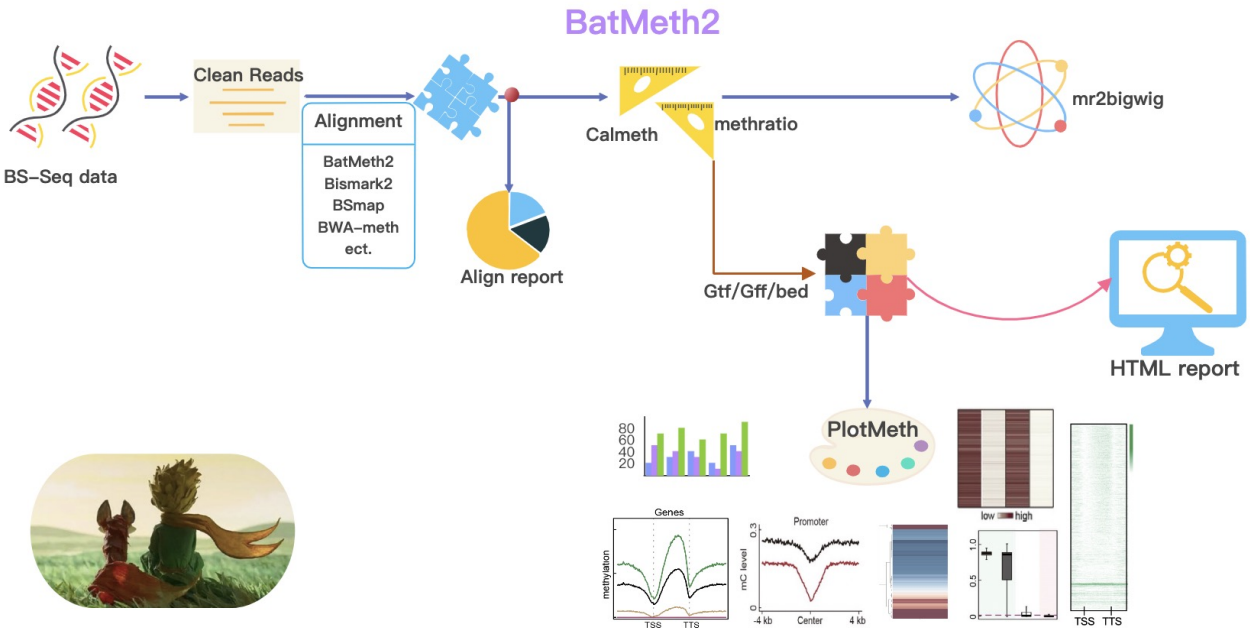
BatMeth2

Release 1.1

momocoding

May 11, 2022

CONTENTS



BatMeth2 is an **easy-to-use, auto-run** package for DNA methylation analyses. In order to complete the DNA methylation data analysis more conveniently, we packaged all the functions to complete an easy-to-use, auto-run package for DNA methylation analysis. During the execution of BatMeth2 Tool, an html report is generated about statistics of the sample.

Installation

- Please download and install the tools (see *Installation*)

The functions you can use BatMeth2 to do:

- *Alignment*: Align bsseq data
- *Calculate DNA methylation level*: Calulate DNA methylation level (ML) across whole genome
- *Calulate mC across predefined regions*: Calulate DNA ML profile or heatmap across gene / TE or peak region
- *Meth2BigWig*: Convert ML txt file to BigWig format, used for IGV visulization
- *DiffMeth*: Perform differential analyses with auto defined regions or predefined regions.
- *PlotMeth*: Plot DNA ML profile, heatmap or boxplot across genes/TEs/etc.

CHAPTER

ONE

CONTENTS

1.1 Installation

BatMeth2: An Integrated Package for Bisulfite DNA Methylation Data Analysis with Indel-sensitive Mapping.

1.1.1 Requirements

- gcc >= v4.8
- gsl
- zlib
- samtools >= v1.3.1
- fastp, raw reads as input need.

The details of requirements can see *Requirements*

1.1.2 Install

```
a) git clone https://github.com/ZhouQiangwei/BatMeth2.git
b) Change directory into the top directory of BatMeth2
    $ cd BatMeth2
c) Type
    $ ./configure
    $ make
    $ make install
e) The binary of BatMeth2 will be created in bin/
```

Tip: For feature requests or bug reports please open an issue on github.

