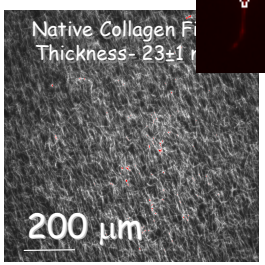
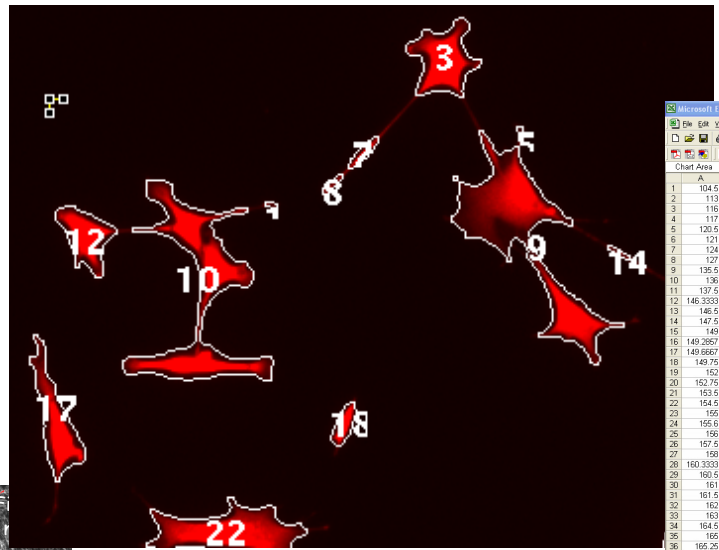
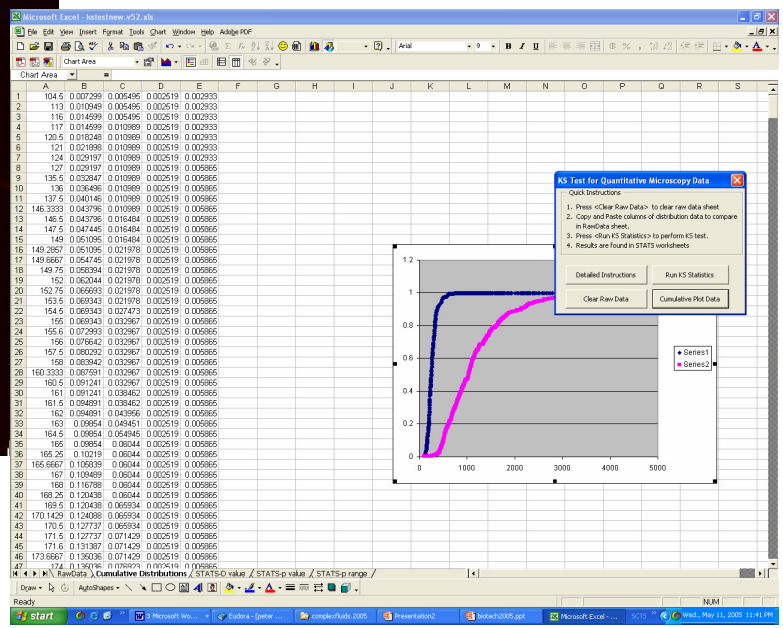
Identifying Statistically Relevant Differences between Distributions of Cell Response







John T. Elliott, Kurt Langenbach, Alex Tona and Anne Plant
NIST/Biotechnology- Cell and Tissue Measurements Group
100 Bureau Drive Gaithersburg, MD. 20899

SBS 2005

Special Interest Group, SBS 2005

Identifying statistically relevant differences between distributions of cell response.

