

Annotate.OMERO.Fig guidelines

We provide an example dataset to run the Annotate.OMERO.Fig and guidelines of how to run it.

The pipeline can be run for two purposes; Firstly can be run as a standalone, independent and generalisable application to systematically cycle through a large set of images and score them based on a list of user-defined questions. These are in the end summarised in a tsv file together with information that are obtained through querying InterMine, with user specified queries. Instructions on how to run and customise the application are given below. Secondly if the user has images stored in OMERO, we have added python scripts to connect with OMERO via python, pull the user specified images and process them to be compatible with the input format to the scoring application. The scripts to do so are also available in this repository.

Setup

You need `python > 3` to run these scripts.

The project depends on the libraries in file "requirements.txt", the user should install them with `pip install "library"`

In case of an `XCRUN: ERROR:`, update the xcode by running the following in the command line: `xcode-select --install`.

Input files

1. Images to test
2. `figure_id_reference`: file containing the image name (without the extension) and mapping it to a unique gene identifier. An example set can be found in this github repository, mapping the figure ID to the CPTI line ID. For the whole set of images in this study the mapping can be found in the file, `figureid_to_geneid_mapper.xlsx` in Zenodo.
3. `Flyline_FBg_whole_genome.csv`: file mapping the CPTI lines to a Flybase gene ID. Contains information for the whole genome, not just for the genes corresponding to images being scored. In general this could be a file containing more information mapped to the unique identifier for each image contained in file `figure_id_reference`.
4. `datasets/testQueries/`: contains txt files specifying the queries that can be asked. More query examples can be found here: <https://www.flymine.org/flymine/querybuilder>.

Output Files

`zegami.tsv`: Output file containing information for the whole genome; also contains the scoring results for the images that were scored.

How to build the Annotate.OMERO.Fig application

1. Move the images you want to test in the folder "src/figures". These should be in jpg format.
2. Modify the file `questions.py` to set the questions you want to ask, instructions are in the comments in the file.

How to run the Annotate.OMERO.Fig application

Run the script from the src directory using the command-line with `python3 questionnaire.py questions.py answers/figures/*.jpg`

The command is of the form: `questionnaire QUESTIONS-FPATH SAVE-DIR [IMG-FPATHS ...]`

To view the first image to score, click on the "Open in Viewer" button on the graphical interface that pops up. To move to scoring the next image click the "save and next" button.

To make sure every image is only scored once in each cycle, if there is already an answer in the `answers/` folder the program can't be run for this image. So, if you want to run the program and score the images again make sure you empty the `answers/` folder.