

# **Genetic Tuning of Single-Cell Morphological Variability**

**Oaz Nir<sup>1</sup>, Chris Bakal<sup>2</sup>, Norbert Perrimon<sup>2</sup>, Bonnie Berger<sup>3</sup>**



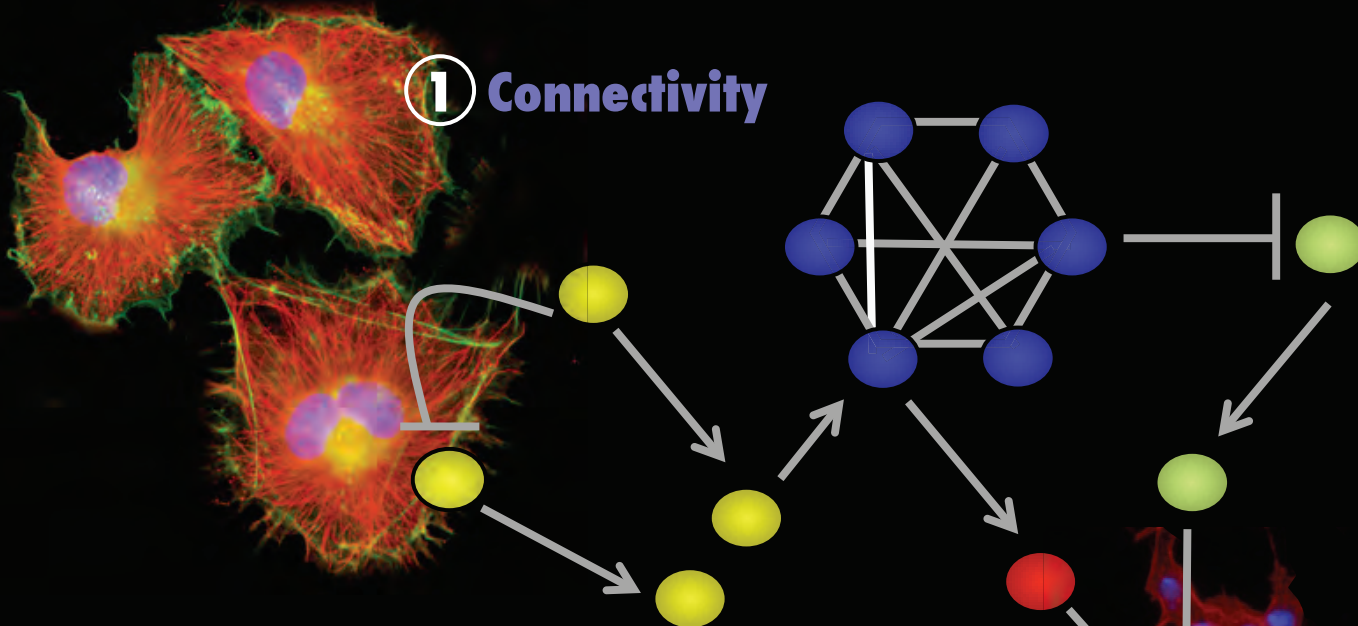
**<sup>1</sup> Harvard/MIT Division of Health Science and Technology and Applied Mathematics, MIT.**

**<sup>2</sup> Department of Genetics, Harvard Medical School, Howard Hughes Medical Institute.**

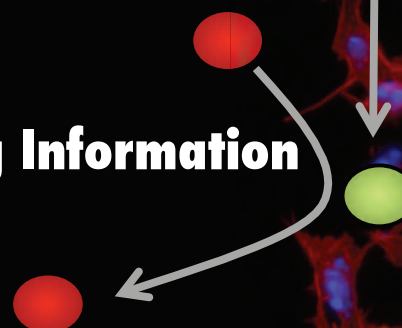
**<sup>3</sup> Applied Mathematics and Computer Science and Artificial Intelligence, MIT, and  
Harvard/MIT Division of Health Science and Technology.**

# Challenges in Describing Signaling Networks

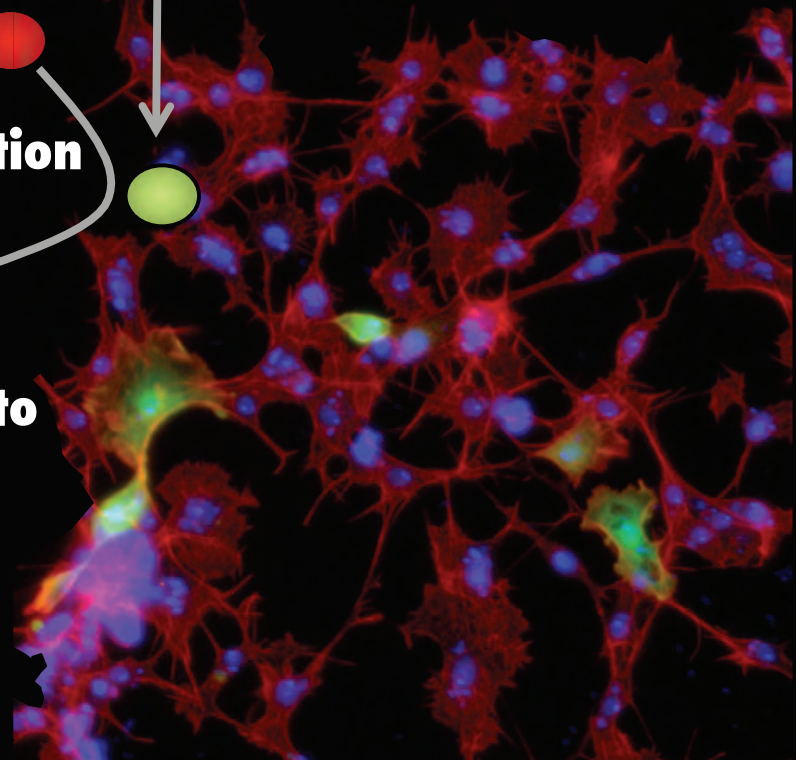
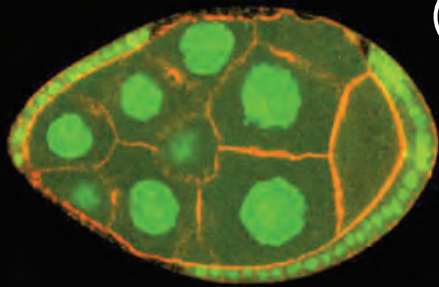
① **Connectivity**



② **Flow of Signaling Information**



③ **Variability of Response to Signaling Information**



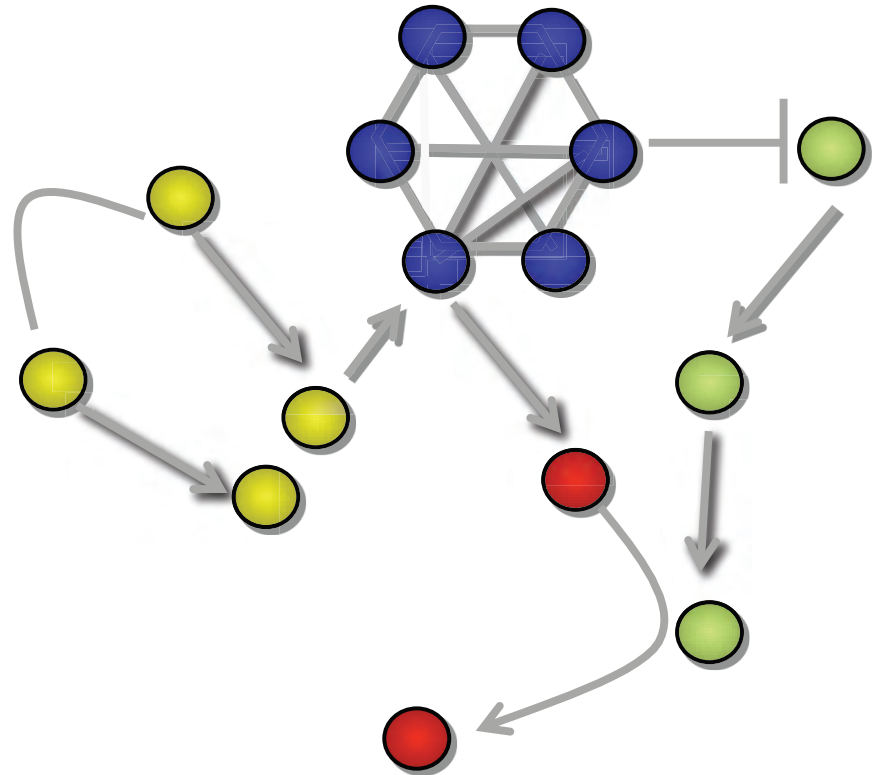
# How Do You Study Morphological Variability in Cellular Processes?

## Outline

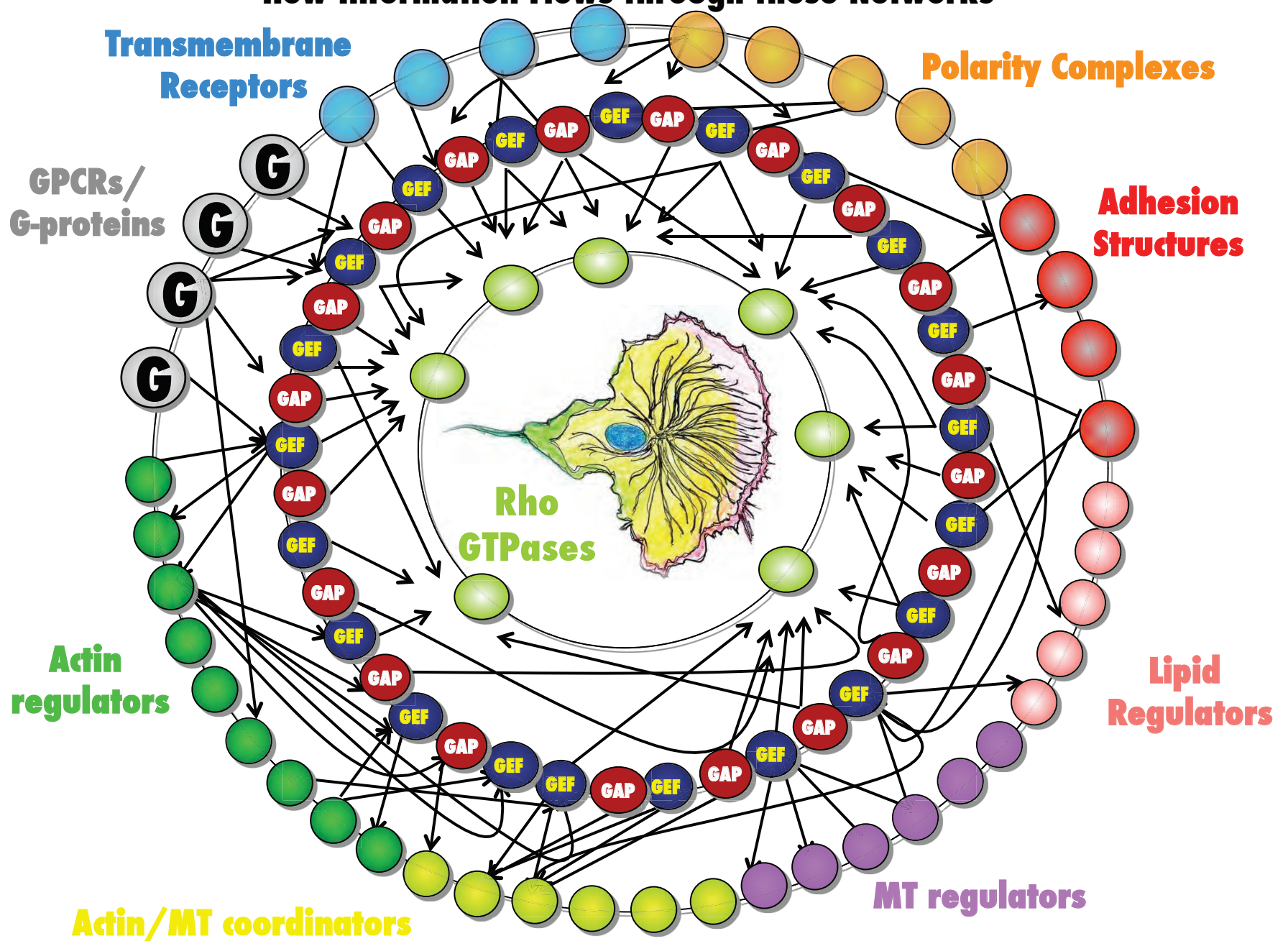
1. New kind of data  
*morphology = signaling state*

