

The Future of Subcellular Pattern Analysis



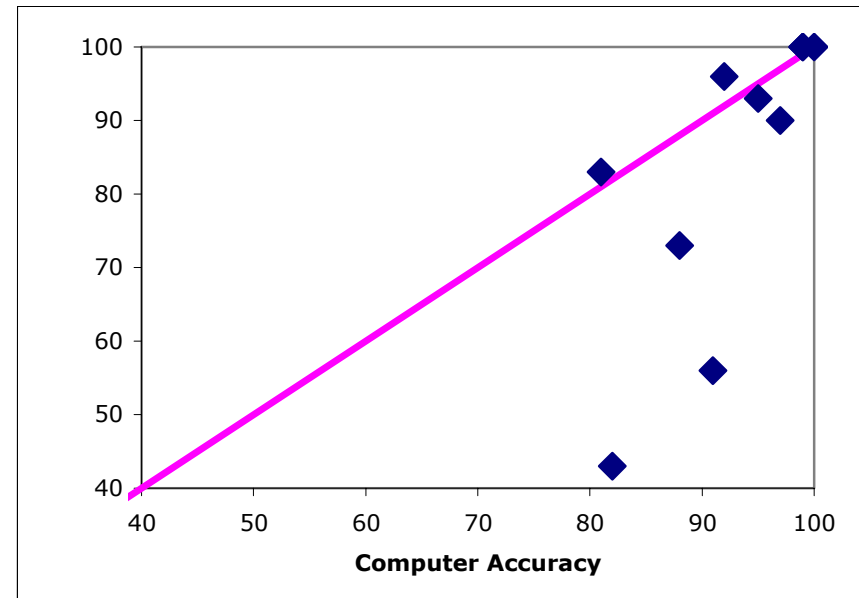
Robert F. Murphy

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



Progress over past 10+ years

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



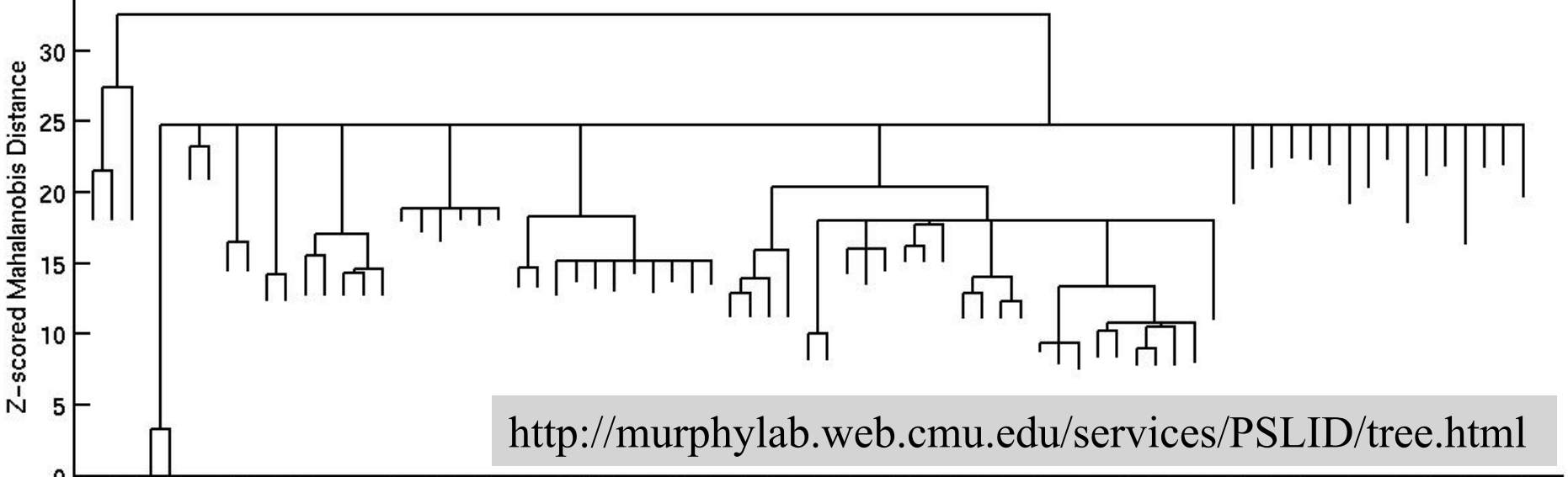


Why do we do HTM/HCS?

