

Installing Packages for TuniCUT

January 19, 2016

1 Installing TuniCUT

Shashank Gandhi

GitHub: [shashank357](#)

Email: shashank.gandhi@caltech.edu

Please follow this step-by-step tutorial to install TuniCUT on any local machine to run predictions.

1.1 Step 1:

Visit Anaconda's official [webpage](#) and look for Mac OSX 64bit **commandline installer**. Download the installer file for python 3.5.

Open Terminal, type:

```
cd ~/Downloads/
```

Press enter. Then type:

```
bash Anaconda3-2.4.1-MacOSX-x86_64.sh
```

Press Enter.

Anaconda will start installing. Enter “yes” whenever prompted. Anaconda includes conda, conda-build, Python, and over 100 automatically installed scientific packages and their dependencies. Like Miniconda, over 200 scientific packages can be installed individually with the “conda install” command.

Now close your terminal (Command+Q) and open it again. In order to check whether anaconda was successfully installed or not, type:

```
conda --version
```

on the terminal window and press Enter.

This version number will appear if anaconda was successfully installed.

1.2 Step 2:

Installing packages on the terminal.

If the terminal window is still open, type:

```
conda update conda
```

and hit enter. Then type:

```
conda update --all
```

and hit enter.

Then run the following installation commands:

```
conda install biopython
```

```
conda install scikit-learn
```

```
conda install seaborn
```

1.3 Step 3:

Running TuniCUT on a local machine:

Make a new directory on your Desktop and name it “Tunicut”. Please make sure that the following files are present in this new directory:

1. TunicutPredictions-v2.py
2. TrainingData.txt
3. CompleteData.csv

Open Terminal again. Type:

'cd ~/Desktop/Tunicut/

and press enter. This will take you to the folder where these files are present.

You can check that all the files are present in this folder by typing:

ls

and pressing enter.

In order to run TuniCUT for your sequences, you need your sequences as a Fasta File. Please note that since the prediction model needs to know exactly what base is present at what location, sequences that contain "NNN" will throw an error (or wont be considered for guideRNA design).

1.3.1 Run TuniCUT using the following command:

```
python TunicutPredictions-v2.py -f YourFastaFile.fa -d YourDatabaseName.db -o
YourOutputFileName.txt -t TrainingData.txt -c CompleteData.csv
```

Here's an example of how the command should look like:

```
python TunicutPredictions-v2.py -f LionelPredictions.fa -d TestJan14.db -o TestOutput.txt -t Training-
Data.txt -c CompleteData.csv
```

YourFastaFile.fa : Name of your fasta file containing sequences.

YourDatabaseName.db : An important feature of Tunicut. In the event that you have a large sequence (eg. entire chromosome) for which you have designed all possible guideRNAs, however, you only wanted guides in a specific region (eg. position 200 to 300, or a gene that you know the coordinates for), this database will come in handy. You can interact with your database and get sequences in specific regions only. To do that, simply type the following after you have run the above mentioned command.

```
sqlite3 YourDatabaseName.db
```

Press Enter.

In order to get sequences with PAM context "NGG", enter:

```
select * from TemplateguideRNA where Start > X and End < Y; , where X and Y are your start
and end coordinates.
```

In order to get sequences with PAM context "CCN", enter:

```
select * from NonTemplateguideRNA where Start > X and End < Y;
```

To exit this environment, simply type **.exit**.

Unfortunately, in the present version of TuniCUT, one has to find NGG and CCN sequences separately. This will be resolved in the next version.

YourOutputFilename.txt : Self explanatory. This is the text file that will have your results. This can easily be imported in microsoft excel if needed. Please ensure that you add ".txt" extension to the filename as it is not added by default.

CompleteData.csv: Contains entire dataset. Run as-is.

TrainingData.txt: Contains Training dataset. Run as-is.

Once you run the command, your results should be present in your current working directory. You can import it in excel or R (based on your personal preference of course).

If you want to install this on windows, please use "spyder" to run these commands as the DOS environment on windows does not support Linux commands.

In case of any question, email [Shashank Gandhi](#).