1.2 Example Data

1.2.1 Data

You can download the test data on *https://drive.google.com/open?id=1SEpvJbkjwndYcpkd39T11lrBytEq_MaC*

Or *https://pan.baidu.com/s/1mliGjbn_33wlQLieqy5YOQ* with extraction code: *kr32*.

Example data contain files:

- input fastq.gz (paired end)
- genome file
- usage code and details
- gene annotation file

1.2.2 Citation

[Zhou Q, Lim J-Q, Sung W-K, Li G: An integrated package for bisulfite DNA methylation data analysis with Indel-sensitive mapping. BMC Bioinformatics 2019, 20:47.](https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2593-4)

Tip: For feature requests or bug reports please open an issue on github.

1.3 Build index

BatMeth2: An Integrated Package for Bisulfite DNA Methylation Data Analysis with Indel-sensitive Mapping.

1.3.1 Genome index

- Have a fasta-formatted reference file ready, and then make the neccessary pairing data-structure based on FM-index.

For WGBS type

```
BatMeth2 index -g genome.fa
```

or for RRBS

```
BatMeth2 index_rrbs -g genome.fa
```

Run *BatMeth2* to see information on usage.

Tip: For feature requests or bug reports please open an issue on github.

1.4 Pipeline

Tip: BatMeth2 can perform one click analysis or the following modules step by step:

- *Alignment*
- *Calculate DNA methylation level*
- *Calulate mC across predefined regions*
- *Meth2BigWig*
- *PlotMeth*
- *DiffMeth*

tool	input files	main output file(s)	main application
Alignment	Single/Paired-end fastq/gz files	alignment sam/bam file	Perform DNA methylation level calculation and SNP/ASM detection
Calculate DNA methylation level	BS-seq align sorted sam/bam file	methratio file (loci/region)	Perform DNA methylation visulization on chromosome, and diff analysis
Calulate mC across predefined regions	methration file from calmeth	methlevel file on genes/TEs etc.	DNA methylation profile or heatmap on genes/TEs/peak regions/etc.
Meth2BigWig	methration file from calmeth	bigwig files (c/cg/chg/chh)	Convert DNA methylation file to bigwig format.
PlotMeth	methy files from calmeth/methyGff	methy profile/heatmap/boxplot	visulization of DNA methylation across samples
DiffMeth	methration file from calmeth	Diff methy cytosines/regions	Perform Differential DNA methylation analysis

1.4.1 BatMeth2 pipeline

An easy-to-use, auto-run package for DNA methylation analyses:

Raw reads:

```
BatMeth2 pipel --fastp ~/location/to/fastp \
-1 Raw_reads_1.fq.gz -2 Raw_read_2.fq.gz \
-g ./batmeth2index/genome.fa \
-o meth -p 8 --gff ./gene.gff
```

Or clean reads:

```
BatMeth2 pipel -1 Clean_reads_1.fq.gz -2 Clean_read_2.fq.gz \
-g ./batmeth2index/genome.fa \
-o meth -p 8 --gff ./gene.gff
```

You can always see all available command-line options via --help:

```
$ BatMeth2 --help
```

- After the program runs successfully, a series of files with '- o' as prefix and DNA methylation level will be generated in the output directory. Please refer to the doc for the specific output file and format details.
- In addition, there will be an HTML report file containing basic information and statistical results of data analysis.

1.4.2 BatMeth2 pipeline main parameters

Build index

Usage: (must run this step first)

1. Build index using for wgbs data

```
$ BatMeth2 index -g genomefile
```

2. Build index using for rrbs data

```
$ BatMeth2 index_rrbs -g genomefile
```

Main Alignment paramaters

[Fastq Quality Conreol]	
--fastp	fastp program location
If --fastp is not defined, the input file should be clean data.	
[Main paramaters]	
-o	Name of output file prefix
-O	
	Output of result file to specified folder, default output to current folder (./)
[Aligners paramaters]	
-g	Name of the genome mapped against
-i	
	Name of input file, if paired-end. please use -1, -2, input files can be separated by commas. eg. -1 readA.fq.gz,readB.fq.gz -2 ..
-1	Name of input file left end, if single-end. please use -i
-2	Name of input file left end
-p	Launch <integer> threads
-n	maximum mismatches allowed due to seq. errors [0-1]

