



Predictive understanding of cell biological systems through kinetic analysis

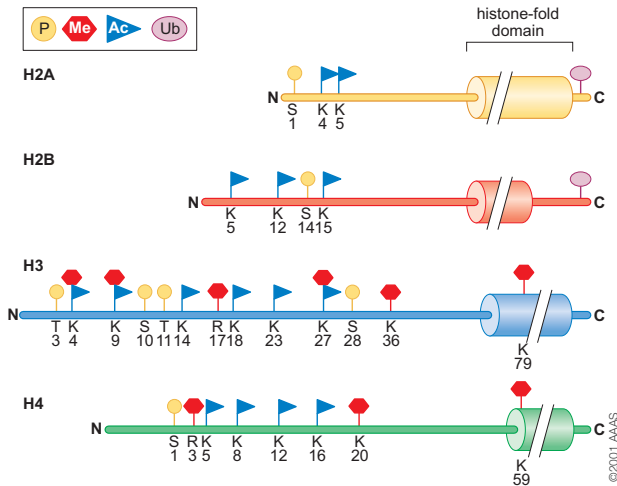
Gregor Neuert

Vanderbilt University

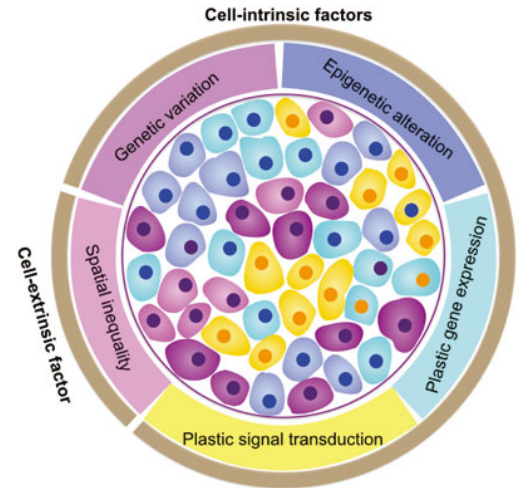
NSF-workshop: "Finding Your Inner Modeler."
University of Alabama at Birmingham
June 13 & 14, 2019

Biology is governed by combinatorial complexity

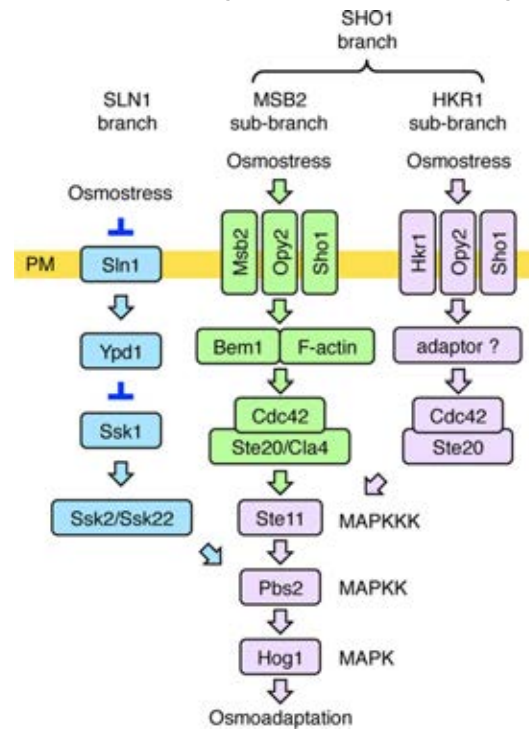
Histone code



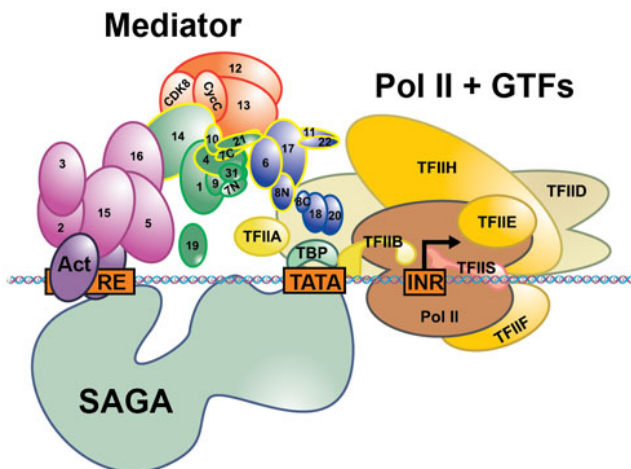
Disease mutations / Cell types







Pathway complexity



Protein complexes

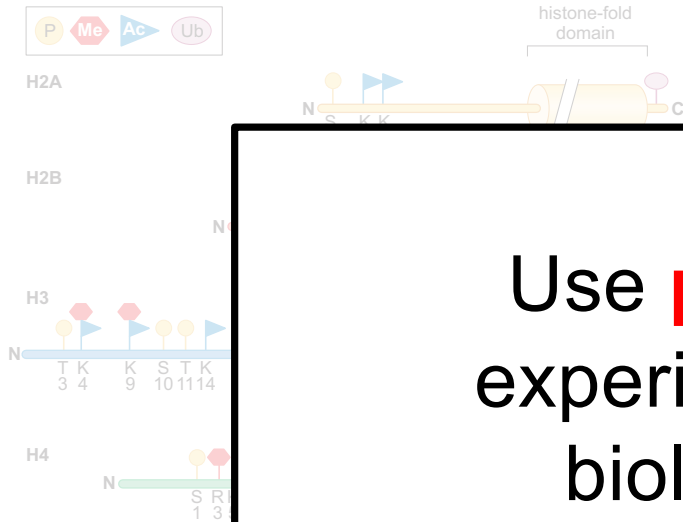


Genotype / Phenotype

	A	A ^y
A	Agouti coat AA 	Yellow Coat AA ^y 
A ^y	Yellow coat AA ^y 	Dead A ^y A ^y 

Biology is governed by combinatorial complexity

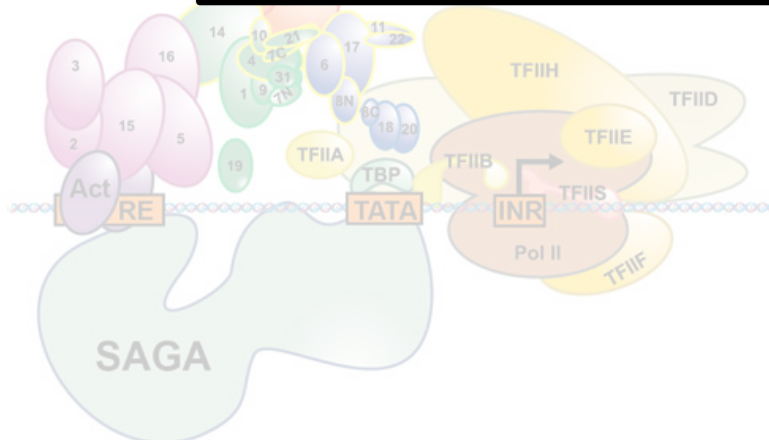
Histone code



Disease mutations /
Cell types

Cell-intrinsic factors

Use **predictive** modeling and experimental design to address biological **complexity** and molecular **mechanism** in health and disease state



A	AA 	AA ^y 
A ^y	Yellow coat AA ^y 	Dead A ^y A ^y 

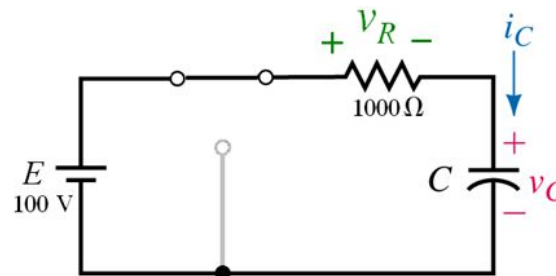
Many physical systems are deterministic and models are very predictive

Homogeneous / Deterministic: variance in experimental data is lower than variance in measurement

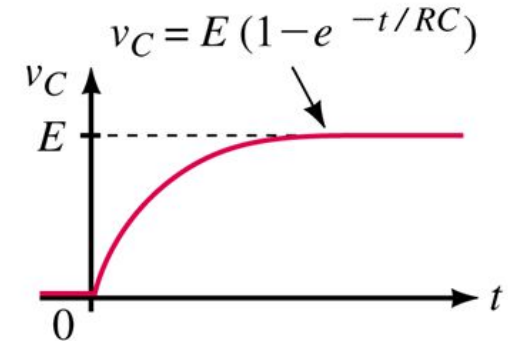
Radio



Simple model



Model simulation and prediction



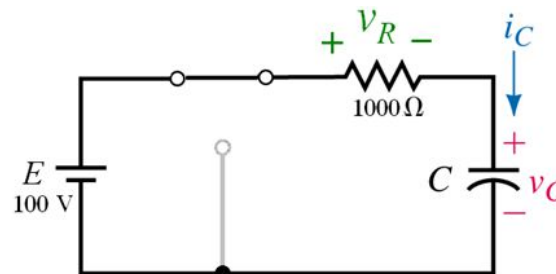
Fundamental differences in modeling cells in comparison to physical objects

Homogeneous / Deterministic: variance in experimental data is lower than variance in measurement

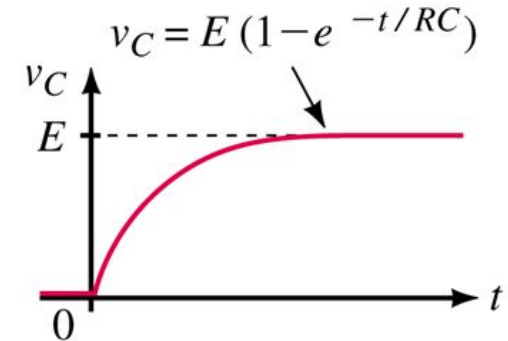
Radio



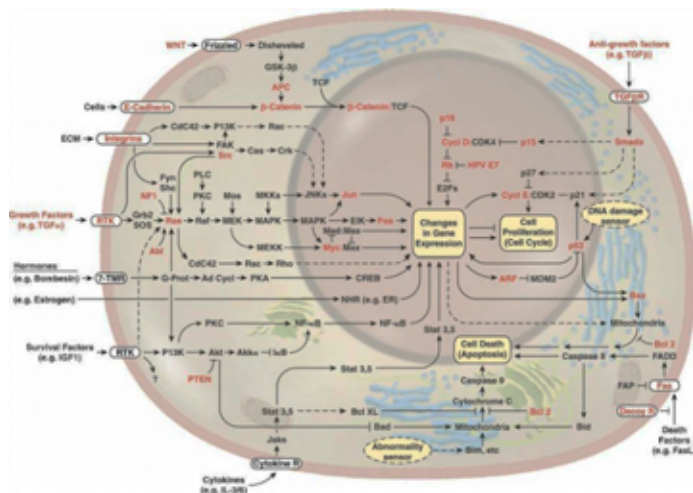
Simple model



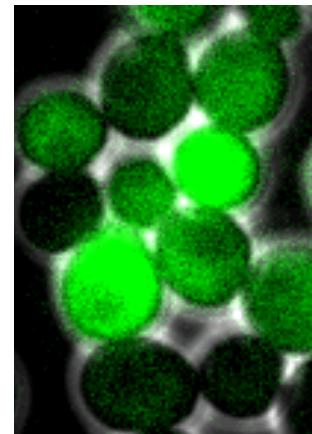
Model simulation and prediction



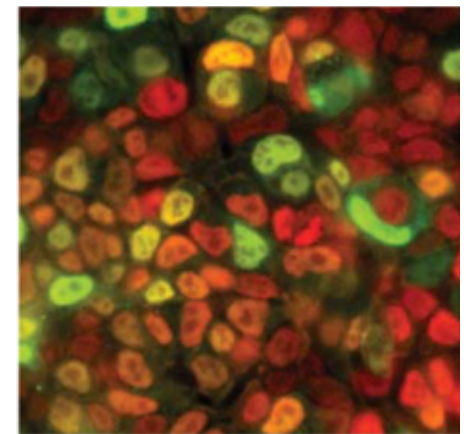
Heterogeneous / Stochastic: variance in experimental data is higher than variance in measurement



Yeast



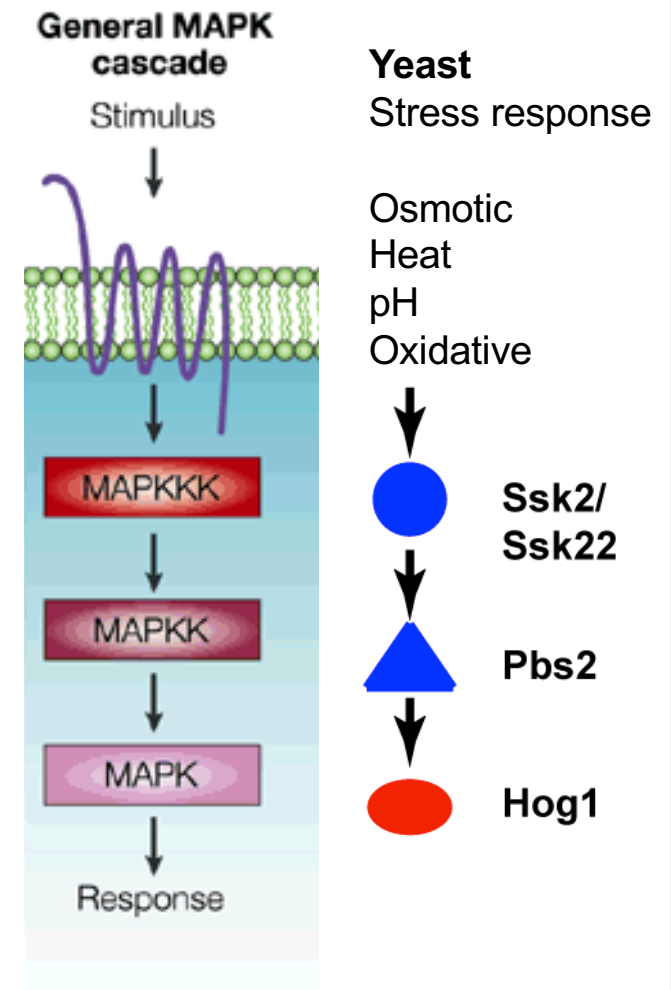
Stem Cells



How to practically approach
predictive modeling in biology?

Yeast are a great model organism for modeling

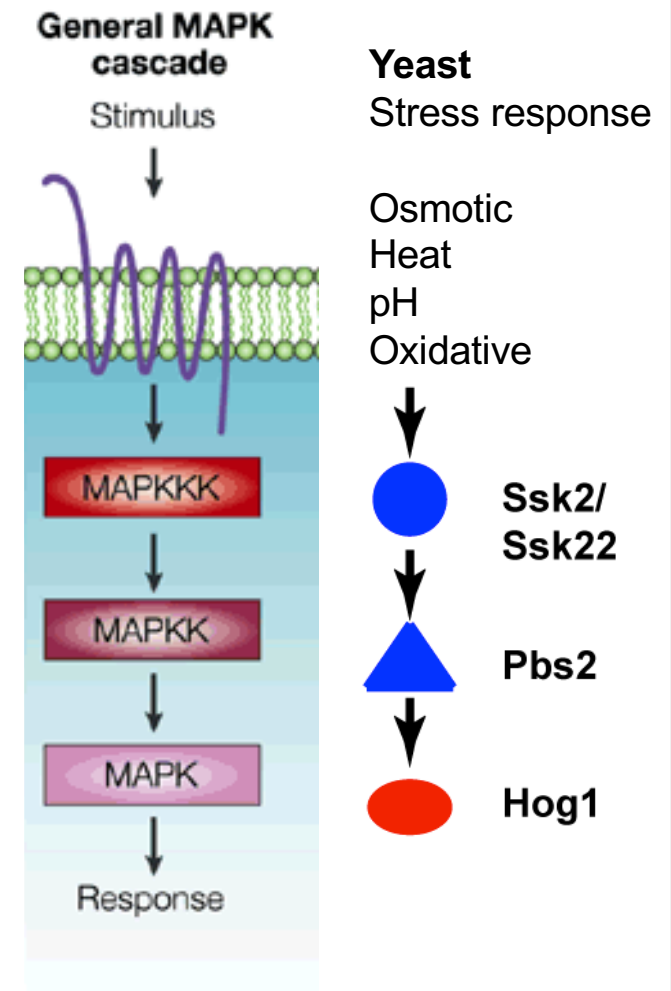
1. Many fundamental biological processes are conserved from yeast to human and have been first discovered in yeast (many Nobel prizes).
2. Yeast cells are much smaller than mouse or human cells which allows to image the same data in 1 day instead of 10-30 days at the same resolution.
3. Yeast cell cycle (90 min) vs mouse/ human cells (16 – 48h) => faster turnaround.
4. Yeast genetics is still much more controlled, efficient and cheaper than CRISPR in mouse or human cells.



