**READ-ME Translation Code**

**Version Information:**

* Code was created and tested on Python version 3.11.4.
* The code requires you to import the numpy package, the math package, the os package, the csv package, and the sys package.
* This program can be run on any computer and does not need specific hardware.

**Installation Guide:**

* After downloading the Python code file the program can be run immediately assuming that the computer it was downloaded on has a recent version of python installed and the packages listed above.

**Instructions for Use:**

**Run Time:**

* The run time is dependent on the number of images and the number of mRNA and peptide spots in those images. For large batches of images the program sometimes has to be run overnight. For the sample code it should take no longer then 5 minutes.

**Sample Data and Results:**

* Sample data and expected results to test the code can be found in the Marie-Vera Lab’s GITHUB (https://github.com/LR-MVU/neuron). To replicate the results you must follow the suggested parameters below.

**Instructions:**

* After downloading the code off the GITHUB and running the spot detection for both the mRNA and peptide spots using FISH-QUANT you can press run on the Translation Code.
* After you run the code you will be asked to input various pieces of information.

**Questions asked by the program:**

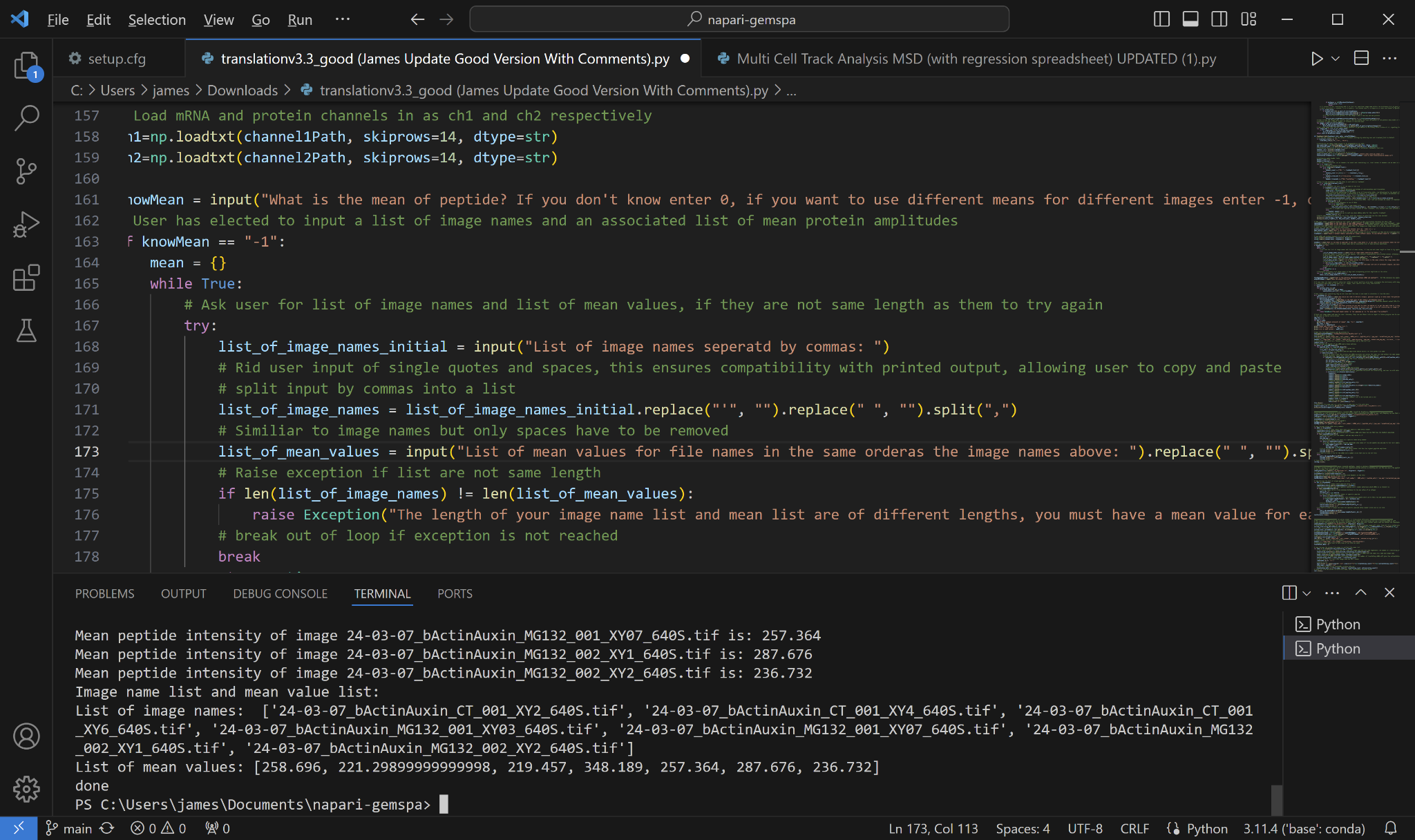
* **What is the full path of your mRNA channel?** Input the full file-path (without quotation marks) for your mRNA txt file produced by FISH-QUANT spot detection.
* **What is the full path of your peptide channel?** Similar to above but for the peptide spot detection txt file.
* **What is the protein channel path (ex. '640S.tif')** Input the channel path (no quotation marks), ending of your protein image file names. For example if the protein image file’s name is 24-03-07\_bActinAuxin\_CT\_001\_XY2\_640S.tif and the mRNA file name is 24-03-07\_bActinAuxin\_CT\_001\_XY2\_55S.tif then the protein channel is 640S.tif (the difference between the protein image name and mRNA image name). For sample data use 640S.tif.
* **What is the protein channel path (ex. '555S.tif'):** Similar to above but the mRNA channel (no quotation marks), ending of your mRNA image file names. For sample data use 555S.tif.
* **Input treatment labels separated by commas without spaces. To use default input 0:** This is used to create the treatment spreadsheets. Use the difference in naming between the different conditions. For example if the image names are 24-03-07\_bActinAuxin\_CT\_001\_XY6\_640S.tif and 24-03-07\_bActinAuxin\_MG132\_001\_XY03\_640S.tif the two conditions would be CT and MG132 and would be imputed “CT,MG132” (without the quotation marks). For the sample code use “Ctrl,MG132” (without the quotation marks).
* **What is the mean of peptide? If you don't know enter 0, if you want to use different means for different images enter -1, otherwise enter the mean for all images:** If you do not know the mean enter 0 and the program will calculate it for you (input 0 for the sample code). If you know the list of mean values you would like to use enter -1 and the program will prompt you for the image name list and the mean list. If you want to use a single mean value for all image names enter that mean values

