Colorado State University

Inferring Predictive Signal-Activated Gene Regulation Models from Noisy Single-Cell Data

Brian Munsky

Chemical and Biological Engineering
School of Biomedical Engineering
Colorado State University

May 19, 2015

munsky@engr.colostate.edu

Presented at the SIAM Conference on Applications of Dynamical Systems

Acknowledgements

Colorado State University

Science AAAS

REVIEW

Using Gene Expression Noise to Understand Gene Regulation

Brian Munsky,[1]* Gregor Neuert,[2]* Alexander van Oudenaarden[2,3]

Phenotypic variation is ubiquitous in biology and is often traceable to underlying genetic and environmental variation. However, even genetically identical cells in identical environments display variable phenotypes. Stochastic gene expression, or gene expression "noise," has been suggested as a major source of this variability, and its physiological consequences have been topics of intense research for the last decade. Several recent studies have measured variability in protein and messenger RNA levels, and they have discovered strong connections between noise and gene regulation mechanisms. When integrated with discrete stochastic models, measurements of cell-to-cell variability provide a sensitive "fingerprint" with which to explore fundamental questions of gene regulation. In this review, we highlight several studies that used gene expression variability to develop a quantitative understanding of the mechanisms and dynamics of gene regulation.

www.sciencemag.org SCIENCE VOL 336 13 APRIL 2012

Science AAAS

Systematic Identification of Signal-Activated Stochastic Gene Regulation

Gregor Neuert,[1,2]* Brian Munsky,[3]* Rui Zhen Tan,[1,3,4] Leonid Teytelman,[1] Mustafa Khammash,[4,7]† Alexander van Oudenaarden[1,2,8]†‡

Although much has been done to elucidate the biochemistry of signal transduction and gene regulatory pathways, it remains difficult to understand or predict quantitative responses. We integrate single-cell experiments with stochastic analyses, to identify predictive models of transcriptional dynamics for the osmotic stress response pathway in *Saccharomyces cerevisiae*. We generate models with varying complexity and use parameter estimation and cross-validation analyses to select the most predictive model. This model yields insight into several dynamical features, including multistep regulation and switchlike activation for several osmosensitive genes. Furthermore, the model correctly predicts the transcriptional dynamics of cells in response to different environmental and genetic perturbations. Because our approach is general, it should facilitate a predictive understanding for signal-activated transcription of other genes in other pathways or organisms.

1 FEBRUARY 2013 VOL 339 SCIENCE www.sciencemag.org

Gregor Neuert, Vanderbilt

Alexander van Oudenaarden, Hubrecht

Acknowledgements

Colorado State University

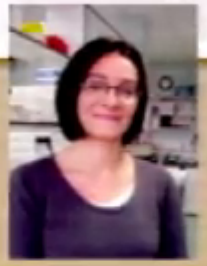
Cell Reports
Report

A. Senecal, et al., Transcription factors modulate c-Fos transcriptional bursts, *Cell Reports*, **8**:1, 75-83, 2014.

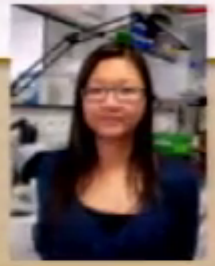
Transcription Factors Modulate c-Fos Transcriptional Bursts

Adrien Senecal,[1,2,3] Brian Munsky,[4] Florence Proux,[1] Nathalie Ly,[1] Floriane E. Braye,[1] Christophe Zimmer,[5] Florian Mueller,[1,5,*] and Xavier Darzacq[1,2,6,*]

Adrien Senecal



Florence Proux



Nathalie Lao

Floriane Brayé

Christophe Zimmer

Florian Müller

Xavier Darzacq

Outline

Colorado State University

1. **Introduction - Information from transcript fluctuation**
2. Measuring and modeling single-cell and single-molecule responses
3. Case studies:
   i. Kinase-activated gene transcription in budding yeast.
   ii. Kinase-activated gene transcription in human cells.
4. Concluding remarks

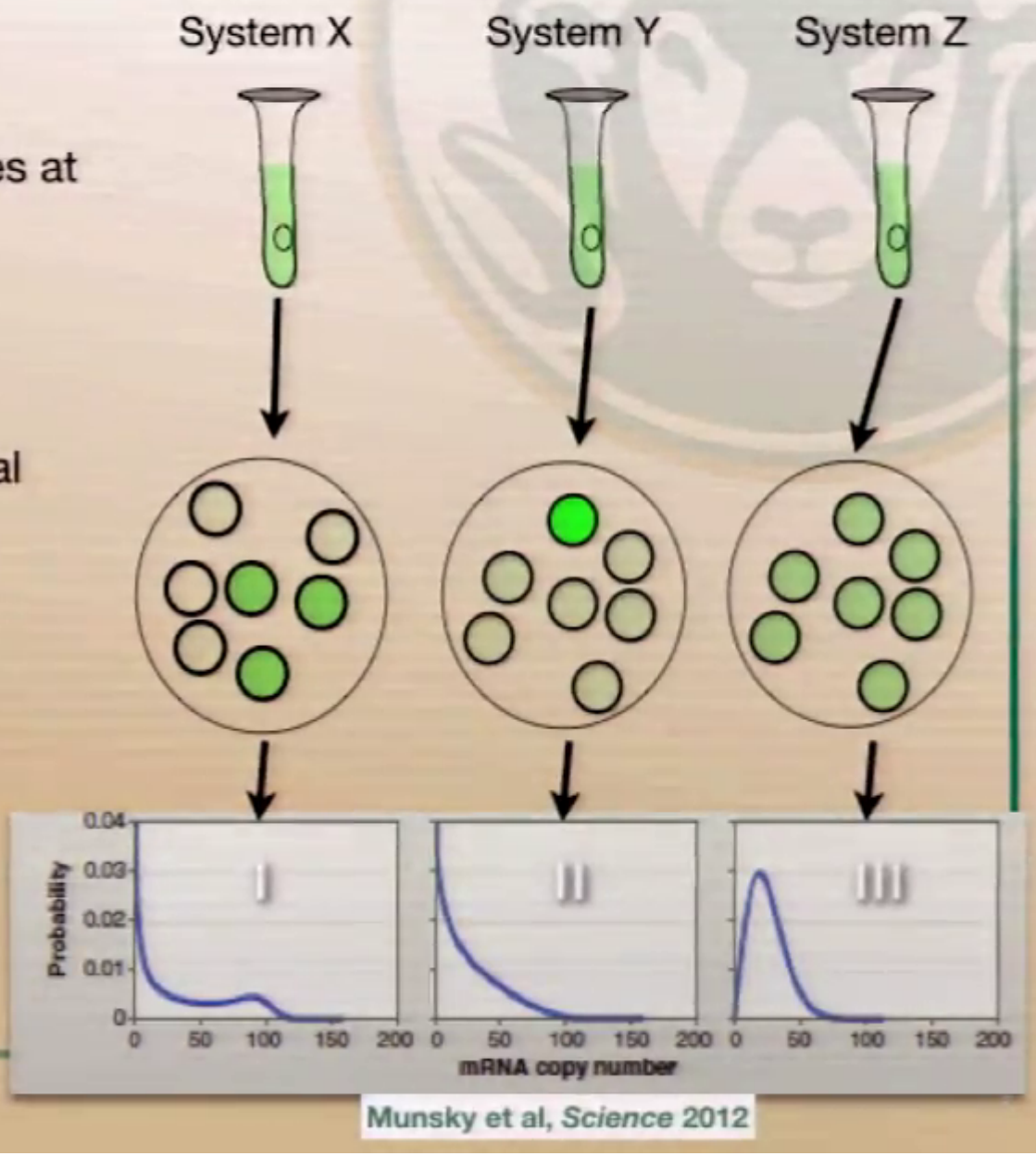
Information in fluctuation

Colorado State University

Different systems (species, inputs, mechanisms, …) may express genes at equal *average* levels.

Single-cell measurements may reveal hidden response differences.

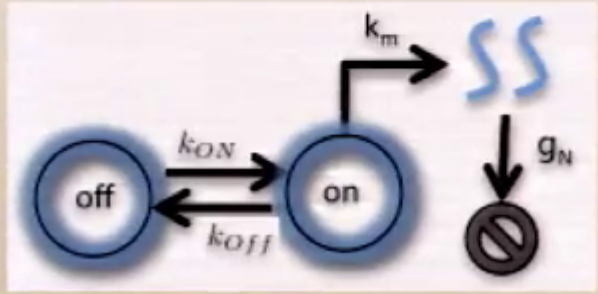
Collective responses can exhibit distinctive "fluctuation fingerprints".



Munsky et al, *Science* 2012

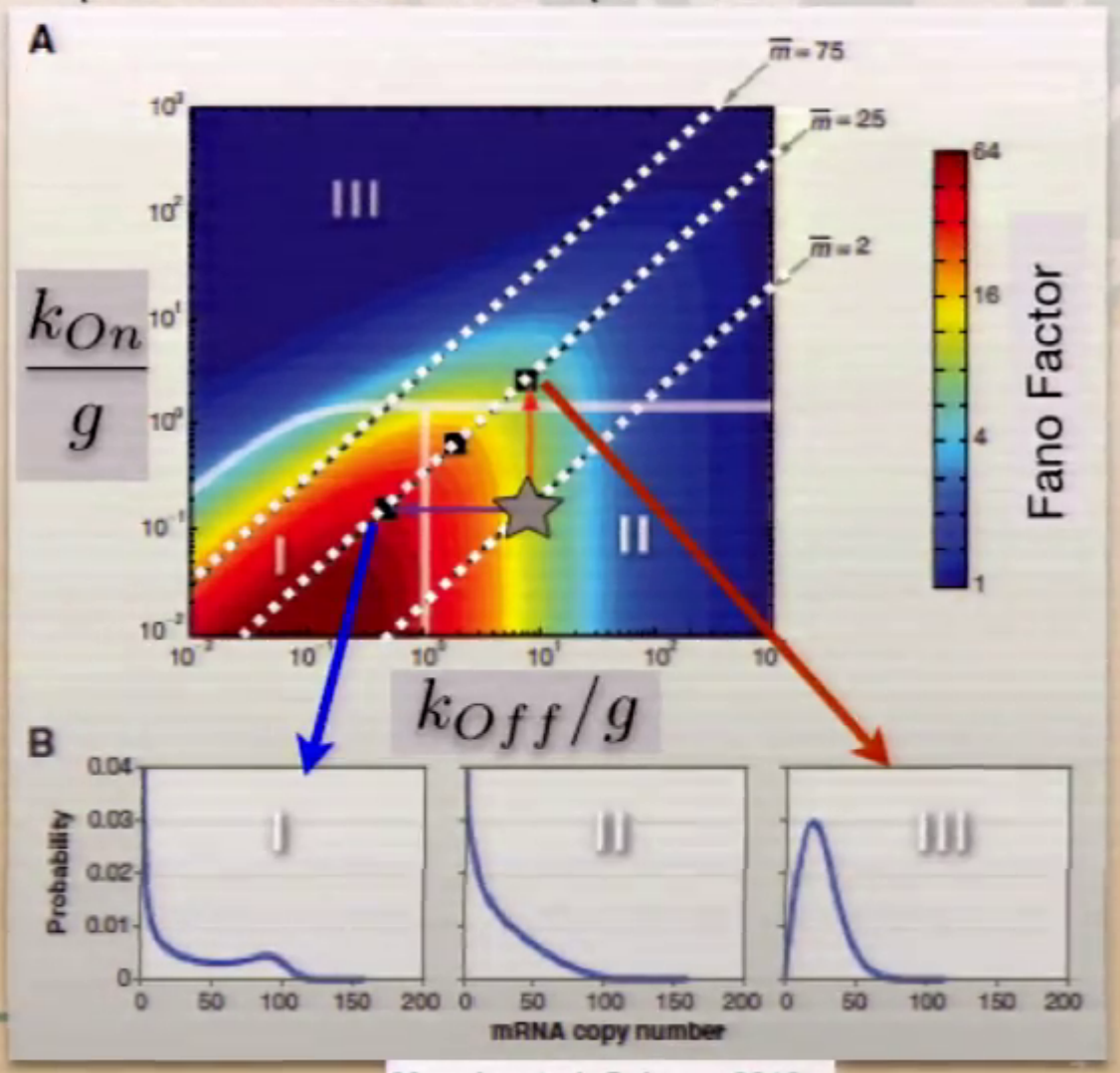
Fluctuations may indicate gene regulation mechanisms

- Consider the bursting gene expression model:



- Compute the expression mean and variability as functions of all parameters.
- Tuning k_{Off} or k_{On} can increase expression, but:
- **Tuning k_{Off} increases variability.**
- **Tuning k_{On} decreases variability.**

