**cuRRBS**

*customized Reduced Representation Bisulfite Sequencing*

**0. Getting started.**

*cuRRBS* was written in bash, Python 2.7.11 and R 3.2.3. We can not guarantee that the software will work for different versions of the languages, although it is very likely that it will.

First of all, you will need to install Python and R distributions, that can be found here:

<https://www.python.org/downloads/>

<https://mirrors.ebi.ac.uk/CRAN/>

Afterwards, you should download or clone the *cuRRBS* repository in your machine.

Files needed to run the software

* Pre-computed files [COMPULSORY]. Downloaded or self-generated (see below).
* enzymes\_to\_pre\_compute.txt [OPTIONAL]. In case the pre-computed files are self-generated.
* Isoschizomer and methylation-sensitivity annotation file [COMPULSORY]. e.g. isoschizomers\_CpG\_annotation.csv . Explain meaning of 1,0,NA.
* enzymes\_to\_check.txt [COMPULSORY]. Make sure that all the pre-computed files are in the corresponding directory (otherwise errors will appear).

NOTE: many of these files used for our systems can be found in the “/utils” folder.

**1. Pre-computing the enzymatic digestions.**

*cuRRBS* makes use of some pre-computed files that contain the *in silico* digestions of the genome of interest. These files only need to be generated once for a given genome, which takes around 8 hours.

In order to save some time to the user, the pre-computed files for the human genome (hg38) can be downloaded from the following link:

<https://www.ebi.ac.uk/~dem44/cuRRBS_pre_computed_files/>

The pre-computed files must be unzipped in the working directory where the software will generate its files (see later in point 2). e.g. if I create a working directory folder in “~/Desktop/run\_cuRRBS”, the folder containing the pre-computed files should be “~/Desktop/run\_cuRRBS/pre\_computed\_files/”.

Advanced users

It is possible to generate your own pre-computed files from a genome that it is not available in the previous link. Check “cuRRBS/src/pre\_compute\_digestions.py” for more information. In any case, the pre-computed files should be stored in the working directory as described above.

The “enzymes\_to\_pre\_compute.txt” file should contain all the individual enzymes that will be used afterwards by the software (e.g. in our case, the first member of all the isoschizomer families that contain at least one methylation-insensitive enzyme). The script “iso\_annotation\_2\_pre\_computed\_enzymes.R” generates the “enzymes\_to\_pre\_compute.txt” file for a given annotation file (e.g. in our case, for the annotation file “isoschizomers\_CpG\_annotation.csv”).

NOTE: make sure that the necessary dependencies are properly installed.

**2. Configure the software.**

Install following dependencies:

Add cuRRBS to PATH:

|  |
| --- |
| The last step is adding the Somatypus directory to your PATH environment variable, so |
|  |

|  |
| --- |
| that the somatypus command can be called from the command line. You can do this either |
|  |

|  |
| --- |
| by editing your ~/.bashrc file with a text editor (e.g. nano) and adding the line: |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| export PATH=/\*FULL/PATH/TO\*/somatypus-x.x/src:$PATH |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| Or just by appending the relevant line to the file with: |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| cd path/to/somatypus-x.x |
|  |

|  |
| --- |
| echo "export PATH=$PWD/src:\$PATH" >> ~/.bashrc |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| Either way, it is then necessary to source the .bashrc file for the changes to be |
|  |

|  |
| --- |
| applied: |
|  |

|  |
| --- |
|  |
|  |

source ~/.bashrc