Kristen L. et. Al. Nature Reviews Molecular Cell Biology, 2002

Yeast are a great model organism for modeling

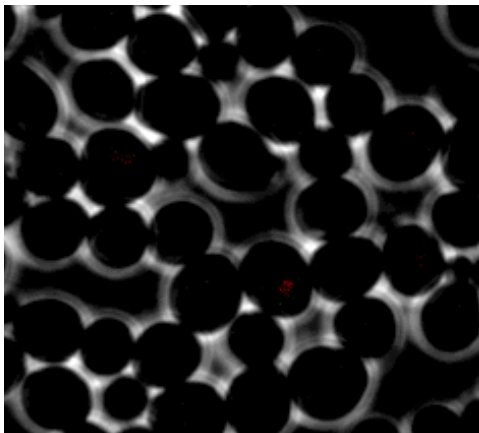
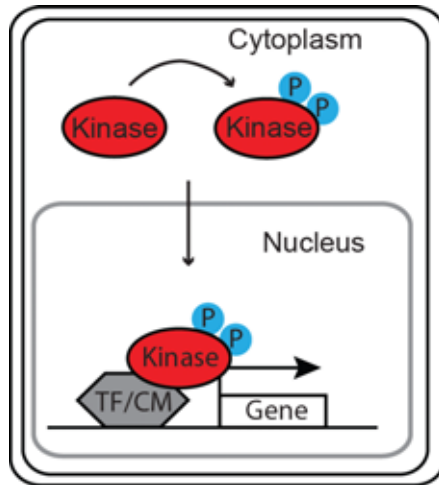
1. Many fundamental biological processes are conserved from yeast to human and have been first discovered in yeast (many Nobel prizes).
2. Yeast cells are much smaller than mouse or human cells which allows to image the same data in 1 day instead of 10-30 days at the same resolution.
3. Yeast cell cycle (90 min) vs mouse/ human cells (16 – 48h) => faster turnaround.
4. Yeast genetics is still much more controlled, efficient and cheaper than CRISPR in mouse or human cells.
5. What is a representative normal human cell?
6. How do we know results are not tissue or disease specific but rather of general interest?
7. What is a biological replica in a patient sample?
8. Are cell lines more informative than yeast cells?
9. Do we want to cure mice or humans?
- 10. We are interested in understanding fundamental questions of how to model biological systems.**



Kristen L. et. Al. Nature Reviews Molecular Cell Biology, 2002

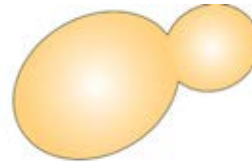
Single cell approaches to quantify signal transduction and transcription

Signal transduction



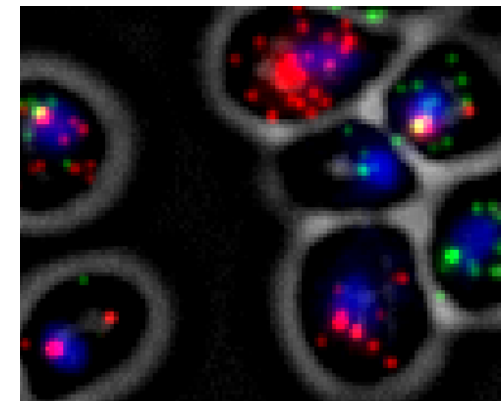
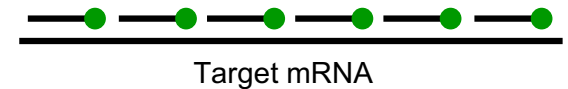
Live cell time lapse microscopy

Osmotic stress / NaCl



Transcription

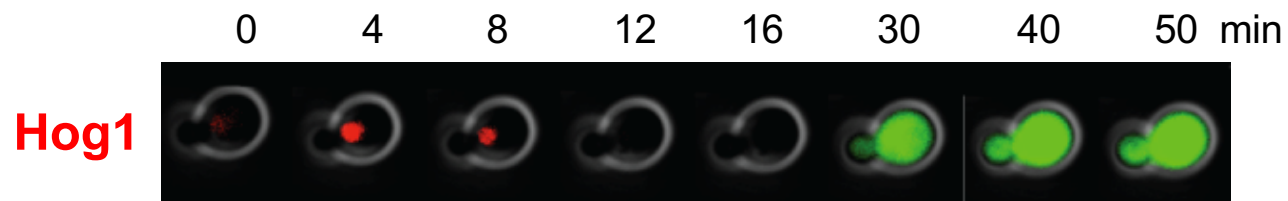
30 – 60 probes, 20nt each



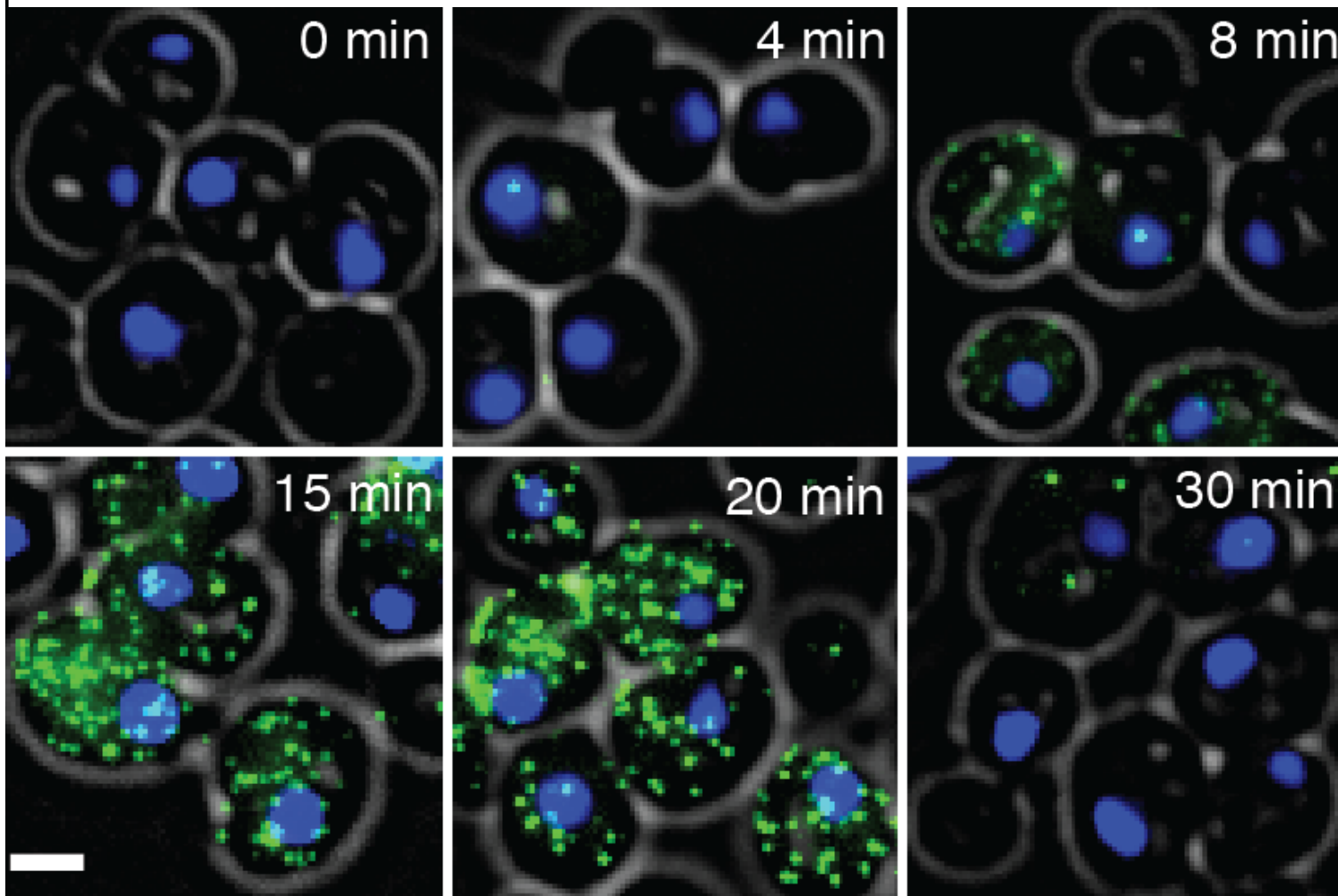
- Femino et al., Science 1998
- Raj et al., Nature Method 2008
- Bumgarner, Neuert et al., Molecular Cell 2012
- Munsky, Neuert et al., Science 2012
- Van Werven, Neuert et al., Cell 2012
- Neuert, Munsky et al., Science 2013

Fluorescent in-situ hybridization with single molecule resolution

Homogeneous signal transduction results in heterogeneous transcription



Endogenous mRNA expression of the *STL1* gene

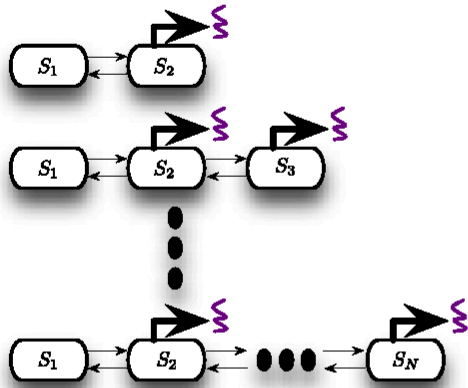


- Yeast cell boundary
- DAPI stained nucleus
- STL1* mRNA

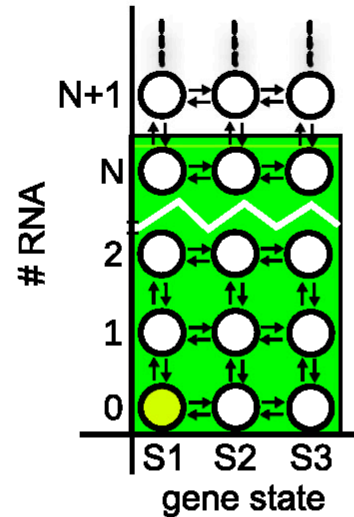
3 genes
2 different conditions
WT and three mutants
each 16 time points
2-3 biological replica
~ 160.000 cells

FSP: Solving the stochastic chemical master equation

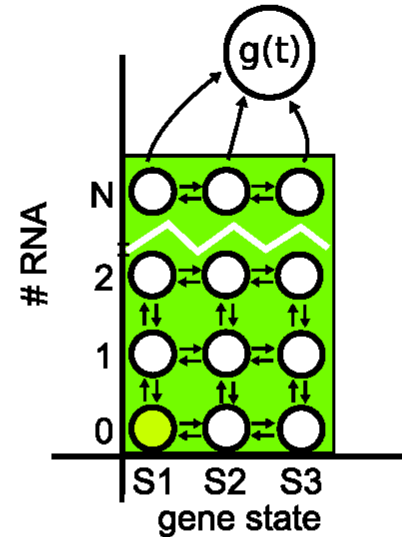
A N-state gene regulatory model



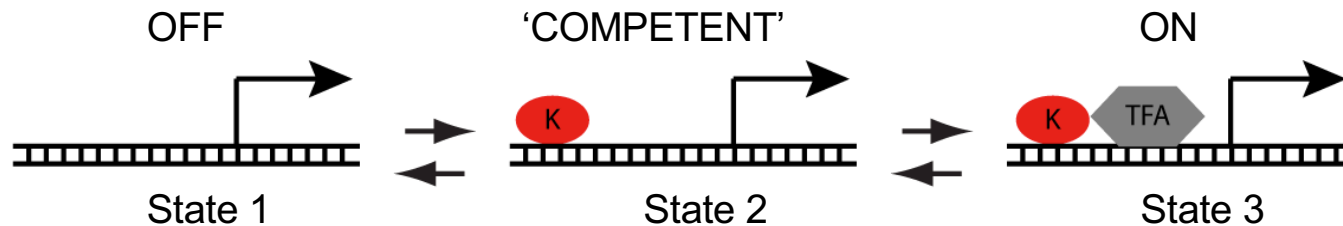
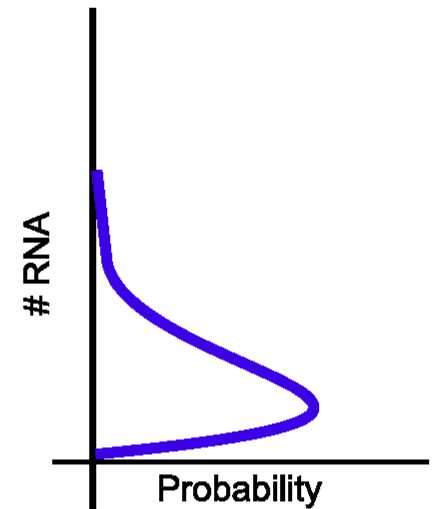
B Continuous time Markov chain



C Finite State Projection (FSP)



D Total RNA Probability Mass Function



Solve chemical master equation: description of the probabilities over time

Determine Probability density of the gene regulatory model:

$$P_{i,m} = P(\text{state} = i, \text{mRNA} = m)$$

Solving CME very fast with finite state projection

Munsky et al., JCP, 2006; Munsky, Fox, Neuert, Methods, 2015; Fox, Neuert, Munsky, JCP, 2016

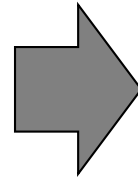
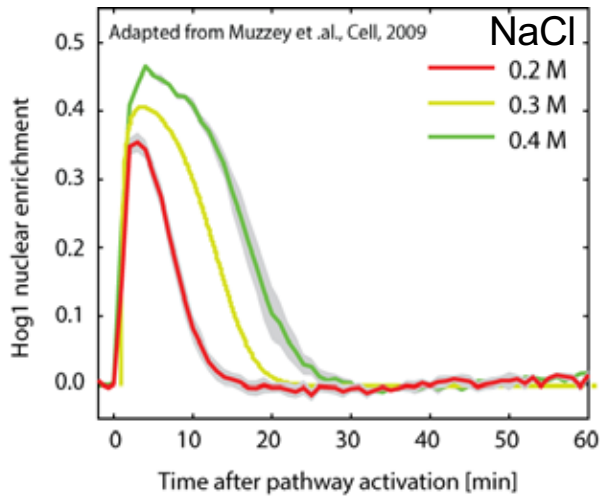


In collaboration with:

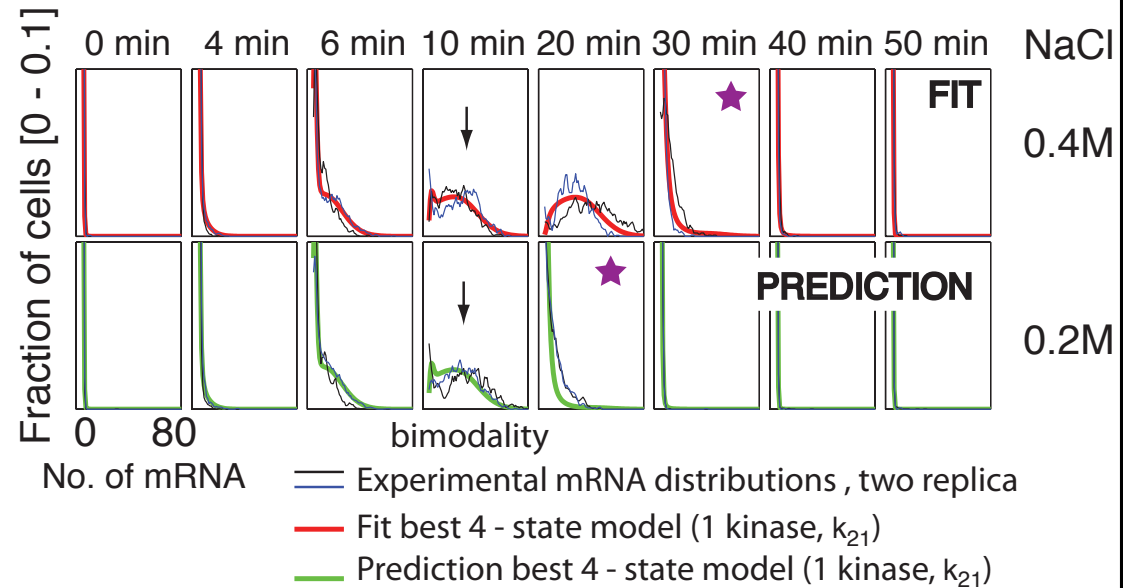
Dr. Brian Munsky
Assistant Professor
Colorado State University

