Searching for new antibiotics using digital images of infected worms

Mark-Anthony Bray, Megan E. Rokop, Carolina Wählby and Anne E. Carpenter

Background

The increasing resistance of bacteria to antibiotics

The use of antibiotic compounds to target and destroy bacteria is of critical importance to medical treatment of infectious diseases. Unfortunately, the increasing rise of antibiotic-resistant bacteria poses an alarming challenge, particularly with respect to hospital-acquired infections. As antibiotics decrease in efficacy, patients are ill for longer periods of time, leading to increased mortality for the individual and higher risk of epidemic infection for the surrounding community, along with increased costs in healthcare.

Traditional approaches to antibiotic discovery

One approach to this problem is to uncover new antibiotics to replace the ineffective compounds. Traditionally, such novel drugs are discovered using assays which measure either the inhibition of growth or killing of a pathogen (i.e. a disease-causing bacterial species). However, this method has yielded a low return on investment of time and expense with few viable drug candidates to date; accordingly, leading pharmaceutical companies have largely abandoned this approach. The reason for this failure lies in part in the inefficiency of *in vitro* and whole cell assays (that is, testing potential antibiotics on either bacteria growing in isolation, or bacteria infecting cells growing in culture). These experiments often lead to discovery of chemical compounds that turn out to be either toxic to humans, or are insufficiently novel (i.e. they have the same mechanism of action as current antibiotics).

C. elegans *as a model organism for antibiotic research*

Another, more promising, approach is to test potential antibiotics on suitably chosen "model organisms" that have been infected with pathogenic bacteria. A model organism is defined as a non-human organism that can be studied to draw conclusions about how biological systems in humans function. The roundworm *Caenorhabditis elegans* (*C. elegans*) has proven to be extraordinarily useful for this role. It has a development cycle, proceeds through morphogenesis and growth to an adult stage, ages and then dies. It has a nervous system, muscles for motion (along with muscle degeneration upon aging), a metabolic system and (most importantly for our purposes) an immune system. Furthermore, its short life-cycle, ease of growth in culture dishes and small size (1 mm long) makes it amenable to high-throughput testing (or "screening") of thousands of chemical compounds simultaneously.
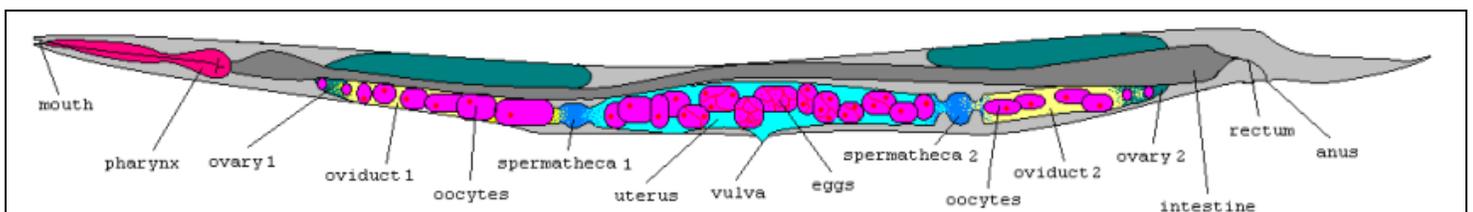

Figure 1: Anatomic diagram of *C. elegans*. Total length of an adult worm is ~1mm.

For this exercise, you will be studying the effects of infecting *C. elegans* worms with the *E.faecalis* bacteria. In humans, this bacterial species is a leading cause of hospital infections, and it is also often antibiotic-resistant. Your goal is to determine which chemical compounds (from a large collection of potential antibiotics) can save (or "rescue") *C. elegans* from infection by *E.faecalis*, by applying various chemical compounds to the worms and measuring how many worms survive infection in the presence of each compound.

An image-based approach for C. elegans *analysis*

To begin this experiment, we grow the worms in culture. We culture the worms in a small test tube (a "well"), in a multi-well plate that consists of 384 wells fused into a single plate. (Each plate is arranged in a 12-row by 24-column format: the letters A – P and the numbers 1-24 indicate the rows and columns, respectively.) Into each well of the plate, we place a fixed number of worms, and we then infect the worms with the *E. faecalis* bacteria. We then dispense a different chemical compound in each well, and allow infection (and possibly "rescue" by the chemical) to occur. Since these compounds are of unknown effect, certain wells are designated as *control wells* to provide a basis for comparison: some wells are treated with known antibiotics (positive controls) while others contain no chemical compounds (negative controls). The worms are then stained with SYTOX, a fluorescent stain that becomes brighter when worms die, and each well is imaged with an automated microscope.

