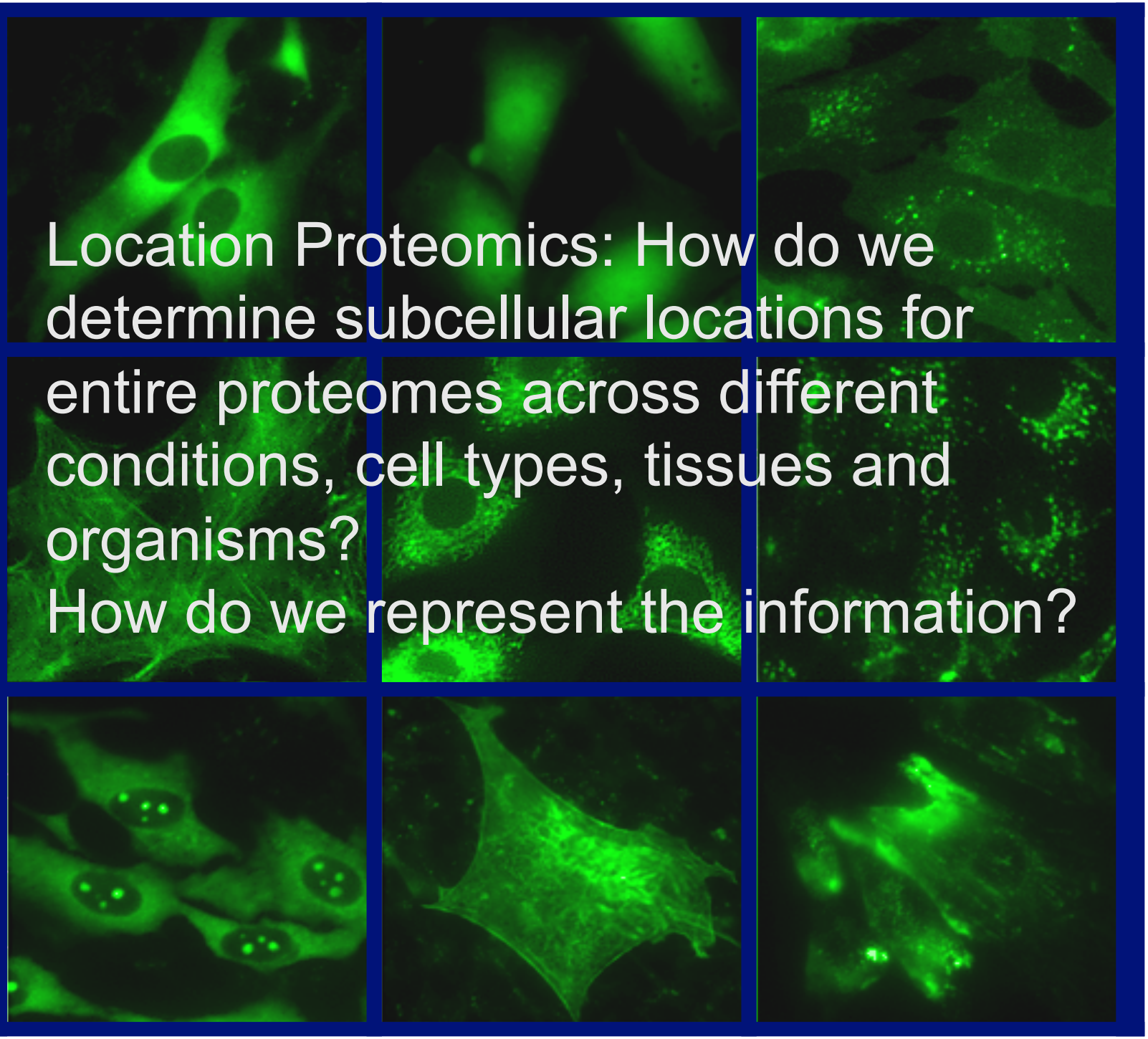
Proteome-Scale Analysis of Subcellular Patterns in Tissue Microarray Images

Robert F Murphy,
Arvind Rao,
Estelle Glory-Afshar

RAY AND STEPHANIE LANE
Center for Computational Biology

Carnegie Mellon



Images from Randtag project: Murphy, Jarvik, Berget

Outline

- Status of automated subcellular pattern analysis in cultured cells
  - Classification
  - Pattern unmixing
  - Generative models of patterns
- Application to tissue images
  - Classification
  - Comparison

Automated Analysis of Subcellular Location

- Problem is hard because different cells have different **shapes, sizes, orientations**
- Organelles **not found in fixed locations**
- Use **numerical features** to describe patterns

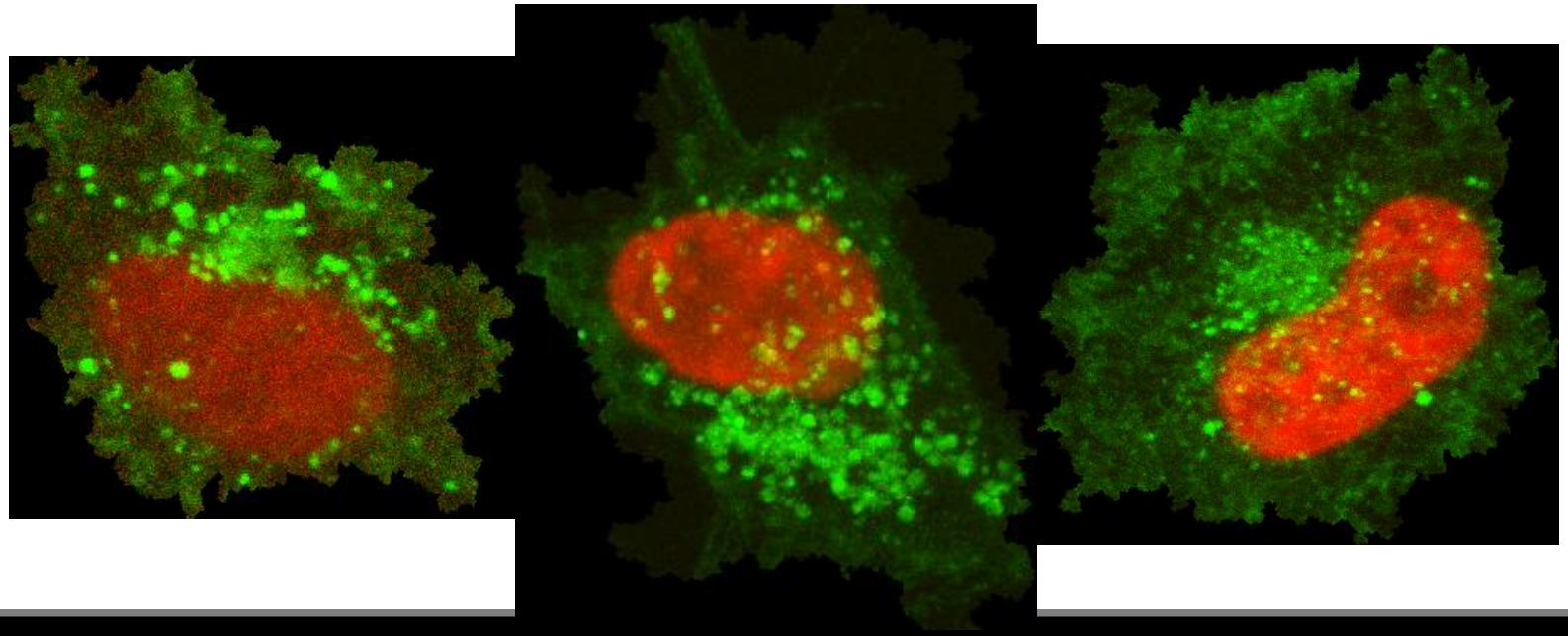
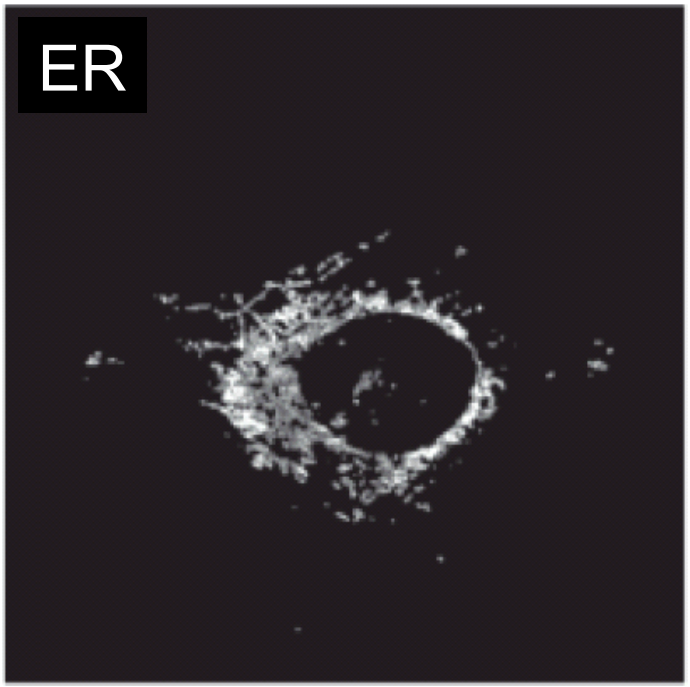
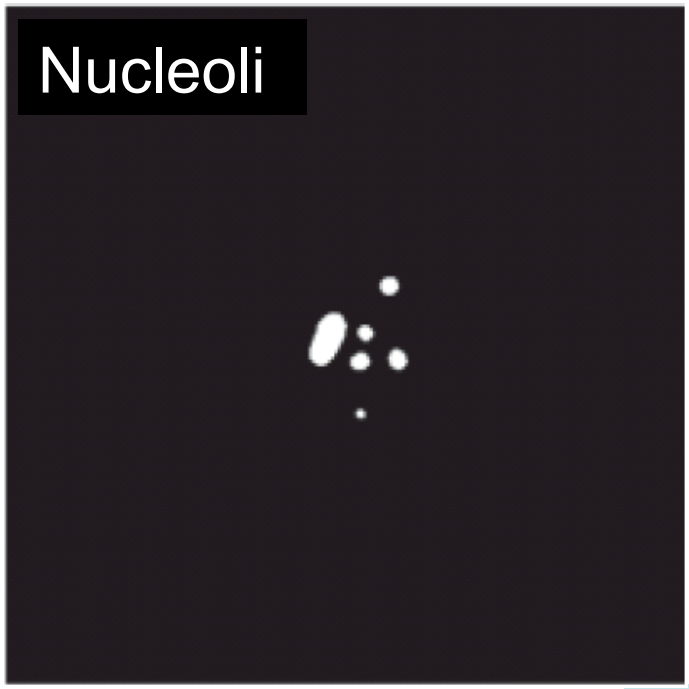