2. How to measure  
morphological variability  
from this data

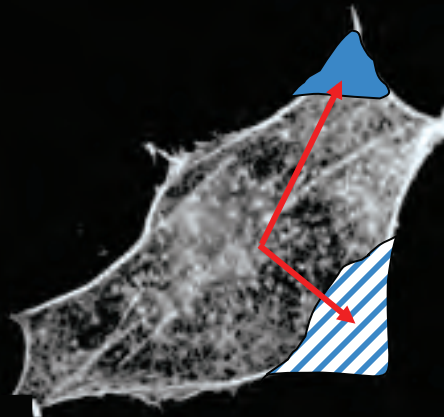
3. Understanding  
variability in the context  
of signaling network  
architecture



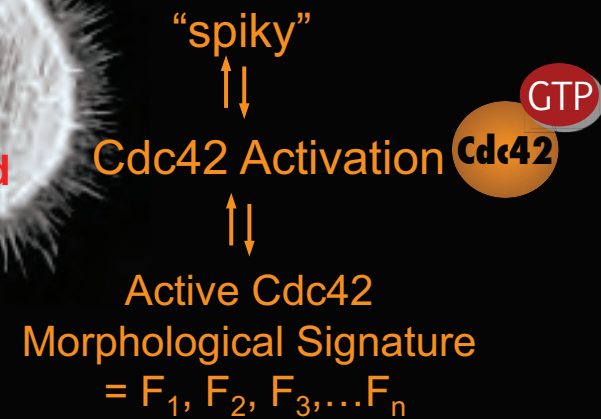
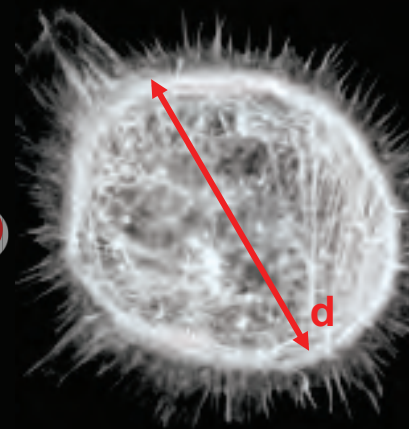
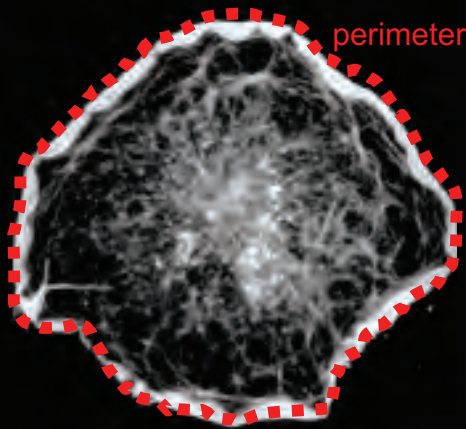
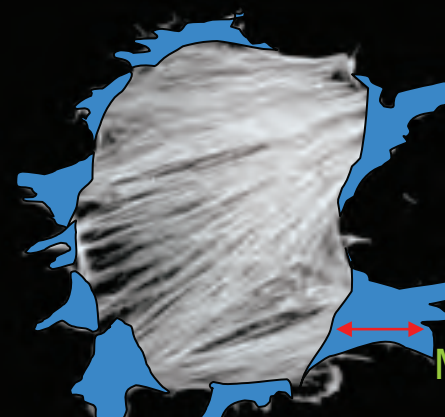
# Understanding how Signaling Networks that Regulate Morphology are Organized and How Information Flows through these Networks



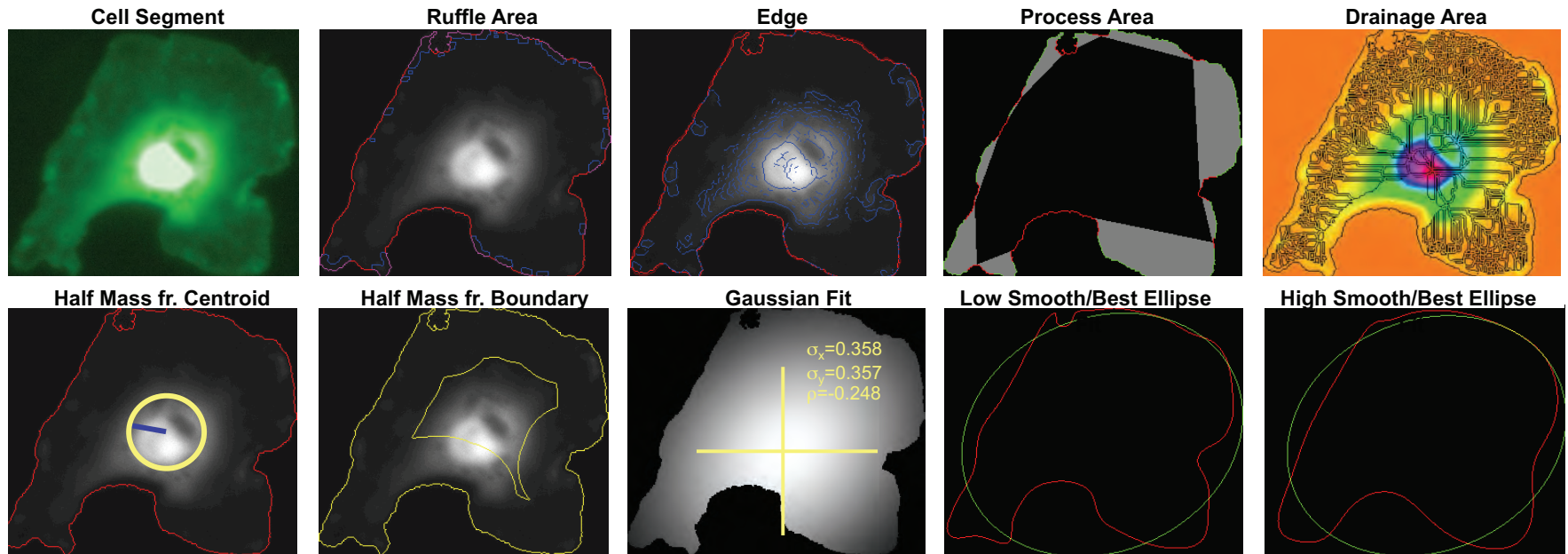
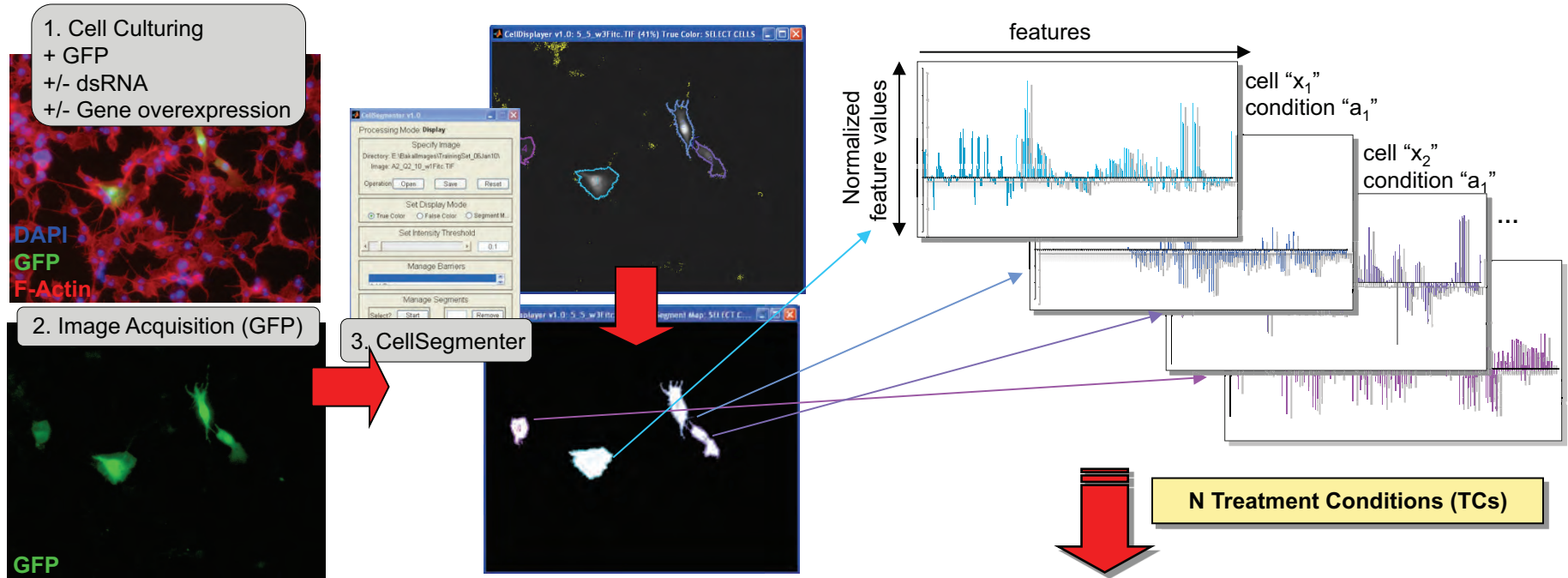
# Cell Morphology = Signaling State



