

# High Content 2016

September 12th-14th  
3rd Annual Conference

Joseph B. Martin Conference Center at Harvard Medical School, Boston, MA

## Advanced Tools for Data Analysis: CellProfiler

Mark-Anthony Bray, Ph.D  
Novartis Institutes of Biomedical Research  
Cambridge, Massachusetts, USA  
mark.bray@novartis.com

**START HERE:** <https://github.com/braymp/sbi2>

# Overview and Requisites

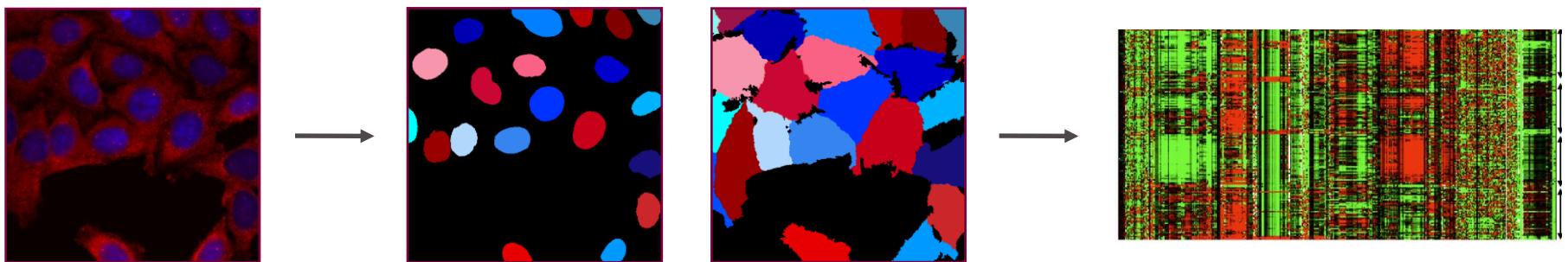
---

- Agenda: Hands-on demo using **CellProfiler** for high-content analysis
- This session assumes that you either:
  - Attended the Introductory HCS image analysis session
  - Have a good working knowledge of the basics of image analysis

# CellProfiler: Overview



- **Process** large sets of images
- **Identifies and measures** objects
- **Export** data for further analysis

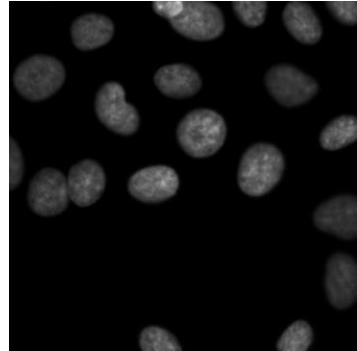


**Goal:** Provide powerful image analysis methods with a user-friendly interface

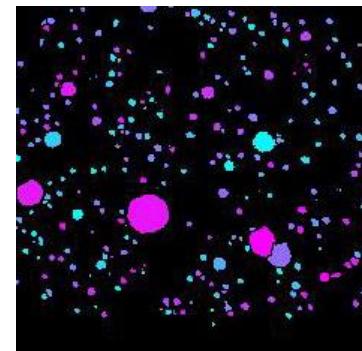
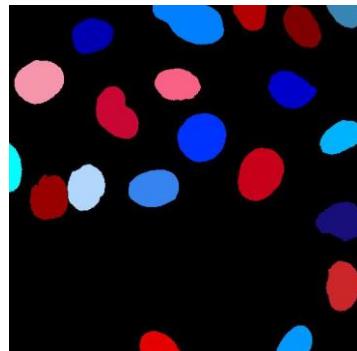
**Philosophy:** Measure everything, ask questions later

# Typical CellProfiler Pipeline Workflow

- For image-based assays, the basic objective is always to
  - Identify cells/organisms
  - Measure feature(s) of interest

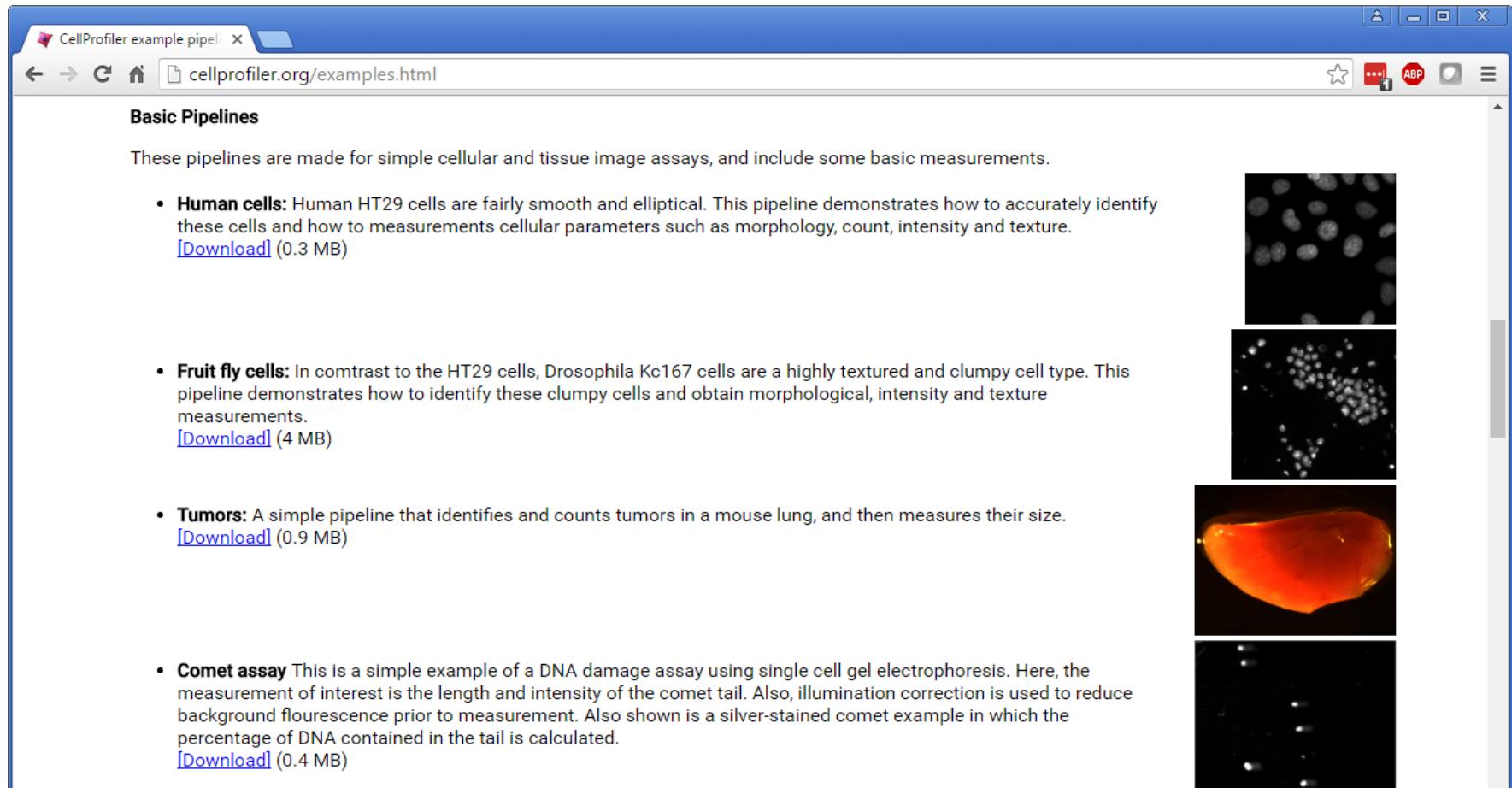


- The uniqueness of each assay comes in
  - Deciding what compartments to identify and how to identify them
  - Determining which measure(s) are most useful to identify interesting samples

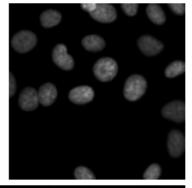
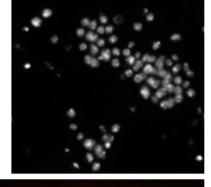
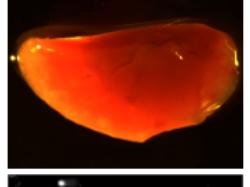


# Typical Starting Point: An Existing Pipeline

Go to [www.cellprofiler.org/examples.html](http://www.cellprofiler.org/examples.html) for many examples



The screenshot shows a web browser window titled "CellProfiler example pipeline". The address bar contains "cellprofiler.org/examples.html". The main content area is titled "Basic Pipelines" and describes simple cellular and tissue image assays. It lists four pipeline examples:

- Human cells:** Human HT29 cells are fairly smooth and elliptical. This pipeline demonstrates how to accurately identify these cells and how to measure cellular parameters such as morphology, count, intensity and texture. [\[Download\]](#) (0.3 MB) 
- Fruit fly cells:** In contrast to the HT29 cells, Drosophila Kc167 cells are a highly textured and clumpy cell type. This pipeline demonstrates how to identify these clumpy cells and obtain morphological, intensity and texture measurements. [\[Download\]](#) (4 MB) 
- Tumors:** A simple pipeline that identifies and counts tumors in a mouse lung, and then measures their size. [\[Download\]](#) (0.9 MB) 
- Comet assay:** This is a simple example of a DNA damage assay using single cell gel electrophoresis. Here, the measurement of interest is the length and intensity of the comet tail. Also, illumination correction is used to reduce background fluorescence prior to measurement. Also shown is a silver-stained comet example in which the percentage of DNA contained in the tail is calculated. [\[Download\]](#) (0.4 MB) 

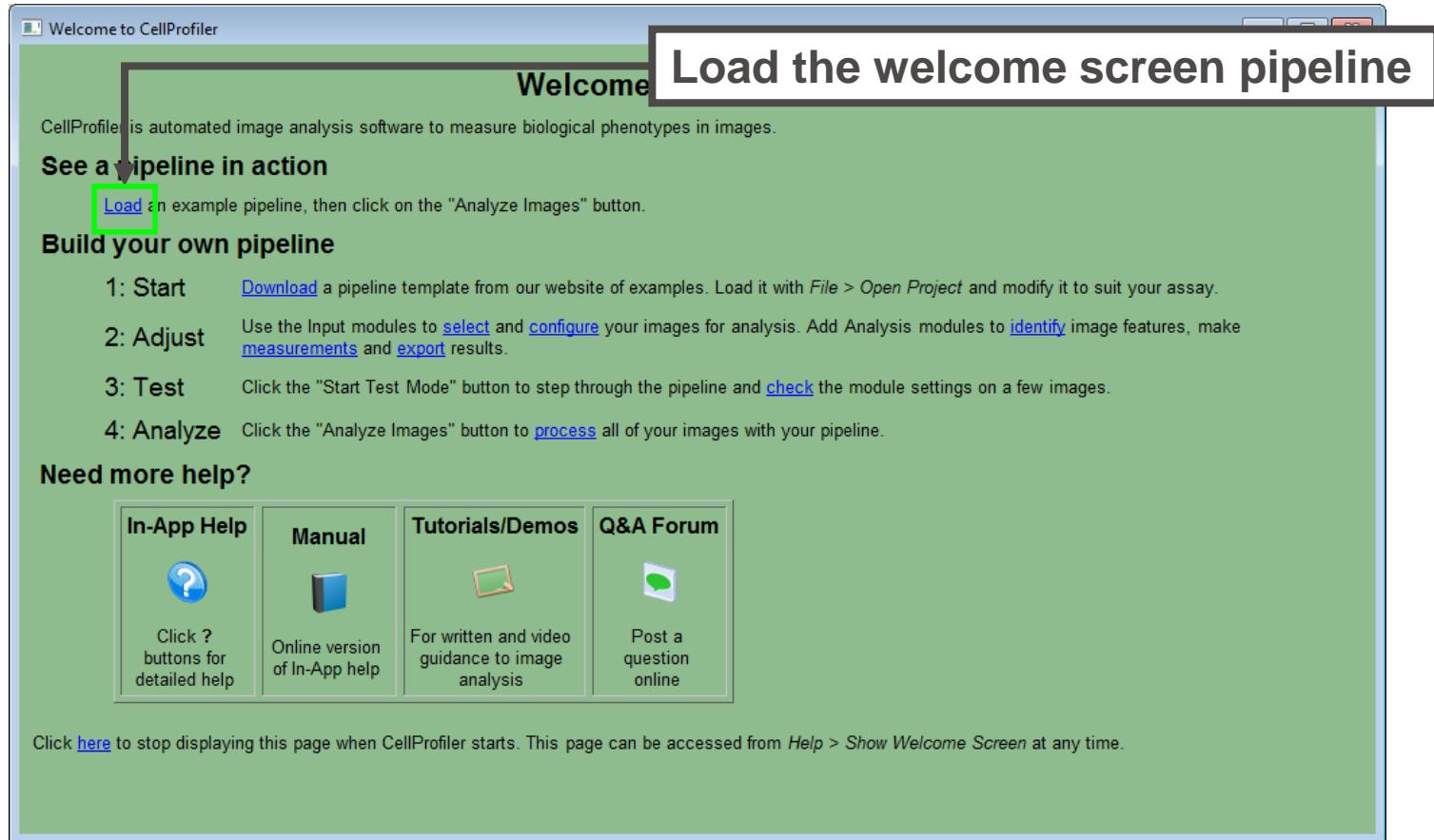
## Specialized pipelines

In addition to cellular object and feature identification, these pipelines include some of the more specialized modules in CellProfiler for image pre-processing or measurement.

- Cell/particle counting, and scoring the percentage of stained objects:** CellProfiler is commonly used to count cells or other objects as well as percent-positives, by measuring the per-cell staining intensity. This pipeline shows how to do both of these tasks, and demonstrates how various modules may be used to accomplish the same result. 

# Typical Starting Point: An Existing Pipeline

- Open CellProfiler



# The CellProfiler Interface

CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cpproj (C:\Trunk\ExampleImages\ExampleSBSImages)

**Pipeline**

**Input modules**

- Images
- Metadata
- NamesAndTypes
- Groups

**Analysis modules**

- CorrectIlluminationApply
- CorrectIlluminationApply
- IdentifyPrimaryObjects**
- IdentifySecondaryObjects
- IdentifySecondaryObjects
- IdentifyTertiaryObjects
- IdentifyTertiaryObjects
- MeasureCorrelation
- MeasureObjectIntensity
- MeasureObjectSizeShape
- MeasureTexture
- CalculateMath

**Output**

**Module notes**

Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation.

**Module settings**

Select the input image **CorrBlue** (from CorrectIlluminationApply #06)

Name the primary objects to be identified **Nuclei**

Typical diameter of objects, in pixel units (Min,Max) **8 28**

Discard objects outside the diameter range?

Discard objects touching the border of the image?

Threshold strategy **Global**

Thresholding method **MoG**

Approximate fraction of image covered by objects? **0.2**

Select the smoothing method for thresholding **Automatic**

Threshold correction factor **1.2**

Lower and upper bounds on threshold **0.04 1.0**

Method to distinguish clumped objects **Intensity**

Method to draw dividing lines between clumped objects **Intensity**

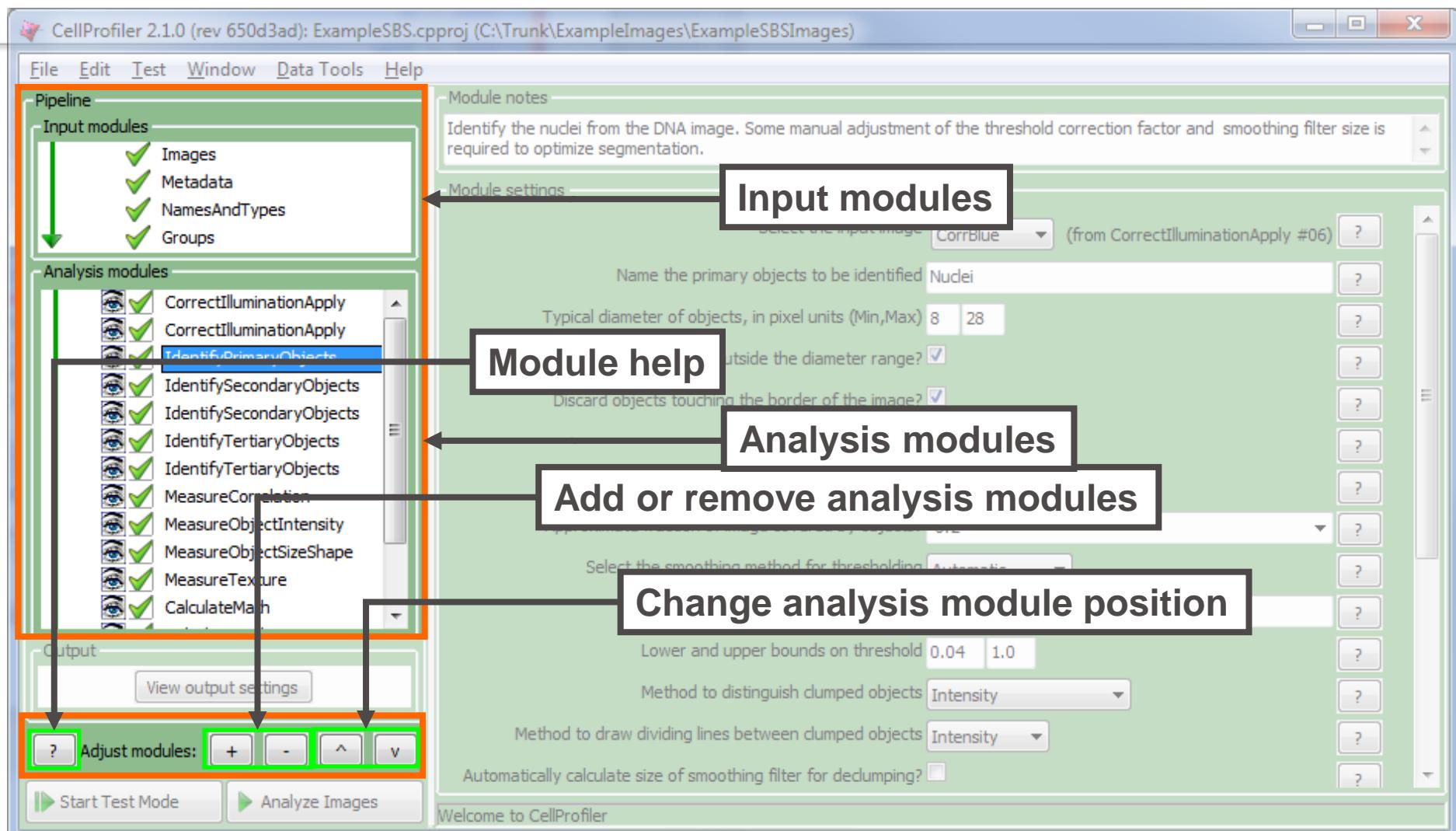
Automatically calculate size of smoothing filter for dedumping?

**Adjust modules:**

**Start Test Mode**

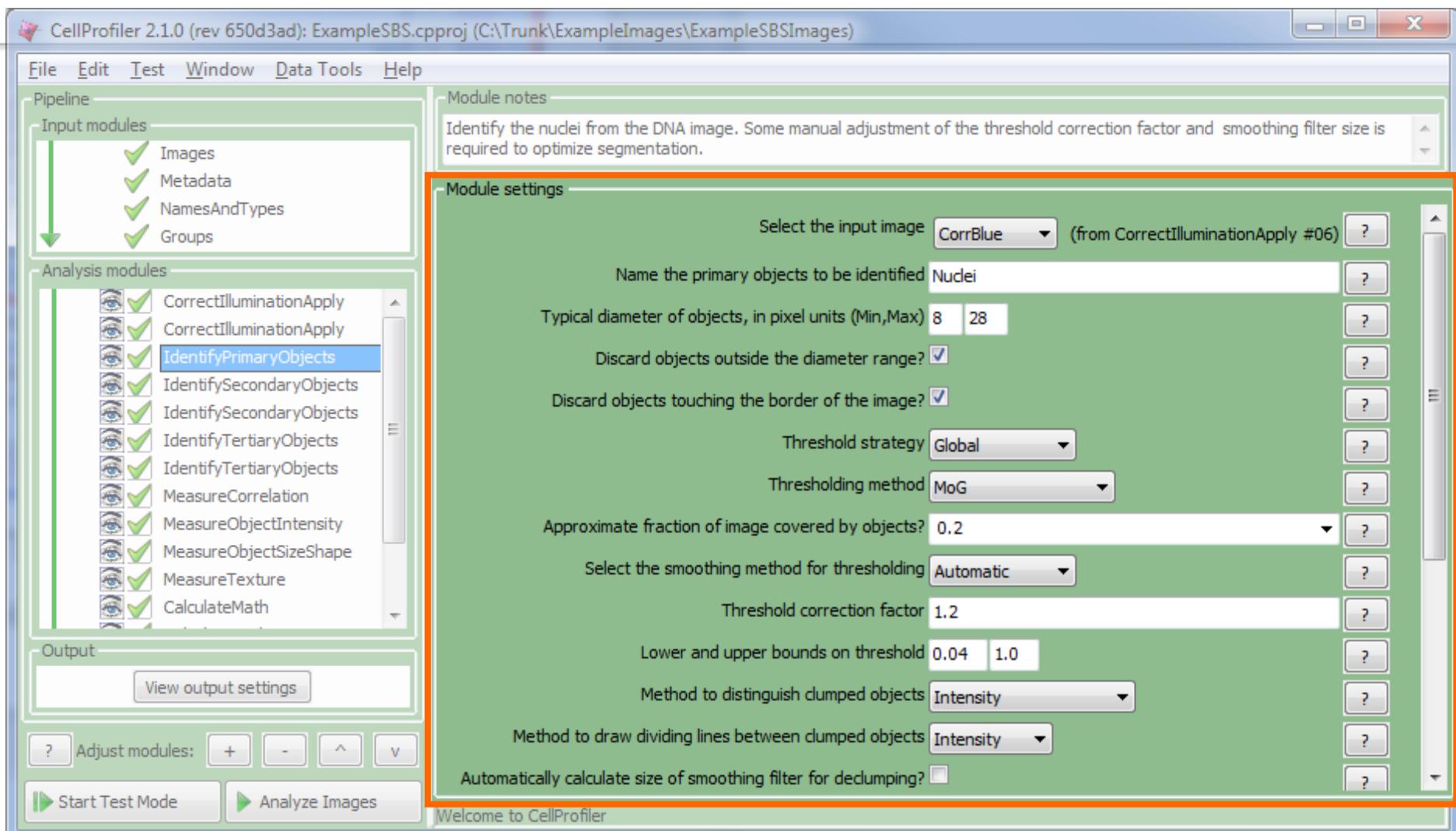
Welcome to CellProfiler

# The CellProfiler Interface



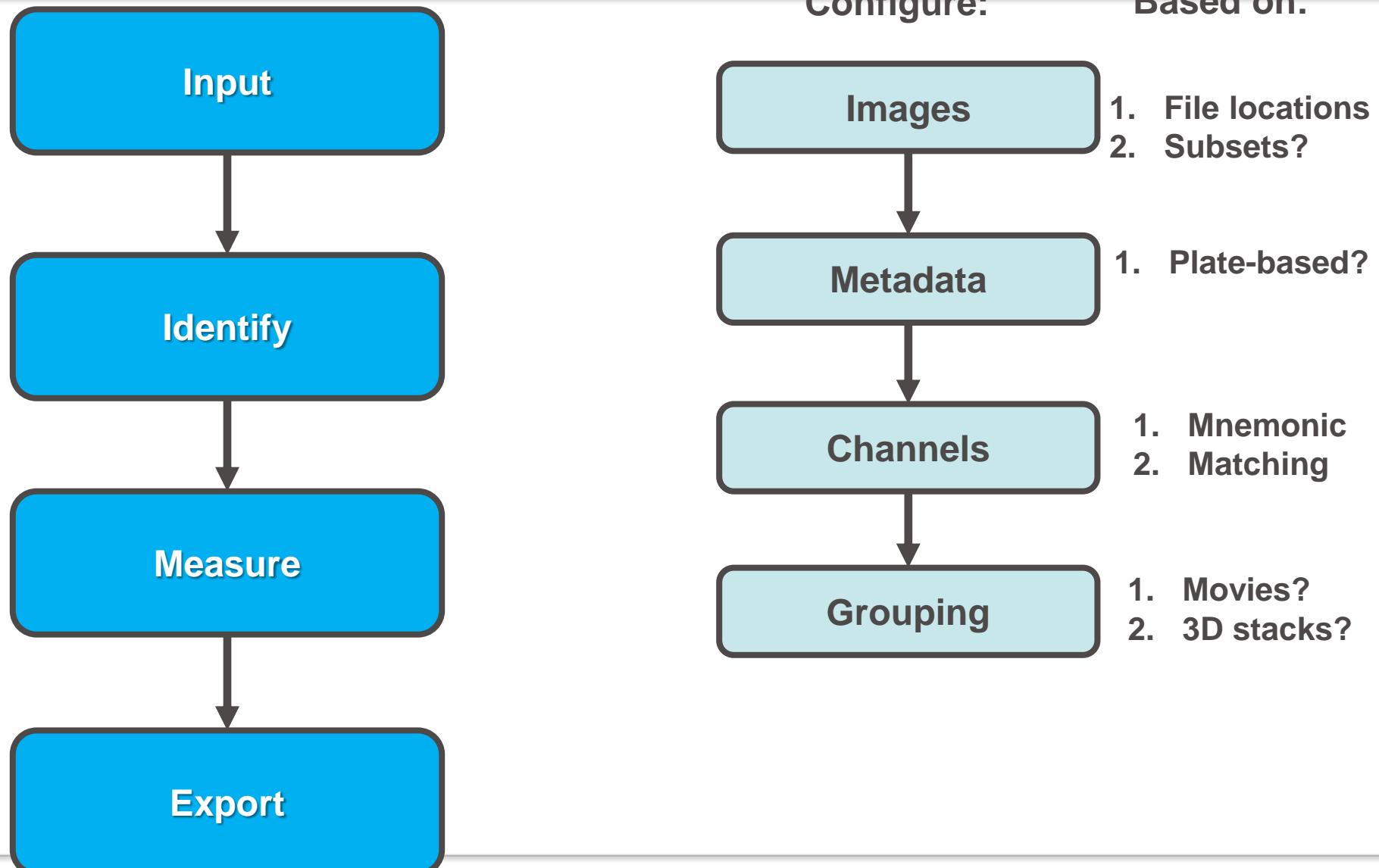
- **Pipeline panel:** Displays modules in pipeline
  - Modules executed in order from top to bottom, extensive help for each