**If -1 was entered the program will ask you further questions**

* + **List of image names separated by commas:** provide the program a list of image names that your protein channel has. This functionality should really only be used when copying in mean values previously generated and printed by the program. Example input: '24-03-07\_bActinAuxin\_CT\_001\_XY2\_640S.tif', '24-03-07\_bActinAuxin\_CT\_001\_XY4\_640S.tif', '24-03-07\_bActinAuxin\_CT\_001\_XY6\_640S.tif', '24-03-07\_bActinAuxin\_MG132\_001\_XY03\_640S.tif', '24-03-07\_bActinAuxin\_MG132\_001\_XY07\_640S.tif', '24-03-07\_bActinAuxin\_MG132\_002\_XY1\_640S.tif', '24-03-07\_bActinAuxin\_MG132\_002\_XY2\_640S.tif'
  + **List of mean values for file names in the same order as the image names above:** similar to the above question should be copied from previous program run’s print statement. Example input:

258.696, 221.29899999999998, 219.457, 348.189, 257.364, 287.676, 236.732

**Screenshot of output from where the example input above were copied:**

****

* **What is the distance threshold between mRNA and peptide?** This can be adjusted based on your specifications but we have been using 700nm (written into the program as 700) to account for chromatic aberration (400nm) and the sometimes large size of peptide that can cause the center to be further from mRNA than expected. Input 700 for the sample code.
* **Save the output file as:** Self explanatory (can be any phrase), we have been using result\_test.
* **How many proteins do you want to take the median of to get the amplitude of a single protein:** This question is asking how many proteins you would like to average per image to get the median amplitude of a single peptide. Averaging could potentially increase accuracy but at a cost of time spent running the program. Use 100 for the sample code.
* **What is the minimum number of points that need to be not translating for image to be included in analysis (minimum number of points averaged to get single peptide amplitude):** Lowest value possible would be 1, which would include any image that has one protein that is not translating that can be taken as the value of a single peptide. You can use a higher number for better accuracy. Any image that does not meet this requirement will not be included in the spreadsheets. Use 1 for the sample code.

**Explanation of files produced by program:**

* **fishQuantTranslation\_Results.txt** and **fishQuantTranslation\_Results.csv** (the same file in different formats)includes all pairs of mRNA and protein spots that are within the translating threshold distance before they are filtered for repeats.
* **Translating\_mRNA\_no\_repeats.txt** includes all pairs of mRNA and protein spots that are within the translating threshold distance after they have been filtered for repeats (detailed explanation in below pseudo-code).
* **untranslating.txt** includes all mRNAs that have not been assigned a nascent peptide spot and therefore are considered not to be translating.
* **cell\_translatingmRNA.txt** and **cell\_translatingmRNA.csv** (the same file in different formats) includes the number of translating and not translating mRNA per cell for each image analysed.
* **count.txt** includes the number of mRNAs per cell per image that have 0,1,2 … 20 nascent peptides being translated which is determined by the amplitude of the protein spot.
* **count\_by\_cell.csv** is similar to **count.txt** above but is in csv format and includes the number of translating and not translating mRNAs per cell.
* **count\_by\_image.csv** is similar to **count\_by\_cell.csv** but is by image rather than by cell within the images.
* **count\_by\_cell\_percent.csv** is similar to **count\_by\_cell.csv** above but has the number of mRNAs that have 0,1,2 … 20 nascent peptides being translated as a percent of total translating mRNAs.
* **count\_by\_image\_percent.csv** is similar to **count\_by\_cell\_percent.csv** but is by image rather than by cell within the images.
* **control\_heat\_count\_by\_image.csv** and **control\_heat\_translating\_by\_image.csv** reorganise the data found in **count\_by\_image\_perecnt.csv** to make it easier to plot in prism.

