Automated construction of generative models from time series cell images: Tools for more complete analysis of perturbagen effects

Robert F Murphy

Ray & Stephanie Lane Professor of Computational Biology and
Professor of Biological Sciences, Biomedical Engineering and Machine Learning
Carnegie Mellon University
External Senior Fellow, FRIAS, University of Freiburg

RAY AND STEPHANIE LANE
Center for Computational Biology
Carnegie Mellon

ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

FRIAS
FREIBURG INSTITUTE
FOR ADVANCED STUDIES

LIFE SCIENCES - LIFENET

Overview

- Input assumed to be 2D or 3D movie (single or multichannel)
- Want to detect and model perturbations
  - Temporal pattern feature changes
  - Object type composition changes
  - Object type proximity changes

TEMPORAL PATTERN FEATURES

BIOINFORMATICS ORIGINAL PAPER

Vol. 26 no. 13 2010, pages 1630–1636
doi:10.1093/bioinformatics/btq239

Systems biology

Advance Access publication May 19, 2010

Automated analysis of protein subcellular location in time series images

Yanhua Hu[1,2], Elvira Osuna-Highley[1,3], Juchang Hua[1,2,4], Theodore Scott Nowicki[1,2], Robert Stolz[5], Camille McKayle[5] and Robert F. Murphy[1,2,3,4,6,*]

[1]Center for Bioimage Informatics, [2]Department of Biological Sciences, [3]Department of Biomedical Engineering, [4]Department of Machine Learning, Carnegie Mellon University, Pittsburgh, PA 15213, [5]Division of Science and Mathematics, University of the Virgin Islands, St Thomas, VI 00803 and [6]Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

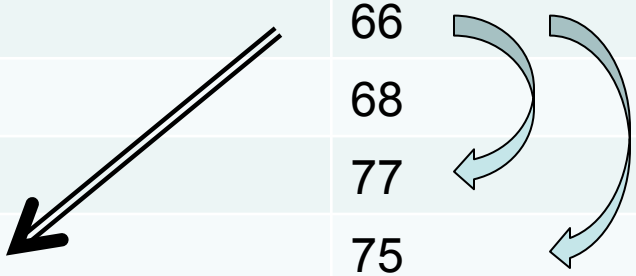
Descriptive approach

- Calculate features that measure temporal patterns
  - Object tracking
  - Temporal texture
  - Normal (optical) flow
  - Fourier transform
  - Autoregression
- Use for classification or clustering

Results

- Evaluate ability to classify movies of 12 different GFP-tagged proteins

Temporal feature type	Without static features	With static features
None	-	66
Object tracking	nd	68
Temporal texture	66	77
Normal flow	75	75
Fourier transform	60	69
Autoregression	nd	59



Speed and utility

- Object tracking (slow)
- Temporal texture (fast, accuracy++)
- Normal flow (slow, accuracy++)
- Fourier transform (fast)
- Autoregression (fast)

- Can distinguish or group movies by their temporal patterns using these features
- As with most feature-based methods, limited ability to interpret differences

OBJECT TYPE COMPOSITION MODELS

Movie Analysis via Object Type Changes

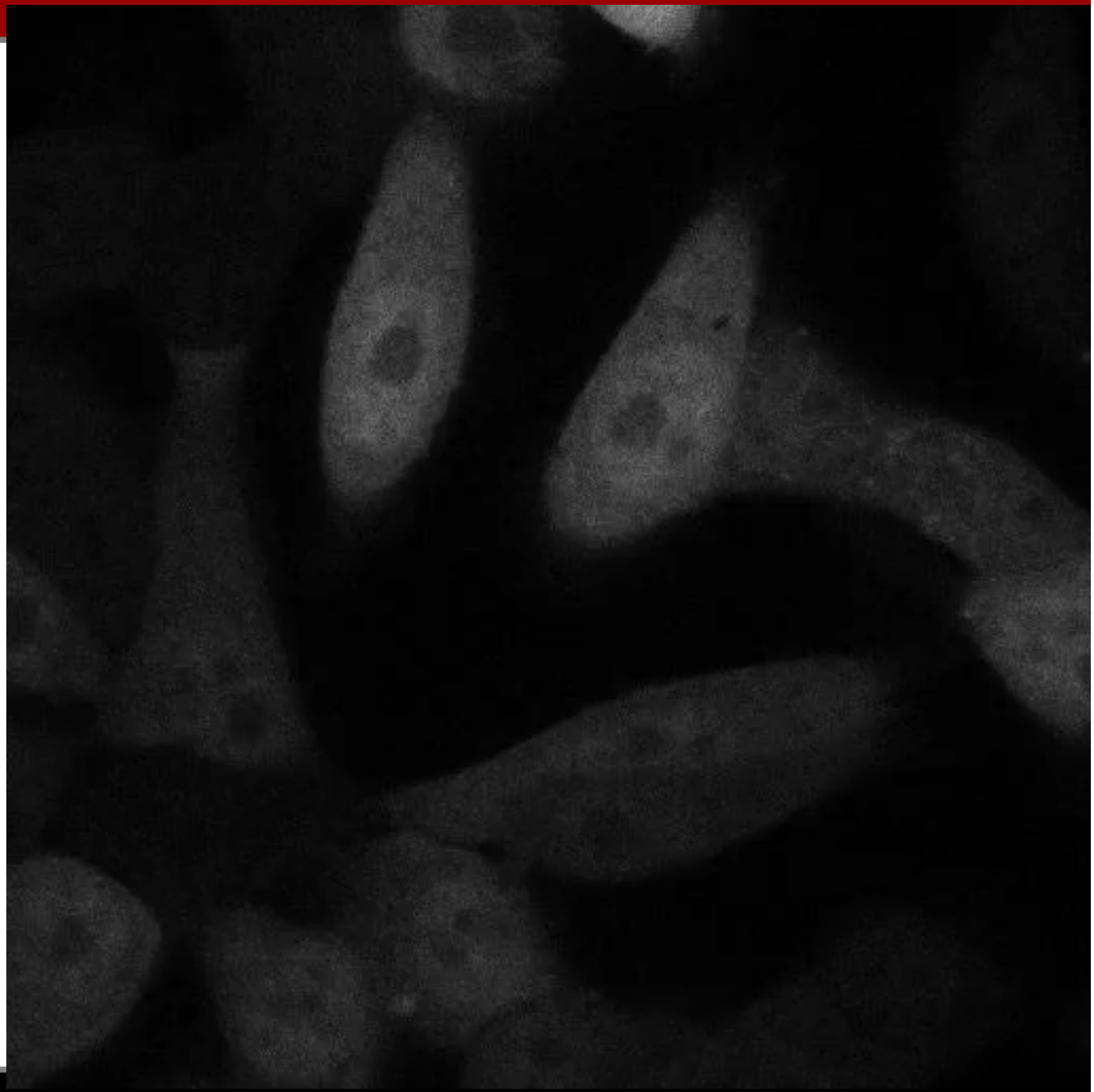
- HeLa cells expressing GFP-tagged growth factor receptor-bound protein 2 (Grb2)
- TGF added at t=0
- 3D movie over 9.2 minutes
- 8 sec / frame
- Alexander Sorkin group,
  Univ. Pittsburgh School Medicine

HeLa cells expressing growth factor receptor-bound protein 2(Grb2)

TGF added at t=0

single slice from 3D movie

A. Sorkin U. Pitt

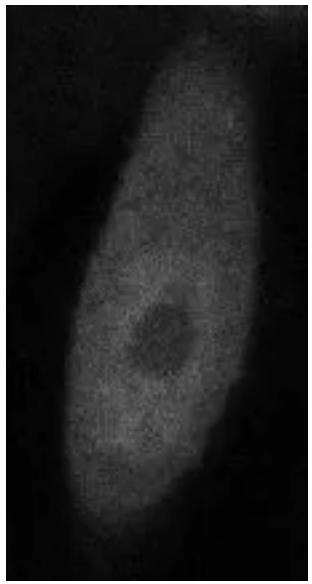


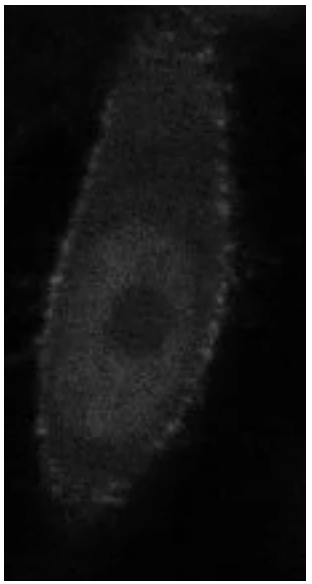
Three “patterns” from visual analysis

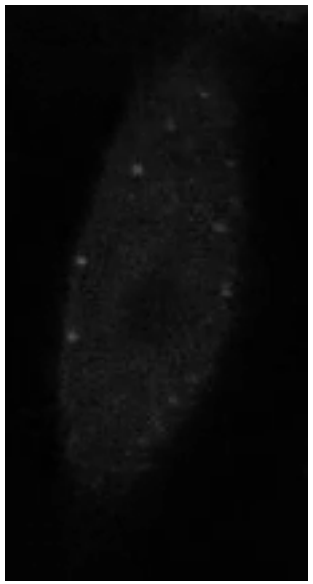
Cytosol

Plasma membrane/
Coated Pits/Vesicles

Internal
Vesicles







Goal

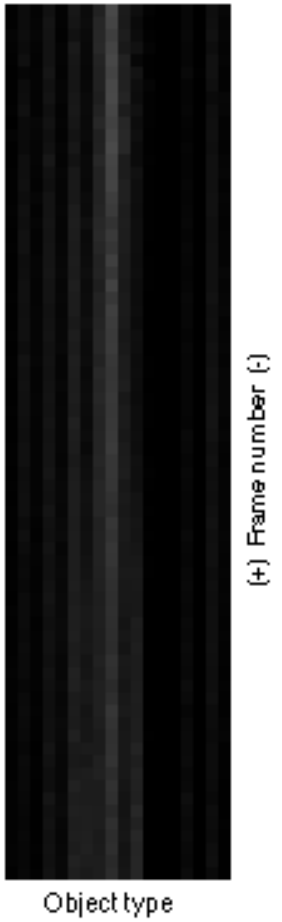
- Analyze temporal dependence of pattern changes with minimal assumptions
- Major assumption: Patterns representable by composition of objects
- “Bag of visual words”
- Can be calculated “on the fly”