Calmeth paramaters

--Qual	calculate the methratio while read QulityScore >= Q. default:20
--redup	REMOVE_DUP, 0 or 1, default 1
--region	Bins for region meth calculate , default 1000bp.
-f	
	for sam format outfile contain methState. [0 or 1], default: 0 (dont output this file).
--coverage	>= <INT> coverage. default: 4
--binCover	>= <INT> nCs per region. default: 3
--chromstep	>= <INT> nCs per region. default: 3
	Chromosome using an overlapping sliding window of 100000bp at a step of 50000bpdefault step: 50000(bp)

MethyGff/Annoation paramaters

--gtf/--gff/--bed/--bed4/--bed5	
	gtf / gff / bed files, bed: Chr start end; bed4: Chr start end strand; bed5: Chr start end id strand;
-d/--distance	
	DNA methylation level distributions in body and <INT>-bp flanking sequences. The distance of upstream and downstream. default:2000
--step	
	Gene body and their flanking sequences using an overlapping sliding window of 5%of the sequence length at a step of 2.5% of the sequence length. So default step: 0.025 (2.5%)
-C	<= <INT> coverage. default:1000

1.4.3 Output files

Output file format and details see "https://github.com/GuoliangLi-HZAU/BatMeth2/blob/master/output_details.pdf".

Output report details see "https://www.dna-asmdb.com/download/batmeth2.html" .

Tip: For feature requests or bug reports please open an issue on github.

1.5 Alignment

1.5.1 BatMeth2 align

Single-end-reads

DNA methylation sequencing single-end data alignment:

```
An example usage is:
    batmeth2 -g /data/index/genome/genome.fa -i Read.fq.gz -o outPrefix -p 10
```

Paired-end-reads

DNA methylation sequencing paired-end data alignment:

```
An example usage is:
    batmeth2 -g /data/index/genome/genome.fa -1 Read_R1_left.fq.gz -2 Read_R2_right.fq.
↪gz\
    -o outPrefix -p 10
```

Parameters

[Main paramaters]	
--inputfile/-i	bs-seq input fastq files, fastq format or gzip format
--genome/-g	Name of the genome mapped against, MUST build index first *Build index*
--outputfile/-o	Name of output file prefix
--threads/-p	Launch <integer> threads
--non_directional	Alignments to all four bisulfite strands will be reported. Default: OFF.
--insertsize/-s	inital insert size, default 600, will be aoto detected by input files
--std/-d	standard deviatiion of reads distribution, will aoto detected by input
--flanksize/-f	size of flanking region for Smith-Waterman
--swlimit	try at most <integer> sw extensions
--indelsize	indel size
--NoInDels/-I	not to find the indels result
--help/-h	Print help

Note: To use BatMeth2, you need to first index the genome with *Build index*.

Tip: For feature requests or bug reports please open an issue on github.

1.6 Calculate DNA methylation level

1.6.1 Calmeth

Calculate DNA methylation level from alignment files, you can obtained single-base cytosine DNA methylation results, and the chromosome region DNA methylation levels files.

```
An example usage is:
  with bam file:
    calmeth [options] -g genome.fa  -b alignment.sort.bam -m output.methrario.txt
  with sam file:
    calmeth [options] -g genome.fa  -i alignment.sort.sam -m output.methrario.txt
```

Important: The bam or sam file MUST sorted by *samtools sort*.

1.6.2 Paramaters

[Main paramaters]	
-m/--methratio	[MethFileNamePrefix] Predix of methratio output file
--genome/-g	Name of the genome mapped against, MUST build index first *Build index*
-i/--input	Sam format file, sorted by samtools sort.
-b/--binput	Bam format file, sorted by samtools sort.
-Q [int]	caculate the methratio while read QulityScore >= Q. default:20
-n [float]	Number of mismatches, default 0.06 percentage of read length. [0-1]
-c|--coverage	>= <INT> coverage. default:4
-nC	>= <INT> Cs per region. default:1
-R/--Regions	Bins for DMR caculate , default 1000(1kb) .
--binsfile	
	DNA methylation level distributions in chrosome, default output file: {Prefix}.methBins.txt
-s/--step	
	Chrosome using an overlapping sliding window of 100000bp at a step of 50000bp. default step: 50000(bp)
-r/--remove_dup	REMOVE_DUP, default:true
-f|--sam [outfile]	
	f for sam format outfile contain methState.
--sam-seq-beforeBS	Converting BS read to the genome sequences.
--help/-h	Print help

