Image Data Exploration and Analysis Software

David Basiji, William Ortyn, Cathleen Zimmerman, Richard Bauer, David Perry, Richard Esposito, Thaddeus George, Brian Hall, and Philip Morrissey

Amnis Corporation, 2505 Third Avenue, Suite 210, Seattle, WA, USA 98121

Abstract

High resolution multispectral imaging of cells can simultaneously generate morphologic, textural and photometric information, along with detailed molecular content and distribution information via fluorescence imaging. A multimode image set of an individual cell may offer several hundred quantitative parameters that provide insights into cell structure, function and lineage. Multiple cellular processes, including inter- and intra-cell signaling, regulation of proteins, morphological changes, and internalization of surface binding events, can occur along a continuum where some cells exhibit significant responses to stimuli while others are unaffected. Therefore, even a single biological process often requires the observation and analysis of tens of thousands of cells in order to apply statistical rigor to the analysis. The multispectral imaging of tens to hundreds of thousands of cells can present a data overload problem. The tools for the analysis of such data must allow the biologist to quickly exclude artifacts, "zero-in" on the specific cell type(s) under investigation, and provide sufficient discriminating power to allow the biologist to apply their specific expertise and easily communicate their findings to others. Here we provide a description of the capabilities and operation of the IDEAS® image data exploration and analysis software. We demonstrate how IDEAS automatically calculates the standard intensity-based features common to flow cytometry while also quantitating numerous morphologic and photometric features used in confocal and other forms of microscopy. We show how the biologist´s specific knowledge can be incorporated into an analysis via the creation of custom image analysis features. Finally, we describe "similarity" algorithms, which cross-correlate different images of the same cell to quantitate nuclear translocation and co-localization events.

Figure 1: IDEAS Statistical Image Analysis User Interface

The IDEAS user interface is designed to be familiar to users having either flow cytometric or image analysis backgrounds. An Image Gallery (left) allows the inspection of any and all cells in a data file, with flexible options for pseudo coloring, contrast, brightness, and compositing of imagery. In the example pseudopod formation assay shown below, each cell is represented by a row of images. Cells were imaged in darkfield (blue) and brightfield (gray). In addition, they were labeled with the DRAQ5® DNA binding dye to produce a nuclear fluorescence image (magenta) and the PODO cell surface marker (orange). The last two images in each row show the position of the PODO marker relative to the nucleus, as well as the localization of the PODO marker relative to the actin-associated marker NHERF-1 (green).

The Workspace to the right accommodates enlarged individual cell images and experiment notes, as well as the standard dot plots histograms, and gating tools found in flow cytometric data analysis packages . Tabular data are displayed for individual objects or entire populations at the top of the Workspace.

The Image Gallery is linked to the Workspace such that selecting a dot in a plot displays the associated cell imagery at the top of the Image Gallery, while selecting a cell image in the gallery highlights the associated dot in all plots containing that cell. These capabilities allow the inspection of outliers and provide visual feedback in the determination of gate boundaries. The linkage of imagery to plots gives IDEAS a "virtual cell sort" capability. Drawing a gate on a dot plot or histogram defines a sub-population, which can then be inspected in the Image Gallery. Conversely, populations can be defined by selecting individual cell images in the Image Gallery and then "back gated" onto plots to determine where cells of interest fall within histograms or dot plots.

