DESCRIPTION.md

Introduction

The images used for this exercise are taken from scans of breast cancer tissue.

The lab stained on such tissue different proteins of interest and acquired multiple images through an Imaging mass cytometry system.

Each image correspond to one monochromatic channel, each channel corresponds to one of the marked proteins.

Channels can be overlaid for visualization.

The breast cancer tissue consists of multiple cells, which can be distinguished using a single cell segmentation mask.

The single cell segmentation mask is an image of same size of the acquired stained channels; it assigns to every pixel the unique ID of the cell it belongs to.

The single cell segmentation mask is also used to identify the region of the image corresponding to a given cell.

Assignement

The challenge will assess both the coding skills and the organization skills. The candidate should:

1. Develop a library that fulfills the problem statement below
   - Use your choice of programming language
   - As the code is supposed to be reused by different members of the lab, the code should be organized in a way that could be used as both a command run from the command line and as a library that can be imported within existing code.
   - propose an API for the library
2. Create an online versioning repository that contains:
   - the source code of the library
   - tests to validate the library
   - documentation and examples on how to use the library
   - installation script for the following platforms - Linux (your choice of distro), Mac|Windows (optional - any distribution/version of your choice).
   - documentation on how to deploy and install the library
3. Setup an automated test environment to run the test every new update of the library code in the repository.

Problem statement

3 intensity channels images have been acquired from a breast cancer tissue scan.

The channels are represented as 3 separate .tiff files.

Pixels in each .tiff file measure the intensity of the corresponding channel.

1. Implement a channel overlay.

- Overlay the 3 channels in a single RGB image where each color R,G,B, corresponds to the intensity of each single channel.
- Save the overlay image in .png format on a local filesystem.

Input

- input data are found in `./data/images`

- HistoneH3(Yb176Di).tiff
- E-cadherin(Er167Di).tiff
- Fibronectin(Dy163Di).tiff

Output

- A single image in .png format that represents the overlay channels.

Note

- Output image should be comparable with the one found in `./results/1.png`

2. Highlight single cells

Given a single cell segmentation mask image, overlay it to the output image produced at previous step, and highlight the individual cells.

Input

- Output of the step 1
- Single cell segmentation mask is found in `./data/single-cell-mask/single_cell_mask.tiff`

Output

- Output a single image in .png format with the highlighted cells

Note

- Output image should be comparable with the one found in `./results/2.png`

3. Compute the mean of each channel

For each cell indentified in step 2, compute the mean of each channel R,G,B.

Input

- Output of step 2

Output

- Output a .csv file. Each row should contains: Cell_id - from the single cell mask -, channel#1 mean value, channel#2 mean value, channel#3 mean value

Reference

- The images provided in this exercise are a subset of The Nature Methods breast cancer raw data set