John T. Elliott, Kurt L. Langenbach, Alex Tona and Anne L. Plant

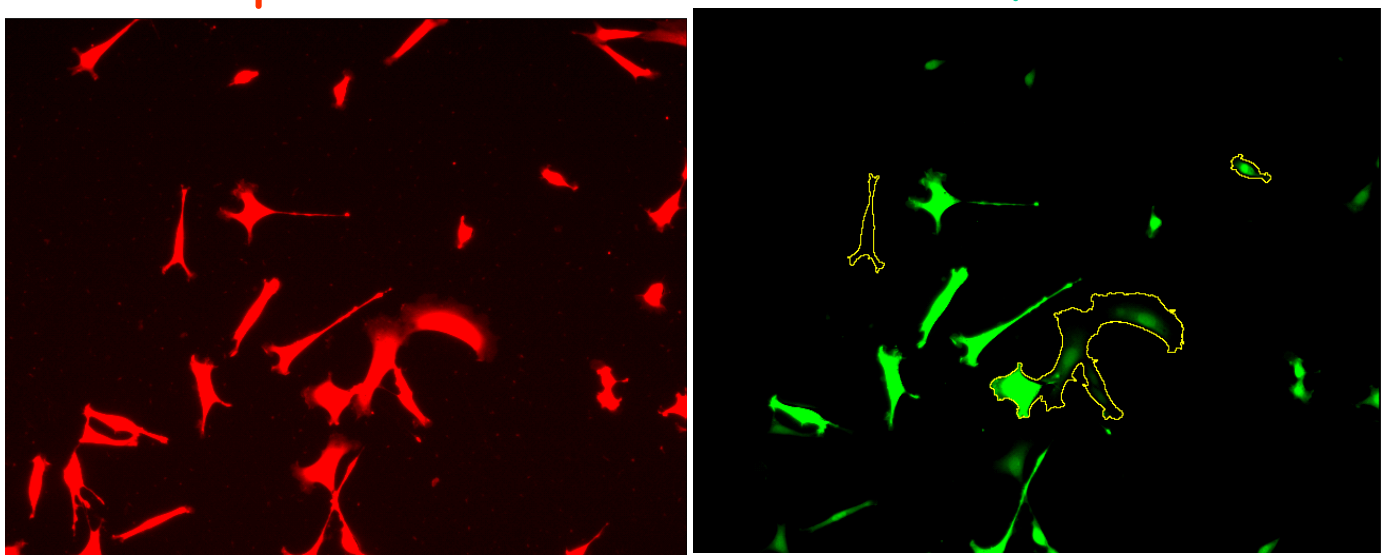
Cell and Tissue Measurement Group
NIST/Biotechnology Division
100 Bureau Drive
Gaithersburg, MD 20899
Jelliott@nist.gov

Advances in automated microscopy technologies, cell-based assays, and image analysis tools have allowed quantitative automated microscopy to emerge as an in vitro screening tool for a variety of applications. Unlike assays which require lysis of a population of cells and determination of an average cell response, microscopy-based assays can provide cell response data from individual cells. An interesting challenge in collecting and analyzing data from individual cells is that the cells in a population exhibit a range of biological responses even when the population is derived from a single clone and the extracellular environment for each cell is identical. To quantify differences between cell populations, comparisons of distribution data are required. By employing highly reproducible and analytically validated extracellular matrix protein-coated cell culture surfaces, and robust staining techniques, we are able to collect microscopic images of cells that can be used as reference data for image analysis development and statistical testing procedures. Currently, we are examining the use of the Kolmogorov-Smirnov test (KS test) to parameterize differences between distributions and to determine the statistical relevance of the difference. Our results show how factors such as image segmentation, cell number and experimental variabilities influence the statistical testing results. This information can provide guidance for developing control experiments that generate information about the quality of the experimental techniques and the image analysis procedures.

