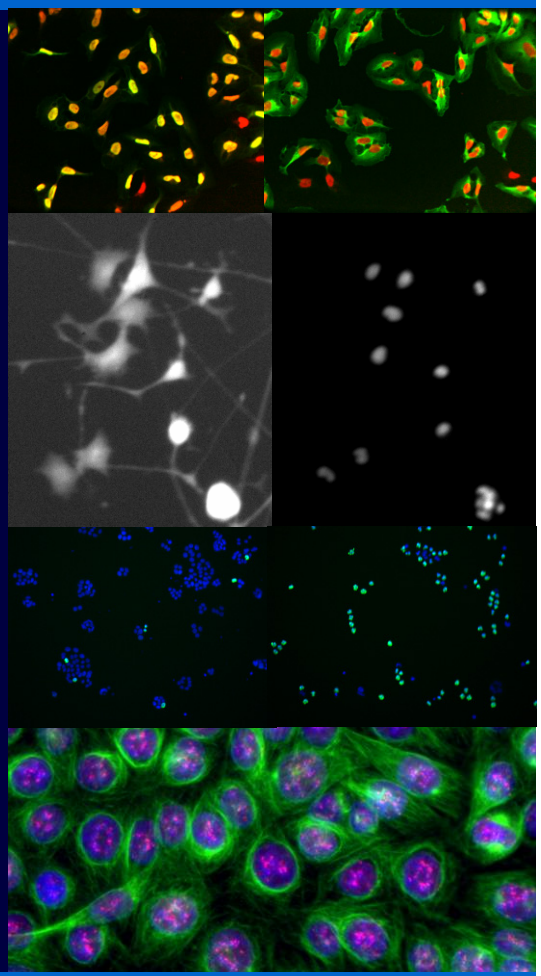
MILLENNIUM®

Breakthrough science. Breakthrough medicine.SM

Techniques in High-Content Screening and Assay Development Workshop

# Developing a Robust Automated Image-Based Assay

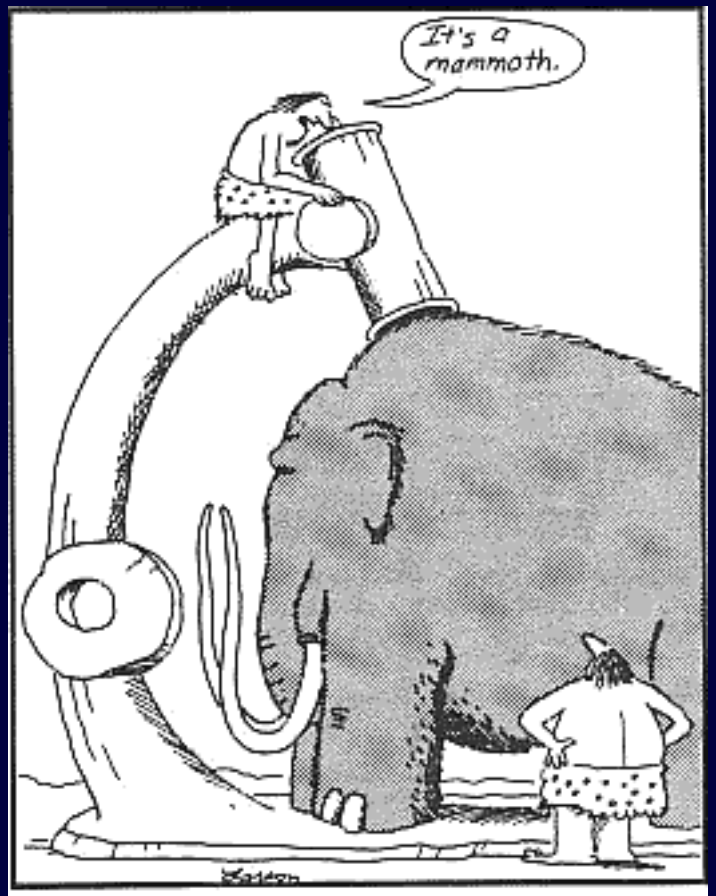


Doug Bowman,
Sr. Mgr Imaging Sciences
Millennium Pharmaceuticals

Drug Discovery Technology
August 7, 2006

# The Big Picture

The Far Side *by Gary Larson*

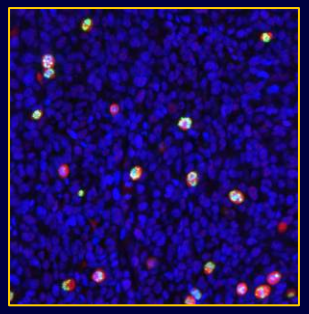


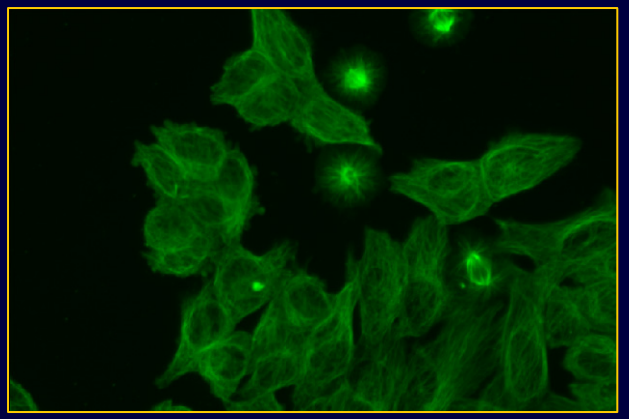

Early microscope

- Cell-based assay
- Immunofluorescence
- Technologies / Instrumentation
- Screening Assay
- People

# HCS Assays using Immunofluorescence

- IF Staining
  - Fix cells to substrate
  - Permeabilize membranes
  - Block
  - Add primary
  - Add secondary
- Image Acquisition
- Image Analysis

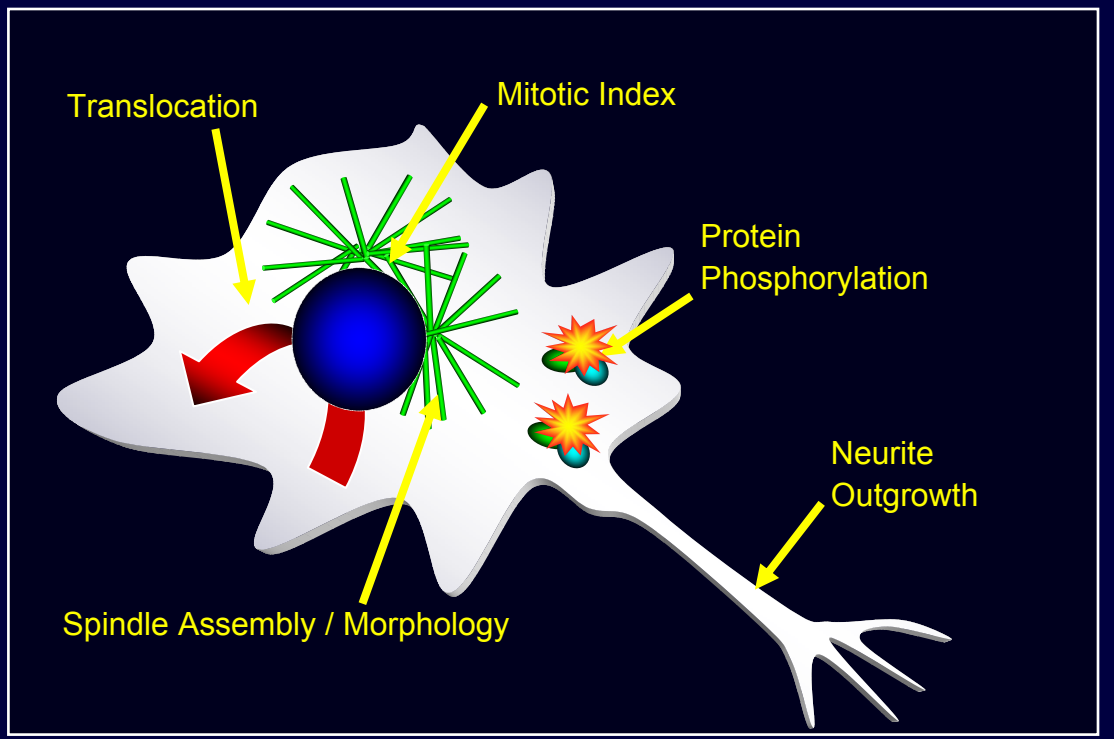




# Cellular “High Content” Imaging

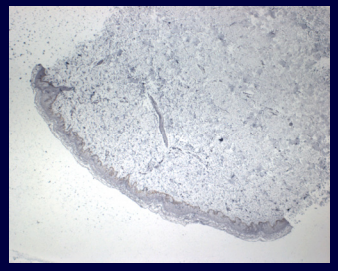
- Provide physiological context for assaying drug activity via the quantitative analysis of cellular and sub-cellular events.
- Analysis also allows for the classification of cellular phenotypes based on morphological parameters.

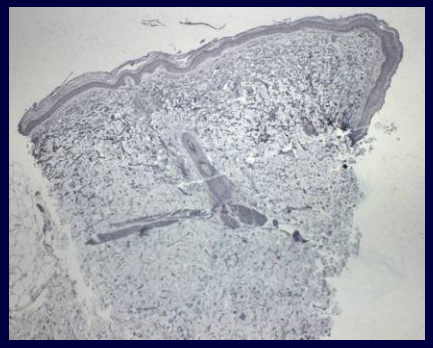
“HIGH CONTENT”

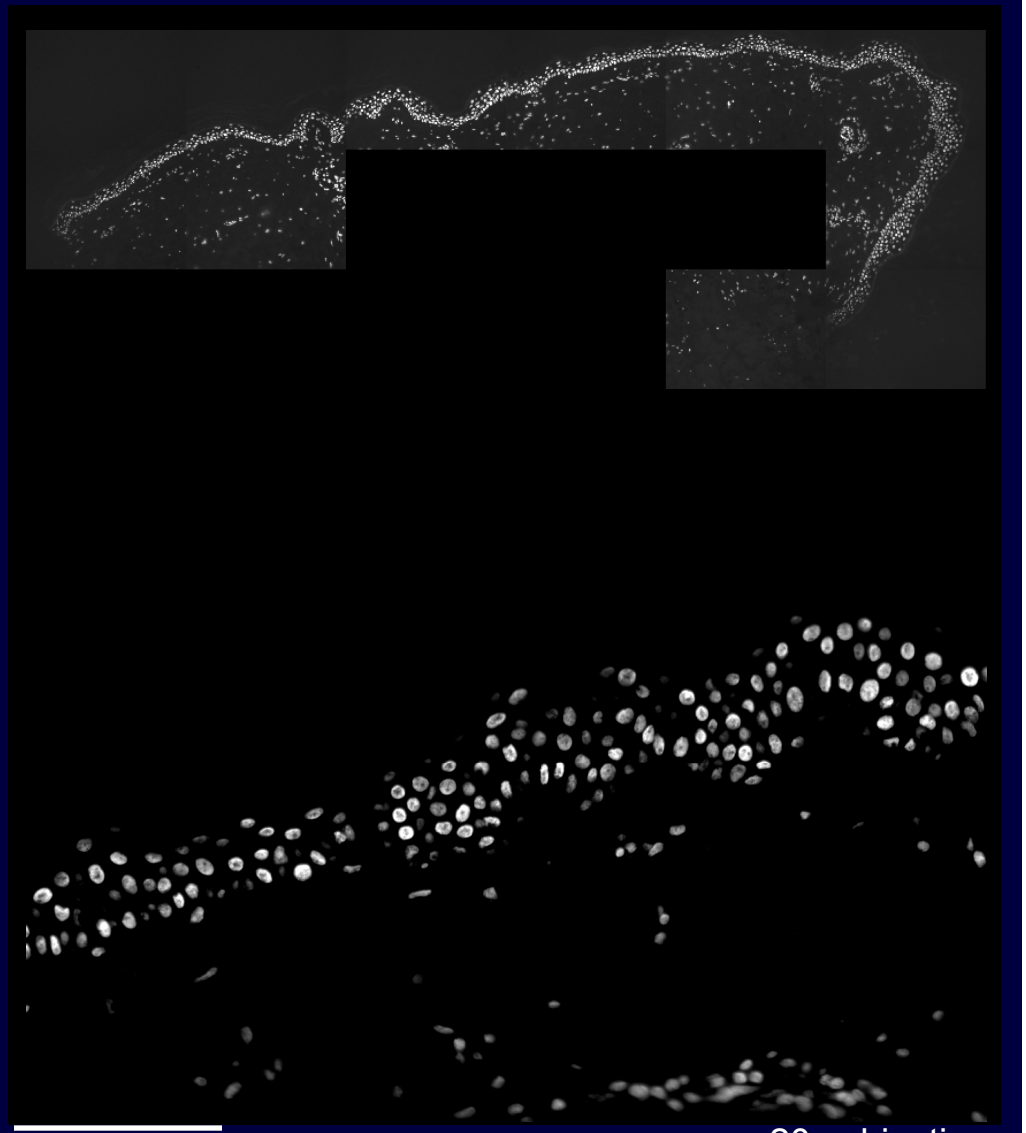


# Tissue…

- High content also!
  - Many of same assays
- Similar issues
- High resolution scanning



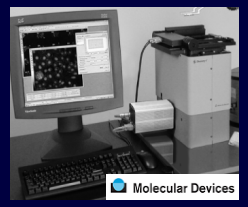




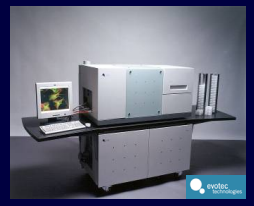
# Technologies

- Instrumentation
  - Liquid handling robotics
  - Automated microscope-based systems capable of acquiring high resolution images

Discovery-1




Opera



Automated Microscope

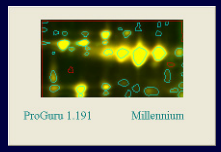
InCell1000

ArrayScan

- Software algorithms

MetaXpress

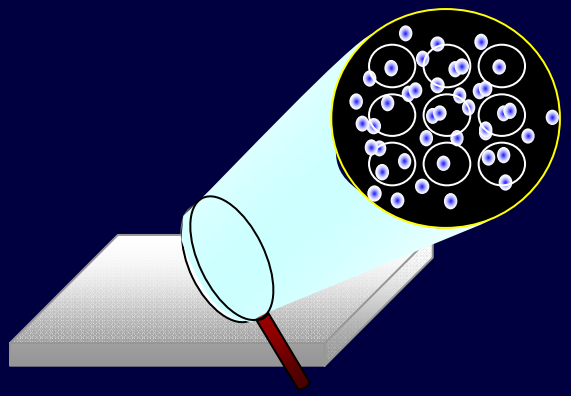
ProGuru




Acapella

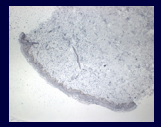
# Image-based assays

- Large volume of data



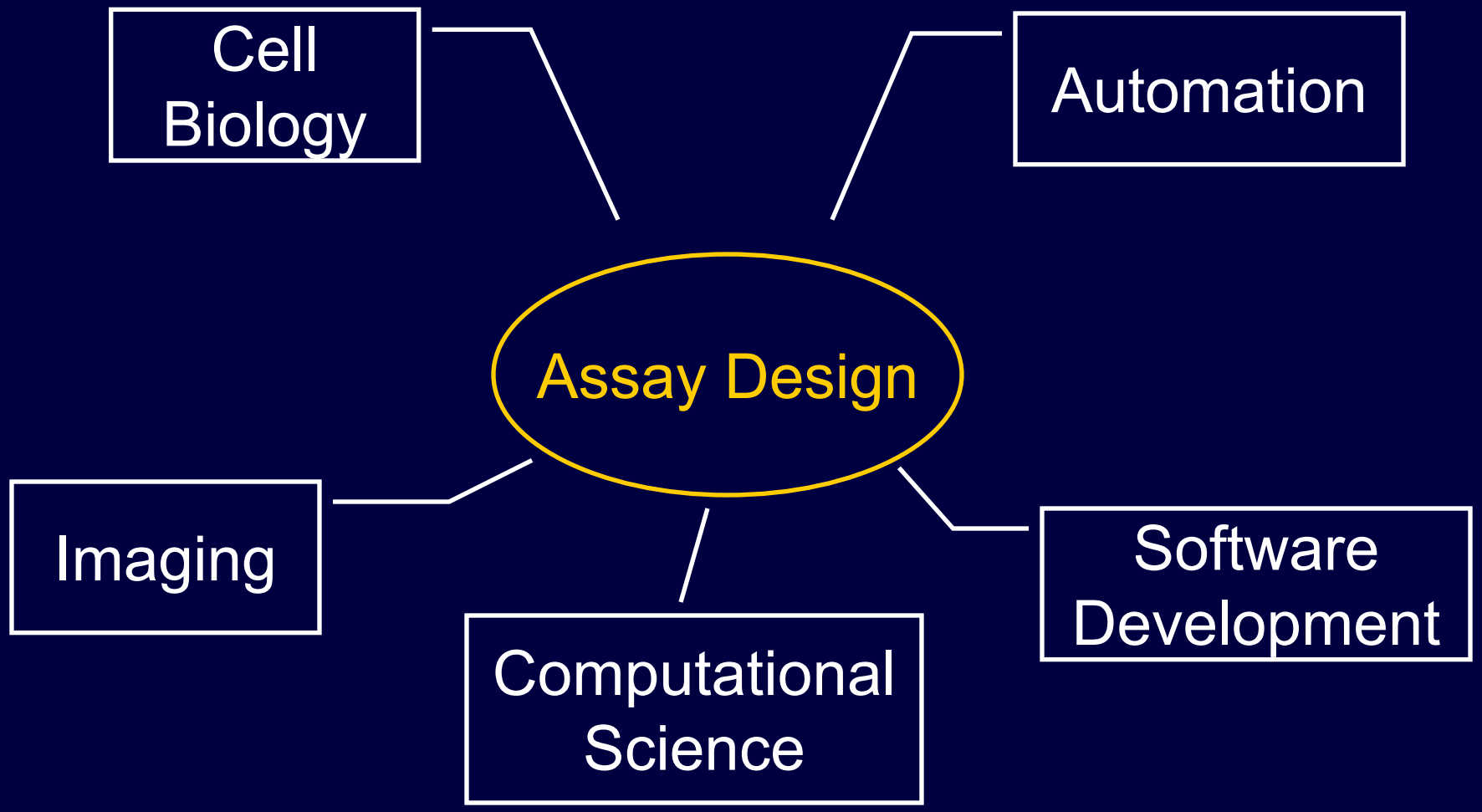
= 2λ / image * 9 sites / well * 96 wells / plate
= 1728 images / plate
= 1GB+ / plate
3 plates / hr (Opera)  1 plate / hr (Discovery-1)

= 2λ / image * 5 sites / sample * 3 mice / timepoint
= 30 images / timepoint
= 400+MB / 5 timepoints



= 3λ / image * 15 sites / membrane
= 30 images / sample
= 120MB / sample

# Working Group



Developing a Robust Automated Image-Based Assay:

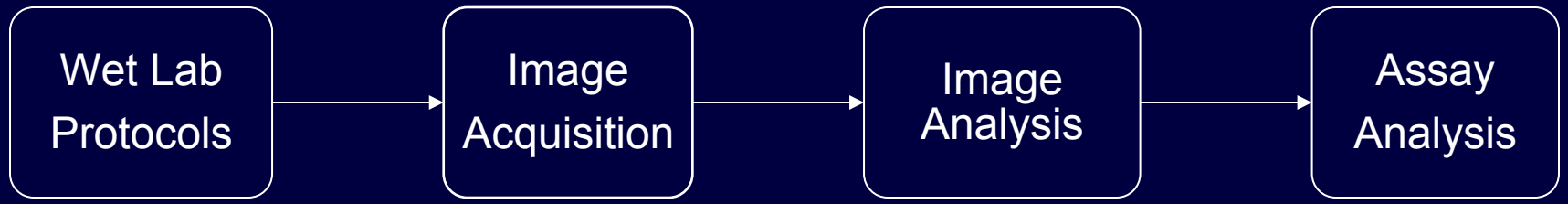
# Guidelines

- Use automation where possible
- Visualize images
- Develop robust process controls
- Use data-driven approach

# A data-driven approach to assay design

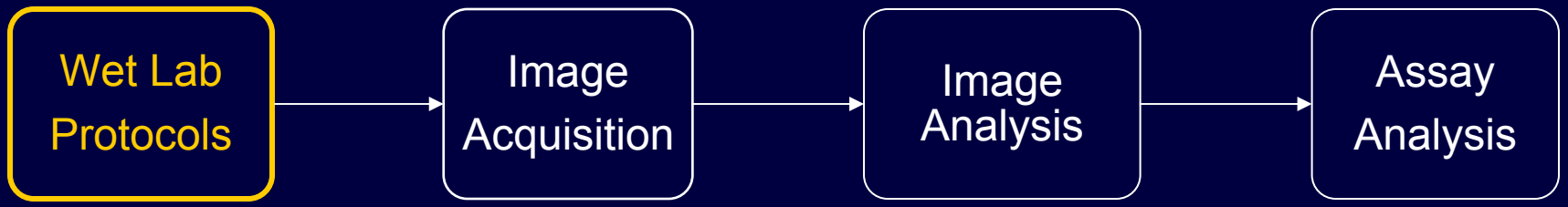
- Dissect components of the assay
- Using control experiments
  - Identify drivers of variability
    - Minimize them experimentally or account for them statistically
  - Optimize parameter selection (robust)
- Assess performance of assays in project-relevant context

# HCS Assay Components



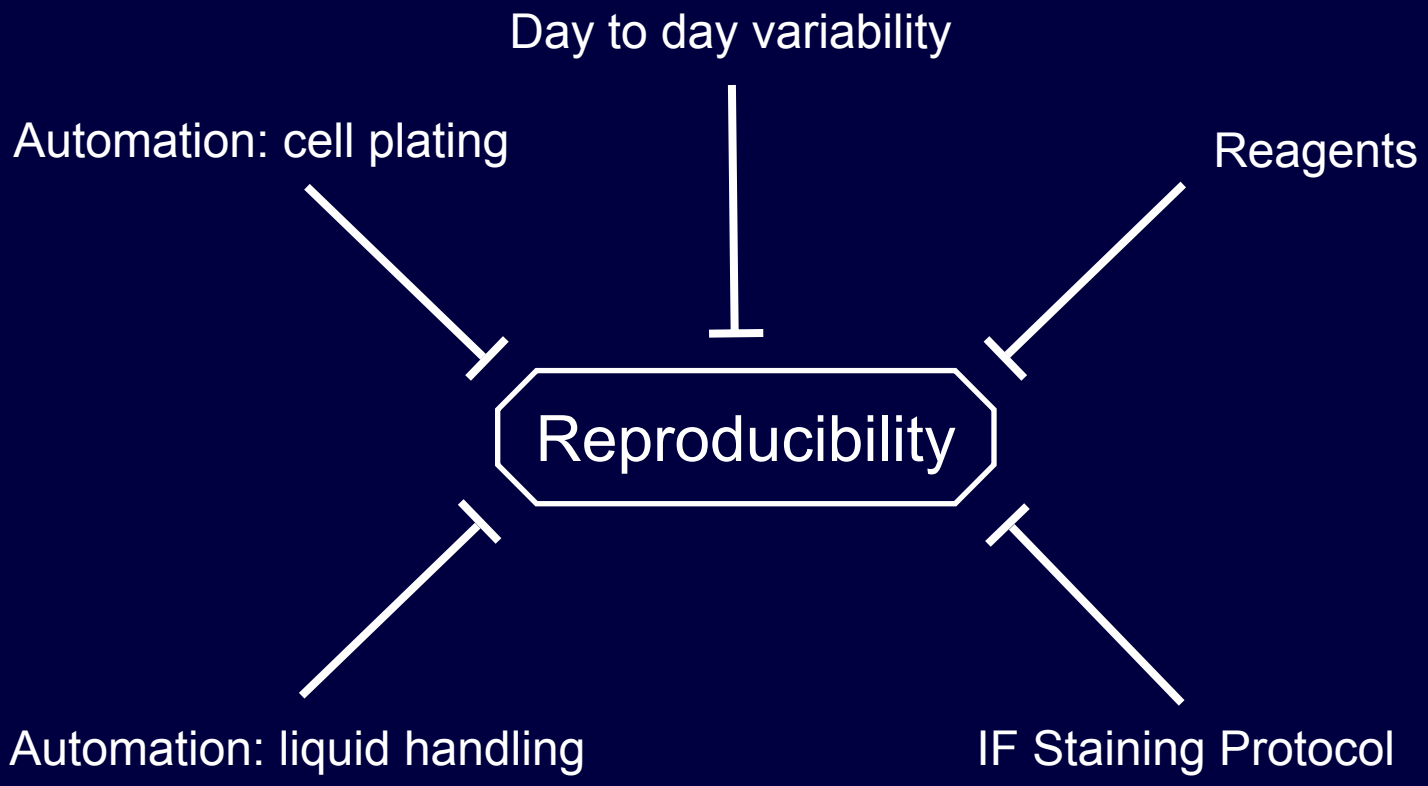

- Dissect components
- Using control experiments
  - Identify drivers of variability
  - Minimize them experimentally or account for them statistically

# HCS Assay Components




- Dissect components
- Using control experiments
  - Identify drivers of variability
  - Minimize them experimentally or account for them statistically

# Wet Lab Protocols



Wet Lab Protocols:

# Testing liquid handling using Artel system

- Accuracy and Reproducibility
  - Compound addition / dilution
  - Cell Plating
  - Antibody addition / protocols

| | | CVs |
|---|---|---|
| 100ul | manual (12 channel Multipipette) | 4.43% |
| 100ul | Multidrop | 1.67% |
| 50ul | Platemate | 0.75% |
| 50ul | Tecan | 0.80% |
| 10ul | Precision 2000 | 1.34% |


Artel MVS

Older Multidrop tube cassettes found inaccurate – *relative inaccuracy ~25% and CVs ~2%*

New cassettes purchased and validated – relative inaccuracy ~2% and CVs ~0.7%

Wet Lab Protocols:

# IF Staining Protocol

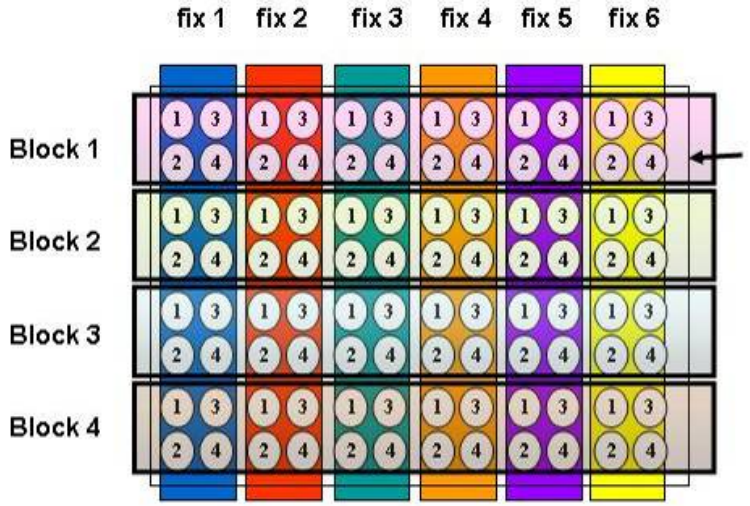
Facilitate a systematic comparison of a large number of assay parameters to maximize data quality while minimizing the time required to find optimal parameters.