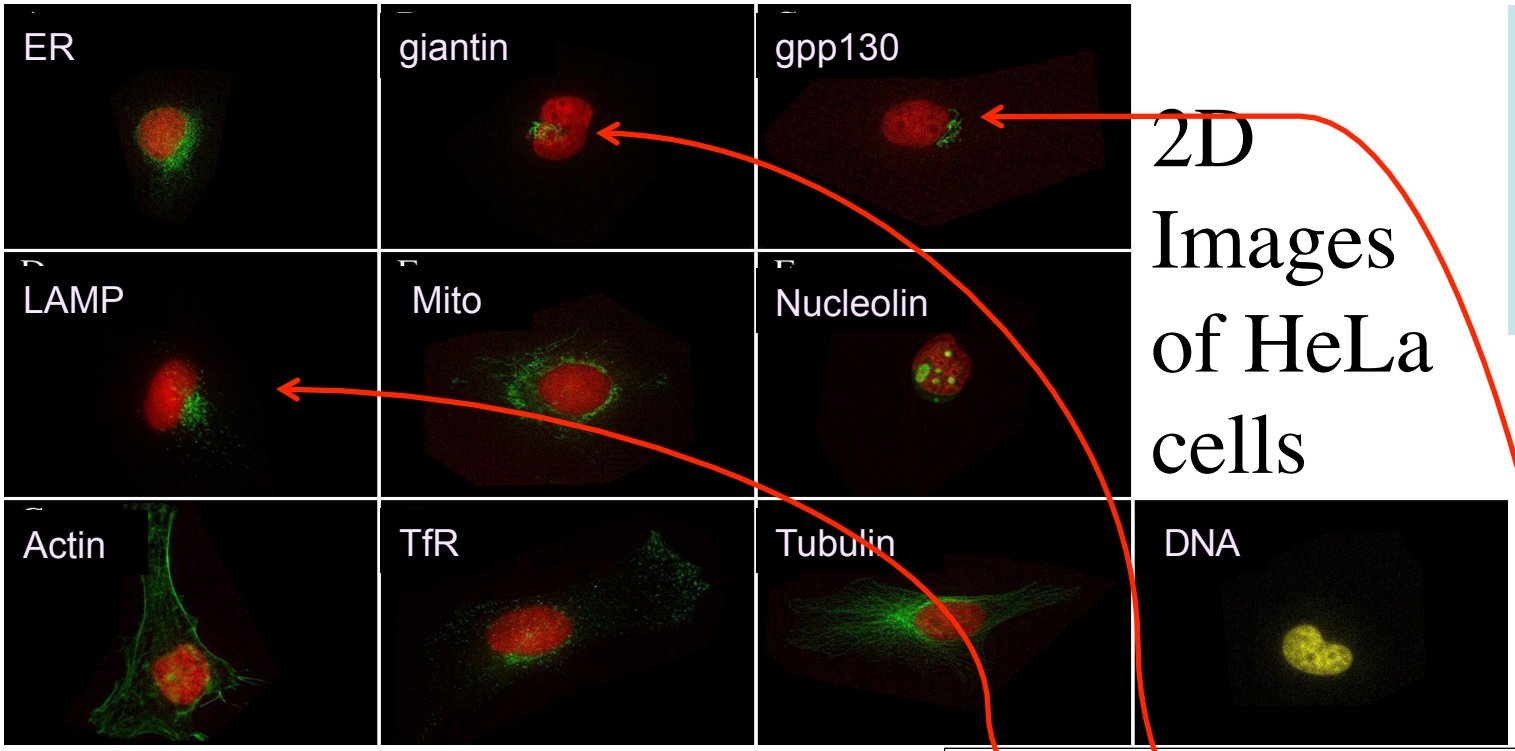
Illustration: 2D Morphological Features to Distinguish Patterns





108	# of objects	6
83	Average size of objects	232
31	Average distance to COF	4

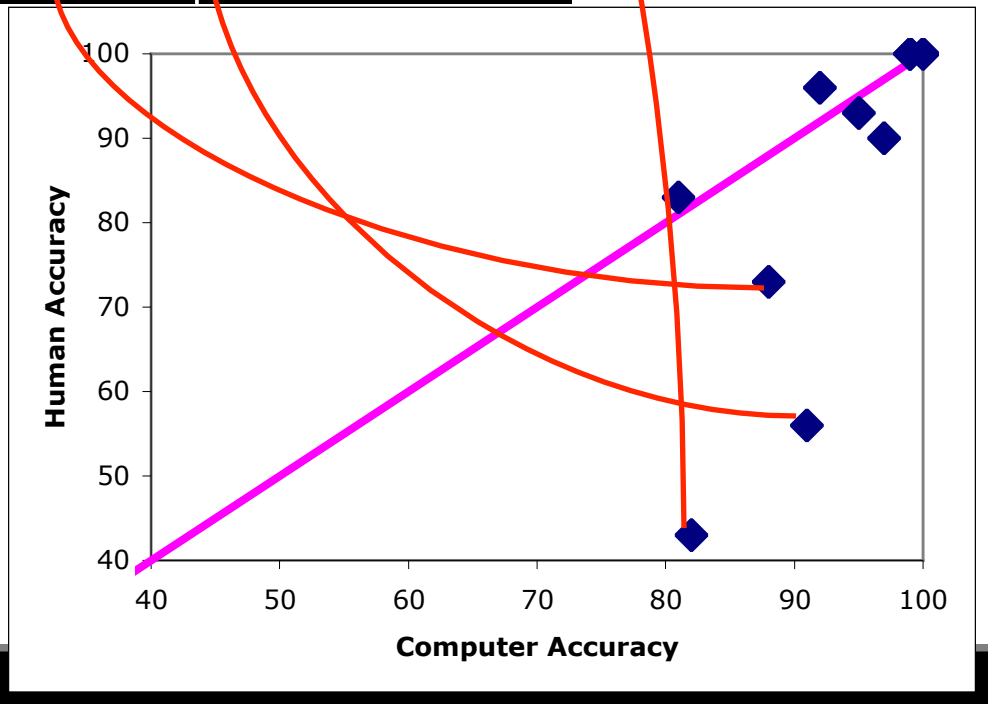
Any of these could be used to distinguish these two classes



2D Images of HeLa cells

Boland et al 1998; Murphy et al 2000; Boland & Murphy 2001; Murphy et al 2003; Huang & Murphy 2004

Subcellular Pattern Classification: Computer vs. Human



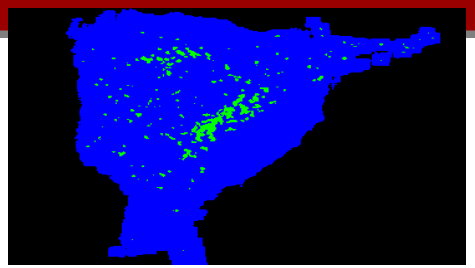
Pattern unmixing

- Many proteins (or other macromolecules) may be found in more than one organelle
- Features “see” each combination of organelles as a new pattern
- Can we “unmix” such mixed patterns?

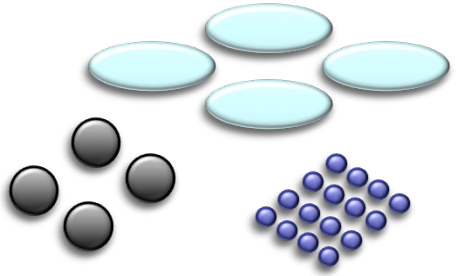
Supervised Unmixing Assumptions

- have markers found in only one subcellular location (fundamental pattern)
- fundamental pattern can be represented by frequencies of distinct object types (e.g., 10% small round and 90% long skinny)
- mixed pattern formed by adding together the objects from two or more fundamental patterns - *no new object types created*

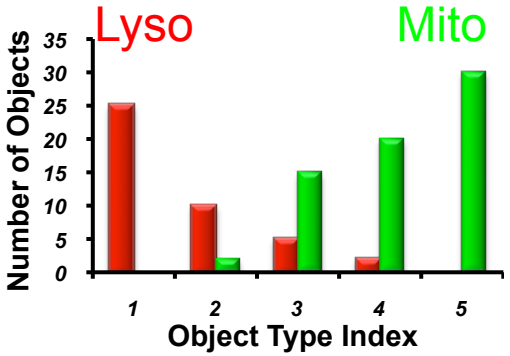
Find objects in each image



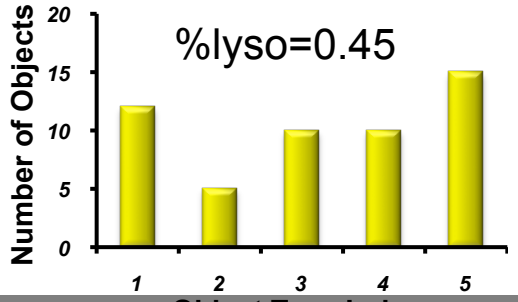
Learn object types from pure samples

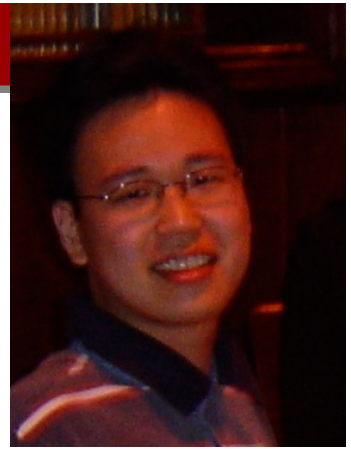


Learn model for distributions of object types in pure samples



Determine pattern fractions for mixed images





Learning object types

- Find all objects in all images of fundamental types
- Describe each object by features such as size, ellipticity, distance from nucleus
- Cluster objects to find types
- Represent each fundamental pattern as probabilities of observing each object type