Expression 'Noise' versus parameters



Munsky et al, *Science* 2012

Outline

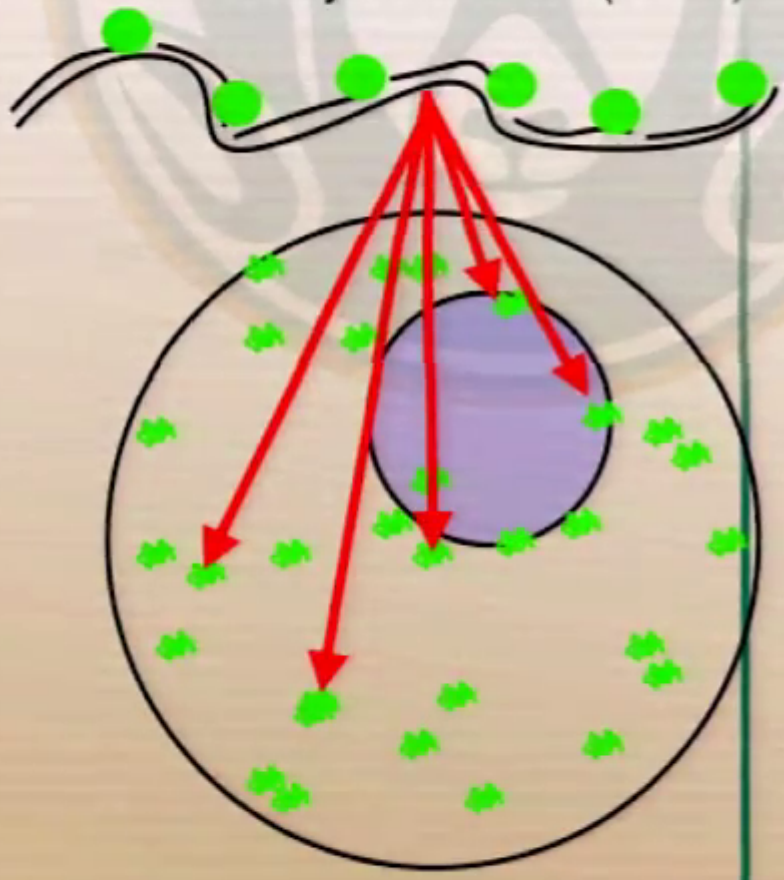
Colorado State University

1. Introduction - Information from transcript fluctuation
2. **MEASURING and modeling single-cell and single-molecule responses**

Single-Molecule FISH (smFISH)

Colorado State University

- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.



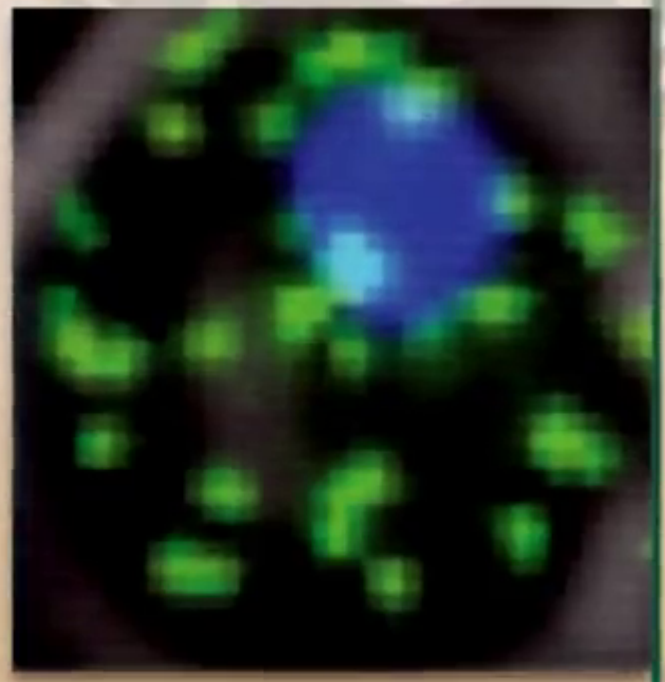
(Neuert, Munsky, *et al*, 2013)

Single-Molecule FISH (smFISH)

Colorado State University

- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.

48 (20bp) probes/mRNA,
Tetramethylrhodamine (TMR)

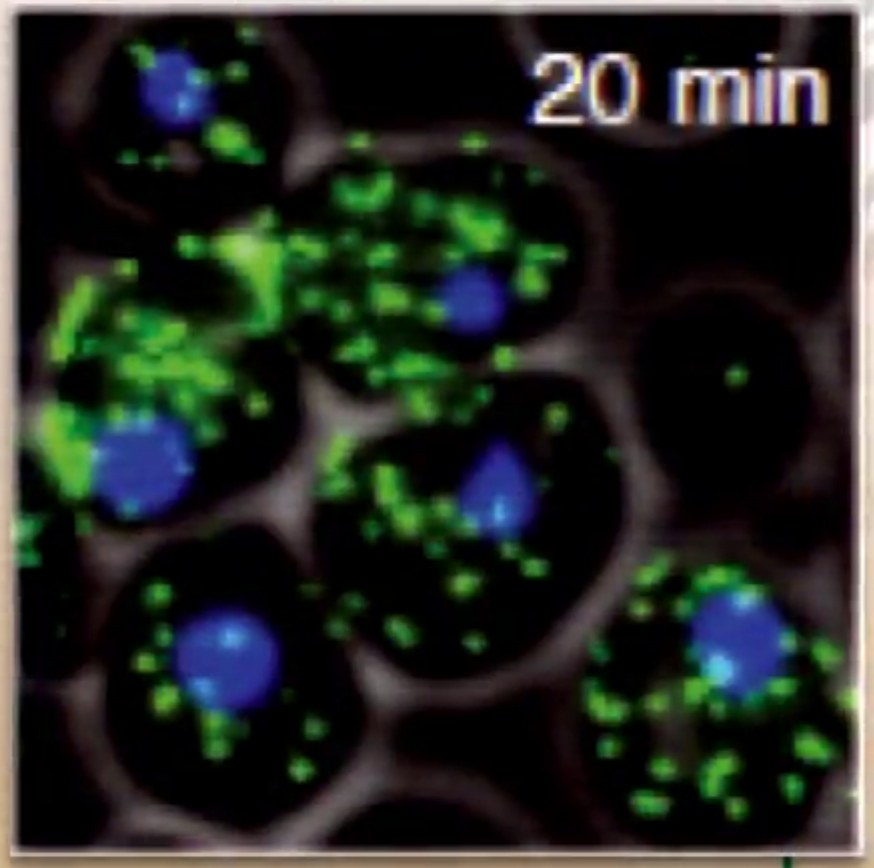


(Neuert, Munsky, *et al*, 2013)

Single-Molecule FISH (smFISH)

Colorado State University

- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.
- Spatial localization enable inter- and intra-nuclear detection.



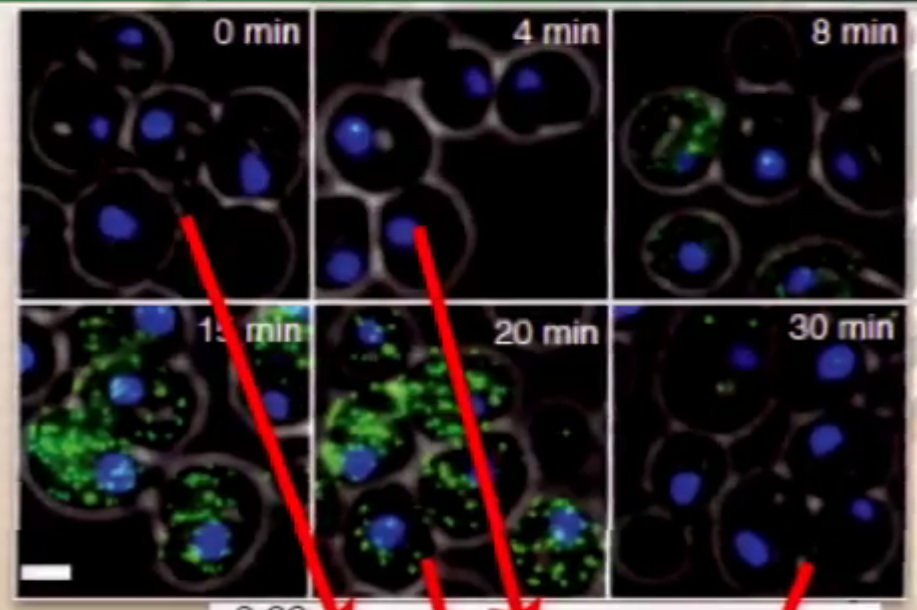
(Neuert, Munsky, *et al*, 2013)

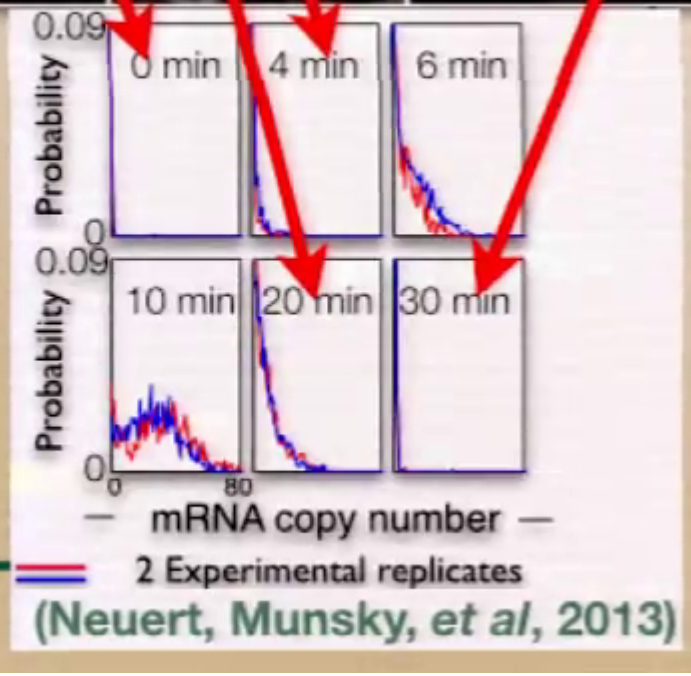
Single-Molecule FISH (smFISH)

Colorado State University

- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.
- Spatial localization enable inter- and intra-nuclear detection.
- Fast time resolution (1-2 min).

Statistics are repeatable and therefore predictable!

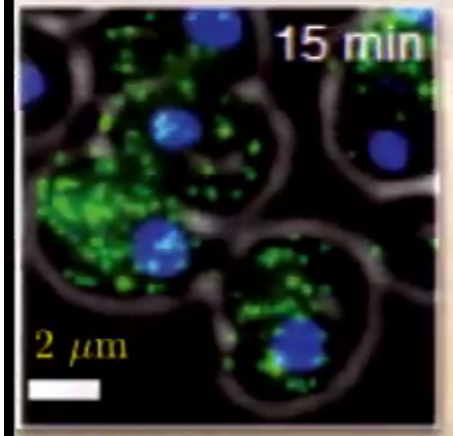




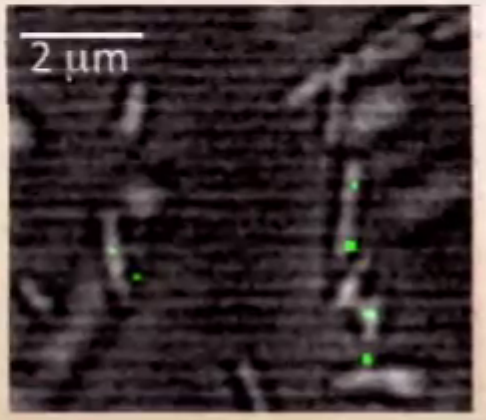
(Neuert, Munsky, *et al*, 2013)

Single-Molecule FISH (smFISH)

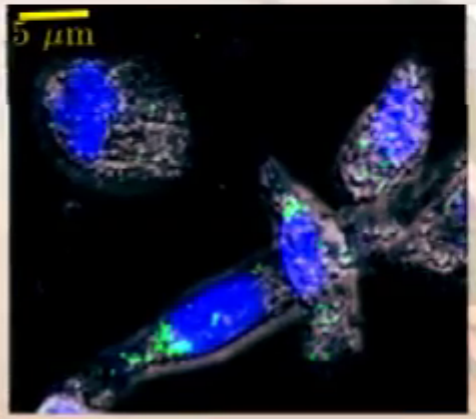
Colorado State University



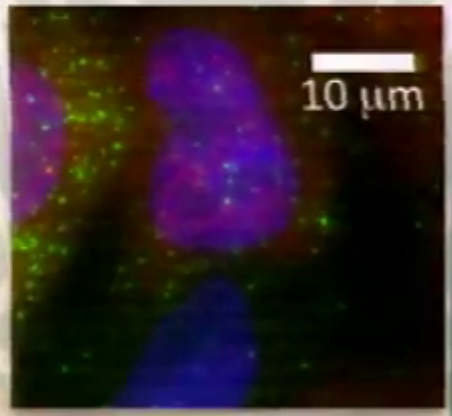
STI1 mRNA in *Saccharomyces cerevisiae* (budding yeast)
-G. Neuert (VU)