- (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



<http://murphylab.web.cmu.edu/services/PSLID/tree.html>

Protein name

Unknown-39
Unknown-36
Unknown-35
Unknown-31
Unknown-28
Unknown-27
Tpm4
Efl-gam
2610301D06R1k-2
Cnm2
Kars
8430422M09R1
Rpl36
Unknown-19
Unknown-15
Unknown-14
Ddx3
Canx
Unknown-7
Unknown-11
SimilarToSiahbp
Ewsh
Hmg2-1
Hmg1-2
Hmg1-1
Nfix-1
Rtn3-1
**Rps6
Tubb2-1
Ltbpl-Pex12
Tctex1
Sh3d3
**Hmg2-2
Rps11
2610301D06R1k-1
Rp23-278K23
**Pr1m2
Anxa2
Unknown-4
Unknown-3
Anxa5
Unknown-41
Unknown-40
Unknown-37
Unknown-23
Unknown-22
Unknown-18
Txn1
Unknown-13
Rps25
Unknown-26
Unknown-21
Unknown-34
Unknown-33
U1HG
Unknown-24
Lgals1-2
Unknown-20
Timm23-5133400D
Mrps18b
Dial
**Epar
Atpsa1-1
Atdp-1
Atpsa1-2
Unknown-38
Sdrp
Sep15
Unknown-12
Glut1
Unknown-2
**Rpl32
Unknown-25

Human description

Nucl1+Cytopl1+PlasMemb
cytoplasmt+ nucleus
cytoplasmt+ nucleus
nucleus+ w cytoplasm
cytoplasmt+ w cytoplasm
cytoplasmt+ w cytoplasm
cytoplasm
cytoplasm
weak uniform
weak uniform
cytoplasm
cytoplasm
cytoplasm+nuclMemb
cytoplasmt+ nucleus
nucleus+ w cytoplasm
nucleus+ w cytoplasm
nucleus+ w cytoplasm
nucleus
nucleus
nucleus
nucleus
nucleus+ER
ER
Mito
cytoskeleton
cytoplasm
cytoskeleton
cytoskeleton
cytoplasm
cytoplasm+ nucleus
cytoplasm
cytoplasm
cytoplasm
cytoplasm
cytoplasm
cytoplasm
NucLMemb+ w uniform
cytoplasm
cytoplasm
cytoplasm
sm cytop part+ w nucl
cytoplasmt+ nucleus
uniform
uniform
uniform
uniform
uniform
uniform
uniform
small cytop
small cytop
small cytop
nucl+nuclMemb
Mito
Mito
Mito
small cytop
Mito
Mito
cytoplasm
cytoplasm
Mito
Mito
cytoskeleton
cytoplasmt+Pl
cytoplasm
nucleolar
nucleolar
nucleolar

From databases

Mito+Rib+Unk
Nuc
Mito
Mito
unknown
Rib+Unk

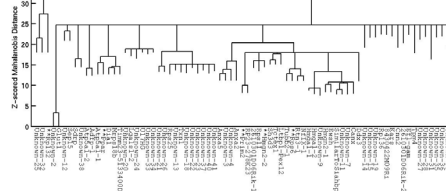


The problem

**Cell Type
(Order 10^2)**

**Condition
(Order 10^2)**

Protein (Order 10^6)



**Plus: Time scale from subsecond
to years**



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)

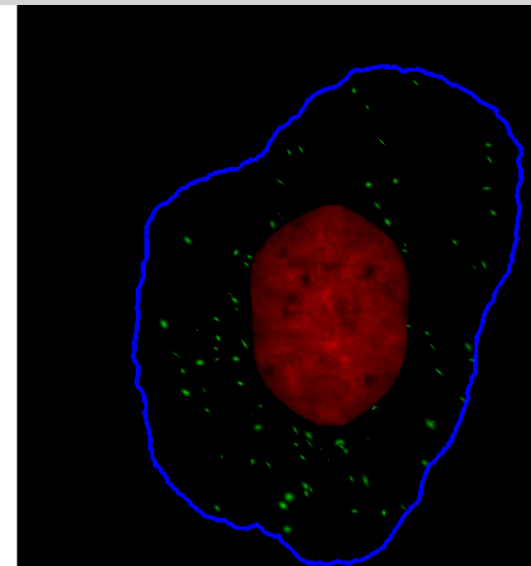
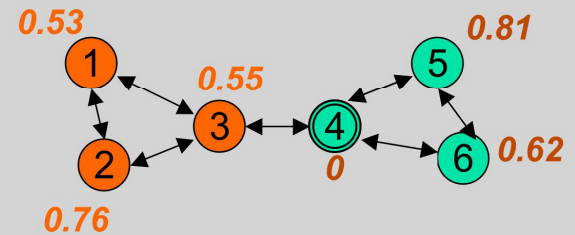


How do we really analyze subcellular location?

- 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





Human Cytome Project?

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