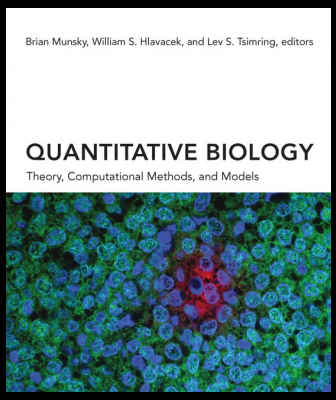
UQ-Bio Summer School 2022
Measuring, Modeling and Predicting Gene Regulation Dynamics at the Single-Cell Level

Brian Munsky

Chemical and Biological Engineering
School of Biomedical Engineering
Colorado State University

munsky@colostate.edu

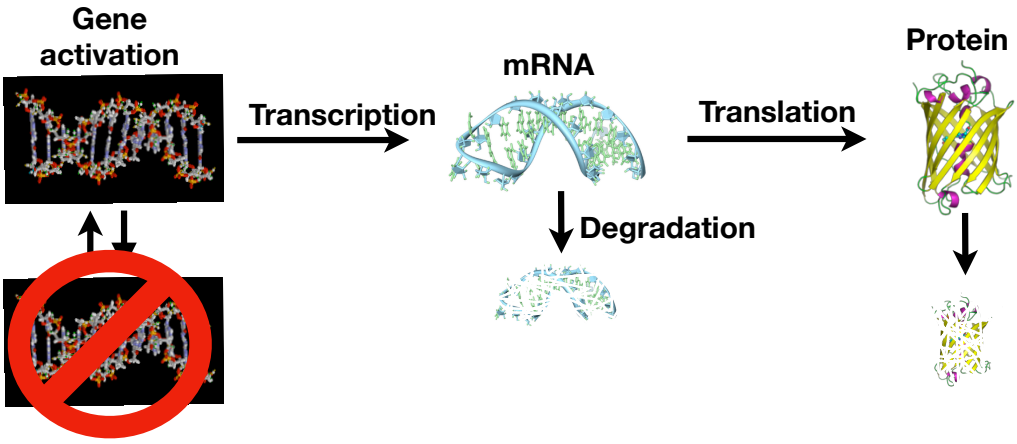
June 1, 2022



NIGMS W.M. KECK FOUNDATION NSF CAREER Award

munsky@colostate.edu

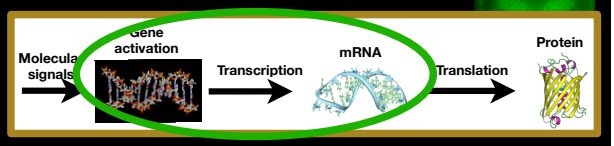
Introduction - The Central Dogma of Molecular Biology



Genetically identical cells in identical environments produce **stochastic, spatial, temporal fluctuations**.

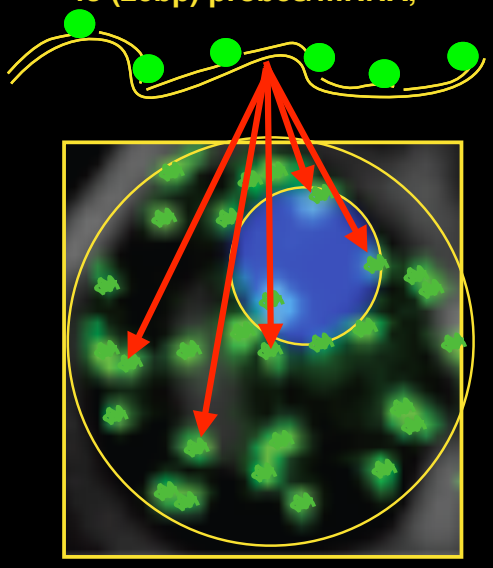
Our goal is to **measure, model,** and **predict every** stage of these fluctuations.

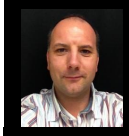
Measuring single-cell transcription using Single-Molecule Fluorescence in situ Hybridization (smFISH)



- SM-FISH allows quantification of endogenous transcription response:
  - **Number** of *individual* mRNA per cell,
  - **3D Location** of *individual* mRNA,
  - **DNA transcription site** activity,

