Quick Start Guide: Find the Best Feature

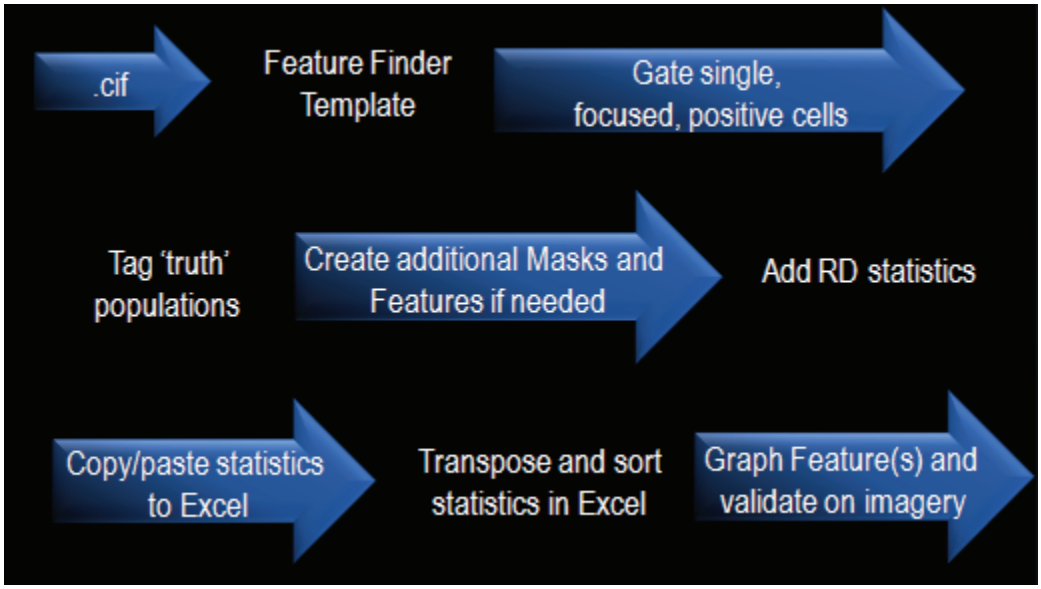
A guide to choosing the best feature

IDEAS provides multiple image-based parameters that enable objective discrimination of cells based on their appearance. In addition, IDEAS calculates a statistic (Rd) that measures the degree of separation a given parameter provides between two populations. This provides a method for choosing parameters that best discriminate cells.

This guide describes the process used to find features that effectively measure image-based phenotypes with minimal knowledge of the feature set. A general description of the steps is followed by a specific example.

General:

1. Open a data file using the Feature Finder template and begin analysis (single, focused, positive).
2. Create two tagged 'truth' populations of cells that represent the morphological phenotypes you wish to separate.
3. Create any additional masks and features (not found in the template) you think may help distinguish the differences you see between the truth populations.
4. Calculate the statistical separation (RD) between the two populations afforded by the features.
5. Rank the features, one category at a time, according to RD and plot the features with the highest RD for the truth populations.
6. Validate by applying the feature to the base population and verify the results visually. Use independent controls if available and on multiple files and experiments.
7. Refine truth populations if necessary and repeat the process for each phenotype you see in the experiment.



Example:

Find the best features that distinguish whole blood neutrophils from lymphocytes based on differences in nuclear imagery. In this example, cells were stained with a nuclear dye (Channel 7) and collected with BF in Channel 5.

