Microscopy data analysis practical - smFISH

Philipp Savakis, Evelina Tutucci

January 10, 2024

1 Overview

In this practical, you will explore and analyse microscopy image data from a single molecule fluorescence *in situ* hybridisation (smFISH) experiment on single cells of *Saccharomyces cerevisiae*. This type of experiment allows the visualisation of individual mRNA molecules. You are handed fluorescence data from one location of a microscopy sample. The data consists of of three z-stacks, corresponding to the following channels:

- DAPI (bulk staining of nucleic acids - useful to find the location of the nucleus)

- CY3 (visualises a specific mRNA - which one it is, you need to infer from the data)

- CY5 (visualises another specific mRNA - which one it is, you need to infer from the data)

2 Setting up

2.1 required software and data

- anaconda (install from website). You can install pycharm and/or spyder, but neither are required for this practical.

- image data folder from canvas (find the zip file pertaining to your group, download and unzip)

2.1.1 setting up virtual environment for napari

The software that you will be mainly interacting with is napari, a python based image viewer. For a clean setup, we will create a new python virtual environment. A virtual environment ensures that

- we use an appropriate python version

- all the dependencies (i.e. required software packages) are there

- the dependencies for this project do not interfere with other software on your machine

We will delegate the work of creating/organising the virtual environment to anaconda.

2.1.2 navigation to correct folder

Under windows, hit the windows key and type anaconda. You should see

```
anaconda prompt
```

as a choice. Open anaconda prompt.

You should see a command line interface that looks something like:

```
(base) C:\Users\YourUserName> _
```

Use the command line to navigate to the unzipped data folder you downloaded from canvas. There, navigate to the *biophotonics* folder. type

```
dir
```

to explore the directory's contents. You should find a file named

```
requirements.txt
```

On mac, navigate to the biophotonics folder in the group folder you downloaded from canvas. Right-click on the folder and select *open in new terminal window*. You should see a command line interface that looks something like:

```
(base) User@user: ~/whereever/you/donwloaded/the/folder/to/biophotonics$ _
```

2.1.3 creation of environment

The *(base)* in your command line interface tells you that you are in the base environment - we have not created a custom environment yet. Now, create a new environment by typing the following:

```
conda create -n biophotonics python=3.9 -y
```

This tells anaconda to create a new environment, which is named *biophotonics* and which will be set up using python version 3.9. We include the -y flag to confirm installation of the libraries that anaconda will download. You can leave this out for fuller control, but then will have to manually confirm installations.

After the environment is created, we need to activate it. Type:

```
conda activate biophotonics
```

On, windows, your command line should now show

```
(biophotonics) C:\The\location\of\your\group\folder> _
```

On mac, you should see:

```
(biophotonics) User@user: ~/whereever/you/donwloaded/the/folder/to/biophotonics$ _
```

We will now make use of *pip* for installation of most packages. Convince yourself that the folder you have navigated to contains a file named requirements.txt. Now, type

```
pip install -r requirements.txt
```

This should download a bunch of python libraries. After all is done, install the napari viewer, by typing:

```
pip install napari[all]
```