48 (20bp) probes/mRNA,

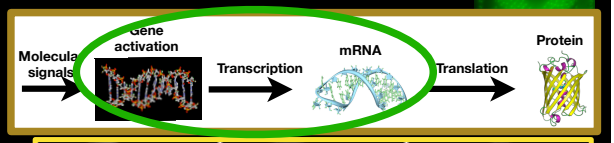




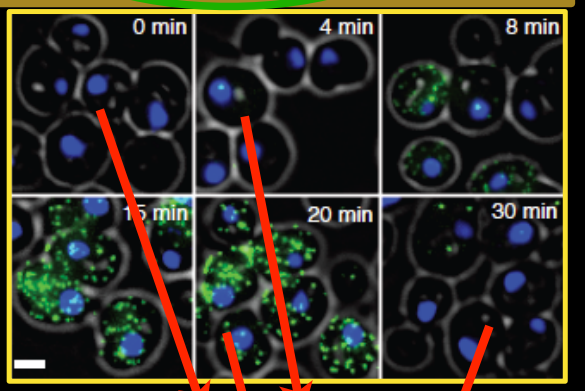
Gregor Neuert, Vanderbilt

Neuert, Munsky, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018

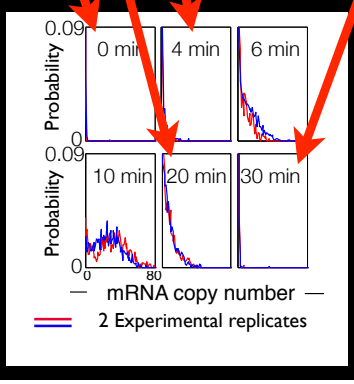
Measuring single-cell transcription using Single-Molecule Fluorescence in situ Hybridization (smFISH)

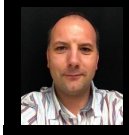


- SM-FISH allows quantification of endogenous transcription response:
  - **Number** of *individual* mRNA per cell,
  - **3D Location** of *individual* mRNA,
  - **DNA transcription site** activity,
  - Fast (1-2 minute) time resolution,
  - 100s or 1000s of cells per time point or condition.



smFISH provides highly ***reproducible & quantitative*** statistics of (noisy) single-cell responses.

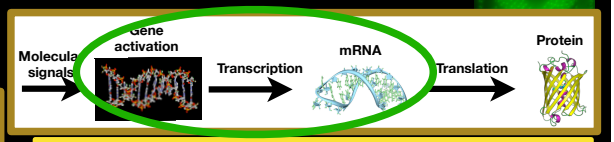


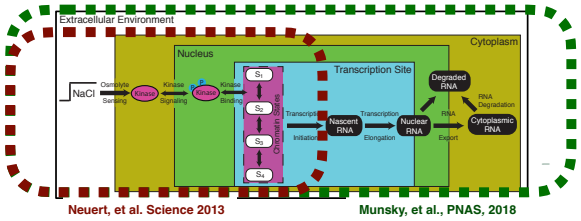


Gregor Neuert, Vanderbilt

Neuert, Munsky, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018

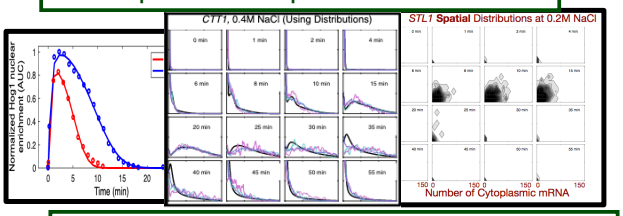
Measuring single-cell transcription using Single-Molecule Fluorescence in situ Hybridization (smFISH)



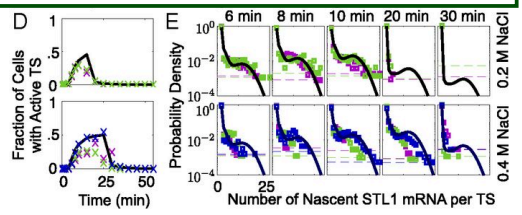


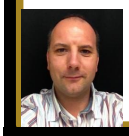
Neuert, et al. Science 2013 — Munsky, et al., PNAS, 2018

