

BSpice.Clustering Documentation

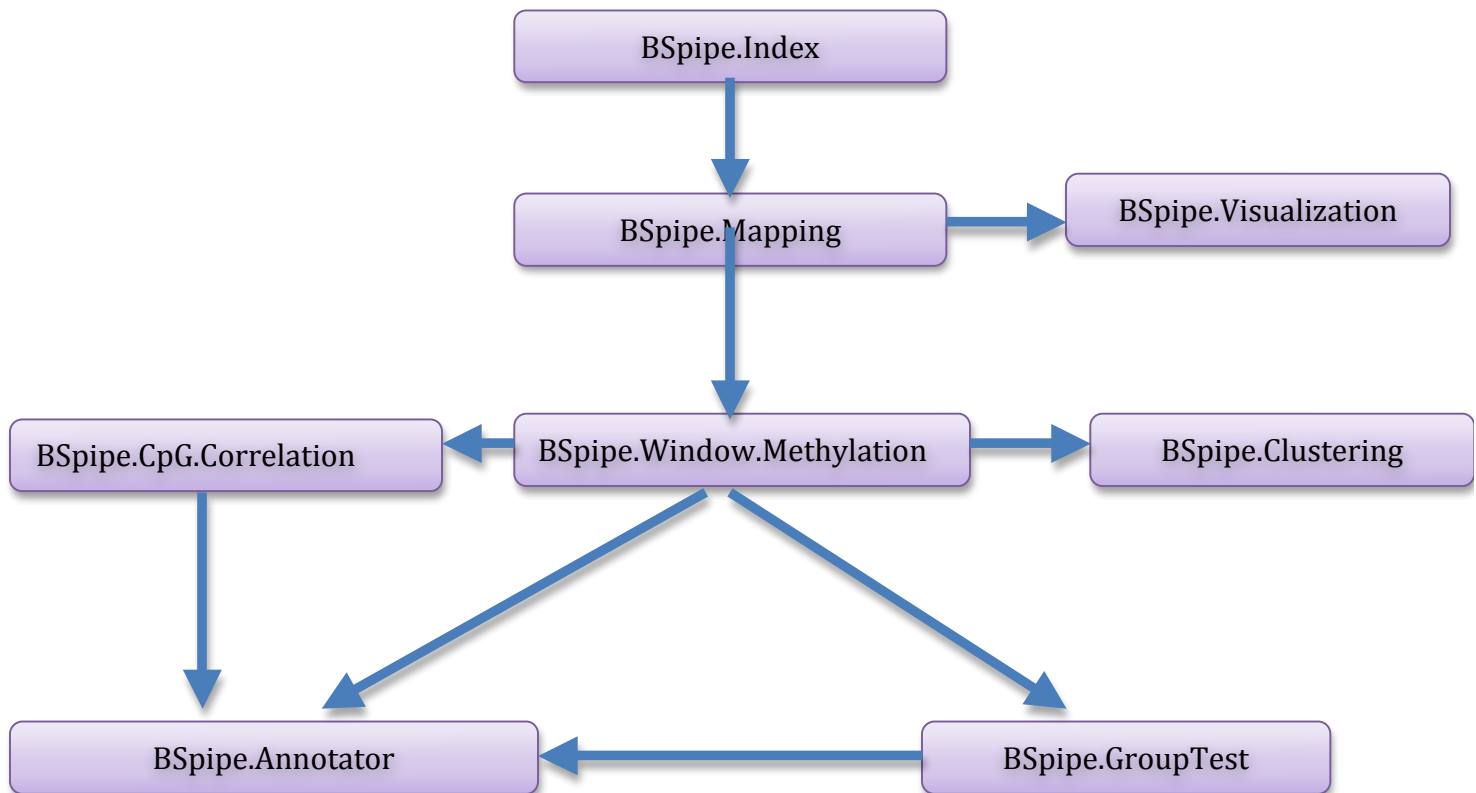
Description: Performs unsupervised clustering on methylation measured across windows.

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 including mapping, annotation, calculation of CpG correlation etc.

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



BSpice.Clustering is specific utility under BSpice software that performs unsupervised clustering.

Parameters:

Name	Type	Description
Input.excel.file	file (.xls)	The output .xls file generated by BSpice.Window.Methylation
Minimum.CpGs	int	Minimum number of CpGs in a window. The program filters windows with CpGs less than this value [default - 5]
Minimum.methylation.difference	float	Minimum methylation difference . The program filters windows with methylation difference less than this value [default -0.25]
Minimum.standard.deviation	int	Minimum standard deviation of methylation for clustering [default - 2]
Number.of.windows	int	Number of windows to be used while clustering
Sample.configuration.file	file (.conf)	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
```

- **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 program creates heatmaps in png format and xls format.