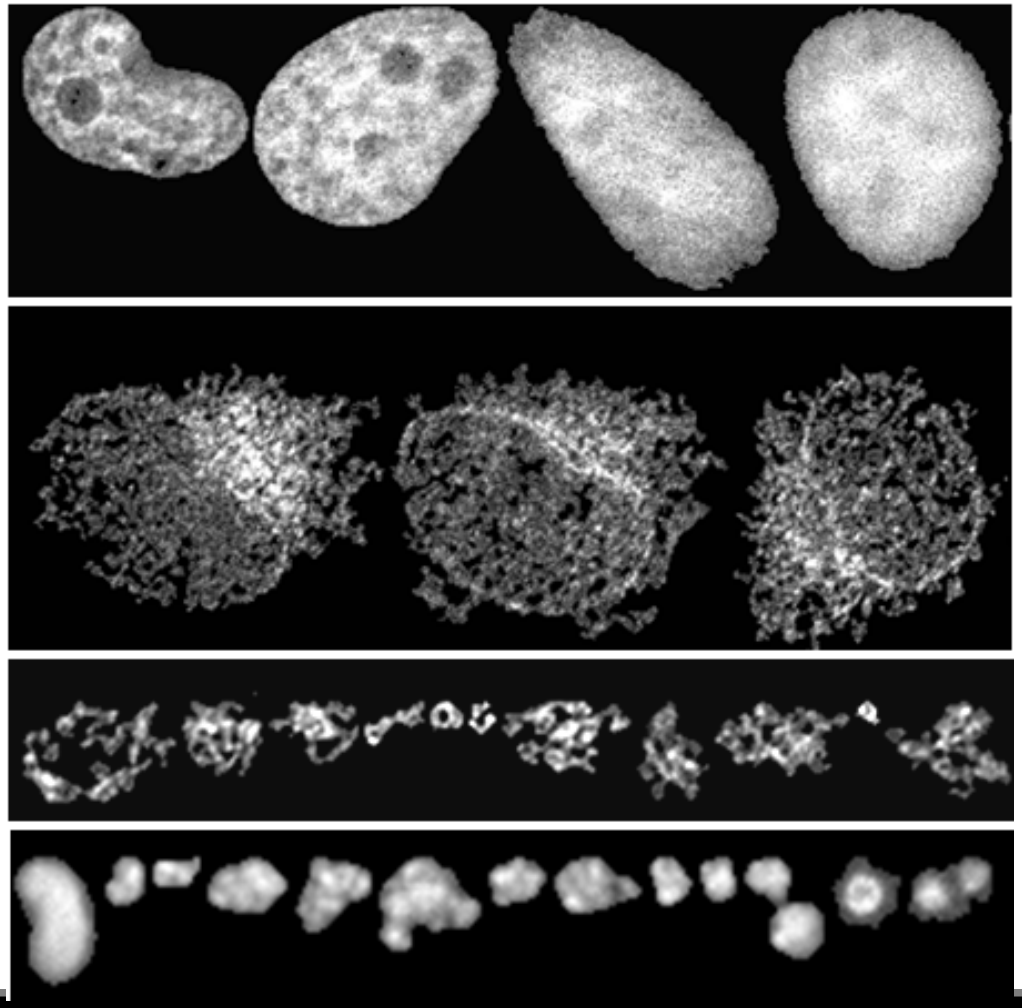
Examples of Object Types

Type A

Type B

Type C

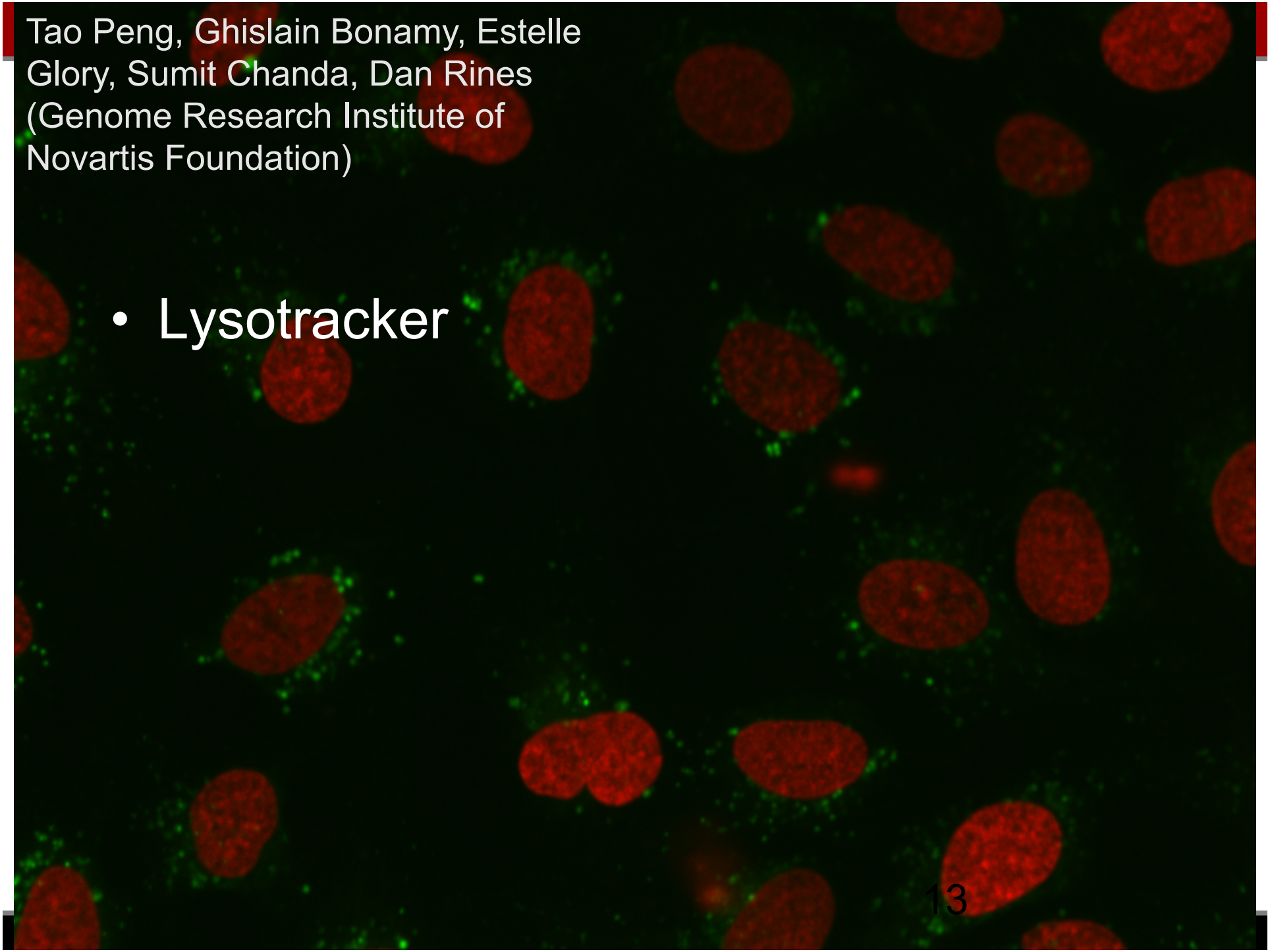
Type D



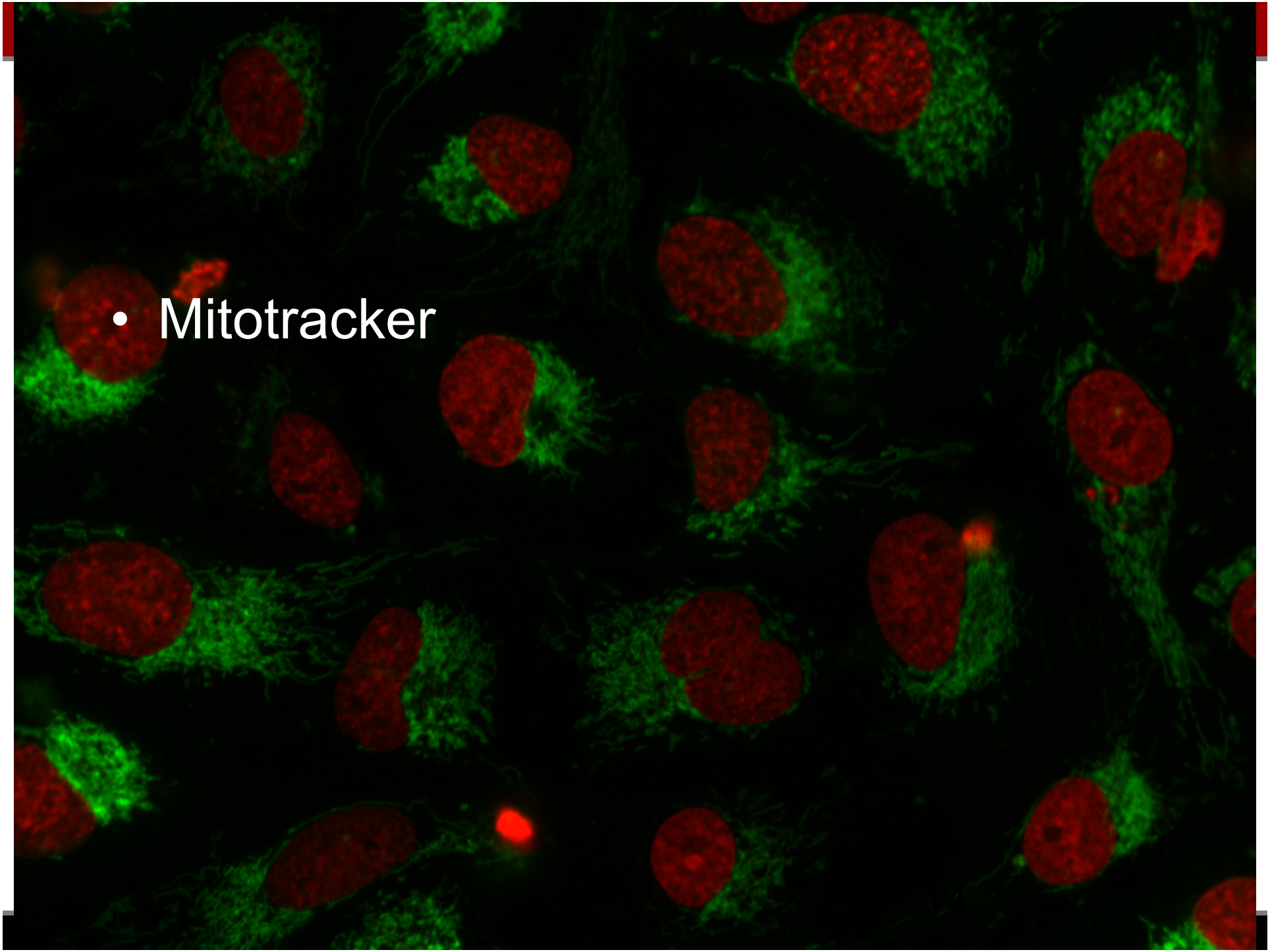
Test samples

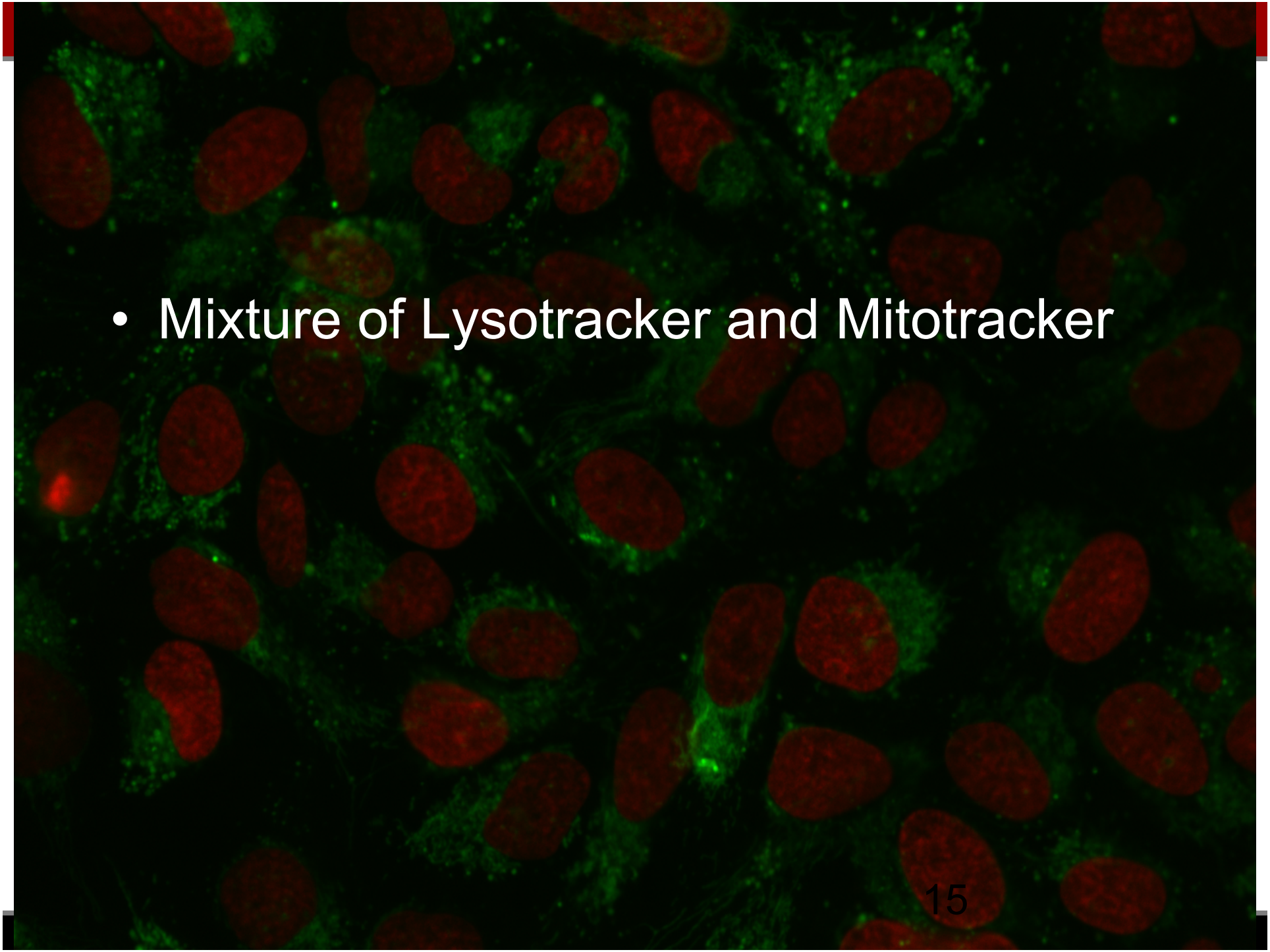
- How do we test a subcellular pattern unmixing algorithm?
- Need images of known mixtures of pure patterns – difficult to obtain “naturally”
- Created test set by mixing different proportions of two probes that localize to different cell parts (lysosomes and mitochondria)

