**Image Processing Pipeline**

Pipeline described below is created for analyzing single molecule counting data coming from embryos in different genetic backgrounds. Python scripts are in bold and input and output files are italicized. Input and output Excel files shared across the scripts are color-coded.

Below is the file structure of this package:

|\_\_source: Source code for this package. Any “.py” or “.m” file should be stored here.

|\_\_preprocess\_imaris\_input: script to process imaris output .xls and/or .csv

|\_\_wildtype

|\_\_\_\_\_\_input: where experimental data files (named in the format: WT<embryo\_index>.xlsx) are placed.

|\_\_\_\_\_\_output:

|\_\_\_\_\_\_\_\_\_\_embryo<embryo\_index>: folders containing the results of analysis of each individual embryo.

Whenever you try to all input to this program, please follow this structure and add data in the right place in order to avoid the program crashing or not doing the proper analysis.

In this package, we used the following:

* Python 3
* Python modules: [scipy](https://www.scipy.org/install.html), numpy (included in scipy), [xlrd](https://pypi.python.org/pypi/xlrd), [xlwt](https://pypi.python.org/pypi/xlwt), [matplotlib](https://matplotlib.org/users/installing.html)

If you are analyzing a new set of embryos, you can either use the command-line arguments to directly run the Python and MATLAB scripts below or write a separate Python script like wildtype\_analysis.py, deltac\_analysis.py, or deltad\_analysis.py that calls the scripts in order as explained below. We provided you with these files; if you add/delete/modify experimental data, you may want to look at these files and change accordingly.

**wildtype\_analysis.py**

* Calls other Python scripts to analyze all embryos in a given genetic background
* The following input values need to be specified at the top of the script. These values are fixed throughout our study, therefore, we did not allow them to be specified through command-line arguments. If you change these values in your study, you can change them in the code of wildtype\_analysis.py; the code is thoroughly commented.
  + - prepare input sample list SampleInfo.xls
  + Column 1 : file name from imaris
  + Column 2 : python name
  + Column 3-4 : Left/right angle

--- not in used, could be random number

* + Column 5-8 : somite background
  + name of the folder for all input and output files (folder) --we assume that all raw experimental data files to be in input folder; all result figures and Excel files will be saved in output folder
* The following scripts are called to analyze the embryos:

1. **embryo\_analysis.py** for each embryo
   * Takes raw experimental data of an embryo and creates two Excel files containing data needed for future analyses
   * input: Excel file containing raw experimental data of an embryo
   * output: *slices.xls(background subtracted Her), sliceInfo.xls(raw Her count & border info for each slices), cells.xls* (inside ./wildtypefulldataset/output/embryo<embryo\_index> folder)
   * command-line argument format: python embryo\_analysis.py -i <input Excel file> -d <output directory> -a <initial angle> -dA <angle change rate> -n <number of sections> -f <0 or 1 to specify input format> -s <half threshold shift> -m1<background for her1> -m7<background for her7>
     1. angles: angels for slicing the embryos. Angles are measured counterclockwise from the positive x-axis. For wild-type, the angles for each embryo should be given through left\_angles and right\_angles variables in wildtype\_analysis.py .
     2. number of sections: The Excel file containing raw experimental data is divided into multiple (e.g. four) sections. Open the Excel file and count the number of sections (defined in the first row).
     3. input format: In some cases (mostly if the data is new), embryo data needs to be split into left and right and shifted along the x-axis so that the sections do not overlap. Open each input Excel file and determine whether the embryo has been divided into left and right regions. If the file does not specify whether a section is in the left or right region, it means the data has not been split in half and shifted yet. If this is the case, write “-f 0”. If the file does specify whether a section is in the left or right region, write “-f 1”.
   * raw data format: The first line contains header information. 6 columns together contain data for one section and sections are laid out horizontally. Within a section, each line corresponds to a single cell. Second, third, and fourth columns correspond to cell position coordinates (x,y,z). Fifth and sixth columns correspond to the number of *her1* and *her7* mRNA molecules/concentration detected within the cell. Sections may contain different number of cells, resulting in different number of rows.
   * **note:** This script needs regions.py and slices.py (both written by us) in the source folder. If a slice has fewer than three cells, we ignore the slice and write “Too few cells to analyze” to the *slice.xls*.

**Complete list of scripts in alphabetical order**

|  |  |
| --- | --- |
| Script | Description |
| regions.py | Defines class Region. Needed for embryo\_analysis.py. |
| shared.py | Needed for almost all python scripts. Includes function ‘ensureDir’ that creates an output folder if it doesn’t exist yet and two functions ‘isFloat’ and ‘isInt’ that check if a given string can be converted into a float or an integer. |
| slices.py | Defines class Slice. Needed for regions.py (and therefore embryo\_analysis.py). |
| wildtype\_analysis.py | Calls all scripts that analyze and combine data from all wild-type embryos. |

Written in July 2016 by Soo Bin Kwon (skwon94@ucla.edu), updated in Aug 2020 by Oriana Zinani ([Oriana.zinani@cchmc.org](mailto:Oriana.zinani@cchmc.org))

|  |  |
| --- | --- |
| genotype | angle curves used in the code for different genotypes |
| WT | y = 0.0052x + 54.487 |
| her1her7HET | y = 0.015x + 55.494 |
| her1her7HOM | y = 0.0052x + 54.487 |
| b567\_WT | y = 0.0054x + 65.295 |
| b567\_HET | y = 0.0572x + 47.289 |
| Gene-paired embryos | y = 0.022x + 54.056 |
| Gene-unpaired embryos | y = 0.0607x + 42.264 |
| WT\_21.5C | y = 0.0513x + 40.591 |
| WT\_28C | y = 0.0621x + 32.643 |