Figure 2: Schematic of automation of sample preparation.
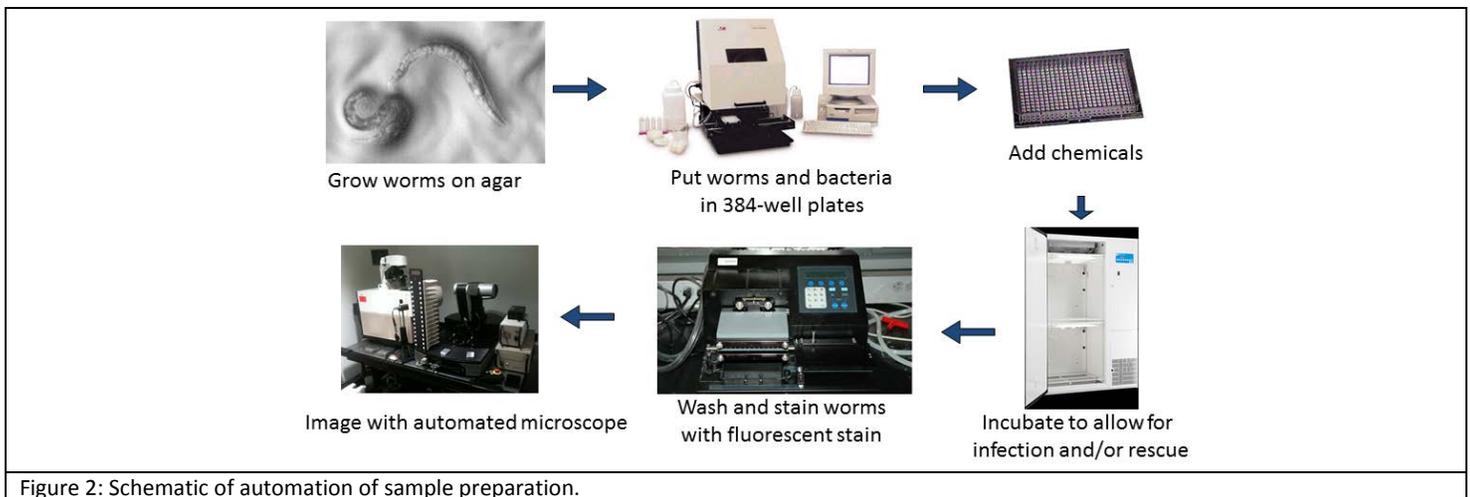
Activity 1: Automatic analysis of C. elegans *images using a fluorescent stain for indicating dead worms*

Your assignment is to perform a "live/dead" assay to determine which compounds "rescue" the worms from infection by the harmful bacteria. The traditional method of assessing worm death is to physically poke the worm with a probe and look for a response (an approach that is not feasible for a screening assay involving thousands of compounds). Counting the number of dead (SYTOX-stained) worms in wells by eye is also tedious. Instead, you will have an automated microscope and a computer perform the task.

In order to analyze the image data, you will use CellProfiler, an open-source image analysis package. This section contains instructions necessary to analyze images of infected roundworms treated with different chemical compounds. Each well in the 384-well plate will have two pictures of the worms in this well: a bright-

field image that shows all the worms, and a fluorescence image in which the dead worms are shine brightly (Fig. 3).
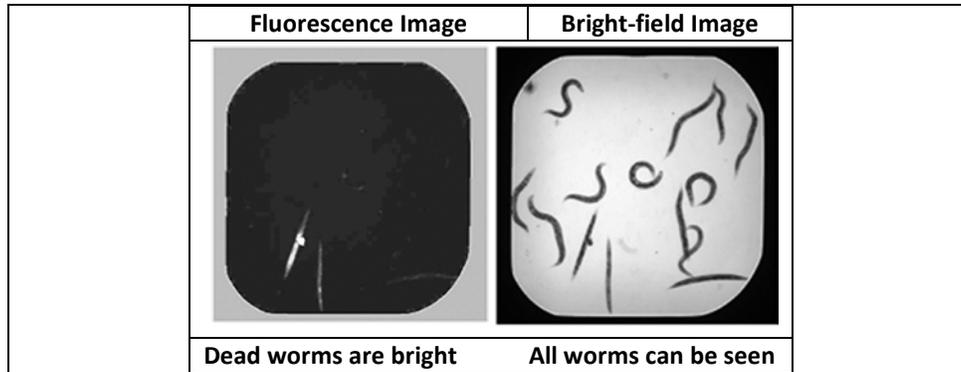

Figure 3: Example well with two SYTOX-stained worms and the corresponding bright-field image.

- To begin running the program, start your computer and select CellProfiler from the Start Menu (Windows) or the Applications folder (Mac).

- On the right of the CellProfiler interface, you will see a large panel with the words "Drop files and folders here." Using File Explorer (Windows) or Finder (Mac), drag and drop the folder that contains the images of the worms folder into this panel.

- At the bottom of the CellProfiler interface, click the "View output settings" button. In the panel that appears to the right, click the folder button to the right of the "Default Output Folder" box and browse to the Desktop.

- In the main menu (at the top of the screen on Mac, or the top of the CellProfiler interface on Windows), select *File > Import Pipeline > From file…*, and browse to the "Pipelines" folder.

- Double-click the file 'Worm_Analysis_MeasureFluorescence.cppipe' to select and load it. This file is called a pipeline which consists of individual modules; each module performs a unique image processing step and can be arranged such that they are executed in sequential order.

- Click the button labeled "Analyze images" at the bottom-right of the CellProfiler interface to start the pipeline. The software will try to find all the worms in a well and measure the amount of fluorescence from the worms. The chemical compounds that are most effective at "rescuing" the worms from infection will be represented by the wells with the least amount of fluorescence (i.e. the most worms that are still alive after infection).

The pipeline will run beginning with the first of 15 images, and the window will update 15 times, each time the next image is analyzed. The full run may take a few minutes.

