OMERO and PSLID

Robert F. Murphy
Lane Center for Computational Biology
Carnegie Mellon University
PSLID mission and overview
Determination of Protein Location from Microscopy Images

• Location Proteomics: Providing information about subcellular protein location
  → (dynamic) proteome map to understand normal and abnormal cell behaviors
  → clues for interaction/ regulation of proteins
  → clues for protein function

• Data driven: Determination of location for all proteins from microscopy images
CD-tagging collection

- **Proteome-scale, live cell imaging of tagged proteins**
- Using **CD-tagging** (developed by Jonathan Jarvik and Peter Berget): Infect population of cells with a retrovirus carrying DNA sequence that will “tag” in a random gene *in its genomic context*

- Isolate separate clones, each of which produces express one tagged protein

- Use RT-PCR to identify tagged gene in each clone

- Collect many live cell images for each clone using fluorescence microscopy
PSLID - Protein Subcellular Location Image Database

Go to the PSLID database containing CD-tagged NIH 3T3 cell clones

Go to the PSLID Public Database Release 4 (released August 1, 2008)

Download datasets contained within PSLID

Download SLIC feature calculation and pattern analysis software used in PSLID and full PSLID software

Supported by the National Institute of General Medical Sciences, Grant GM075205
• 4441 GFP-tagged 3T3 cell clones

• 15-25 images were acquired by automated microscopy (IC100) for each clone for a total of ~140,000 images. Of those, ~50,000 were classified as unusable (out of focus, debris containing, too few cells) so only ~90,000 images appear in the final sets.

• ~300 distinct proteins identified; tagged protein in remaining clones being identified by sequencing

• 667 images from confocal microscopy images for interesting clones

• Release 2 to be release on July 11, 2010 (ISMB meeting)
PSLID principles

• Fundamental unit is a set
• Elements of sets strongly typed
  – 2D, 3D, …
  – image, cell, object…
• Analysis results stored in DB, linked to element
• Meta analysis
• Search by image content
## Summary of RandTag Database

**Dataset type** = 2D Image

**Gene Assignment Confidence:** High: Matching sequence read from both primers

<table>
<thead>
<tr>
<th>Targets</th>
<th>Clones</th>
<th>Locations from Other Databases</th>
<th>Locations from Image Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit</td>
<td>3T3-CDtag-BS2C3</td>
<td>unknown, mitochondrion; membrane, proton-transporting two-sector ATPase complex, proton-transporting two-sector ATPase complex; catalytic domain; mitochondrion, cytoplasm, peroxisome, plasma membrane, endoplasmic reticulum, nuclear envelope, nucleus, nucleolus, mitochondrial inner membrane</td>
<td>cytoplasmic</td>
</tr>
<tr>
<td>caldesmon 1</td>
<td>3T3-CDtag-CW1G5</td>
<td>cytoskeleton, cytoplasm; membrane fraction, actin cap; nucleus, cytoplasm, peroxisome, extracellular region, endoplasmic reticulum</td>
<td>cytoskeleton</td>
</tr>
</tbody>
</table>

*OME workshop, June 15-16, 2010, Paris*
## PSLID Randtag collection

The image shows a webpage from the PSLID service, which appears to be a database for selecting genes based on specific cellular locations. The table below represents the selection options:

<table>
<thead>
<tr>
<th>Location</th>
<th>Cytoplasm</th>
<th>Cytoskeleton</th>
<th>ER</th>
<th>Golgi</th>
<th>Lysosome</th>
<th>Mitochondria</th>
<th>Nuclear</th>
<th>Nucleoli</th>
<th>Plasma Membrane</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoplasm</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytoskeleton</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>ER</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Golgi</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Lysosome</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleoli</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma Membrane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

This table is likely used in the context of the OME workshop, which took place on June 15-16, 2010, in Paris.
Many ways to access and display the data

- Look at overview of the database content
  → **Database summaries (by target, by subcellular location)**

- Automate the query, to include the link in a website for example
  → **use the URL query**
  By Target:
  
  http://pslid3.cbi.cmu.edu/RandTag_develop/search.jsp?target=40S 20ribosomal 20protein 20S28

  By experiment:
  
  http://pslid3.cbi.cmu.edu/RandTag_develop/search.jsp?cell_name=3T3-CDtag-BM1B10

  Returning the answer as HTML or XML format

- browsing the collection by using the web pages
  → View **plates**
  → View Summary of the **well**
  → View all the **images** of the well
  → View **Set**
Search results for Cell_type: 3T3-CDtag-BM1B10, Image Type: 2D Stain, Target: ALL

90 images returned (50 images shown) from the query.

View the summary of set temp3100_128687f458090c87b76f494c98b049a8.