1.6.3 Output files

```
1. prefix.methratio.txt
2. prefix.methBins.txt
3. prefix_Region.CG/CHG/CHH.txt
4. prefix.mCdensity.txt
5. prefix.mCcatero.txt
```

1.6.4 Output file format

```
1. methratio
    Chromosome Loci Strand Context C_count CT_count methlevel eff_CT_count rev_G_count␣
↪rev_GA_count MethContext 5context
    # ex. Chr1    61      +       CHH     3       11      0.286364        10.5    20      ␣
↪ 21      hU      ATCTT
    # C_count      The number of C in this base pair.
    # CT_count     The number of coverage in this base pair.
    # eff_CT_count Adjust read coverage based on opposite strand.
    # rev_G_count  The number of G in the reverse strand.
    # rev_GA_count The number of coverage in the reverse strand.
    # MethContext  M/Mh/H/hU/U, M means the methylation level  80%, etc
2. methBins
    Chrom BinIndex methlevel context
```

(continues on next page)

(continued from previous page)

```
    # ex. Chr1    1       0.113674        CG
    # The BinIndex is defined by -s paramater in calmeth.
    # This file can be used for visualization the DNA methylation level acorss the␣
↪chromosome.
3. Region
    chrom regionStart strand context c_count ct_count
    # ex. Chr1    1001    +       CG      1       227
    # The bins methylation level output file (BS.mr_Region.C*.txt) can be used to do DMR␣
↪detection.
4. mCdensity
    CG/CHG/CHH C count in [0, 1%) [1%, 2%) ... [49%, 50%) ... [99%, 100%]
    # According to the DNA methylation level, the number of cytosine sites at different␣
↪methylation levels was counted from 0 to 100.
5. mCcatero
    Average DNA methylation level including mC, mCG and other states.
```

Tip: For feature requests or bug reports please open an issue on github.

1.7 Calulate mC across predefined regions

- *methyGff*
- *Paramaters*
- *Output files*
- *Output format*

1.7.1 methyGff

Through GTF, GFF or bed files, the methylation level of the designated region and upstream and downstream was calculated, and the methylation level matrix was generated. The generated methylation level file and matrix file can be used to generate profile and Heatmap visualization.

- The methratio file calculated by calmeth, the format is *chrom pos strand context nC nCover methlevel*.

For exsample: *chr1 34 - CHG 2 14 0.142857*

```
An example usage is:
  with gtf file:
    methyGff -B -o gene.meth -G genome.fa -gtf gene.gtf -m output.methrario.txt

  with multiple gtf file:
    methyGff -B -o expressed.gene.meth unexpressed.gene.meth \
        -G genome.fa  -gtf expressed.gene.gtf unexpressed.gene.gtf -m output.methrario.
↪txt

  with bed file:
```

(continues on next page)

(continued from previous page)

```
  methyGff -B -o gene.meth -G genome.fa -b gene.bed -m output.methrario.txt

with multiple bed file:
  methyGff -B -o expressed.gene.meth unexpressed.gene.meth \
      -G genome.fa -b expressed.gene.bed unexpressed.gene.bed -m output.methrario.txt
```

Important: The number of input gtf/gff/bed files must be the same as the number of output prefixes..