Description of the pipeline modules and their function

As the pipeline runs, you can read a description of each module's function in the pipeline below and look at the corresponding module display window:

- o *Images:* Loads each pair of images (bright-field and fluorescence) into CellProfiler.
- o *CorrectIlluminationCalculate*: Finds the background image (the bright-field image minus the worms).
- o *CorrectIlluminationApply:* Corrects for uneven lighting by removing the background image from the bright-field image.
- o *ImageMath:* The worm identification requires bright objects on a dark background, so we invert the pixel values in the corrected bright-field image to make the foreground bright and the background dark. We will show an example of this image later as you will be examining it in the exercise.
- o *ApplyThreshold*: Identify the well from the inverted image.
- o *Morph:* Clean up the well by shrinking it slightly to get rid of bright edge pixels.
- o *Crop*: Crop the color-inverted image with the shrunken well image to get our final image.
- o *IdentifyPrimaryObjects:* Identify the worms in the cropped and color-inverted image.
- o *MeasureImageIntensity:* Measure a collection of intensity statistics obtained from the fluorescence image using only the worm pixels. One collection of statistics is evaluated **per image**.
- o *MeasureObjectSizeShape:* Measure a collection of area statistics evaluated **per worm**, for each image. These measurements will be used later in Activity #2.
- o *RescaleIntensity, OverlayOutlines, DisplayDataOnImage*, *ConvertObjectsToImage, Resize* and *Tile*: These modules scale the intensities, superimpose the outlines and colored labels of the worms onto the images resize and tile the images together to prepare them for output.
- o *SaveImages*: Save the final image to a file for later review.
- o *ExportToSpreadsheet*: Output the final set of measurements to a text file.

Additional details are included in the "Module Notes" at the top of each module.

- Once the pipeline is complete, you will have two new files in the Default Output Folder:

  - o A single spreadsheet file labeled "Activity1_Fluorescence_Data.csv" (containing measurements for each image); and

  - o An image file corresponding to each of your 15 input images (highlighting the detected worm objects). Each well is represented with a pair of panels: (1) The bright-field image with each worm outlined with a color is shown on the left. (2) Each object that was detected as a "worm" by the computer is shown in a distinct (arbitrary) color on the right. The well number is shown in yellow in the middle of each pair of images, and each detected object is numbered.

- Open the spreadsheet file "Activity1_Fluorescence_Data.csv" (located in the Default Output Folder, your desktop) in either Excel or a regular text editor. Each row corresponds to an individual image; each column is a **per image** measurement. The column headers show the name of the measurements collected.

Description of the fluorescence measurements collected

Six fluorescence intensity measurements are produced by the *MeasureImageIntensity* module, each of which is given a special nomenclature in the spreadsheet:

o   *Maximum (Intensity_MaxIntensity_FluorescenceImage_WormObjects)*: The intensity of the brightest pixel contained within any worm.

o   *Mean (Intensity_MeanIntensity_FluorescenceImage_WormObjects)*: The mean (or average) intensity of all the pixels contained within a worm.

o   *Median (Intensity_MedianIntensity_FluorescenceImage_WormObjects)*: The median intensity of all the pixels contained within a worm.

o   *Minimum (Intensity_MinIntensity_FluorescenceImage_WormObjects)*: The intensity of the dimmest pixel contained within any worm.

o   *Standard deviation (Intensity_StdIntensity_FluorescenceImage_WormObjects)*: The standard deviation of the intensities from all the pixels contained within a worm.

o   *Total (Intensity_TotalIntensity_FluorescenceImage_WormObjects)*: The sum of the intensities of all the pixels contained within a worm.

- Looking at the example image (Fig. 3), which of the six fluorescence measurements listed above would be most useful for determining the efficacy of each of the 15 chemical compounds tested in our experiment?
- For the intensity measurement you selected above, record the value for that measurement (for each image) in the chart below (note that Metadata_WellNumber is the last column in the spreadsheet). Fill in the parentheses with the name of the column in the spreadsheet that you chose as most useful for this analysis.

Well number (Metadata_WellNumber)	Selected fluorescence intensity measure ( )
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

- According to the measurement you recorded above, which wells contained chemical compounds that best rescued the worms from infection by the harmful bacteria?
- Which well(s) do you think were most likely negative controls?

Using an image-based screening approach like this, a study tested 37,214 chemical compounds for their ability to rescue the worms from an otherwise lethal *E. faecalis* infection (Moy et al, ACS Chem Biol, 2009). The screen found 80 known antibiotics or compounds known to have antimicrobial effects. However, the screen also uncovered 28 compounds which did not have previously documented antimicrobial activity but nonetheless cured the worms. Furthermore, six of these appeared to rescue the *C. elegans* from infection not by inhibiting the growth of the bacteria, but by boosting the host worm immunity. This study helps validate the whole-animal "model organism" approach because it found novel antibiotics, as well as a novel mechanism of boosting the immune system that would not have been revealed without using whole organisms in the screen.

Activity 2: Determining new criteria for distinguishing live vs. dead worms

The SYTOX fluorescent stain works because cell membranes of dead worms become permeable, thus allowing the SYTOX stain (which cannot pass through the membranes of live worms) to enter the worms; this allows SYTOX to function as a dead worm indicator. The SYTOX stain is not perfect, however; on occasion, other objects may also absorb the SYTOX stain: (a) live worms, (b) the bacterial cells, (c) large debris present in the well as an artifact of the worm culturing. In addition, the amount of fluorescent staining one sees with SYTOX is not linearly proportional to the number of dead worms.