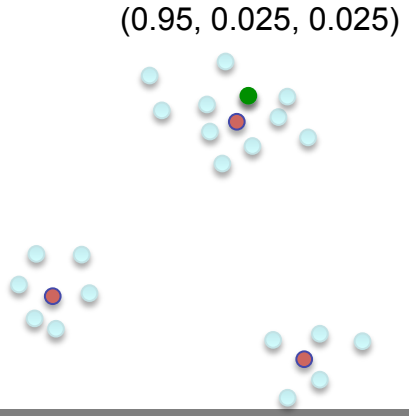
Method

- Segment objects with adaptive thresholding
- Cluster objects by geometric features
- Describe frame as a vector of object type proportions
- Cluster vectors (specify number of clusters)
- Fraction of each pattern contained in model as normalized inverse relative distance to each cluster centroid
- Cluster fraction vectors

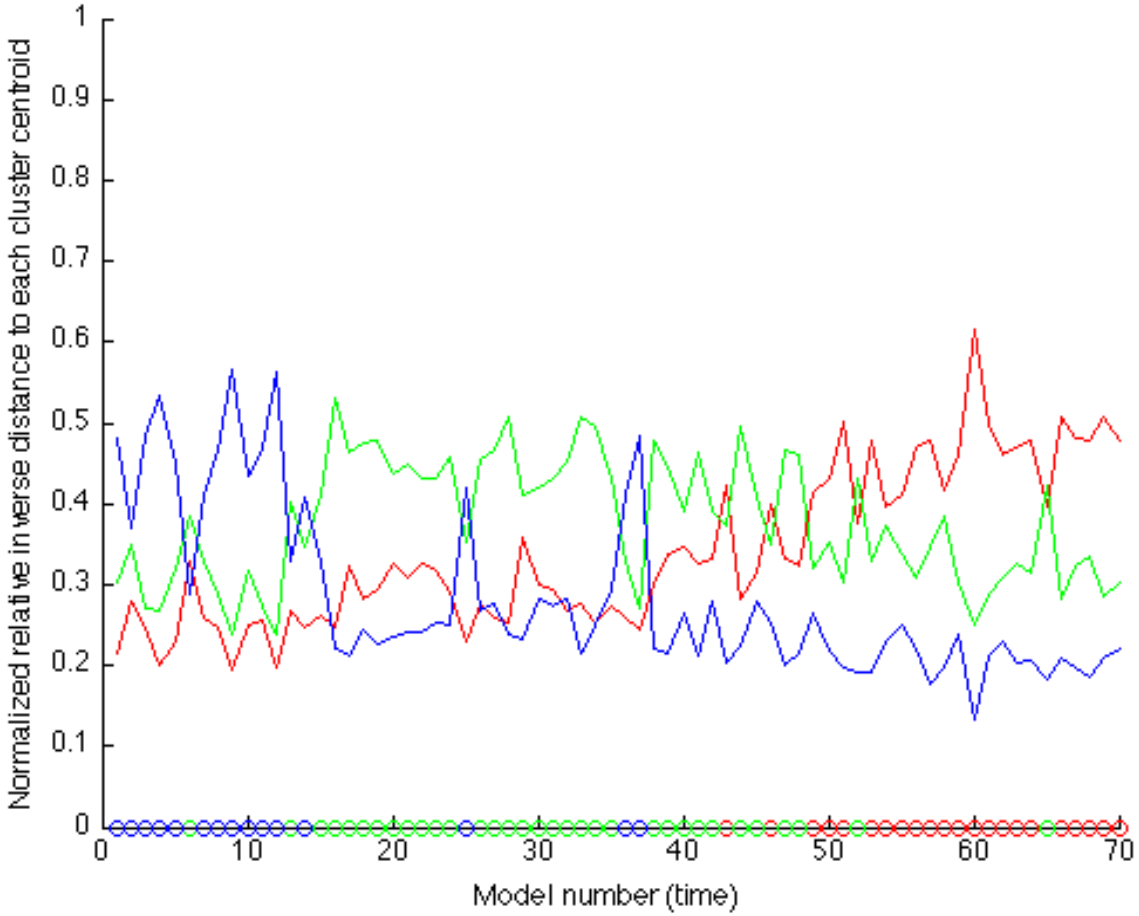
$$ird_i = \left(\sum_{j=1}^{k} d_j \right) / d_i$$

$$nird_i = ird_i / \sum^{k} ird_j$$





Capturing phases of pattern changes



Greg Johnson

Devin Sullivan

- Can find and visualize temporal pattern changes
- Still descriptive

OBJECT TYPE TRANSITION MODELS

BIOINFORMATICS ORIGINAL PAPER

Vol. 27 no. 13 2011, pages 1854–1859
doi:10.1093/bioinformatics/btr286

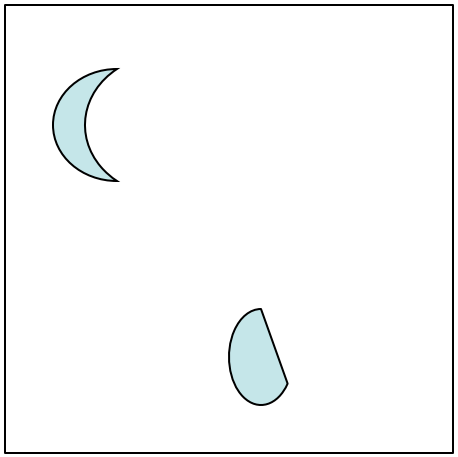
Systems biology

Advance Access publication May 9, 2011

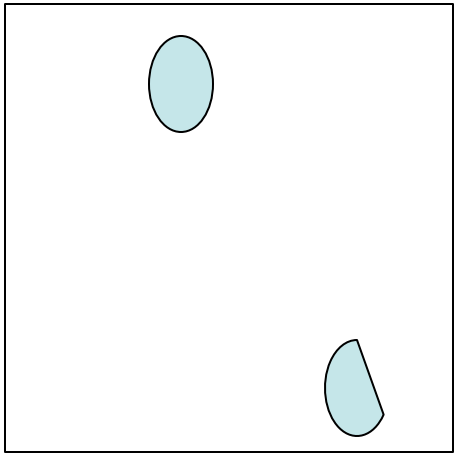
Model building and intelligent acquisition with application to protein subcellular location classification

C. Jackson[1,2], E. Glory-Afshar[1,2], R. F. Murphy[1,2,3,4,5] and J. Kovačević[1,2,3,6,*]

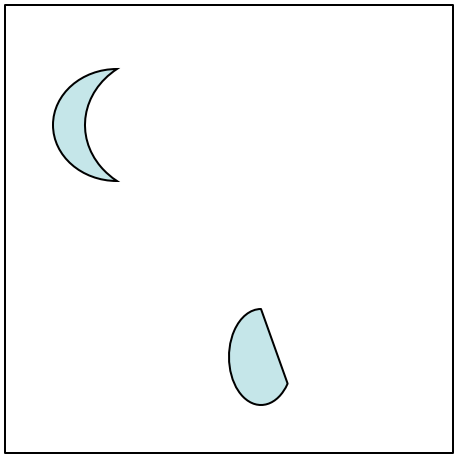
[1]Center for Bioimage Informatics, [2]Department of Biomedical Engineering, [3]Lane Center for Computational Biology, Carnegie Mellon University, 5000 Forbes Ave., [4]Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Ave., [5]Machine Learning Department, Carnegie Mellon University and [6]Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA



t=0



t=1

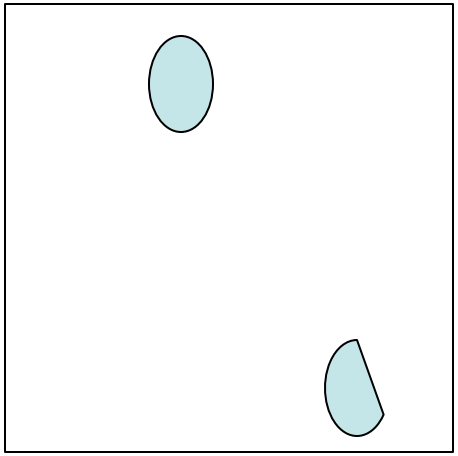


t=0

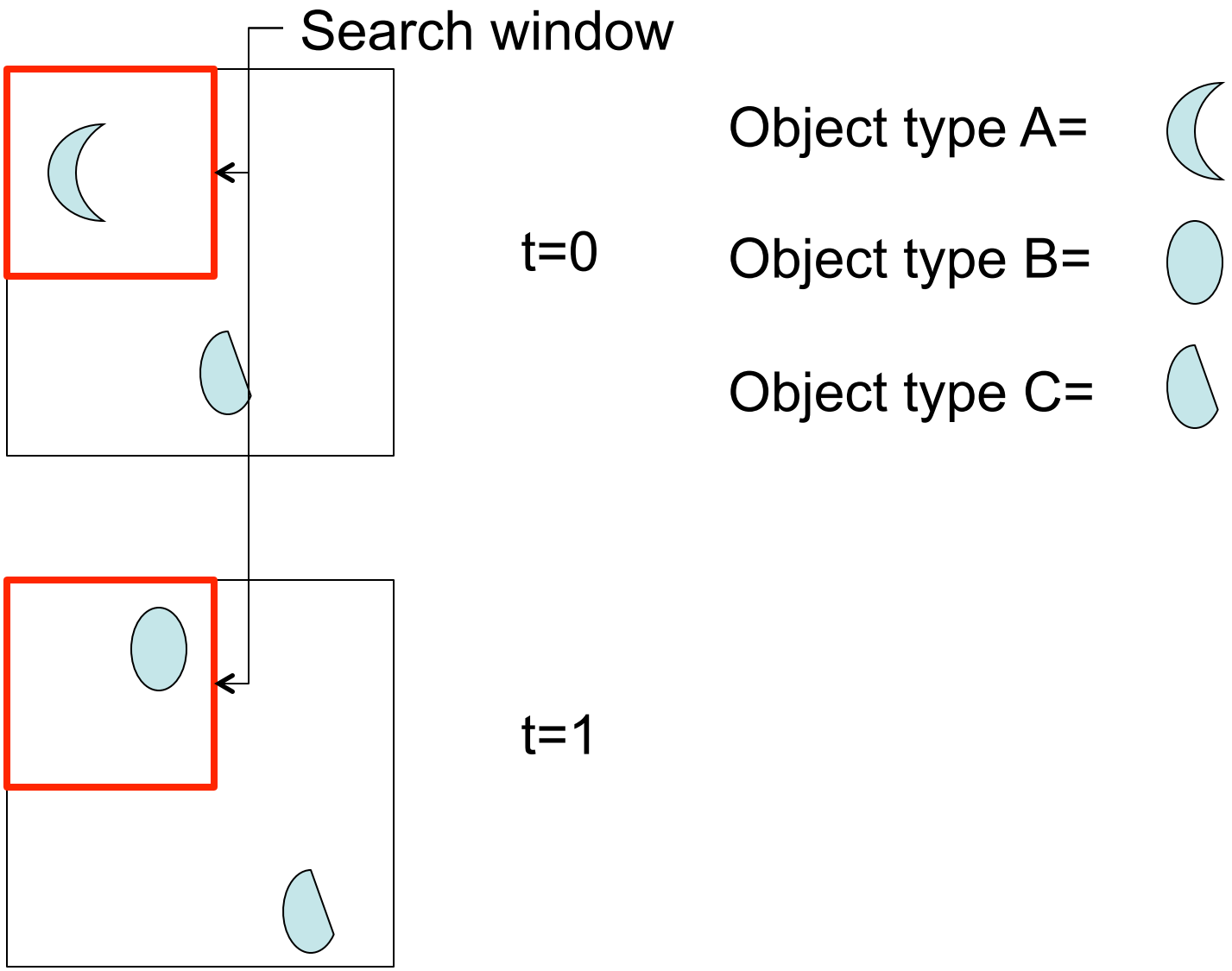
Object type A=

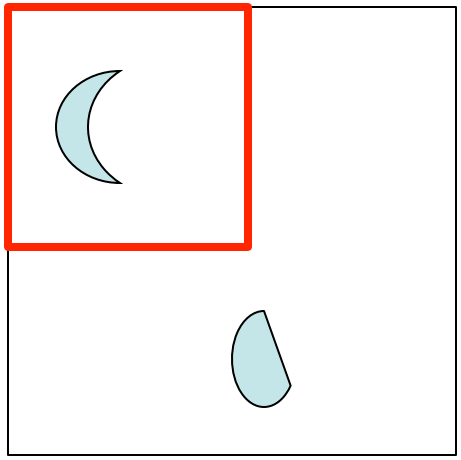
Object type B=

Object type C=



t=1


Search window
t=0
t=1
Object type A=
Object type B=
Object type C=

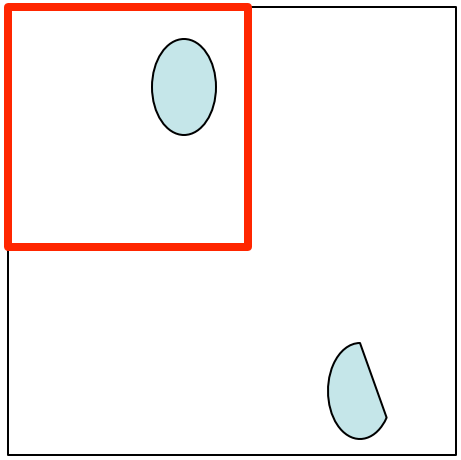


t=0

Object type A=

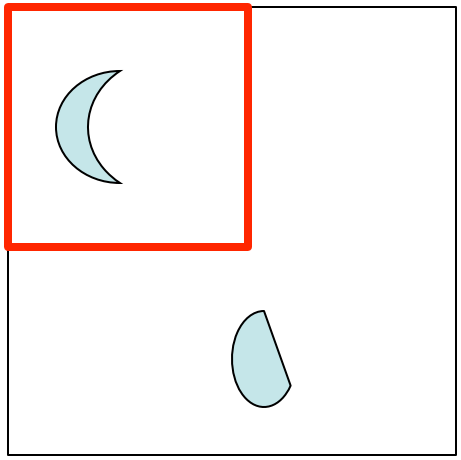
Object type B=

Object type C=



A ⇨ B

t=1

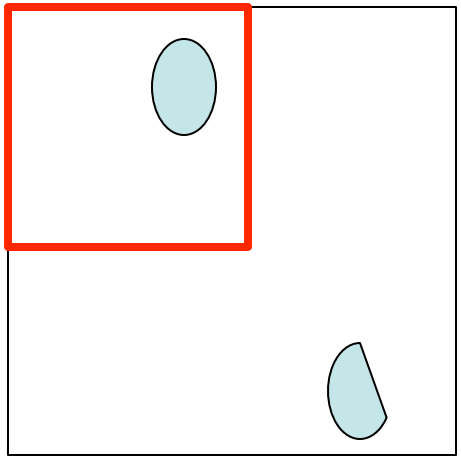


t=0

Object type A=

Object type B=

Object type C=

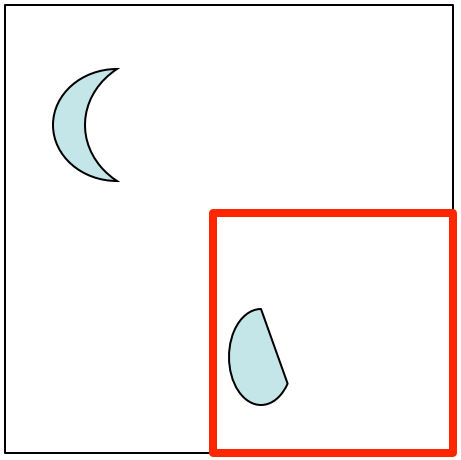


t=1

A ⇨ B

To object type

From object type	A	B	C
A		1	
B			
C			

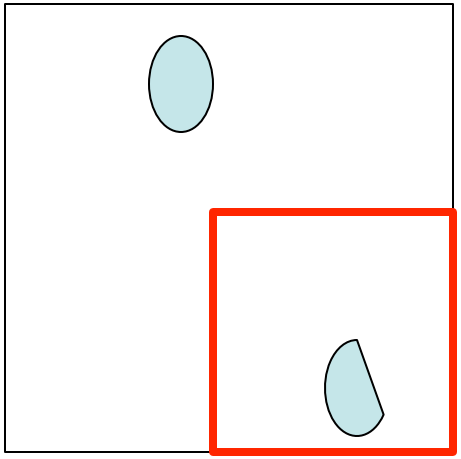


t=0

Object type A=

Object type B=

Object type C=



t=1

A ⇨ B

C ⇨ C

To object type

From object type

	A	B	C
A		1	
B			
C			1

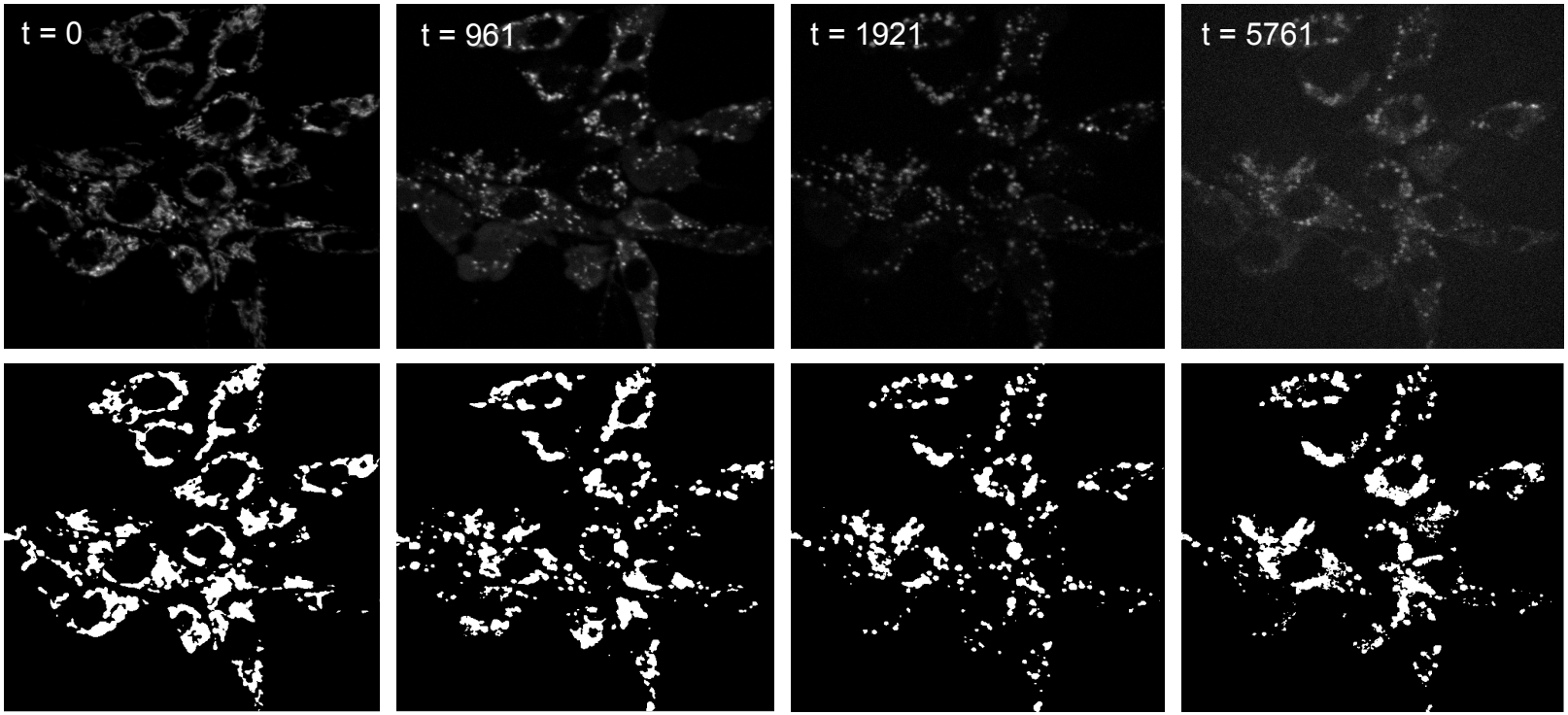
Modeling mitochondrial response to hyperosmotic stress

