Quantitative Image Analysis 8

Edited by: Zvi Kam, Weizmann
For Advance Light Microscopy course

Quantitative Image Analysis

- The digital image
- Biological feature extraction
- quantification
- Statistics

General post-genomic goals in cell biology:

Accumulate a coherent picture of the integrated behavior and coordinated activities of cellular components under normal and perturbed conditions. Test role of many components, identify the relevant ones for various cellular functions. Assign weights to their contributions in normal and diseased cells.

Since such integration must involve multiple labs:

Technical goal:

Establish a universal framework for compilation and multi-lab integration of cell-level informatics:

"*equivalence to BLAST in sequence data*"

- Classical biology: "descriptive". 200 years of cataloging life complexity led to functional understanding.

- Molecular biology: "specific". Identify life molecules, motifs (structure-function), control mechanisms.
Life in concrete terms.

- After 50 years of molecular biology, understanding is searched at higher "systems" level.

How to integrate "specific" with the "descriptive"?
How to link phenotypes with molecules?

We can combine the power of "descriptive biology" with Molecular biology by making the descriptions QUANTITATIVE. This is the key requirement to convert molecular network graphs into FUNCTIONAL MODELS with predictive capability for perturbations and diseases.

Quantitative light microscopy is most fit for cellular studies: rich multi-dimensional information (space, time-dynamics, "colors"). Provides excellent link between molecular behavior and cellular functions.

This is the potential power of high-throughput microscopy

Gels - average protein content and modifications

FACS - cell-by-cell content (and gross cell shape)
<10 parameters per cell

Microscopy - per-cell content, intra-cellular localization,
detailed structure of cellular organelles
50 parameters per cell in one experiment ["well"]
500 parameters for complementary labeling

+ drugs and genetic perturbations
space and time

This is a promising basis for cell-level informatics

but…

Variability between cells imposes many samples (high-throughput).

Should be acquired without giving up information (definition of sub-cellular structures, compartments and organelles, localization of molecules).

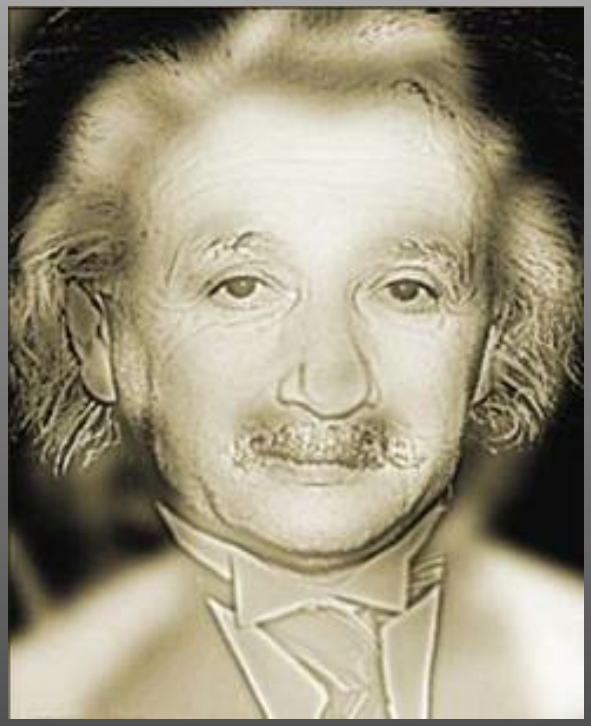
I. Acquisition technology:

II. Analysis:

High-throughput high-definition microscopy

I. Acquisition technology:

A key feature is the image **resolution**.

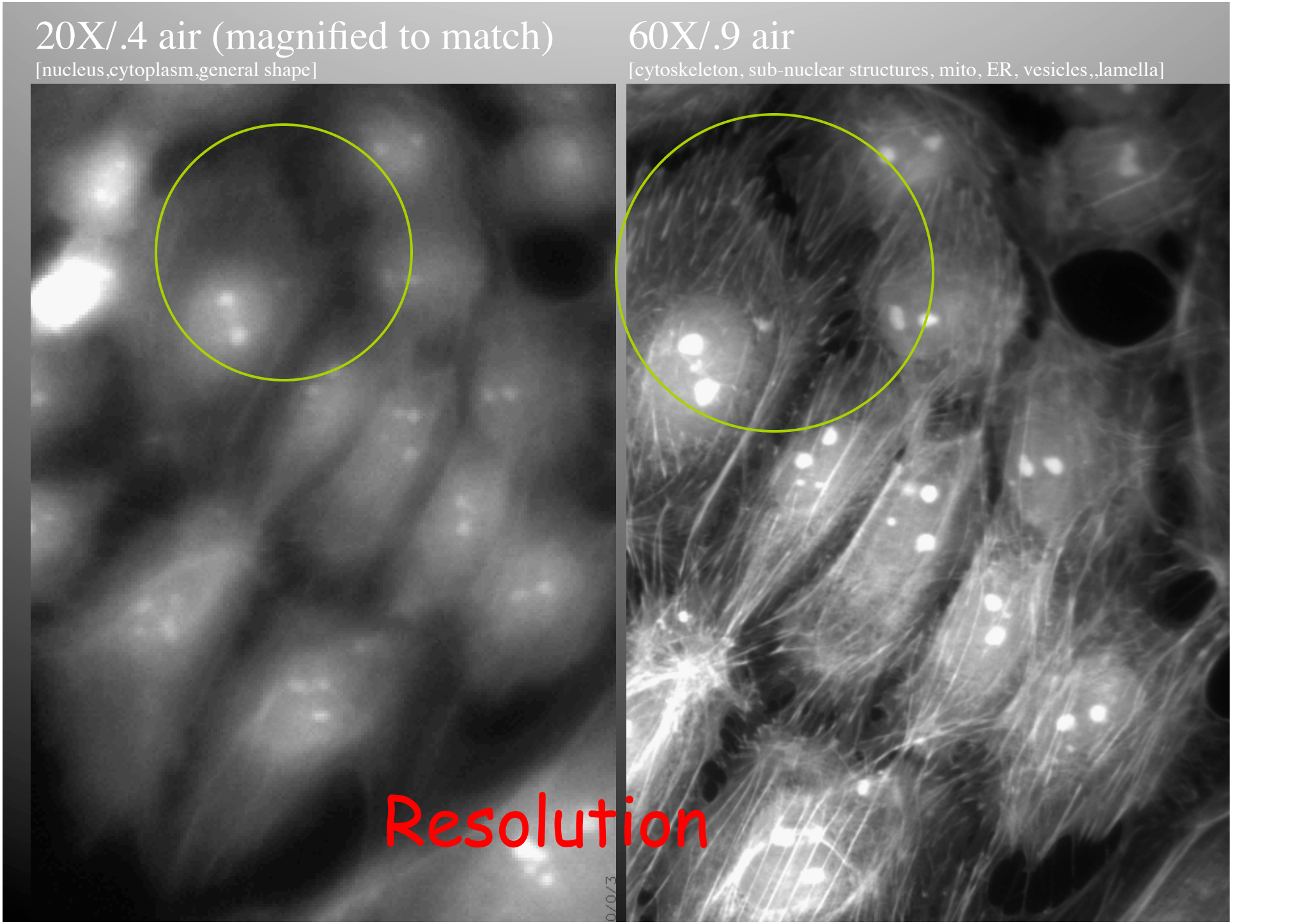


20X/.4 air (magnified to match)

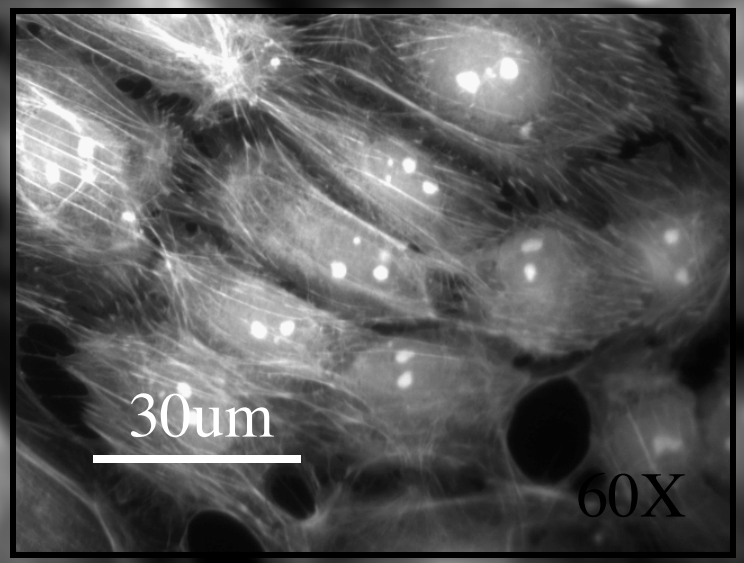
[nucleus,cytoplasm,general shape]

60X/.9 air

[cytoskeleton, sub-nuclear structures, mito, ER, vesicles,,lamella]



The penalty



(+thin depth of focus)

The Screening Microscope that we put together:

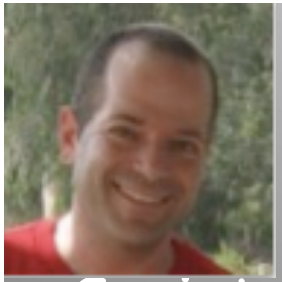
Olympus IX71 special model with objective changer
Prior XY stage
Very fast laser AutoFocus for all objectives till 100X air [water coming]
Sedat quad and Live Tripple with 10mSec filter wheels
Andor camera
Environmental chamber (+/-0.2°C) CO_2 via permiable membrane [3 days]
Directly coupled LED illumination [Lumincor] [~2x Hg Arc]
4Terra disk
Scripting language, time, plate positions, color and F/T looping
Multiple Z 3D or Sweeping-Focus 2.5D
Low-Mag search -> Hi-Mag acquire mode
Cell Phone SMS alerts

Today (in 1h) we shall not attempt to teach or even review methods in image processing:

- Very intensive field – academic and industrial
- Covers extremely broad types of imaging modalitires
- Hundreds of dedicated and related journals
- Thousands of papers and conferences each year

Instead try to set the principles in application to microscopy.

- Special niche in image processing:
  - fluorescence – extracts the features of interest
  - multi-color multi-modal images of specimens
  - multi-parametric quantification – almost unique



The YuvGAP Solution

YuvGAP - **Y**our **U**ltimate **V**ersatile **G**eneral **A**nalysis **P**latform

- Script-driven analysis pipe-line structure with unlimited depth looping capability
- Image analysis is broken into modular steps each with input/output images and parameters
- Modules sorted into categories for simplicity
- Fast data communication between modules [shared memory] and interactive user monitoring
- Scripts can be called by other scripts:
  hierarchy of basic, low- and high-level scripts
  names use biological terminology – readability
  + on-line help, manuals and examples
- Define cell "objects" with morphological, intensity and textural attributes
- Open to host new routines
- Import images and metadata for different file formats

How to extract Quantified features from images

Image analysis: a field of optimized receipts
Can observe repeated processing categories:

Preprocessing *(image improvement)*

field correction
de-noising
background estimation and subtraction
contrast enhancement
textures (fibers, points, edges, graininess)

Segmentation *(extraction of features, define ROI)*

threshold
connected components (binary, waterShed, multiscale, fiberScore, vesicleScore)
segment aggregation and multi-mask merging

Quantification *(of morphology & fluorescence)*

area, perimeter, extents, convexHall, secondMoments, higherMoments (Zernike)
fluorescence colors total and average intensities, background subtracted
textural energies

plus

Statistics *(Images, Analyzed results)*
Visualization *(Images, Analyzed results)*

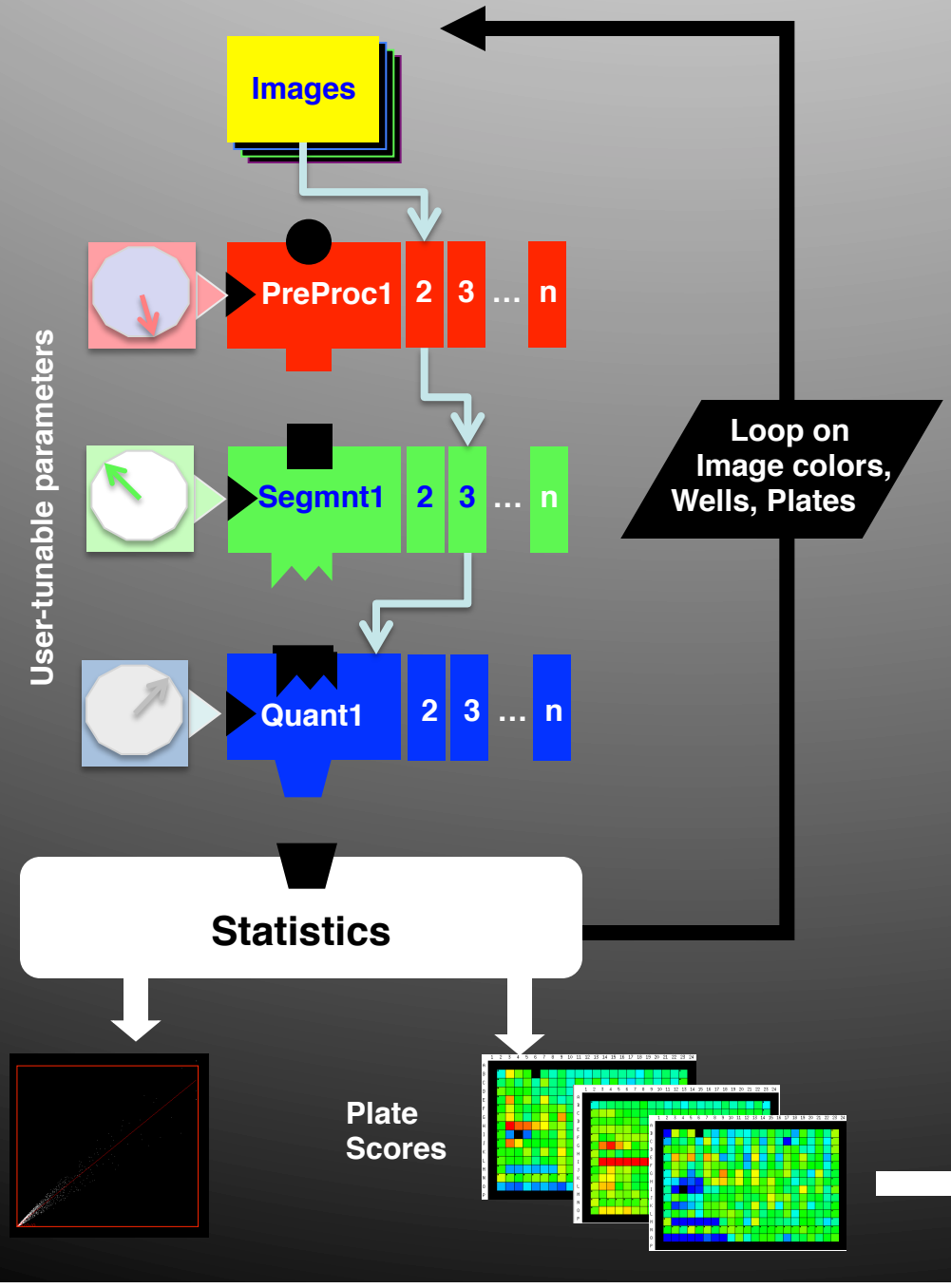
Image processing as a cook-book of receipts

Breaking the analysis into sequential steps:

In each step we can chose from modules
in the corresponding category

Interactive optimization of the analysis steps
and tune user-defined parameters
Automation: creates a pipe-line looping on many images
multi-dimensional structure of the experiment:
"meta-data"

Image processing pipe-line as a jig-saw puzzle



Multicolor,
time dependent
images

Preprocessing steps
(illumination correction,
de-noising,
Background subtraction)

Objects segmentation
(nuclei, cells, organelles,
cytoskeleton, vesicles)

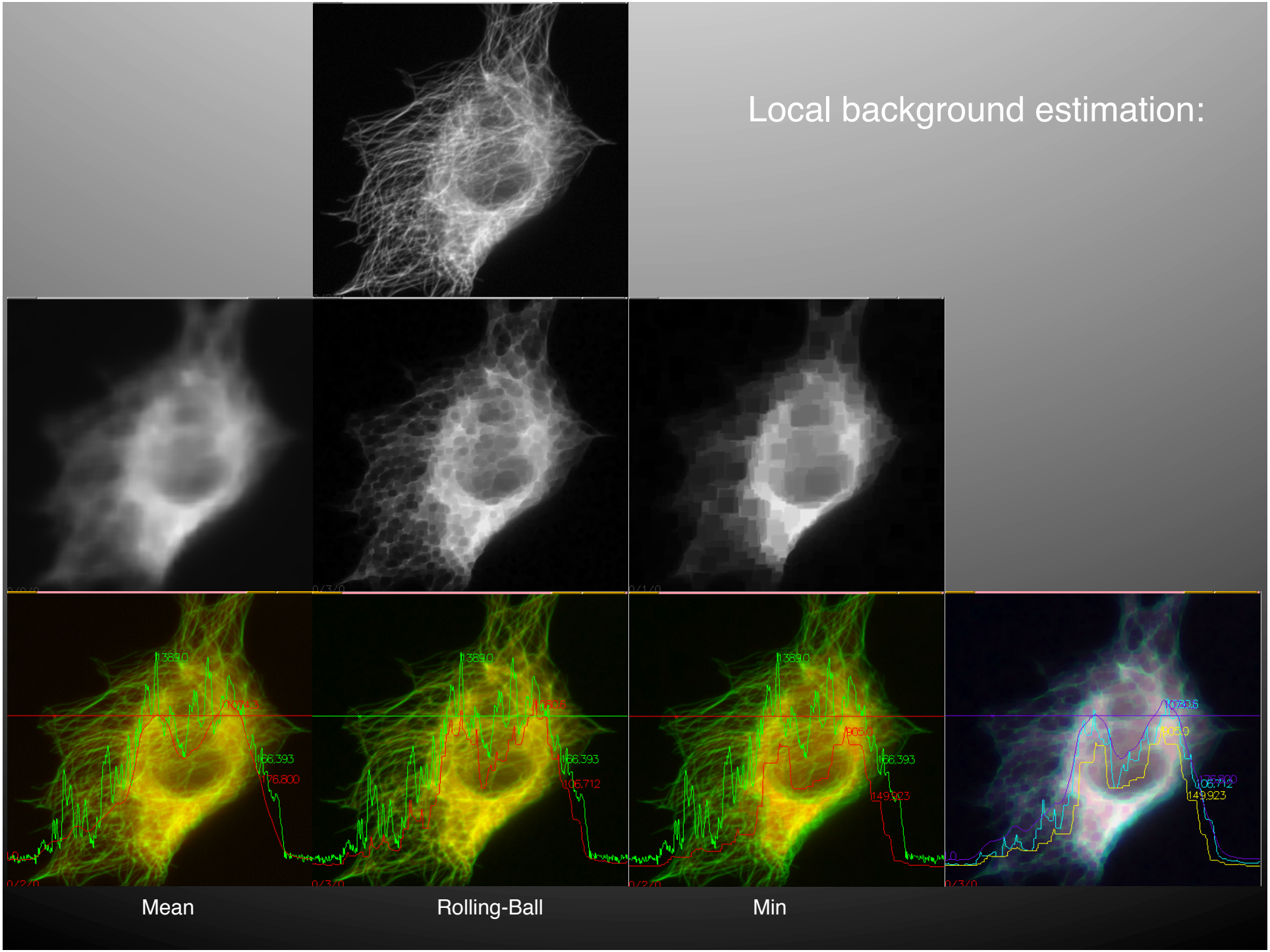
Quantify features
(morphology, intensity, texture)

Accumulation of quantified data,
Statistical analysis
Scoring of effects compared to control

I. Preprocessing

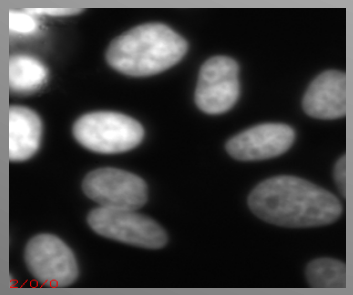
Correction of non-uniform illumination
Bleach correction (time-lase)
De-noising
Color decomposition
Deconvolutions
Background estimation and subtraction
Smoothing, flattening, contrast enhancement.
Edges, Fibers and Points enhancement.
Textures: directional polarity, graininess, density (cells, intracellular organelles), tissue organization
Combination of fluorescence, phase and DIC images

Local background estimation:

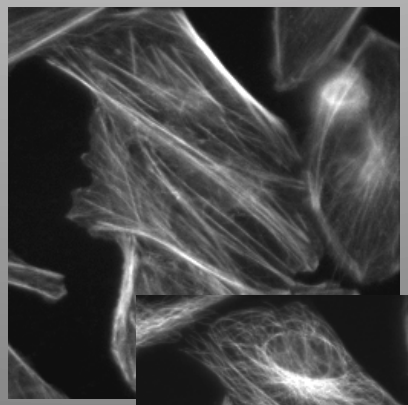


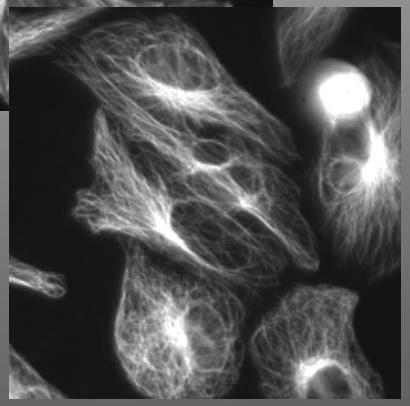
Preprocessing help segmentation (quantify original images)

Original

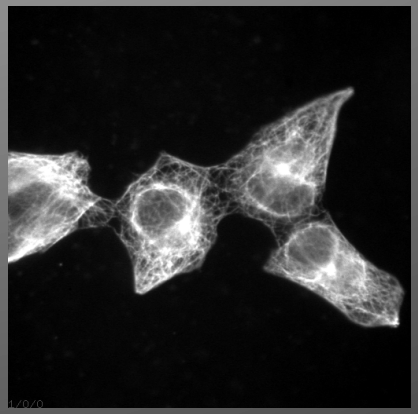


Smoothing to define

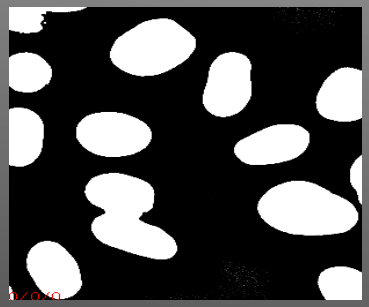




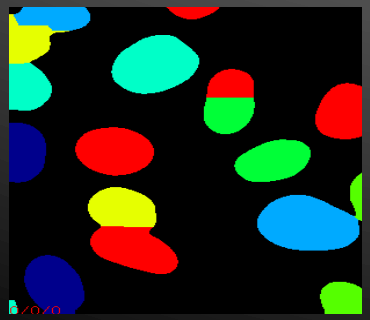
Rolling-Ball
Subtracting local background

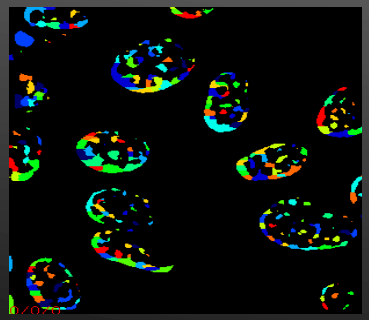


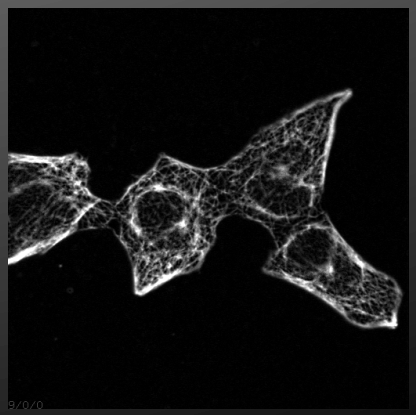
Threshold before and after field flttening