# The CellProfiler Interface



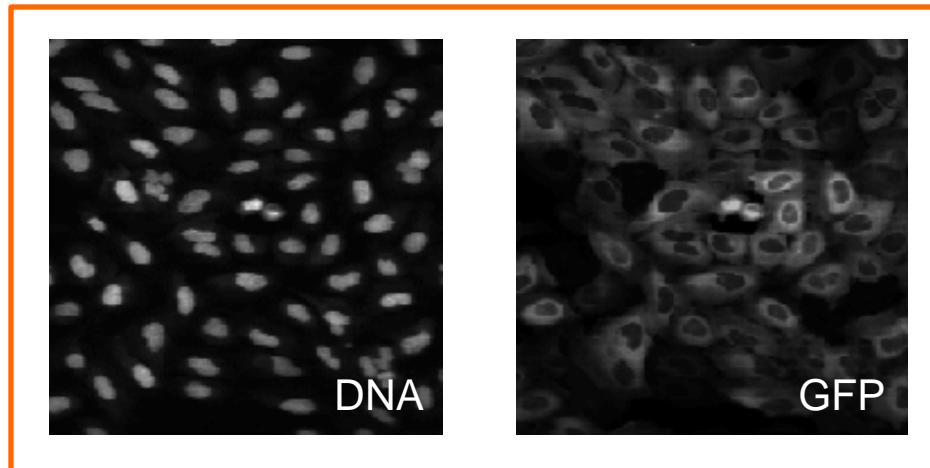
- **Settings panel:** View and change settings for each module
  - Clicking on a different module updates the settings view

# Typical CellProfiler Workflow



# Creating A CellProfiler Project

- Use the **Input** modules to create an *image set*



## An *image set*

- Most commonly defined as 1 or more channels at one imaging location
- Other definitions depend on the assay

- Add **Analysis** modules to pre-process the images, identify objects, make measurements
- Add **Export** modules to write out images/measurements to disk or database

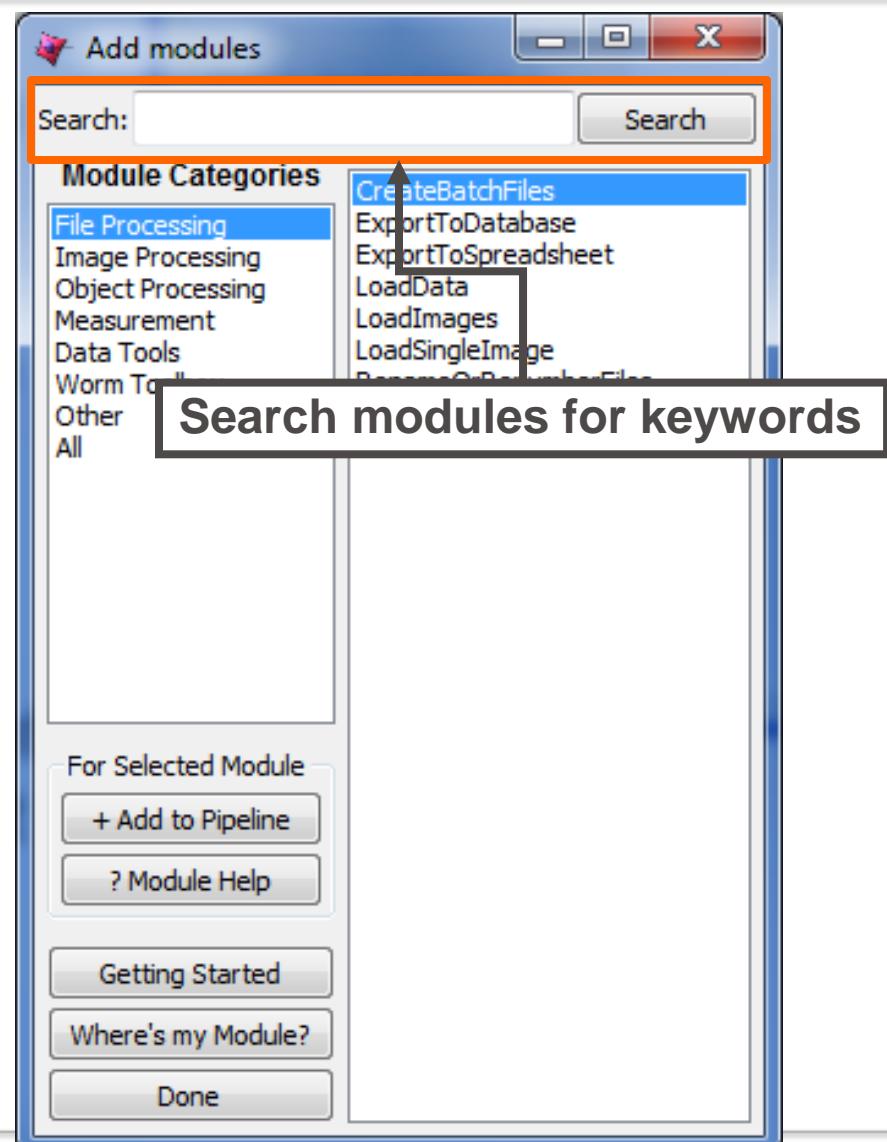
# Loading and Configuring Your Images

---

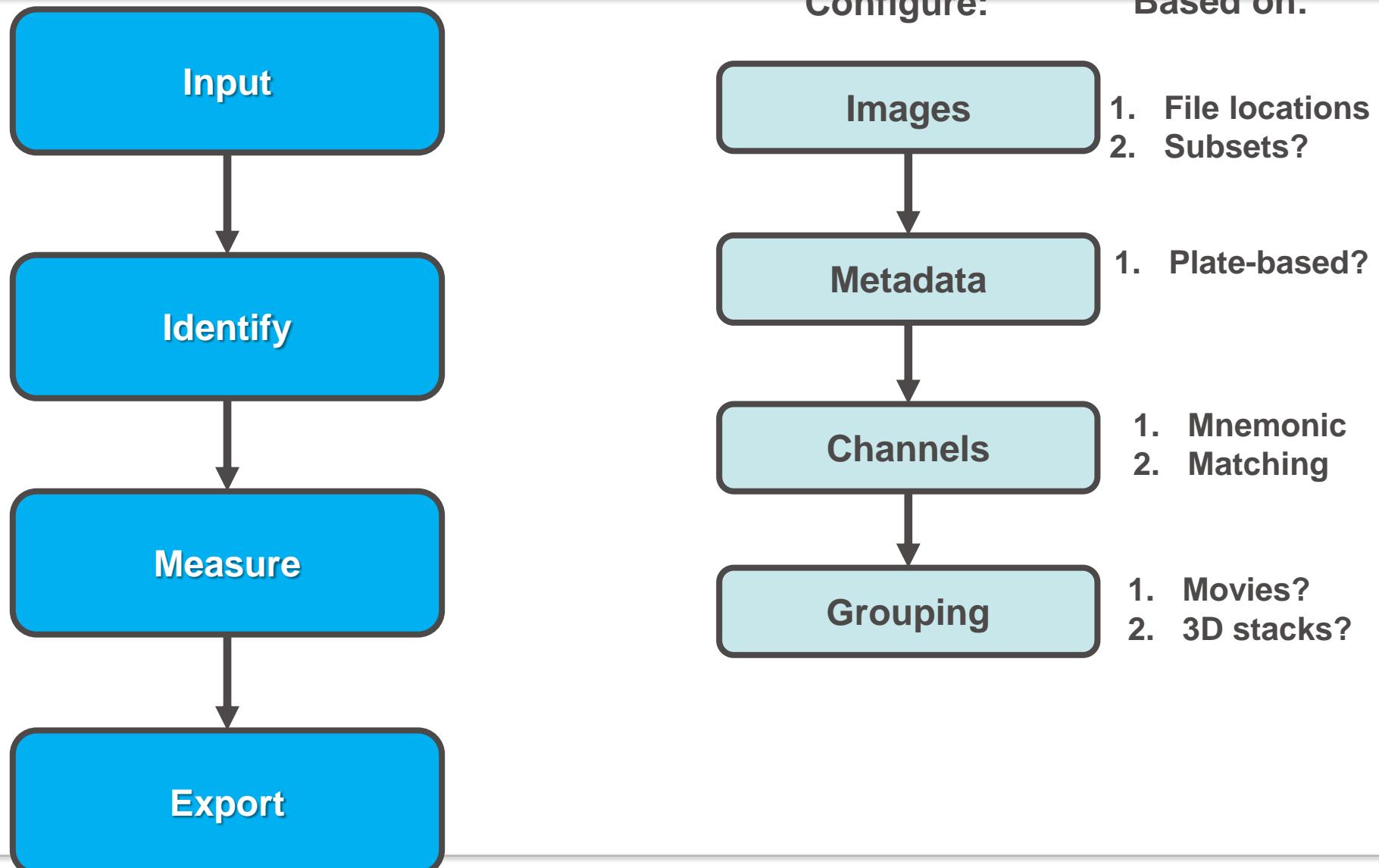
- Browse to the files extracted from the ZIP file you downloaded from GitHub
- **Images** module
  - Right-click on the file list panel; select “Clear file list” from context menu
  - Drag-and-drop the “images” folder into the file list panel module
- **NamesAndTypes** module: Make the following changes
  - **OrigBlue**: [File] [Does] [Contain] “**d0.png**”
  - **OrigGreen**: [File] [Does] [Contain] “**d1.png**”
  - **OrigRed**: [File] [Does] [Contain] “**d2.png**”
- **Crop** module: Adjust the settings
  - Left and right rectangle positions:
  - Top and bottom rectangle positions:

# Module Categories

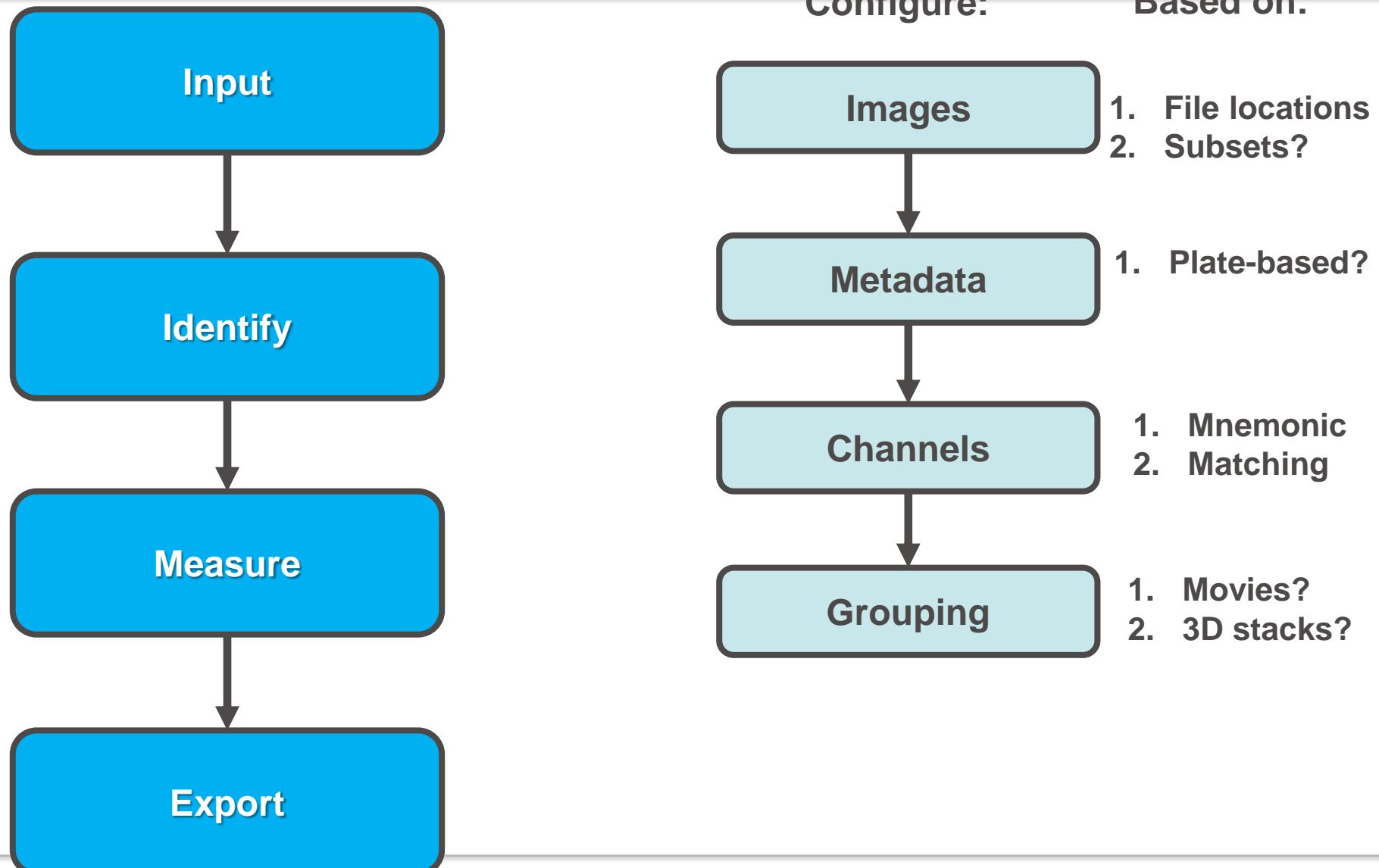
- **File processing:** Image input, file output
- **Image processing:** Often used for pre-processing prior to object identification
- **Object processing:** Identification, modification of objects of interest
- **Measurement:** Collection of measurements from objects of interest
- **Data Tools:** Measurement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



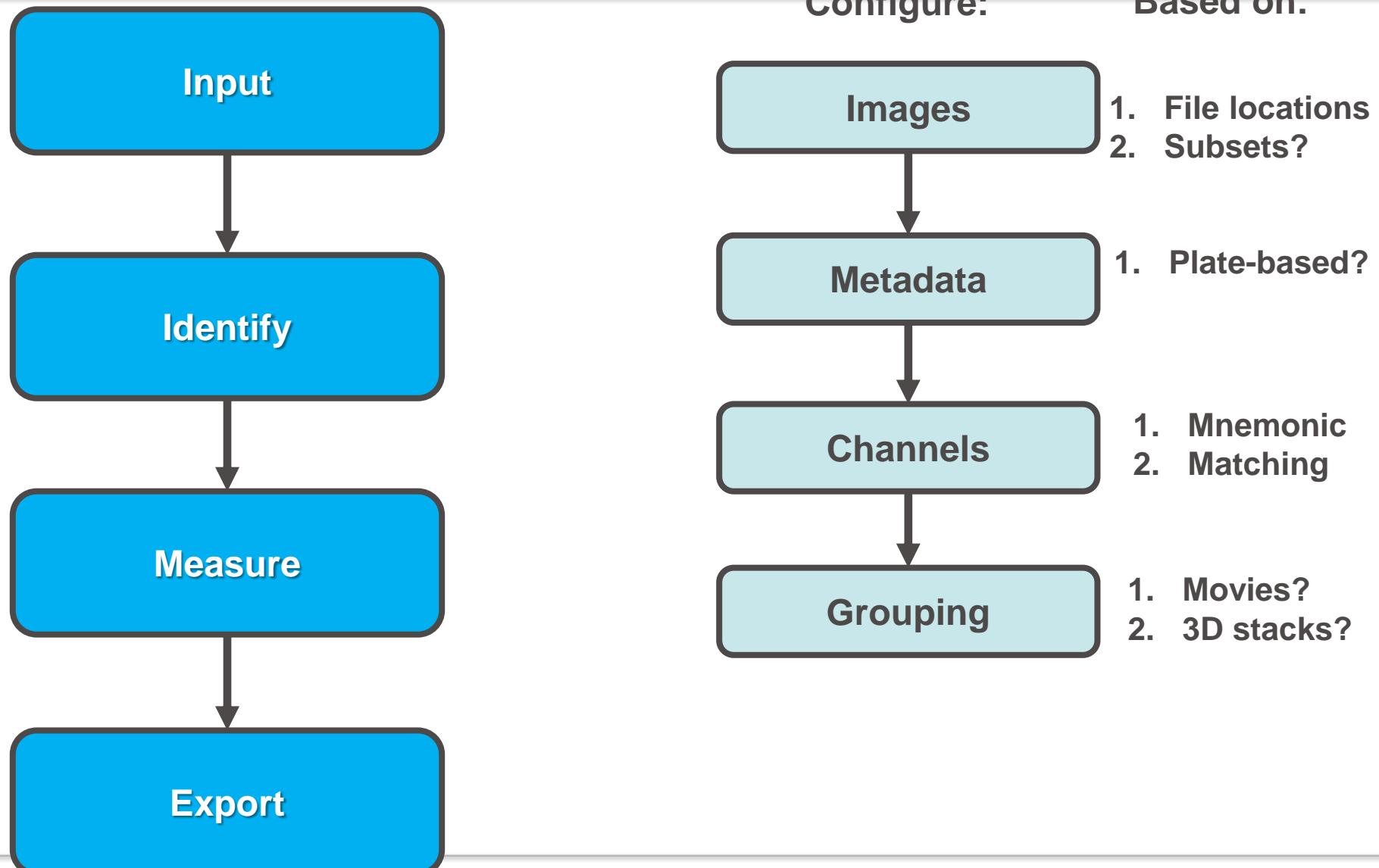
# Typical CellProfiler Workflow



# Typical CellProfiler Workflow



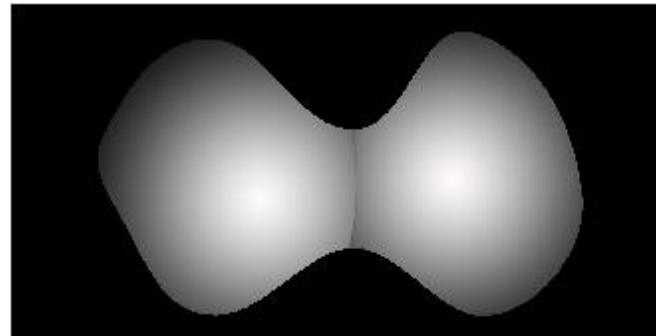
# Typical CellProfiler Workflow



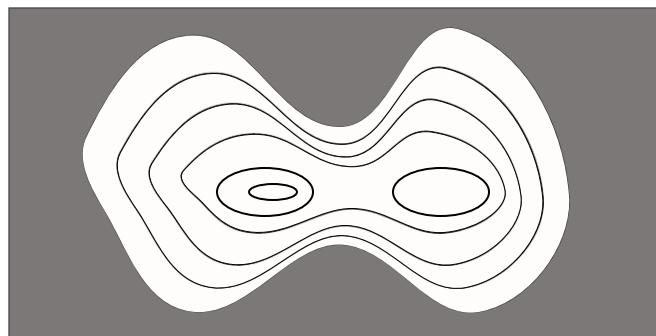
# Object Identification

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

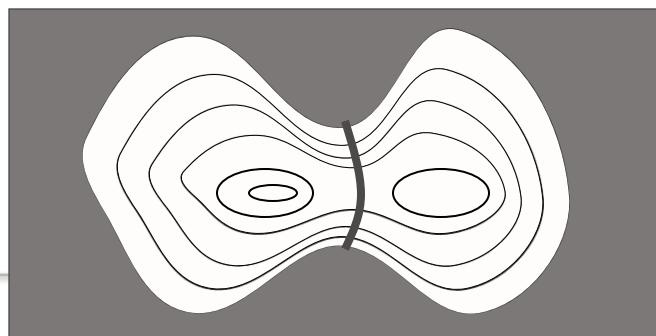
**Step 1:** Distinguish the foreground from the background by picking a good threshold



**Step 2:** Identify objects as regions brighter than the threshold



**Step 3:** Cut and join objects to “improve” their shape



# Primary Object Identification

The screenshot shows the CellProfiler 2.1.0 software interface with the title bar "CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cpproj (C:\Trunk\ExampleImages\ExampleSBSImages)". The menu bar includes File, Edit, Test, Window, Data Tools, and Help. The Pipeline panel on the left lists Input modules (Images, Metadata, NamesAndTypes, Groups) and Analysis modules (CorrectIlluminationApply, IdentifyPrimaryObjects, IdentifySecondaryObjects, IdentifyTertiaryObjects, MeasureCorrelation, MeasureObjectIntensity, MeasureObjectSizeShape, MeasureTexture, CalculateMath, CalculateMath, CalculateStatistics). The "IdentifyPrimaryObjects" module is selected. The Module notes section states: "Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation." The Module settings section contains the following parameters:

- Select the input image: CorrBlue (from CorrectIlluminationApply #06)
- Name the primary objects to be identified: Nuclei
- Typical diameter of objects, in pixel units (Min,Max): 8 28
- Discard objects outside the diameter range?
- Discard objects touching the border of the image?
- Threshold strategy: Global
- Thresholding method: MoG
- Approximate fraction of image covered by objects: 0.2
- Select the smoothing method for thresholding: Automatic
- Threshold correction factor: 1.2
- Lower and upper bounds on threshold: 0.04 1.0
- Method to distinguish clumped objects: Intensity
- Method to draw dividing lines between clumped objects: Intensity
- Automatically calculate size of smoothing filter for declumping?
- Size of smoothing filter: 5

At the bottom, there are buttons for "View output settings", "Adjust modules" (+, -, ^, v), "Start Test Mode", and "Analyze Images". A "Welcome to CellProfiler" message is also present.