On mac, type

```
python -m pip install ''napari[all]''
```

To install the napari plugin for this course, type:

```
pip install -e .
```

Then, open napari, by typing:

```
napari
```

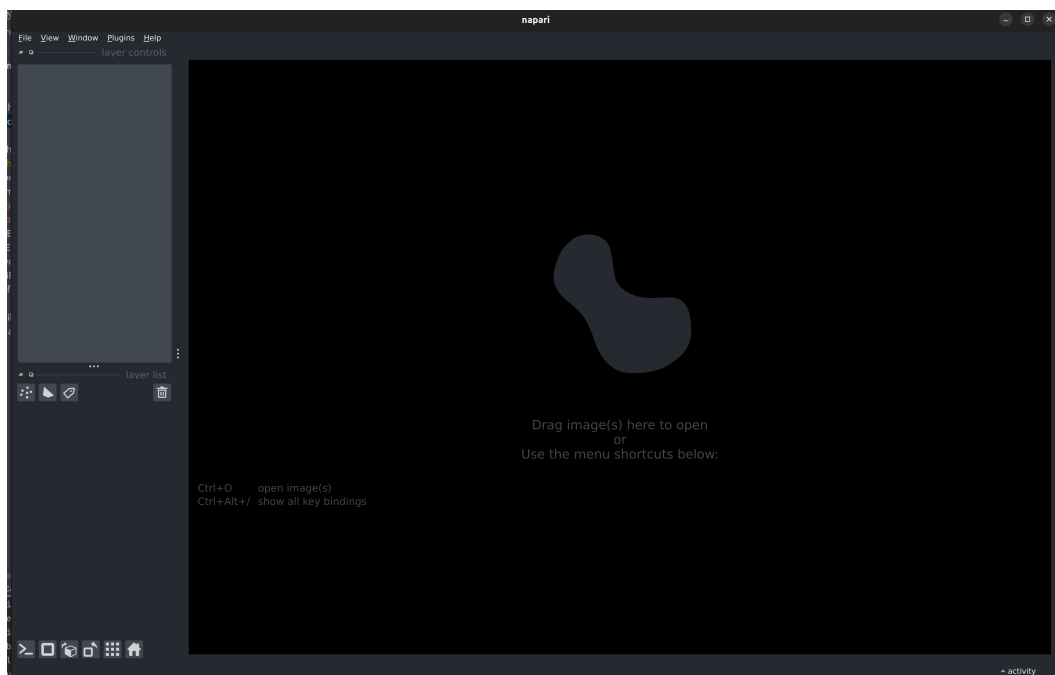If all is well, you should be looking at this window:

Figure 1: napari user interface

3   data exploration with napari

First, let's look at the data that was provided to you. Open the images by dragging them into the napari viewer. each .tif file in your /data/ folder has the following dimension: 41 x 2304 x 2304. This means that each image consists of 41 images, with height and with of 2304 pixels each:

You can visualise different z-layers through the slider at the bottom of the image.

By default, napari makes no distinction between distances in the x,y,z direction. Most microscopes, however, have lower resolution in z than they do in x/y. This means that two pixels in z direction are further apart than two pixels in x/y direction.

To tell napari that this is the case, we need to adjust layer scaling. We can do this by pressing the

```
>_
```

button on the bottom left of the napari interface.

This should open a console window. In there, type:

```
viewer.layers[0].scale = [3, 1, 1]

viewer.layers[1].scale = [3, 1, 1]

viewer.layers[2].scale = [3, 1, 1]
```

Napari has powerful 3D visualisation functionality, which will allow us to explore the data in 3D. Press the square button next to the console button (or use the shortcut ctrl+y) in the napari interface. Now, the image should be transformed into a 3D image. Use your mouse to move around in the image. Use the menu to the left to manipulate colour representations and transparency/opactiy of the data. Can you create a 3D-overlay, which shows the mRNA spots and the nuclei? Take a screenshot for your report.

When you are done exploring the data, restore the uniform scaling by typing:

```
viewer.layers[0].scale = [1, 1, 1]

viewer.layers[1].scale = [1, 1, 1]

viewer.layers[2].scale = [1, 1, 1]
```

into the napari console. Exit 3D visualisation mode through ctrl+y

4 image segmentation with napari

In this practical, we want to say something about the number of transcripts in individual cells. To do this, we need to find individual cells first. To find them, we could label each individual cell by hand, but if we want to finish this practical today, we can use something slightly less labour-intensive:

- Find the *Plugins* menu item at the top of the napari user interface.

- Open the biophotonics/Random forest plugin.

