

# Interferon-stimulated transcriptional bursting in space and time

C.W. Seitz

June 12, 2022

## **1 Introduction: Interferon-stimulated gene expression**

Motivate the study of interferon stimulated transcriptional bursting. Why is interferon stimulated gene expression important and what are the potential consequences of non-constitutive gene expression?

## **2 Transcriptional bursting**

Summarize statistical signatures of transcriptional bursting (Poisson versus non-Poisson gene expression)

## **3 Promoter switching in mammalian nuclei**

Summarize the origins of bursting at the promoter sequence

## **4 Testing the ergodic hypothesis**

Discuss ergodic and non-ergodic regimes of transcriptional bursting. This determines the statistical equivalence of ensemble snapshots and time-series data

## **5 Diffusion of nascent RNA in live-cell nuclei**

Most models thus far have focused on inferring kinetic parameters at a genomic locus with the fate of nascent transcripts left unaddressed. This could be, in part, because evidence has surfaced that transcripts in the nucleus primarily exhibit Brownian motion on their way to the nuclear pores at the nuclear membrane (Singer). At the same time, the pulsation of RNA production followed by Brownian motion makes the spatial arrangement of RNA near the transcription site non-trivial, due to the overlap of multiple distributions, complexities

of RNA elongation and cleavage, etc.. This also is in conflict with the nuclear to cytoplasmic export model used in compartment models of gene expression (Munsky). A first goal is to verify that motion in the nucleus is indeed Brownian while transcription is bursty in live cells (Quintero). These can be analyzed simultaneously, determining the validity of model assumptions along the way. Quintero et al. have published the dynamics at the transcription site for 14 cells imaged at 7Hz, which could be quite useful for testing ergodicity of transcriptional bursting. Two of these cells show 50 nascent transcripts which can be tracked and their dynamics characterized, providing insight for spatio-temporal models. Their data also correlates RNAPII binding with RNA elongation, which may be helpful information for dynamical models.

If RNA copy number is an ergodic process, is the diffusion also ergodic? What type of diffusion do we observe and why?

## 6 Bayesian inference for spatial models of gene expression

In-vivo data is ideal for kinetic parameter inference as it requires minimal assumptions and is compatible with mature dynamical modeling techniques, such as hidden Markov models (Rattray). However, studying transcription in-vivo requires a much more intensive biological component with respect to in-situ methods. Also, in-situ methods are scalable in the sense that many unique RNA species can be fluorescently tagged simultaneously, providing a substrate for the study of coregulation of genes.

Due to these complications, kinetic inference using in-situ datasets has been well-studied in the literature, particularly by fitting models of promoter switching to ensemble snapshot time sequences (Munsky). Bayesian inference of kinetic parameters has become quite popular for in-situ datasets as this class of methods does not require detailed time-series data of transcription. Under the so-called ergodic assumption, the statistics of ensemble snapshots and time-series data are identical and useful for parameter inference. One can imagine that, when cells are indistinguishable in this way, the probability density over transcript counts is the same for the ensemble as for the time-resolved cell. In any case, computing the likelihood function for these data is generally intractable, and we have to make use of supplementary methods. Munsky et al. have performed a nice comparison study for fitting models where the likelihood which has been reduced to first and second order moments or computed explicitly by solving a chemical master equation (CME). Samples from the posterior on kinetic parameters are found in all cases via a MCMC algorithm, giving parameter uncertainties. Likelihood-free methods such as approximate Bayesian computation (ABC) avoid computing the likelihood, instead defining a notion of distance between the observed data and simulated trajectories using the Gillespie algorithm.

Compartment models define promoter states and export of RNA out of the

nucleus. Inference on compartment models is performed by using the finite state projection of the chemical master equation to compute the likelihood of parameters. RNA export rates are not diffusion-based in existing models, can we do this?

## **A Kolmogorov’s forward equation**

## **B Photon statistics for fluorescence microscopy**

## **C Single molecule localization: The Cramer-Rao bound**

Discuss parameter uncertainty, Fisher information and CR bound for RNA localization. This is important for achieving accuracy in particle detection and diffusion models