- Many options for thresholding, cut and join methods, etc.

# Primary Object Identification

The screenshot shows the CellProfiler 2.1.0 software interface. The pipeline workspace is visible on the left, showing the following sequence of modules:

- Input modules: Images, Metadata, NamesAndTypes, Groups.
- Analysis modules: CorrectIlluminationApply (x2), IdentifyPrimaryObjects (selected), IdentifySecondaryObjects (x2), IdentifyTertiaryObjects (x2), MeasureCorrelation, MeasureObjectIntensity, MeasureObjectSizeShape, MeasureTexture.
- Output: View output settings, Adjust modules, Start Test Mode, Analyze Images.

The right panel displays the "Module notes" for the "IdentifyPrimaryObjects" module, which states: "Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation." Below this are "Module settings" for identifying nuclei:

- Select the input image: CorrBlue (from CorrectIlluminationApply #06).
- Name the primary objects to be identified: Nuclei.
- Typical diameter of objects, in pixel units (Min,Max): 8, 28.
- Discard objects outside the diameter range?
- Discard objects touching the border of the image?
- Threshold strategy: Global.
- Thresholding method: MoG.
- Approximate fraction of image covered by objects: 0.2.
- Select the smoothing method for thresholding: Automatic.
- Method to distinguish clumped objects: Intensity.
- Method to draw dividing lines between clumped objects: Intensity.
- Automatically calculate size of smoothing filter for declumping?
- Size of smoothing filter: 5.

A large black box highlights the text "Use Test Mode for previewing object identification results". A black arrow points from the "View output settings" button in the pipeline to the "Start Test Mode" button at the bottom left of the main workspace. The "Start Test Mode" button is also highlighted with an orange border.

# Test Mode

CellProfiler 2.1.0 (rev 4e283c4): ExampleSBS.cpproj (C:\Trunk\ExampleImages\ExampleSBSImages)

**Pipeline**

**Input modules**

- Images
- Metadata
- NamesAndTypes
- Groups

**Analysis modules**

- CorrectIlluminationApply
- CorrectIlluminationApply
- IdentifyPrimaryObjects
- IdentifySecondaryObjects**
- IdentifySecondaryObjects
- IdentifyTertiaryObjects
- IdentifyTertiaryObjects
- MeasureCorrelation
- MeasureObjectIntensity
- MeasureObjectSizeShape
- MeasureTexture
- CalculateMath

**Output**

[View output settings](#)

Adjust modules: [+](#) [-](#) [^](#) [v](#)

[Run](#) [Step](#)

[Exit Test Mode](#) [Next Image Set](#)

**Module notes**  
Identify the cells by using the nuclei as a "seed" region, then growing outwards until stopped by the image threshold or by a neighbor. The Propagation method is used to delineate the boundary between neighboring cells.

**Module settings**

Select the input image: CorrGreen (from CorrectIlluminationApply #05) [?](#)

Select the input objects: Nuclei (from IdentifyPrimaryObjects #07) [?](#)

Name the objects to be identified: PropCells [?](#)

Select the method to identify the secondary objects: Propagation [?](#)

Threshold strategy: Global [?](#)

Thresholding method: Otsu [?](#)

Two-class or three-class thresholding?: Two classes [?](#)

Minimize the weighted variance or the entropy?: Weighted variance [?](#)

Select the smoothing method for thresholding: No smoothing [?](#)

Threshold correction factor: 1.0 [?](#)

Lower and upper bounds on threshold: 0.02 1.0 [?](#)

Regularization factor: 0.05 [?](#)

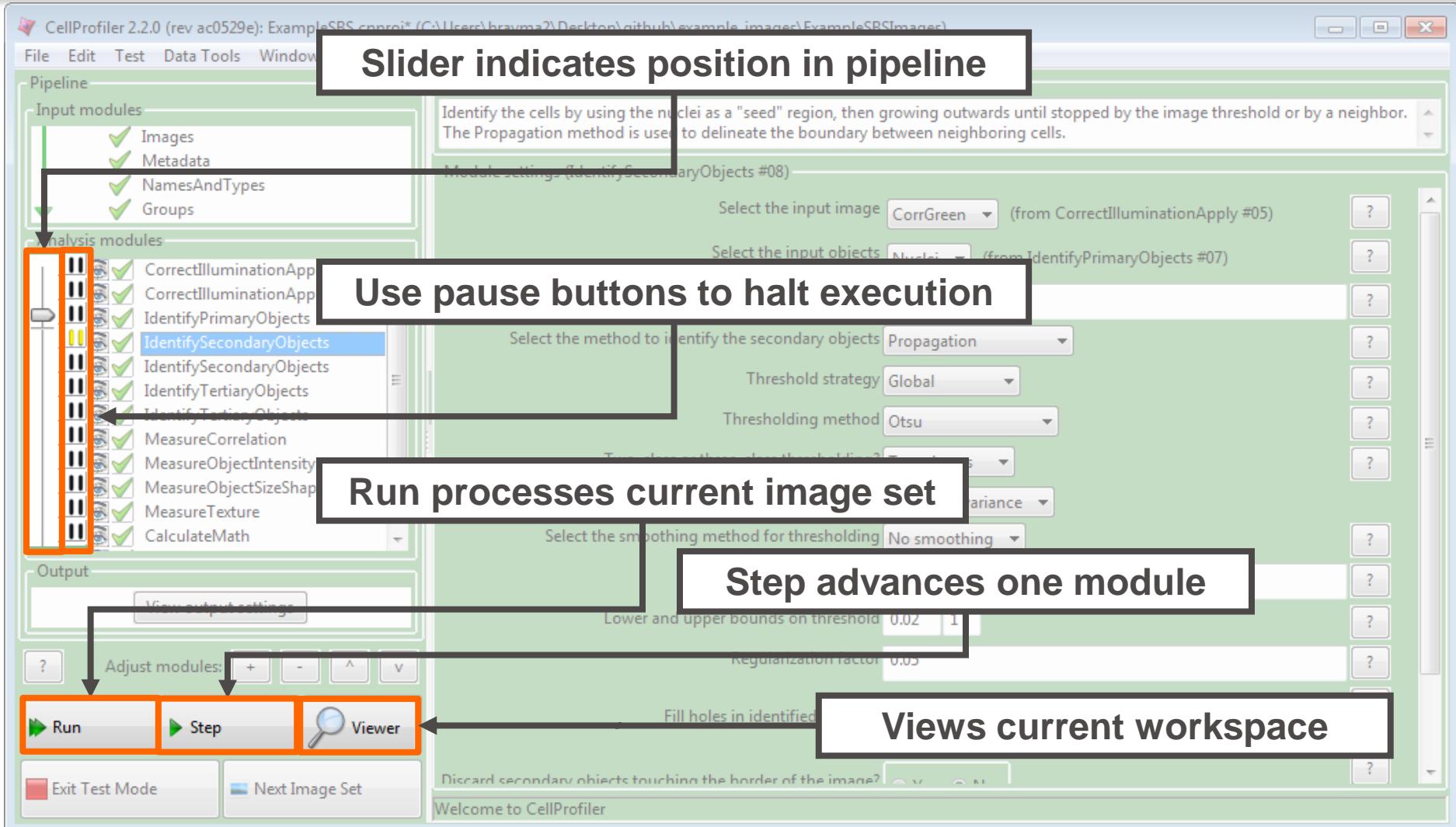
Fill holes in identified objects?  [?](#)

Discard secondary objects touching the border of the image?  [?](#)

Retain outlines of the identified secondary objects?  [?](#)

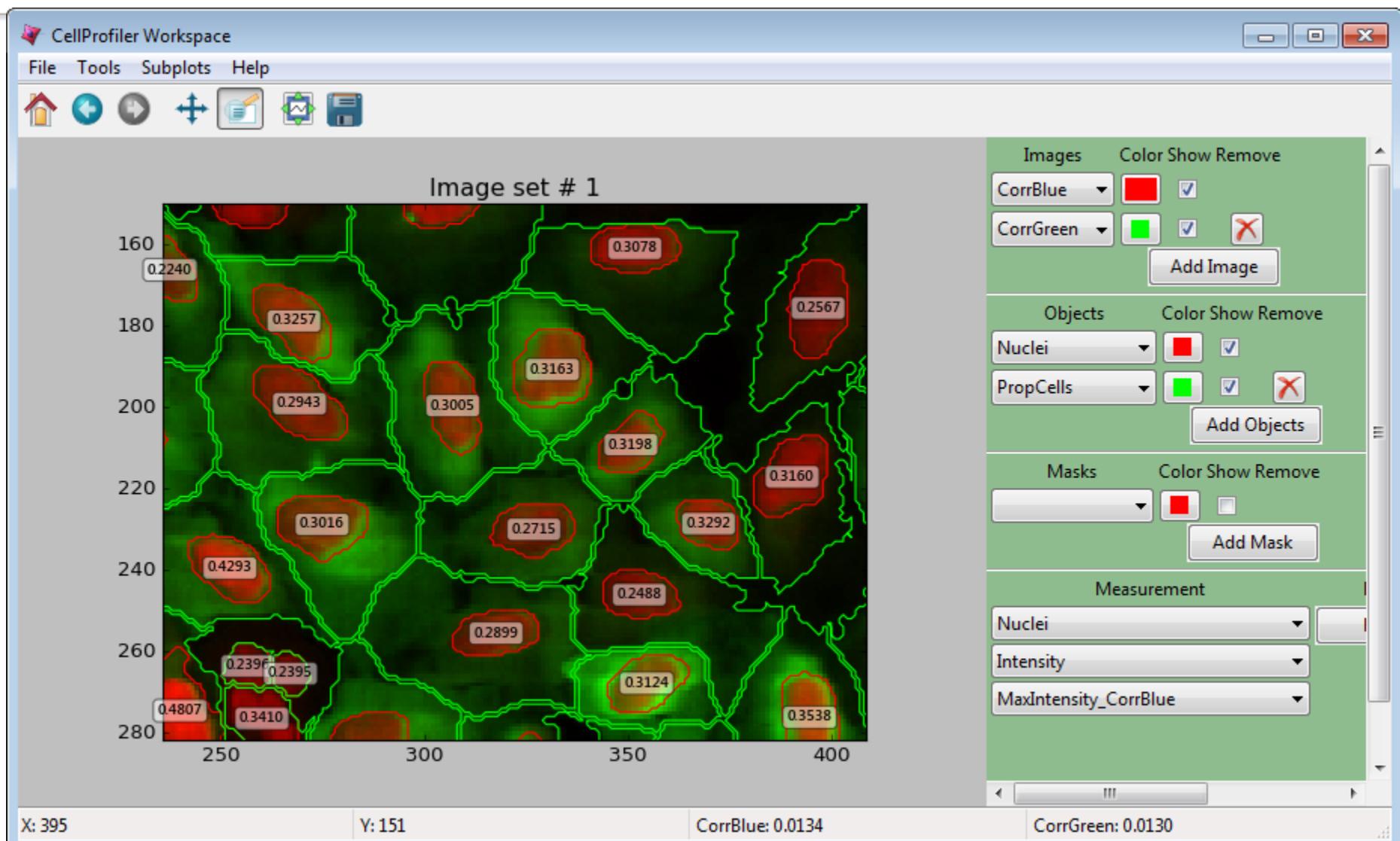
Welcome to CellProfiler

# Test Mode

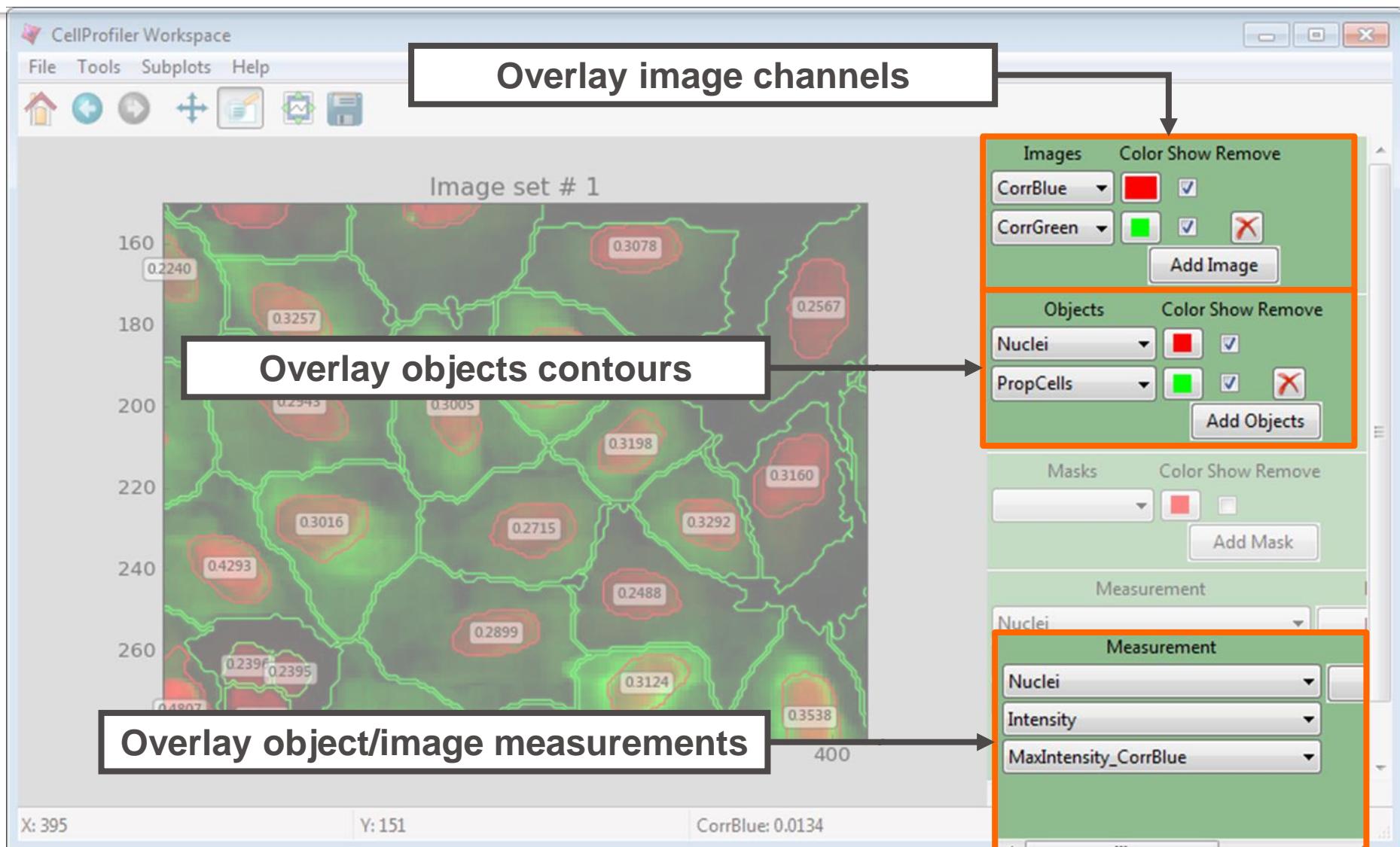


- Use the Test menu item for more options

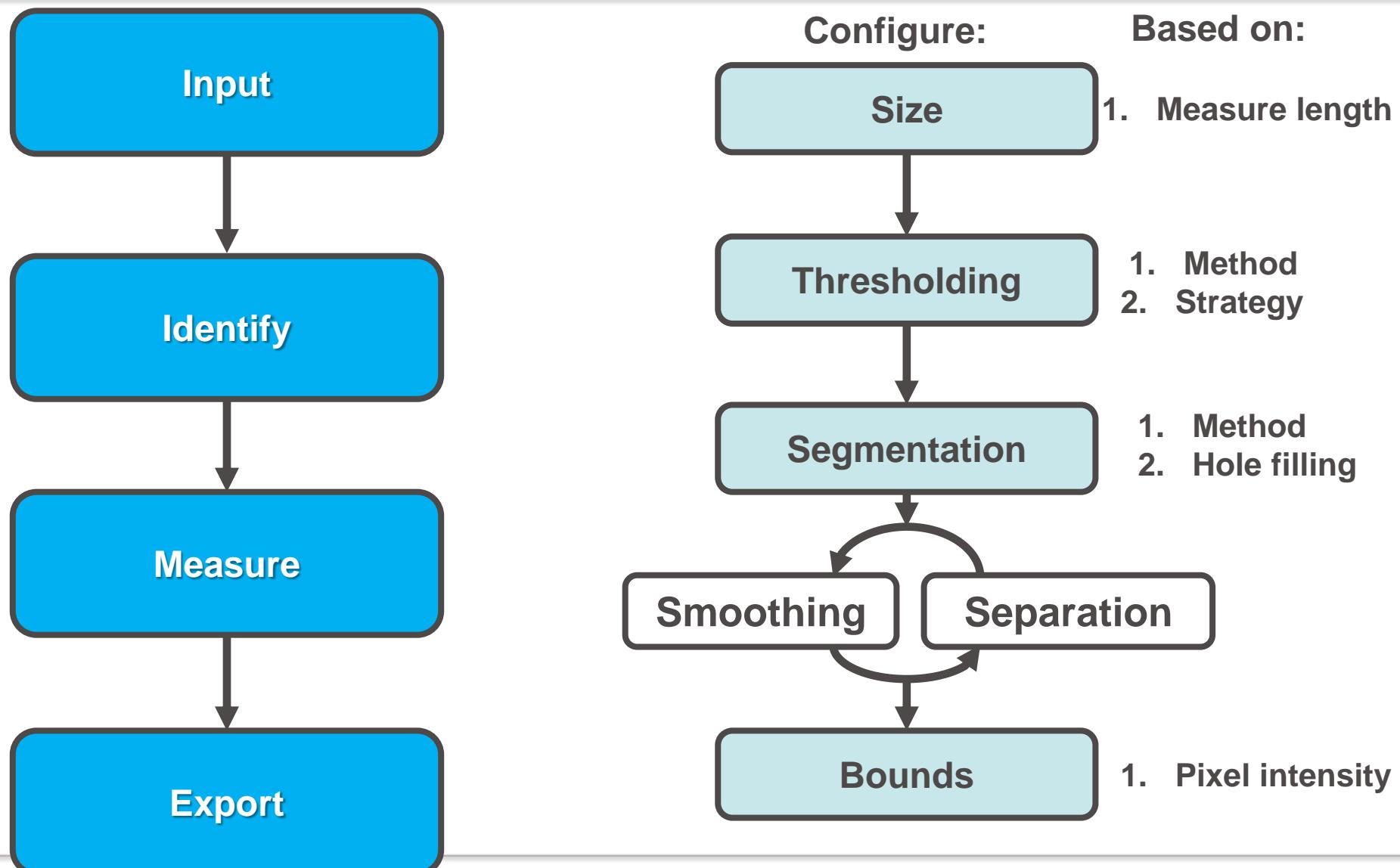
# Viewing the Current Workspace



# Viewing the Current Workspace



# Typical CellProfiler Workflow



# Filtering Invalid Objects

The screenshot shows the CellProfiler 2.1.0 software interface with the following details:

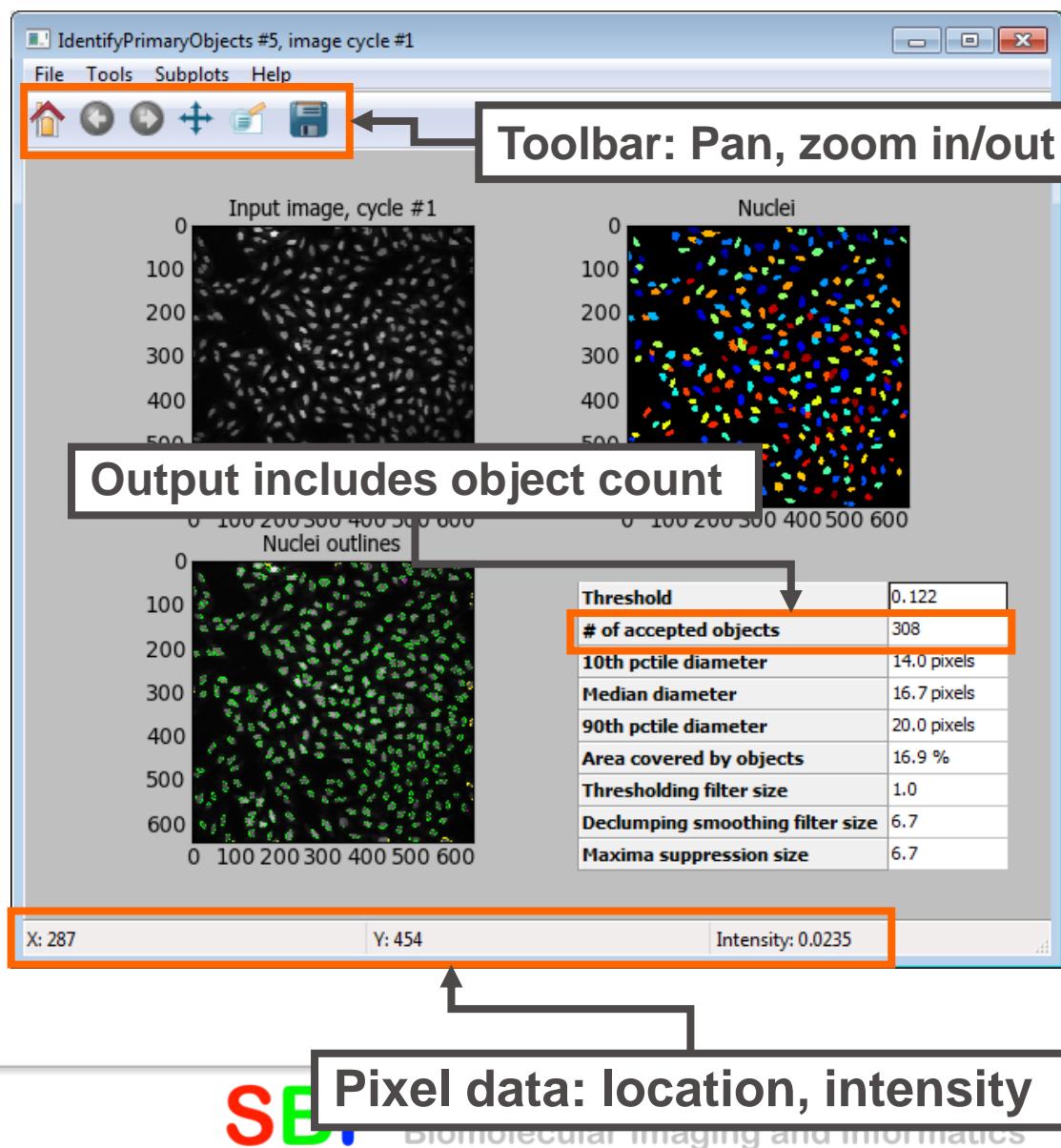
- Input modules:** Images, Metadata, NamesAndTypes, Groups.
- Analysis modules:** CorrectIlluminationApply (x2), IdentifyPrimaryObjects (selected), IdentifySecondaryObjects (x2), IdentifyTertiaryObjects (x2), MeasureCorrelation, MeasureObjectIntensity, MeasureObjectSizeShape.
- Module notes:** Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation.
- Module settings (IdentifyPrimaryObjects):**
  - Select the input image: CorrBlue (from CorrectIlluminationApply #06).
  - Name the primary objects to be identified: Nuclei.
  - Typical diameter of objects, in pixel units (Min,Max): 8, 28.
  - Discard objects outside the diameter range?
  - Discard objects touching the border of the image?  (highlighted with an orange box and an arrow pointing to it from the text below).
  - Threshold strategy: Global.
  - Thresholding method: MoG.
  - Approximate fraction of image covered by objects?: 0.2.
  - Lower and upper bounds on threshold: 0.04, 1.0.
  - Method to distinguish clumped objects: Intensity.
  - Method to draw dividing lines between clumped objects: Intensity.
  - Automatically calculate size of smoothing filter for declumping?
  - Size of smoothing filter: 5.
- Output:** View output settings, Adjust modules (+, -, ^, v), Start Test Mode, Analyze Images.
- Welcome to CellProfiler**

