

## ***Profiling Steps for “Cell Painting, an image-based assay for morphological profiling”***

The steps to produce per-well morphological profiles from single cell measurements are given below. A previously published dataset<sup>1</sup> is used as an example.

1. Install programs
  - a. Install miniconda from <http://conda.pydata.org/miniconda.html>  
`bash Miniconda-latest-Linux-x86_64.sh`
  - b. Install Python packages  
`conda create -n cellpntg MySQL-python numpy scipy matplotlib scikit-learn nose  
mock python-dateutil progressbar ipython  
source activate cellpntg  
pip install verlib`
  - c. Get CellProfiler Analyst  
`git clone https://github.com/CellProfiler/CellProfiler-Analyst.git  
cd CellProfiler-Analyst  
git checkout 598789b  
export PYTHONPATH=`pwd``
2. Upload CellProfiler measurements to a MySQL database. Note that this step can be skipped if the CellProfiler module *ExportToDatabase* was used to write the per-cell measurements directly to a MySQL database.
  - a. Setup a MySQL database server to upload measurements.
  - b. Download the MySQL dump files (.sql) of the measurements; these are contained in Supplementary Data 4.
  - c. From command-line, execute these commands to upload measurements to the MySQL database, where [user] and [password] is the user name and corresponding password to the SQL database server, and [database\_name] is the name of the database on the server which the measurements will be stored.  
`mysql -u [user] -p[password] [database_name] < seedseq_per_image.sql  
mysql -u [user] -p[password] [database_name] < seedseq_per_nuclei.sql  
mysql -u [user] -p[password] [database_name] < seedseq_per_cells.sql  
mysql -u [user] -p[password] [database_name] < seedseq_per_cytoplasm.sql  
mysql -u [user] -p[password] [database_name] < seedseq_per_object.sql`
3. Update configuration file to access the database.
  - a. Download the CellProfiler Analyst configuration file (.properties) which points to this data in the database; this is contained in Supplementary Data 4.
  - b. In the file seedseq.properties, edit the fields db\_host, db\_user, db\_passwd, image\_table and object\_table to the appropriate values.
  - c. Create a folder called “input” and save the file seedseq.properties to this folder.
  - d. Create a folder called “code” at the same folder level as “input”.
  - e. Execute steps 4-6 from within the “code” folder on the command line.
4. Create local cache of CellProfiler generated measurements
  - a. From command line, execute this command to generate a local cache of the measurements.  
`python -m cpa.profiling.cache -r \  
../input/seedseq.properties \  
../input/cache ""`
5. Create normalization parameters for the measurements.
  - a. From command line, execute this command to generate normalization parameters.  
`python -m cpa.profiling.normalization \  
../input/seedseq.properties`

```
../input/seedseq.properties \  
../input/cache ""
```

6. Create per-well profiles

- a. From command line, execute this command to generate per-well profiles.

```
python -m cpa.profiling.profile_mean \  
--method=median \  
--normalization=RobustStdNormalization \  
-o profiles.csv \  
-c -g \  
../input/seedseq.properties \  
../input/cache \  
Well
```

- b. Compare the profiles.csv file with the CSV provided as Supplementary Data 5 to ensure that the steps were executed correctly. Note that the ordering of rows will likely be different depending, and numerical values might differ slightly due to differences in floating point errors.

## References

1. Singh, S. *et al.* Morphological Profiles of RNAi-Induced Gene Knockdown Are Highly Reproducible but Dominated by Seed Effects. *PLoS One* (2015). doi:10.1371/journal.pone.0131370