FIT Spatial and Temporal Stochastic Data

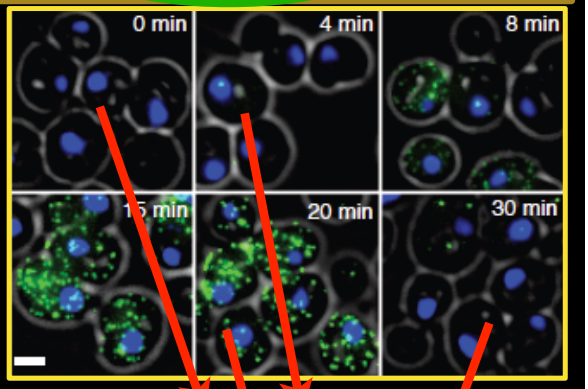


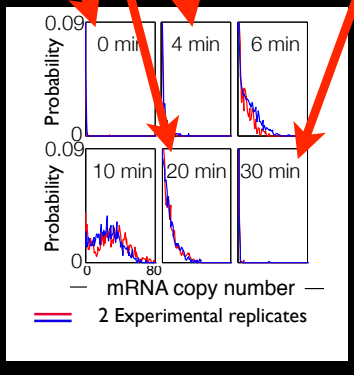
PREDICT Transcription Site Activity





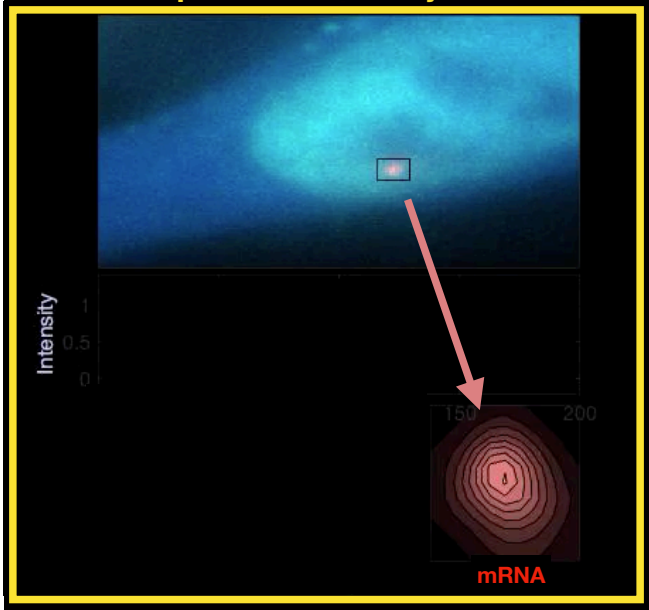
Gregor Neuert, Vanderbilt

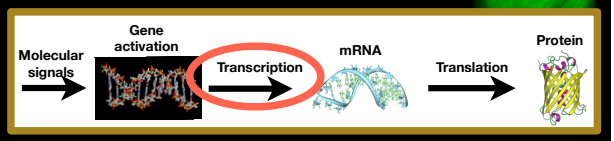




Neuert, Munsky, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018

RNAP2 Recruitment, Phosphorylation and Transcription is a Bursty Process





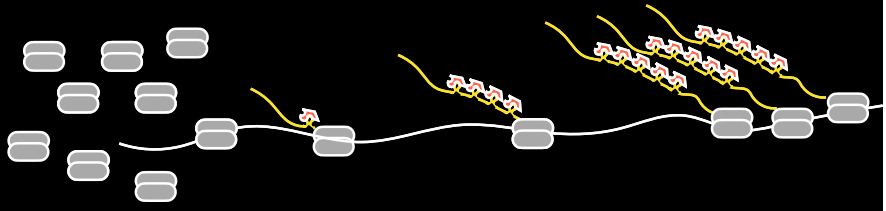
- Using MS2/MCP labeling, we observe *live nascent RNA transcription*.

MS2/MCP Label

Linda Forero-Quintero

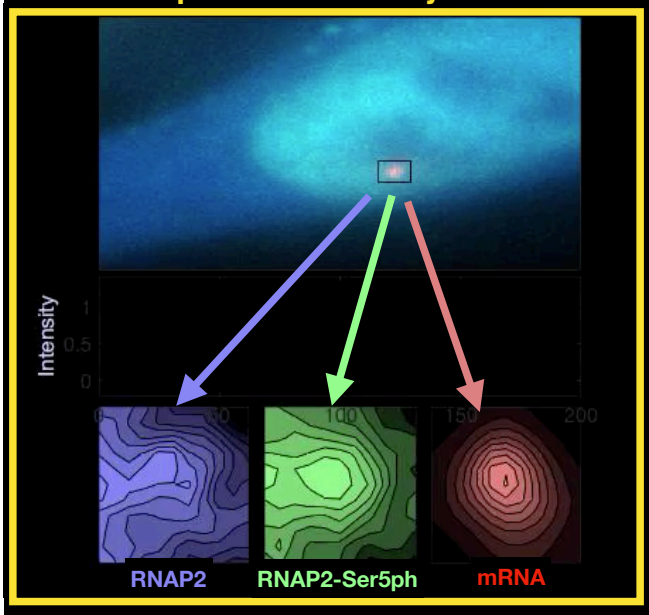
Will Raymond

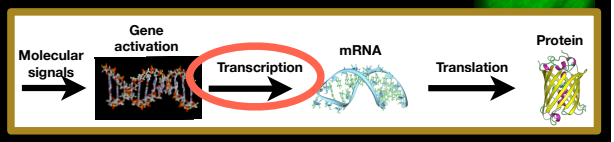
Tim Stasevich



Forero, Raymond et al, *Nat. Comms.*, 2021

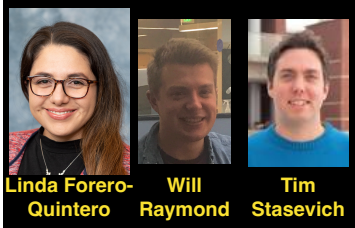
RNAP2 Recruitment, Phosphorylation and Transcription is a Bursty Process



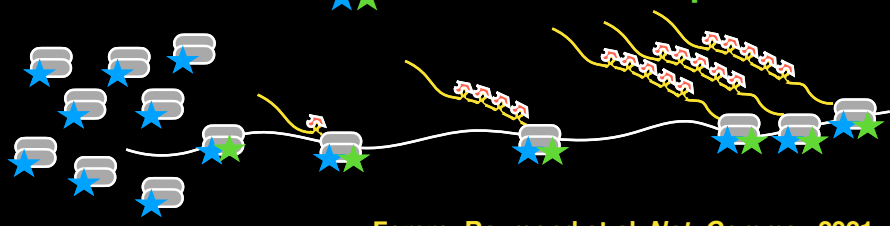


- Using MS2/MCP labeling, we observe *live nascent RNA transcription*.
- Fragmented antibody (FAB) probes allow us to quantify RNA Polymerase II (RNAP2) before (green) and after (green+blue) Ser5 phosphorylation.