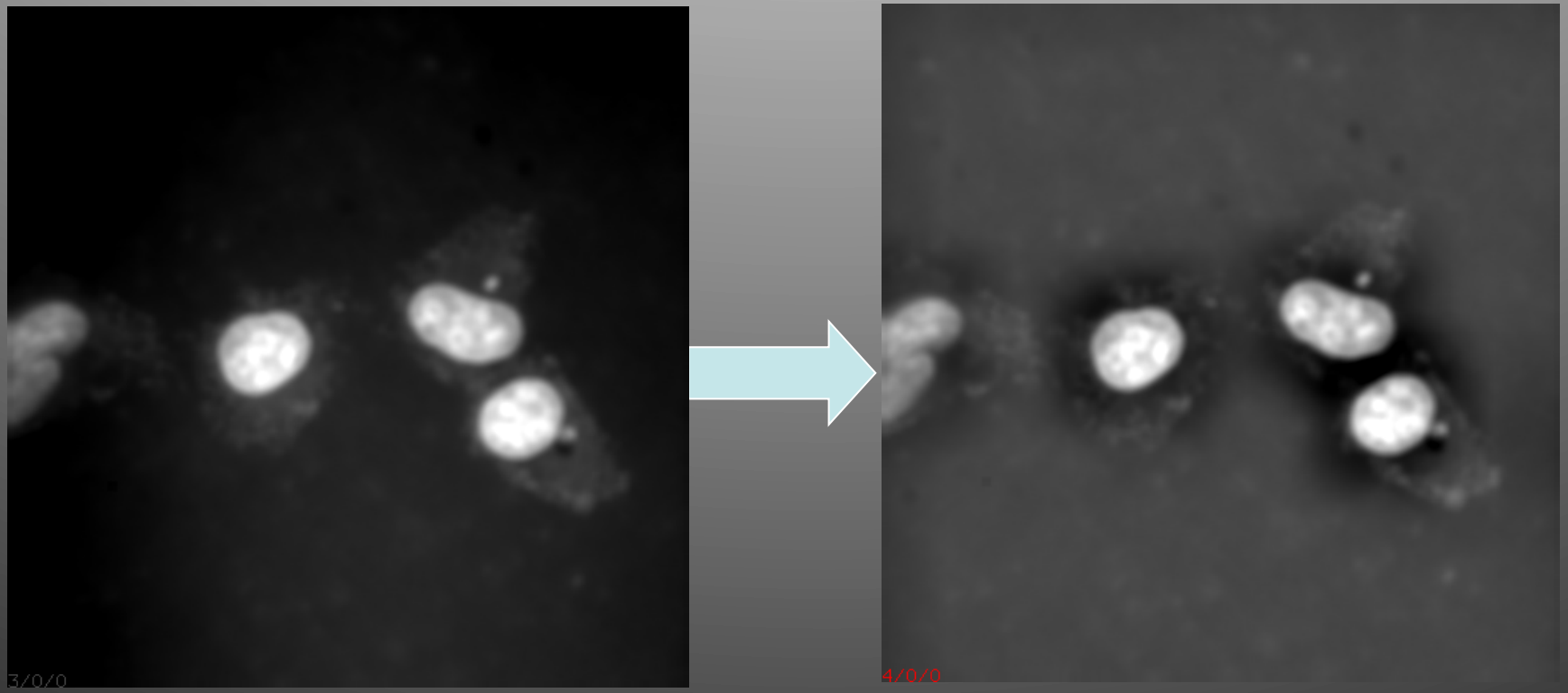
Segmentation of nuclei and intra-nuclear texture



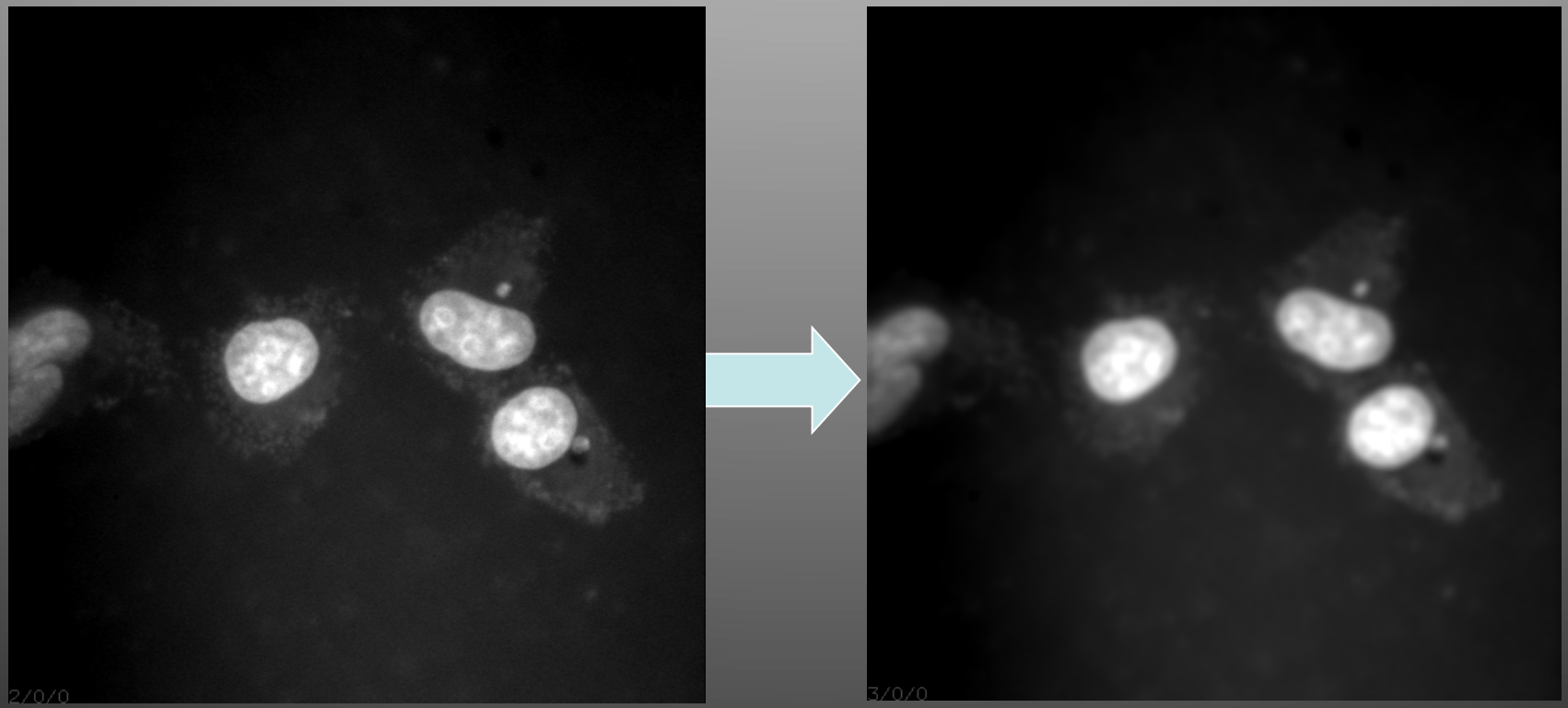




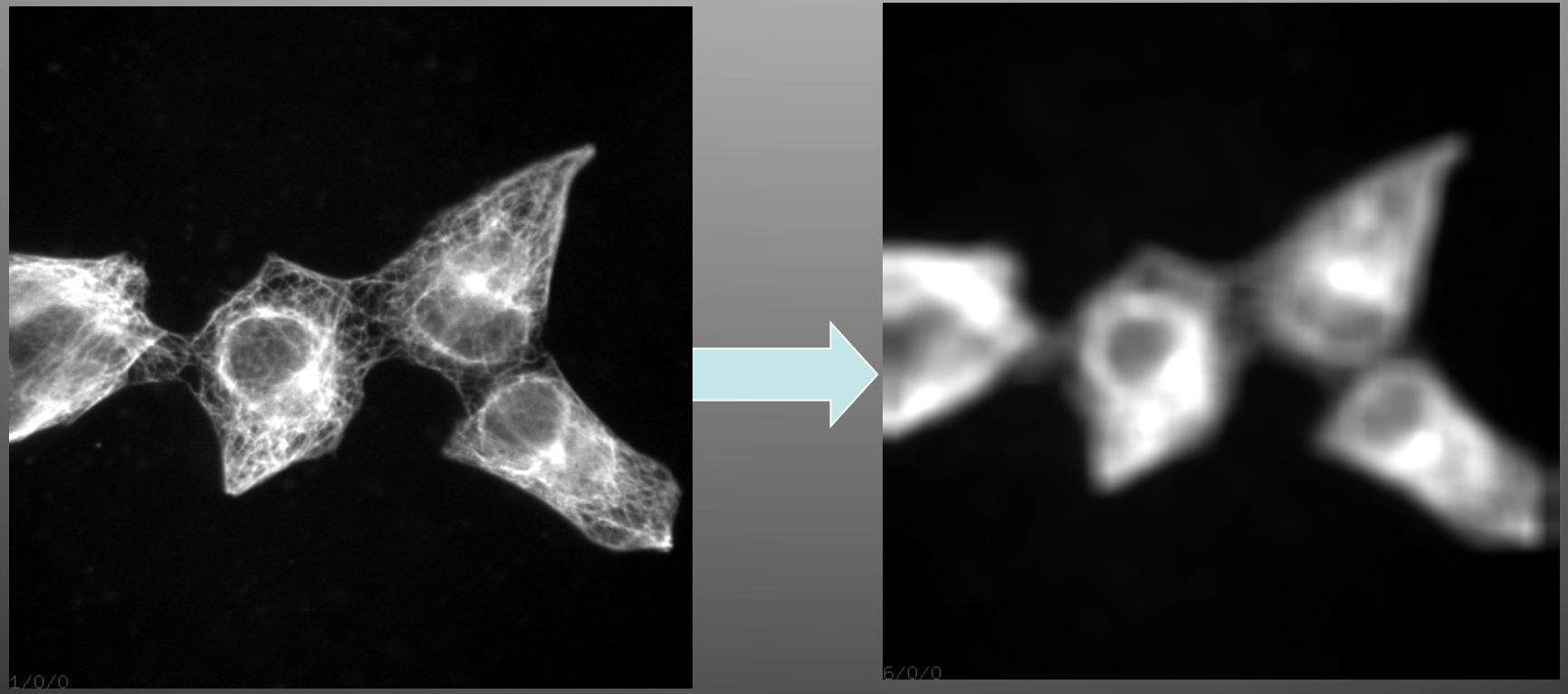
High Pass Filter



Low Pass Filter (Nuclei)



Low Pass Filter (Cells)



II. Segmentation (feature extraction)

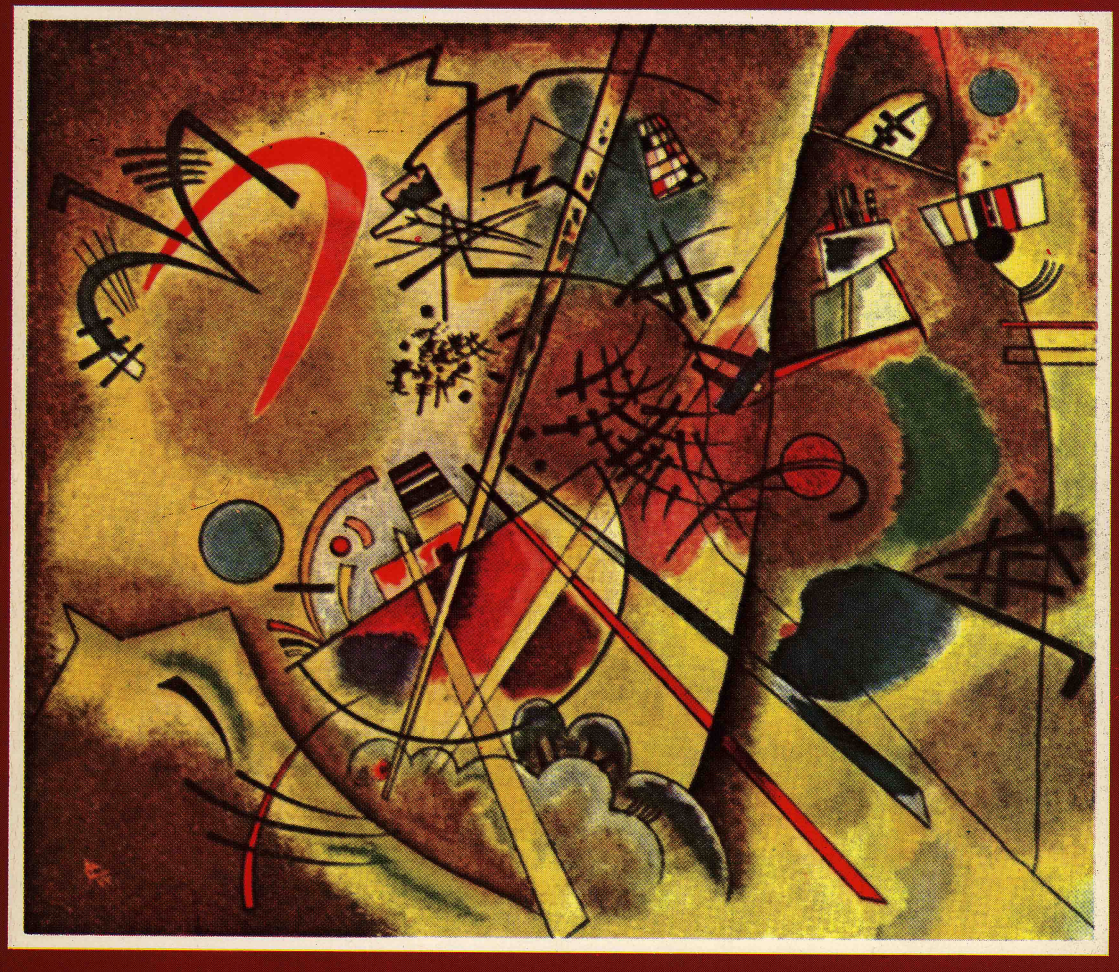
The most challenging step in the analysis!

Threshold estimate
Binary threshold
Contiguous component analysis (objects)
Watershed & SeededWaterShed
Segmentation of “points” (vesicles, MT ends, telomers)
Fiber segmentation and skeletonization
Multi-scale (pyramide) segmentation
Aggregation and Splitting of segments to define cells (e.g based on nuclei)
Thinning and cleaning, definings skeleton, outlines, ridges etc.

Segmentation

Wassily Kandinsky

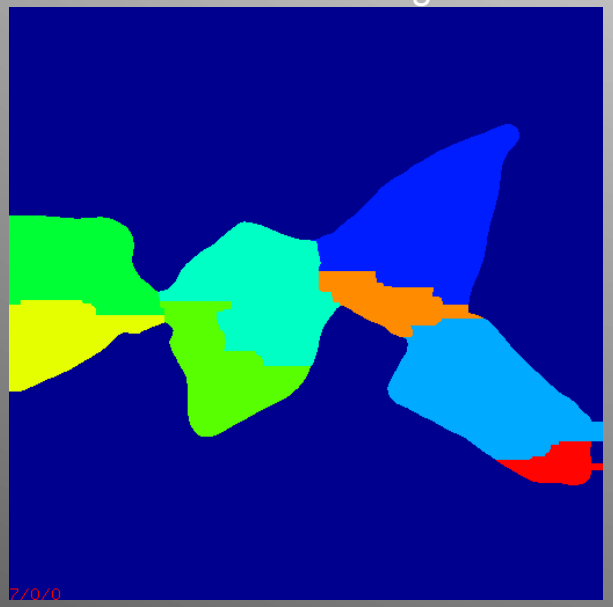
POINT AND LINE TO PLANE



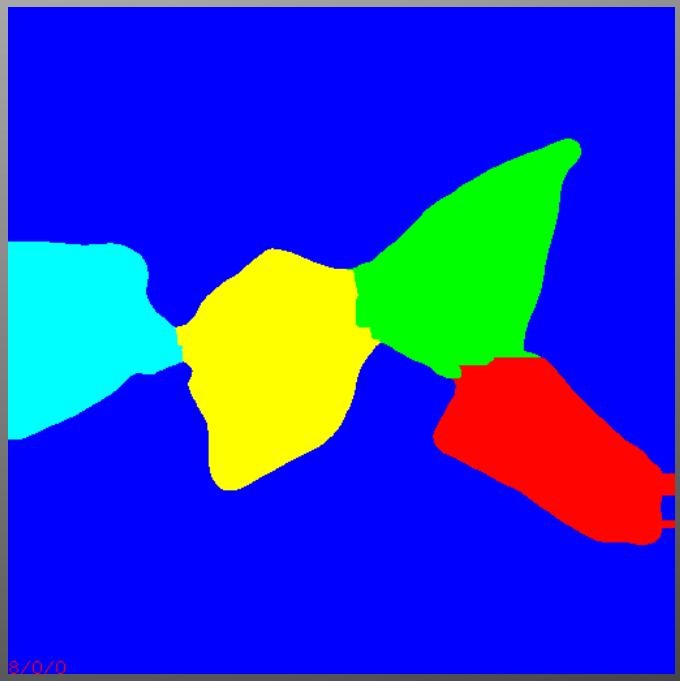
Original image. Cell coverage. Individual cells.