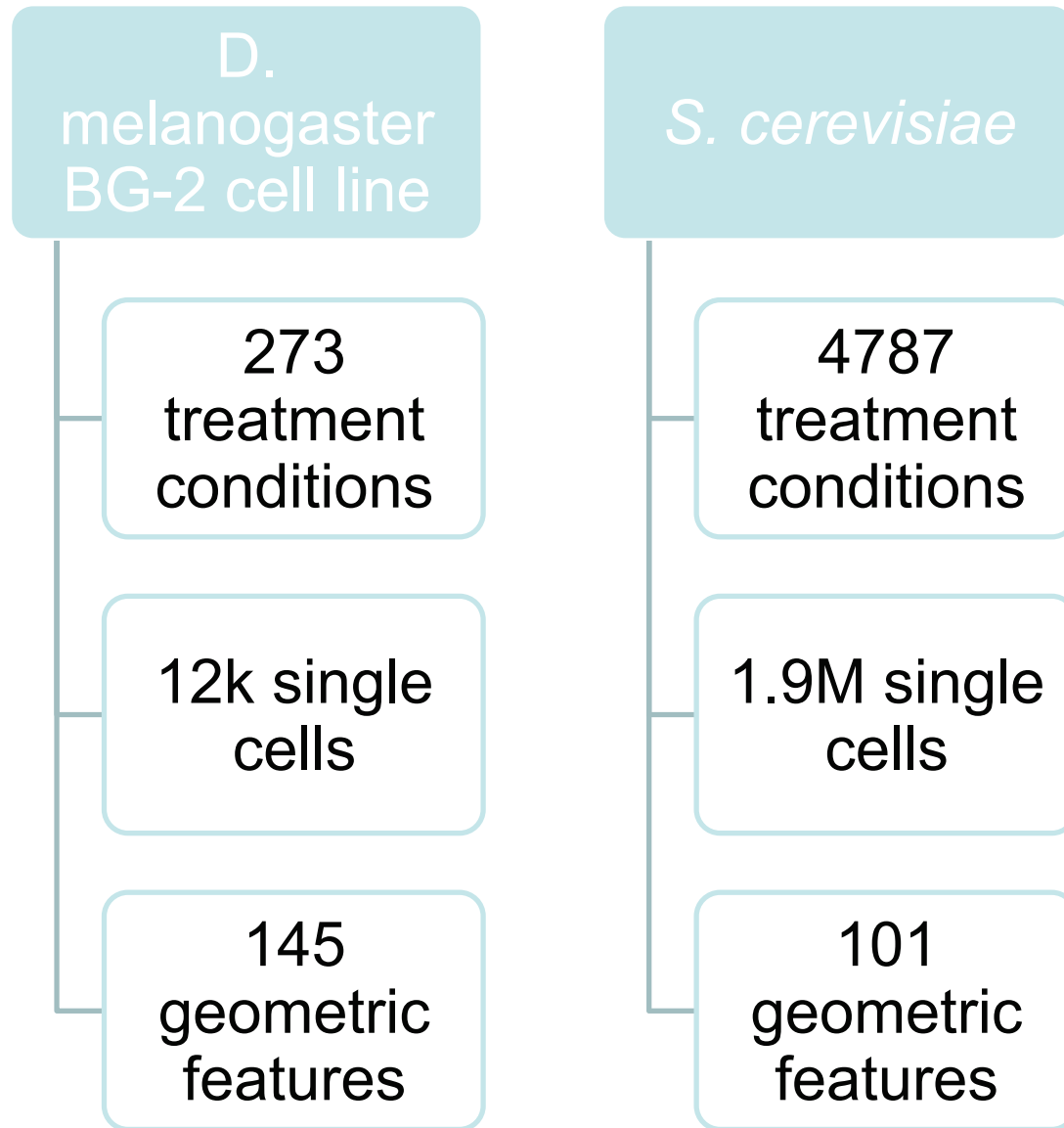
normal



# Acquiring Morphological Signatures from Complex Images

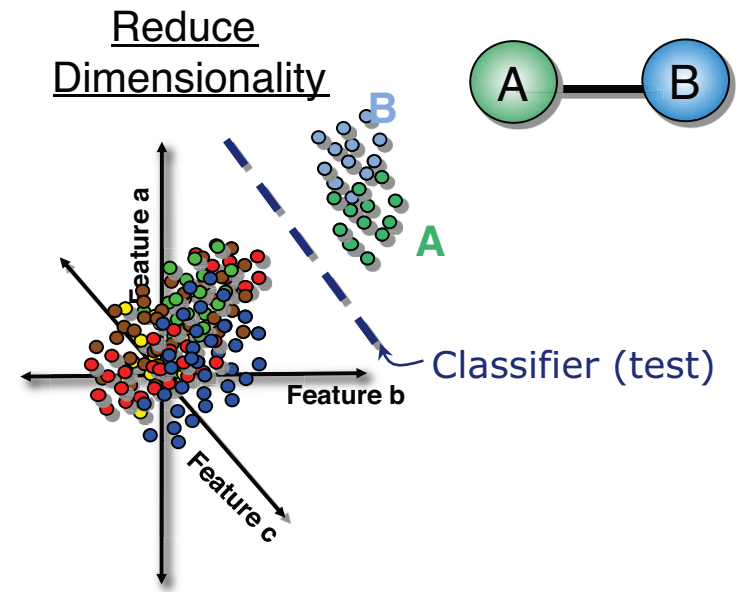
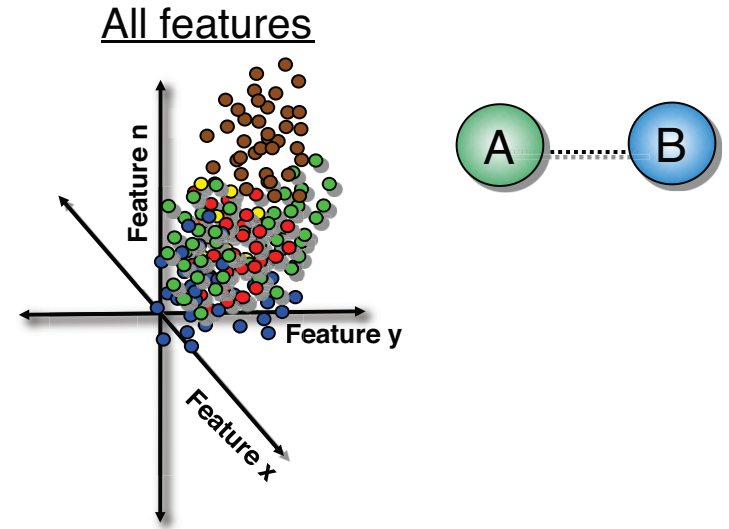
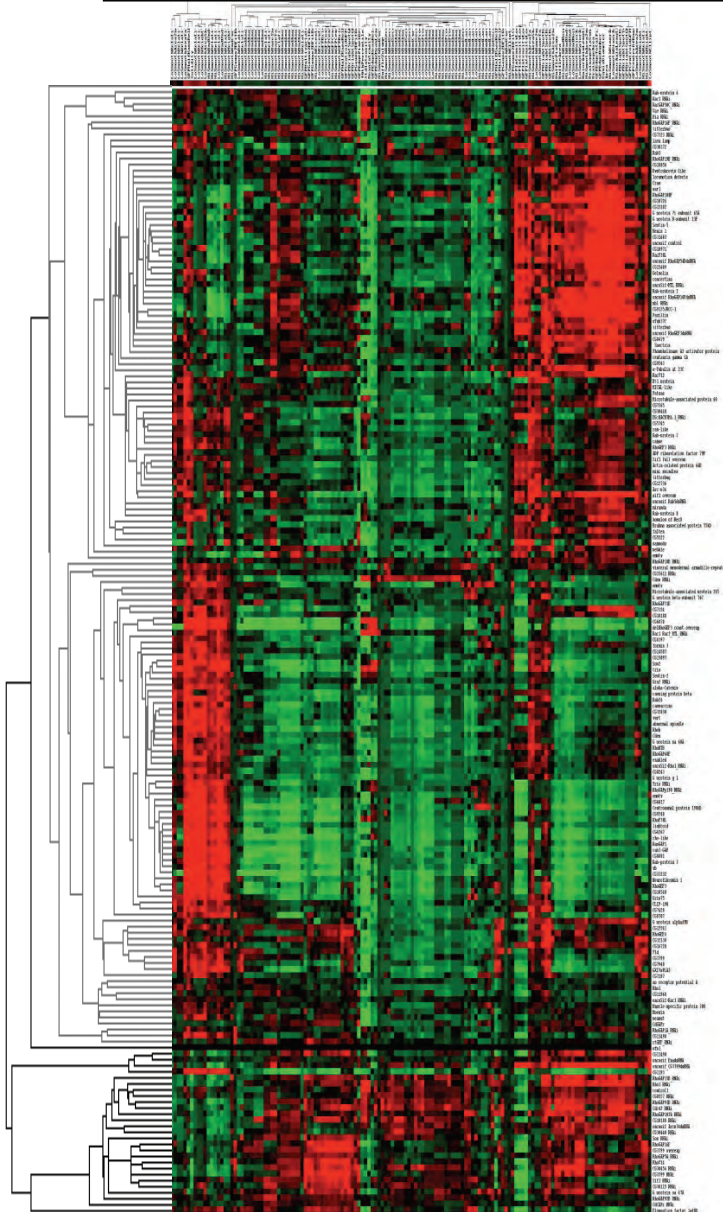


# Data Summary



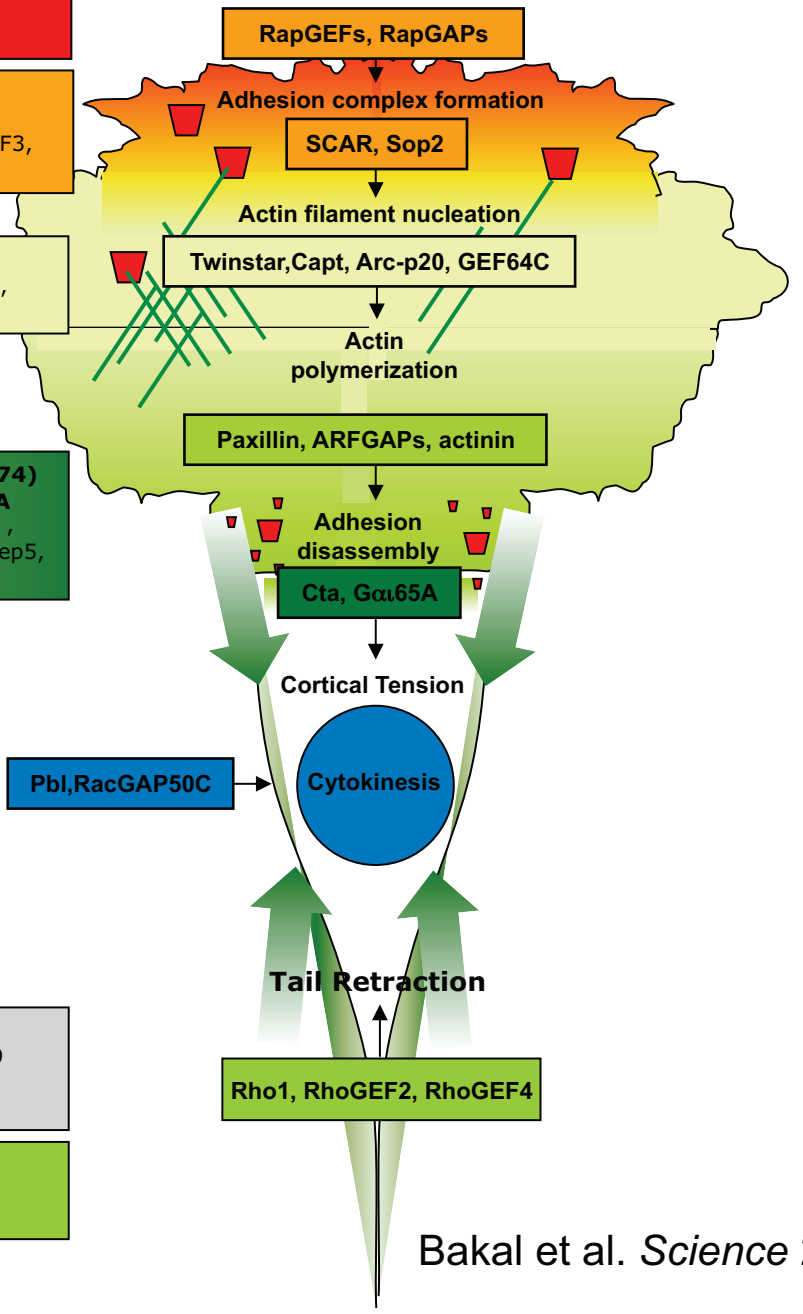
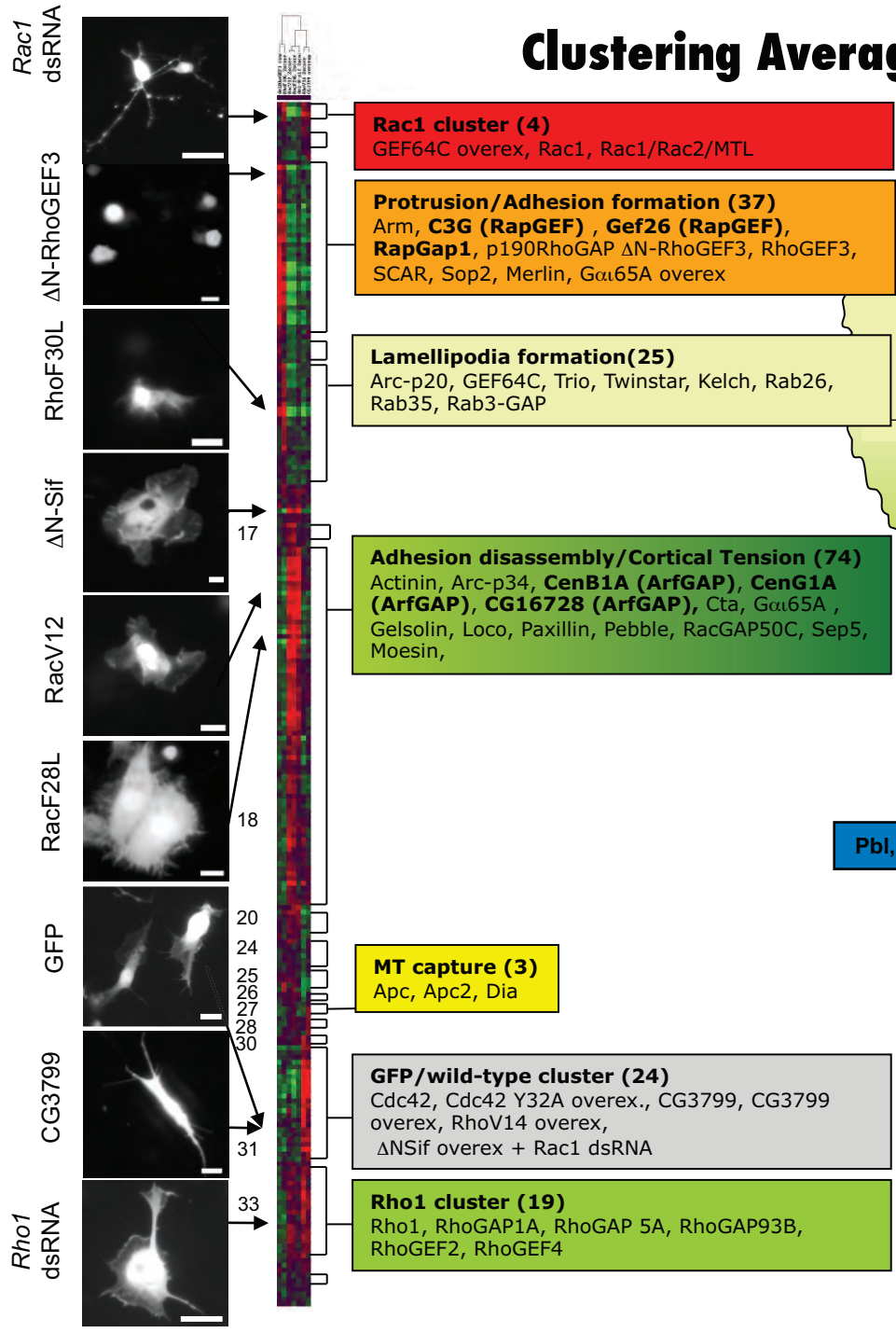
# Raw Morphological Data is Not Interpretable: Dimensionality Reduction

