Analytical Solutions to Chemical Master Equations

Abhyudai Singh

Assistant Professor

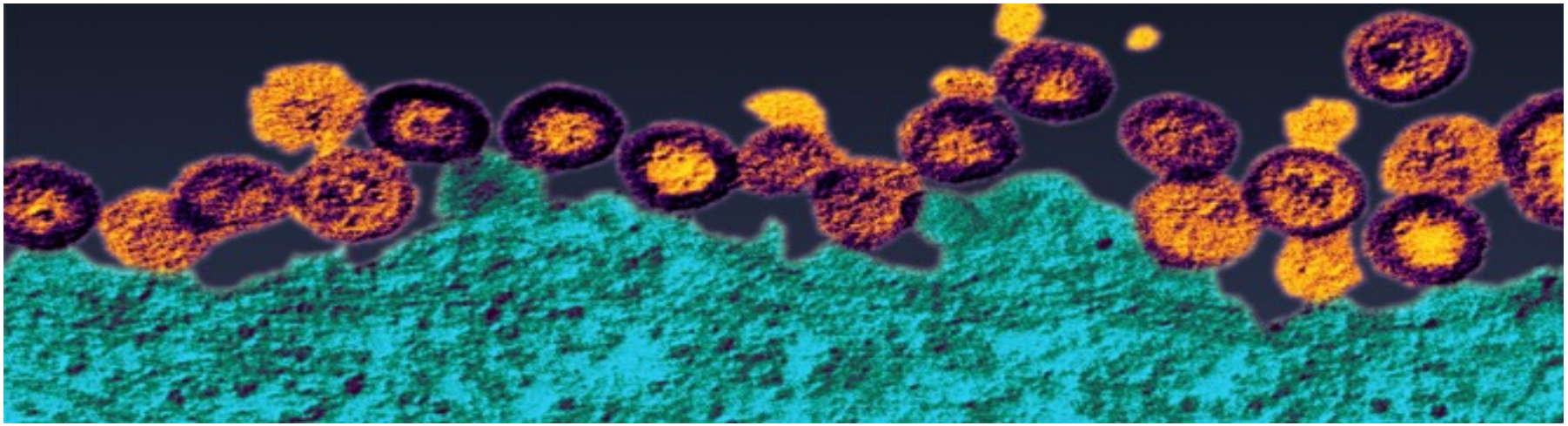
Electrical and Computer Engineering

Biomedical Engineering

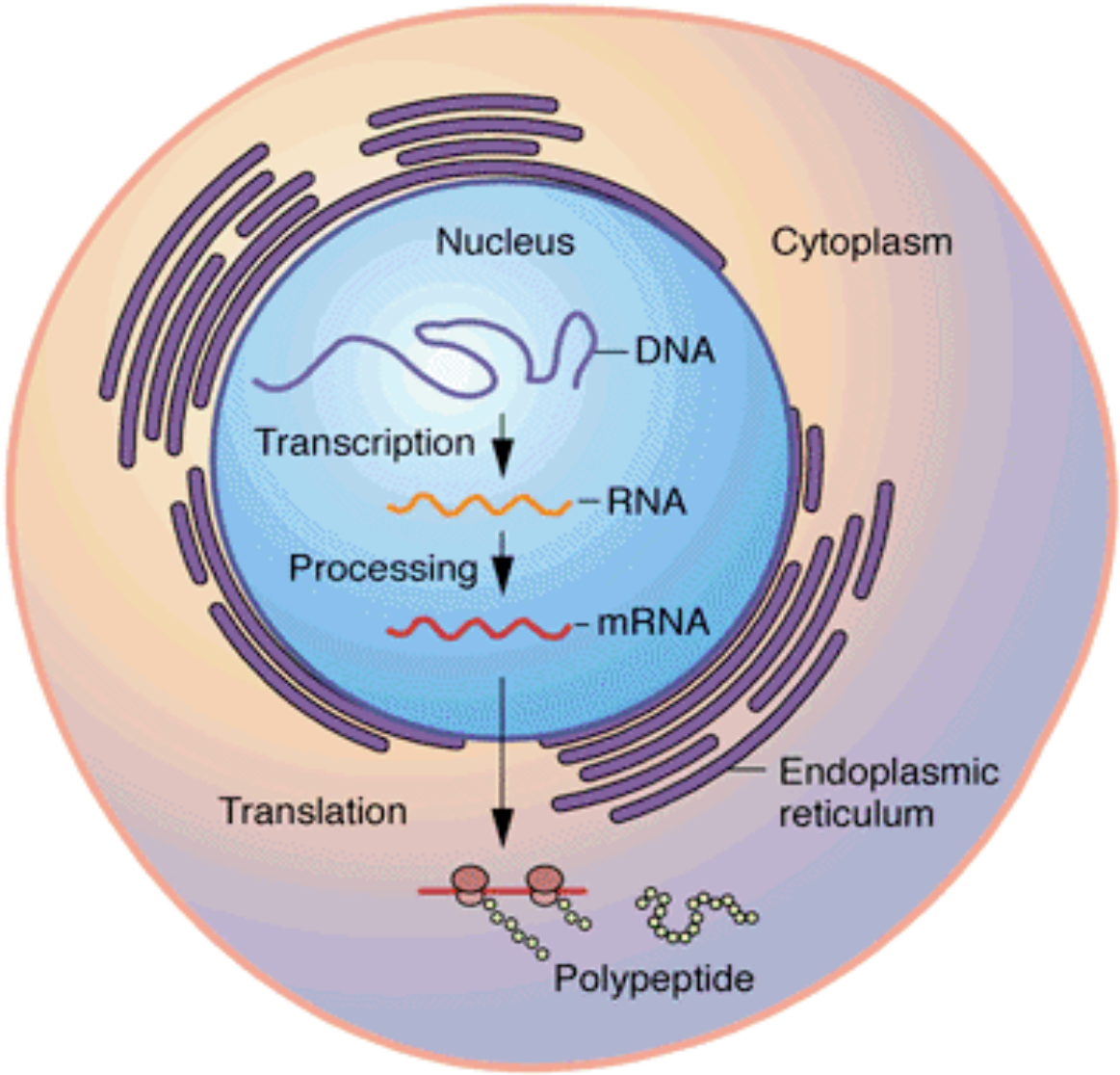
Mathematical Sciences

University of Delaware, Newark, DE

http:/udel.edu/~absingh/

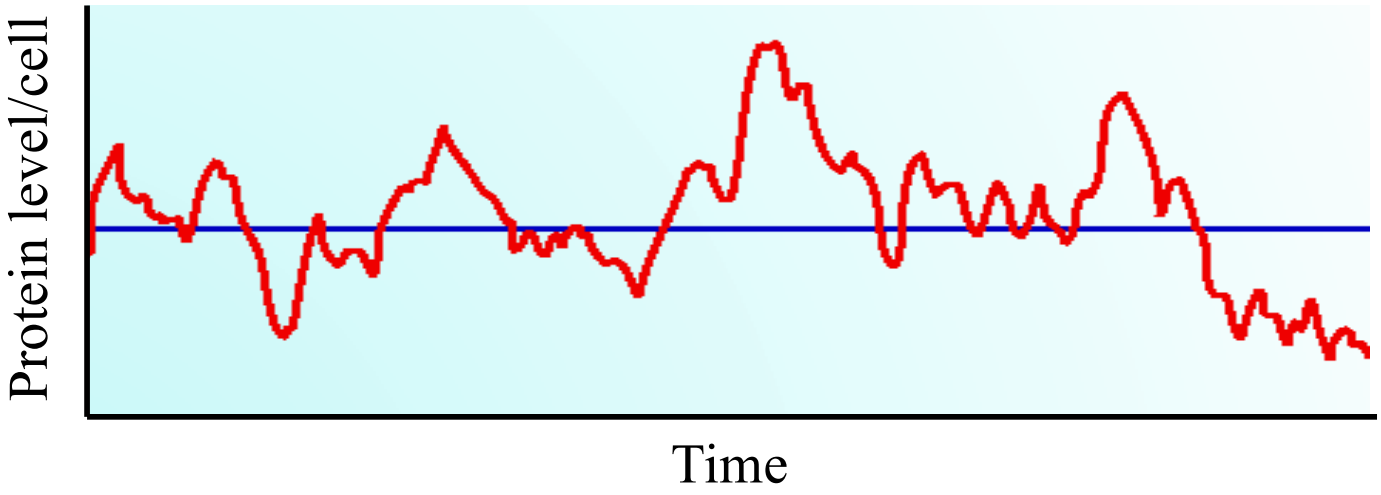


Gene expression: Production of proteins from genes through transcription and translation



Inside cells gene-expression is a stochastic process

- Timing of biochemical reactions is inherently stochastic
- Low-copy numbers of genes/mRNAs/proteins inside cells



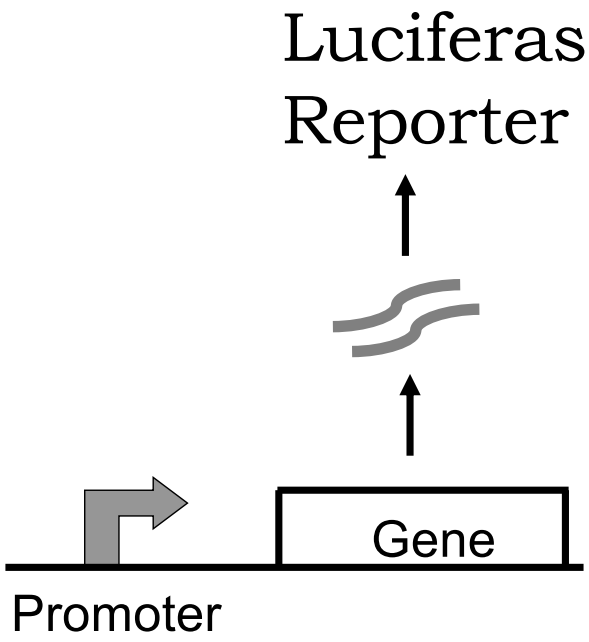
Suter et al. Science 2011, Taniguchi et al. Science 2010, Elowitz et al. Science 2002, Raser et al. Science 2005, Raj et al. Cell 2009, Blake et al. Nature 2003, Bar-Even et al. Nature Genetics 2006

