chris wiggins

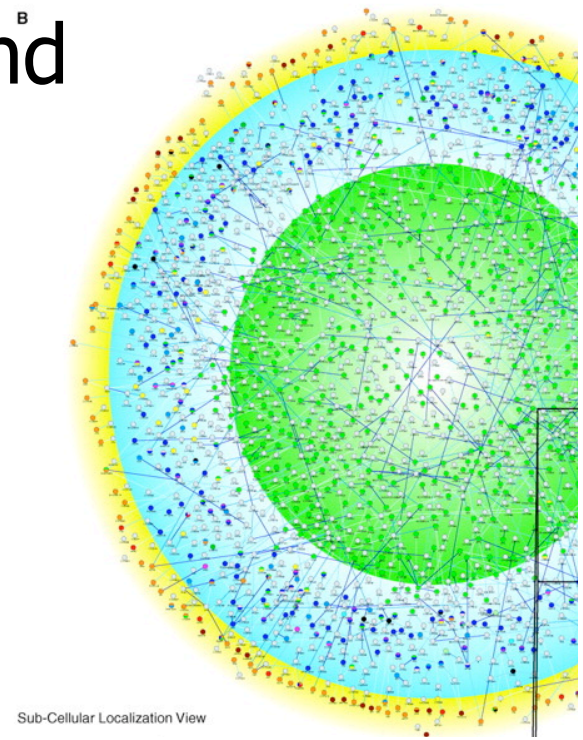
department of applied physics and applied mathematics

+

center for computational biology and bioinformatics

columbia university

chris.wiggins@columbia.edu



schedule:

Morning Session, May 04, 2007 (Friday) (Davis Auditorium)		
9:00-9:20	Chris Wiggins	Inverse problems and imaging for systems biology
9:30-10:00	Fred Chang	Dynamic assembly of cells: rings, filaments and tubules
10:10-10:40	Ruben L. Gonzalez	Single-molecule enzymology of complex biochemical reactions
10:50-11:20	Rafa Yuste	Reverse engineering the cortical microcircuit

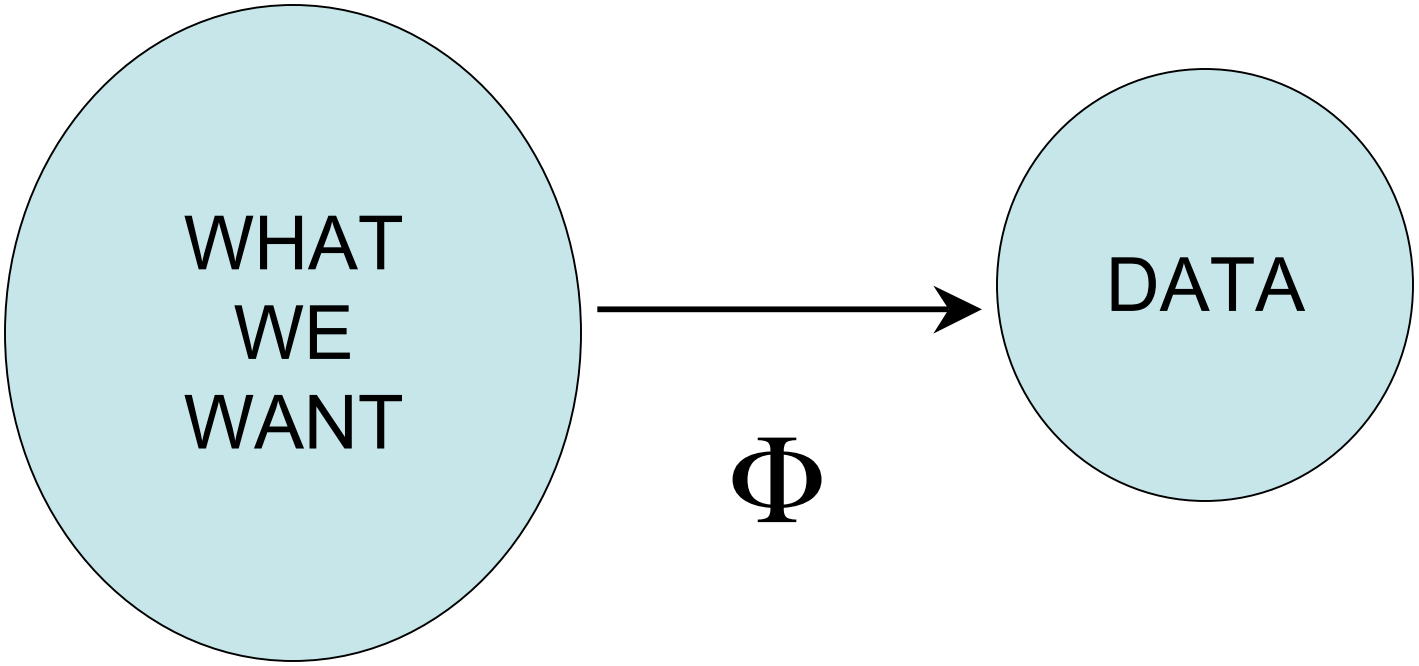
outline:

1. context
2. problems
   - inference problem
   - image problem
   - inference problem w/images
3. call to arms

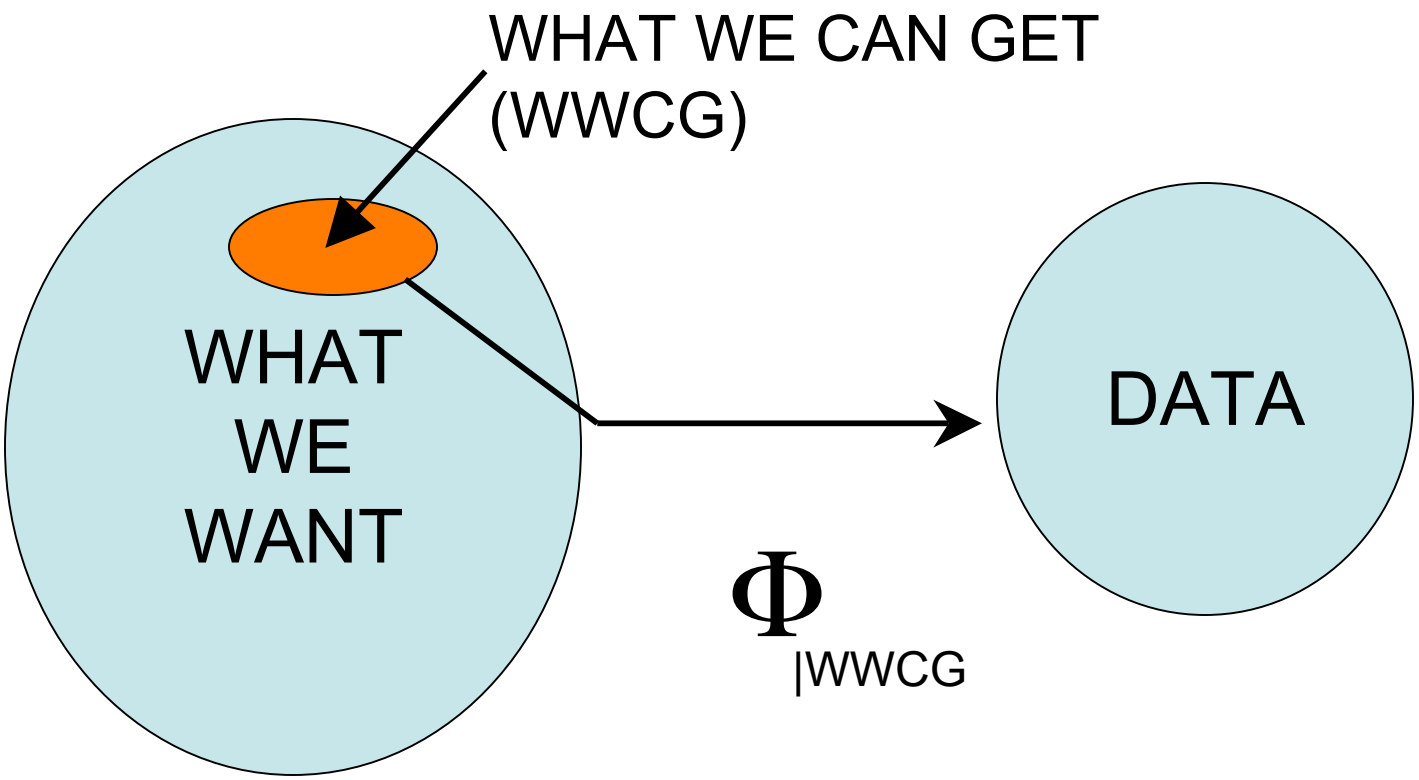
biological pathologies:

1. no “basic equation”:
   - model-building
   - validation
   - complexity control
2. goal of imaging
   - diagnosis/natural science vs.
   - image enhancing

