

Optimal detection, separation, and analysis of FISH images

Aparna Rajpurkar and Maggie Engler

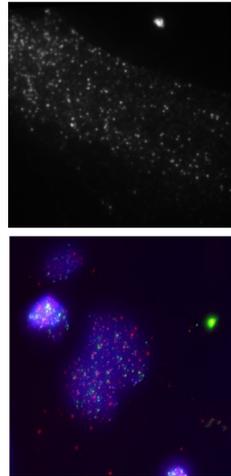
Department of Genetics, Department of Electrical Engineering, Stanford University

Background

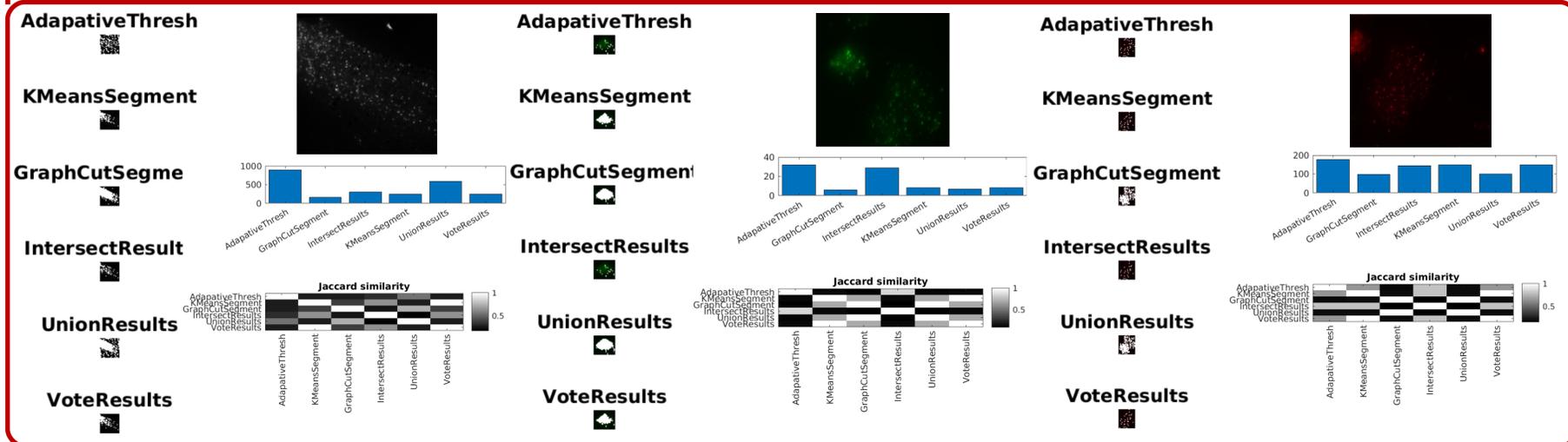
Fluorescent in situ hybridization (FISH) is a broadly applied experimental technique in biology to detect presence, localization, and abundance of specific nucleotide-based units such as genes, chromosome ends, or mRNA transcripts.

In the past, most FISH images have been analyzed manually. However, the advent of highly multiplexed FISH strategies such as MERFISH (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 generality on new datasets, and assess the real use case performance and biological relevance of results.



Comparison



Algorithms

Adaptive Threshold

We compute the output images from locally adaptive thresholding over the range of possible sensitivity factors (from zero to one), and count the number of spots identified at each value. Then, the sensitivity factor where the number of identified spots is least sensitive to changes in sensitivity should be the value at which we are including the lowest amount of noise.

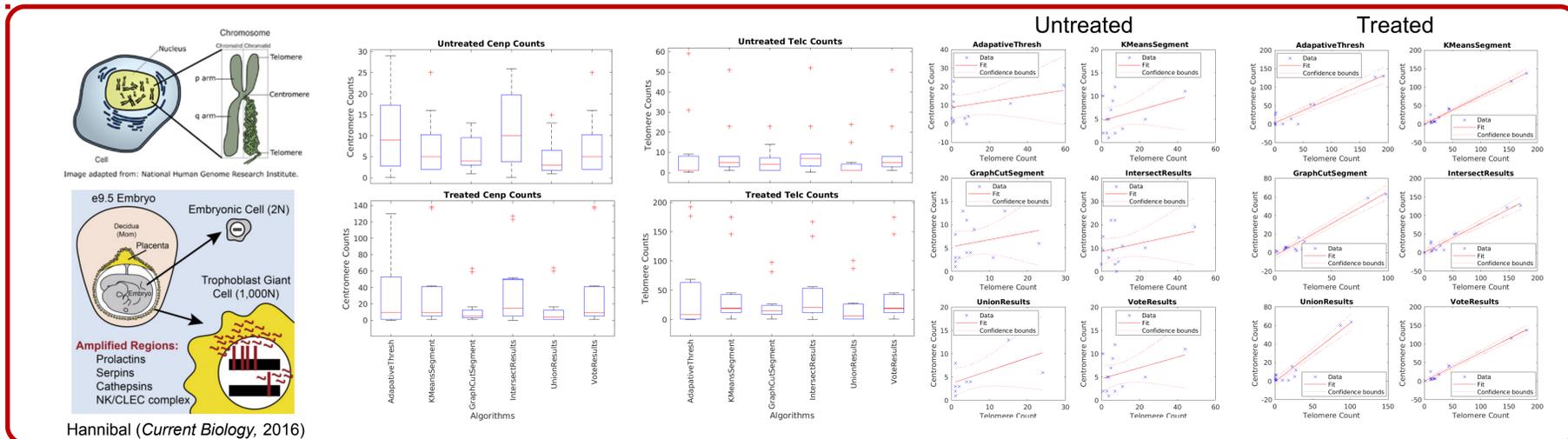
K-means Clustering

We implement k-means segmentation—separation of the input images into k clusters, with k as an input parameter—and return the processed image. Because we are performing an object extraction task, we already know the number of clusters should be $k = 2$ for foreground and background pixels, and we do not need to iterate over this parameter. The optimal output image will be the binarized image with the lowest overall sum of the squared “within-group” error, meaning the difference in intensity.

Graph Cuts Segmentation

For a third segmentation method, we consider the relatively new algorithm for image segmentation, from graph cuts based on energy minimization. Graph Cuts uses an initial data cost matrix, which defines the cost of assigning a specific label (0 or 1) to a pixel, and a smoothness cost, which in the binary case is a 2-by-2 matrix s where $s_{i,j}$ describes describes the cost of labeling neighboring pixels with labels i and j . The data cost matrix is defined to be 1 if the intensity of a pixel is greater than the average intensity and 0 otherwise.

Biological Relevance



Conclusion

In conclusion, we implemented 3 different algorithms with the goal of detecting FISH fluorescence spots as well as combinations of their results, and tested their performance on new datasets in a realistic experimental use case. We found that Kmeans and GraphCuts performed similarly, but Adaptive Thresholding worked best on all data types tested and the high background noise of some samples caused other algorithms to lose accuracy. However, trends in the data with regards to treated and untreated samples are preserved in all cases. Therefore, we recommend Adaptive Thresholding as the optimal strategy to detect and count FISH fluorescence spots on real biological data.

Data sources: Kay Kobak: mouse placenta data, Chen, Boettiger, and Moffit et al. (2015): MERFISH data. References: Chen, Boettiger, and Moffit et al (2015), Tsanov et al (2016), Wu et al (2015), Patel et al (2016), Boykov et al (2001), Kolmogorov et al (2002), Boykov and Kolmogorov et al (2004), Bagon (2006).