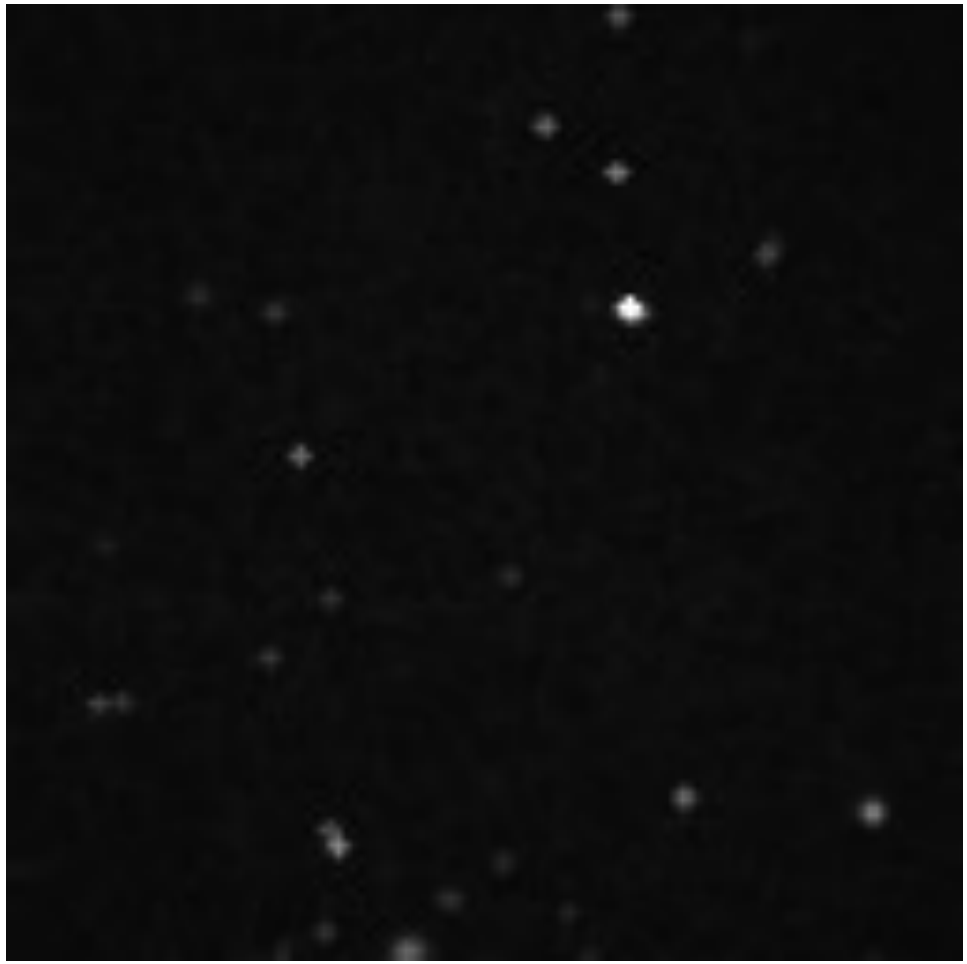
Science
AAAS

Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics

David M. Suter *et al.*

Science **332**, 472 (2011);
DOI: 10.1126/science.1198817



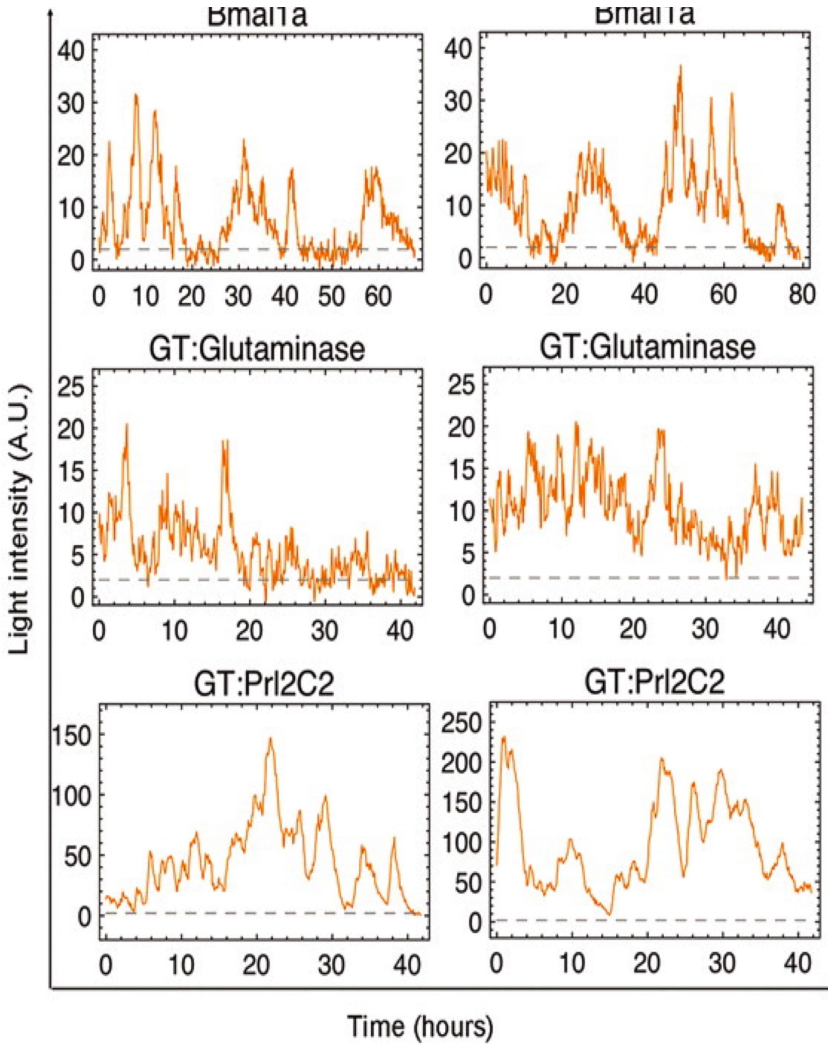


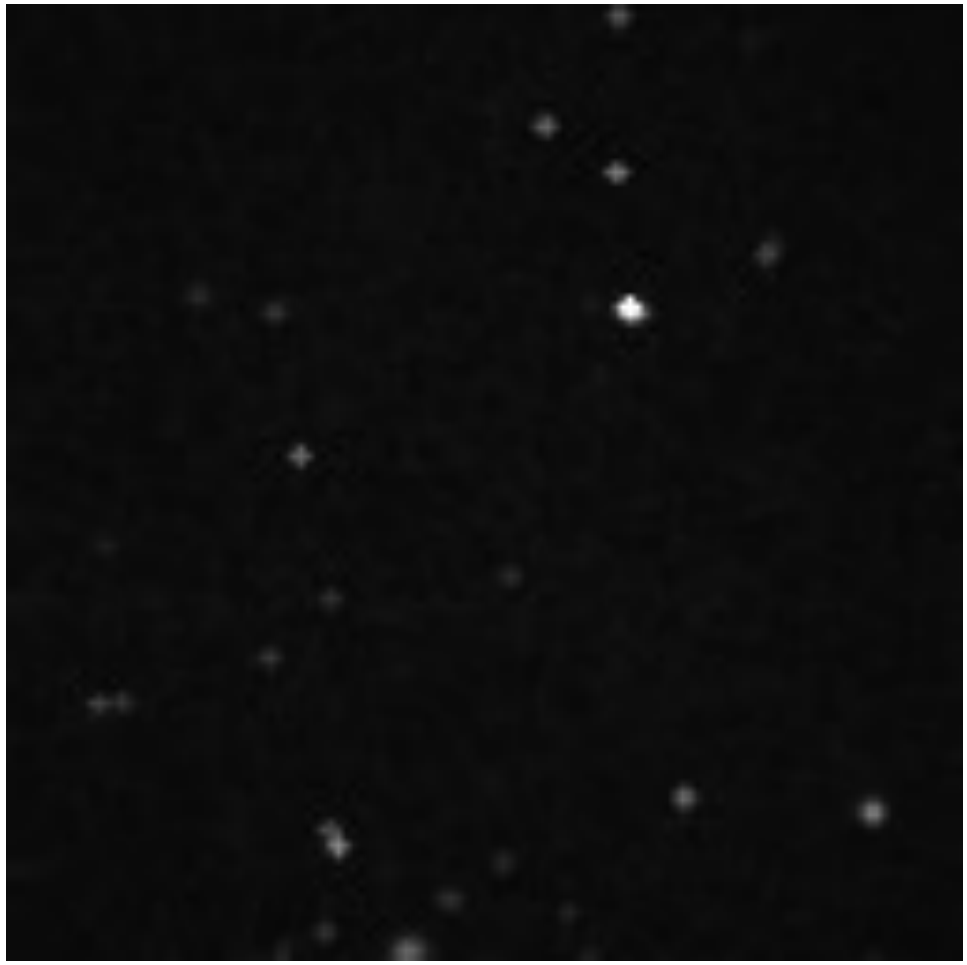
Science
AAAS

Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics

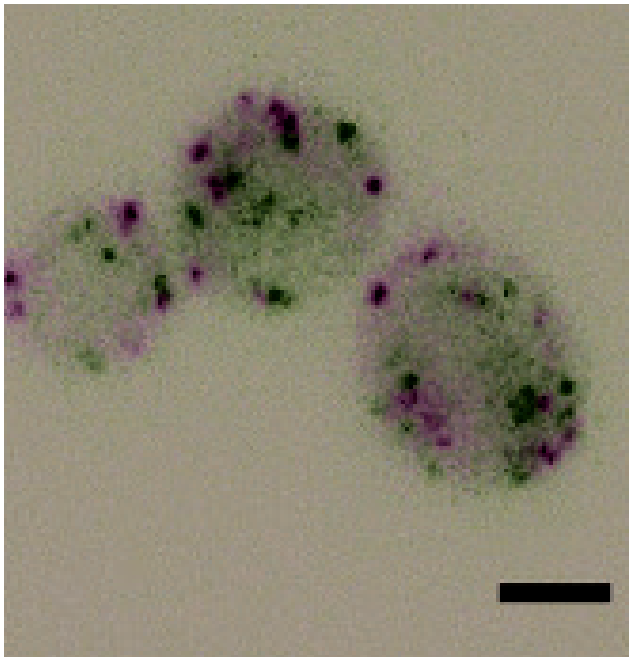
David M. Suter *et al.*

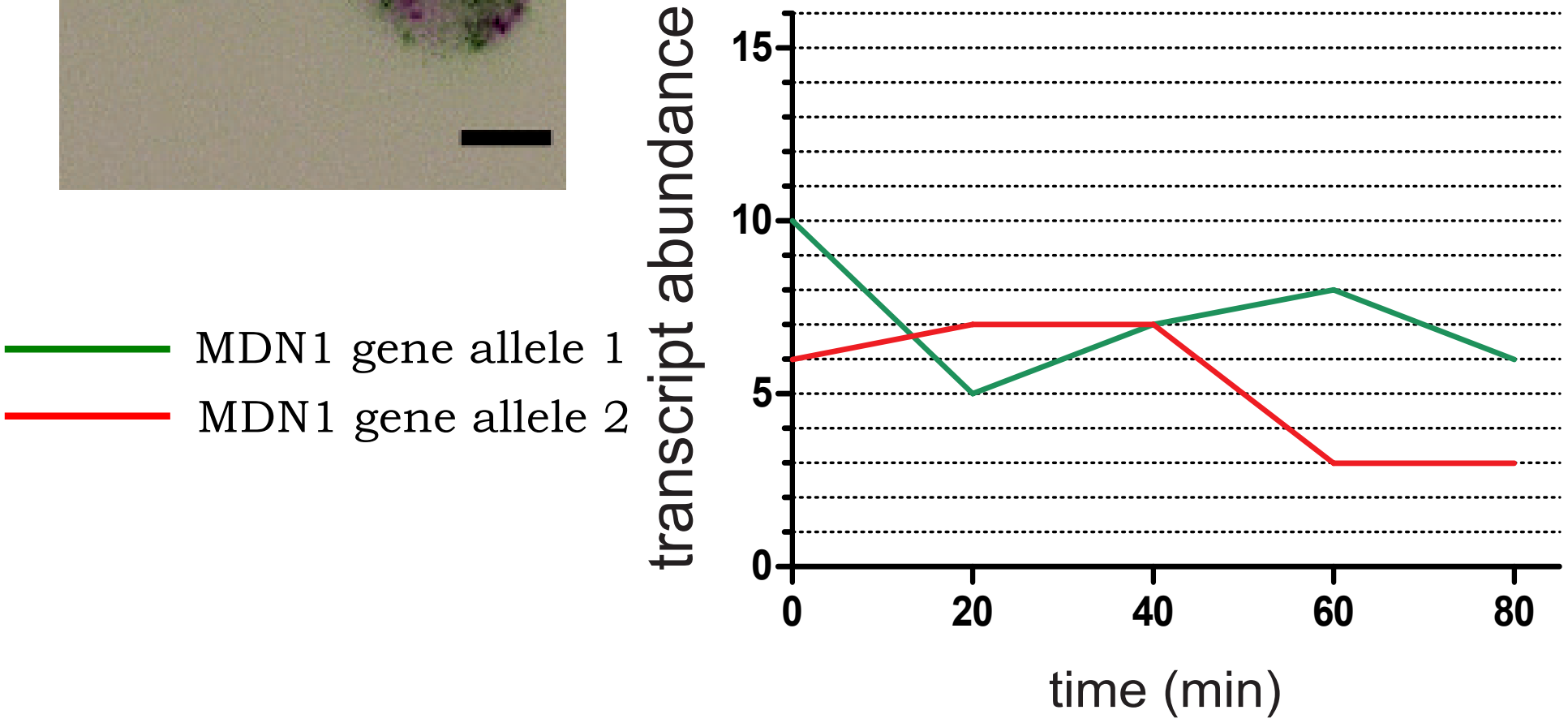
Science **332**, 472 (2011);
DOI: 10.1126/science.1198817





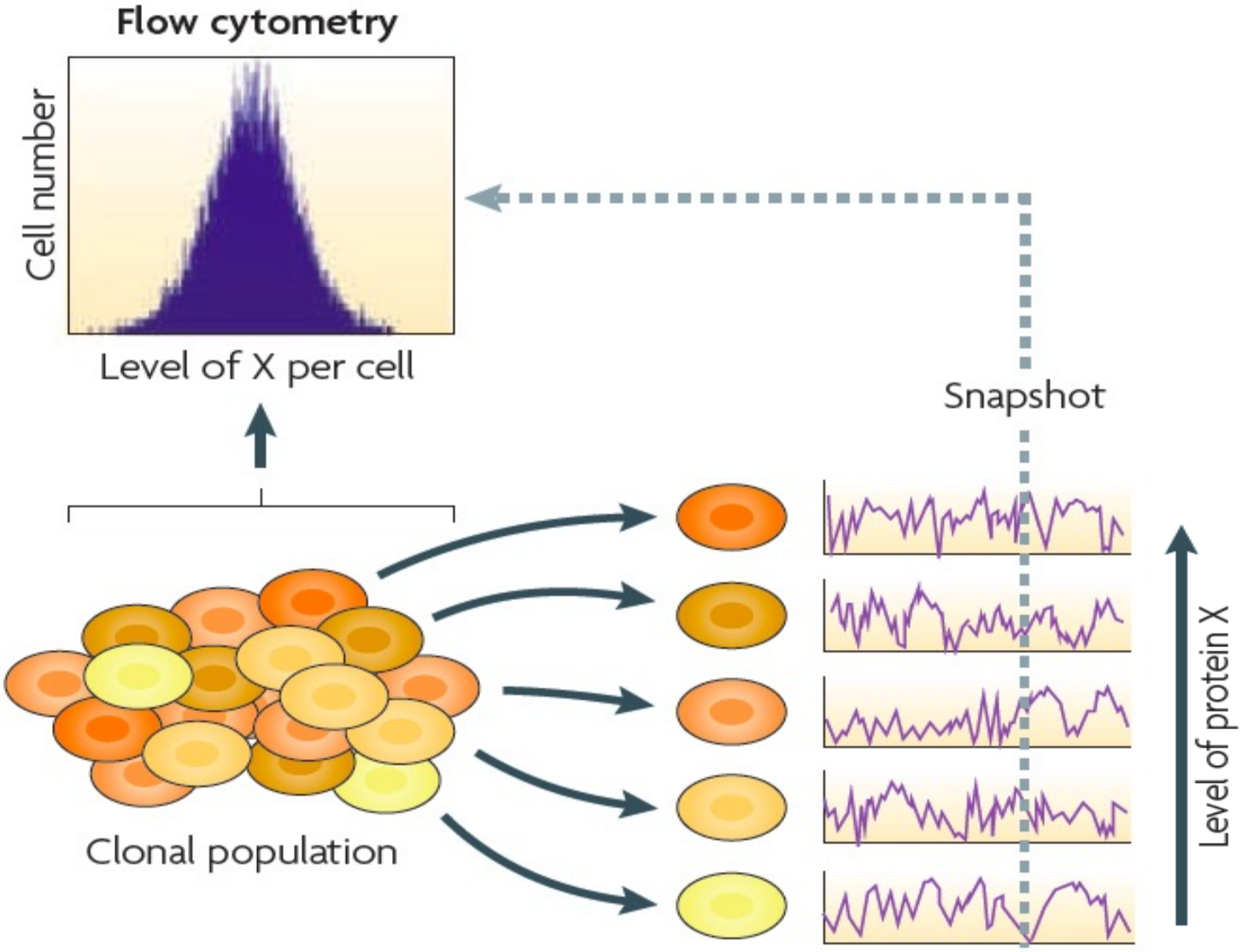
Single-cell measurements reveal randomness in mRNA copy numbers





Hocine et al Nature Methods 2012

Stochastic expression creates non-genetic heterogeneity in protein levels



Brock et al Nature Reviews Genetics 2009

Functional roles of expression noise

- Increased expression noise associated with diseased states

  Raj et al. Cell 2009 , Fraser et al. Plops Biology 2004

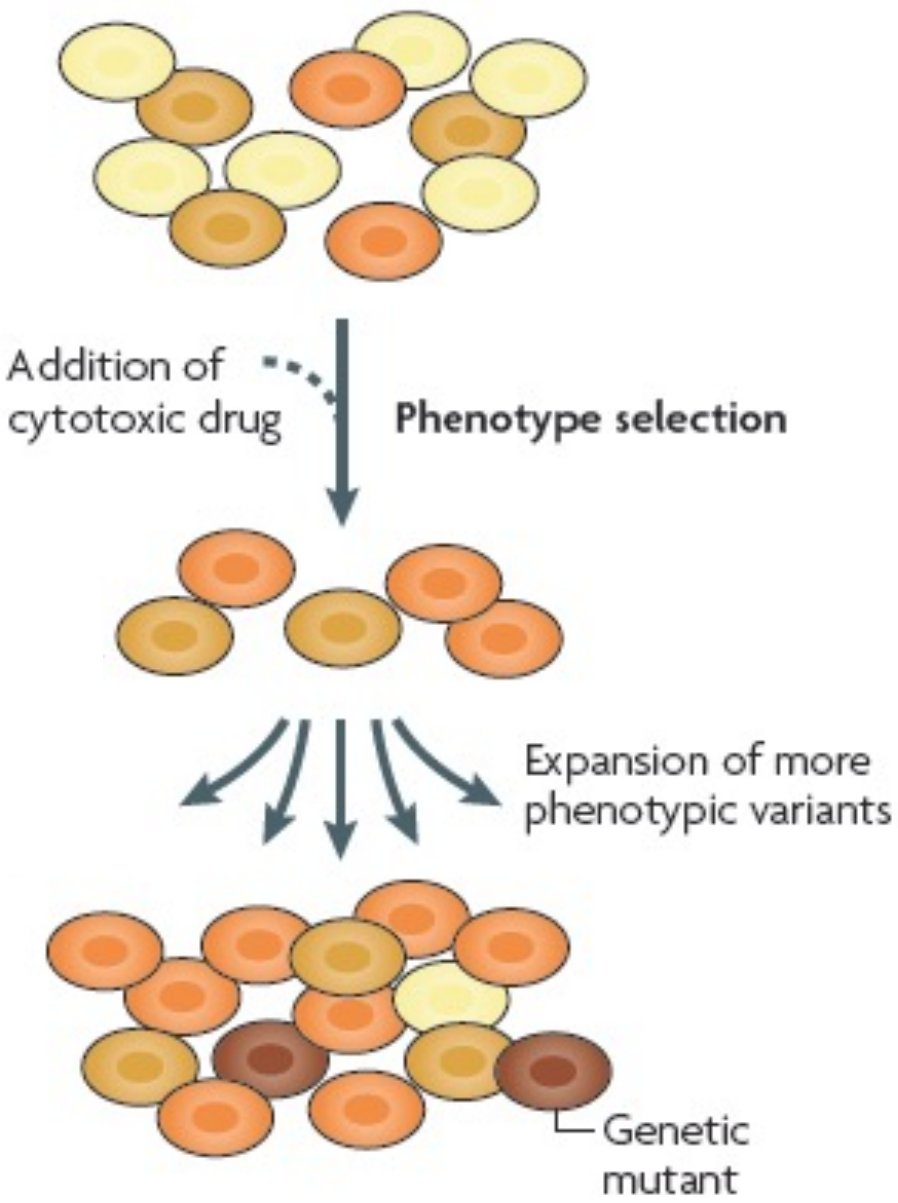
- Stochastic gene expression drives cell fate decisions

  Chang et al. Nature 2008, Losick et al. Science 2008, Maamar et al. Science 2007

- Expression noise implicated in E. Coli. antibiotic resistance and mutation-independent selection of tumors

  Brock et al. Nature Genetics 2009, Balaban et al. Science 2004

Expression noise drives antibiotic resistance



Brock et al. Nature Genetics 2009

Fundamental questions in noise biology

Biological consequences of noise in gene-expression?

Tools for modeling stochastic fluctuations in protein levels?

