User Guide of CpG_MPs v1.1.0

CpG_MPs is developed by Perl script wchich could be run in Linux or Window system. Users need to install Perl in the Window system before running the program. CpG_MPs mainly includes four modules including Data Normalization, Identification of Methylation Patterns of regions, Identification of Differentially and Conserved Methylated Regions, and Calculation the sequence features of the regions of DNA methylation patterns.

(1) Data Normalization

This module help you to convert methylation files into ".wig" format to display in ucsc.".wig" format is also the normal format which CpG_MPs requires.

USAGE: Normalization [OPTION] {--m_u <c1,c2>|--r_t <c1,c2>} [-i <file|folder>] [-o <folder>]

ARGUMENTS:

-h,--help,--man

Display the helpfile.

OPTION:

--depth

A depth threshold,DEFAULT:5.A site is valid when its depth is more than the threshold.

--head

The methy files has a header.

(The command requires only one in these following two options.)

-m_u,--methy_unmethy

The column of methylated counts and unmethylated counts.

-r_t,--ratio_total

The column of methylation level(from 0 to 1) and total counts(i.e. read depth).

I/O:

-i,--infile

The input methy file.Also can be a folder which has methy files splited by chromosome.REQUIRED

-o,--outdir

The output folder.REQUIRED

EXAMPLE:

BRIEF:

Normalization --methy_unmethy 5,6 -i test.txt -o out_folder

Split "test.cov" into "out_folder/" by chromosomes. methy counts is in col 5,unmethy counts is in col 6.

(the format of test.txt seems like ".cov" files comes from Bismark .)

OUTPUT:

Wig format will like this (chr1.wig):

track type=wiggle_0 name="Methylation Level" visibility=full color=20,150,20 altColor=150,20,20 windowingFunction=mean
variableStep chrom=chr1
3003660 1
3004613 0.87804
3004636 0.86956

****

Bismark and **BSMAP** is two popular BS-Seq align tools. This module can implement normalization for ".cov" files from Bismark(after v0.10.0) and ".bed" files from BSMAP.

Bismark: Normalization -m_u 5,6 -i H1.cov -o out
BSMAP: Normalization --head -r_t 4,5 -i H1.bed -o out

(2) Identification of Methylation Patterns of regions

This module help you to get methylation patterns from the methylation files after data normalization.

USAGE: CpG_MPs [options] [-i <folder>] [-o <folder>]

ARGUMENTS:

-h,--help,--man

Display the helpfile.

OPTION:

--methylation_level_U

The threshold of methylation level for unmethylated hotspot region. The default methylation_level_U is set as 0.3.

--methylation_level_M

The threshold of methylation level for methylated hotspot region. The default methylation_level_M is set as 0.7.

--hotspot_CpG_number

The successive CpG number (with methylation level ≤ methylation_level_U or ≥methylation_level_M) thresholdis used to find hotspot region (including unmethylated and methylated hotspot region).

The default hotspot_CpG_number is set as 3.

--abnormal_CpG_number

The CpG number with methylation level more than 0.5 is allowed to be ≤abnormal_CpG_number when extending the unmethylated hotspot region.

The CpG number with methylation level less than 0.5 is allowed to

be ≤abnormal_CpG_number when extending the methylated hotspot region. The default is set as 1.

--ucsc

export extra ".bed" files in outdir to display on UCSC.

I/O:

-i,--indir

The input methy folder (comes from CpG_MPs_normalization).REQUIRED

-o,--outdir

The output folder.REQUIRED

EXAMPLE:

CpG_MPs --ucsc -i NP_methy -o NP_pattern

Get methylation patterns from NP_methy folder.The result is in NP_pattern.

Export extra ".bed" files.

