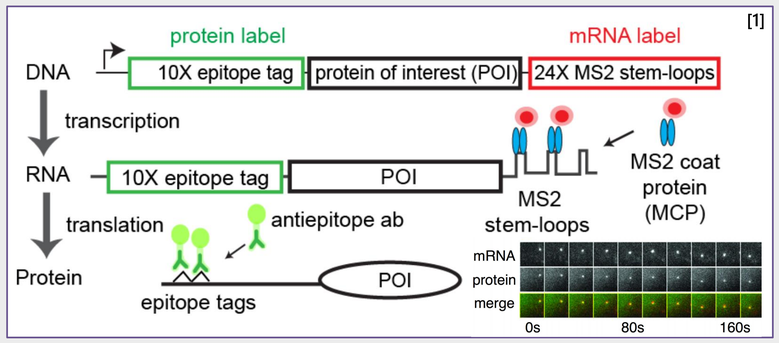
Nascent Chain Tracking Multiplexing: Dynamic Labelling

Will Raymond[1], Luis Aguilera[2], Sadaf Ghaffari[4], Timothy Stasevich[3], Brian Munsky[1,2]

[1]School of Biomedical Engineering, [2]Department of Chemical & Biological Engineering, [3]Department of Biochemistry and Molecular Biology, [4]Department of Computer Science
Colorado State University, Fort Collins, CO, USA 80523

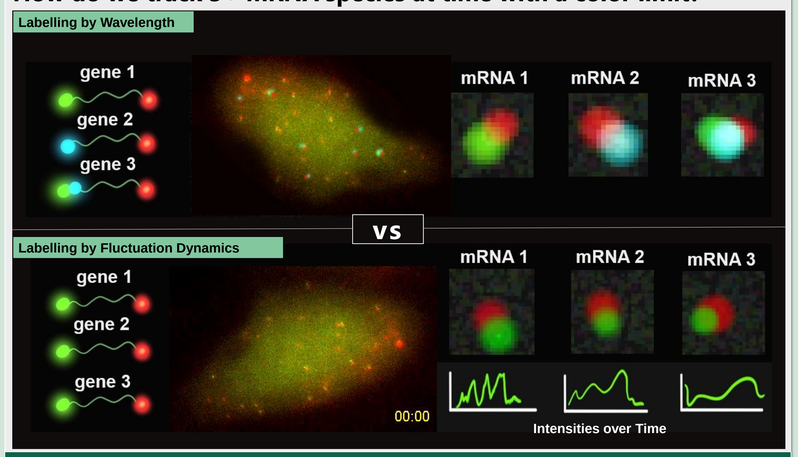
Introduction

Nascent Chain Tracking (NCT) is an experimental technique that probes single mRNA transcript dynamics via the use of a dual fluorescent probe system. Growing polypeptide chains translate repeated epitope regions for a fluorophore conjugated fragmented antibody (fAB), while a 3'UTR tag region provides stem loop repeats for fluorophore conjugated MCP/PP7 to bind. Both fluorescent channels are tracked in time with a constant RNA signal and a fluctuation protein signal, indicating active translation.



NCT is currently limited in the amount of mRNA species it can label and track over time, a liminted amount of different wavelength fluorophores are available to the microscopist. We are aiming to expand this "labelling" palette via a pipeline of simulated NCT videos and machine learning classifiers to label spots via their dynamics, instead of their wavelength.

How do we track 3+ mRNA species at time with a color limit?



Simulated Data Packages

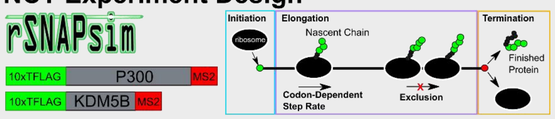
RNA Sequence to Nascent Protein Simulator (rSNAPsim) is a Python module for rapid design of single-RNA translation experiments utilizing a TASEP based stochastic simulation of NCT intensity dynamics. rSNAPsim provides the intensity trajectories to rSNAPed (RNA Sequence to Nascent Protein Experiment Design) which convolutes the signals with a microscope realistic experiment distortion via adding a real blank cell background, Brownian motion, tracking, and matching signal to noise ratio to the blank cell. rSNAPed distorts and recaptures the intensity trajectories for a more realistic downstream ML classification.

[2] rSNAPsim is open source and available to the public and provides a myriad of functionality to design experiments, simulate intensity data, analyze intensity data, and fit several relevant models to NCT experiments. rSNAPsim also provides several utility libraries for sequence manipulation and arbitrary codon optimization.