- fix
- perm
- block
- primary ab dilution
- secondary ab dilution
- buffers
- wash buffers
- number of washes
- time/temp of incubation….etc

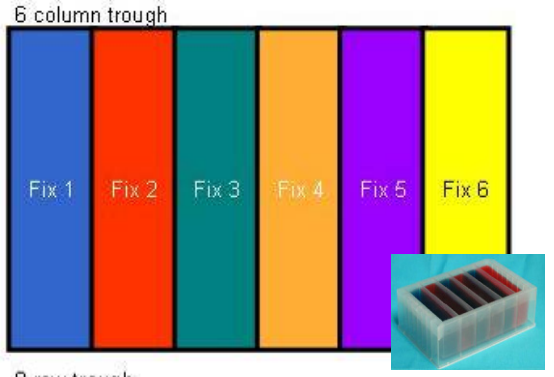
# Antibody Optimization

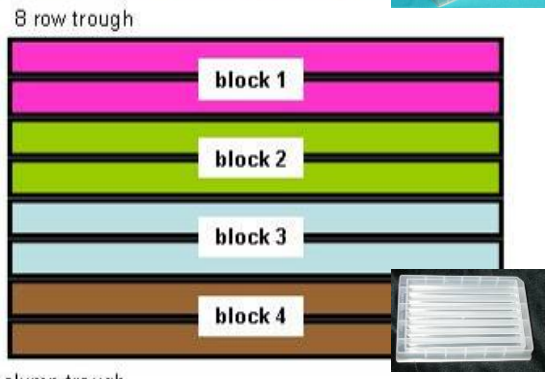
**Matrix of 6 fixatives - 4 blocks - 4 ab dilutions to assay**

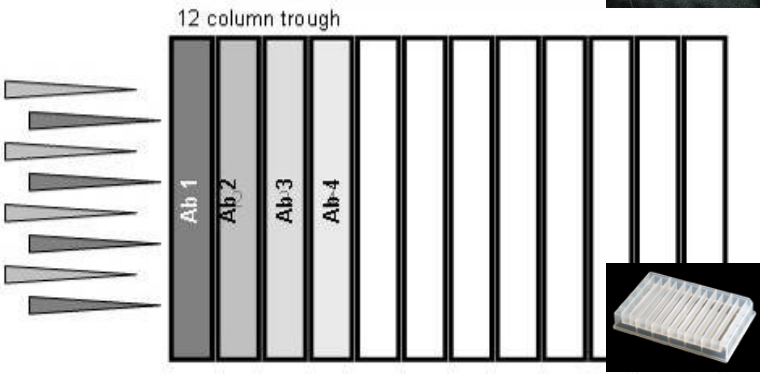
***96 individual conditions***




4 Ab dilutions in quadrants

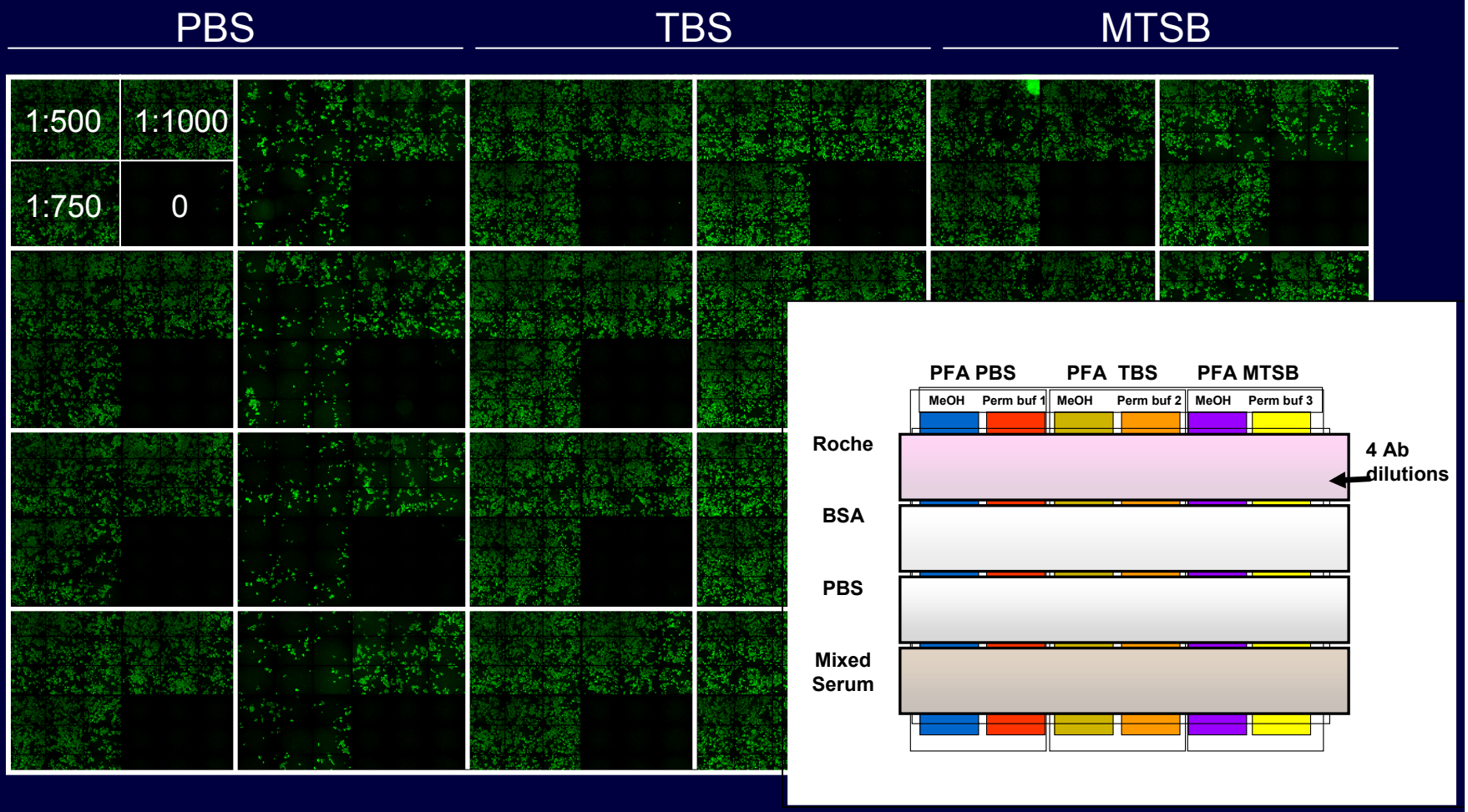


Wet Lab Protocols:

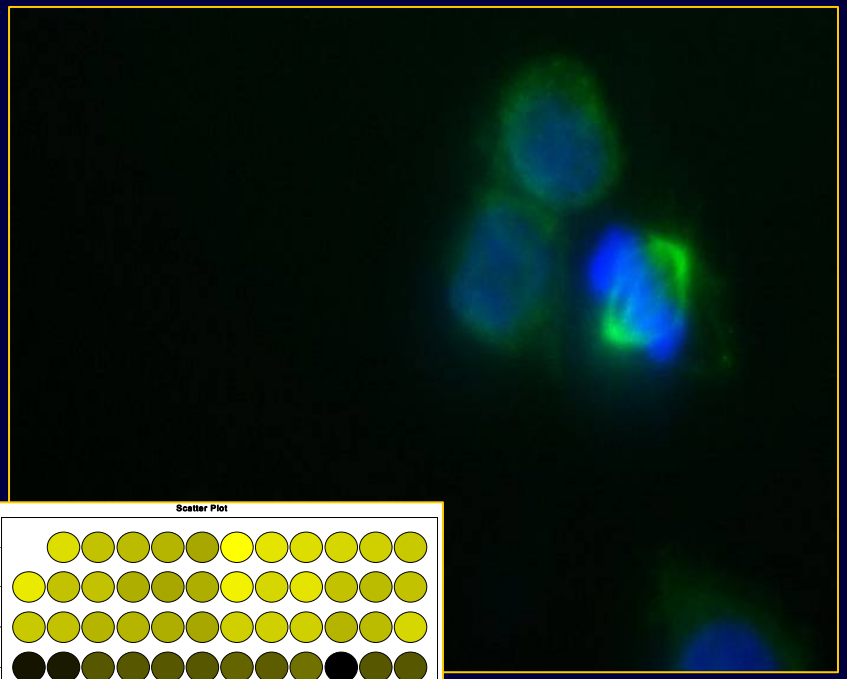
# Antibody Optimization



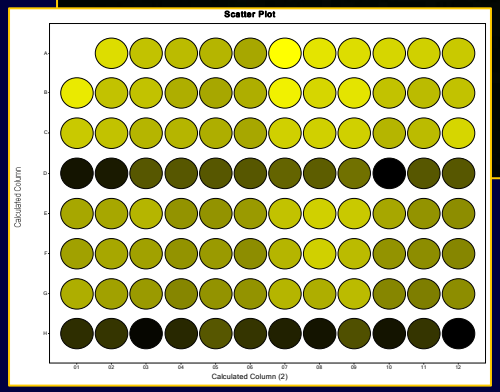
Wet Lab Protocols:

# Antibody Optimization

Before

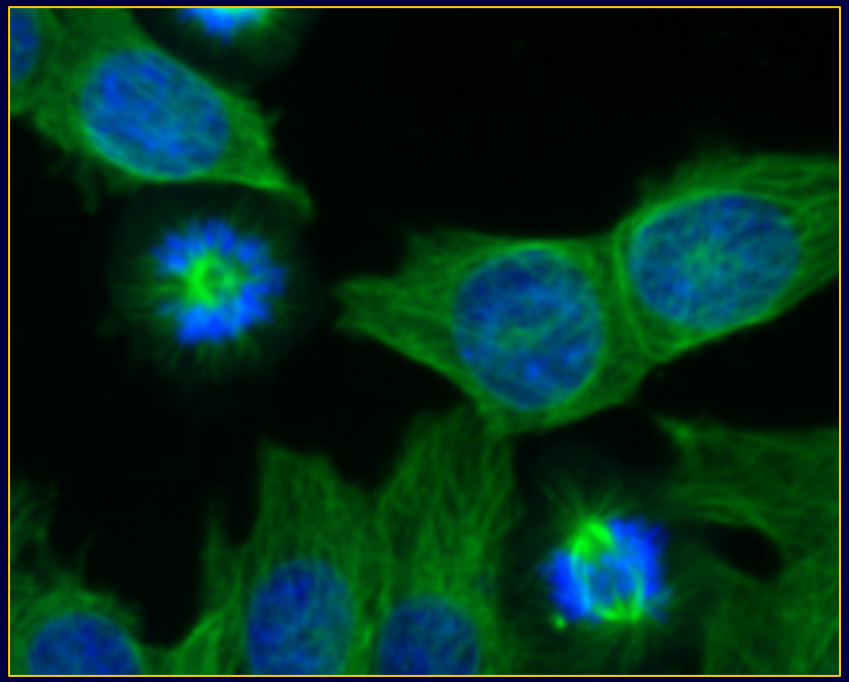


DNA / Tubulin



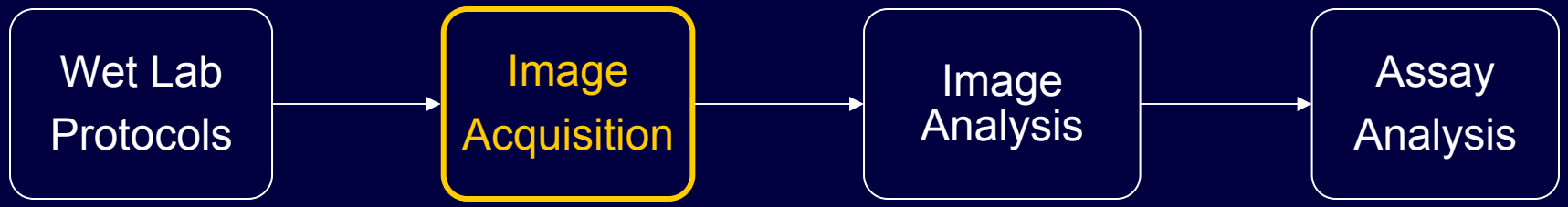

S/N

After



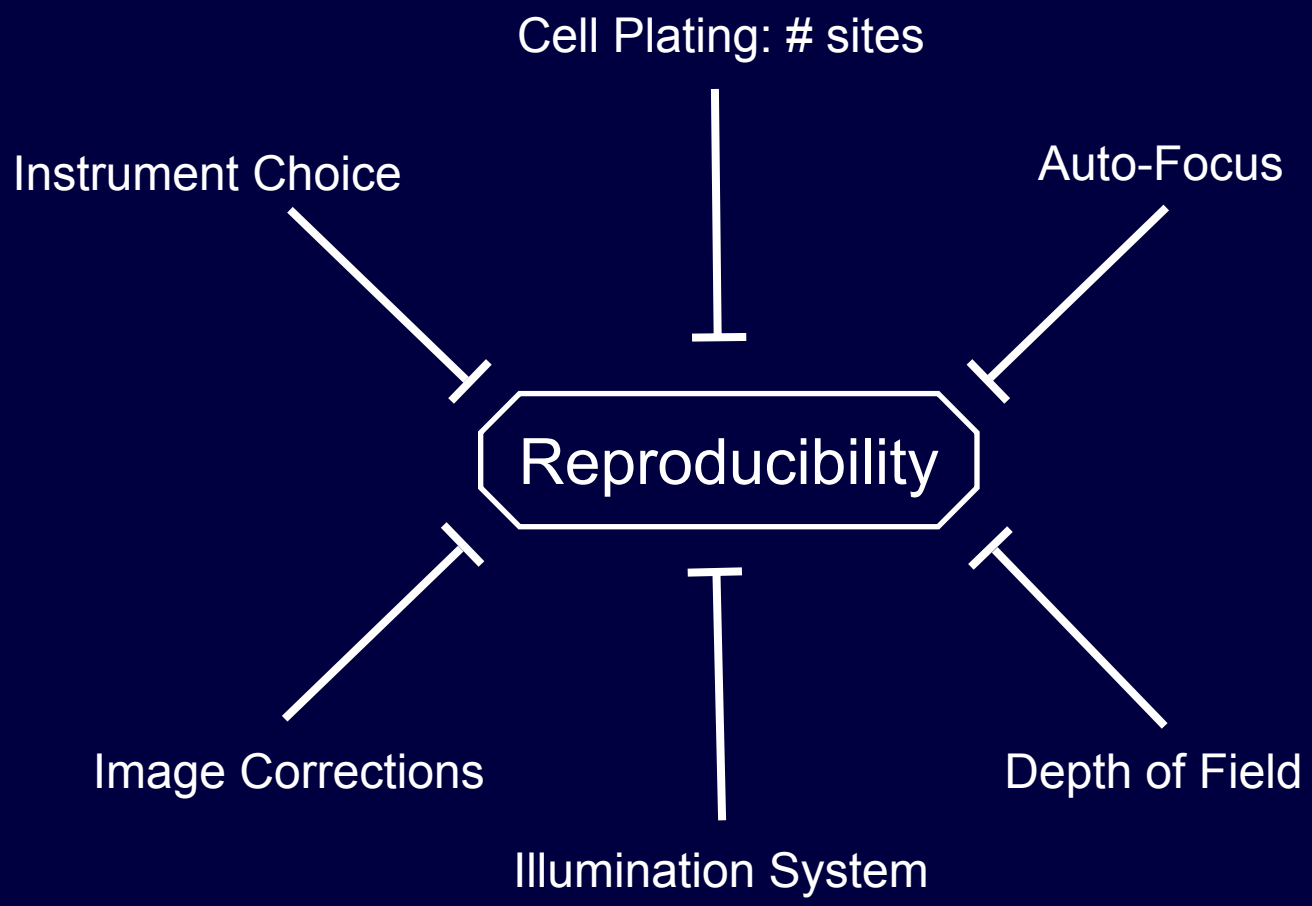
- Comparison of > 400 parameters (~2 weeks)
- Enhances analysis of spindle morphology
- Save $ Antibody concentration

# HCS Assay Components




- Dissect components
- Using control experiments
  - Identify drivers of variability
  - Minimize them experimentally or account for them statistically

# Image Acquisition



# Image Acquisition

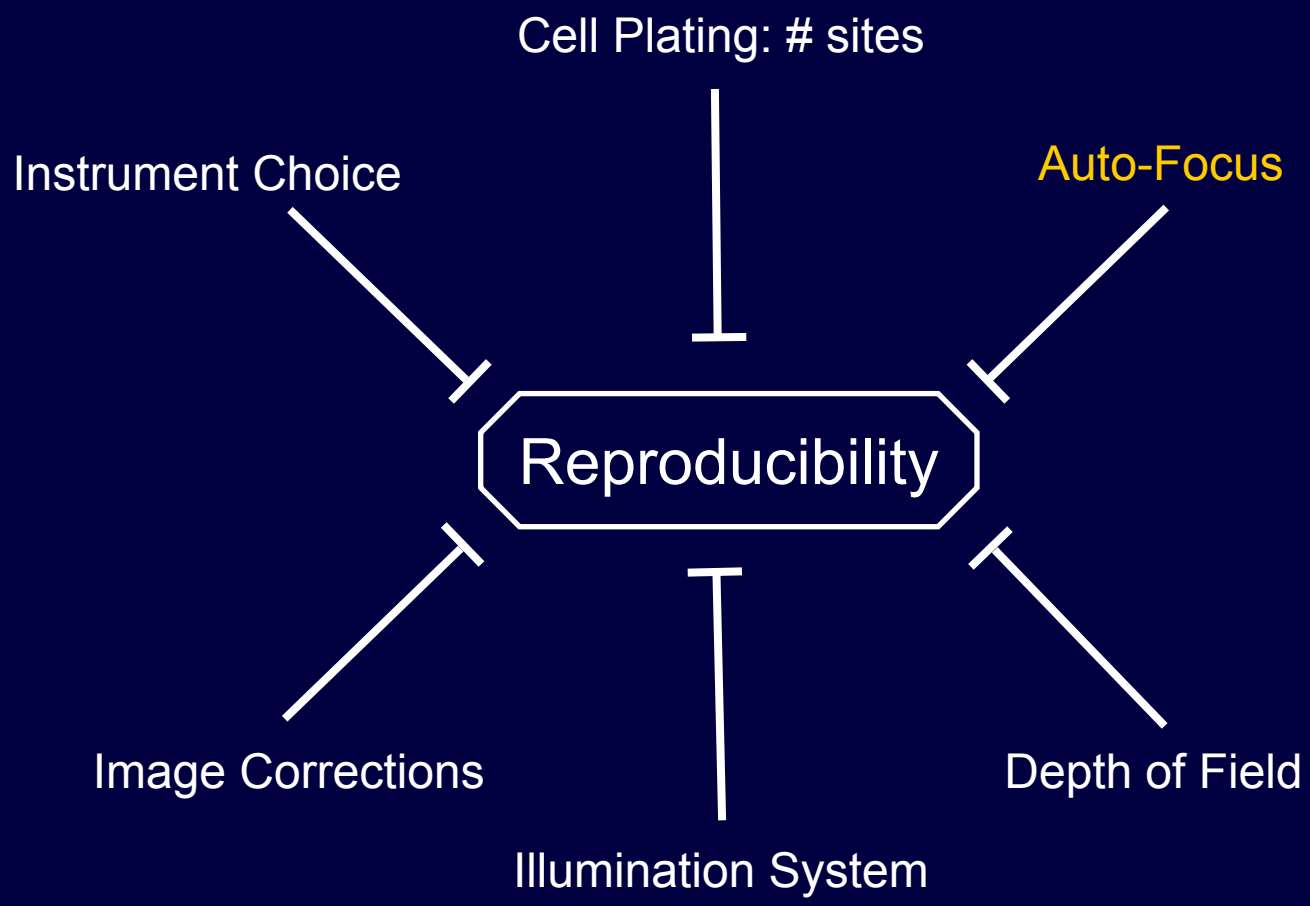
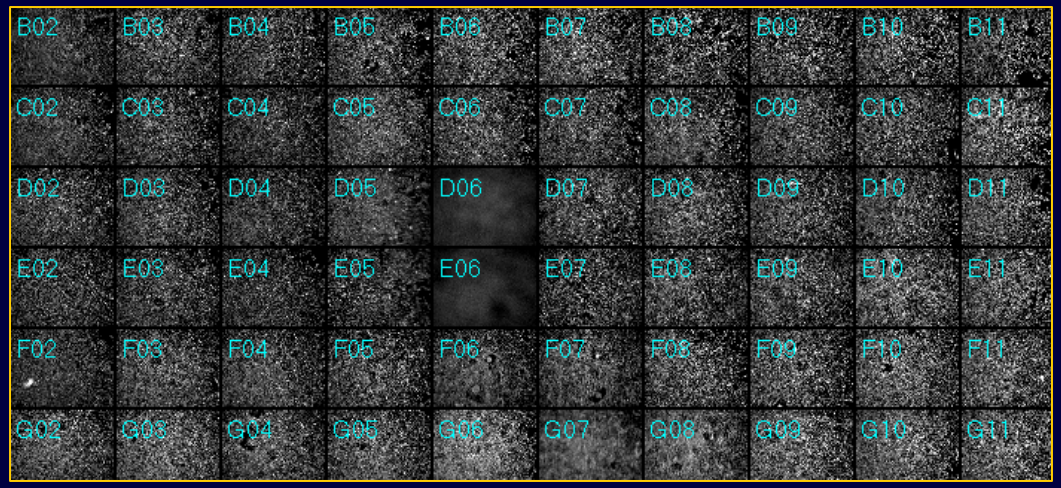


Image Acquisition:

# AutoFocus

- Image-based and Laser-based



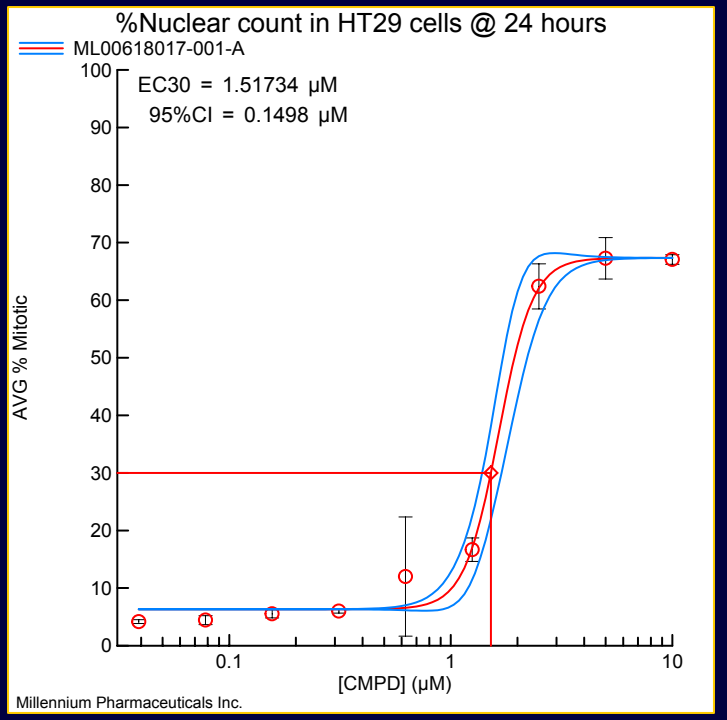
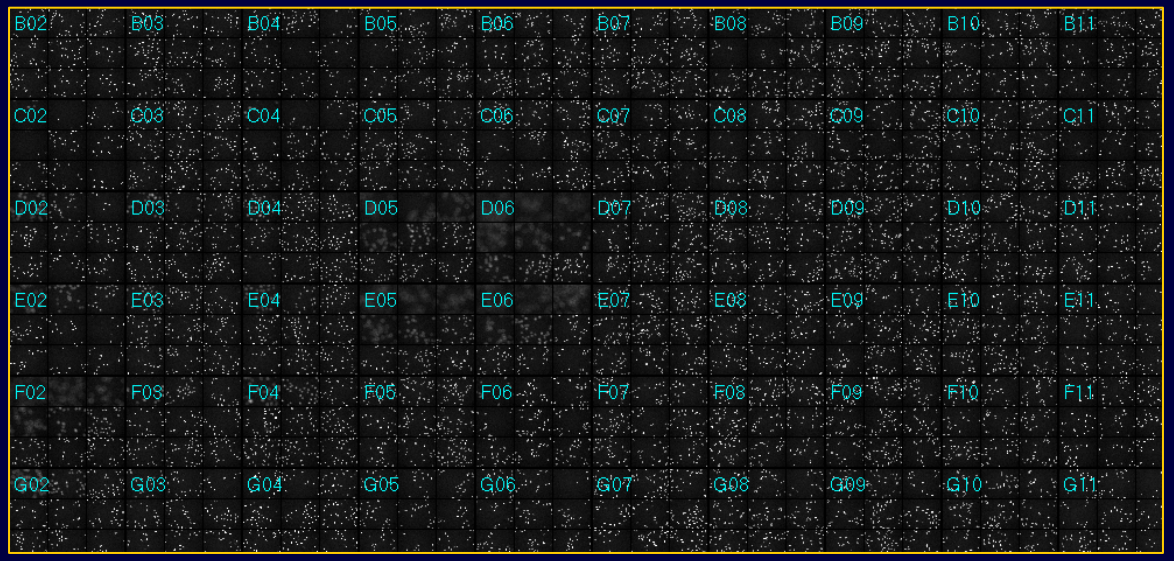
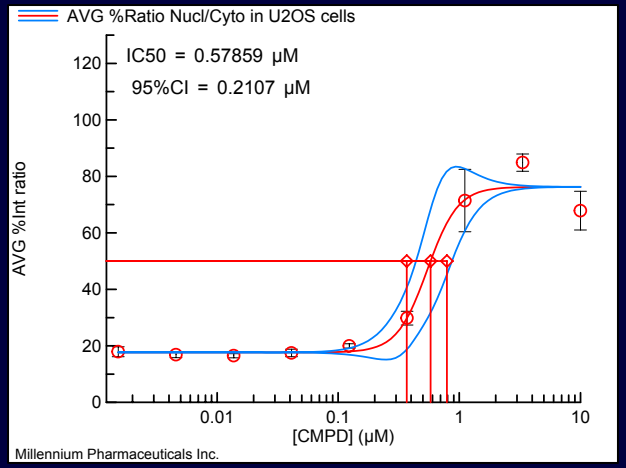



Image Acquisition:

# AutoFocus




60 wells, 9 sites / well

Image Acquisition:

# AutoFocus

- Minimize experimentally
  - Instrument parameters
  - Plate selection
- Identify, Remove: Auto-focus Classifier
  - Classify several hundred images by hand
  - Use image processing measures to extract quantities that vary based on edges and intensity
  - Train a classifier to call out-of-focus wells

Image Acquisition:

# AutoFocus

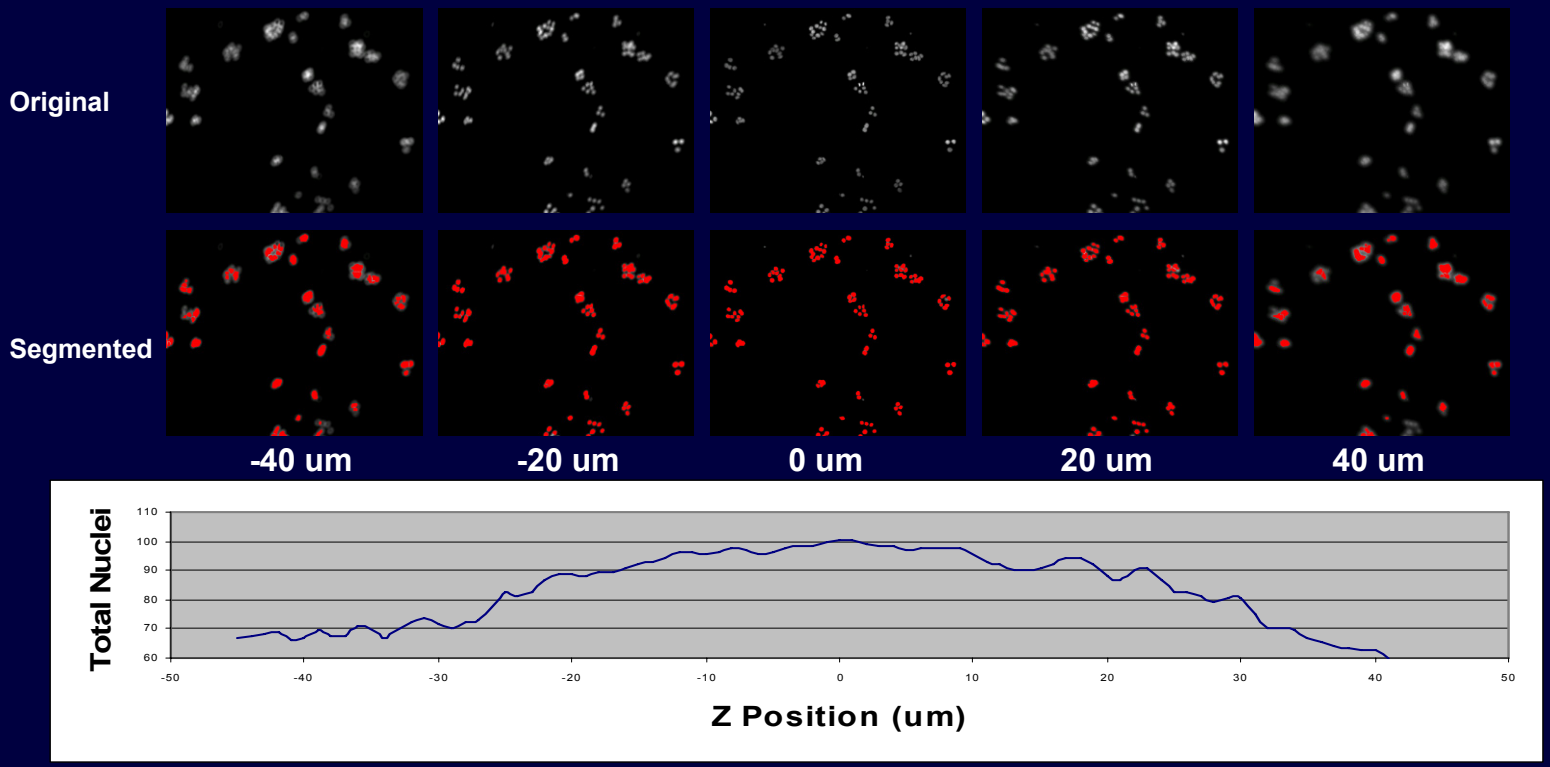


Image Acquisition:

# AutoFocus

- BayesNet classifier learns differences between in-focus and out-of-focus images
- Algorithm maximizes separation of classes on training set by:
  - automatically assigning weights for each predictor
  - automatically selecting threshold

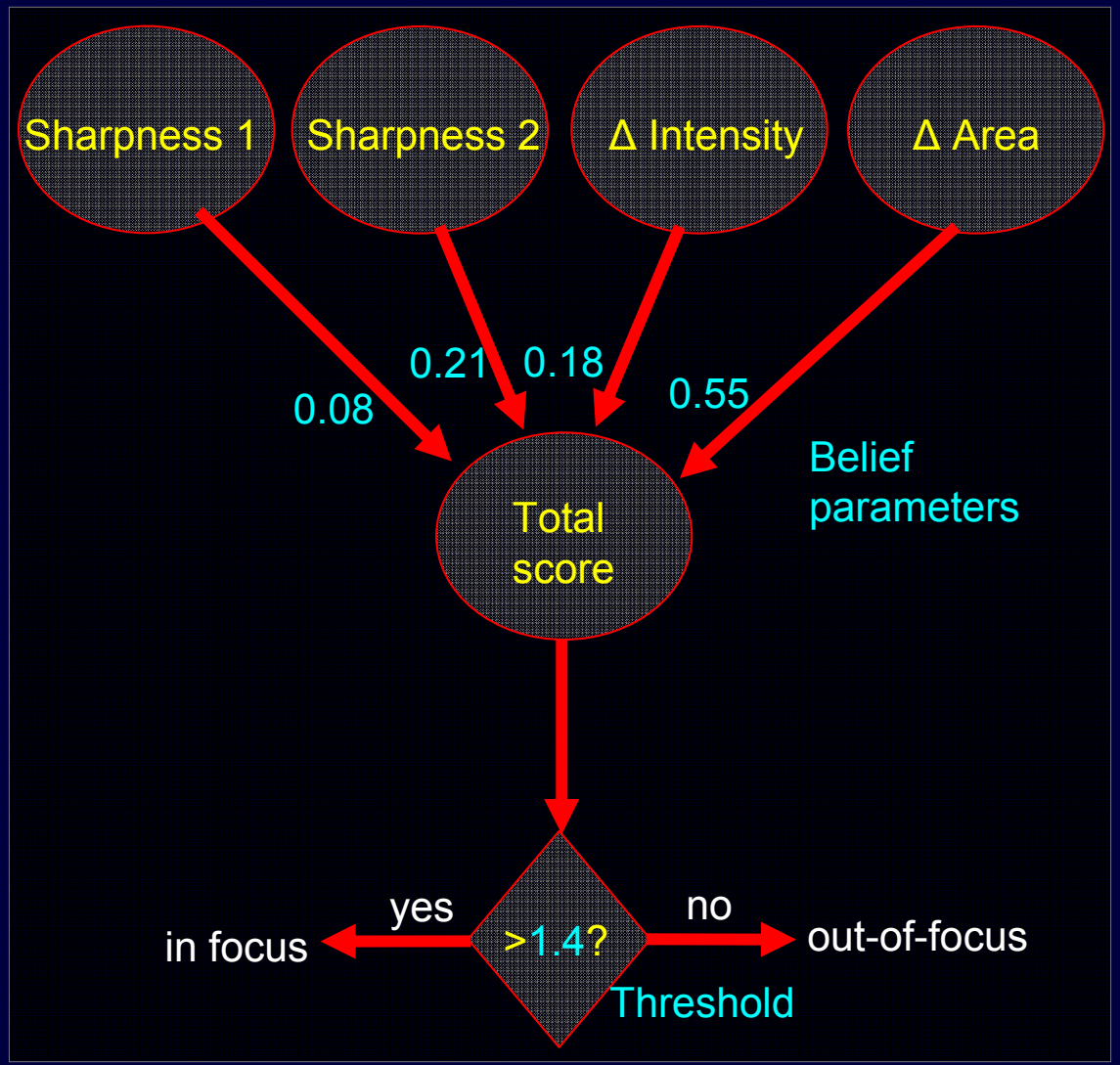


Image Acquisition:

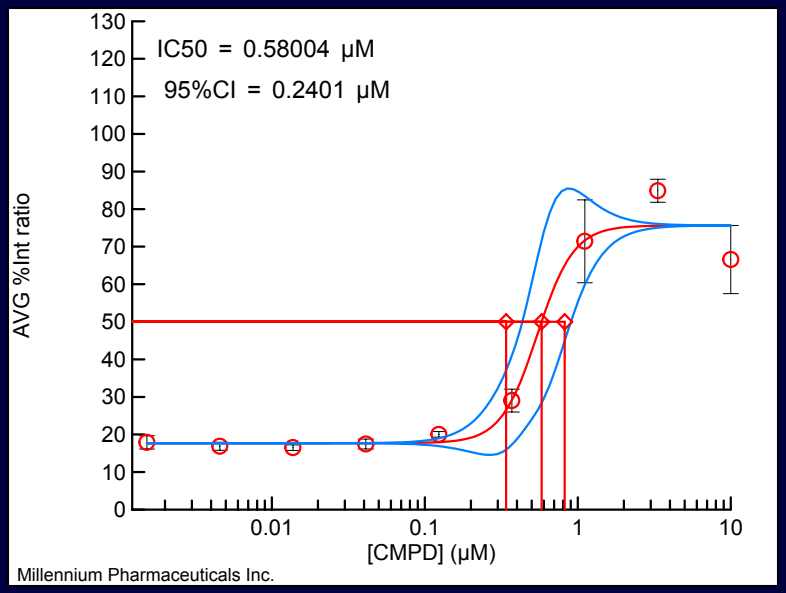
# AutoFocus

- Results of BayesNet assessed using cross-validation:
  - Algorithm is trained on 9/10$^{th}$s of the data and tested on remaining 1/10$^{th}$
  - Process is repeated 10 times
  - Error rates are averaged
- 93% to 98% of the wells were called correctly (for 3 different cell types).
- Nearly all classification errors occurred in 'marginal' category
- Zero-R (baseline classification error)= 32%

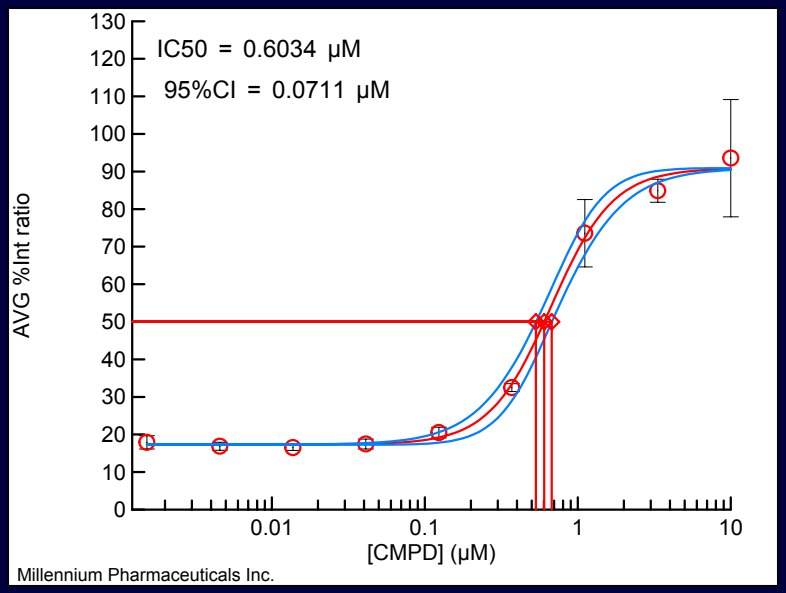
Image Acquisition:

# AutoFocus

## Filtering out-of-focus images via machine learning




Original Data




Filtered Data

# Image Acquisition

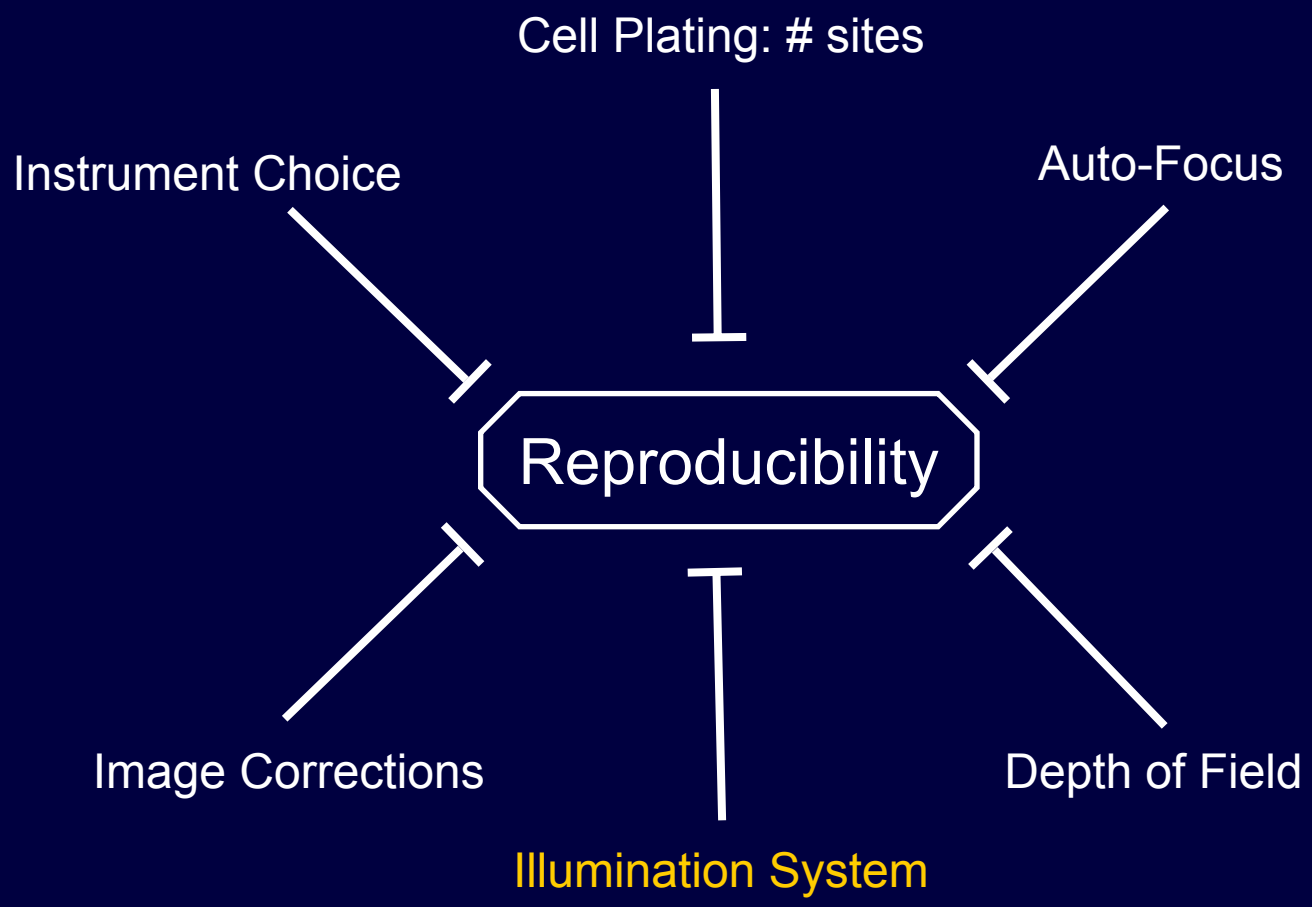
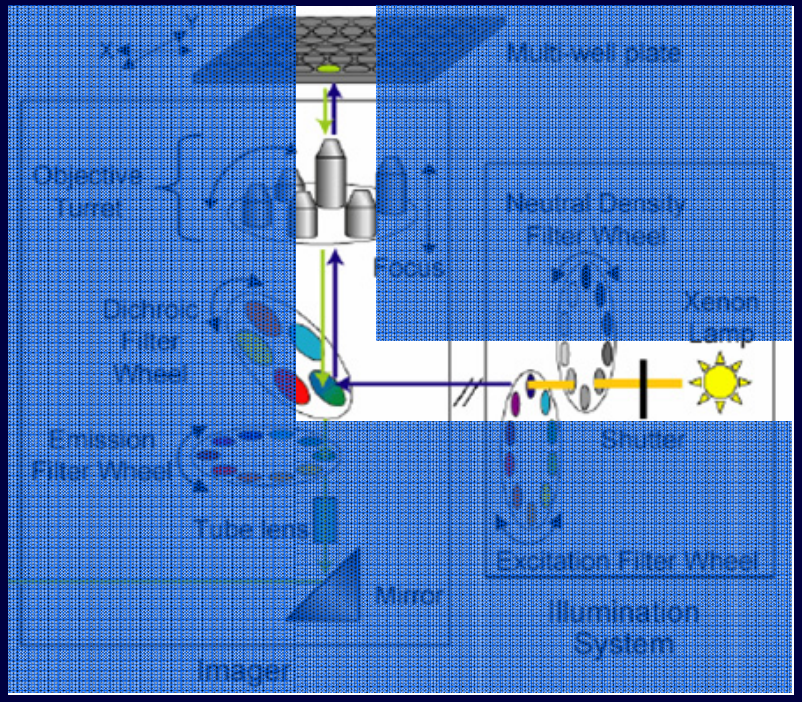
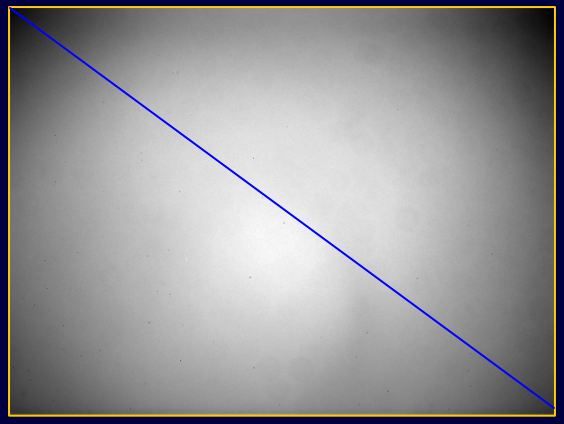


Image Acquisition:

# Illumination system variation

- What is it?
  - Fluorescence illumination system



Uniform sample

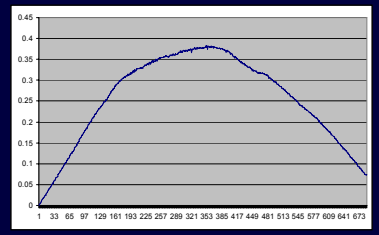
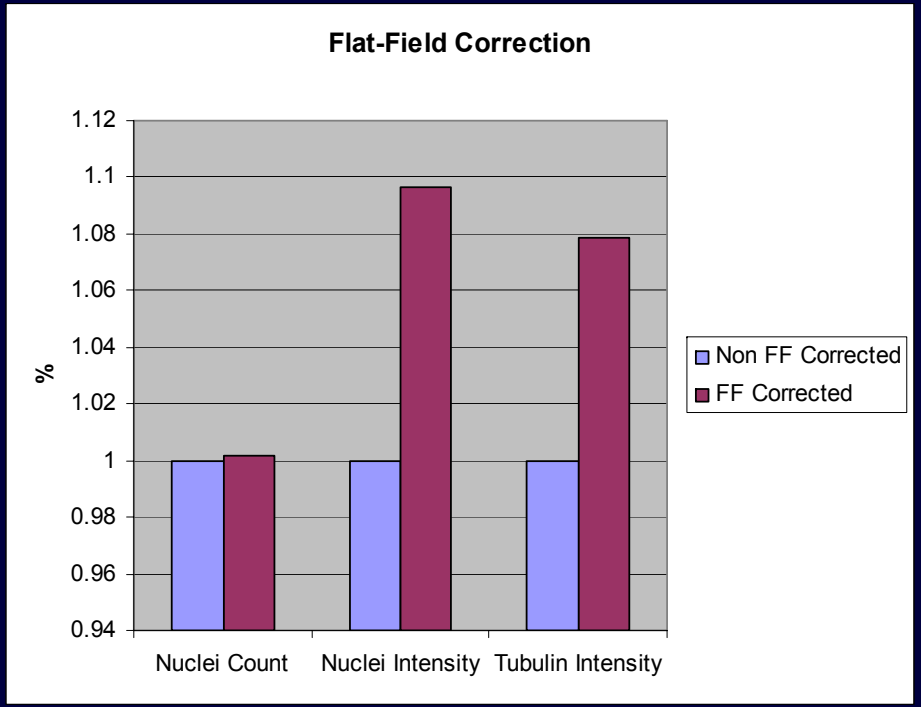


Image Acquisition:

# Illumination system variation

- Is it a problem?
  - ↓ Cell counting assays
  - ↑↑ Cell intensity assays
  - ↑↑ Phenotyping