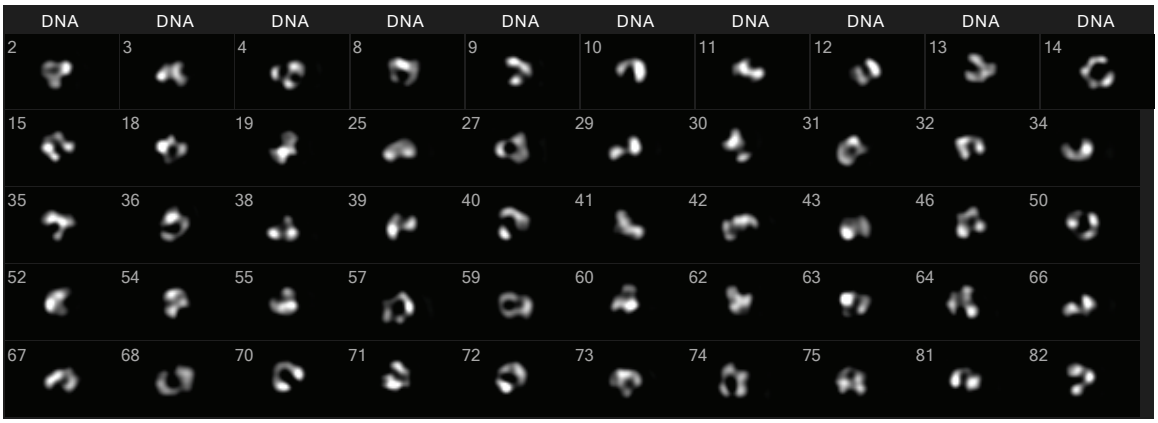
1. Open the data file and apply the Feature Finder template (.ast).* Gate focused cells on the Gradient RMS graph, single cells on the Brightfield Area versus Aspect Ratio graph,.

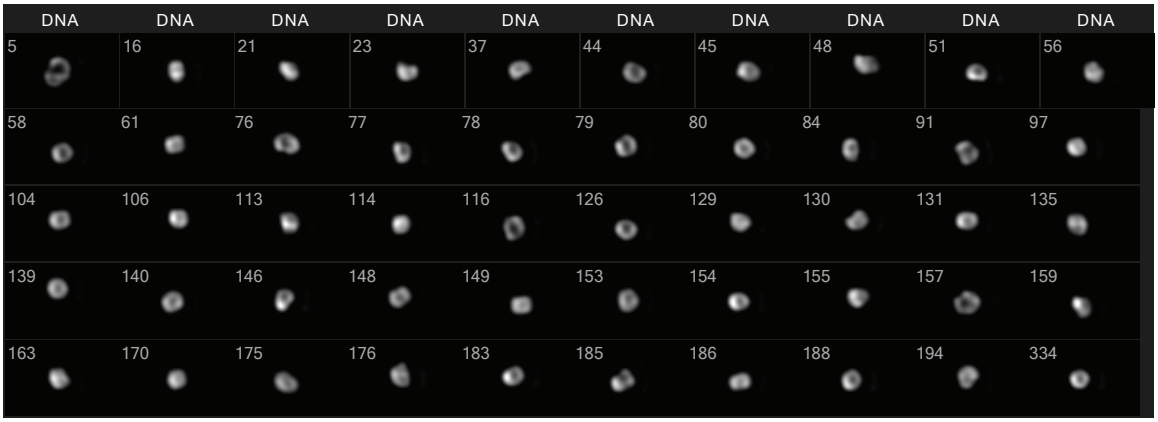
*The Feature Finder template is installed with IDEAS on the C drive.
The path is C>Program Files>Amnis Corporation>IDEAS Application v5.0>FeatureFinder_5.0.ast

2. Create the neutrophil (lobed) and lymphocyte (round) truth populations from within the single, focused cell gate using the tagging tool.

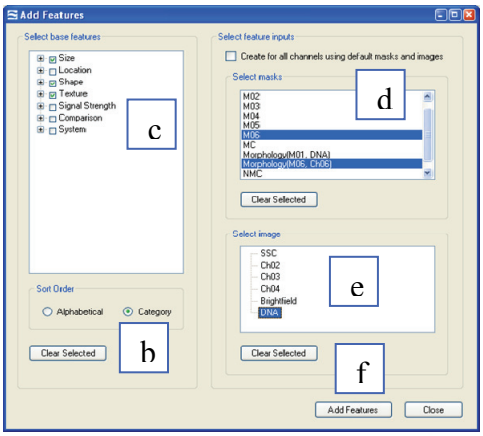
Note: When selecting truth populations, choose images that represent the range of each truth. For example, because neutrophils can have two or more lobes, we chose cells that have two, three, four, etc lobes. Lymphocytes have round or oblong nuclei, so we included examples of each. Also, the more cells the better: we chose 50 of each cell type.

