Miguel Chavez Pachas - **Final Project Proposal**

Project Description: One important aspect of my current work is the identification and analysis of lipid droplet organelles within a cell via immunofluorescence staining and imaging. In order to extract as much relevant and useful information, it is imperative to analyze a sufficient number of samples to be able to make a proper conclusion from the data. Currently, the detection, segmentation, quantification, and statistical analysis process is manually labor intensive, especially since the process is extremely repetitive. Additionally, the variability in analysis adds subjectivity when executed manually. I propose a program that can visualize, detect, quantify, and normalize , extract some relevant information(diameter of LD, distance from nucleus, number of LDs per cell, for example) and generate statistical reports to simultaneously analyze and plot graphs along with relevant statistical data for each treatment. The program would be optimized to detect, quantify, and analyze lipid droplets due to their circularity and clearly defined shape features, however in principle this could be applied to any fluorescence signal.

Program Inputs:

File Types: image files (**.tiff, .jpg, .bmp, .czi** (for metadata))

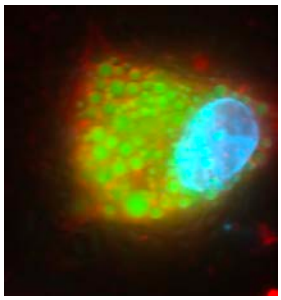
Note: there will be some constraints that the user will have to abide to such as naming conventions for each file (experimentname_date_identifier_channelwavelength), the requirement of a minimum of two channels (DAPI for nuclear count = cell count), (Lipid Droplet/compartment channel). This also requires the user to standardize the exposure time across all samples that are to be directly compared for intensity. However, if .czi files can be implemented, this would remove a lot of the manual work since the LUTs, metadata concerning scale and other data is already embedded in that file type which would save time and result in less manual steps.

These input file formats would be opened, processed, analyzed , and plotted with several existing libraries such as **numpy, skimage, pyplot, pillow**.

User Input: just establishing where directory where files to be analyzed are located (however if required, possibly a simple standardized script system that can be edited by the user to further simply the process of indicating certain parameters, file path

Program Outputs:

- Plot(s) **.png** containing a bar graph for each treatment group “Control”, Sample1, Sample2.etc (Treatment group (x), MFI - Lipid Droplet channel(y) or Total Lipid Droplet Count normalized to DAPI (y)
- a **.csv** file containing the plotted data in x,y (treatment, mean fluorescence intensity(normalized to DAPI) or total number of lipid droplets) coordinates, format and relevant statistical test results (error bars for example)



Project Risk List:

- Although we have done a great deal of plotting, processing images will be something new however I have been recently reading lots of documentation and utilizing resources to understand the libraries I would need to incorporate
- Although the tools exist to handle this analysis, the automation and adaptability of the program will be the most difficult parts; especially achieving a program that is high throughput and that requires the least amount of user intervention (although I am open to incorporating parameters for user to change if I get stuck on how to automate certain parts – for example checking histograms and requiring user input on threshold masks etc.)
- Although I recognize some aspects of my proposed program may be difficult to incorporate or automate, I think there are aspects I could simplify or some features that can be left out that can make the final project more manageable, and I remain very open to doing that.
- I recognize that to fully automate image processing, the use of machine learning or training sets to extrapolate thresholding/masks might provide a more adaptable system but my experience in integrating libraries like TensorFlow are limited which might hinder me achieving that