

BSpice.Window.Methylation Documentation

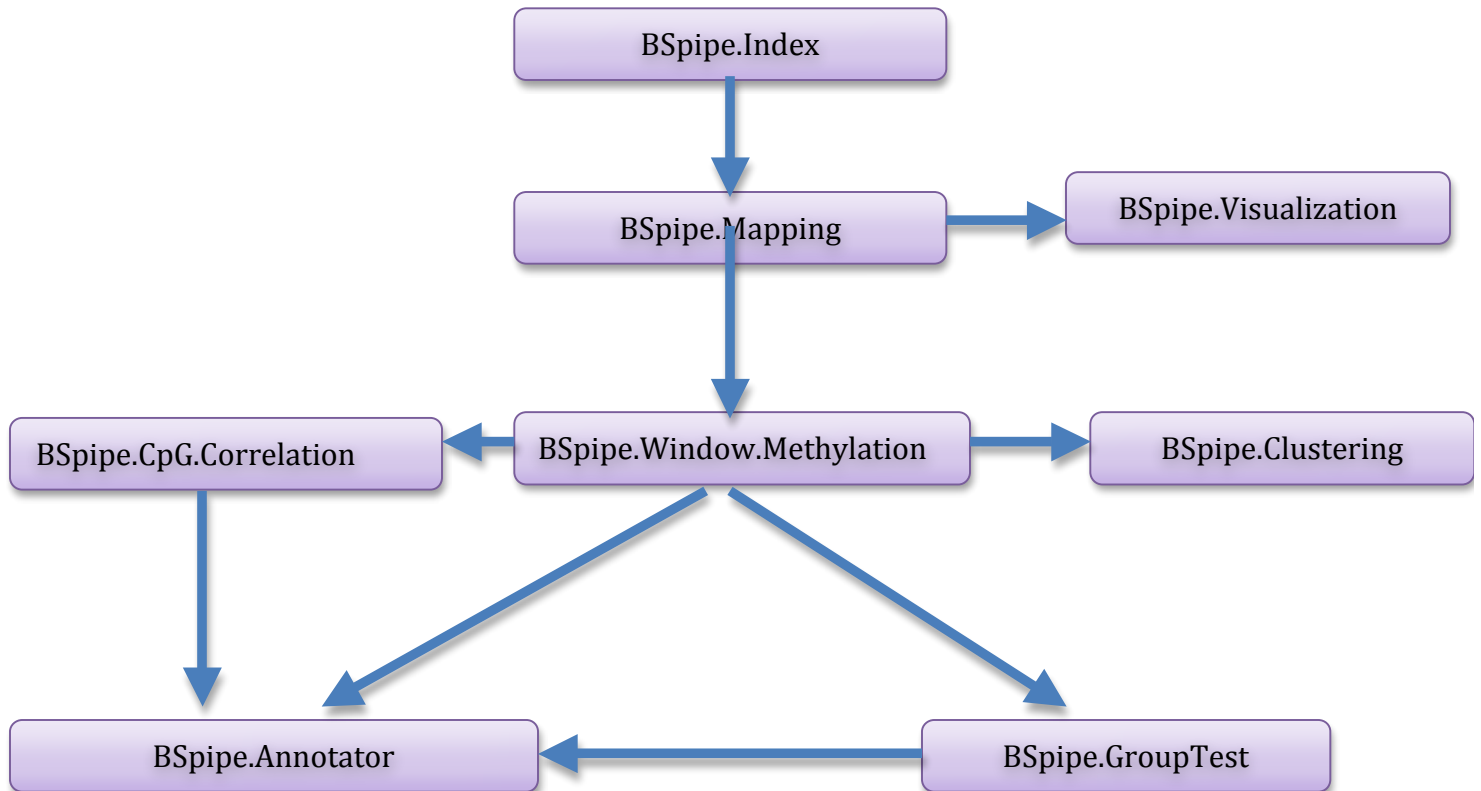
Description: Measures methylation across windows in samples mapped by BSpice

Author: GHSU COMICS

BSpice Version: 1.0

Bisulfite sequencing is a powerful technique to study DNA cytosine methylation. Bisulfite treatment followed by PCR amplification specifically converts unmethylated cytosines to thymine. Coupled with next generation sequencing technology, it is able to detect the methylation status of every cytosine in the genome. BSpice is an efficient bisulfite mapping software. It comes with a couple of utilities for annotation, calculation of CpG correlation etc.