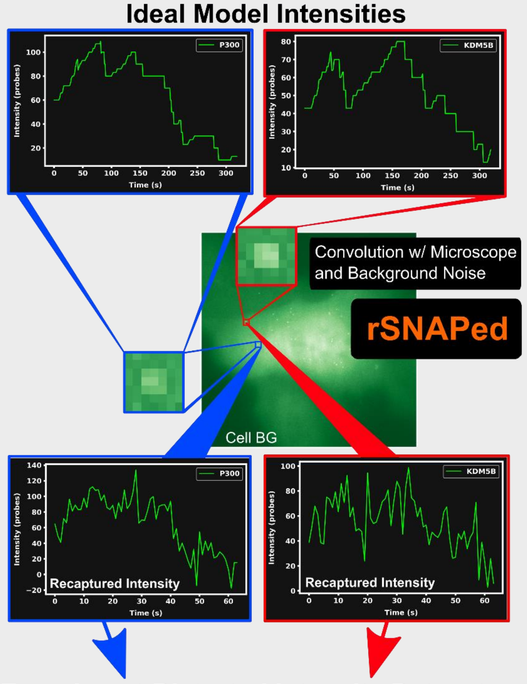
https://github.com/MunskyGroup/rSNAPsim

Simulation Pipeline

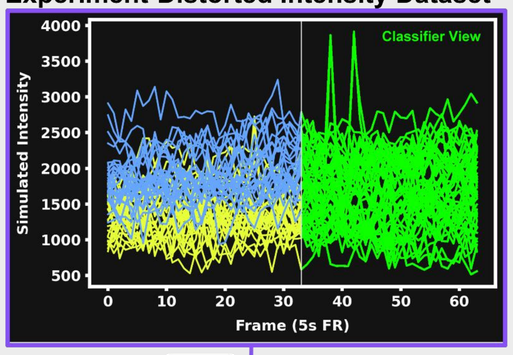
NCT Experiment Design



Ideal Model Intensities

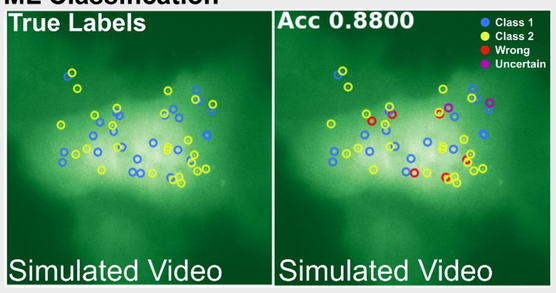


Experiment-Distorted Intensity Dataset



CNN

ML Classification



Results

CNN

Simulated datasets were generated to linearly explore five parameters of interest to NCT experiment design: mRNA lengths, frame Rates, frames used, ribosomal elongation rate, ribosomal initiation rate.

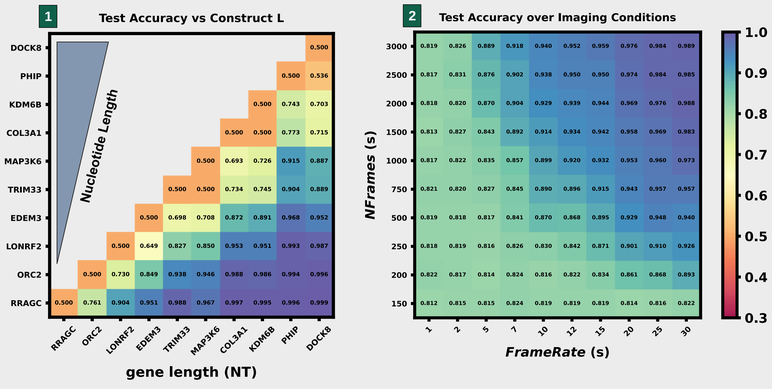
Constant Parameter Values

1	2	3	4	5
5000 spots, 50/50 split class one class 2, 80/20 Train test split				
Ke both genes = 5.33 aa/s				
Ki both genes = 0.06 1/s				
		Imaging conditions: 64 frames 5s frame rate		
	Gene 1: P300 (8268 NT) Gene 2: KDM5B (5658 NT)			

Parameter types: Variable | Constant

1	2	3	4	5
mRNA length	mRNA length	mRNA length	mRNA length	mRNA length
Frame rate	Frame rate	Frame rate	Frame rate	Frame rate
Frame count	Frame count	Frame count	Frame count	Frame count
Initiation rate	Initiation rate	Initiation rate	Initiation rate	Initiation rate
Elongation rate	Elongation rate	Elongation rate	Elongation rate	Elongation rate

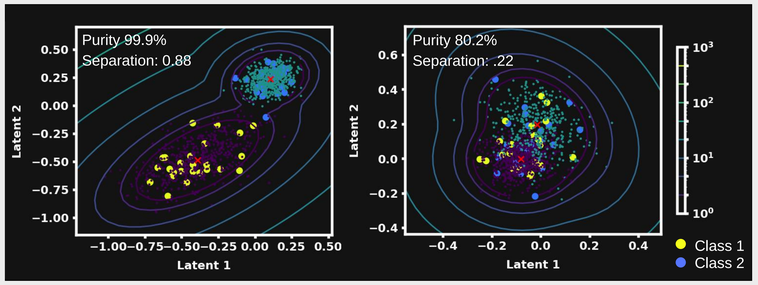
Biological Variation and Experimental Design Exploration



ML NCT spot classification is possible with feasible NCT video conditions (~ 5 minutes of video). Classification accuracy goes up with the amount of information per video (ribosome flow) and increasing temporal resolution per ribosomal dwell time (Figures 2 and 3). Additionally with identical translation parameters, mRNA construct length difference enough to create detectable fluctuation difference (Figure 1). For classification in conditions where there is a sufficient construct length difference, initiation and elongation rates must not create a condition where fluctuations match, e.g. If the longer mRNA has a higher elongation rate than the shorter mRNA, intensity dynamics become similar (Figures 4 and 5).

Future Work

Unsupervised approaches can also be explored with the simulated NCT pipeline. Unsupervised classification is closer to experimental conditions where microscopists eschew multiple color labels and only use one protein channel. Currently, we are exploring classification with randomly generated NCT videos with varying experimental conditions. We hope to provide a path to experimentalists looking to perform NCT multiplexing experiments.



References and Acknowledgements

(1) Morisaki, Tatsuya, et al. Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**.6292 (2016): 1425-1429.
(2) Aguilera, Luis. Raymond, Will. Fox, Zachary. Stasevich, Timothy. Munsky, Brian. Computational design and interpretation of single-RNA translation experiments. *PLoS computational biology* 15(10). (2019).

NSF CAREER Award
W.M. KECK FOUNDATION
NIH NIGMS