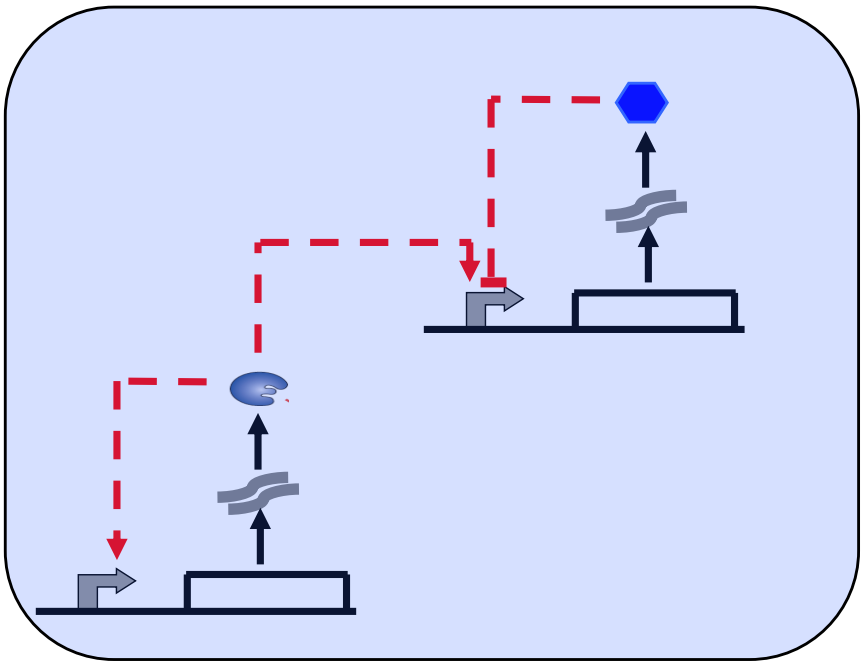
Regulatory mechanisms for buffering stochastic fluctuations?

Can fluctuations reveal information about underlying circuits?

Outline

- Background on stochastic modeling
- Solving CME using generating functions
- Inferring gene expression parameters from single-cell data
- Effect of transport delays on protein noise levels
- Cell-fate regulation in HIV

Genetic circuits as a set of chemical reactions



M chemical reactions

N chemical species $X_1, \ldots, X_N$

# of molecules $x_1, \ldots, x_N$

Stochastic formulation of biochemical reactions

Chemical reaction	Probability reactions occurs in $(t,t+dt]$	Change in population count of chemical species
$\mathrm{X}_1 \xrightarrow{c} \mathrm{X}_2$	cx_1dt	$x_1 \mapsto x_1 - 1 \quad x_2 \mapsto x_2 + 1$
$\mathrm{X}_2 + \mathrm{X}_2 \xrightarrow{c} \mathrm{X}_3$	$cx_2(x_2 - 1)dt/2$	$x_2 \mapsto x_2 - 2 \quad x_3 \mapsto x_3 + 1$
$\mathrm{X}_3 + \mathrm{X}_4 \xrightarrow{c} \mathrm{X}_5$	cx_3x_4dt	$x_3 \mapsto x_3 - 1 \quad x_4 \mapsto x_4 - 1 \quad x_5 \mapsto x_5 + 1$

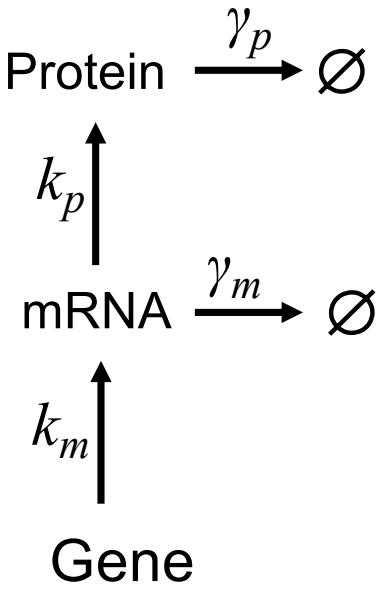
McQuarrie J. Applied Prob. 1967

$$x(t) = [x_1(t), x_2(t), \ldots, x_N(t)]$$

Population count of chemical species

Goal is to obtain the probability distribution of $x(t)$

Stochastic model for gene expression

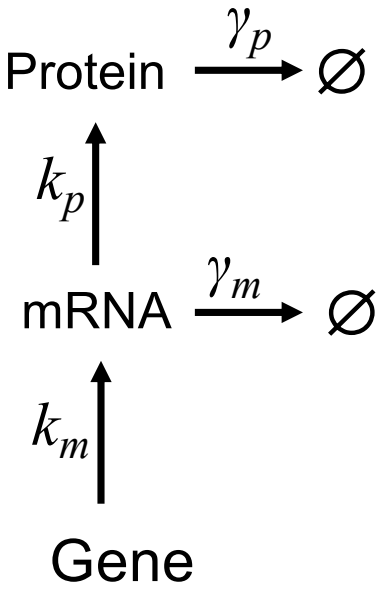


$p(t)$: Protein count at time t

$m(t)$: mRNA count at time t

Event	Reset in population count	Prob. event occurs in $(t, t+dt]$
Transcription	$m(t) \to m(t)+1$	$k_m dt$
mRNA degradation	$m(t) \to m(t)-1$	$\gamma_m m(t) dt$
Translation	$p(t) \to p(t)+1$	$k_p m(t) dt$
Protein degradation	$p(t) \to p(t)-1$	$\gamma_p p(t) dt$

Chemical Master Equation (CME)



$p(t)$: Protein count at time t

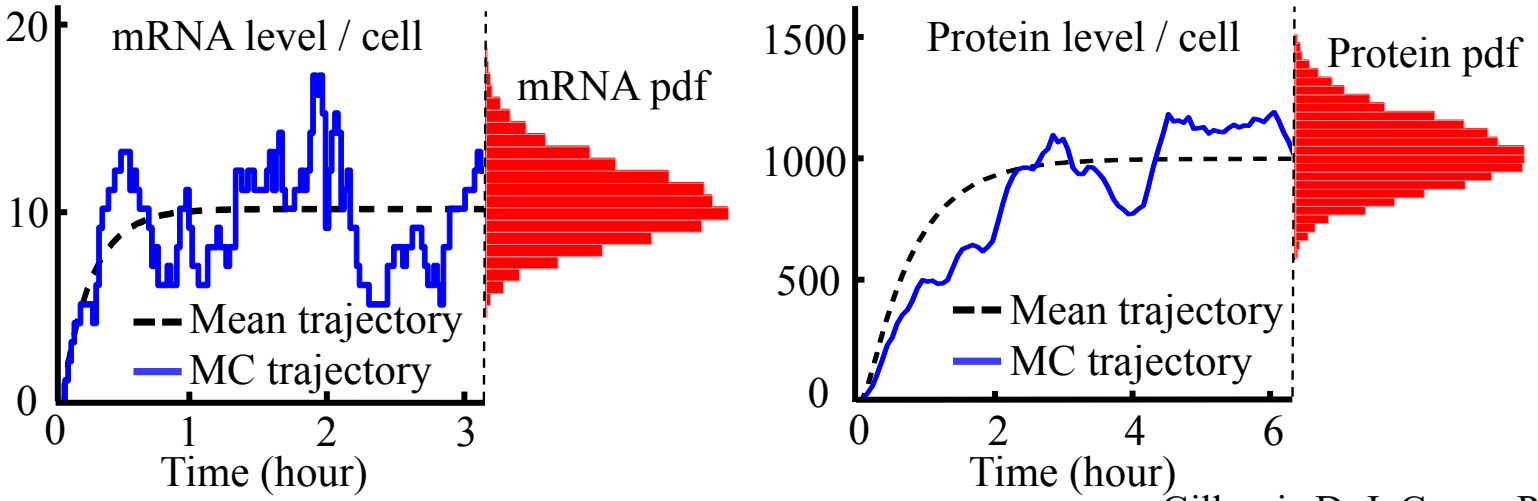
$m(t)$: mRNA count at time t

$P_{i,j}(t)$: Probability$(m(t) = i, p(t) = j)$

$$\frac{dP_{i,j}(t)}{dt} = k_m P_{i-1,j}(t) + \gamma_m (i+1) P_{i+1,j}(t) + k_p i P_{i,j-1}(t)$$
$$+ \gamma_p (j+1) P_{i,j+1}(t) - P_{i,j}(t)(k_m + \gamma_m i + k_p i + \gamma_p j)$$

State of art in solving CME

- Analytical solutions available in some cases
- Finite-state projection algorithm — Munsky & Khammash. J. Chemical Physics 2006
- Various Monte Carlo Simulation Techniques



Gillespie D. J. Comp. Physics 1977

- Moment Closure Techniques for predicting statistical moments

Singh & Hespanha IEEE Trans. Automatic Control 2011

Simple birth process

$$\varnothing \xrightarrow{k} X$$

$$\frac{dP_i(t)}{dt} = k\left(P_{i-1}(t) - P_i(t)\right), \quad P_0(0) = 1$$

$$G(z,t) := \sum_{i=0}^{\infty} z^i P_i(t)$$

$$\frac{\partial G(z,t)}{\partial t} = \sum_{i=0}^{\infty} z^i \frac{dP_i(t)}{dt} = k\sum_{i=0}^{\infty} z^i P_{i-1}(t) - k\sum_{n=0}^{\infty} z^i P_i(t)$$

$$\sum_{i=0}^{\infty} z^i P_{i-1}(t) = z\sum_{i=0}^{\infty} z^{i-1} P_{i-1}(t) = z\sum_{i=0}^{\infty} z^i P_i(t) = zG(z,t)$$

Probability distribution function of birth process

$$\varnothing \xrightarrow{k} X$$

$$\left.\begin{array}{l} \dfrac{\partial G(z,t)}{\partial t} = k(z-1)G(z,t) \\ P_0(0) = 1 \Rightarrow G(z,0) = 1 \end{array}\right\} \Rightarrow G(z,t) = e^{k(z-1)t}$$

$$P_i(t) = \frac{1}{i!}\left.\frac{\partial^i G(z,t)}{\partial z^i}\right|_{z=0} \Rightarrow P_i(t) = \frac{(kt)^i}{i!}e^{-kt}$$

Birth and death process

$$\varnothing \xrightarrow{k} X$$

$$X \xrightarrow{\gamma} \varnothing$$

$$\frac{dP_i(t)}{dt} = kP_{i-1}(t) + \gamma(i+1)P_{i+1}(t) - (k+\gamma i)P_i(t), \quad P_0(0) = 1$$

$$\sum_{i=0}^{\infty} iz^i P_i(t) = z\sum_{i=0}^{\infty} iz^{i-1} P_i(t) = z\frac{\partial G(z,t)}{\partial z}$$

$$\frac{\partial G(z,t)}{\partial t} = k(z-1)G(z,t) - \gamma(z-1)\frac{\partial G(z,t)}{\partial z}, \quad G(z,0) = 1$$

$$G(z,t) := \sum_{i=0}^{\infty} z^i P_i(t) \Rightarrow G(1,t) = 1$$

Probability distribution function of birth-death process

$$\varnothing \xrightarrow{k} X$$

$$X \xrightarrow{\gamma} \varnothing$$

$$G(z,t) = e^{\frac{k}{\gamma}(z-1)(1-e^{-\gamma t})}$$

$$\lim_{t\to\infty}(G(z,t)) = e^{\frac{k}{\gamma}(z-1)}$$

$$P_i(t) = \frac{(k/\gamma)^i}{i!}(1-e^{-\gamma t})^i e^{-\frac{k}{\gamma}(1-e^{-\gamma t})}$$

$$\lim_{t\to\infty}(P_i(t)) = \frac{(k/\gamma)^i}{i!} e^{-\frac{k}{\gamma}}$$

Simple decay process

$$X \xrightarrow{\gamma} \varnothing$$

$$\frac{dP_i(t)}{dt} = \gamma\left((i+1)P_{i+1}(t) - iP_i(t)\right), \quad P_{x_0}(0) = 1$$

$$\Rightarrow \frac{\partial G(z,t)}{\partial t} = \gamma(1-z)\frac{\partial G(z,t)}{\partial z} \qquad \begin{aligned} G(z,0) &= z^{x_0} \\ G(1,t) &= 1 \end{aligned}$$

$$G(z,t) = \left(1 + (z-1)e^{-\gamma t}\right)^{x_0} \qquad P_i(t) = \binom{x}{i}\left(1 - e^{-\gamma t}\right)^{x_0 - i} e^{-\gamma t i}$$

Burst-birth and death process

$$\varnothing \xrightarrow{k} B \times X \qquad X \xrightarrow{\gamma} \varnothing$$

$$\left(\text{Probability of } B = j\right) = \alpha_j$$

$$\frac{dP_i(t)}{dt} = k\sum_{j=0}^{i} \alpha_j P_{i-j}(t) + \gamma(i+1)P_{i+1}(t) - \left(k+\gamma i\right)P_i(t)$$

$$\Rightarrow \frac{\partial G(z,t)}{\partial t} = k(\alpha(z)-1)G(z,t) - \gamma(z-1)\frac{\partial G(z,t)}{\partial z}$$

$$G(z,0) = 1$$

$$G(1,t) = 1$$

Probability distribution function of bursty birth-death process

Proteins are produced in geometric bursts

$$\alpha_j = \rho^j (1-\rho) \qquad \alpha(z) = \frac{(1-\rho)}{1-\rho z}$$

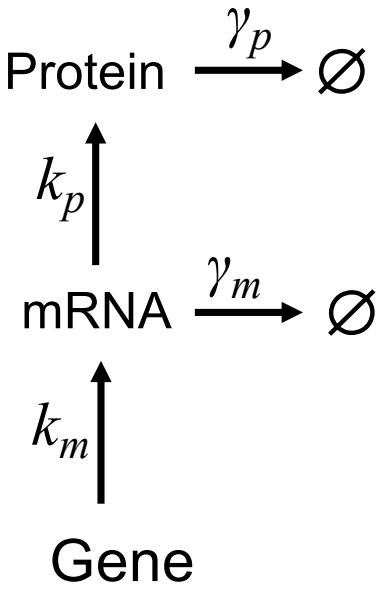
$$G(z,t) = \left(\frac{1 + \frac{\rho}{1-\rho}(1-z)e^{-\gamma t}}{1 + \frac{\rho}{1-\rho}(1-z)} \right)^{\frac{k}{\gamma}} \qquad \lim_{t\to\infty}(G(z,t)) = \frac{(1-\rho)^{k/\gamma}}{(1-\rho z)^{k/\gamma}}$$

$$\lim_{t\to\infty}(P_i(t)) = \frac{\Gamma((k/\gamma)+i)}{\Gamma(i)\Gamma(k/\gamma)}(1-\rho)^{k/\gamma}\rho^i$$

Outline

- Background on stochastic modeling
- Solving CME using generating functions
- Inferring gene expression parameters from single-cell data
- Effect of transport delays on protein noise levels
- Cell-fate regulation in HIV

Chemical Master Equation (CME)



$p(t)$: Protein count at time t

$m(t)$: mRNA count at time t

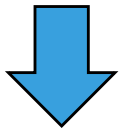
$P_{i,j}(t)$: Probability$(m(t) = i, p(t) = j)$

$$\frac{dP_{i,j}(t)}{dt} = k_m P_{i-1,j}(t) + \gamma_m (i+1) P_{i+1,j}(t) + k_p i P_{i,j-1}(t)$$
$$+ \gamma_p (j+1) P_{i,j+1}(t) - P_{i,j}(t)(k_m + \gamma_m i + k_p i + \gamma_p j)$$

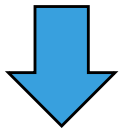
Solution CME for stochastic gene expression

$$\frac{dP_{i,j}(t)}{dt} = k_m P_{i-1,j}(t) + \gamma_m (i+1) P_{i+1,j}(t) + k_p i P_{i,j-1}(t)$$
$$+ \gamma_p (j+1) P_{i,j+1}(t) - P_{i,j}(t)(k_m + \gamma_m i + k_p i + \gamma_p j)$$

$$G(y,z,t) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} y^i z^j P_{i,j}(t)$$



Linear first-order PDE on the generating function



Exact solution for generating function

Bokes et al. J. Math Biology 2011

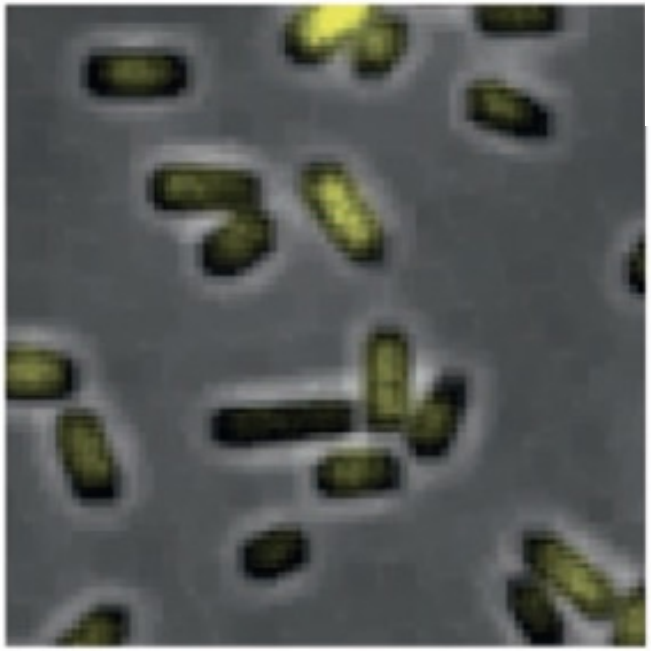
Solution CME for stochastic gene expression