Tao Peng, Ghislain Bonamy, Estelle Glory, Sumit Chanda, Dan Rines (Genome Research Institute of Novartis Foundation)



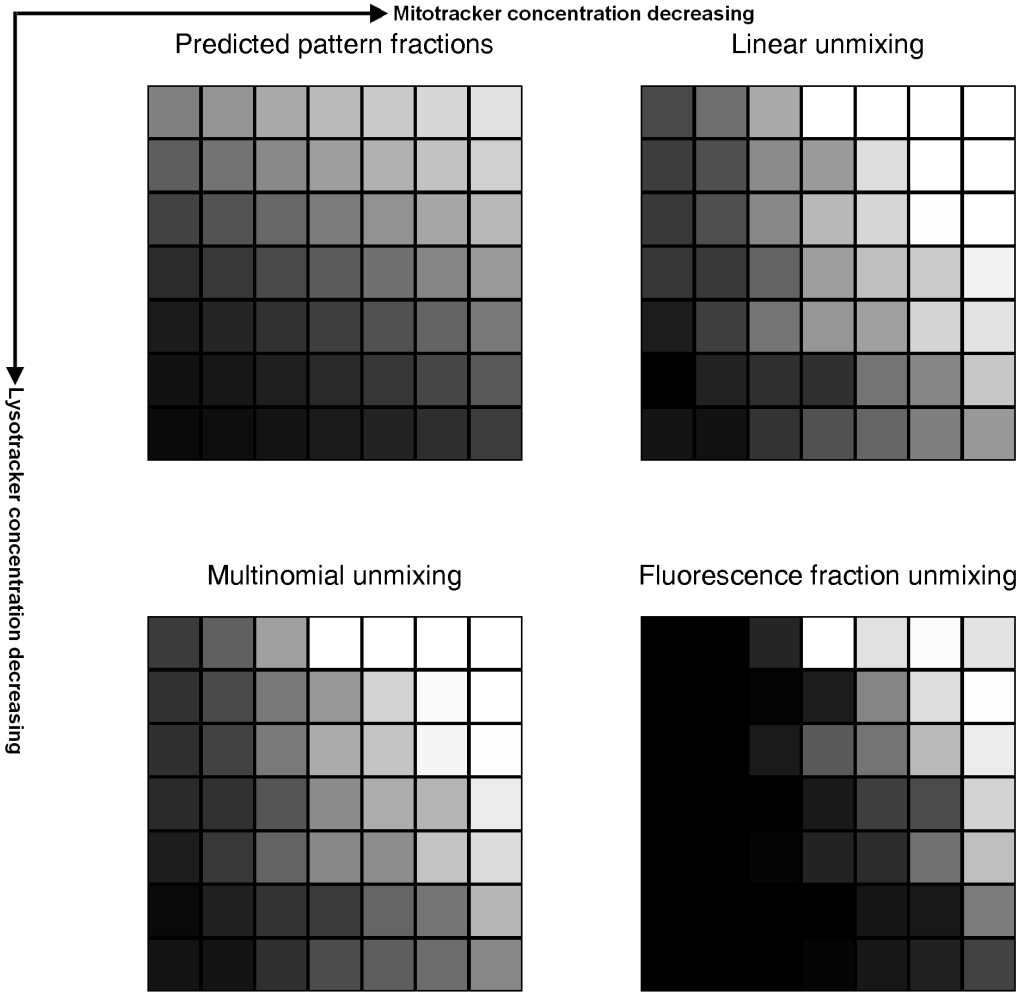
- Lysotracker


• Mitotracker

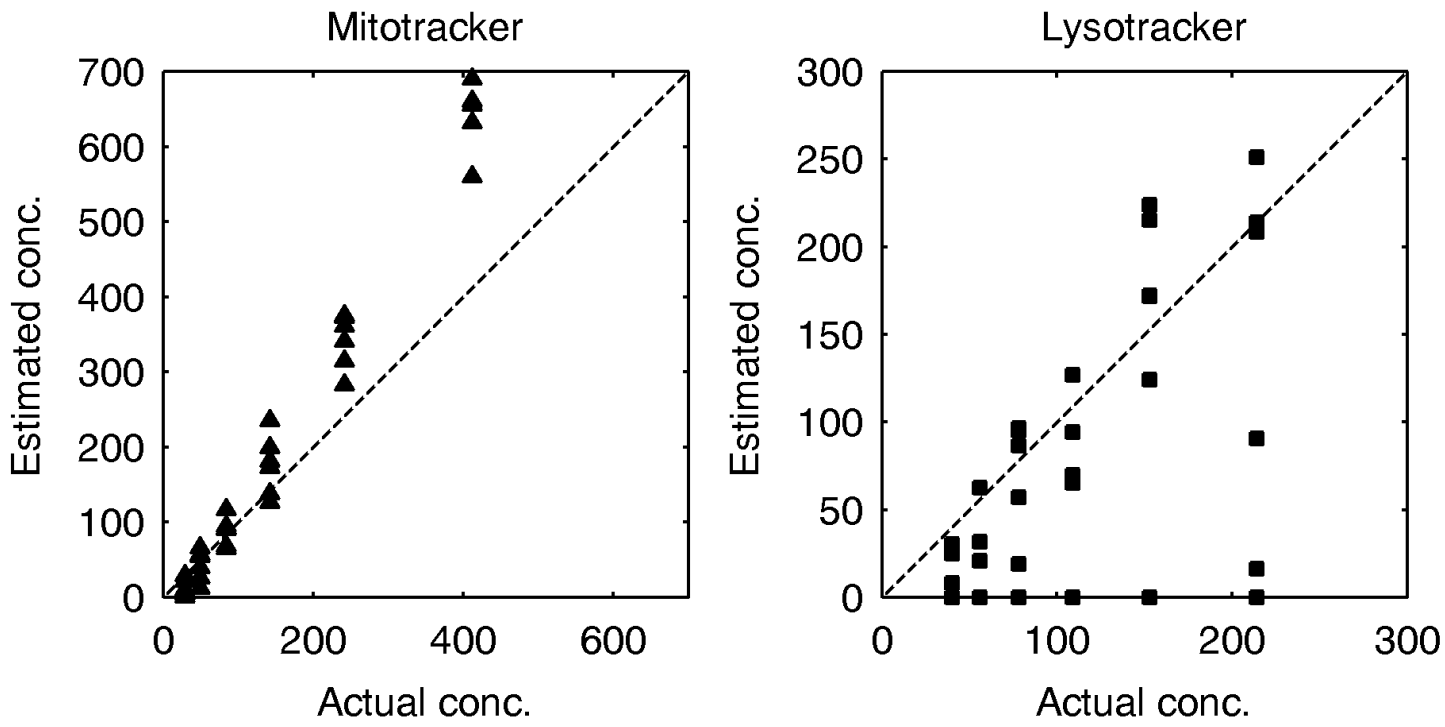

• Mixture of Lysotracker and Mitotracker

Proceedings of National Academic of Sciences, January 2010

Supervised unmixing results



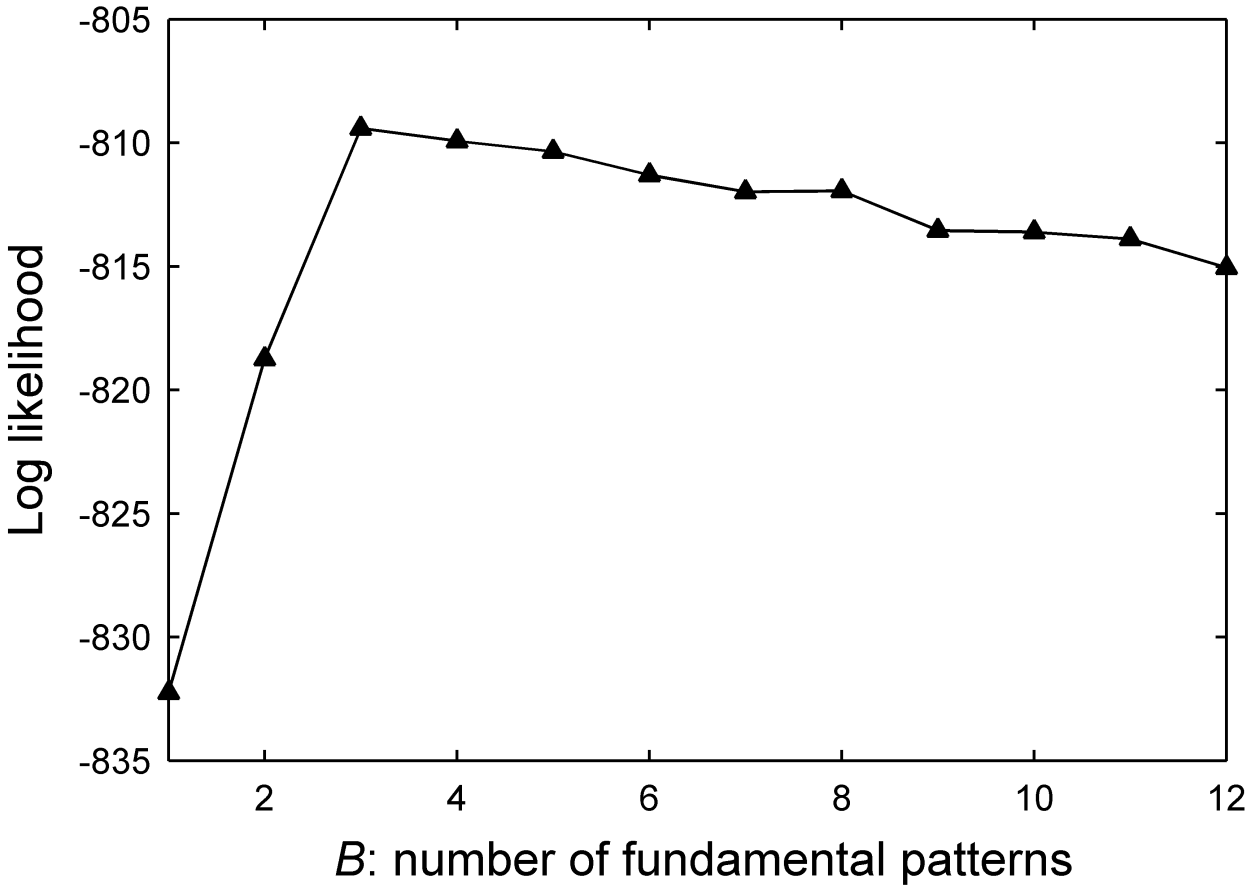
Supervised unmixing results



Unsupervised unmixing

- May not know all fundamental patterns, and/or may not have markers that are entirely in just one
- Solution: Simultaneously estimate the fundamental patterns and the fractions

