Automated pipeline to process single-molecule Fluorescence In Situ Hybridization (smFISH) images.

Luis U. Aguilera[1,2], Joshua Cook[1,3], Eric Ron[1], Linda Forero-Quintero[1], Timothy J. Stasevich[2] , and Brian Munsky[1,3]

***Abstract*— Mathematical modeling of cellular pathways is used widely in biology to comprehend molecular mechanisms thoroughly. These models require large amounts of experimental data to be validated correctly. Fluorescence in Situ Hybridization (FISH) is a well-established technology used to detect the transcriptional expression of any gene. Here, we work towards implementing a fully automated image processing pipeline to process FISH data by using Cellpose [1] and BIG-FISH [2]. The resulting code allows to correctly process hundreds of images for segmentation and spot quantification.**

I. Introduction

Fluorescence in situ hybridization (FISH) is used to detect and visualize individual mRNA molecules within cells. smiFISH (single-molecule inexpensive FISH) consists of designing primary probes that contain the target sequence against the desired gene plus a common sequence that is then annealed to a complementary sequence labeled with a fluorescent dye [3]. These probes allow the mRNA to be detected as individual puncta when viewed using a fluorescence microscopy. Processing FISH data is time-consuming and requires extensive user input. Recently, advances in deep learning and image processing have allowed the development of novel tools to quantify massive biological datasets in an automated manner. Here, we developed an open-source Python code to process automatically large sets of FISH data. The image processing pipeline allows hundreds of images to pass through the pipeline for segmentation and spot detection to occur on each image separately.

The individual steps of the pipeline are data collection from the microscope, which is then transferred to a Network-Attached Storage (NAS). The data is then sent to a computational cluster where segmentation is achieved using Cellpose [1], and spot detection is performed using Big-FISH [2]. The data is then organized into a CSV file containing information about cell size and the detected spots' location. Additionally, a metadata file is created to detail the parameters used in the code and the necessary version of the used libraries. Finally, a pdf file is created to show the individual steps of the pipeline and allow the user to supervise the complete process. All results are uploaded to the NAS for interpretation by the experimentalist.

LUA, LFQ and BM were supported by NIGMS grant R35-GM124747. JC and ER were supported by NSF Grant 1941870.

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

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

[3]School of Biomedical Engineering, Colorado State University, Fort Collins, CO, USA. Email: brian.munsky@colostate.edu

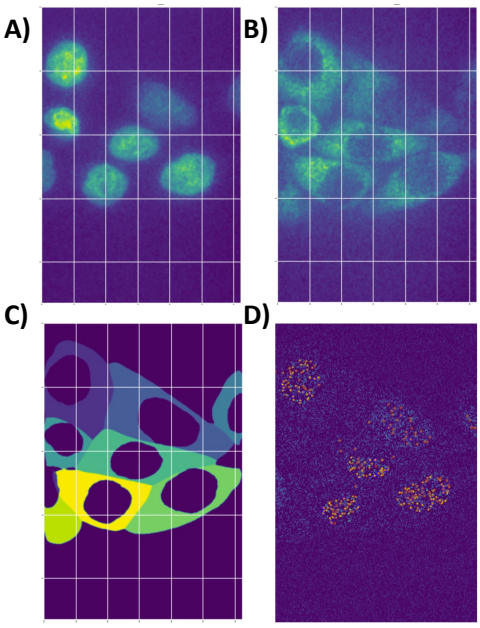

Figure 1 shows a maximum projected image of HeLa cells in which endogenous GAPDH was targeted using smiFISH. In A), the cell nucleus is stained using DAPI. In B), the mature and nascent mRNAs and transcription sites of endogenous GAPDH are marked using smiFISH probes targeting GAPDH exons pre-labeled with cy5. C) Shows the segmentation of the nucleus and cytosol using Cellpose[1]. Finally, in D) FISH puncta are detected using Big-FISH[2]. The complete code is accessible here: `https://github.com/MunskyGroup/FISH_Processing`.

References

[1] 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.

[2] A. Imbert, W. Ouyang, A. Safieddine, E. Coleno, C. Zimmer, E. Bertrand, T. Walter, and F. Mueller, "Fish-quant v2: a scalable and modular tool for smfish image analysis," *RNA*, pp. rna–079 073, 2022.

[3] N. Tsanov, A. Samacoits, R. Chouaib, A.-M. Traboulsi, T. Gostan, C. Weber, C. Zimmer, K. Zibara, T. Walter, M. Peter *et al.*, "smifish and fish-quant–a flexible single rna detection approach with super-resolution capability," *Nucleic acids research*, vol. 44, no. 22, pp. e165–e165, 2016.