- How would each of these circumstances (a), (b), & (c) above each affect the expected results? Would each of these three conditions lead you to conclude that the chemical compound in that well was better than it really is at rescuing worms or worse than it really is at rescuing worms? Explain your answers for all three circumstances.
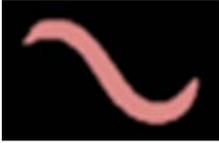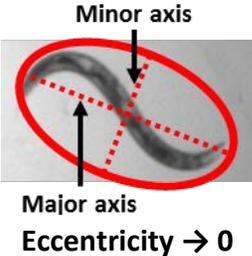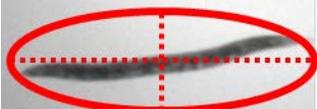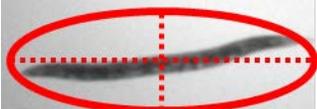
Because of the problems with the SYTOX fluorescence stain, let's now explore using a different visual feature to distinguish live versus dead worms. This will work because there is another intriguing feature of the worms upon death that can be exploited for our purpose.

- Open the image file 'Worm_color_image_automatic.png' that the pipeline created. Each well is represented with a pair of panels: (1) The bright-field image with each worm outlined with a color is shown on the left. (2) Each object that was detected as a "worm" by the computer is shown in a distinct (arbitrary) color on the right. The well number is shown in yellow in the middle of each pair of images, and each detected object is numbered.

- Visually compare the top three highest scoring wells with the top three lowest scoring wells from the chart you filled in on Page #5. Describe an obvious visual feature (not relying on the SYTOX fluorescent stain) that distinguishes live and dead worms.
- We need to automate this process, and allow the computer to determine whether a worm has this visual feature or not. Name some quantitative measures you think might be amenable to automatic measurements, so that the computer could distinguish live versus dead worms based on this new visual feature.

- From the Default Output Folder, open the spreadsheet file 'Activity2_Morphology_Data.csv' in either Excel or a regular text editor. Each row corresponds to an individual image, and each column represents a **per image** measurement based on measurements of individual worms in that image.

Description of the morphological measurements collected

Six morphological measurements are calculated by CellProfiler. For each measurement, the value is computed for each worm, and then the median across all worms is calculated for each image; this median value is stored in the spreadsheet. Each of these median measurements is given a special nomenclature in the spreadsheet and is described below:

Area *Median_WormObjects_AreaShape_Area*	The number of pixels in one object (i.e. "worm")	
Eccentricity *Median_WormObjects_AreaShape_Eccentricity*	How round or oblong an ellipse would be if it was fitted to that object. Equals 0 for a circle (the most round), and equals 1 for a line segment (the most oblong).	**Minor axis** **Major axis** **Eccentricity → 0**
MajorAxisLength *Median_WormObjects_AreaShape_MajorAxisLength*	The length in pixels of the long axis of an ellipse that best fits the object.	
MinorAxisLength *Median_WormObjects_AreaShape_MinorAxisLength*	The length in pixels of the short axis of an ellipse that best fits the object.	 **Eccentricity → 1**
Perimeter *Median_WormObjects_AreaShape_Perimeter*	The total number of pixels around the boundary of one object (i.e. "worm").	
Solidity *Median_WormObjects_AreaShape_Solidity*	Describes the extent to which the border is convex or concave. Computed as the area of the object divided by the area of the smallest polygon that encloses the object. Equals 1 for a convex shape and < 1 for a concave shape	**Solidity → 1** **Solidity < 1**
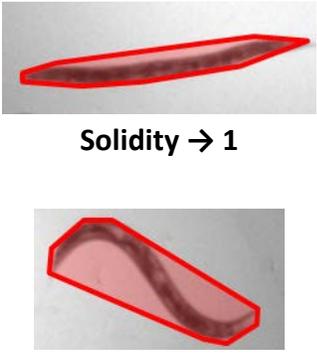
- From the chart you recorded on Page #5 above, select the wells that gave the three highest values and the wells that gave the three lowest values. Record these wells below, and also record the six measurements shown for each well (using 3 significant digits is sufficient).

Well number (Metadata_WellNumber)	Area	Eccentricity	Major Axis Length	Minor Axis Length	Perimeter	Solidity
1st highest:						
2nd highest:						
3rd highest:						
1st lowest:						
2nd lowest:						
3rd lowest:						

- Based on the measurements above, which morphological measurement is the most helpful for distinguishing live from dead worms? What are your criteria for making this decision?
- For the measurement you selected above, record the value for that measurement (for each image) in the chart below. Fill in the parentheses with the name of the column in the spreadsheet that you chose as most useful for this analysis.

Well number (Metadata_WellNumber)	Selected measure ( )
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

- Pick a cutoff value that will separate the five highest scoring images from the ten lowest scoring images.

As you have noted above, live worms tend to be sinusoidal or "wavy" in appearance, whereas upon death, they appear rod-like. As the worms become bloated with bacteria and die, the worms undergo a "rigor mortis"-like process where the body becomes stiff and straight. Therefore, using worm shape to distinguish live versus dead worms is an alternative to a fluorescence approach using the SYTOX stain.

Activity 3: Investigating the limitations of using the shape-based approach

To use the shape-based approach, we must first overcome one of the major challenges of worm image analysis: the inaccurate identification of touching and crossing worms leading to inaccurate shape measurements.

- Because the SYTOX stain sometimes stains live worms, bacteria and debris, open the image file 'Worm_color_image_automatic.png' that you opened before. Recall that each well is represented with a pair of panels. The bright-field image with each worm outlined with a color is shown on the left. Each object that was detected as a "worm" by the computer is shown in a distinct color on the right. The well number is shown in yellow in the middle of each pair of images, and each detected object is labeled with a unique number.