**Pseudo-Code:**

**Functions:**

**Write\_csv:**

* Takes a numpy array or python list of data, a python list of headers, and a file path as input and produces a csv file.

**ProtAvg:**

* Takes as input a numpy array of mRNA spots, a numpy array of protein spots, a distance limit between an mRNA and protein spot for the mRNA to be considered translating in nanometers, the number of mature proteins not localised to an mRNA per image to be averaged to get the amplitude of a single protein, and the minimum number of mature proteins that must be averaged for an image to be included.
* For each image determine which protein spots are not within the threshold distance given to the function.
* Average the first *X* number of spots where *X* is the maximum number of spots to average per image given to the function and set by the user. A limit is used to decrease the run time of the program.
* Return a dictionary where the image names are the keys and the median amplitude of mature proteins is the value. Exclude images from the dictionary that do not have enough mature proteins to average, a value given to the function and set by the user.

**Treatment\_data:**

* This function takes the final data from later on in the program and reorganises it by treatment type to make it easier to graph in prism. This function is not necessary for the program to run successfully.

**Main Program:**

**Get information and calculate median protein amplitudes:**

* Get information from the user. A detailed description of this information can be found in the instructions section above on how to run the code.
* The user can either input a single value for median single protein amplitude to be used across all images, a list of median single protein amplitude values and a corresponding list of image names in order to have different values for different images, or instruct the program to call the protAvg function to determine the median single protein amplitude values for each of the images. The program will print back the median amplitude values being used for each of the images for the users reference.

**Create initial translating mRNA-protein pair and untranslating mRNA txt and csv files:**

* Open txt files for both translating mRNA-protein pair and untranslating mRNA and write headers into txt file
* Loop through every mRNA and protein pair and determine whether the distance between the mRNA and protein spot is less than the translating threshold given by the user. If it is then add the pair to the initial translating mRNA-protein pair txt file. If an mRNA has no protein within the translating threshold distance add it to the untranslating txt file.

**Filter entries where the same mRNA spot is within the translating range of multiple protein spots:**

* Load in txt file created in last step
* Loop through all unique mRNA entry numbers and determine the brightest protein spot associated with each mRNA and write that into the new filtered translating mRNA-protein pair txt file.

**Filter entries where the same protein spot is within the translating range of multiple mRNA spots:**

* Load in txt file created in last step
* Loop through all unique peptide entry numbers and determine the mRNA that is the closest distance to the peptide and write that into the final filtered translating mRNA-protein pair txt file
* Add all mRNAs that no longer have any associated proteins to the txt file of untranslating mRNAs created in an earlier step.

**Determine the number of translating and untranslating mRNAs per cell:**

* Load final txt file created in the last step
* Loop through all cells within all images and count the number of mRNA-protein pairs for that specific cell, this is the number of translating mRNA in that cell.
* Subtract the number of translating mRNA in every cell from the total number of mRNA in the cell from the original mRNA FISH-QUANT file to get the number of untranslating mRNA.
* Do the same thing by images rather than by cells.
* Save info for by image in a txt file and save both by image and by cell as python lists for the next step.

**Determine the number of proteins being translated and create final spreadsheets**

* Load in the final filtered mRNA-protein txt file from two steps ago.
* Loop through all cells within all images and determine the number of peptides are being translated for each mRNA by looking at the protein spot amplitude. The number of mRNAs with each amount of peptides being translated is what is being determined.
* If the protein amplitude is below 0.5 times the median mature peptide amplitude (MMPA) for that image the number of peptides being translated is assigned 0, if it is between 0.5-1 times the MMPA it is assigned 1, if it between 1-(0.74\*2) times the MMPA then it is assigned 2, and beyond this the threshold is between 0.74N-(0.74(N+1)) where N is 2 or above and the assigned number of peptides being translated is N+1. Anything above 20 nascent peptides is combined together.
* Similar to the previous translating vs. untranslating step this process is repeated on a per image basis.
* The number of proteins being translated data is combined with the number of translating vs. untraslating mRNA data and written out as CSV files. There are 4 CSV files written; the first CSV file has the count of the number of mRNAs at each number of peptides being translated per cell, the second is the same as the first but by image, the third is the same as the first but as a percentage of total translating mRNA rather count numbers, and the fourth is the same as the third but by image rather than by cell. A more detailed explanation of these files is available above in the instructions section.
* Additional CSV files are then produced by the treatment\_data function to make the data easier to graph in prism. This is not necessary for the program to run.