prism Documentation

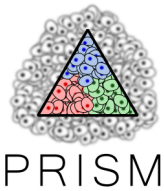
Release 1.0.1

Dohoon Lee

Jun 22, 2021

Getting started

Methylation **P**attern-based, **R**eference-free **I**nference of **S**ubclonal **M**akeup.



CHAPTER 1

Getting started

CHAPTER 2

What is PRISM?

Note: This section is under construction.

CHAPTER 3

How PRISM works

Note: This section is under construction

CHAPTER 4

Installation

Note: This section is under construction.

4.1 Installing via pip

PRISM is distributed as a python package named `subclone-prism`, so it can be easily installed via PyPI.

```
$ pip install subclone-prism
```

4.2 Installing via bioconda

PRISM will support installation via bioconda soon.

4.3 Installing from source

The latest version of PRISM can be installed by directly cloning the github repository.

```
$ git clone https://github.com/dohlee/prism.git
$ cd prism
$ python setup.py install
```

CHAPTER 5

Requirements

Note: This section is under construction.

PRISM requires BAM file of RRBS reads, aligned by Bismark.

PRISM strictly requires mapping results of Bismark, a bisulfite read mapping tool. Also note that PRISM only applies to RRBS data, and unfortunately, the feasibility of PRISM to the data from other methylation profiling techniques such as whole genome bisulfite sequencing (WGBS), methylated DNA immunoprecipitation sequencing (MeDIP-Seq), or methyl-CpG binding domain-based capture sequencing (MBDCap-Seq) has not been verified.

CHAPTER 6

Using PRISM

CHAPTER 7

Quickstart

Install PRISM with pip.

```
$ pip install subclone-prism
```

Run analysis.

```
------------------------------
Extract epiloci from BAM file.
------------------------------
$ prism extract -i sample.bam -o sample.met

---------------------------------------------------------------------
Preprocess epiloci to get finer estimates of epigenetic subclones
and to rescue more fingerprint epiloci from noisy methylation data.
---------------------------------------------------------------------
$ prism preprocess -i sample.met -o sample.corrected.met

-----------------------------------------------
Infer the subclonal composition of the sample.
-----------------------------------------------
If you want 1-sample deconvolution, run:
$ prism deconvolute -i sample.corrected.met -o sample.prism.result
or if you want 2-sample deconvolution, run:
$ prism deconvolute -i sample1.corrected.met sample2.corrected.met -o sample.prism.
↪result

-----------------------------------------
Scatterplot for visualization of the result.
-----------------------------------------
$ prism scatter -i sample.prism.result -o sample.png

---------------------------------------------------------------------
Annotation of fingerprint epiloci for functional characterization of
discovered epigenetic subclones.
```

(continues on next page)

(continued from previous page)

```
----------------------------------------------------------------------
$ prism annotate -i sample.prism.result -o sample.prism.annotated.result \
--beds annotation_a.bed annotation_b.bed \
--annotation-names ANNOTATION-A ANNOTATION-B
```

The PRISM analysis is done in three steps: `extract` - `preprocess` - `deconvolute`.

CHAPTER 8

Extract

Note: This section is under construction.

`prism extract` command extracts viable epiloci from a BAM file. In other words, it extracts genomic regions harboring a sufficient number of mapped reads (>= d) with a sufficient number of CpGs (>= c). A resulting file with those epiloci information is generated, whose file extension will be `.met` afterwards. To extract epiloci with default settings (d = 20, c = 4), simply run the command below:

```
$ prism extract -i sample.bam -o sample.met
```

If you want to specify your own cutoffs for the required sequencing depth and the number of CpGs, say, d = 15 and c = 3, run as follows:

```
$ prism extract -i sample.bam -o sample.met -d 15 -c 3
```

CHAPTER 9

Preprocess

Note: This section is under construction.

`prism preprocess` command corrects for the errors in methylation patterns in order to amplify the number of fingerprint epiloci and calibrate for the subclone size estimates.

```
$ prism preprocess -i sample.met -o sample.corrected.met
```

You may benefit from multithreading with `-t/--threads` option.

```
$ prism preprocess -i sample.met -o sample.corrected.met -t 30
```

This step is resource intensive, so by default PRISM tries to pre-filter the epilocus that is not likely to be a fingerprint epilocus. This pre-filtering of course can be turned off by `-f/--no-prefilter` option and this indeed gives a better estimates of subclones. However, please be warned, depending on your data size, this step will last long. To help you deciding whether or not to apply prefiltering, with 30 threads (`-t 30`) ~200M met file took about 5 hours to be preprocessed without prefiltering.

```
$ prism preprocess -i sample.met -o sample.corrected.met --no-prefilter -t 30
```

For a more detailed description about all options, run `prism preprocess -h`.

CHAPTER 10

Deconvolute

Note: This section is under construction.

`prism deconvolute` command infers the subclonal composition of the sample. Simply give methylation pattern-corrected epiloci file.

```
$ prism deconvolute -i sample.corrected.met -o sample.prism.result
```

Another feature of PRISM is that is can utilize two or more samples from a single tumor at the same time, and jointly infer subclonal composition. To provoke joint-analysis, specify a list of `corrected.met` files.

```
$ prism deconvolute -i sample1.corrected.met sample2.corrected.met -o sample.prism.
↪result
```

For a more detailed description about all options, run `prism deconvolute -h`.

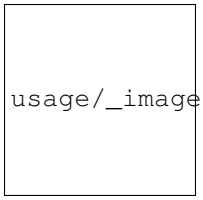
CHAPTER 11

Scatter

Note: This section is under construction.

`prism scatter` command generates a scatterplot of the PRISM analysis result. You need a result of `prism deconvolute`. The dimension of analysis (i.e., the number of samples you gave to `prism deconvolute` command) should not be more than three to visualize it. Note that the file extension of output file should be a general one for image files such as png, jpg, or pdf.

```
$ prism scatter -i sample.prism.result -o sample.png
```



CHAPTER 12

Annotate

Note: This section is under construction.

CHAPTER 13

Related information

CHAPTER 14

Citing PRISM

Note: This section is under construction.

CHAPTER 15

References

Note: This section is under construction.

CHAPTER 16

Quickstart

Install PRISM with pip.

```
$ pip install subclone-prism
```

Run analysis.

```
------------------------------
Extract epiloci from BAM file.
------------------------------
$ prism extract -i sample.bam -o sample.met

---------------------------------------------------------------------
Preprocess epiloci to get finer estimates of epigenetic subclones
and to rescue more fingerprint epiloci from noisy methylation data.
---------------------------------------------------------------------
$ prism preprocess -i sample.met -o sample.corrected.met

-----------------------------------------------
Infer the subclonal composition of the sample.
-----------------------------------------------
If you want 1-sample deconvolution, run:
$ prism deconvolute -i sample.corrected.met -o sample.prism.result
or if you want 2-sample deconvolution, run:
$ prism deconvolute -i sample1.corrected.met sample2.corrected.met -o sample.prism.
↪result

-------------------------------------------
Scatterplot for visualization of the result.
-------------------------------------------
$ prism scatter -i sample.prism.result -o sample.png

---------------------------------------------------------------------
Annotation of fingerprint epiloci for functional characterization of
discovered epigenetic subclones.
```

(continues on next page)

(continued from previous page)

```
-----------------------------------------------------------------------
$ prism annotate -i sample.prism.result -o sample.prism.annotated.result \
--beds annotation_a.bed annotation_b.bed \
--annotation-names ANNOTATION-A ANNOTATION-B
```

The PRISM analysis is done in three steps: `extract` - `preprocess` - `deconvolute`.