MS2/MCP Label
RNAP2-CTD
RNAP2-CTD-Ser5ph

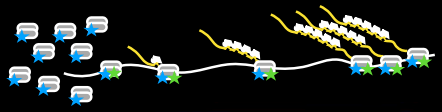


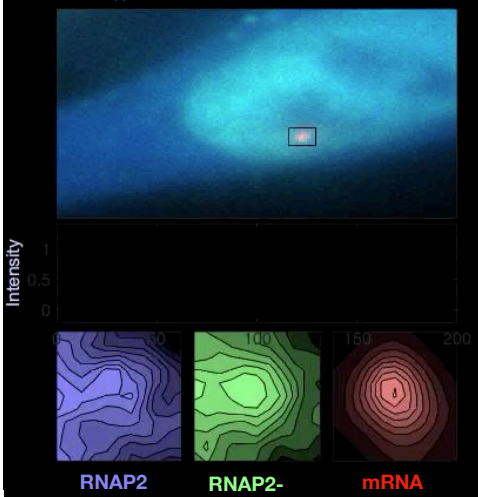
Linda Forero-Quintero | Will Raymond | Tim Stasevich



Forero, Raymond et al, *Nat. Comms.*, 2021

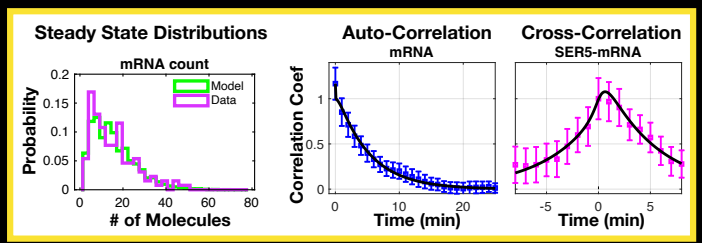
RNAP2 Recruitment, Phosphorylation and Transcription is a Bursty Process





MCP Label **RNAP2-CTD**
RNAP2-CTD-Ser5ph

- Each RNAP2/transcription trajectory is different due to the intrinsic noise of that particular process.
- Although we cannot expect a model to match any individual trajectory, we can ask that models match key statistics for the signals $I(t)$:
  - intensity joint distributions: $P(I)$
  - intensity auto- and cross-covariance: $\mathbb{E}\{(I(t)-\mu_I)(I(t+\tau)-\mu_I)^T\}$



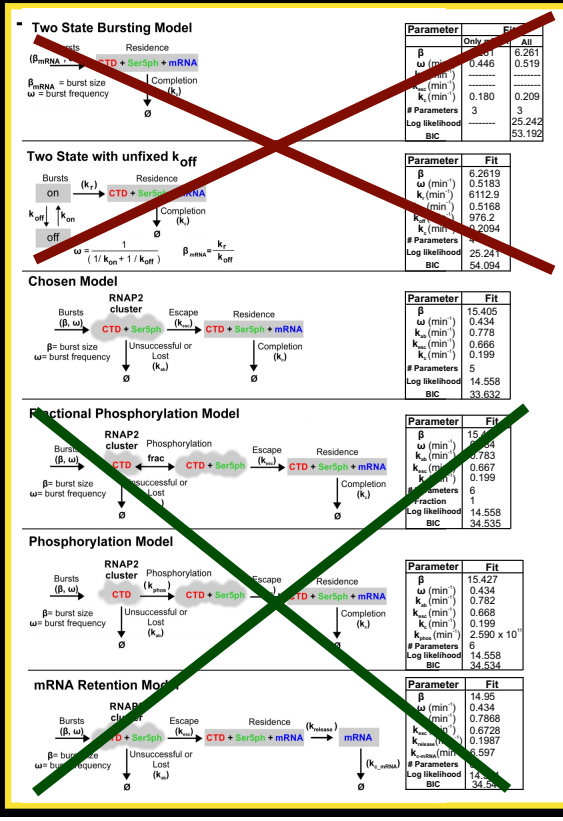
A good model should match all these statistics with a single (hopefully simple) set of mechanisms and parameters.

Linda S. Forero-Quintero | Tim Stasevich

Forero, Raymond et al, *Nat. Comms.*, 2021

Selecting the best model to reproduce RNAP2 and MS2 dynamics

- We tried many different extensions to the 2-state bursting gene expression model, fit them to data and evaluated their uncertainty using the Bayesian Information Criteria (BIC).



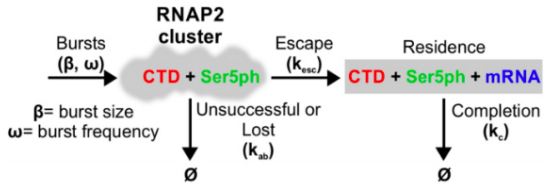
Models that lacked a separate RNAP2 recruitment and transcription escape events were too simple and could not reproduce the data.