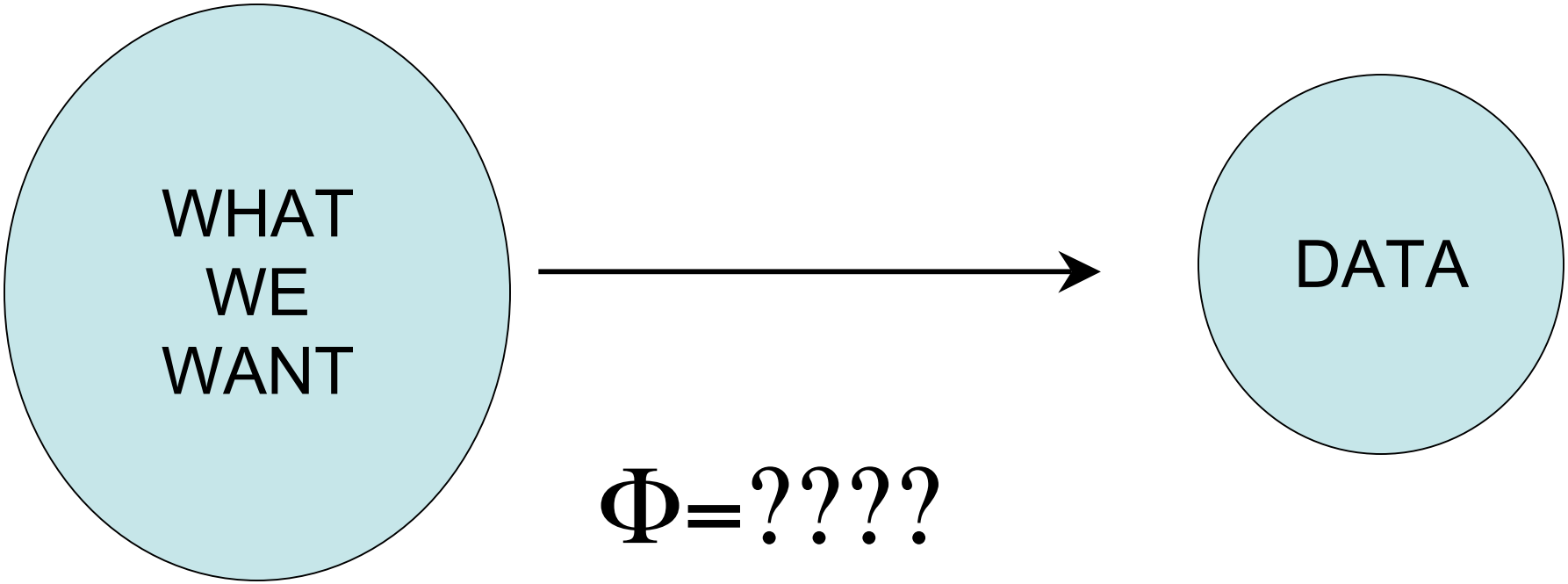
Problem: Φ is often SMOOTHING



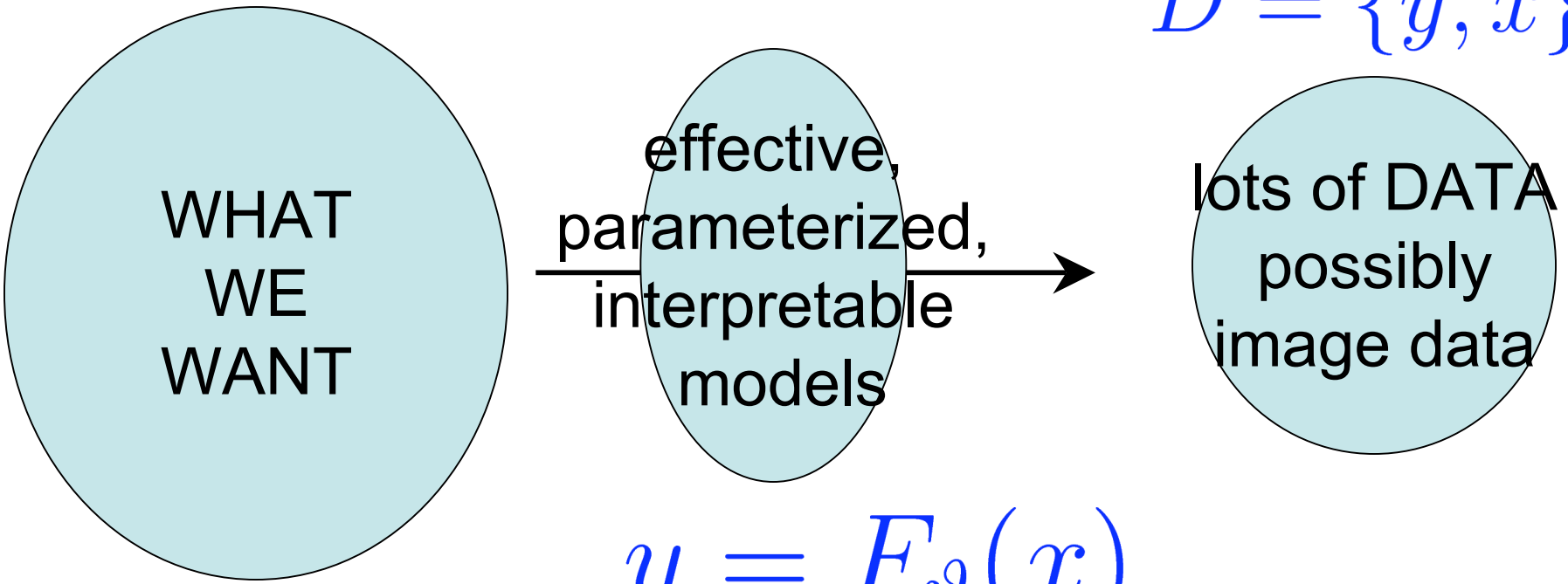
Solution: Simplify the problem



biological problem: we have NO CLUE what Φ IS



solution: learn parameters in effective or generative model



$$D = \{y, x\}$$

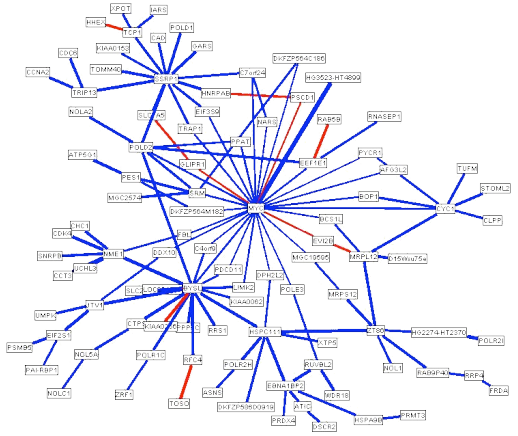
$$y = F_\vartheta(x)$$

$$p(D) = \sum_z p(D|z)p(z)$$

biological inverse problems

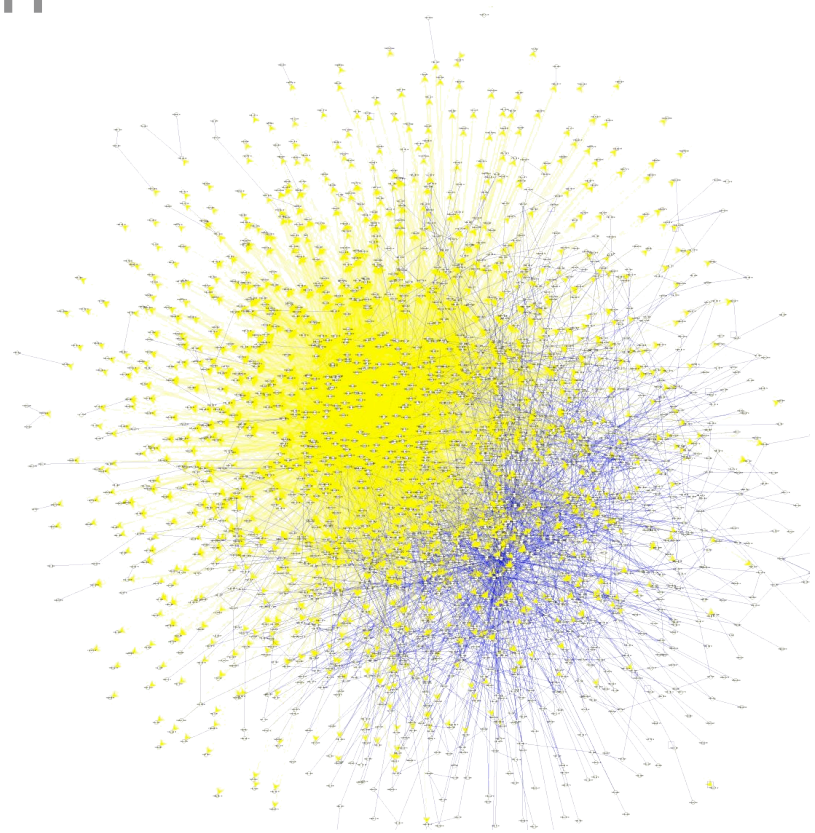
1. parameter learning,
   e.g., "z"
2. structure learning
   (a generic "systems"
   problem)

$$p(D) = \sum_z p(D|z)p(z)$$



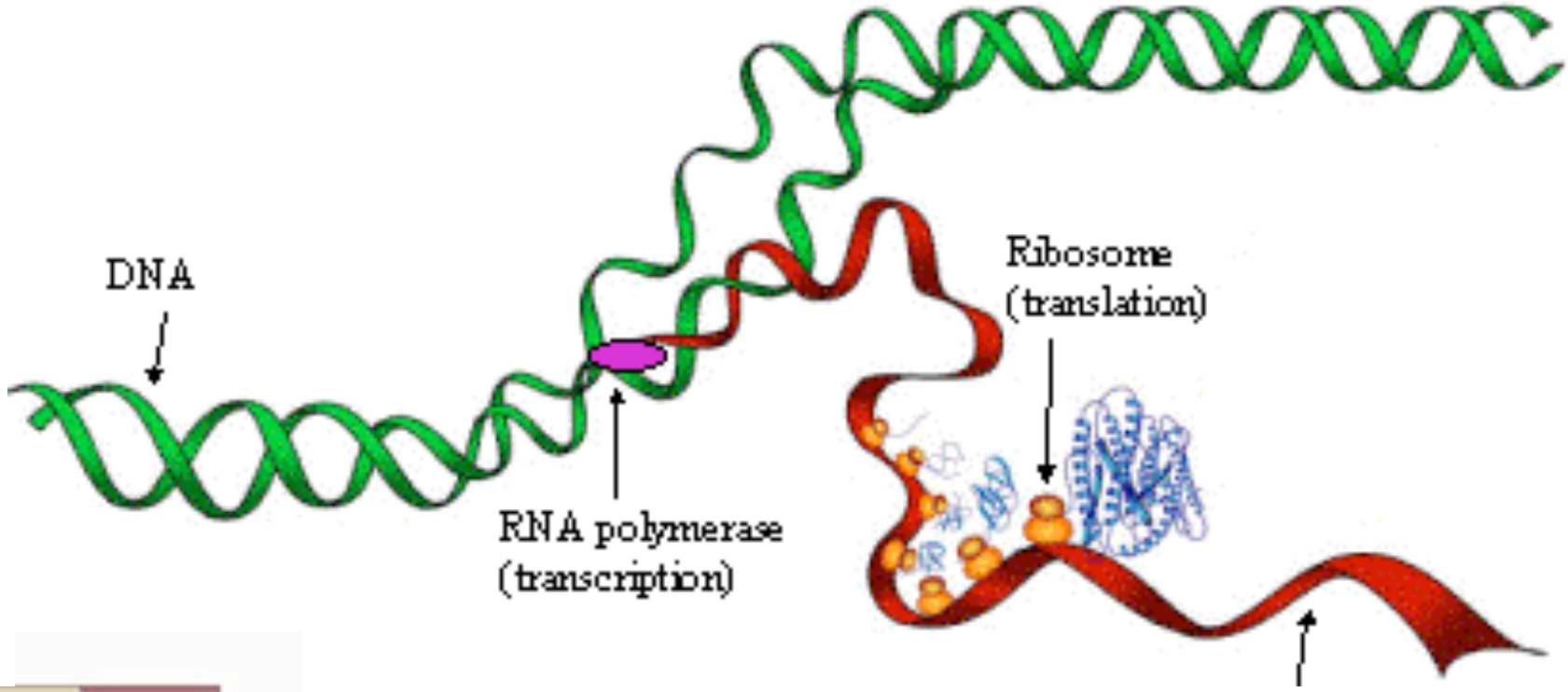
biological networks?

