The Future of Subcellular Pattern Analysis

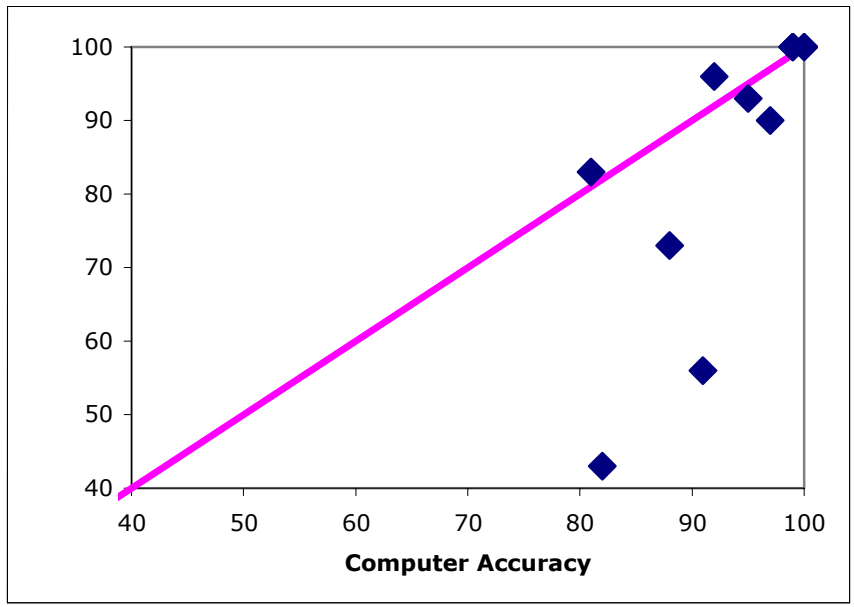
Robert F. Murphy

Departments of Biological Sciences and Biomedical Engineering, Machine Learning, and

Carnegie Mellon

Progress over past 10+ years

- Have many systems to compare and classify subcellular patterns
- Performance better than humans



Why do we do HTM/HCS?

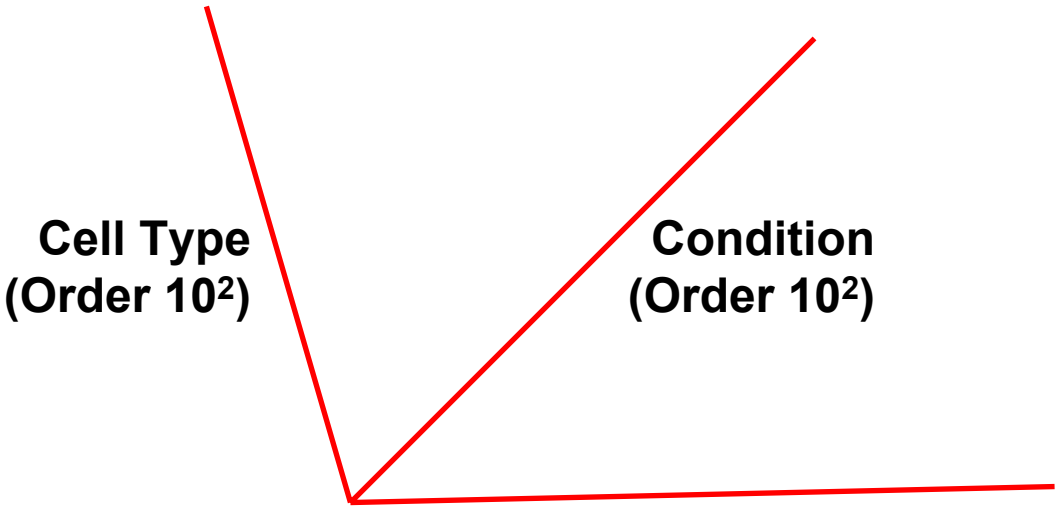
- (Minor) To learn drug effects using *artificial* constructs (e.g., target-probe chimeras)
- (Major) To learn about changes in *natural* subcellular distributions of probes

Transition to unsupervised learning

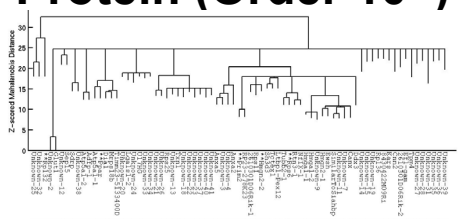
- Can only go so far with supervised learning - reach the limits of what we know
- Alternative is to use unsupervised learning - clustering

Z-scored Mahalanobis Distance
0
5
10
15
20
25
30
http://murphylab.web.cmu.edu/services/PSLID/tree.html
Protein name
Human description
From databases

The problem



Protein (Order 10^6)



Plus: Time scale from subsecond to years

Center for Bioimage Informatics
from image to knowledge
Carnegie Mellon

Orthogonal data sources

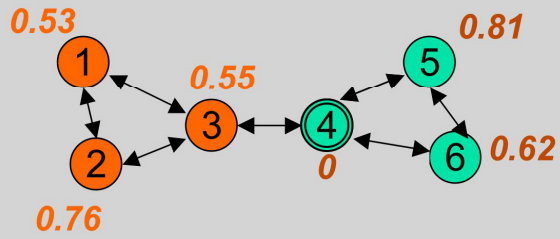
- Cytochemical images like Protein Atlas (fixed cells, one color)
- Sequential multicolor immunofluorescence like MELK (fixed cells, many colors)
- GFP-tagged proteins (live cells, one to few colors)

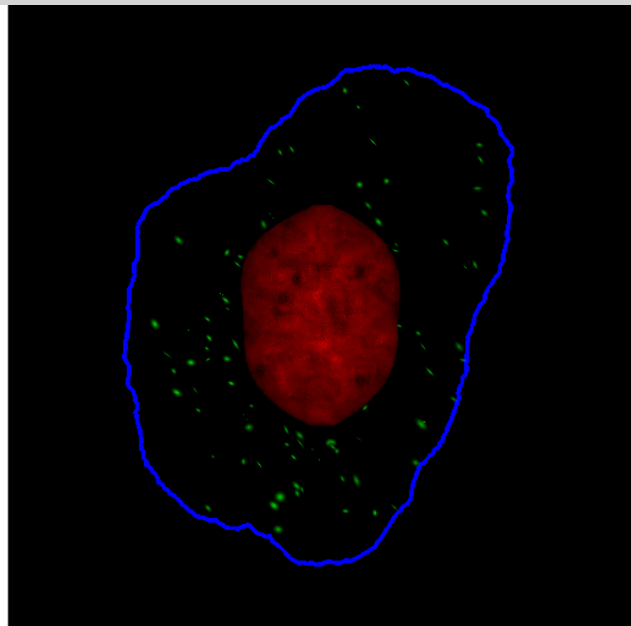
How do we really analyze subcellular location?

- Classification and comparison good for focused questions but there are too many questions to ask
- Need intelligent (optimized) data collection: probabilistic methods to integrate available data, make predictions and suggest experiments

Part of the solution?

- Graphical models to allow inference from multiple cells in tissue
- Generative models to synthesize information from image sets





Center for Bioimage Informatics
from image to knowledge
Carnegie Mellon

Human Cytome Project?

- Scope of problem argues for cooperation on grand scale
- New inference and synthesis methods