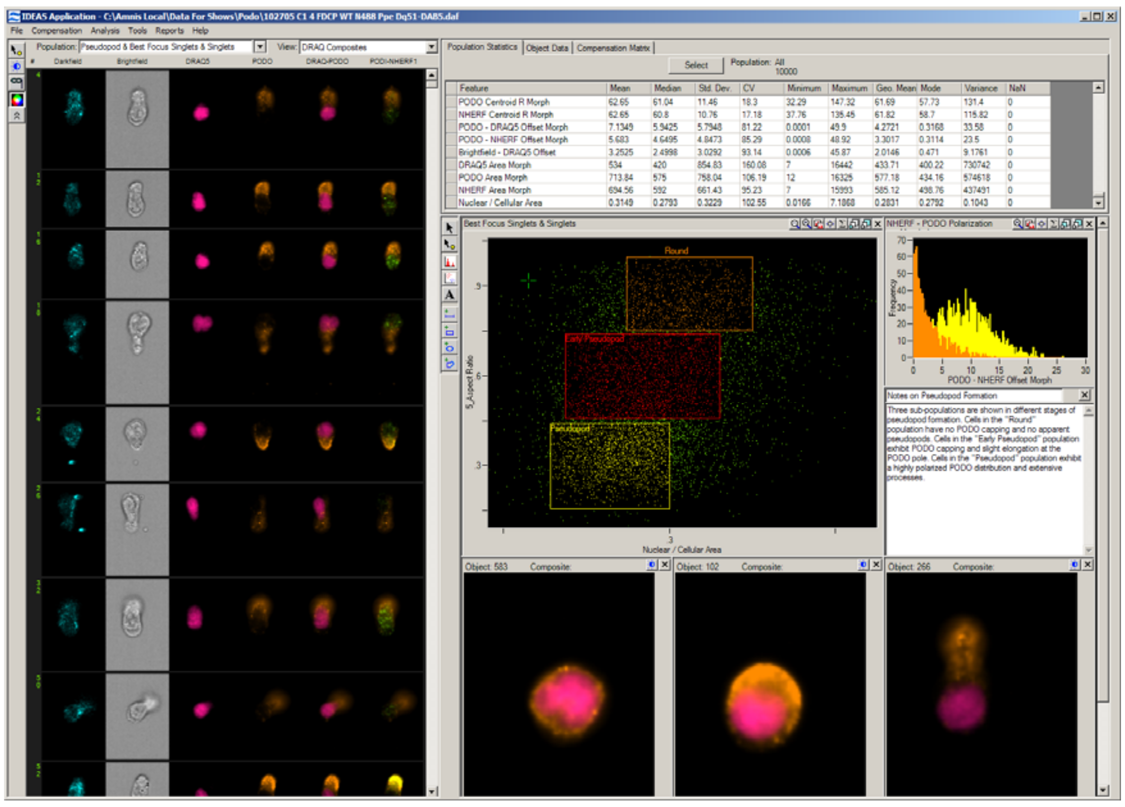


Figure 2: Quantitative Image Feature Set

When an ImageStream data set is loaded into IDEAS, the software automatically generates binary masks that define the extent of each image and then calculates over 40 quantitative features per image. If an assay is designed to use all six image channels of the ImageStream system, nearly 250 features per cell will be calculated. The basic feature set quantitates the signal intensities and distributions within a cell, as well as its size, shape, texture, and other morphologic characteristics.

