Nascent Chain Tracking Signal Classification

Raymond William[1], Aguilera Luis[1], Ghaffari Sadaf[3], Fox Zachary R.[4,5] and Munsky Brian[1,2]

***Abstract*— mRNA translation is the necessary cellular process of decoding messenger-RNA strands into usable proteins; Within the past decade, nascent chain tracking (NCT) fluorescent microscopy techniques have allowed experimentalists to observe translation at single molecule resolution for consistent time-series measurements. However, NCT experiments so far have been limited to observation of one to two species of mRNAs at a time. In this work, we provide a thorough computational exploration of NCT signal classification via machine learning classifiers on simulated NCT data-sets containing two disparate mRNA signals.**

***Index Terms*— Machine learning, nascent chain tracking, single molecule fluorescence, multiplexing, experiment design**

I. Background Information

Nascent chain tracking (NCT) is a fluorescence microscopy technique that utilizes co-localized probes to record mRNA translation at a single molecule resolution. The technique consists of encoding repeated protein epitopes within the CDS of mRNA where fluorescently conjugated fragmented antibodies can bind on growing nascent chains to give a protein signal. Additionally, a 3'UTR tagging system like MS2 provides a constant RNA signal. Together the co-localized channels confer the ability to track and record ribosomal elongation of individual mRNAs within a single diffraction limited spot. NCT has been used in the past eight years to investigate a myriad of dynamics such as mRNA decay, IRES activity, and ribosomal frameshifting [1].

NCT suffers from a limited number of color channels to utilize for tagging mRNAs of interest, and, as of yet, simultaneous observation of multiple mRNA species within the same cell has not been achieved. Here, we present a machine learning based exploration of classifying simulated NCT signals over several critical NCT experimental parameters to provide experimental design guidance.

II. CNNs can achieve high accuracy for realistic NCT settings on simulated data

Ribosomal elongation rate, ribosomal initiation rate, movie frame rate, frame count, and mRNA lengths were selected as critical NCT experimental design parameters. Parameter sweeps were set up to test classification accuracy over five different linear combinations of the design parameters. Simulated movies were generated by combining the NCT-TASEP based modelling from rSNAPsim [2] and realistic background imaging from rSNAPed [pending]. A final data size of 2500 signals for two classes for 3000 seconds for 445 parameter combinations was used (∼37,083 hours of simulated video).

Simple convolutional neural networks (CNNs) were trained to differentiate NCT signals from two different mRNA species for realistic ranges of imaging conditions. For a framerate of five seconds and 64 total frames, (∼ five minutes of video), ML performance regularly eclipses 75% validation accuracy across parameter combinations. Accuracy with CNNs was contingent on the frequency content and intensity moments of the given signals to differentiate. Manifolds where signal moments were too similar provides poor classification as expected, however, parameter combinations with minimal parameter differences are possible to differentiate given enough signal resolution and length.

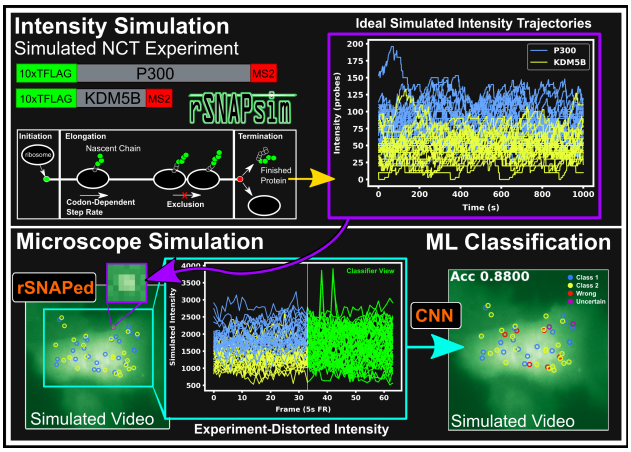

Fig. 1: NCT multiplexing classification pipeline

Our pipeline can provide a first past simulated data-set that can quantify NCT Multiplexing experiment feasibility while using the same protein channel colors, saving lab time and resources for the cost of computational effort and opening up color channels for other fluorescence tagging.

References

[1] C. A. Cialek, A. L. Koch, G. Galindo, and T. J. Stasevich, "Lighting up single-mrna translation dynamics in living cells," *Current Opinion in Genetics and Development*, vol. 61, pp. 75–82, 4 2020.

[2] L. U. Aguilera, W. Raymond, Z. R. Fox, M. May, E. Djokic, T. Morisaki, T. J. Stasevich, and B. Munsky, "Computational design and interpretation of single-rna translation experiments," *PLoS computational biology*, vol. 15, p. e1007425, 10 2019.

This work was funded by NIH-NIGMS grant R35GM124747

[1]School of Biomedical Engineering, Colorado State University, Fort Collins, CO 80523, USA. Email: wsraymon@rams.colostate.edu

[2]Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO 80523, USA.

[3]Department of Computer Science, Colorado State University, Fort Collins CO, 80523, USA.

[4]T-CNLS, Theoretical Division, LANL, Los Alamos NM, 87545, USA.

[4]CCS-3, Computer, Computational and Statistical Sciences Division, Theoretical Division, LANL, Los Alamos NM, 87545, USA.