Labeled Masks

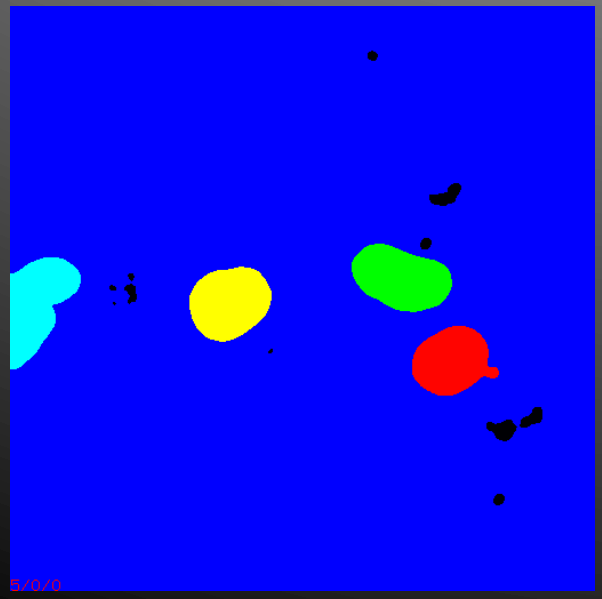
Cell mask – over segmented



Segment Merging

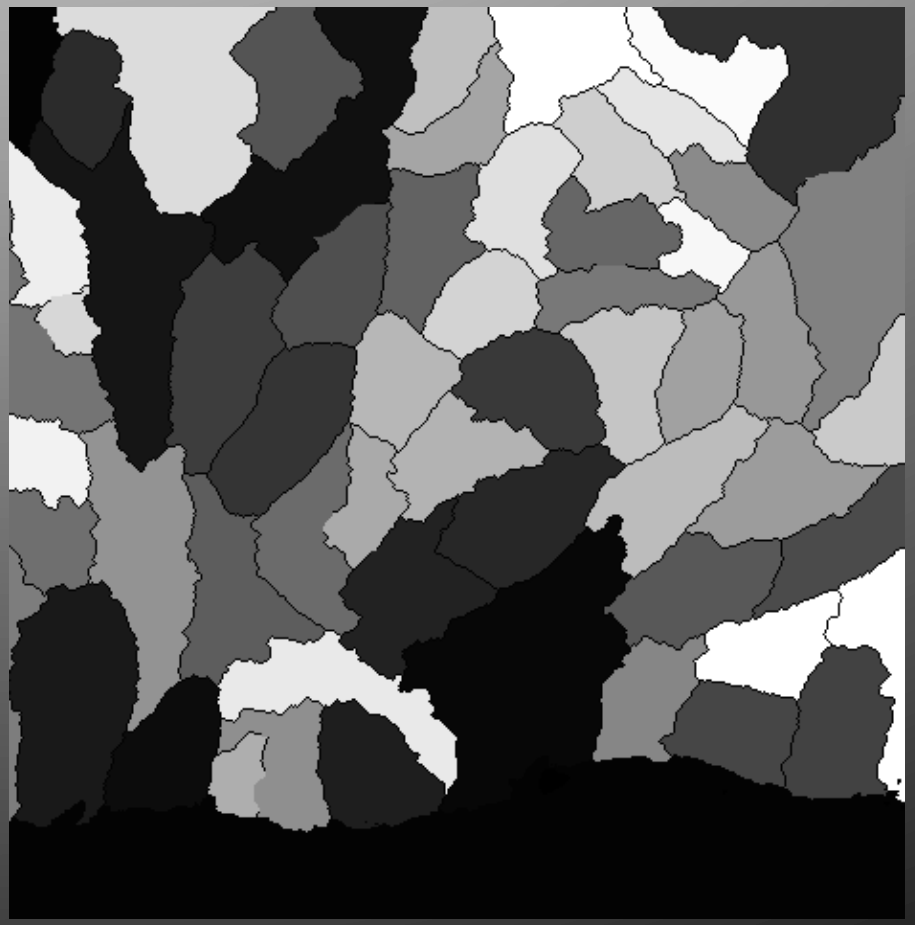


Corrected cell mask

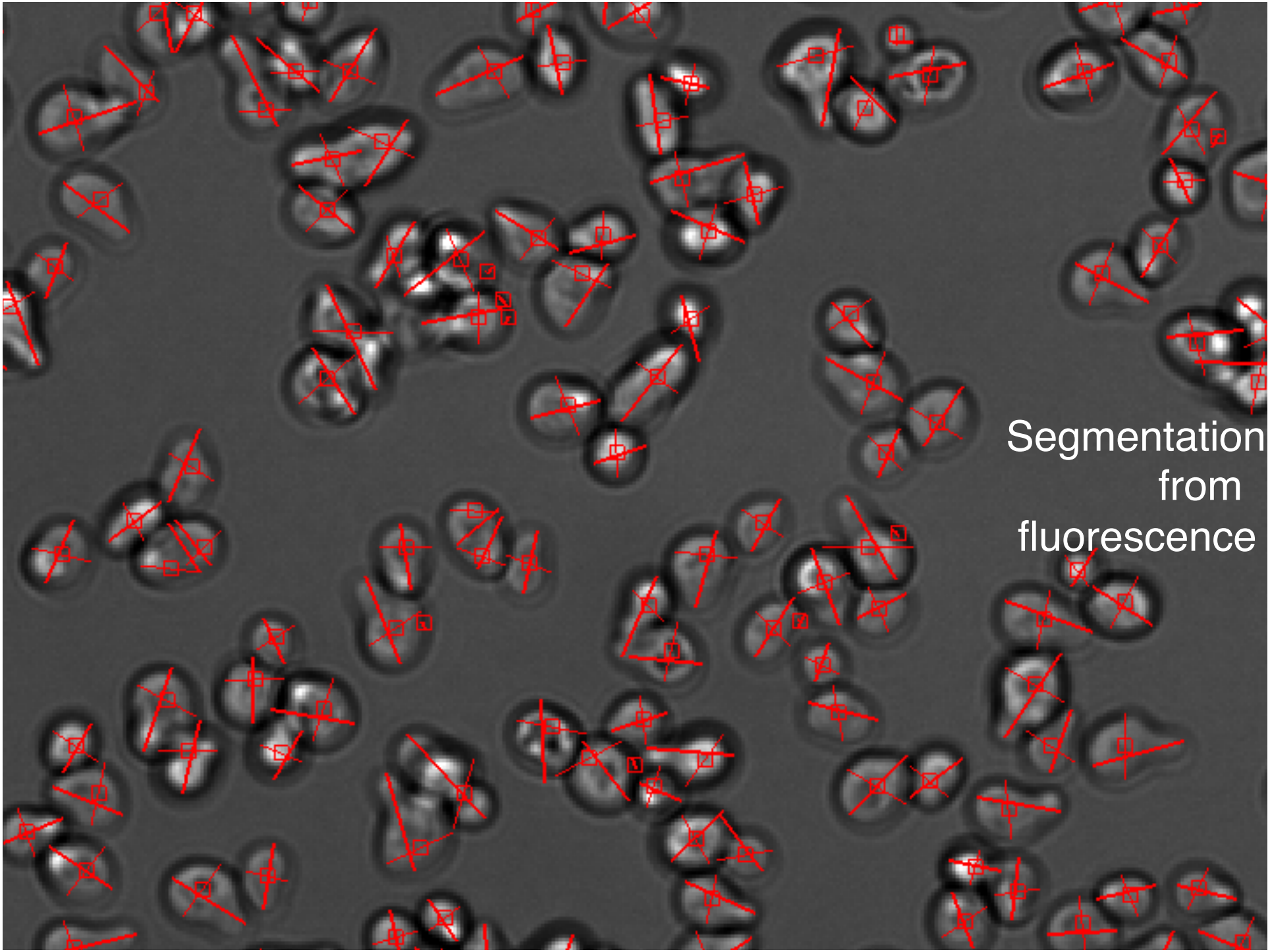


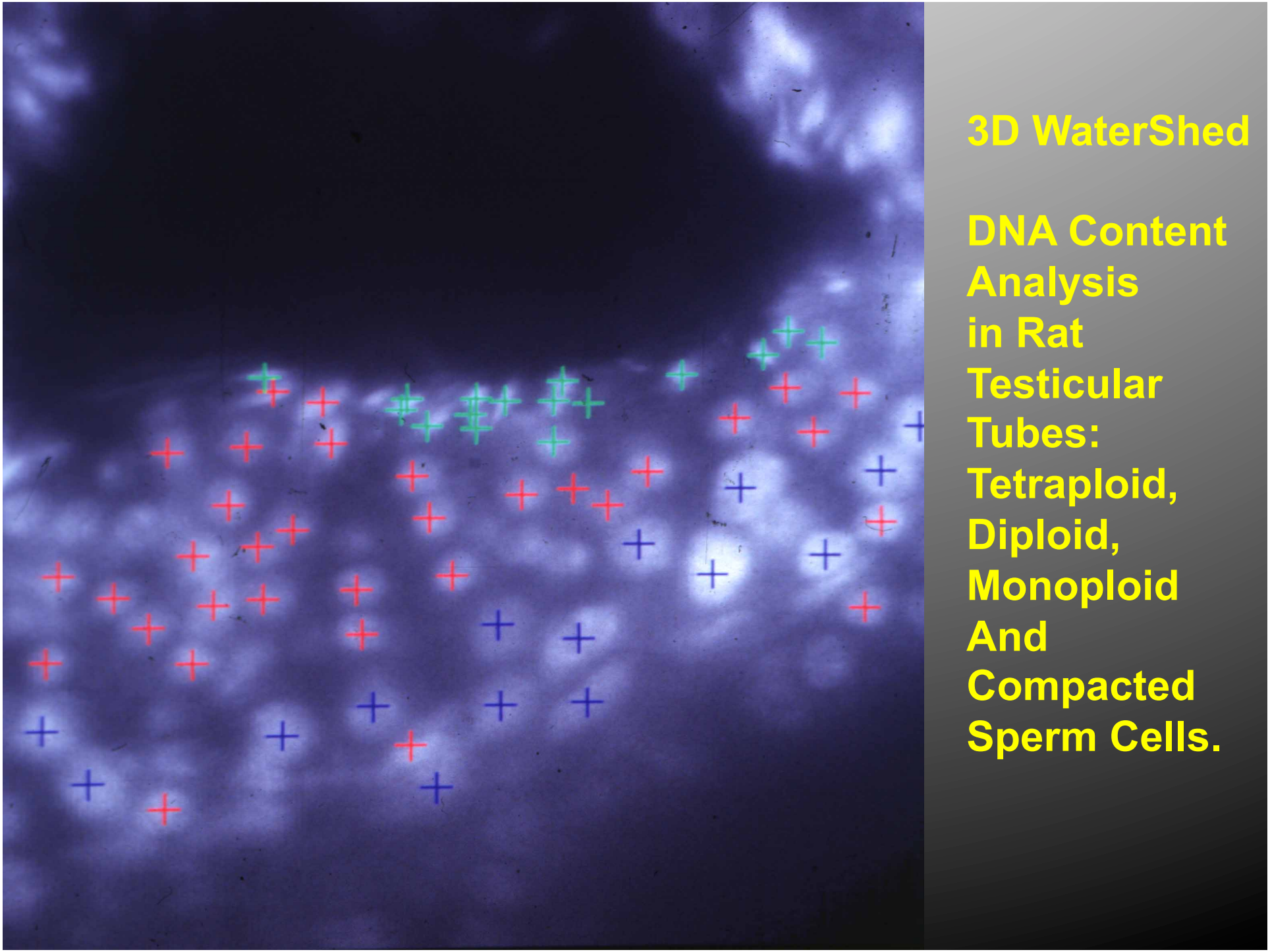
Nuclei mask

Segment Splitting

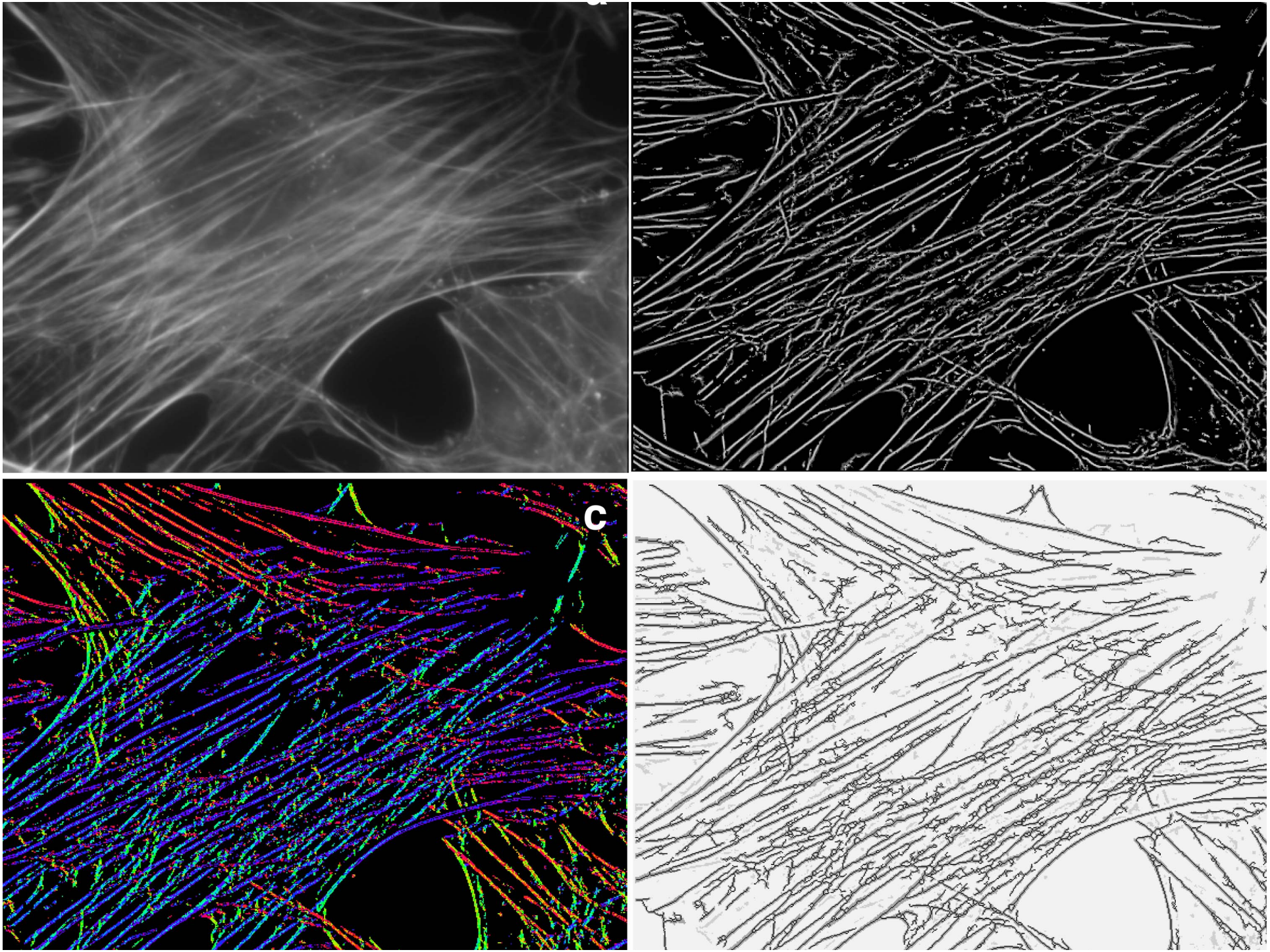


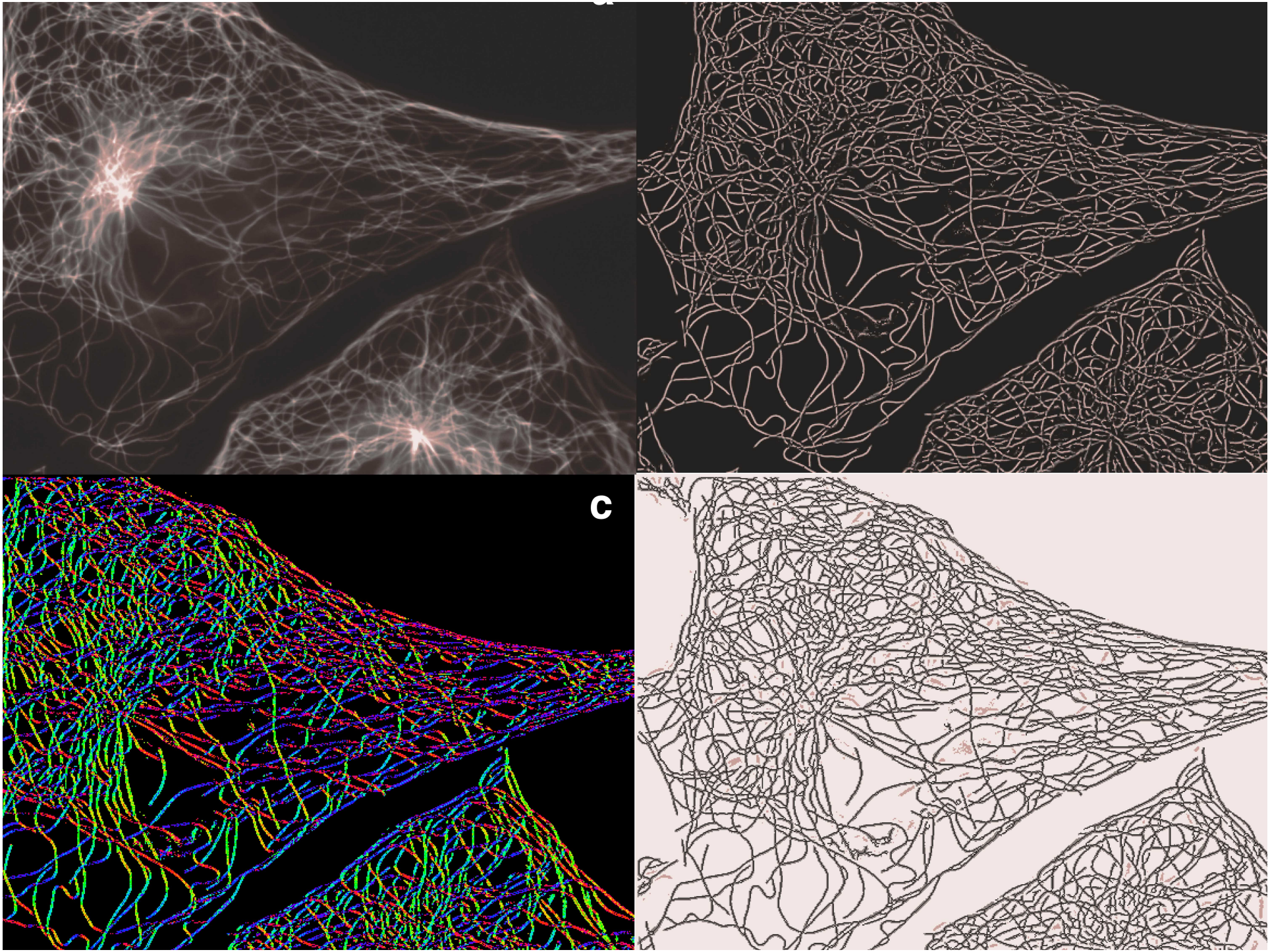
Seeded Segmentation

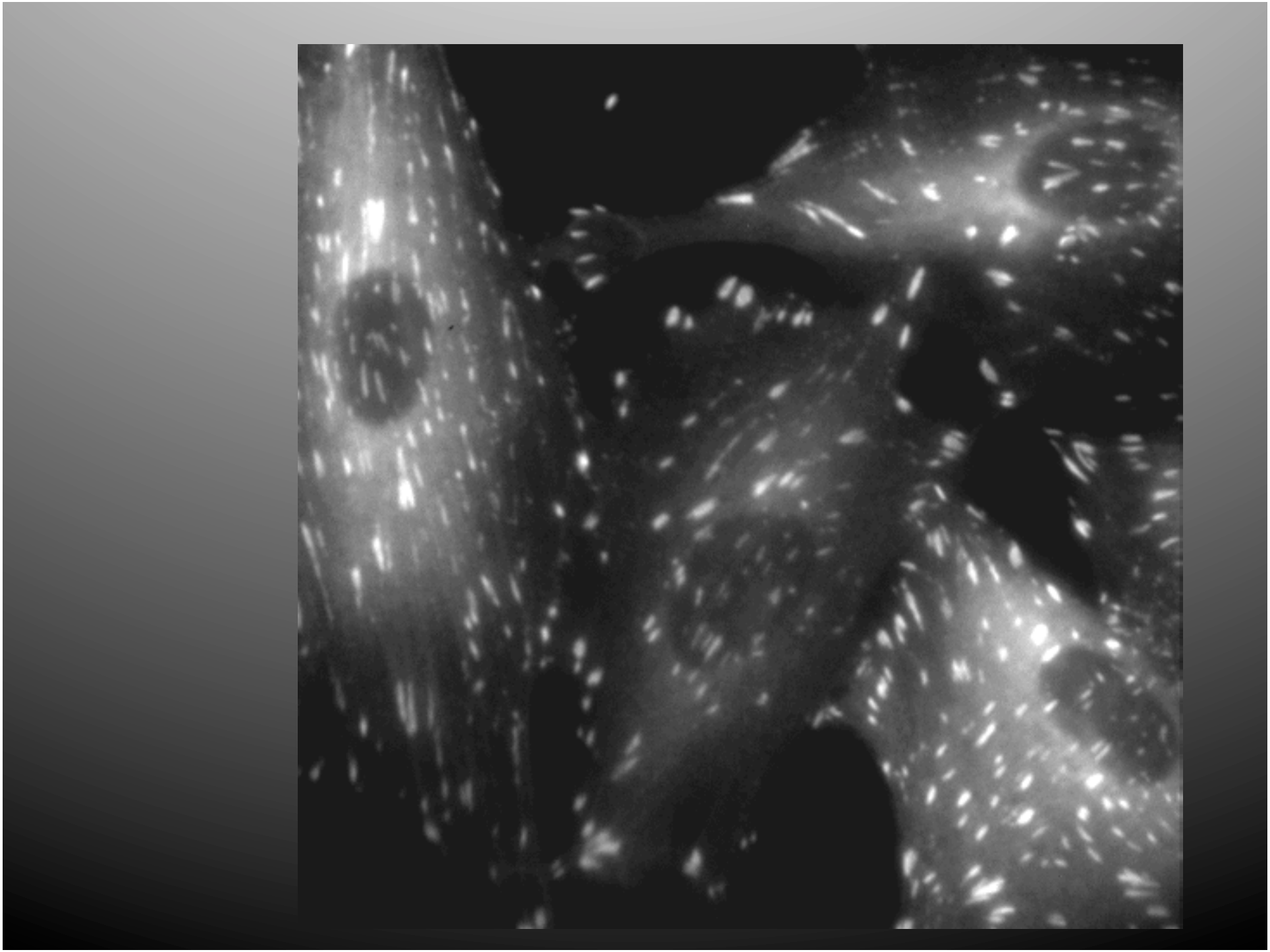

Segmentation
from
fluorescence

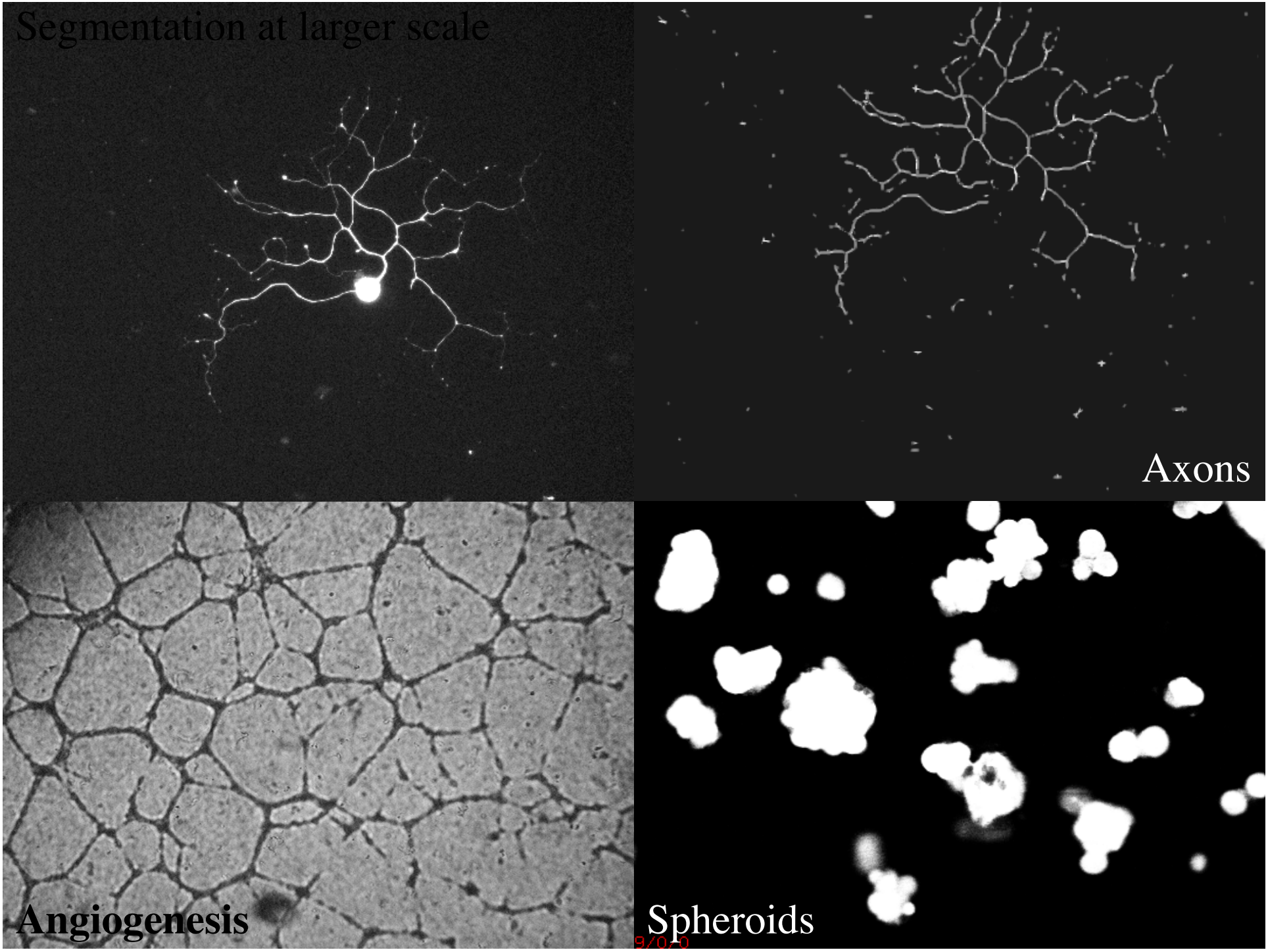

3D WaterShed
DNA Content Analysis in Rat Testicular Tubes: Tetraploid, Diploid, Monoploid And Compacted Sperm Cells.

