

Predictive spatial models of gene regulation via Bayesian inference

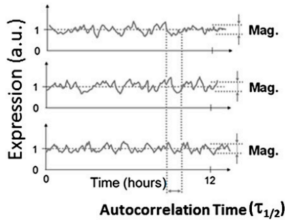
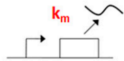
Clayton W. Seitz

June 9, 2022

Gene expression is stochastic and non-constitutive

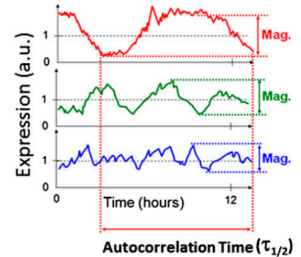
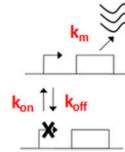
A

**Constitutive
gene expression**



B

**Episodic 'bursty'
gene expression**



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

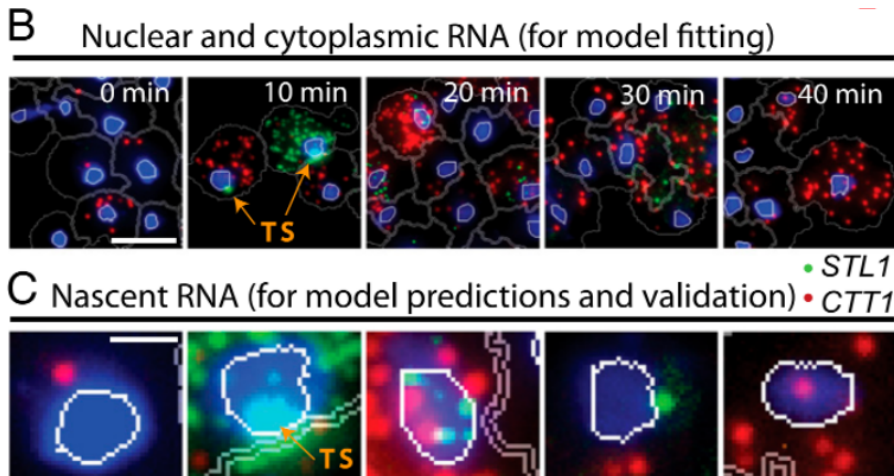


Figure 1: Munsky et al., PNAS 2018

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)

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

A compartment model for spatial gene expression

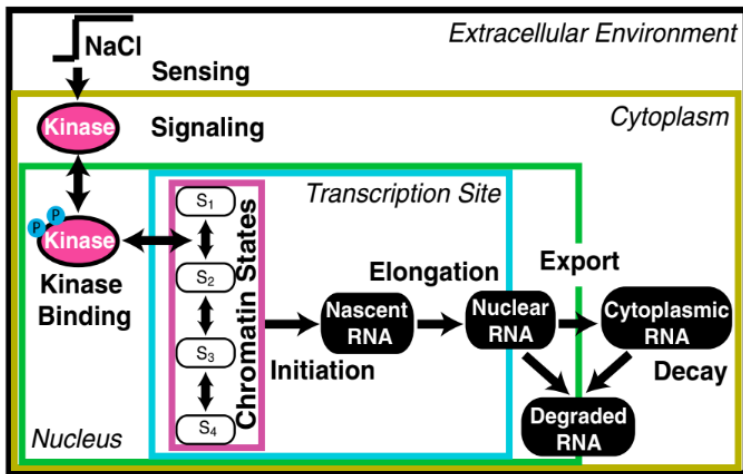
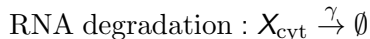
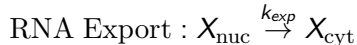
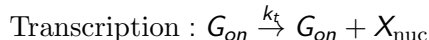
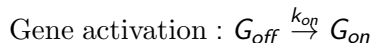


Figure 2: Munsky et al., PNAS 2018

A compartment model for spatial gene expression

Let X represent an arbitrary RNA transcript of gene G



Raw data collected post induction can be used to infer parameters

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

Brownian RNA transport in the nucleus

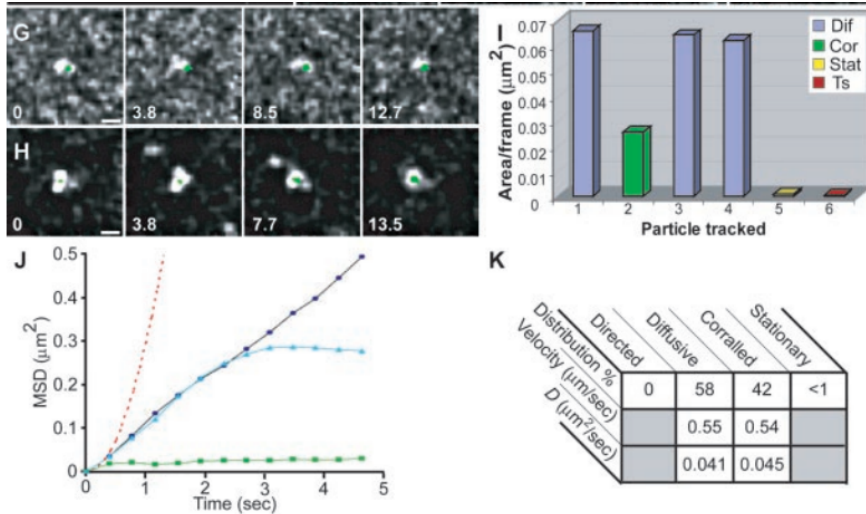


Figure 3: Shav-Tal et al., Science 2004

Rate of nuclear export is determined by diffusion

Particles have been 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

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

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 \mathbf{x}^* for which the promoter is active

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

with $\mu = \lambda t$

Bayesian parameter inference using ensemble snapshots

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

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

with $\mathbf{x}_t = \{x_1, \dots, x_n\}$ 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 θ 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 $\Delta t \rightarrow 0$ 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