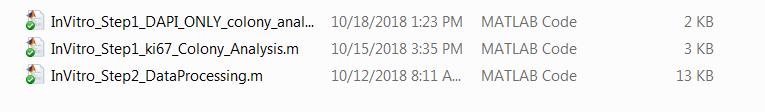
Read me.

# Introduction.

This document is a readme file to accompany the MATLAB scripts found at <https://github.com/sbhoyar1/ColonySizeDistributions> in the folders labeled InVitro\_Scripts and InVivo\_Scripts.

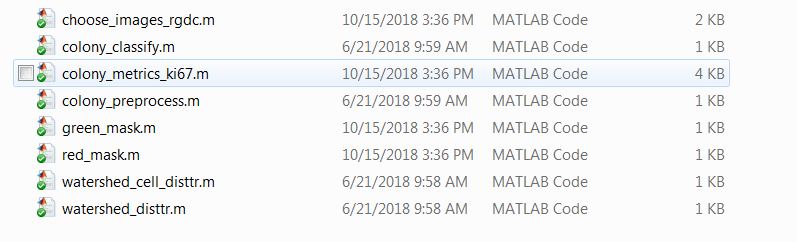
# In vitro Processing

The in vitro processing scripts can be found in the folder :\\MATLAB Files\InVitro\_Scripts.



The scripts are designed to process images that were taken in 6 well plates. The images from each 6 well plate must be kept in the same folder as the ‘InVitro\_Step1\_ki67\_Colony\_Analysis.m’ script. The individual image file names must be “<well>\_<channel>\_D\_XX\_<details>.tif” where ‘XX’ refers to the a timepoint in days.

All the functions located in :\\MATLAB Files\InVitro\_Functions\_Referred should be placed in the MATLAB functions folder.



## Preprocessing

In the ‘Preprocessing’ section, the user must specify the binarization threshold for the functions ‘green\_mask.m’ and ‘red\_mask.m’ to detect green and red fluorescent cells respectively. For DAPI and Ki67 channels, the threshold is not expected to vary significantly from the default value specified, if imaging parameters ensure that bright regions are just below saturation. More details can be found in the tutorial.

## Output

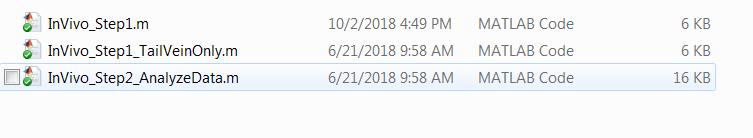
For each well, the ‘InVitro\_Step1\_ki67\_Colony\_Analysis.m’ script generates an output excel spreadsheet labeled ‘<Well>\_Data.xlsx’ containing colony level metrics including nuclear counts. These spreadsheets are inputs for the second MATLAB script. Next run the second script, ‘InVitro\_Step2\_DataProcessing.m’ You will need to have ‘<Well>\_Data.xlsx’ in the same folder as the script itself in order for the script to run.

The colonies are classified as red or green with the aid of the parameter labeled ‘Redfrac’. It quantifies what fraction of the total emission in the colony is due to RFP-expression. The parameter ‘redfrac\_limit’ is the limit that defines the boundary between red and green colonies and can be set manually in the ‘InVitro\_Step2\_DataProcessing.m’ script. The remainder of the script will generate graphical representations of the data. The graphical data will include plots of the probability distribution functions of the colony sizes as well as the distribution of Ki-67. The user may obtain the colony size data to produce custom plots as well. More details are provided in the tutorial.

# In Vivo Processing.

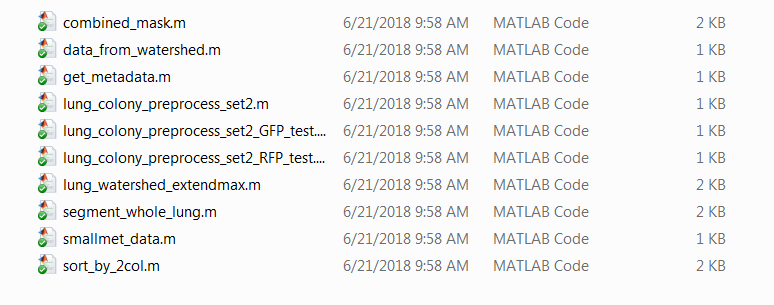
## Folder Structure

The script should be in the same folder as the image folders. Each image folder must contain the stitched GFP, RFP and DAPI of a given tissue section, in our case lung tissue. The ‘.tif’ images must have the channel name in the filename of the image. There must be only one image of each channel in a given image folder. The naming structure of the image folder is ‘M\_XX\_D\_YY\_<details> where XX is the mouse number and YY is the time-point. ‘M’ and ‘D’ are labels identifying sample number (in our case mouse number) and D can be used for any additional descriptor. We usually use the length of the experiment as a second descriptor.

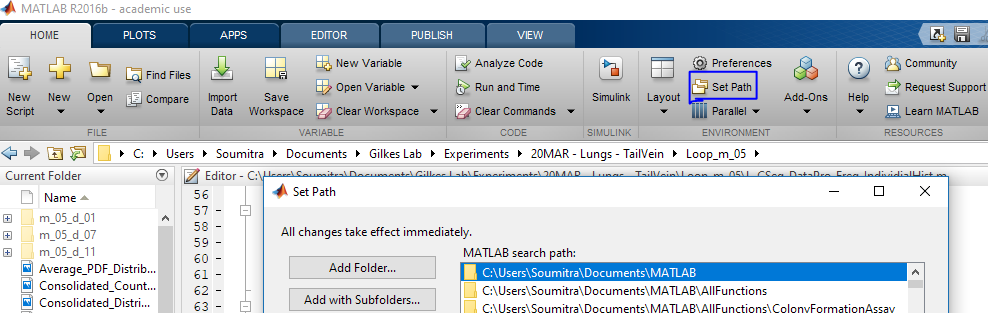


## Script Details

The in vivo processing requires several functions that should be placed in the MATLAB functions folder. All the supplied functions should be kept in a folder that is added to the MATLAB list of accessed folders. This can be done using the ‘set path’ button on the MATLAB home screen.



The ‘set path’ button is shown in the blue box:



In the ‘Preprocessing’ section, the ‘lung\_colony\_preprocess\_set2.m’functions are used. The threshold is set manually in the function based on the segmentation results. A few images are used to manually check segmentation, and the thresholds may be adjusted accordingly, before the entire set of images is processed. For additional details, please see the tutorial section.

In ‘Part 4’ describing the segmentation of small mets, the function ‘smallmet\_data’ the user can modify two numerical inputs – the gamma value, and the threshold. These can be used to adjust the segmentation. However, the default values provided are general values that work well in most applications.

The ‘Save Images’ section may be used to generate figures or cartoons of segmented regions.

The output is a single .xlsx file with details of each met, grouped in two sheets, RFP\_data and GFP\_data. This file is the input for Step 2.

## Step 2.

In the second step, executed by the script ‘InVivo\_Step2\_AnalyzeData.m’, the following salient points must be kept in mind.

The variable ‘cell\_area\_in\_pixels’ should be input and is based on the pixel to micron ratio of the microscope and objective used to capture the image.

The variable ‘MinCellLimit’ may be used to eliminate mets below a certain size, if we expect small artifacts in our images. We found that setting this value to 3 provide a good signal to noise ratio. The output of step 2 is a histogram that displays colony size distributions, with the size of the mets on the X-axis and the probability on the Y-axis. A probability density function is also obtained and plotted. Options for normalization are provided. Please see additional specifics described in the tutorial.