ISAC XXIII International Congress Poster 150183

Limits Of Miniaturization In High-Content Analysis: How Better Data Extraction Can Reduce Cell Number Requirement

Ilya Ravkin, Ph.D., Consultant, www.ravkin.net

Abstract

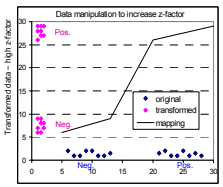
There is always a desire to miniaturize assays, including cell-based assays analyzed by imaging. This becomes a necessity when the number of available cells is limited, as with primary cells, or, by design, as in multiplexed cell arrays. Among the factors that limit miniaturization is increased variability of data due to small number of cells. Different image analysis algorithms have varying ability to extract, from images, the most stable cell features characteristic of a given assay. The algorithm most effective at extracting characteristic information will yield more precise data and will reduce the number of cells required.

Measures of Assay Quality

In high throughput drug screening it is common to evaluate the quality of assays by a statistical parameter that depends on the dynamic range and the variability of the assay. Several such parameters have been introduced; Z-factor (Ref. 1) being the most popular. These measures have proven to be very useful in capturing and comparing variability caused by assay biology and by instrumentation (e.g., pipetting).

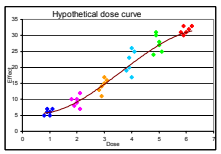
$$Z = 1 - 3\left(\frac{SD_{pos} + SD_{neg}}{|M_{pos} - M_{neg}|}\right)$$

SD is standard deviation, *M* is mean, *pos* and *neg* are the two extreme states of the assay, which define its dynamic range



Assays based on imaging introduce several new variables: imaging resolution, size of the imaged area and the data extraction algorithm. Having a quality measure, such as the Z-factor, allows optimization of factors that can be controlled, e.g., the best data extraction algorithm. In addition to introducing new variables, cellular imaging assays lead us to reconsider the quality measure itself. An assay measure derived from an image may be computationally very complex. It may contain operations that saturate the values from the positive and negative states of the assay, thus artificially reducing variability. This may happen unintentionally. Moreover, the Z-factor can be manipulated intentionally by applying a mathematical transformation that maps all positive values into a single value and all negative values into another single value, resulting in a Z-factor of 1. Generalizing the quality measure to a dose-curve eliminates the likelihood of artificial manipulation. Depending on the dose-response model used, there could be variations in how the Z-factor is generalized. We call this family of generalized measures V-factors. These measures consider all points in the dose-curve with equal weights. Below are two variants of V-factor that are most useful.

$$V_L = 1 - 6\left(\frac{SD_{of\_fit}}{|M_{pos} - M_{neg}|}\right)$$



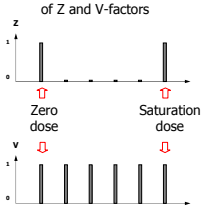
$$SD_{of\_fit} = \sqrt{\frac{1}{n}\sum(f_{logistic} - f_{experiment})^2}$$

Formula for logistic model

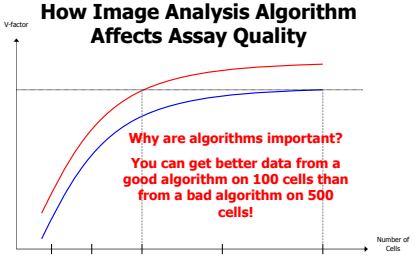
$$V_A = 1 - 6\left(\frac{Avg\_SD}{|M_{pos} - M_{neg}|}\right)$$

Alternative definition without a model

Weights of different concentrations in calculation of Z and V-factors



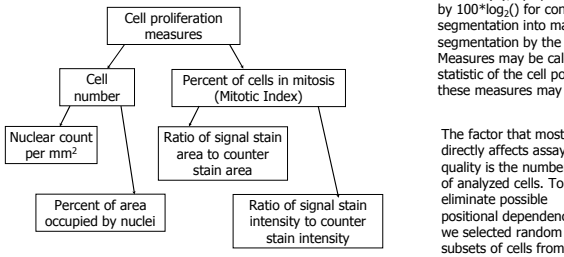
How Image Analysis Algorithm Affects Assay Quality

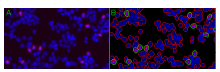


Typical curve of assay quality vs. cell number

Study of algorithm and parameter dependency of assay quality is easier to implement than studies of other dependencies because it does not require a biological sample or imaging equipment and can be done on fixed sets of images asynchronously by different people. We have collected several sets of images which are now publicly available through the Society for Biomolecular Screening (*www.ravkin.net/SBS/D&IA_SIG.htm*). Most images used in this poster are taken from these sets.

