

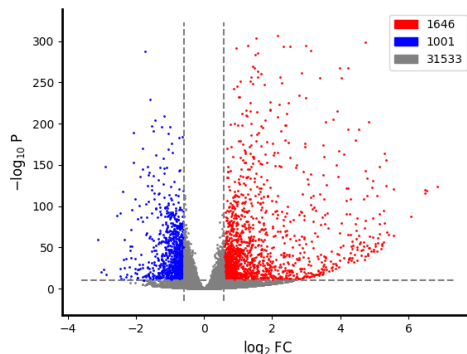
# Establishing a quantitative framework for analyzing IFN $\gamma$ induction in HeLa cells

Clayton W. Seitz

June 28, 2022

# Interferon- $\gamma$ induces differential gene expression in HeLa cells

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



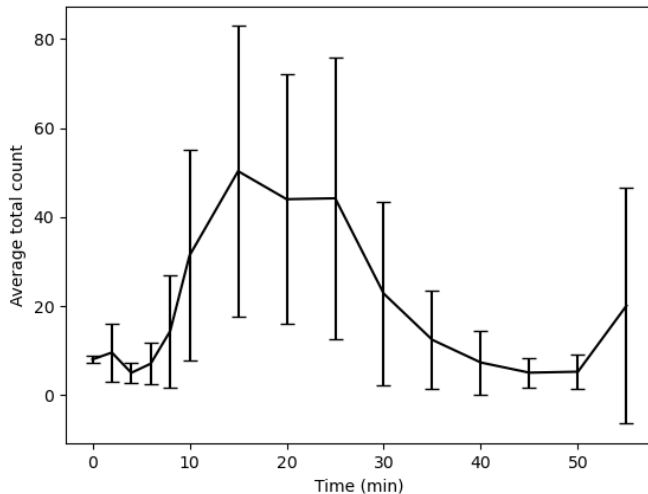
Siwek et al. *Activation of Clustered IFN $\gamma$  Target Genes Drives Cohesin-Controlled Transcriptional Memory* Cell 2020

# Key questions

- ▶ Does IFN $\gamma$  induce transcriptional bursts in HeLa cells?
- ▶ Which genes?
- ▶ What are the parameters of the burst (size, frequency, etc.)?
- ▶ In general, it possible to correlate spatial patterning with transcriptional bursting, using only ensemble snapshots (FISH)?

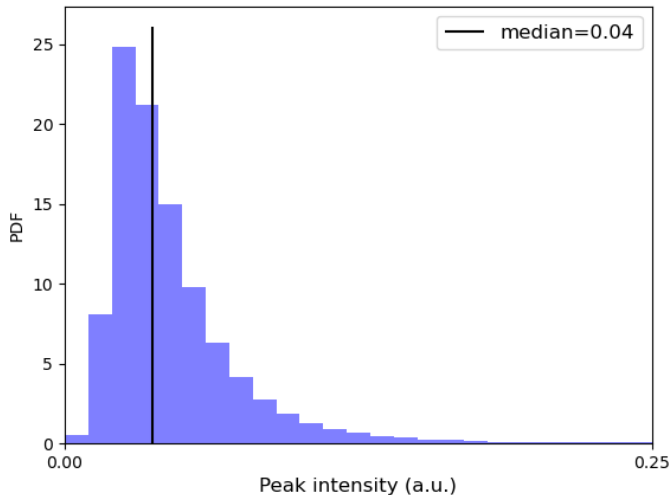
Need to establish analysis methods on a testing dataset...

## Significant 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

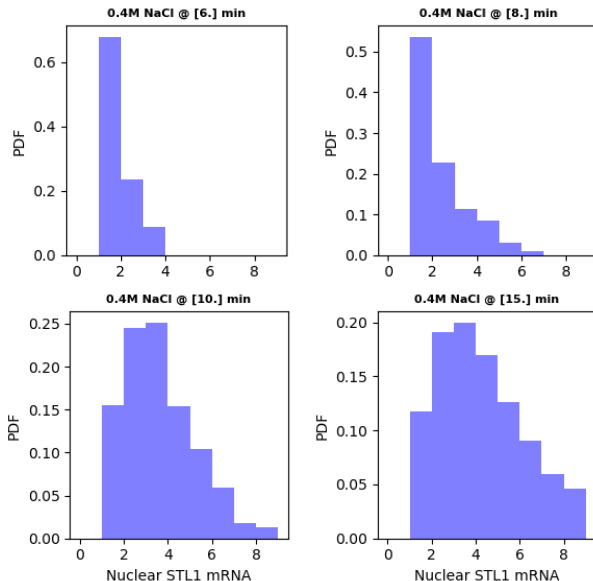
# Assessing STL1 mRNA count variability at the transcription site