- When you look at the bright-field images, in which wells do you see "tangles" of two or more worms touching or lying on top of each other?
- Record the number of tangles in each well into the 2nd column of the chart below
- For each image that has tangles, a *tangle* is defined as multiple (2+) worms that the computer recognizes as only 1 worm (Fig. 4).
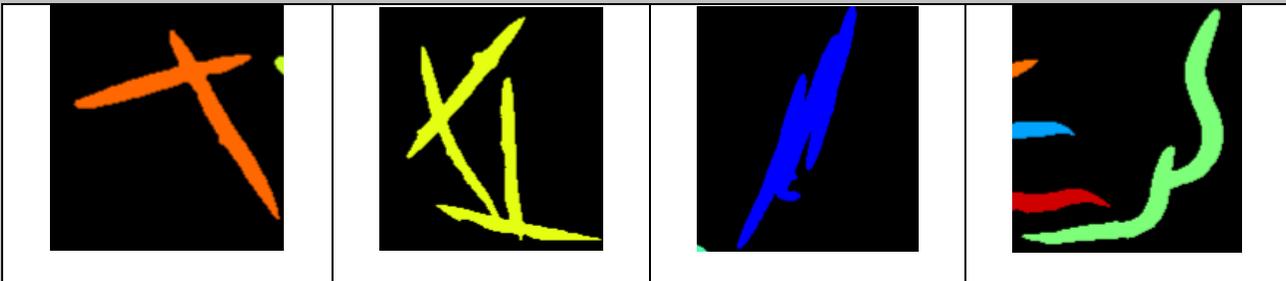

Figure 4: Examples of "tangled" worms recognized by the computer as one object , which should be removed from consideration.

Record the number of actual worms that comprise each tangle for each well.

Well number	Number of tangles	Number of worms per tangle
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

- Would this issue drastically affect your ability to accurately determine which chemical compounds work best to rescue worms from infection, using the shape-based approach? Why or why not?

It turns out that using the shape-based approach depends strongly on how well separated the worms are, and how well the computer recognizes touching worms as separate objects. Scientists have thus established ways to train the computer to recognize tangles of worms and recognize them as being groups of multiple worms (as opposed to one worm).

Activity 4: Automatic analysis of C. elegans *images using shape-based features*

Often in these experiments, worms that are touching each other are very difficult to separate using automated analysis methods. The next activity will involve a new method where you will create a worm "model" that you will use to train the computer to recognize touching worms as separate objects. Once you have trained the computer to do so (in a process called "machine learning"), you will then apply that model to the data we used earlier, to separate the worms. You will then use the shape-based approach to analyze the same data we used earlier to determine which chemical compounds are most effective at rescuing the worms from infection.

Part 1: Create a worm "model" so that the computer can recognize each worm separately

For this exercise, you will create a pipeline that will allow us to train the computer to identify worms as separate objects.

This pipeline contains some of the same modules as the first pipeline, with the exception of an *EditObjectsManually* and *UntangleWorms* module.

o *EditObjectsManually*: An interface where the user can click on unwanted objects to remove them from the analysis.
o *UntangleWorms:* In the "Training mode," this module uses examples of separated, individual worms to create a model of a prototypical worm, so that later the computer can "learn" to identify touching worms as separate objects.

A special note on the *EditObjectsManually* module*:* For technical reasons, the worm identification requires bright objects on a dark background, so we invert the pixel values in the bright-field image to make the foreground bright and the background dark. Because of this necessary color-inversion step, in this section, the bright-field images you will look at will be inverted (from dark-on-light to light-on-dark) as shown here:
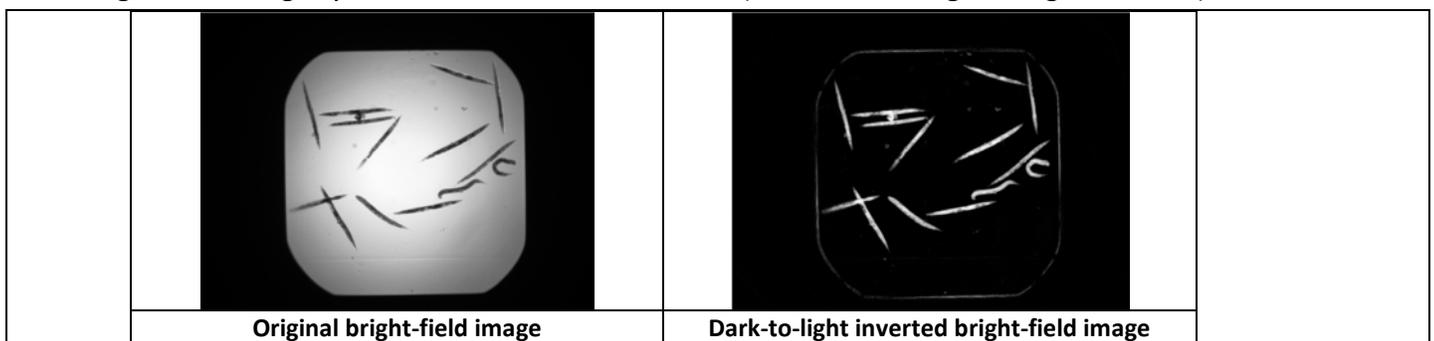

| Original bright-field image | Dark-to-light inverted bright-field image |

Figure 5: Panels showing the color inversion of the images used for manual worm object selection.