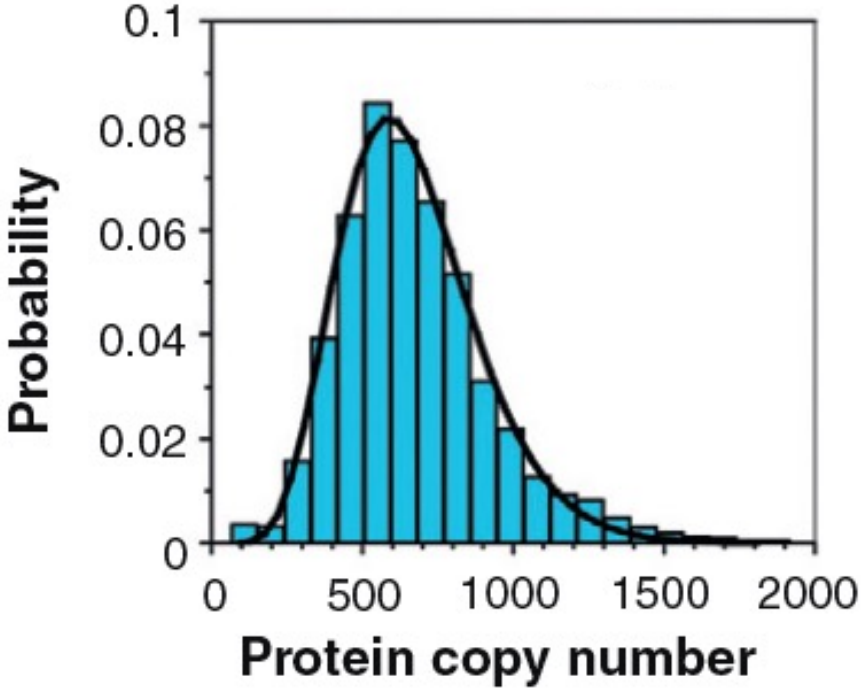
$$\frac{dP_{i,j}(t)}{dt} = k_m P_{i-1,j}(t) + \gamma_m (i+1) P_{i+1,j}(t) + k_p i P_{i,j-1}(t)$$
$$+ \gamma_p (j+1) P_{i,j+1}(t) - P_{i,j}(t)(k_m + \gamma_m i + k_p i + \gamma_p j)$$

$$G(z) = \exp\left[\theta((1+\xi-\xi z)^{-\eta} - 1)\right]$$

$$\eta = \frac{\gamma_p}{\gamma_m} \quad \xi = \frac{k_p}{\gamma_m}\frac{1}{(1+\eta)^2} \quad \theta = \frac{k_m}{\gamma_p}\frac{(1+\eta)^2}{\eta}$$

Parameter inference from protein distribution





$$G(z) = \exp\left[\theta((1+\xi-\xi z)^{-\eta}-1)\right]$$

$$\eta = \frac{\gamma_p}{\gamma_m} \quad \xi = \frac{k_p}{\gamma_m}\frac{1}{(1+\eta)^2} \quad \theta = \frac{k_m}{\gamma_p}\frac{(1+\eta)^2}{\eta}$$

$\frac{\gamma_m}{\gamma_p}$, $\frac{k_p}{\gamma_p}$ & $\frac{k_m}{\gamma_p}$ can be inferred from experiment distribution

Parameter inference from protein distribution

k_m (Transcription rate); k_p (Translation rate); γ_m (mRNA degradation)

All rates normalized by protein degradation rate

Parameters	Real value	1000 data	2000 data
k_m	10	9.2	9.4
γ_m	5	4.4	4.9
k_p	100	82	87

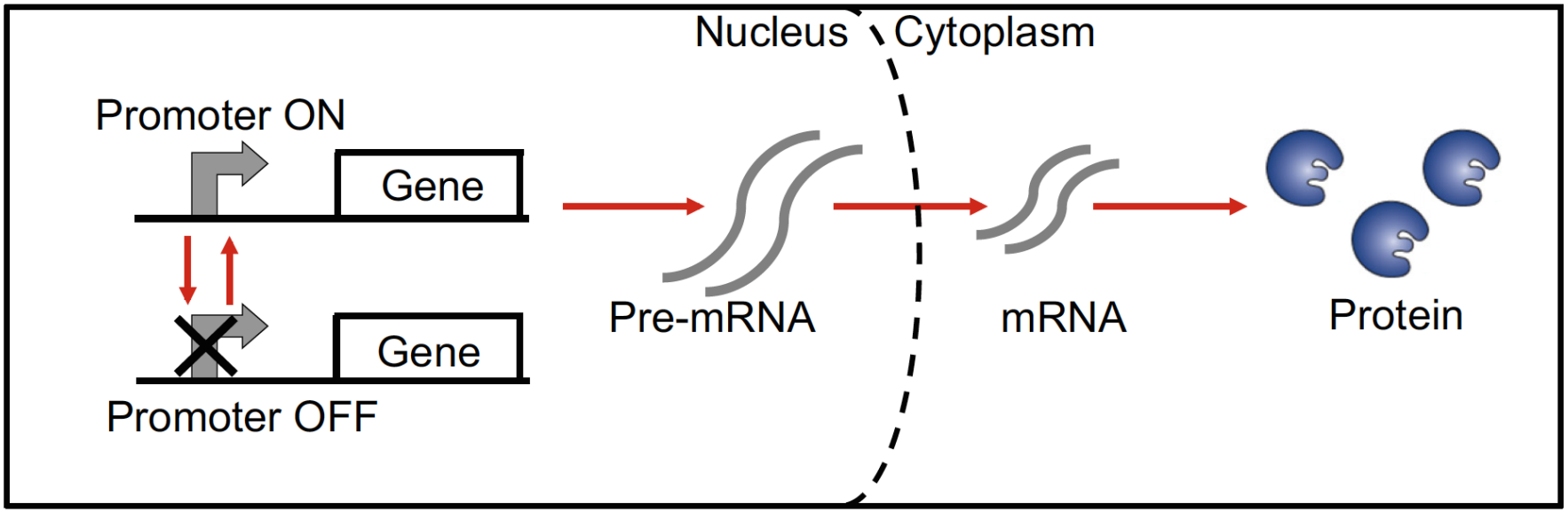
Parameters	Real value	1000 data	2000 data
k_m	2	1.72	1.88
γ_m	1	0.8	0.92
k_p	100	91	97

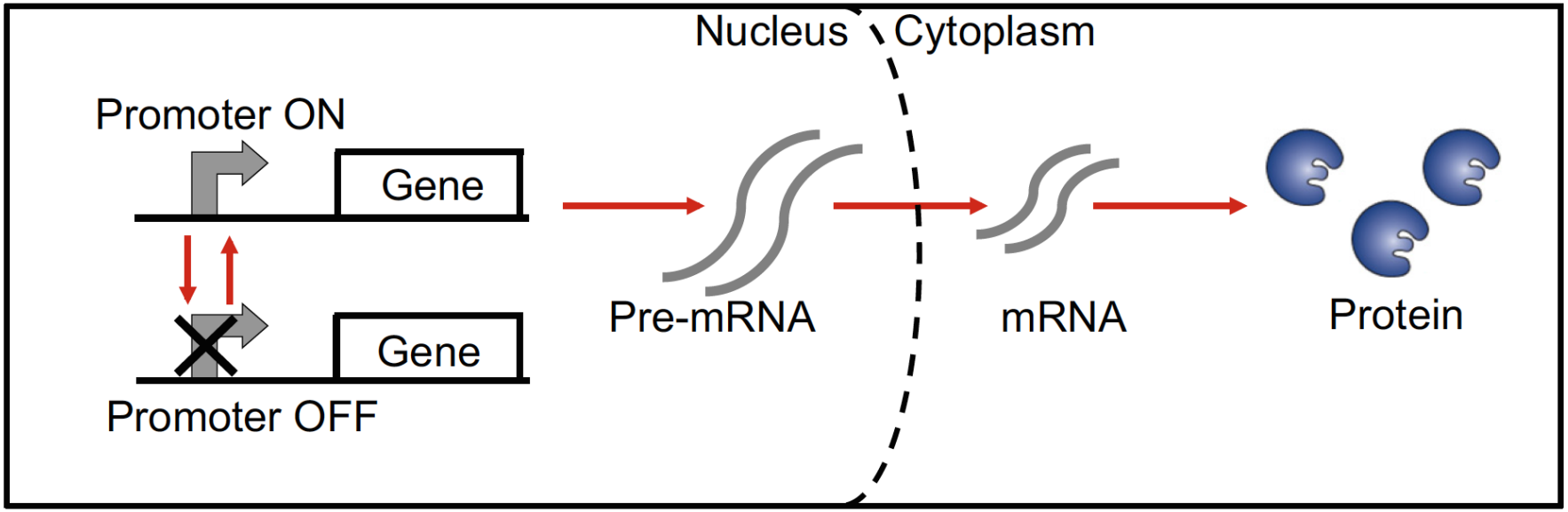
Taniguchi et al. Science 2010; Newman et al. Nature 2006; Bar-Even et al. Nature Genetics 2006

Biophysical Journal Volume 103 September 2012 1087–1096

Consequences of mRNA Transport on Stochastic Variability in Protein Levels

Abhyudai Singh[†*] and Pavol Bokes[‡]
[†]Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware; and [‡]Department of Applied Mathematics and Statistics, Comenius University, Bratislava, Slovakia





$$Gene \xrightarrow{k_m} Gene + B \times M_n \quad P(B = z) = \alpha_z$$

$$M_n \xrightarrow{\gamma_e} M_c \quad M_c \xrightarrow{\gamma_c} \varnothing$$

Chemical Master Equation (CME)

$$Gene \xrightarrow{k_m} Gene + B \times M_n \quad P(B = z) = \alpha_z$$

$$M_n \xrightarrow{\gamma_e} M_c \quad M_c \xrightarrow{\gamma_c} \varnothing$$

$$\begin{aligned}\frac{dP(m_n, m_c, t)}{dt} = k_m \Bigg(& \sum_{z=0}^{m_n} \alpha_z P(m_n - z, m_c, t) - P(m_n, m_c, t) \Bigg) \\ & + \gamma_e((m_n + 1)P(m_n + 1, m_c - 1, t) \\ & - m_n P(m_n, m_c, t)) + \gamma_c((m_c + 1) \\ & \times P(m_n, m_c + 1, t) - m_c P(m_n, m_c, t)),\end{aligned}$$

Transforming the CME into a PDE

$$G(x,y,t) = \sum_{m_n=0}^{\infty} \sum_{m_c=0}^{\infty} x^{m_n} y^{m_c} P(m_n, m_c, t)$$

$$\varphi(u,v,t) = \ln G(1+u, 1+v, t)$$

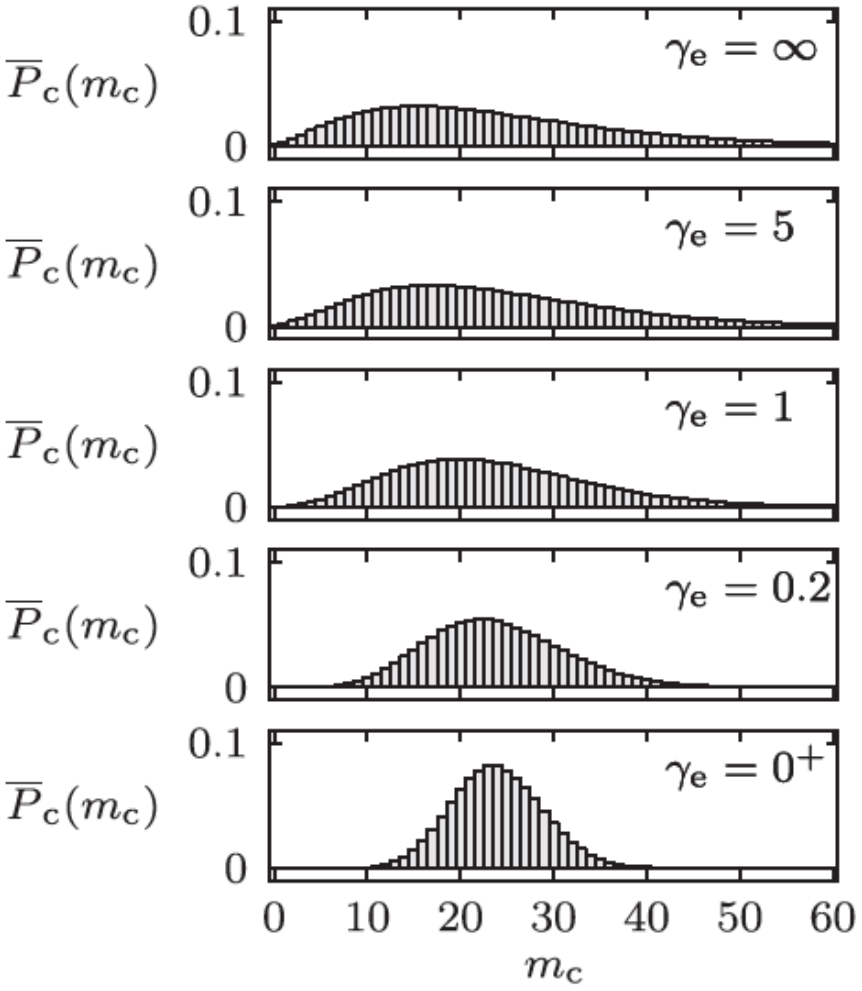
Factorial-cumulant generating function

$$\frac{\partial \varphi}{\partial t} = k_m(M(u) - 1) + \gamma_e(v-u)\frac{\partial \varphi}{\partial u} - \gamma_c v \frac{\partial \varphi}{\partial v}$$

Can be solved using the method of characteristics

Singh & Bokes. Biophysical Journal 2012

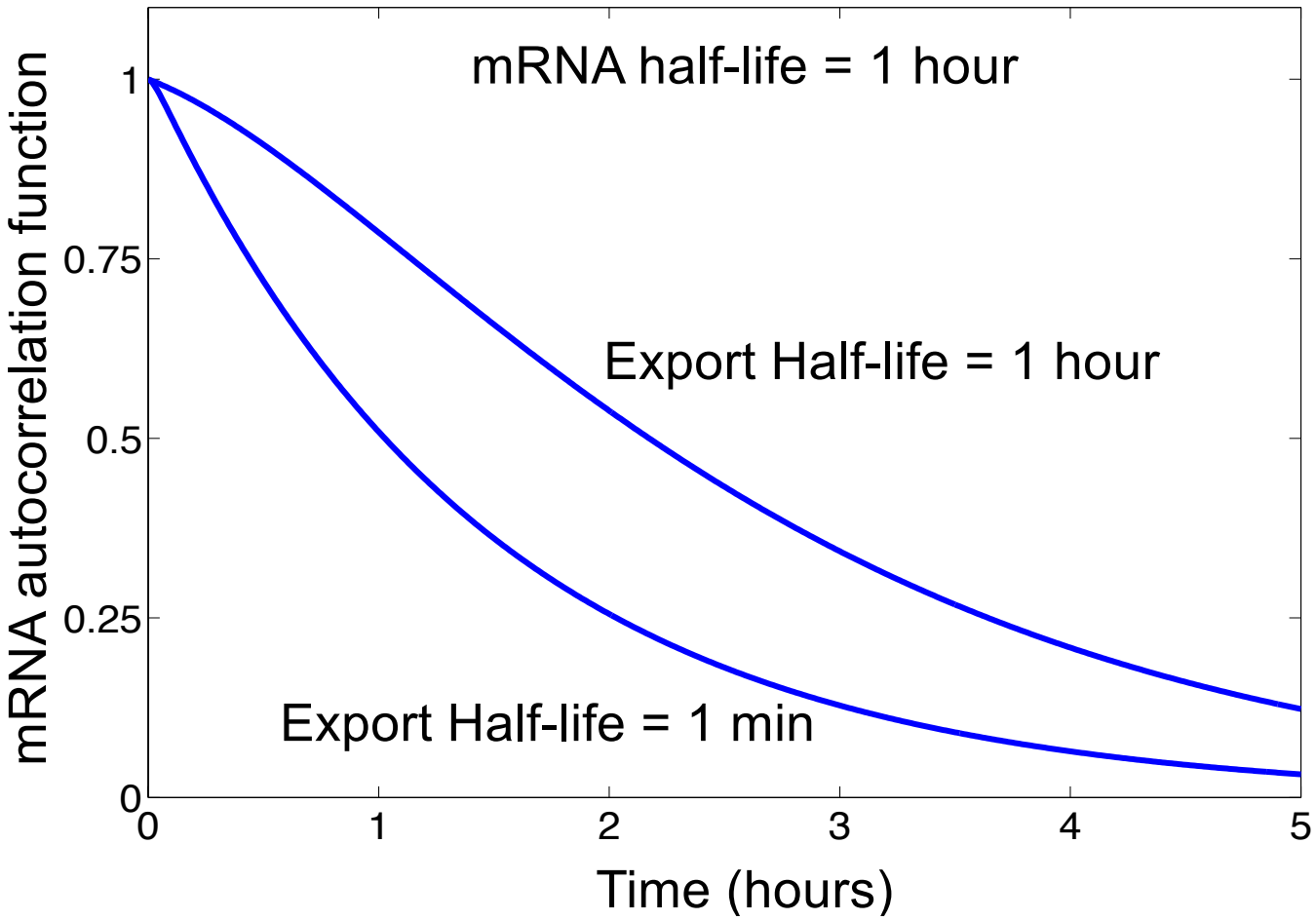
Transport delay reduces fluctuations in cytoplasmic mRNA levels