145 phenotypic features



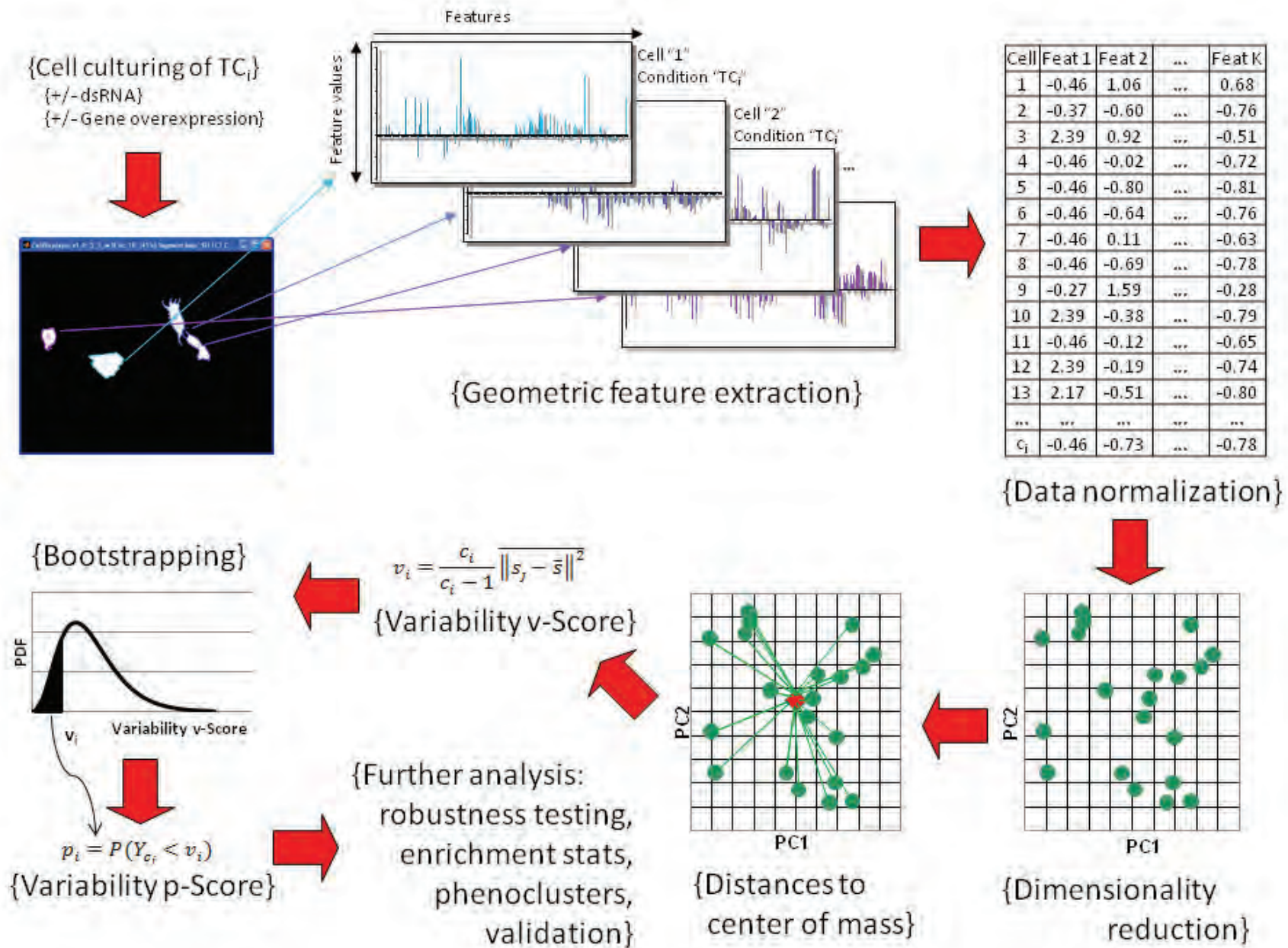


# Clustering Average Morphologies



Bakal et al. *Science* 2007

# Back to Single-Cell Populations: Measuring Morphological Variability



# Computation of PCs

- Principal components calculation
  - Number of dimensions:  $d \sim 100$
  - Number of samples:  $n \sim 2,000,000$
- Covariance matrix method
  - Eigenvalue decomposition of covariance matrix
  - $O(d^3+d^2n)$
- Fixed-point algorithm method
  - Sharma & Paliwal, *Pattern Recognition Letters* 2007
  - $O(d^2hL+d^2n)$

Fast PCA algorithm for computing leading eigenvectors

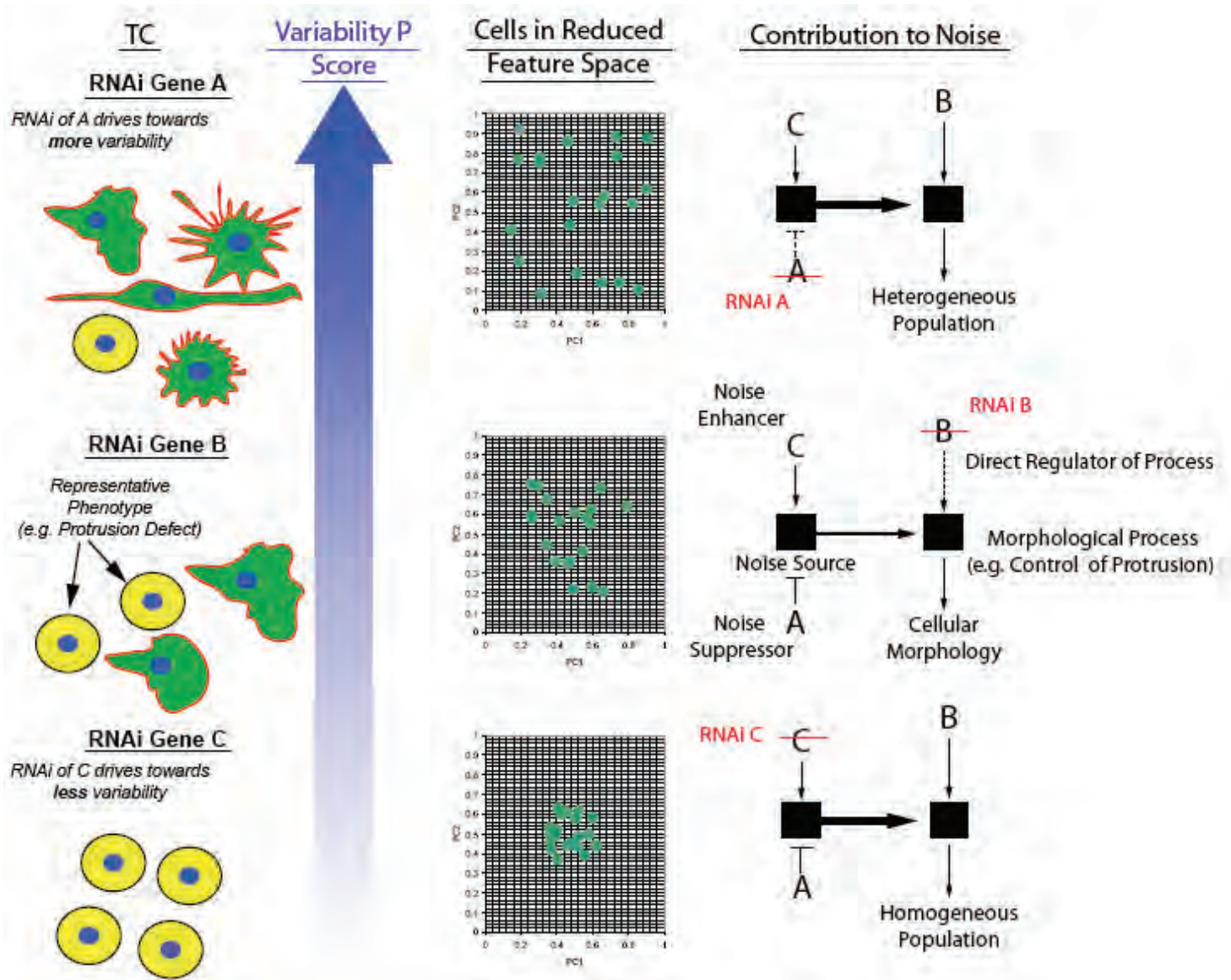
---

1. Choose  $h$ , the number of principal axes or eigenvectors required to estimate. Compute covariance  $\Sigma_x$  and set  $p \leftarrow 1$
2. Initialize eigenvector  $\varphi_p$  of size  $d \times 1$  e.g. randomly
3. Update  $\varphi_p$  as  $\varphi_p \leftarrow \Sigma_x \varphi_p$
4. Do the Gram-Schmidt orthogonalization process

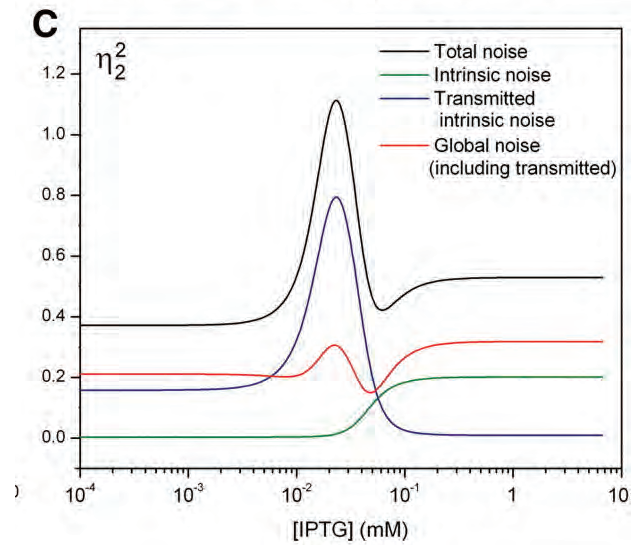
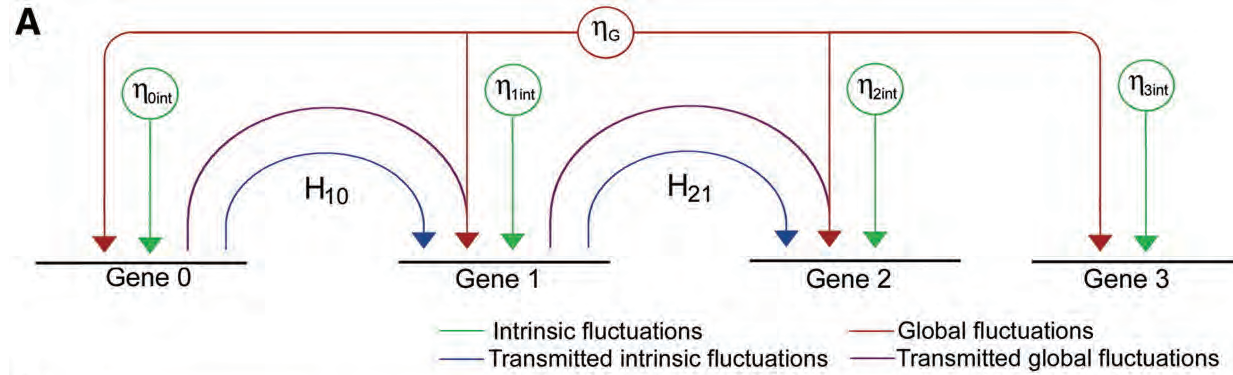
$$\varphi_p \leftarrow \varphi_p - \sum_{j=1}^{p-1} (\varphi_j^T \varphi_p) \varphi_j$$