Segmentation of Sub-cellular organelles

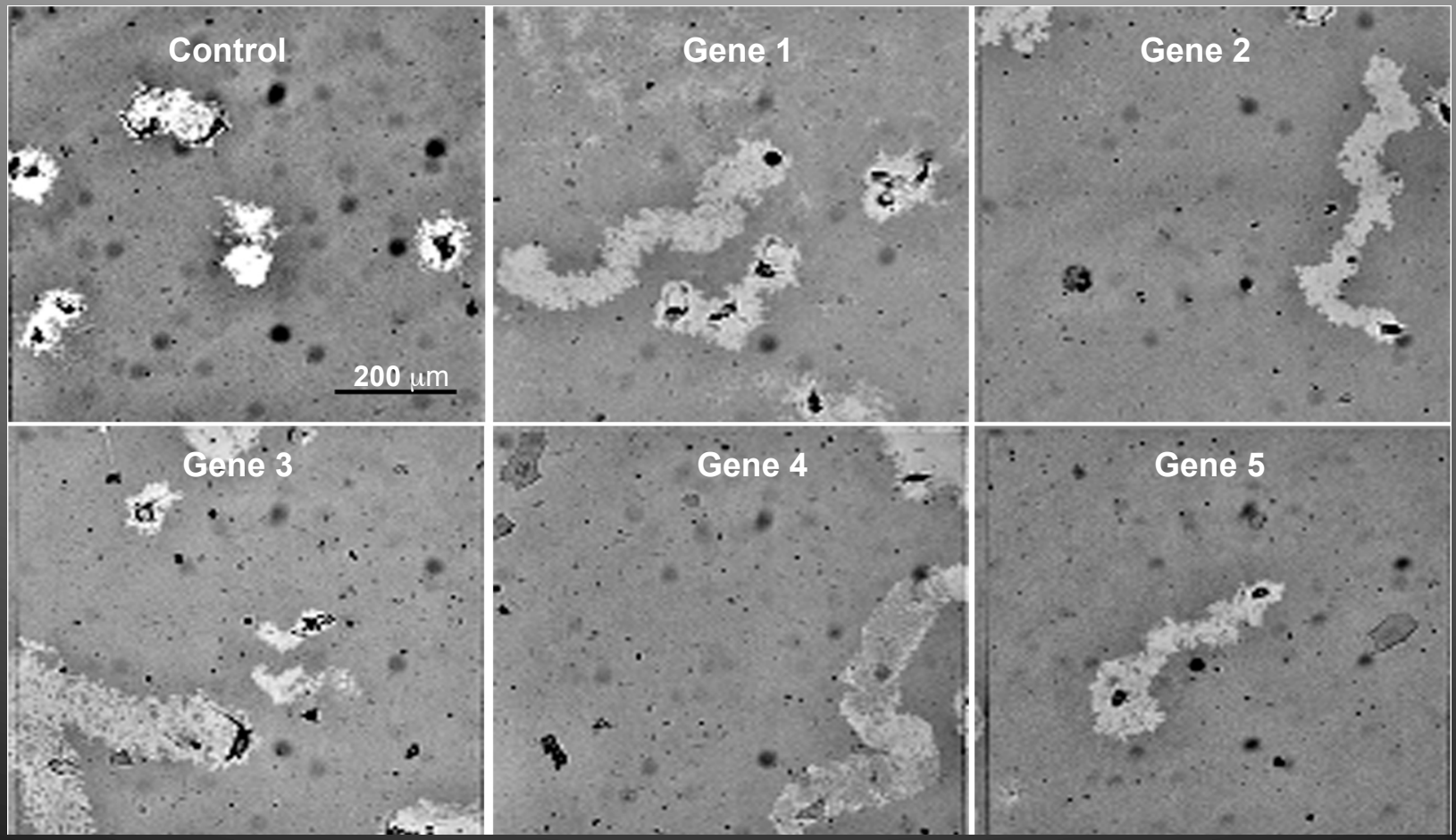

c

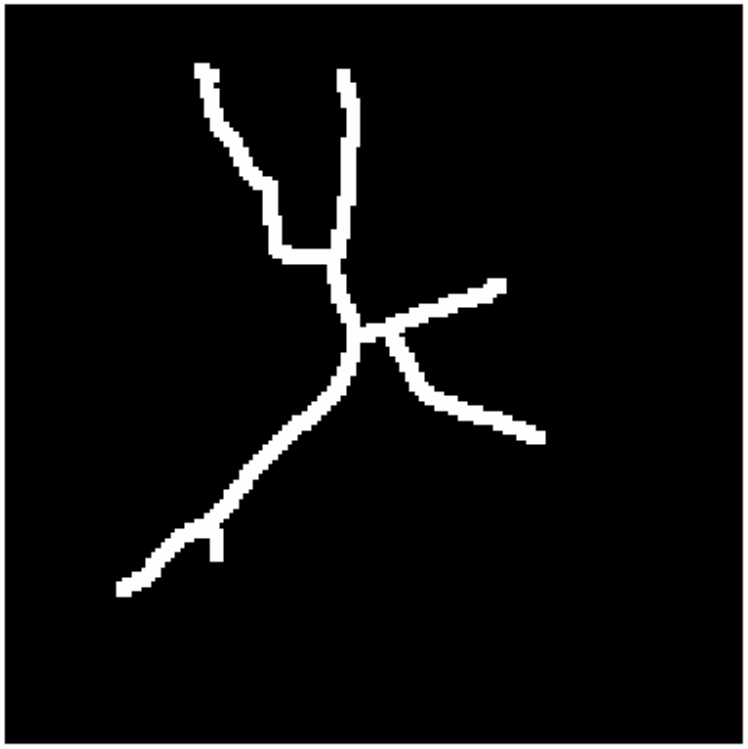

c



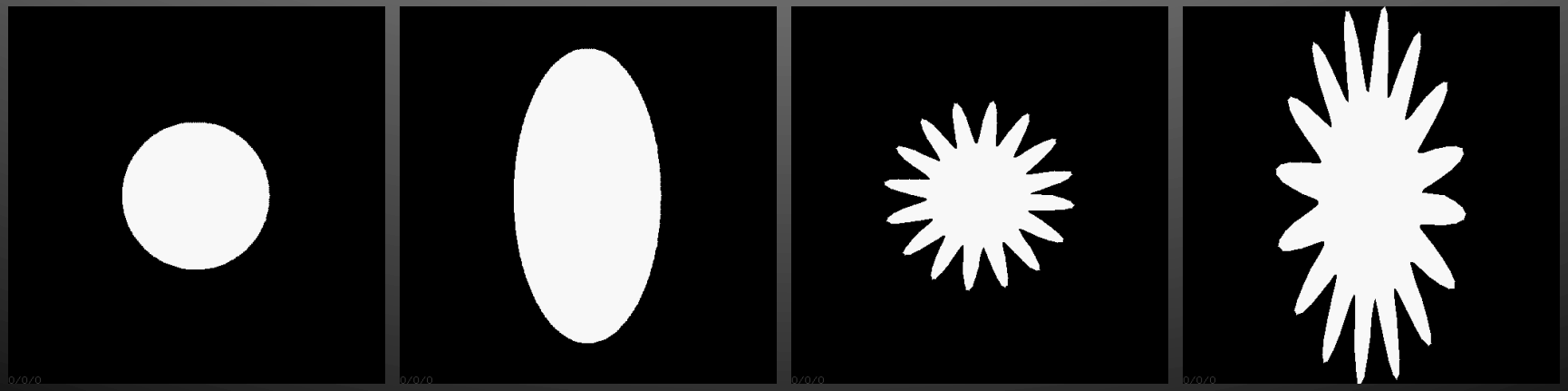

Segmentation at larger scale
Axons
Angiogenesis
Spheroids

Screening genes Affecting Cell Migration using phagokinetic tracks

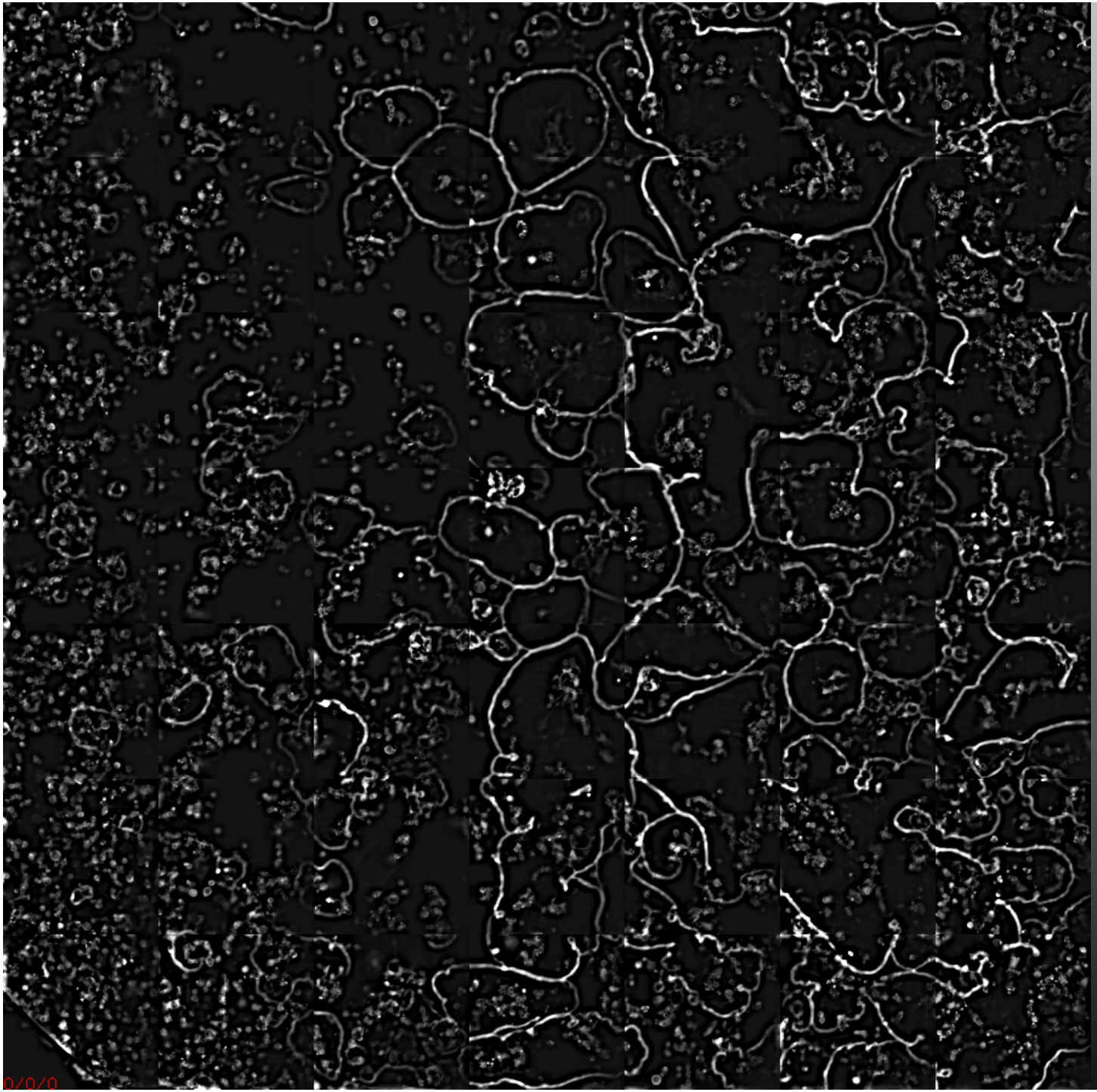




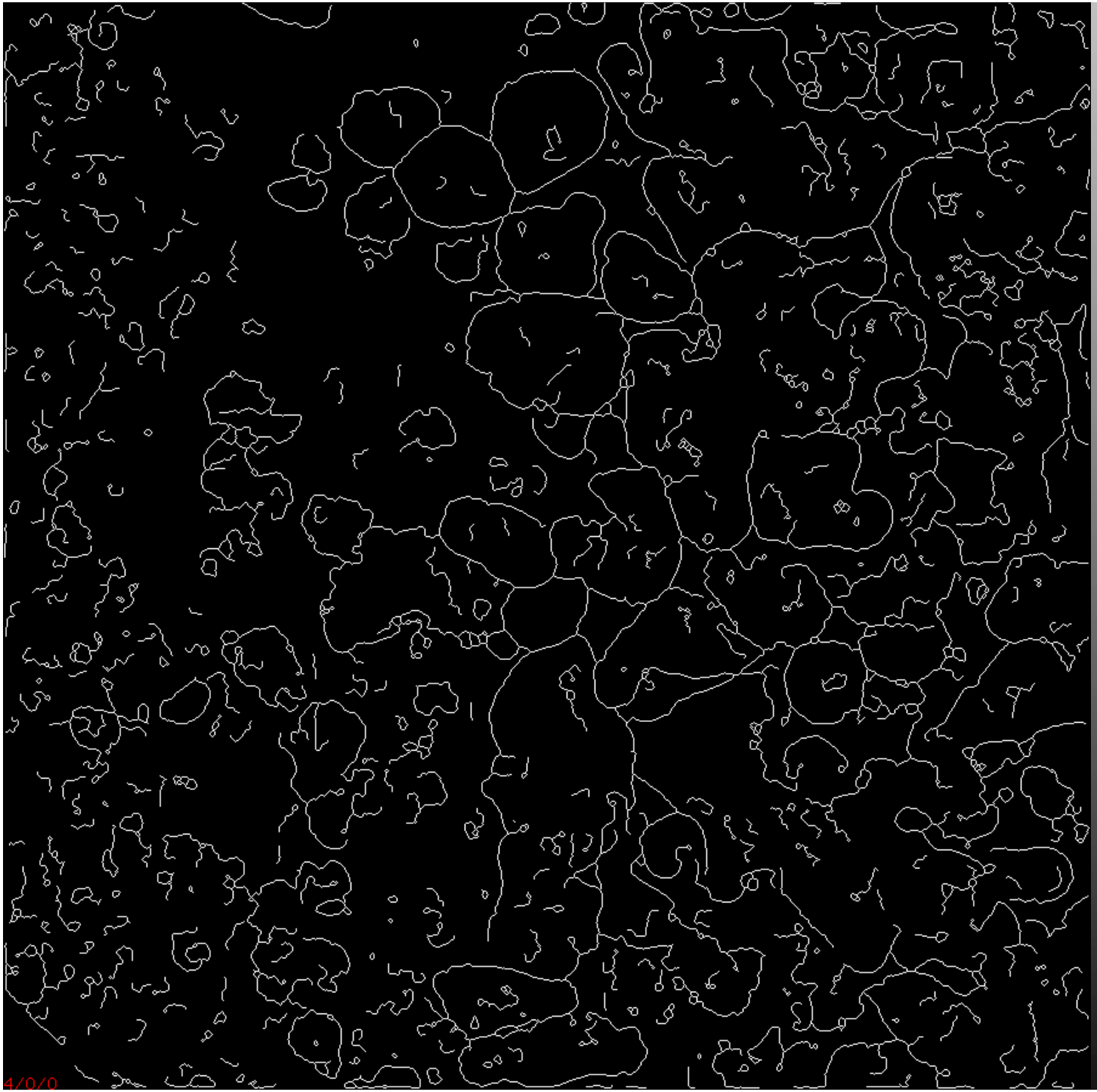
(Gorelick, Galun, Sharon, Basri & Brandt)



(G. Dunn)



Osteoclasts
Sealing-zone
formation



Osteoclasts
Sealing-zone
Attributes:
Number
Area
Perimeter
Density
Elongation
…
…
Protein conc.
in rings

III. Quantification

Morphological attributes of objects:

Center of gravity, Area, Perimeter, Extent (min/max x,y),
Best fitted ellipsoid, long and short axes.
Convex hall (area, perimeter),

Multi-color intensity attributes:

Total integrated intensity inside object, average intensity,
Background subtracted total intensity,
Intensity-weighted center of gravity and fitted ellipsoid.
Textural measures.
Pixel-by-pixel and average color intensity ratios

Tracking objects in time-lapse movies
Time-dependence of attributes
(of averages, or of tracked cell-by-cell features)
Concentration dependence of drug effects

Attributes calculated for segmented cells and sub-cellular structures

Anti-clockwise rotation angles are positive.

Mark geometrical attributes initially calculated in pixel units, and DelPix=1.

When they are converted to physical units (microns for Perimeters, square microns for areas, etc) DelPix is reassigned to pixel dimensions (microns).

% Mark Fluorescence intensity-derived attributes, initially calculated in digital image counts and setting Inorm[w]=1.

Upon normalization Inorm[w] is reset to the normalization factor for color component w, typically derived from the controls.

Basic sub-names:

Flag,SN,ID,Score

Area, Perim, CmX, CmY [or Cx,Cy?], RadGyr, Lax, Sax, Orient, AxRatio +CH,

TotInt, AvgInt, Bkgnd, TotInt_b, AvgInt_b, +1, +2

O (over), Colocal,

Ns,Y,X0,X1

Attribute Name	**Description**
<u>Global attributes for an image or set of images (e.g. montage of tiled images)</u>	
N	Number of primary objects (e.g. cells)
FracArea	Fractional objects area coverage within the image
% In,Im,Imn[w]	Minimal, Maximal and Mean fluorescent intensity for all image color components, w.
DelPix	Physical Pixel dimension (=1, reassigned after conversion of pixels to microns)
Inorm[w]	Intensity normalization factor (=1, reassigned after intensity normalization)
% Bkgnd[w]	Background intensity (averaged out of objects area)
Ns[n]	Number of Run-Length segments, k [for k=0; k<Ns[n]] for primary object n [for n=0; n<N]
CellIY, CellIX0, CellX1[n,k]	Objects region of interest Run-Length code, in image pixels

Morphological attributes for each primary object , (for n=0; n<N]

CellFlag[n]	Typically=1; Excluded object (e.g. outliers): negative Flag (-1, -2 etc. indicating exclusion criteria)
CellSN[n]	Serial object number (typically sorted by area)
CellID[n]	Object identification number, e.g. time-tracked identification number
CellICx, CellICy[n]	Geometric “center of mass” x&y image pixel coordinates
CellIXs, CellIXe,	Rectangular x-extent (start and end extreme x coordinates in pixels)
CellIYs, CellIYe[n]	Rectangular y-extent (start and end extreme y coordinates in pixels)
# CellArea[n]	Object area (in pixels, later converted to physical units)
# CellPerim[n]	Perimeter (number of boundary pixels, 4-neighbor connectivity, typically 15% less then for 8-conn)
# CellRadGyr[n]	Geomertical Radius of Gyration
# CellLax, CellSax[n]	Long [A] and short [B] axes (half diameters) of best fitted ellipsoid from geometrical 2nd moments
CellOrient [n]	Orientation Angle of long axis with image coordinate system (anti clock-wise angle with X)
CellAxRatio[n]	Axial ratio=Elongation=Aspect ratio=A/B=1/R, Ellipsoid perimeter~pA{3(1+R)-sqrt[(1+3R)*(3+R)]}
# CellAreaCH[n]	Convex Hull area (in pixels, later converted to physical units)
# CellPerimCH[n]	Convex Hull Perimeter
CellShapFact[n]	Shape factor = Perim/sqrt(pA) >1 if boundary disperse from best fitted Ellipse [Area= 4πAB]
CellSolid[n]	Solidity = sqrt(CellArea/p)/ (CellPerim/ p)
CellRound[n]	Roundness = CellArea/4pA^2
CellSmFact [n]	Smooth factor =CellArea/(4pAB) >1 if divert from ellipsoid
CellDispers[n]	Disperssion = [Graham Dunn]
CellEccent[n]	Eccentricity= sqrt(1-R^2) =0 for a sphere, ->1 for elongated ellipse
CellFormFac[n]	Form Factor= p Area/Perim^2
CellPolarity[n]	Cell Polarity range:[0,1]
CellRough[n]	Roughness = Perim/PerimCH
CellSolidCH[n]	Solidity from Convex Hull= Area/AreaCH
CellConvx[n]	Convexity= PerimCH/Perim
CellCompact[n]	Compactness= sqrt(Area/4p)/A
CellEulerNo	Euler number (number of holes)
# CellRadG[n]	Radius of gyration (Geometrical)
# CellPerimCH[n]	Convex hull perimeter
# CellDensity	Local density of neighboring cells [list methods of calculation]
CellEntropy	Local measure of organization for cell neighborhoods

Fluorescent intensity-derived primary objects attributes for each color component, w

Cm1x,Cm1y[n,w]	Fluorescent intensity weighted center of mass in image X&Y pixels
# D1x,D1y[n,w]	Distance between geometrical and intensity weighted centers
# CellRadGyr1[n]	Intensity-weighted Radius of Gyration
# CellLax1, CellSax1[n,w]	Long and short axes from intensity-weighted second moments
CellOrient1[n,w]	Angle of long axis with X coordinate
CellAxRatio1[n,w]	Axial ratio = Long/Short axes
%CellMinInt,CellMaxInt[n,w]	Minimum and Maximum intensities in the cell area
% CellTotInt1[n,w]	Total (integrated) fluorescent intensity for each labeled color
% CellAvgInt1[n,w]	Average fluorescent intensity for each labeled color
% CellBcgnd1[n.w]	Local background intensity
% CellTot_b1[n,w]	Background-subtracted total intensity
% CellAvg_b1[n,w]	Background-subtracted average intensity
% CellTxr1[n,w]	Total Textural energy (for a kernel) in object (e.g. Haralick, Zernike, Variance)
% CellTxrOArea1[n,w]	Textural energy per unit area
CellTxrOint1[n,w]	Textural energy per average intensity
CellZernk[n,i]	Zernike moment
CellHaralick[n,i]	Haralick textural coefficients, 14 for each resolution x number of resolutions (R.Murphy)

Fluorescent intensity-derived primary (cytoplasm) to secondary (nucleus) objects intensity ratios etc

TotNucOCyt[n,w]	Total fluo intensity Nucleus to Cytoplasm ratio = TotInt2[n,w,0]/(TotInt1[n,w]- TotInt2[n,w,0])
TotNucOCyt_b[n,w]	Background-subtracted Total fluorescence intensity Nucleus to Cytoplasm ratio
TotNucOCyt[n,w]	Total fluorescence intensity Nucleus to Cytoplasm ratio
TotNucOCyt_b[n,w]	Background-subtracted Total fluorescence intensity Nucleus to Cytoplasm ratio
AvgNucOCyt[n,w]	Average fluorescence intensity Nucleus to Cytoplasm ratio
AvgNucOCyt_b[n,w]	Background-subtracted Average fluorescence intensity Nucleus to Cytoplasm ratio
Colocal[n,w1,w2]	Colocalization factors of two colors

Attributes for cytoskeleton morphology and associated fluorescence intensity

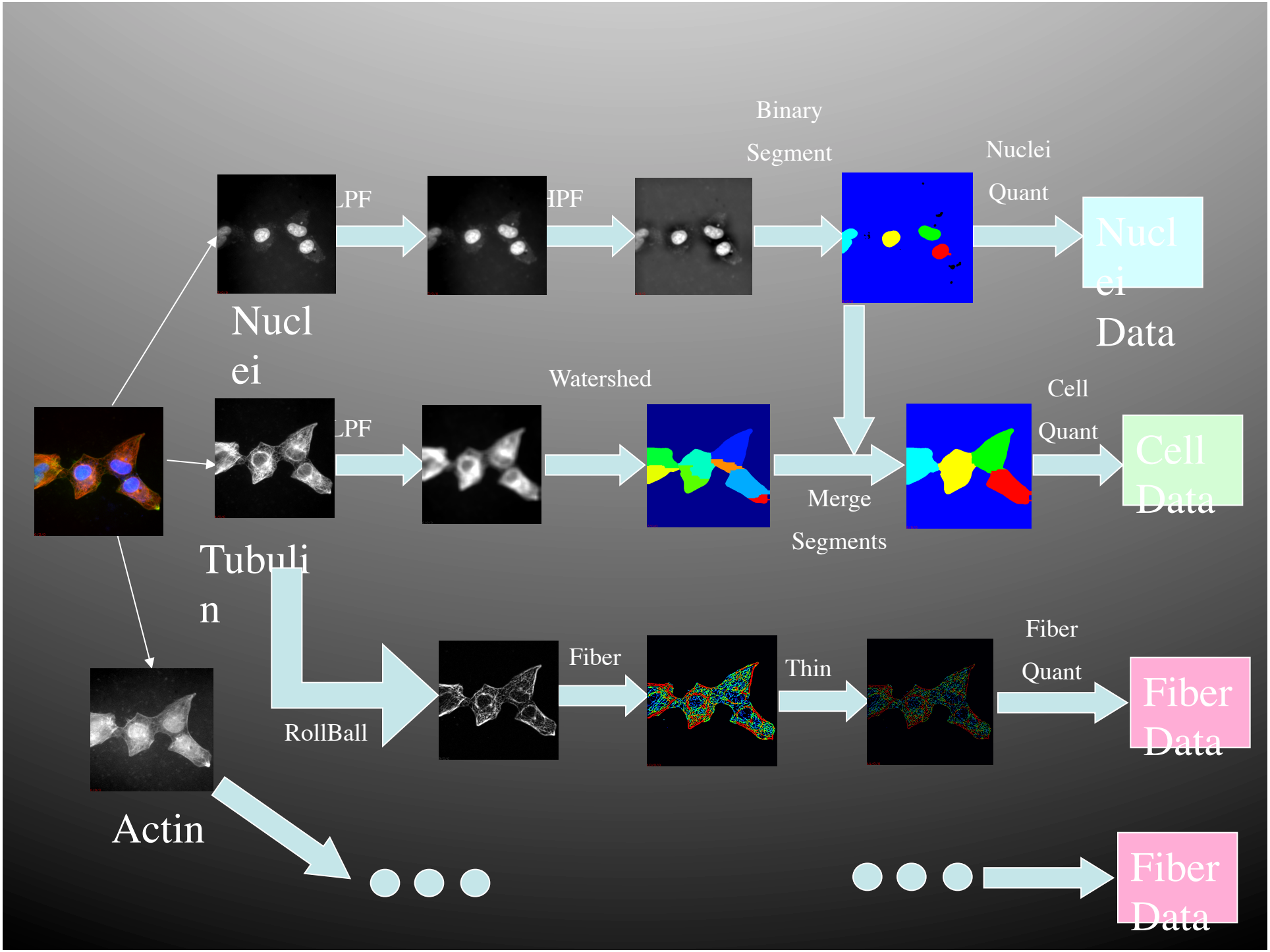
# FibArea[n]	Fibers area
# FibLen[w]	Total fibers length
% TotFib[w]	Fiber-associated total intensity
% TotFib_b[w]	Background-subtracted Fiber-associated total intensity
% AvgFib_b[w]	Background-subtracted Fiber-associated average intensity
% FibOLen[w]	Background-subtracted Fiber-associated Intensity per unit length
FibPol[w]	Fiber Polarization Factor =