Measuring Cell Response

Cell Shape

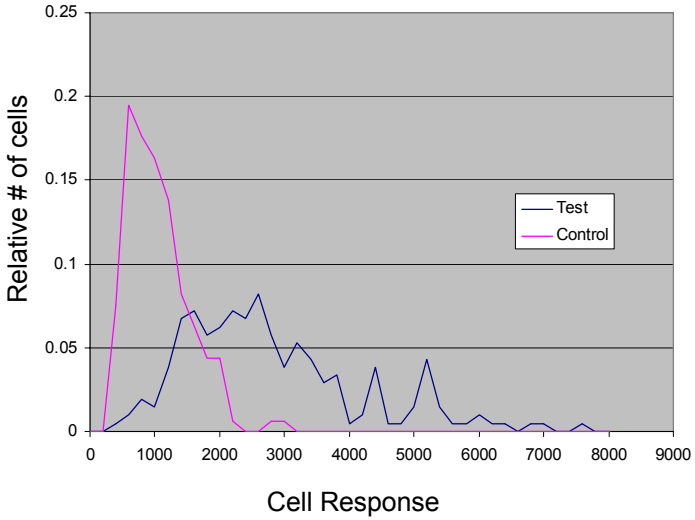
Protein Expression



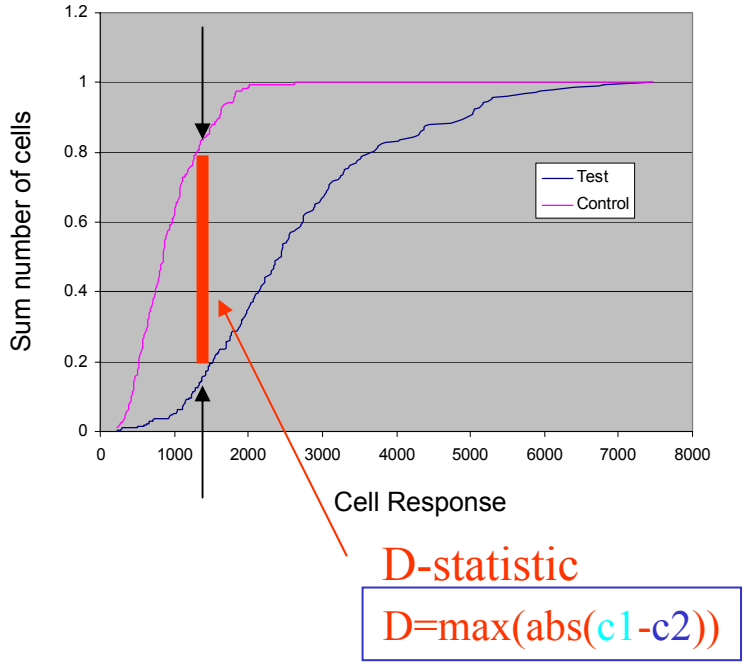
Clonal population of NIH 3T3 fibroblast expressing GFP.

- Response is not identical in every cell.
- Cell response data is in the form of a population distribution.
- The question: "Is the test cell response significantly different than the control cell response?"

KS Test and the D-Statistic



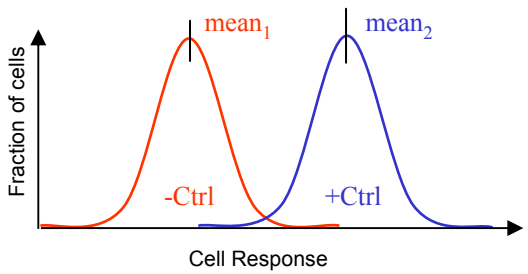
Prepare Cumulative Distribution

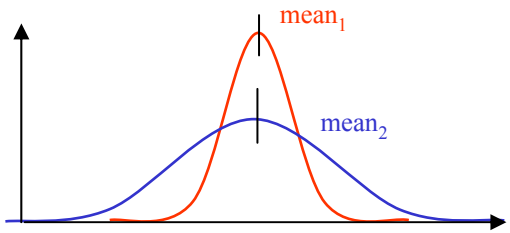


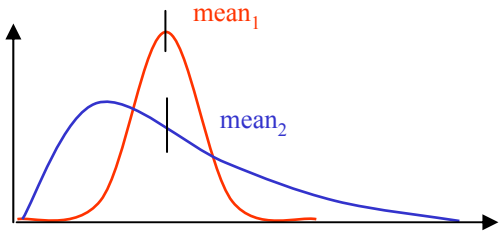
- Kolmogorov-Smirnov (KS) test provides a metric for comparing distributions of data.
- The D-statistic is the maximum absolute vertical distance between two cumulative distributions.
- It is sensitive to changes in distribution position and shape.
- It varies from 0 to 1.