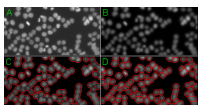
Cell Number Dependency for Cell Proliferation Assay. Comparison of 4 Measures



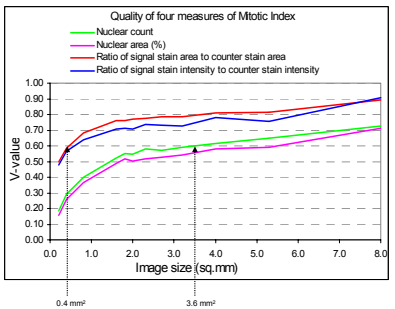


A - Image of Mitotic Index Assay.
counter stain – blue
mitotic phase stain - red

B - Adaptive threshold contours.
counter stain contour - red
signal stain contour - green

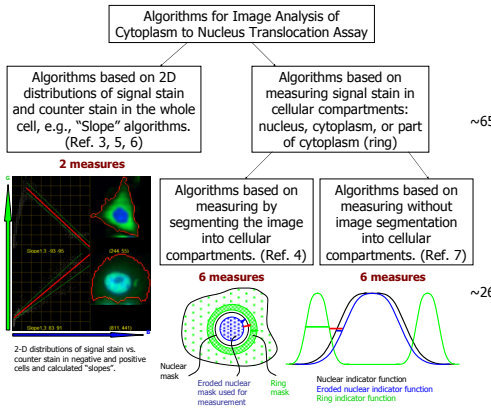


Counting of nuclei:
A - image of counter stain
B - smoothed image
C - smoothed image with adaptive threshold contours
D - contours with watershed separation lines inside



V-factors of four measures of cell proliferation at objective magnification 2X as a function of image size. There is a clear advantage in using relative measures: the image size at which V-factor reaches 0.6 is 0.4mm² for the ratio-of-areas measure, and 3.6mm² for the nuclear-count measure.

Cell Number Dependency for Cytoplasm to Nucleus Translocation Assay. Comparison of 10 Measures

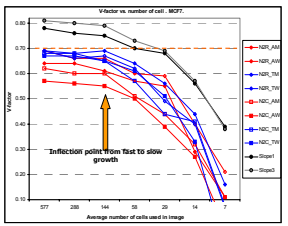


2-D distributions of signal stain vs. counter stain in negative and positive cells and calculated "slopes".

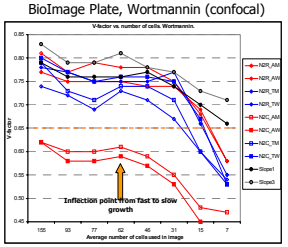
A family of assay measures can be constructed as ratios of the average (A) or total (T) amounts of stained protein in the three compartments: nucleus (N), cytoplasm (C) and ring (R). The ratio measures are scaled by $100*\log_2()$ for convenience . The measures may be produced by segmentation into masks of nucleus and cytoplasm (M) or without segmentation by the use of indicator functions (W), see Ref. 7. Measures may be calculated for each cell; the well measure is some statistic of the cell population, e.g., mean or median; alternatively, these measures may be calculated on the whole image.

The factor that most directly affects assay quality is the number of analyzed cells. To eliminate possible positional dependency, we selected random subsets of cells from the whole image, not contiguous image fragments. At a certain number of cells, the quality curve changes from fast to slow growth. The curves in each plot are shifted only in Y, but not in X, which means that this number depends on images, but not on algorithms. Note that some algorithms do not reach the quality of other algorithms at any number of cells.