- 3T3 cells expressing GFP-tagged mitochondrial protein
- Pre-equilibrated with Hoechst 33342 to mark nucleui
- Add 5M NaCl to increase NaCl concentration by ≈74mM

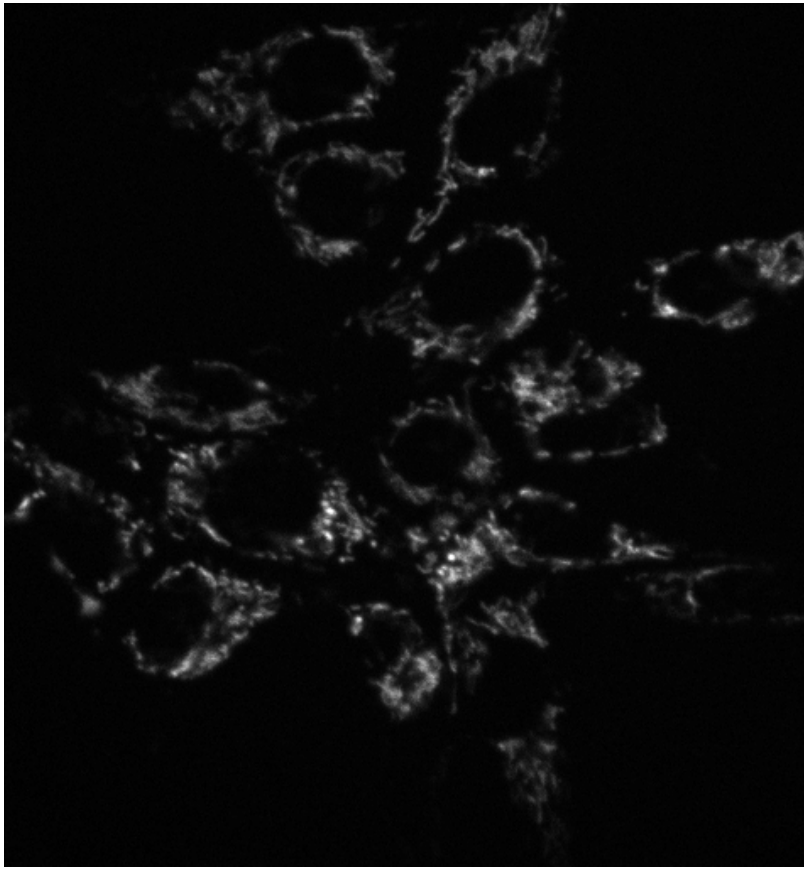
- Model Components
  - $\boldsymbol{m_{\lambda}}$ – *k-by-1* vector representing the proportion of objects of type λ
  - $\boldsymbol{m_{\lambda,\lambda}}$ – *k-by-k* matrix representing the proportion of objects of type λ that have a **nearby** object of type λ' in the subsequent frame
  - $\boldsymbol{m_{\lambda,0}}$ – *k-by-1* vector representing proportion of objects of type λ with no nearby objects in the subsequent frame
  - $\boldsymbol{m_{0,\lambda}}$ – *k-by-1* matrix representing the proportion of objects of type λ that have **appeared** from no nearby object of type λ' in the previous frame

Adaptive image acquisition protocol

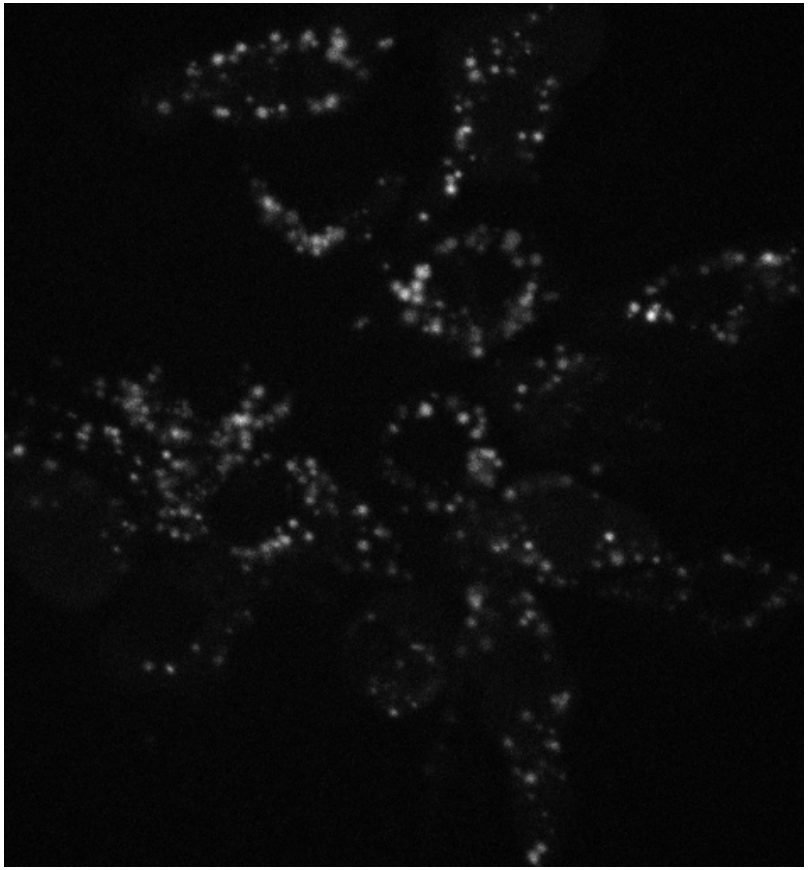
For each time point *t*
Wait until next *t*
Do
Image 5 frames
Add to $model_t$
while $model_t$ error > error threshold



Example images



Before



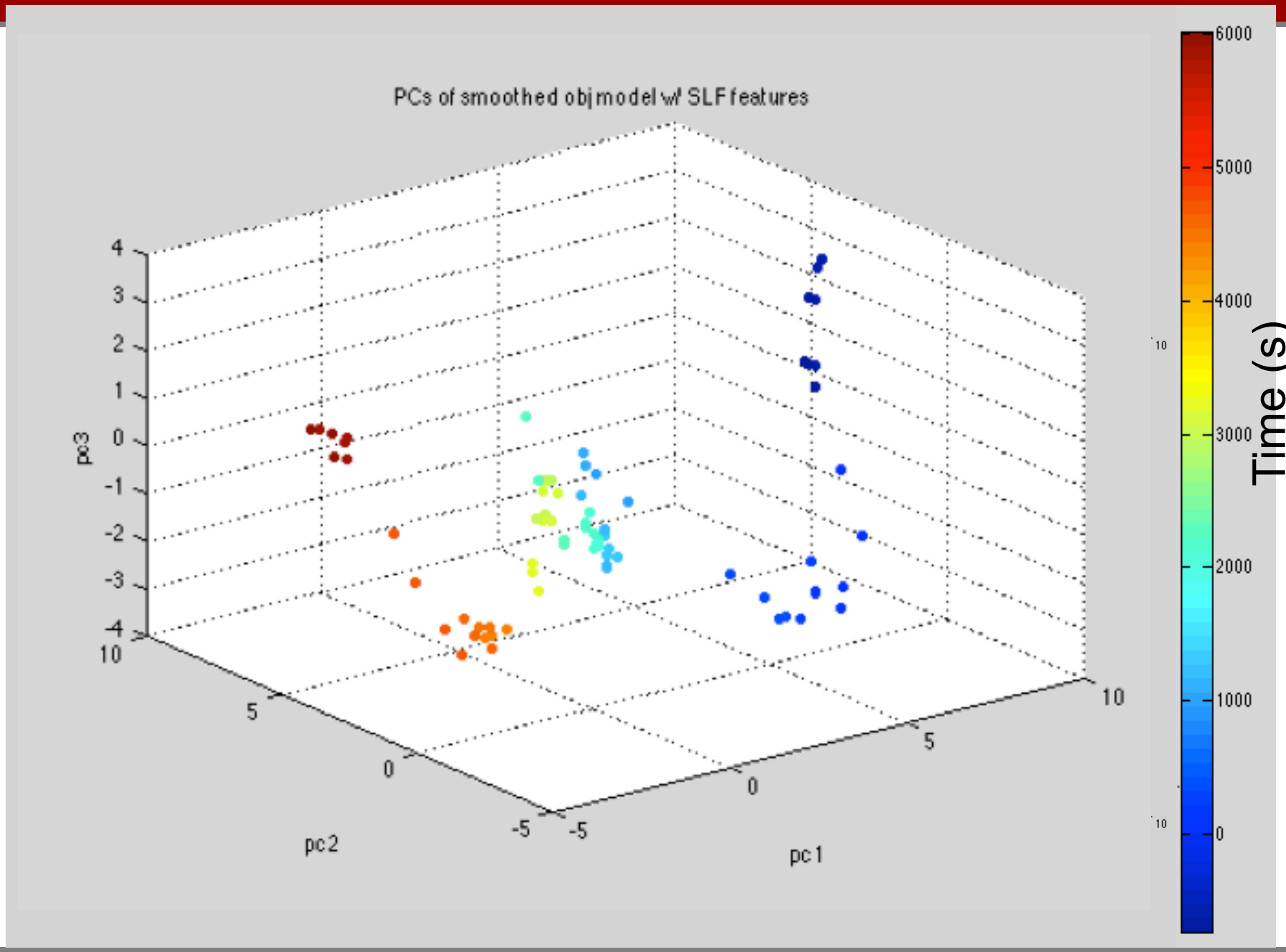
After

Features for each frame pair

- Proportion of object types (7)
- Proportion of object transition types for each frame pair (63)
  - Object to object
  - Object disappear
  - Object appear
- 3D SLF (85)

- Z-scored all features

- Kalman smoothing


PCs of smoothed obj model w/ SLF features
pc3
pc2
pc1
Time (s)
4
3
2
1
0
-1
-2
-3
-4
10
5
0
-5
-5
0
5
10
6000
5000
4000
3000
2000
1000
0

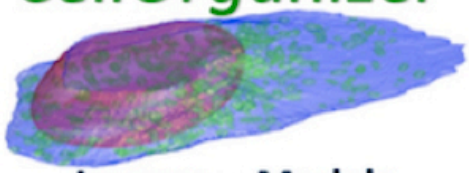
Conclusions: Temporal models

- Various approaches available
  - Extend features to capture spatiotemporal information
  - Learn changes in object composition over time
  - Learn generative model of how objects change, appear, disappear

Update on static generative models

- Have previously described methods for building generative models of nuclei, cell shape, organelle pattern
- Recently extended to 3D
- Collected tools under CellOrganizer framework

CellOrganizer



Images ⇔ Models

RAY AND STEPHANIE LANE Center for Computational Biology

Carnegie Mellon University

Center for Bioimage Informatics
from image to knowledge

Home
People
Publications
Downloads

The *CellOrganizer* project provides tools for

- learning generative models of cell organization directly from images
- storing and retrieving those models in XML files
- synthesizing cell images (or other representations) from one or more models

Model learning captures variation among cells in a collection of images. Images used for model learning and instances synthesized from models can be two- or three-dimensional static images or movies.