<table>
<thead>
<tr>
<th>Image</th>
<th>Cell Name</th>
<th>Organism</th>
<th>Experiment</th>
<th>Protocol</th>
<th>Target</th>
<th>Microscopy  &amp; Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>3T3-CDtag-BM1B10</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>unidentified</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>Image</td>
<td>3T3-CDtag-BM1B10</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>dna</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>Image</td>
<td>3T3-CDtag-BM1B10</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>none</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>Image</td>
<td>3T3-CDtag-BM1B10</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>unidentified</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>Image</td>
<td>3T3-CDtag-BM1B10</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>dna</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>Image</td>
<td>3T3-CDtag-BM1B10</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>none</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
</tbody>
</table>
URL search: by target

Search results for Image Type: 2D Static, Target: 40S ribosomal protein S28

22 images returned (30 images shown) from the query.

View the summary of set temp3105_128687F45B090C87B76F404C98E049A8

<table>
<thead>
<tr>
<th>Image</th>
<th>Cell Name</th>
<th>Organism</th>
<th>Experiment</th>
<th>Protocol</th>
<th>Target</th>
<th>Microscopy &amp; Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>214</td>
<td>3T3-CDtag-BM1B9</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td><strong>40S ribosomal protein S28</strong></td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>232</td>
<td>3T3-CDtag-BM1B9</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>40S ribosomal protein S28</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>235</td>
<td>3T3-CDtag-BM1B9</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>40S ribosomal protein S28</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>238</td>
<td>3T3-CDtag-BM1B9</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>40S ribosomal protein S28</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>244</td>
<td>3T3-CDtag-BM1B9</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>40S ribosomal protein S28</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
<tr>
<td>247</td>
<td>3T3-CDtag-BM1B9</td>
<td>Mus musculus</td>
<td>3T3-CDtag-BM1</td>
<td>3T3 RandTag Protocol 2</td>
<td>40S ribosomal protein S28</td>
<td>CellLab IC100 Image Cytometer</td>
</tr>
</tbody>
</table>
URL search: by target, returned in XML format

- <pslid_target_list>
  - <results_summary>
    <num_records>199</num_records>
  </results_summary>
  - <target_result>
    - <target>
      1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase delta-1 (Phosphomonoester phospholipase C)(Phospholipase C-delta-1)(PLC-delta-1)(PLC-III)
    </target>
    <number_of_stacks>3</number_of_stacks>
  </target_result>
  - <target_result>
    - <target>
      28 kDa heat- and acid-stable phosphoprotein (PDGF-associated protein)(PAP)(PDGF-A-associated protein 1)(PAP1)
    </target>
    <number_of_stacks>4</number_of_stacks>
  </target_result>
  - <target_result>
    - <target>
      40S ribosomal protein S11
    </target>
    <number_of_stacks>4</number_of_stacks>
  </target_result>
  - <target_result>
    - <target>
      40S ribosomal protein S24
    </target>
    <number_of_stacks>4</number_of_stacks>
  </target_result>
  - <target_result>
    - <target>
      40S ribosomal protein S28
    </target>
    <number_of_stacks>22</number_of_stacks>
  </target_result>
</pslid_target_list>
**Experiment:** 3T3-CDtag-CZ1  
**Microscope:** IC100

<table>
<thead>
<tr>
<th>Plate</th>
<th>Well</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>A1</td>
<td>Unidentified protein BL105: 25</td>
</tr>
<tr>
<td>A1</td>
<td>A2</td>
<td>Unidentified protein BL105: 25</td>
</tr>
<tr>
<td>A1</td>
<td>A3</td>
<td>Unidentified protein BL105: 25</td>
</tr>
<tr>
<td>A1</td>
<td>A4</td>
<td>Unidentified protein BL105: 25</td>
</tr>
<tr>
<td>A1</td>
<td>A5</td>
<td>Unidentified protein BL105: 25</td>
</tr>
<tr>
<td>A1</td>
<td>A6</td>
<td>Unidentified protein BL105: 25</td>
</tr>
<tr>
<td>B1</td>
<td>B1</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>B1</td>
<td>B2</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>B1</td>
<td>B3</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>B1</td>
<td>B4</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>B1</td>
<td>B5</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>B1</td>
<td>B6</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>C1</td>
<td>C1</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>C1</td>
<td>C2</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>C1</td>
<td>C3</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>C1</td>
<td>C4</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>C1</td>
<td>C5</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>C1</td>
<td>C6</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>D1</td>
<td>D1</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>D1</td>
<td>D2</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>D1</td>
<td>D3</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>D1</td>
<td>D4</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>D1</td>
<td>D5</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>D1</td>
<td>D6</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>E1</td>
<td>E1</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>E1</td>
<td>E2</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>E1</td>
<td>E3</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>E1</td>
<td>E4</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>E1</td>
<td>E5</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
<tr>
<td>E1</td>
<td>E6</td>
<td>Unidentified nuclear protein BL105: 25</td>
</tr>
</tbody>
</table>
Images of one well