Vitra Plate, MCF7 Cells (widefield)

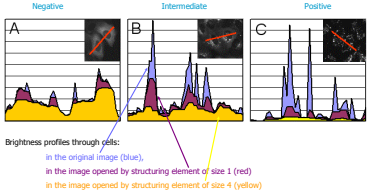


BioImage Plate, Wortmannin (confocal)

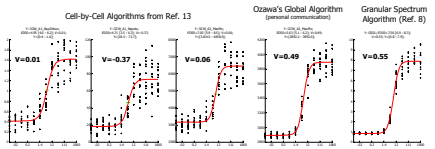


Cell-By-Cell vs. Global Analysis and Cell Number Dependency for Transfluor Assay. Comparison of 5 Measures

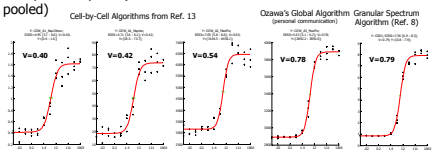
The basis of our method is the concept known in mathematical morphology as size distribution (Ref. 9), granulometry (Ref. 11), pattern spectrum (Ref. 10) or granular spectrum (Ref. 8). This distribution is produced by a series of openings of the original image with structuring elements of increasing size. At each step the volume of the open image is calculated as the sum of all pixels. The difference in volume between successive steps of opening is the granular spectrum. The distribution is normalized to the total volume (integrated intensity) of the image. The following diagram shows how openings of increasing size affect images with different granularity.



~65 cells per data point (CompuCyte image set)



~260 cells per data point (same as above but scan fields of the same well are pooled)



Conclusions

1. The key to determining cell number requirement and comparing data extraction methods is a quantitative measure of assay quality.
2. V-factor is less susceptible to computational artifacts than Z-factor and gives a more realistic measure of assay performance.
3. The assay quality curve as a function of the number of cells is not linear. It has areas of fast growth and of slow growth, finally reaching saturation. This behavior is replicated on different assays and different algorithms.
4. Different assays may have different cell number requirements. Different algorithms for the same assay may have different cell number requirements. Different implementations of conceptually the same algorithm may have different cell number requirements.
5. Measures that intuitively "make sense" do not always give the best statistical quality. Specifically, cell-by-cell analysis is not necessarily better than global analysis.

References

1. J.-H. Zhang, T.D.Y. Chung, K.R. Oldenburg "A simple statistical parameter for use in evaluation and validation of high throughput screening assays", J. Biomol. Screening 4: pp. 67-73, 1999
2. I. Ravkin "Quality Measures for Imaging-based Cellular Assays" SBS 2004 conference poster #P12024 (*)
3. I. Ravkin, V. Temov, A.D. Nelson, M.A. Zarowitz, M. Hoopes, Y. Verhovsky, G. Ascue, S. Goldbard, O. Beske, B. Bhagwat, H. Marciniak "Multiplexed high-throughput image cytometry using encoded carriers", Proc. SPIE Vol. 5322, pp. 52-63, 2004 (*)
4. US Patent 6,671,624 "Machine readable storage media for detecting distribution of macromolecules between nucleus and cytoplasm in cells".
5. I. Ravkin, V. Temov, "Image analysis without segmentation - A new method to measure cytoplasm to nucleus translocation", poster, High Content Analysis Conference, San Francisco 2005. (*)
6. US Patent application 20050186554A1 "Image analysis and assay system" (*)
7. I. Ravkin, V. Temov "Comparison of several classes of algorithms for cytoplasm to nucleus translocation" SBS 2005 Conference poster #P02025 (*)
8. I. Ravkin "Statistical properties of algorithms for analysis of cell images", IBC's Inaugural Conference on High-Content Analysis, November 17-18, Washington DC. (*)
9. J. Serra, Image Analysis and Mathematical Morphology, Vol. 1. Academic Press, London, 1989
10. Maragos, P. "Pattern spectrum and multiscale shape representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, N 7, pp. 701-716, 1989
11. L. Vincent "Granulometries and Opening Trees", Fundamenta Informaticae, 41, No. 1-2, pp. 57-90, IOS Press, 2000.
12. I. Ravkin, V. Temov "Bit representation techniques and image processing", Applied Informatics, v.14, pp. 41-90, Finances and Statistics, Moskow, 1988 (in Russian)
13. K. Ozawa, C. Hudson, K. Wille, S. Karaki, R. Oakley "Development and validation of algorithms for measuring G-protein coupled receptor activation in cells using the LSC-based imaging cytometer platform", Cytometry Part A, Volume 65A, Issue 1, pp.69-76, 2005

(*) Available at http://www.ravkin.net/