Objective identification of predictive model

Input:
Signal transduction dynamics



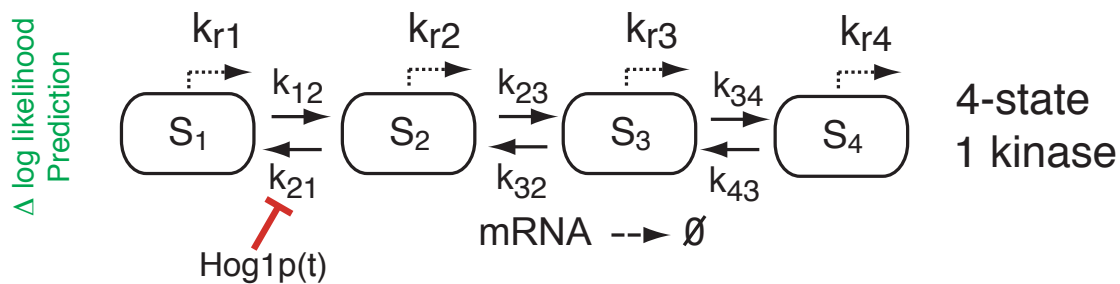
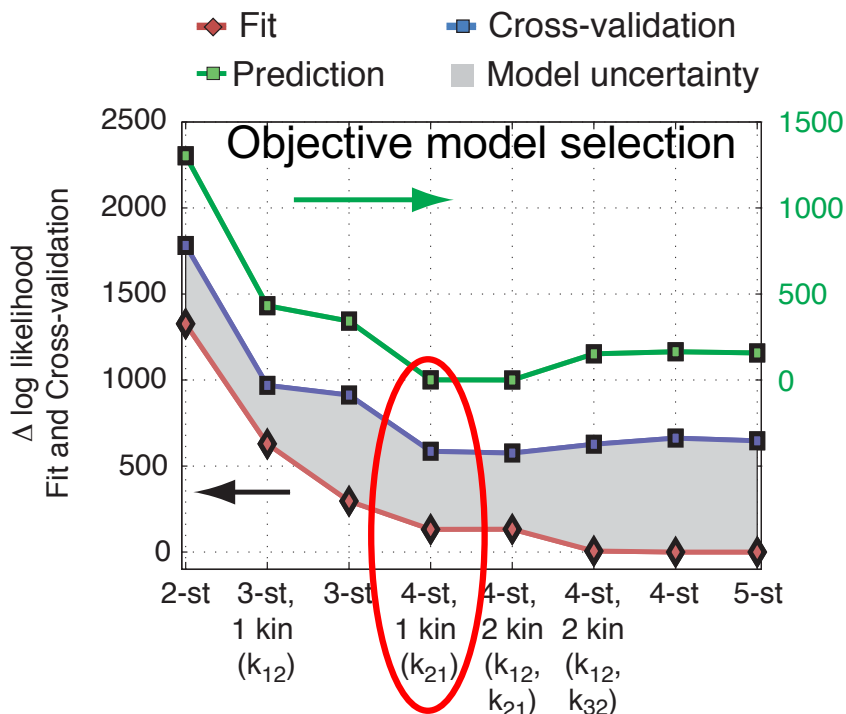
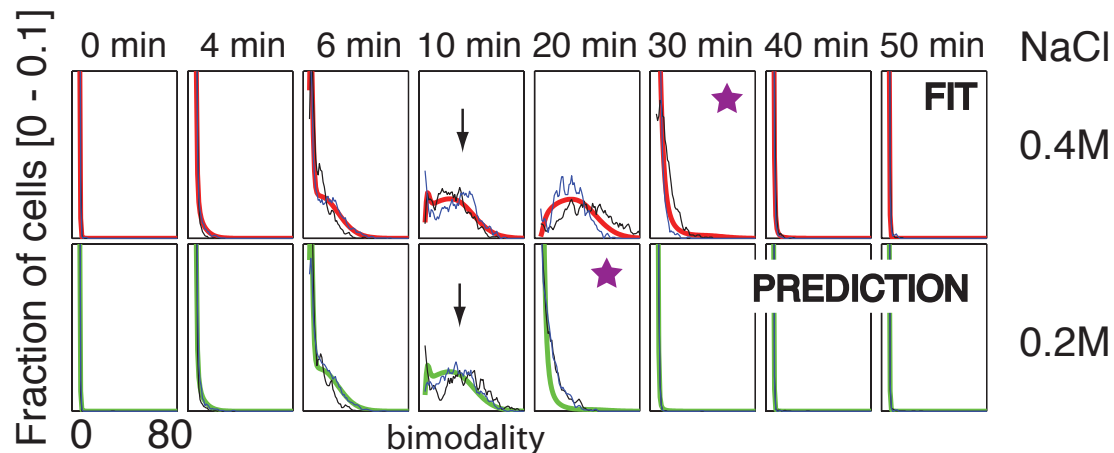
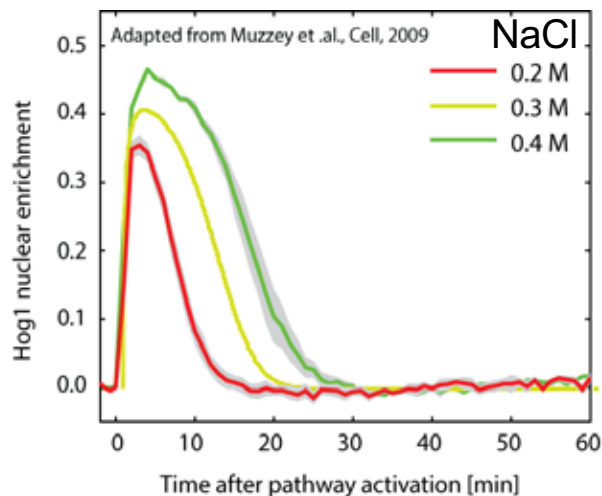
Output:
mRNA expression dynamics



Objective identification of predictive model

Input:
Signal transduction dynamics

Output:
mRNA expression dynamics

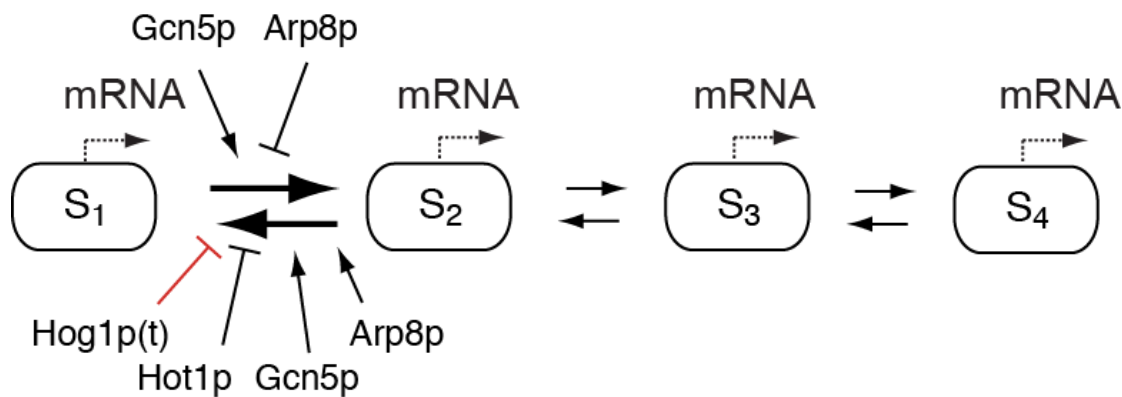


Model identification:

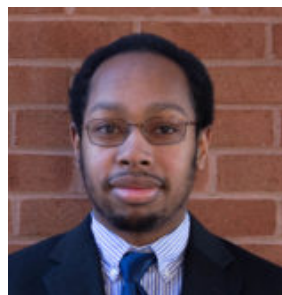
- Number of states
- Biologically meaningful rate values
- Type of regulation (activating / repressing / switch / gradual)

Munsky et al., JCP, 2006; Neuert, Munsky, et al., Science 2013;
Munsky, Fox, Neuert, Methods, 2015; Fox, Neuert, Munsky, JCP, 2016

Chromatin mutants modulate specific rates



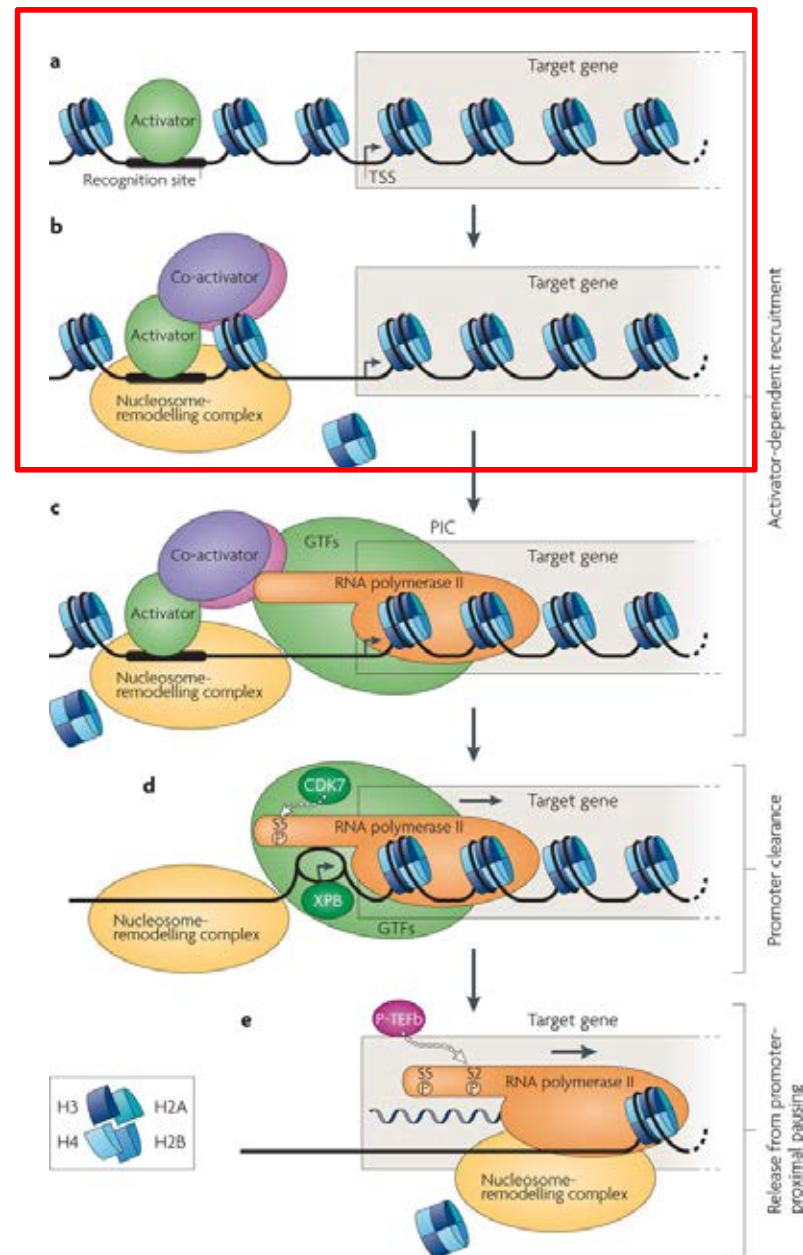
Ben Kesler



Jason Hughes



Rohit Venkat



Are **single cell** experiments
more predictive than cell
population experiments?



Dr. Brian Munsky
Assistant Professor
Colorado State University

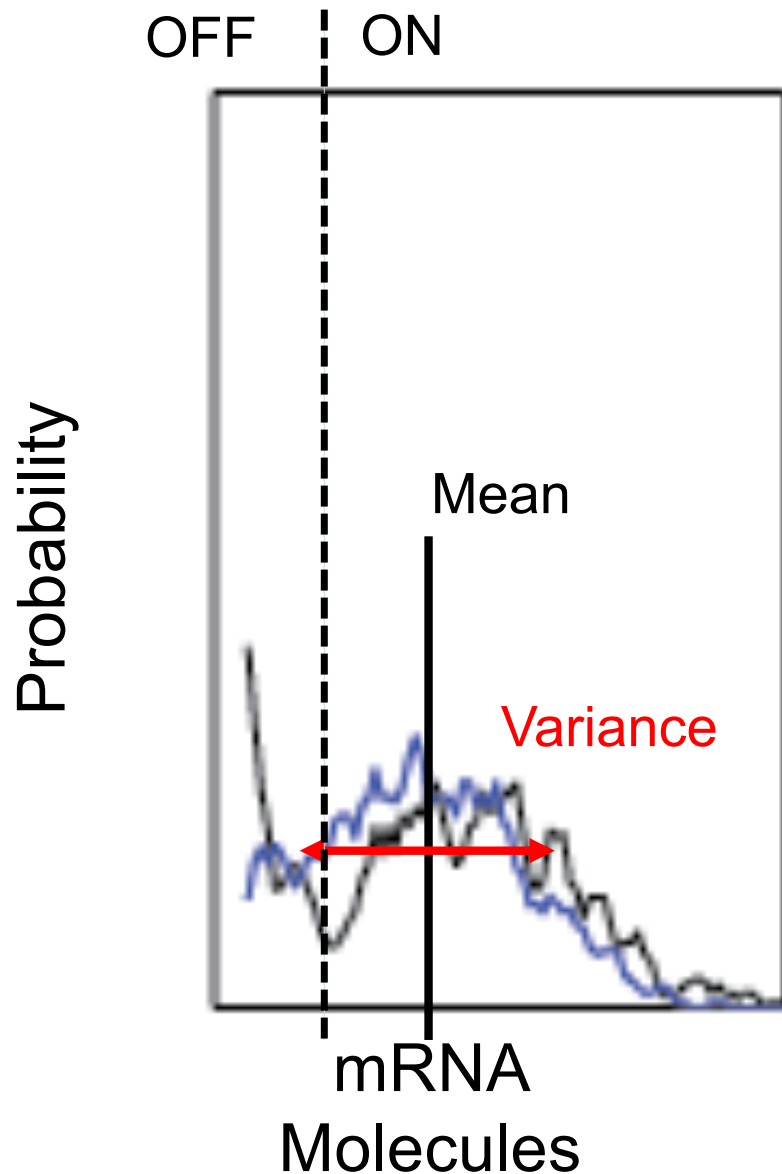


Zachary Fox
Graduate Student
Colorado State University

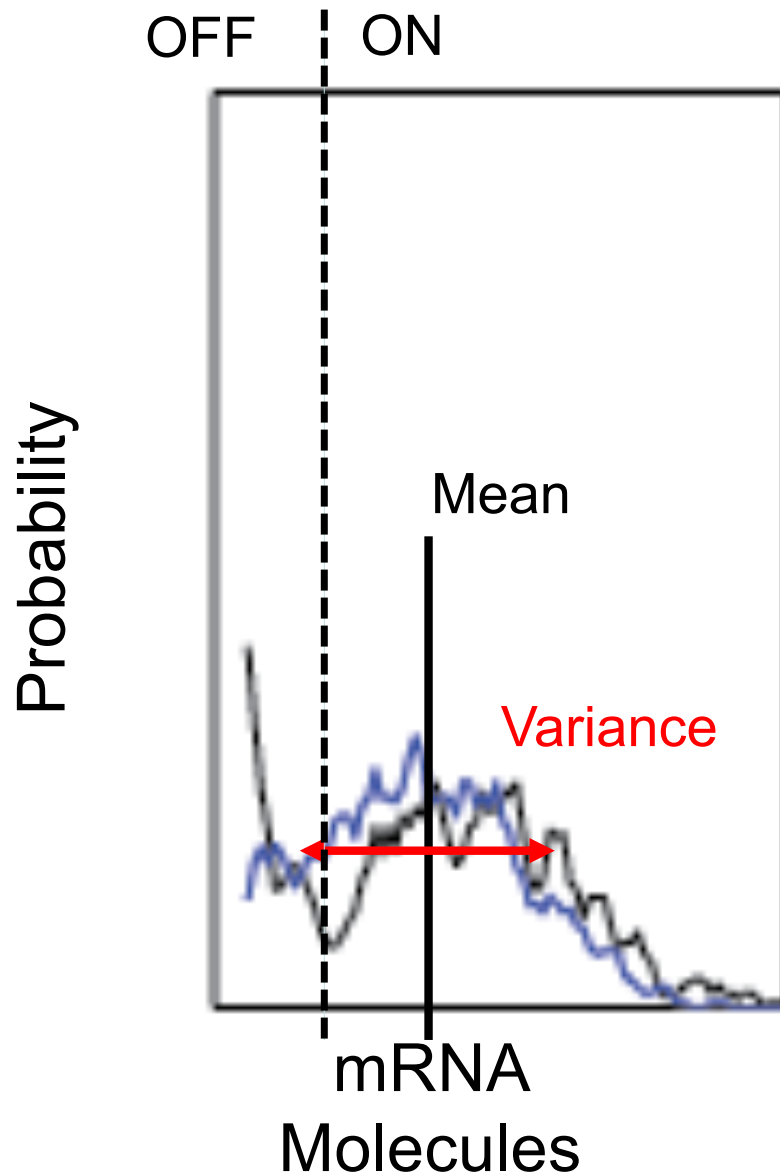


Guoliang Li,
Post-doctoral fellow
Vanderbilt University

Single cell data contains information on the mean, variance, fraction of cells and full distribution