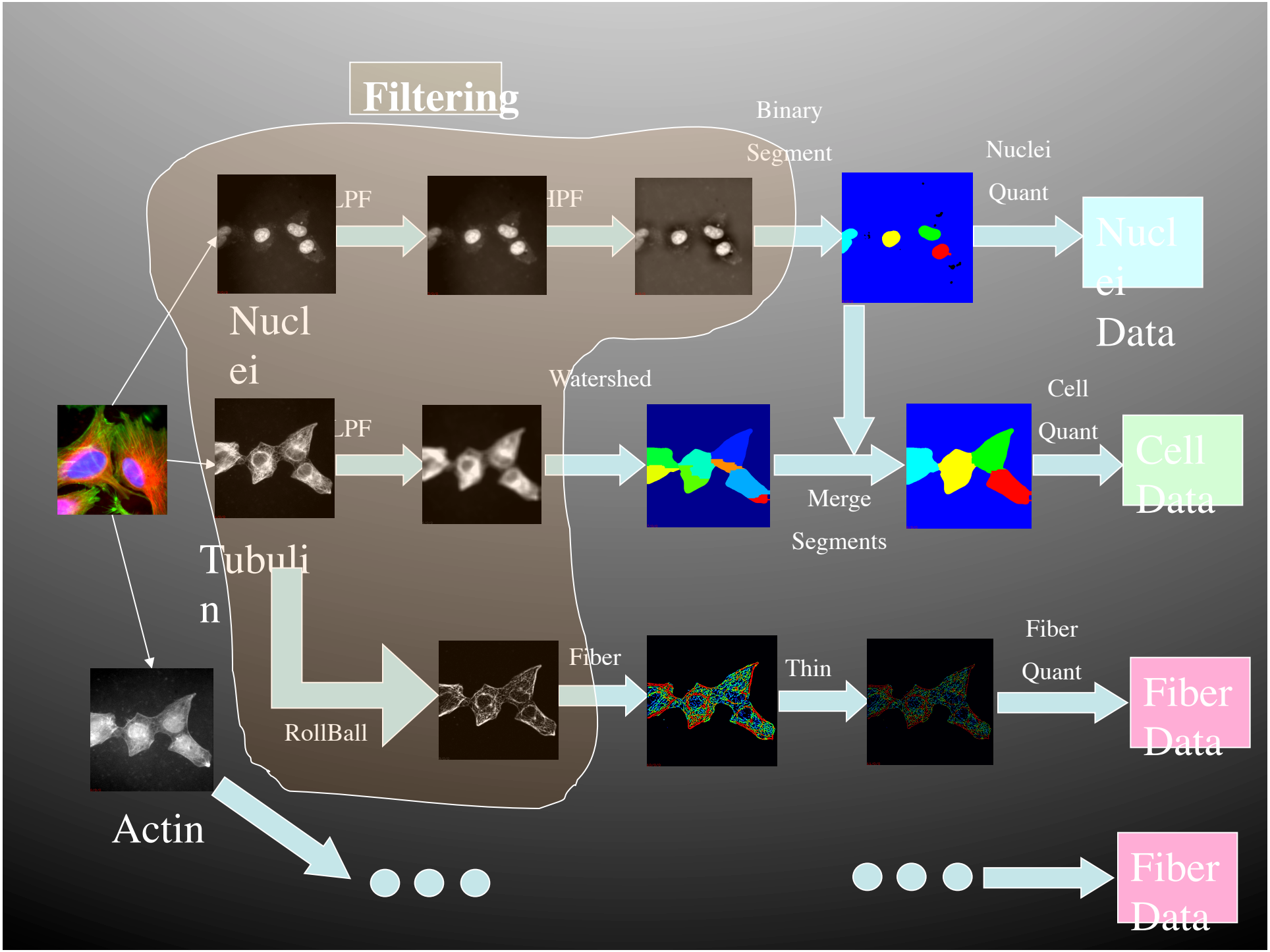
Attributes for cytoskeleton morphology and associated fluorescence intensity

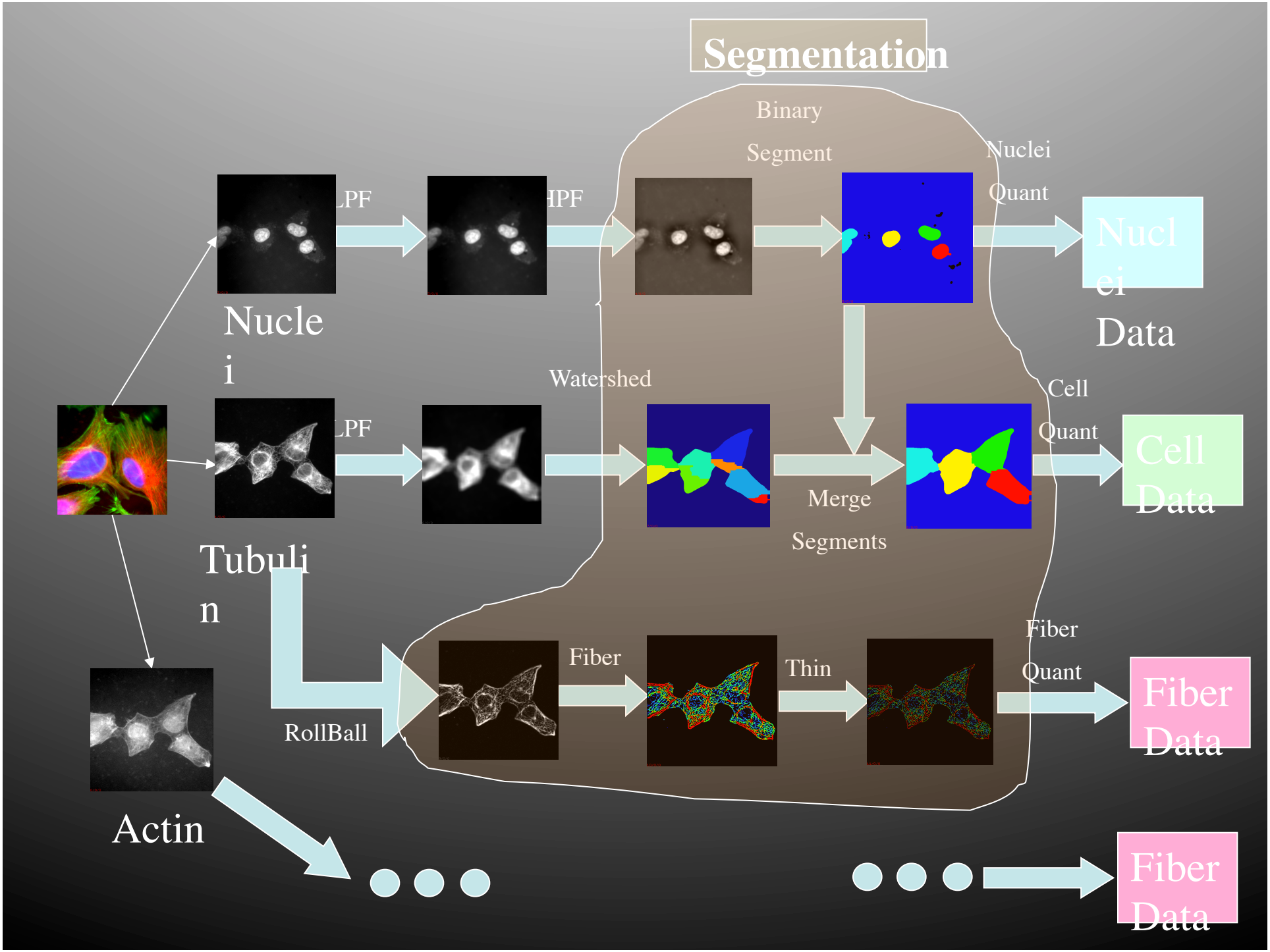
# FibArea[n]	Fibers area
# FibLen[w]	Total fibers length
% TotFib[w]	Fiber-associated total intensity
% TotFib_b[w]	Background-subtracted Fiber-associated total intensity
% AvgFib_b[w]	Background-subtracted Fiber-associated average intensity
% FibOLen[w]	Background-subtracted Fiber-associated Intensity per unit length
FibPol[w]	Fiber Polarization Factor =

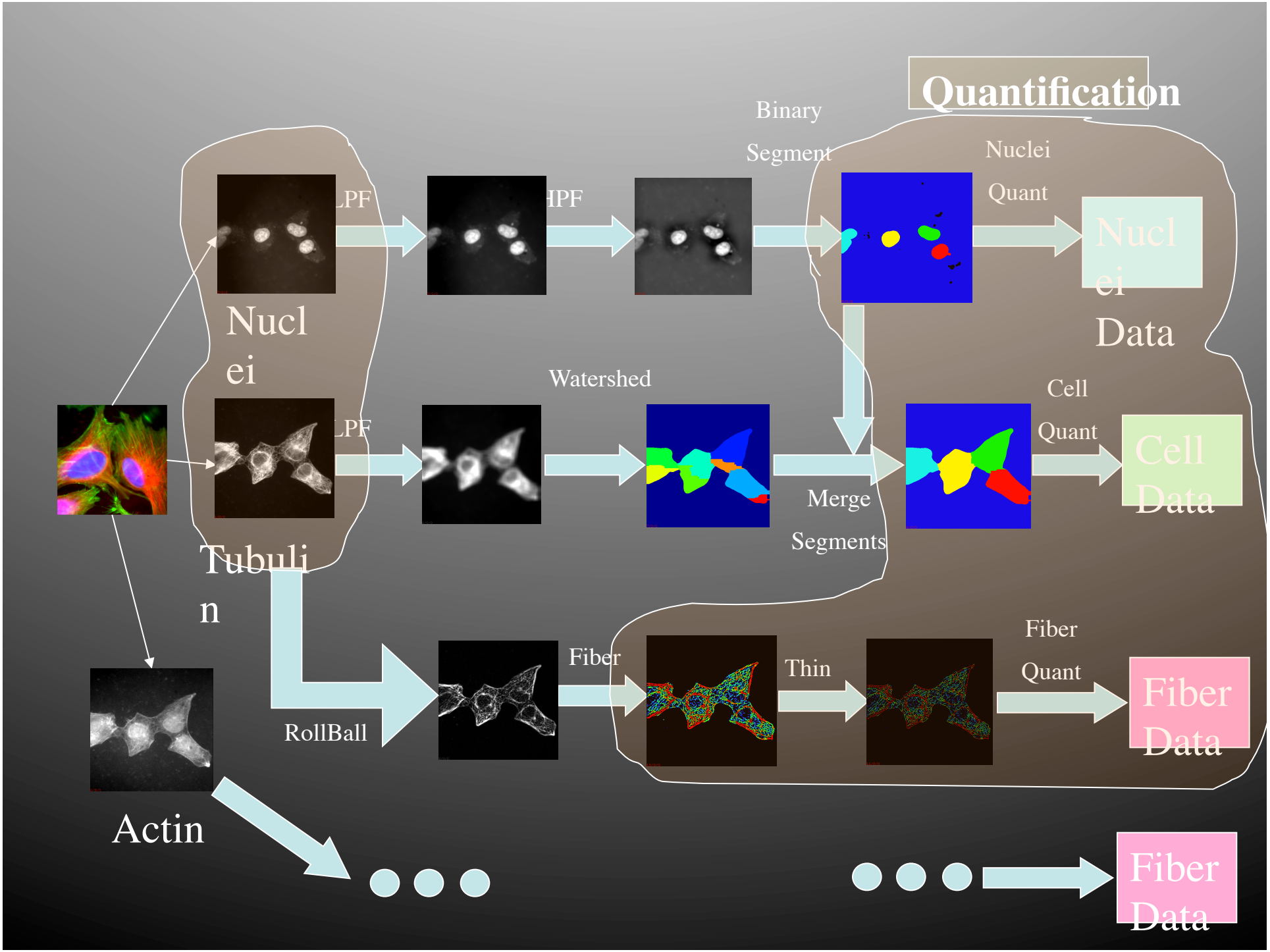
Attributes for secondary objects(organelles , m=0 ..<M) in primary object (cell number n) color w
(e.g. Nucleus, Focal Adhesions, Golgi, Mitochondria, Endoplasmic). Statistics is accumulated for each primary object.

M[n]	Number of secondary objects per primary object
# C2x,C2y[n,w,m]	Geometric "centers of mass" x&y coordinates (in pixels)
# D2x,D2y[n,w,m]	Distance between intensity weighted centers of secondary and primary objects
X2s,X2e,Y2s,Y2e[n,w,m]	Rectangular extent (start and end extreme x&y coordinates in pixels)
# Area2[n,w,m]	Secondary Objects area
# Perim2[n,w,m]	Secondary Objects Perimeter
# Lax2, Sax2[n,w,m]	Long and short axes (half diameters) of best fitted ellipsoid from geometrical second moments
Orient2[n,w,m]	Angle of long axis with respect to image coordinate system (anti clock-wise angle with X)
AxRatio2[n,w,m]	Axial ratio = Lax2/Sax2
rAng2[n,w,m]	Angle between long axis and line connecting secondary object center to primary center
pAng2 [n,w,m]	Angle with respect to cell polarity axis
# Dp2[n,w,m]	Shortest distance to primary object boundary
NDE2[n,w,m]	Normalized distance between cell edge and cell center = Dp2/(Dp2+Dc2)
# OrgArea[n]	Total area of secondary objects in each primary objects
# OrgPerim[n]	Total perimeter for secondary objects in each primary objects
# OrgRadGyr[n]	Intensity-weighted Rad Gyr for all secondary object segments in a cell =organelle compactness
# OrgDx,OrgDy[n]	Distance between secondary objects center and cell center
% OrgMinInt,OrgMaxInt[n,w]	Minimum and Maxumum intensities in the cell area
% OrgTotInt[n,w]	Total (integrated) fluorescent intensity for each labeled color
% OrgAvgInt[n,w]	Total (integrated) fluorescent intensity for each labeled color
% OrgBcgnd[n.w]	Local background intensity [from dilated masks, extent out of segment etc]
% OrgTot_b[n,w]	Background-subtracted total intensity
% OrgAvg_b[n,w]	Background-subtracted average intensity
% TotInt2[n,w,m]	Total (integrated) fluorescent intensity for each labeled color
% AvgInt2[n,w,m]	Total (integrated) fluorescent intensity for each labeled color
% Bcgnd2[n.w,m]	Local background intensity
% Tot_b2[n,w,m]	Background-subtracted total intensity
% Avg_b2[n,w,m]	Background-subtracted average intensity
% Txr2[n,w,m]	Total Textural energy (for a textural kernel and scale) in objects (e.g. Haralick, Zernike, Variance)
% TxrOArea2[n,w,m]	Textural energy per unit area
TxrOInt2[n,w,m]	Textural energy per average intensity


Binary
Segment
Nuclei
Quant
LPF
HPF
Nucl
ei
Nucl
ei
Data
Watershed
Cell
Quant
LPF
Cell
Data
Merge
Segments
Tubuli
n
Fiber
Quant
Thin
Fiber
RollBall
Fiber
Data
Actin
Fiber
Data


Filtering
Binary
Segment
Nuclei
Quant
LPF
HPF
Nucl
ei
Nucl
ei
Data
Watershed
Cell
Quant
LPF
Cell
Data
Merge
Segments
Tubuli
n
Fiber
Quant
Thin
Fiber
Data
Fiber
RollBall
Actin
Fiber
Data


Segmentation
Binary
Segment
Nuclei
Quant
LPF
HPF
Nucle
i
Nucl
ei
Data
Watershed
Cell
Quant
LPF
Cell
Data
Merge
Segments
Tubuli
n
Fiber
Quant
Fiber
Thin
RollBall
Fiber
Data
Actin
Fiber
Data


Quantification
Binary
Segment
Nuclei
Quant
LPF
HPF
Nucl
ei
Nucl
ei
Data
Watershed
Cell
Quant
LPF
Cell
Data
Merge
Segments
Tubuli
n
Fiber
Quant
Fiber
Thin
RollBall
Fiber
Data
Actin
Fiber
Data

The quantified attributes for each segmented cell, and statistical summery.

Data --------- t=1 well [8,2] -----------

Obj#	1:Flag	2:ID	3:CM x	4:CM y	5:Nuc Area	6:Nuc Perim	7:Nuc TotInt	8:Nuc Bck Int	9:Nuc TotInt-Bck	10:Nuc AvgInt-Bck
Obj(1):	0.50	1.00	482.91	284.84	3000.00	0.00	21711800.00	1081.20	18468214.00	6156.07
Obj(2):	0.50	2.00	504.85	426.09	1767.00	0.00	8259100.00	1081.20	6348628.00	3592.89
Obj(3):	1.00	3.00	437.14	407.42	1435.00	0.00	5999945.00	1081.20	4448430.00	3099.95
Obj(4):	1.00	4.00	408.92	406.38	1091.00	0.00	4058477.00	1081.20	2878893.00	2638.77
Obj(5):	1.00	5.00	444.67	240.75	1060.00	0.00	6121366.00	1081.20	4975299.00	4693.68
Obj(6):	1.00	6.00	464.22	256.76	987.00	0.00	6898878.00	1081.20	5831738.50	5908.55
Obj(7):	1.00	1.00	87.35	473.15	2511.00	0.00	14955397.00	998.77	12447481.00	4957.18
Obj(8):	0.50	2.00	57.06	497.60	1860.00	0.00	10360923.00	998.77	8503207.00	4571.62
Obj(9):	0.50	1.00	18.39	262.55	2256.00	0.00	10496238.00	1137.73	7929519.00	3514.86
Obj(10):	1.00	2.00	71.41	138.73	2124.00	0.00	9361633.00	1137.73	6945094.50	3269.82

. . . . . .

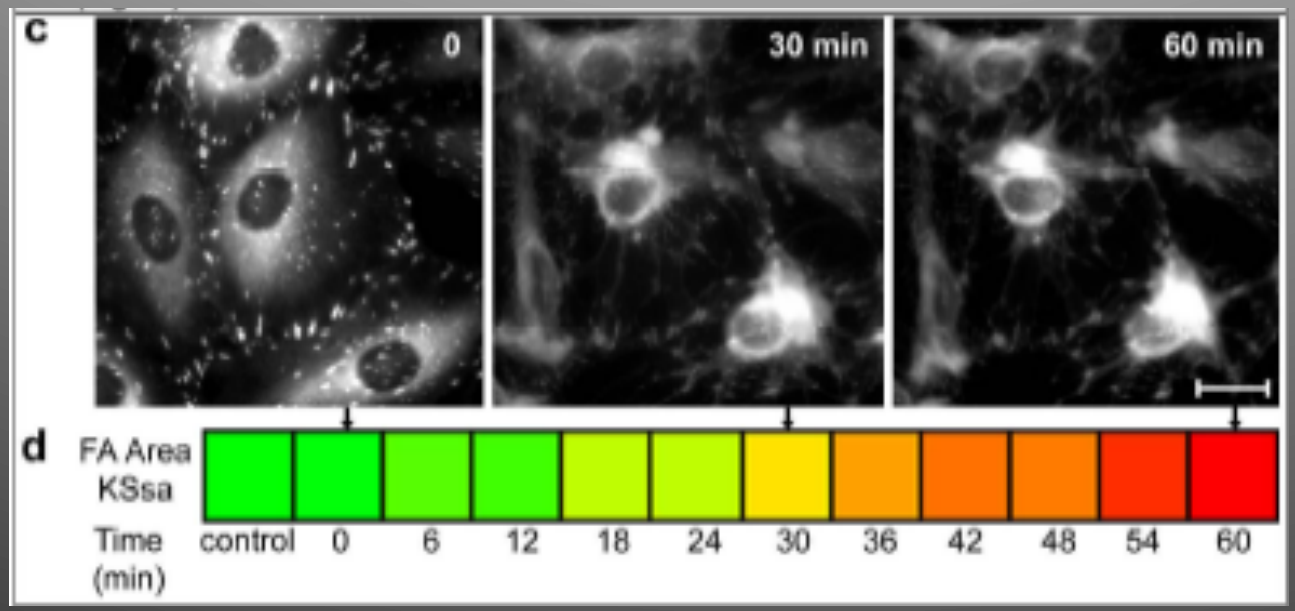
Sum'ry	Flag	ID	CM x	CM y	Nuc Area	Nuc Perim	Nuc TotInt	Nuc Bck Int	Nuc TotInt-Bck	Nuc AvgInt-Bck
min	0.50	1.00	4.45	4.30	301.00	0.00	646614.00	968.76	231931.34	770.54
max	1.00	46.00	509.24	508.82	3850.00	0.00	29520566.00	1546.80	25924334.00	10587.49
sum	1970	2.428e+04	5.725e+05	5.794e+05	2.474e+06	0	1.438e+10	2.679e+06	1.145e+10	1.022e+07
avgC	0.88	10.80	254.68	257.72	1100.52	0.00	6398129.67	1191.94	5093010.74	4546.99
avgF	1.00	12.33	290.62	294.09	1255.83	0.00	7301012.95	1360.14	5811719.87	5188.65
stdvC	0.22	7.84	153.17	156.95	479.23	0.00	3490259.00	122.93	2989047.25	1274.58
stdvF	0.90	13.35	297.19	301.75	1200.34	0.00	7288207.50	1198.26	5905350.00	4722.26
perc10	0.50	2.00	36.95	37.97	612.00	0.00	3391689.00	1064.37	2562691.00	3044.01
perc90	1.00	22.00	468.49	475.22	1776.00	0.00	10615037.00	1358.64	8697990.00	6252.31

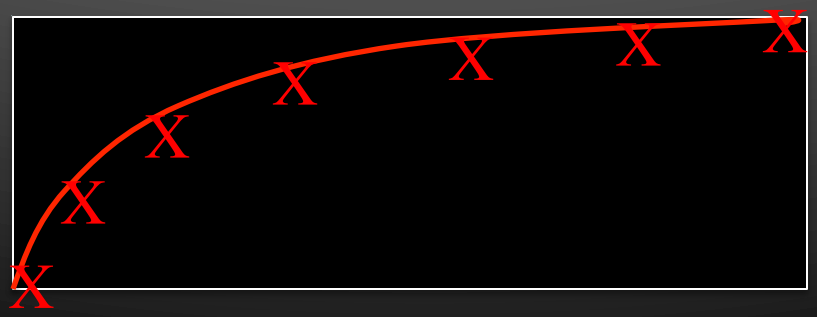
count=2248 sum flags=1970.0

TeraBytes of images -> MegaBytes of data

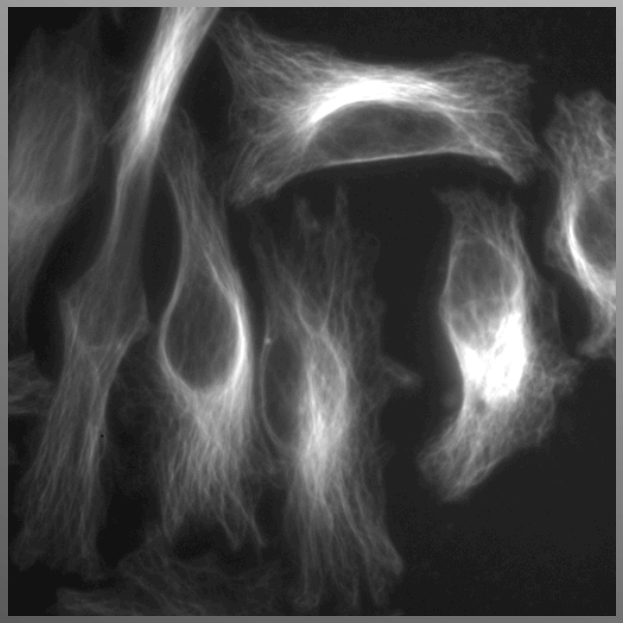
Multi-sample analysis:
Time dependent effects

Averaged (fixed preps) or cell-by-cell properties
(live time-tracking)

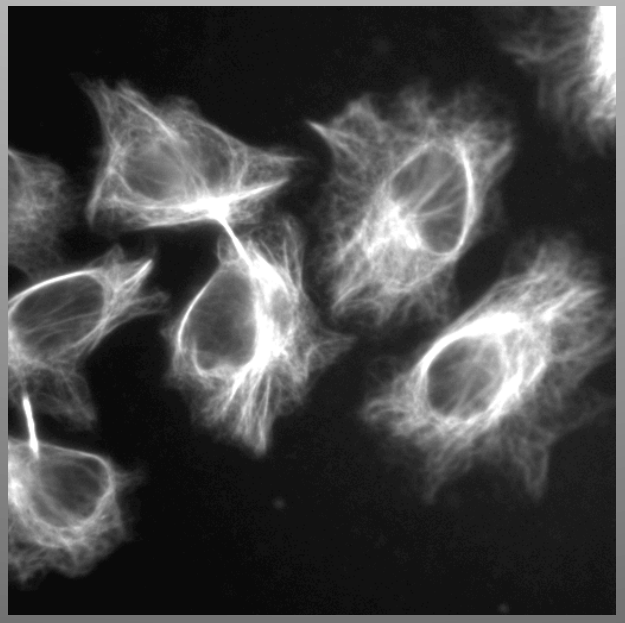




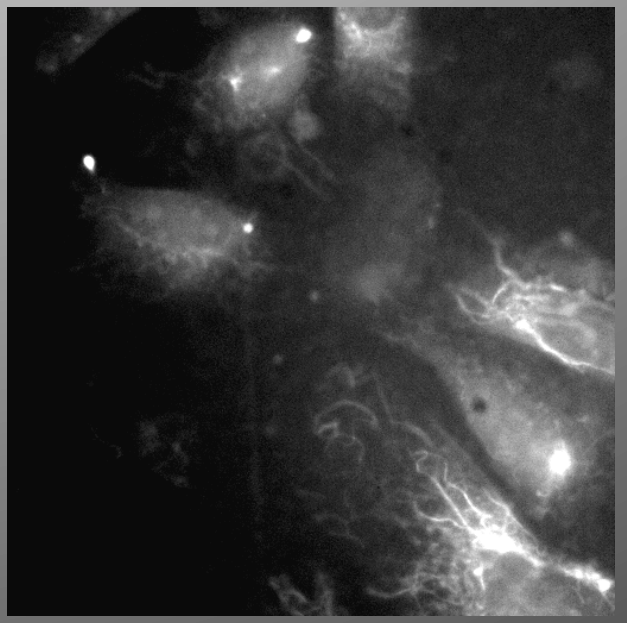
Multi-sample analysis: Concentration-dependent effects