OUTPUT:

output will like this (chr13.patterns):

Chrosome	Start_position	End_position	Methylation_pattern	Length	Average_methylation_level	Variance
chr13	3000397	3127036	1	126640	0.897	0.120
chr13	3127725	3357222	1	229498	0.913	0.107
chr13	3357289	3360770	-1	3482	0.047	0.078
chr13	3361137	3405393	1	44257	0.902	0.114

The ucsc ".bed" format will like this:

track itemRgb="On" name="Methylation Region" description="methylation patterns regions" visibility=3

chr13	3000436	3038679	MR	0	+	3000436	3038679 58,181,74
chr13	3038888	3127036	MR	0	+	3038888	3127036 58,181,74
chr13	3127725	3357354	MR	0	+	3127725	3357354 58,181,74
chr13	3357380	3360770	UMR	0	+	3357380	3360770 238,29,35
chr13	3360809	3405393	MR	0	+	3360809	3405393 58,181,74
chr13	3406351	3477301	MR	0	+	3406351	3477301 58,181,74
chr13	3477396	3477886	UMR	0	+	3477396	3477886 238,29,35
chr13	3478075	3536684	MR	0	+	3478075	3536684 58,181,74
chr13	3536764	3538770	UMR	0	+	3536764	3538770 238,29,35

(3) Identification of Differentially and Conserved Methylated Regions

You can use this tool to get differentially methylation regions from mutiple methylomes.

USAGE: CpG_MPs_Dif_Con [options] [--patterns <folder_string>] [--methy <folder_string>] [-o <folder>]

ARGUMENTS:

-h,--help,--man

Display the helpfile.

OPTION:

--V_value

The value is set to estimate the identified differentially methylation regionor conserved region covered by the extent of samples.

The identified region is more reliable with larger V_value.The default V_value is more than 0.5.

--CpG_number

The identified differentially methylation region or conserved region must contain ⩾CpG_numberCpG sites. The default CpG_number is set as 4.

--ucsc

export extra ".bed" files in outdir to display on UCSC.

I/O:

--patterns

The patterns folder string.Split by ",".REQUIRED. like this: NP_patterns,ESC_patterns

--methy

The input methy folder (comes from CpG_MPs_normalization).REQUIRED

Split by ",".like this NP_methy,ESC_methy

-o,--outdir

The output folder.REQUIRED

EXAMPLE:

CpG_MPs_Dif_Con --ucsc --patterns NP_patterns,ESC_patterns,PGC_patterns --methy NP_methy,ESC_methy,PGC_methy -o NP_ESC_PGC

Get differentially methylation regions and conserved methylation regions from NP_methy folder.

The result is in NP_ESC_PGC.Export extra ".bed" files.

OUTPUT:

output will like this (chr1.dmr):

Chromosome	Start_position	End_position	Length	Average_methylation_level	V_value	Entropy
chr1	3943631	3943805	174	0.673_0.094_0.673	1	0.192695486142842
chr1	4588643	4589123	480	0.233_0.706_0.233	1	0.308232831955781
chr1	4679290	4679477	187	0.897_0.273_0.897	1	0.156921500047012

chr1 4759803 4760445 642 0.561_0.261_0.561 1 0.694229364256945
chr1 4774603 4774873 270 0.917_0.152_0.917 1 0.0759645439391204

(4) Calculation the sequence features of the regions of DNA methylation patterns

This tool help you to get sequence feature for methylation patterns.

USAGE: Seq_Feature [--ref <folder>] [-i <folder>] [-o <folder>]

ARGUMENTS:

--help,--man,-h

display the helpfile.

I/O:

--ref

the reference genome folder(.fa).REQUIRED

-i,--indir

the input directory, methylation patterns folder. REQUIRED

-o,--outdir

the output directory.REQUIRED

EXAMPLE:

BRIEF(only required arguements):

Seq_Feature --ref ref_genome_folder -i patterns/ -o seq_feature/

get sequence feature for methylation patterns in patterns/.

OUTPUT:

the output will like this(chr1.seqf):

Chrosome Start_position End_position Methylation_pattern Length Average_methylation_level Variance GC_content CpG_OE

chr1 3000574 3329885 1 329312 0.897 0.118 0.389521183558449 0.146981950106983
chr1 3334584 3470553 1 135970 0.889 0.110 0.380260351548136 0.120034873515113
chr1 3473945 3539119 1 65175 0.890 0.115 0.388937476026084 0.138771040623581
chr1 3540334 3638079 1 97746 0.909 0.093 0.384077097784053 0.125952462499578