README.1.pyIRIS

Hao Yu

ISS_pyIRIS

Information Refining for *In situ* Sequencing
(Python3 version)

---

1. Introduction

This software is used to transform the fluorescent signal of *In Situ* Sequencing (ISS) into base sequences (barcode sequences), as well as the quality of base and their coordinates.

---

2. Installation

2.1. The development environment passed our test

- macOS 10.14.x
- CentOS 6.x and 7.x
- Ubuntu 16.04.x and 18.04.x
- Python3 3.7.x with models of
  - numpy
  - scipy
  - opencv-python 3.x or opencv-contrib-python 3.x

We didn't test our software on Microsoft Windows platform.

2.2. Installing Python3 models

We suggest to invoke pip3 to install Python3 models, like:

```
pip3 install numpy
pip3 install scipy
pip3 install opencv-python==3.4.x
          or opencv-contrib-python==3.4.x
```

Our software don't restrict the version of numpy and scipy, but we restrict the version of opencv-python or opencv-contrib-python to version 3.4.x

2.3. The directory structure of pyIRIS

---

3. Processable data type

We prepared 2 types of strategy to parse the different techniques of *in situ* sequencing, which published by Rongqin Ke and Chee-Huat Linus Eng.

Here, Ke's data (R Ke, Nat. Methods, 2013) is the major data type for our software, in this data, the barcode are composed of 5 types of pseudo-color, representing the A, T, C, G bases and background.

In the type of Eng's data, Each image is composed of 4 channels, of which, the first 3 channels means blobs with distinction by 3 pseudo-colors and the last one means background. Then, each continuous 4 images are made a Round, also named a Cycle. So, there are 12 pseudo-colors in a Round. For example, the Eng's data (CH Eng, Nat. Methods, 2017) include 5 Rounds, 20 images, 80 channels in any of shooting region.

3.1. The directory construction form of data

3.1.1. Directory construction form of R Ke

The Directory construction form of R Ke we recommended is like following table:

```
data directory
|---cycle 1
|   |---Y5.tif
|   |---FAM.tif
|   |---TXR.tif
|   |---Y3.tif
|   |---DAPI.tif
|
|---cycle 2
|   |---Y5.tif
|   |---FAM.tif
|   |---TXR.tif
```

```
|   |---Y3.tif
|   |---DAPI.tif
|
|---cycle 3
|   |---(...)
|
|---cycle 4
|   |---(...)
|
(...)
```

3.1.2. Directory construction form of CH Eng

The Directory construction form of CH Eng we recommended is like following table:

```
data directory
|---hyb1
|   |---pos1.tif
|   |---pos2.tif
|   |---pos3.tif
|   |---(...)
|
|---hyb2
|   |---pos1.tif
|   |---pos2.tif
|   |---pos3.tif
|   |---(...)
|
|---hyb3
|   |---pos1.tif
|   |---pos2.tif
|   |---pos3.tif
|   |---(...)
|
|---hyb4
|   |---pos1.tif
|   |---pos2.tif
|   |---pos3.tif
|   |---(...)
|
|---hyb4
|   |---(...)
|
```

```
|---hyb5
|   |---(...)
|
|---hyb6
|   |---(...)
|
|---hyb7
|   |---(...)
|
|---hyb8
|   |---(...)
|
(...)
```

---

4. Usage

You need only input the directories of cycle, like one of the following command if your data according to **Ke's**:

```
python3 pyIRIS.py --ke 1 2 3 4
python3 pyIRIS.py --ke {1..4}
```

You need input the the image files in each cycle, like one of the following command if your data according to **Eng's**:

```
python3 pyIRIS.py --eng 1/pos1.tif 2/pos1.tif 3/pos1.tif 4/pos1.tif
python3 pyIRIS.py --eng 1/pos2.tif 2/pos2.tif 3/pos2.tif 4/pos2.tif
python3 pyIRIS.py --eng {1..20}/pos1.tif
```

---

5. Result

There are two result files will be generated in your present directory, the base calling result and the background image file. Binding barcode info file, these two file would be used for following analysis of mating software, 'DAIBC'.

```
present directory
|---basecalling_data.txt
|---background.tif
```

5.1. The format of 'basecalling_data.txt'