10

To start the analysis, follow the steps below:

- In the main menu, select *File > Import Pipeline > From file…*, and browse to the worm folder. Double-click the file 'Worm_Analysis_CreateWormModel.cppipe' to select and load it.

- Select *Window > Close All Open Windows* to close all current windows.

- Click the "Analyze images" button.

- Once the pipeline gets to the *EditObjectsManually* module, it will display a window titled "Choose objects to keep". At this point, it will pause, because it is waiting for you to check if any tangles have been improperly recognized as single worms.

- Remove any tangles (by clicking on them) in each image that have been identified as one worm even though they multiple worms. Also make sure to remove any non-worm objects or "broken" worms as you do not want to include them in your model. Each time you are done with an image, click the "Done" button, so that the next image pops up (see details below). As the pipeline continues, you will repeat the above steps for each image in the analysis. In between each of the 15 images, you will see windows pop up that show the resulting training data that will be used later by the computer to learn from you what an individual worm looks like.

- Once the pipeline is complete, you will have a "MyWormModel.xml" file containing the worm model in the Default Output Folder. Make sure not to move this file elsewhere.

Instructions for why you are creating the "worm model":

o You will be working to exclude objects that are aggregates of worms touching each other so the computer learns that these are unacceptable. Any collection of pixels that share the same color is considered by the computer to be one object. Since the untangling procedure assumes that the training set consists of the individual worms, aggregates of multiple worms need to be removed from consideration.

Instructions for how you are creating the "worm model":
o Click on a colored outlined object in the display to remove objects that aren't single worms (that is, tangles; see Fig. 4).
o Once you click on an object in the display, you will see that the outline changes from solid to dashed. If you make a mistake, click on the object again; the outline will change back to solid.
o If a click doesn't remove an object, it may be that the object is so thin that the mouse click is missing it. Zoom in and try clicking on it again.
o There are some tools on the bottom toolbar that may be helpful to see the individual worms:

The 1st icon lets you reset the view back to the original view.
The 2nd and 3rd icons let you step backwards and forwards through any changes you made to the view.

The 4th icon lets you move in any direction in the display by clicking and dragging.

The 5th icon lets you zoom into the display by dragging and drawing a box to zoom in on.

o If the worms are hard to see, you can resize the window to see the panels better. You can also uncheck the "Display image" box to see the labeled objects without the inverted image.

Part 2: Apply the worm model to the images, in order to train the computer with this new model

This next pipeline is almost identical to the first pipeline you ran in "Activity #1", but now you are analyzing the data using the worm-shape-based approach, as opposed to the fluorescence SYTOX-based approach. Because you are using the shape-based approach this time, you will be applying the worm model you just created, so that the computer knows how to identify touching worms as separate objects. Even though you are running the analysis on the 15 images you used for training, normally the analysis would be performed on thousands of images; thus the automated approach is vastly faster and less tedious than the standard approach of scoring by eye.

We will also be incorporating the *ClassifyObjects* module to classify each object in an image on the basis of a measurement and a cutoff that you choose. In our case, we will be classifying the objects (i.e. worms) by the cutoff you chose for worm shape at the very end of "Activity #2."

To start the shape-based image analysis, follow the steps below:

• In the main menu, select *File > Import Pipeline > From file…*, and browse to the worm folder. Double-click the file 'Worm_Analysis_ UntangleWormsAndMakeMeasurements.cppipe' to select and load it.

• Select *Window > Close All Open Windows* to close all current windows.

• In the main CellProfiler window, you will see the *ClassifyObjects* module in the left-hand list, which has a red warning next to it. Click on *ClassifyObjects* so that its settings are shown on the right. The reason why a warning is shown is that it is waiting for you to enter the measure that you decided earlier was best for distinguishing live from dead worms. To configure the settings for this module, do the following:
  ▪ Select the measurement you decided to use at the end of "Activity #2"; select it from the "Measurement" drop-down box. This box is blank until you select a measurement, and then the red text and warning go away.
  ▪ Enter the cutoff that you decided to use at the end of "Activity #2"; type it into the setting labeled "Enter the custom thresholds separating the values between bins".

• In the main CellProfiler window in the left-hand list, you will also see the *CalculateMath* has a red warning next to it. Click on *CalculateMath* so that its settings are shown on the right. The reason why a warning is shown is that it is waiting for a measurement given by the ClassifyObjects module. To configure the settings for this module, select either *Classify_**Bin1**_NumObjectsPerBin* or *Classify_**Bin2**_NumObjectsPerBin* from the "Measurement" drop-down box.
  ▪ Choose "Bin1" if worms that are *below* the measurement/cutoff you selected are alive (i.e., rescued from infection).
  ▪ Choose "Bin2" if worms that are *above* the measurement/cutoff you selected are alive.

This box is blank until you select a measurement, and then the red text and warning go away.

- Click the "Analyze images" button. This may take a few minutes. If you watch the display for each of the 15 images as they update, you can see the computer now correctly recognizing tangles as being comprised of multiple overlapping worms.

- Once the pipeline is complete, you will have a spreadsheet for Activity #4 in the Default Output Folder, along with an image file 'Worm_outline_untangled.png' outlining the worms that have been "untangled." (Several additional files are also generated, but they can be ignored.)