Ysr35 sRNA in *Yersinia Pseudotuberculosis* (339nt)
-D. Shepherd (CU Denver)



Traf6 mRNA in THP1 cells
-D. Shepherd (CU Denver)



c-Fos mRNA (green) and p-p38 kinase (red) in U2OS cells
-A. Senecal (CNRS)

smFISH has been applied to many different RNA in many different organisms

Outline

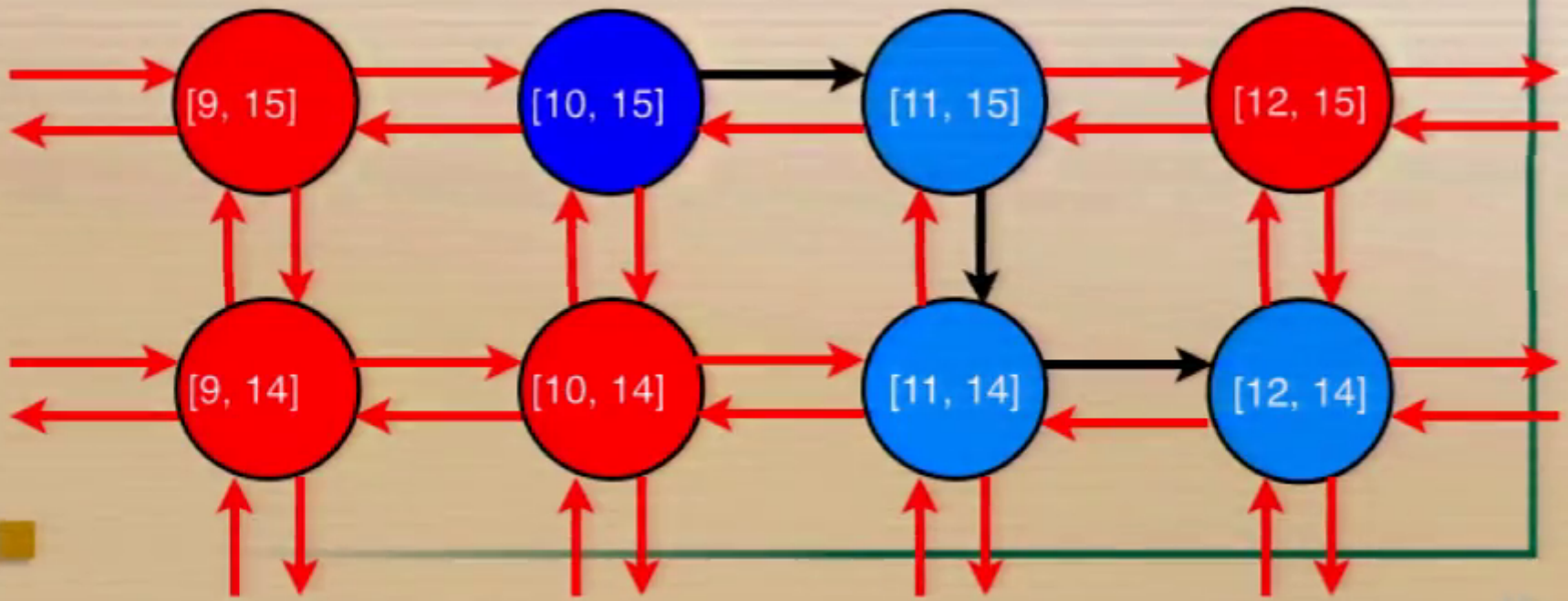
Colorado State University

1. Introduction - Information from transcript fluctuation
2. **Measuring and MODELING single-cell and single-molecule responses**

A Markov description of single-cell gene regulation

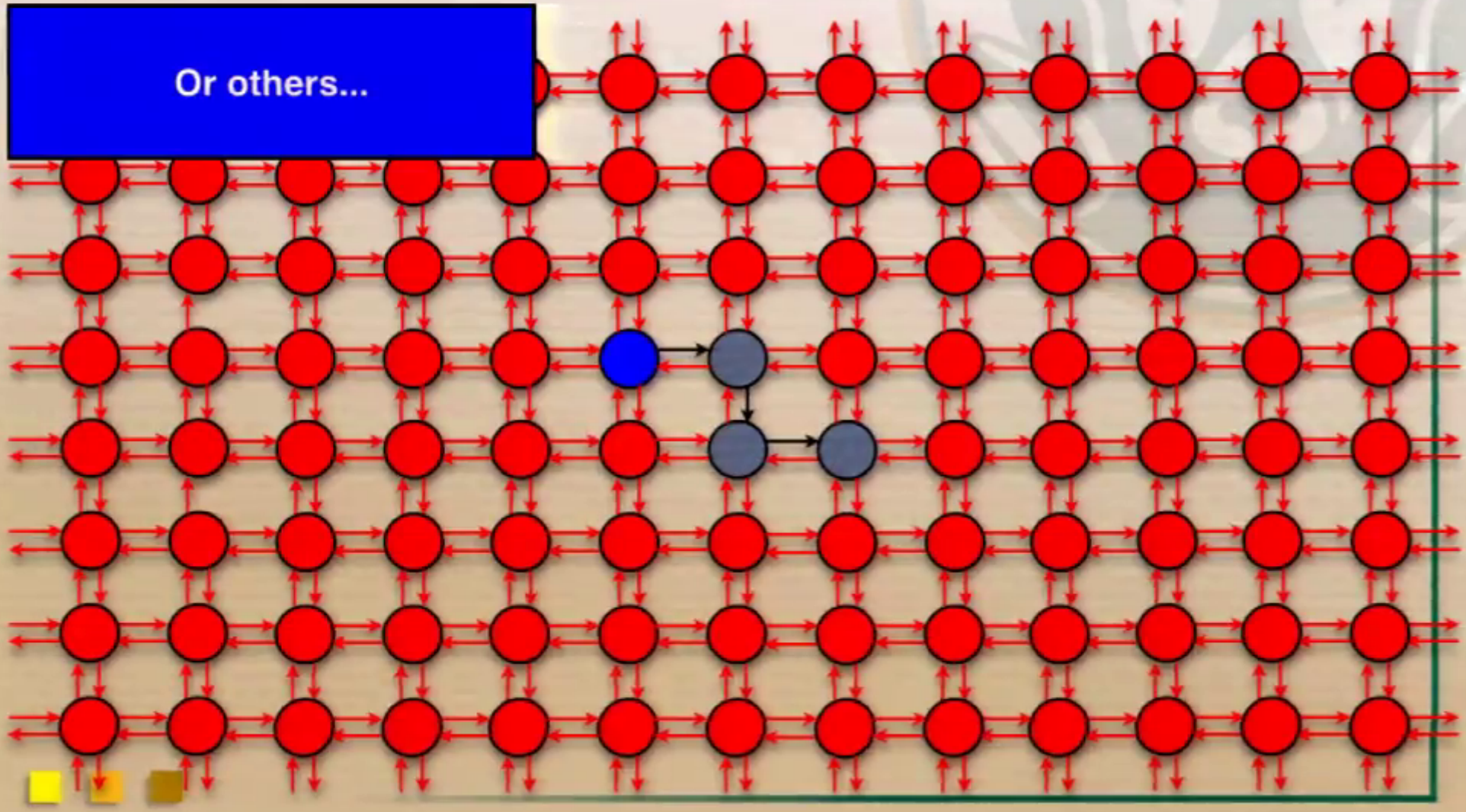
Colorado State University

- At any time, the state of the system is defined by its integer population vector: $\mathbf{x} \in \mathbb{Z}^N$
- Reactions are transitions from one state to another.
- These reactions are random, others could have occurred:



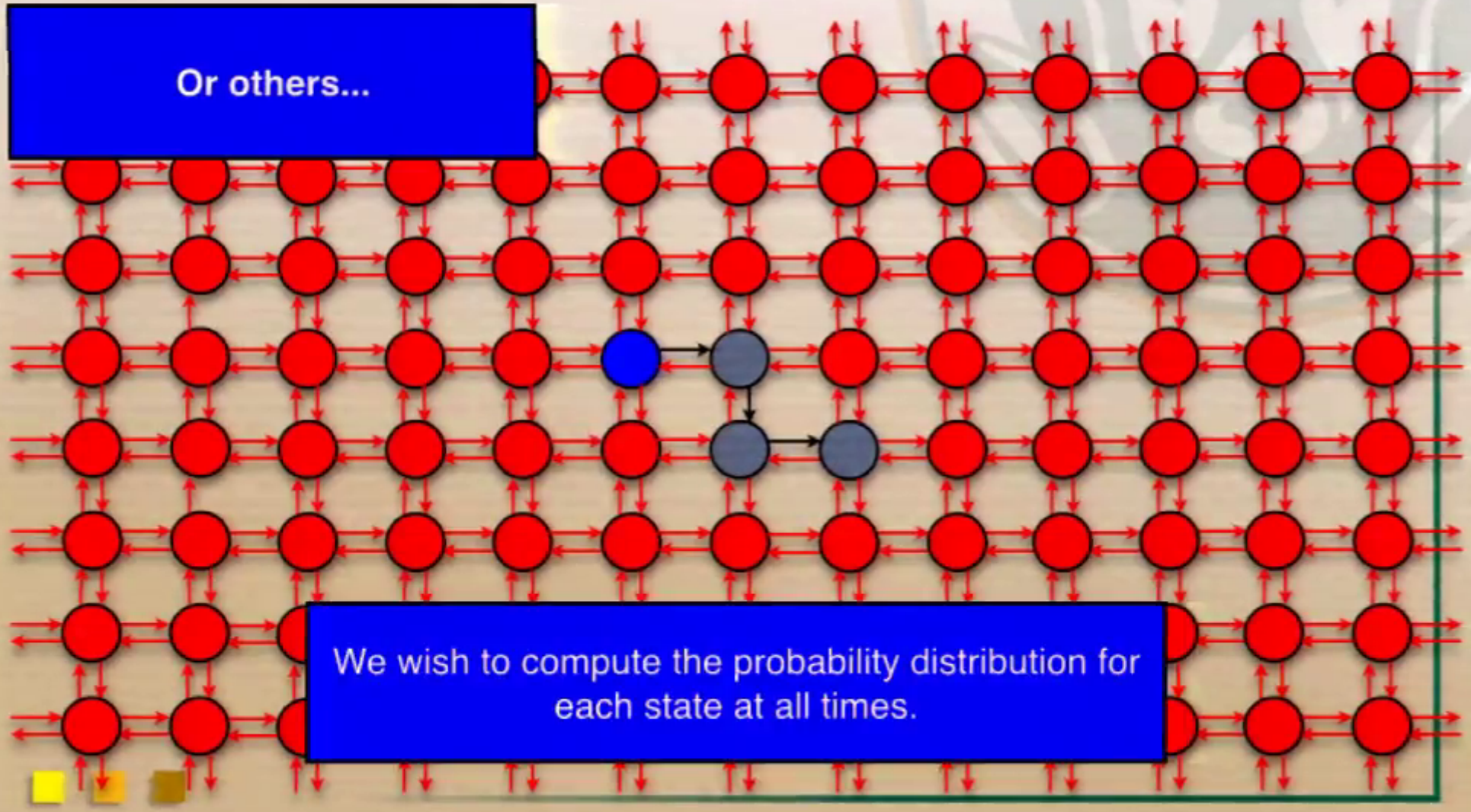
A Markov description of single-cell gene regulation