**Discard objects that fail size criterion or touch the image border**

- See FilterObjects module for more advanced filtering options

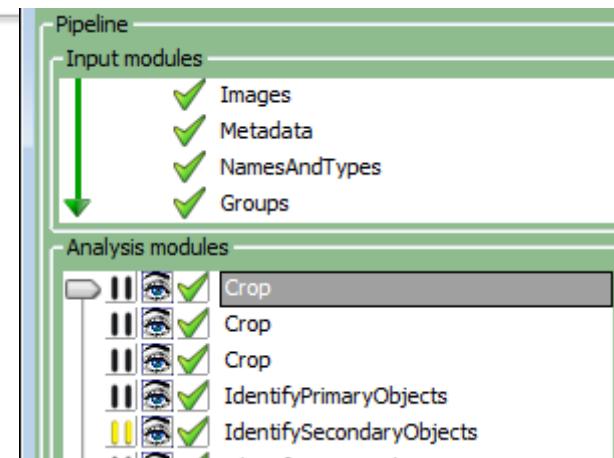
# Primary Object Identification

- Segmented objects are colored
  - Shows if each object has been identified and separated properly
- Outlines: Valid objects
  - Green:** Valid
  - Yellow:** Invalid – Touching border
  - Red:** Invalid – Size criterion

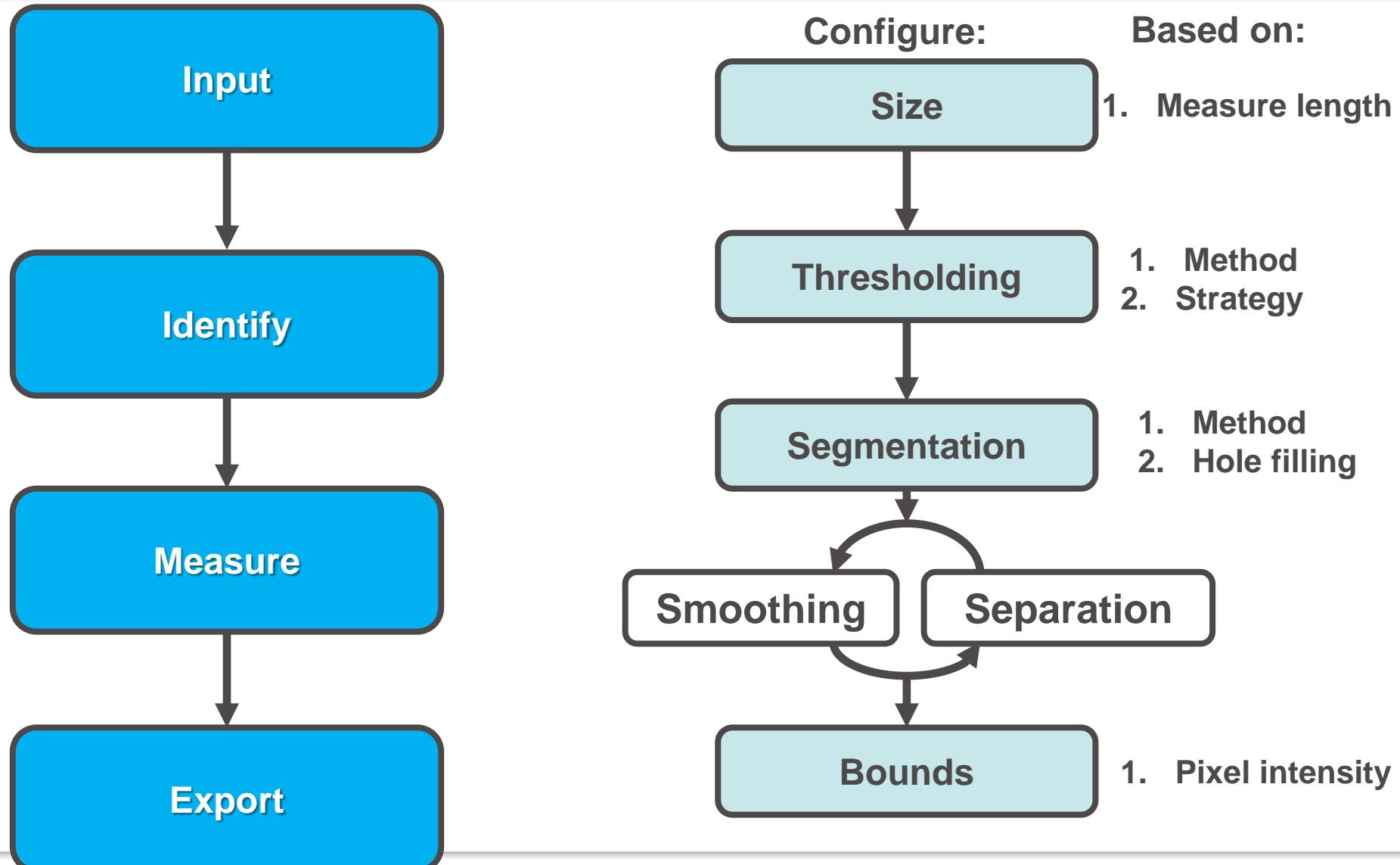


# Primary Object Identification

- Press the Test mode button
- Adjust **Crop** module settings
  - Left and right rectangle positions: 250, end
  - Top and bottom rectangle positions: 1, 250
- Set a pause after **IdentifyPrimaryObjects**, run the pipeline
- Experiment with the exclusion settings
  - Measure a few nuclei with the measuring tool
  - Toggle the border touching criteria, re-run the module
  - Toggle the size exclusion criteria, re-run the module
- What min/max diameter setting would be most appropriate in this case?



# Typical CellProfiler Workflow



# Image Thresholding

CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cppproj (C:\Trunk\ExampleImages\ExampleSBSImages)

**File Edit Test Window Data Tools Help**

**Pipeline**

**Input modules**

- Images
- Metadata
- NamesAndTypes

**IdentifyPrimaryObjects** (highlighted)

- CorrectIlluminationApply
- CorrectIlluminationApply
- IdentifySecondaryObjects
- IdentifySecondaryObjects
- IdentifyTertiaryObjects
- IdentifyTertiaryObjects
- MeasureCorrelation
- MeasureObjectIntensity
- MeasureObjectSizeShape
- MeasureTexture
- CalculateMath
- CalculateMath
- CalculateStatistics

**Output**

**View output settings**

**Adjust modules:** + - ^ v

**Start Test Mode** **Analyze Images**

**Module notes**  
Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation.

**Module settings**

Select the input image CorrBlue (from CorrectIlluminationApply #06) ?  
The primary objects to be identified Nuclei ?

Typical diameter of objects, in pixel units (Min,Max) 8 28  
Discard objects outside the diameter range?  ?  
Discard objects touching the border of the image?  ?

**Threshold strategy** Global ?  
**Thresholding method** MoG ?  
Approximate fraction of image covered by objects? 0.2 ?  
Select the smoothing method for thresholding Automatic ?

Threshold correction factor 1.2  
Lower and upper bounds on threshold 0.04 1.0  
Method to distinguish clumped objects Intensity ?  
Method to draw dividing lines between clumped objects Intensity ?  
Automatically calculate size of smoothing filter for declumping?  ?  
Size of smoothing filter 5 ?

Welcome to CellProfiler

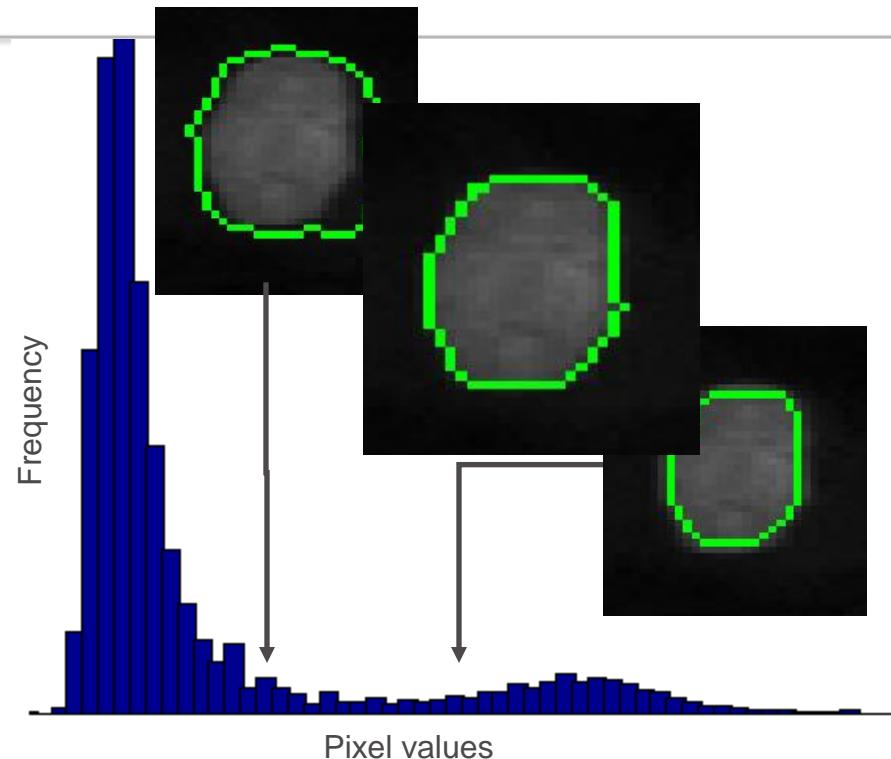
**Adjust the thresholding method**

The screenshot shows the CellProfiler 2.1.0 software interface. The pipeline on the left includes 'IdentifyPrimaryObjects' (which is highlighted with a black box), 'CorrectIlluminationApply', 'IdentifySecondaryObjects', 'IdentifyTertiaryObjects', 'MeasureCorrelation', 'MeasureObjectIntensity', 'MeasureObjectSizeShape', 'MeasureTexture', 'CalculateMath', and 'CalculateStatistics'. The 'IdentifyPrimaryObjects' module has several settings in the module settings panel: 'Select the input image' set to 'CorrBlue' (from 'CorrectIlluminationApply #06'), 'The primary objects to be identified' set to 'Nuclei', 'Typical diameter of objects, in pixel units (Min,Max)' set to '8 28', 'Discard objects outside the diameter range?' checked, 'Discard objects touching the border of the image?' checked, 'Threshold strategy' set to 'Global', 'Thresholding method' set to 'MoG', 'Approximate fraction of image covered by objects?' set to '0.2', 'Select the smoothing method for thresholding' set to 'Automatic', 'Threshold correction factor' set to '1.2', 'Lower and upper bounds on threshold' set to '0.04 1.0', 'Method to distinguish clumped objects' set to 'Intensity', 'Method to draw dividing lines between clumped objects' set to 'Intensity', and 'Automatically calculate size of smoothing filter for declumping?' unchecked. A large red box surrounds the 'Threshold strategy', 'Thresholding method', and 'Select the smoothing method for thresholding' settings. A black arrow points from the 'Discard objects touching the border of the image?' checkbox to the red box. A callout box labeled 'Adjust the thresholding method' is positioned over the 'IdentifyPrimaryObjects' module in the pipeline.

# 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?*

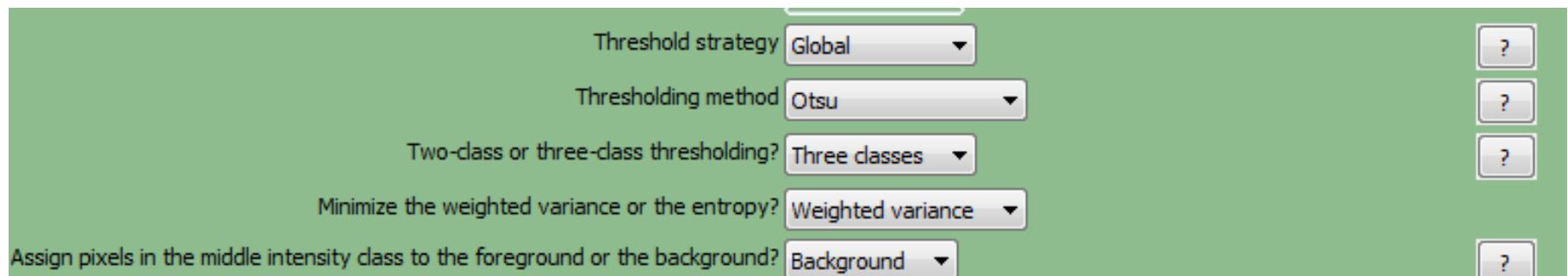


Method: Pick the method that provides the best results

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

# Image Thresholding

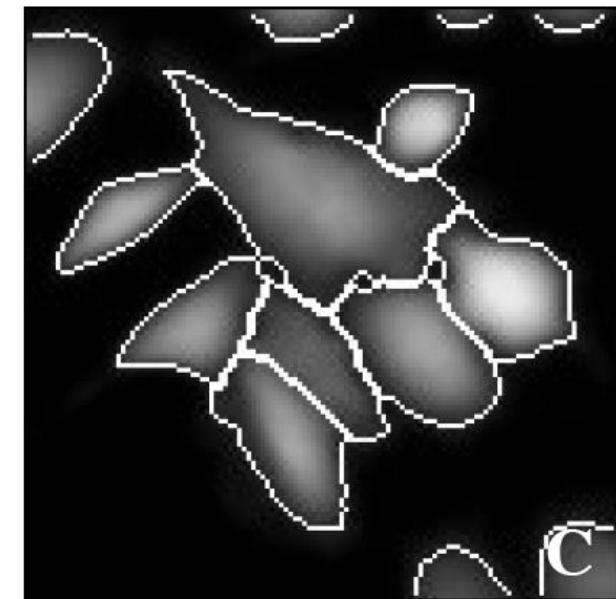
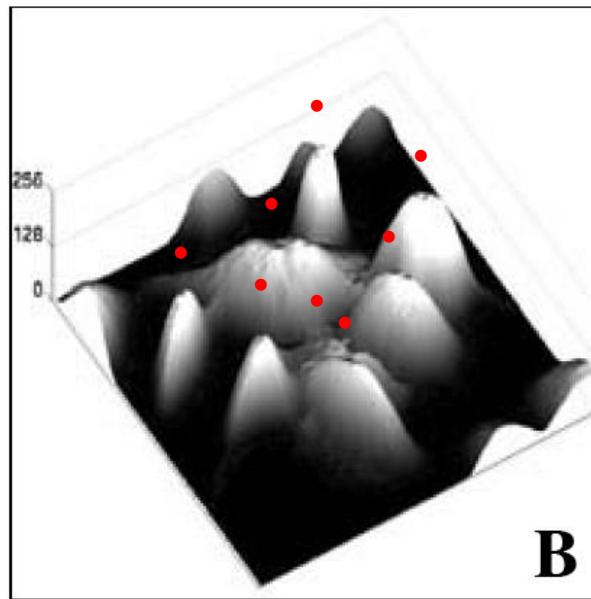
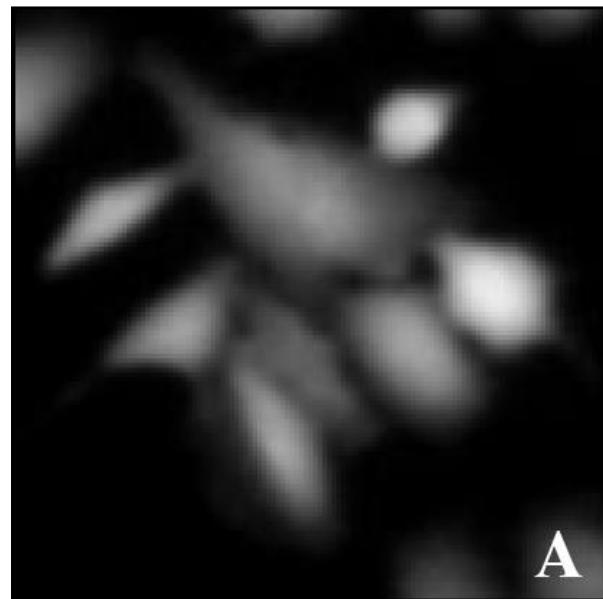
- Experiment with the threshold settings
  - Start with the “Thresholding method” setting
  - Try adjusting the “Thresholding strategy”, if time permits



- Good: Otsu (two classes), MOG (fraction = 0.05), RidlerCalvard
- Too lenient: Background, RobustBackground, Kapur
- Too strict: MCT

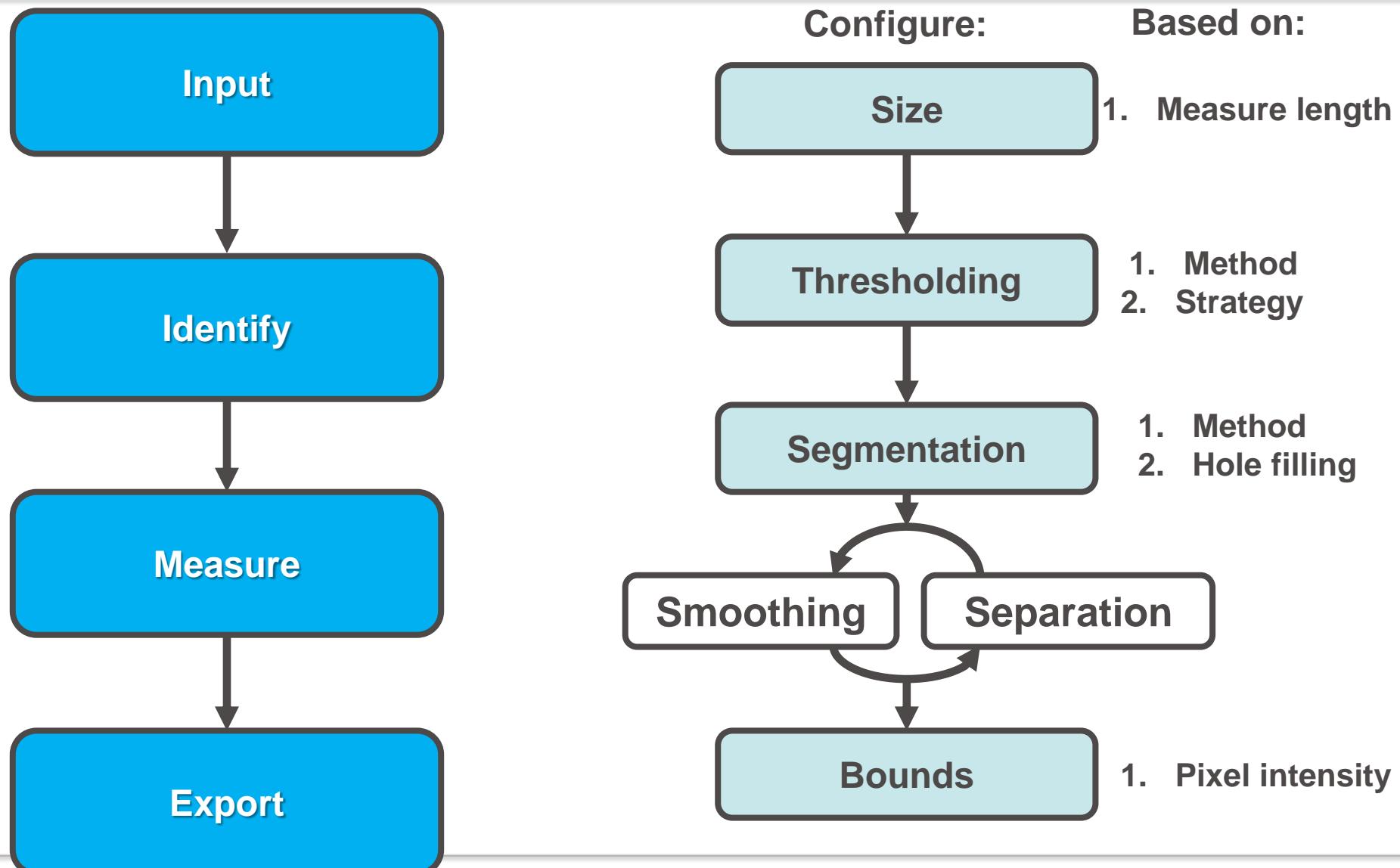
# Separating Touching Objects

- Once the foreground objects have been identified, what next?



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

# Typical CellProfiler Workflow



# Separating Touching Objects

CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cppproj (C:\Trunk\ExampleImages\ExampleSBSImages)

**Adjust settings to “de-clump” objects**

Module notes  
Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation.

**Pipeline**

**Input modules**

- Images

**Analysis modules**

- CorrectIlluminationApply
- CorrectIlluminationApply
- IdentifyPrimaryObjects**
- IdentifySecondaryObjects
- IdentifySecondaryObjects
- IdentifyTertiaryObjects
- IdentifyTertiaryObjects
- MeasureCorrelation
- MeasureObjectIntensity
- MeasureObjectSizeShape
- MeasureTexture
- CalculateMath
- CalculateMath
- CalculateStatistics

**Output**

**Module notes**

Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation.

Thresholding method: MoG

Approximate fraction of image covered by objects: 0.2

Select the smoothing method for thresholding: Automatic

Threshold correction factor: 1.2

Method to distinguish clumped objects: Intensity

Method to draw dividing lines between clumped objects: Intensity

Automatically calculate size of smoothing filter for declumping?

Size of smoothing filter: 5

Automatically calculate minimum allowed distance between local maxima?

Automatically calculate minimum allowed distance between local maxima?

Speed up by using lower-resolution image to find local maxima?

Retain outlines of the identified objects?

Fill holes in identified objects?