5. Normalize  $\varphi_p$  by dividing it by its norm:  $\varphi_p \leftarrow \varphi_p / |\varphi_p|$
  6. If  $\varphi_p$  has not converged, go back to step 3
  7. Increment counter  $p \leftarrow p + 1$  and go to step 2 until  $p$  equals  $h$
-

# Genetic Tuning of Morphological Variability

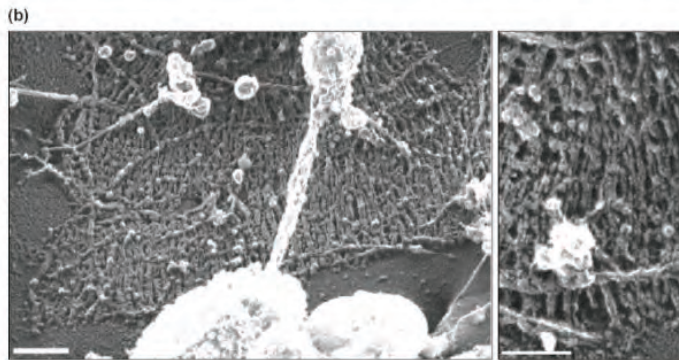
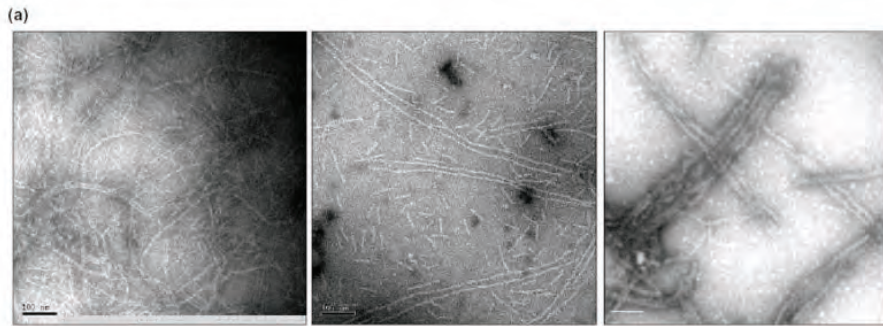


# Noise propagation in networks as measured by transcription – same result for morphology?

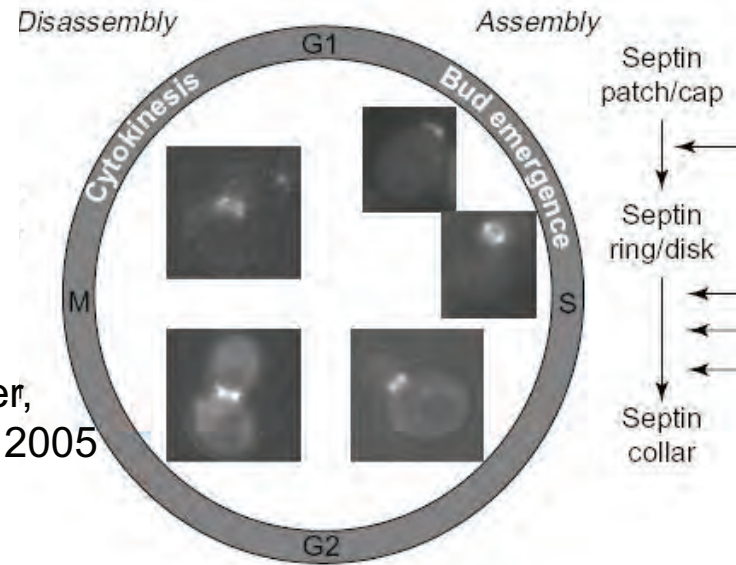
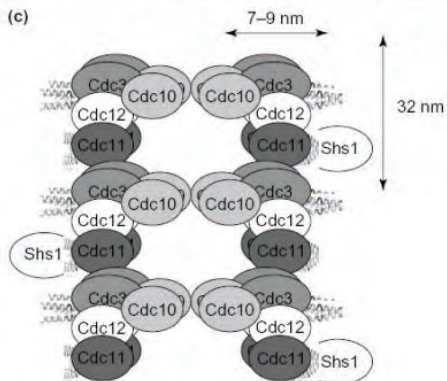


Pedraza & van Oudenaarden, *Science* 2005

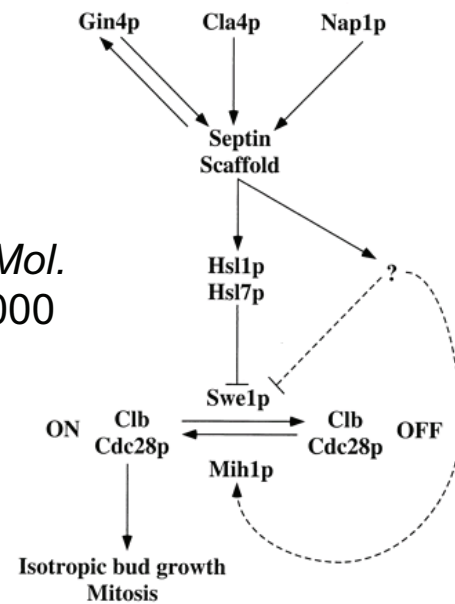
# Studying Variability in a Cellular Process: Septin Ring Formation in Yeast



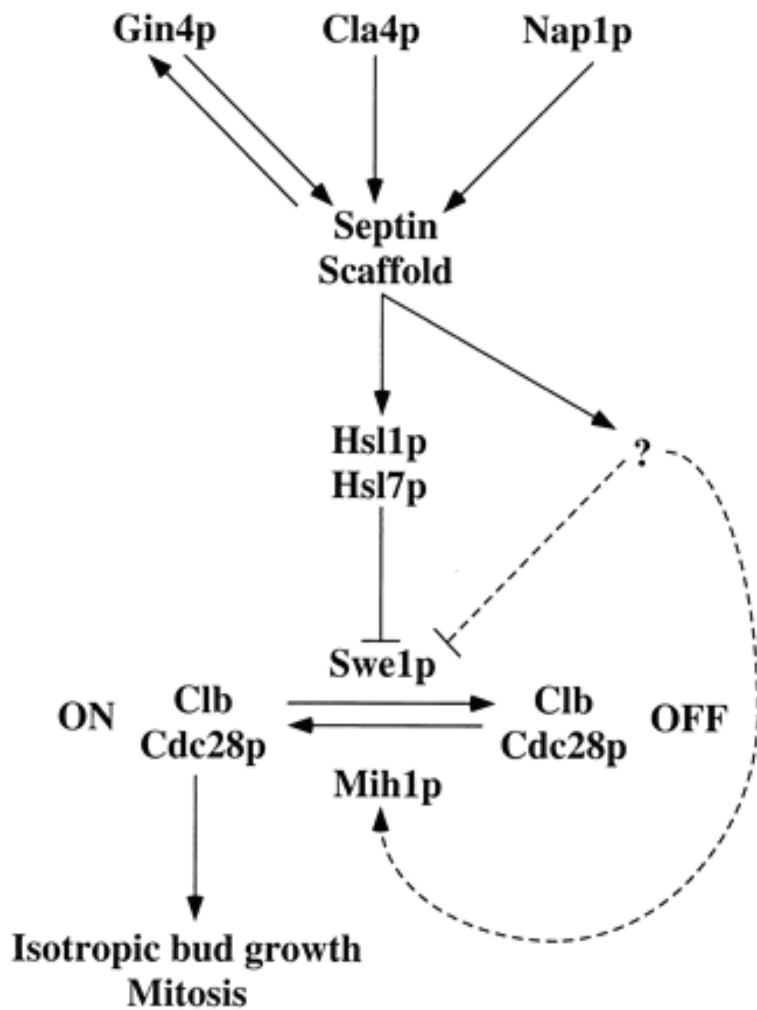
Versele & Thorner, *Trends Cell Bio.* 2005



Longtine et al., *Mol. and Cell. Bio.* 2000



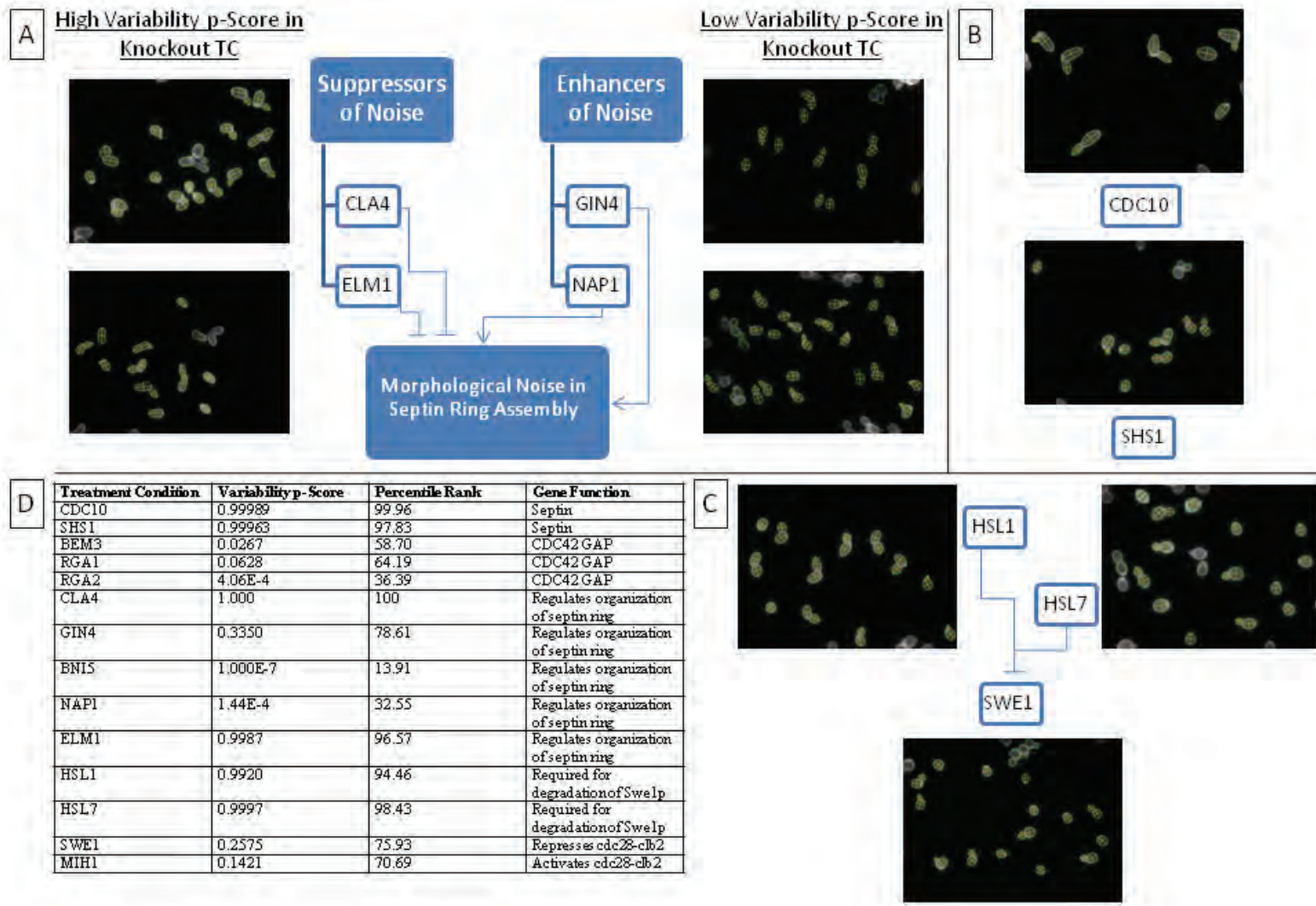
# Studying Variability in a Cellular Process: Septin Ring Formation in Yeast



<u>Hypothesis</u>	<u>Gene Knockout</u>	<u>Variability</u>
1	Cdc42, Cdc24	Lethal/high
2	Septins	Lethal/high
3	Cla4, Gin4, Bni5, Elm1	High
4	Hsl1, Hsl7	High
5	Swe1	Low
6	Cdc42 GAPs	Low
7	Cdc28	Lethal/high