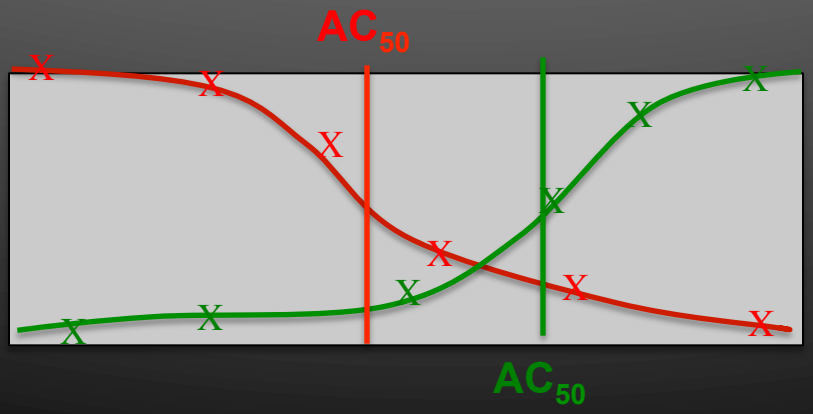
Control (DMSO 1.2%)



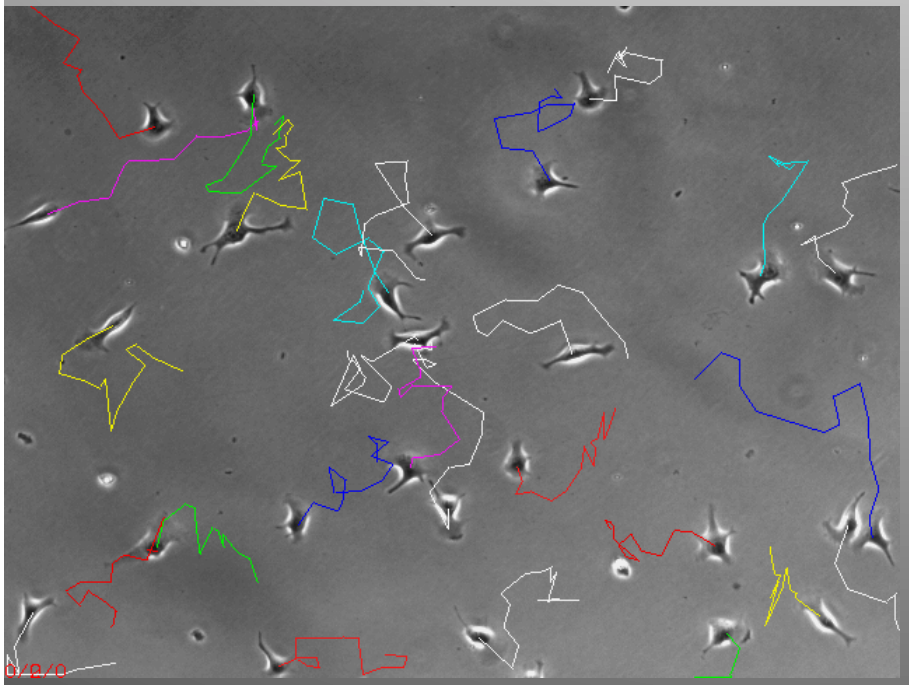
Taxol (10 uM)

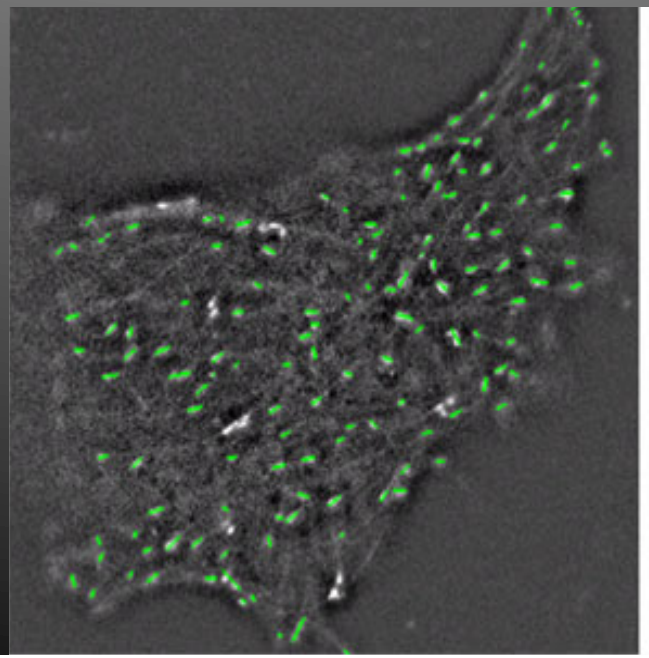


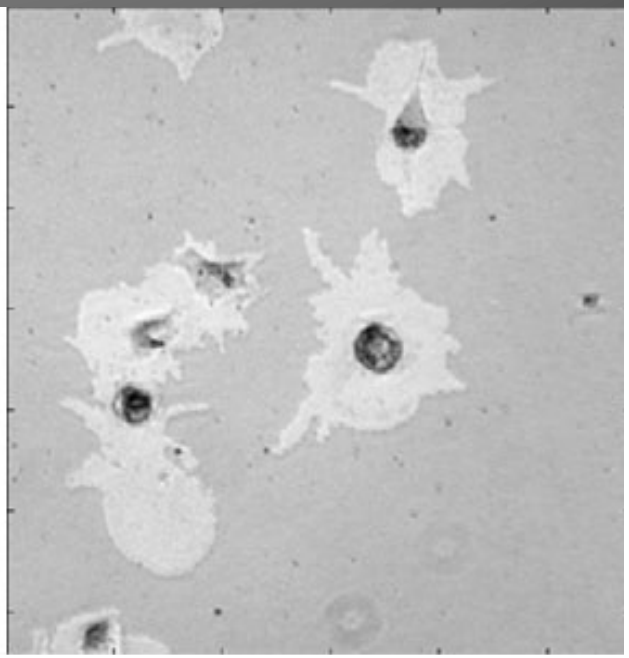
Nocodazole (10 uM)



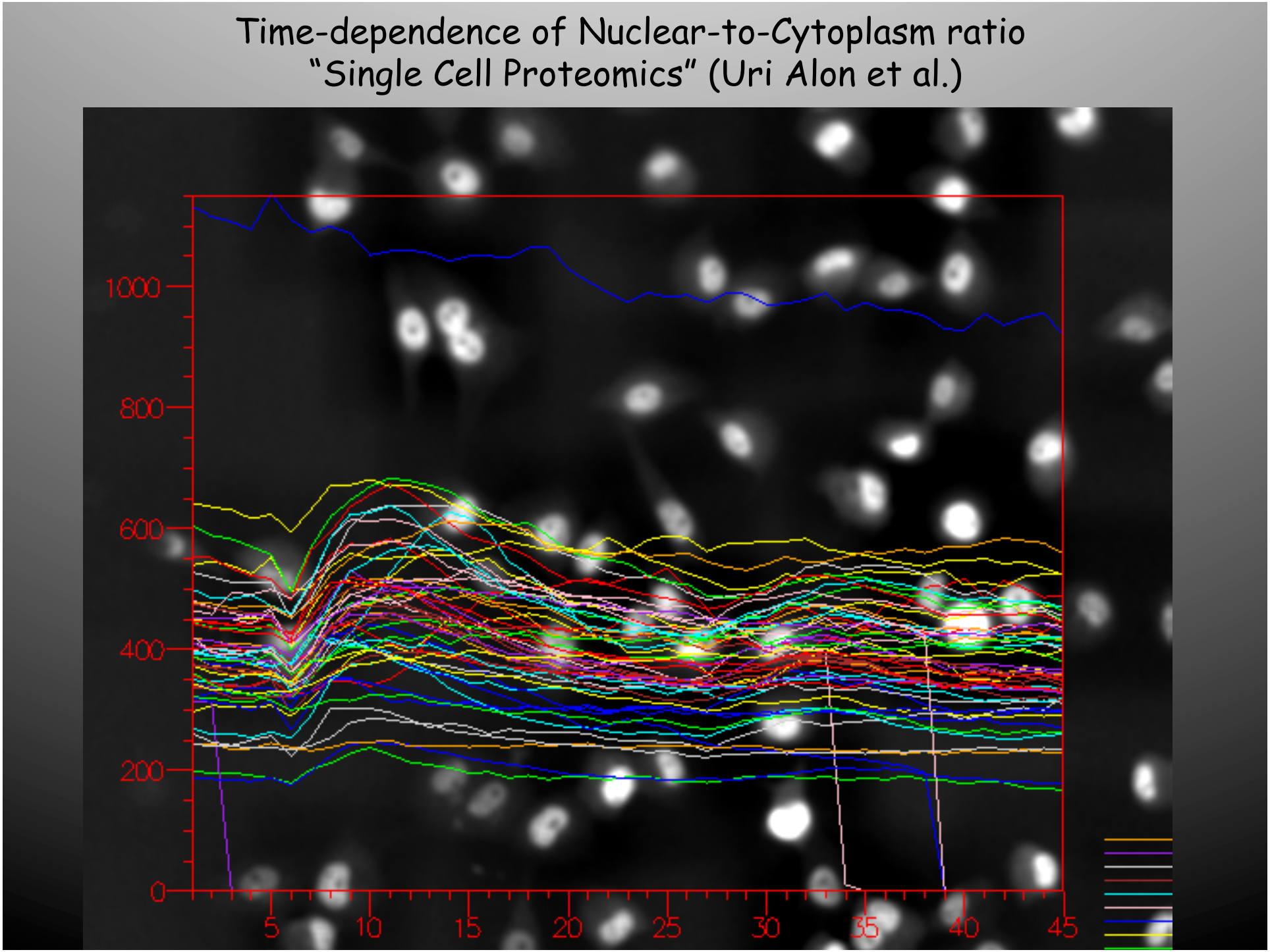
Tracking (cell locomotion, microtubule ends, phagokinetic tracks)



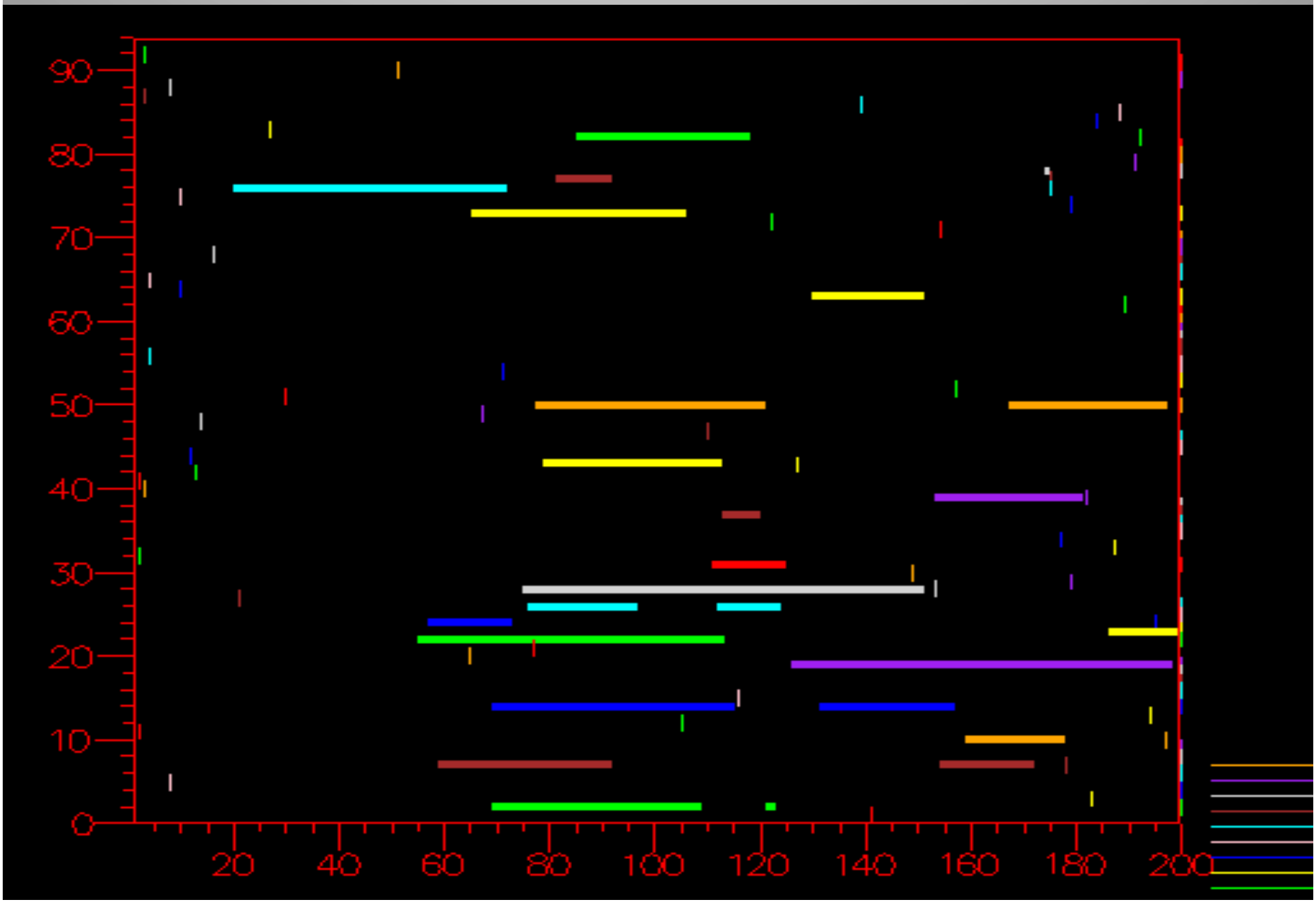


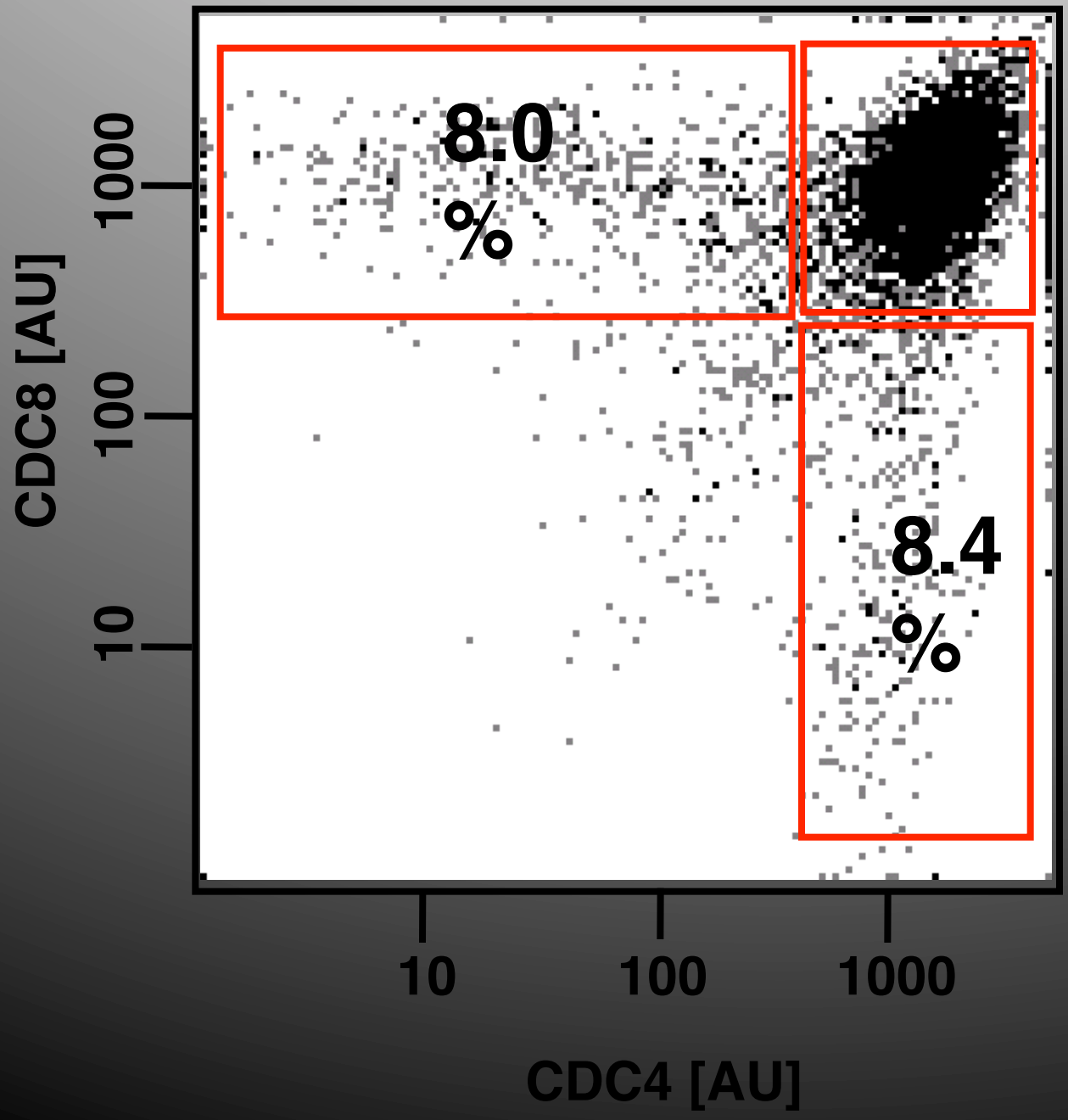


Time-dependence of Nuclear-to-Cytoplasm ratio
"Single Cell Proteomics" (Uri Alon et al.)



Cell by cell time dependent property (T-cell adhesion in flow)

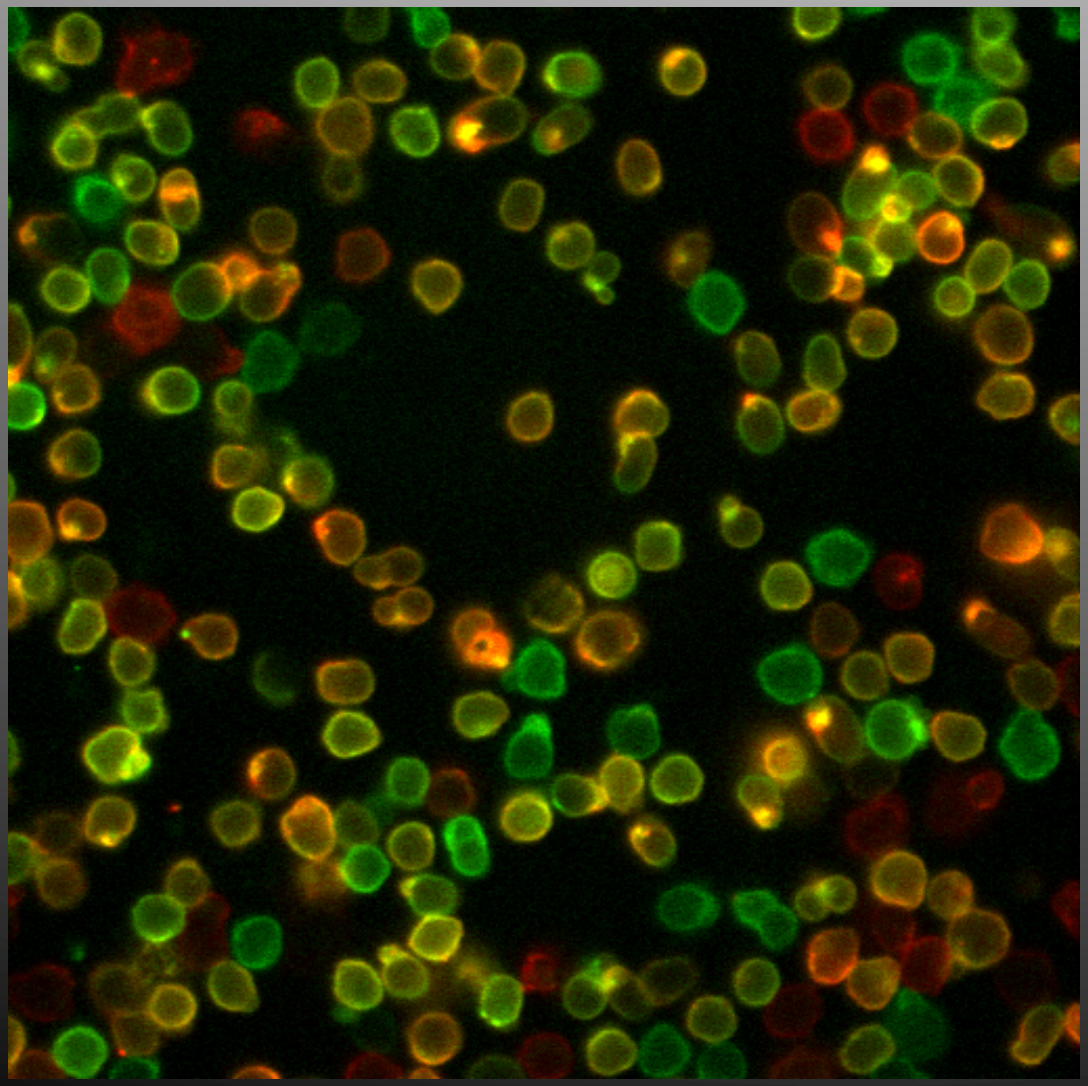




FACS-like ScatterPlot From 6500 cells in 49 images

X&Y Log-Scales

80% are CDC4&8 positive.

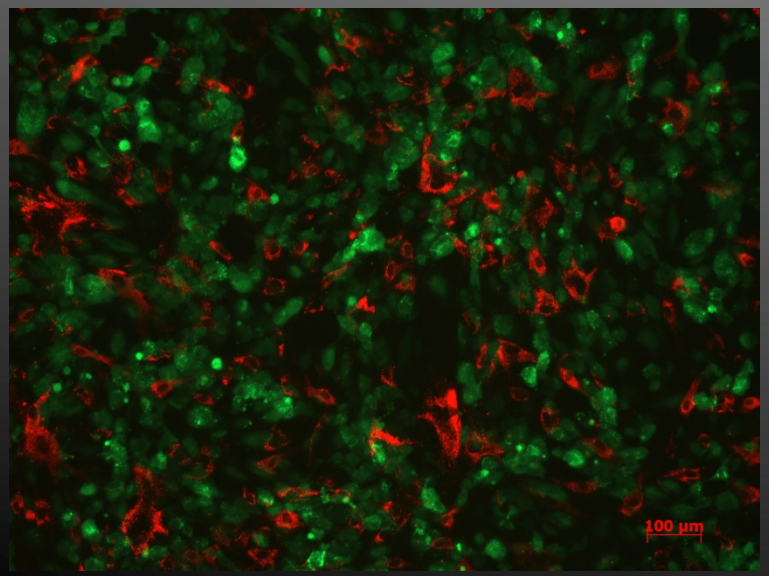


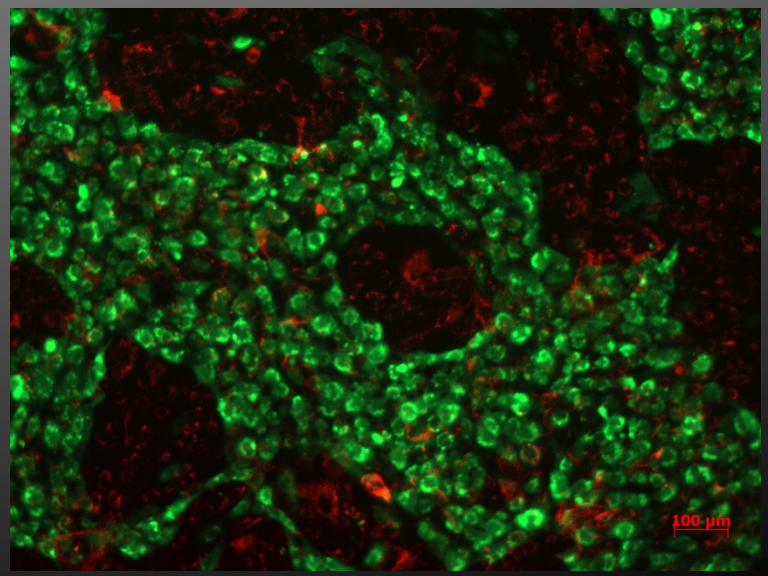
FITC (CDC4)
PE (CDC8)
Double-lablled
Thymocytes:
Typical field
Imaged with
60X/.95
Objective.