We select the "Goldilocks Model" that minimizes the Bayesian Information Criteria (BIC)
(Akaiki I.C. yields identical result)

Models that had too many parameters (e.g., extra phosphorylation steps or distinct mRNA elongation/processing steps) could not be fully constrained by available data.

Will Raymond

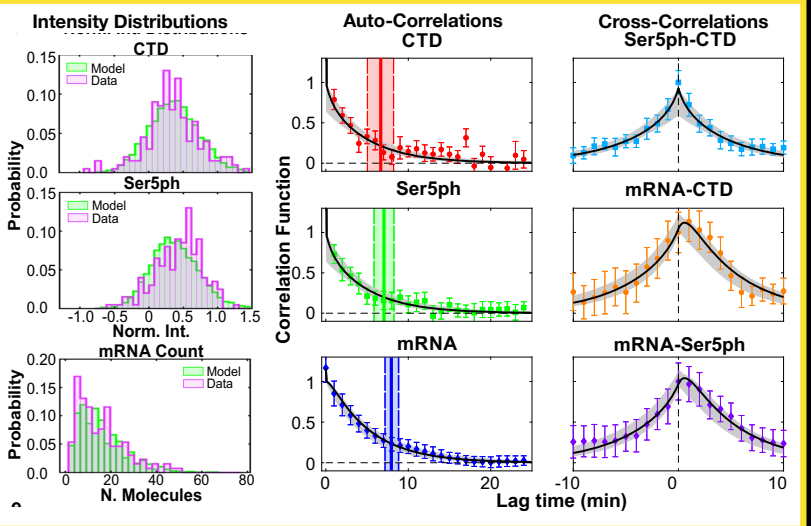
After fitting many models with different states and mechanisms, we selected (BIC/AIC) a simple model with four reactions and five parameters:



Parameter	Fit (95 % CI)
β	15.4 (11.8-61.8)
ω (min^{-1})	0.43 (0.33-0.66)
k_{ab} (min^{-1})	0.78 (0.62-44.22)
k_{esc} (min^{-1})	0.67 (0.59-6.55)
k_c (min^{-1})	0.20 (0.17-0.22)

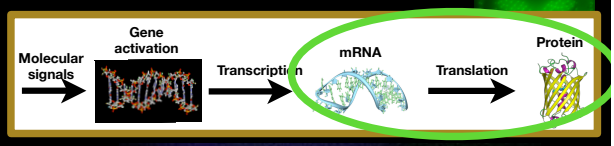
- Bursty recruitment of RNAP2 with frequency ω and average size β
- Aborted attempt with rate k_{ab}
- Successful escape with rate k_{esc}
- RNA completion with rate k_c

Despite its simplicity, this model captures the distributions, auto- and cross-correlations of total RNAP2, RNAP2-Ser5ph, and nascent mRNA at the transcription sites.

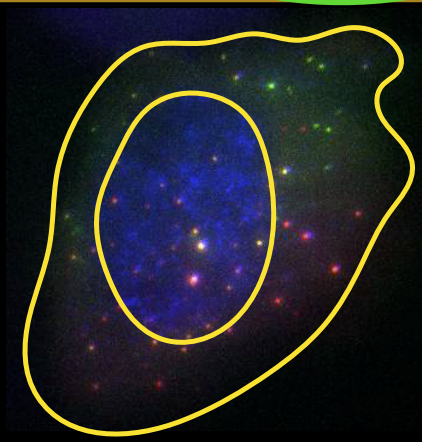


Forero-Quintero, Raymond et al., *Nat Comms,* 2021

Observing and simulating canonical and non-canonical translation from a single mRNA.



- FABs can also be used to quantify *Nascent Protein translation* from single mRNA in living cells.
- Different colors can be used to observe different open reading frames or different ribosomal entry sites.
  - red — mRNA
  - green — translating proteins



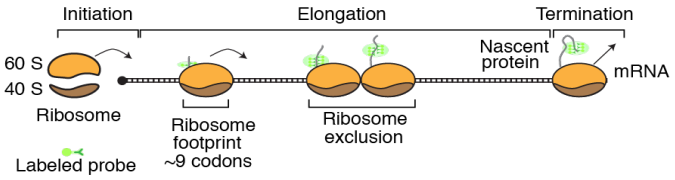
MCP Label
MS2 Hairpin
FAB Labels
SM Peptide

Kenneth Lyon **Luis Aguilera** **Amanda Koch** **Tim Stasevich** **Tatsuya Morisaki**

Lyon, Aguilera, et al, *Molecular Cell*, 2019
Aguilera, Raymond, et al, *PLoS Comp Biol*, 2019
Koch, Aguilera et al, *Nat. Struct, Mol. Biol.*, 2020

Canonical Translation Dynamics can be captured by a simple Codon-Dependent TASEP Model.

Nascent protein dynamics are captured by a *Totally Asymmetric Simple Exclusion Process* (TASEP).