Single cell data contains information on the mean, variance, fraction of cells and full distribution

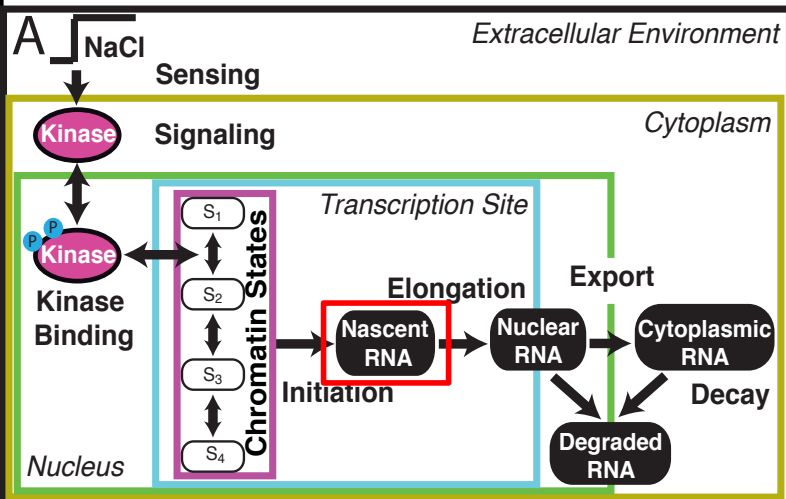
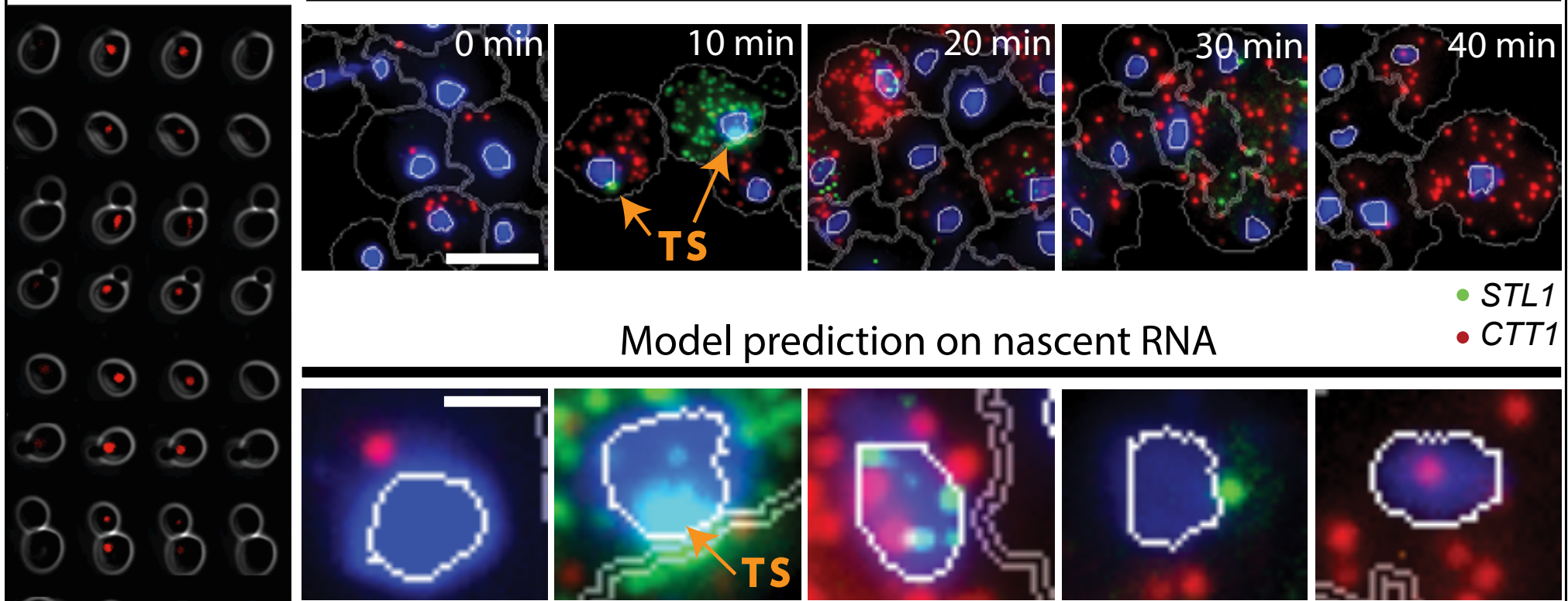


Same Experiments	Same Model	Modeling Assumption
Mean	1. Moment of the distribution	Central Limit Theorem needs to be fulfilled
Variance	2. Moment of the distribution	
ON-Fraction	From FSP	No assumption on the shape of the distribution
Distribution	FSP	

Single cell experimental and modeling framework

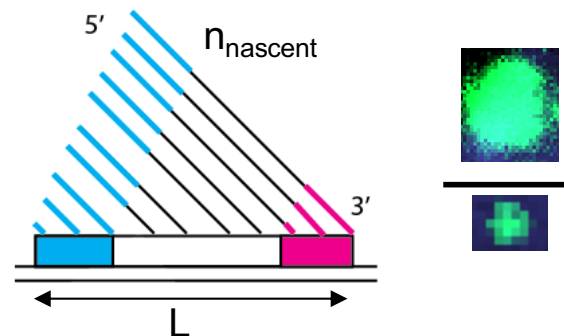
Hog1 Time →

Model fitting on nuclear and cytoplasmic RNA



Transcription site =
Number of nascent transcripts

$$\left\langle \frac{I_{TS}}{I_{mature}} \right\rangle = \langle n_{nascent} \rangle = \frac{k_{i\max}L}{2k_{elong}}$$



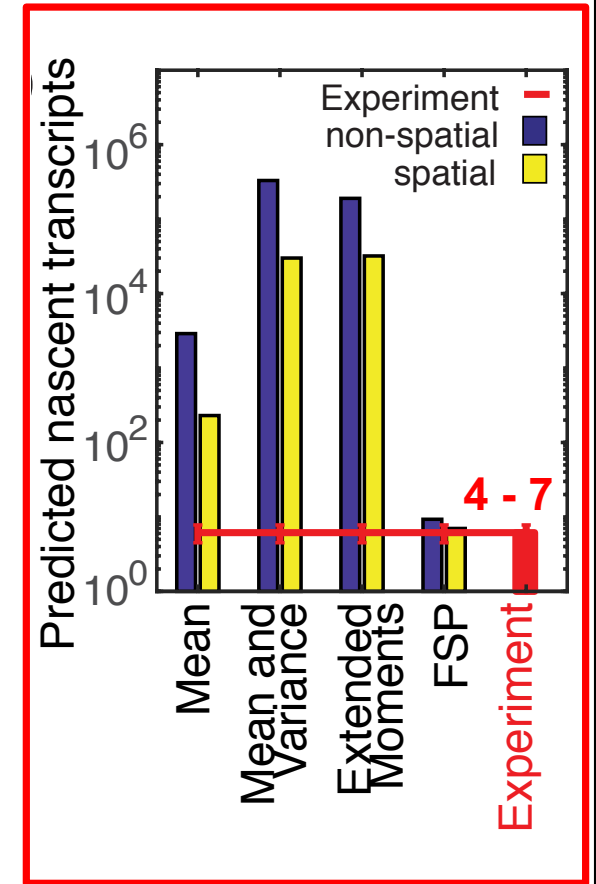
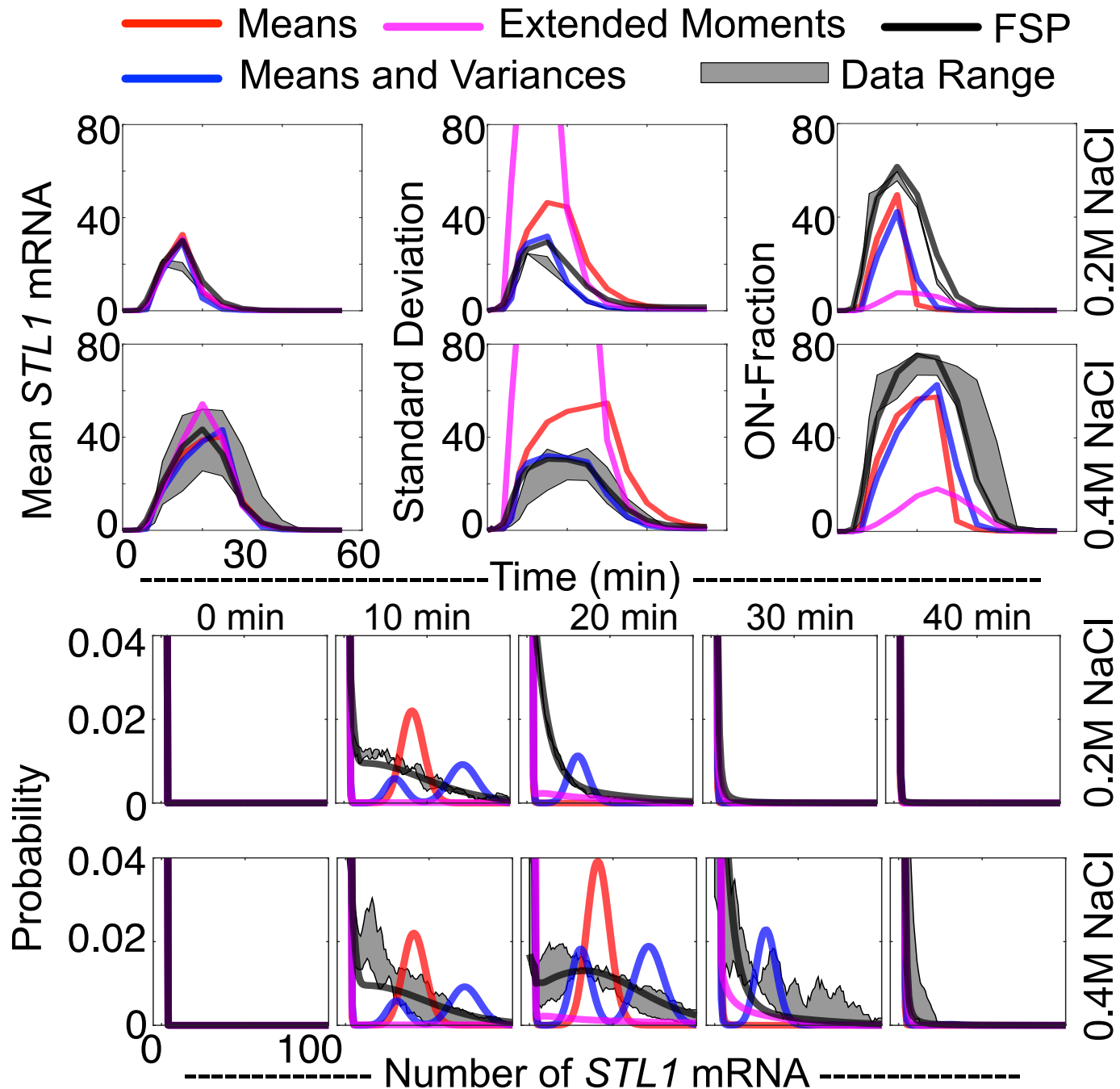
Use CTT1 data to get k_{elong}

$$k_{elong} = 91 \pm 9 \text{ (simplified)}$$

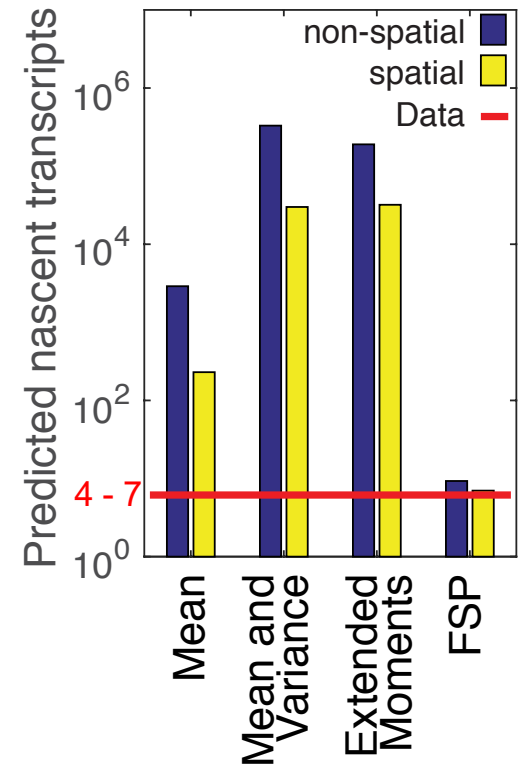
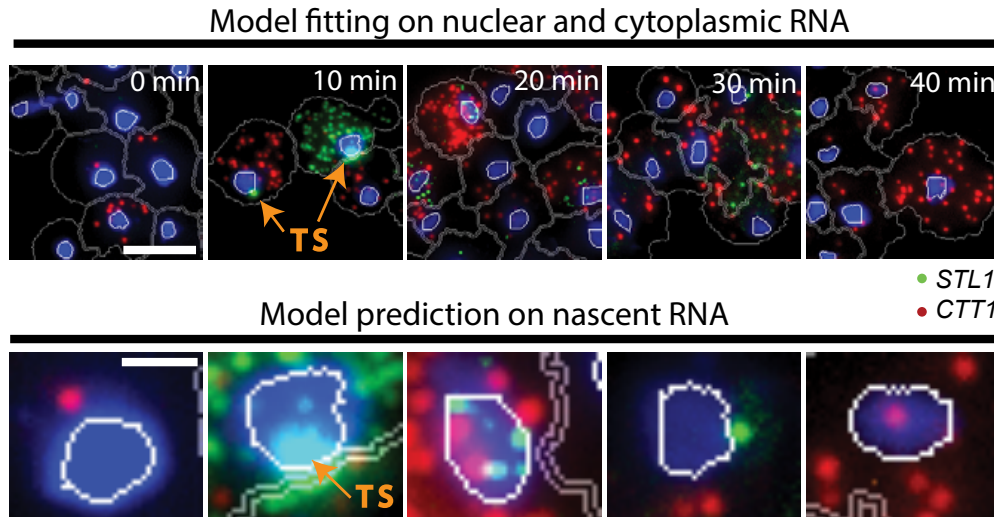
$$k_{elong} = 63 \pm 13 \text{ (full model)}$$