Longtine et al., *Mol. and Cell. Bio.* 2000

# Studying Variability in a Cellular Process: Septin Ring Formation in Yeast





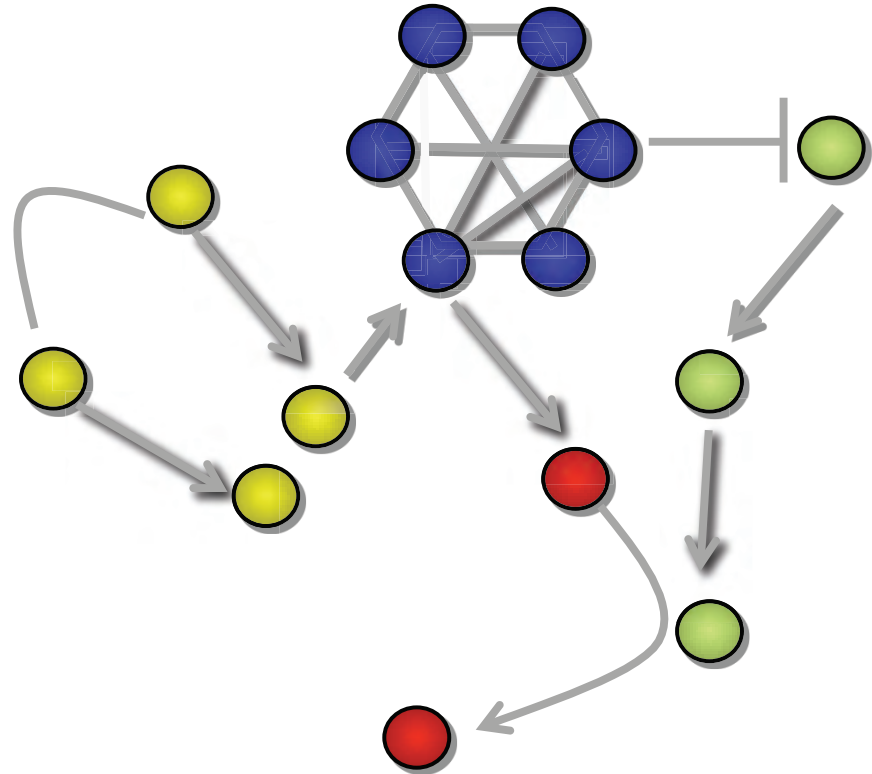
# How Do You Study Morphological Variability in Cellular Processes?

## Outline

1. New kind of data  
*morphology = signaling state*

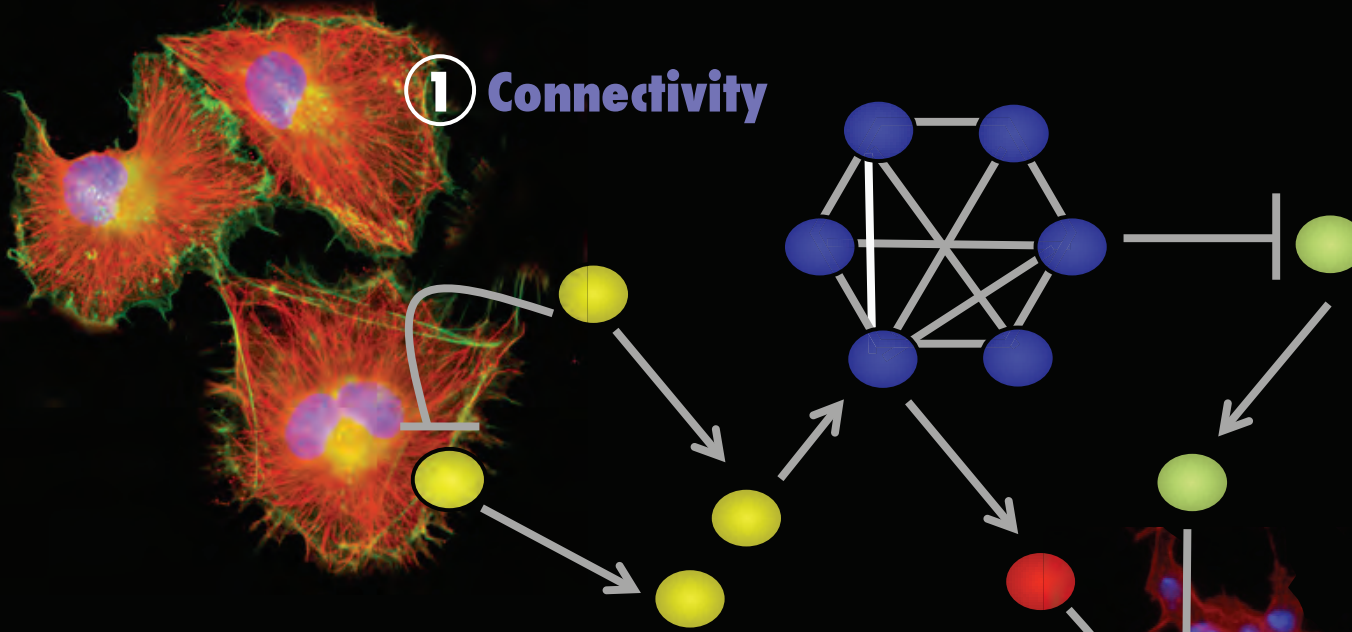
2. How to measure  
morphological variability  
from this data

3. Understanding  
variability in the context  
of signaling network  
architecture

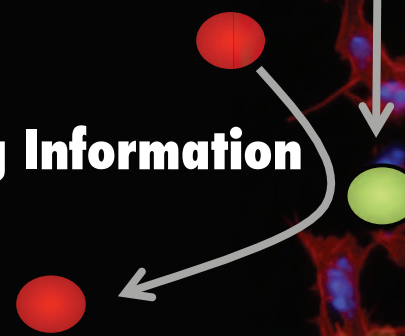


# Challenges in Describing Signaling Networks

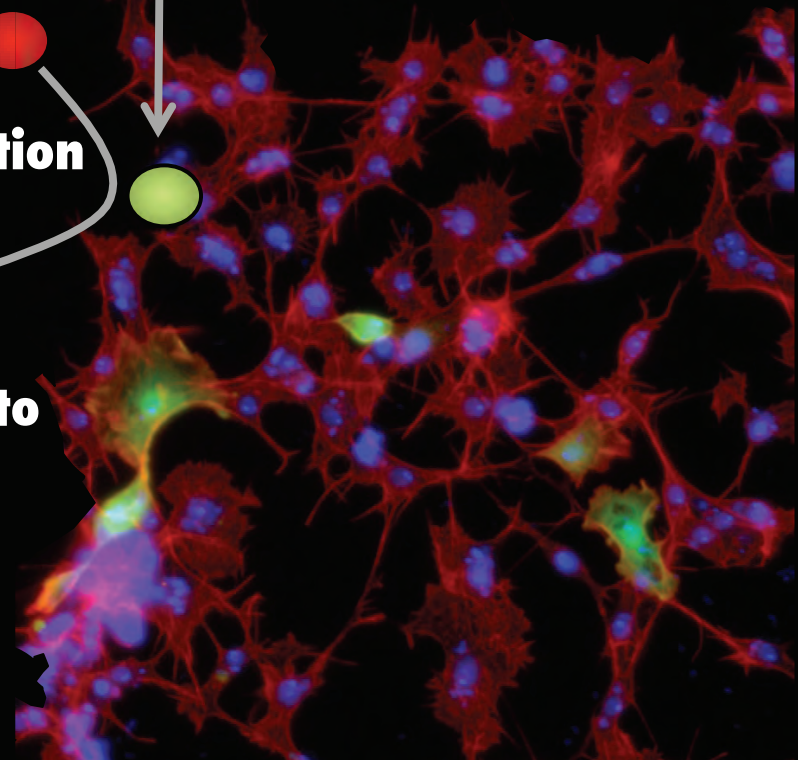
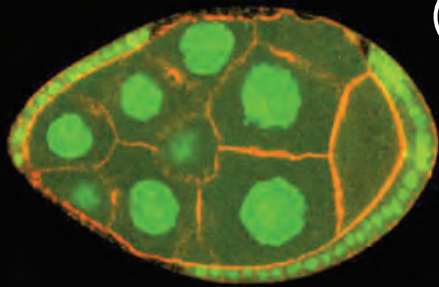
① **Connectivity**



② **Flow of Signaling Information**



③ **Variability of Response to Signaling Information**

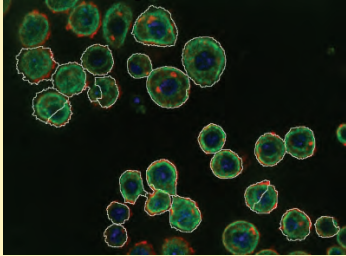


# Future Directions: Expanding to More Genes

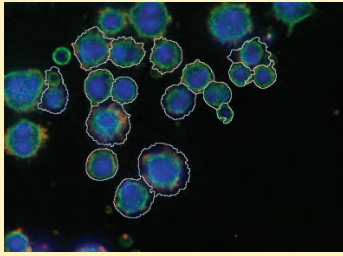
## Local Networks of Kinases and Phosphatases That Regulate Cell Shape

Cytokinesis Defect

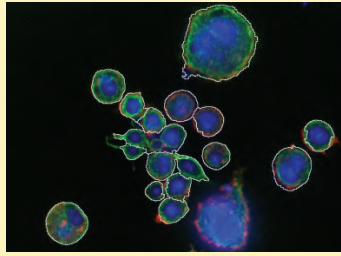
*CycA*



*Sti*

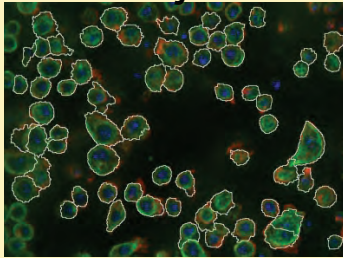


*ial*

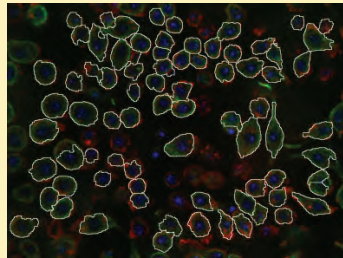


Teardrop Cluster

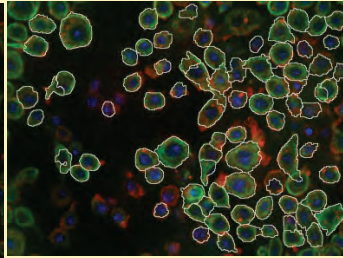
*acj6*



*CG7097*

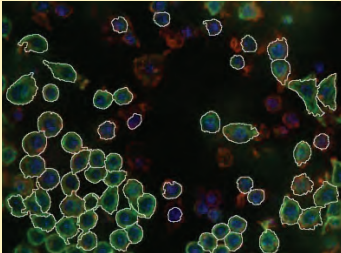


*Pka C1*

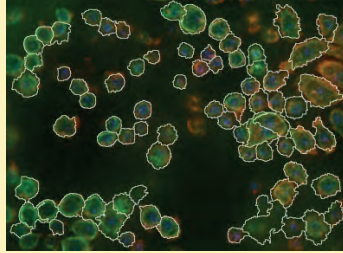


Size Defect

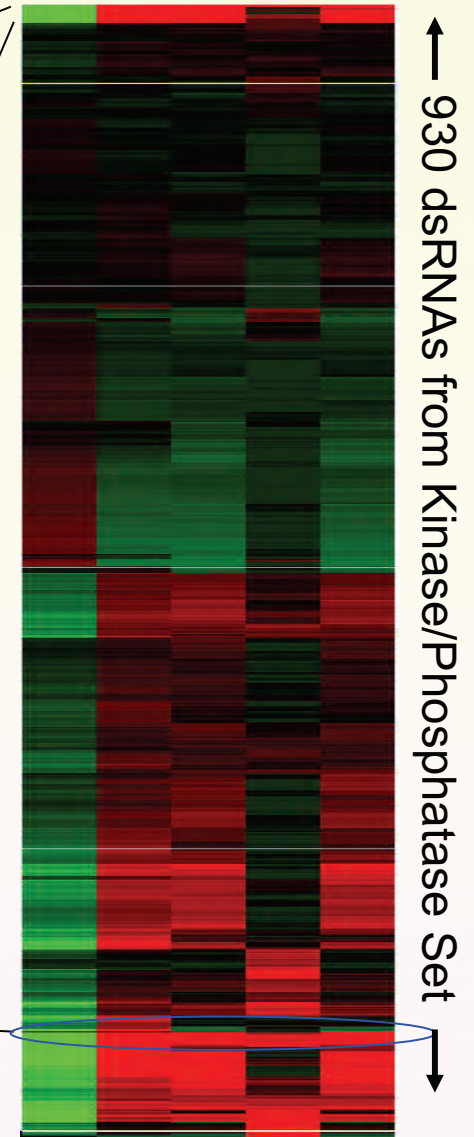
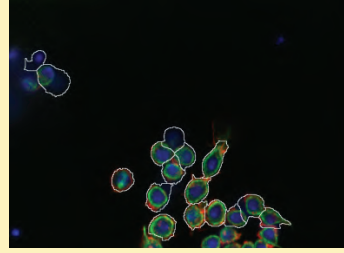
*rok*