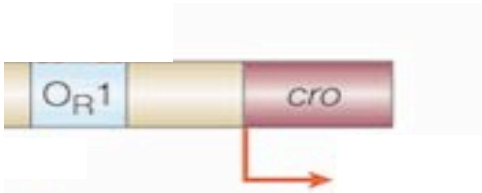
- protein-protein
- transcriptional regulation
- signaling



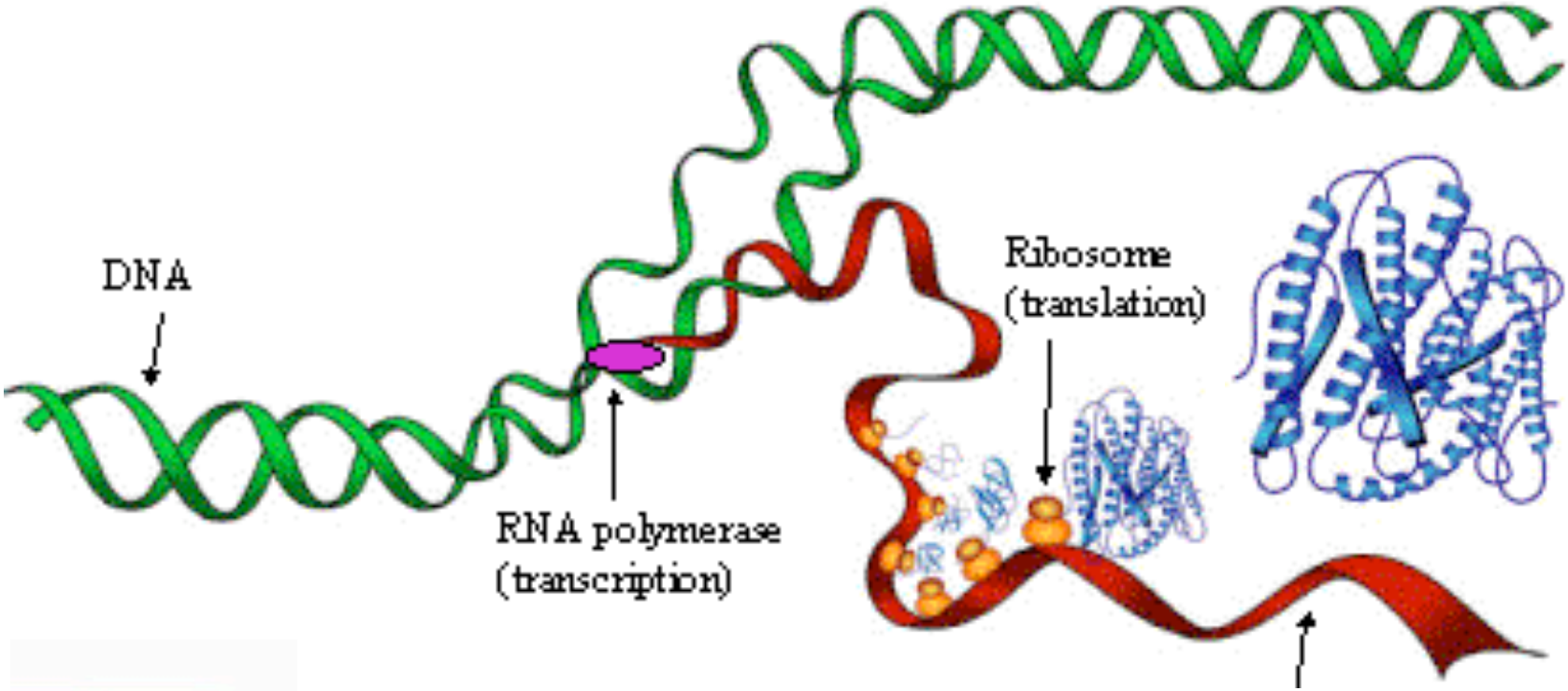
biology as told by a theorist

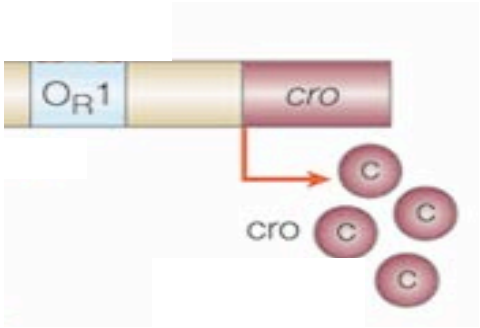
D->R->P



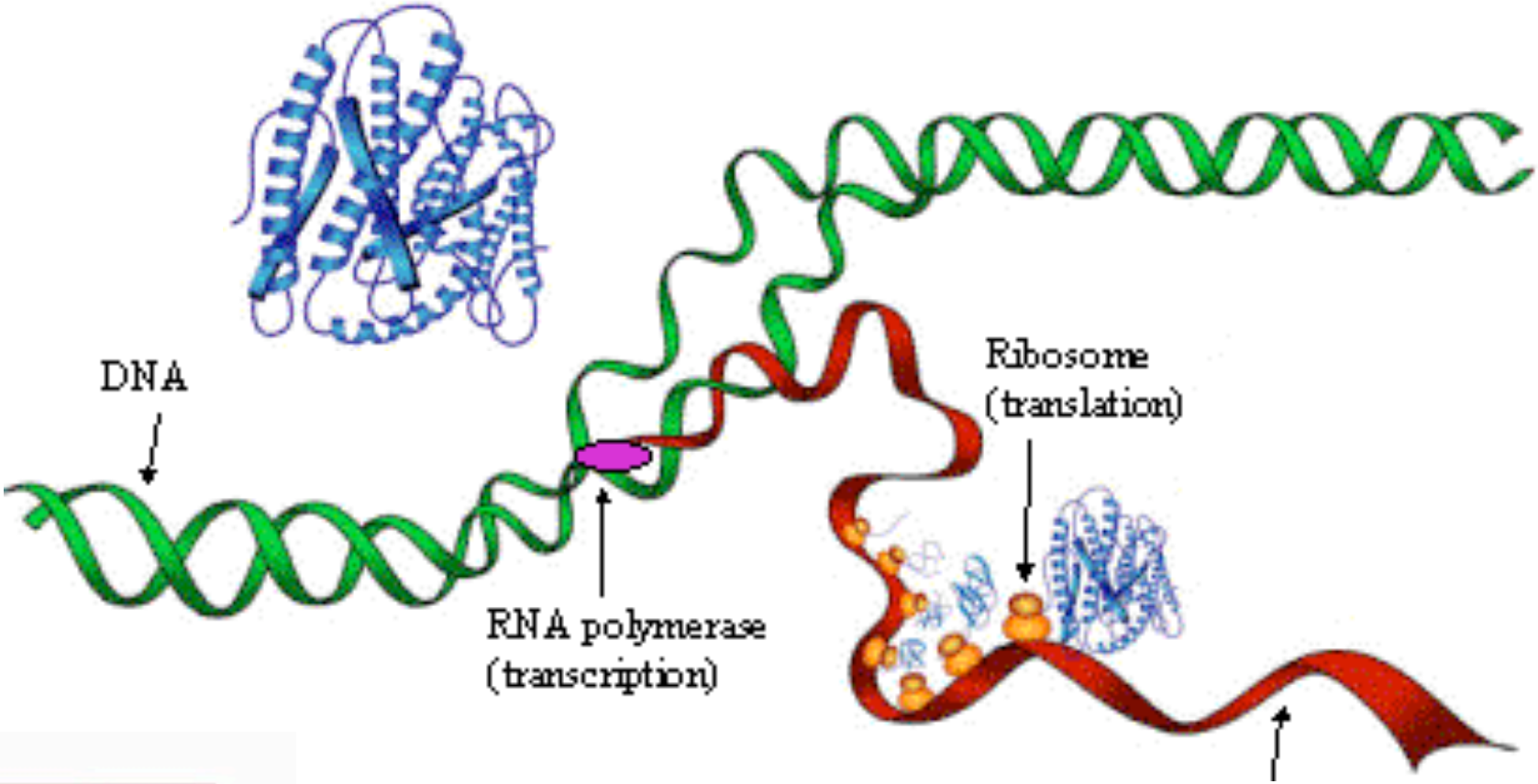


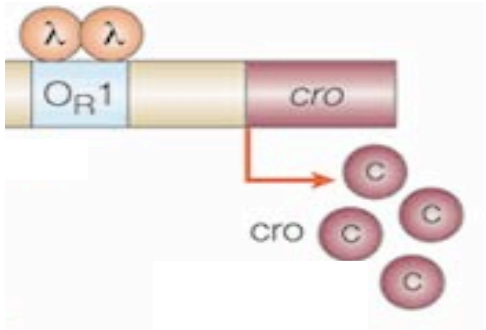
biology as told by a theorist



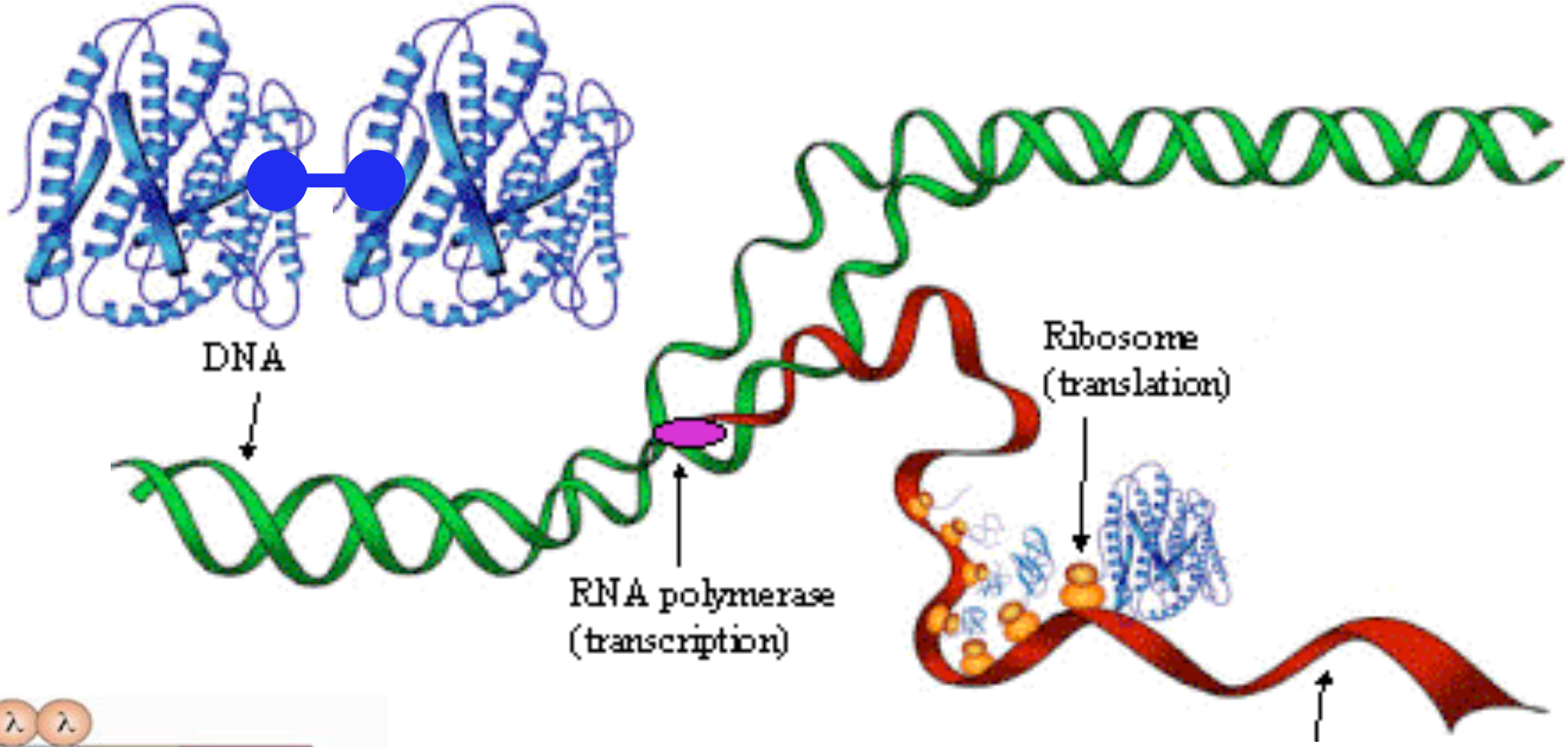


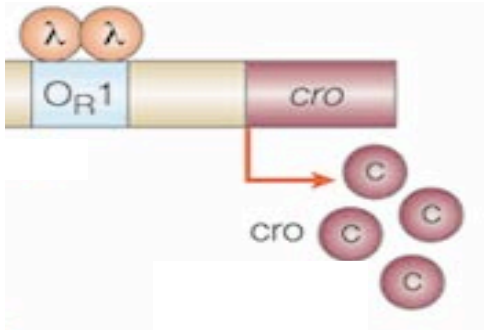