Munsky et al., PNAS, 2018

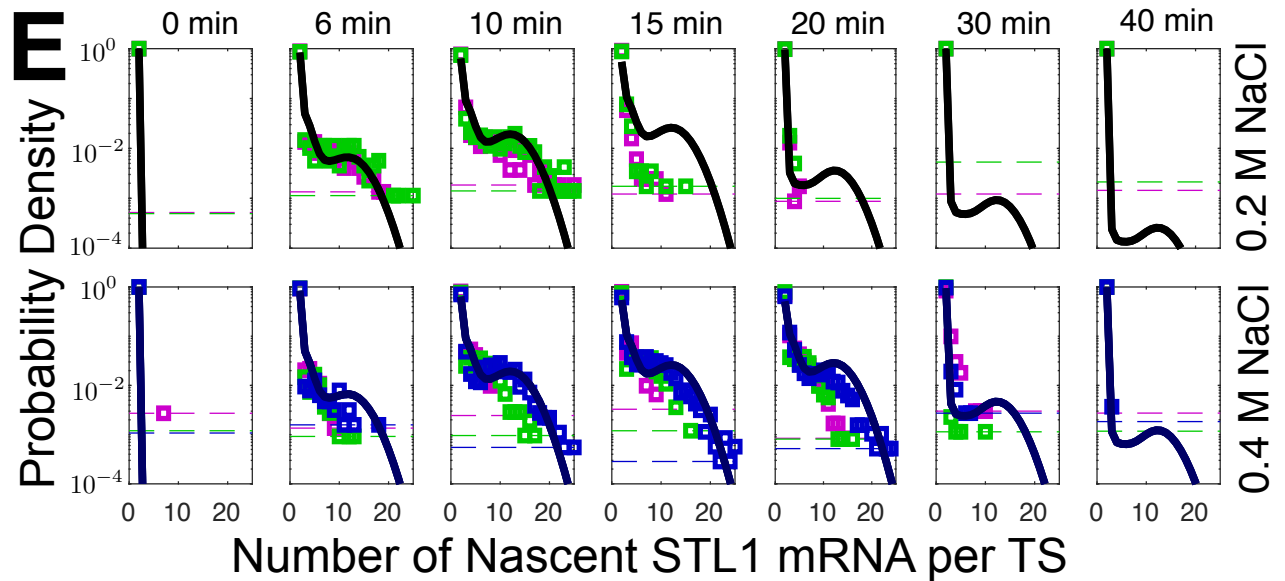
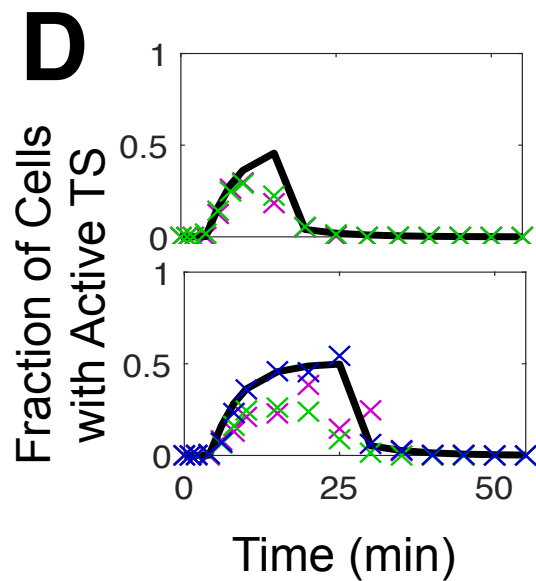
Good fits do not result in good predictions



Fitted distribution result in superior predictions of nascent transcription



Predictions for *STL1* transcription site activity versus time and NaCl.



Inferred parameters are highly reproducible

Parameter comparison (within 8%): Degradation rate

$\gamma_{STL1} = 4.9E-3$ 1/s (Neuert et al. Science 2013)

$\gamma_{STL1} = 5.3E-3$ 1/s (Munsky et al., PNAS, 2018)

$\gamma_{CTT1} = 2.0E-3$ 1/s (Neuert et al. Science 2013)

$\gamma_{CTT1} = 2.1E-3$ 1/s (Munsky et al., PNAS, 2018)

Based on:

New geographic location

New lab

New microscope

Different yeast strain

New reagents

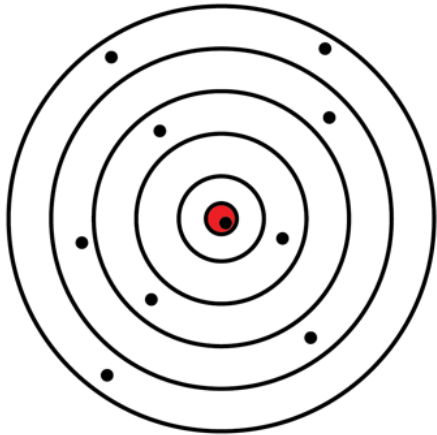
New person doing the experiment

Improved image processing software

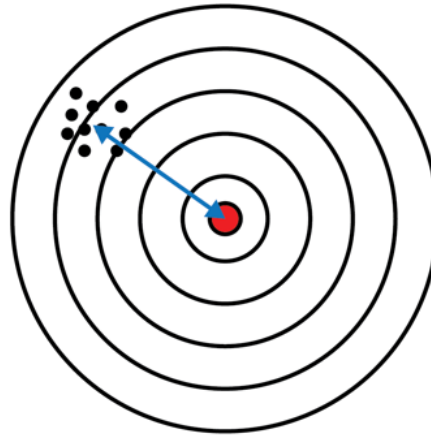
Why are **single cell** experiments
more predictive than cell
population experiments?

Quantify contribution to bias and uncertainty of parameter estimation in modeling

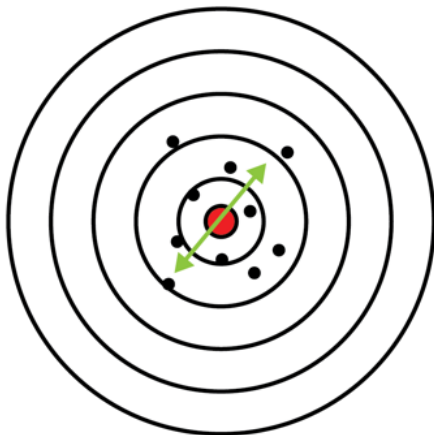
Not identifiable



Biased less uncertain

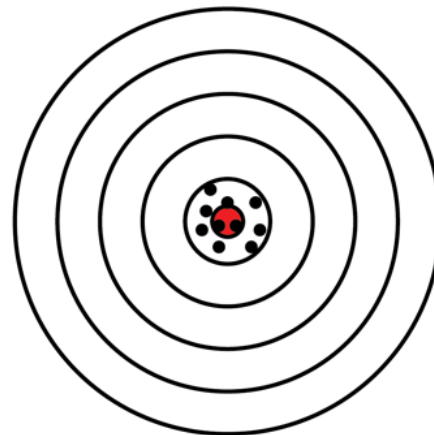


Unbiased uncertain



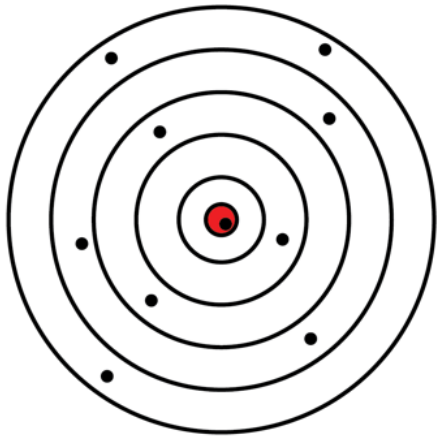
bias
uncertainty

Unbiased less uncertain

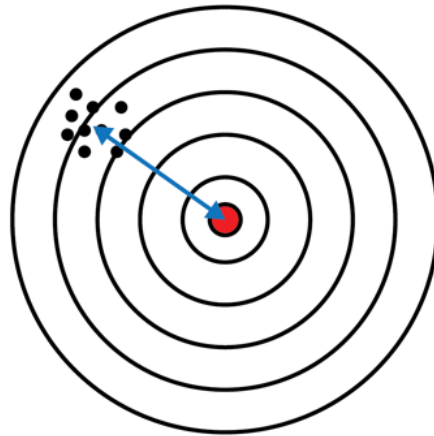


Quantify contribution to bias and uncertainty of parameter estimation for different models / data

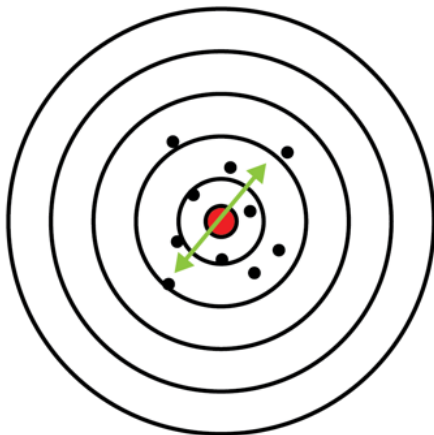
Not identifiable



Biased less uncertain

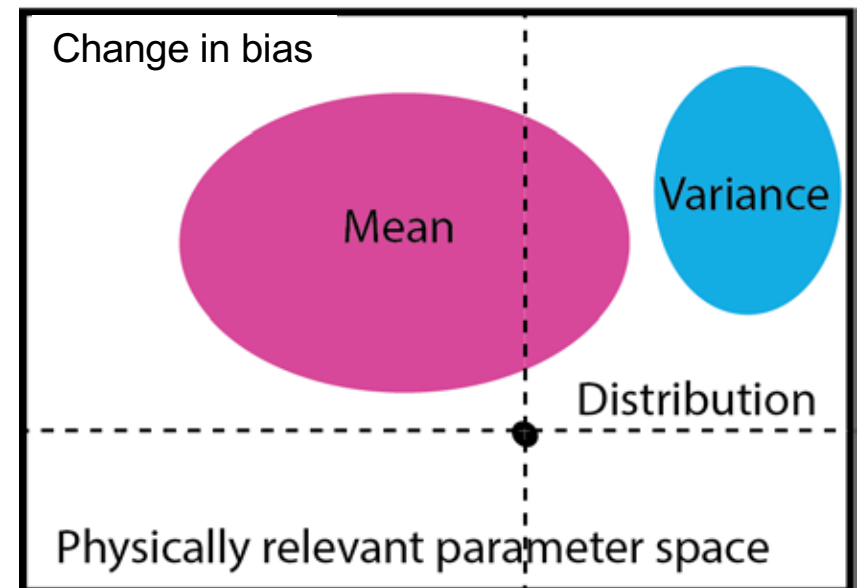
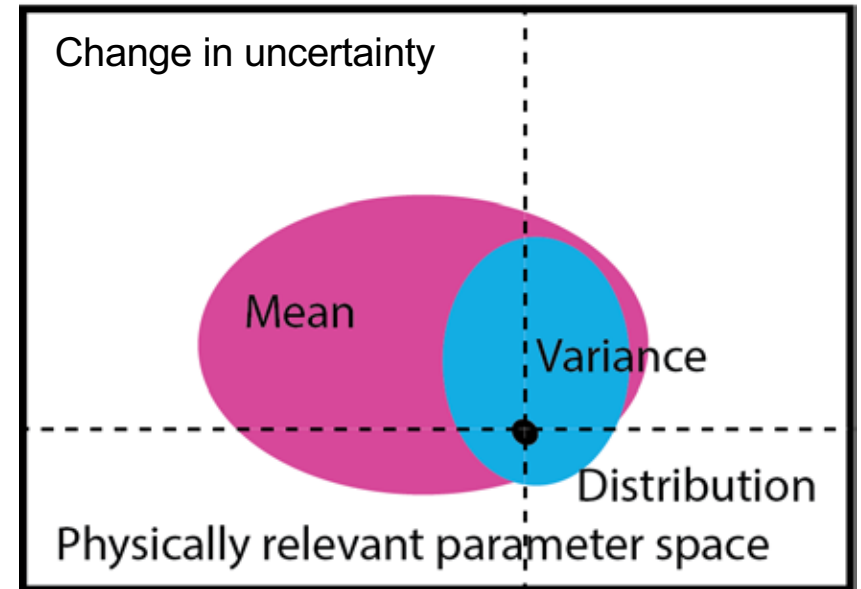
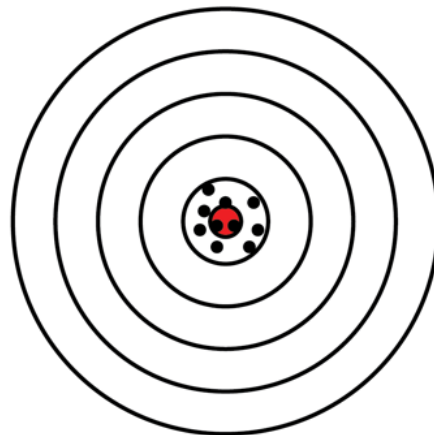