Learning the number of fundamental patterns

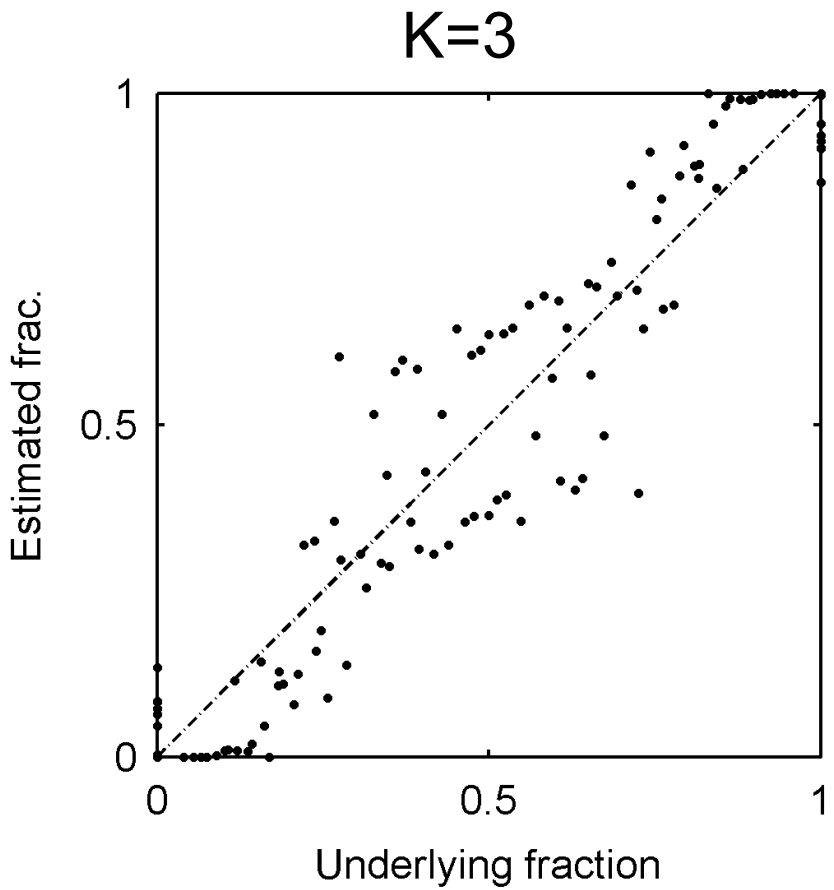


Finding fundamental patterns

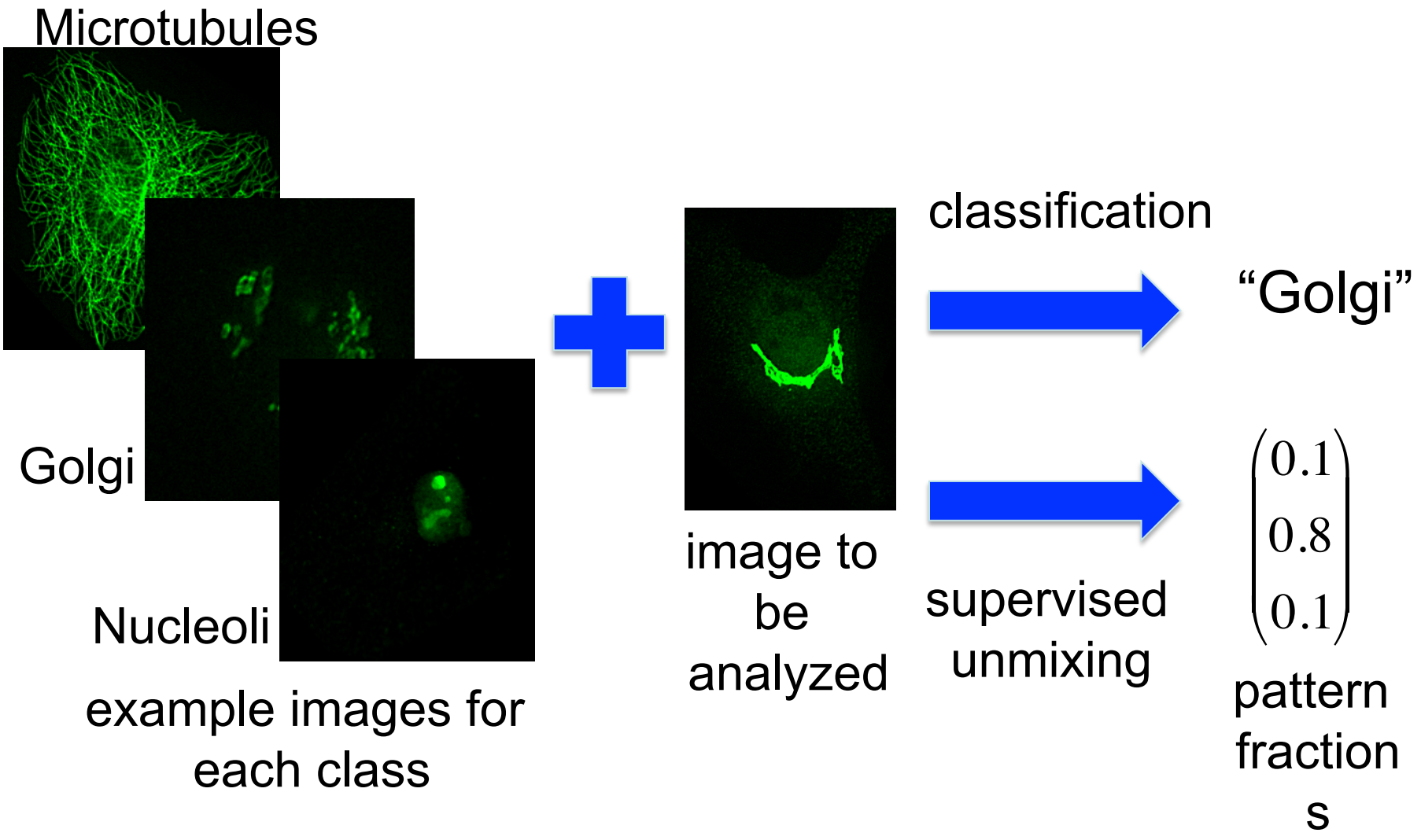
- Best fit obtained with three fundamental patterns
- Using those three patterns, calculate unmixing % for the “pure” images

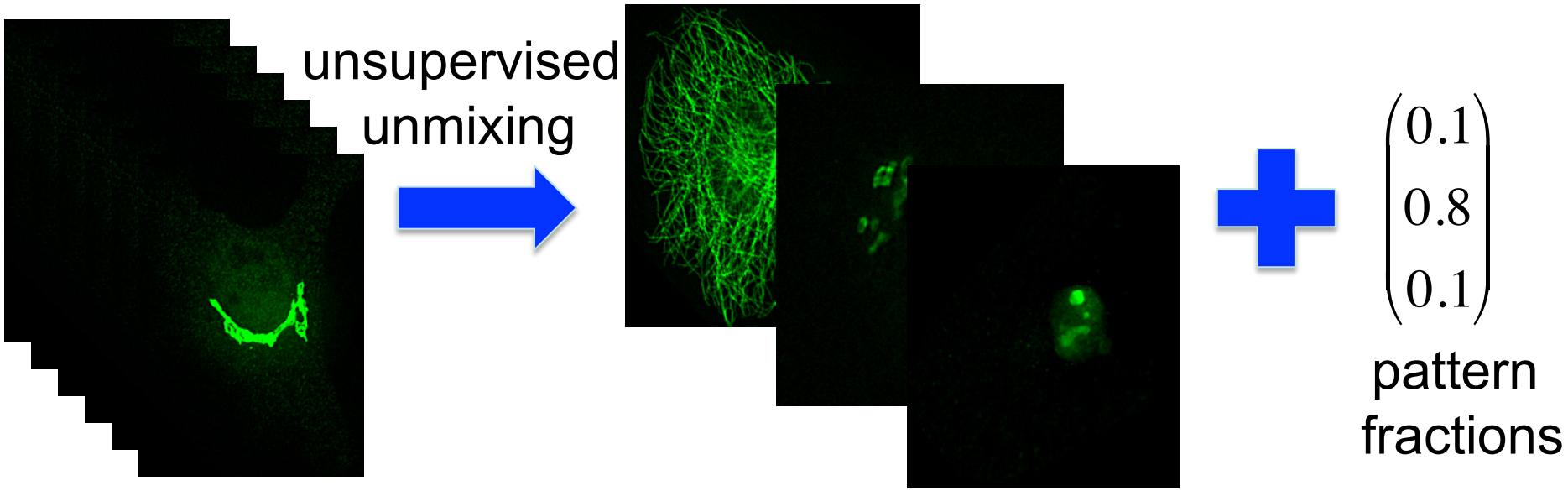
	Pure Mito Images	**Pure Lyso Images**
Pattern 0	0.0	0.0
Pattern 1	8.8	99.9
Pattern 2	91.2	0.1

Unsupervised unmixing results



Luis Coelho
Tao Peng

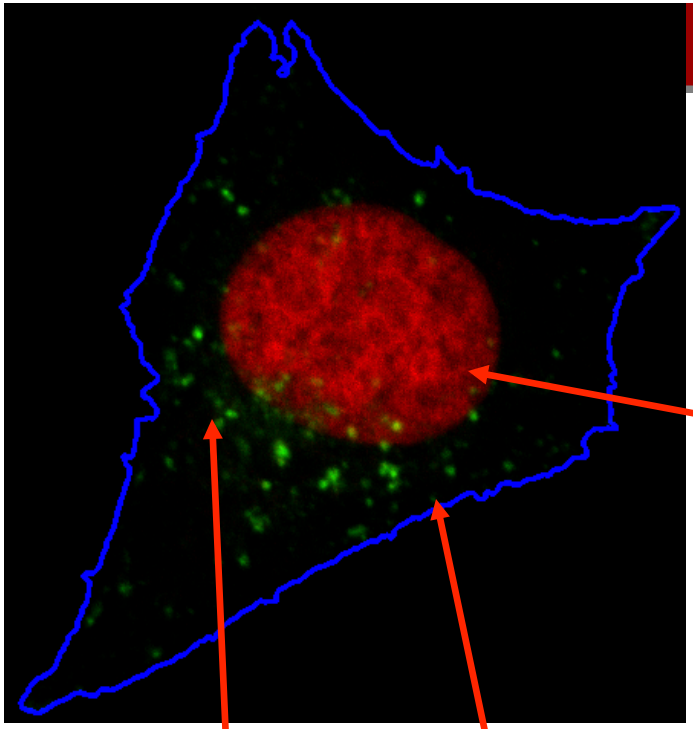

Microtubules
Golgi
Nucleoli
example images for each class
image to be analyzed
classification
"Golgi"
supervised unmixing
$\begin{pmatrix} 0.1 \\ 0.8 \\ 0.1 \end{pmatrix}$
pattern fractions


unsupervised
unmixing
0.1
0.8
0.1
pattern
fractions
collection
of images
fundamental classes

Generative models for communicating patterns

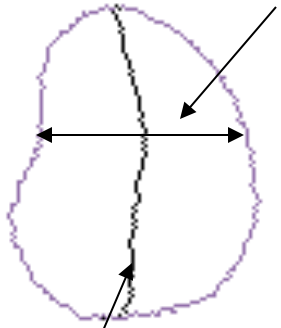
- How do we communicate results learned about subcellular patterns?
- Proposal: Use generative models learned from images to capture pattern and *variation* in pattern