Truth populations:



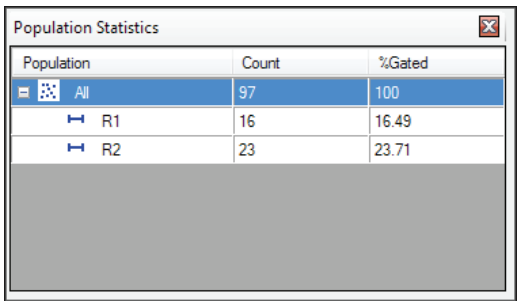


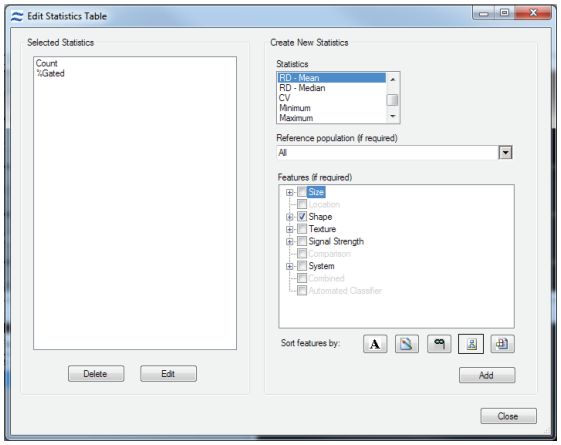
3. Note the differences in nuclear size, shape, and texture between the two populations.
4. Go directly to step 6 if you use the Feature Finder template. Otherwise, create the Morphology mask for the nuclear image using the mask manager.
5. Create features from the Size, Shape, and Texture categories using the DNA Morphology and the default DNA channel mask.
    a. Open the Feature manager from the Analysis menu and click Add Multiple Features.
    b. In the Add Features window select Category as the 'Sort Order'.
    c. Check Size, Shape and Texture boxes for the base features.
    d. Select masks M07 and Morphology(M07,Ch07)
    e. Select the Ch07 (DNA) image.
    f. Click Add Features.
    g. The list of features will be displayed.



h. Click OK to add the features.
i. Features that already exist will not be added. Click OK.
j. Close both windows and the features will be calculated.

6. Add a Population Statistics table to the analysis area.
   a. Click on the Population Statistics tool. ∑
   b. Right-click in the Population Statistics table and choose 'Edit Statistics Table'.
   c. Select the statistic RD –Mean. This is a measure of how well two populations are separated from each other based on their Means and standard deviation.
   d. Select one of the tagged populations as the reference population.
   e. Sort features by Category and select a category, in this case 'Shape'.
   f. Sort features by Images used and de-select all but the image of interest, in this case Ch7.
   g. Click 'Add Statistics'.
      Note, in this example both populations are in the same file so we can calculate RD within IDEAS. If the two populations to compare are in different files, the RD must be calculated after exporting the Mean and SD statistics. The RD for a feature is the difference in the Means divided by the sum of the SDs for the two populations. Or simply merge the two files together and analyze as above.