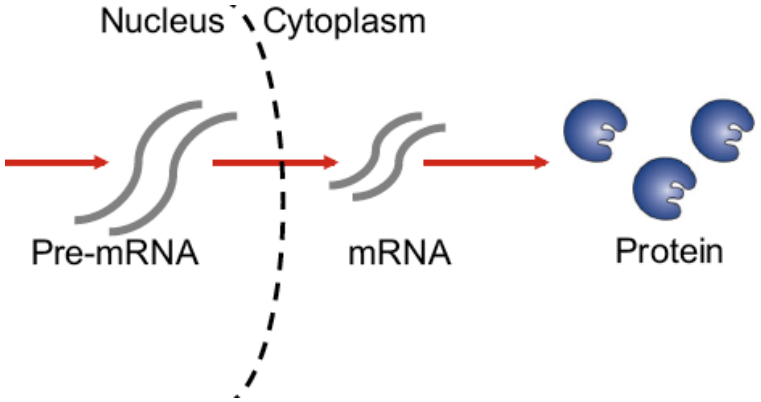


Transport delay extend the duration of fluctuations

$$Gene \xrightarrow{k_m} Gene + B \times M_n$$

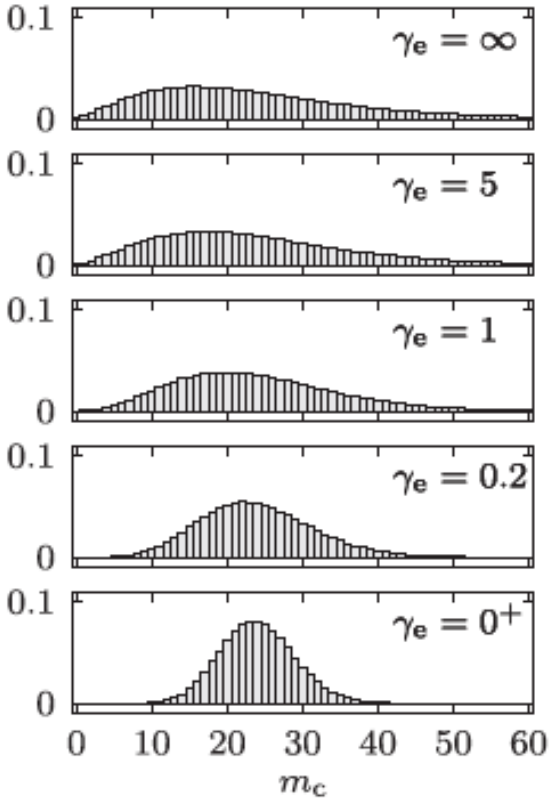
$$M_n \xrightarrow{\gamma_e} M_c \qquad M_c \xrightarrow{\gamma_c} \varnothing$$





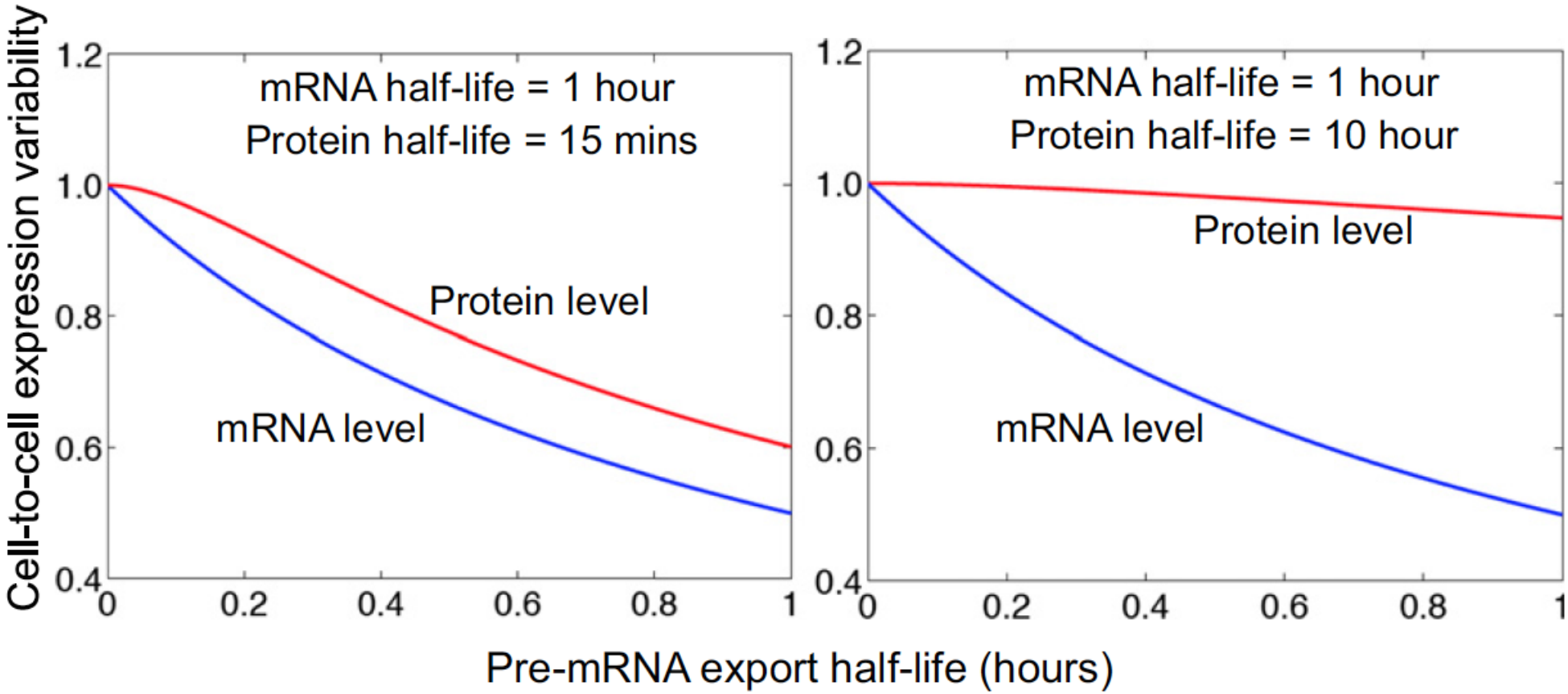
Longer mRNA fluctuations

Smaller mRNA fluctuations



No change in Protein variability

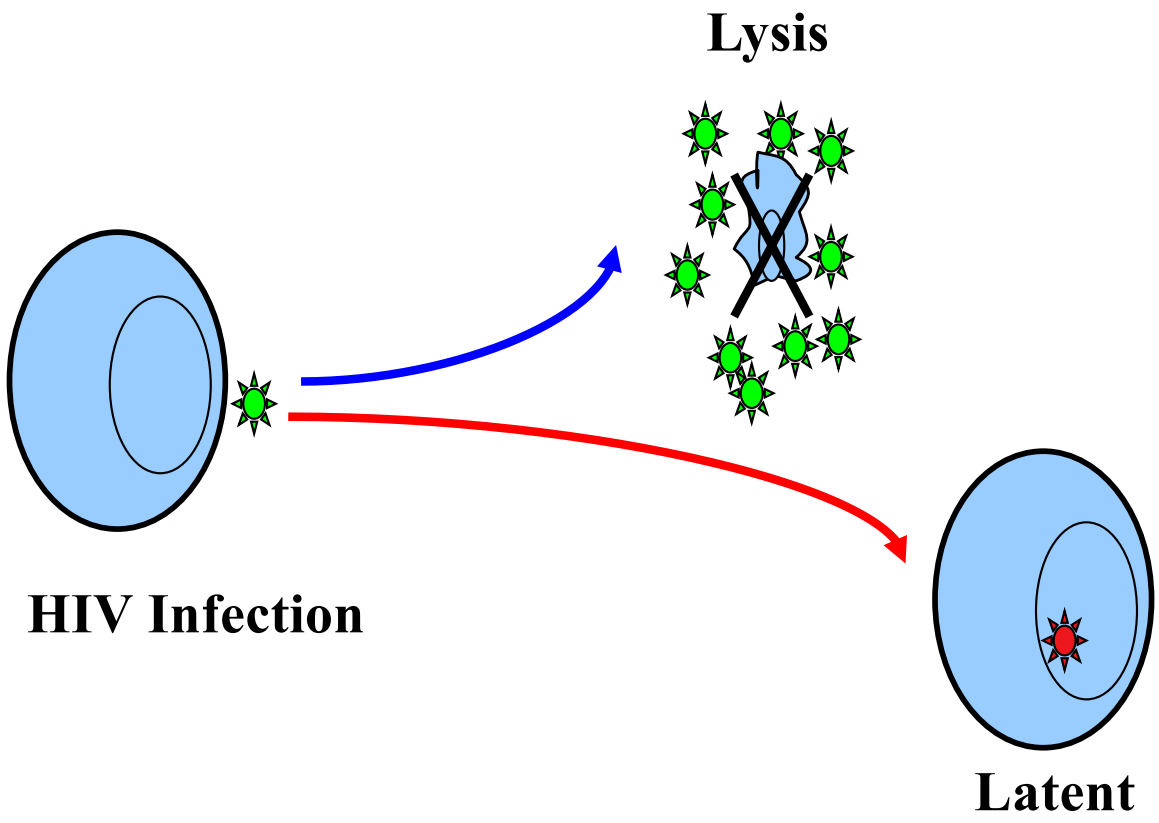
Singh & Bokes. Biophysical Journal 2012


Cell-to-cell expression variability
1.2
1.0
0.8
0.6
0.4
0
0.2
0.4
0.6
0.8
1
mRNA half-life = 1 hour
Protein half-life = 15 mins
Protein level
mRNA level
mRNA half-life = 1 hour
Protein half-life = 10 hour
Protein level
mRNA level
Pre-mRNA export half-life (hours)

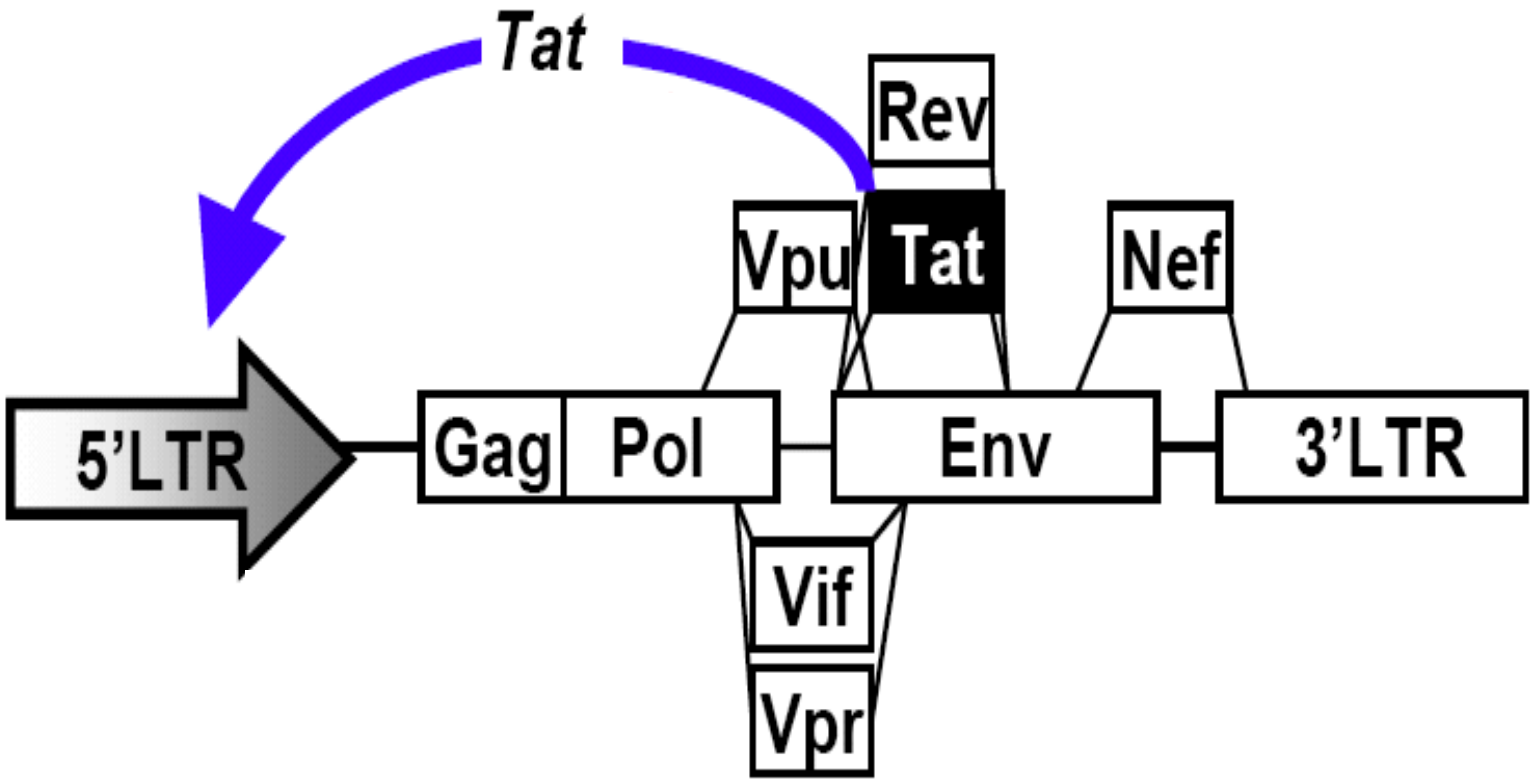
Outline

- Background on stochastic modeling
- Solving CME using generating functions
- Inferring gene expression parameters from single-cell data
- Effect of transport delays on protein noise levels
- Cell-fate regulation in HIV

HIV cell fate decision between active replication and latency



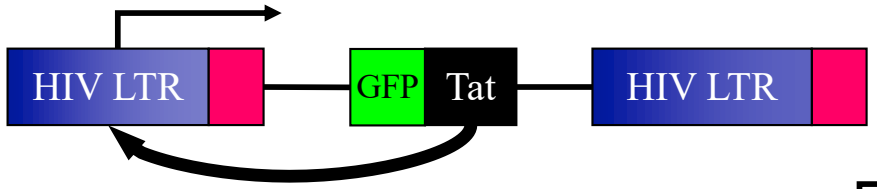
How does HIV's genetic circuit drive cell-fate decision?