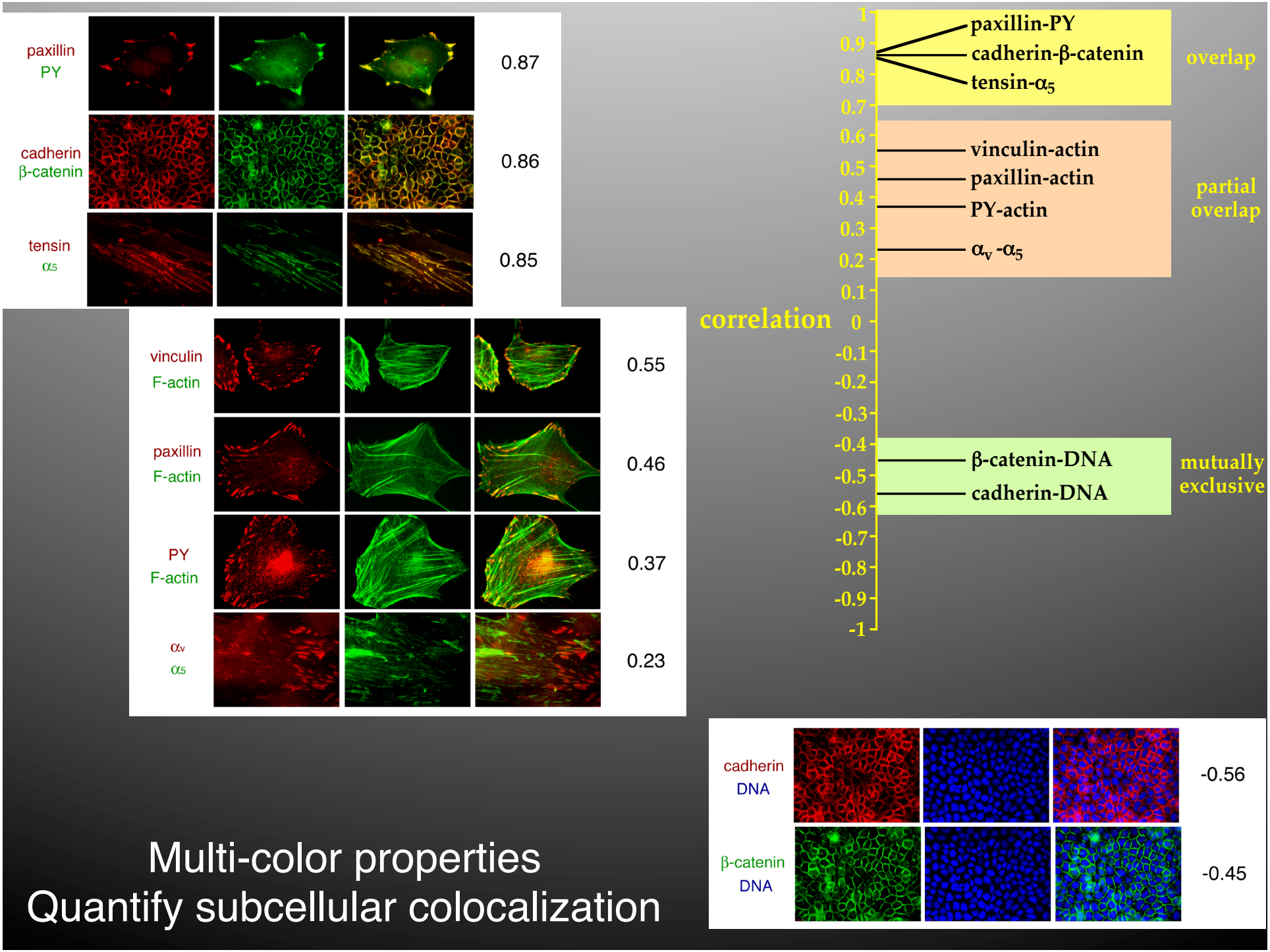
Shape attributes,
Distribution of dye,
Textural attributes,
Color colocalization

(Shulamit Levenberg et al.)

1. Cell Number & Density:
   uniform, random, bi-lobed (lumen vs, assembled tissue)
2. Entropy of cell distribution
3. Textural parameters at various scales
4. Separation or intercalation of hff & Huvec cells
5. Cell Area (smaller projection in tightly packed tissue?)
6. **Superstructural organization**
   lumen size distribution, maximum lumen radius, sharpness of borders
7. Time dependence of the above attributes.







Multi-color properties
Quantify subcellular colocalization

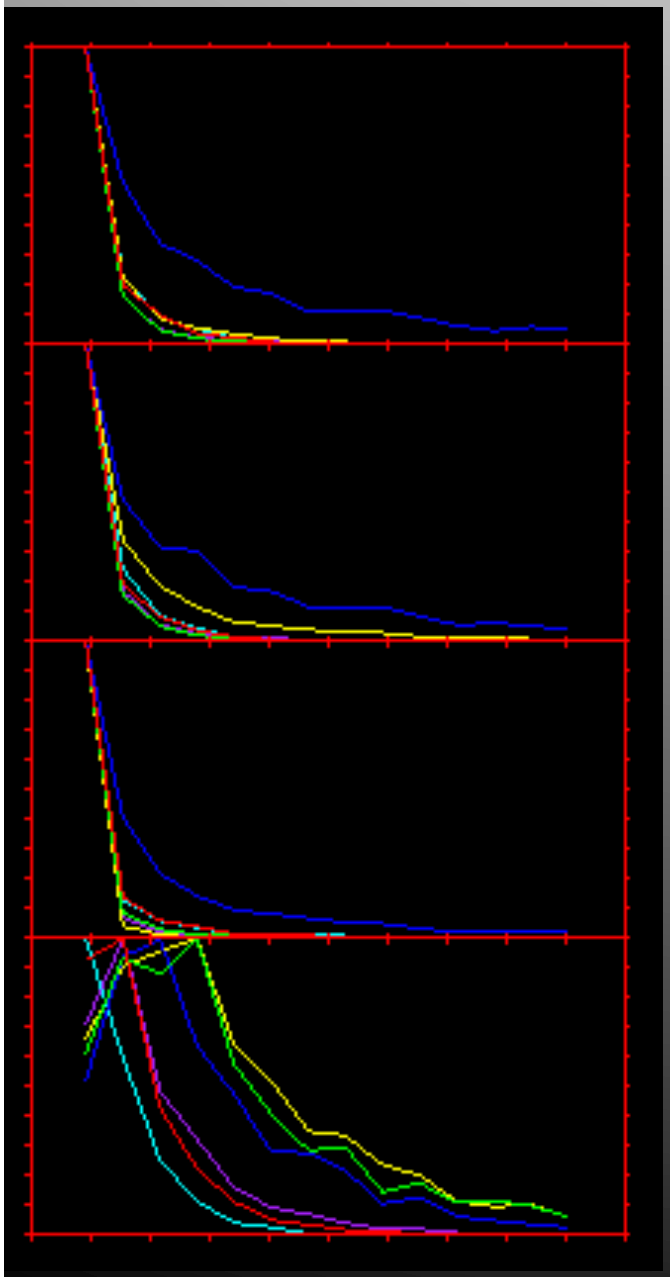
V. Statistics

Averages and Standard Deviations of attributes
Median, Percentiles, Absolute Deviates - better measures for non-normal distributions
Histograms, Scatter plots, regressions, correlation coefficients
Non-parametric comparisons (KS, WR)
Clusters
Principal Component analysis
Multi-parametric comparisons between controls and test images.
Elimination of outliers and artifacts.

Sum'ry	Flag	ID	CM x	CM y	Nuc Area	Nuc Perim	Nuc TotInt	Nuc Bck Int	Nuc TotInt-Bck	Nuc AvgInt-Bck
min	0.50	1.00	4.45	4.30	301.00	0.00	646614.00	968.76	231931.34	770.54
max	1.00	46.00	509.24	508.82	3850.00	0.00	29520566.00	1546.80	25924334.00	10587.49
sum	1970	2.428e+04	5.725e+05	5.794e+05	2.474e+06	0	1.438e+10	2.679e+06	1.145e+10	1.022e+07
avgC	0.88	10.80	254.68	257.72	1100.52	0.00	6398129.67	1191.94	5093010.74	4546.99
avgF	1.00	12.33	290.62	294.09	1255.83	0.00	7301012.95	1360.14	5811719.87	5188.65
stdvC	0.22	7.84	153.17	156.95	479.23	0.00	3490259.00	122.93	2989047.25	1274.58
stdvF	0.90	13.35	297.19	301.75	1200.34	0.00	7288207.50	1198.26	5905350.00	4722.26
perc10	0.50	2.00	36.95	37.97	612.00	0.00	3391689.00	1064.37	2562691.00	3044.01
perc90	1.00	22.00	468.49	475.22	1776.00	0.00	10615037.00	1358.64	8697990.00	6252.31

count=2248 sum flags=1970.0

1-attribute statistics



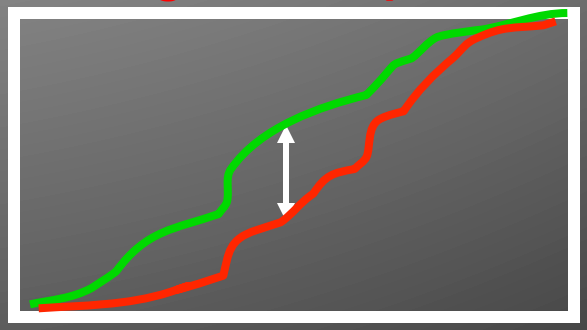
Most stunning feature:
diversity implies broad distributions
They are anything but normal ones!

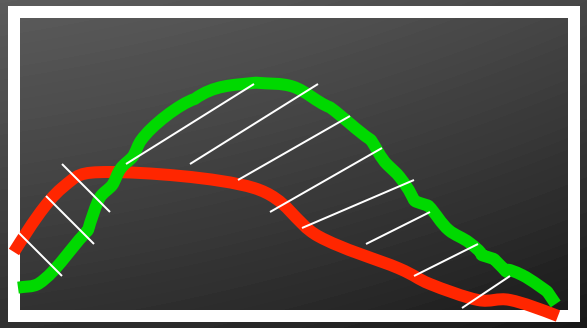
~~Avergaes~~
~~Standard Deviations~~

Non-parametric statistics

Percentiles
Wilcoxon RankSum
Kolmogorov-Smirnov

Histogram comparison





"survival of the outliers"

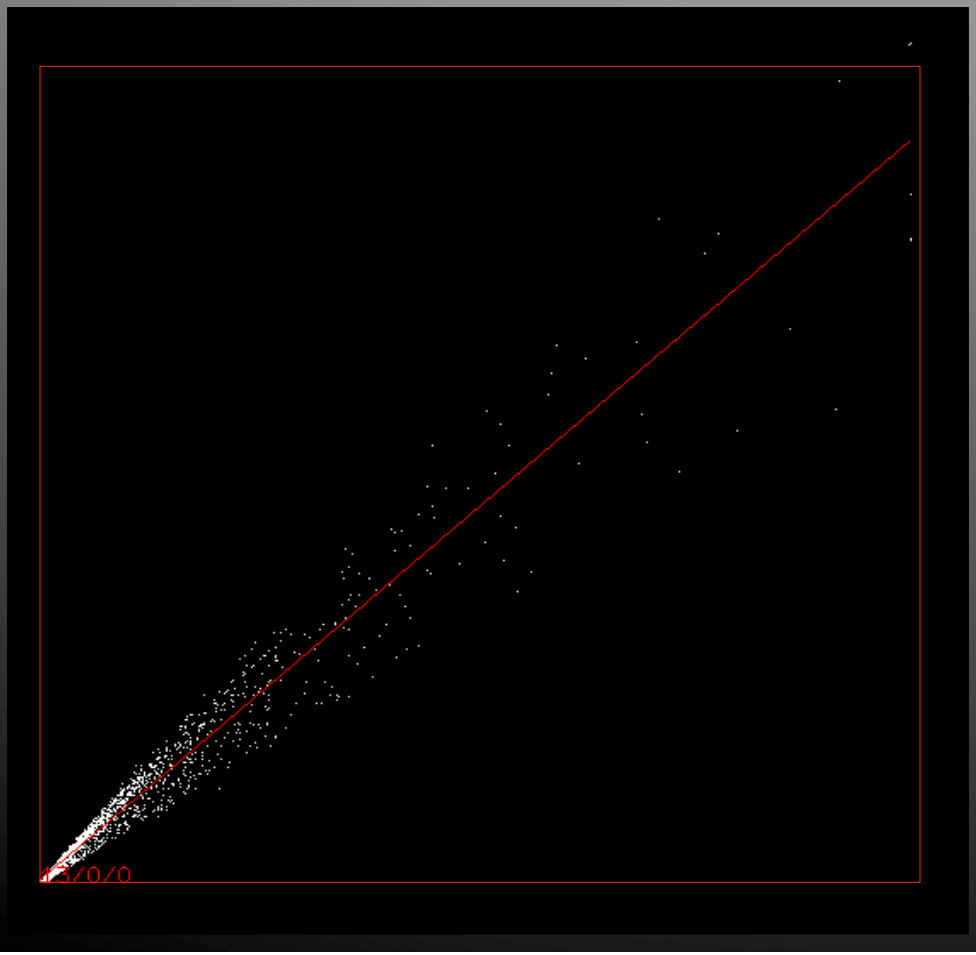
K-Means, K-Medians and Hieranchical clustering
Self-organizing Mats
Principal components

Z-values
p-values
Correlation Coefficients
Mean absolute deviation (MAD)
Student-Newman-Keuls (SNK) test
Wald-Wolfowitz runs test
Mann-Whitney U test [==Wincoxon ranksum]
Kruskal-Wallis analysis of ranks
Median test
Pitman's permutation test
Spearman R (rank) correlation
Kendall's Tau
Siegel-Tukey test

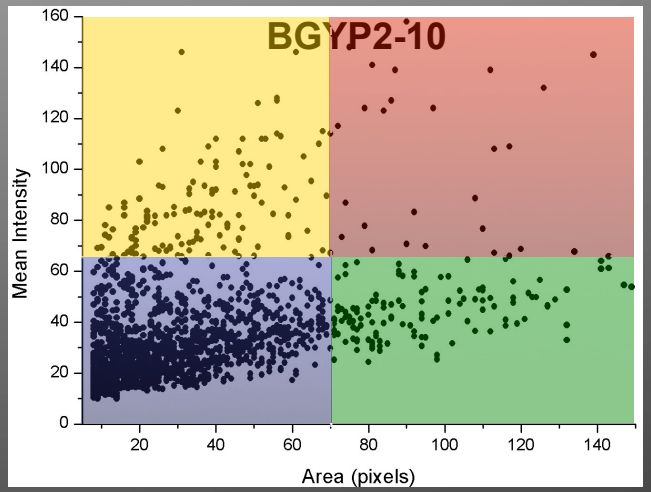
2-attribute statistics

Another feature: **correlations between attributes**

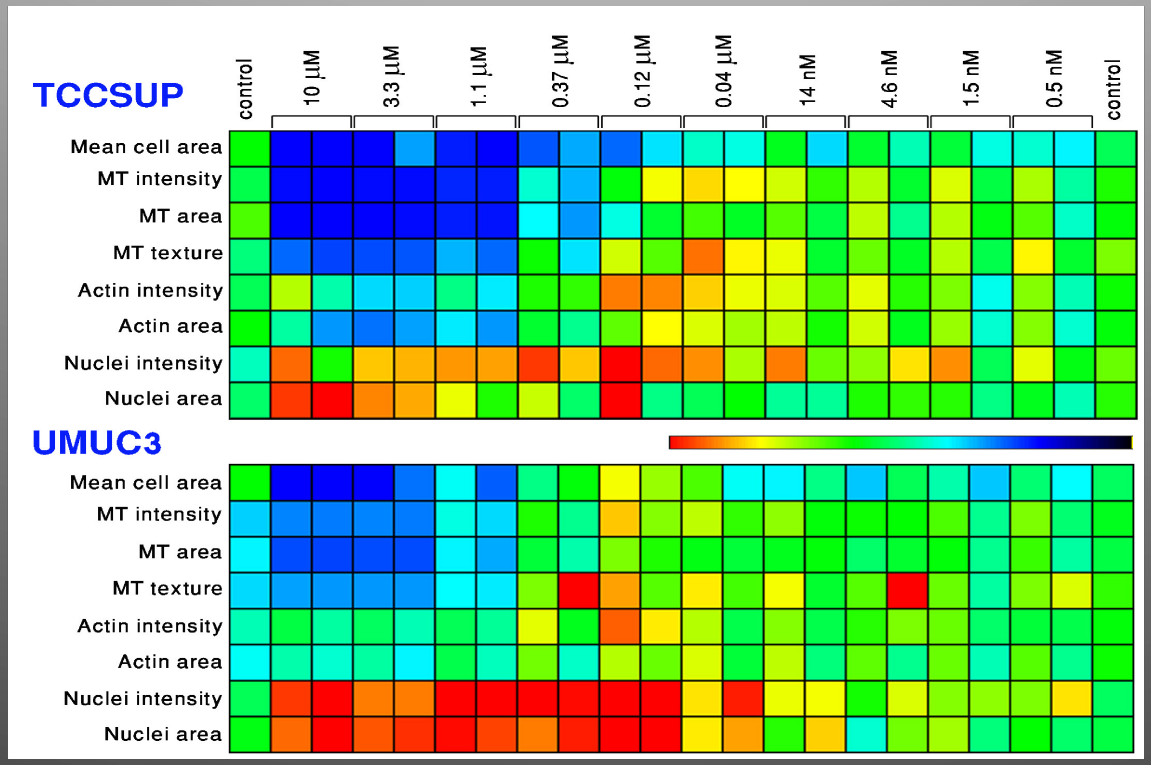
ScatterPlots
(regression, covariance)



Quarters
(extended percentiles)



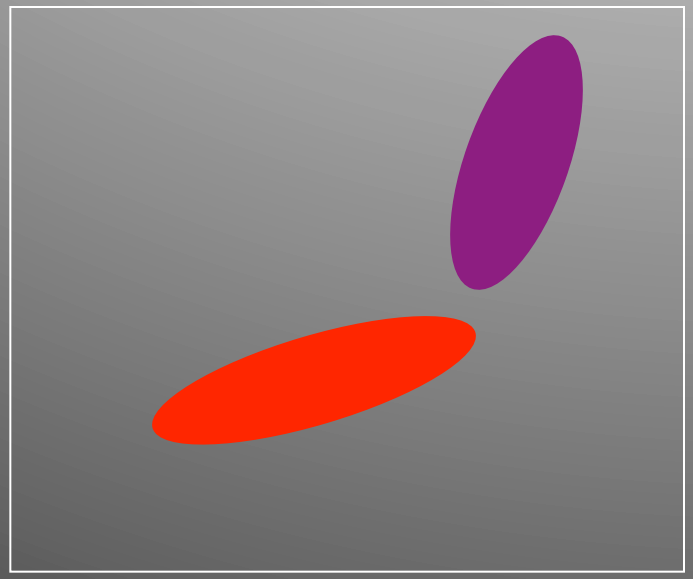
n-attributes



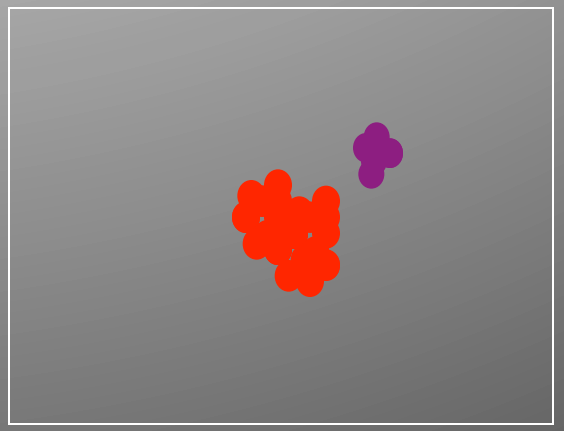
Drug titration responses as reported by 8 different attributes: correlated, anticorrelated different scales
Systems cell biology approach would benefit from unbiased Report of any effect, and not a search for a specific one.

n-attribute statistics

Principal Component Analysis



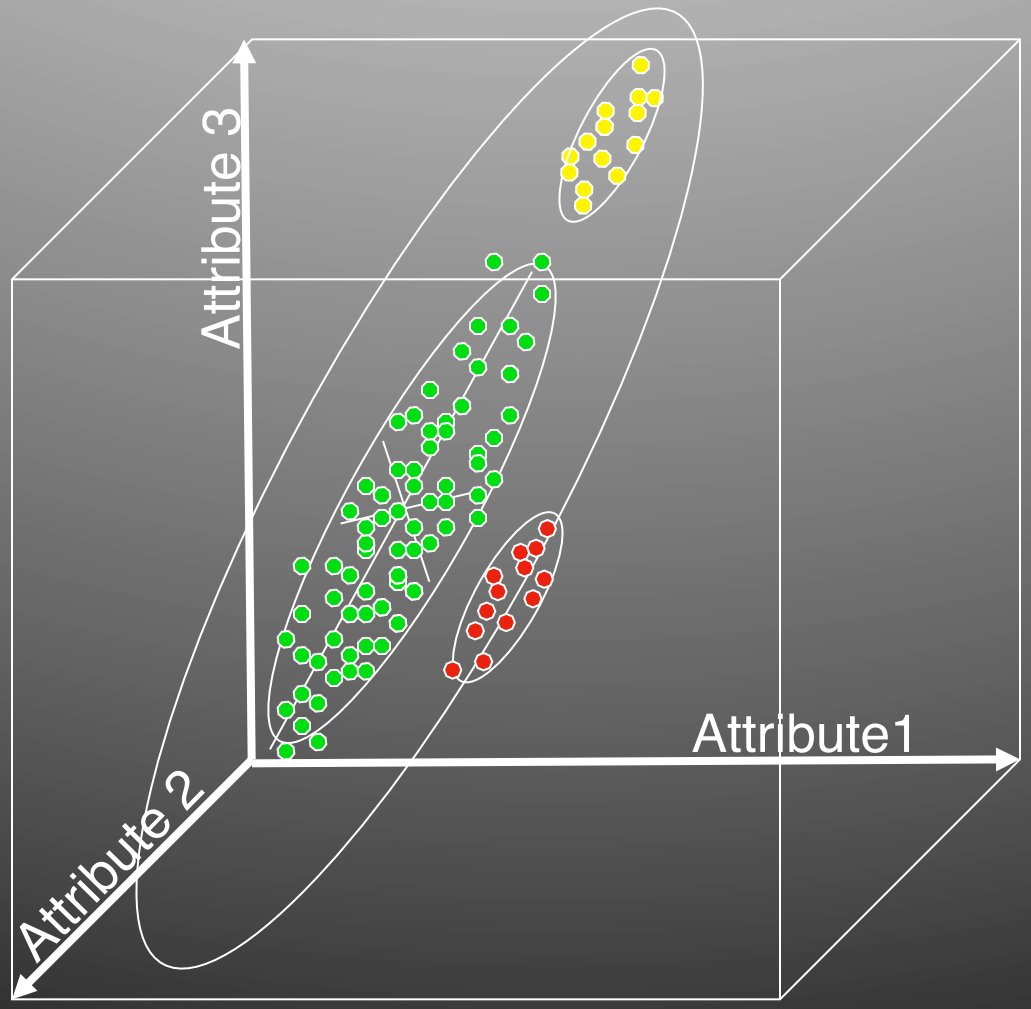
Clusters
Classifiers (e.g. NN)