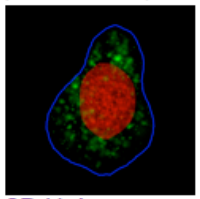
Current components of *CellOrganizer* can learn models of

- cell shape
- nuclear shape
- chromatin texture
- vesicular organelle size, shape and position
- microtubule distribution.

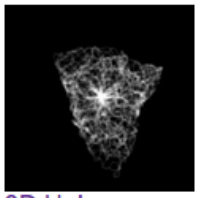
These models can be *conditional* upon each other. For example, for a given synthesized cell instance, organelle position is dependent upon the cell and nuclear shape of that instance.

Cell types for which generative models for at least some organelles have been built include human HeLa cells, mouse NIH 3T3 cells, and Arabidopsis protoplasts. Planned projects include mouse T lymphocytes and rat PC12 cells.

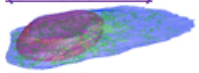
Synthesized Cell Images
(click to view)



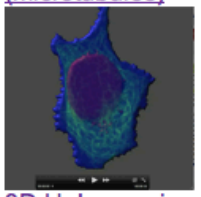
2D HeLa (endosomes)



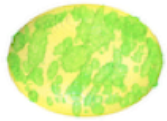
3D HeLa (microtubules)



3D HeLa (mitochondria)



3D HeLa movie



3D protoplast (chloroplasts)

Support for *CellOrganizer* has been provided by grants GM075205 and GM090033 from the National Institute of General Medical Sciences, by a Forschungspreis from the Alexander von Humboldt Foundation, and by the School of Life Sciences of the Freiburg Institute for Advanced Studies.

Alexander von Humboldt Stiftung/Foundation

ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG

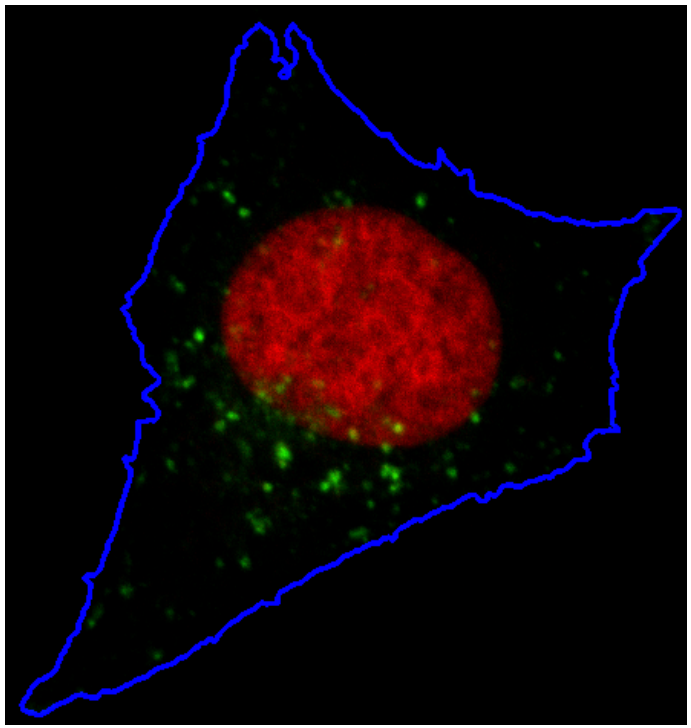
FRIAS FREIBURG INSTITUTE FOR ADVANCED STUDIES

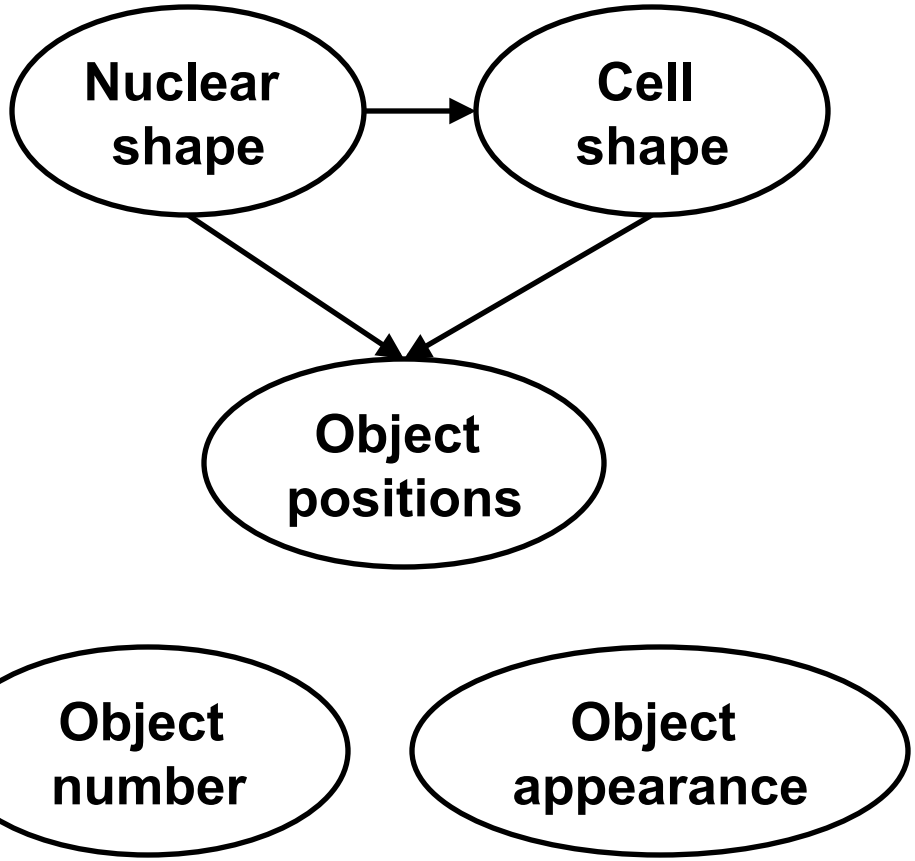
LIFE SCIENCES - LIFENET

Overview

- Choose parametric or non-parametric way of representing a particular component (nucleus, cell shape, lysosome, microtubule) in a single cell (may be conditional upon other components)
- Combine results from many cells to build statistical model of variation -> model
- Randomly sample from model -> instance

