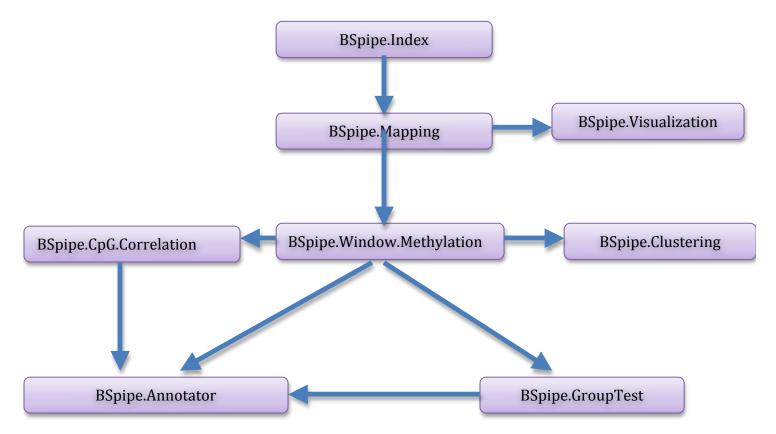# BSpipe.GroupTest Documentation

Description: Performs statistical tests across groups of samples.
Author: GHSU COMICS
BSpipe 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. BSpipe 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 BSpipe package.

BSpipe.Group is specific utility under BSpipe software that performs statistical tests between groups.

Parameters:

Name	Type	Description
Input.excel.file	file (.xls)	The .xls file created by BSpipe.Window.Methylation module.
Sample.configuration	tab delimited txt (.conf)	Refer next section (for multiple I/P files)
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]
Cutoff.Pvalue	float	Cutoff P value for differential methylation [default- 0.01]

Cutoff.fdr	float	Cutoff FDR for differential methylation [default – 0.05]
Heatmap.windows	int	Number of windows for heatmap [default -100]
Statistical.test	text	Choice of statistical tests to be performed. This includes Students T test, Wilcoxon Rank Sum Test, Rao-Scott Chi Square Test and Pairwise Fisher's Exact Test [default – Students T test]

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

An example of a sample configuration file is shown 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 performs the statistical test of choice and creates an excel file with the results. It also creates a histogram of p-values and adjusted p-values (png files). Also a volcano plot is created based on the adjusted p-values.

The above set of files are created for all locations being tested and filtered locations based on cutoff value.