# Image Acquisition

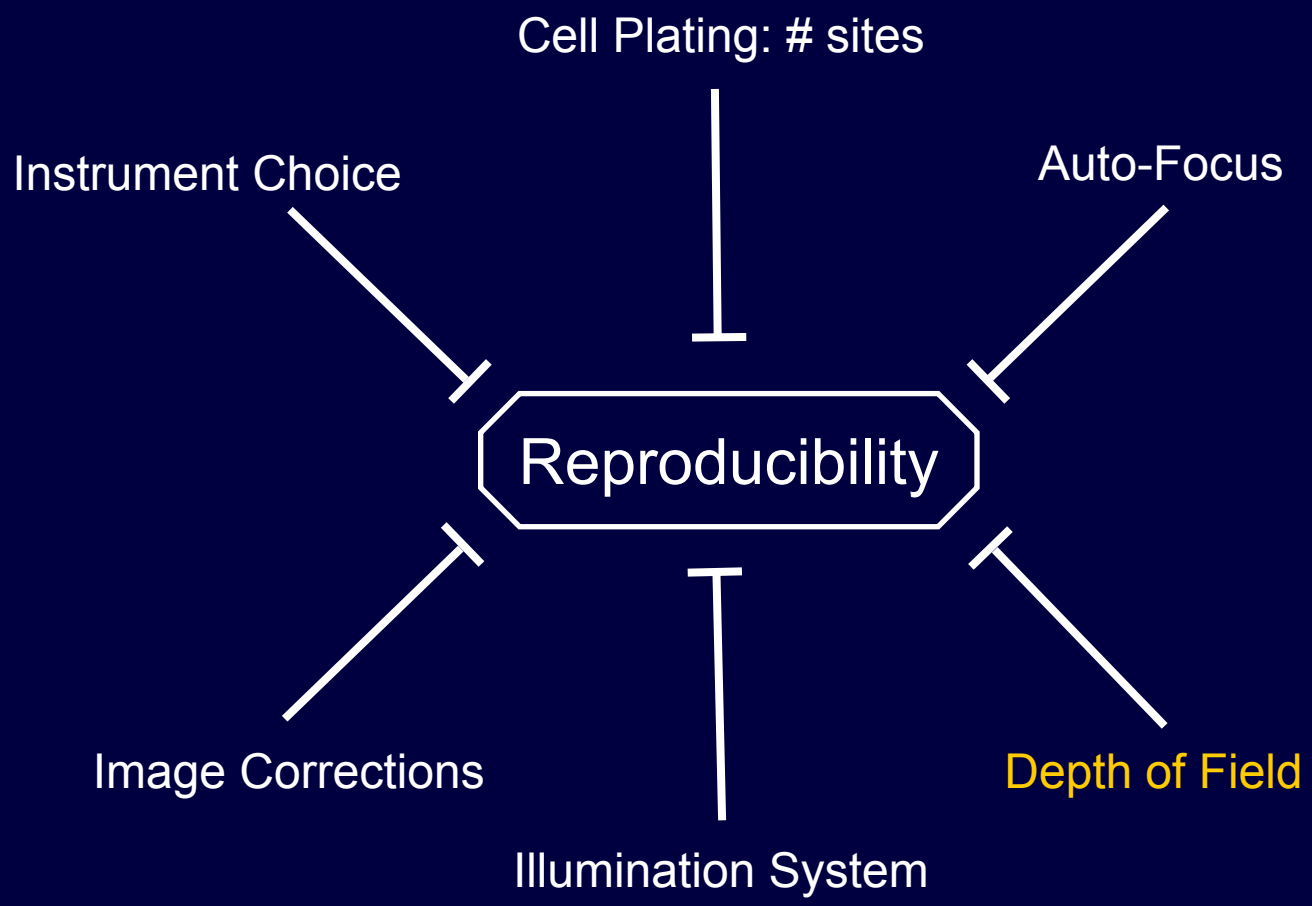
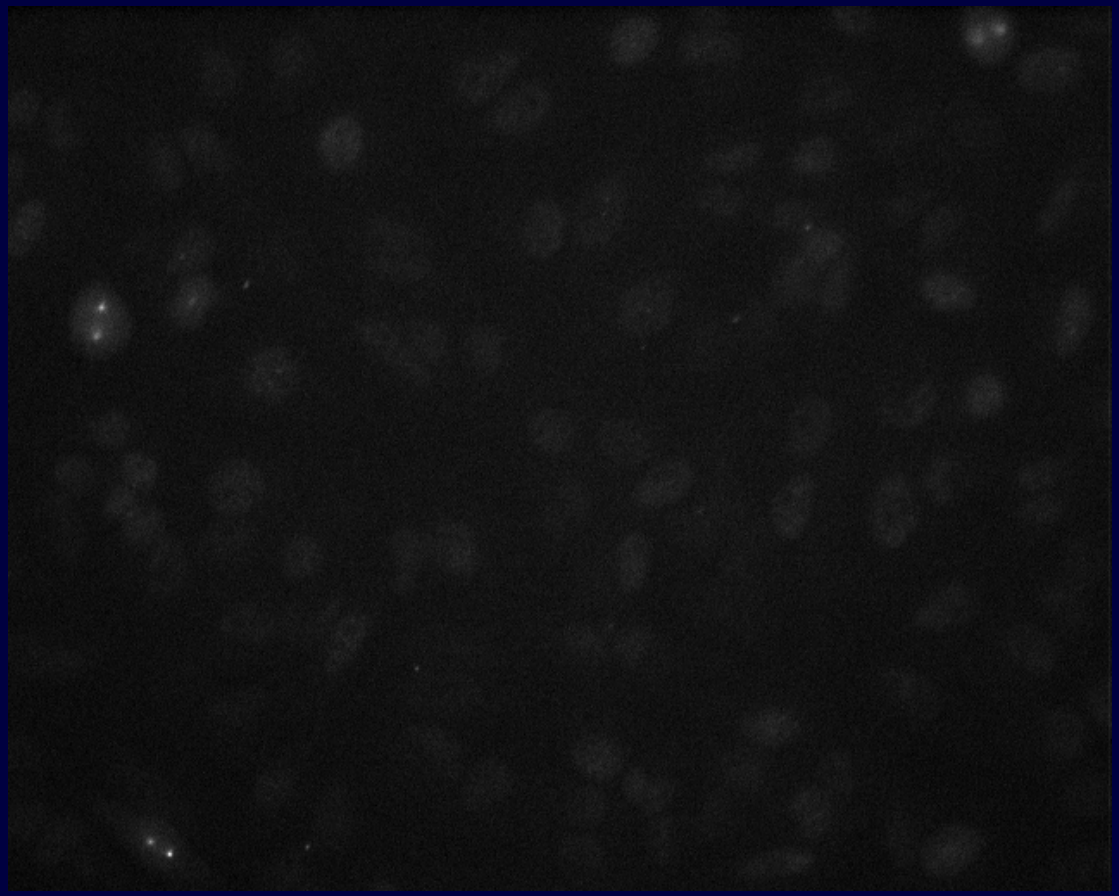


Image Acquisition:

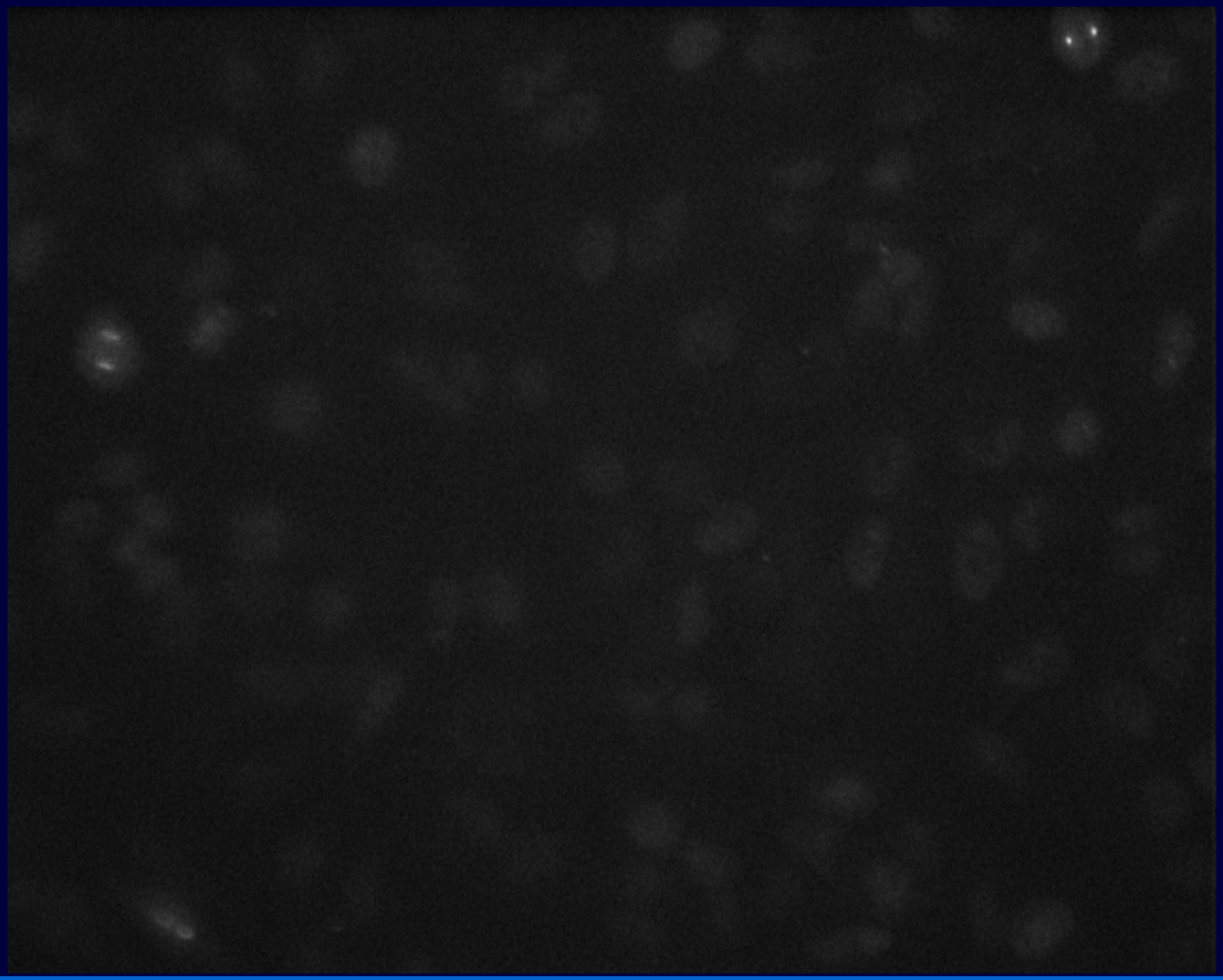
# Depth of Field

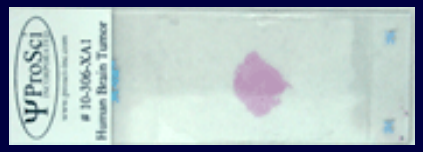


20x

Image Acquisition:

# Depth of Field

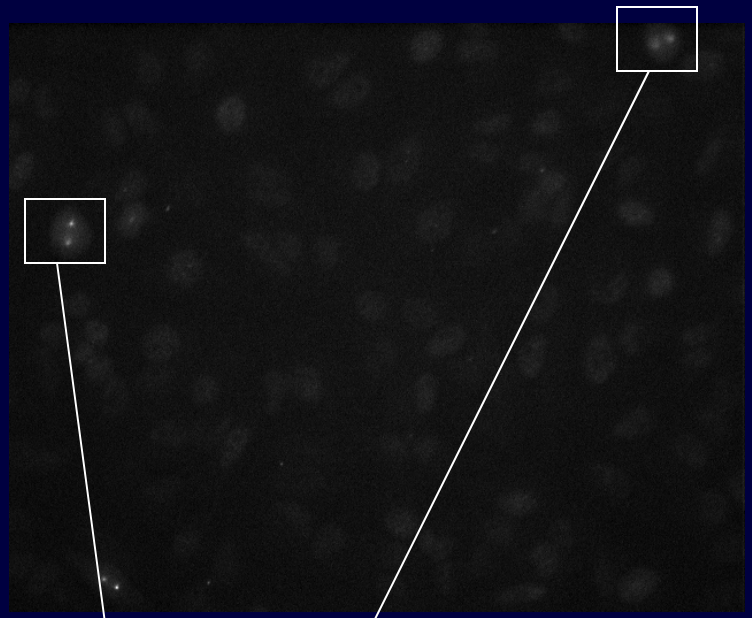


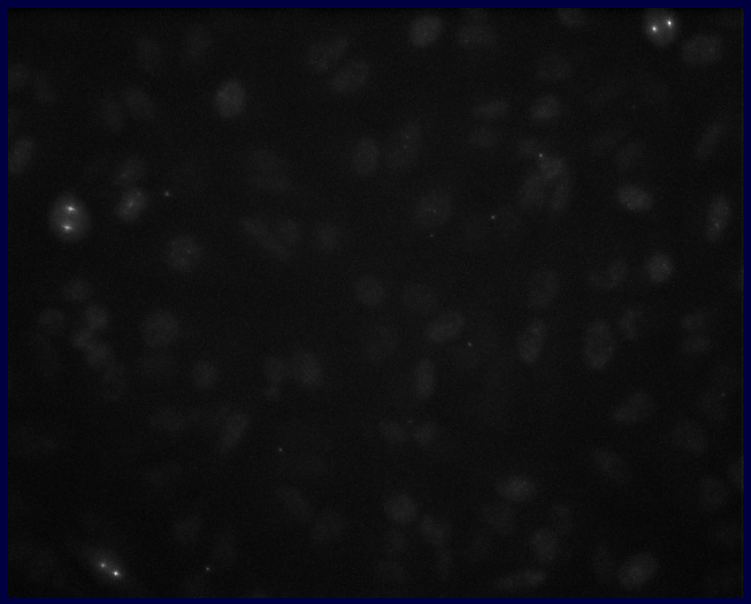

**- 4.0 um**

Image Acquisition:

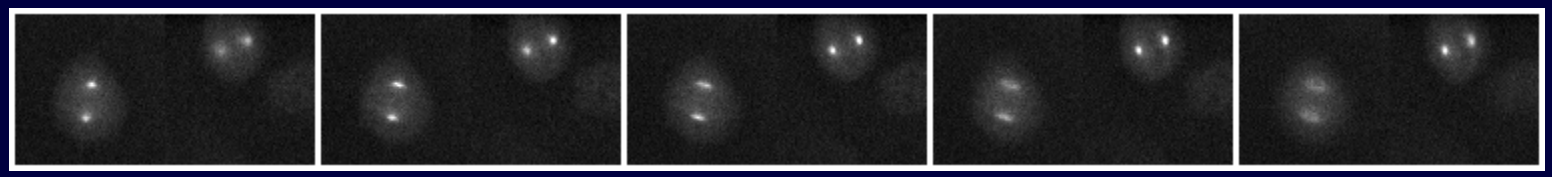
# Depth of Field



Single Plane



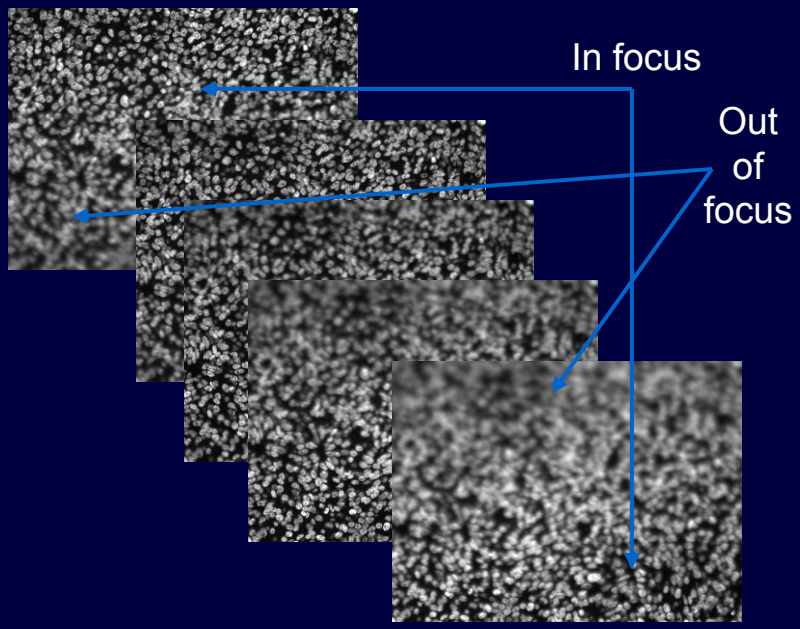
“Best Focus”



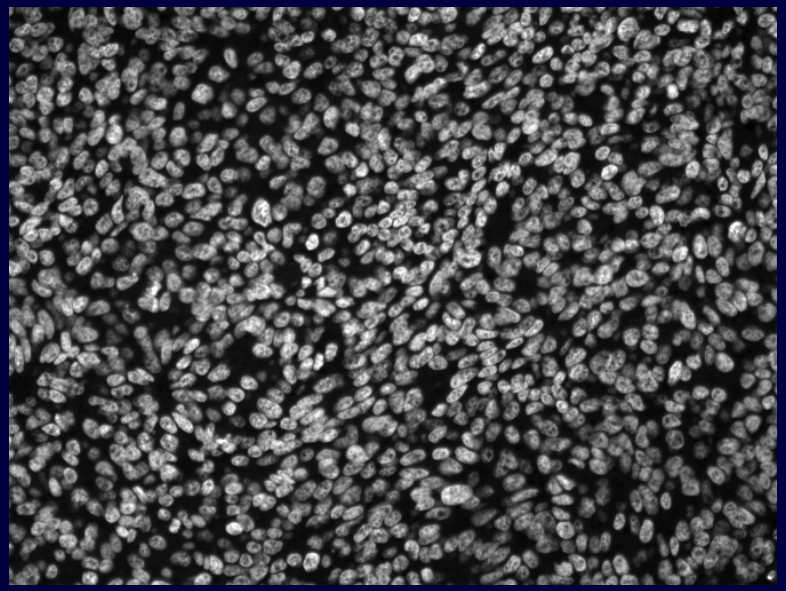
Z = -4.0um  Z = -2.0um  Z = 0.0um  Z = 2.0um  Z = 4.0um

# Tissue Imaging

- Challenge: depth of focus of the objective is less than the flatness of the specimen. This results in portions of a single image plane being out of focus.

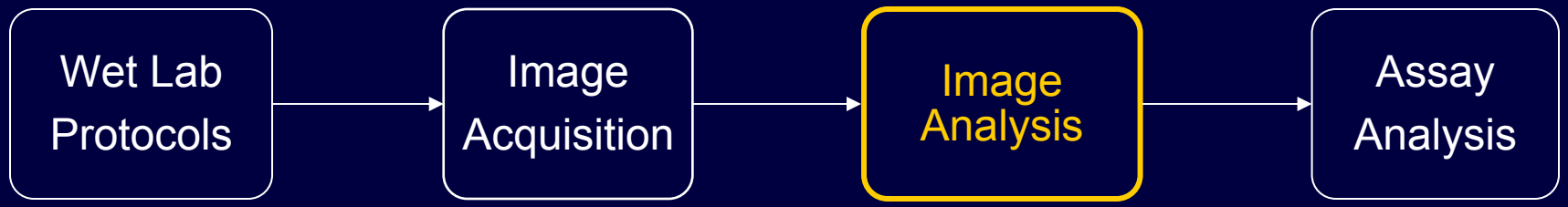



Multiple optical sections @ 2.0um intervals



"Best Focus" Image

# HCS Assay Components




- Dissect components
- Using control experiments
  - Identify drivers of variability
  - Minimize them experimentally or account for them statistically

# Image Analysis

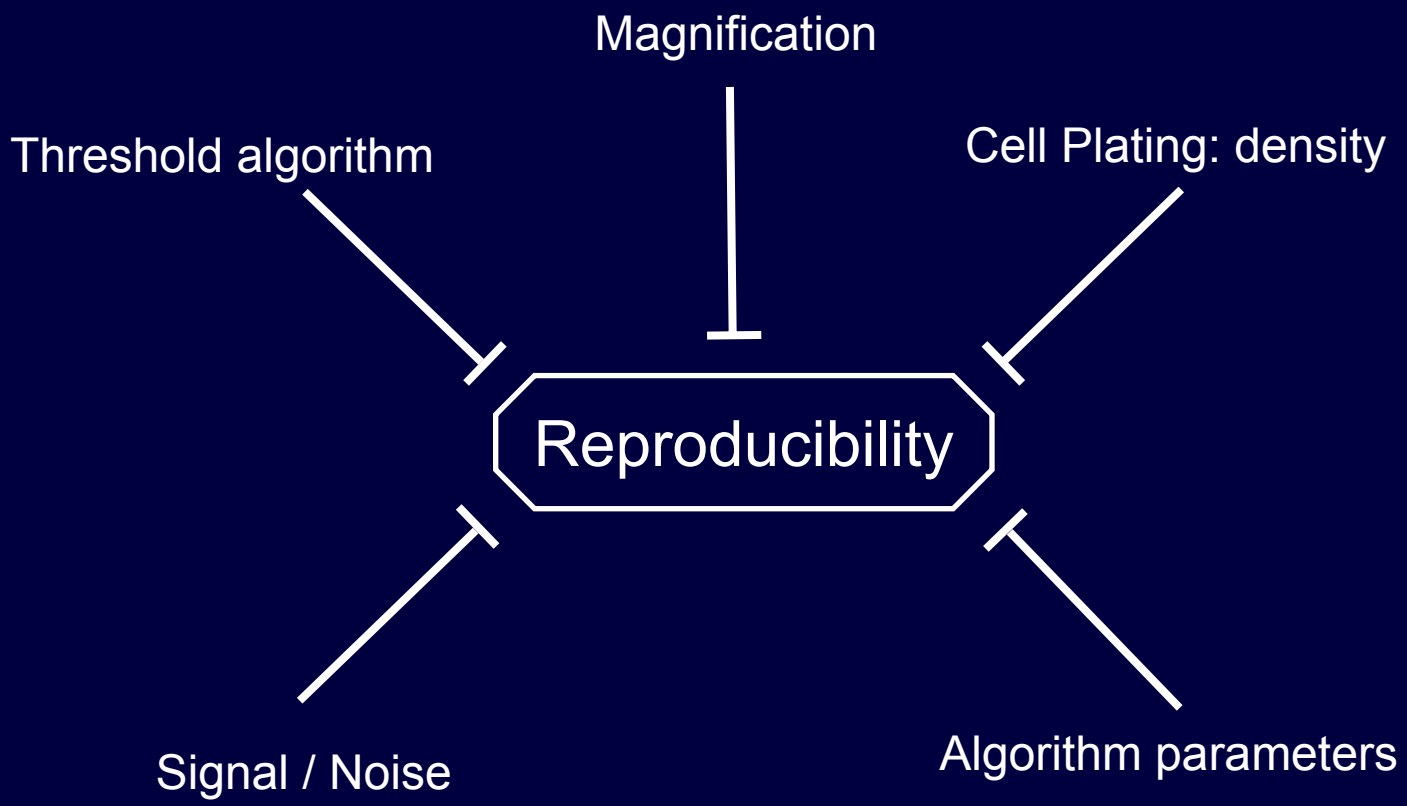


Image Analysis:

# Cell Plating Density

- Assess
  - Optimal cell plating density to minimize segmentation errors
- DOE:
  - Serial dilution
  - DAPI
  - 6 replicates

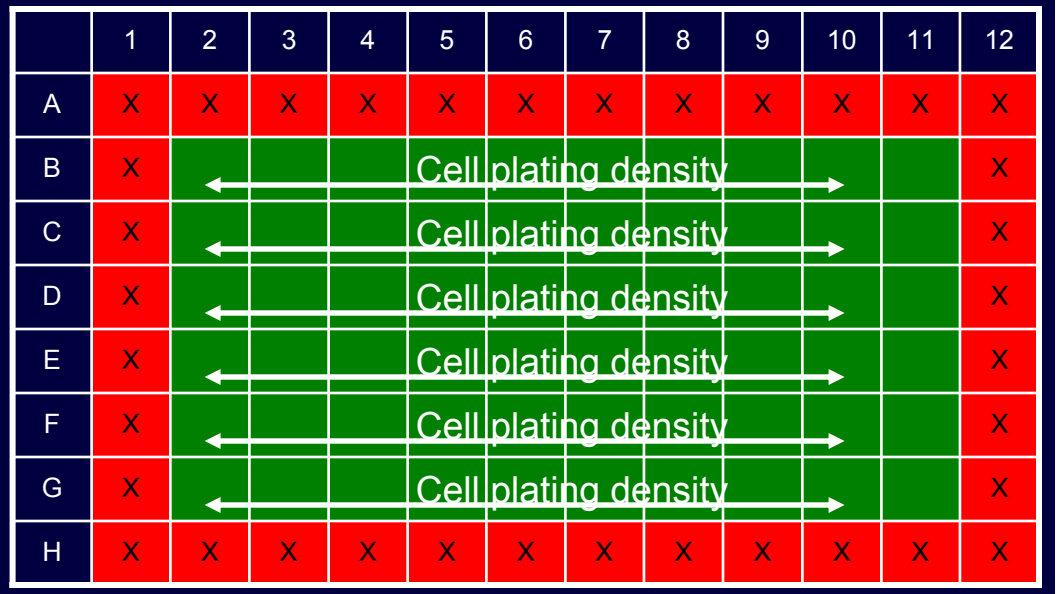
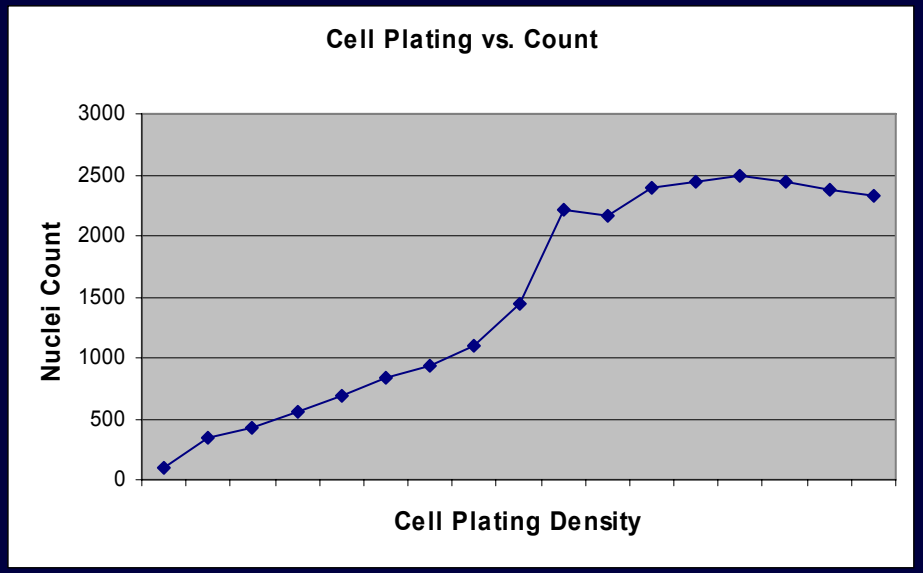
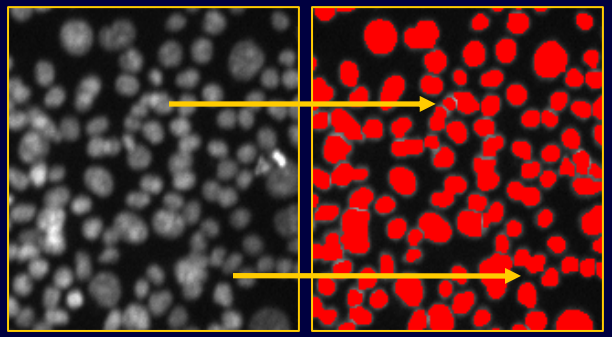


