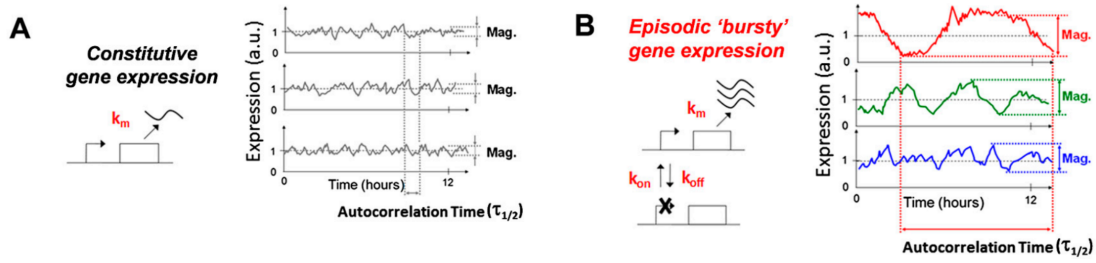
# Predictive spatial models of gene regulation via Bayesian inference

Clayton W. Seitz

June 12, 2022

# Gene expression is stochastic and non-constitutive




**Single-state models**

- Transcription occurs at a fixed rate
- mRNA counts are Poisson
- Underestimates variance in mRNA counts

**Two-state models**

- Promoter can be on or off
- mRNA counts are not Poisson

Chromatin structure is complicated e.g., loops. Multiple *cis*-regulatory elements can contribute to promoter switching and transcription control

# STL1/CTT1 induction with NaCl in yeast

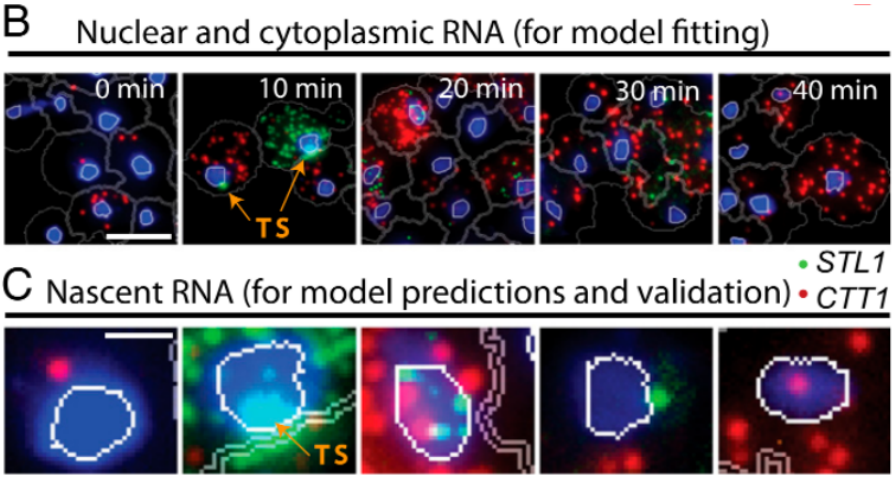



**Figure 1:** Munsky et al., PNAS 2018

# A quick note on ergodicity and ensemble snapshots

- Ergodicity = statistics of the ensemble and a single process are the same
- Implies the parameters of the process are the same for every cell
- Ensemble snapshots useful due to experimental constraints (multiplexing)

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^n = \int (x - \mu)^n p(x) dx$$

# A compartment model for spatial gene expression

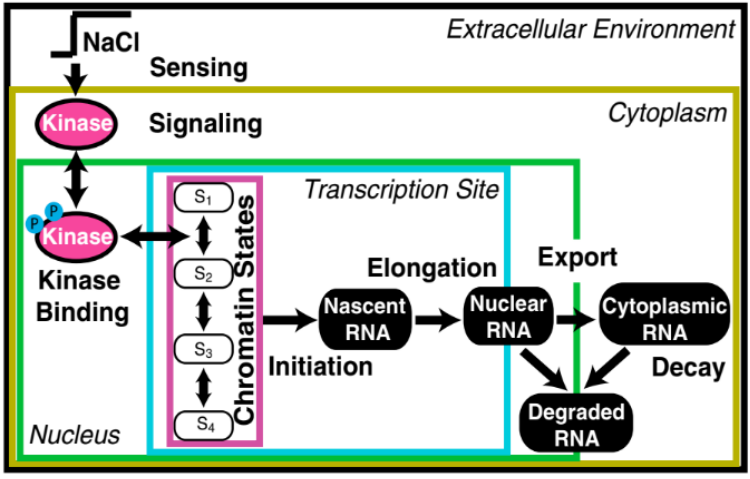



**Figure 2:** Munsky et al., PNAS 2018

# A compartment model for spatial gene expression

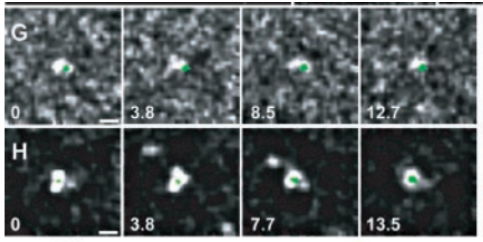
Let $X$ represent an arbitrary RNA transcript of gene $G$