biology as told by a theorist



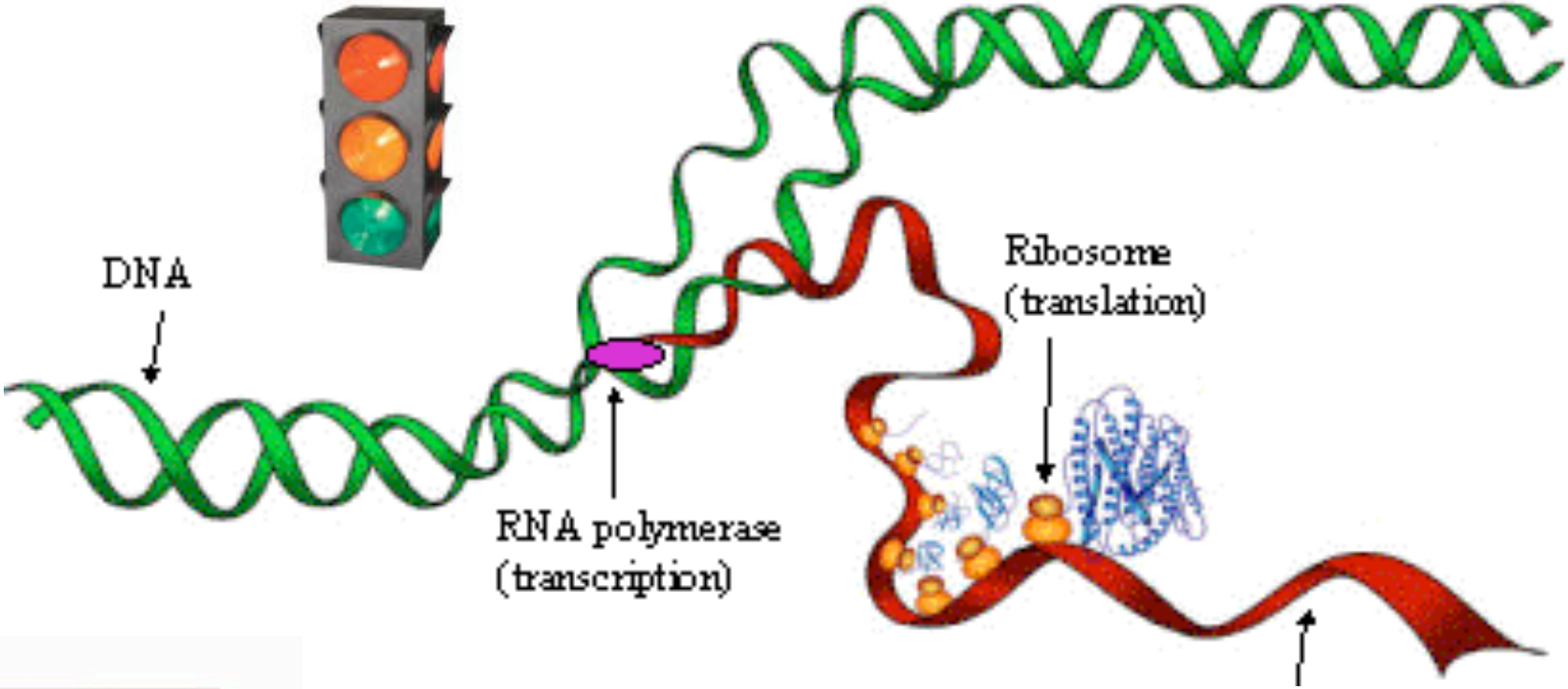


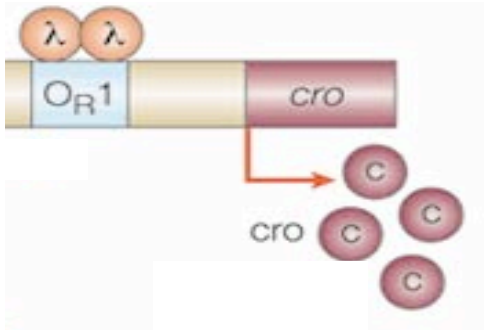
biology as told by a theorist



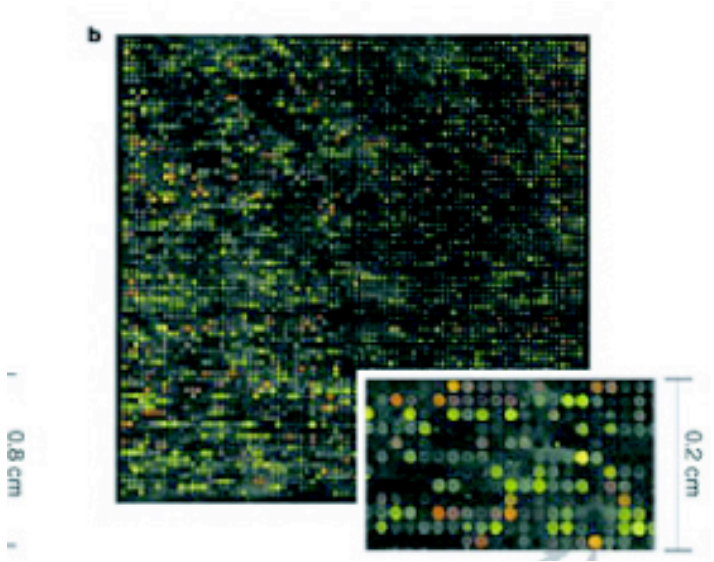


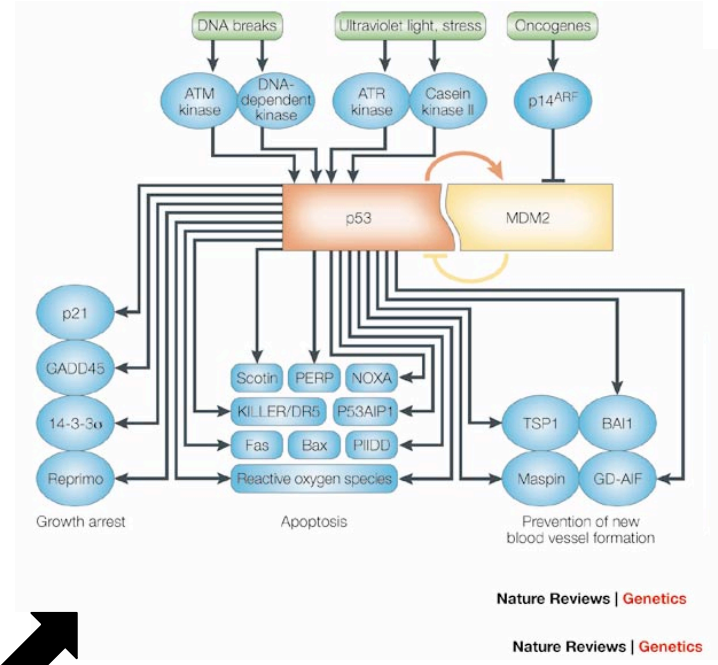
biology as told by a theorist





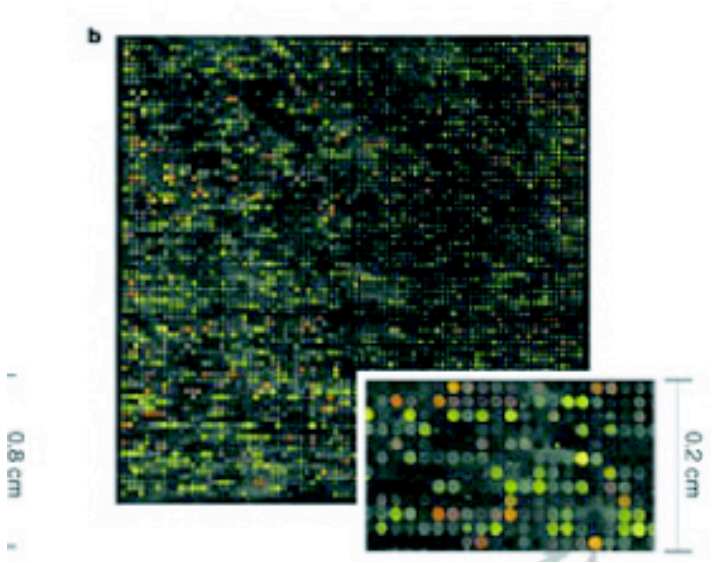
learning networks from biology

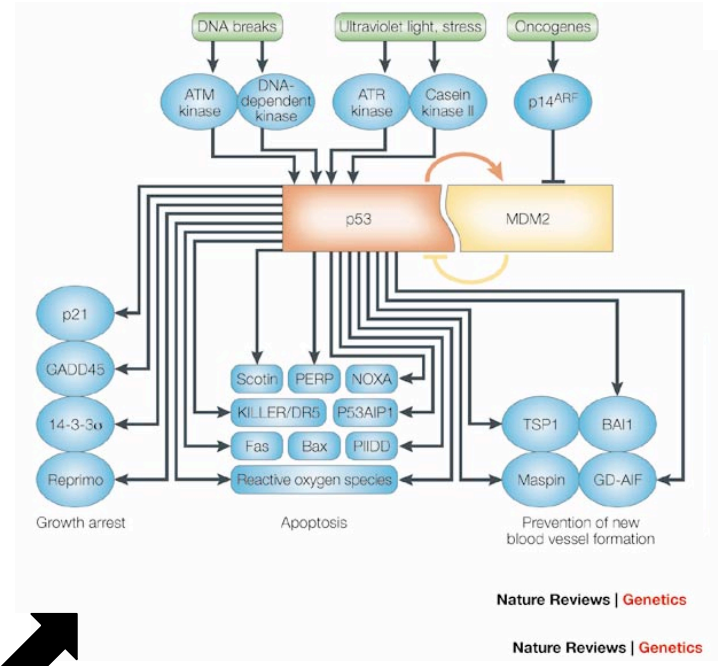




?