Advantage of using a D-statistic over mean value differences

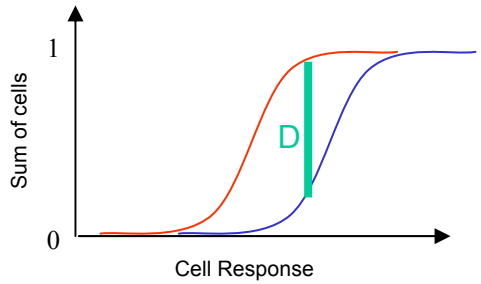
Response Distributions

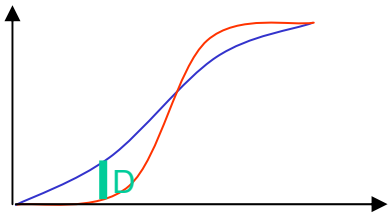


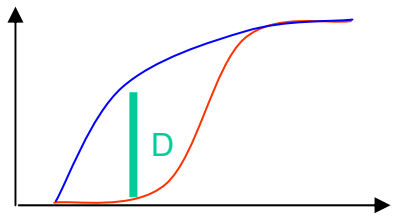




Cumulative Distributions





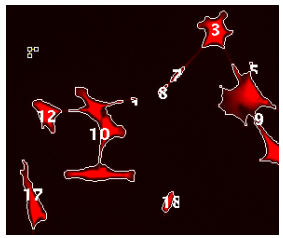


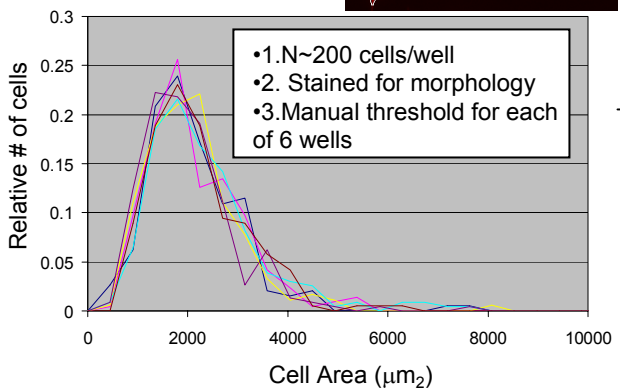
Measurable Difference?

	Mean Value	D-statistic
	Yes	Yes
	No	Yes
	No	Yes

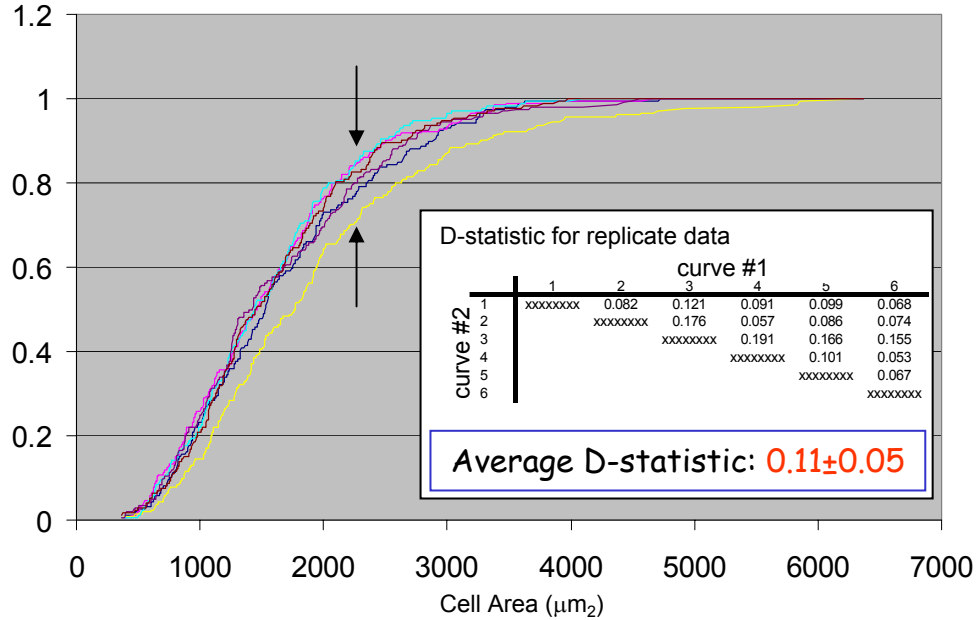
Experimental Noise using D-statistic with Replicate Data