Colorado State University



A Markov description of single-cell gene regulation

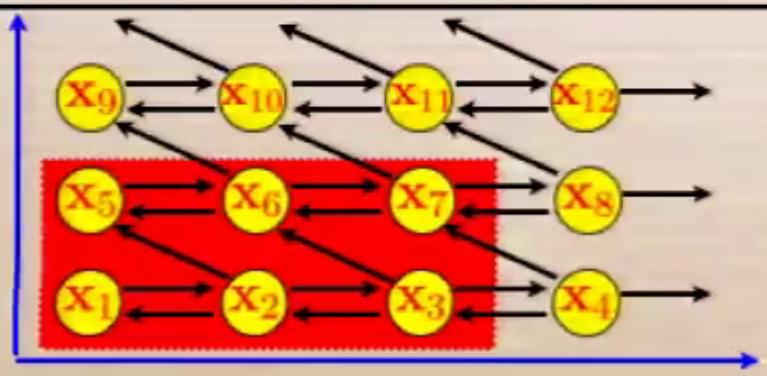
Colorado State University



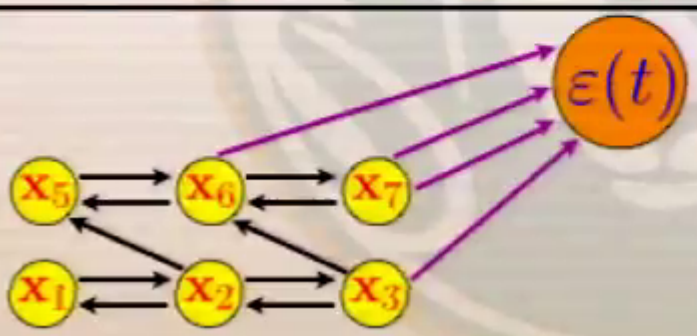
The finite state projection approach

Colorado State University

The Full System



The Projected System (FSP)



Full Master Equation

$$\begin{bmatrix} \dot{\mathbf{P}}_J \\ \dot{\mathbf{P}}_{J'} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_J & \mathbf{A}_{JJ'} \\ \mathbf{A}_{J'J} & \mathbf{A}_{J'} \end{bmatrix} \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix}$$

Dimension = $\#(J) + \#(J')$ = Infinite

FSP Master Equation

$$\begin{bmatrix} \dot{\mathbf{P}}_J^{FSP} \\ \dot{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_J & \mathbf{0} \\ -\mathbf{1}^T\mathbf{A}_J & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \varepsilon(t) \end{bmatrix}$$

Dimension = $\#(J) + 1 = 7$

The FSP Theorem

(Munsky, *JCP* '06)

$$\mathbf{P}_J(t) \geq \mathbf{P}_J^{FSP}(t) \quad \text{and}$$

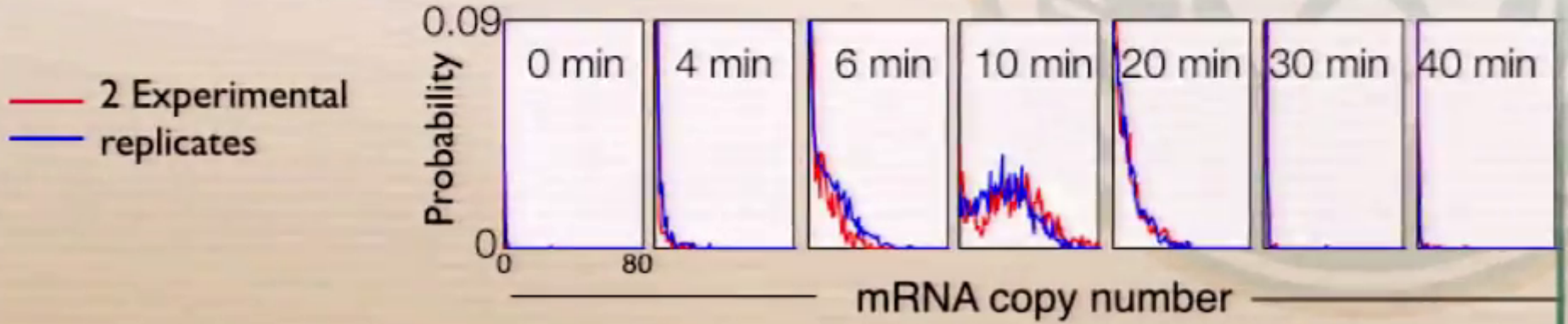
$$\left\| \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'} \end{bmatrix} - \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \mathbf{0} \end{bmatrix} \right\|_1 = \varepsilon(t)$$

Software and tutorial available at:
http://www.engr.colostate.edu/~munsky/Software.html

Inferring parameters from single-cell measurements.

Colorado State University

Although single-cell reactions may be ***Stochastic***, their statistics follow a ***Deterministic*** set of ODE's *(i.e., the CME).*

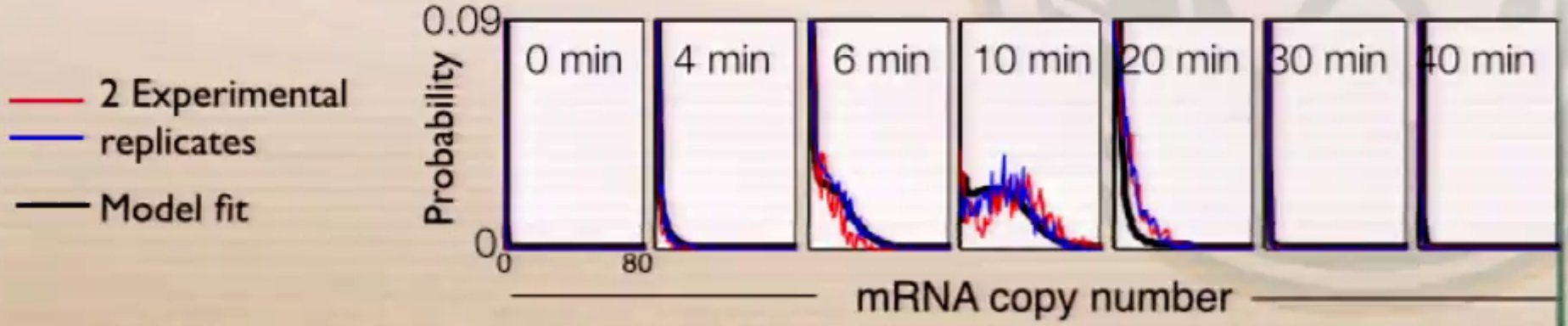


We can fit and potentially predict these statistics.

Inferring parameters from single-cell measurements.

Colorado State University

Although single-cell reactions may be ***Stochastic***, their statistics follow a ***Deterministic*** set of ODE's *(i.e., the CME).*



We can fit and potentially predict these statistics.

Fitting metrics:

Low cell counts --> maximum likelihood.*

$$\log L(\{n_i\}|\Lambda) = \sum_{i=1}^{N} \log p_{n_i}(\Lambda) = \sum_{j=0}^{\max\{n_i\}} m_j \log p_j(\Lambda)$$

High cell counts --> Kullback Leibler Divergence.*

$$D_{KL}(Q||P) = \sum q_j(\log q_j - \log p_j) = -\sum q_j \log(p_j) + C = -\frac{1}{N}\sum m_j \log(p_j) + \frac{C}{N}$$

**Minimization is equivalent for constant sample sizes per condition and time point.*

Outline

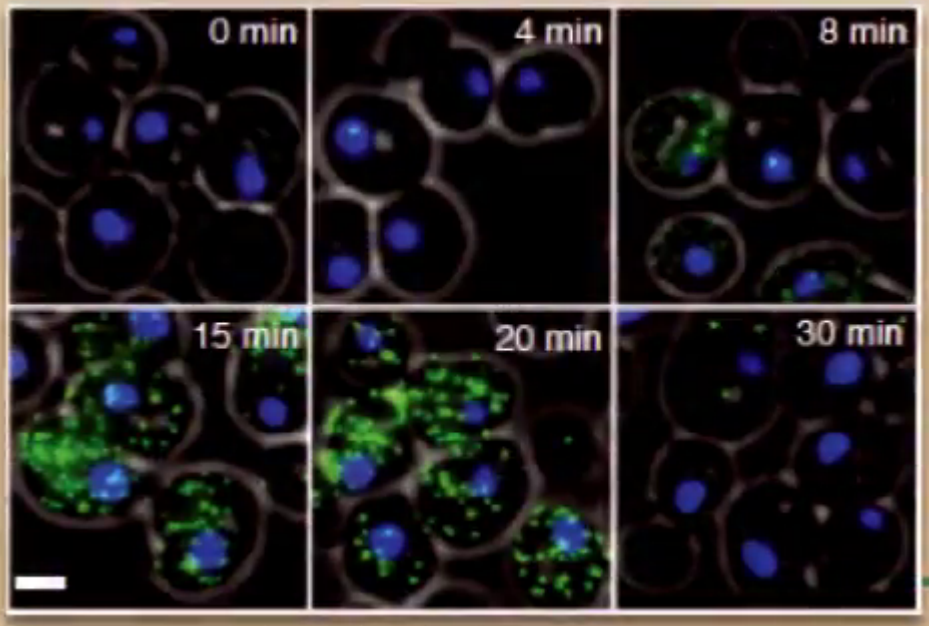
1. Introduction - Information from transcript fluctuation
2. Measuring and Modeling single-cell and single-molecule responses
3. **Case Studies:**
   i. **Predicting kinase-activated gene regulation dynamics in Saccharomyces cerevisiae (budding yeast)**
   ii. **Quantitative modeling for c-Fos mRNA burst dynamics in U2OS cells.**

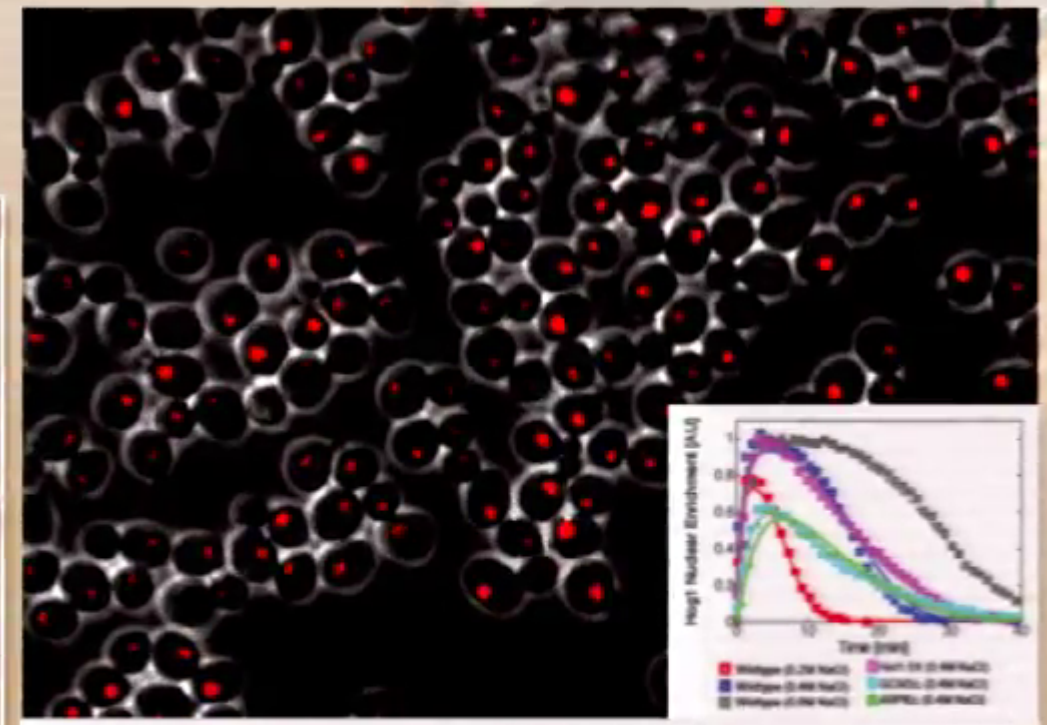
Signal-activated gene regulation

(Osmotic shock response in yeast)

Colorado State University

- 0.2M NaCl is added at t=0.
- Hog1 (red) activates in 1-2 min.
- ... and remains active for ~12 min.
- *Stl1* mRNA appear at 4 min.
- ... and are gone by 25 min.