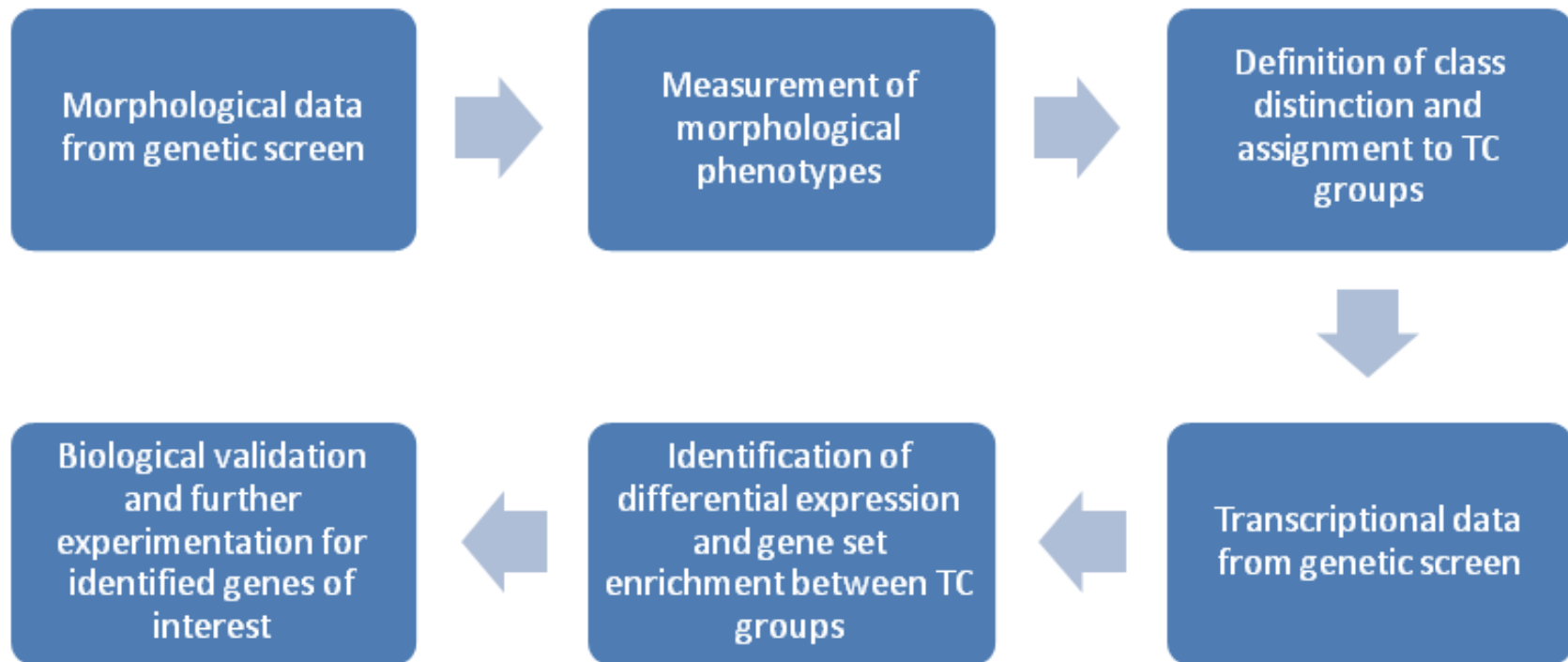
*EDTP*



*CG32666*



# Integration with Other High-throughput Data Sources



## Acknowledgments:



Chris Bakal

Bonnie  
Berger

John Aach

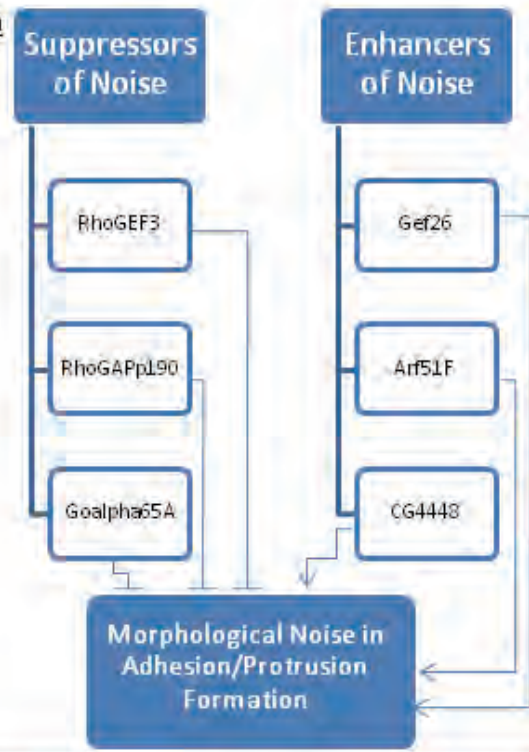
Norbert  
Perrimon

George  
Church



A

High Variability p-Score in Knockout TC



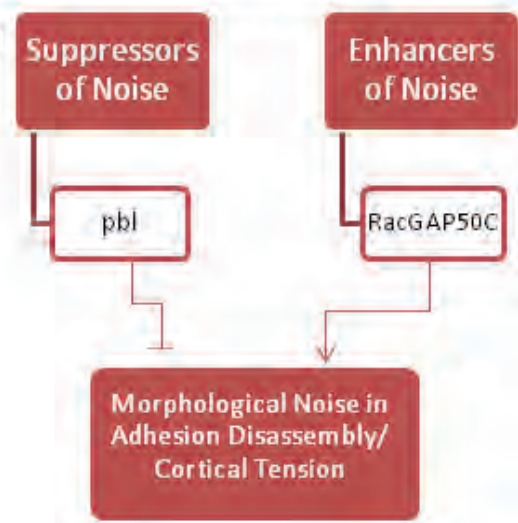
Low Variability p-Score in Knockout TC

B

Treatment Condition	Variability p-Score
Gα65A <sub>overex</sub>	0.0000
Gef26	0.0000
delRhoGEF3 <sub>const<sub>overexp</sub></sub>	0.0000
Arf51F	0.0004
CG4448	0.0010
CG9699	0.0015
C3G	0.0151
CG9426	0.0157
Merlin	0.0179
...	...
CG1583	0.8851
CG7846	0.9261
RhoGEF3	0.9598
RhoGAPp190	0.9780

C

High Variability p-Score in Knockout TC



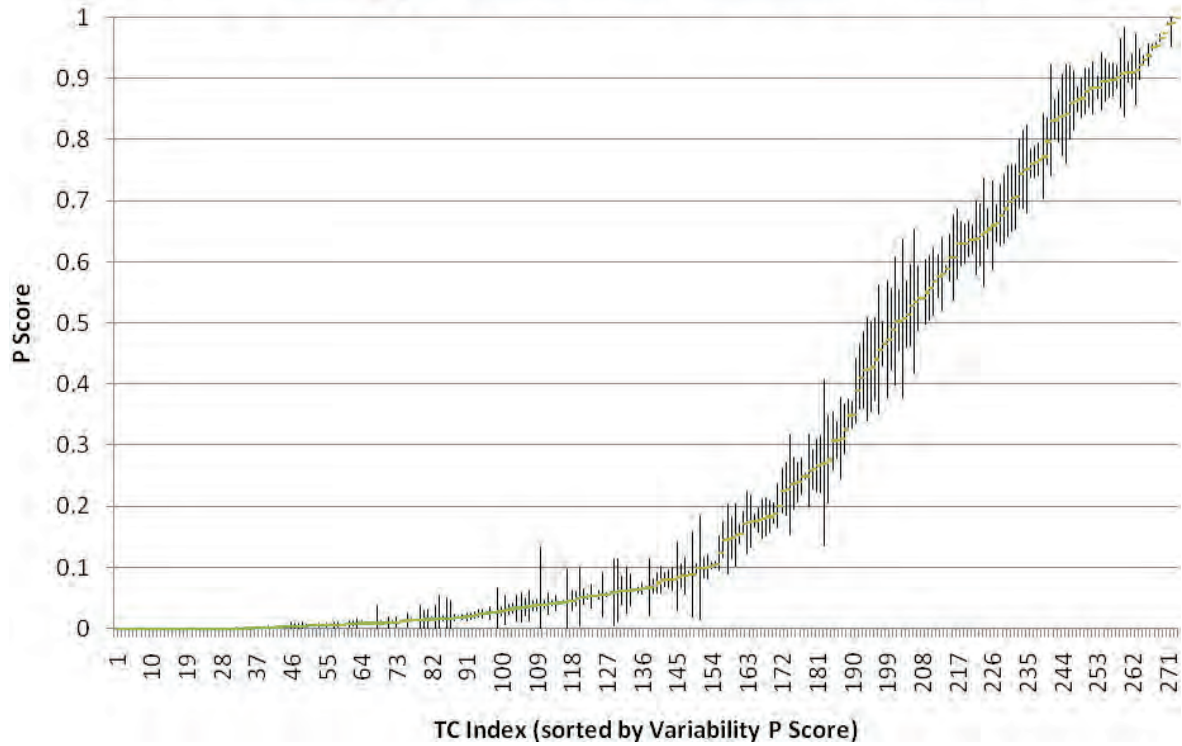
Low Variability p-Score in Knockout TC

D

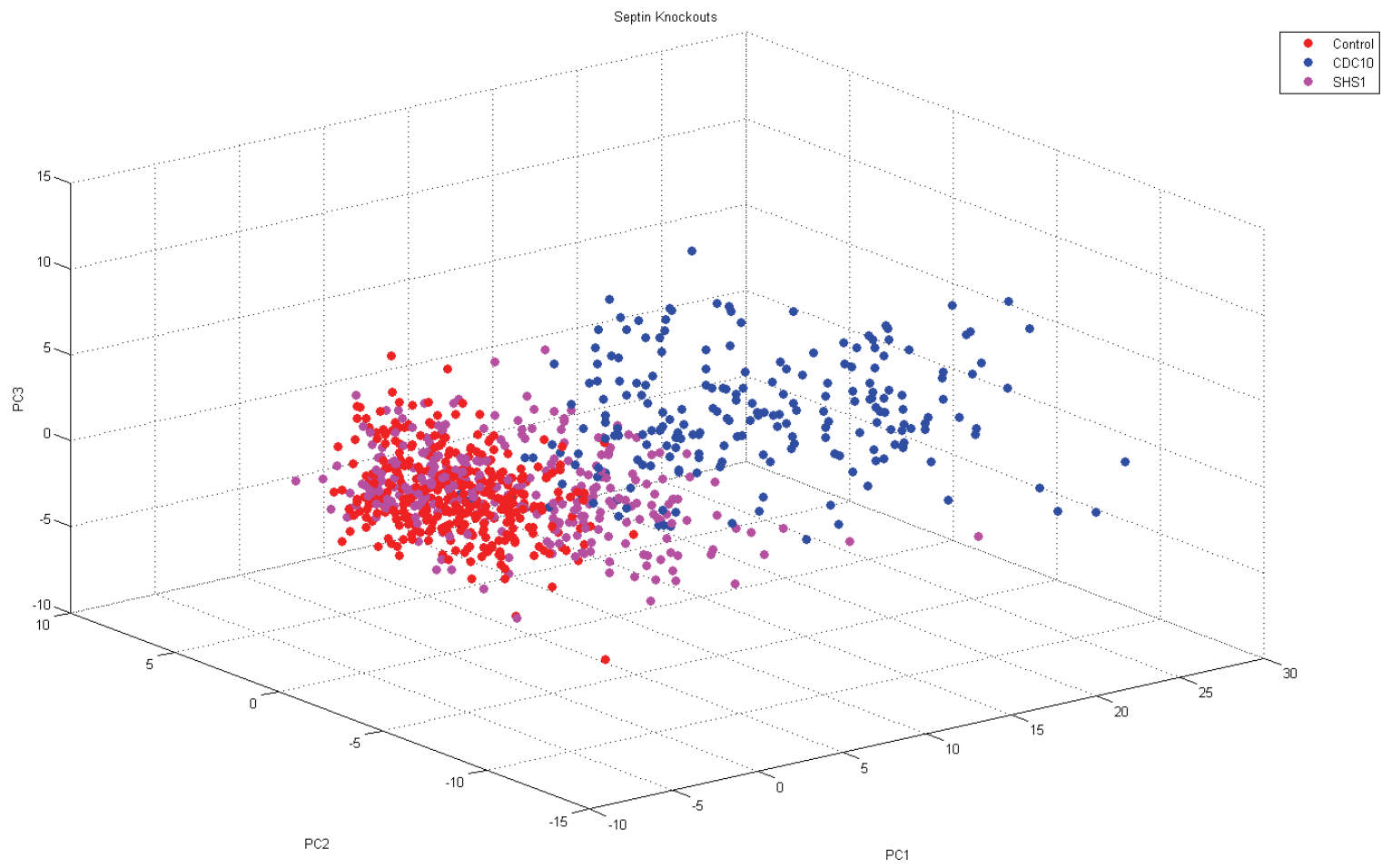
Treatment Condition	Variability p-Score
RacGAP50C	0.0000
oncoSif+MTL <sub>RNAi</sub>	0.0000
CG12102	0.0001
RacV12	0.0001
oncosif <sub>control</sub>	0.0002
RacF28L	0.0003
cenB1A	0.0005
...	...
Septin-5	0.9023
CG30158	0.9759
pbl	1.0000

