RNA Sequence to NAscent Protein Experiment Designer (rSNAPed)

Luis U. Aguilera[1,2], Timothy J. Stasevich[2] and Brian Munsky[1]

***Abstract*— We evaluate computational methods to automate image processing for single-molecule gene expression experiments. Our results are integrated into an open-source Python package, rSNAPed: RNA Sequence to NAscent Protein Experiment Designer. rSNAPed uses standard libraries such as Cellpose, TrackPy, and rSNAPsim. rSNAPed can be used to generate synthetic data by simulating single-molecule translation spots on top of real non-expressing cells. These simulations result in realistic images that create complex scenarios to challenge and quantify our image processing code's quality. rSNAPed automatically processes large data sets to quantify important biophysical parameters.**

***Index Terms*— Single-molecule data, Image processing, simulations, experimental design.**

I. Introduction

NAscent chain tracking technology (NCT) allows to performing of single-molecule gene expression experiments [1]. NCT has been a fundamental tool for elucidating important biological processes, including ribosomal frameshifting, IRES-translation, and single-molecule virus dynamics [2]. Notwithstanding NCT's importance, processing NCTs data is still a time-consuming and labor-intensive endeavor, resulting in small sample size data sets that could potentially influence the variability of the results and drive inaccurate statistical conclusions.

In this project, we introduce RNA Sequence to NAscent Protein Experiment Designer (rSNAPed), a new open-source python library to fully automate the analysis of NCT data. rSNAPed core functionality is achieved by generating a large number of simulated data sets that closely match the complexity of the images obtained under the microscope. These data sets are generated using more challenging conditions, including low signal-to-noise ratio and large particle concentrations. Then, simulated data sets are fed into the code to train threshold-selection algorithms, and in this way reduce human intervention. A fully automated image processing pipeline is achieved using these trained algorithms. Finally, mechanical models are generated to quantify the system's biophysical parameters and match the experimental data.

This work was funded by the W.M. Keck Foundation and NIH grants 5R35GM124747 and R35GM119728.

[1]Department of Chemical and Biological Engineering, Colorado State University Fort Collins, CO 80523, USA. Email: luis.aguilera@colostate.edu

[2]Department of Biochemistry and Molecular Biology and Institute for Genome Architecture and Function, Colorado State University, Fort Collins, CO 80523 USA.

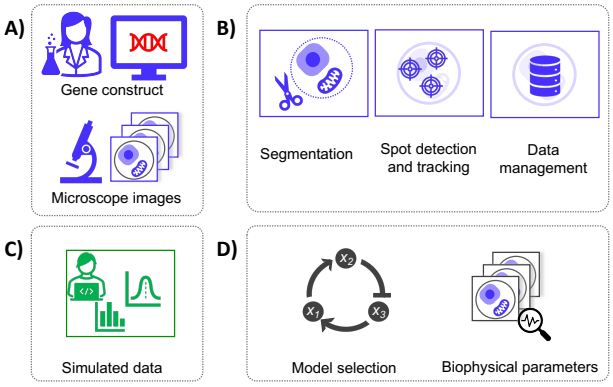

Figure 1 shows the steps used in rSNAPed. A) the software requires the DNA gene sequences of the studied gene and the microscope images. B) In the first step, the software performs automated cell segmentation using Cellpose [3]. Subsequently, single-molecule spots are detected using TrackPy [4]. Transcription or translation spots are quantified and classified. C) The software generates a mechanistic model representing the studied biological processes, and this model is solved using stochastic dynamics using rSNAPsim [5]. D) Finally, real data and statistical conclusions are integrated to estimate relevant biophysical parameters for the studied biological processes.

References

[1] T. Morisaki, K. Lyon, K. F. DeLuca, J. G. DeLuca, B. P. English, Z. Zhang, L. D. Lavis, J. B. Grimm, S. Viswanathan, L. L. Looger *et al.*, "Real-time quantification of single rna translation dynamics in living cells," *Science*, vol. 352, no. 6292, pp. 1425–1429, 2016.

[2] 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 & development*, vol. 61, pp. 75–82, 2020.

[3] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature Methods*, vol. 18, no. 1, pp. 100–106, 2021.

[4] D. B. Allan, T. Caswell, N. C. Keim, C. M. van der Wel, and R. W. Verweij, "soft-matter/trackpy: Trackpy v0.5.0," Apr. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4682814

[5] 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 live-cell, single-rna translation experiments," *PLoS Computational Biology*, vol. 15, no. 10, p. e1007425, 2019.