Generative Cell Models



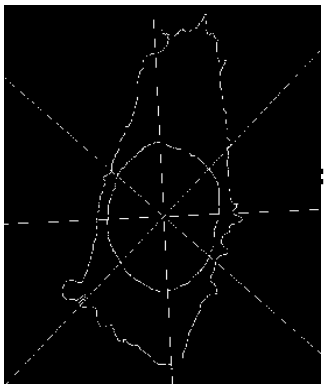
Nucleus

width



Medial axis

Cell membrane

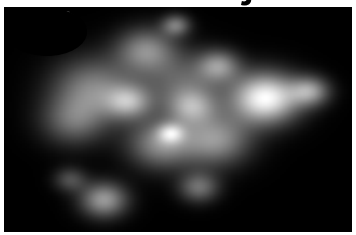


$$\frac{d_1 + d_2}{d_2}$$

Model parameters

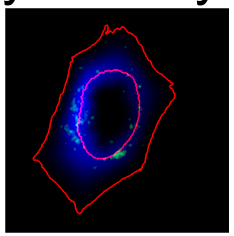
Protein distribution

Gaussian objects



+

Probability Density Function



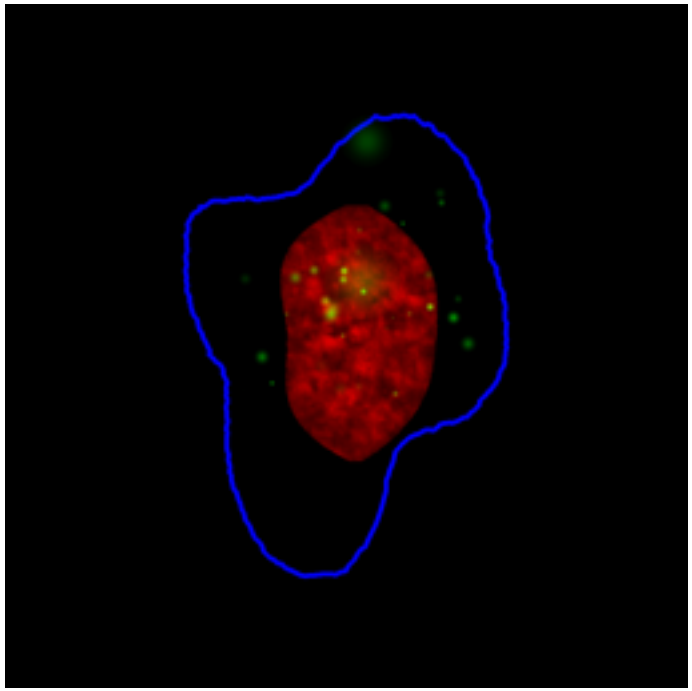
Zhao & Murphy 2007

Generative Cell Models

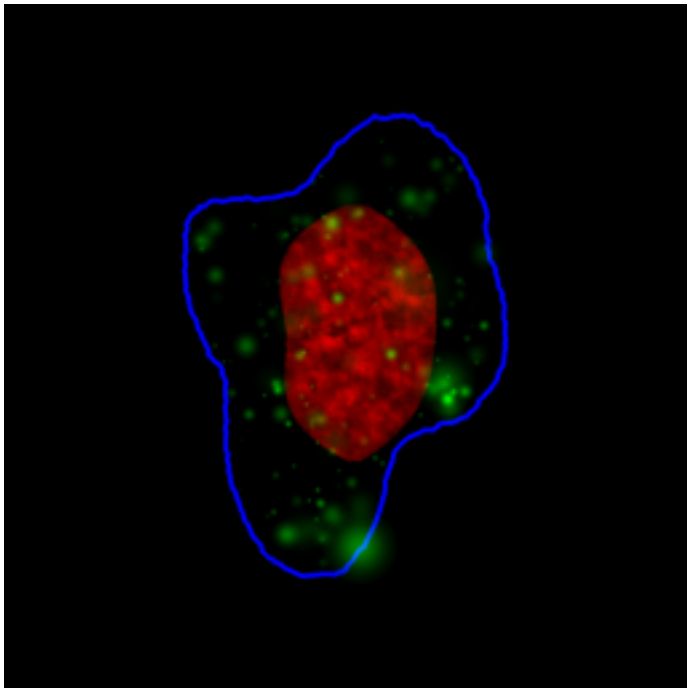
- Training
  - Learn the distributions of model parameters over many cells for each component of the model
- Synthesis
  - Randomly sample parameters from the distributions
  - Construct components using the sampled parameters

Synthesized nuclear shapes

Synthesized Images



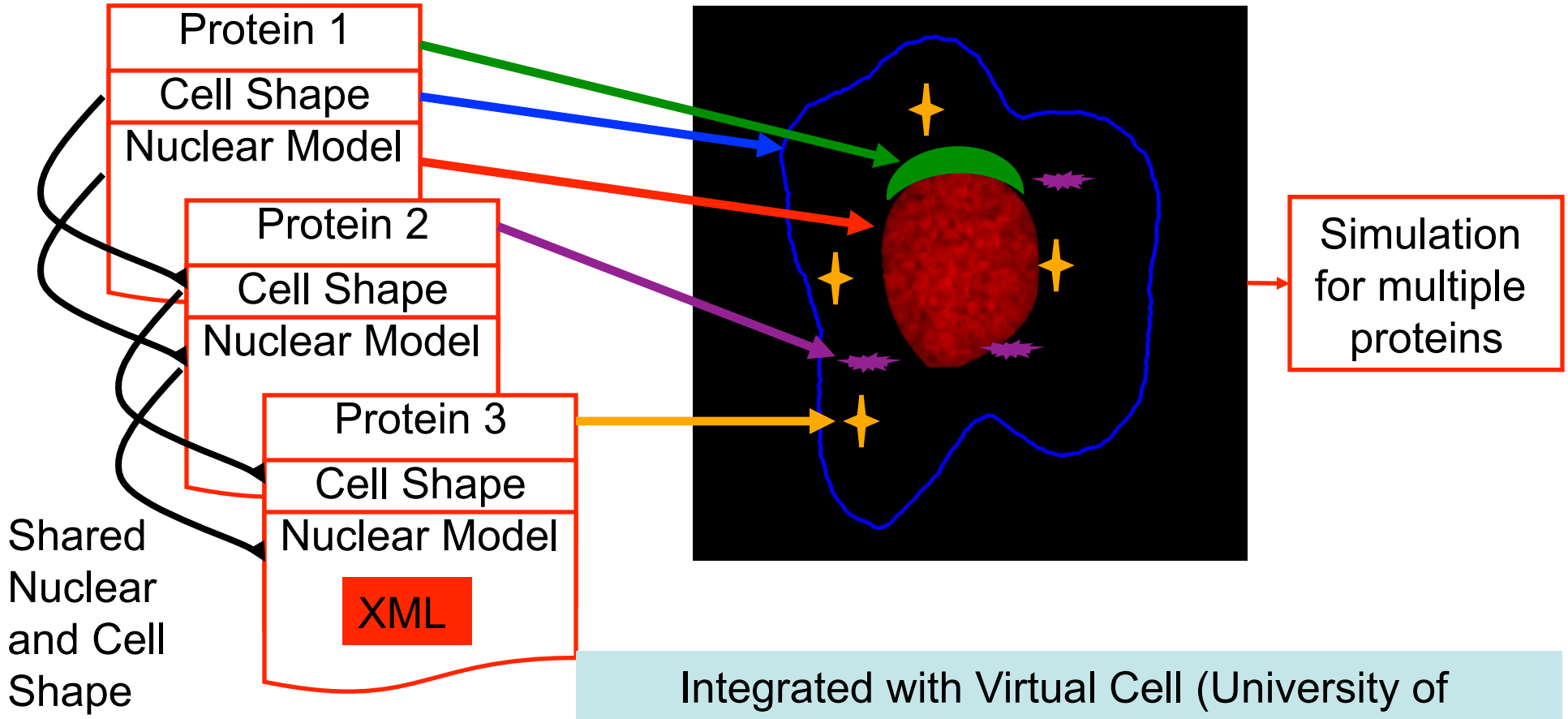
Lysosomes



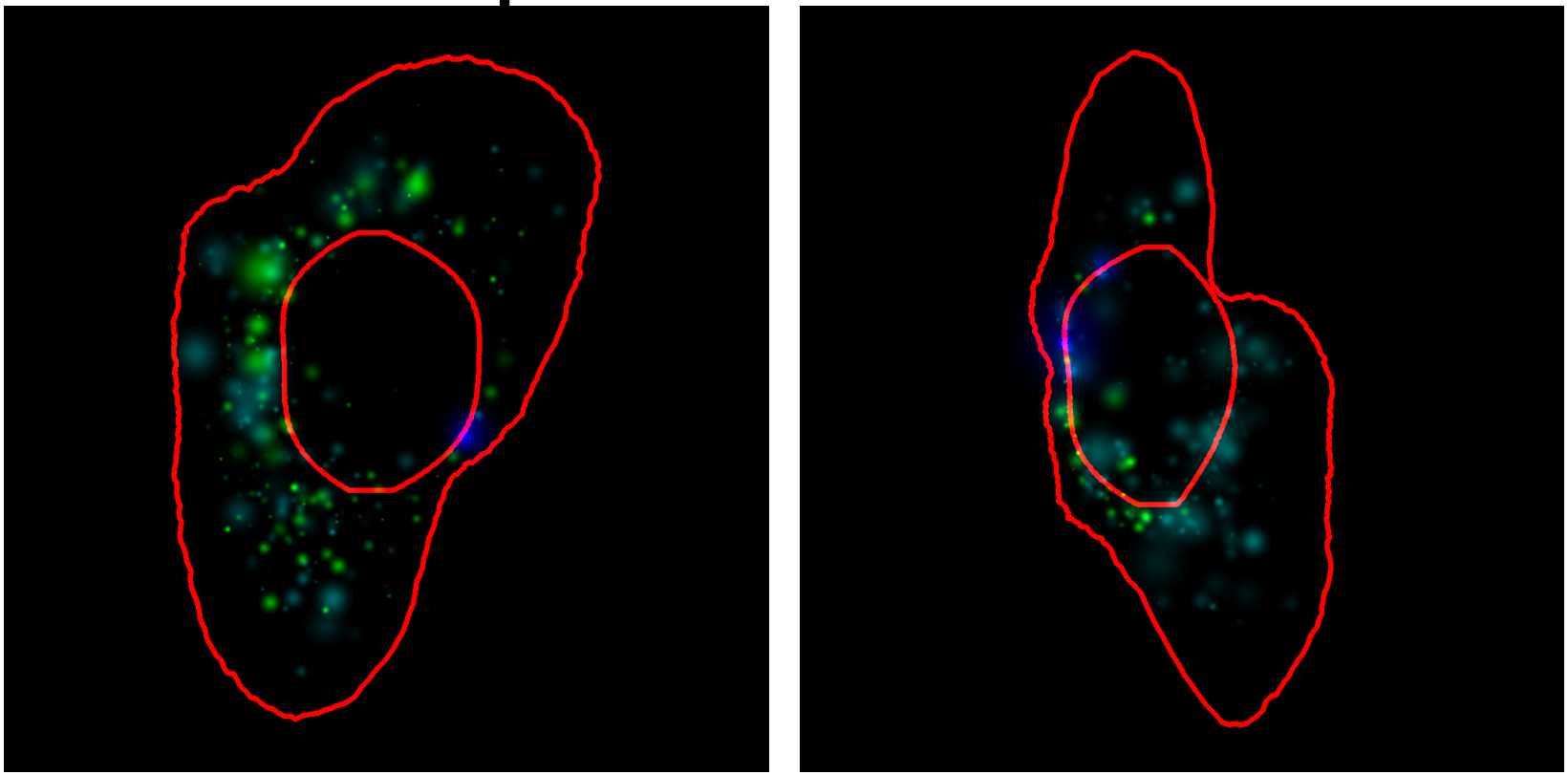
Endosomes

- SLML toolbox - Ivan Cao-Berg, Tao Peng, Ting Zhao
- Have portable tool for generating images from model