Unbiased uncertain



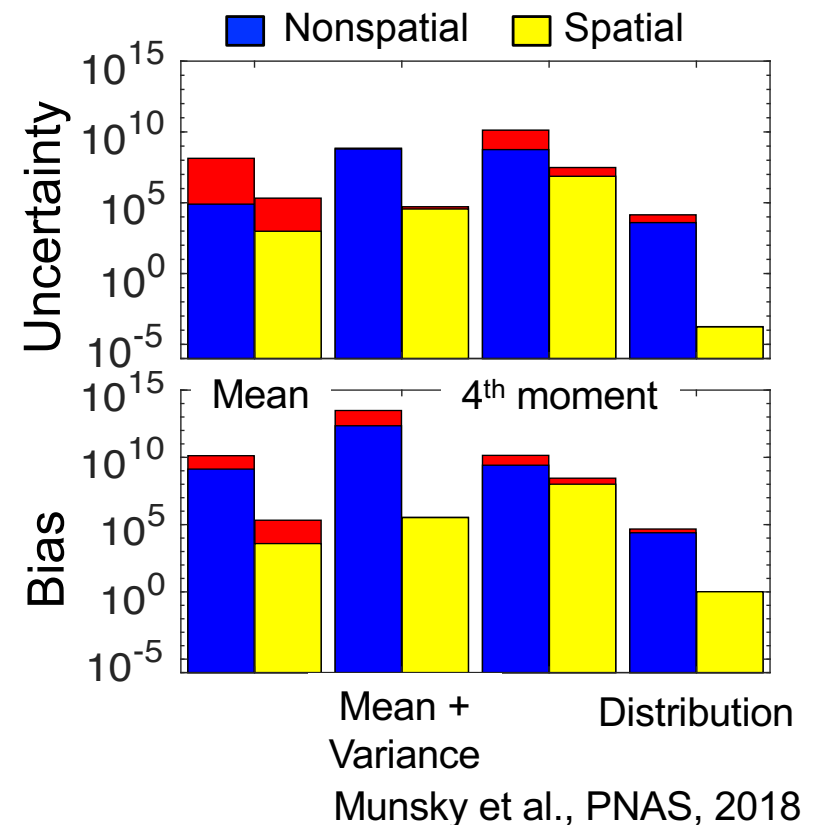
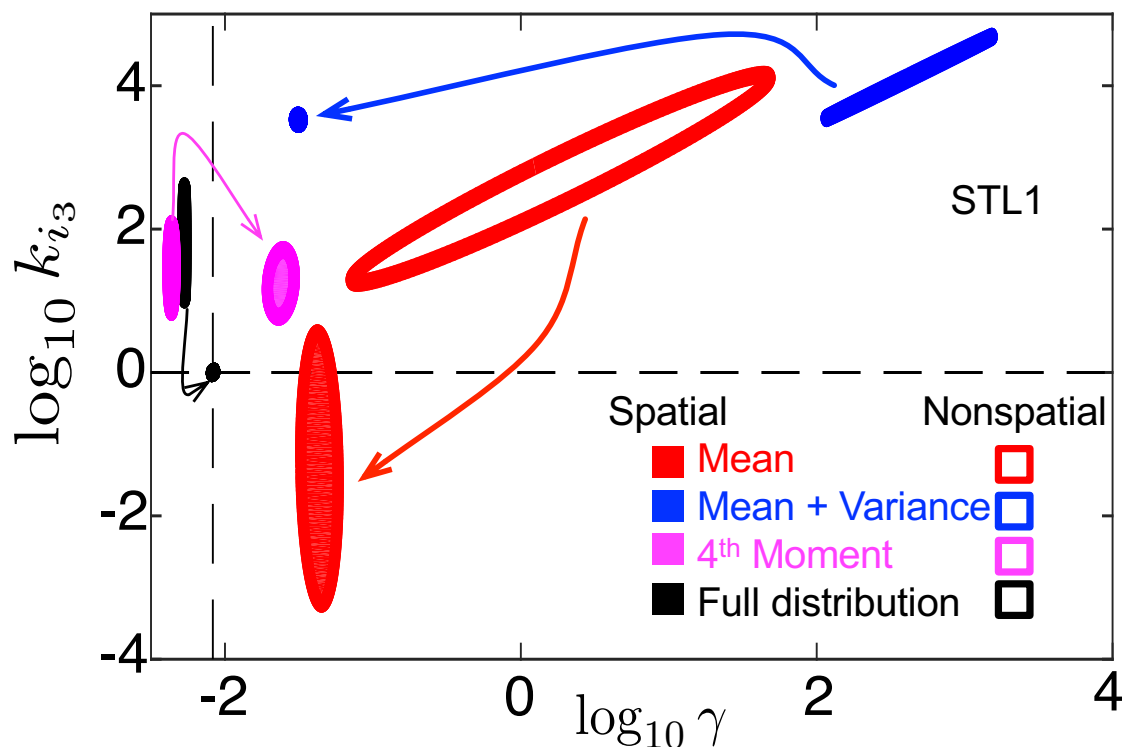
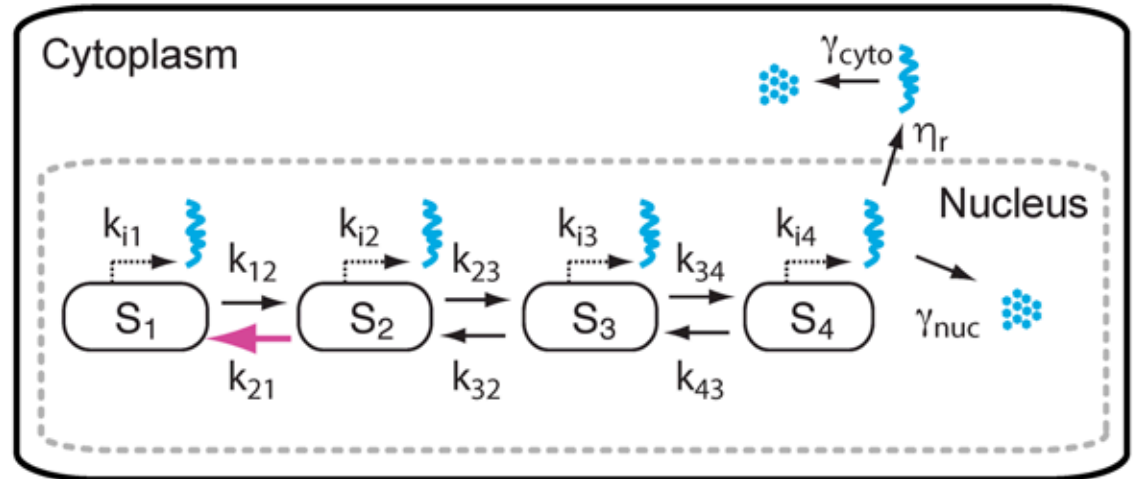
bias
uncertainty

Unbiased less uncertain



Distributions reduce bias and uncertainty

Using **Markov Chain Monte Carlo** simulation with **Metropolis Hastings** sampling to estimate **parameter bias** (distance from true parameters) and **parameter uncertainty** (width of parameter distribution)

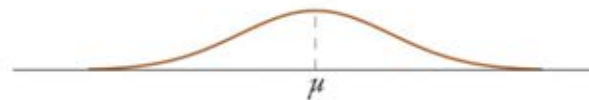


Why is there such a **strong bias**
in parameters using **moments**
instead of distributions?

Using moments requires that the Central Limit Theorem is fulfilled

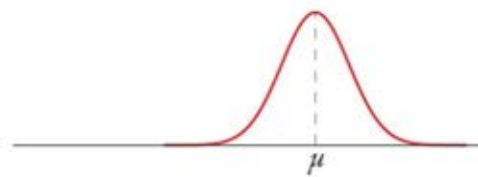
The central limit theorem (CLT) states that if one makes sufficiently many quantitative observations from the same underlying distribution then the average of those observations would be normally distributed with a deviation given by the standard error of the mean.

Population
distribution



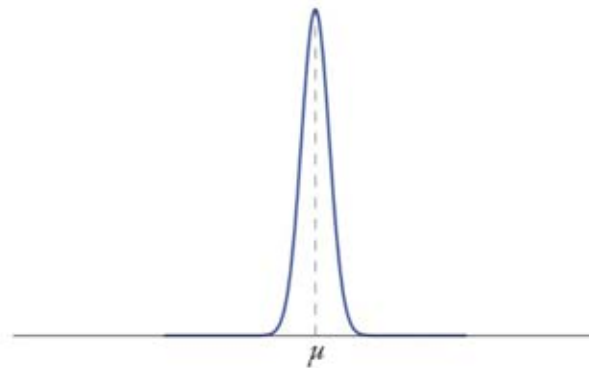
Number of proteins

Sampling distribution
of the **mean from
5 measurements**



Mean number
of proteins

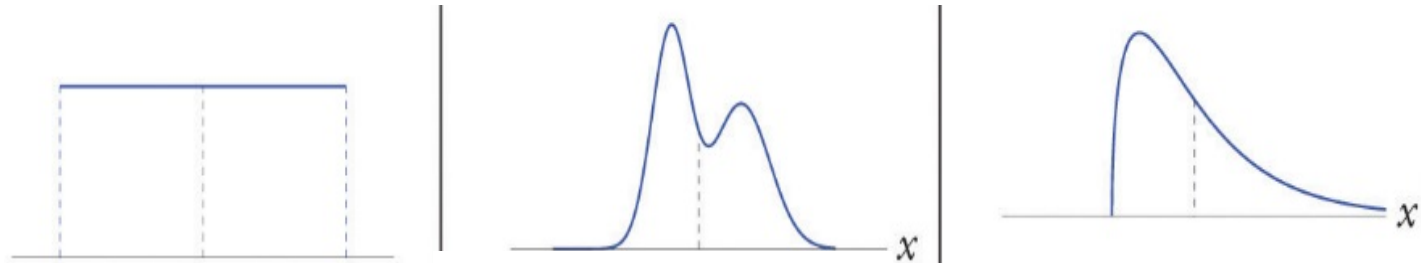
Sampling distribution
of the **mean from
30 measurements**



Mean number
of proteins

Nonsymmetrical distributions require more measurements to satisfy the CLT

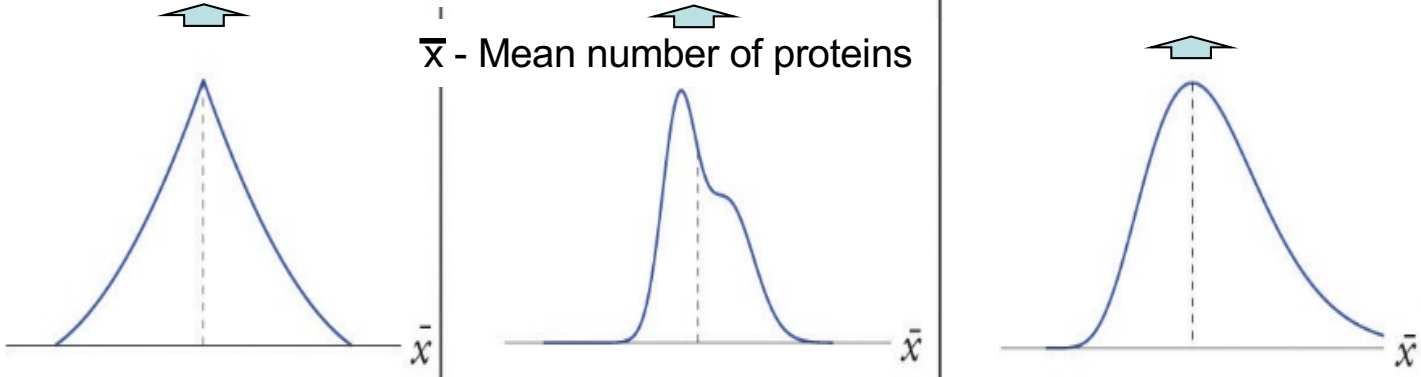
Population distribution



x - Number of proteins

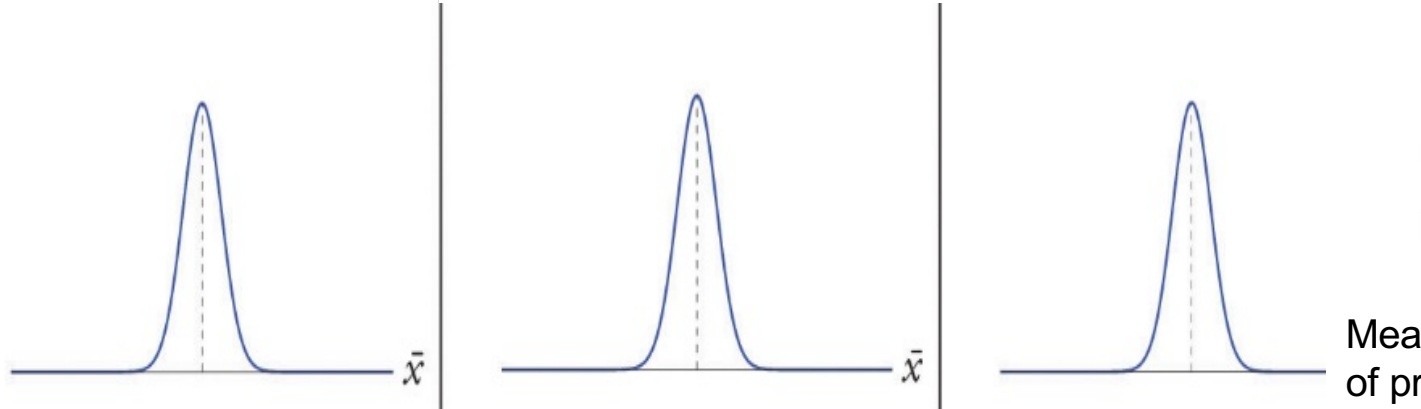
\bar{x} - Mean number of proteins

Sampling distribution of the mean \bar{x} from 5 measurements



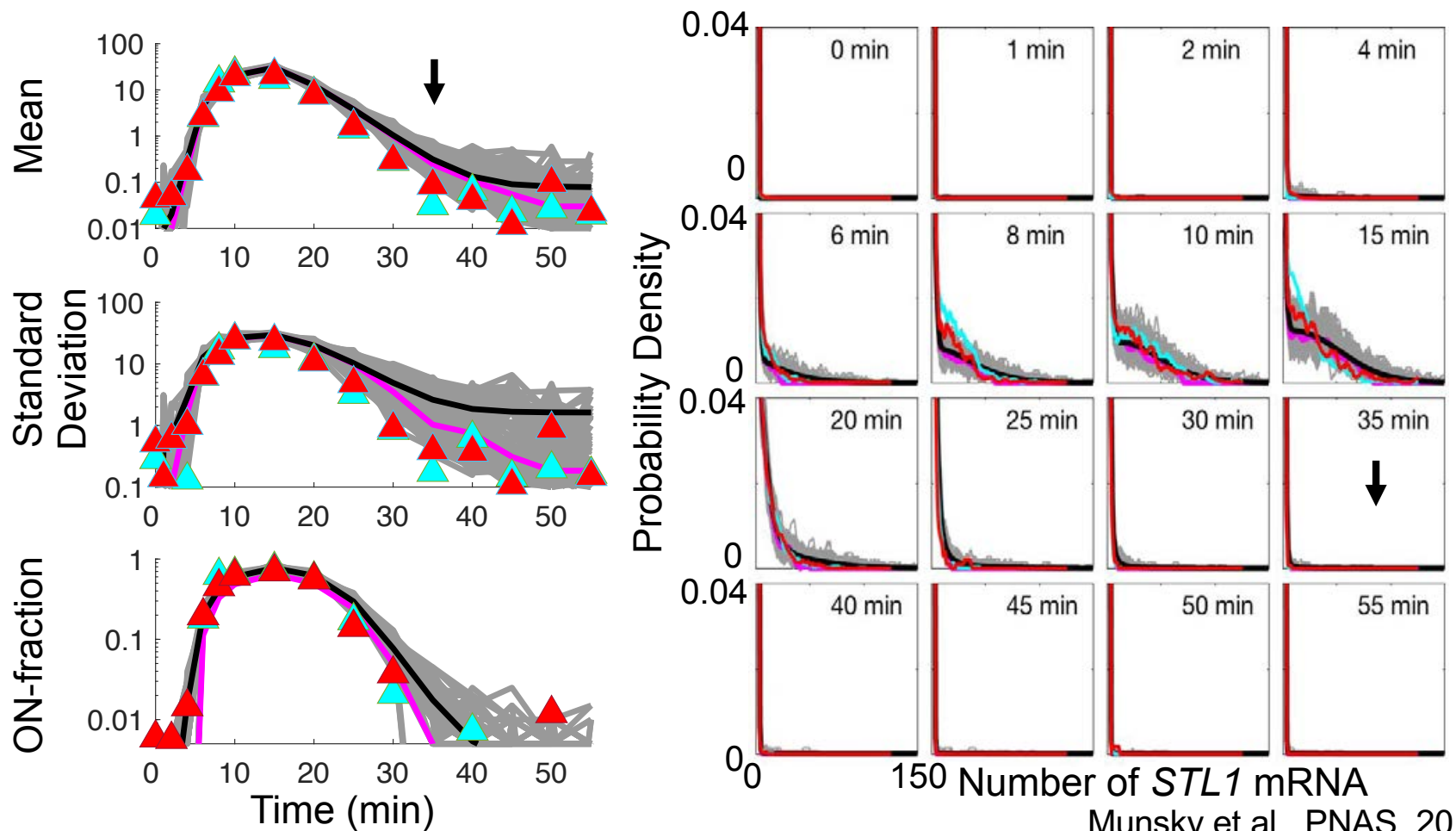
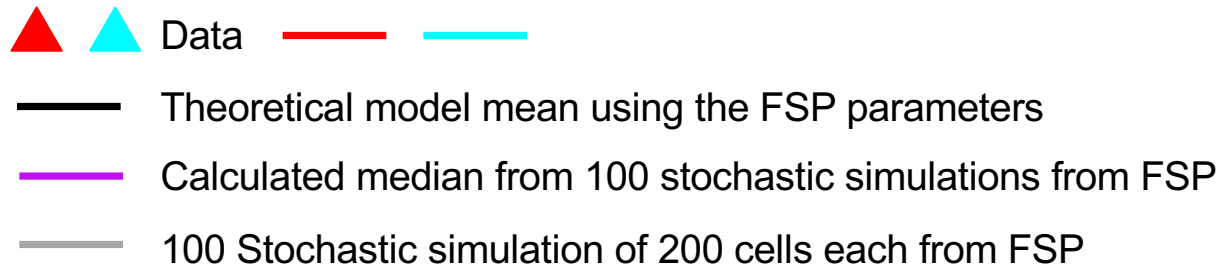
\bar{x} - Mean number of proteins