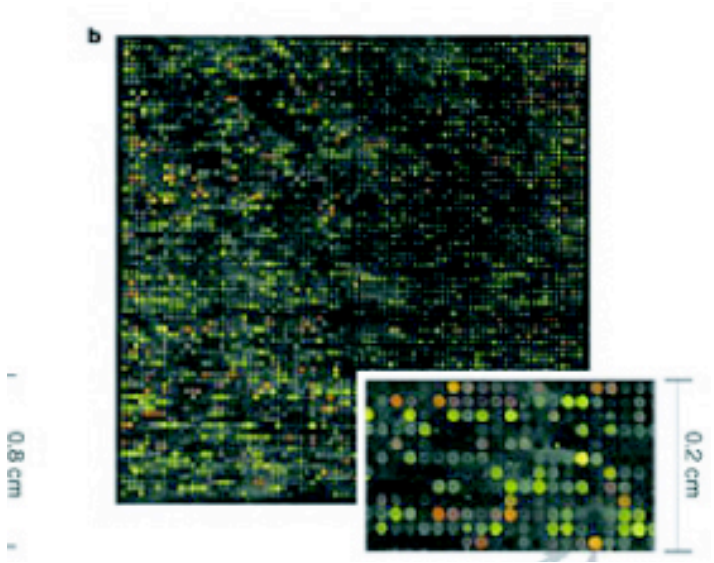
learning networks from biology

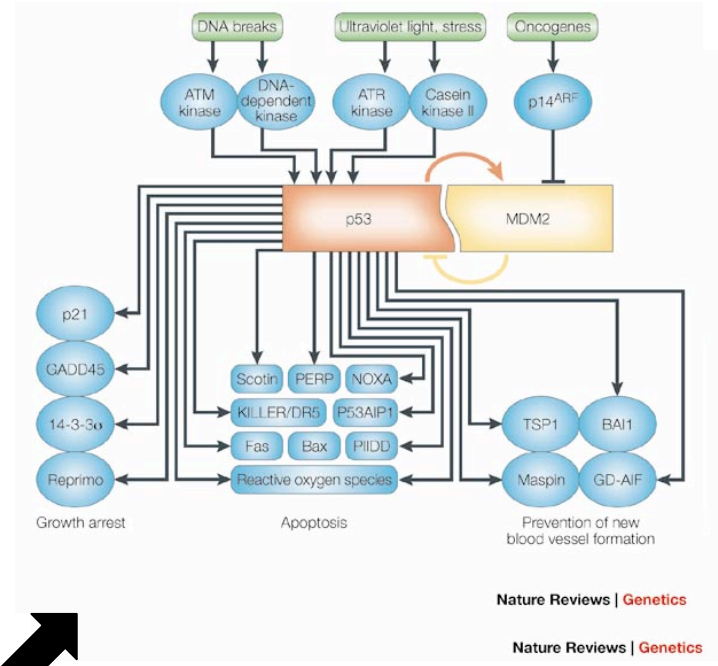




$$y = F_\vartheta(x)$$

learning networks from biology



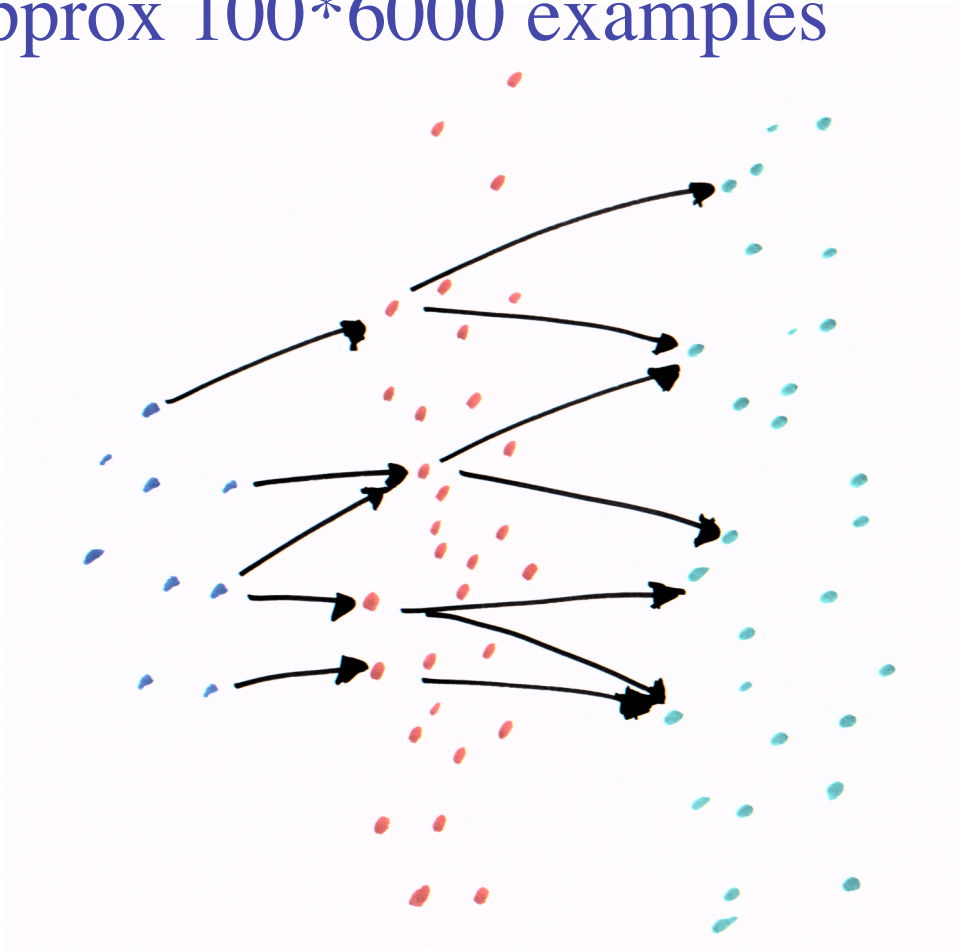


$$y = \sum_t \alpha_t \sigma_{\pi,\mu}$$

boosting: base on biological rules

parents - “motifs” - children

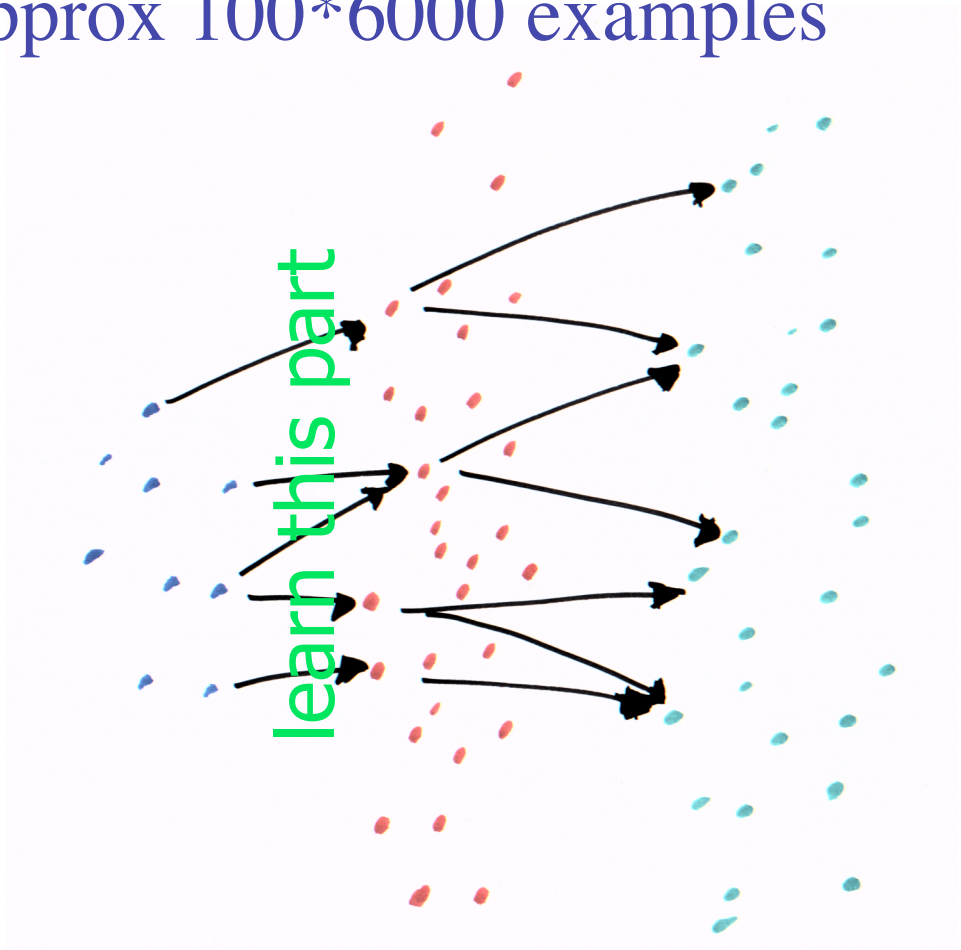
- 10M-dimensional feature space
- approx 100*6000 examples