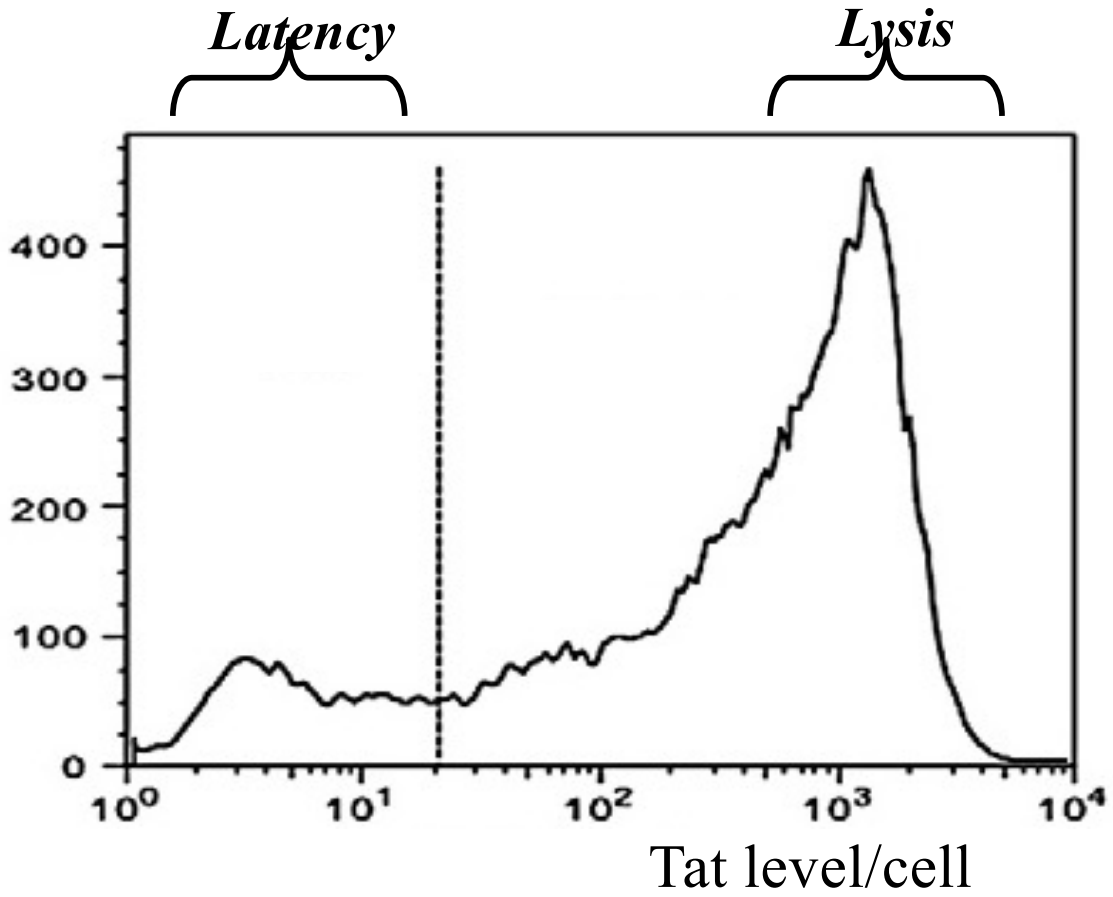
Outline

- Stochastic Tat expression drives HIV into latency
- Makes makes Tat expression stochastic?
- Stochastic modeling of the Tat-feedback circuit
- Therapies for purging the latent reservoir

Stochastic expression coupled with positive feedback circuitry can drive HIV latency

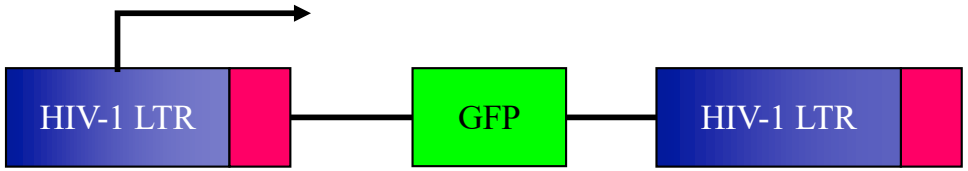


Infected Jurkat T with virus → Single-cell sorting → Expand into clonal populations



Chavez et al. PLOS Pathogens
Razooky et al. Cell 2015
Weinberger et al. Cell 2005
Weinberger et al. Nature Genetic
Pearson et al. J. Virology 20
Jordan et al. EMBO 2003

Monitoring stochastic expression from the HIV promoter

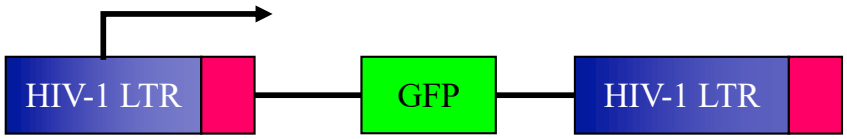


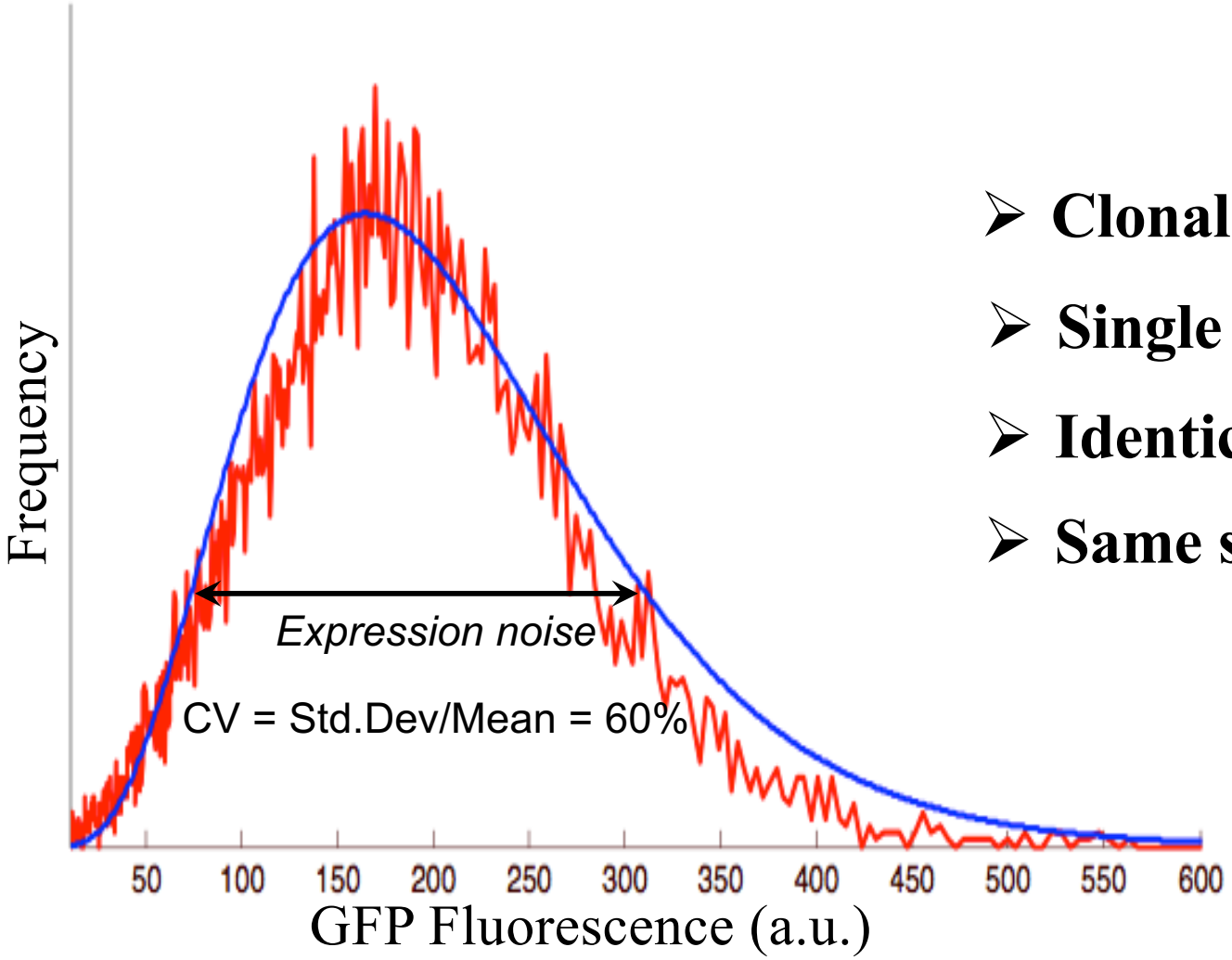
Infect Jurkat T cells with virus

Single-cell sorting

Expand into isoclonal populations

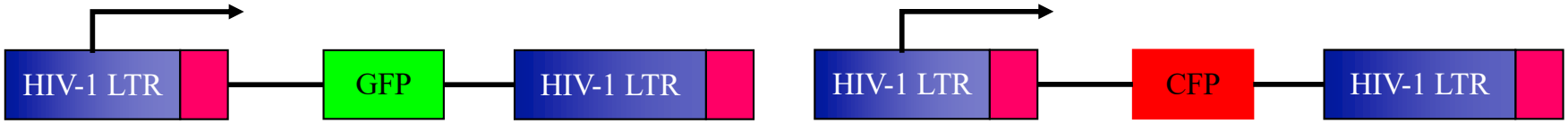
HIV encodes a promoter with high noise in gene-expression

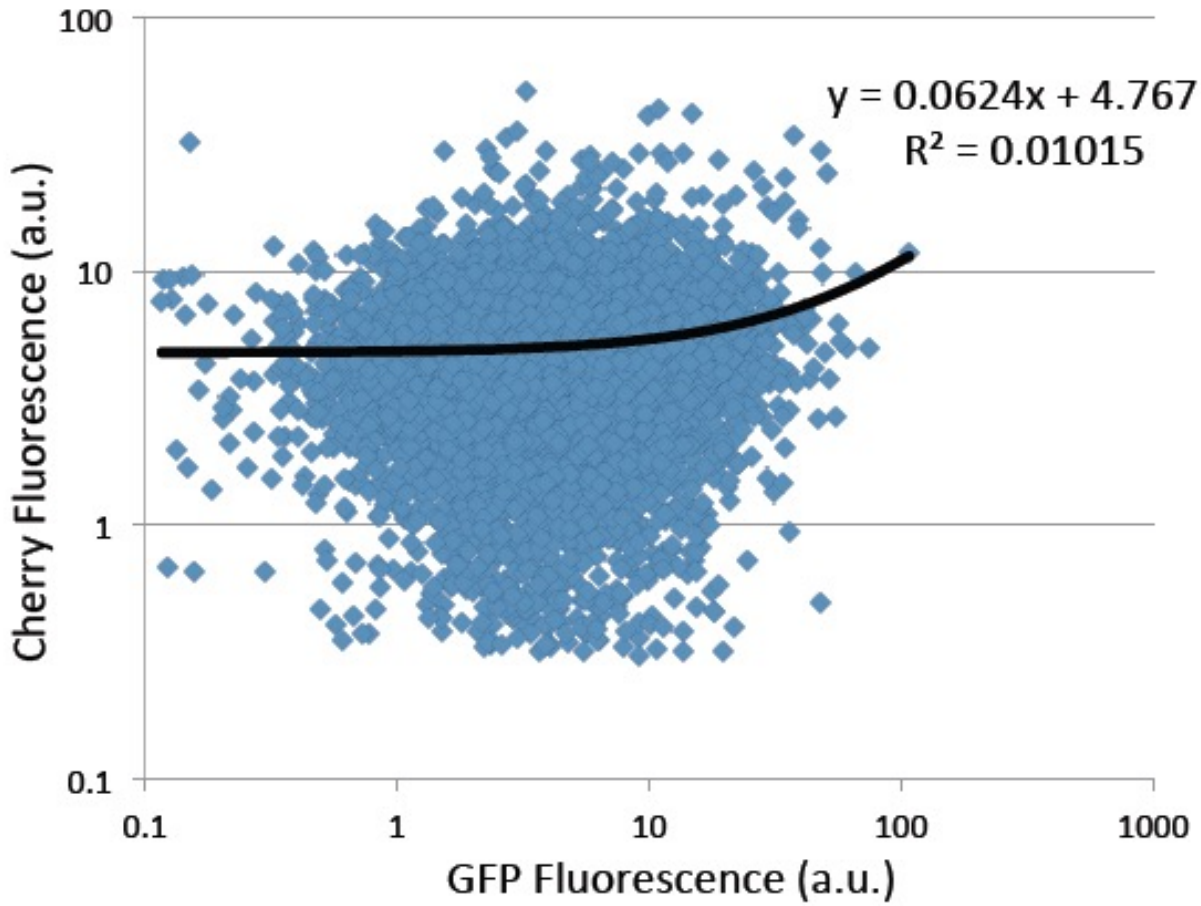




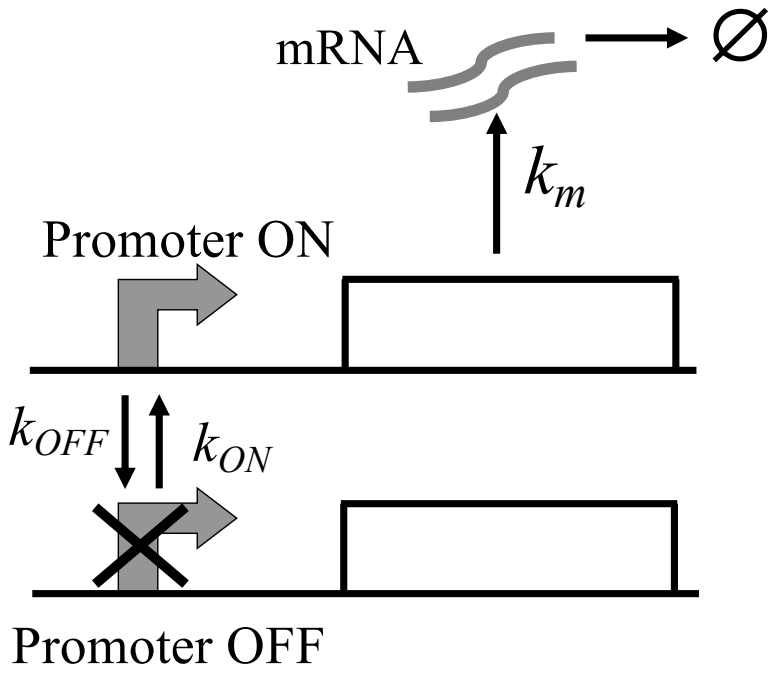
- **Clonal cell population**
- **Single promoter copy**
- **Identical integration site**
- **Same shape and size**

HIV gene expression noise is intrinsic





Transcriptional bursting at the HIV LTR creates gene-expression variability

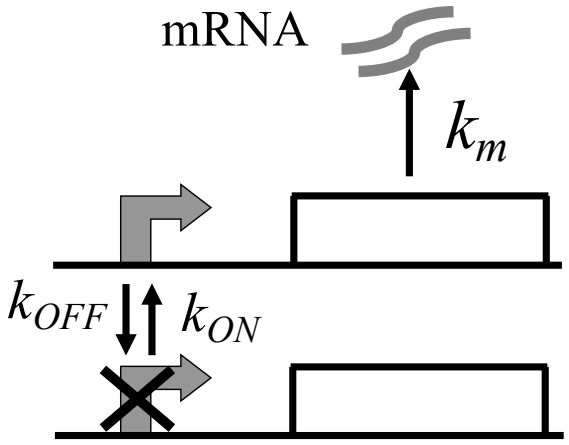


Transcriptional burst frequency=k_{on}

Transcriptional burst size=k_m/k_{off}

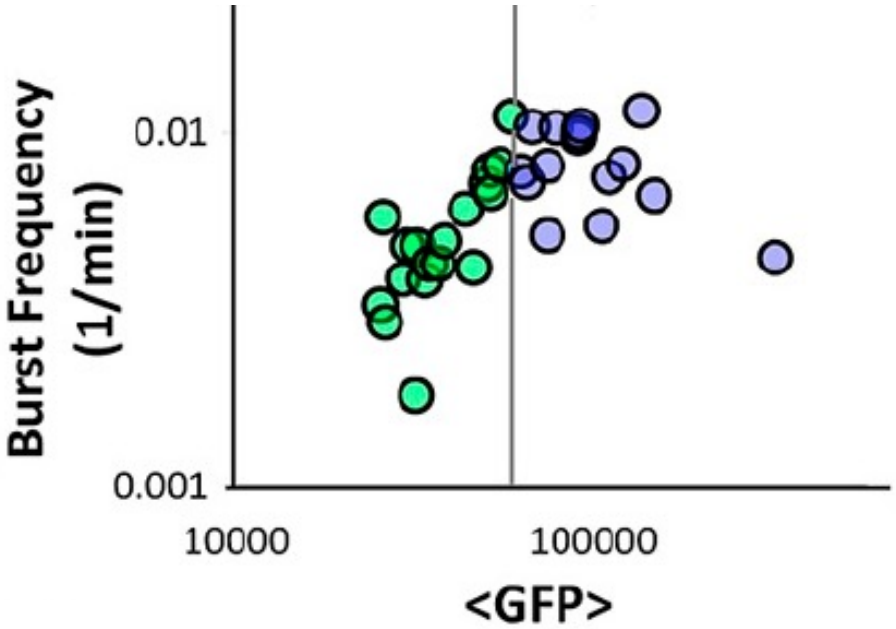
Singh et al. Biophysical Journal 2010
Singh et al. Nature MSB 2012
Dar et al PNAS 2012