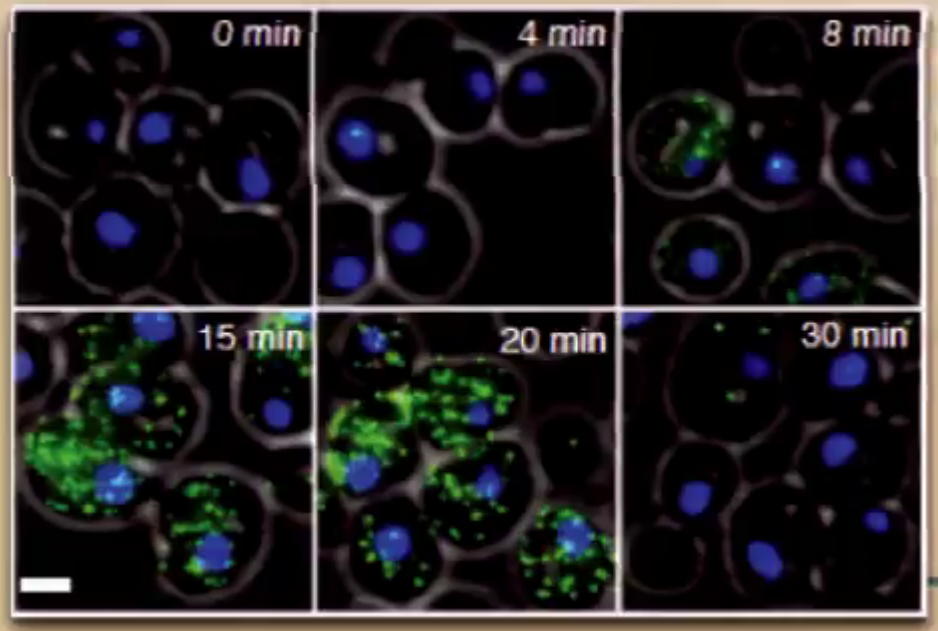
0-20 minutes after 0.2M NaCl shock
(Neuert, Munsky, et al, 2013)

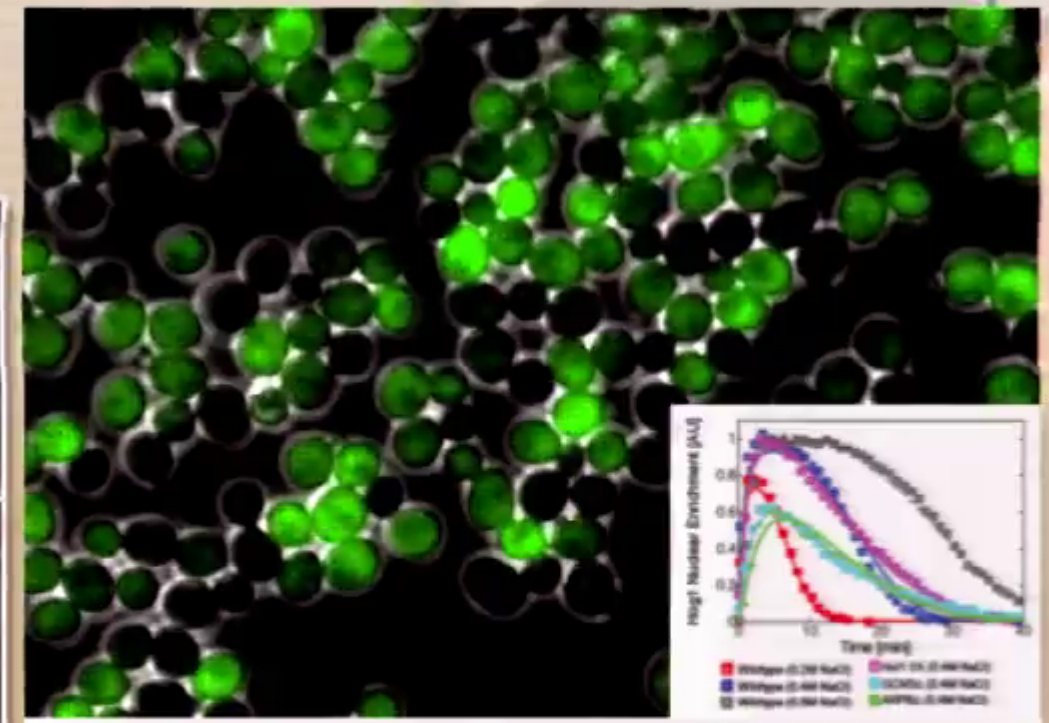
Signal-activated gene regulation

(Osmotic shock response in yeast)

Colorado State University

- 0.2M NaCl is added at t=0.
- Hog1 (red) activates in 1-2 min.
- ... and remains active for ~12 min.
- *Stl1* mRNA appear at 4 min.
- ... and are gone by 25 min.
- Stl1-GFP appear at ~30 min.





0-60 minutes after 0.2M NaCl shock
(Neuert, Munsky, et al, 2013)

Possible model structures:

The Hog1 kinase (★) activates STL1, but how?

Colorado State University

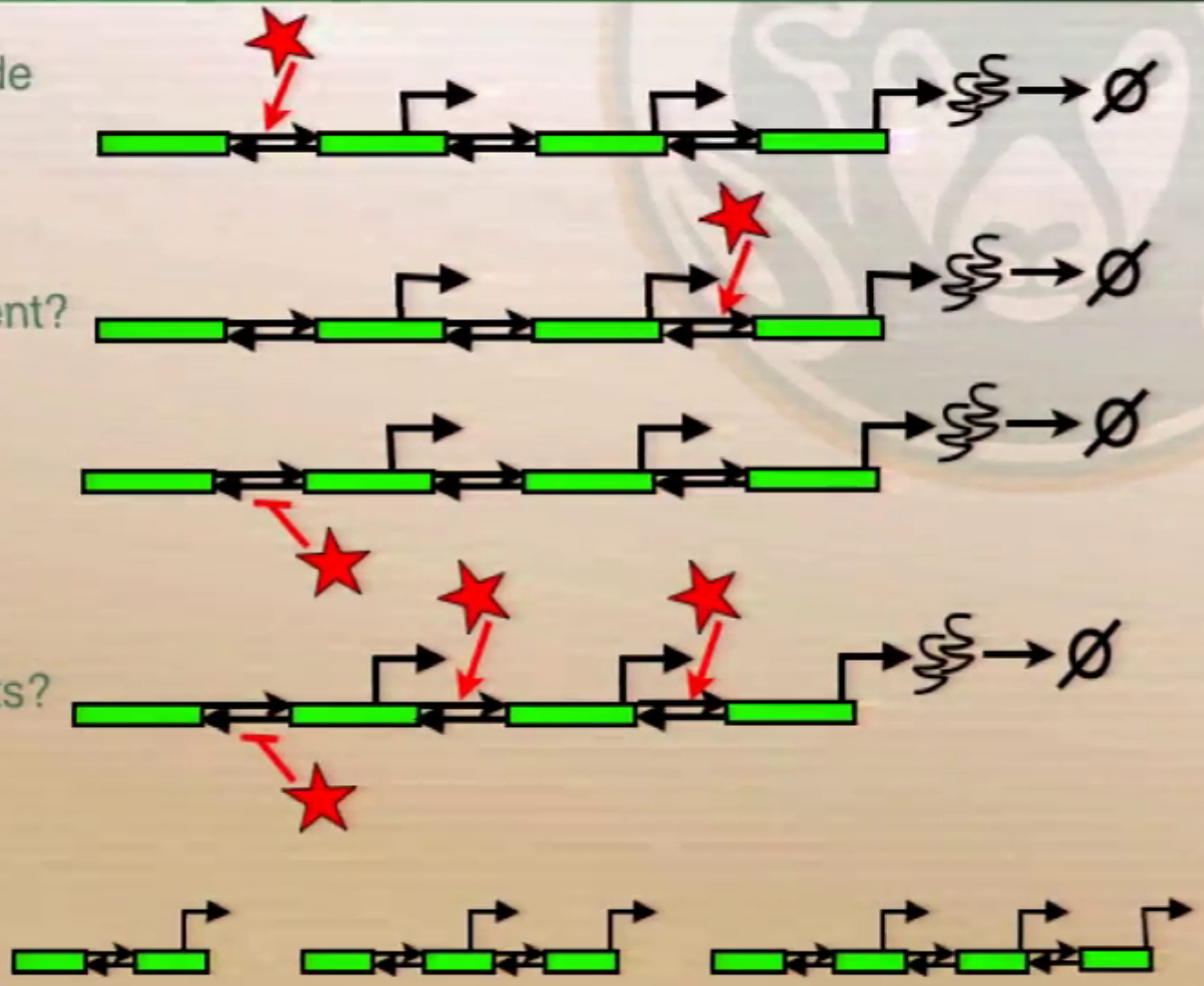
Is it the first of a cascade of activation events?

...the last activation event?

Does it repress a deactivation event?

Are there multiple effects?

How many states are needed?

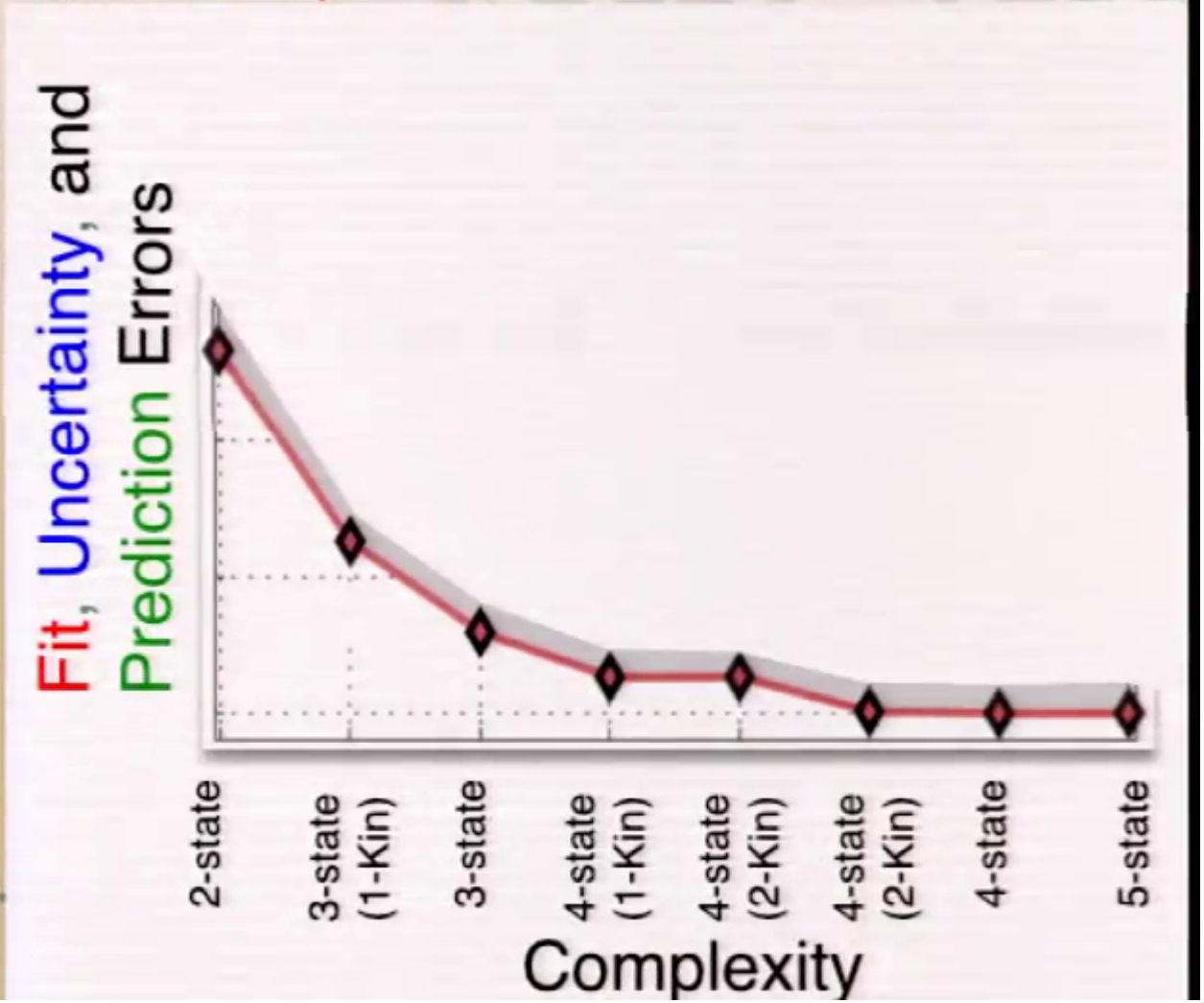


Evaluating model structures of varying complexity

Colorado State University

We fit different 2-, 3-, 4- and 5- state model structures to wild-type data at 0.4M osmotic shock.

More states (and parameters) yield better fits,...