Morphology	
Area:	Number of pixels in the image mask
Perimeter:	Number of pixels on the outer edge of the image mask
Combined Mask Area:	Number of pixels in the union of all of a cell's image masks
Aspect Ratio:	Ratio of the image mask's width to its height
Aspect Ratio Intensity:	Intensity-weighted ratio of the image's width to its height
Centroid X:	Center of the image mask along the X-axis
Centroid X Intensity:	Intensity-weighted center of the image along the X-axis
Centroid Y:	Center of the image mask along the Y-axis
Centroid Y Intensity:	Intensity-weighted center of the image along the Y-axis
Major Axis:	Maximum width of the image mask in any axis
Major Axis Intensity:	Intensity-weighted maximum width of the image mask in any axis
Minor Axis:	Minimum width of the image mask in any axis
Minor Axis Intensity:	Intensity-weighted minimum width of the image mask in any axis
Object Rotation Angle:	Angle relative to the Y-axis of the major axis
Object Rotation Angle Intensity:	Angle relative to the Y-axis of the intensity-weighted major axis
Compactness:	Degree of deviation of the image mask from a perfect circle
Elongatedness:	Lobicity of the image mask
Negative Curvature:	Number of invaginations on the perimeter of the image mask
Signal Strength	
Total Intensity:	Summed intensity of the pixels in the image mask
Mean Intensity:	Total Intensity divided by image mask area
Minimum Intensity:	Lowest pixel intensity within the image mask
Peak Intensity:	Highest pixel intensity within the image mask
Background Mean Intensity:	Average intensity of the pixels outside the image mask
Background StdDev Intensity:	Standard deviation of the intensity of the pixels outside the image mask
Combined Mask Intensity:	Summed intensity of the pixels in the combined image mask
Intensity:	Background-subtracted Combined Mask Intensity
Texture	
Frequency:	Variance of the pixel intensities
Gradient Max:	Maximum rate of change of the image intensity profile
Gradient RMS:	Root mean square of the rate of change of the image intensity profile
Spot Count:	Number of connected regions present in a mask.
Spot Large Max:	Highest intensity amongst all image spots that are 14-28 pixels wide
Spot Large Total:	Summed intensity of all image spots that are 14-28 pixels wide
Spot Medium Max:	Highest intensity amongst all image spots that are 7-13 pixels wide
Spot Medium Total:	Summed intensity of all image spots that are 7-13 pixels wide
Spot Small Max:	Highest intensity amongst all image spots that are less than 7 pixels wide
Spot Small Total:	Summed intensity of all image spots that are less than 7 pixels wide
Spot Raw Max:	Highest intensity amongst all image spots in all width ranges
Spot Raw Total:	Summed intensity of all image spots in all width ranges
Image Comparisons	
Similarity:	Degree to which two images are linearly correlated within a masked region, ranges from -infinity to +infinity
Similarity Normalized:	Pearson's correlation coefficient of the gray levels of the images within the masked region, ranges from -1 to 1
Similarity Bright Detail:	Same as Similarity but restricted to small, bright regions
Similarity Bright Detail Normalized:	Same as Similarity Normalized but restricted to small, bright regions
Object Information	
Camera Line Number:	Unique line number associated with each row of image data
Camera Timer:	Timestamp associated with each row of image data
Flow Speed:	Calculated flow speed at time of object passage
Object Number:	Unique serial number assigned to each object in a data file

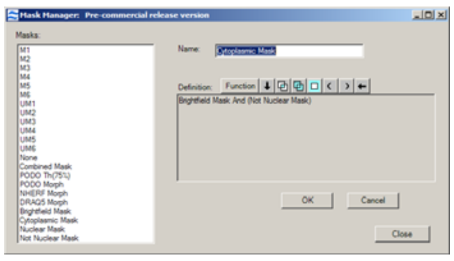
Figure 3: Custom Image Masks

The user can define custom image masks that augment the masks that IDEAS generates automatically. Custom masks can be used to define areas of overlap between different probes, to delineate specific sub-cellular regions without the need for a dedicated probe, and other purposes. The example below illustrates the definition of a cytoplasmic mask, using the brightfield and nuclear fluorescence images. Once defined, a cytoplasmic mask is generated for every cell in the data file and can be used as the basis for the calculation of additional features.

(A) The Mask Manager provides binary operators that allow the user to create a wide variety of masks by combining existing masks using Boolean logic. A cytoplasmic mask is created by subtracting the nuclear image mask from the brightfield mask, which covers the whole cell. The expression: **Brightfield Mask AND (NOT Nuclear Mask)** inverts the nuclear mask and combines it pixel-by-pixel with the brightfield mask.

(B) The image and mask components of the binary expression for the cytoplasmic mask are shown (left to right): brightfield image, brightfield mask, nuclear image, nuclear mask, and the resulting cytoplasmic mask.