Sampling distribution of the mean from 30 measurements



Mean number of proteins

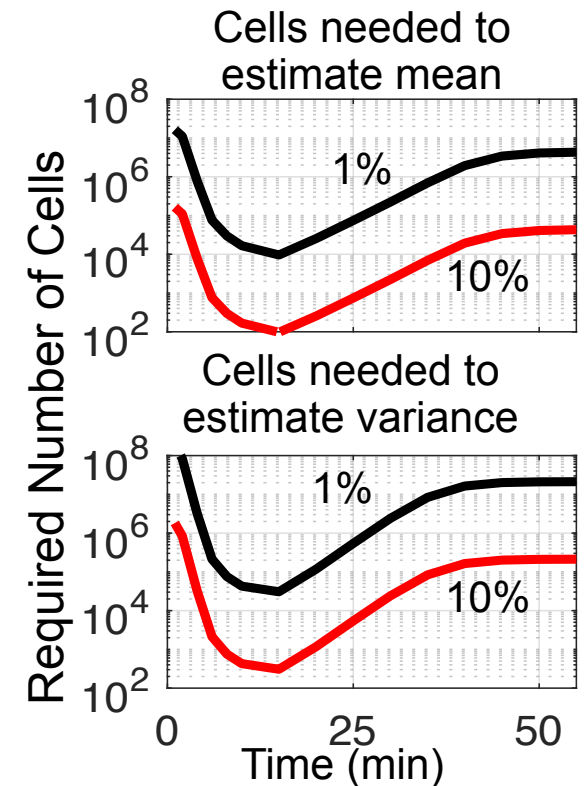
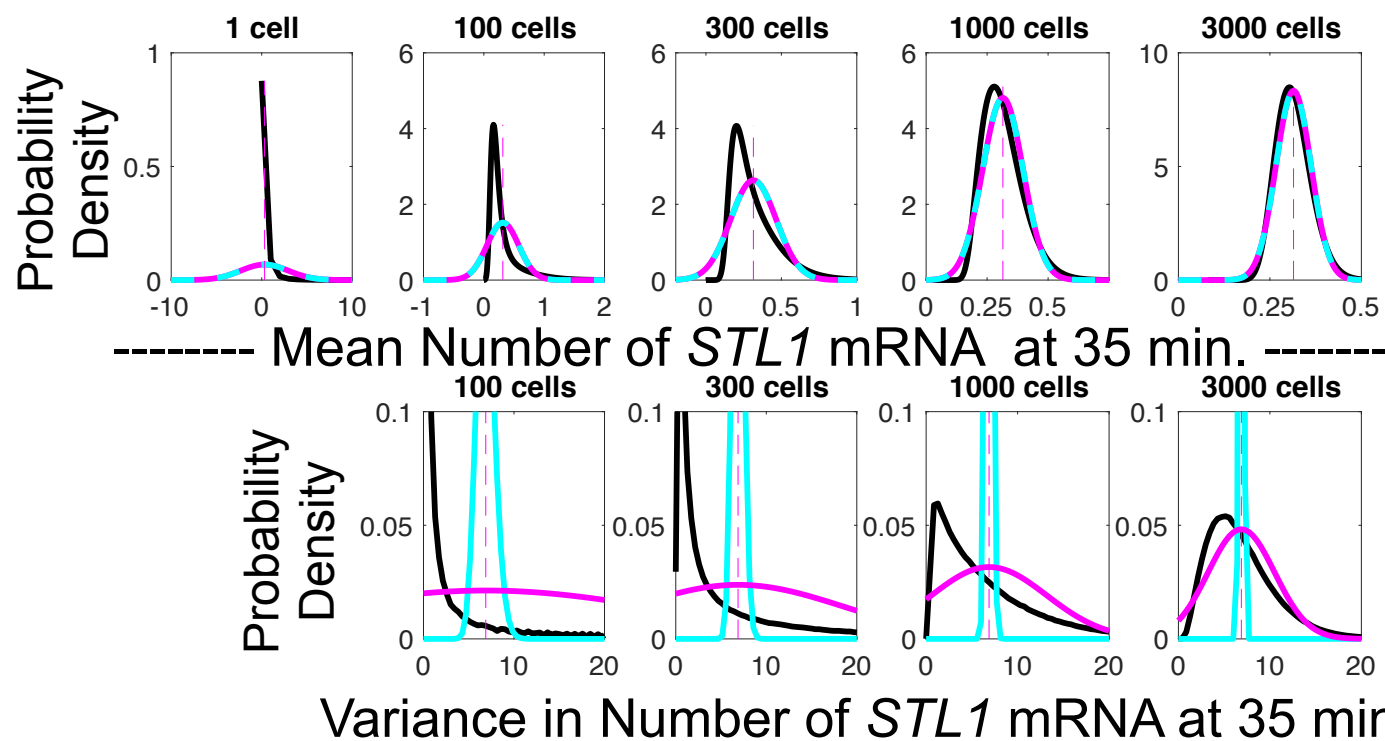
Nonsymmetrical distributions are a result of high expressing cells violating the CLT



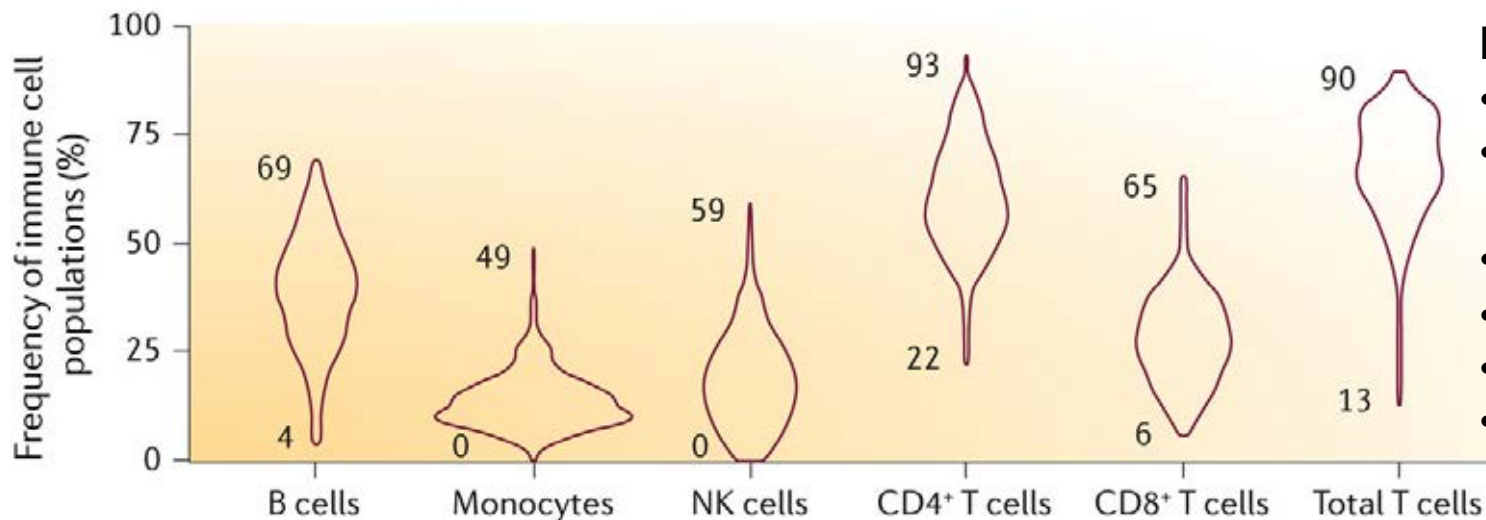
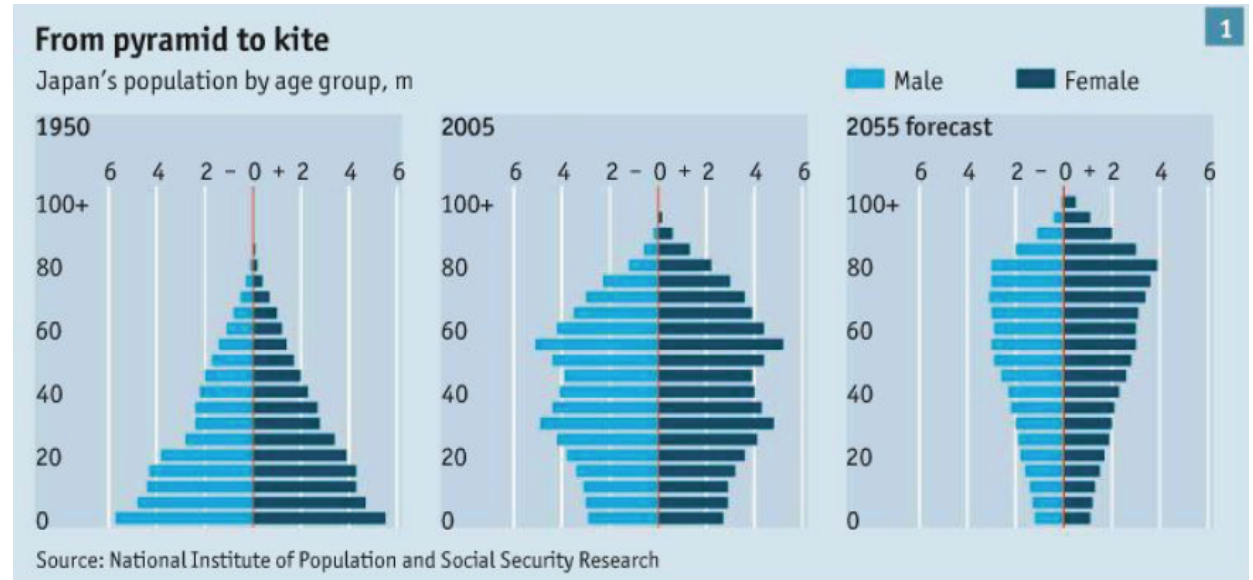
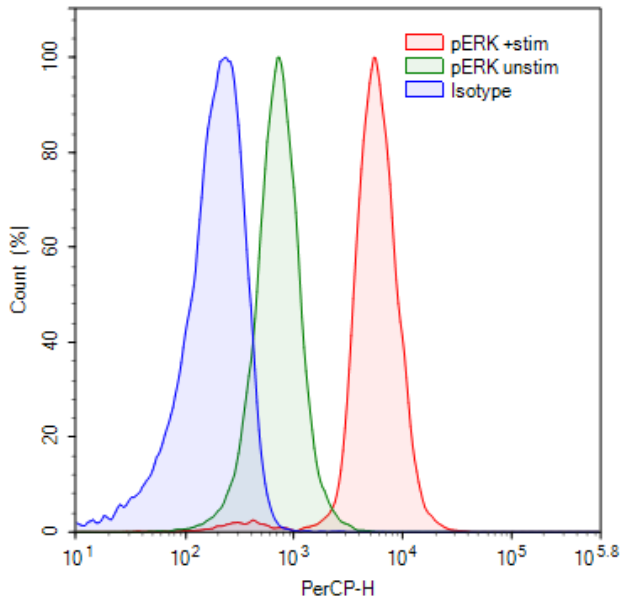
Nonsymmetrical distributions due to outlier cells require very large number of cells to fulfill the Central Limit Theorem (CLT), which is a requirement to apply mean, variance or higher moment approximations

Gaussian approximation
Extended moment analysis
Full distribution

Munsky et al., PNAS, 2018



General implications for identification of parametric model from any positive nonsymmetricly distributed data set



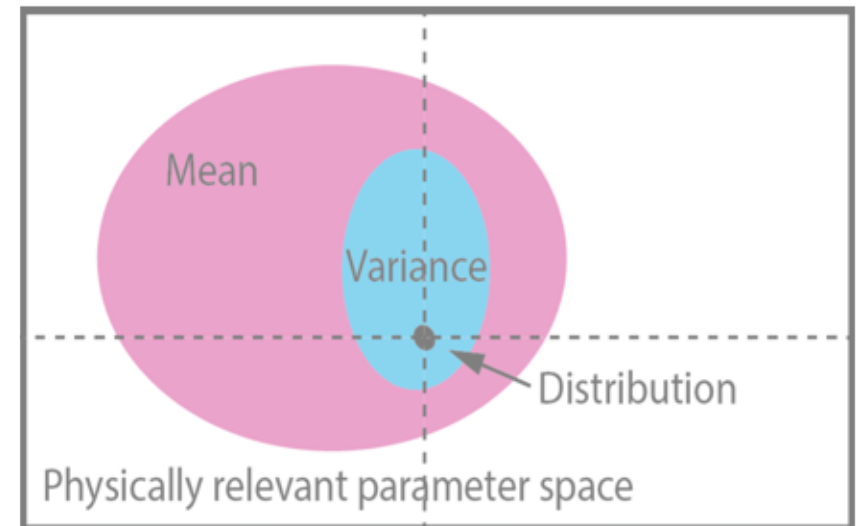
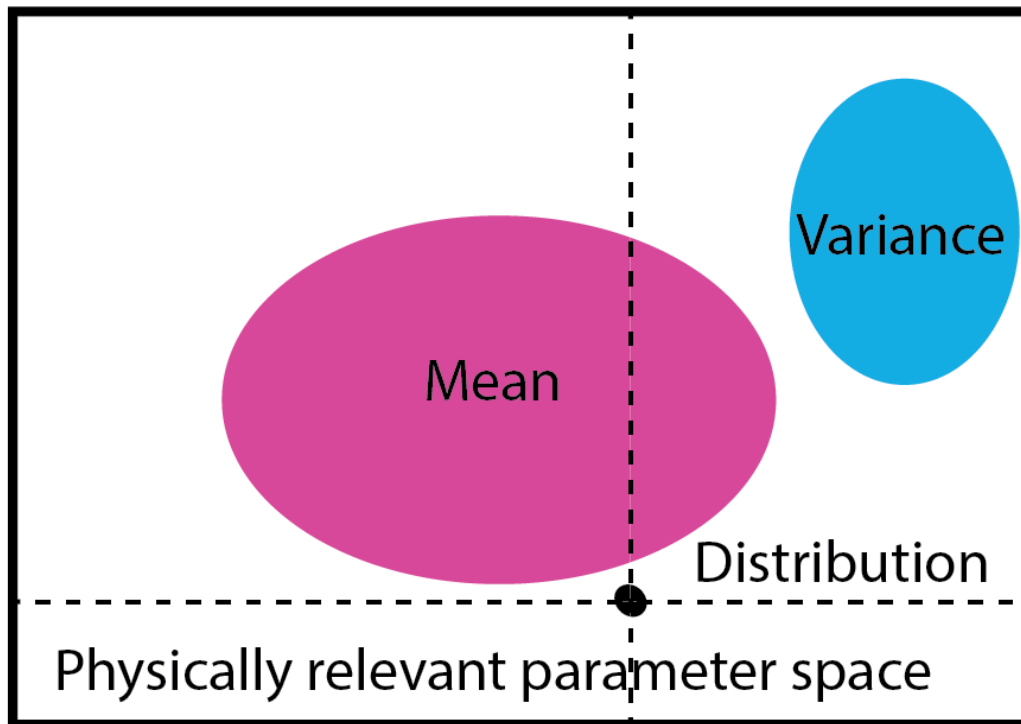
High expressing cells:

- Mammalian cells
- High expressing rare cells
- Mutant cells
- Rare cell type
- Different genotype
- Resistant individual