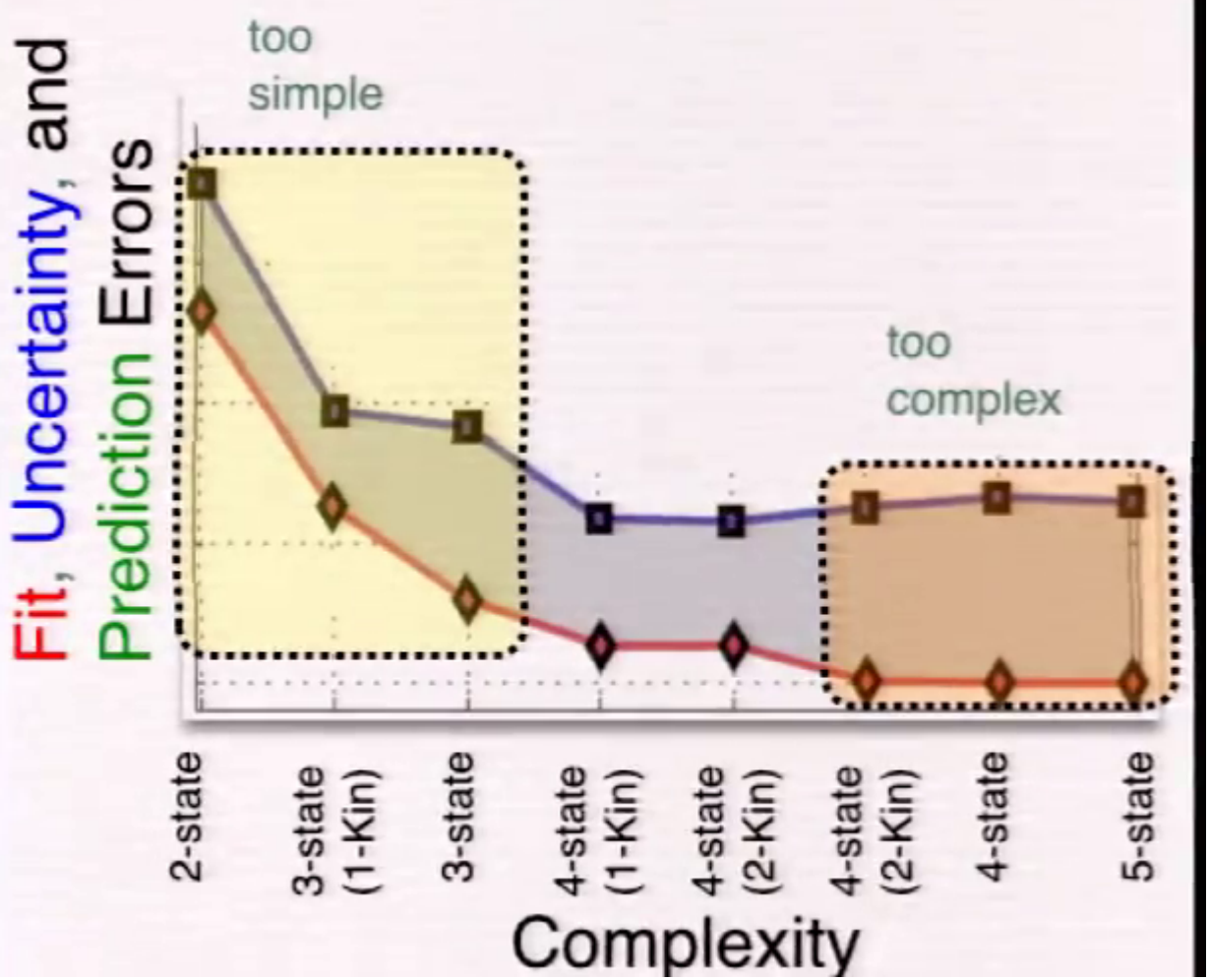
Evaluating model structures of varying complexity

Colorado State University

We fit different 2-, 3-, 4- and 5- state model structures to wild-type data at 0.4M osmotic shock.

More states (and parameters) yield better fits,...

but they also give rise to greater uncertainty.

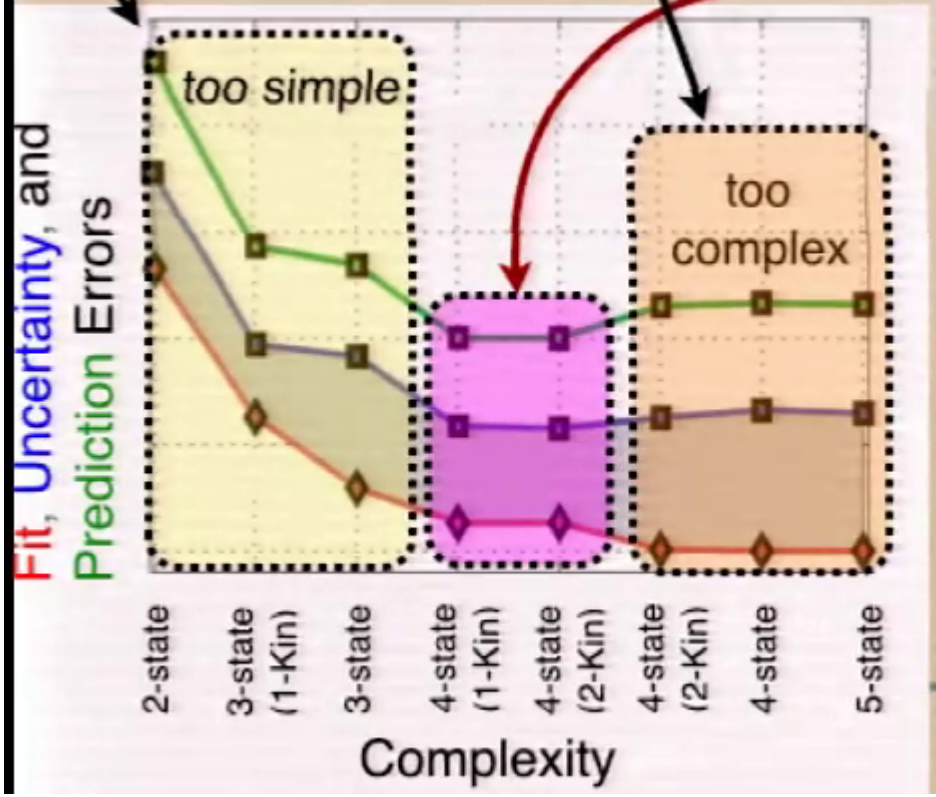


Evaluating model structures of varying complexity

Colorado State University

Overly-simple models cannot match the data. → Inaccurate predictions.

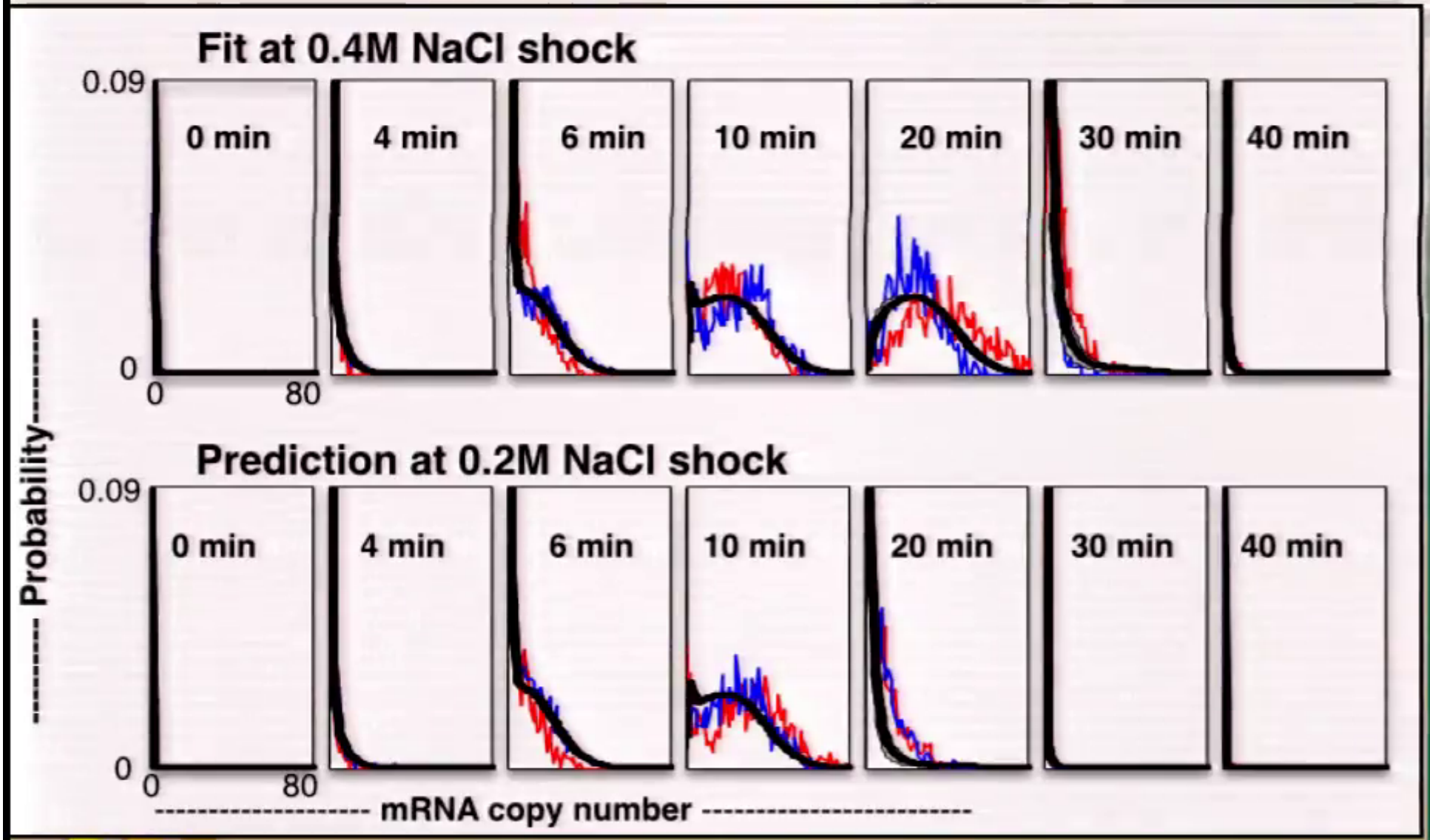
Overly-complex models are poorly constrained. → Imprecise predictions.

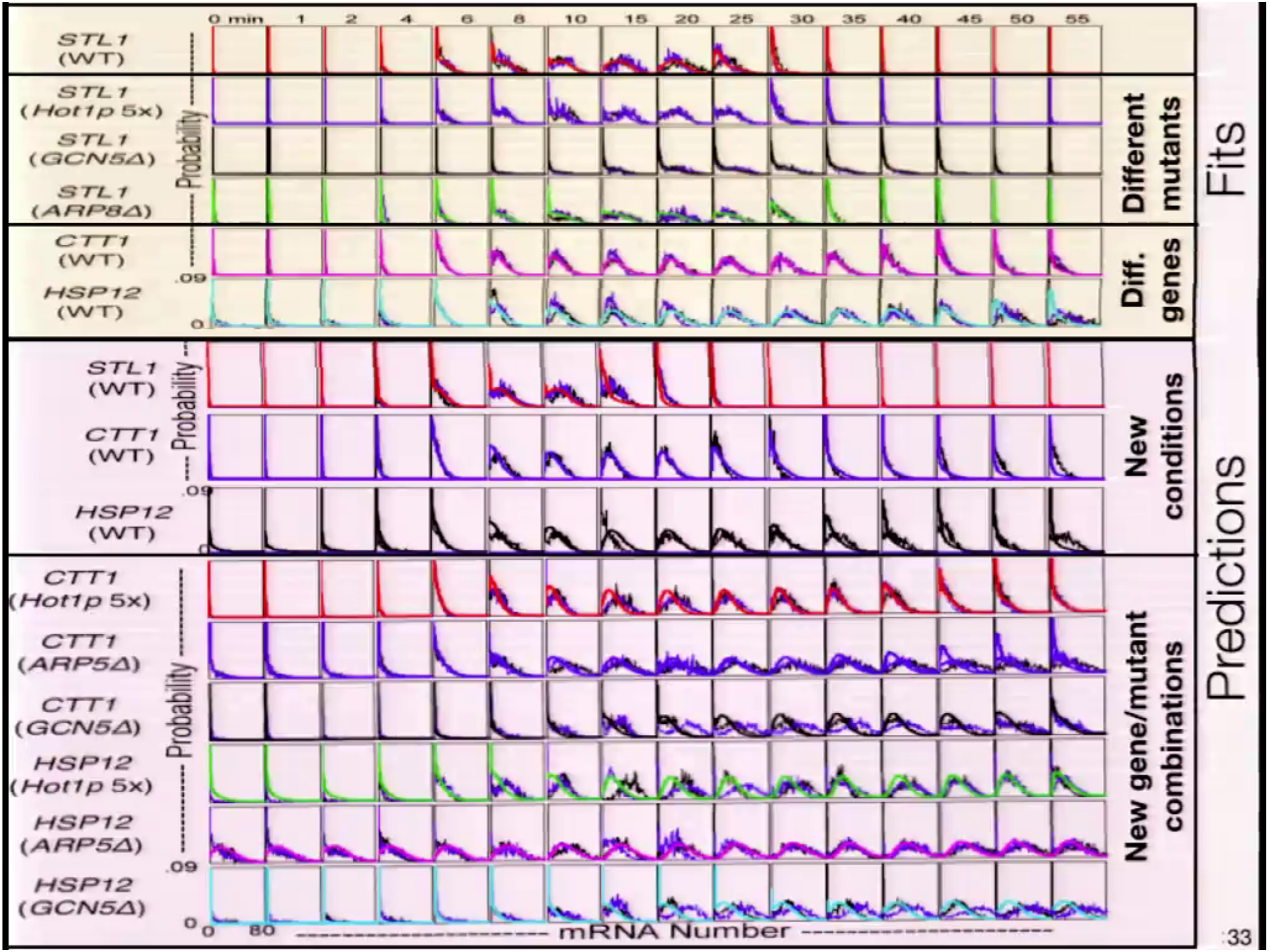


Cross-validation analysis provides an excellent a priori estimate of predictive power.