The TASEP Model has two parameters:

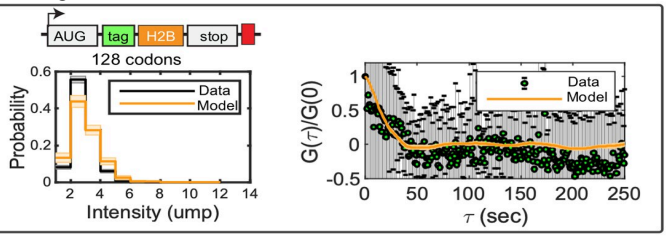
- initiation rate
- average elongation rate*

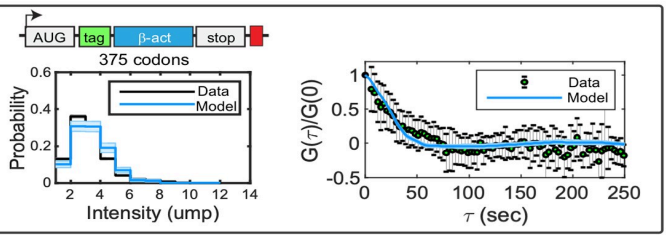
With just these two parameters, the model captures:

- the distribution of nascent proteins per mRNA in units of mature protein.
- the auto-covariance of the protein translation signal.

**Codon-dependent translation rates are defined by the Codon Adaptation Index.*

Steady-State Distributions **Auto-Covariances**

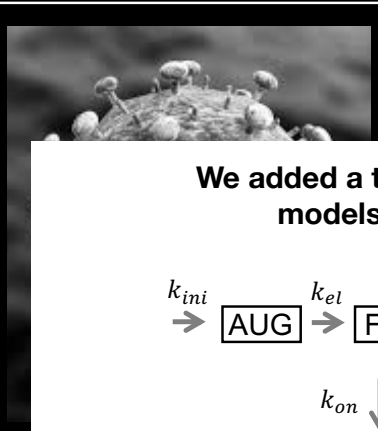





Luis Aguilera **Will Raymond**

Aguilera, Raymond, et al, *PLoS Comp Biol*, 2019

Extending the TASEP to include bursts of non-canonical translational frame-shifting



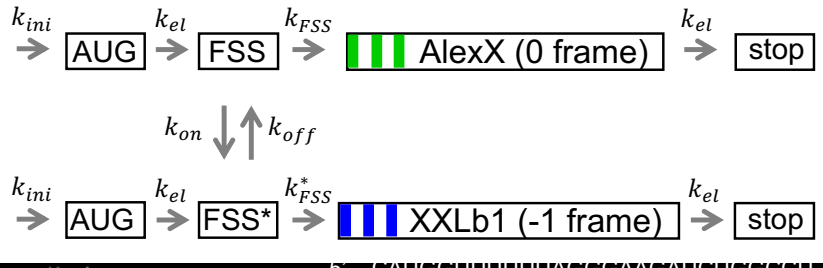
Two for the price of one!

HIV and other viruses use frameshift stimulatory

te two

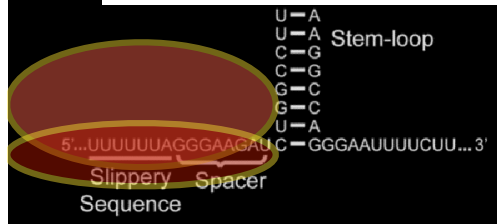
HIV-1

We added a third color in the -1 frame and extended models to allow bursts of frame-shifting.



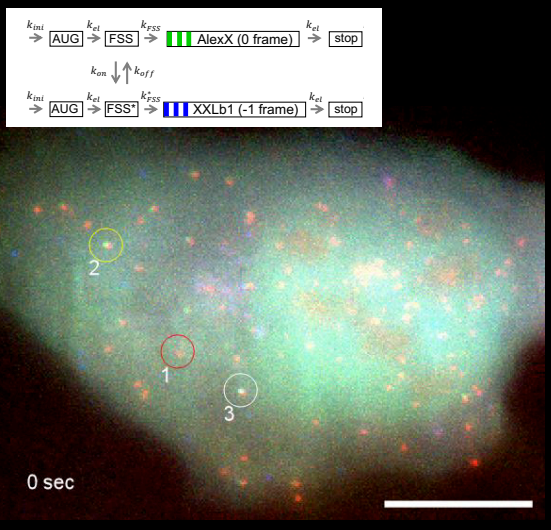
5'... CAUGGUUUUUUAGGGAAGAUCUGGCCU .... 3'