(A)



(B)

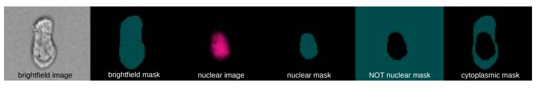


Figure 4: Custom Image Analysis Features

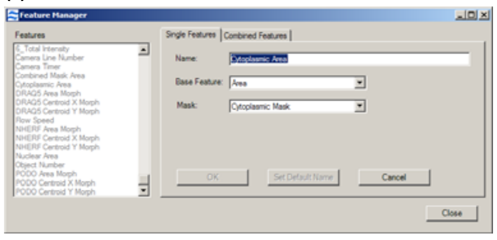
The Feature Manager allows the user to define custom image analysis features to augment the basic IDEAS feature set. New features are created by mathematically combining existing features, by analyzing properties of masks created with the Mask Manager, or by using masks to define specific regions of interest within images and quantitating sub-cellular characteristics. A simple example is the creation of a "nuclear to cytoplasmic area" feature for the pseudopod formation assay.

(A) No cytoplasm-specific stain was employed in the assay, but the cytoplasmic area can be calculated by specifying that a new "area" feature reference the Cytoplasmic Mask defined previously in Figure 3. The number of pixels in the Cytoplasmic Mask of each cell determines its cytoplasmic area.

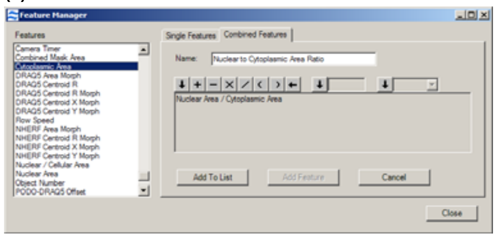
(B) The nuclear area feature was automatically generated by IDEAS from the nuclear image. Using the Feature Manager's expression builder, the Nuclear Area can be divided by the Cytoplasmic Area to produce the N/C ratio.

(C) Once the N/C ratio is defined in IDEAS, it is automatically calculated for all cells in the data file and can be used to generate dot plots and histograms. As shown in this overlay histogram, pseudopod-forming cells (yellow) tend to have smaller N/C ratios than round cells (red).

(A)



(B)



(C)

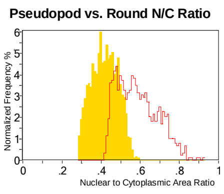
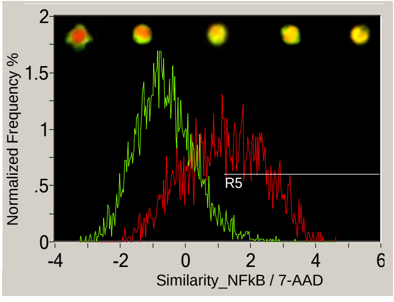


Figure 5: Image Similarity Algorithm

In addition to custom masks and features, IDEAS includes image similarity algorithms for the quantitative localization and co-localization of probes. The similarity algorithms perform a pixel-by-pixel cross-correlation between any two images of a given cell. This capability can be used, as shown in the histogram below, to quantitate the degree of translocation of transcription factor NF-κB from the cytoplasm to the nucleus. Imagery at the top of the histogram of cells having a low similarity score exhibit NF-κB (green) surrounding the nucleus (red), while cells with high similarity scores show significant red-green overlap. Control (green) and treated (red) population histograms can be overlaid and gates set at specified confidence intervals to determine the fraction of responding cells.



Conclusion

IDEAS is designed to combine the sophisticated masking and quantitative features of traditional image analysis software with powerful statistical analysis and plotting and gating tools of flow cytometry data analysis packages. Since ImageStream data files can contain hundreds of thousands of images, a major design goal was to provide an intuitive interface that could be learned quickly by biologists with little or no image analysis experience. Similarity algorithms add a simple, robust, and novel method of quantitating co-localization of probes within cells. Ease of use is further enhanced via re-usable analysis templates and the ability to batch process multiple data files.

AMNIS CORPORATION 2505 Third Avenue, Suite 210 Seattle, WA 98121 PHONE 206.374.7000 www.amnis.com