Cell morphology -projected cell area





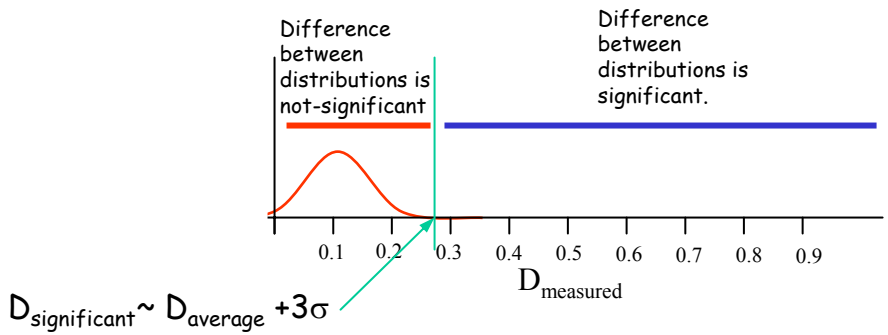
Generate Cumulative Distribution



•Computing the average D-statistic between replicates provides information about the minimum D that can be detected.

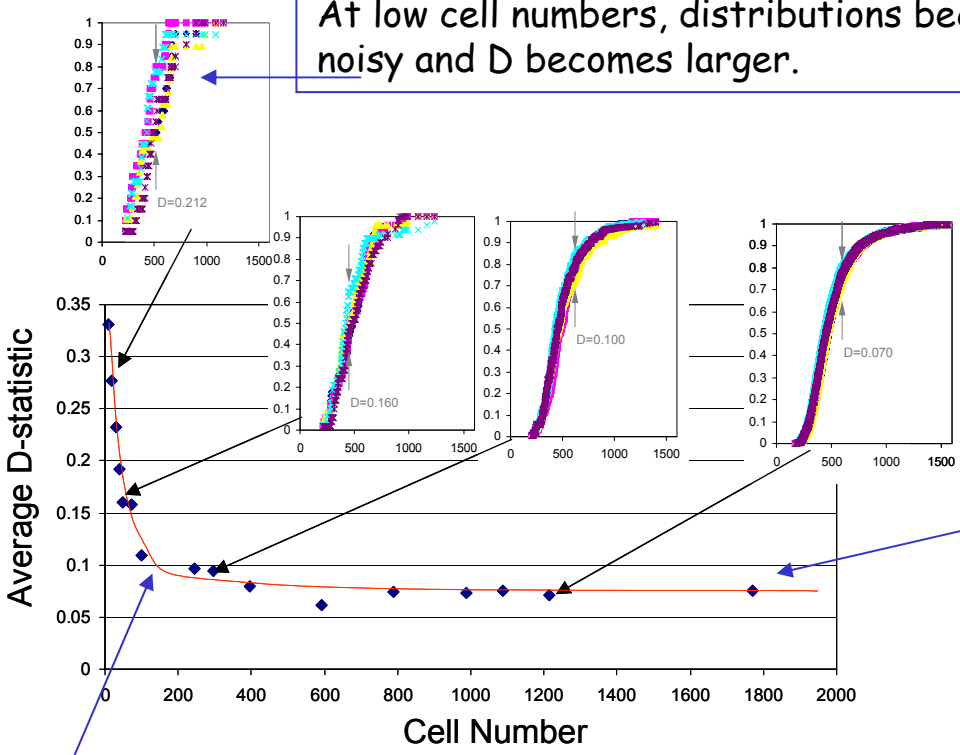
Noise Elements

- Pipetting variabilities
- Staining
- Segmentation (threshold)
- Cell number
- Well Variabilities
- Instrument



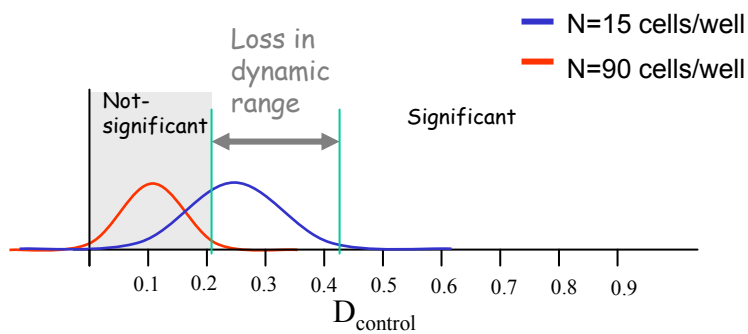
Cell number influences replicate D-statistics.