Ensure that you are in 2D view mode. For computational reasons, we need to restrict cell segmentation to a single 2304 by 2304 image. We can generate this image from the z-stack, by generating a max projection. We will use the DAPI channel for segmentation. So,

- select the DAPI layer

- adjust the z-layers you want to combine (*z level start/finish*)

- create the max projection with the *max projection* button.

This should create the following layers:

- `Labels`

- `FISH_DAPI max proj`

Now,

- Select

  `Labels`

- Familiarise yourself with the interface on the left.

- use the drawing tools to label a part of the image.

- for background use colour 1 (brown in the default colouring)

- for cell interior (nucleus and surroundings) use colour 2 (light blue in the default colouring)

- for the border between cell and background use colour 3 (purple in the default colouring)

Figure 2: interactive generation of training examples for random forest algorithm

Make sure that you use exactly these colors, or the underlying segmentation routine will do funny things. Once you are happy with your artwork

- select

  `FISH_DAPI max proj`

- hit the button

  `classify selected layer`

This creates a new layer named

`segmentation probabilites`

This layer represents the random forest algorithm's belief that a given pixel in the image corresponds to a given class (background, cell interior, or border between cell and background).

Given this belief landscape, we need to now find regions in the image that belong to the same object. We will use the watershed algorithm to achieve this. The watershed algorithm needs seedpoints, which we can generate from the DAPI channel. The rationale is that high fluorescence in the DAPI channel corresponds to a nucleus, and nuclei are typically at the centroid of a cell.

- Find the *Plugins* menu item at the top of the napari user interface.

- Open the biophotonics/watershed plugin.

- select the FISH_DAPI layer.

- adjust the intensity/distance parameters and press *find nuclei*

- remove superfluous layers (in case you had to use *find nuclei* more than once)

- check the nuclei layer you generated and add/remove points as needed

- select the *segmentation probabilities* layer

- click the *watershed* button

- this creates a *segmentation* layer.

- use the drawing tools to adjust cells as needed.

- save the segmentation image to disk (button on the right)

Saving the segmentation generates two files, one .tif file which holds the actual segmentation, and a .pkl file, which holds information about all the unique cell labels. We need this information later to indentify cells that are in the experiment but that have no mRNA counts.

5   Finding individual mRNAs

Normally, we would take the 3D data in all its glory and perform preprocessing and 3D gaussian fitting to capture point light sources (corresponding to single mRNAs) as precisely as possible. This is unfortunately beyond the capabilty of most laptops. Therefore, we have implemented a 2D spot finder for this practical.

- Find the *Plugins* menu item at the top of the napari user interface.

- Open the biophotonics/ spot finder plugin.

- select the FISH_CY5 layer and create a max projection

- discard the Labels layer that was created here

Now, to make individual spots more prominent, we will use a trick, in which the image is copied, and gaussian filters with different standard deviations are run over the indivudal copies. Then the difference between the images is computed (difference of gaussian, DoG).

You can adjust the gaussian filter through the sigma parameters, and then hit *compute difference of gaussian*. Play around with the values and inspect the resulting images.

Now, after having transformed the image with DoG, look at pixel intensities.

- Select the *difference of gaussian* layer

- move your cursor over the image

- at the bottom left of the napari user interface, you should see numbers corresponding to the xy position, and to the intensity at a given point