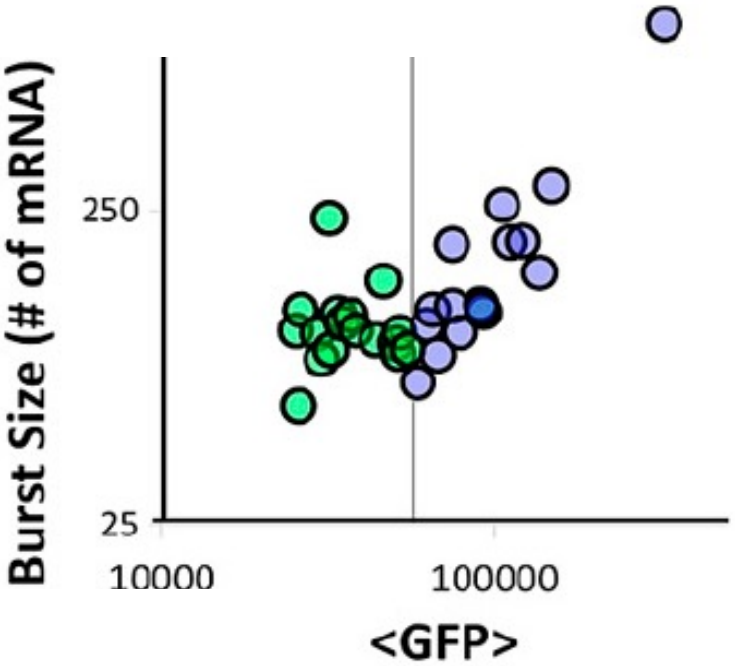
Site of integration in the human genome modulates burst frequency and size



Transcriptional burst size=k_m/k_{off}

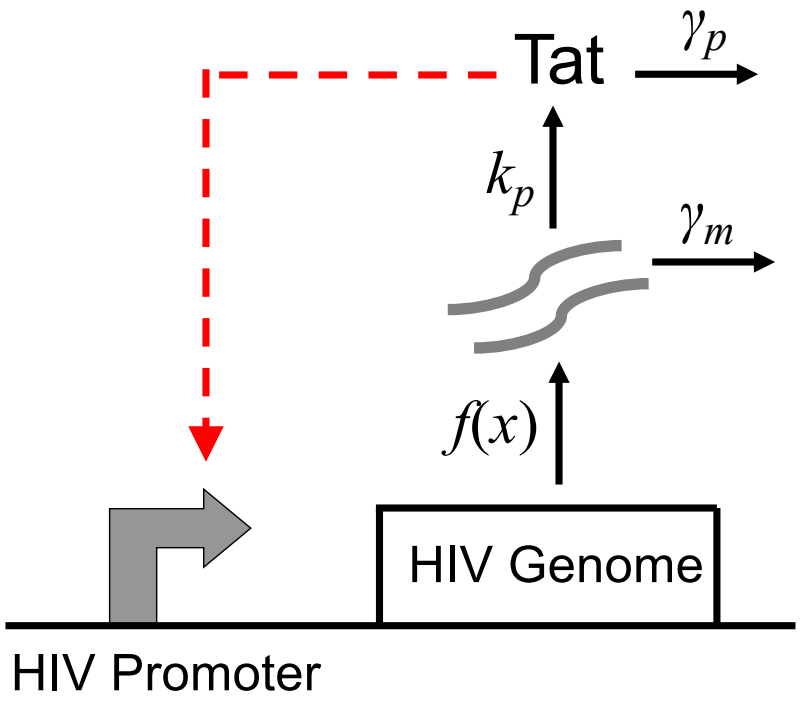
Transcriptional burst frequency=k_{on}





Proceeding of National Academy of Science 2012

Deterministic model of Tat feedback circuit



$m(t)$: mRNA count at time t

$x(t)$: Tat protein count at time t

$$\frac{dm}{dt} = f(x) - \gamma_m m$$

$$\frac{dx}{dt} = k_p m - \gamma_p x$$

k_m : Maximum transcription rate (High Tat)

$k_m b$: Basal transcription rate (No Tat)

c : Positive feedback strength

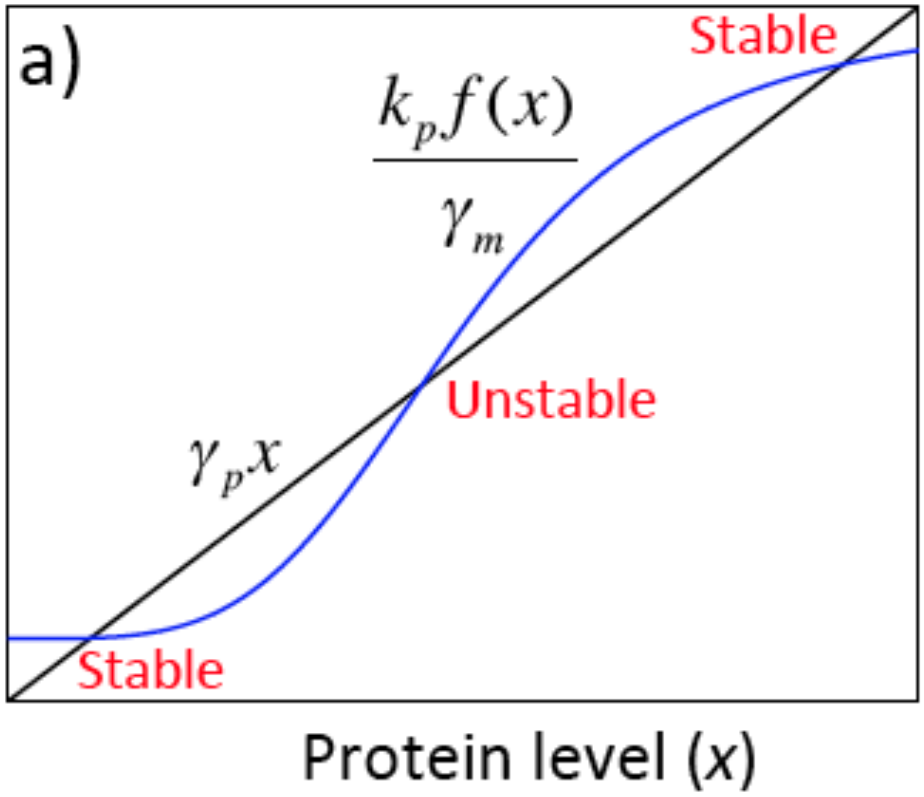
H : Hill Coefficient

$$f(x) = k_m \frac{b + (cx)^H}{1 + (cx)^H}$$

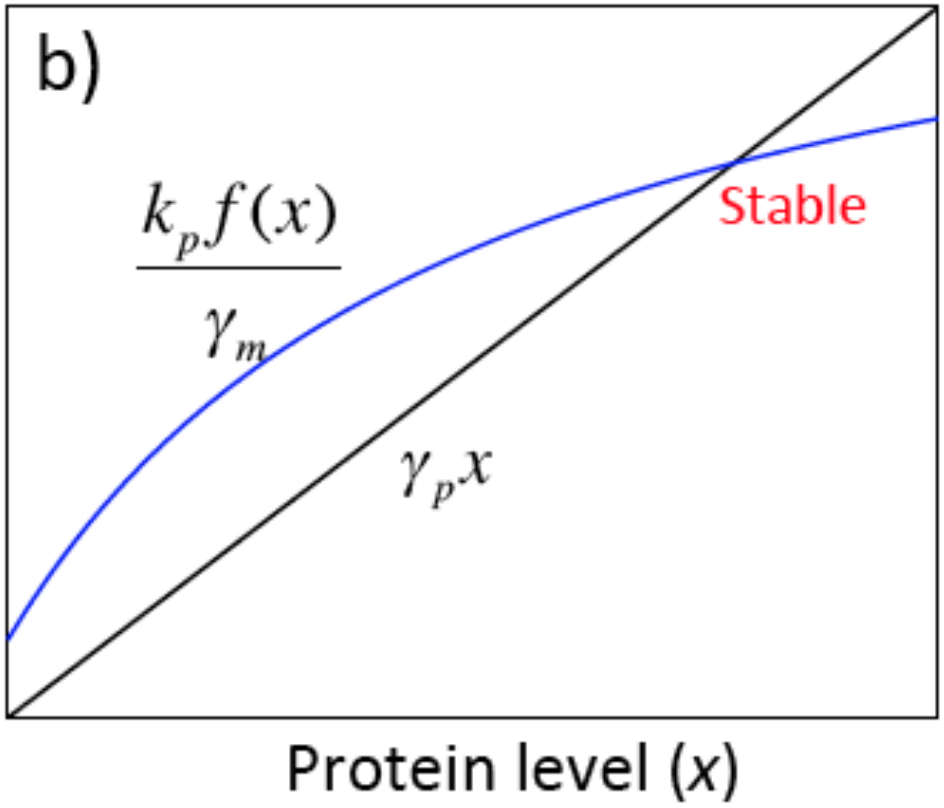
$$\frac{k_p}{\gamma_m} f(x) = \gamma_p x$$

High Hill coefficient is necessary for bistability

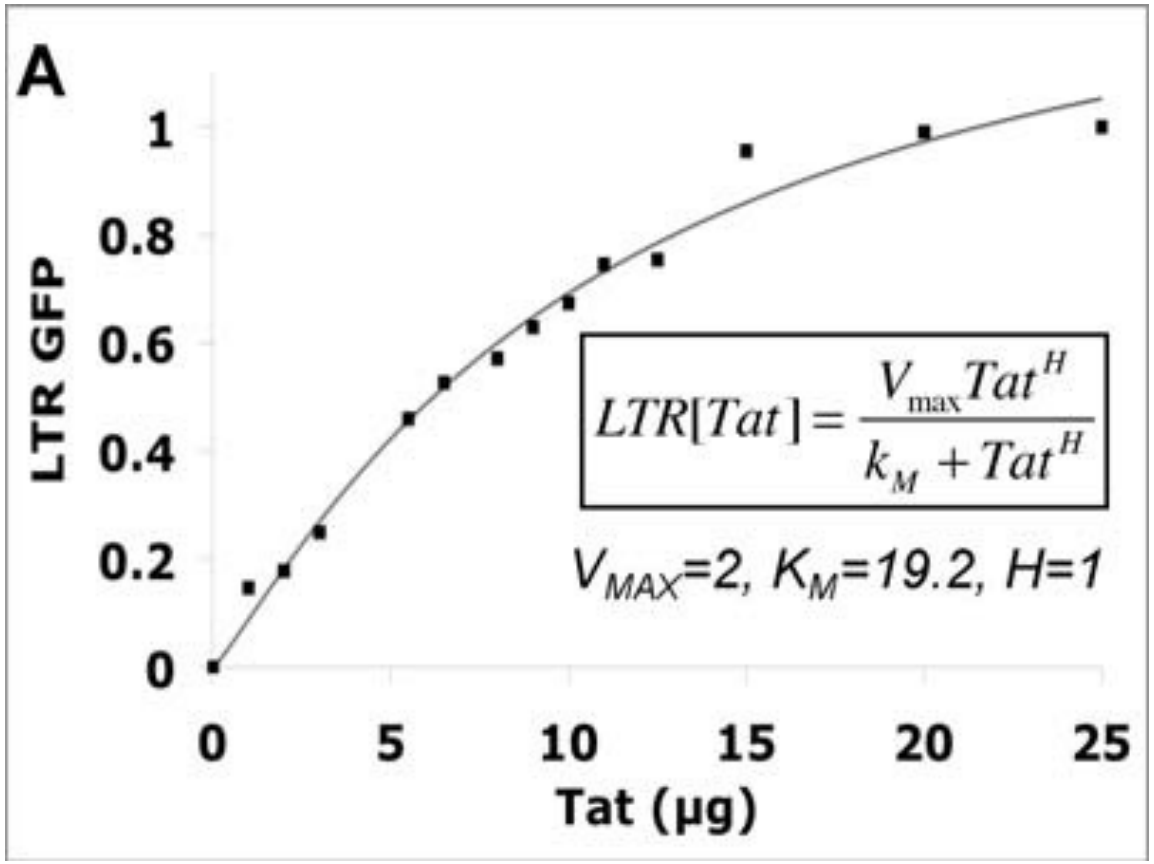
H>1 (Bi-stability)



H=1 (Mono-stability)

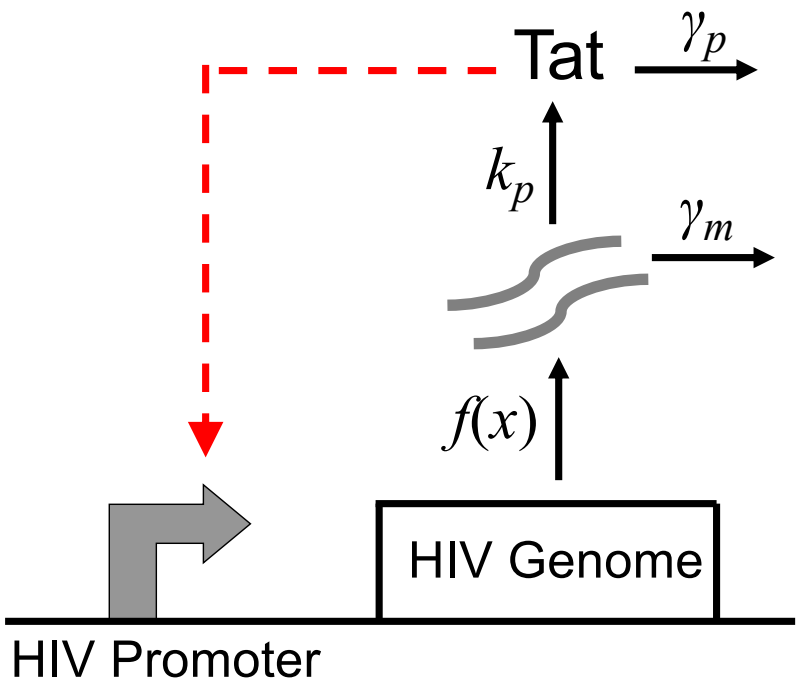


Tat positive feedback circuit lacks bistability



Weinberger et al. PLos Biology 2007

Stochastic model of Tat feedback circuit



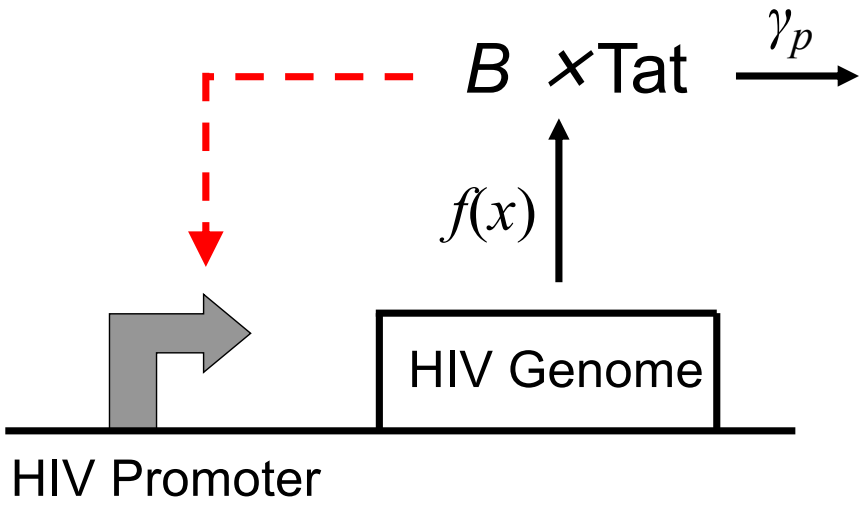
$m(t)$: mRNA count at time t

$x(t)$: Tat protein count at time t

Event	Reset in population count	Probability event will occur in $(t, t+dt]$
Transcription	$m(t) \to m(t)+1$	$f(x(t))dt$
mRNA degradation	$m(t) \to m(t)-1$	$\gamma_m m(t)dt$
protein translation	$x(t) \to x(t)+1$	$k_p m(t)dt$
protein degradation	$x(t) \to x(t)-1$	$\gamma_p x(t)dt$