The following is the flowchart of use of utilities in BSpice package.



BSpire.Window.Methylation is specific utility under BSpire software that measures methylation across windows.

Parameters:

Name	Type	Description
Input.file	*.bam / *.bed.gz	This can be any of the output files from BSpire.Mapping module. The program ultimately uses the entire directory as input.
Sample.configuration	tab delimited txt (.conf)	Refer next section (for multiple I/P files)
Window.size	int	The window size to be considered in basepairs [default - 200]
Step.size	int	The step size to be considered in basepairs [default - 200]
Project.name	text	Project name

RRBS.name	text	RRBS name in the reference conf file. This is an optional parameter. Use this only if you want to filter out CpGs present in restriction sites.
Reference.configuration.file	tab delimited txt (.conf)	This is an optional parameter. Use this only if you want to filter out CpGs present in restriction sites. [Refer next section]

Configuration files:

- **Sample Configuration file**

This option is used when multiple samples with groups have to be mapped. The sample conf file is a tab delimited file in the following format:

Column 1: Full path of input fastq file

Column 2: File type (either 1 / 2 if paired file or 1 if single reads)

Column 3: Sample Name

Column 4: Sample Group Name

A snapshot of sample configuration file is provided below.

```
GNU nano 1.3.12
M5.fastq      1      M5      Normal
M9.fastq      1      M9      Normal
M13.fastq     1      M13     Normal
M15.fastq     1      M15     Normal
M45.fastq     1      M45     Normal
M48.fastq     1      M48     Normal
MG18.fastq   1      MG18    Cancer
MG14.fastq   1      MG14    Cancer
MG8.fastq    1      MG8     Cancer
MG16.fastq   1      MG16    Cancer
MG10.fastq   1      MG10    Cancer
MGN.fastq    1      MGN     Cancer
```

- **Reference Configuration file**

This file saves information of the reference file. For convenience, BSpipeline can be run with the reference fasta file, which creates all necessary files and a reference configuration file.

The file is organized as below:

```
ref  reference name
seq  reference fasta (full path)
length  reference fasta.length (full path)
tc    reference c.fasta (full path)
ag    reference g.fasta (full path)
index_tc  program name  bsbowtie reference.c prefix
index_ag  program name  bsbowtie reference.g prefix
mspi  path to mspi.bed file.
```

A sample reference configuration file for hg19 is shown below.

```
GNU nano 1.3.12
ref      chr19
seq      chr19.fa
length  chr19.fa.length
tc       chr19.c.fa
ag       chr19.g.fa
index_tc          bowtie  chr19.c
index_ag          bowtie  chr19.g
mspi    chr19.mspi.bed
█
```

- **Running jobs on SGE**

Users can submit mapping jobs to the Sun Grid Engine (SGE), by including the SGE configuration file (optional). This option is not set by default. Administrators can edit the wrapper shell script to include the sge.conf file using the prefix '-c' in the command line.

N.B: For convenience, sge.conf file is available with the uploaded scripts in genepattern, which can be edited by the administrators.

Output files:

The output of BSpine.Window.Methylation is an .xls file, which contains information for all samples in the sample configuration file, with each line representing a window. For each sample there are 3 columns, first column being the number of windows applicable for the range. The second column refers to the number of methylated reads in that window range and the third column refers to the number of unmethylated reads in that window range,