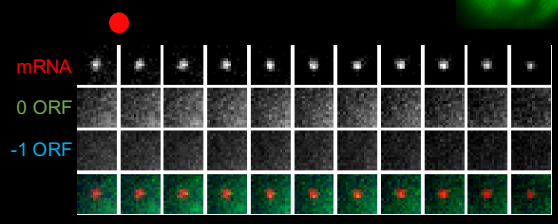
Original Protein

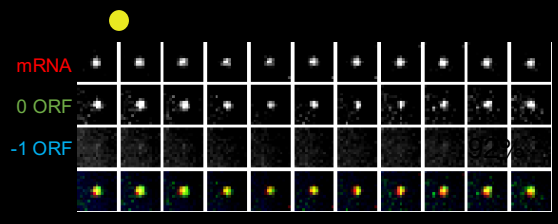


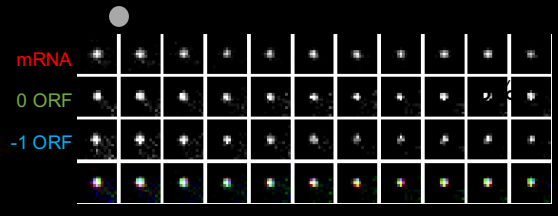
Lyon, Aguilera, et al, *Molecular Cell*, 2019

Extending the TASEP to include bursts of non-canonical translational frame-shifting









Kenneth Lyon

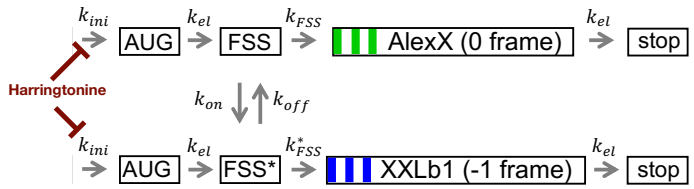
Luis Aguilera

Tim Stasevich

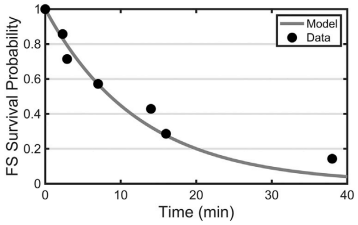
Tatsuya Morisaki

Lyon, Aguilera, et al, *Molecular Cell*, 2019

Extending the TASEP to include bursts of non-canonical translational frame-shifting



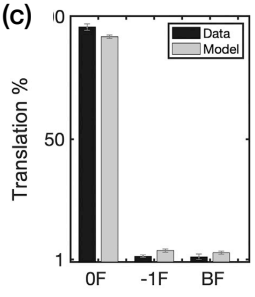
(a)



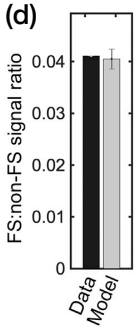
A single **bursting frame-shift model** captures:

(b)

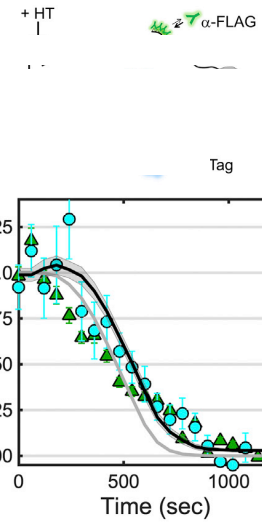
(c)



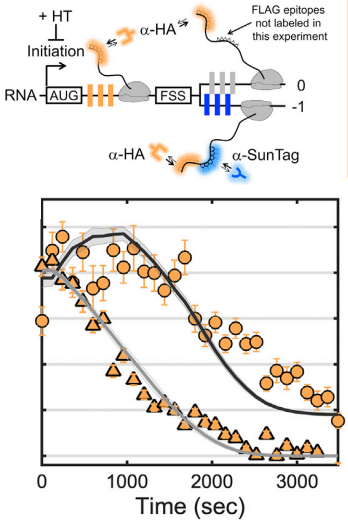
(d)



(e)

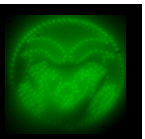


(f)



Lyon, Aguilera, et al, *Molecular Cell*, 2019

Observations From Measuring and Modeling Single-cell Processes



- **The Central Dogma is a *Noisy Process*, where mRNA and proteins are created in *discrete bursts*.**
- **Transcription and Translation bursts can be measured in *real time*, at *single-molecule resolution*, and in *living cells*.**
- ***Simple* discrete stochastic models are sufficient to quantitatively reproduce and often predict *every step* of these processes.**
- **By *testing multiple models* in different stress or drug response conditions, it is possible to gain insight into which bursting mechanisms are affected under what experimental perturbations.**

That's all great, BUT…

✳ Single-cell experiments are **expensive**, **noisy,** and there are **vast numbers** of possible experiment designs or user-supplied inputs.

Experiment Design Considerations

- **Number of cells**
- **Sampling times or periods**
- **Choice of fluorophore(s)**
- **Number and placement of probes**
- **Choice of which genes, mRNA, or protein to measure**
- **Inducer/drug concentrations and delivery times**

Measurement Error Considerations

- **Microscope resolution**
- **Image processing errors (segmentation, spot detection, track linking)**
- **Photobleaching**
- **Autofluorescence**
- **Camera exposure time**
- **Light source power and wavelength and optical filters**
- **Delays due to drug/inducer diffusion or nuclear import**

✳ To squeeze as much ***information*** as possible out of each experiment we need to use the most appropriate computational analyses.