Take home message

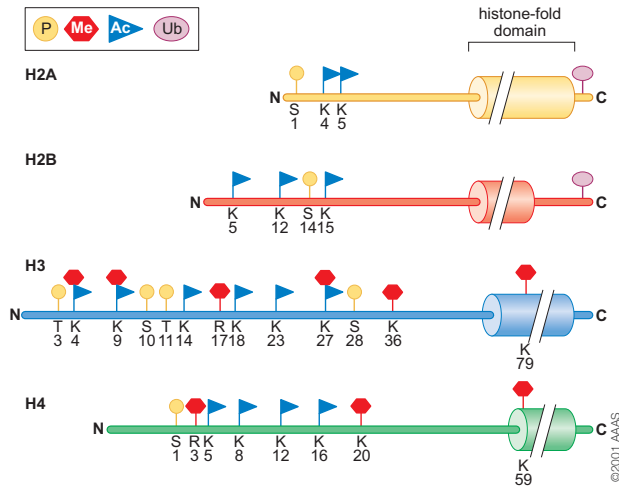
Model inference from **means or variances** result in **biased parameter sets** and **poor predictions** because **data is positive and non-symmetrical distributed** and therefore **modeling assumptions are violated**

and not only because of over fitting

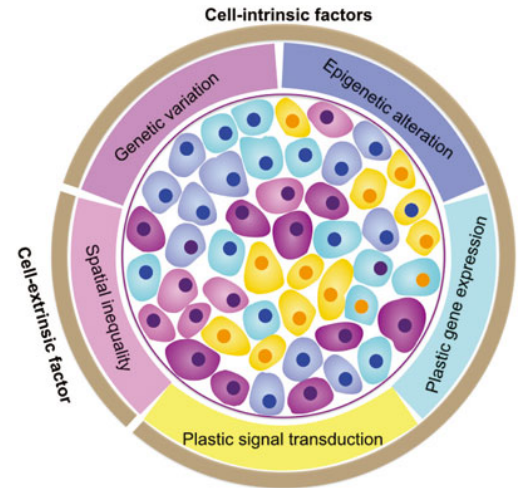


Biology is governed by dynamic processes

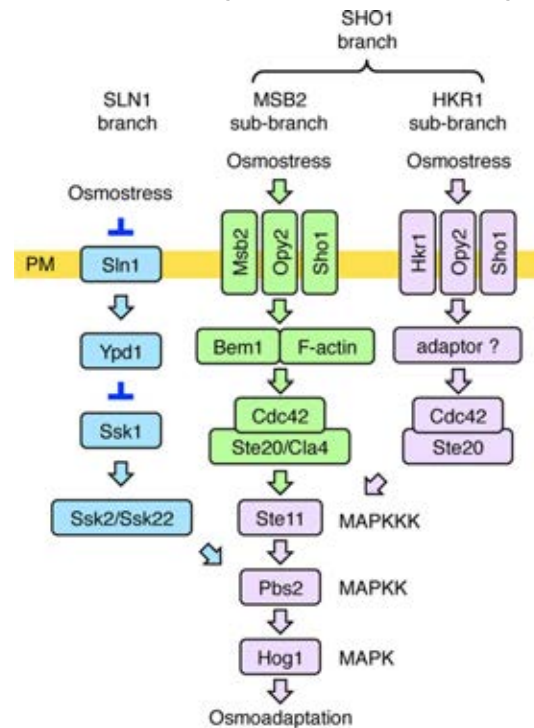
Histone code



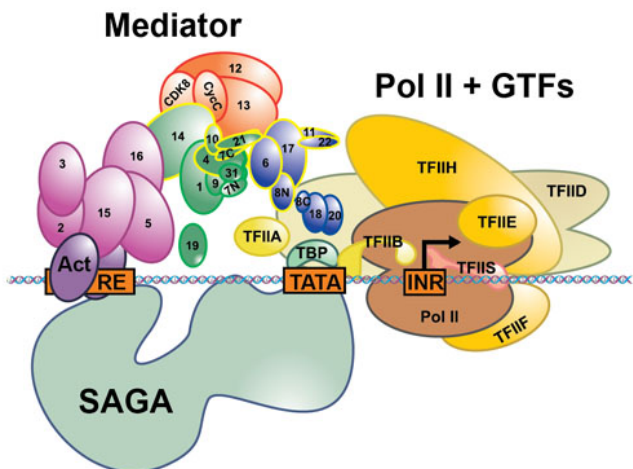
Disease mutations / Cell types







Pathway complexity



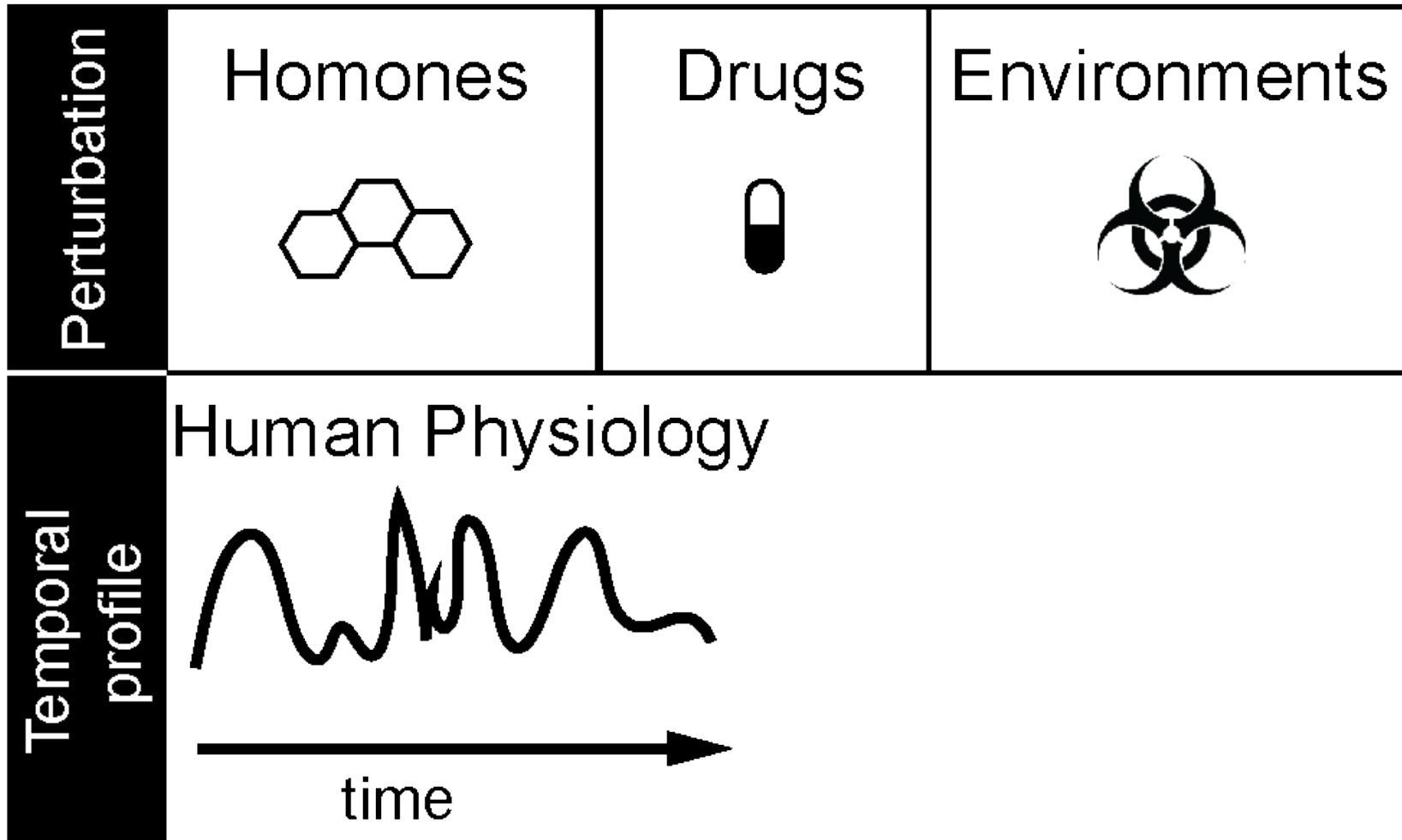
Protein complexes



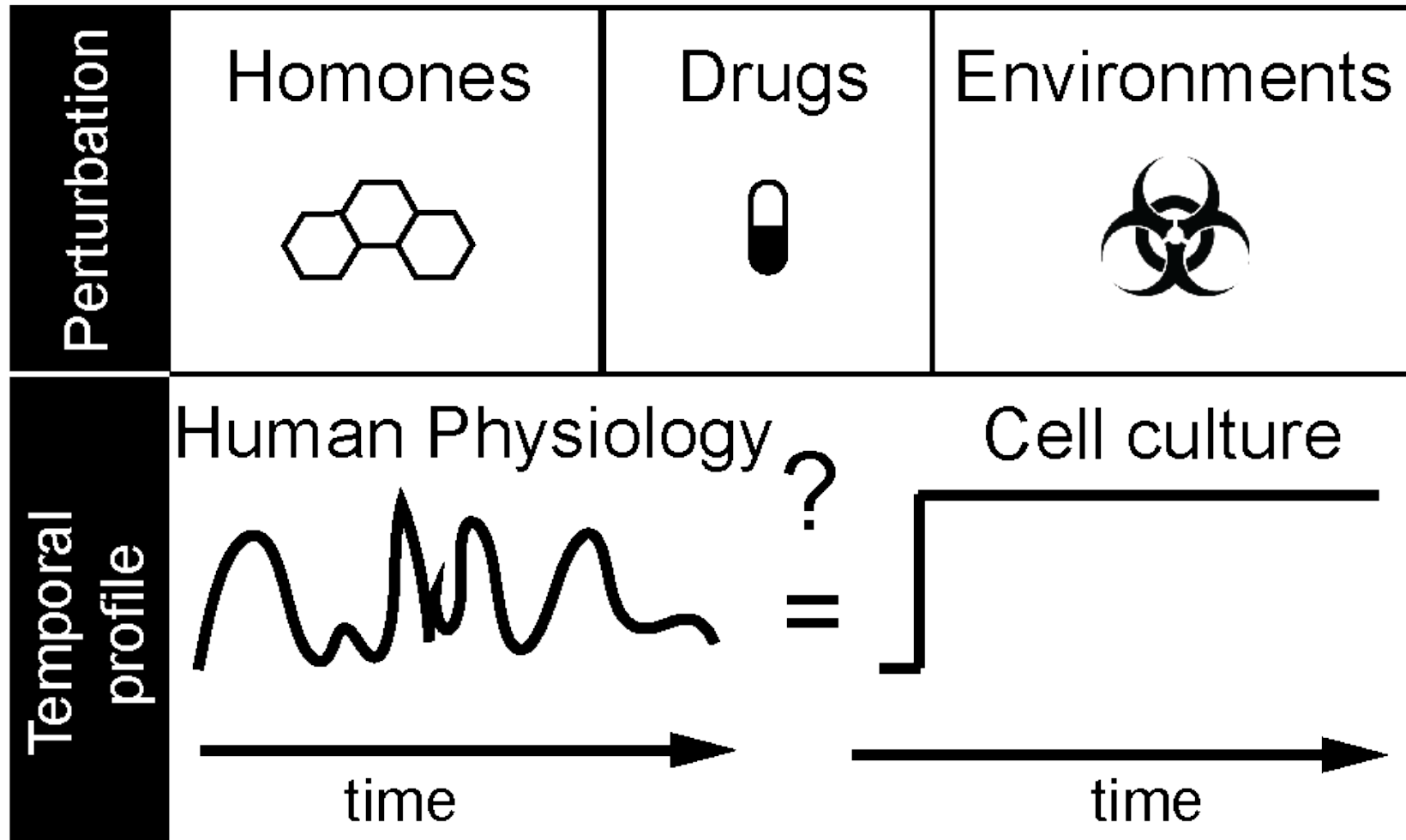
Genotype / Phenotype

	A	A ^y
A	Agouti coat AA 	Yellow Coat AA ^y 
A ^y	Yellow coat AA ^y 	Dead A ^y A ^y 

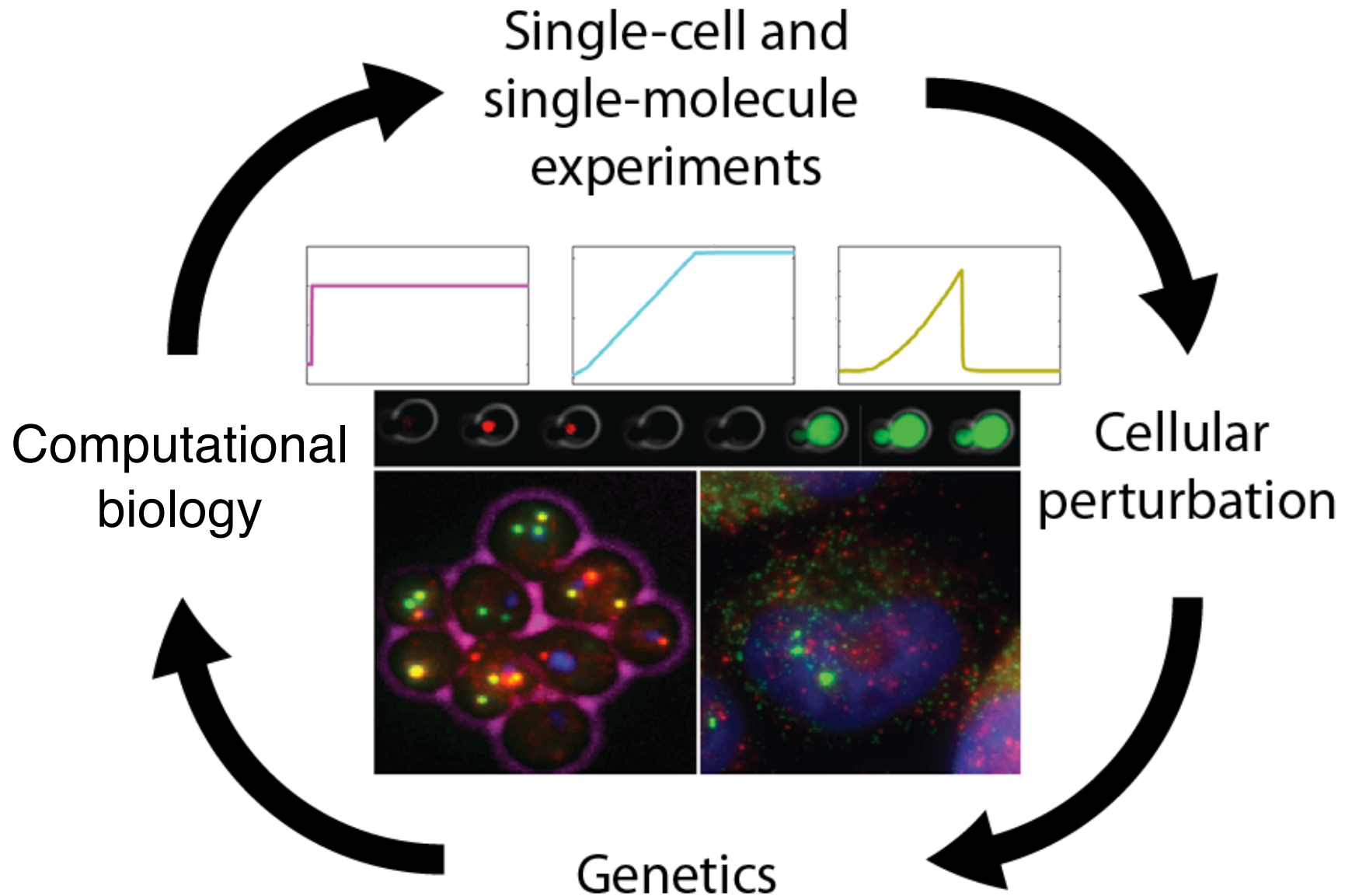
Cell environments in humans change kinetically



Majority of current biomedical research focuses on static environments



Multidisciplinary approach to study dynamic cell signaling and gene regulation



Are **biological mechanisms**
and phenotypes specific to
kinetic environments?

How do **human cells** respond to
kinetic environments?

Why are **kinetic environments**
more informative to built **predictive**
models?

Acknowledgements

Post-docs



Hossein Jashnsaz
Poster # 14

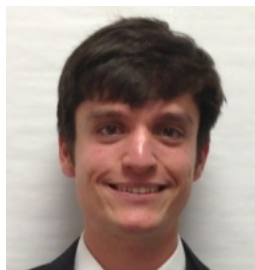


Amanda Johnson
Poster # 8



Guoliang Li
Now Assistant
Professor

Graduate students



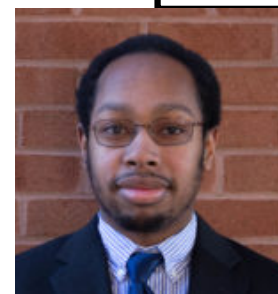
Ben Kesler



Alexander Thiemicke



Rohit Venkat



Jason Hughes



Joseph Cleland

Poster # 17, 19

In collaboration with:

Colorado State University



Brian Munsky
Assistant Professor



Zachary Fox
Graduate student

Funding



NIH DIRECTOR'S
NEW INNOVATOR
AWARD



VANDERBILT
UNIVERSITY

DP2 GM114849

R01 GM11589203

U2C CA233291

AHA, VICC

5T32GM008320-28 (MBTG)

5T32GM008554-21 (CBMS)

5T32NS007491-16 (TPICTP)

1T32LM012412-01 (BIDS)

5T32DK101003-05 (ITED)