Development of an automated quantification algorithm for determining fluorescence distribution in yeast cells (September 1, 2015 to August 31, 2016)

Participants:

Student	Faculty Advisor	Department	University
Wen-Chieh Hsieh (Hsieh@purdue.edu)	R. Claudio Aguilar (claudio@purdue.edu)	Biological Sciences	Purdue
Leqi Liu (lliu03@brynmawr.edu)	Jia Tao (taoj@tcnj.edu)	Computer Sciences	Bryn Mawr college

Details of dates and attendance of team meetings:

S.No	Dates	Venue	Participants
1	Dec 8 2015	Purdue University, IN	Wen-Chieh Hsieh Leqi Liu (online)
2	Dec 18 - 20 2015	Bryn Mawr college, PA	Wen-Chieh Hsieh Leqi Liu

Details of money spent:

Grant awarded: $2000

Money spent: $ 813.81

Money remaining: $ 1186.19

Presentation:

Wen-Chieh Hsieh, Leqi Liu, Jia Tao and Claudio Aguilar. Development of an automatic quantification algorithm for determining fluorescence distribution in yeast cells. (2015 NSF site visit)

Project Introduction

The initial step of creating effective therapeutic approaches is to understand the disease causing mechanism in detail, which in many cases involves acquiring and analyzing massive amounts of biological data. However, conventional, human-involved analysis methods present serious disadvantages when processing high-information content, image data, such as user-introduced error/biases and low throughput. For instance, examining protein subcellular localization typically requires quantifying hundreds of cells per experiment through visual examining and manually outlining areas from images collected by microscopy. The procedure is not only laborious but also imprecise that would restrict the efficiency of data analysis process, and more importantly, the accuracy of the interpretation after the data analysis.

The objective of this student-initiated proposal is ***to develop an automated quantification algorithm to extract and analyze quantitative biological information related to endocytosis of membrane proteins*** *to facilitate data analysis and reduce user-introduced errors.*

Endocytosis is one of the key processes for regulating signaling pathways that are crucial for maintaining normal cellular functions and tissue development. For example, internalization of DLL1 protein is the triggering step of the Notch signaling pathway that when misregulated causes multiple developmental diseases and cancer. Therefore, understanding the regulatory mechanism of membrane protein internalization is a necessary step prior to discovering therapeutic approaches to those diseases. Budding yeast is a powerful model system to study endocytosis. Here we use a green-fluorescent-protein (GFP) tagged membrane protein to determine the level of endocytosis via quantifying the cellular distribution of fluorescence signal collected through fluorescence microscopy.

Working strategy:

Stage 1: Identify single cells from microscopic images

Yeasts are small unicellular organisms that we routinely use as a model to study distribution of biological products, such as proteins. In microscopic images, there could be multiple yeast cells (Figure 1). In order to collect biological information from each individual cells, we need to distinguish the cells from the background, and further locate their coordinates. We use “R” as a platform for developing the algorithm for image analysis.

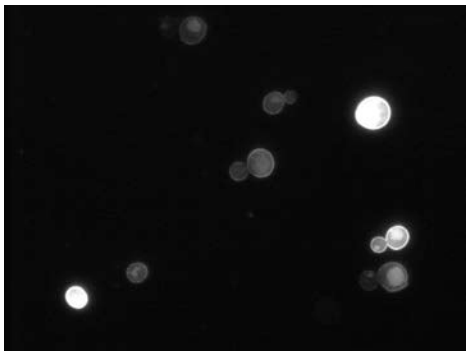

Fig.1 Original image

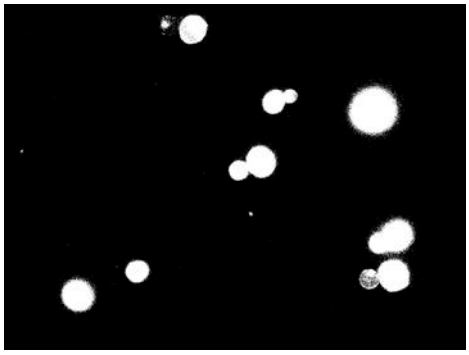

Fig.2 Filtered image

Based on the pixel intensity, we set a serial of thresholds to the original images, and then a filtered image that could remove background noise will be automatically selected. The selection is based on a pattern we found: Among sets of threshold images series that identified the same number of pixel clusters, the first image of the longest and the earliest set represents the original image the best. The resulting image selectively preserves the signal from the cells (figure 2).

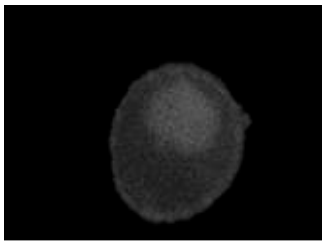

Fig.3 Single cell image

In such an image, adjacent pixels within a certain area are clustered into components. Each component covers the area of a single cell from the original image. Nevertheless, some of components contain excessive areas that contributed by the halo of the membrane, especially when the cells are bright. Therefore, we further defined precisely the outline of the cells. The strategy is to apply serial of threshold to the individual cells until the halo disappears. We generated a curve by plotting the total intensity of each threshold image over the serial numbers of the corresponding images. From that curve, we further plot the fourth derivative of the original curve and one of the peaks on the new curve indicates serial number of the threshold image that defines the cells the best. Throughout the process, the coordinates and the pixel intensities of each cell can be retrieved from the original image to determine the value of total pixel intensity of a cell (figure 3).

Stage 2: Analyze protein distribution within each single cell

Since proteins fulfill their functions at specific intracellular locations, tracking their distribution is an important way to study their activities. For example, in our case, the target membrane protein triggers downstream cellular events when it is internalized to the cell. Therefore, precisely quantify the fraction of protein being internalized could help us to understand the level of activation.

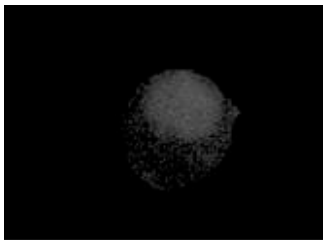

Fig.4 Intracellular part

The initial step is to separate the signal in two cell components: internal and membrane-localized. Since the size of intracellular part of a cell is proportional to the size of the whole cell, we “shrink” the cell by a factor equal to the average membrane thickness to get the total signal inside of the cell. Then, we use the same threshold-strategy as in stage 1 to eliminate background noise within the cells. This way, we can extract the intracellular signal contribution of the target protein (Figure 4). The obtained pixel intensities (Total cell, intracellular part, background) can be substitute into an equation to calculate the distribution of the target protein with an equation:

$$Intracelluar\ localization\ = \frac{Intracellular - Background}{Total - Background}$$

Current status:

We have established the core of the algorithm. Currently, we are intensively testing the program; to improve the accuracy of the results and to troubleshoot the errors.