✳ We also need systematic tools to ***choose*** experiments to ***minimize uncertainty*** about the mechanisms or parameters of interest.

Acknowledgments and Collaborators (2019-2022)

Hog Signaling/Transcription Activation
Gregor Neuert, Vanderbilt
Guoliang Li, Vanderbilt
Hossein Jahnsaz, Vanderbilt
Jason Hughes, Vanderbilt
Douglas Shepherd, CU Denver

Single-mRNA Translation
Tim Stasevich, CSU
Tatsuya Morisaki, CSU
Kenneth Lyon, CSU
Amanda Koch, CSU
Linda Forero Quintero, CSU

Cell Heterogeneity in Algal Biofuels:
Babetta Marrone, LANL
Claire Sanders, LANL
Elaheh Alizadeh, CSU

Cell-free Single-BioMolecule Kinetics:
Sabrina Leslie, McGill University
Shane Scott, McGill University

MultiScale Bayesian FSP tools:
Ania Baetica, CalTech
Thomas Catanach, Sandia

Soil Microbiome Machine Learning
John Dunbar Lab, LANL

Inflammation driven transcription
James Werner, LANL
Dan Kalb, LANL

q-bio Training and Textbook:
William S Hlavacek, LANL
Lev S. Tsimring, UCSD
Marek Kimmel, Rice University
Former students of the 2010-2017 q-bio Summer Schools

Postdocs:
Luis Aguilera
Linda Forero Quintero
Huy Vo

Graduate Students:
Zachary Fox
Lisa Weber
Mohammad Tanhaemami
Michael May
Jaron Thompson
Will Raymond
Eric Ron

Undergrads:
Elliot Djokic
Charis Ellis
Torin Moore
Zach Haigh
Charlotte Mitchell
Joshua Cook
Rachel Keating
Stuart McKnight
Katie Davis

Funding:
WMKF
NIH/NIGMS
DOE/BER
NSF CAREER

munsky@colostate.edu

UQ-Bio — Daily Schedule

Mondays - Fridays

- **09:00 - 10:00 - General Lecture (seminar style)**
- **10:15 - 12:15 - Python Tutorial**
- **12:15 - 14:00 - Lunch break (NOTE - the CSU Dining Commons are only open until 13:30).**
- **14:00 - 15:00 - Scheduled work time - please work in teams to discuss workbook assignments.**

Monday, Wednesday, Friday

- **15:30 - 17:00 - Hack-a-thon Session**

Tuesday and Thursday

- **15:30 - 17:00 - Career Discussion Panel**

UQ-Bio Schedule

Module 1 - Fluorescence Microscopy for Single-Cell Analysis (June 1-3)

- **Tutorial 1.1** - Basic Image Manipulation in Python (June 1, 10:15)
- **Lecture 1** - Dr. Linda Forero - Fluorescence Labeling Techniques for Single-Cell Experiments (June 2, 09:00)
- **Tutorial 1.2** - Single-Cell Segmentation in Python (June 2, 10:15)
- **Lecture 2** - Prof. Douglas Shepherd - Single-Cell Microscopy (June 3, 09:00)
- **Tutorial 1.3** - Single-Molecule Detection and Tracking (June 3, 10:15)

Module 2 - Basic Statistical Analyses (June 5-6)

- **Lecture 3** - Dr. Christian Meyer (June 6, 09:00)
- **Tutorial 2.1** - Basic Probability and Statistics (June 6, 10:15)
- **Lecture 4** - Prof. Anushree Chatterjee (June 7, 09:00)
- **Tutorial 2.2** - Basics of Supervised Machine Learning (June 7, 10:15)
- **Lecture 5** - Prof. Ashok Prasad (June 8, 09:00)

UQ-Bio Schedule

Module 3 - Simulating Biochemical Reaction Dynamics (June 8-10)

- **Tutorial 3.1** - Propensity Functions, Stoichiometries and ODE Analyses of Biochemical Reactions (June 8, 10:15)
- **Lecture 6** - Prof. Soham Ghosh (June 9, 09:00)
- **Tutorial 3.2** - Stochastic Simulations for Single-Cell Gene Regulation Dynamics (June 9, 10:15)
- **Lecture 7** - Dr. Tatsuya Morisaki (June 10, 09:00)
- **Tutorial 3.3** - Totally Asymmetric Simple Exclusion Process Models for Protein Translation (June 10, 10:15)

UQ-Bio Schedule

Module 4 - Inferring Models from Single-Cell Data (June 13-15)

- **Lecture 8** - Dr. Zachary Fox (June 13, 09:00)
- **Tutorial 4.1** - Chemical Master Equations and the Finite State Projection Analysis (June 13, 10:15)
- **Lecture 9** - Prof. Elizabeth Read (June 14, 09:00)
- **Tutorial 5.2** - Maximum Likelihood Estimation for Single-Cell Dynamics (June 14, 10:15)
- **Lecture 10** - Kaan Öcal (June 15, 09:00)
- **Tutorial 5.3** - Bayesian Estimation of Models for Single-Cell Gene Regulation (June 15, 10:15)