- find, which values are typical for background pixels

- do the same for spot pixels

- find a threshold value inbetween two separate spots from background

- hit the threshold button

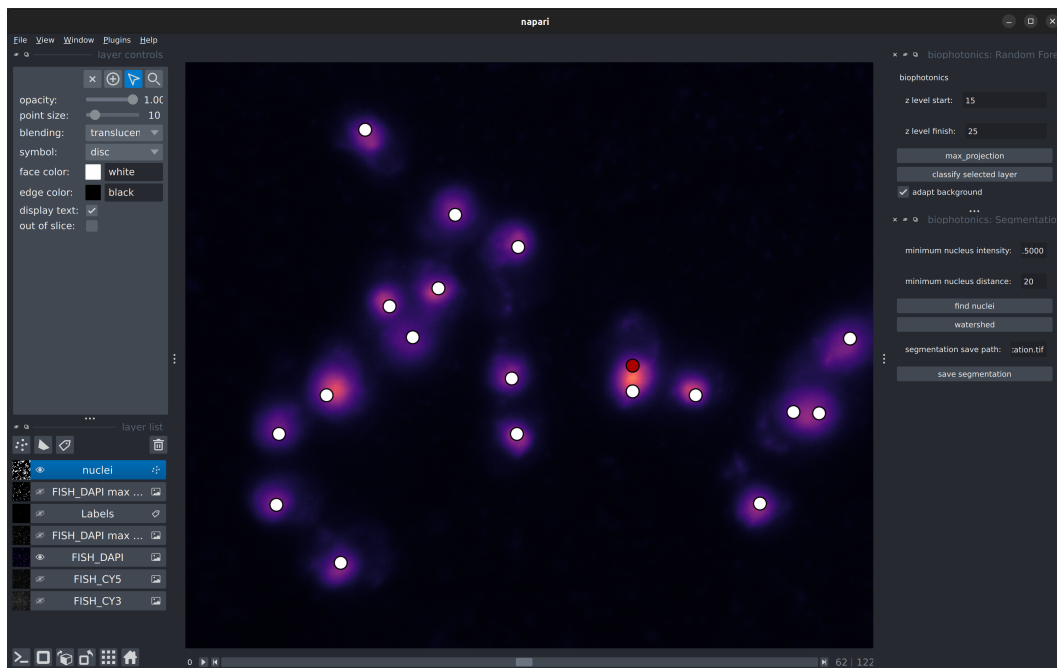- adjust the newly created coords layer as needed, e.g.:

Figure 3: nuclei coordinates after nuclear detection. Sometimes two spots are identified. Adjust these as needed.

- now, ensure that all the right layers are selected: spot-intensity layer should be your CY-max projection, and spot position layer is the layer that you created during thresholding.

- save the spots to a pkl file using the *save spot information* button

- repeat this section for the other channel, FISH_CY3

6 data processing

You should now be the proud creator of two .pkl files. These files are in a storage format that allows data/object exchange between different python instances. Your file system likely will not know what to do with them.

To look at the content of your files,

- open a new instance of anaconda prompt (windows) or terminal (mac/linux)

- no need to activate the napari environment here - the base environment should have all we need

- navigate to the folder *notebooks* in the biophotonics folder you downloaded from canvas

- type

  ```
  jupyter notebook
  ```

- open

  ```
  data_plotting.ipynb
  ```

- follow the steps laid out in the notebook

7 Finally

Find out which mRNAs were fished for in your group.

8 Report guidelines

The report should be divided into four sections:

1. Aim of the study

2. Methods

3. Results

4. Conclusions

You should

- describe the biological question and the imaging methods employed to create the datasets that you received.

- describe segmentation and spot detection (show pictures of the results)

- discuss problems and limitations of the approach

- add histograms for cell counts per cell and for mRNAs per transcription site and any further analysis that you think is important for this practical.

- discuss results and define gene expression mode of the genes you were assigned.

Criteria	Ratings	Pts
This criterion is linked to a learning outcome	This area will be used by the assessor to leave comments related to this criterion	
Structure and clarity of the narrative The report contains clearly defined parts: abstract, introduction, methods, results, discussion, references All these parts have appropriate content The storyline makes sense, there is logic in the transitions Report is complete and not excessively long		30
Correct representation of facts The facts taken from the literature or from your own knowledge are true Literature references are used		20
Data analysis and conclusions Data analysis is done properly and obtained results are shown You critically analyze obtained results and draw logical conclusions		30
Layout and figures The figures are well visible and explained, and clearly illustrate your text Proper use of fonts, margins, line spacings, etc		10
Grammar and spelling Minor mistakes and typos are not counted		10
		100