boosting: add 1 edge at at time

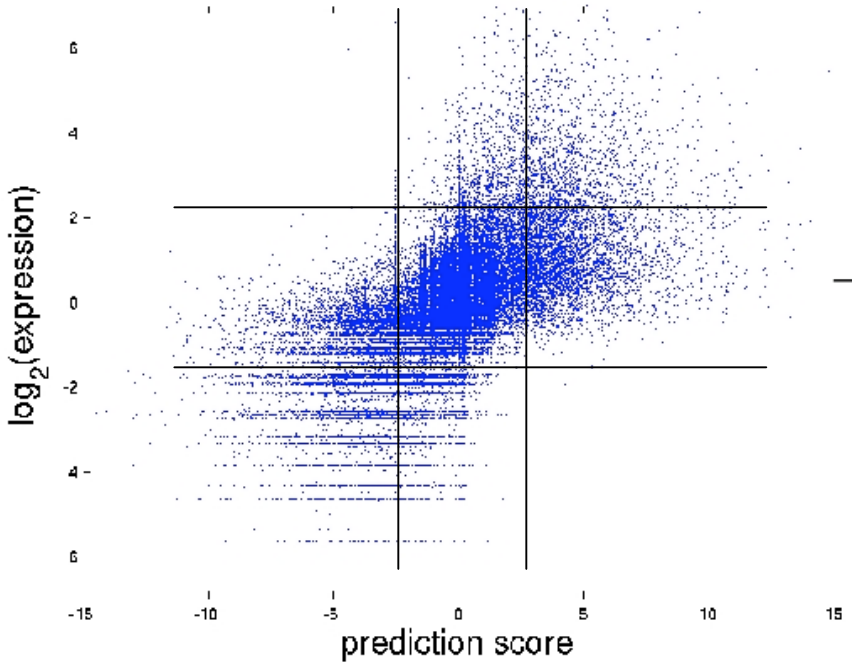
parents - “motifs” - children

- 10M-dimensional feature space
- approx 100*6000 examples



basic notions: fitting vs. overfitting

- "10-fold cross-validation" yields test loss of 13.6%



		Predicted Bins		
		Down	Baseline	Up
True Bins	Down	16.5%	8.9%	1.5%
	Baseline	9.3%	32.4%	6.3%
	Up	2.8%	9.9%	12.0%

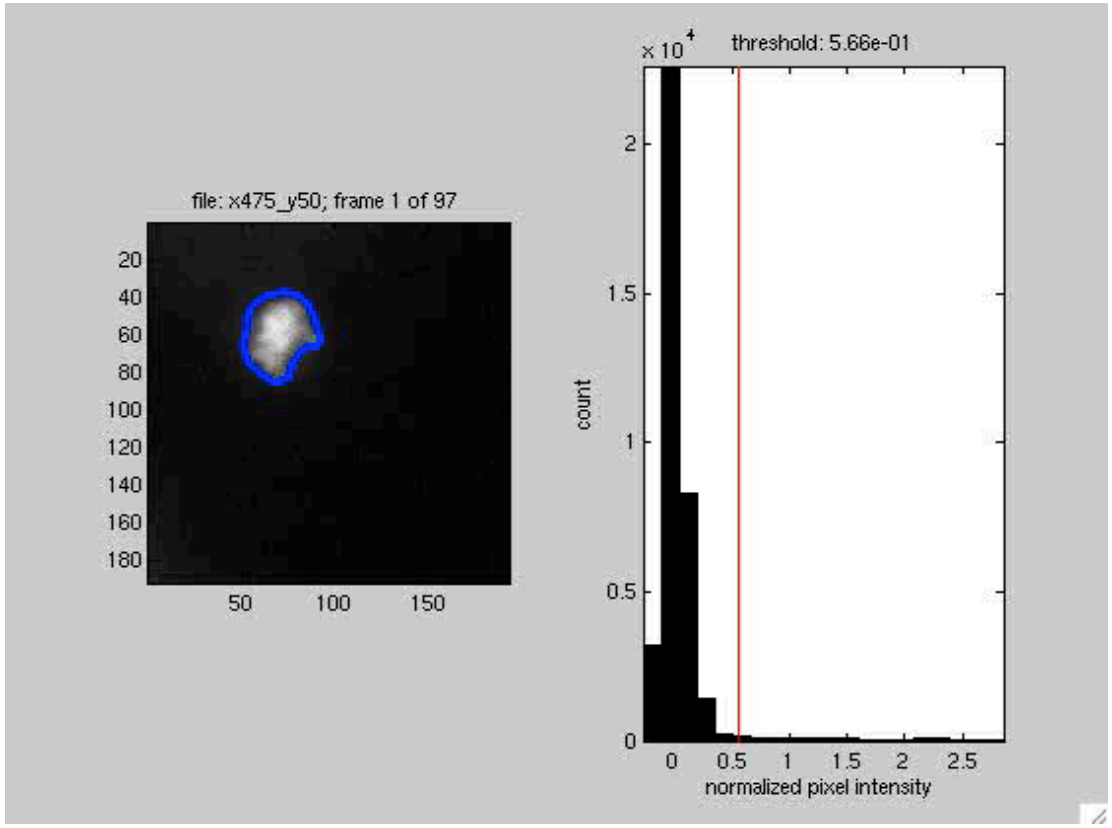
- Empirical estimate of generalization error
- not chi squared (not training data, and not normal)
- M. Middendorf, A. Kundaje, CW, Y. Freund and C. Leslie

outline:

1. context
2. problems
   - inference problem
   - image problem
   - inference problem w/images
3. call to arms

segmentation: PDEs, graph-cuts, or...

$$p(D) = \sum_z p(D|z)p(z)$$

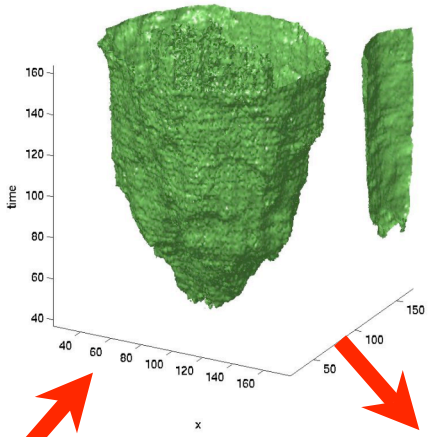


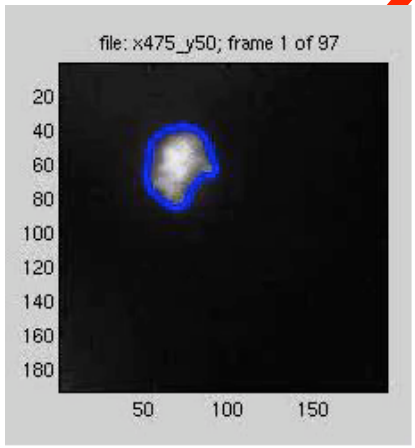
movie: tasha sims, dustin lab; harry xenias+ben dubin-thaler, sheetz lab

noVel: jake hofman

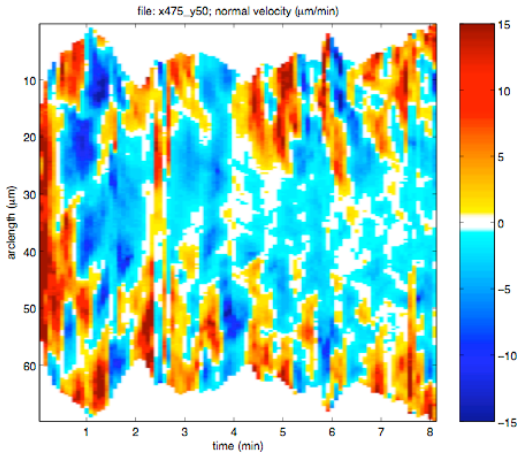
automated, statistical, “high-throughput” (whole movies) quantitative motility assays

1. identify cell edge
(statistical segementation)

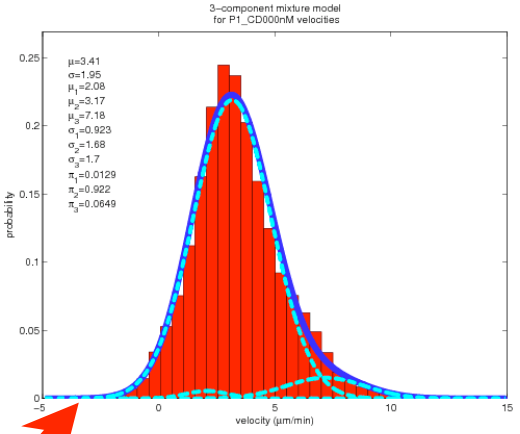




2. evaluate normal velocity
(optical flow)