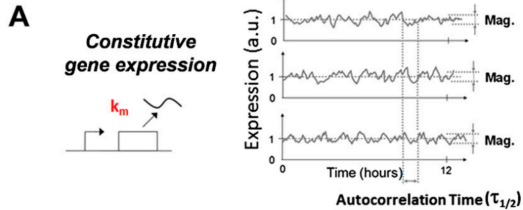
The median of the mRNA intensity distribution is used to determine the number of nascent RNA at the transcription site (TS)

# Assessing STL1 mRNA count variability at the transcription site

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

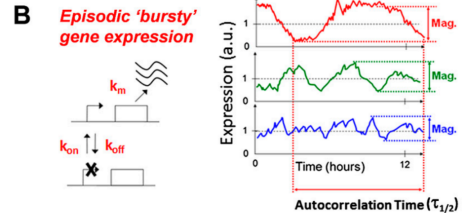


# Gene expression is stochastic and non-constitutive



## Single-state models

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

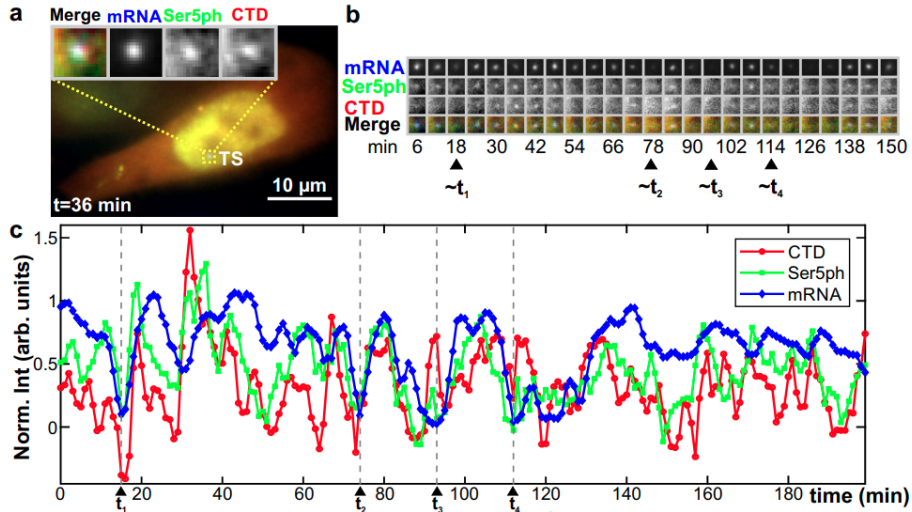


## Multi-state models

- ▶ 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 and non-constitutive (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

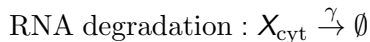
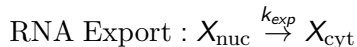


# Ensemble averages and variances do not fully explain underlying transcription dynamics

- ▶ Transcription is stochastic, meaning that RNA counts can only be understood in terms of a probability distribution
- ▶ High variance in mRNA counts suggests more complicated underlying dynamics which are not evident in ensemble averages
- ▶ We cannot assume that cells are bursting synchronously
- ▶ Bursting phase has implications for correlating bursting with spatial organization via ensemble data (FISH)
  - ▶ Classification of cells based on  $P(X_{nuc}, X_{cyto}, X_{TS})$ ?
  - ▶ Can also look at the evolution of the spatial feature distributions

## A compartment model for IFN $\gamma$ induced gene expression

Let  $X$  represent an arbitrary RNA transcript of IFN $\gamma$  induced gene  $G$ . Assume two chromatin states (on and off)



Raw data collected post induction can be used to infer parameters

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

# Bayesian parameter inference using ensemble snapshots

Likelihood-based methods can infer  $\theta$  from ensemble snapshots (FISH data)

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

One way is through maximum a posteriori estimation (MAP):

$$\theta^* = \operatorname{argmax}_{\theta} P(X|\theta)$$

A more robust (but harder) way is via Bayesian inference, which lets us infer  $\theta$  from  $X$  while quantifying the uncertainty in our estimate:

$$P(\theta|x) \propto P(x|\theta)\pi(\theta) = \pi(\theta) \prod_t P(x_t|\theta)$$

The likelihood  $P(X, t)$  is the solution to the chemical master equation at time  $t$

## 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

$$\begin{aligned} P(x_i, t) &= \sum_j T_{ji}(x_i, t | x_j, t - \Delta t) P(x_j, t - \Delta t) \\ &= \sum_k T_k(x_i, t | x_i - \nu_k, t - \Delta t) P(x_i - \nu_k, t - \Delta t) \end{aligned}$$

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)

$$\frac{dP(x, t | x_0)}{dt} = \sum_k T_k(x - \nu_k) P(x - \nu_k, t) - T_k(x) P(x, t)$$