Reduced model of Tat feedback circuit

$$\text{Probability}(B = z) = (1-\alpha)^z \alpha, \quad z = \{0,1,2,3,\ldots\}$$



$x(t)$: Tat protein count at time t

Event	Reset in population count	Probability event will occur in $(t, t+dt]$
Transcription	$x(t) \rightarrow x(t) + B$	$f(x(t))dt$
protein degradation	$x(t) \rightarrow x(t) - 1$	$\gamma_p x(t) dt$

Shahrezaei and Swain PNAS 2008

Analytical derivation of the steady-state Tat pdf

$P(i,t)$: Prob. of i Tat molecules at time t

$$\frac{dP(i,t)}{dt} = \sum_{z=0}^{i-1} (1-\alpha)^{i-z-1} \alpha f(z) P(z,t), \qquad i = \{0,1,2,\ldots\}$$
$$+\gamma_p (i+1) P(i+1,t) - (f(i) + \gamma_p i) P(i,t)$$

$$\frac{\bar{P}(i+1)}{\bar{P}(i)} = \frac{k_m (b+ci) + \gamma_p i (1-\alpha)(1+ci)}{\gamma_p (i+1)(1+ci)}$$

Connecting shape of Tat distribution with model parameters

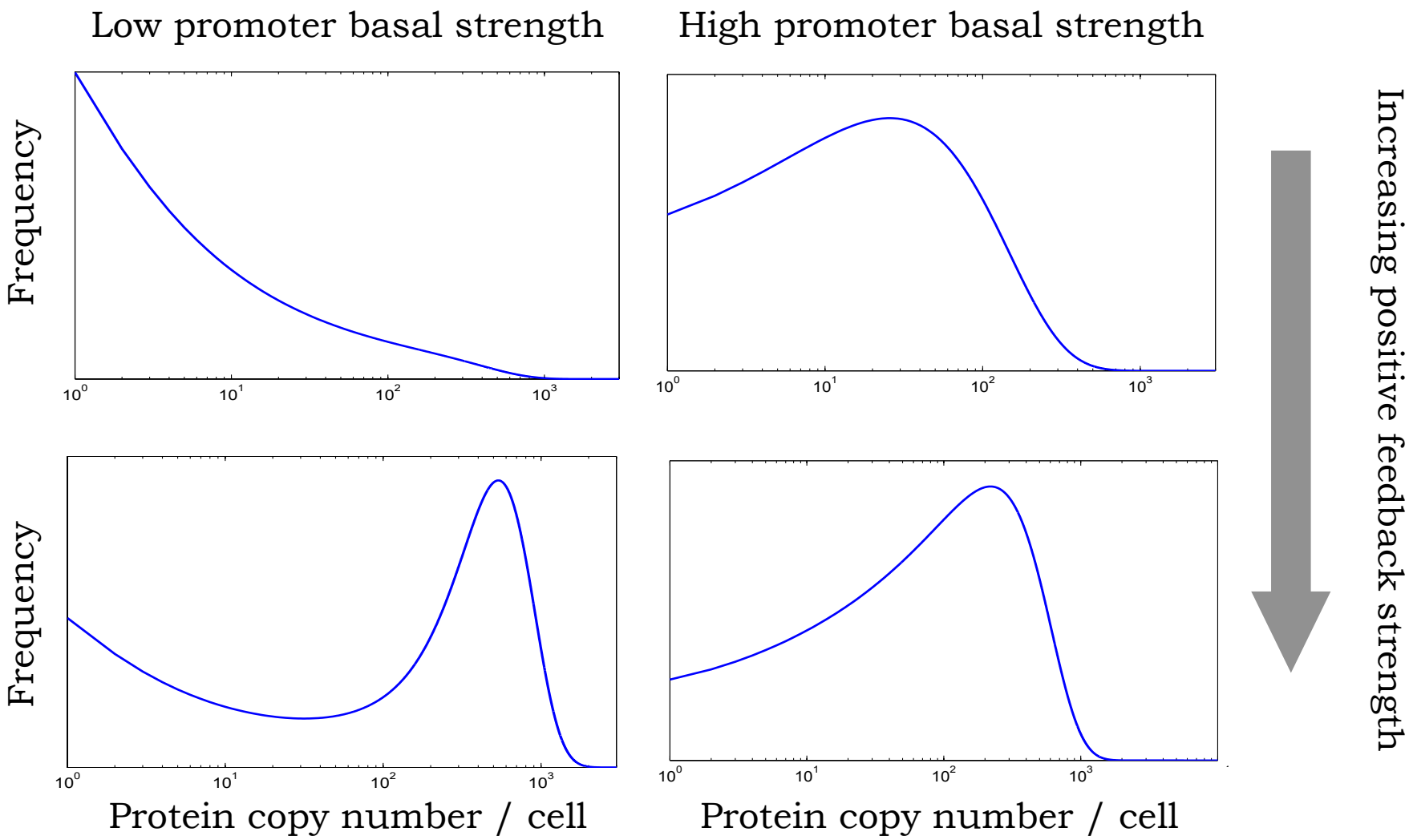
Theorem : Let $\bar{P}(i)$ be the steady-state Tat level distribution. Assuming $k_m > \gamma_p$

a) If $bk_m < \gamma_p$ and feedback strength $c < c^*$ then $\bar{P}(i)$ is unimodal with a zero mode.

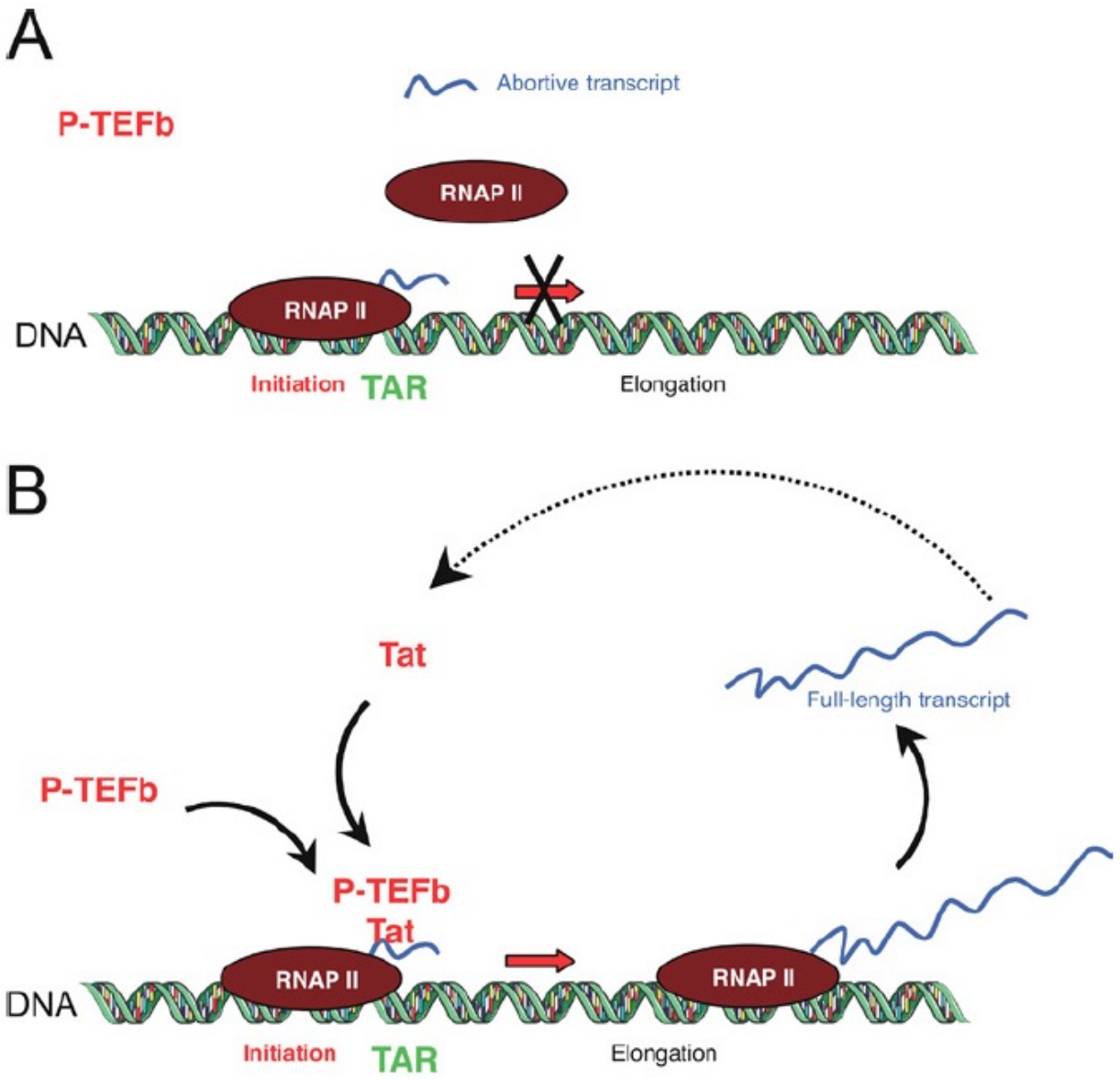
b) If $bk_m < \gamma_p$ and feedback strength $c > c^*$ then $\bar{P}(i)$ is *bimodal* with a zero and a non-zero mode.

c) If $\gamma_p < bk_m$, then $\bar{P}(i)$ is unimodal with a non-zero mode.

A low basal rate of production is necessary for bimodality



HIV promoter has a low basal rate of transcription

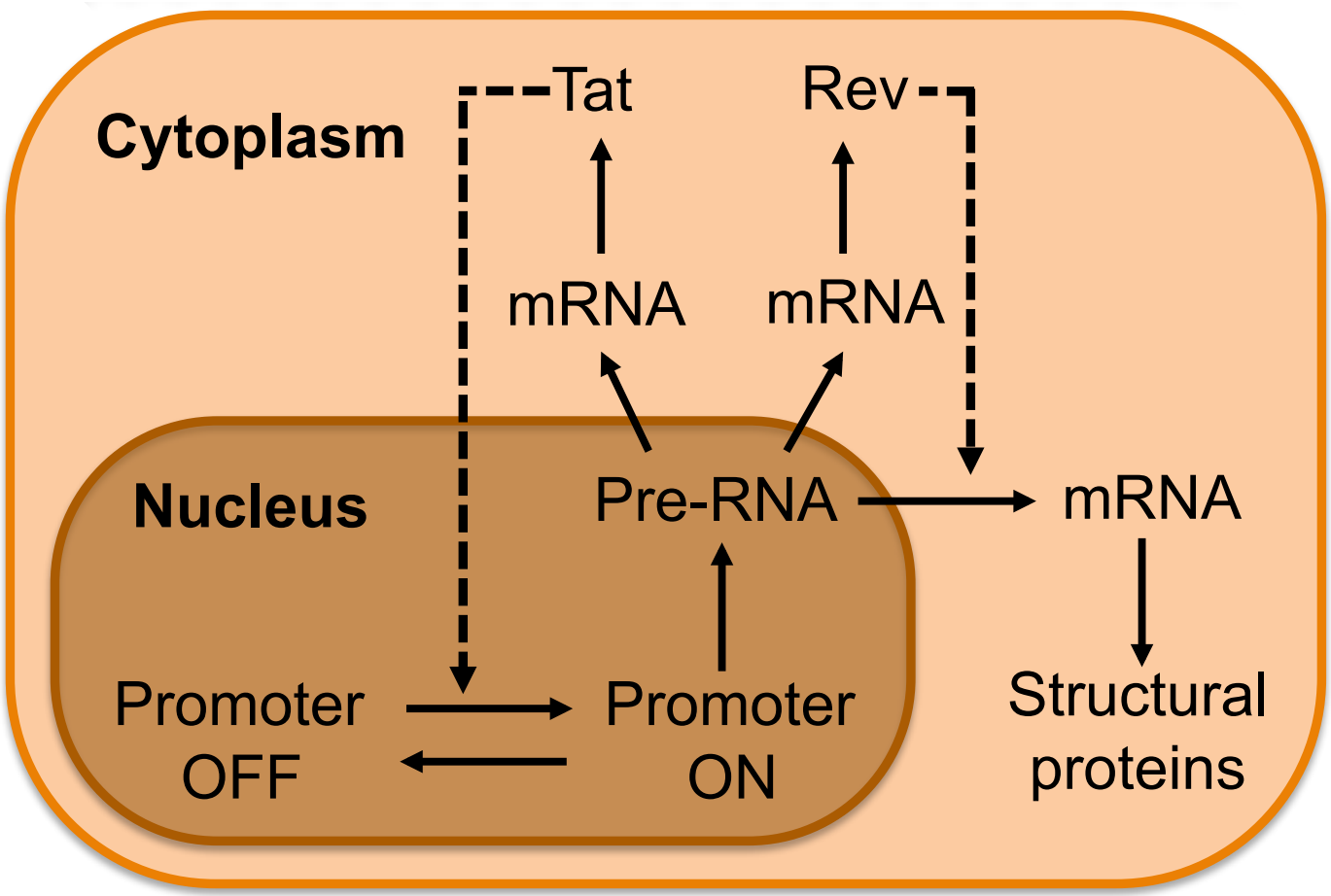


Barrandon et al.Biol. Cell 2008

Summary

- Analytical solution for steady-state Tat probability distribution
- Monostable system can generate bimodal distributions
- Bimodality requires a low basal rate of production and strong positive feedback
- Tat feedback circuit is in a regime to exhibit bimodality
- Feedback architecture has evolved to drive HIV into latency

Future Work


Cytoplasm
Tat
Rev
mRNA
mRNA
Nucleus
Pre-RNA
mRNA
Promoter
OFF
Promoter
ON
Structural
proteins

"Kick and kill" strategy to purge the latent reservoir

Current strategy

- Reactivate latent cells using small-molecule drugs (HDAC inhibitors)
- Cells are killed by the immune system or the budding virus.

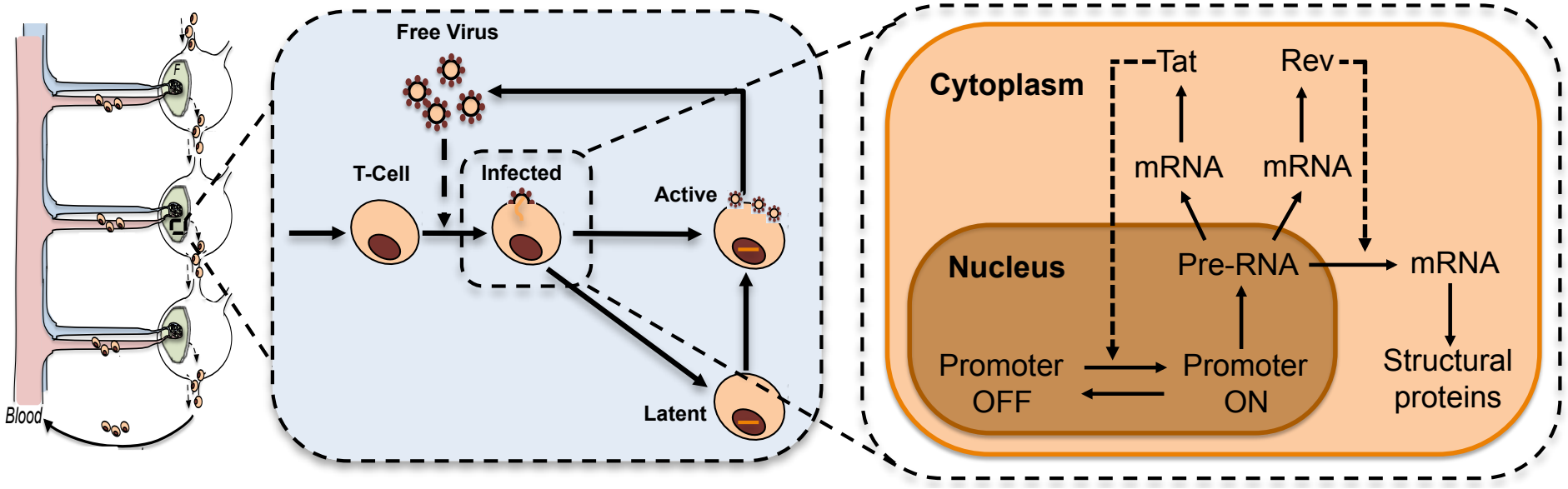
New strategy

- Reactivate latent cells.
- Second Tat-analog drug to bias infected cells against entering latency

Luo et al Royal Society Interface

Buzon et al Nature Medicine 2

Kick and kill strategy to purge the latent reservoir







SAN DIEGO