At low cell numbers, distributions become noisy and D becomes larger.



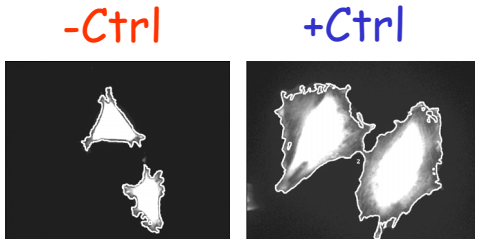
At high cell numbers, D approaches the limit of the D measurement.

Using this cell number (N~100) will maximum sensitivity of assay.

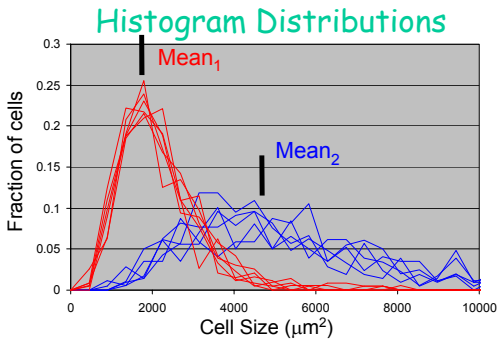


- The number of cells required to establish a minimum average D-statistic for replicate measurements is optimal for assay parameters.

Using D-statistic to determine the quality of an assay (Z-factor, cell morphology)

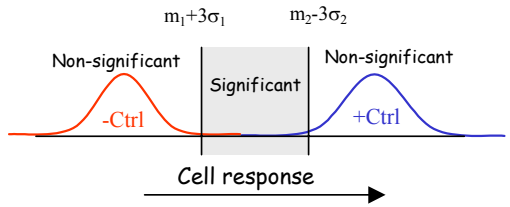


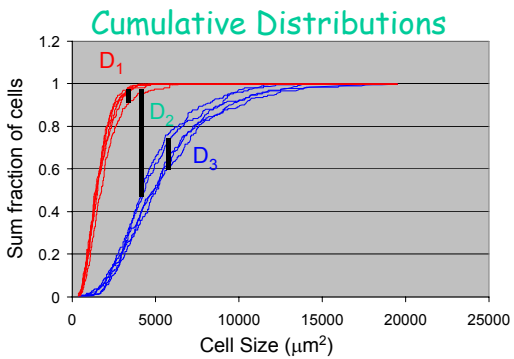
$$Z = 1 - \frac{(3\sigma_1 + 3\sigma_2)}{|m_1 - m_2|}$$



$Mean_1$=1686±148 (n=6)
$Mean_2$=5282±404 (n=5)

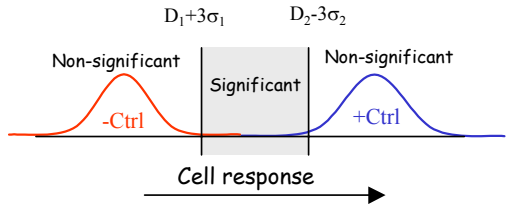
Z=0.53 (~200 cells/well)





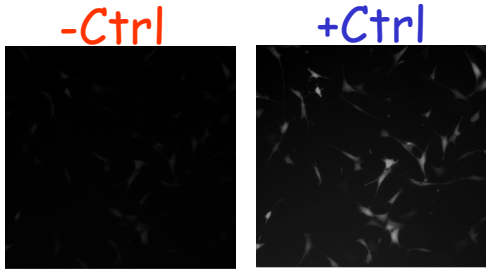
D_1=0.12±0.05 (n=6)
D_2=0.75±0.04 (n=30)
D_3=0.12±0.05 (n=5)