Handling of objects if excessive number of objects identified: Continue

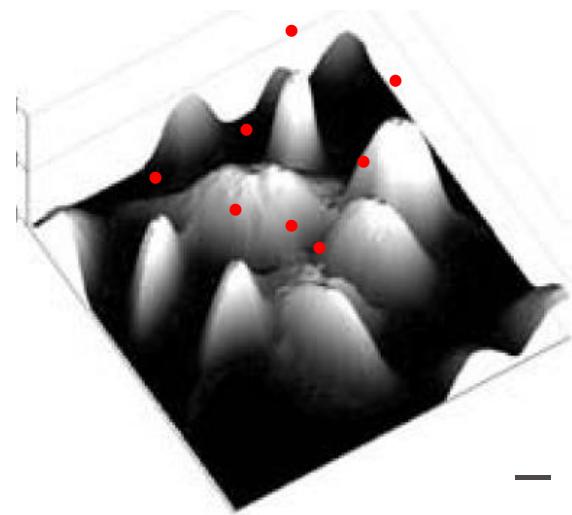
Start Test Mode Analyze Images

Welcome to CellProfiler

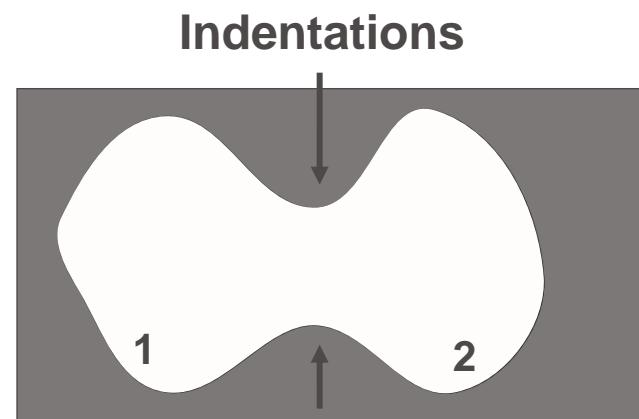
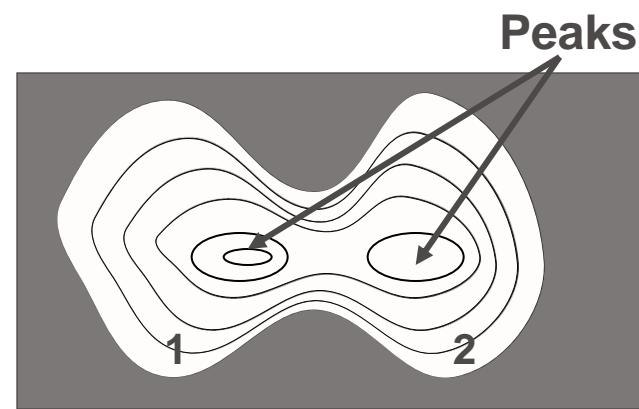
# Separating Touching Objects

## Clump identification: Two options

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

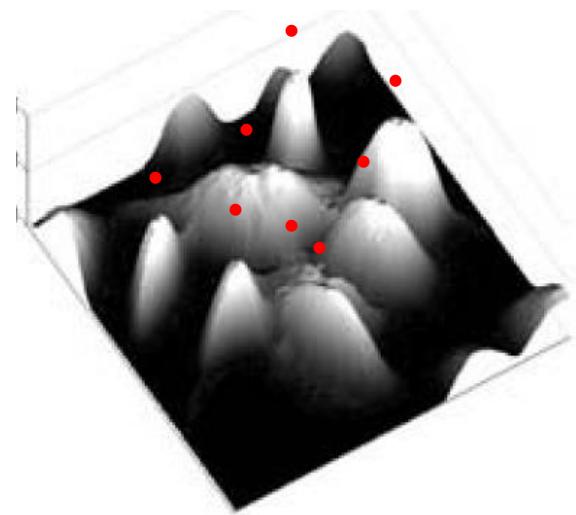


- **Shape:** Works best if objects have indentations where clumps touch (esp. if objects are round)

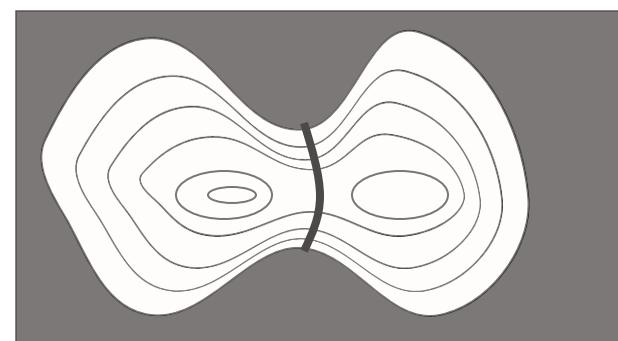
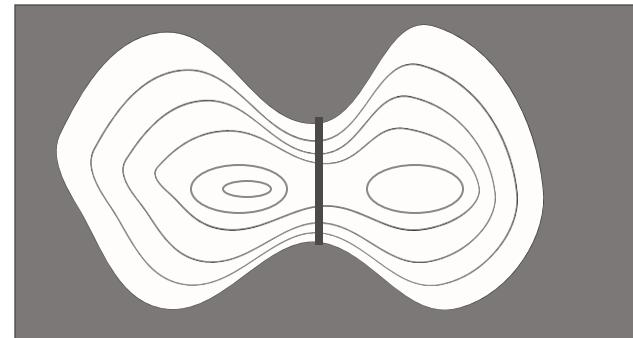


# Separating Touching Objects

## Drawing boundaries: Two options



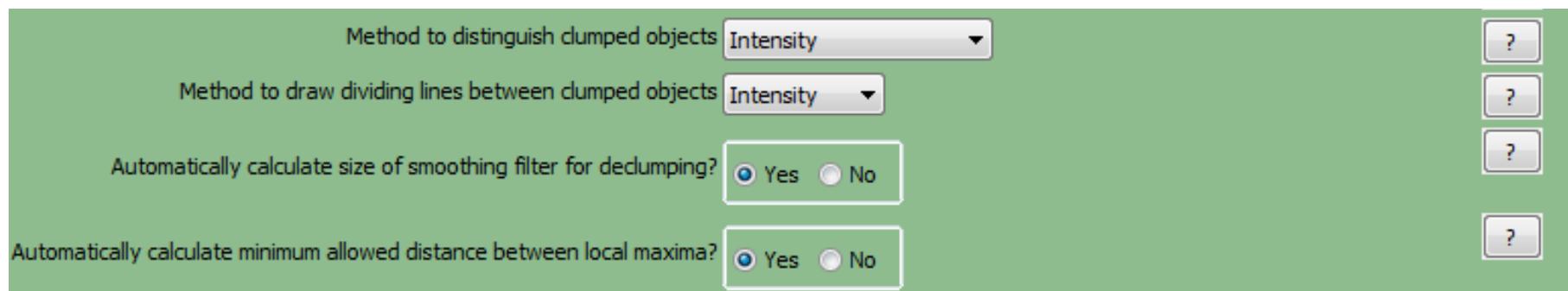
- **Distance:** Draws boundary lines midway between object centers
- **Intensity:** Draws boundary lines at dimmest line between objects



Remember to use *Test Mode* to view results of setting combinations

# Separating Touching Objects

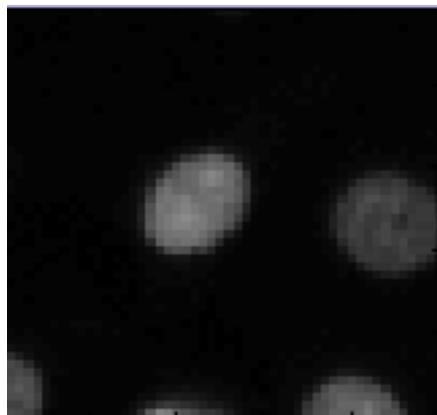
- Experiment with the declumping settings
  - Stick with the “Intensity” and “Shape” methods for now
  - Scan the whole image and look for differences
  - Try the others (“Laplacian of Gaussian”, “Propagate”) if time



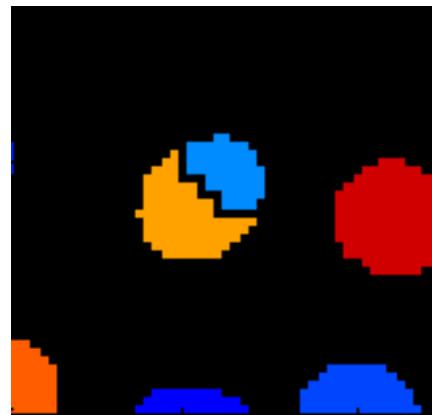
- Decent: Shape/Shape, Shape/Intensity
- Not good: Intensity/Shape

# Separating Touching Objects

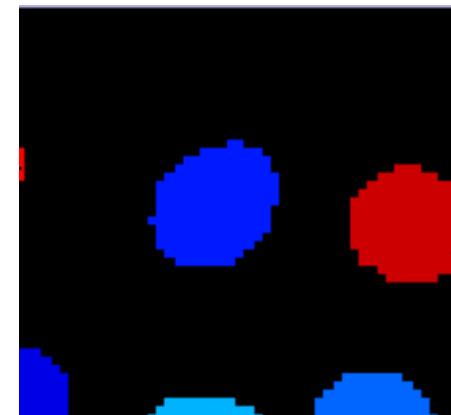
- **Additional separation settings:** Adjust these settings if objects are being incorrectly split into pieces or merged together



Original image



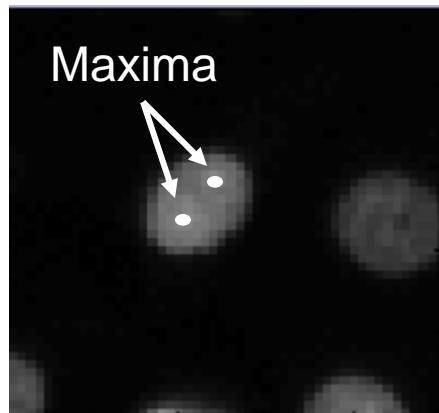
Smoothing filter  
size = 4



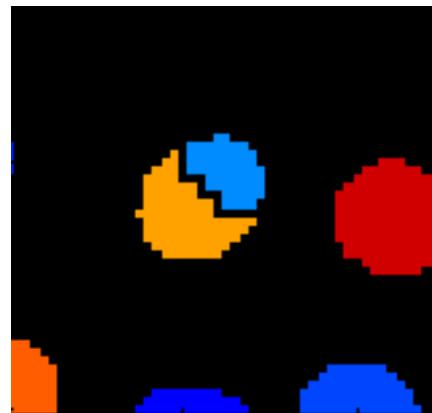
Smoothing filter  
size = 8

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

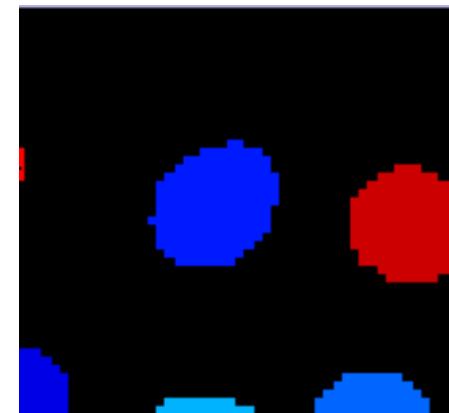
# Separating Touching Objects



Original image



Maxima  
distance = 4

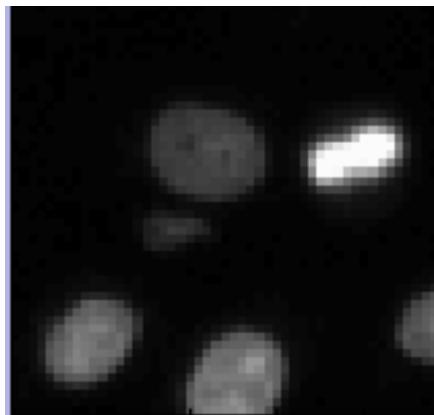


Maxima  
distance = 8

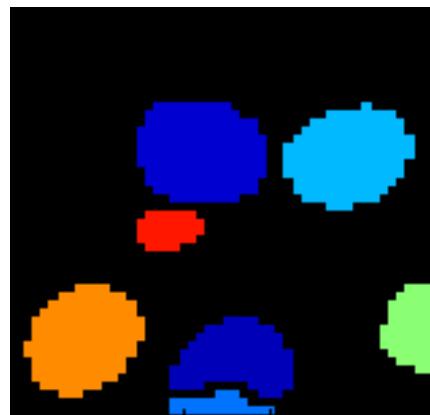
- **Suppress Local Maxima**
  - Smallest distance allowed between object intensity peaks to be considered one object rather than a clump
  - Decrease to reduce improper merging of objects in clumps

# Separating Touching Objects

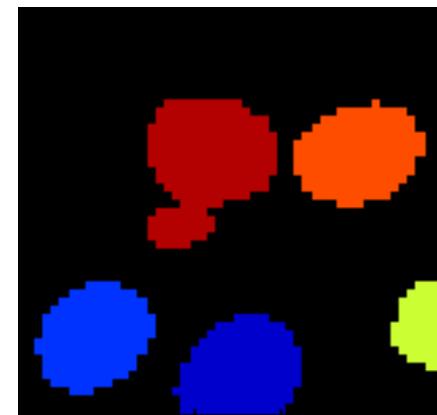
However....



Original image



Smoothing filter  
size = 4

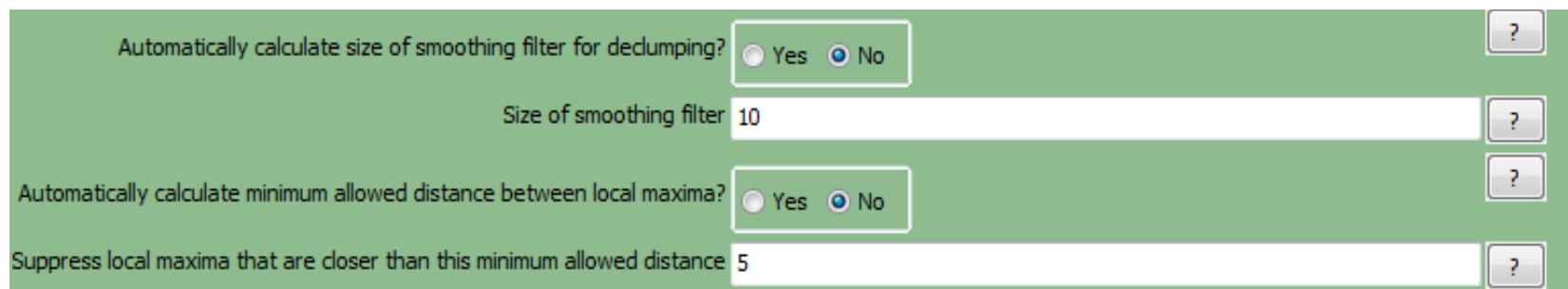


Smoothing filter  
size = 8

- Adjusting can produce more improper segmentation than it solves
- The proper settings are usually a matter of trial and error
  - The automatic settings are a good starting point, though

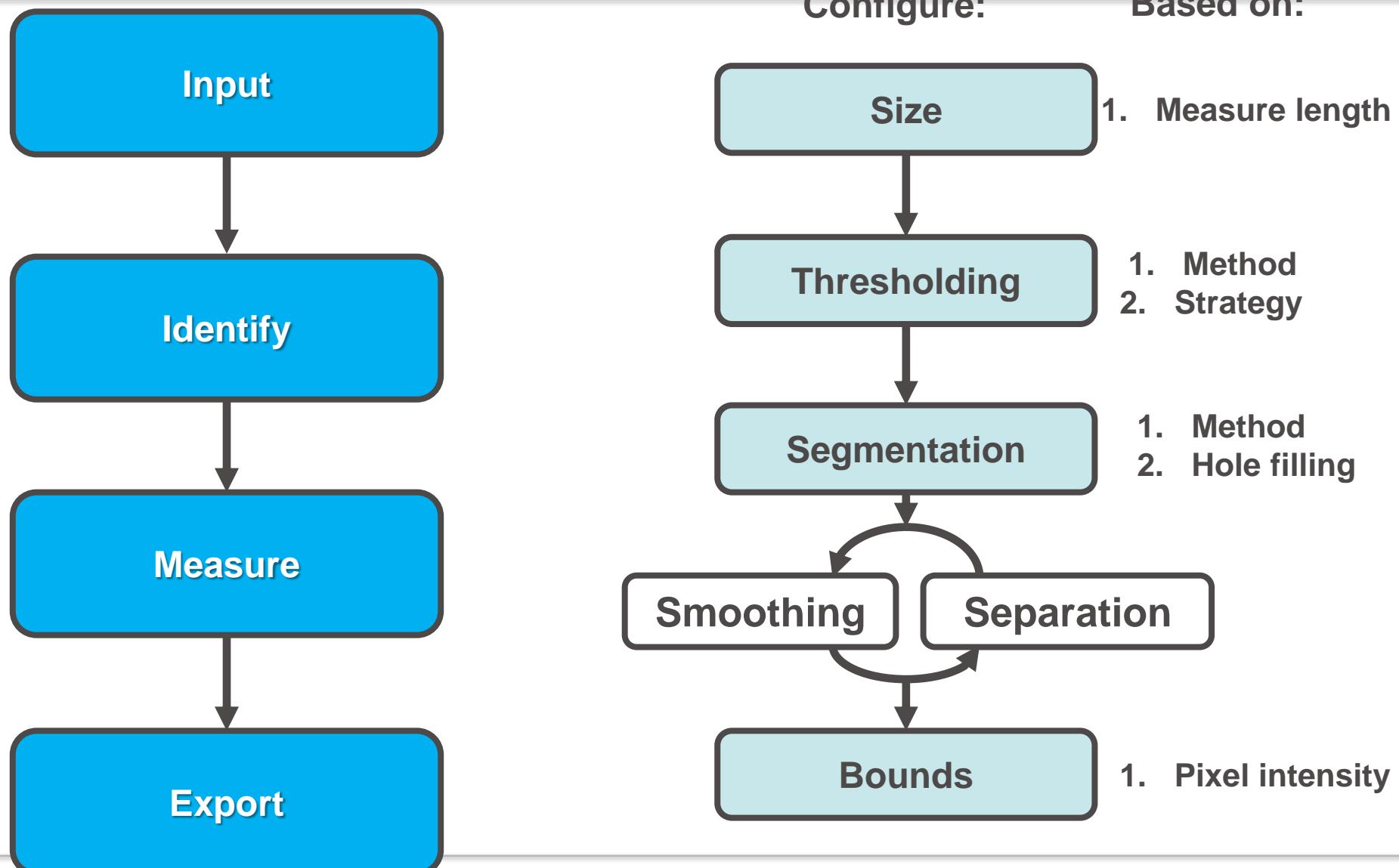
# Separating Touching Objects

- Experiment with smoothing filter, minimum allowed distances
  - Note the current values
  - Suggestion: Step from 2 to 12 pixels for both settings
  - Use “Measure length” tool to ball-park minimum allowed distance
  - Scan the whole image and look for differences



- - Greater than ~10 for smoothing: Under-segments nuclei
  - Greater than ~7 for minimum distance: Under-segments nuclei

# Typical CellProfiler Workflow



# Further Identification Adjustments

CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cppproj (C:\Trunk\ExampleImages\ExampleSBSImages)

**File Edit Test Window Data Tools Help**

**Pipeline**

**Input modules**

- Images
- Metadata
- NamesAndTypes
- Groups

**Analysis modules**

- CorrectIlluminationApply
- CorrectIlluminationApply
- IdentifyPrimaryObjects**
- IdentifyTertiaryObjects
- MeasureCorrelation
- MeasureObjectIntensity
- MeasureObjectSizeShape
- MeasureTexture
- CalculateMath
- CalculateMath
- CalculateStatistics

**Module notes**

Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation.

**Module settings**

Select the input image CorrBlue (from CorrectIlluminationApply #06) ?

Name the primary objects to be identified Nuclei ?

Typical diameter of objects, in pixel units (Min,Max) 8 28 ?

Discard objects outside the diameter range?  ?

Touching the border of the image?  ?

Threshold strategy Global ?

Thresholding method MoG ?

Approximate fraction of image covered by objects? 0.2 ?

Select the smoothing method for thresholding Automatic ?

Threshold correction factor 1.2 ?

Lower and upper bounds on threshold 0.04 1.0 ?

Method to distinguish clumped objects Intensity ?

Method to draw dividing lines between clumped objects Intensity ?

Automatically calculate size of smoothing filter for declumping?  ?

Size of smoothing filter 5 ?

**Make threshold adjustments**

**Output**

**View output settings**

Adjust modules: + - ^ v

Start Test Mode Analyze Images

Welcome to CellProfiler

The screenshot shows the CellProfiler 2.1.0 software interface. The pipeline on the left includes 'IdentifyPrimaryObjects' (selected), 'CorrectIlluminationApply' (x2), 'IdentifyTertiaryObjects', 'MeasureCorrelation', 'MeasureObjectIntensity', 'MeasureObjectSizeShape', 'MeasureTexture', 'CalculateMath' (x2), and 'CalculateStatistics'. The 'Module notes' section for 'IdentifyPrimaryObjects' states: 'Identify the nuclei from the DNA image. Some manual adjustment of the threshold correction factor and smoothing filter size is required to optimize segmentation.' The 'Module settings' section for 'IdentifyPrimaryObjects' includes fields for 'Select the input image' (set to 'CorrBlue' from 'CorrectIlluminationApply #06'), 'Name the primary objects to be identified' (set to 'Nuclei'), 'Typical diameter of objects, in pixel units (Min,Max)' (set to 8 and 28), 'Discard objects outside the diameter range?' (checked), 'Touching the border of the image?' (checked), 'Threshold strategy' (set to 'Global'), 'Thresholding method' (set to 'MoG'), 'Approximate fraction of image covered by objects?' (set to 0.2), 'Select the smoothing method for thresholding' (set to 'Automatic'), 'Threshold correction factor' (set to 1.2), 'Lower and upper bounds on threshold' (set to 0.04 and 1.0), 'Method to distinguish clumped objects' (set to 'Intensity'), 'Method to draw dividing lines between clumped objects' (set to 'Intensity'), 'Automatically calculate size of smoothing filter for declumping?' (unchecked), and 'Size of smoothing filter' (set to 5). A large black callout box covers the 'IdentifyPrimaryObjects' module. A red box highlights the 'Threshold correction factor' and 'Lower and upper bounds on threshold' settings, with a black arrow pointing to the 'Threshold correction factor' field.