- In the Default Output Folder, open the spreadsheet file "Activity4_Morphology_Data.csv" in either Excel or a regular text editor. Each row corresponds to an individual image, and each column represents a **per image** measurement. The columns contain the same median shape measurements as before, but there are also two additional measurements resulting from the new *ClassifyObjects* module.

These new columns are labeled *Classify_**Bin1**_NumObjectsPerBin* and *Classify_**Bin2**_NumObjectsPerBin*. Each column contains the number of one type of worm (live versus dead) in that well, which the computer has tallied based on the measurement and cutoff you entered. "**Bin1**" represents the worms that are below the measurement/cutoff you selected, and "**Bin2**" represents those above the cutoff.

- Record the number of "live" versus "dead" worms (i.e. from the two **Bin** numbers described above) for each well in the experiment and percentage that are live, in the chart below. Fill in the parentheses with the name of the column you used for this analysis, because which bin corresponds to which type of worm depends on the measurement and cutoff that you have chosen.

Well number (Metadata_WellNumber)	Number of live worms ( )	Number of dead worms ( )	Percent live (Math_PercentLive)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			

- According to the numbers you recorded above, which wells contained chemical compounds that best rescued the worms from infection by the harmful bacteria?
- Which well(s) do you think were most likely negative controls?

Activity 5: Designing a chemical genetic screening experiment for a different biological process or disease

The general approach used here, testing the effects of a large number of chemical compounds to discover new medical treatments or help us understand biological processes, is known as "*chemical screening.*" By looking at the effects of thousands of chemical compounds on cells or organisms affected by a disease, these screens may form the beginning stages of the (very long) drug discovery process, as in our example performed above.

Alternatively, large "*genetic screens*" can be done where, instead of each well of the 384-well plate being treated with a different chemical compound, each well is treated with a different genetic perturbation. As a result, the cells or organisms in each well are either **lacking the function** of just 1 of the ~20,000 genes in the human genome (in what is called an "RNAi screen"), or they are **over-expressing a high level of the function** of just 1 of the 20,000 genes in the human genome (in what is called an "overexpression screen"). These genetic studies help scientists determine which genes have functions that are associated with a variety of human diseases. These genetic screens are generally referred to as RNAi screens and over-expression screens, respectively.

Whenever scientists design an experiment involving image-based analysis, they must decide whether they want to do a chemical screen or either type of genetic screen, based on their experimental goals. This is just one of many decisions they must make, in terms of experimental design.

The most major decisions to make before initiating a screen include:

(1) What biological process or diseases are you investigating in your screen, and why?

(2) What kind of organism/cell will you use, and why?

(3) What are your positive and negative controls?

(4) What growth conditions or treatments are required so that the cells or organisms are in the proper state to study the biological process or disease of interest?

(5) Are you going to do a chemical screen or a genetic screen?   *If you are going to do a chemical screen*: which chemicals are you going to test (e.g. newly synthesized chemicals that have never been studied before, all FDA-approved compounds, natural products from nature, a combination of all three of these collections, etc) and why? *If you are going to do a genetic screen*: will you be doing an RNAi screen or an over-expression screen, and why?

(6) What visible features will you looking for, and how would an automated approach help you measure these features? Will you be making **per image** or **per cell/organism** measurements? How many different

To follow up on the final criteria/question in this list, consider this exercise. In our example, where we sought new antibiotics against *E. faecalis*, a number of measurements were calculated, both in a **per image** approach and in a **per worm** approach. Only two types of images were analyzed: originally we used one color of fluorescence to look at SYTOX live/dead staining after defining the worm using bright-field images, and later we used bright-field images only, for worm shape analysis.

Choosing the number of images (bright-field, and fluorescence, and how many colors of fluorescence) that will be taken per well is important for determining whether your focus is on speed of the screen, or on complexity of the measurements – essentially, breadth versus depth. Choosing breadth or depth gives a scientist different benefits and drawbacks to the study.

Multiple colors of stains can be used in the same cells or organisms, such that scientists can stain multiple parts (organs) in an organism, or multiple parts (organelles) in a cell. In these experiments, an even larger number of measurements (and thus a very large amount of data) can be collected using an automated microscope. Scientists often find the rich information from these measurements helpful in accurately identifying samples of interest.

Automated image analysis software such as CellProfiler has been used not just to discover new antibiotics using model organisms such as *C. elegans*. This software has also been used to analyze data from hundreds of other chemical and genetic screens. Below are some examples.

Discovering potential new antibiotics to treat tuberculosis	
Similar to the *C. elegans* and *E. faecalis* example, tuberculosis (TB) is another disease of concern due to increasingly antibiotic-resistant bacteria. In this assay, the human cells (blue) are infected with the tuberculosis bacteria (green) and are screened for compounds which kill the bacteria, while leaving the human cells intact. In this way, new antibiotics to treat TB can hopefully be discovered.	TB infection, no treatment TB infection + chemical