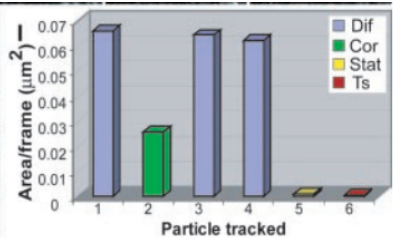
$$\begin{aligned}
\text{Gene activation} &: G_{off} \xrightarrow{k_{on}} G_{on} \\
\text{Gene inactivation} &: G_{off} \xrightarrow{k_{off}} G_{on} \\
\text{Transcription} &: G_{on} \xrightarrow{k_t} G_{on} + X_{\text{nuc}} \\
\text{RNA Export} &: X_{\text{nuc}} \xrightarrow{k_{exp}} X_{\text{cyt}} \\
\text{RNA degradation} &: X_{\text{cyt}} \xrightarrow{\gamma} \emptyset
\end{aligned}$$

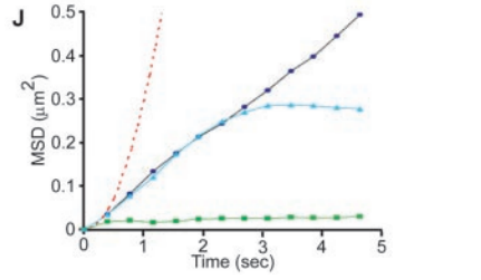
Raw data collected post induction can be used to infer parameters

$$\theta = (k_{on}, k_{off}, k_t, k_{exp}, \gamma)$$

# Brownian RNA transport in the nucleus




K

| Distribution % / Velocity (μm/sec) / D (μm²/sec) | Directed | Diffusive | Corralled | Stationary |
|---|---|---|---|---|
| | 0 | 58 | 42 | <1 |
| | | 0.55 | 0.54 | |
| | | 0.041 | 0.045 | |

**Figure 3:** Shav-Tal et al., Science 2004

# Rate of nuclear export is determined by diffusion

Particles have beens shown to be distinguishable (non-ergodic dynamics)

The probability current $J(r, t)$ determines the rate of RNA export

Even if diffusion is Brownian motion, the export rate parameter is not constant:

$$\frac{dP(x, y, t)}{dt} = D \cdot \nabla^2 P = D \frac{dJ(x, y, t)}{dt}$$

There is also the birth of new particles affecting the density

# An Ising model of promoter switching

Promoter activation often requires chromatin reorganization, binding of specific TF combinations. We can imagine a random (?) walk through the binding phase space

$$\mathcal{H} = -\frac{1}{2}\sum_{i,j} J_{ij} x_i x_j - \sum_j h_j x_j \quad P(\boldsymbol{x}) = \frac{1}{Z}\exp\left(-\beta\mathcal{H}(\boldsymbol{x})\right)$$

We know $p(x_j = 1) = \exp(-\beta H(x_i = 1))$. Then, we define

$$h_j = -\frac{1}{\beta}\log p(x_i = 1) = -\frac{1}{\beta}\log\frac{[x_j]}{K_j + [x_j]}$$

Suppose that there is a single state or set of states $\boldsymbol{x}^*$ for which the promoter is active

$$\lambda = \frac{Z_{on}}{Z_{on} + Z_{off}} \quad \rightarrow \quad P(n|\mu) = \frac{\mu^n}{n!}\exp(-\mu)$$

with $\mu = \lambda t$

# Bayesian parameter inference using ensemble snapshots

Suppose we have a series of ensemble snapshots of an *in-vitro* population:

$$\mathbf{x} = \{\mathbf{x}_0, ..., \mathbf{x}_t\} \quad \mathbf{y} = \{\mathbf{y}_0, ..., \mathbf{y}_t\}$$

with $\mathbf{x}_t = \{x_1, ..., x_n\}$ and similarly for $\mathbf{y}$. Under perfect measurements $\mathbf{x} = \mathbf{y}$

We would like to use $\mathbf{x}$ to fit a dynamical model $\mathcal{M}(\theta)$. Bayesian inference lets us infer $\theta$ from $\mathbf{x}$ while quantifying the uncertainty in our estimate:

$$P(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta) = \pi(\theta)\prod_t f(\mathbf{x}_t|\theta)$$

The likelihood $f(\mathbf{x}_t|\theta)$ is often difficult to define or intractable to compute due to the curse of dimensionality, making even MLE a challenge

# Kolmogorov's forward equation (chemical master equation)

Dynamics on biochemical reaction networks are inherently stochastic and the state space is discrete. We can only write probabilities over the state space

$$\begin{aligned} P(\mathbf{x}_i, t) &= \sum_j T_{ji}(\mathbf{x}_i, t|\mathbf{x}_j, t - \Delta t) P(\mathbf{x}_j, t - \Delta t) \\ &= \sum_k T_k(\mathbf{x}_i, t|\mathbf{x}_i - \nu_k, t - \Delta t) P(\mathbf{x}_i - \nu_k, t - \Delta t) \end{aligned}$$

where $T_k$ is the probability of a reaction channel $k$ firing in the interval $(t, t + \Delta t)$.

Taking the limit $\Delta t \to 0$ one can derive the forward Kolmogorov equation or chemical master equation (CME)

$$\frac{dP(\mathbf{x}, t|\mathbf{x}_0)}{dt} = \sum_k T_k(\mathbf{x} - \nu_k) P(\mathbf{x} - \nu_k, t) - T_k(\mathbf{x}) P(\mathbf{x}, t)$$

# References I