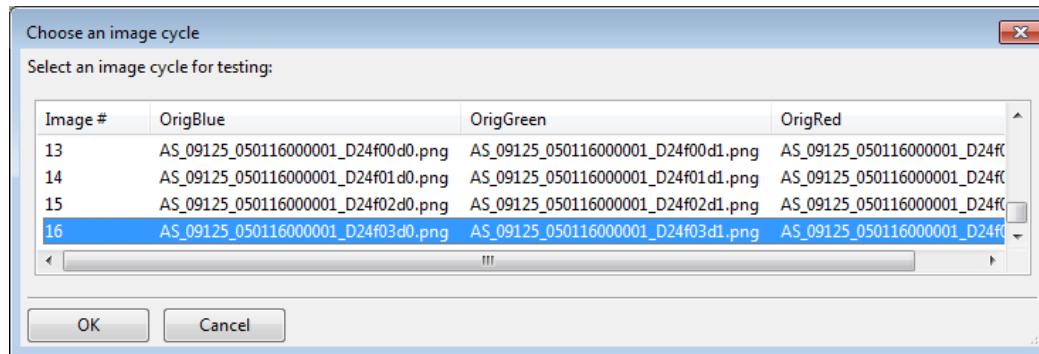
# Further Identification Adjustments

---

- **Correction factor**
  - Multiplication factor applied to threshold
  - Adjusts threshold stringency/leniency
  - Setting this factor is empirical
  
- **Upper/lower bounds**
  - Set safety limits on automatic threshold to guard against false positives
  - Helpful for unexpected images: Empty wells, images with dramatic artifacts, etc

## Further Identification Adjustments

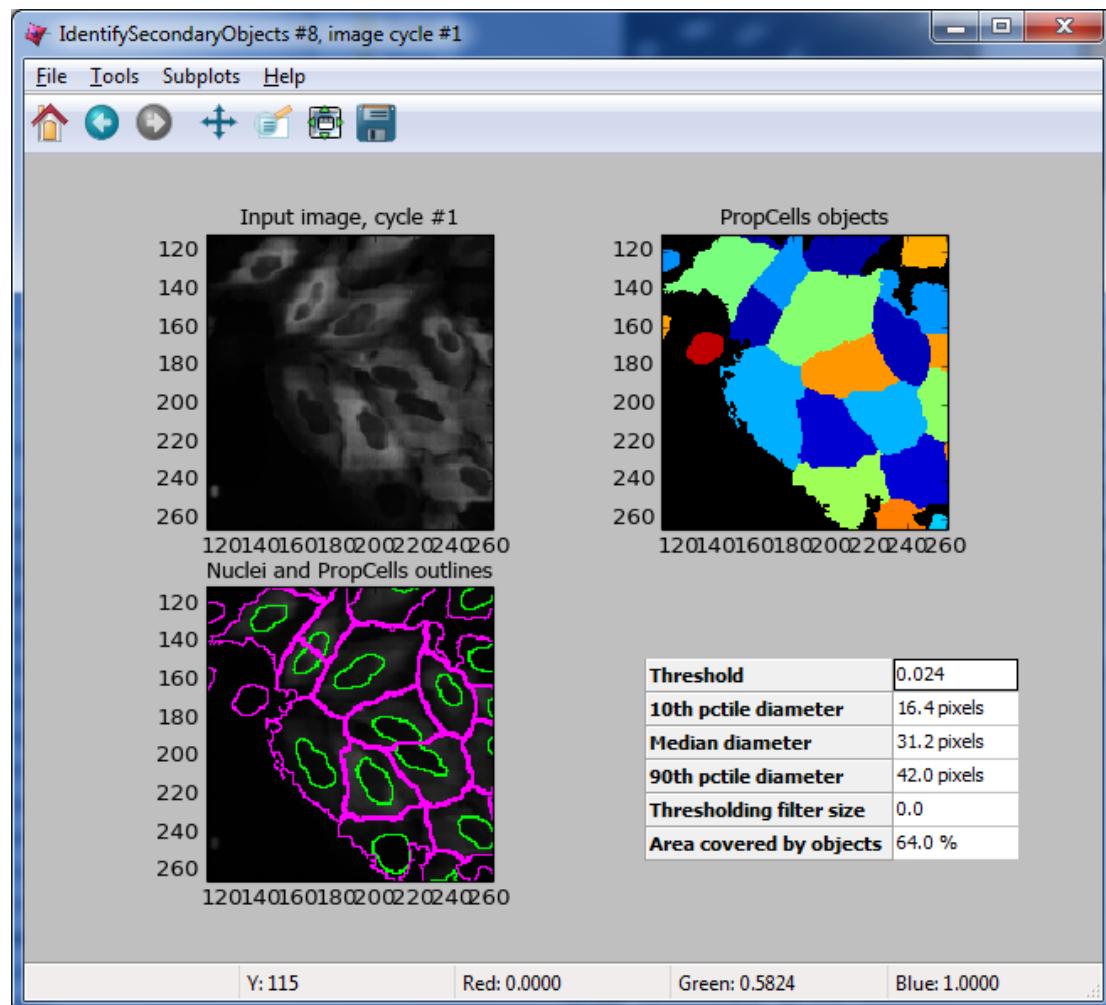
- From *Test > Choose Image set*, select the last image set



- Run the pipeline. How does the nuclei identification look? Why?
- What appears to be a good lower bound?
  - Using the intensity tool
  - Using the histogram
- Adjust the lower bound, re-run the module
- Confirm your settings: Use Test menu to go back to the 1<sup>st</sup> image, run the pipeline

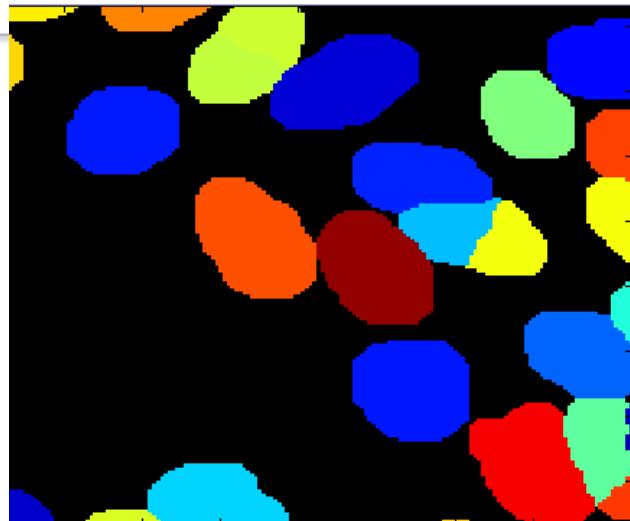
# Secondary Object Identification

- **Goal:** Identify cell boundaries by “growing” primary objects
  - Nuclei typically more uniform in shape, more easily separated than cells
- **Approach:** Segment nuclei → Seeds for cell segmentation by using a cell stain channel

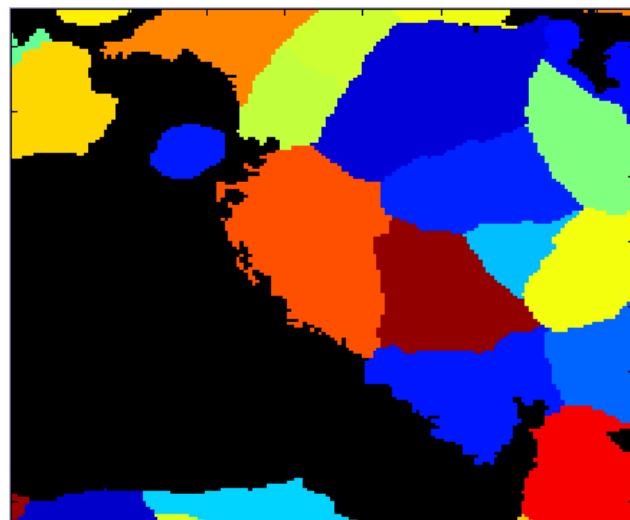


# Secondary Object Identification: Methods

- Distance-N: Ignores image information
  - Useful when cell stain is absent
- Watershed, Propagate, Distance-B: Uses image information
  - Finds dividing lines between objects and background / neighbors

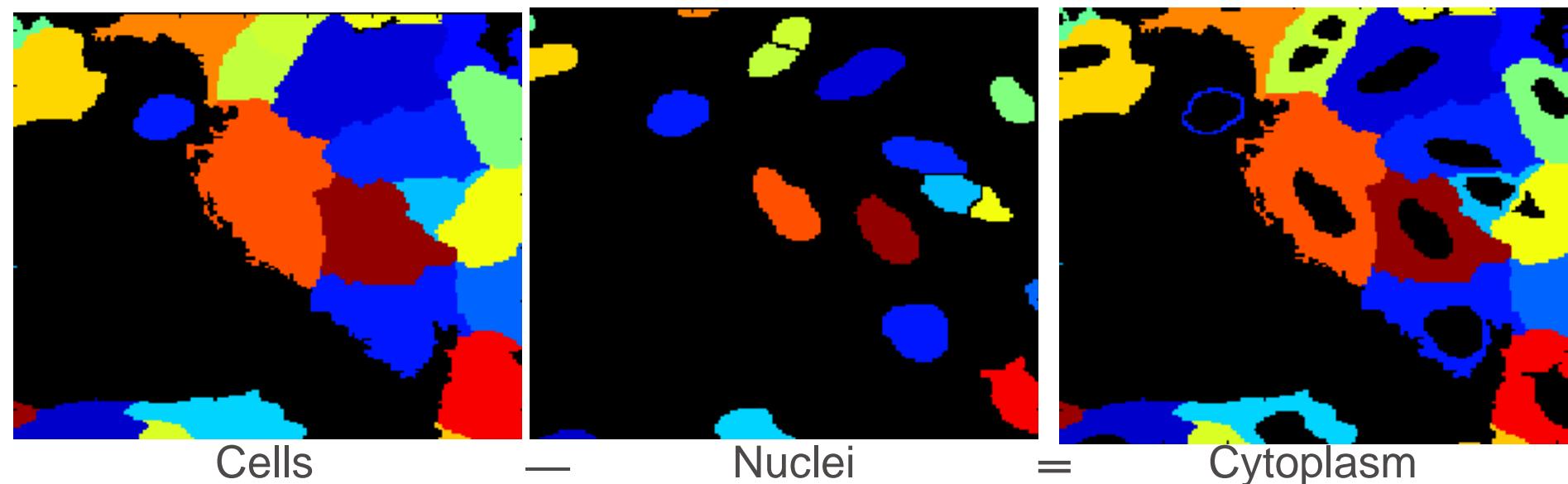


Distance-N



# Tertiary Object Identification

- **Goal:** Identify *tertiary objects* by removing the primary objects from secondary objects
  - “Subtract” the nuclei objects from cell objects to obtain cytoplasm



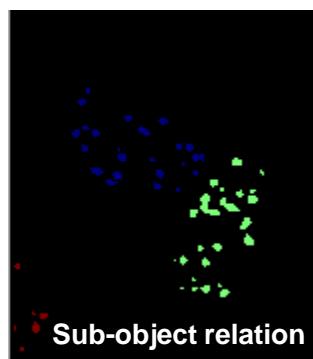
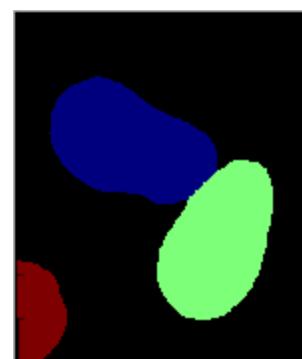
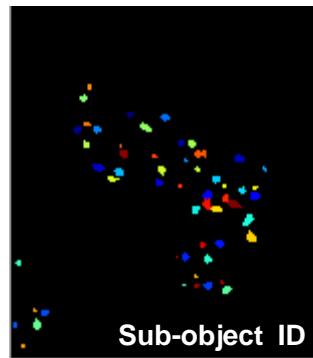
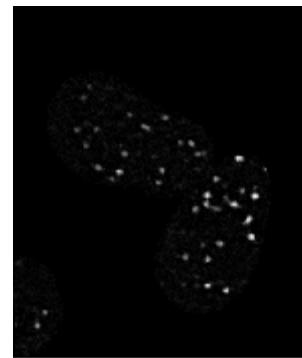
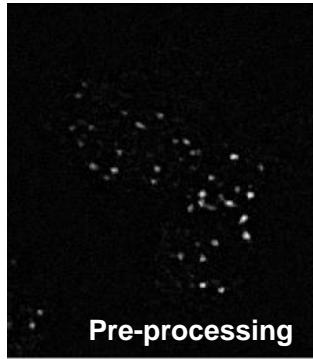
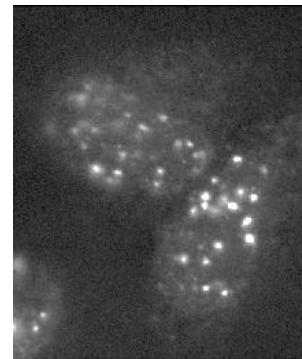
# Secondary and Tertiary Object Identification

---

- In *IdentifySecondaryObjects*, change the input image to “CropRed”
- Experiment with the secondary object identification methods
- Try different thresholding methods, if time permits
- Which method works best in this case?
  - Good: Propagation, Distance-B
  - Decent: Watershed-Image, Watershed-Gradient
  - Not good: Distance-N
- Press Step button to execute *IdentifyTertiaryObjects*

# Identifying Subcellular Structures

- With appropriate markers, other subcellular compartments can be labeled
- These can be identified using the same methods already mentioned
- Consider using enclosing object as mask for better pre-processing, thresholding
- Make sure to assign subfeatures to enclosing objects

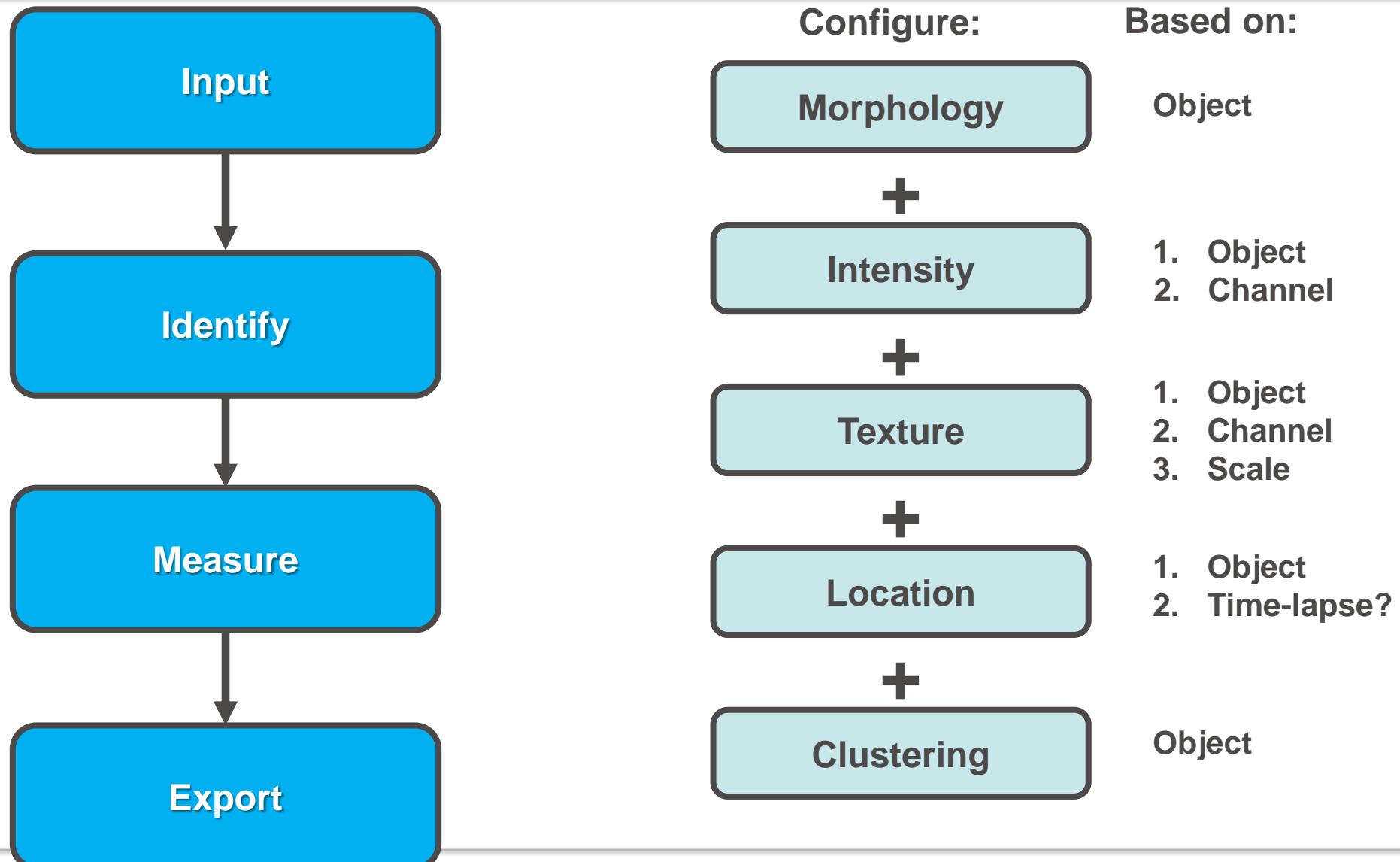


# Identifying Subcellular Structures

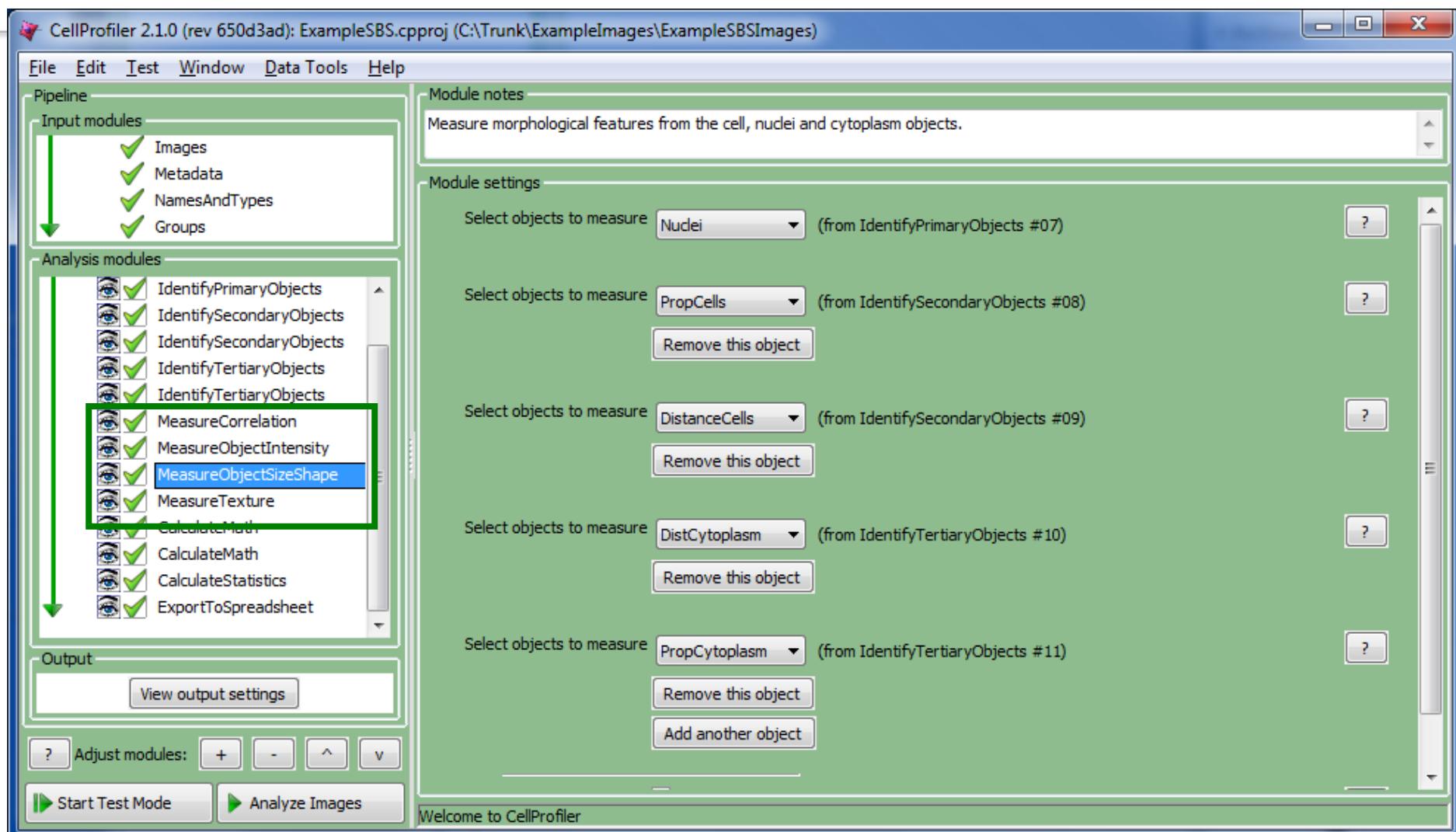
---

- Add another *IdentifyPrimaryObjects* module
  - Position it after *IdentifyTertiaryObjects*
  - Re-enter Test mode if you need to
  - Adjust settings
    - Select “CropGreen” as the input image
    - Enter “pH3” as the primary object name
- Add *RelateObjects* module
  - Position it after *IdentifyPrimaryObjects*
  - Adjust settings
    - Set child objects as “pH3”
    - Set parent objects as “Nuclei”
- Set new pause after *RelateObjects*, and run the pipeline

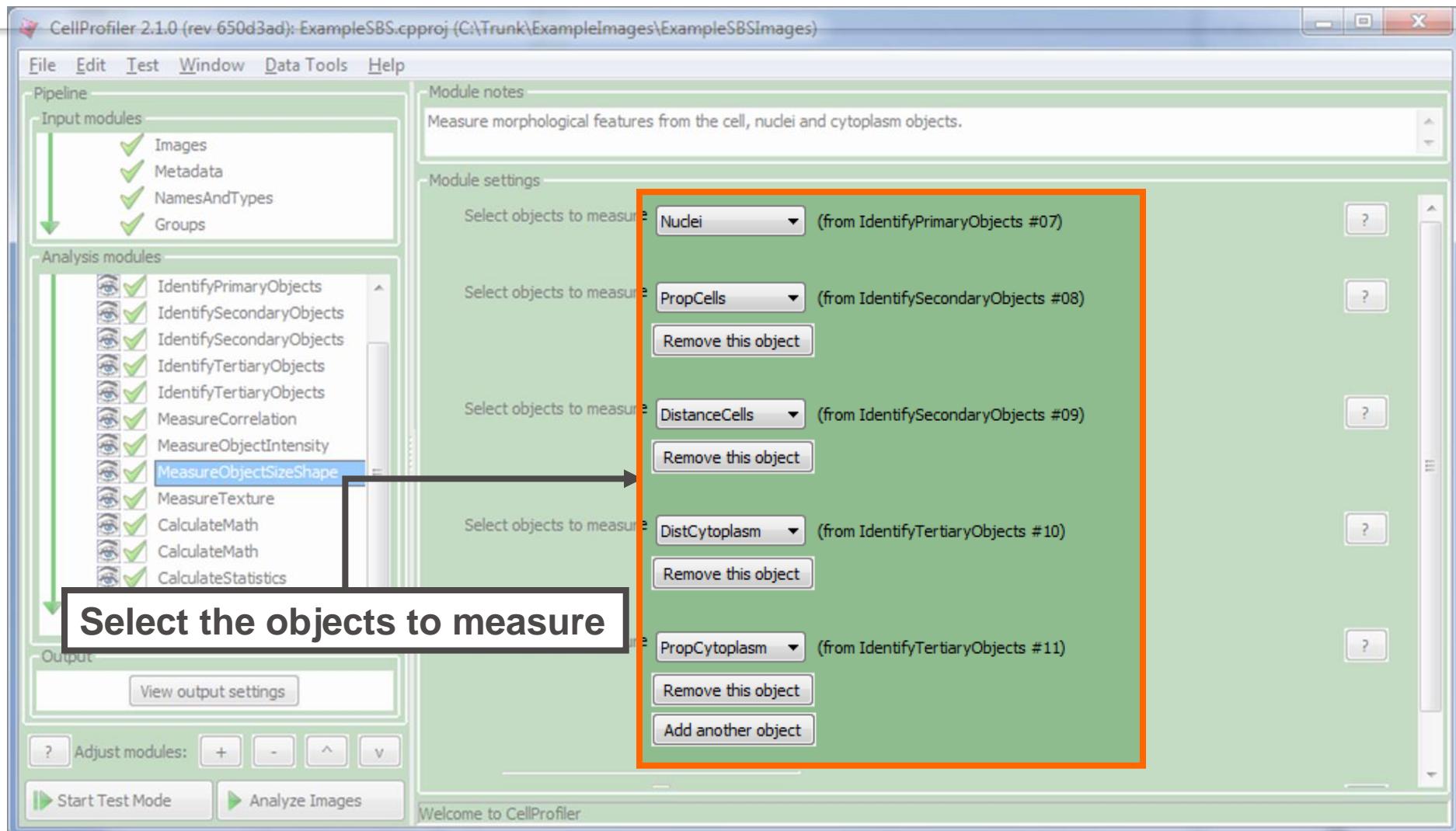
# Typical CellProfiler Workflow



# Measurement Modules: Morphology



# Measurement Modules: Morphology



# Measurement Modules: Intensity

CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cppproj (C:\Trunk\ExampleImages\ExampleSBSImages)

**Pipeline**

**Input modules**

- Images
- Metadata
- NamesAndTypes
- Groups

**Analysis modules**

- IdentifyPrimaryObjects
- IdentifySecondaryObjects
- IdentifySecondaryObjects
- IdentifyTertiaryObjects
- IdentifyTertiaryObjects
- MeasureCorrelation
- MeasureObjectIntensity**
- MeasureObjectSizeShape
- MeasureTexture
- CalculateMath
- CalculateMath
- CalculateStatistics
- ExportToSpreadsheet

**Output**

**Module notes**

Measure intensity features from nuclei, cell and cytoplasm objects against the corrected DNA and GFP images.