Fits and predictions for *STL1* regulation

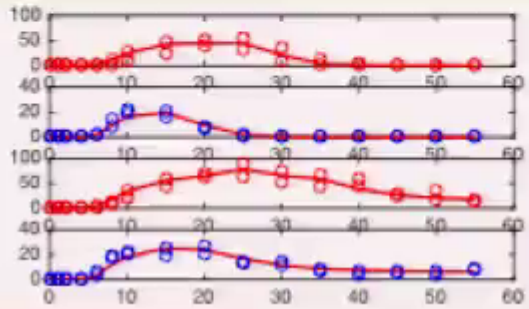
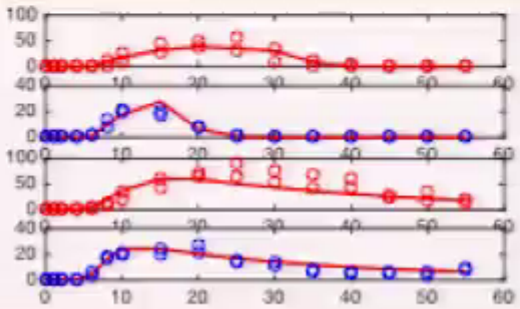
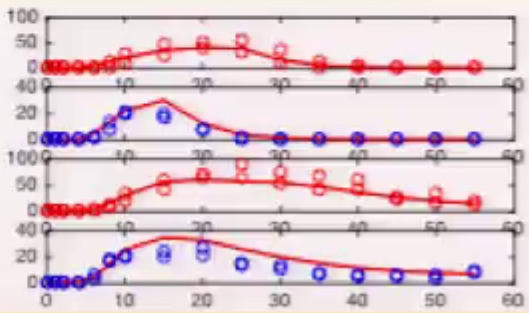
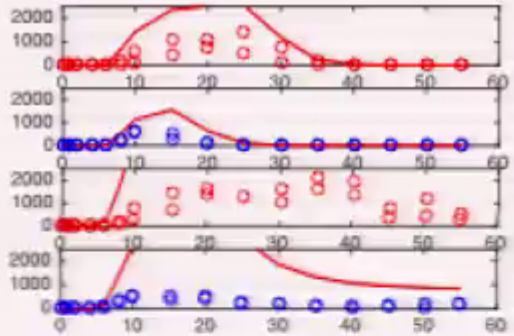
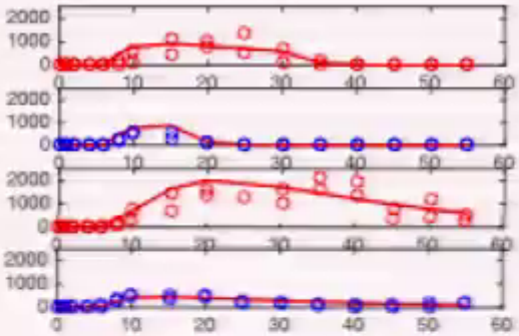
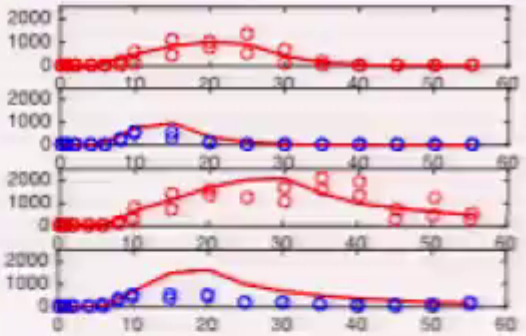
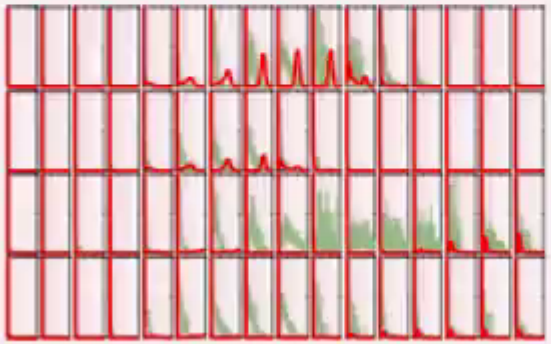
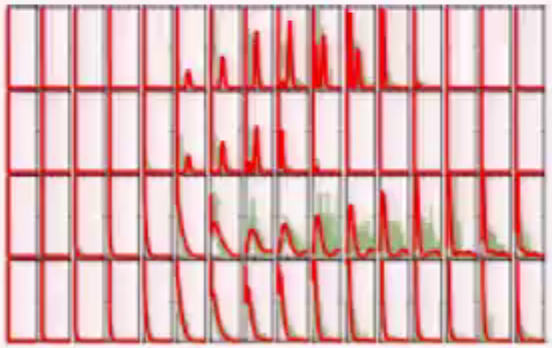
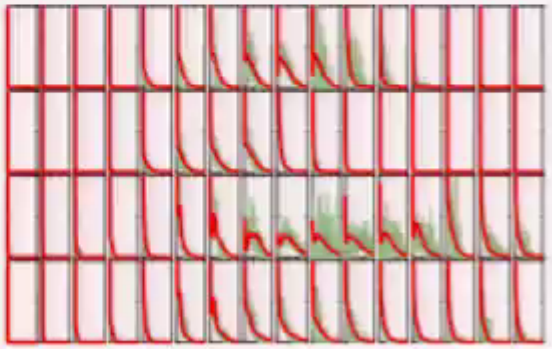
Colorado State University




0 min
1
2
4
6
8
10
15
20
25
30
35
40
45
50
55
STL1 (WT)
STL1 (Hot1p 5x)
STL1 (GCN5Δ)
STL1 (ARP8Δ)
CTT1 (WT)
HSP12 (WT)
Probability
.09
0
Different mutants
Diff. genes
Fits
STL1 (WT)
CTT1 (WT)
HSP12 (WT)
Probability
.09
0
New conditions
CTT1 (Hot1p 5x)
CTT1 (ARP5Δ)
CTT1 (GCN5Δ)
HSP12 (Hot1p 5x)
HSP12 (ARP5Δ)
HSP12 (GCN5Δ)
Probability
.09
0
0
80
mRNA Number
New gene/mutant combinations
Predictions

Results from fitting different statistics.

Colorado State University

	Fit Means	Fit Mean & Variance	Distributions
Plot Means			
Plot Variance			
Plot Distributions			

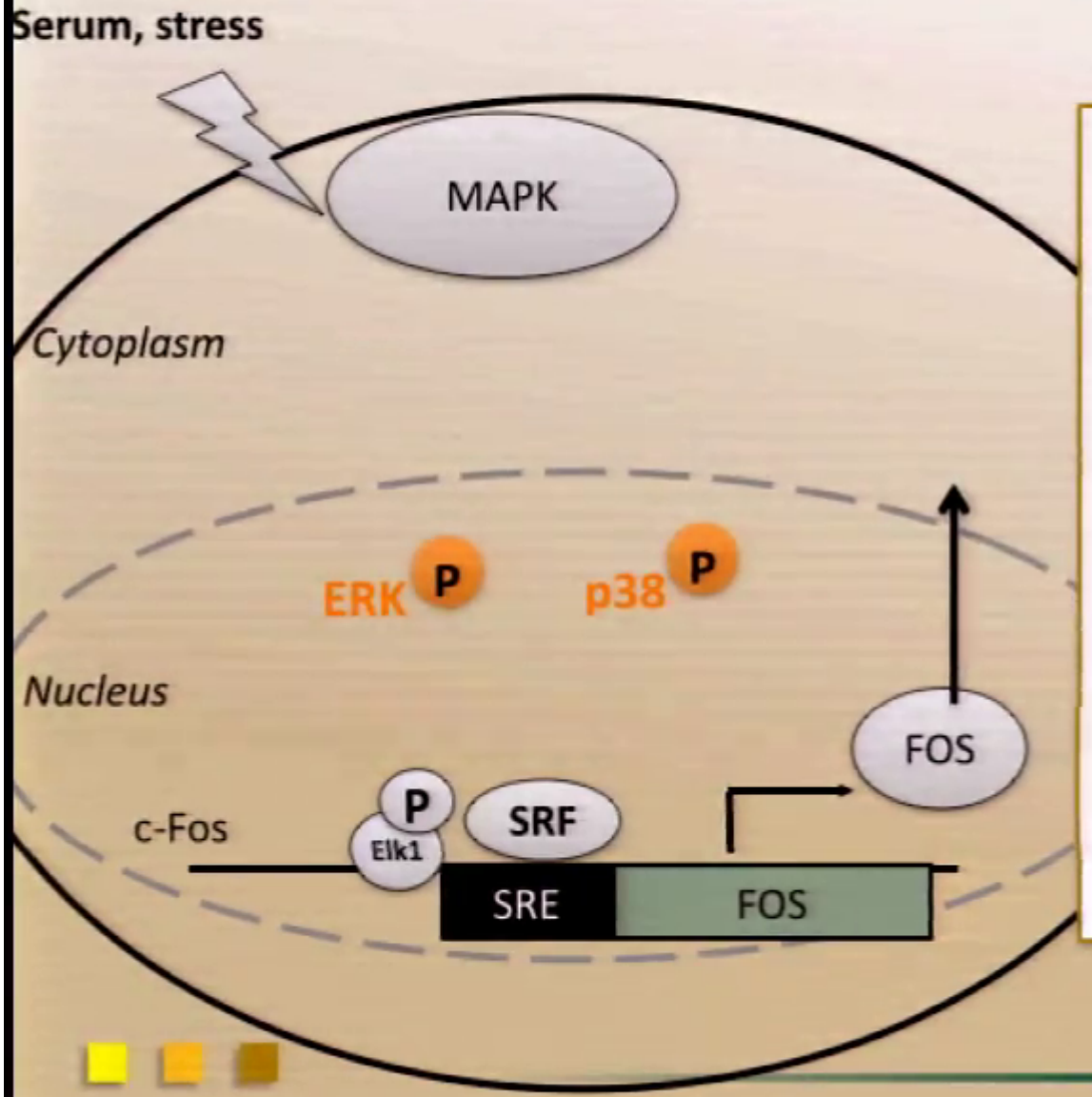
Outline

Colorado State University

1. Introduction - Information from transcript fluctuation
2. Measuring and Modeling single-cell and single-molecule responses
3. **Case Studies:**
   i. **Predicting kinase-activated gene regulation dynamics in Saccharomyces cerevisiae (budding yeast)**
   ii. **Quantitative modeling for c-Fos mRNA burst dynamics in U2OS cells.**

A glance at the activation of c-Fos

Colorado State University



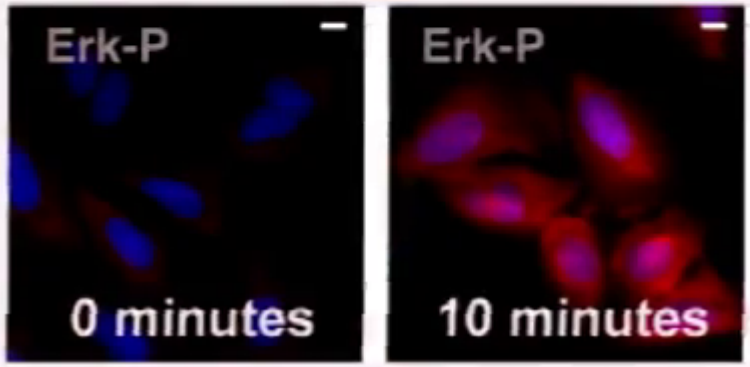
Chain of events:

- Serum (zinc stress) activates phosphorylation of ERK (p38).
- ERK/p38 translocate to nucleus & phosphorlyate Elk1.
- Elk1 and serum response factor (SRF) bind to serum response element (SRE).
- The activated promoter now transcribes c-Fos mRNA.
- c-Fos affects differentiation, proliferation, survival, …

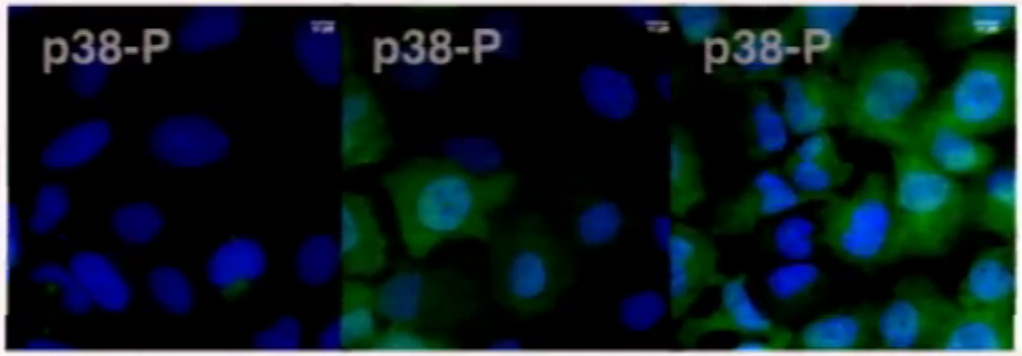
Dynamics of MAPK Signal Induction

Colorado State University

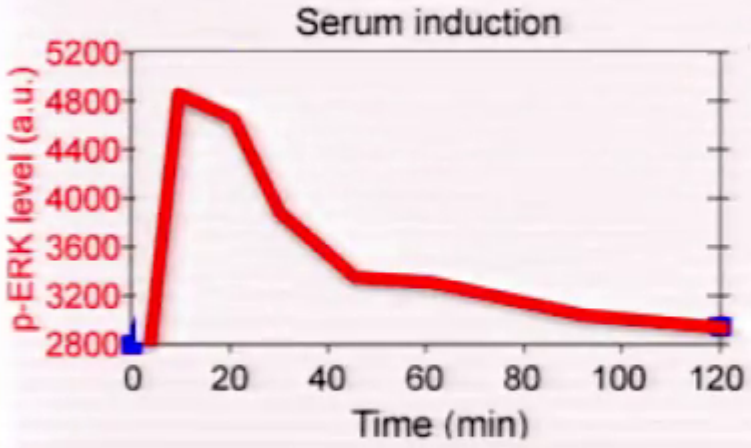
Erk and p38 phosphorylation following stimulus

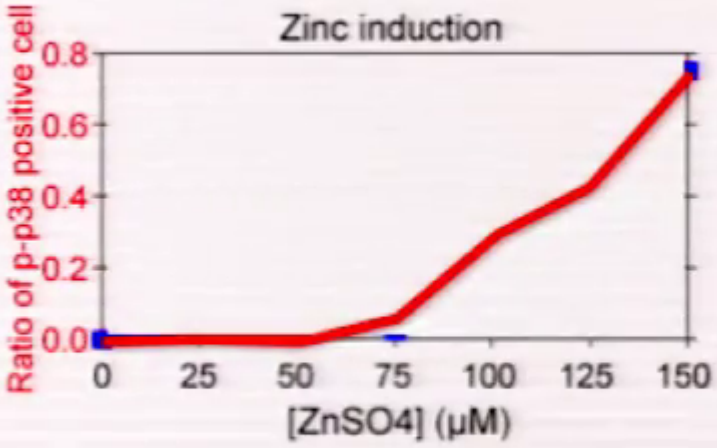


serum induction



0μM ZnSO4 75μM ZnSO4 4h 125μM ZnSO4



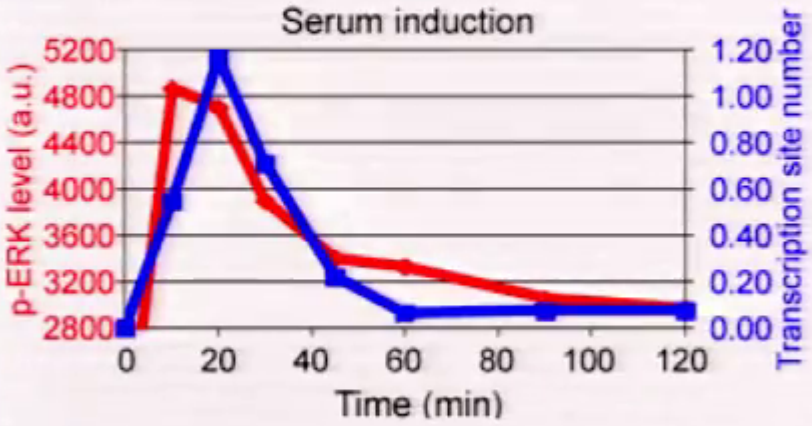


Signaling Affects Transcription Site Activation

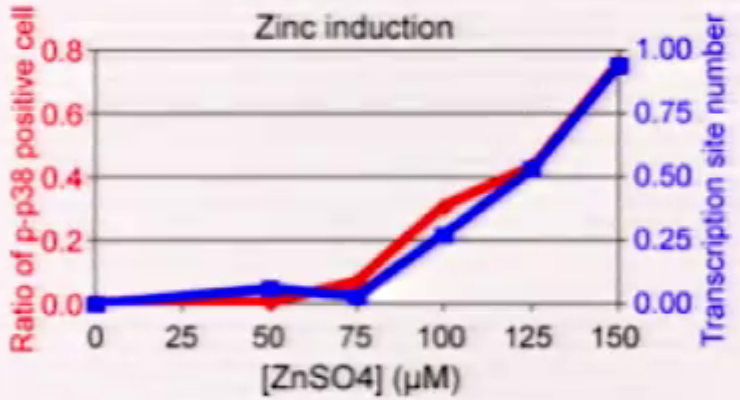
Colorado State University

The number of ATS's varies randomly from cell to cell.

The *average* number of ATS's tracks MAPK induction dynamics.



Serum activation

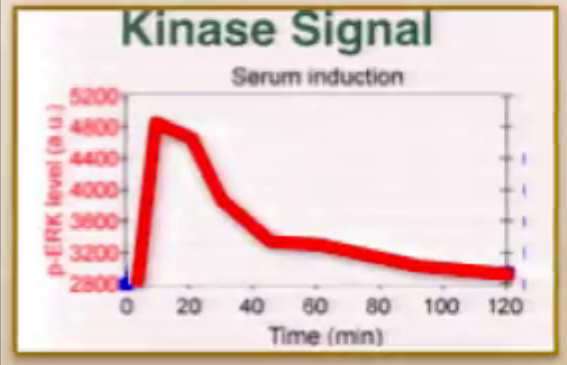


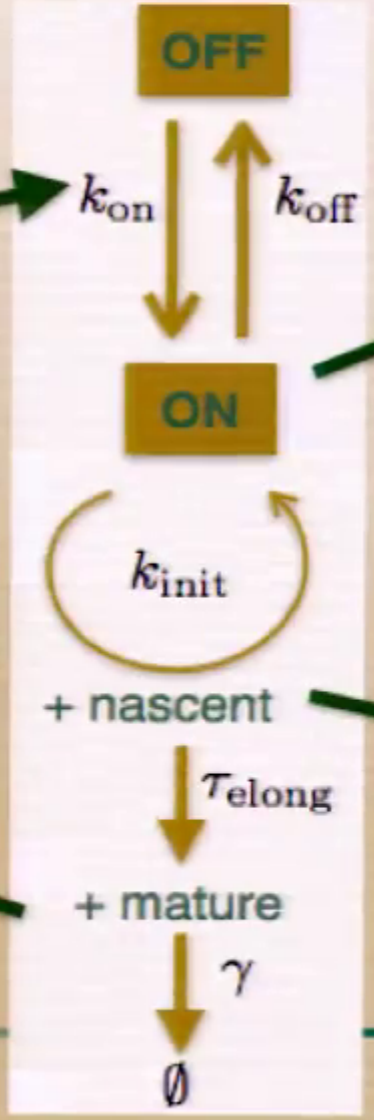
Zinc activation

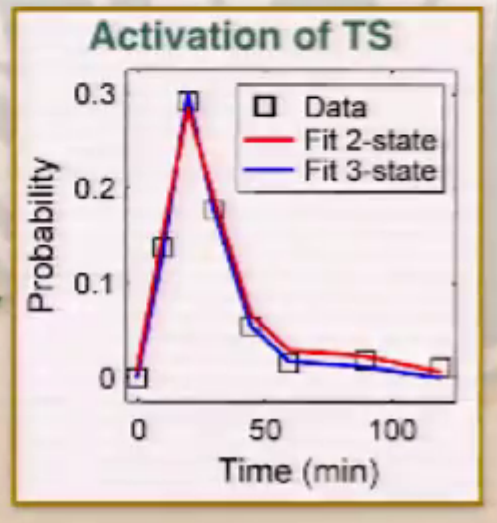
Quantitative fits to signal-activated transcription dynamics

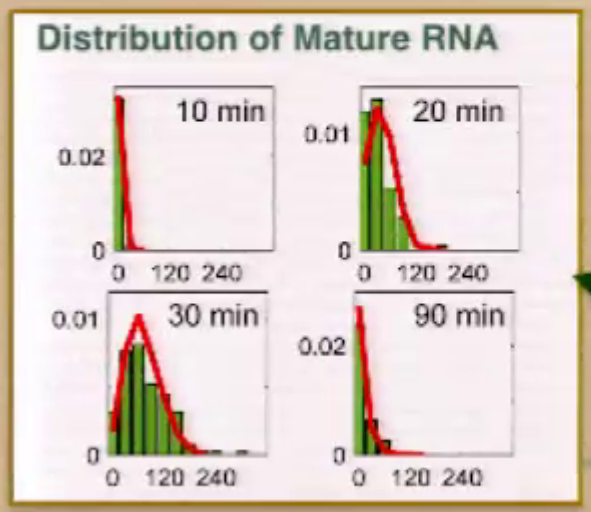
Colorado State University

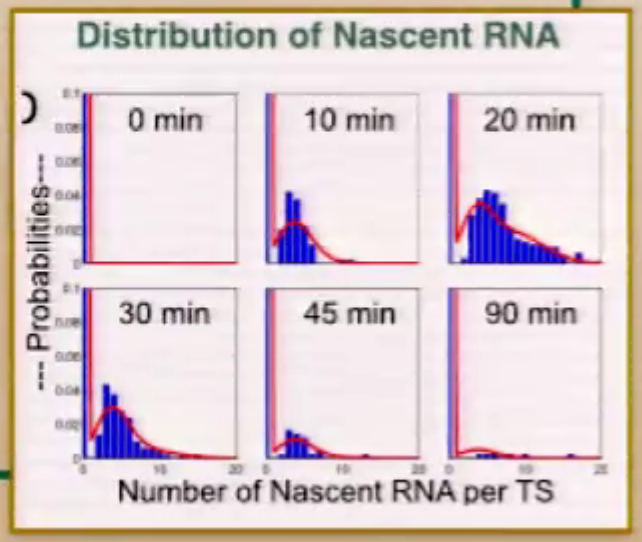
- Burst saturation model accurately captures c-Fos dynamics.











Summary and Conclusions

Colorado State University

Stochastic, temporal, and spatial fluctuations affect single-cell dynamics

These complicate modeling and disrupt transcription control.

But statistics of these fluctuations follow deterministic rules:

Cells exhibit distinct, measurable `fluctuation fingerprints'.

Can reveal subtle gene regulation mechanisms & parameters.

Can be predicted with high accuracy and precision.

Uncertainty Quantification reveals when models are too simple, too complex, or just right (i.e., the Goldilocks Model).

We have identified **predictive quantitative models** of transcription regulation for many natural and synthetic genes in several organisms.

Prediction is the first step toward design, optimization and control.

References:

Colorado State University

...wherein dynamic stimuli and single-cell measurements reveal gene regulation mechanisms

1. *Munsky, *Neuert, van Oudenaarden, Using Gene Expression Noise to Understand Gene Regulation, ***Science***, **336**:6078, 183–187, 2012.

2. *Neuert, *Munsky, *et al*, Systematic Identification of Signal-Activated Stochastic Gene Regulation, ***Science***, **339**:6119, 584-587, 2013.

3. Senecal, Munsky, *et al*, Transcription Factors Modulate c-Fos Transcriptional Bursts, ***Cell Reports***, **8**:1, 75-83, 2014.