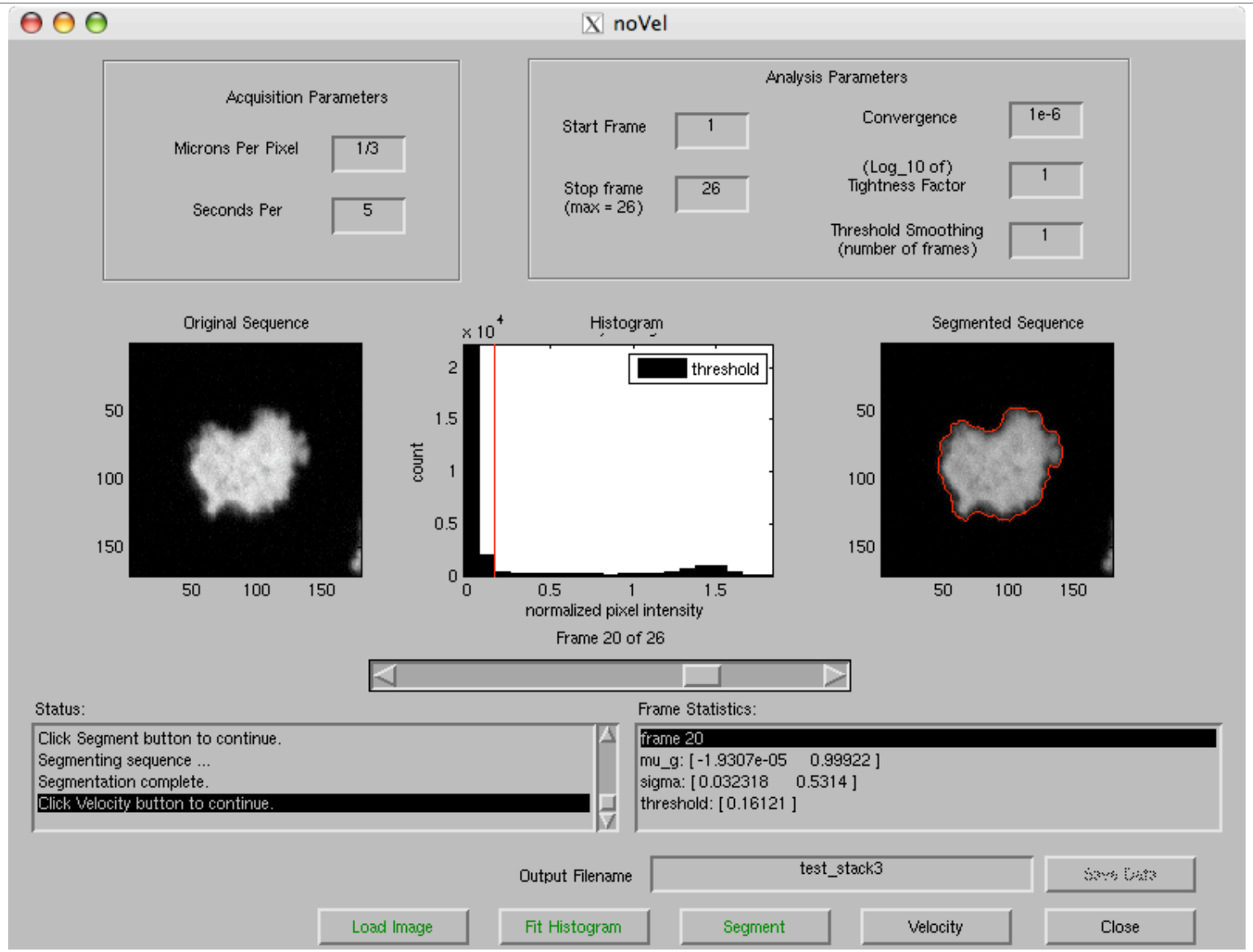


3. map onto surface
(differential geometry)

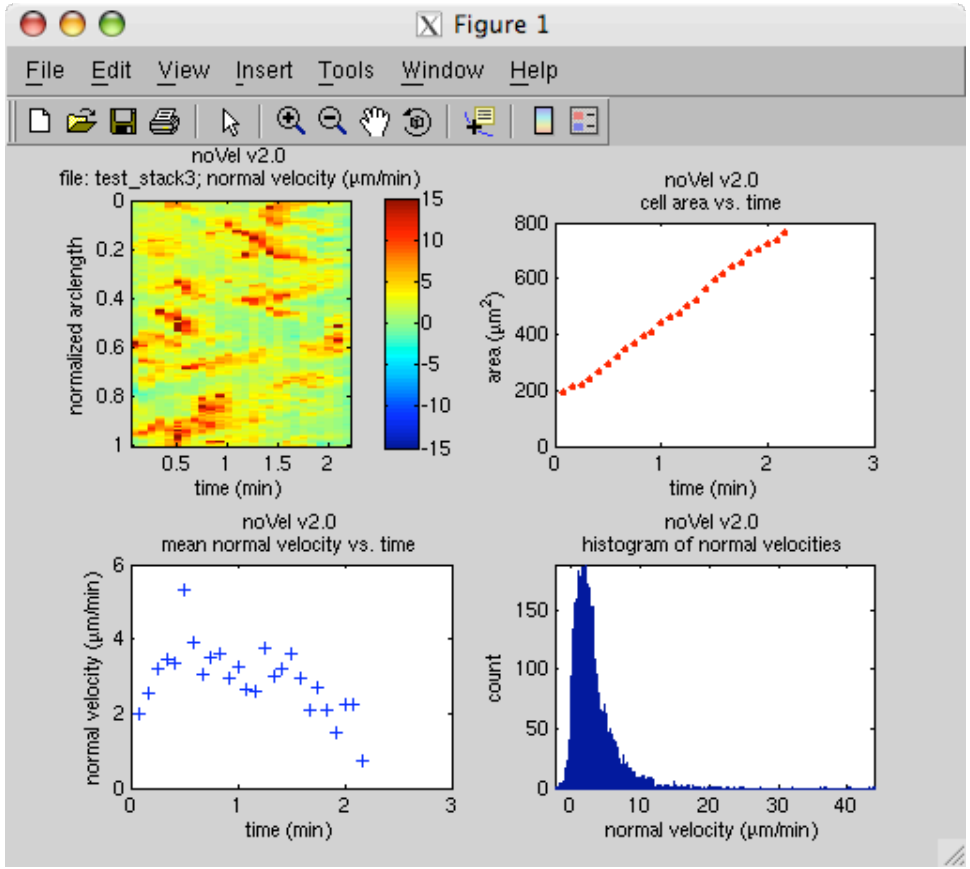


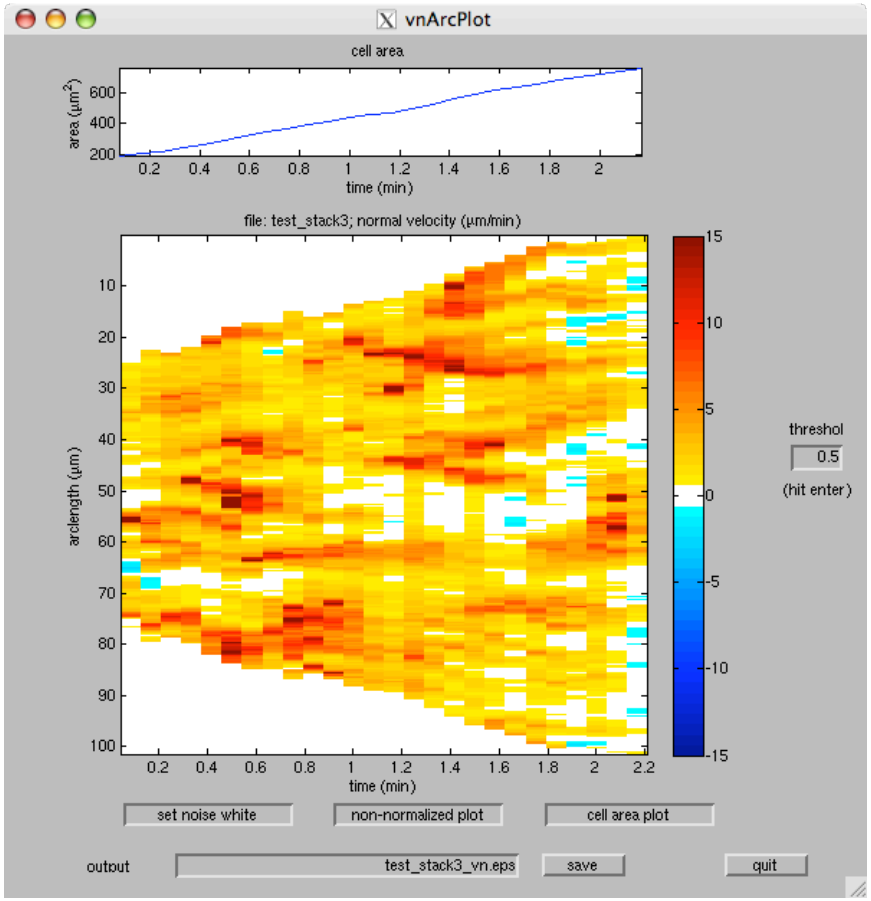
4. analyze velocity statistics
(mixture modeling)

noVel: software interface



noVel: software interface (cont'd)

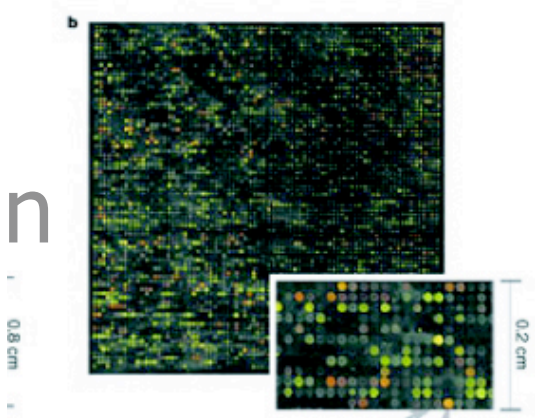
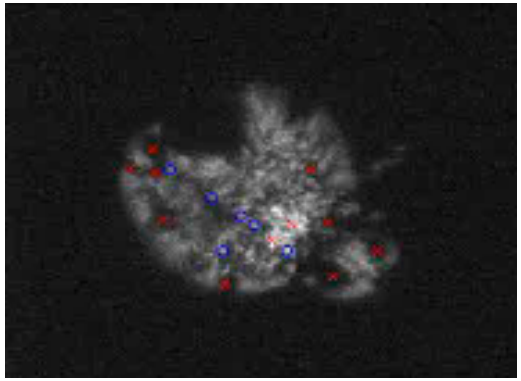