<b>GO Term</b>	<b>P value for enrichment</b>	<b>Yeast genes</b>
Chromosome organization	2.42e-07	RAD52 CYC8 DEF1 CTF8 BUD32 SPT10 CTF4 NPL6 SPC72 RTT109 RAD54 SCP160 HTL1 EST1 CIK1 RAD50 DCC1
Response to DNA damage stimulus	3.18e-07	RAD27 RAD52 RNR1 DEF1 SAC3 CTF8 SPT10 CTF4 NPL6 RTT109 RAD54 HTL1 HOF1 RAD50 DCC1
Cellular component organization	2.39e-03	BEM1 RPB4 RAD52 CYC8 EDC3 DEF1 SAC3 SHE4 CTF8 BUD32 IRC25 CDC10 SPT10 CTF4 NPL6 SPC72 RTT109 RAD54 SCP160 HTL1 EST1 CHC1 CLA4 CIK1 BEM2 RAD50 DCC1
Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	2.44e-03	CTK2 RPB4 RNR4 RAD27 RAD52 CYC8 RPA12 SWI4 EDC3 ADH1 RNR1 DEF1 SAC3 SQS1 SHE4 BUD32 IRC25 SPT10 CTF4 NPL6 RTT109 RAD54 SCP160 HTL1 EST1 ANP1 RAD50
Anatomical structure homeostasis / telomere maintenance / telomere organization	4.23e-03	RAD52 DEF1 BUD32 RAD54 EST1 RAD50
Organelle organization	5.71e-03	RAD52 CYC8 DEF1 SAC3 SHE4 CTF8 BUD32 SPT10 CTF4 NPL6 SPC72 RTT109 RAD54 SCP160 HTL1 EST1 CLA4 CIK1 BEM2 RAD50 DCC1
Cell cycle	7.79e-03	SAC3 CTF8 CTF4 SPC72 CLA4 CIK1 DCC1
Mitosis / nuclear division / phase of mitotic cell cycle	9.52e-03	SAC3 CTF8 CTF4 SPC72 CLA4 CIK1 DCC1
<b>GO Term</b>	<b>P value for enrichment</b>	<b>Yeast genes</b>
Mitochondrial translation	3.32e-03	ISM1 MSW1 MRPS12 MRPL23 NAM2 RSM25 RSM18 MRPL11 IFM1 RTC6 RSM19 MRPL32 MRPS8 MRPL10 MRPL16 MRPL7 MRP10 SWS2 MRPL13 RSM27 MSE1 GRS1 MSR1 PET112

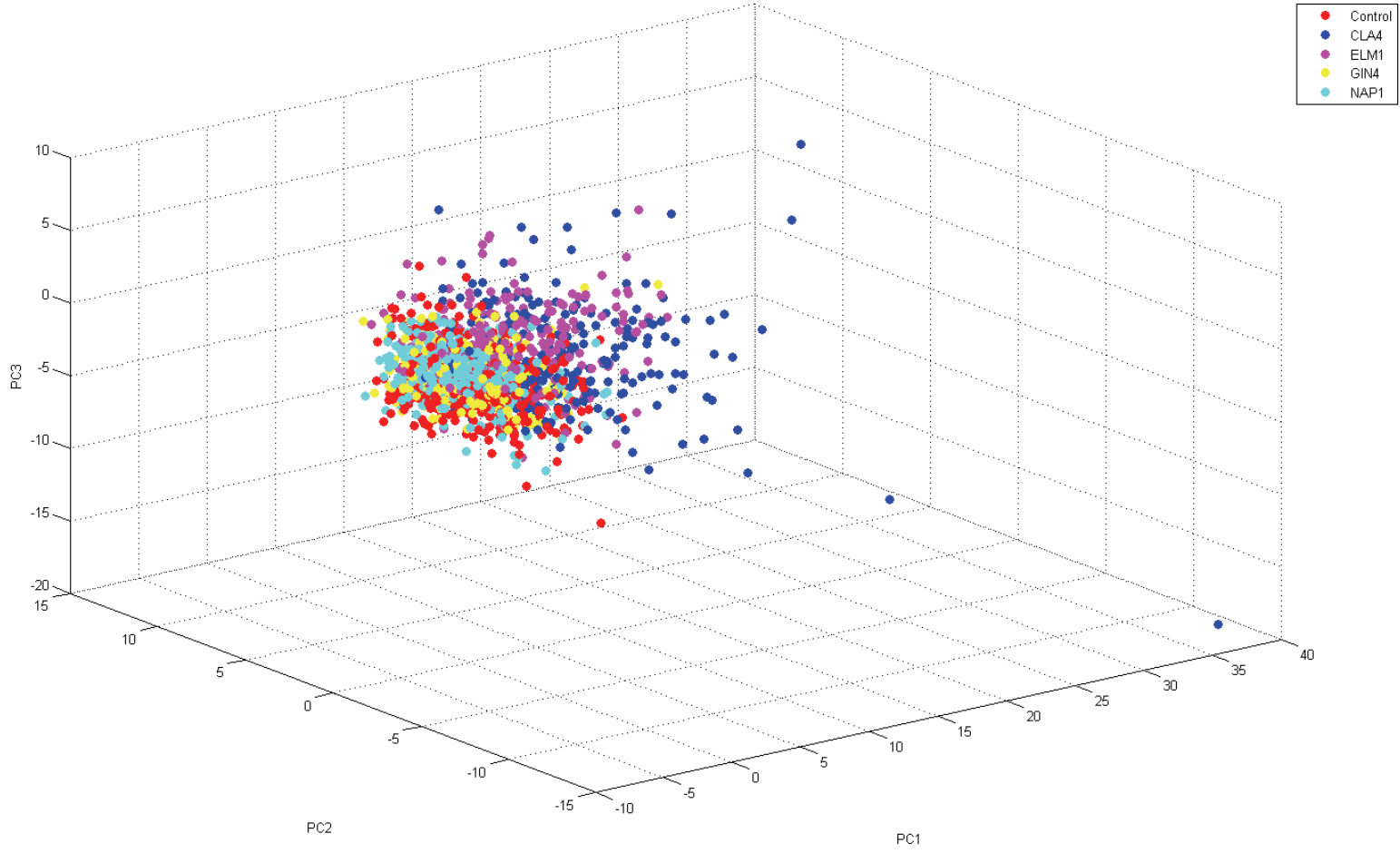
### Variability p-Scores with Standard Errors

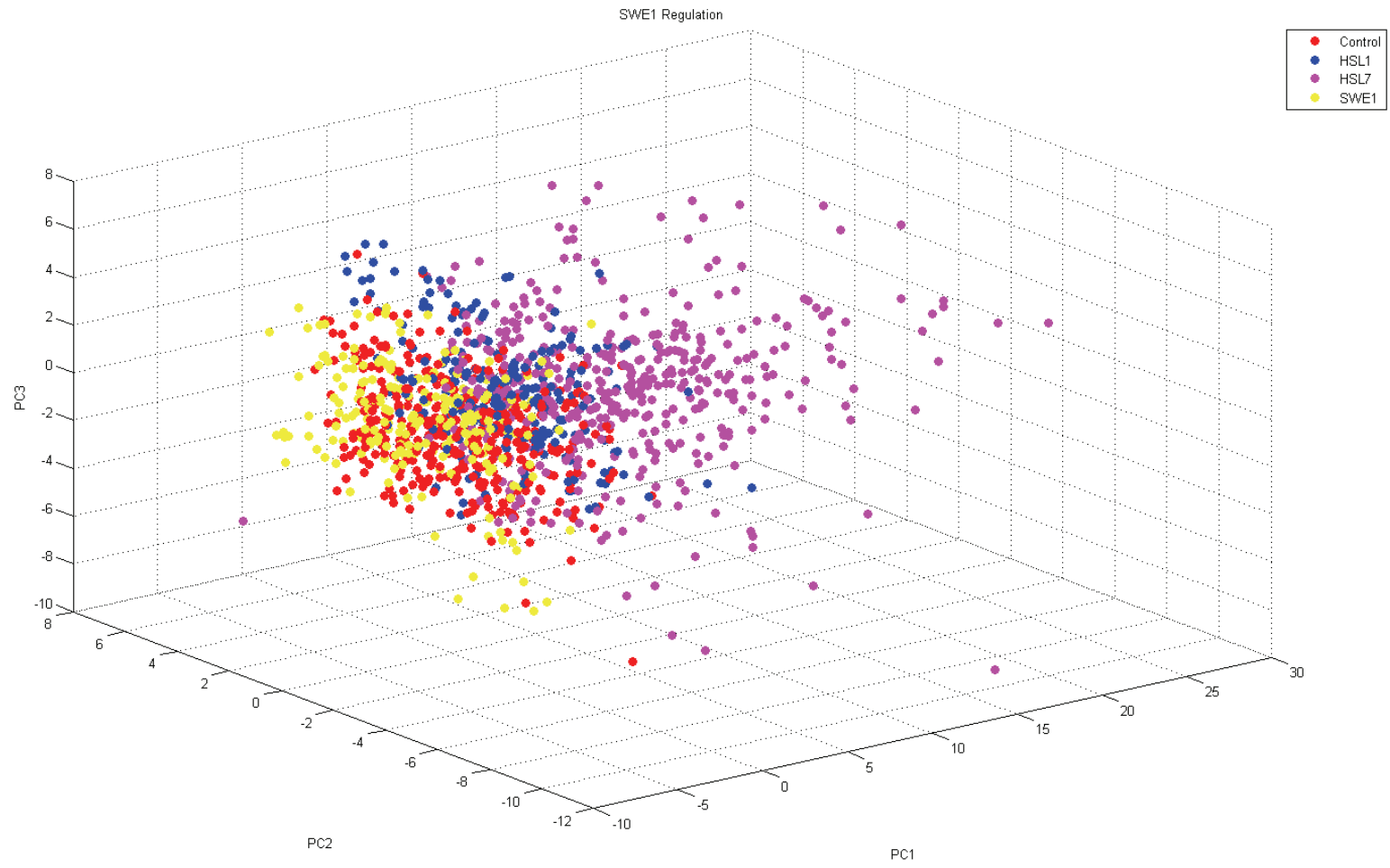






Regulation of Septin Assembly





<b>Method</b>	<b>1 PC</b>	<b>2 PCs</b>	<b>3 PCs</b>	<b>5 PCs</b>	<b>10 PCs</b>	<b>NNs</b>
<b>1 PC</b>	-	3.15E-12	6.74E-10	6.74E-10	3.15E-12	1.75E-04
<b>2 PCs</b>	3.15E-12	-	3.15E-12	6.74E-10	6.74E-10	1.75E-04
<b>3 PCs</b>	6.74E-10	3.15E-12	-	3.15E-12	3.15E-12	1.75E-04
<b>5 PCs</b>	6.74E-10	6.74E-10	3.15E-12	-	2.50E-13	1.75E-04
<b>10 PCs</b>	3.15E-12	6.74E-10	3.15E-12	2.50E-13	-	1.75E-04
<b>NNs</b>	1.75E-04	1.75E-04	1.75E-04	1.75E-04	1.75E-04	-