

Optimal detection, separation, and analysis of FISH images

Aparna Rajpurkar - arraipur@stanford.edu

Maggie Engler - mengler@stanford.edu

(We will not be using an Android device.)

Background

Fluorescent in situ hybridization (FISH) is a broadly applied experimental technique in biology to detect presence, localization, and abundance of specific nucleotide-based units such as genes, chromosome ends, or mRNA transcripts. The structure of interest is marked by an oligonucleotide probe that hybridizes to the DNA or RNA and carries fluorophores, which are detected by fluorescence microscopy. Researchers can then visualize the localization of RNA within the cell to gain insights into RNA colocalization and post-processing, for example. In the past, most FISH images have been analyzed manually. However, the advent of highly multiplexed FISH strategies such as MERFISH¹ (multiplexed error robust FISH) makes automatic processing methods that can efficiently achieve high accuracy and precision on large datasets essential.

Goal

To design and implement a broadly-applicable algorithm to optimally detect individual fluorescence points from multiplexed FISH images, test its accuracy and precision on publically available data, and apply it to new datasets.

Datasets and strategy

We will be focusing on the example datasets from the original MERFISH paper. We will implement, test and compare different strategies to separate and identify individual points of fluorescence, such as:

1. A thresholding approach followed by detection of continuous regions.
2. Modelling FISH signal as Gaussian point spread functions.²
3. (Potentially) A machine-learning based approach to classify noise and signal.³

We will use these strategies to develop an optimal algorithm to identify localization of individual RNAs and decode their identity through the MERFISH 16-bit binary code, and test against published results, comparing our accuracy in retrieving correct RNA identities as well as the limit of our precision in separating nearby fluorescence points.

Upon successful development of our detection strategy, we can then apply our algorithm to new/unpublished datasets such as:

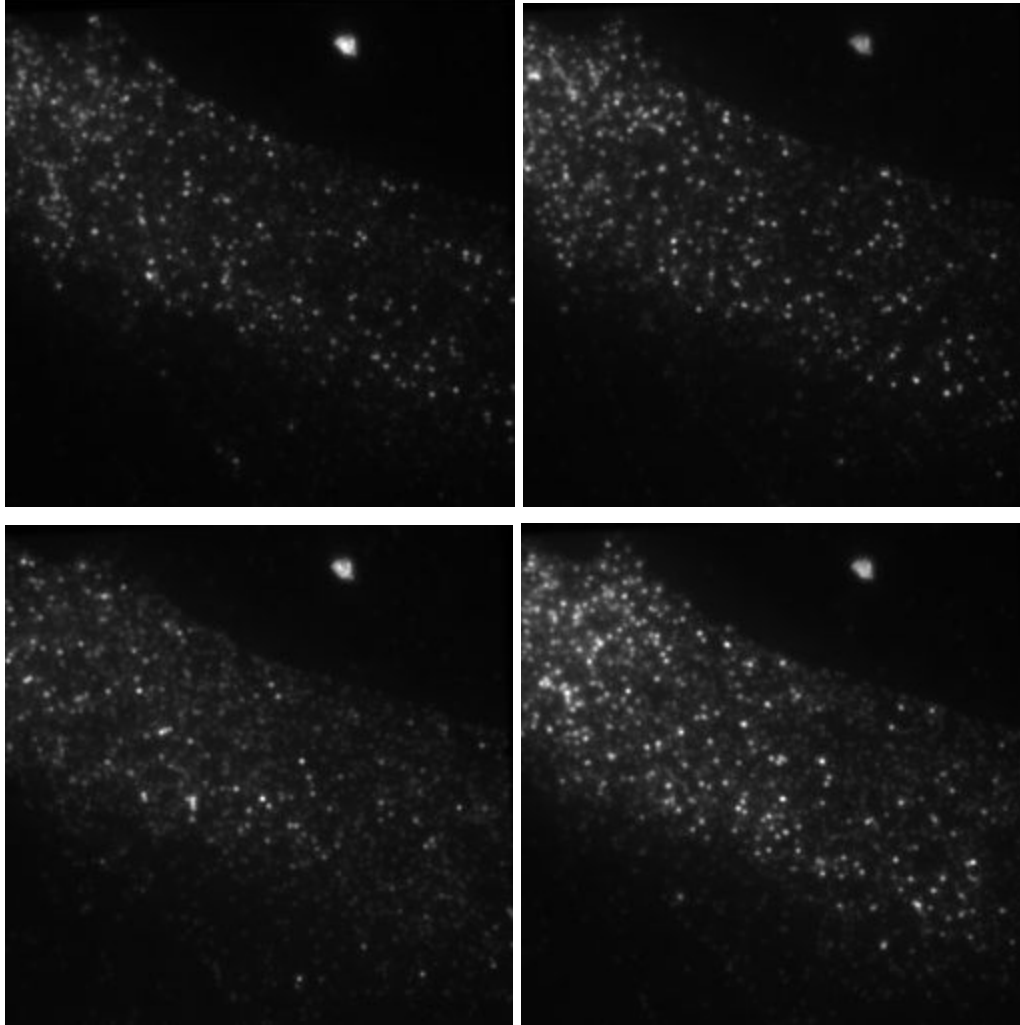
1. Z-stack images of the mouse placenta, showing chromosomal ends marked by FISH (higher noise). We will apply our algorithm to count number of chromosome ends and gain insight into over-replication of trophoblast DNA during development of the placenta.
2. (Potentially) MERFISH images of whole-tissue mouse thymus, where the challenge would be to decode individual RNAs across many cells in an epithelial layer (cells very

close together, lower resolution). This will allow us to gain insight into T cell development and morphological organization of a critical immune organ.

These additional datasets would require tuning of parameters to handle their unique challenges, but require the same core functionality from our algorithm while demonstrating its broad applicability to many different questions in biology.

Example images

MERFISH



References

1. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science (New York, NY)*. 2015;348(6233):aaa6090. doi:10.1126/science.aaa6090.

2. Tsanov N, Samacoits A, Chouaib R, et al. smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Research*. 2016;44(22):e165. doi:10.1093/nar/gkw784.
3. Wu AC-Y, Rifkin SA. Aro: a machine learning approach to identifying single molecules and estimating classification error in fluorescence microscopy images. *BMC Bioinformatics*. 2015;16:102. doi:10.1186/s12859-015-0534-z.