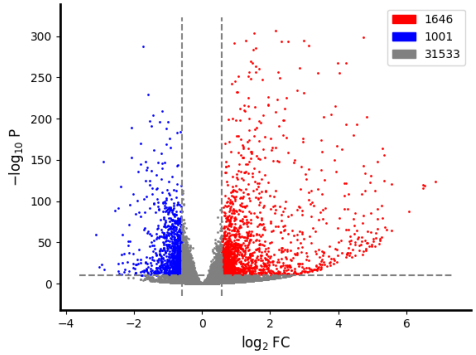
Establishing a quantitative framework for analyzing inducible gene expression in HeLa cells

Clayton W. Seitz

July 4, 2022

Interferon-γ induces differential expression of many genes

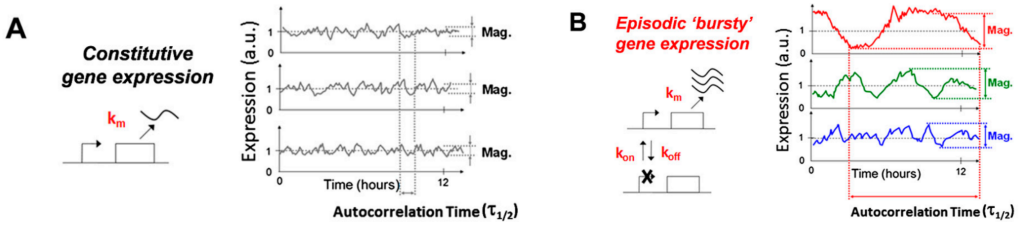
Single cell transcriptome measurements of polyA mRNA for naïve HeLa cells (N=90), induced with interferon gamma (50ng/mL) for 24h



Siwek et al. *Activation of Clustered IFNγ Target Genes Drives Cohesin-Controlled Transcriptional Memory Cell 2020*

This is just a birds eye view of whats really going on...

Promoter models are necessary for non-constitutive gene expression



Single-state models

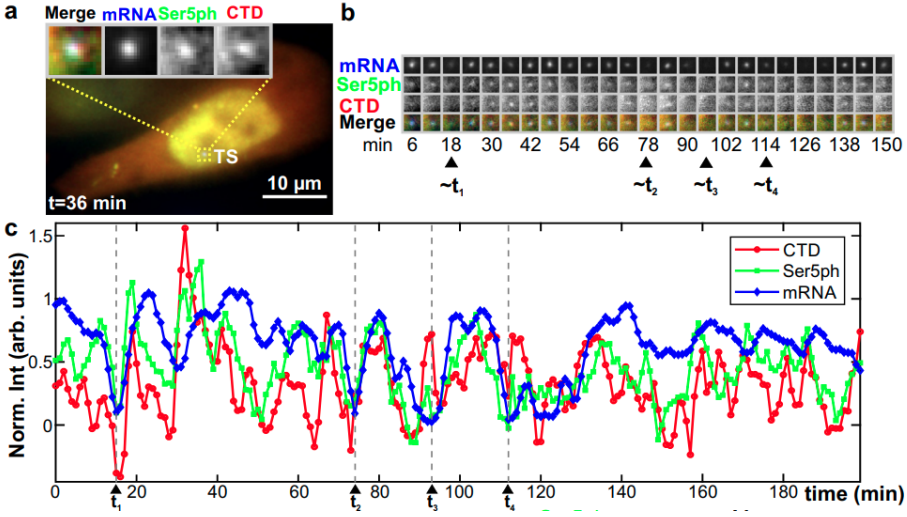
- RNAs are 'born' at a fixed rate
- RNA counts are Poisson

Multi-state models

- Promoter can be in multiple states (switching behavior)
- RNA counts are not Poissonian

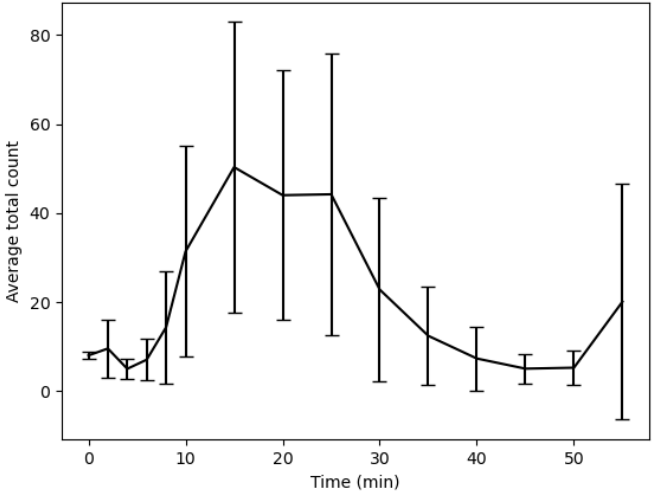
Single-state models tend to underestimate variance in RNA counts

Gene expression is stochastic (live-cell MS2-MCP)



Forero-Quintero, et al. *Live-cell imaging reveals the spatiotemporal organization of endogenous RNAPII phosphorylation at a single gene.* Nat Commun 2021

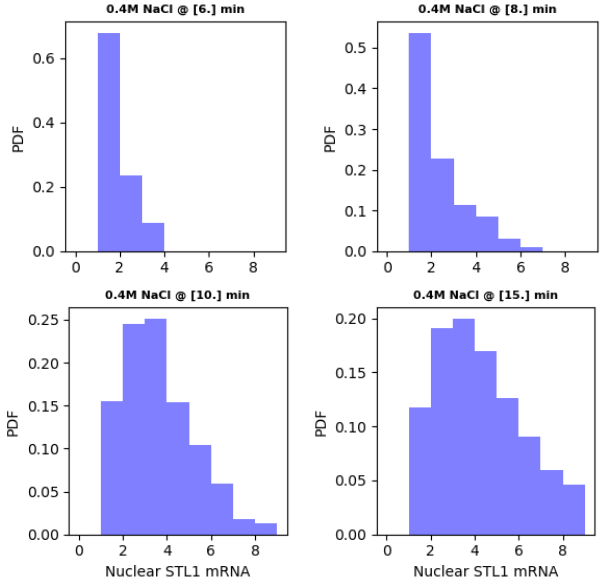
Example: variability in STL1 mRNA counts per cell at 0.4M NaCl



Error bars represent standard deviations from the mean
Cells marked ON for $>$ 3 STL1 mRNA in yeast

Assessing STL1 mRNA count variability at the transcription site

- Brightest spot in the nucleus defined as putative TS
- TS marked ACTIVE if $I > 2 * \mathrm{med}$
- Nascent mRNA count is $\mathrm{round}(I/\mathrm{med})$
- Count variability suggests asynchrony



A spatial model for induced gene expression

Let X represent an arbitrary RNA transcript of an induced gene G. Assume two promoter states (on and off)

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

Raw data collected post induction can be used to infer parameters

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

Bayesian inference of model parameters

It is well-known that using just means and variances gives poor estimates of the model parameters (Munsky et al. PNAS 2018)

Let $\theta = (k_{on}, k_{off}, k_t, k_{exp}, \gamma)$. Using Bayes Rule:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)P(\theta)} \propto P(X|\theta)P(\theta)$$

Can infer θ if we know the likelihood $P(X|\theta)$ (the hard part) and specify a prior $P(\theta)$

Generally we have to resort to Monte Carlo methods to find $P(X|\theta)$

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

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

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)

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