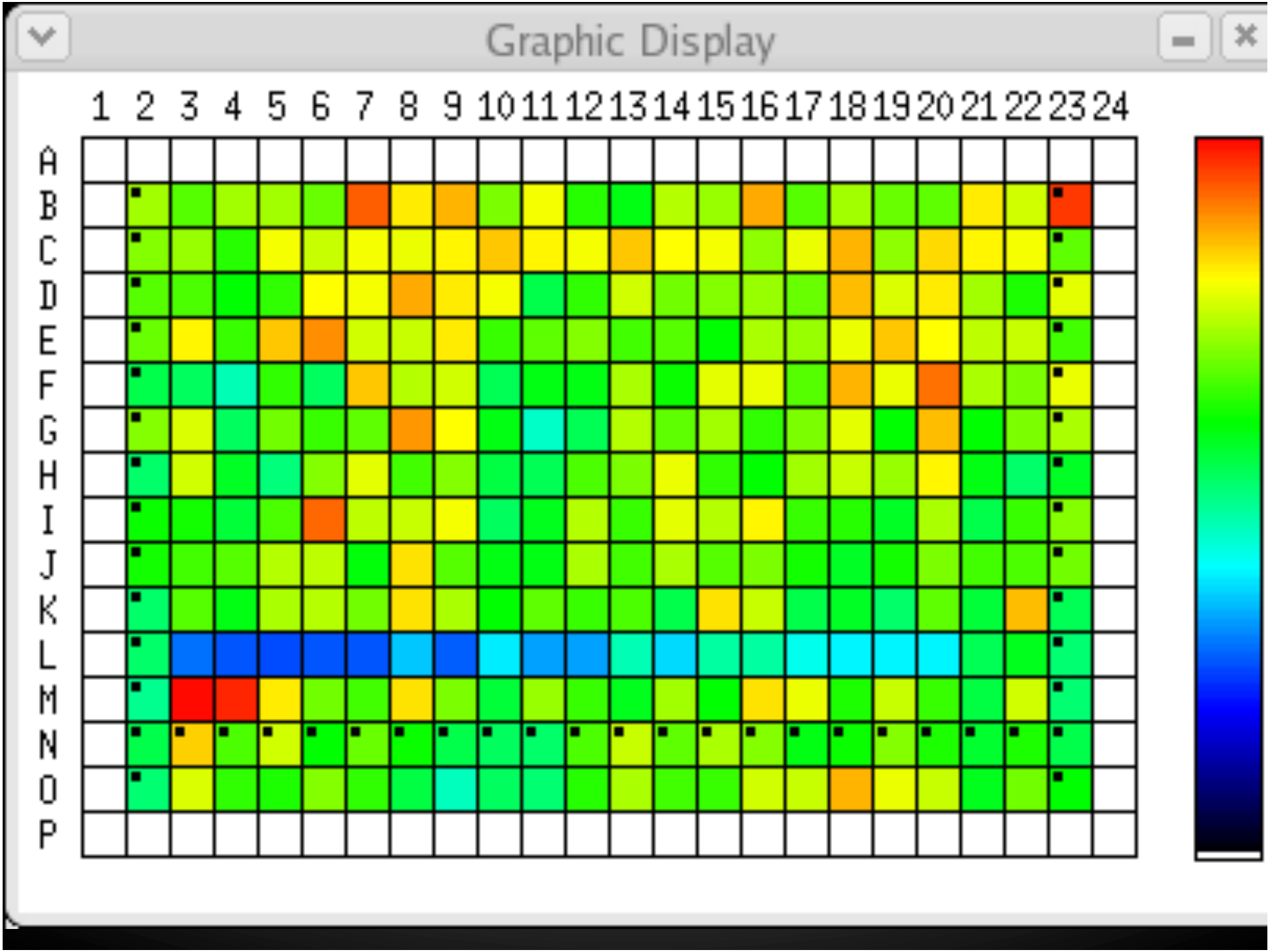
Both depends on distances in n-dimensional space.

1. Euclidian metrics will be biased by attribute scales.
2. Will be distorted by correlations between attributes

Multi-parametric score



Mahalanobis distance - balanced score, n-contributions


Graphic Display
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
A B C D E F G H I J K L M N O P

TeraBytes of images -> MegaBytes of data

-> one plate picture

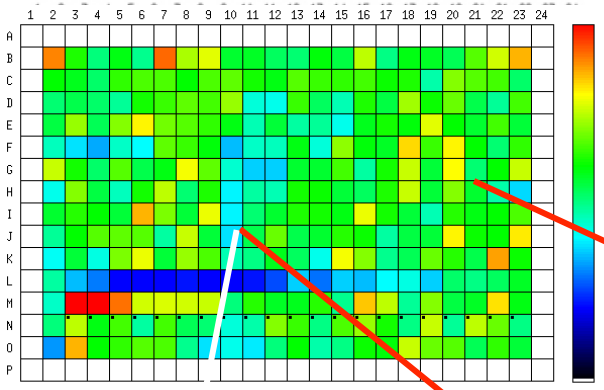
How to "mine" the data

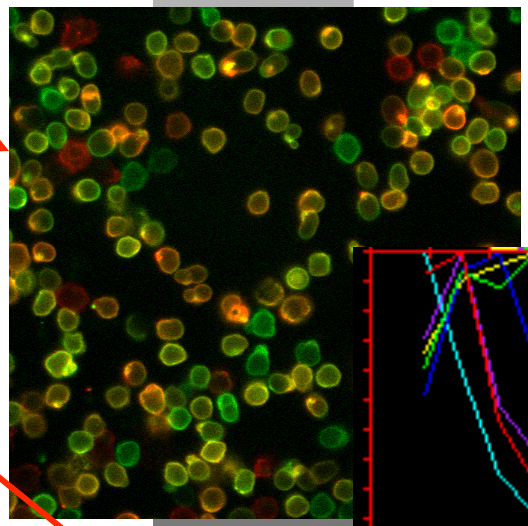
How to "mine" the data

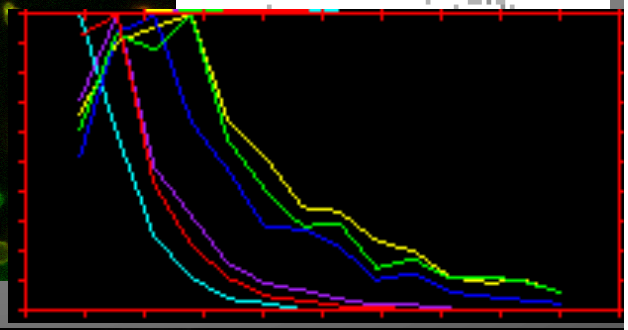
Given a list of cells and their intracellular structures, ["objects"], and for each a set of quantified parameters, ["attributes"] we want to:

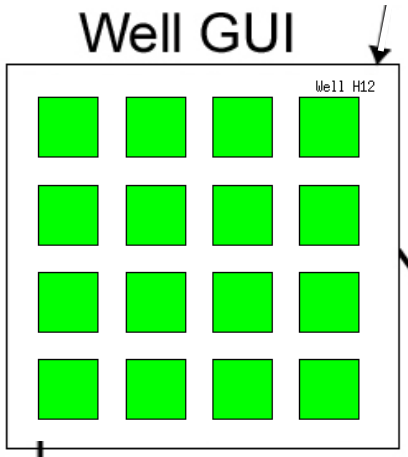
1. Characterize their multi-parametric distributions: Histograms, Scatter Plots, multi-parametric scores
2. Define gold standard "signatures" (from controls) and score the deviation from standard.
3. Identify "outliers" in broad distributions
4. Identify the attributes most contributing to differences from control (unbiased screen design principle)
5. Visualize the cells that show large deviations from controls, and display scores for objects. (two-way "data mining": from images to scores and back

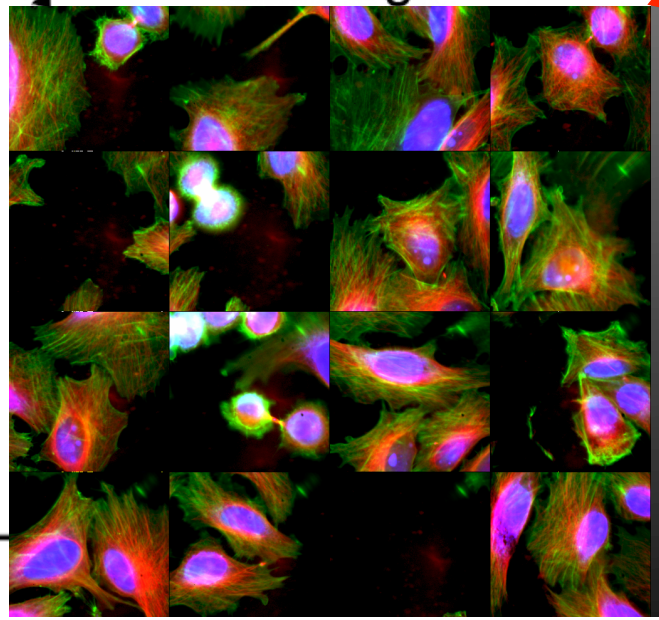
VISUALIZATION:
Well and field links

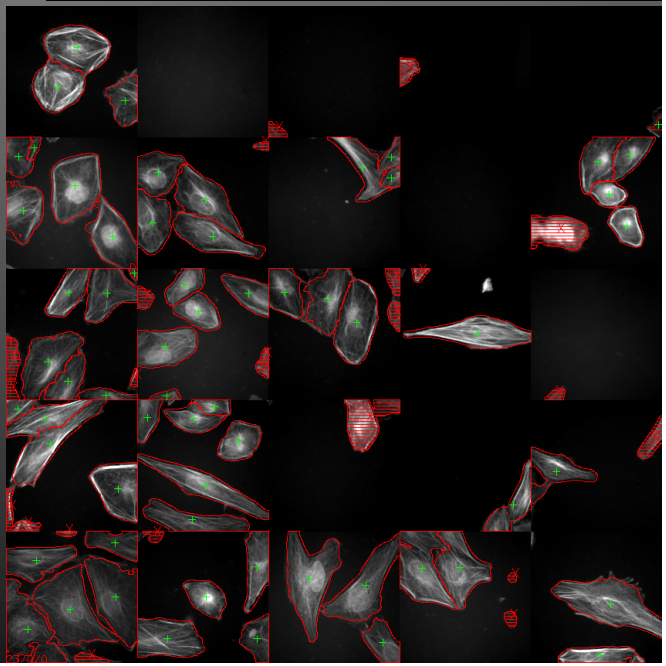


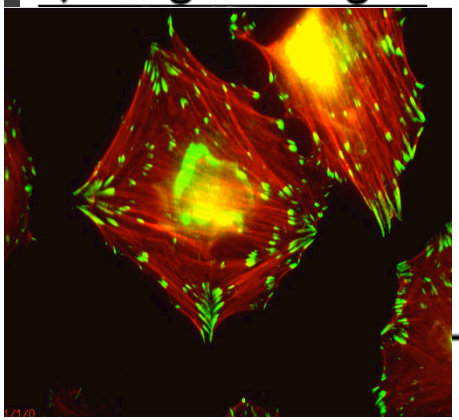


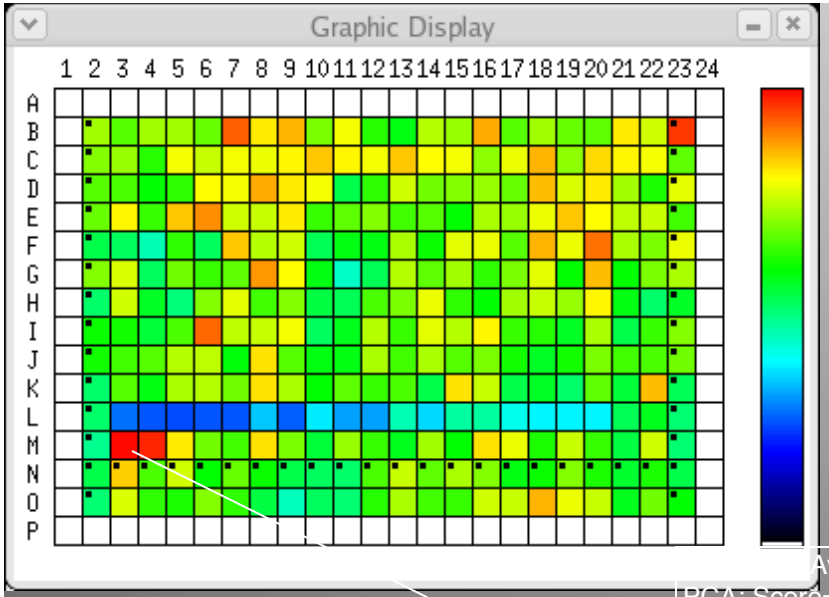












From scores to the original attributes Internal Standards from controls: same SCORES everywhere

Averaged Analysis row,col=[7,23]
PCA: Score=0.076851 Mahalanobis distance=0.172384
Total number of attributes in Data=10, Number of selected attributes = 10
. . . . . Attribute by order of contribution to Score . . . . . .

SlctdAttNames:	Nuc TotInt-Bck	Nuc TotInt	Nuc AvgInt-Bck	Nuc Area	Nuc Bck Int	Nuc Perim
AttribsAverag:	7272459.5	19110168.0	1789.0	3704.4	3169.1	0.000000
BalncdDistnce:	0.239678	0.097599	0.236849	-0.025436	0.002244	0.000000
ContribMahala:	1.434907	0.991043	0.500642	0.065788	0.015461	0.000000
NormContrMaha:	0.790036	0.545652	0.275645	0.036222	0.008513	0.000000

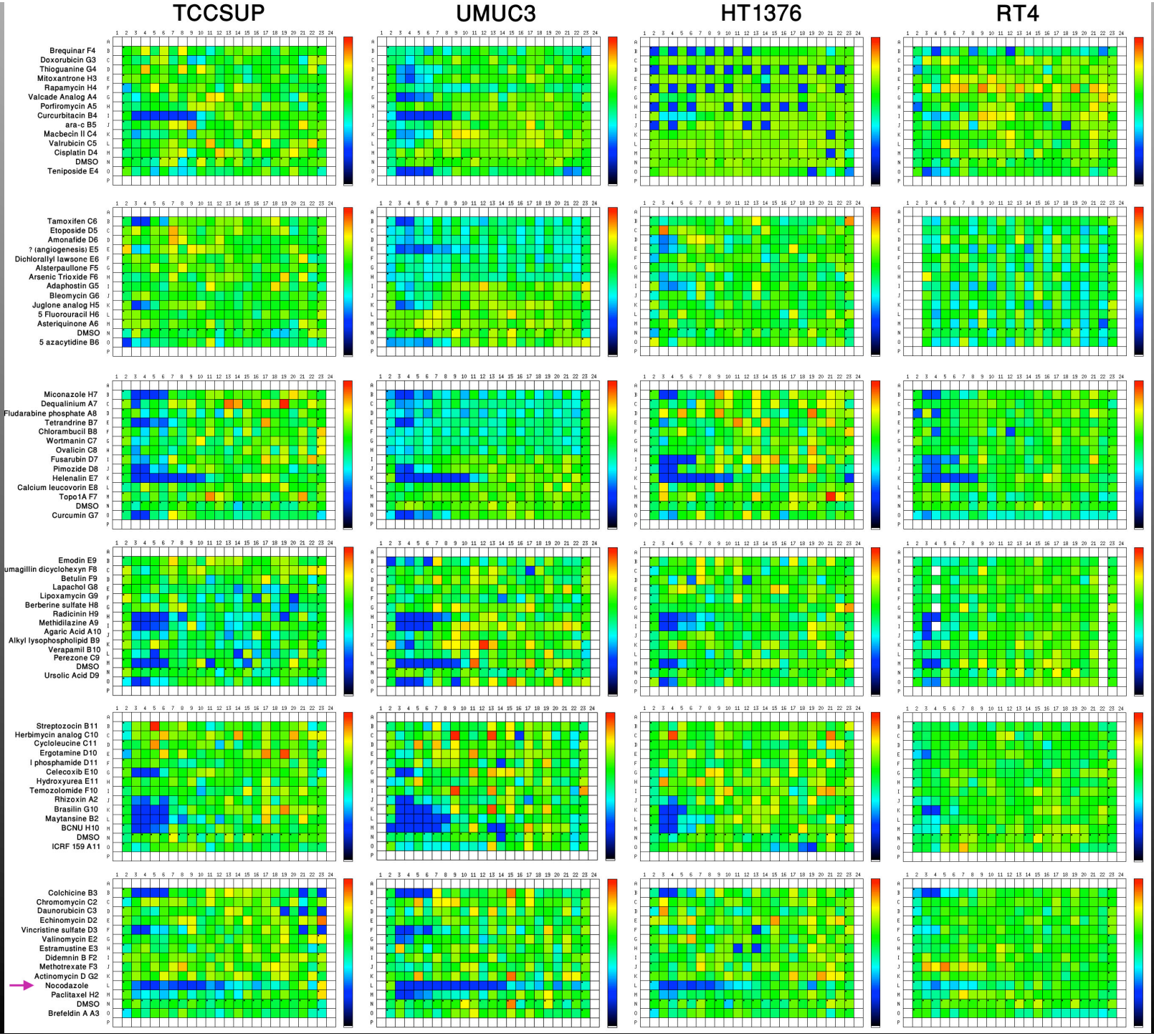
PCA: Well Averaged Analysis row,col=[13,3]
PCA: Score=0.407145 Mahalanobis distance=85.246056
Total number of attributes in Data=58, Number of selected attributes = 39
. . . . . Attribute by order of contribution to Score . . . . . .

SlctdAttNames:	Bck AvgInt	Cell AvgInt-Bck	Cell AvgInt	Cell AvgInt	Fib CellInt-Bck	Bck AvgInt	Fib CellInt-Bck	Cell Area	Fib AvgInt-Bck
AttribsAverag:	1646.994751	776.218811	1816.007202	1227.823120	169.012543	1192.571045	35.253063	15637.710938	382.967682
BalncdDistnce:	0.376617	0.892196	0.409922	0.157536	0.734470	0.147590	0.493979	-0.125828	0.761477
ContribMahala:	590.989319	517.178223	306.110199	142.900635	139.575912	126.761696	17.162472	10.854239	10.396242
NormContrMaha:	0.674765	0.590490	0.349503	0.163157	0.159361	0.144731	0.019595	0.012393	0.011870

SlctdAttNames:	Fib LenInt-Bck	Fib TotInt	Fib TotInt-Bck	Nuc TotInt	AvgBck	Fib LenInt-Bck	Fib AvgInt-Bck	Cell Area	Fib Len
AttribsAverag:	1072.453125	16405442.000000	3054896.750000	8714288.000000	1085.495361	271.878113	74.495911	15637.710938	2310.066406
BalncdDistnce:	0.681143	0.225454	0.385945	-0.131000	0.005688	0.582624	0.680559	-0.125828	-0.082695
ContribMahala:	5.132737	3.077837	2.949579	2.346499	1.783195	1.659733	1.474340	1.292212	1.003325
NormContrMaha:	0.005860	0.003514	0.003368	0.002679	0.002036	0.001895	0.001683	0.001475	0.001146

SlctdAttNames:	Fib Area	Cell2HallAreaRatio	Nuc AvgInt-Bck	Nuc Bck Int	Fib Len	Nuc TotInt-Bck	Fib TotInt-Bck	CnvxHullArea	LongAxis
AttribsAverag:	7432.044434	0.893977	1011.107727	1096.429565	2688.444824	4251916.000000	590245.125000	17524.896484	64.415306
BalncdDistnce:	-0.107105	0.082102	0.198768	-0.208489	-0.074438	0.038863	0.354402	-0.170906	-0.166488
ContribMahala:	0.886792	0.650772	0.594147	0.544599	0.543456	0.522200	0.485930	0.418834	0.328538
NormContrMaha:	0.001012	0.000743	0.000678	0.000622	0.000620	0.000596	0.000555	0.000478	0.000375

SlctdAttNames:	Solidity	Fib Area	Fib TotInt	ShortAxis	AxialRatio	FramePerim	Perimeter	CnvxHullPerim	AxisAngle
AttribsAverag:	0.896519	8057.922852	10780692.000000	31.039423	2.176268	116.026222	629.696350	492.830414	86.318970
BalncdDistnce:	0.122238	-0.094044	0.020768	-0.041358	-0.093361	-0.140281	-0.133398	-0.108567	-0.034126
ContribMahala:	0.255395	0.251604	0.220695	0.098748	0.049901	0.037017	0.013813	0.002631	0.000841
NormContrMaha:	0.000292	0.000287	0.000252	0.000113	0.000057	0.000042	0.000016	0.000003	0.000001


TCCSUP
UMUC3
HT1376
RT4
Brequinar F4
Doxorubicin G3
Thioguanine G4
Mitoxantrone H3
Rapamycin H4
Valcade Analog A4
Porfiromycin A5
Curcurbitacin B4
ara-c B5
Macbecin II C4
Valrubicin C5
Cisplatin D4
DMSO
Teniposide E4
Tamoxifen C6
Etoposide D5
Amonafide D6
? (angiogenesis) E5
Dichlorallyl lawsone E6
Alsterpaullone F5
Arsenic Trioxide F6
Adaphostin G5
Bleomycin G6
Juglone analog H5
5 Fluorouracil H6
Asteriquinone A6
DMSO
5 azacytidine B6
Miconazole H7
Dequalinium A7
Fludarabine phosphate A8
Tetrandrine B7
Chlorambucil B8
Wortmanin C7
Ovalicin C8
Fusarubin D7
Pimozide D8
Helenalin E7
Calcium leucovorin E8
Topo1A F7
DMSO
Curcumin G7
Emodin E9
umagillin dicyclohexyn F8
Betulin F9
Lapachol G8
Lipoxamycin G9
Berberine sulfate H8
Radicinin H9
Methidilazine A9
Agaric Acid A10
Alkyl lysophospholipid B9
Verapamil B10
Perezone C9
DMSO
Ursolic Acid D9
Streptozocin B11
Herbimycin analog C10
Cycloleucine C11
Ergotamine D10
I phosphamide D11
Celecoxib E10
Hydroxyurea E11
Temozolomide F10
Rhizoxin A2
Brasilin G10
Maytansine B2
BCNU H10
DMSO
ICRF 159 A11
Colchicine B3
Chromomycin C2
Daunorubicin C3
Echinomycin D2
Vincristine sulfate D3
Valinomycin E2
Estramustine E3
Didemnin B F2
Methotrexate F3
Actinomycin D G2
Nocodazole
Paclitaxel H2
DMSO
Brefeldin A A3

Genetic & compounds Perturbations

complementary methods

Drugs – apply fast: Provide time response and causality
of direct effects,
often multiple targets

RNAi - apply slowly. Presumably specific –
but secondary effects develop.

Screen using knockout clones

Cell BioInformatics Data Base

Standardization – Internal controls (positive and negative)
- Absolute scales (xyz dimensions, time)
- calibrated scales (fluorescence)

Integration across labs – complement cellular features from different experiments, cell lines, analysis methods etc.

Summary

* HTP screening microscopy not only hit finding gadget, but platform to assemble cell informatics

•Extend beyond a narrow essay: extract many attributes

* Software allows wide range of feature quantification and flexibility to adopt new ones

* Analysis is sensitive to subtle effects

* Quantitative data is standardized

Summary (cont)

Data mining in large-scale experiments:

Results <--> original images

statistical presentations — montages, multi-color superpositions

Basic significance:

- Link perturbations with cell level responses
- Provide quantitative data for modeling

Biomedical importance:

- Target-free drug discovery
- Drug combination optimization (cell-level synergism)
- Tests on cells may help individualized therapy

Future needs:

- Robust multi-parametric statistical tools
  - outliers and biological variability
- Reverse modeling: convert network graphs plus
  - multiple perturbation experiments results into
  - models

Thanks:

Yuvalal Liron
Yael Paran
Sabina Winograd-Katz
Sarit Batzir

Benny Geiger
Sasha Bershadsky
Uri Alon

Kahn center for Systems Biology
EU


THANX

The End