Experiment: 3T3-CData-CZ1
Microscope: IC100
Set Name: CZ1D5 IC100

View Set Summary

Multi criterion search

Ticket: R7

Logged in as: demo

Enter a search condition

Cell type (=clone name): ALL
Protocol type: ALL
Microscope type: ALL
Experiment title (=Sort): ALL
Sequence Date: ALL
Gene name: Rp17
Gene Assignment Confidence: Low. All sequences

Analysis tools

• Segmentation/feature calculation
• Typical image selection
• Set comparison
• Feature selection
• Classifier training/use
• Stacked/meta classifier training/use
• Clustering to sets
• Generative models training/synthesis
Exporting from PSLID

• For any set, can export
  – Feature file for selected feature set
    • Matlab .mat file
    • CSV
    • Tab delimited
  – Zip file containing all images
    • Original format
    • OME TIFF
      – Images
      – Metadata (OME XML + PSLID custom annotations)
      – Analysis results
    • HDF5 (coming)
Fluorescence Microscopy Annotation Schema

- **Version 1**: February 15, 2002 - database specifications
- **Version 1.1**: September 11, 2002 - database specifications
- **Version 1.2**: May 28, 2003 - database specifications & ER Diagram
- **Version 2.1**: August 25, 2005 - integrated table viewer and ER diagram
- **Version 2.3**: July 17, 2006 - integrated table viewer and ER diagram
- **Version 3.0**: December 15, 2006 - integrated table viewer and ER diagram
- **Version 5.0**: August 1, 2009 - database specifications & ER Diagram
Importing into OMERO

- OME TIFF with PSLID custom annotations
  - Works with current OMERO.importer
  - Can’t load/read custom annotations
  - Should be fixed in OMERO 4.2
- Currently require additional import into PSLID server to capture the rest of the annotations
Adding PSLID tags to OME-XML

• Description of cell lines, reagents and probes in sample
• Features, classifications, and other results from analysis
Sample Description
Target reasoning

• Annotations specify
  – excitation wavelength and emission filters for each image
  – labels (substances, dyes) present in each sample

• Code reasons out which target substance is visualized in each image
The current target for this image is: *Lysosomal membrane glycoprotein 2*

Based on the sample labeling and image collection descriptions, the target of this image can be inferred as follows:

- The dye(s) used in this sample is (are): *Cy5, DAPI*
- The emission filter used for this image is: *Cy5 em filter*
- The wavelength range of this filter is: *665 nm - 740 nm*
- The dye in this sample that matches the filter is: *Cy5 (peak emission 670 nm)*
- The sample is labeled with *Cy5*, which is attached to *goat IgG*.
- The sample is labeled with *goat IgG anti mouse IgG*, which is *goat IgG* whose target is *mouse IgG*.
- The sample is labeled with *mouse IgG anti LAMP2*, which is *mouse IgG* whose target is *Lysosomal membrane glycoprotein 2*.

Therefore, by reasoning, the final target is: *Lysosomal membrane glycoprotein 2*

The reasoned target agrees with the current target.

**Do you want to do such target reasoning for all the images in this set?**
Target reasoning

• PERMITS IDENTIFICATION OF GENERALIZATIONS AND SPECIALIZATIONS

  – CAN SEARCH FOR “TUBULIN” AND FIND ALL ISOFORMS
  – CAN SEARCH FOR “TUBULIN-ALPHA” AND BE ASKED “NO MATCHES, DO YOU WANT TO SEE IMAGES OF TUBULIN-BETA?”

• PERMITS AUTOMATICALLY FIND DIFFERENCES IN REAGENTS AND ASSOCIATE THEM WITH DIFFERENCES IN RESULTS

PSLID analysis results metadata

Diagram:

- Stack
- Set
- Image set
- Image
  - 2D Single Image
    - 2D Single Image feature value
  - 2D Double Image
    - 2D Double Image feature value
- Region Generator
- Mask
- Feature Set
- Feature
  - 2D Double Region Feature Value
  - 3D Double Region Feature Value
- Region
  - 2D Single Region Feature Value
  - 3D Single Region Feature Value
- Region set
  - 1..*
- 3D Region
  - 3D Single Region Feature Value
- 3D Region Set
  - 1..*
- Classifier
  - Subcellular Location
  - Classification

Text:

PSLID client and OMERO server

- PSLID client can connect to OMERO server to select images and create a PSLID set from them
- Currently: duplicated in PSLID server
- Planned: replace “plumbing” in PSLID server with calls to OMERO server
  - Feasibility tests completed
OMERO.client

- Data management
- Browse images
- Measurement
- Interact with PSLID

OMERO.server

Upload/retrieve

PSLID client/server

Invoke PSLID

• 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?
• Assume that we have markers that are found in only one subcellular location (fundamental pattern)

• Assume that each fundamental pattern can be represented by some combination of distinct object types (e.g., 10% small round objects and 90% long skinny objects)

• Assume that a mixed pattern is formed by adding together the objects from fundamental patterns

• Then for each unknown image we can calculate the fraction of fluorescence in each object type, and then estimate how much of each fundamental pattern must be present
Learn object types from pure samples

Learn model for distributions of object types in pure samples

Determine pattern fractions for mixed images
Examples of Object Types

Type A

Type B

Type C

Type D
Supervised unmixing results

Mitotracker

Lysotracker

Estimated conc. vs. Actual conc.
PUnmix and OMERO

• New PUnmix release July 11, 2010

• Versions
  – Matlab source
  – Compiled matlab for MacOS, Windows, Linux

• Supports training on or unmixing images from local disk or multiple OMERO servers
PUnmix GUI
PUnmix GUI
PUnmix after import from OMERO
PUnmix adding local images

Select protein images

Select DNA images (optional)

Unmixing result

Linear unmix
Multinomial unmix
Fluorofraction unmix

PUnmix training mixing model
PUnmix training mixing model
PUnmix training mixing model
SLML Suite

- SLML suite (Subcellular Location Markup Language) which can learn generative models of subcellular patterns from images and can synthesize new images from them
Generative Cell Models

- Cell membrane
- Nucleus
- Gaussian objects
- Protein distribution
- Probability Density Function
- Medial axis
- Model parameters

\[ \frac{d_1 + d_2}{d_2} \]

Zhao & Murphy 2007
Synthesized Images

- Have XML design for capturing model parameters
- Have portable tool for generating images from model

SLML toolbox - Ivan Cao-Berg, Tao Peng, Ting Zhao

- Have portable tool for generating images from model
SLML Suite and OMERO

- Adding same OMERO read interface used by PUnMix
Meta Classification: Graphical Models

Original image

segmentation

feature extraction

| Cell# | feat1 | feat2 | feat3 | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.78</td>
<td>15.7</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.83</td>
<td>23.6</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.71</td>
<td>20.9</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Trained classifier

Result of the classification

ER
Lysosome
Tubulin

Likelihood of each class

<table>
<thead>
<tr>
<th>Cell#</th>
<th>ER</th>
<th>Lyso.</th>
<th>Tub</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.58</td>
<td>0.32</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.40</td>
<td>0.45</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>0.32</td>
<td>0.30</td>
<td>0.38</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Improve labels by loopy belief propagation

### Confusion Matrix after prior updating

<table>
<thead>
<tr>
<th></th>
<th>Nuclear</th>
<th>ER</th>
<th>Golgi</th>
<th>Lyso.</th>
<th>Mito.</th>
<th>Nucleol</th>
<th>Actin</th>
<th>Endo.</th>
<th>Tubulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ER</td>
<td>0</td>
<td>98.8</td>
<td>0</td>
<td>0.8</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Golgi</td>
<td>0</td>
<td>0.2</td>
<td>93.9</td>
<td>0.7</td>
<td>5.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lysosomal</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>99.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mito.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.6</td>
<td>98.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nucleolar</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Actin</td>
<td>0</td>
<td>3.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Endosom.</td>
<td>0</td>
<td>5.6</td>
<td>12.2</td>
<td>3.8</td>
<td>0</td>
<td>5.6</td>
<td>72.5</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>Tubulin</td>
<td>0</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>99.7</td>
</tr>
</tbody>
</table>

- Overall accuracy to recognize the 9 classes = 95.6%
Acknowledgements

• Dr. Michael Boland
• Dr. Meel Velliste
• Dr. Kai Huang
• Dr. Xiang Chen
• Dr. Ting Zhao
• Dr. Shann-Ching Chen
• Dr. Juchang Hua
• Dr. Justin Newberg
• Dr. Baek Hwan Cho
• Tao Peng
• Jenn Backal

Funding
• NSF, NIH, Commonwealth of Pennsylvania

Collaborators/Consultants
• David Casasent, Simon Watkins, Jon Jarvik, Peter Berget, Jack Rohrer, Tom Mitchell, Christos Faloutsos, Jelena Kovacevic, William Cohen, Geoff Gordon