Z_{1-2}=0.57 (~200 cells/well)
Z_{2-3}=0.58

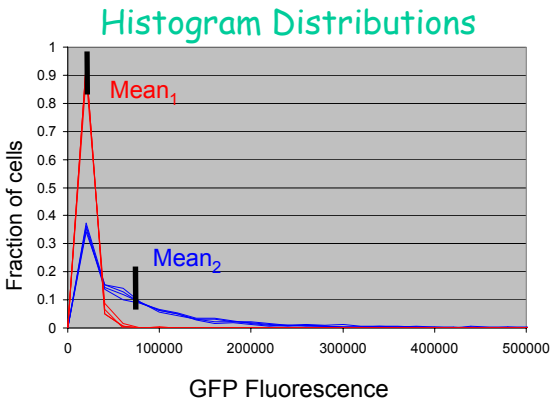


- Projected cell area distributions are nearly normal distributions.
- Z-factors are similar whether using mean values or D-statistic for calculation.

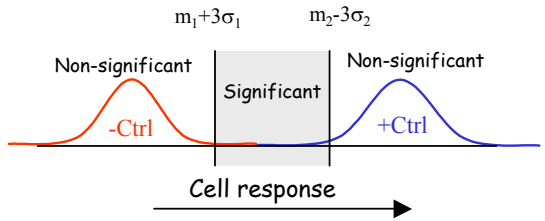
Using D-statistic to determine the quality of an assay (Z-factor, GFP expression)



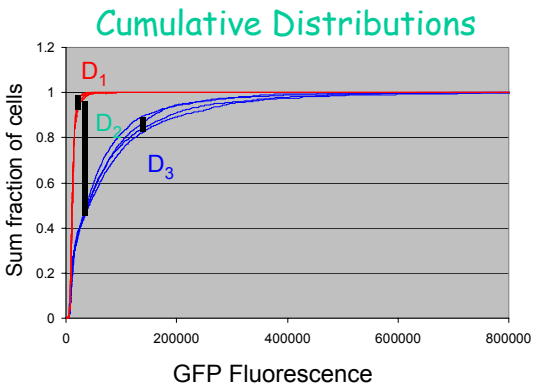
$$Z = 1 - \frac{(3\sigma_1 + 3\sigma_2)}{|m_1 - m_2|}$$



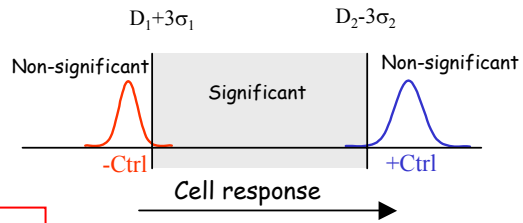
$Mean_1$=12636±304 (n=4)
$Mean_2$=72575±8491 (n=4)



Z=0.56 (~1000 cells/well)



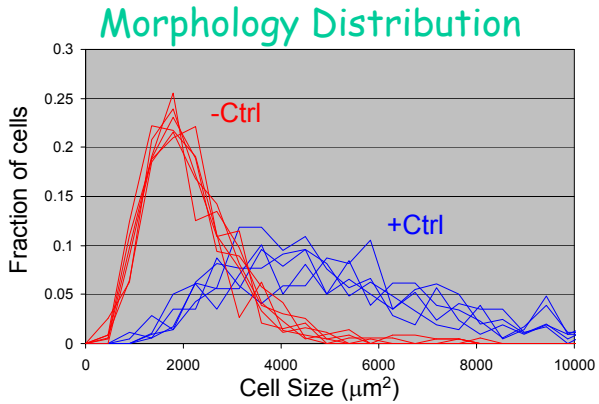
D_1=0.06±0.01 (n=4)
D_2=0.58±0.02 (n=16)
D_3=0.05±0.01 (n=4)