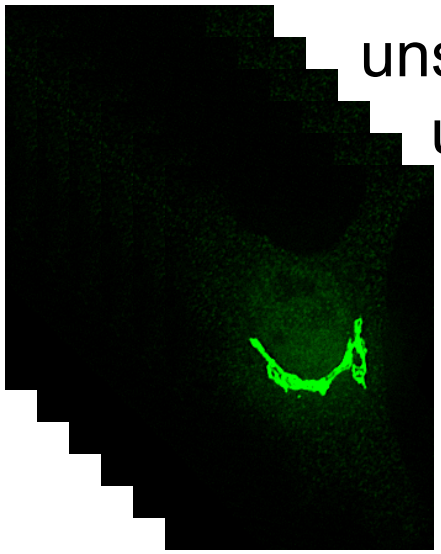
Combining Models for Cell Simulations



Example combinations

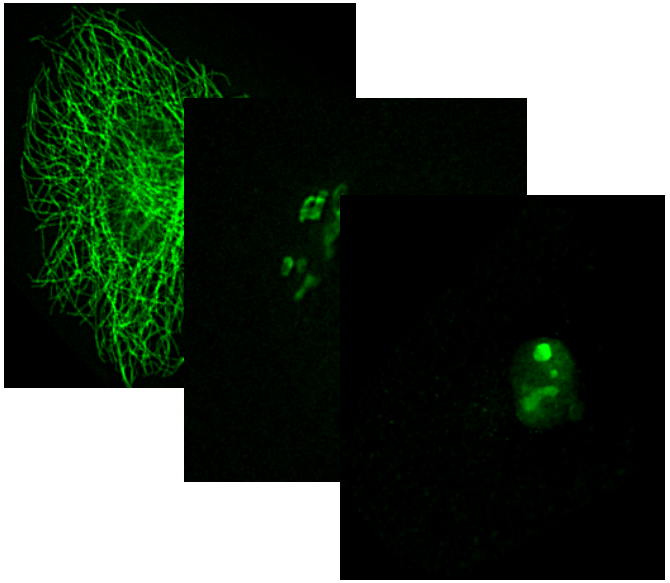


Red = nuclear membrane, plasma membrane
Blue = Golgi
Green = Lysosomes
Cyan = Endosomes



collection
of images

unsupervised
unmixing



fundamental classes

$$\begin{pmatrix} 0.1 \\ 0.8 \\ 0.1 \end{pmatrix}$$

pattern
fractions

$$\begin{pmatrix} w & x & y & z \end{pmatrix} + \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

generative model parameters

Cytometry PART A
Journal of the International Society for Advancement of Cytometry

Communicating Subcellular Distributions

Robert F. Murphy[1,2]*

[1]Lane Center for Computational Biology and Department of Biological Sciences, Carnegie Mellon University

[2]School of Life Sciences, Freiburg Institute for Advanced Studies, Albert Ludwig University of Freiburg

Received 19 May 2010; Accepted 24 May 2010

Grant sponsor: National Institutes of Health; Grant numbers: R01 GM068845, R01 GM075205; Grant sponsor: National Science Foundation; Grant number: EF-0331657.

*Correspondence to: Robert F. Murphy, 5000 Forbes Ave, Pittsburgh, Pennsylvania-15213, U.S.A

Abstract

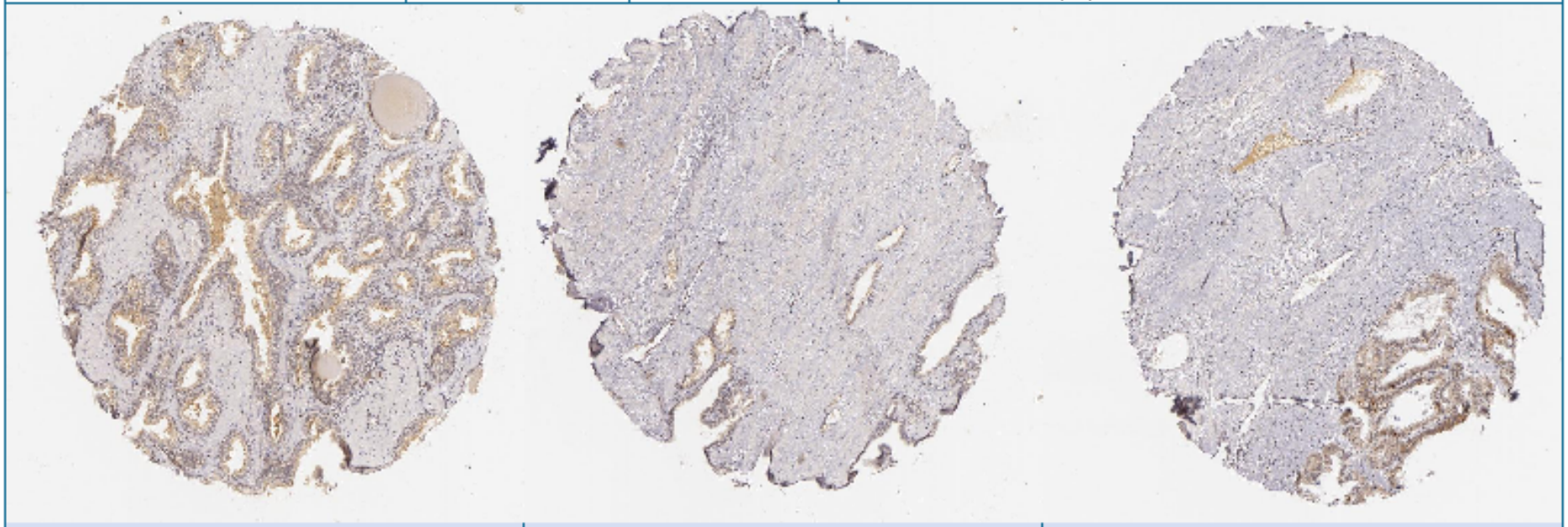
To build more accurate models of cells and tissues, the ability to incorporate accurate information on the distributions of proteins (and other macromolecules) will become increasingly important. This review describes current progress towards determining and representing protein subcellular patterns so that the information can be used as part of systems biology efforts. Approaches to decomposing an image of the subcellular pattern of a protein give critical information about the fraction of that protein in each of a number of fundamental patterns (e.g., organelles). Methods for learning generative models from images provide a means of capturing the essential properties and variation in those properties of cell shape and organelle patterns. The combination of models of fundamental patterns and vectors specifying the fraction of a protein in each of them provide a much better means of communicating subcellular patterns than the descriptive terms that are currently used. Communicating information about subcellular patterns is important not only for systems biology simulations but also for representing results from microscopy experiments, including high content screening and imaging flow cytometry, in a transportable and generalizable manner.

Human Protein Atlas

hpa | the project | protein atlas | dictionary | disclaimer | submission of antibodies

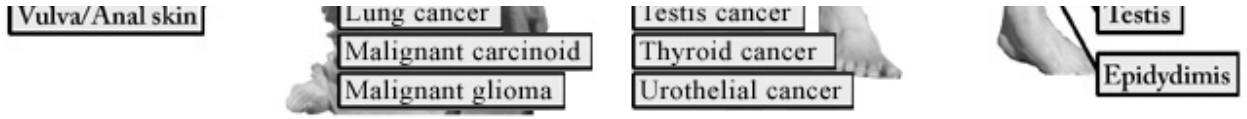
Prostate [CASP8]

Cell Type	Intensity	Quantity	Localization
Glandular cells	weak	>75%	cytoplasmic and/or membranous



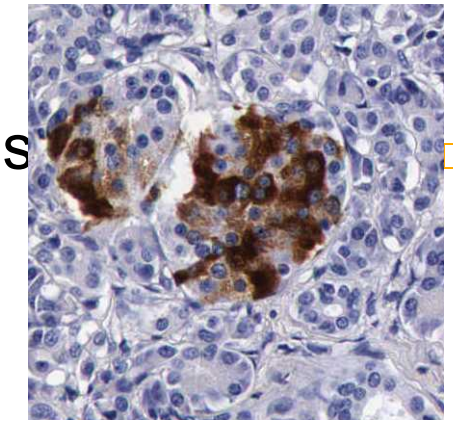
Male, age 51	Male, age 64	Male, age 60

Brown color indicates presence of protein, blue color shows cell nuclei. Image Usage Policy

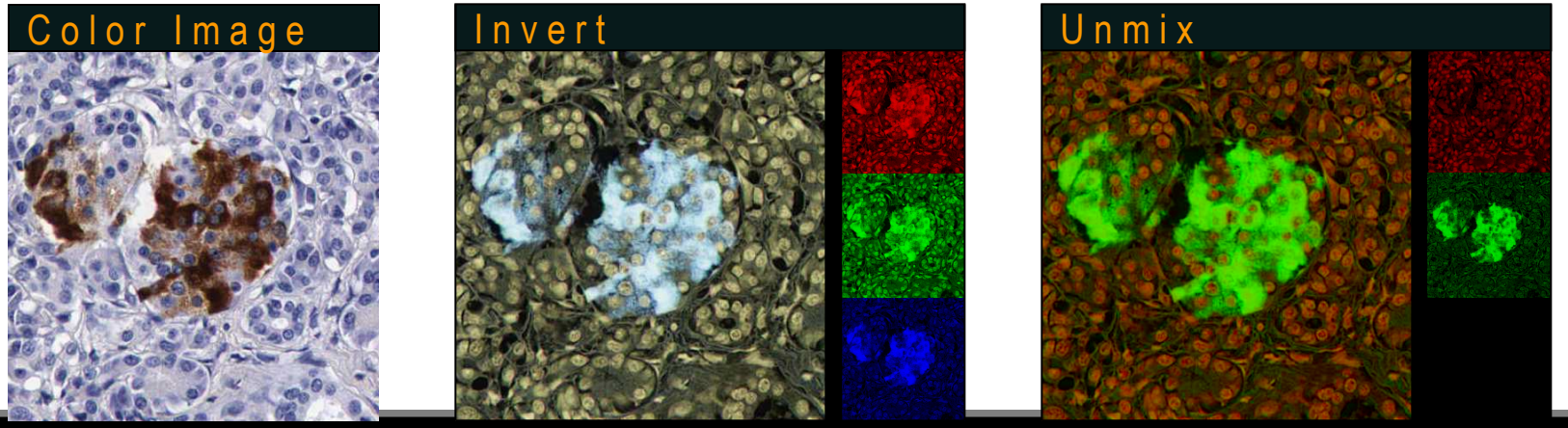


Immunocytochemistry Signal Unmixing

- The Haematoxylin and DAB stains are imaged together
- Each stain contains multiple sources
  - Haem. = $n_1R + n_2G + n_3B$
  - DAB = $n_1R + n_2G + n_3B$
- Use two unmix methods to find w's



R	G	B
28	22	36
28	23	29
28	23	55
28	24	21
28	24	23
28	24	25
28	25	18
28	25	20
28	34	66
28	44	34
28	48	75



Cell segmentation?

