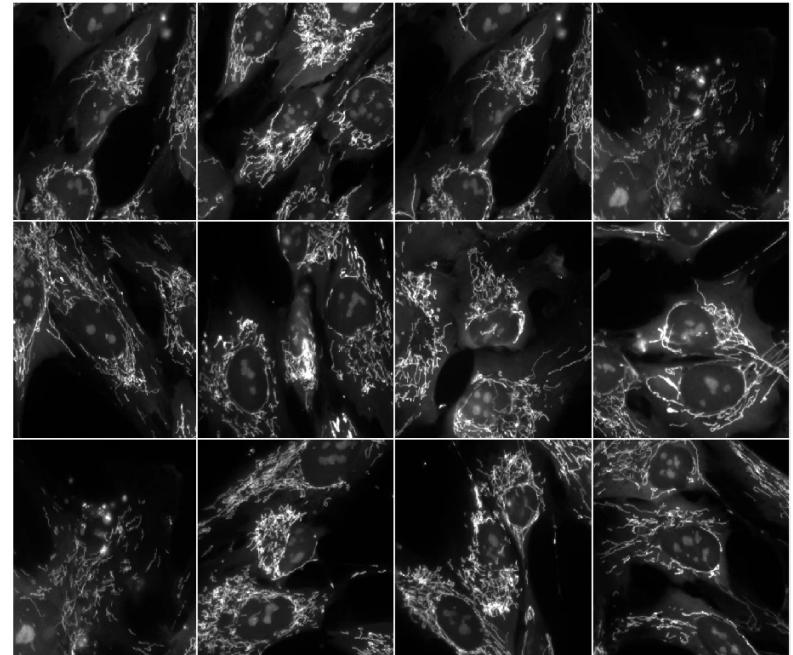


Disease states can be morphologically distinct

Even in subtle ways not obvious to the naked eye



High Scorers



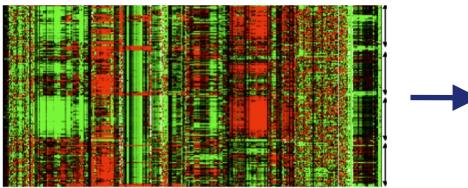
Low Scorers

CellProfiler-Analyst: overview



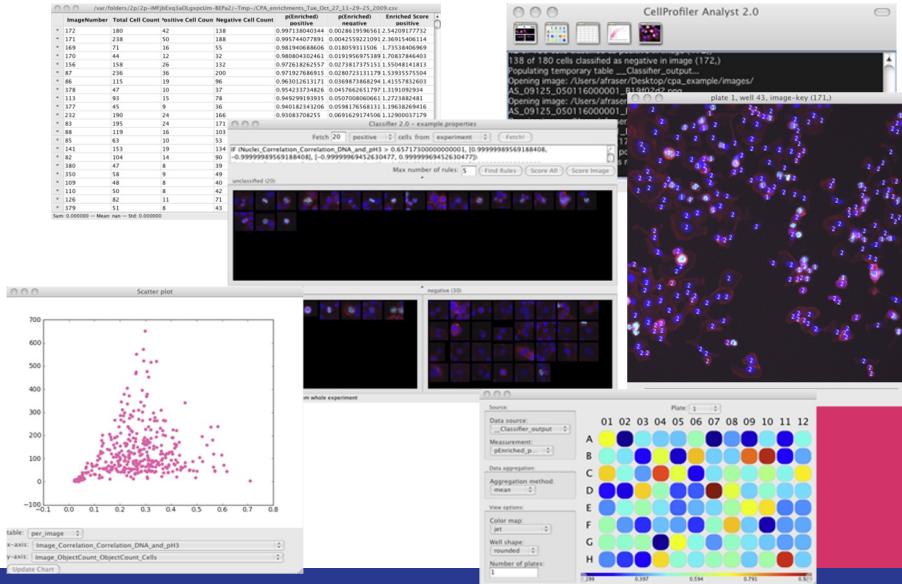
CellProfiler Analyst™

data exploration software



Goal: Provide the user with a powerful suite of image exploration and machine learning methods

- Explore** data large sets of images
- Identify** interesting subpopulations and see the original images
- Identify** interesting phenotypes automatically





How can I learn how
to do this stuff?

Where can I go for
help?

forum.image.sc - Open scientific community forum for bioimage analysis and beyond

▼ Community Partners

All Topics	AGAVE	AICSImageIO	Aydin	BIAFLows	BIII
BIO	BiofilmQ	Bio-Formats	BoneJ	BrainGlobe	Celpose
CellProfiler	CLIJ	CytoData	CytoMAP	Cytomine	DeepLabCut
Fiji	FLIMLib	GerBL	Icy	IDR	ilastik
ImageJ	ImageJ2	ImagePy	ImgLib2	ImJoy	JIPipe
LOBSTER	Mantis Viewer	Mars	MCMICRO	MIA	MIB
μManager	MiTBo	MorphoGraphX	MorphoNet	napari	NEUBIAS
OME	OMERO	OpenIRIS	OpenSPIM	Orbit	Piximi
Py-EM	PYME	Python-Microscope	QuPath	SBEMimage	Scenery
SCIFIO	scikit-image	SciView	Segmenter	SR-Tesseler	StarDist
starfish	SuRVoS	TissUUmmaps	vedo	webKnossos	ZeroCostDL4Mic
... Your Icon Here					

RELATED COMMUNITIES   

SPONSORED BY COBA 

Center for Open Bioimage Analysis

Openbioimageanalysis.org



Tech R&D Projects



Driving Biological Projects



Community Engagement

Gratitude



Cimini Lab members

Beth Cimini
Mario Costa Cruz
Barbara Diaz-Rohrer
Teresa Gao
Melissa Gillis
Nodar Gogoberidze
Serena Larew
Becki Ledford
Levin Moser
Rebecca Senft
Callum Tromans-Coia
Erin Weisbart

IMAGING PLATFORM

Carpenter-Singh Lab members

Anne Carpenter
Shantanu Singh
John Arevalo
Niranj Chandrasekaran
Marzieh Haghghi
Yu Han
Alexander Kalinin
Serena Larew
Becki Ledford
Robert van Dijk



CellProfiler™
cell image analysis software



CellProfiler Analyst™
data exploration software

Many thanks to our
many biology collaborators

Recent major funding for this work provided by:

- CZI Imaging Scientist Fellowship
- NIH NIGMS: MIRA R35 GM122547
- CZI Software Fellows program
- NIH NIGMS: P41 GM135019