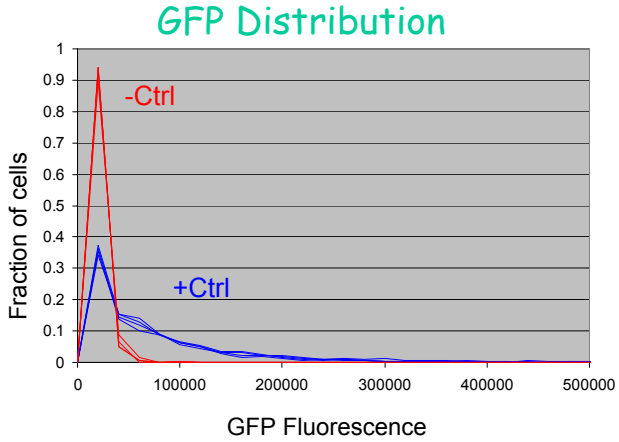


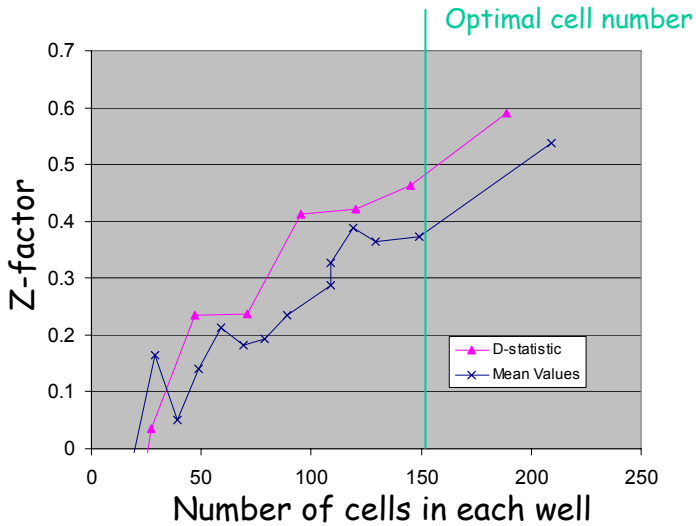
Z_{1-2}=0.82 (~1000 cells/well)
Z_{2-3}=0.83

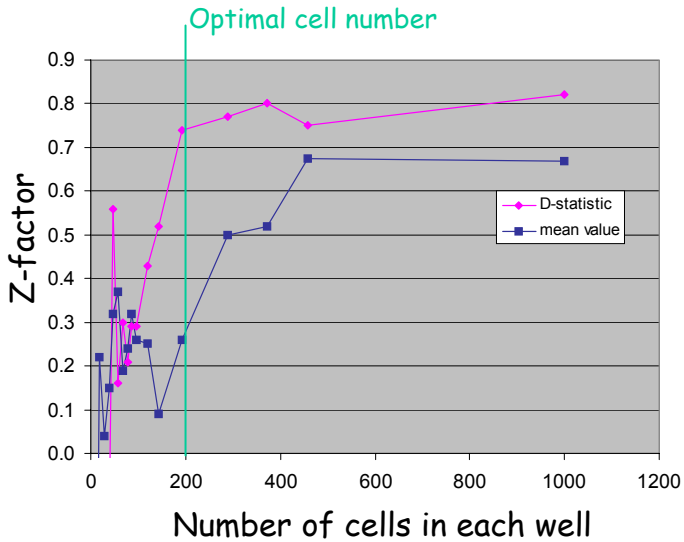
- Using D-statistic to compare distributions significantly improves Z-factor for this assay.

Effect of Cell Number on Z-factor of Test Assays









- Ideal cell number for each assay is estimated as highest Z-factor for lowest number of cells.
- Z-factor becomes noisy at low cell numbers due to undersampling.

Conclusion

- The KS test (D-statistic) can be used as a metric to determine differences between cell response distributions.
- The D-statistic is sensitive to changes in distribution position and shape.
- The average D-statistic between replicate experiments defines the minimum D-statistic that can be reliably measured.
- The "minimum number of cells to use per well" can be determined by plotting the average D-statistic vs. cell number.
- In assays with a broad, overlapping or uneven cell response distributions, the use of a D-statistics may provide a better Z-factor score than that calculated with mean values of cell response.

Tools Web Site

Experimental procedures, reference images, image analysis software/plugins and statistical analysis macros will be available in the Quantitative Cell Biology link at:

http://www.cstl.nist.gov/biotech/Cell&TissueMeasurements/Main_Page.htm