Generative model structure

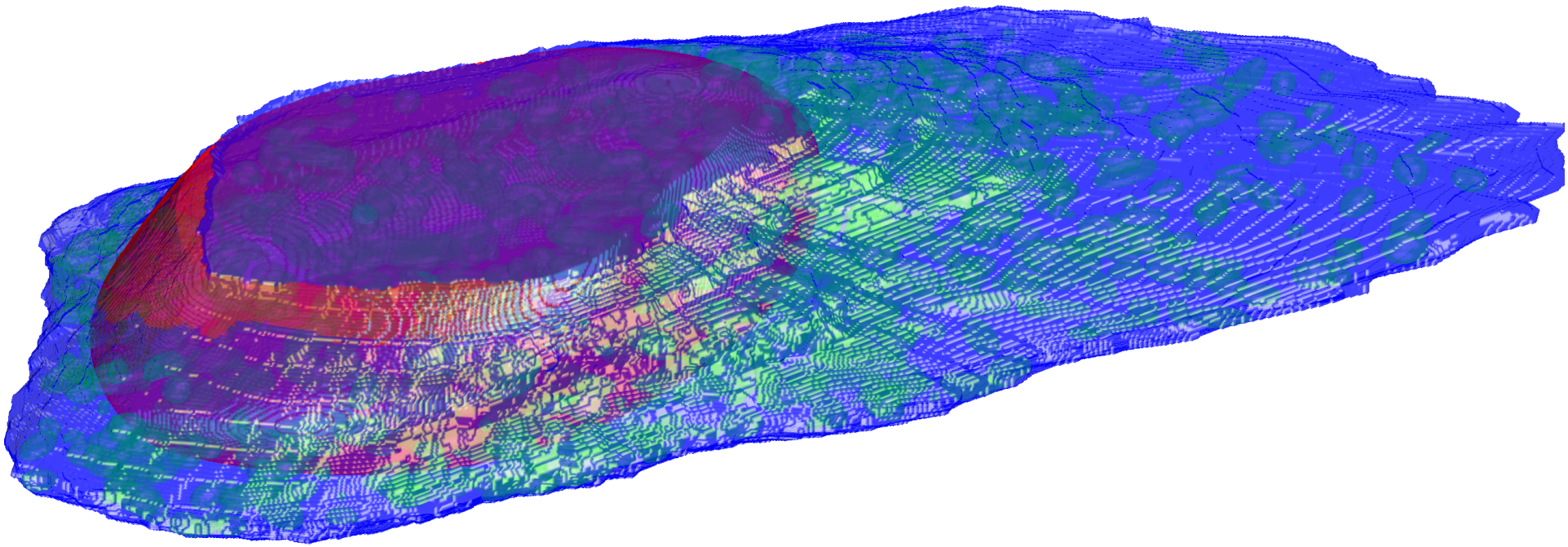




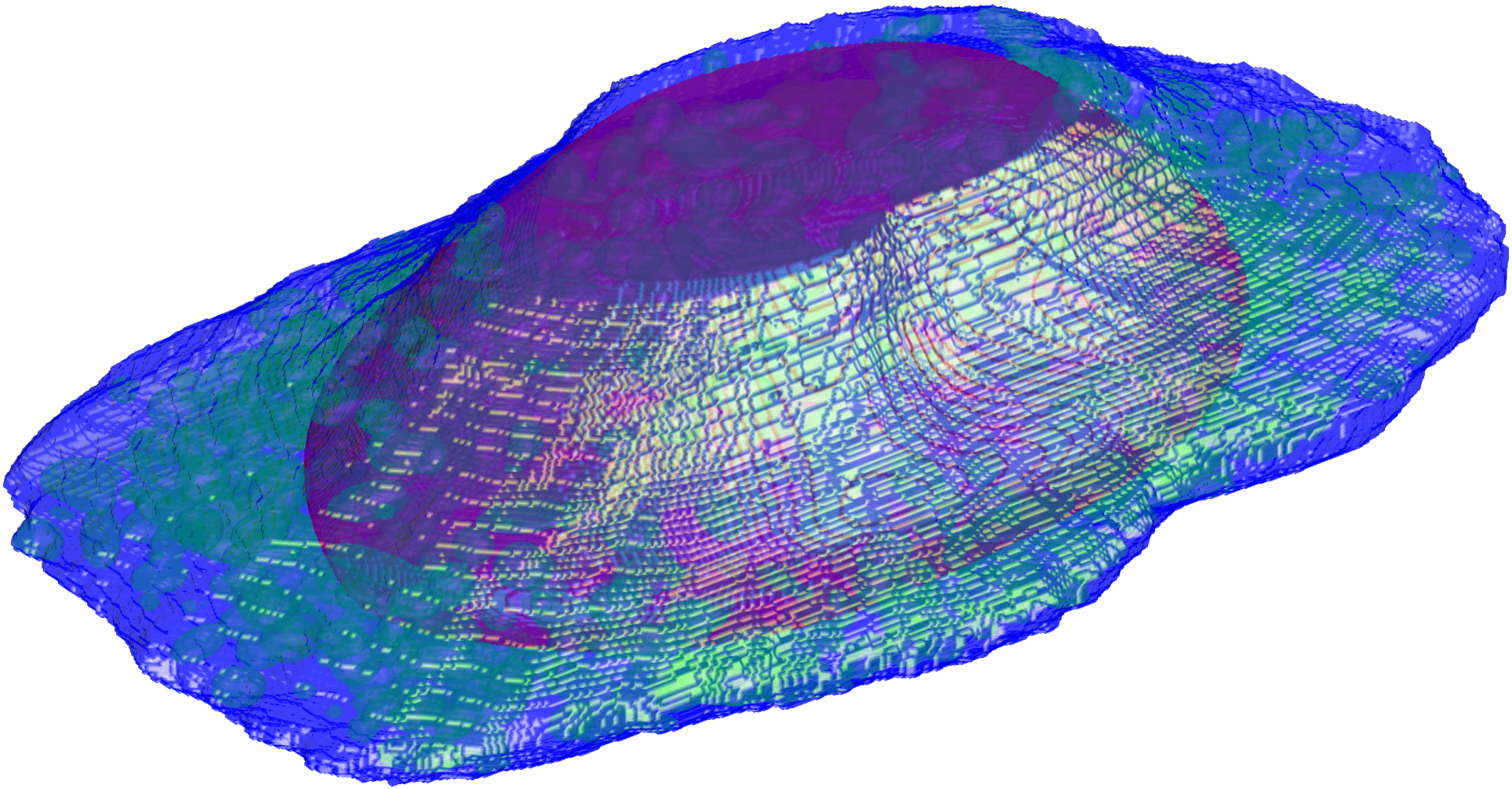
Example 3D instances

3D Endosome Tilt Series

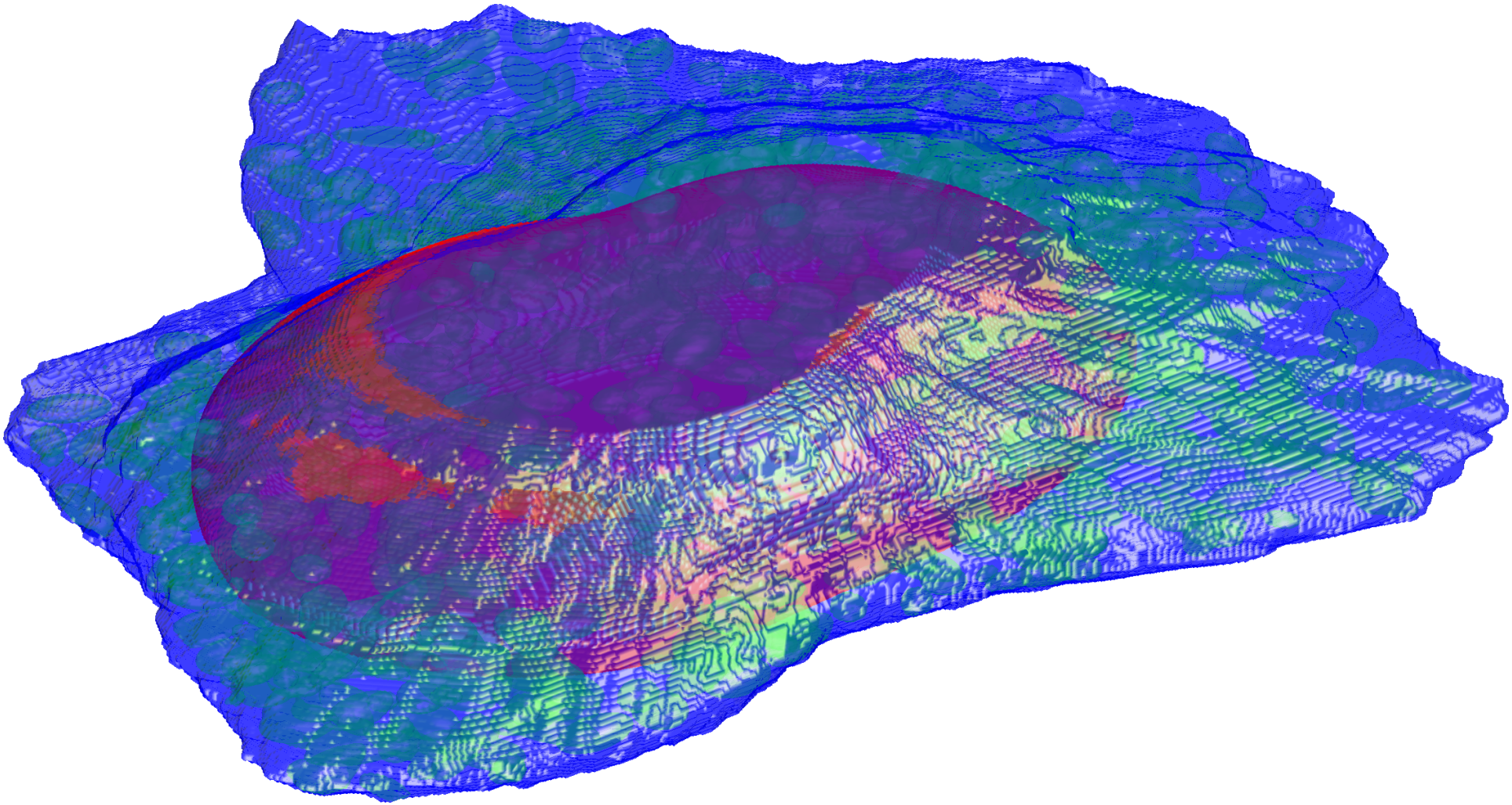
Endosomes



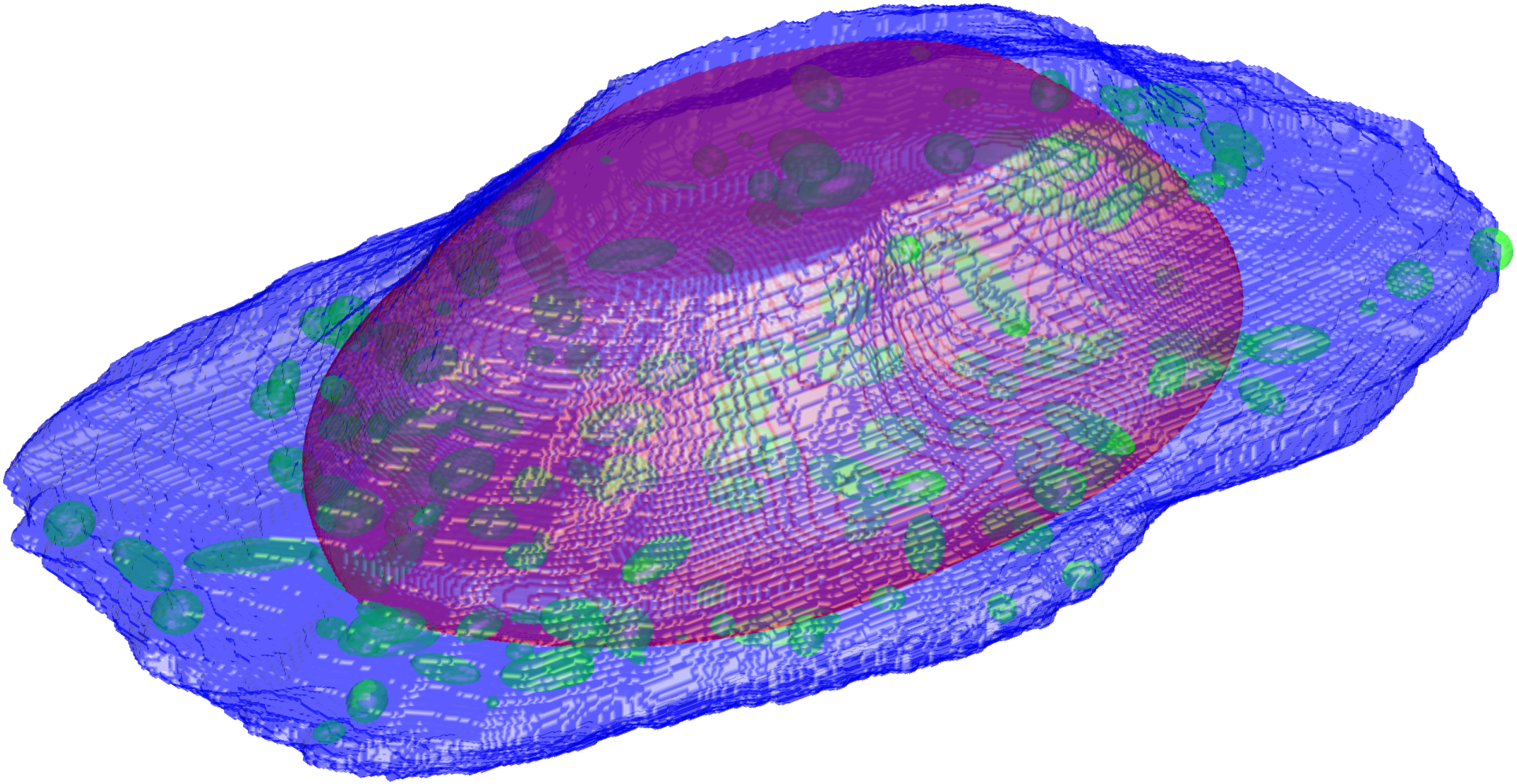
Endosomes



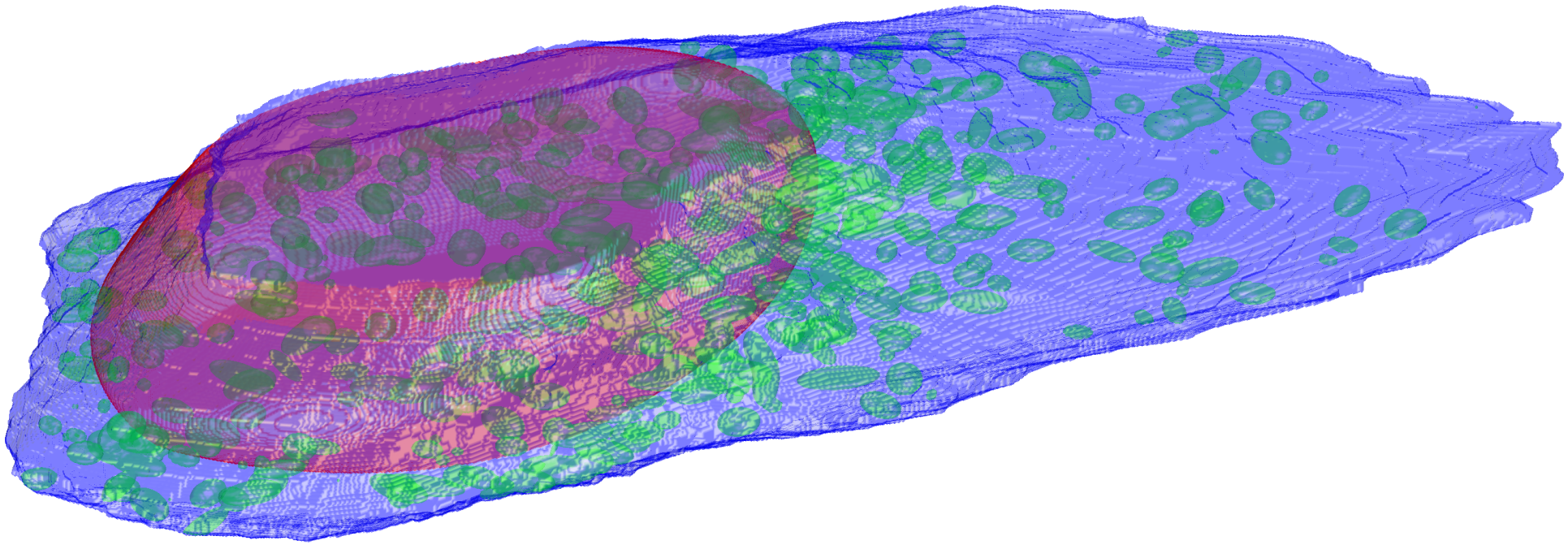
Endosomes



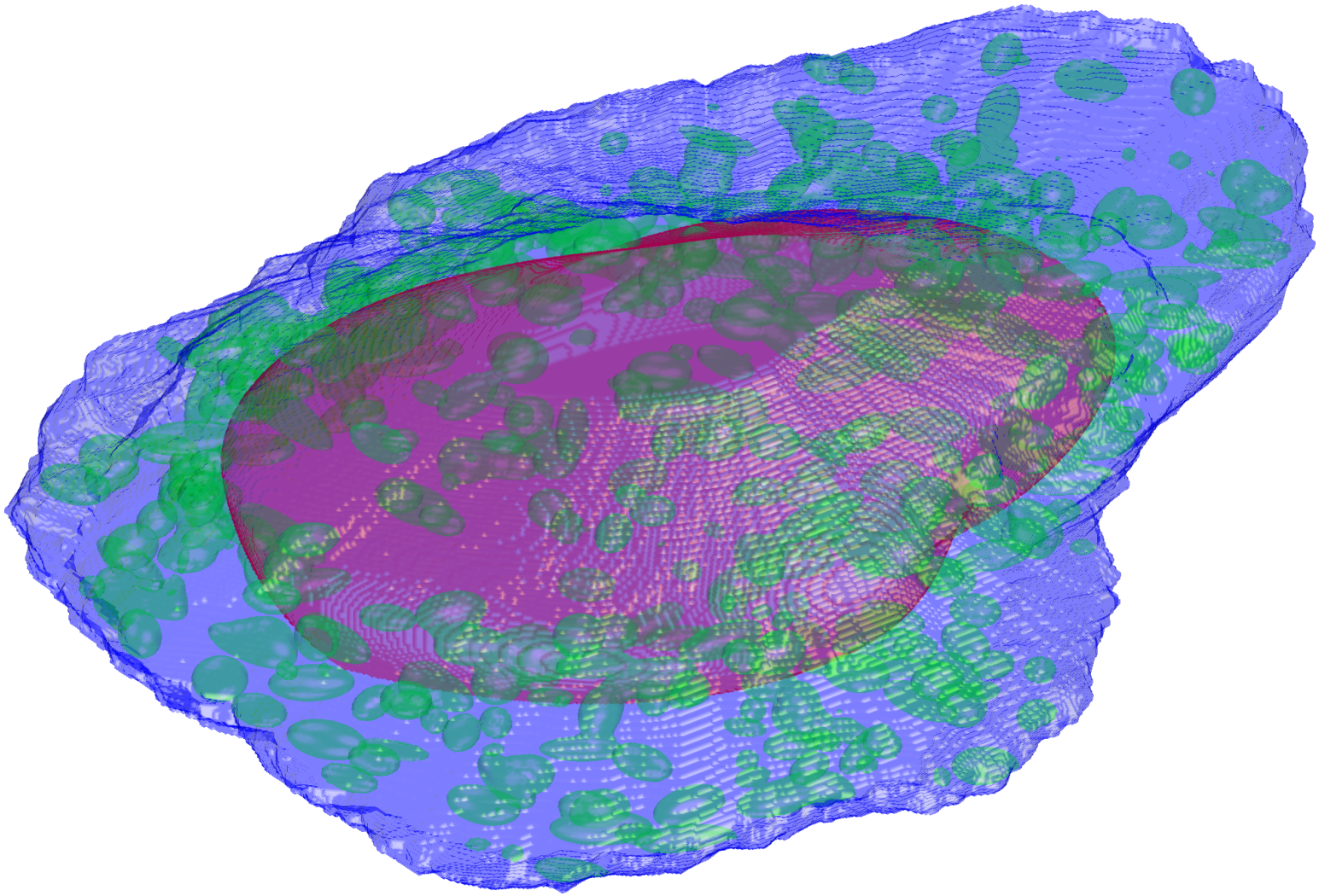
Lysosomes



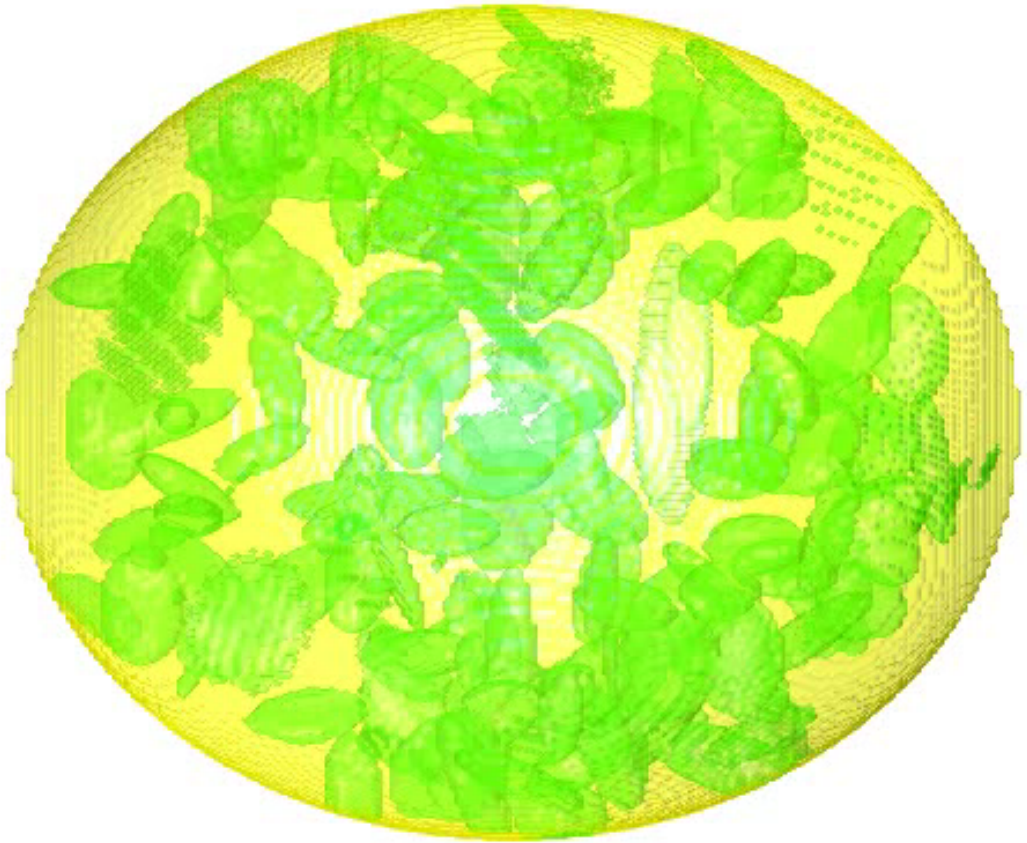
Mitochondria



Mitochondria

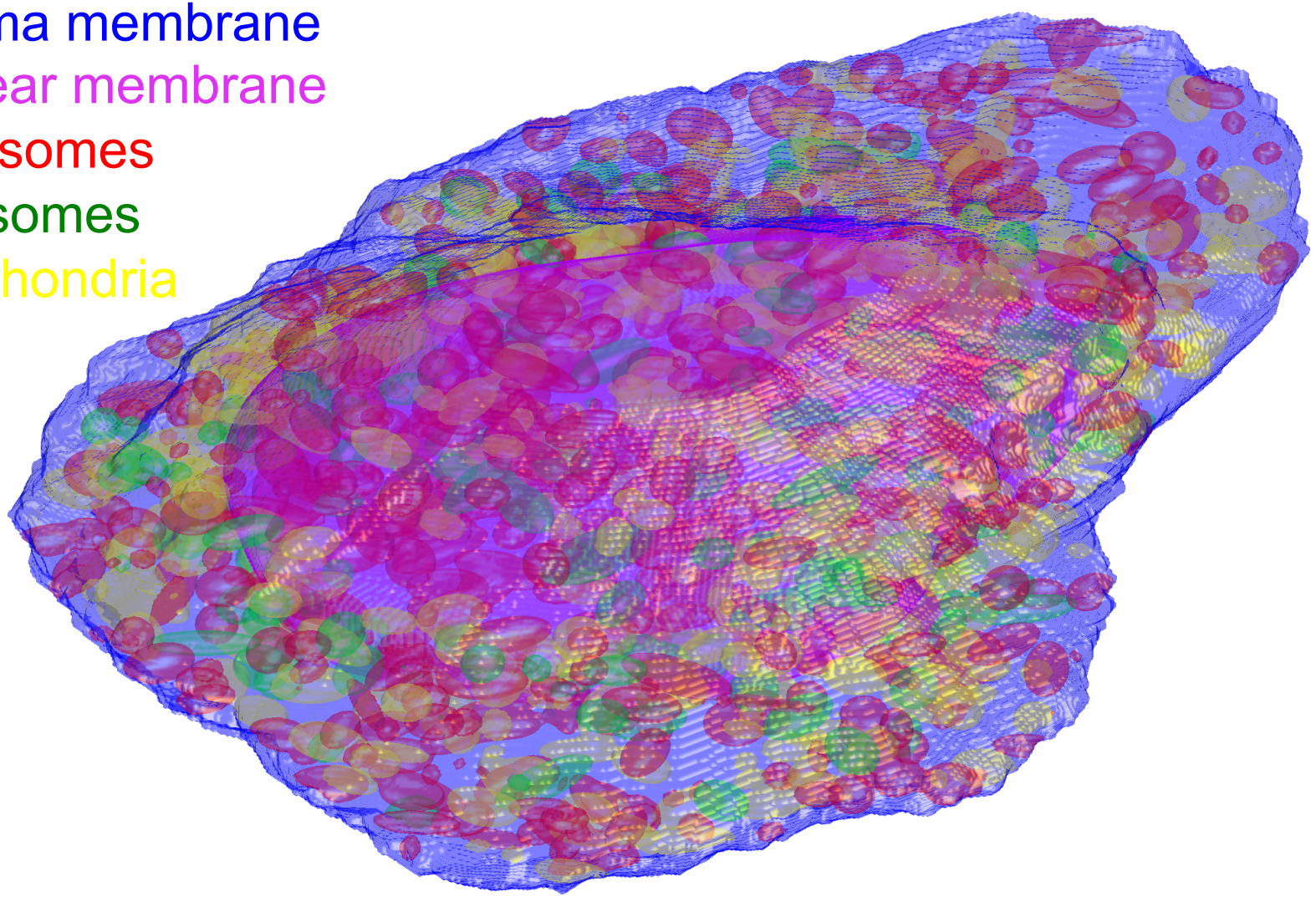


Chloroplasts Rotation Series

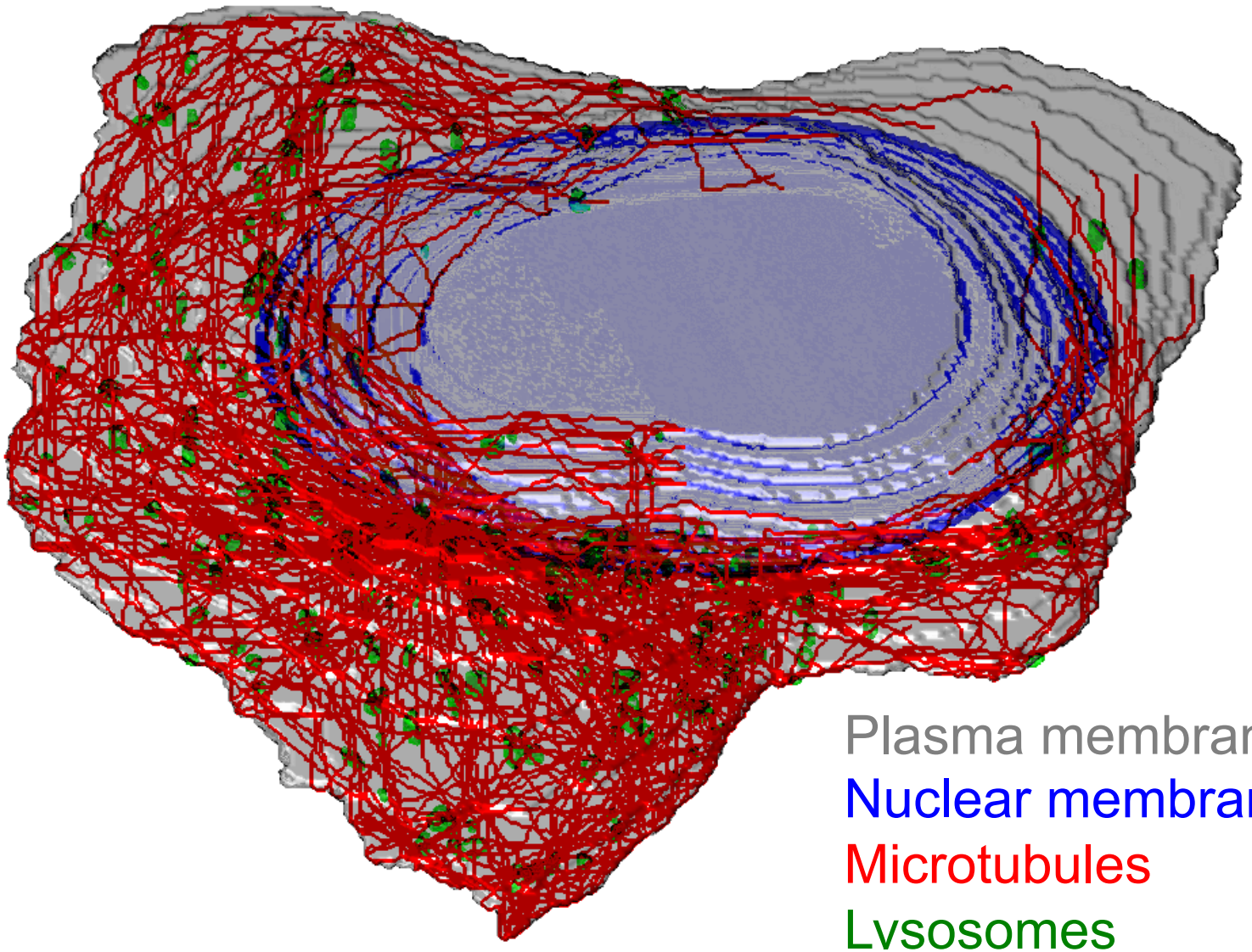