Image Analysis:

# Cell Plating Density



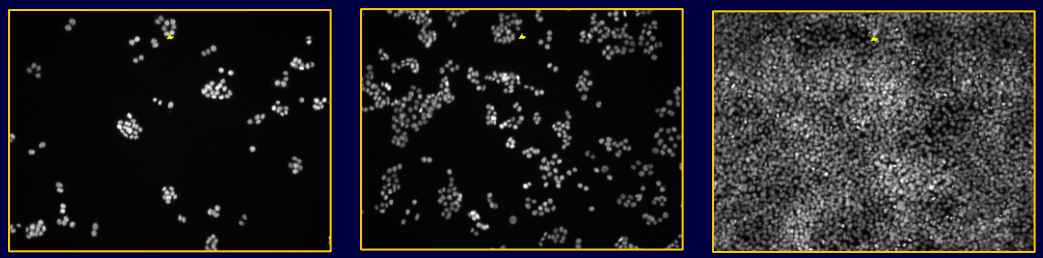
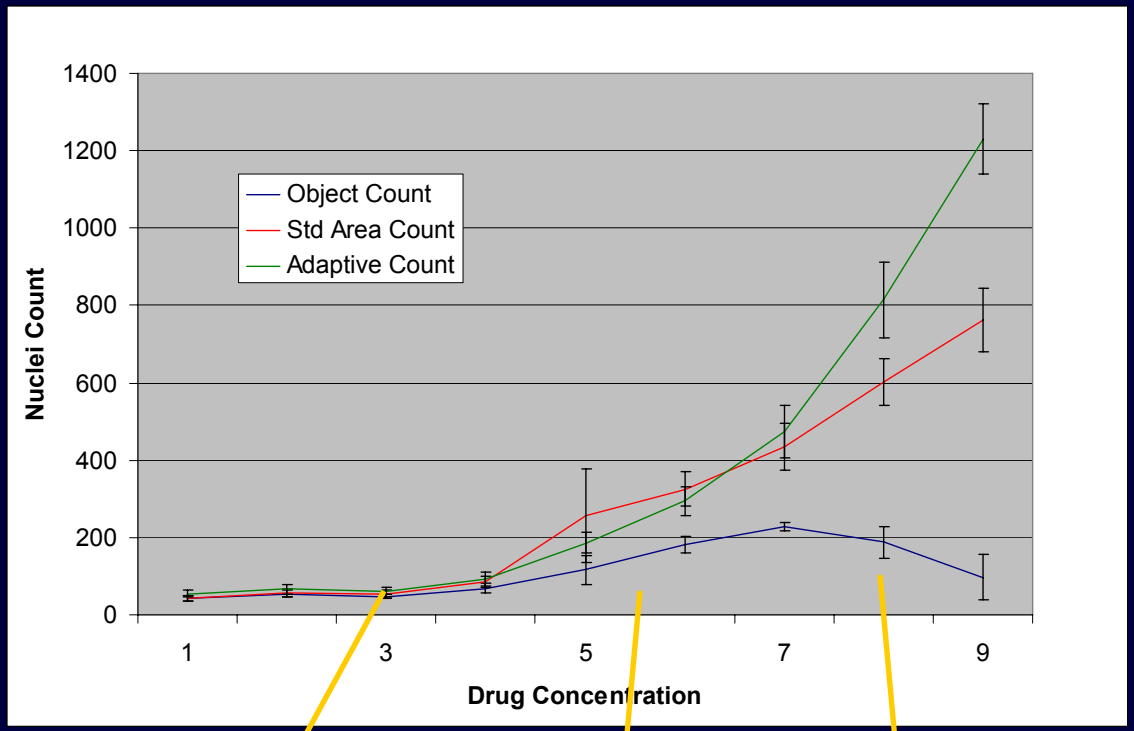
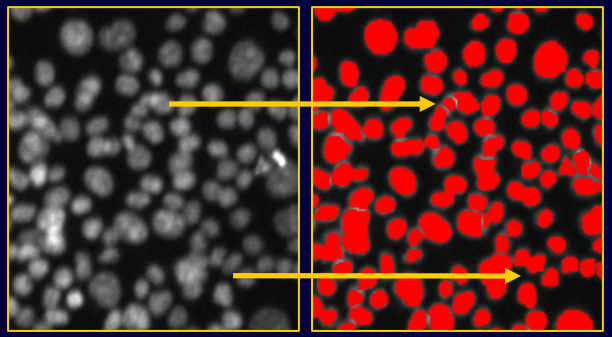


Image Analysis:

# Threshold algorithm



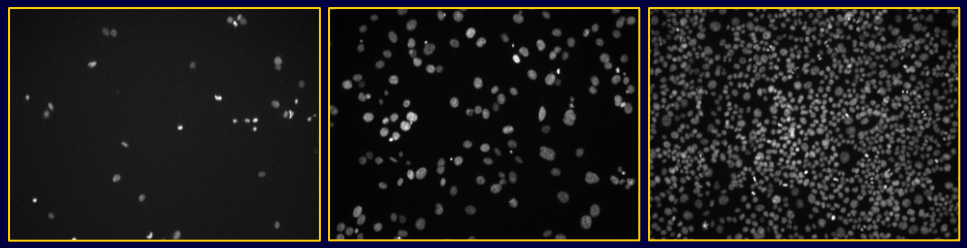


Image Analysis:

# Signal / Noise

- Assess
  - Optimal exposure times to minimize acquisition time and maximize Signal/Noise
    - Image-based autofocus
    - Intensity measurements
    - Segmentation
  - DOE:
  - Uniformity Plate
  - DAPI
  - Multiple exposures

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | X | X | X | X | X | X | X | X | X | X | X | X |
| B | X | | | | | | | | | | | X |
| C | X | | | | | | | | | | | X |
| D | X | | | | | | | | | | | X |
| E | X | | | | | | | | | | | X |
| F | X | | | | | | | | | | | X |
| G | X | | | | | | | | | | | X |
| H | X | X | X | X | X | X | X | X | X | X | X | X |

Image Analysis:

# Signal / Noise

- Single Uniformity Plate, 9 sites/well

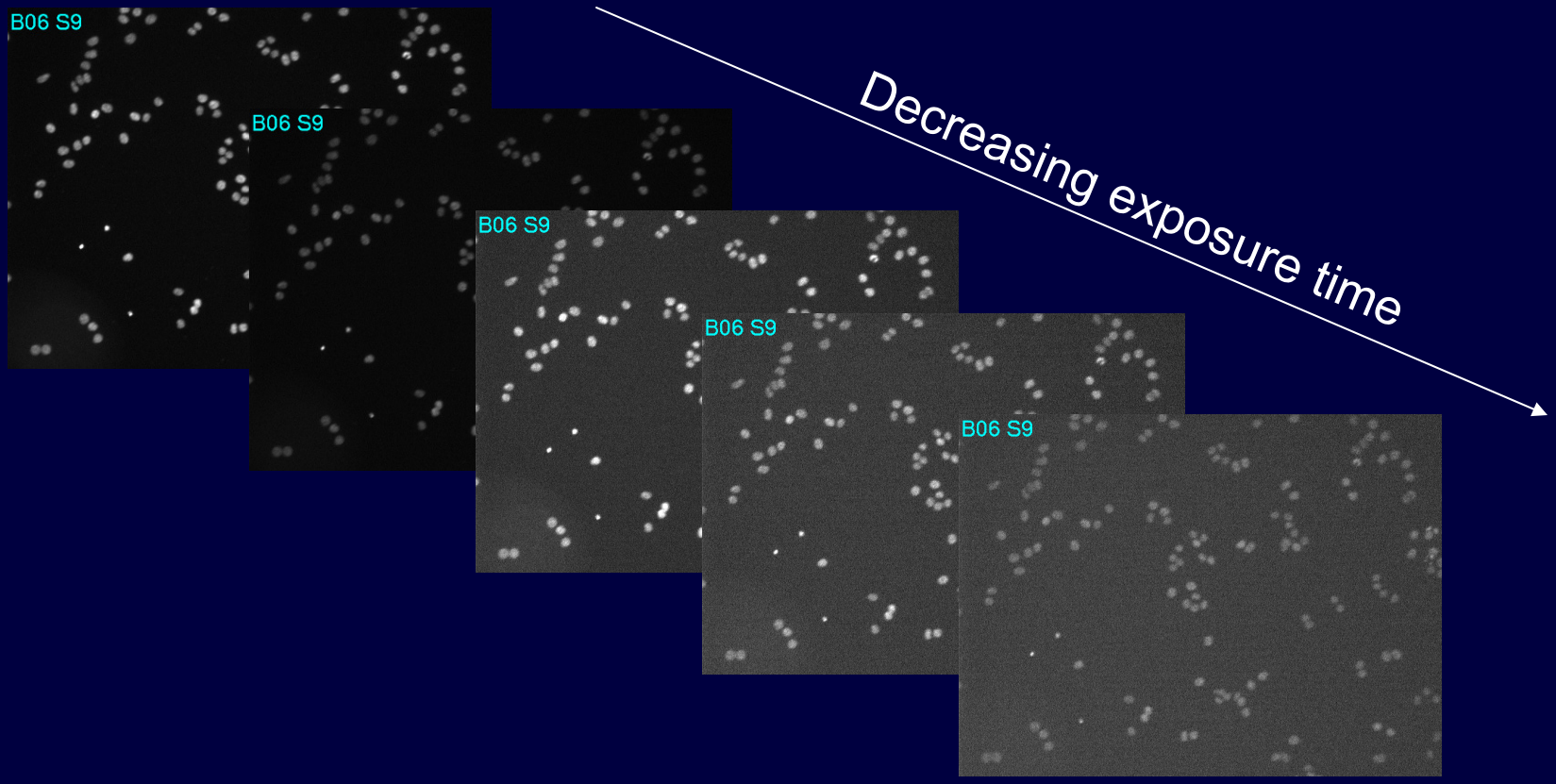
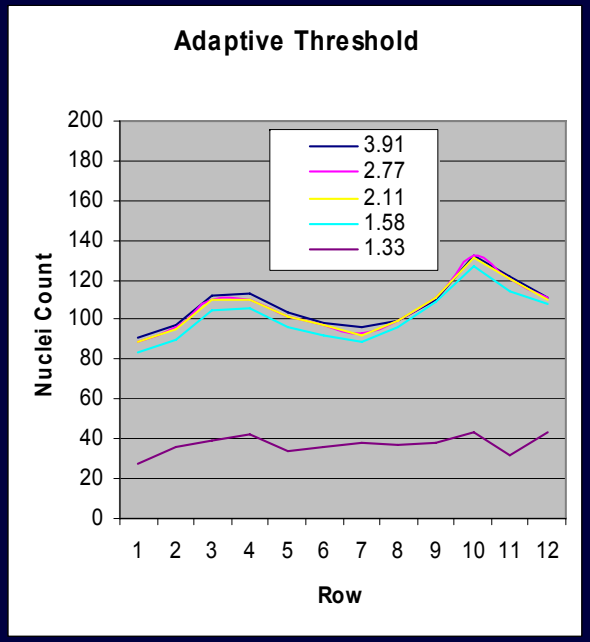
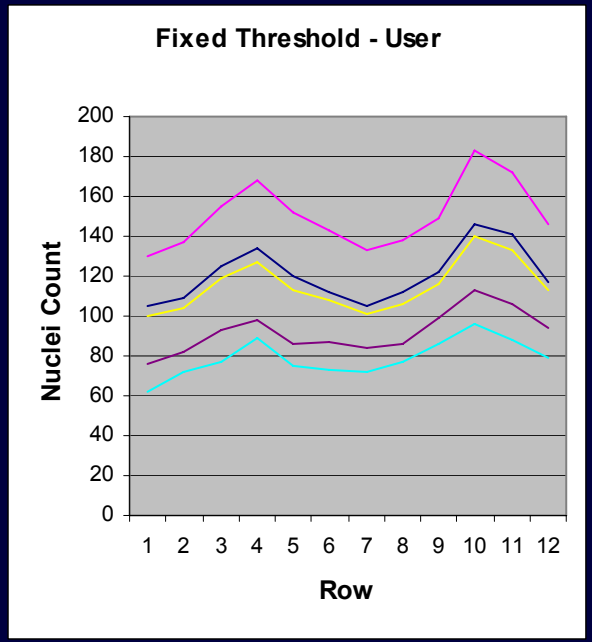


Image Analysis:

# Signal / Noise

- Effect on thresholding



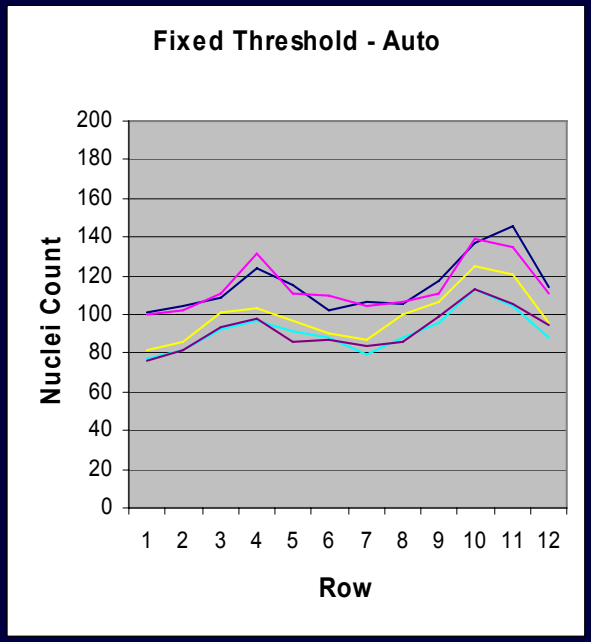
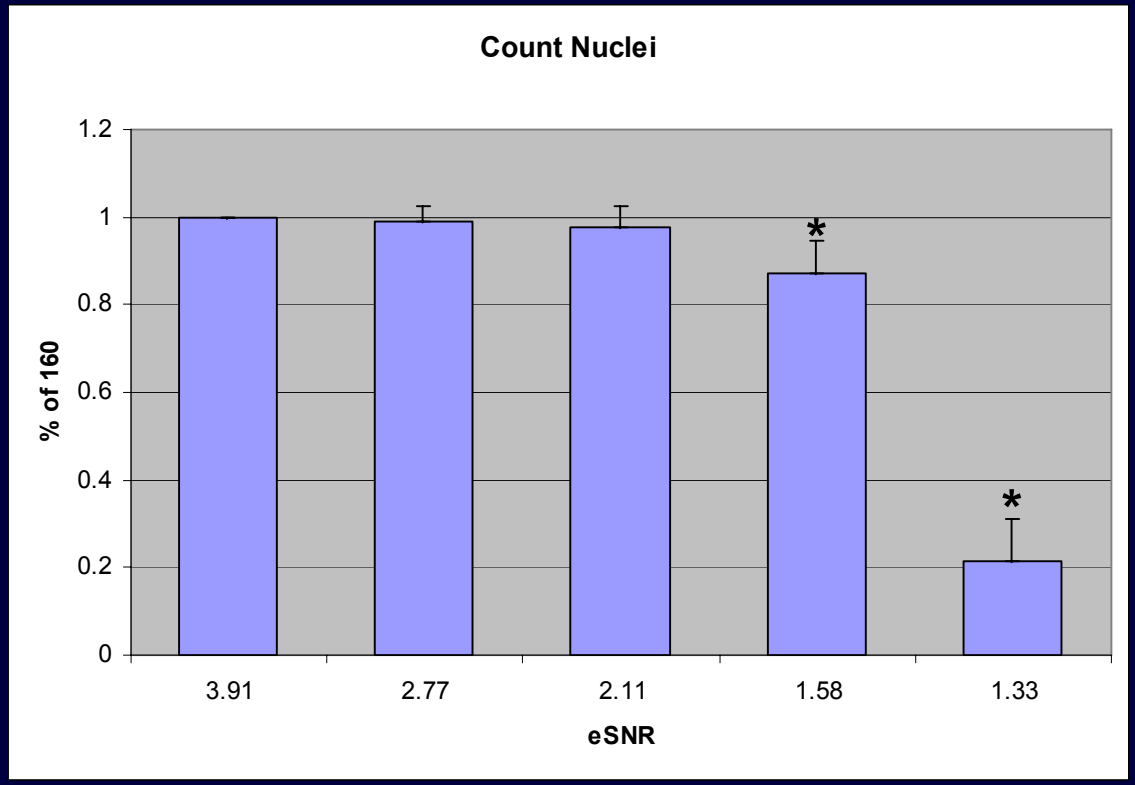



Image Analysis:

# Signal / Noise




*: $p<10^{-100}$,
Welch's two-tailed t-test
Error bars are standard deviations

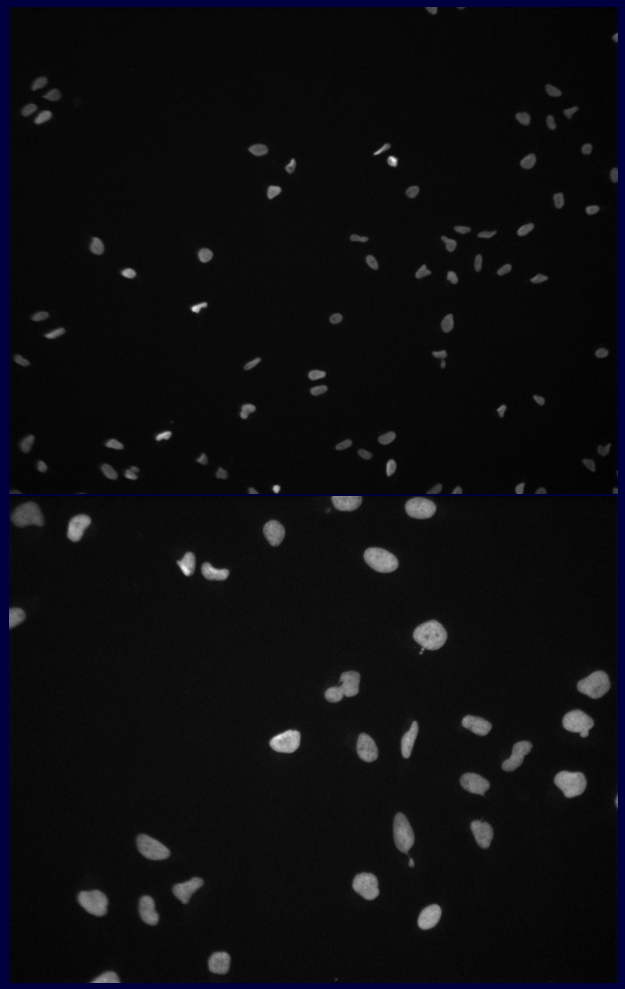
Image Analysis:

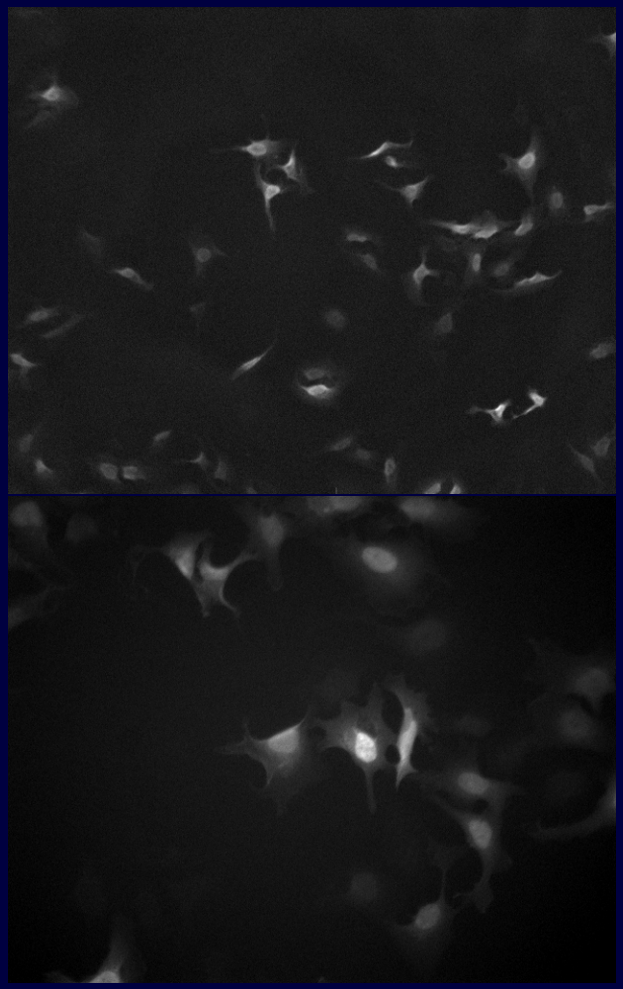
# Magnification

- What objective to use?
  - Segmentation accuracy
  - Dose response curve re-sampling

Image Analysis:

# Magnification





10x

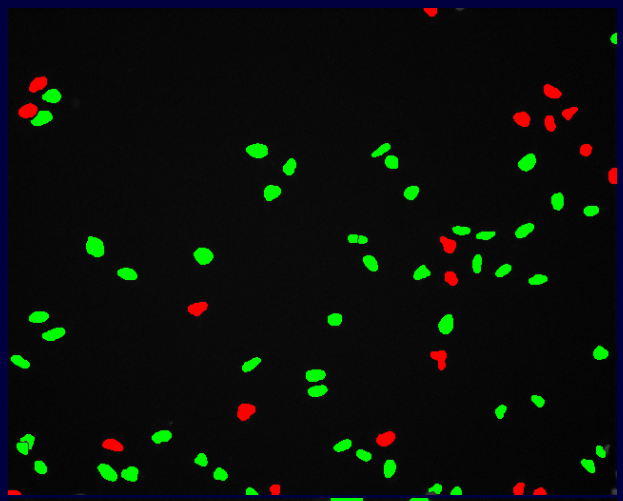
20x

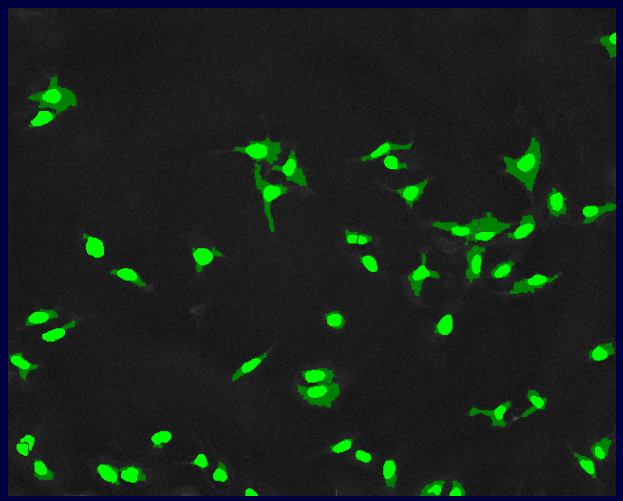
DNA

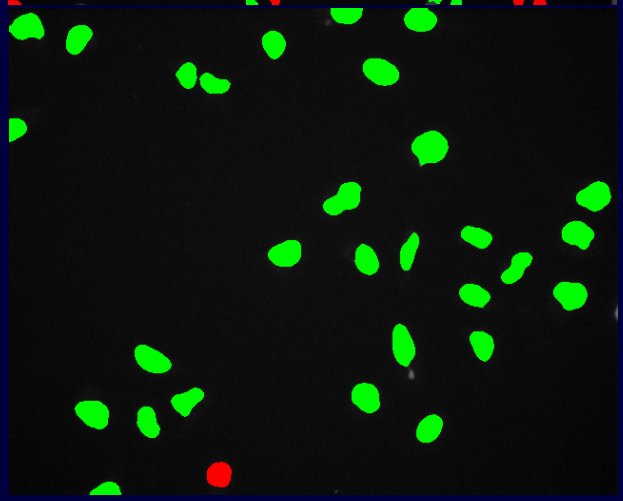
FKHR-GFP

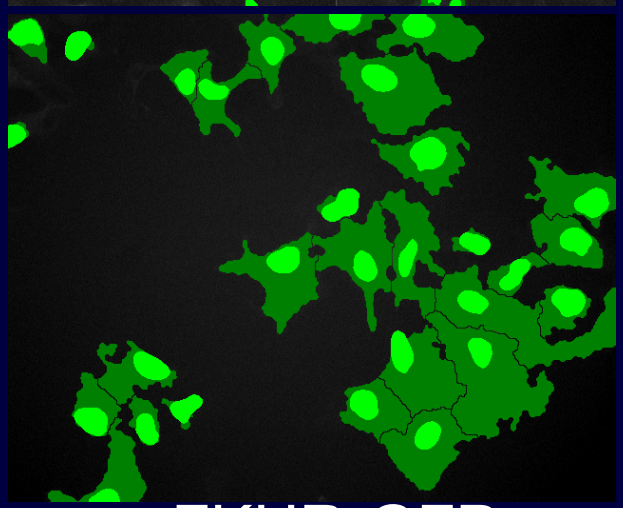
Image Analysis:

# Magnification









10x

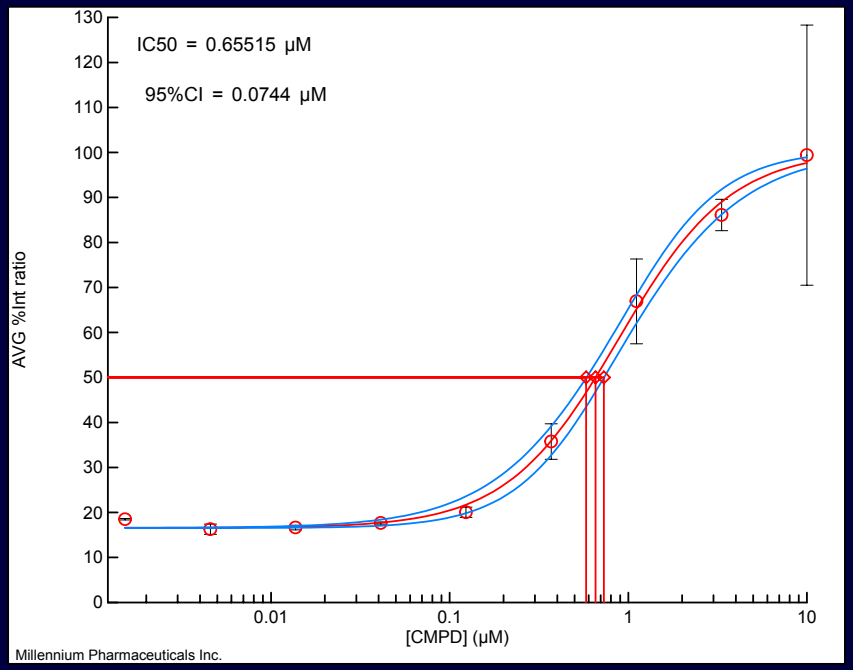
20x

DNA

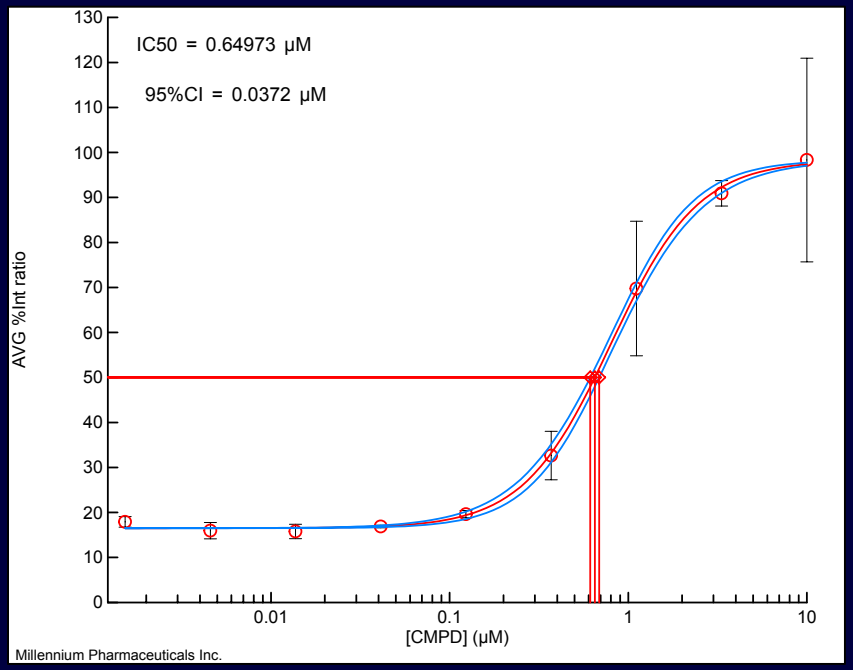
FKHR-GFP

Image Analysis:

# Magnification




10x




20x

Image Analysis:

# Algorithm parameters

- Assess
  - Effect of variability of user-defined parameters
  - Different software algorithms (vendors)
- DOE:
  - Dose Response
  - Multiple algorithm parameters

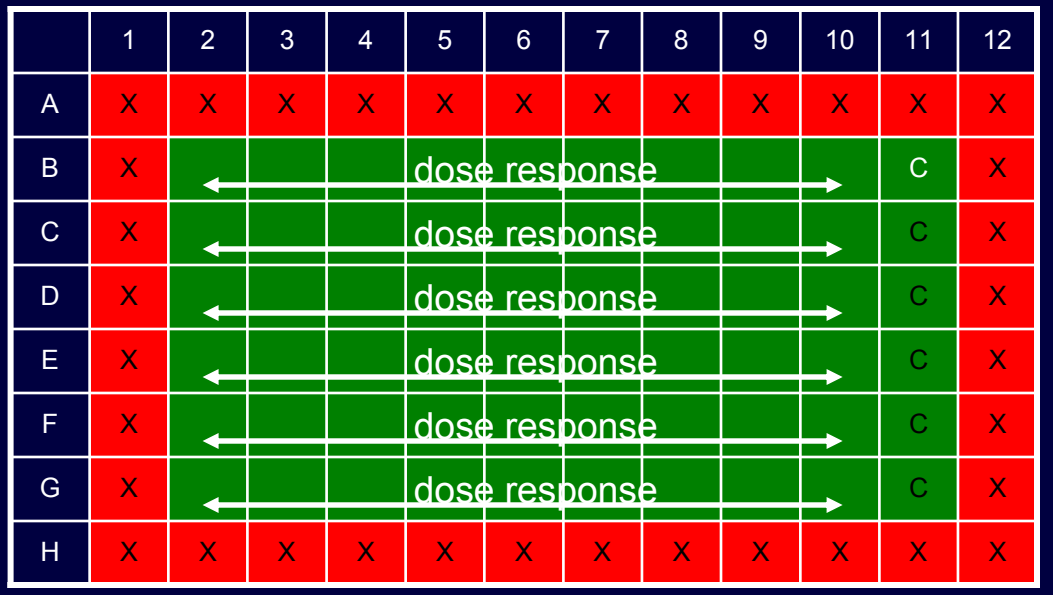
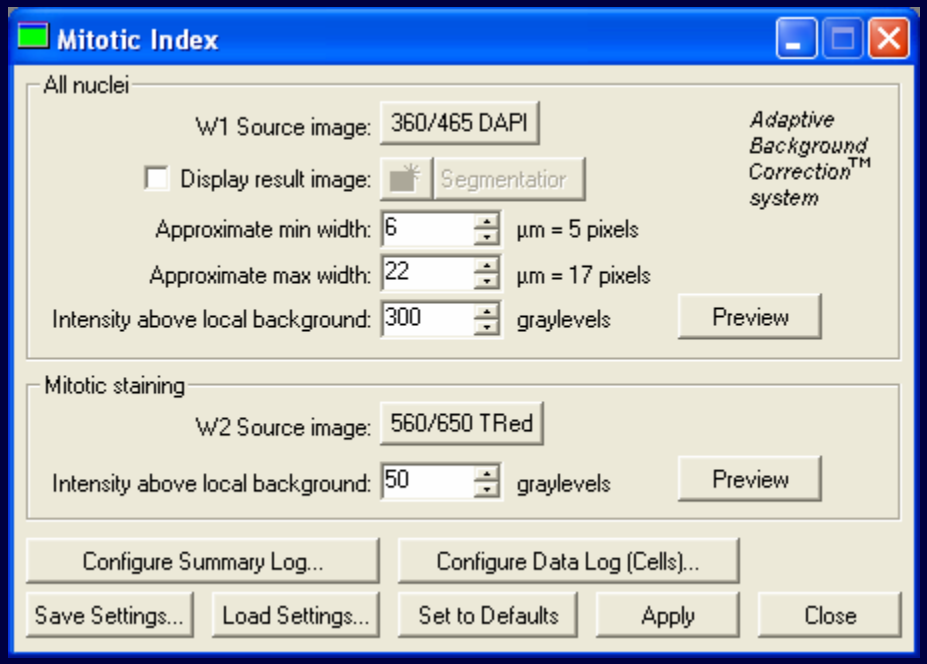
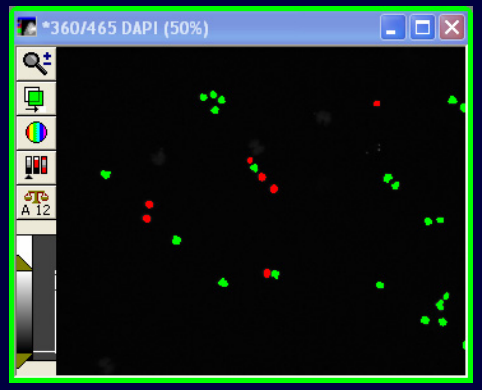


Image Analysis:

# Algorithm parameters

- User variability?



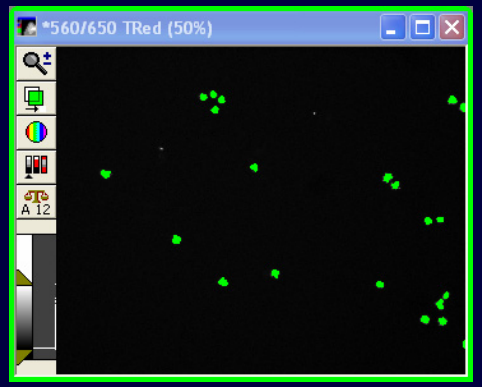
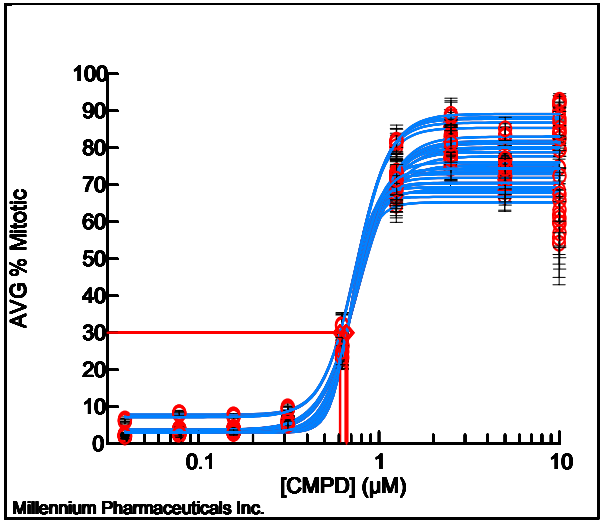
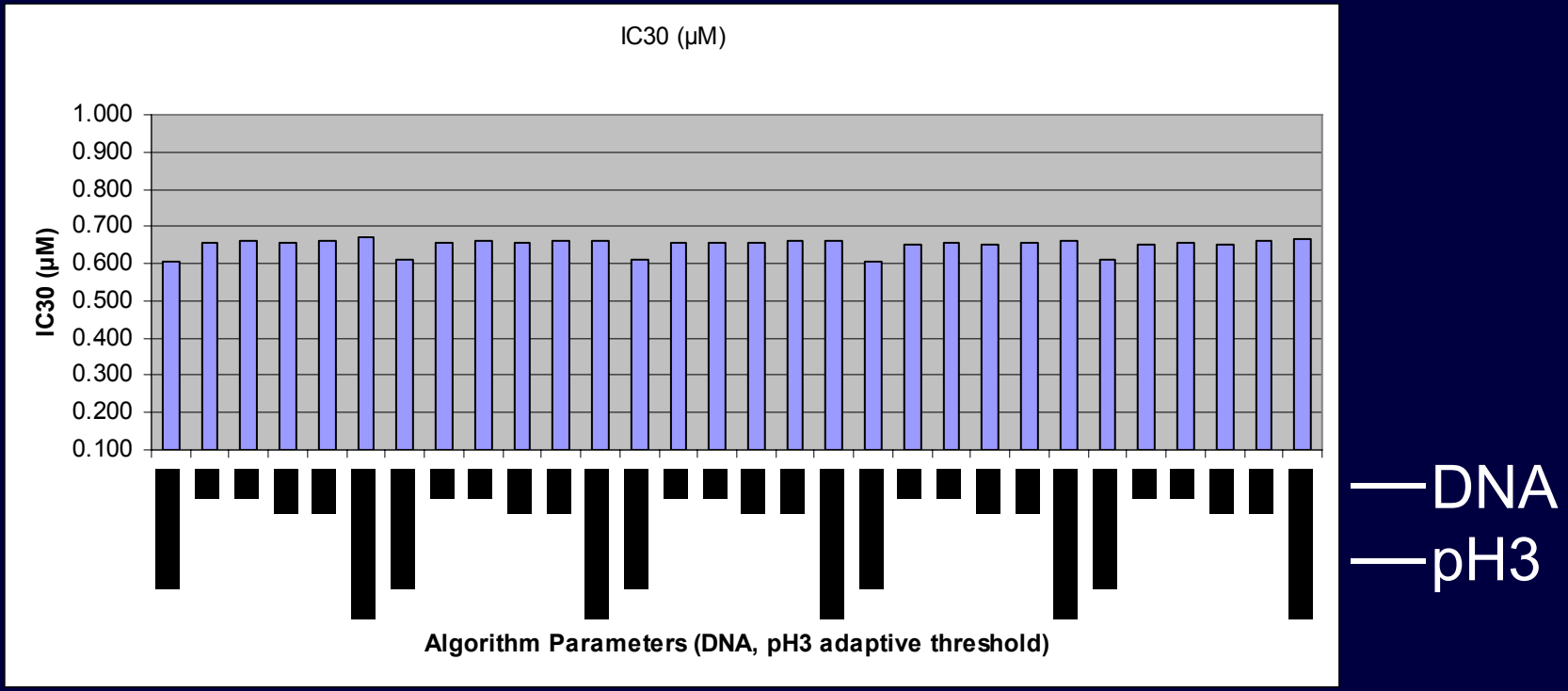



Image Analysis:

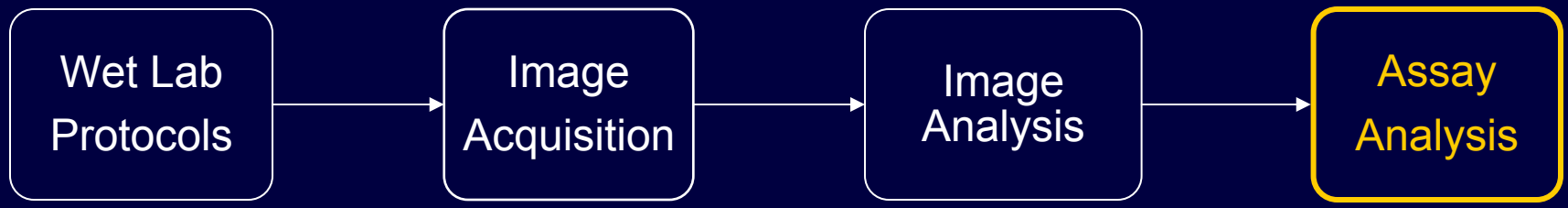
# Algorithm parameters

- Single Plate, 9 sites/well, dose-response
- 30 variations of DNA, pH3 threshold



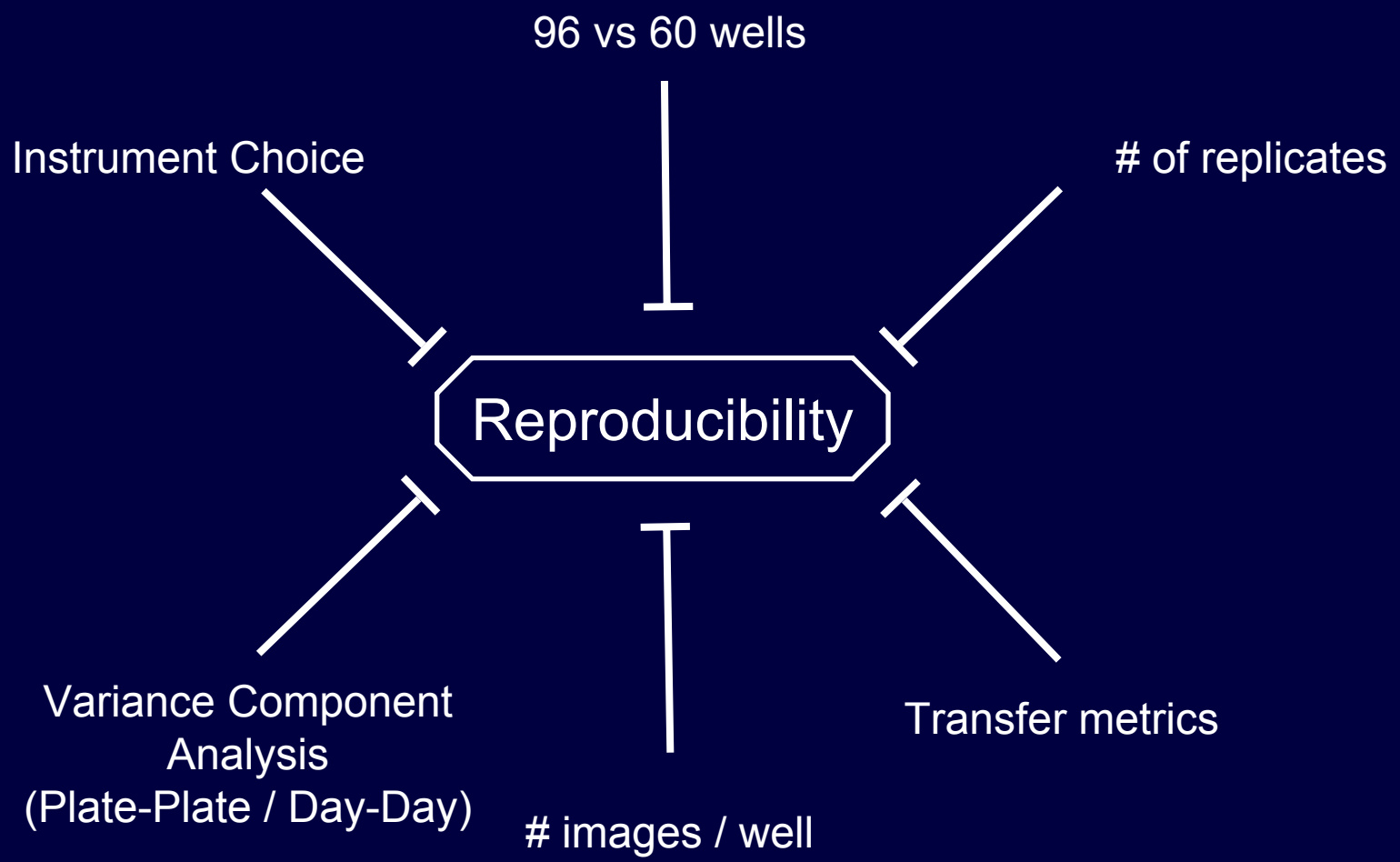



# HCS Assay Components

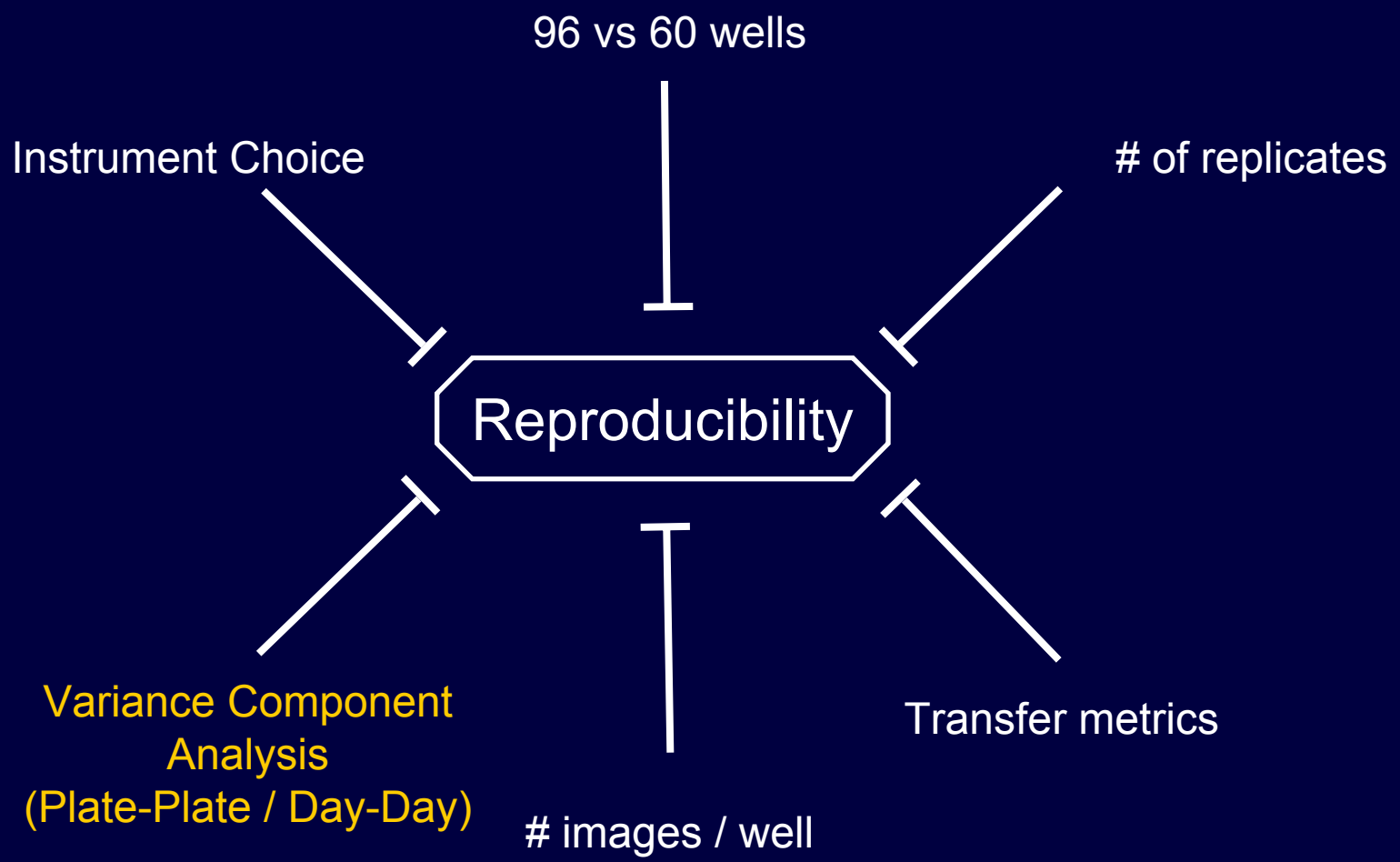



- Dissect components
- Using control experiments
  - Identify drivers of variability
  - Minimize them experimentally or account for them statistically

# Assay Analysis



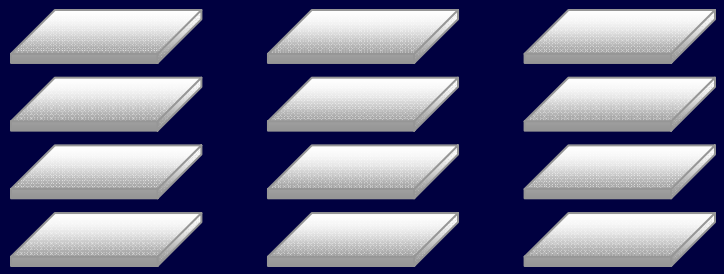
# Assay Analysis



Assay Analysis:

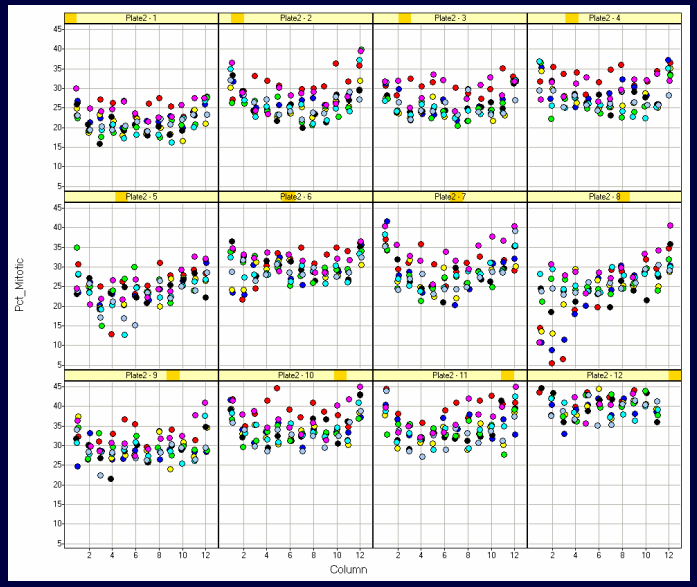
# Variance Component Analysis

- Assess variability
  - Row / column
  - Plate-plate
  - Day-day

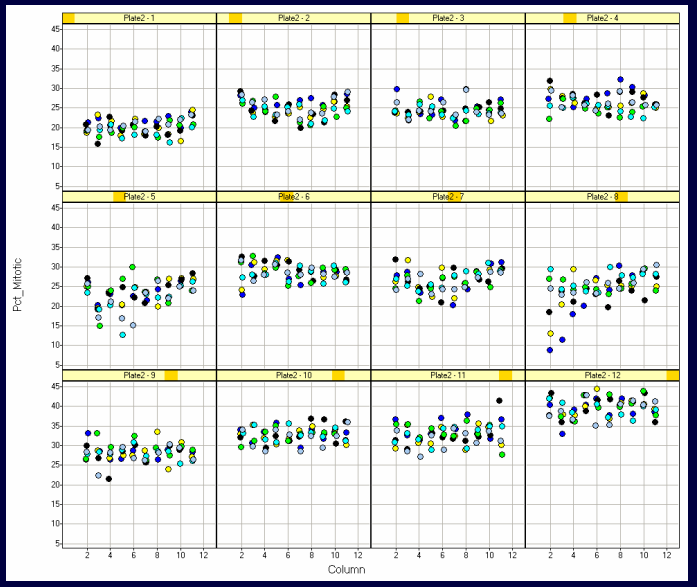



- DOE:
  - Single drug concentration (Uniformity plates)
  - 4 plates x 3 days

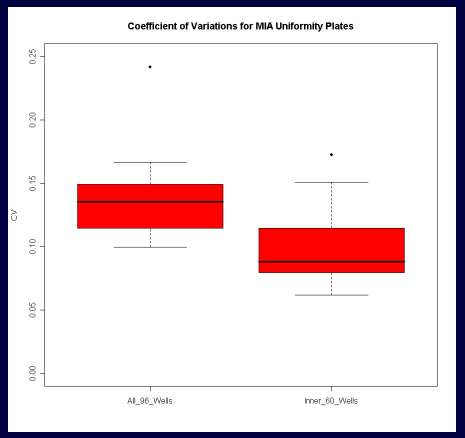
# Assay Analysis:
# Uniformity Data




96 wells




60 wells

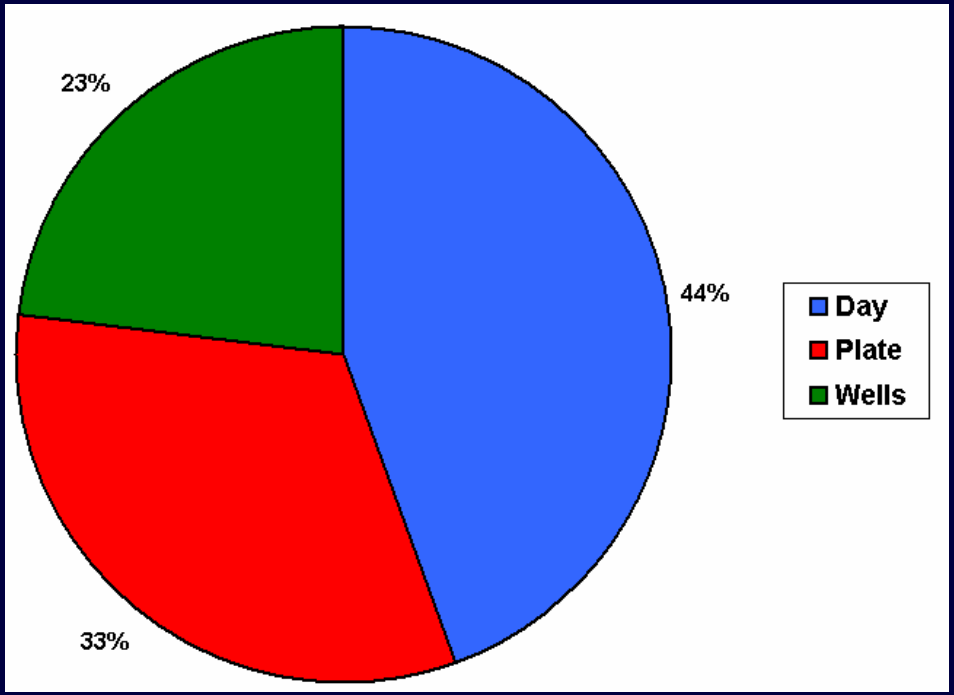



- Inner 60 wells had median CV which was 35% lower than that for all 96 wells

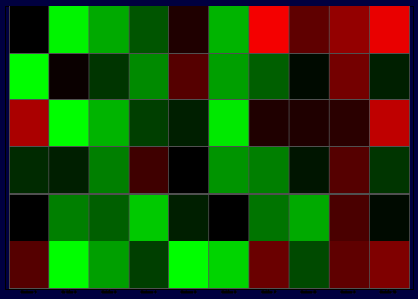
Assay Analysis:

# Uniformity Data

## Variance Component Analysis




12 Plates



Single Plate

Total Mean: 27.7

Total Variance: 32.0 (std dev: 5.7)

Coefficient of Variation: 20.4%

- Variability among days was largest component followed by plate
- Replicate dose response curves should be run on different plates on different days to account for largest amount of variability.

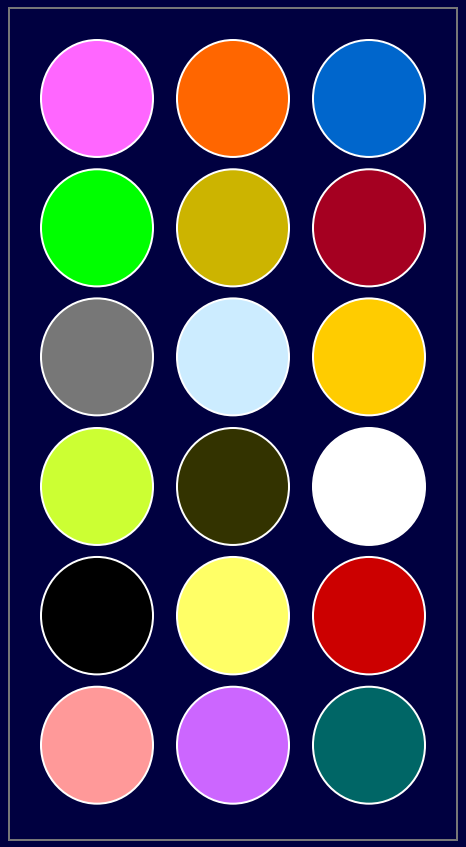
Assay Analysis:

# Re-sampling

- Determine
  - How many sites / well?
  - How many replicate wells?
  - Is the assay robust?
  - Does assay perform well across facilities?
- Sample
  - Uniformity Plates
  - Control Compound
  - Multiple Compounds

# Bootstrapping

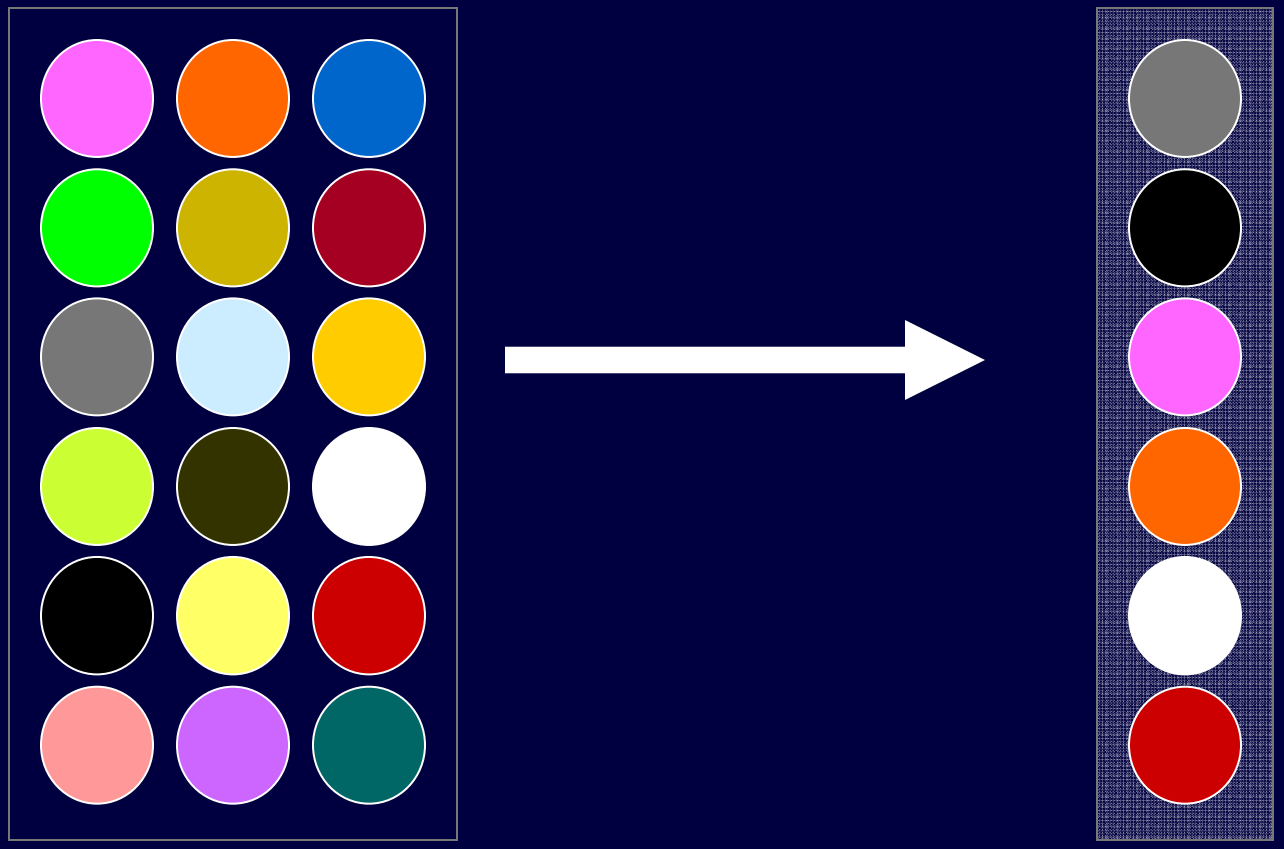
- Method for estimating the sampling distribution of an estimator by re-sampling with replacement from the original sample



Population

# Bootstrapping

- Method for estimating the sampling distribution of an estimator by re-sampling with replacement from the original sample

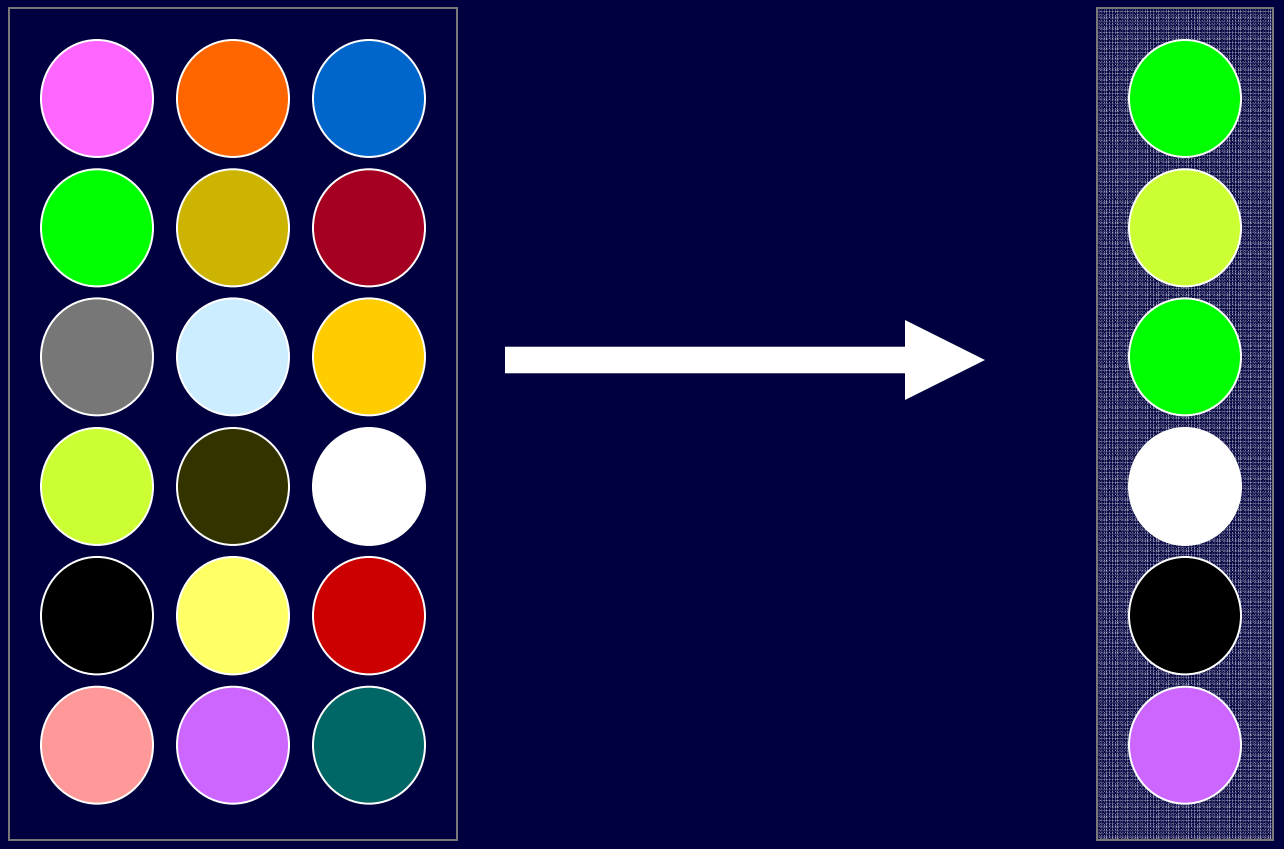


Population

Sample

# Bootstrapping

- Method for estimating the sampling distribution of an estimator by re-sampling with replacement from the original sample

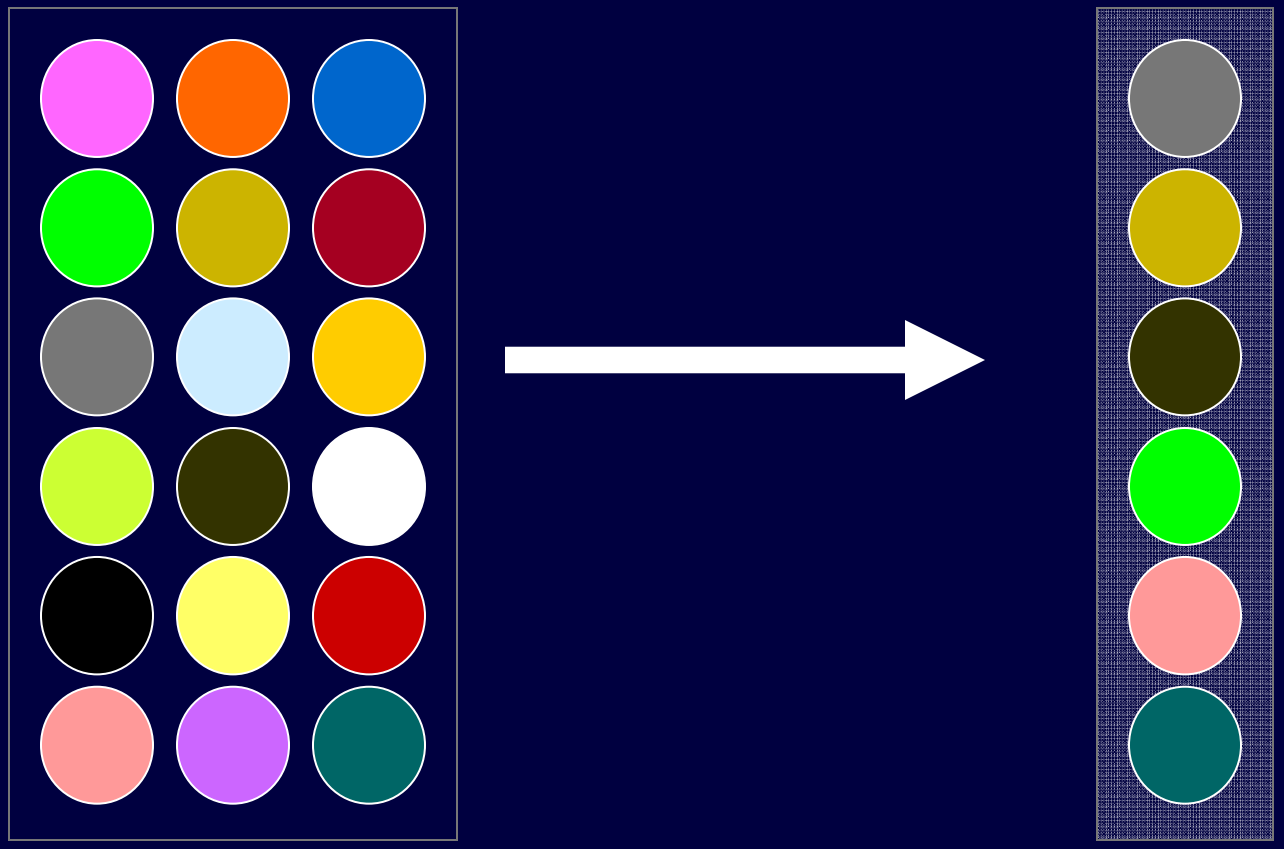


Population

Sample

# Bootstrapping

- Method for estimating the sampling distribution of an estimator by re-sampling with replacement from the original sample

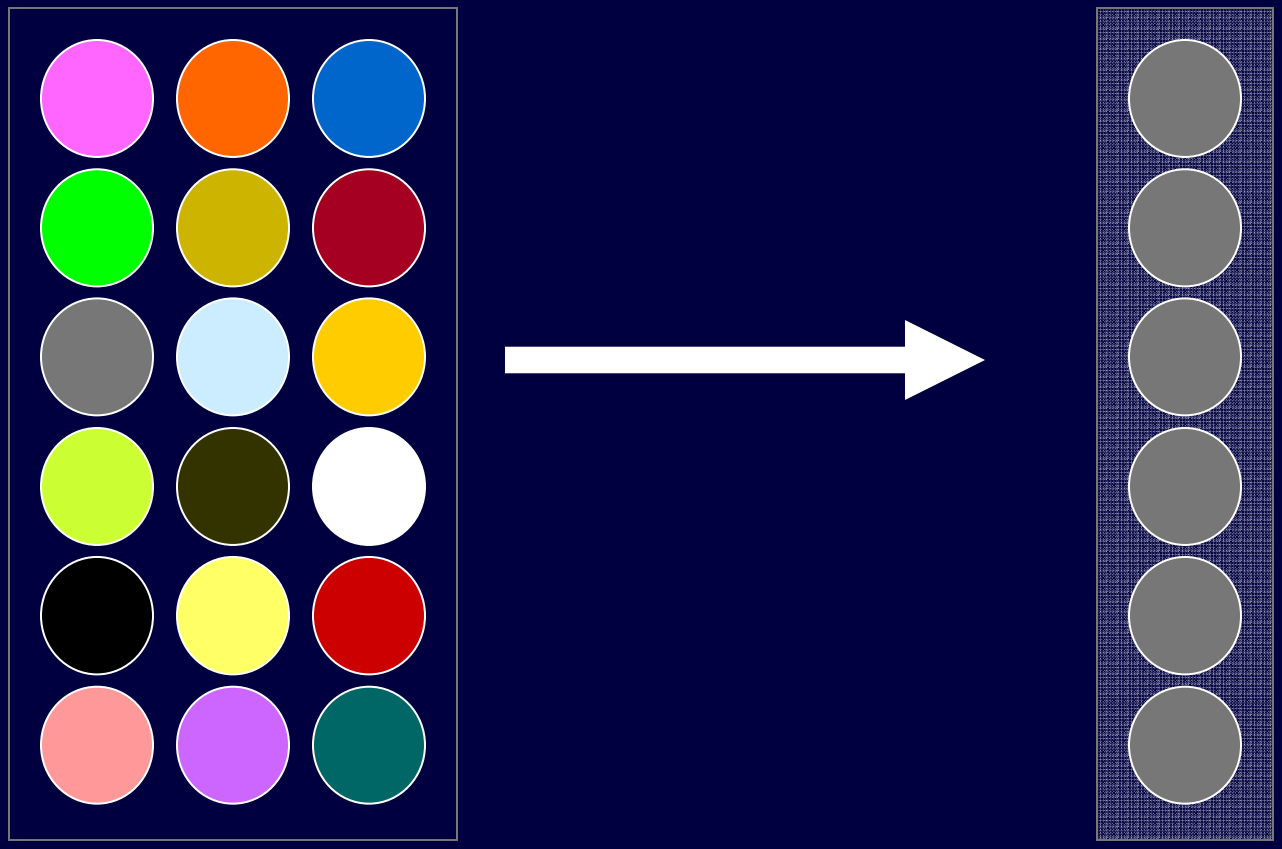


Population

Sample

# Bootstrapping

- Method for estimating the sampling distribution of an estimator by re-sampling with replacement from the original sample



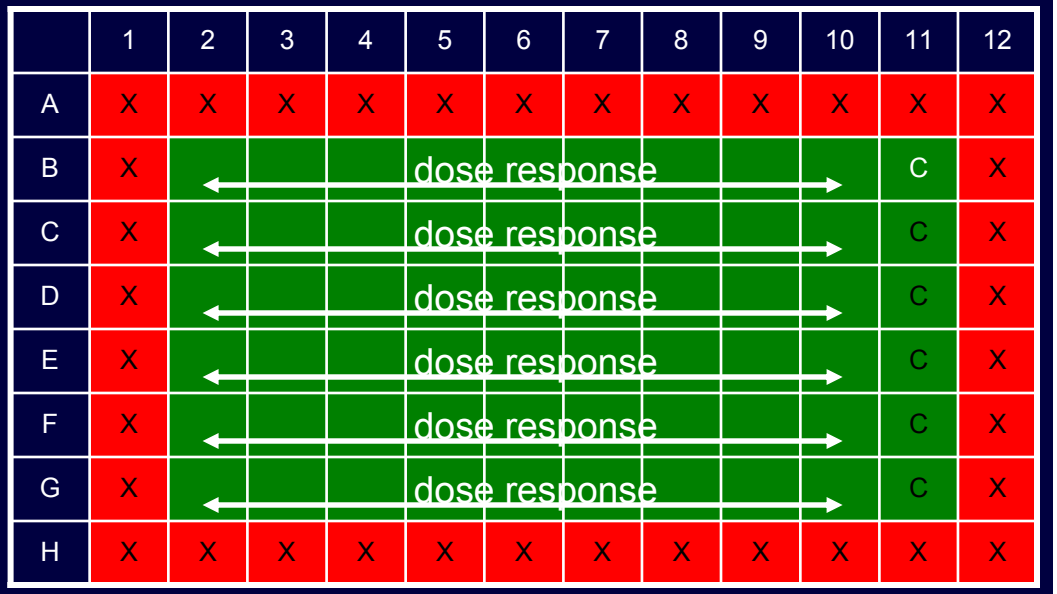
Population

Sample

Assay Analysis:

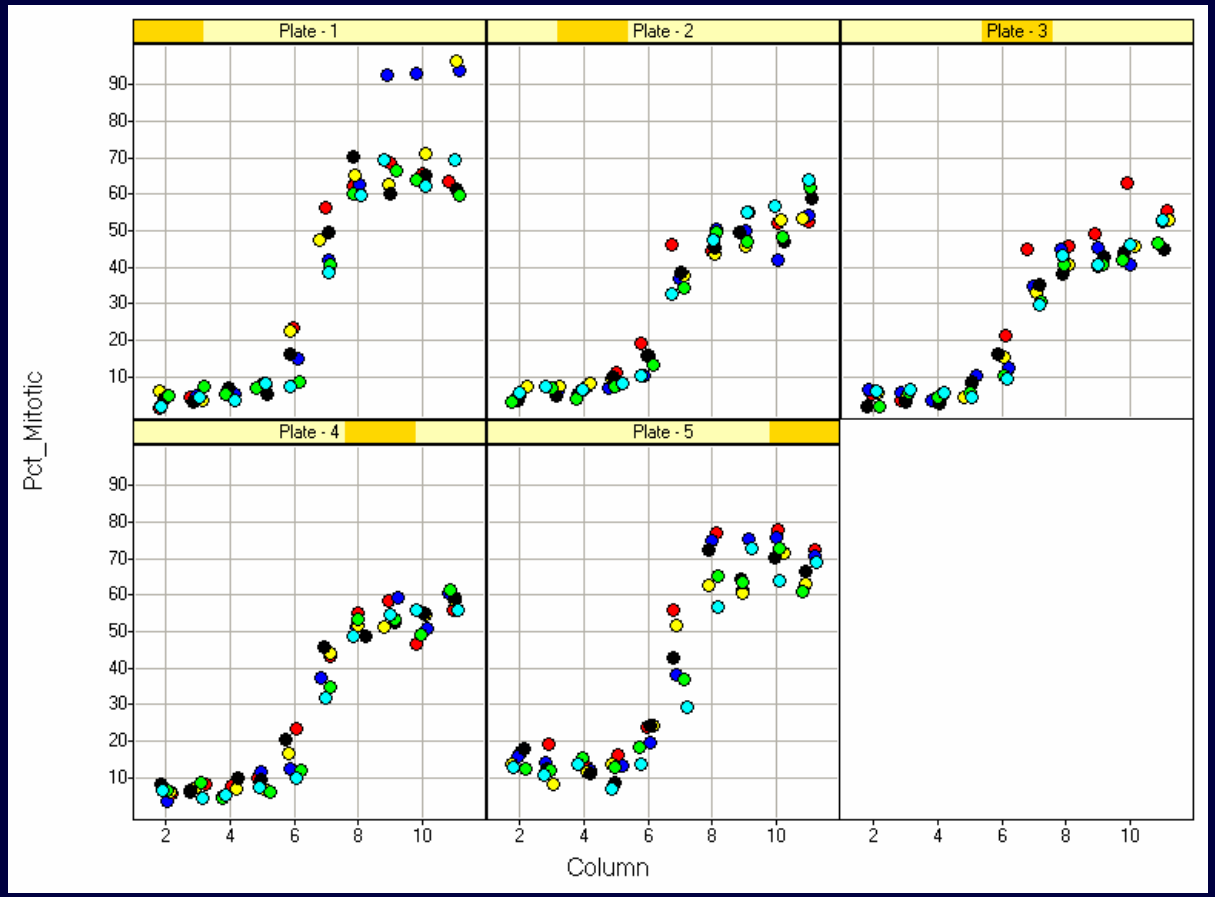
# Control Compound

- Assess
  - Effect of number of replicate curves upon EC30 estimates
  - Effect of number of images per well upon EC30 estimates
- DOE:
  - 6 dose response curves
  - 5 plates



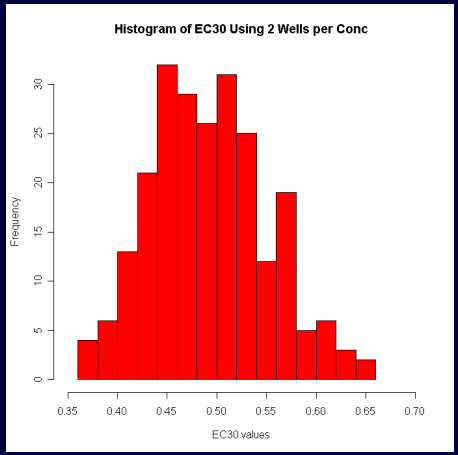
Assay Analysis:

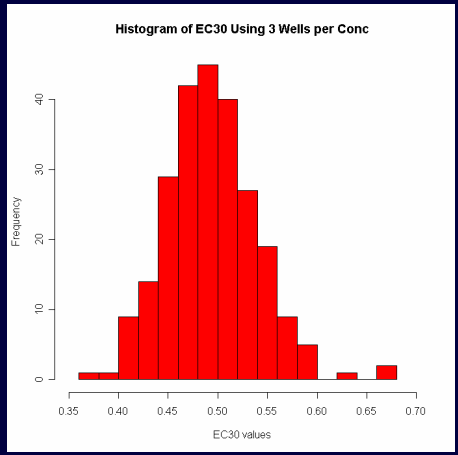
# Control Compound Data

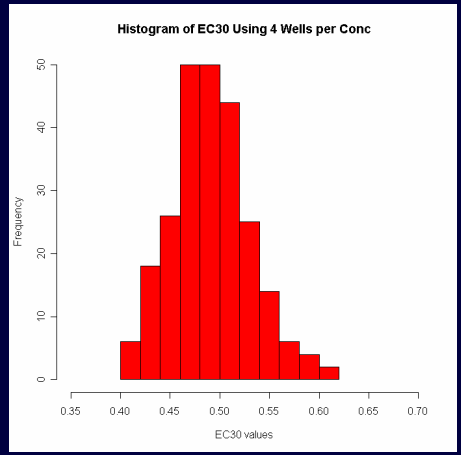


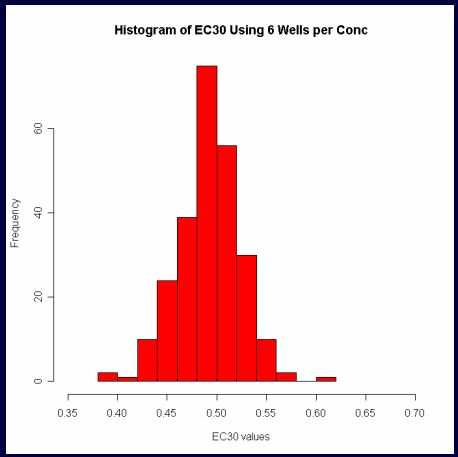
Re-sampling Results

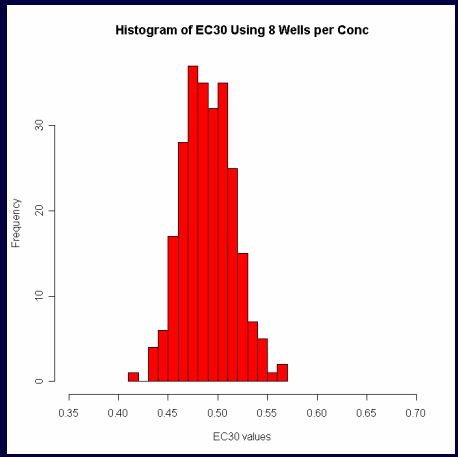
# Distribution of EC30 values

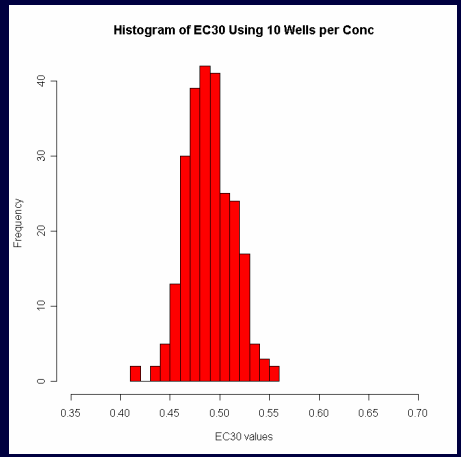


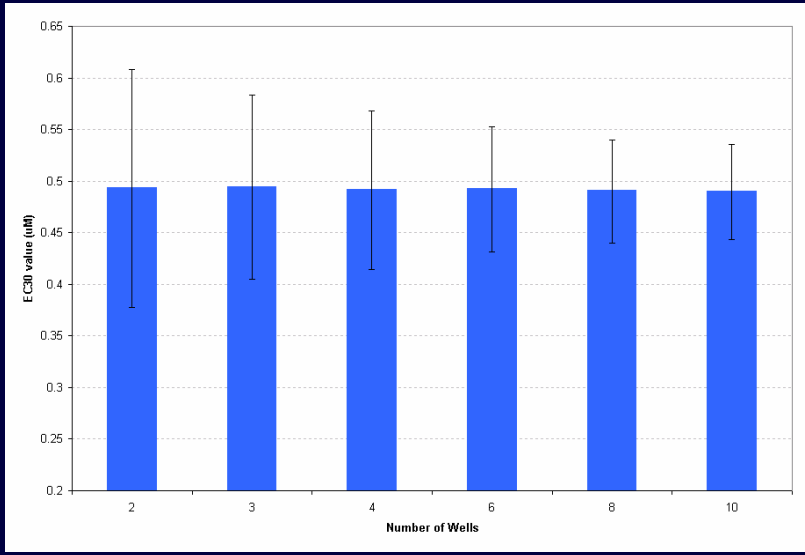



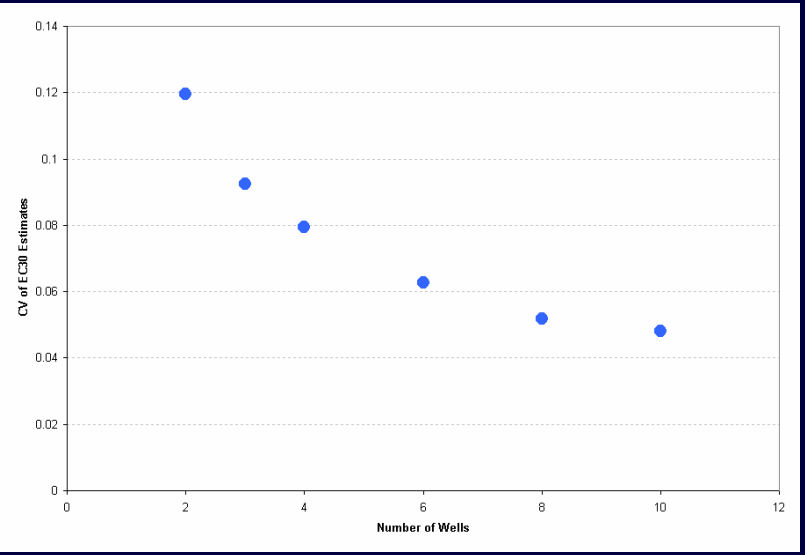



## Re-sampling Results:

# Determine # wells




Error bars are 95% CI



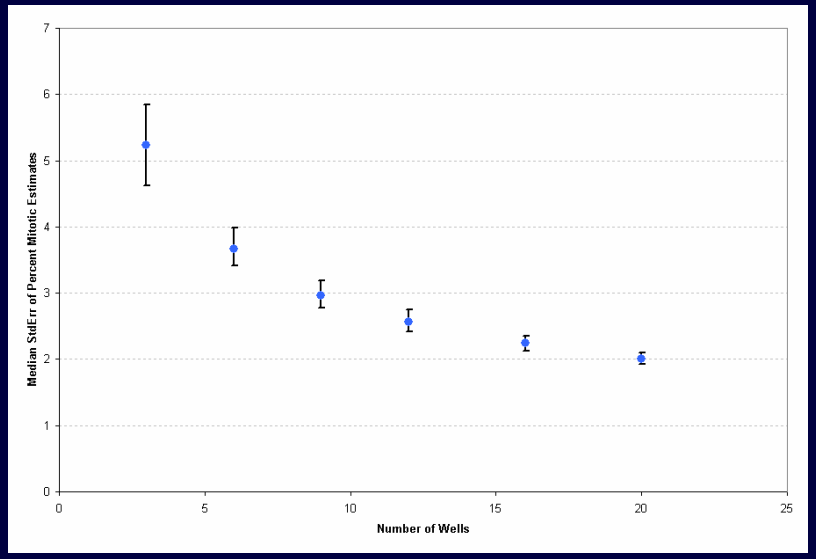

All CVs are less than 12%

-Results incorporate differences among plates and there were four observations with high values.

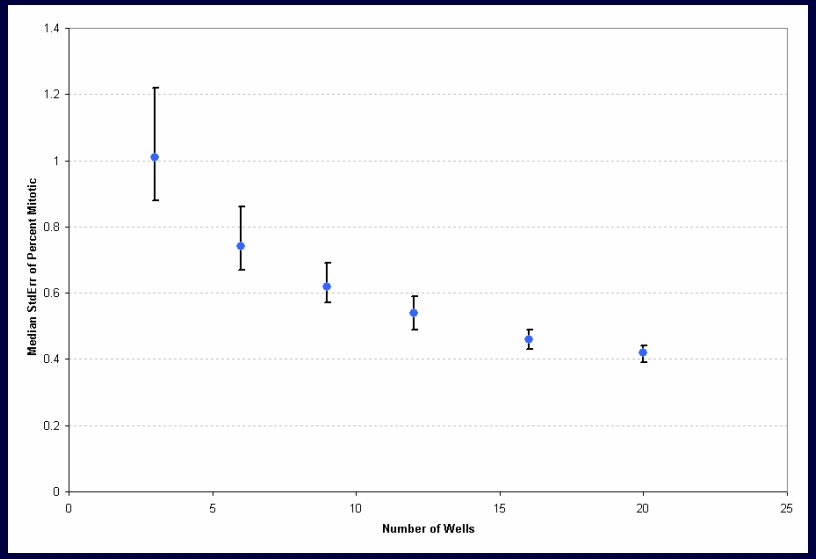
Re-sampling Results:

# Determine # sites/well

10uM

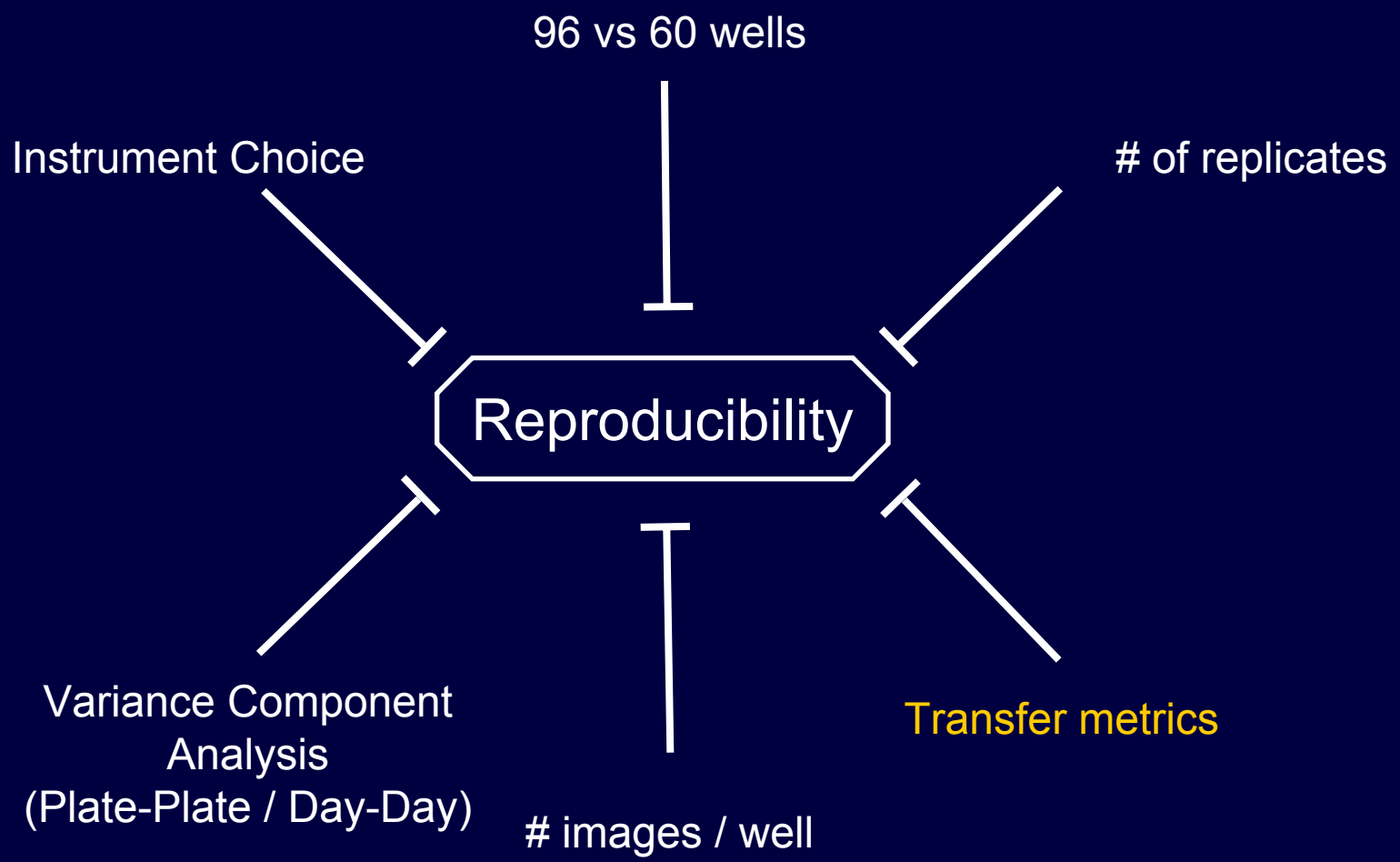



DMSO




Going from 6 images to 9 has almost as much impact on standard error as going from 9 to 16.

# Assay Analysis



Assay Analysis:

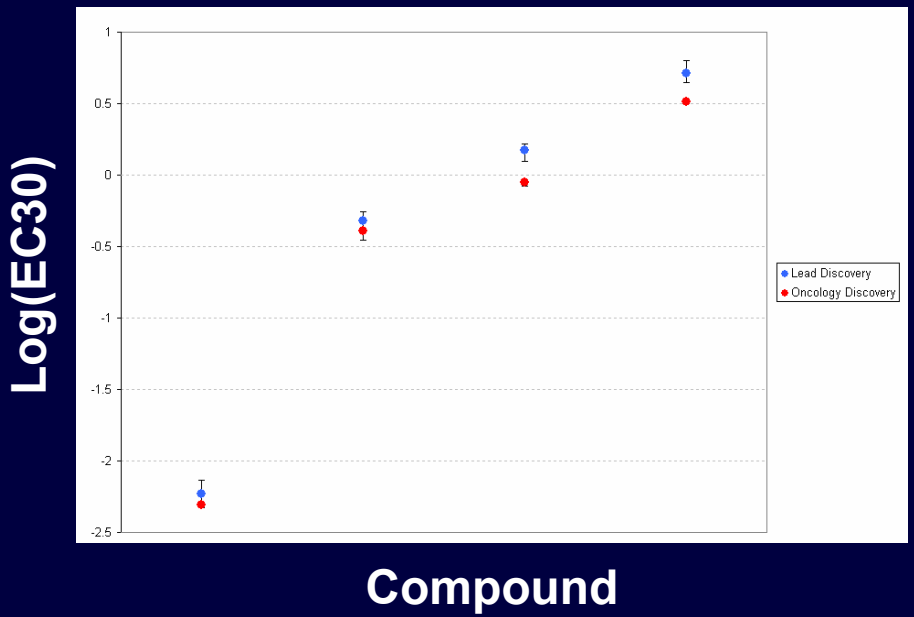
# Transfer

- Assess
  - Compare re-sampled EC30 estimates to estimates generated in Oncology Discovery
- DOE
  - 4 compounds
  - 3 replicates / compound
  - Each compound was run on 4 plates = 12 curves
  - All plates run in two facilities

Assay Analysis:

# Transfer

## Four compound re-sampling results




- Three wells from each concentration were sampled.
- Fit dose response curve and calculated EC30.
- Total of 250 curves
- Examine distribution of EC30 values

•A and B were not significantly different, C and D were approximately 60% higher

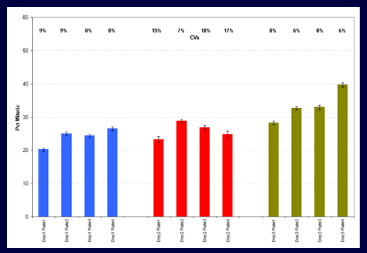
Assay Analysis:

# Instrument selection


Discovery-1

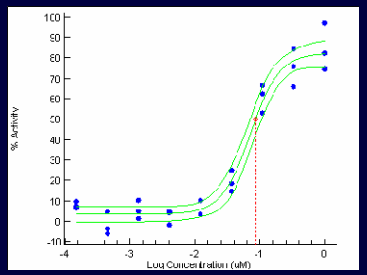
Opera

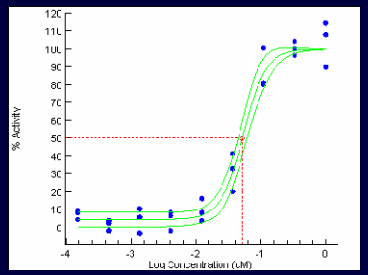
- Uniformity Plates



• Signal/Noise, acquisition time, CVs

- Single Compound



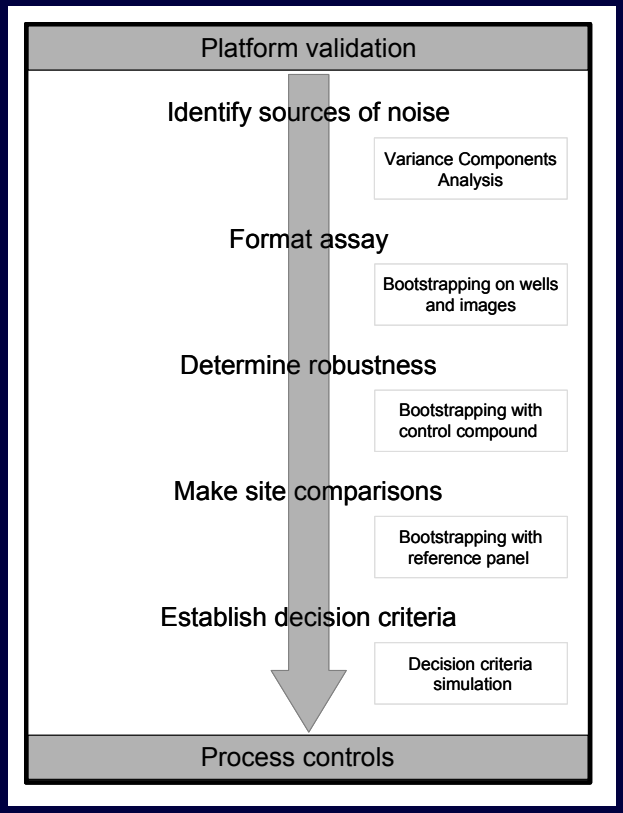




• Z', EC50 CVs, confidence interval

* determination based on signal/noise, speed, robustness, availability

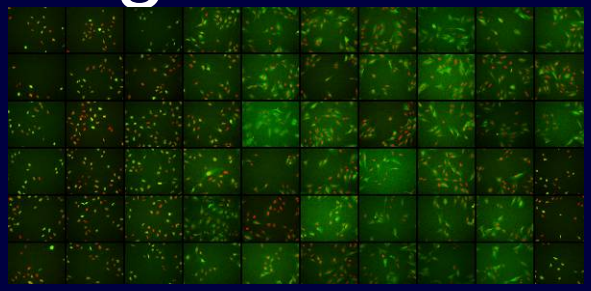
# Overall Assay Evaluation

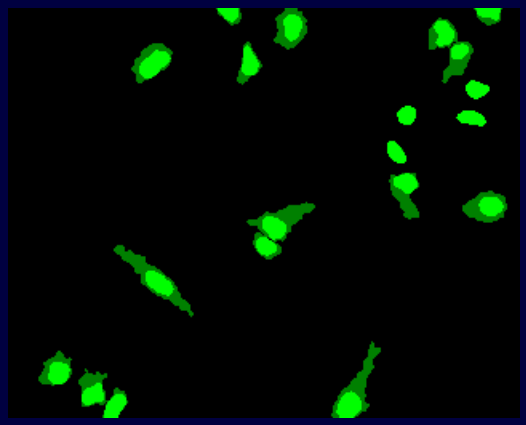
- ~2 months
- Confidence
  - Immunofluorescence staining protocol
  - Automation / liquid handling
  - Instrument / Instrument QC
  - Software Algorithm
  - Robust assay
  - Transfer from Discovery to Platform Group



# Summary

- Look at the images





- Run uniformity plates to identify and minimize noise (biology, instrument, and algorithm)
- Use analysis modules throughout process

# Process Deliverables

- Development Guidelines
  - QC metrics: instruments, algorithm
  - IF staining protocol optimization
  - Image analysis optimization
- Production Guidelines
  - Statistical analysis for transfer
  - Ongoing QC metrics

# Acknowledgements


- Molecular & Cellular Oncology
  - Ben Amidon
  - Jeff Ecsedy
  - Natalie Roy D'Amore
  - Liz Carideo
  - Kara Hoar
  - Claudia Rabino
  - Michelle Tighe-Nestor,
  - Denise Driscoll
  - Deborah Wysong
- Cancer Pharmacology
  - Arijit Chakravarty
  - Katherine Galvin
- Lead Discovery
  - John Ringeling
  - John Donovan
- Computational Sciences
  - Sudeshna Das
  - Mike Pickard

# Breakthrough science. Breakthrough medicine.SM