**Module settings**

Select an image to measure: CorrBlue (from CorrectIlluminationApply #06)

Select an image to measure: CorrGreen (from CorrectIlluminationApply #05)

Remove this image

Add another image

Select objects to measure: Nuclei (from IdentifyPrimaryObjects #07)

Select objects to measure: PropCells (from IdentifySecondaryObjects #08)

Remove this object

Select objects to measure: DistanceCells (from IdentifySecondaryObjects #09)

Remove this object

Select objects to measure: DistCytoplasm (from IdentifyTertiaryObjects #10)

View output settings

Adjust modules: + - ^ v

Start Test Mode Analyze Images

Welcome to CellProfiler

# Measurement Modules: Intensity

Select the image to measure from

Select the objects to measure

The screenshot shows the CellProfiler 2.1.0 software interface. The main window title is "CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cppproj (C:\Trunk\ExampleImages\ExampleSBSImages)". The menu bar includes File, Edit, Test, Window, Data Tools, and Help. The left sidebar lists "Analysis modules" such as IdentifyPrimaryObjects, IdentifySecondaryObjects, IdentifyTertiaryObjects, MeasureCorrelation, MeasureObjectIntensity (which is selected), MeasureObjectSizeShape, MeasureTexture, CalculateMath, CalculateMath, CalculateStatistics, and ExportToSpreadsheet. The "Module settings" panel on the right contains several sections: "Select an image to measure" (set to CorrBlue, from CorrectIlluminationApply #06), "Select objects to measure" (set to Nuclei, from IdentifyPrimaryObjects #07), "Select objects to measure" (set to PropCells, from IdentifySecondaryObjects #08), "Select objects to measure" (set to DistanceCells, from IdentifySecondaryObjects #09), and "Select objects to measure" (set to DistCytoplasm, from IdentifyTertiaryObjects #10). Arrows and boxes highlight the "Select the image to measure from" section and the "Select the objects to measure" section.

# Measurement Modules: Texture

CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cpproj\* (C:\Trunk\ExampleImages\ExampleSBSImages)

**Pipeline**

**Input modules**

- Images
- Metadata
- NamesAndTypes
- Groups

**Analysis modules**

- IdentifyPrimaryObjects
- IdentifySecondaryObjects
- IdentifySecondaryObjects
- IdentifyTertiaryObjects
- IdentifyTertiaryObjects
- MeasureCorrelation
- MeasureObjectIntensity
- MeasureObjectSizeShape
- MeasureTexture
- CalculateMath
- CalculateMath
- CalculateStatistics
- ExportToSpreadsheet

**Module notes**  
Measure texture features from the cell, nuclei and cytoplasm objects.

**Module settings**

Select an image to measure: CorrGreen (from CorrectIlluminationApply #05)

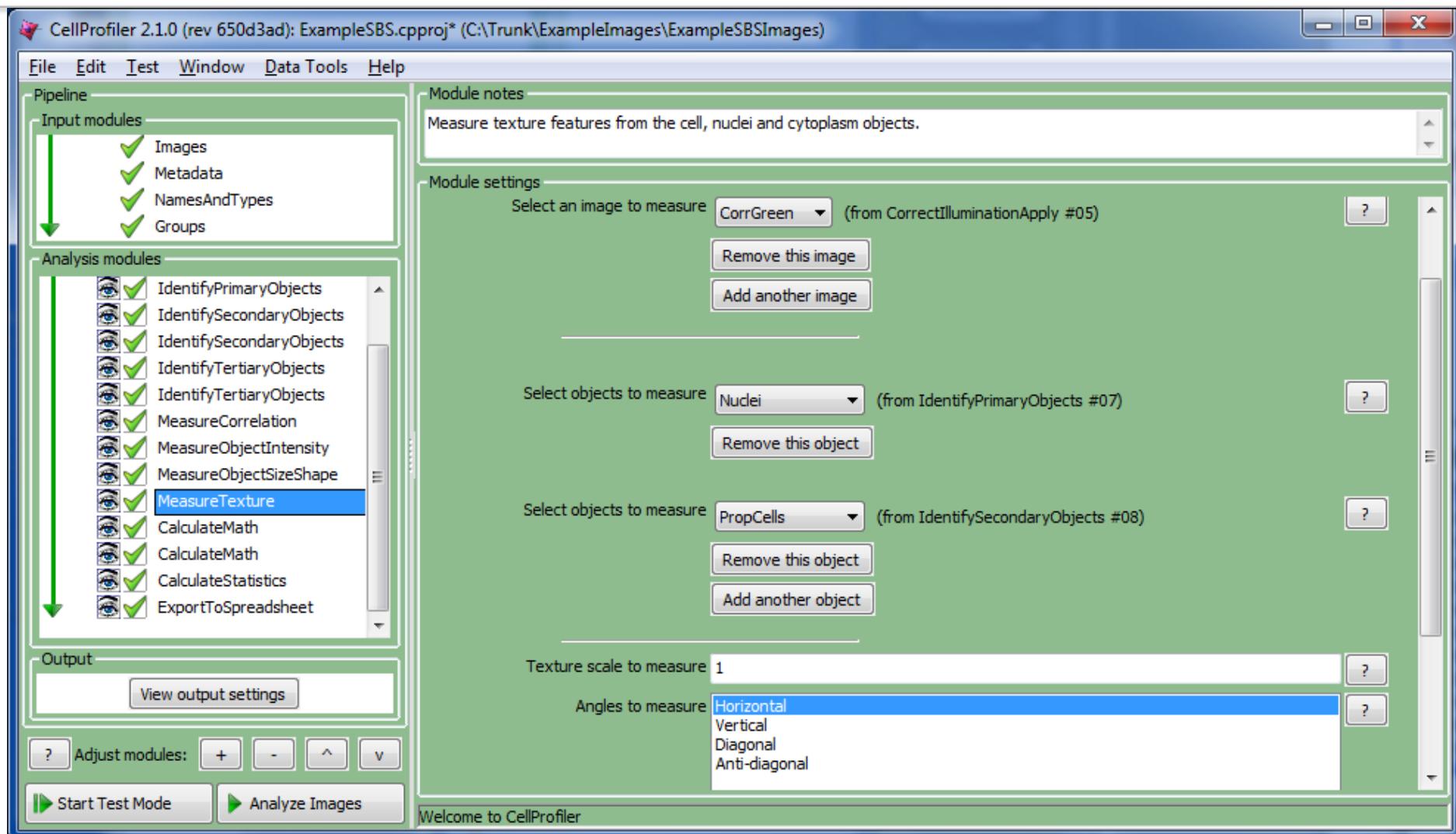
Select objects to measure: Nuclei (from IdentifyPrimaryObjects #07)

Select objects to measure: PropCells (from IdentifySecondaryObjects #08)

Texture scale to measure: 1

Angles to measure: Horizontal (selected), Vertical, Diagonal, Anti-diagonal

Welcome to CellProfiler



# Measurement Modules: Texture

CellProfiler 2.1.0 (rev 650d3ad): ExampleSBS.cproj\* (C:\Trunk\ExampleImages\ExampleSBSImages)

Select the image to measure from

Input modules

- Images
- Metadata
- NamesAndTypes
- Groups

Analysis modules

- IdentifyPrimaryObjects

Select the objects to measure

Module settings:

Select an image to measure: CorrGreen (from CorrectIlluminationApply #05)

Remove this image Add another image

Select objects to measure: Nuclei (from IdentifyPrimaryObjects #07)

Remove this object

Select objects to measure: PropCells (from IdentifySecondaryObjects #08)

Remove this object Add another object

Texture scale to measure: 1

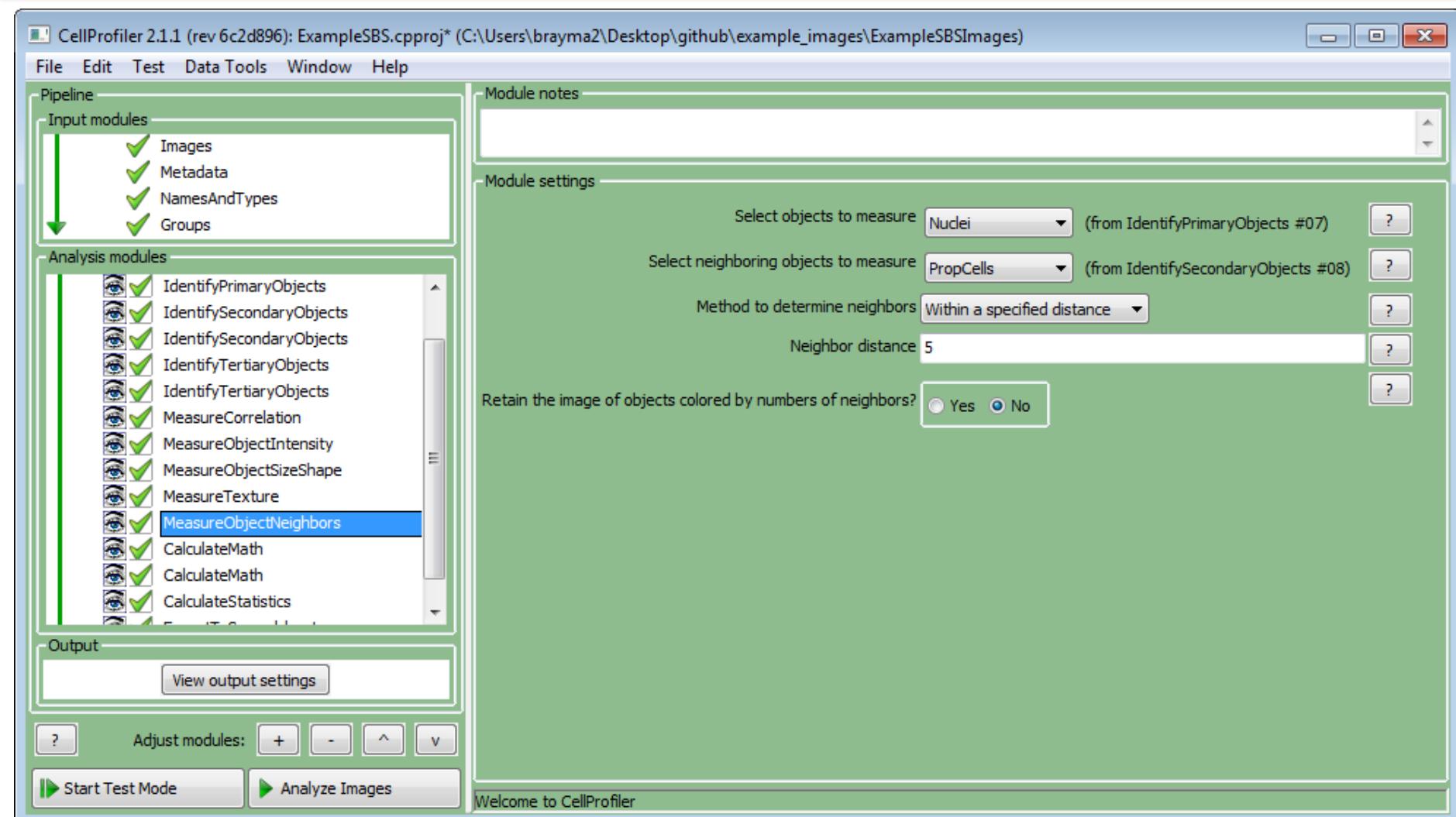
Angles to measure: Horizontal (selected), Vertical, Diagonal, Anti-diagonal

Select the spatial scale

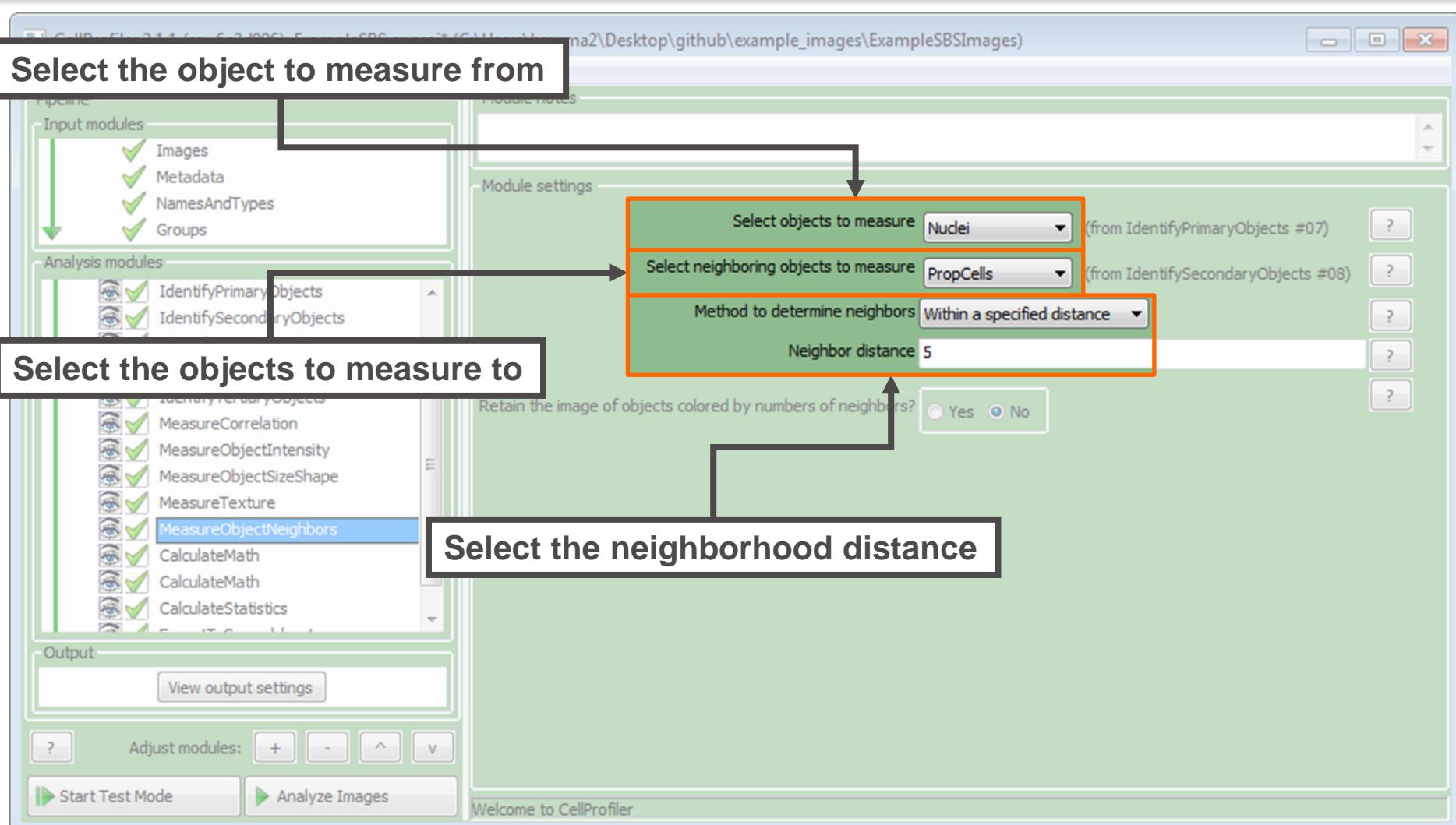
Adjust modules: + - ^ v

Welcome to CellProfiler

# Measurement Modules: Clustering

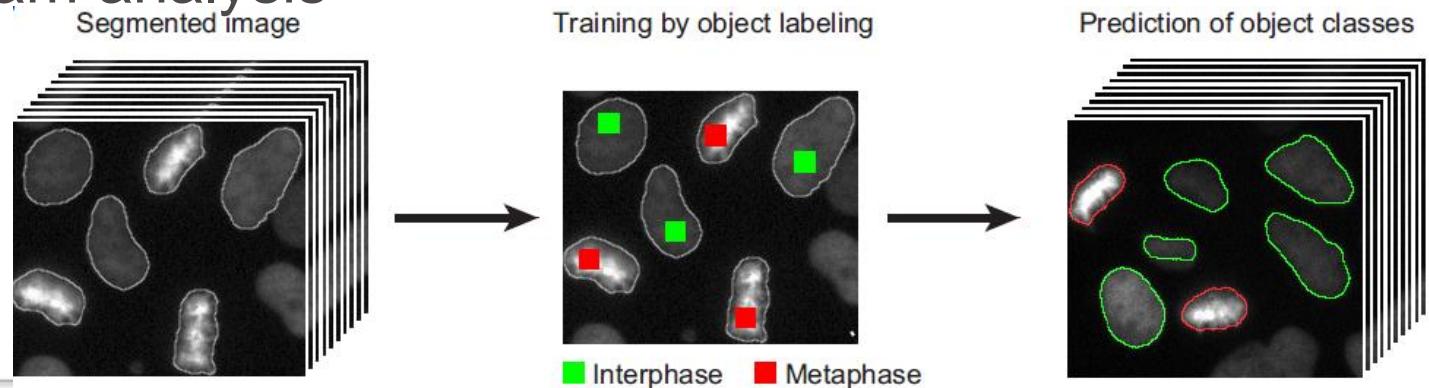


# Measurement Modules: Clustering

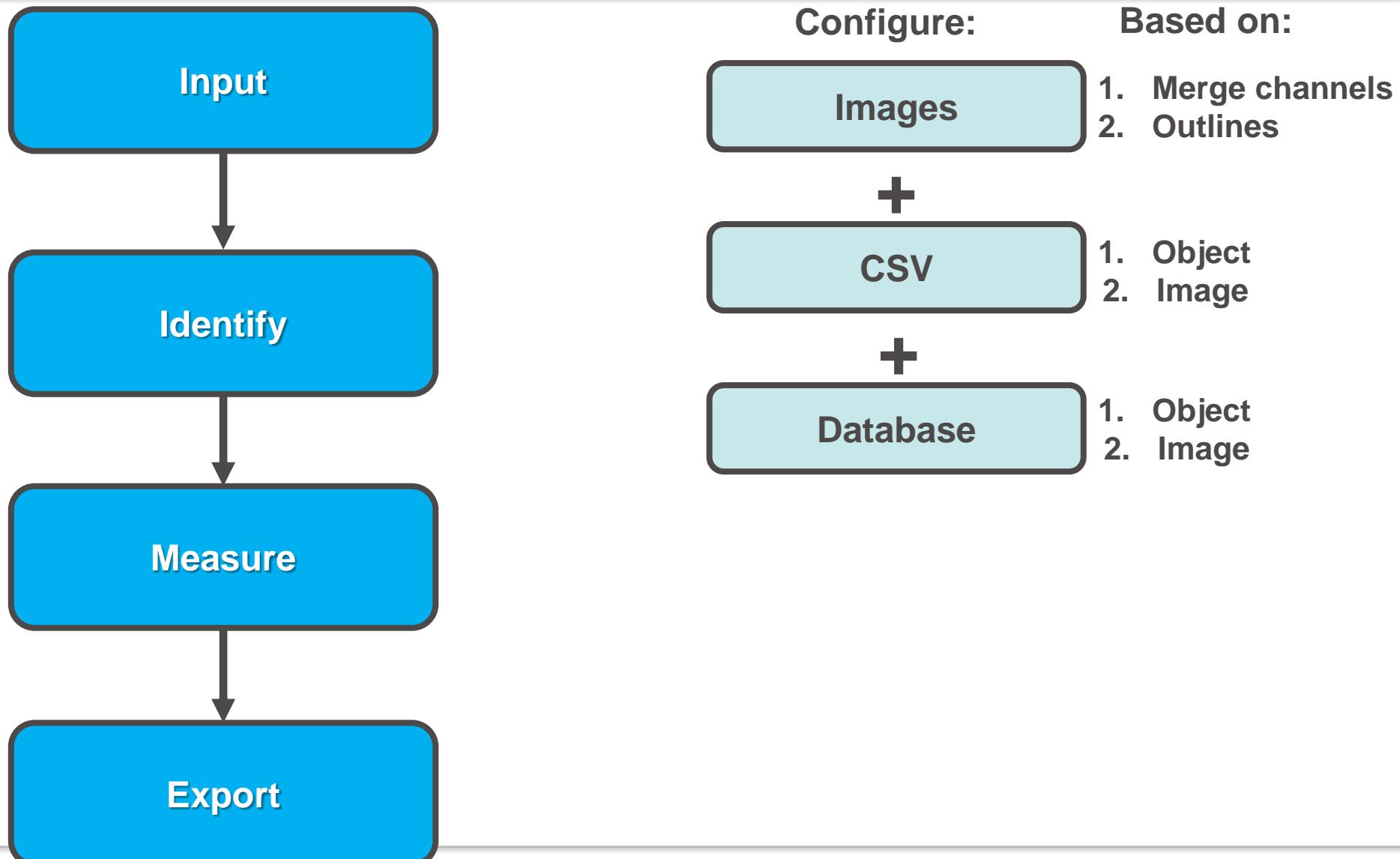


# Combinations of Measurements

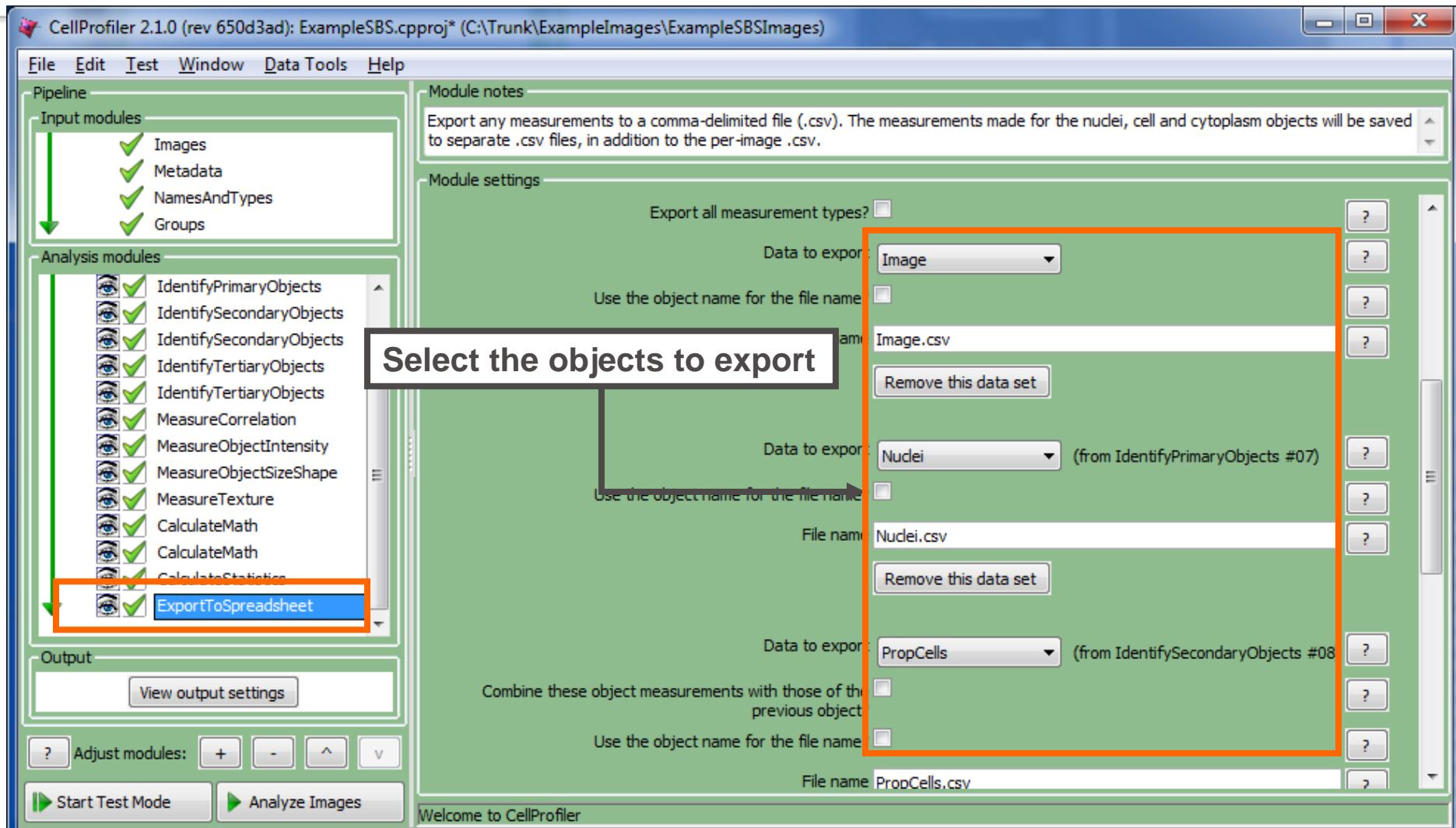
- Phenotype identification may be difficult if hand-selecting from a limited measurement set
- Machine learning (ML) approaches can identify phenotypes from a combination of measurements
- Some measurements (e.g., texture) are hard to interpret as readouts but are excellent fodder for ML approaches to downstream analysis