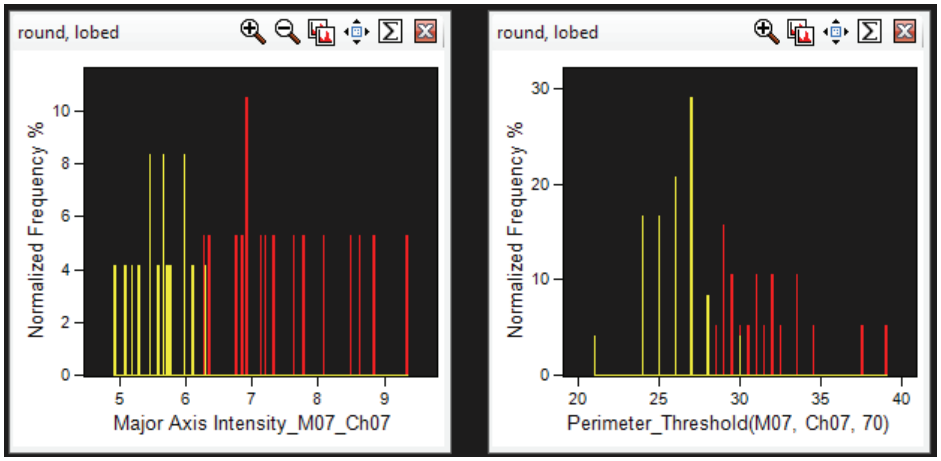


7. Copy and Paste the statistics into an excel spreadsheet.
   a. Launch Excel.
   b. Select the row of statistics in the table for the population you are comparing. Right-click and choose 'Copy Statistics – Transposed.
   c. Paste into the Excel spreadsheet.

8. In excel, sort the features in descending order so that the feature with the highest RD is at the top of the column.

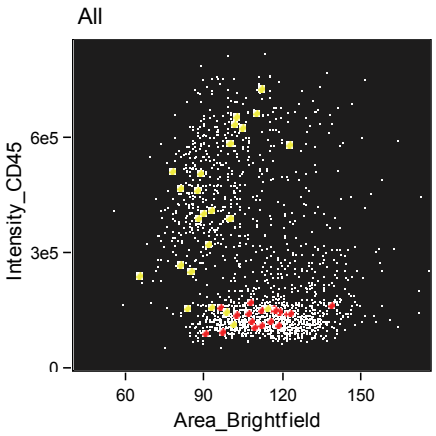
Population	lobed
Major Axis Intensity_M07_Ch07, round, Mean RD	1.508
Perimeter_Threshold(M07, Ch07, 70), round, Mean RD	1.24
Symmetry 3_Object(M07,Ch07,Tight)_Ch07, round, Mean RD	1.229
Lobe Count_Object(M07,Ch07,Tight)_Ch07, round, Mean RD	1.159
Perimeter_Threshold(M07, Ch07, 50), round, Mean RD	1.1
Major Axis_Object(M07,Ch07,Tight), round, Mean RD	1.081
Circularity_Object(M07,Ch07,Tight), round, Mean RD	1.03
Height_M07, round, Mean RD	0.956
Length_M07, round, Mean RD	0.956
Height_Object(M07,Ch07,Tight), round, Mean RD	0.9419
Length_Object(M07,Ch07,Tight), round, Mean RD	0.9419
Area_Threshold(M07, Ch07, 70), round, Mean RD	0.9149
Area_M07, round, Mean RD	0.9144
Perimeter_Object(M07,Ch07,Tight), round, Mean RD	0.9073
Symmetry 4_Object(M07,Ch07,Tight)_Ch07, round, Mean RD	0.8879
Symmetry 2_Object(M07,Ch07,Tight)_Ch07, round, Mean RD	0.8404
Aspect Ratio Intensity_M07_Ch07, round, Mean RD	0.8358
Width_M07, round, Mean RD	0.7556

9. Repeat steps 6-8 for each relevant category.
10. Plot the features with the highest RD from each relevant category for the truth populations.



11. Validate by applying the feature to the base population, independent controls if available and on multiple files and experiments.
12. If you notice false positives and negatives during the validation phase, refine the truth populations and repeat.
13. If you notice more image-based phenotypes, repeat process with appropriate phenotypes

Below is a plot of CD45 versus Size which verifies the location of the lymphocytes and Neutrophils in this standard plot. Truth populations are colored and displayed using a symbol that stands out over the default simple dot. The symbol can be edited in the population manager.



NOTES ON EVALUATING THE FEATURES:

Note: Consider the features that produce the highest RD. If they are intensity based features make sure that the staining was not uneven due to technical issues. If it is a size feature, does it make sense with what you know about the cells?

Since the feature value ranges vary between features this is an approximate comparison and the result should be validated by viewing images across the feature range from the whole population.