- Most analysis of subcellular patterns has used images segmented into single cell regions
- Results on yeast and cultured cells show classification of basic patterns can be achieved without segmentation
- Given difficulty of segmenting tissue images into cells, used field features

Test Dataset from Human Protein Atlas

- Selected 16 proteins from the Atlas
- Two each from all major organelles (class)
- ~45 tissue types for each class (e.g. liver, skin)
- Goal: Train classifier to recognize each subcellular pattern across all tissue types
- Use object and texture features

Justin Newberg

Pattern Classification over 45 tissues

Labels	Prediction							
	ER	Cyto	Endo	Golgi	Lyso	Mito	Nucleolus	Nucleus
ER (131)	**83.2**	7.6	3.1	1.5	2.3	0.8	1.5	0
Cyto (125)	14.4	**64**	3.2	0	10.4	7.2	0	0.8
Endo (111)	8.1	9.9	**75.7**	0	2.7	0	0	3.6
Golgi (126)	1.6	0	0	**87.3**	1.6	0	9.5	0
Lyso (127)	3.9	9.4	1.6	7.9	**75.6**	0	0.8	0.8
Mito (125)	3.2	4	0	3.2	0.8	**85.6**	1.6	1.6
Nucleolus (120)	0.8	0	0	5.8	4.2	1.7	**87.5**	0
Nucleus (117)	0	0.9	8.5	1.7	0	0.9	0	**88**

Overall accuracy 81%

Newberg & Murphy, 2008

Pattern Classification over 45 tissues

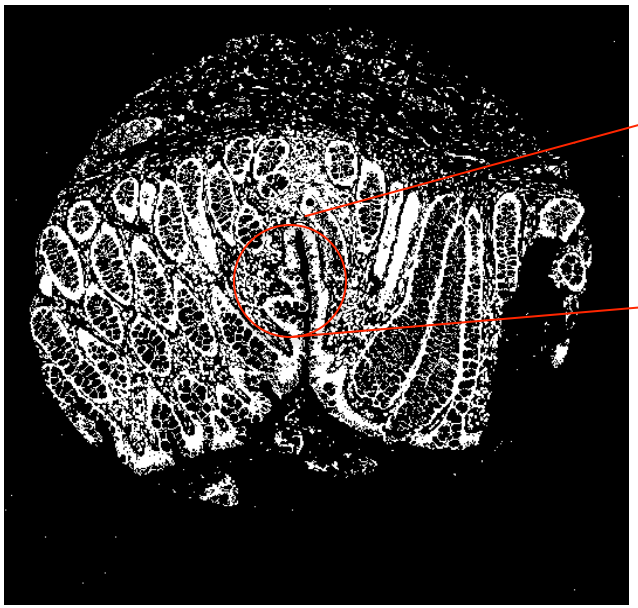
	Prediction							
	ER	Cyto	Endo	Golgi	Lyso	Mito	Nucleolus	Nucleus
ER (53)	**100**	0	0	0	0	0	0	0
Cyto (21)	4.8	**76.2**	0	0	14.3	4.8	0	0
Endo (2)	0	0	**100**	0	0	0	0	0
Golgi (88)	1.1	0	0	**98.9**	0	0	0	0
Lyso (52)	0	1.9	0	0	**96.2**	0	1.9	0
Mito (64)	0	0	0	0	0	**98.4**	1.6	0
Nucleolus (94)	0	0	0	2.1	2.1	1.1	**94.7**	0
Nucleus (78)	0	0	0	0	0	0	0	**100**

Accuracy for 50% of images with highest confidence: 97%

Test large set of proteins for a single tissue

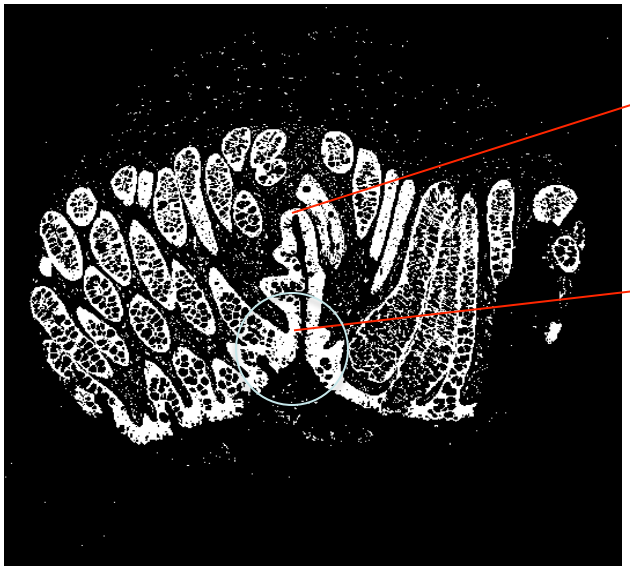
- Analyze images of 1208 proteins from adrenal glands
- Use features that measure texture as well as spatial relationship between protein and hematoxylin staining

Proximity features for a mitochondrial protein



DNA channel

3000 x 3000



Protein Channel

3000 x 3000

Proximity features are computed on a circular window of radius 150 pixels.

Features

- Commute Time of nodes in protein graph to hematoxylin graph
- Cluster Validity indices
- Haralick features
- Spatial Statistics of DNA and Protein point-sets

Classification results for 11 classes

- Cytoplasm
- Endoplasmic Reticulum
- Golgi
- Intermediate Filament
- Lysosome
- Membrane
- Microtubule
- Mitochondria
- Nucleus
- Peroxisome
- Secreted

The class for each protein was obtained from UNIPROT GO annotations (assumed correct)

Obtained classification acccuracy of 91-96%.

Arvind Rao

Automated detection of cancer markers

- Human Protein Atlas also contains images of all common solid tumors
- Comparing patterns of all proteins across all tumors to find proteins whose patterns discriminate tumors from normal tissue

Estelle Glory-Afshar

Comparing normal and tumor images

- Compare images of ~200 proteins in normal and cancerous prostate tissue
- 1 Protein different only in protein pattern: PPARG: Peroxisome proliferator-activated receptor gamma (PPAR-gamma).

[CANCER RESEARCH 58, 3344–3352, August 1, 1998]

Ligand for Peroxisome Proliferator-activated Receptor γ (Troglitazone) Has Potent Antitumor Effect against Human Prostate Cancer Both *in Vitro* and *in Vivo*[1]

Tetsuya Kubota,[2] Kozo Koshizuka, Elizabeth A. Williamson, Hiroya Asou, Jonathan W. Said, Stuart Holden, Isao Miyoshi, and H. Phillip Koeffler

Division of Hematology/Oncology [T. K., K. K., E. A. W., H. A., H. P. K.], Division of Pathology [J. W. S.], Department of Surgery [S. H.], Cedars-Sinai Research Institute, UCLA School of Medicine, Los Angeles, California 90048, and Department of Medicine, Kochi Medical School, Kochi 783-8505, Japan [I. M.]

Effects of ligand activation of peroxisome proliferator-activated receptor γ in human prostate cancer

Elisabetta Mueller*, Matthew Smith†, Pasha Sarraf*, Todd Kroll‡, Anita Aiyer*, Donald S. Kaufman†, William Oh§, George Demetri§, William D. Figg¶, Xiao-Ping Zhou‖, Charis Eng‖, Bruce M. Spiegelman*,**, and Philip W. Kantoff§**

*Dana-Farber Cancer Inst
General Hospital, Harvar
Boston, MA 02115; §Depa
Boston, MA 02115; ¶Med
Cancer Genetics Program

Communicated by Baruj I

Peroxisome proliferat
hormone receptor tha
adipocytes. Activatior
and colon cancer cel
differentiation. In th
expressed in human
rived from these tum
ligands exerts an inhib
cell lines. Further, we
not have intragenic m
the informative tumo
Based on our preclinic
in patients with adva
PPARγ ligand used for

Clinical Study

Peroxisome Proliferator-Activated Receptor-γ Polymorphism, Body Mass and Prostate Cancer Risk: Evidence for Gene-Environment Interaction

Potential biomarkers

- Currently continuing analysis of proteins whose subcellular location features differ between normal and tumor

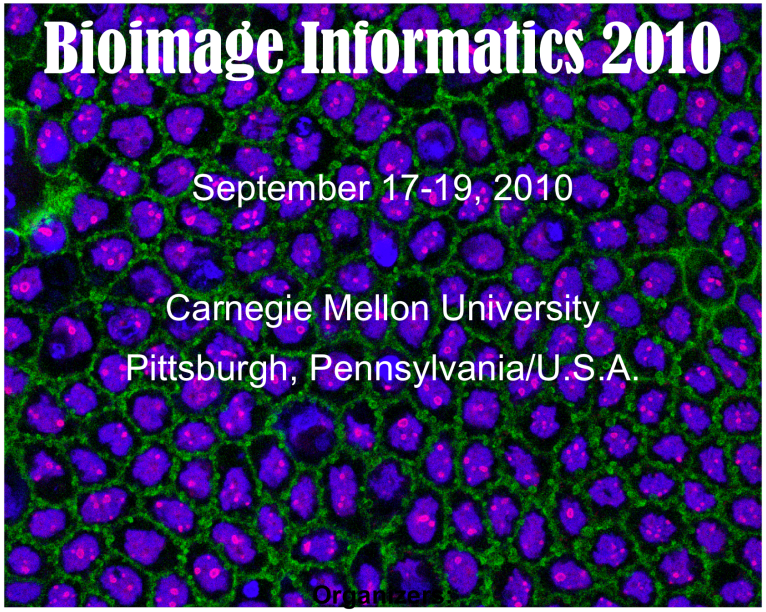
Acknowledgments

- Past and Present Students and Postdocs
  - Michael Boland (Hopkins), Mia Markey (UT Austin), Gregory Porreca (Harvard), Meel Velliste (U Pitt), Kai Huang, Xiang Chen (Yale), Yanhua Hu, Juchang Hua, **Ting Zhao (HHMI Janelia Farm)**, Shann-Ching Chen (Scripps), Elvira Garcia Osuna (CMU), **Justin Newberg**, **Estelle Glory**, **Tao Peng**, **Luis Coelho**, **Aabid Sharif, Rumi Naik, Josh Kangas, Arvind Rao, Jieyue Li**

- Funding
  - NSF, NIH, Commonwealth of Pennsylvania
  - Molecular Biosensors and Imaging Center - TCNP (Waggoner)

- Collaborators/Consultants
  - David Casasent, Simon Watkins, **Jon Jarvik, Peter Berget**, Jack Rohrer, Tom Mitchell, Christos Faloutsos, **Jelena Kovacevic, William Cohen, Geoff Gordon**, B. S. Manjunath, Ambuj Singh, Les Loew, Ion Moraru, Jim Schaff, Paul Campagnola, **Gustavo Rohde, Ghislain Bonamy, Sumit Chanda, Dan Rines, Chris Langmead, Klaus Palme**

Bioimage Informatics 2010



September 17-19, 2010

Carnegie Mellon University

Pittsburgh, Pennsylvania/U.S.A.

Organizers:

Steven Altschuler, University of Texas Southwestern Medical Center

Gaudenz Danuser, Harvard Medical School

Michael Hawrylycz, Allen Institute for Brain Science

Robert F. Murphy, Carnegie Mellon University

Lani Wu, University of Texas Southwestern Medical Center

Hosted by

Center for Bioimage Informatics
from image to knowledge

RAY AND STEPHANIE LANE
Center for Computational Biology
Carnegie Mellon

Local Organizers:

Jelena Kovačević, Gustavo Rohde, Ge Yang

http:// lane.compbio.cmu.edu/bii2010