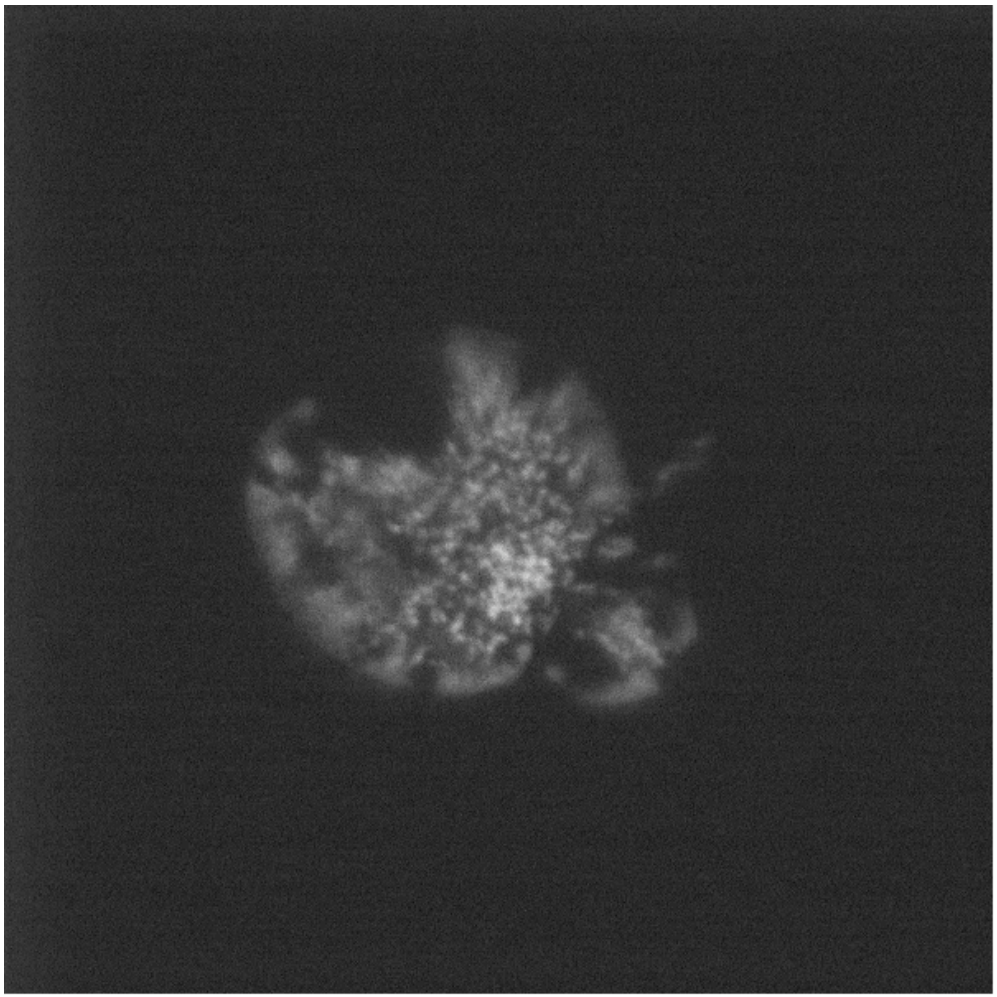
outline:

1. context
2. problems
   - inference problem
   - image problem
   - inference problem w/images
3. call to arms

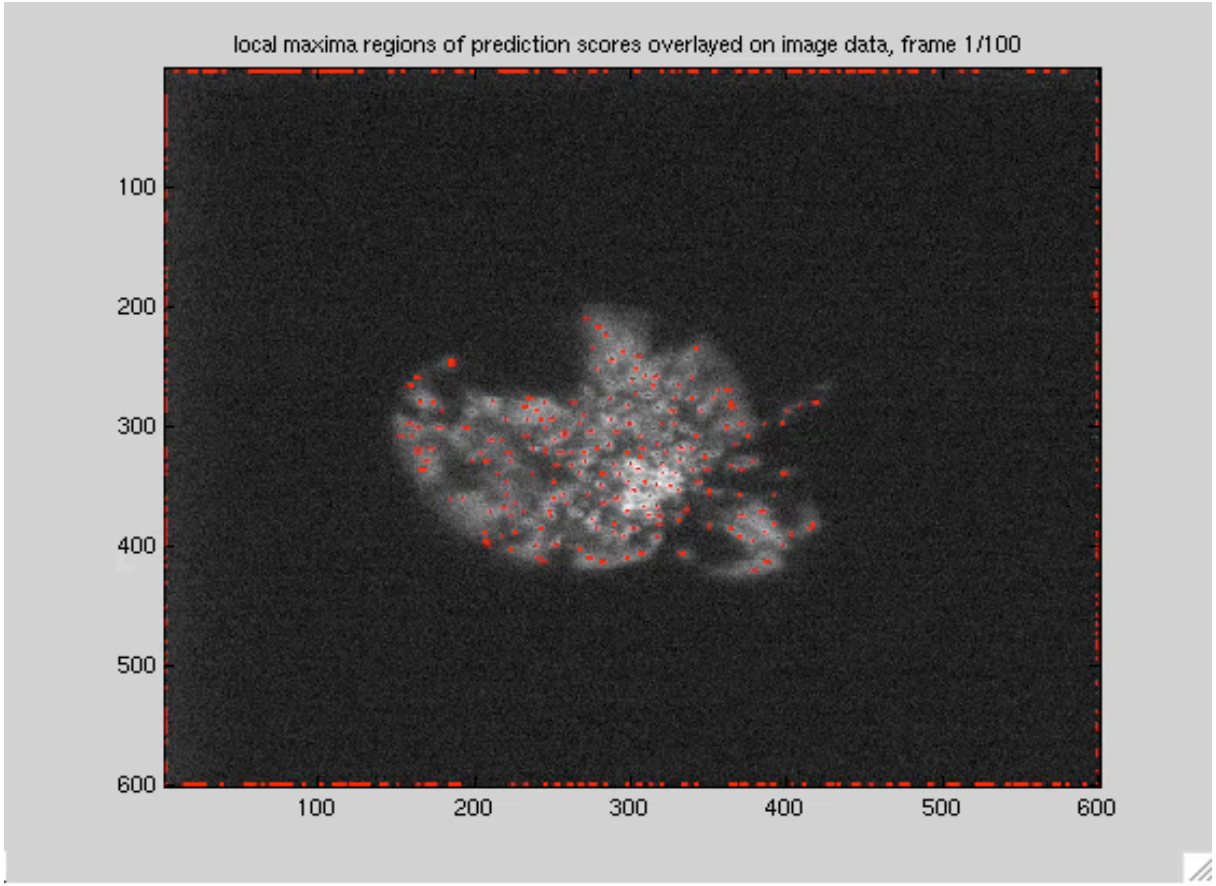
inference + images

what can  do for  ?

foci detection: modern machine learning+microscopy



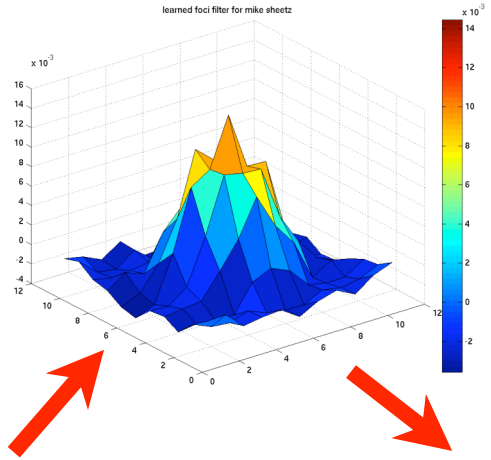
foci detection: modern machine learning+microscopy



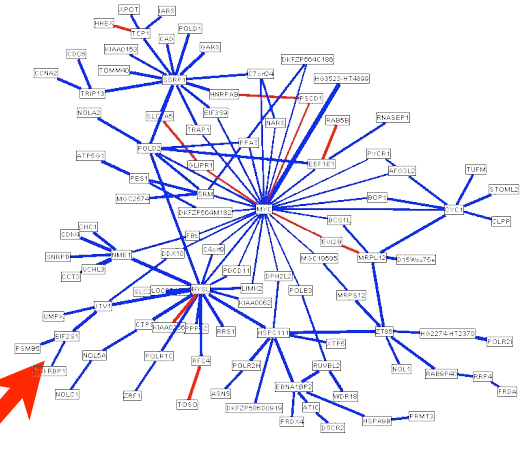
foci detection: simulate biologists, not biology

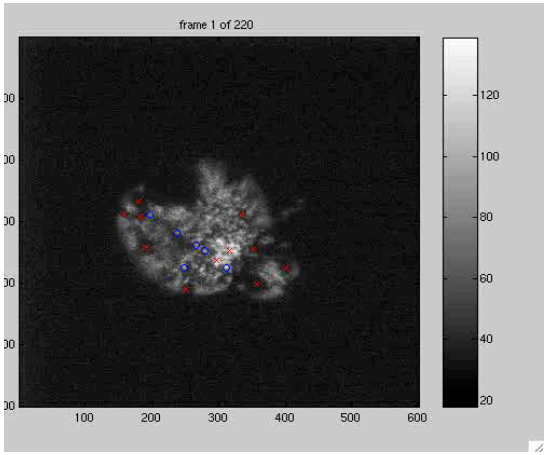
$$\partial_t u = D_u \nabla^2 u + f(u,v)$$
$$\partial_t v = D_v \nabla^2 v + g(u,v)$$

1. label images
(HCI)

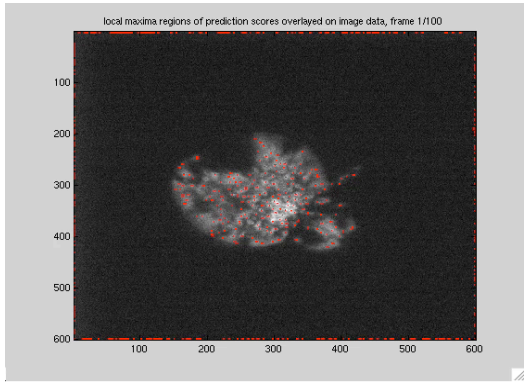


3. correlate with disease states and phenotypic features
(biology)





2. learn detector
(SVMs)



4. develop network + dynamical models
(applied math)

outline:

1. context
2. problems
   - inference problem
   - image problem
   - inference problem w/images
3. call to arms

more info

- inference/reverse engineering
  - ISMB 2004
  - RECOMB 2005
  - open source: www.cs.columbia.edu/compbio/
- bioimage data
  - Cell 2007
  - Biophysical Journal 2007
  - PRL 2006
  - Biophysical Journal 2006
  - Journal of Cell Science 2006
  - open source: cellmap.sourceforge.net
- chris.wiggins@columbia.edu