# Typical CellProfiler Workflow



# Data Export Modules



- User may output images or image measurements

# Data Export Modules

---

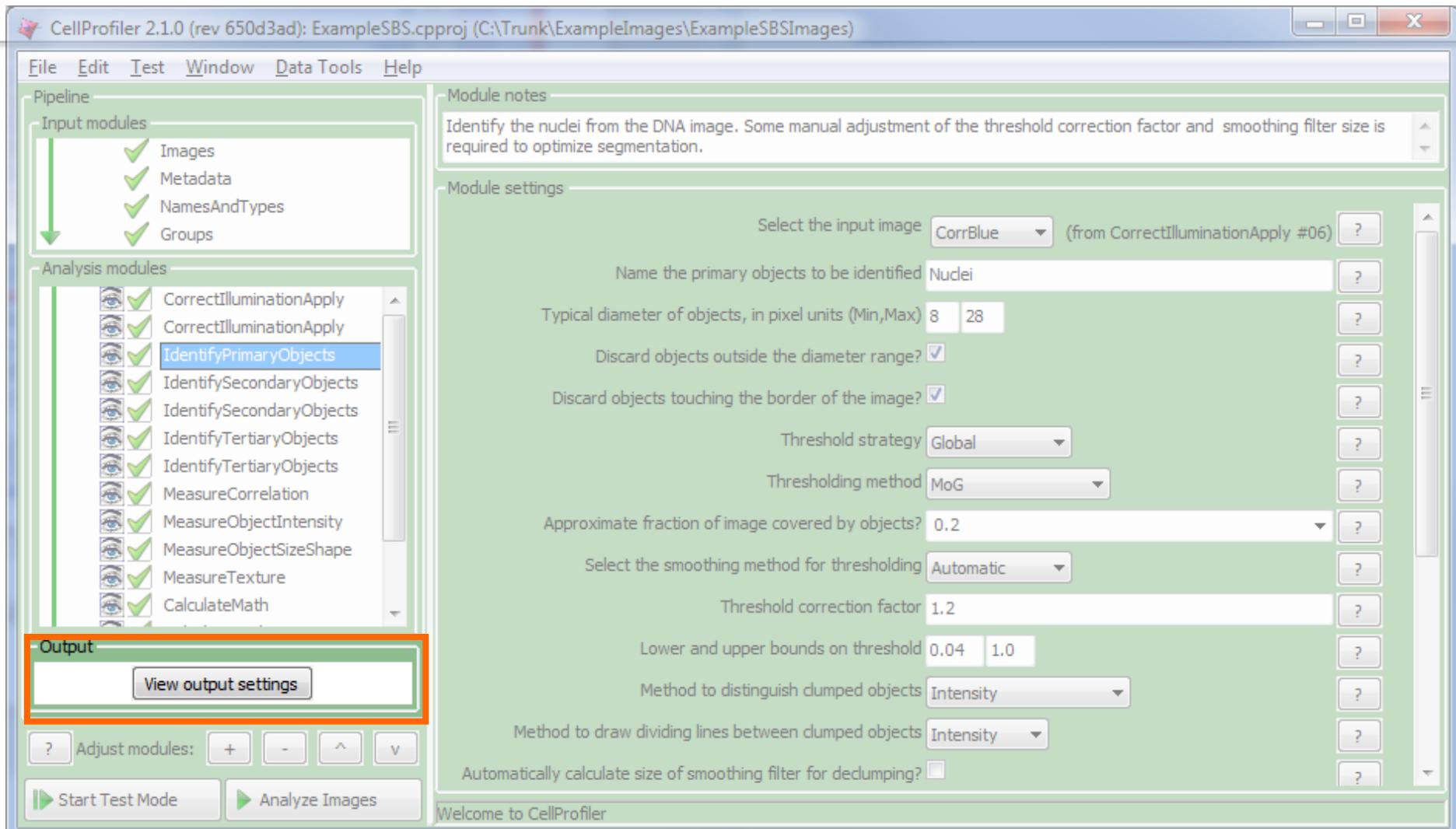
- **Goal:** Retain images of intermediate image processing steps for quality control or save measurements for later analysis and exploration
- **SaveImages:** Writes an image to a file
  - Intermediate images in the pipeline are not saved unless requested
  - Choice of many image formats to write → module can be used as an image format converter
- **ExportToSpreadsheet:** Export measurements as a comma-separated file readable by spreadsheet programs
- **ExportToDatabase:** Export measurements as a per-object and per-table plus configuration file for a MySQL or SQLite database

# Data Export

---

- Remove the **Crop** modules, re-set identification module inputs
- Remove the **SaveImages** module
  - Not yet supported on HCSIA
- Leave settings on **ExportToSpreadsheet** as-is
  - Including this module is required if you want per-well results

# Before The Analysis...



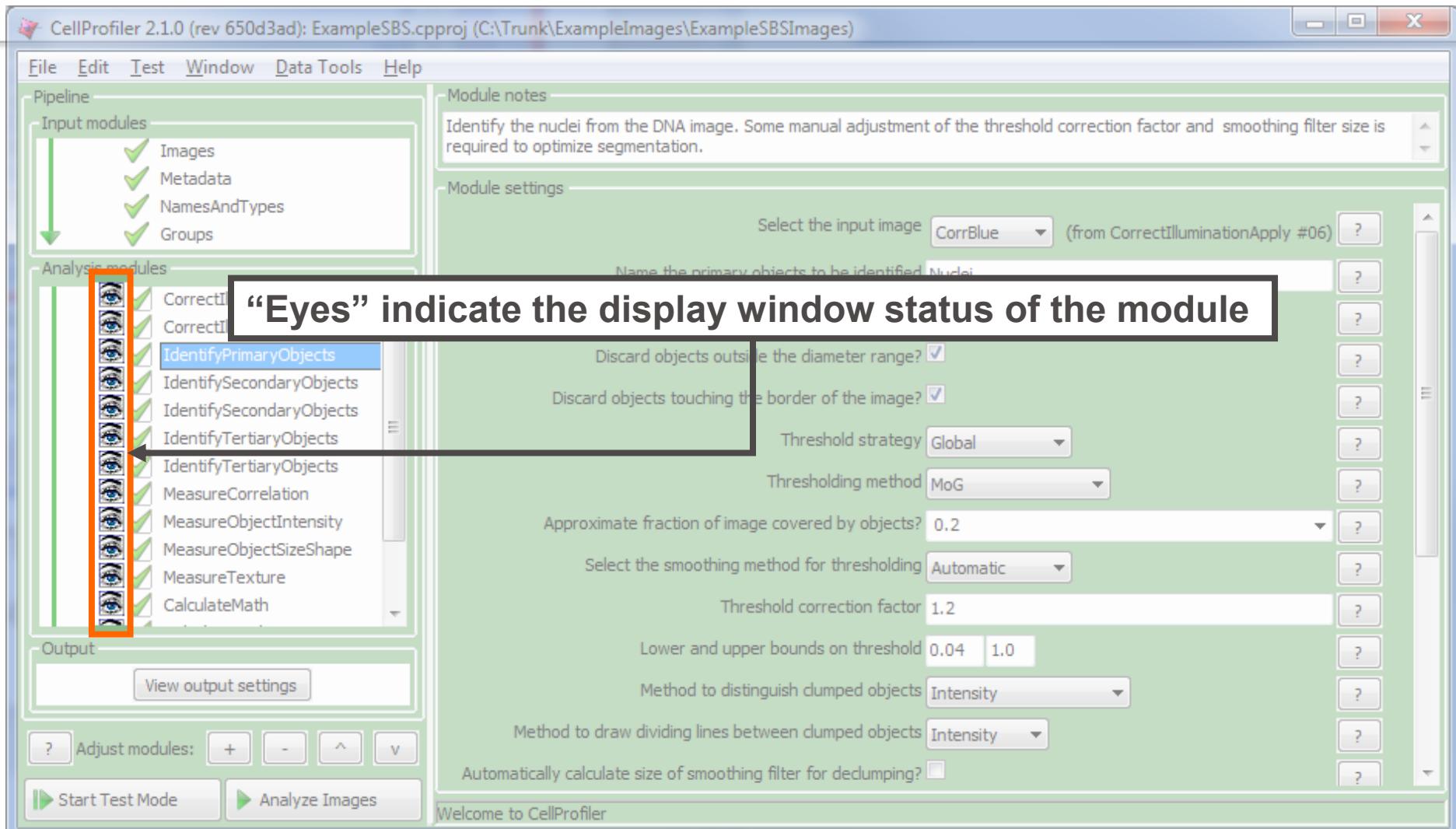
- **Output panel:** Specify the location to place outputs
  - Spreadsheets of measurements, saved images, etc

## Before the Analysis...

---

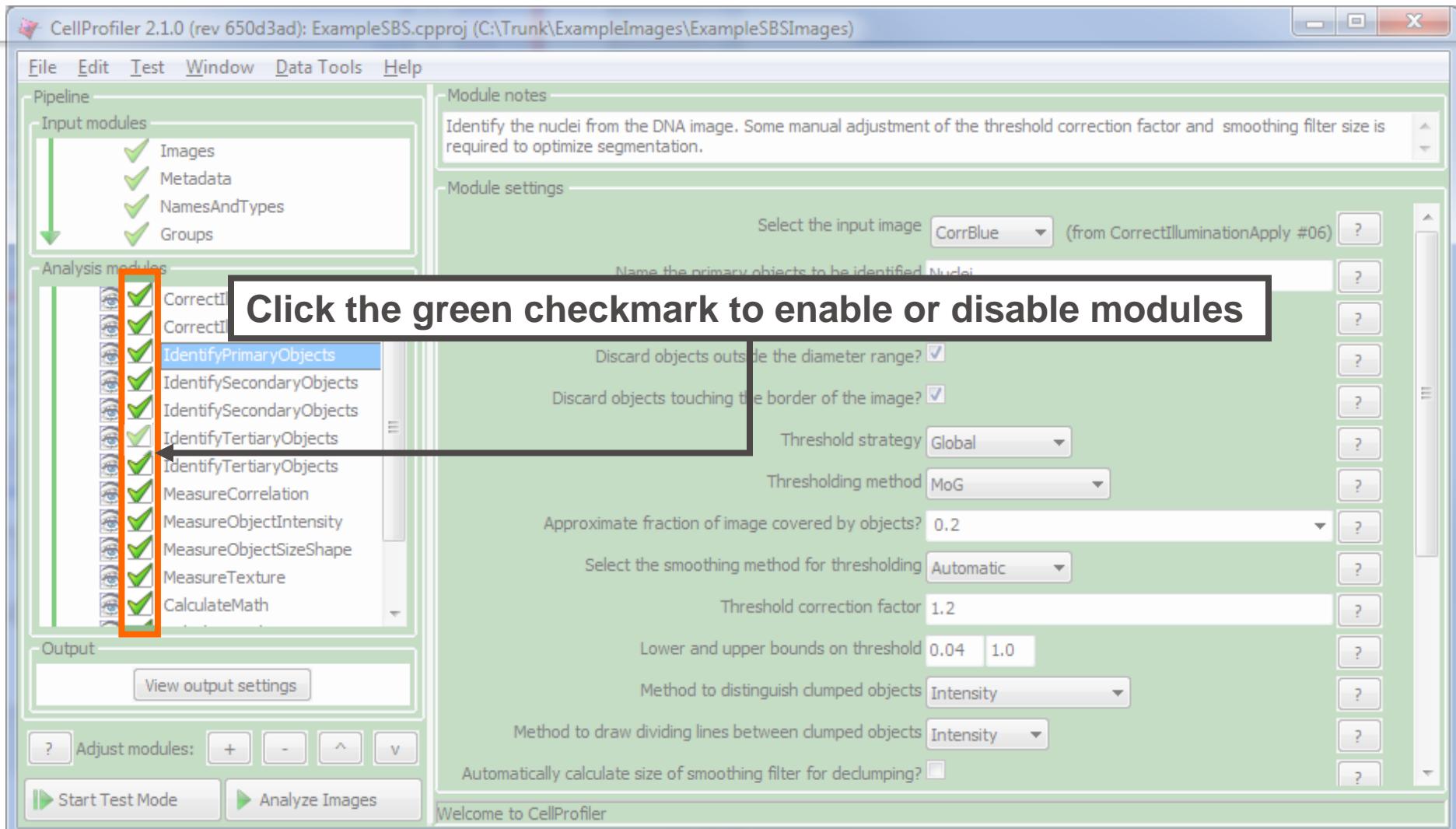
- Saving projects
    - Saves the pipeline, image locations, metadata, etc. (.cpproj)
    - Quick start-up, but not always portable
  - Exporting pipelines
    - Saves just the pipeline (.cppipe)
    - More portable
- vs.
- If publishing, consider submitting your pipeline as **supplemental material**

# Before The Analysis...



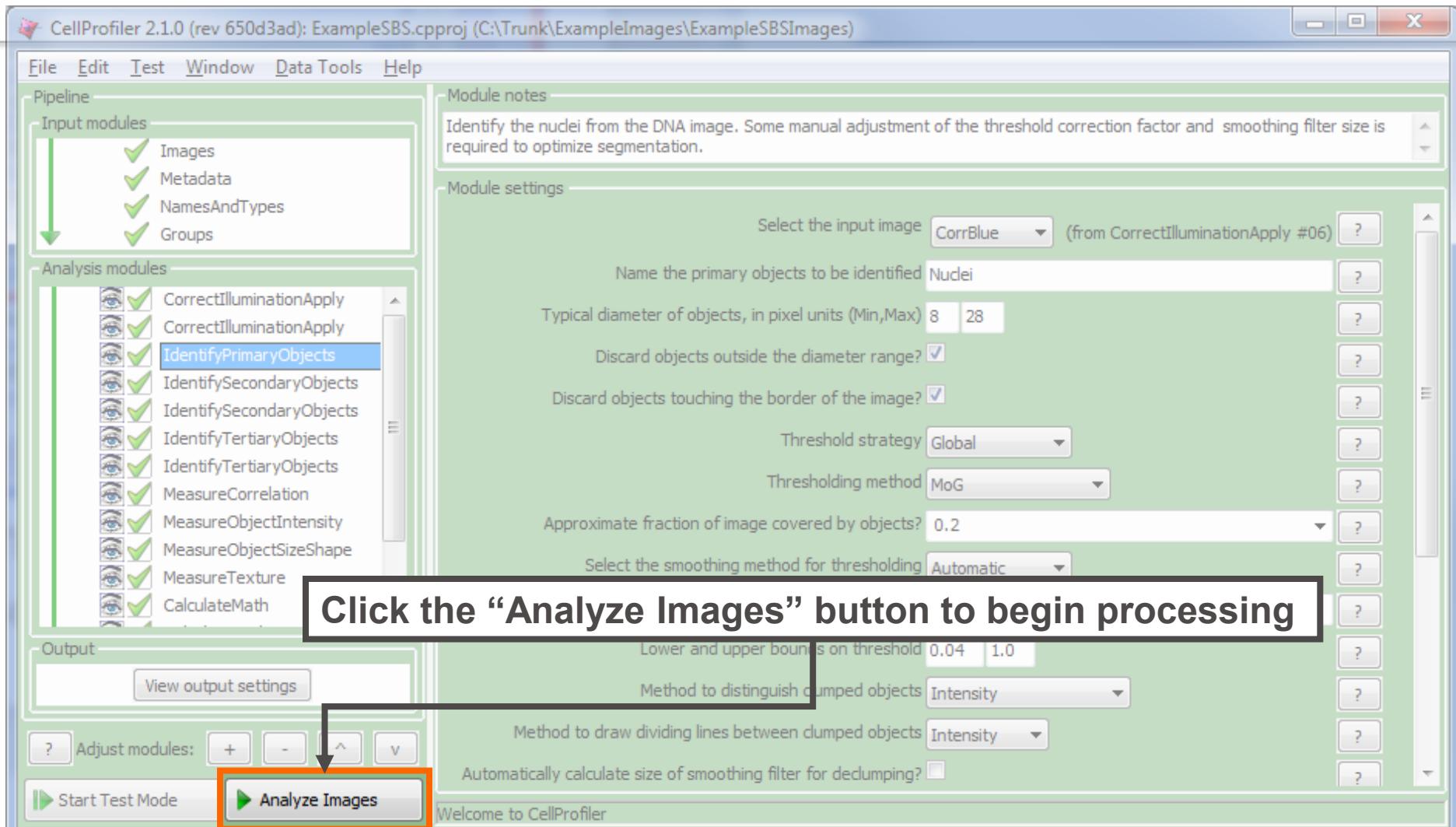
- Close the “eyes” using the Window menu (*Hide all windows on run*)
  - Saves run time and memory usage

# Before The Analysis...



- Disabled module is grayed out, effectively removed from the run

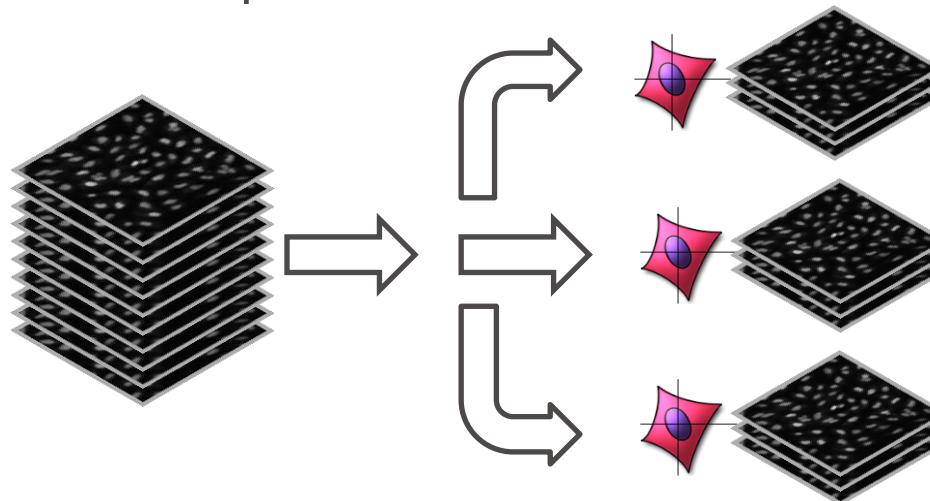
# Before The Analysis...



- Analysis can be paused or halted with additional buttons

# Multiprocessing in CellProfiler

- The more CPUs your computer has, the better...
- Default behavior
  - Number of workers = Number of computing cores
  - Without the GUI (“headless”): One worker, use batch processing to distribute multiple workers



- Number of workers can be set under *File > Preferences*

# Final Notes

- Where to get help
  - Access help from the CellProfiler main window
  - Ask for help on the CellProfiler.org forum

The screenshot shows a web browser window with the URL [forum.cellprofiler.org](http://forum.cellprofiler.org) in the address bar. The page title is "CellProfiler". The navigation bar includes links for "Most Visited", "Main Page - ImagingP...", "Getting Started", "CellProfiler cell image ...", "CellProfiler Forum", and "CellProfiler 2.0 Batch s...". The top right features "Sign Up" and "Log In" buttons, along with a search icon and a menu icon.

The main content area has a header with "all categories", "Latest" (which is highlighted in red), "Top", and "Categories" buttons. Below this is a table with columns: Topic, Category, Users, Replies, Views, and Activity. The first post in the list is titled "Before Posting: Tips" and includes a note about using the search function and attaching files. The second post is titled "Welcome to the CellProfiler forum!" and provides an overview of the software and how to use the forum.

Topic	Category	Users	Replies	Views	Activity
Before Posting: Tips		 	1	60	37m
Welcome to the CellProfiler forum!		 	1	50	8d

# Acknowledgments



 **BROAD**  
INSTITUTE      IMAGING  
PLATFORM

Funding and support for the CellProfiler project has been provided by:



National Institute of  
General Medical Sciences



Massachusetts  
Institute of  
Technology



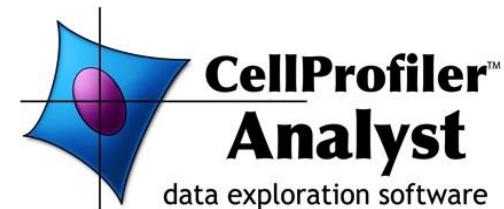
MERCK      NOVARTIS



Free, at [www.cellprofiler.org](http://www.cellprofiler.org):



**CellProfiler™**  
cell image analysis software



**CellProfiler™  
Analyst**  
data exploration software

Contact:  
[imagingadmin@broadinstitute.org](mailto:imagingadmin@broadinstitute.org)

Many thanks to our many  
biology collaborators who  
provide images

Recent funding for this work provided by:

NIH NIGMS (Carpenter: R01 GM089652 and  
Wahlby: R01 GM095672)

The Broad Institute of Harvard and MIT

*The Society of Biomolecular Imaging and Informatics  
(SBI2) is an international community of leaders,  
scientists, and students promoting technological  
advancement, discovery, and education to  
quantitatively interrogate biological models to  
provide high context information at the cellular level.*