Combining Patterns in One Cell

Plasma membrane
Nuclear membrane
Endosomes
Lysosomes
Mitochondria



Multicomponent conditional models



Conclusions: Representation

- Generative model parameters are generalizable, transportable means for comparing and communicating effects of perturbagens across experiments, laboratories, cell types, and technologies

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 (Bristol-Myers Squibb), Juchang Hua, Ting Zhao (Qiushi Acad), Shann-Ching Chen (St. Jude's), Elvira Garcia Osuna (CMU), Justin Newberg (Baylor), Estelle Glory, Arvind Rao (M.D. Anderson), Henry Lin (Microsoft), Tao Peng (Microsoft), Luis Coelho (U Lisbon), Aabid Sharif, Rumi Naik, Josh Kangas, Jieyue Li, Baek Hwan Cho, Taraz Buck, Charles Jackson, Gregory Johnson, Devin Sullivan**
- Funding
  - **NSF, NIH, Commonwealth of Pennsylvania, Alexander von Humboldt Foundation, Freiburg Institute for Advanced Studies, Wellcome Trust**
- 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, Gustavo Rohde, Ghislain Bonamy, Sumit Chanda, Dan Rines, Hagit Shatkay, Chris Langmead, Klaus Palme, Ziv Bar-Joseph**