
EpiClass

Release 2.2.0

Dec 13, 2022

Contents:

1	Overview:	1
2	Quick Install:	3
3	Installation from GitHub:	5
4	The Vignette:	7
5	Quick Usage:	9
5.1	Generate methylation density table ({DREAMtoMD READtoMD}.DT,{timestamp}.csv)	9
5.2	Methylation Density Binary Classifier (MDBC) performance	10
5.3	For more information:	10
6	Indices and tables	13

CHAPTER 1

Overview:

EpiClass

Optimizing and predicting performance of DNA methylation biomarkers using sequence methylation density information.

CHAPTER 2

Quick Install:

We recommend to make conda environment with python==3.7 first:

```
conda create -n name python==3.7 pip  
pip install EpiClass
```


CHAPTER 3

Installation from GitHub:

A copy of the package can be obtained by downloading the git repository:

```
git clone https://github.com/bmill3r/EpiClass
```

Required dependencies can be found in `epiclass_env.yml`. Importantly, `epiclass` is meant for python 3.7.

`samtools` 1.6 was used in its development. It is required for running *epiclass* *READtoMD* on `.bam` sequence alignment files.

We first recommend installing in a fresh python virtual environment, either mediated by `conda` or `virtualenv`, using the `epiclass_env.yml`.

For `conda`:

```
cd EpiClass
conda env create -f epiclass_env.yml
conda activate epiclass
```

Then install:

```
python setup.py build
python setup.py install
```

Alternatively:

```
pip install EpiClass-*.*.tar.gz
```

(found in: `/EpiClass/dist/`)

Check that `epiclass` is installed with:

```
epiclass -V
```

or:

```
epiclass -h
```

CHAPTER 4

The Vignette:

For a deeper insight into how the code works and generating the manuscript figures, check out the vignette and associated jupyter notebooks:

```
https://github.com/bmill3r/EpiClass/blob/master/manuscript_figures/vignette/README_  
↪Vignette.ipynb
```

Quick Usage:

5.1 Generate methylation density table ({DREAM-toMD|READtoMD}.DT.{timestamp}.csv)

From either sequencing alignment reads or DREAMing methylation melt data.

5.1.1 For sequences:

- i. Each file must be its own .sam or .bam file. Point to the directory that contains the files to analyze.
- ii. Alignment files must be from Bismark (<https://www.bioinformatics.babraham.ac.uk/projects/bismark/>).
- iii. Command to process the sequencing reads:

```
epiclass READtoMD -i path/to/files --interval BED/with/genome/coordinates
```

The interval file is a bed file that has genomic coordinates for which methylation density information will be extracted for the samples. If multiple intervals passed, the analysis will be done using pooled information from all of them. Otherwise, can pass “chr:start-stop” for a single genomic region.

5.1.2 For DREAMing data:

- i. Use a raw melt temp .csv file as input. A custom template is provided.
- ii. Command to process the DREAMing melt data:

```
epiclass DREAMtoMD -i rawmelt.csv -temps tempsToMDs.csv -cpg numberCpGs
```

tempsToMDs.csv should be a two column file indicating the methylation density a melting temperature corresponds to for the given locus for which DREAMing was performed. numberCpGs indicates the number of internal CpGs within the given locus for which DREAMing was performed. Both of these are required.

5.2 Methylation Density Binary Classifier (MDBC) performance

Compute the optimal methylation density cutoff for the case and control samples, given the sequence methylation density profiles for the genomic region of interest (or pooled genomic regions):

```
epiclass MDBC -i densityTable.csv -a cases -b controls
```

By default, returns summary tables containing the TPR, 1 - FPR, AUC, and optimal read cutoff for each methylation density cutoff (MDC). Will also return ROCs and boxplots for the optimal MD.

This is done using either the sample read counts or the sample read fractions. As in, the number of reads with a given methylation density or higher, or the sample fraction of reads with a given methylation density or higher.

More extensive plots and sample information can be obtained with additional command flags.

Samples can be normalized based on relative sample fractions, and specific methylation density cutoffs can also be used.

5.3 For more information:

use:

```
epiclass -h  
epiclass READtoMD -h  
epiclass DREAMtoMD -h  
epiclass MDBC -h
```

5.3.1 Troubleshooting:

Overall: 1. Using full, absolute paths is probably a good idea...

5.3.2 Even more detailed information:

Follow the *Documentation* link under ‘Contents’ below:

Documentation

For an extensive example, check out the vignette on the github page:

```
https://github.com/bmill3r/EpiClass/blob/master/manuscript\_figures/vignette/README\_  
↪Vignette.ipynb
```

Check out the bioRxiv preprint:

DOI: 10.1101/579839

License

2019 Brendan F. Miller bmille79 <at> jh <dot> edu

This software is “United States Government Work” under the terms of the United States Copyright Act. It was written as part of the authors’ official duties for the United States Government and thus cannot be copyrighted. This software is freely available to the public for use without a copyright notice. Restrictions cannot be placed on its present or future use.

Although all reasonable efforts have been taken to ensure the accuracy and reliability of the software and associated data, the National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH) and the U.S. Government do not and cannot warrant the performance or results that may be obtained by using this software or data. NHGRI, NIH and the U.S. Government disclaim all warranties as to performance, merchantability or fitness for any particular purpose.

Please cite the authors in any work or product based on this material.

Need Help

Questions? Please contact: bmille79 <at> jh <dot> edu

CHAPTER 6

Indices and tables

- search