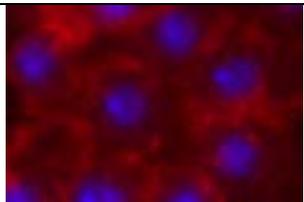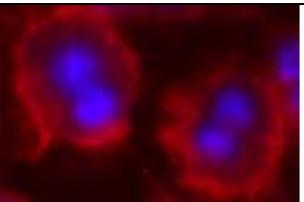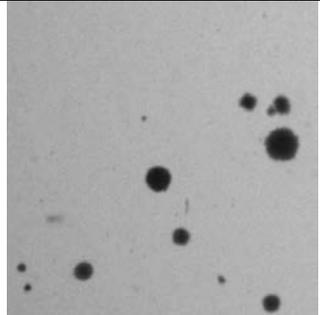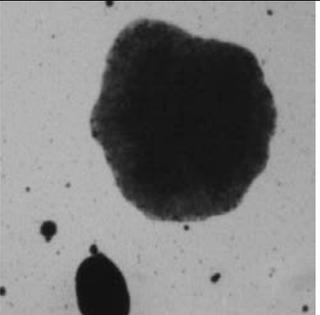| **Discovering potential new treatments for leukemia**	
The goal of this experiment is to identify novel compounds that may lead to potential treatments for a form of leukemia which involves human megakaryocytes, one type of immune system cell. Rather than finding a drug that kills the cancerous cells, the researchers want to find a chemical that causes polyploidy (nuclear division without cellular division), restraining the cancerous cells' uncontrolled growth, and making them more susceptible to conventional cancer therapies.	Leukemia, no treatment Leukemia + chemical

Discovering genes critical for cell division	
Knowledge of the genes and molecules in cells that regulate the process of cell division is important for understanding diseases such as cancer. In this experiment, genes known to be important in the final stage of cell division were targeted using RNAi, which turns off 1 gene per well of the 384-well plate. The goal was to determine which genes – when non-functional – lead to cells becoming binucleated (having two nuclei instead of properly dividing to yield one nucleus per cell). This experiment uses nuclear (blue) and cytoskeletal (red) fluorescent stains to show different cellular compartments.	**Dividing cells, all genes functional** **Dividing cells, – 1 gene**

Discovering genes associated with malignant forms of cancer	
Malignant cancer cells take on the property of being able to grow in an anchorage-independent way (i.e. they can grow in culture in a liquidy soft agar growth medium, whereas non-malignant cells require a solid surface to grow and divide). In this experiment, non-malignant kidney epithelial cells with cancerous properties are genetically perturbed to over-express different human genes, in order to identify genes involved in the process of causing cancers to become malignant. Cells over-expressing genes involved in malignancy are predicted to show an increase in both size and number of colonies when compared to the untreated cells. In this example, we are only using bright-field images for our automated image analysis; no fluorescent images were taken.	**Cells with all genes at normal levels** **Cells + 1 gene over-expressed**

Figure 6: Examples of screening experiments requiring automated image analysis.

Now that you have the tools to design a chemical or genetic screen that uses automated image analysis for gaining insight into a biological process or disease, you can design a screening experiment of your own.

Design a chemical or genetic screen to study a biological process or disease. As you design the screen be sure to address all five questions listed on Page #14, that scientists must consider as they design their experiments.

- Prepare a written report that effectively communicates the design and justification for your screen to your teacher and classmates. Be sure your report describes your answers to these five questions and any other information necessary to understand the goal of the screen and how it would work.

Activity 6: Using the CellProfiler image analysis software to analyze data from another experiment

The "Examples" section on the CellProfiler website (http://cellprofiler.org/examples.shtml) provides images and pipelines for ~20+ different image-analysis-based chemical and genetic screens. With the advent of automated microscopy, and the successful demonstration of sophisticated image analysis methods, the number of quantitative microscopy experiments is rapidly increasing. Because nearly every biology laboratory uses microscopy in some fashion, the modern biologist will be well served by becoming familiar with the software and strategies involved in quantitative image analysis.

- Form groups of 2-4 student members each. Within your group, choose one of the example pipelines from the webpage to work through, and analyze the data from that screen experiment.

- Prepare an oral presentation, with figures, to present to the class, in order to report back about your group's experience using the software for this new example experiment. This presentation should make clear the design and justification of the experiment (using answers to all five questions listed on Page #14). This presentation should also contain figures that display the analyzed data in ways that are effective in summarizing the results and conclusions from the experiment.

Software Used In This Activity

CellProfiler
  Developed by Anne E. Carpenter, Thouis R. Jones and Lee Kamentsky at the Broad Institute of Harvard and MIT. Available for free download at: http://cellprofiler.org

Related Web Sites

CellProfiler: Cell Image Analysis Software: http://cellprofiler.org

WormClassroom: http://www.wormclassroom.org

Additional Resources

Moy TI, Conery AL, Larkins-Ford J, Wu G, Mazitschek R, Casadei G, Lewis K, Carpenter AE, Ausubel FM. High throughput screen for novel antimicrobials using a whole animal infection model. ACS Chem Biol. 2009 July 17; 4(7): 527–533. (Pub Med ID: 19572548)

Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biology 2006; 7:R100. (Pub Med ID: 17076895)

Ljosa V, Carpenter AE. Introduction to the quantitative analysis of two-dimensional fluorescence microscopy images for cell-based screening. PLoS Comput Biol. 2009; 5(12): e1000603. (Pub Med ID: 20041172)

Figure References

Figure 1: http://commons.wikimedia.org/wiki/File:C_elegans_anatomy.png

Figure 2: Courtesy of the Ausubel and Ruvkun laboratories

Figure 3: Adapted from Figure 1 of Moy et al, ACS Chem Biol 2009

Figure 4: Screenshot from CellProfiler

Figure 5: Screenshot from CellProfiler

Figure 6: Images courtesy of Sarah Stanley, Jeremy Wen, John Castoreno and Xiaoxing Wang.

Acknowledgments

The authors would like to acknowledge Annie Lee Conery for collecting the original images and the Ausubel lab for making these images publically available.