1.7.2 Paramaters

Command Format : methyGff [options] -o <OUT_PREFIX> -G GENOME -gff <GFF file>/-gtf <GTF file>/-b <bed file> -m <from Split methratio outfile> [-B][-P]

[Main paramaters]	
-o/--out	Output file prefix
--genome/-G	Name of the genome mapped against, MUST build index first *Build index*
-m|--methratio	DNA methratio output file, generated by the tool calmeth
-c|--coverage	>= <INT> coverage. default:4
-C	<= <INT> coverage. default 600.
-nC	>= <INT> Cs per bins or genes. default:1
-gtf/-gff	Gtf/gff file
-b	Bed file, chrom start end
-b4	Bed file, chrom start end strand
-b5	Bed file, chrom start end geneid strand
-d/--distance	
	DNA methylation level distributions in body and <INT>-bp flanking sequences. The distance of upstream and downstream. default:2000
-B/--body	Calculate the DNA methylation level of per region.
-P/--promoter	Calculate the DNA methylation level of per region's upstream [d]k.
--TSS	Caculate matrix for TSS. [Outfile: outPrefix.TSS.cg.txt]
--TTS	Caculate matrix for TTS. [Outfile: outPrefix.TTS.cg.n.txt]
--GENE	Caculate matrix for TSS. [Outfile: outPrefix.TSS.cg.txt]
--TTS	Caculate matrix for GENE and flank [d]k. [outPrefix.GENE.cg.txt]
-s/--step	
	Gene body and their flanking sequences using an overlapping sliding window of 2% of the sequence length at a step of 1% of the sequence length. So default step: 0.01 (1%)
-bl/--bodyLen	Body length to which all regions will be fit. (default: same as -d)
-S/--chromStep	
	Caculate the density of genes/TEs in chromsome using an overlapping sliding window of 100000bp at a step of 50000bp, must equal "-s" in Split.. default step: 50000(bp)
--help/-h	Print help

1.7.3 Output files

Caution: Output

```
1. prefix.meth.AverMethylevel.txt
2. prefix.meth.Methylevel.txt
3. prefix.meth.TSSprofile.txt
4. prefix.meth.centerprofile.txt
5. prefix.col-0.meth.annoDensity.txt
6. prefix.meth.body.c*.txt
# run methyGff with -B paramater
7. prefix.bdgene.Promoter.c*.txt
# run methyGff with -P paramater
8. prefix.bdgene.TSS.cg.txt
# run methyGff with --TSS paramater
9. prefix.bdgene.TTS.cg.txt
# run methyGff with --TTS paramater
10. prefix.bdgene.GENE.cg.txt
# run methyGff with --GENE paramater
```

1.7.4 Output format

```
1. AverMethylevel
    CG/CHG/CHH meth_of_bin1 bin2 bin3 ... bini binj ... binN
    ## N is defined by -s paramater
2. Methylevel
    CG/CHG/CHH UP/BODY/DOWN Methylevel
    # per line means 1 gene/TE/region
3. TSSprofile
    CG/CHG/CHH meth_of_bin1 bin2 bin3 ... bini binj ... binN
    # DNA methylation level across TSS
    # -d N kb, TSS upstream and downstream N kb
    # -s move step
4. centerprofile
    CG/CHG/CHH meth_of_bin1 bin2 bin3 ... bini binj ... binN
    # DNA methylation level across region center
    # -d N kb, center point upstream and downstream N kb
    # -s move step
5. annoDensity
    chrom pos methlevel strand
    # ex. Chr1    0       0.559940        +-
    # The density of region distributions on chromsome
6. body/Promoter
    chrom regionStart strand context C_count CT_count regionID
    # ex. Chr1    3631    +       CG      45      1314    AT1G01010
    # This file can be used for visualization using `PlotMeth:bt2profile` or␣
↪`PlotMeth:bt2heatmap`
7. TSS/TTS/GENE
    regionID meth_of_bin1 bin2 bin3 ... bini binj ... binN
```

(continues on next page)

(continued from previous page)

```
    ... ...
    # This file is the methylation matrix across all genes, per line represents one␣
↪region (gene/TE/etc)
    # This file can be used for visualization using `PlotMeth:bt2heatmap`
```

Tip: For feature requests or bug reports please open an issue on github.

1.8 Meth2BigWig

BatMeth2: An Integrated Package for Bisulfite DNA Methylation Data Analysis with Indel-sensitive Mapping.

- *methratio2bw*

1.8.1 methratio2bw

- The methratio file calculated by calmeth, the format is *chrom pos strand context nC nCover methlevel*.

For exsample: *chr1 34 - CHG 2 14 0.142857*

For bigWig with strand information

```
python batmeth2_to_bigwig.py -sort -strand genome.fa.fai prefix.methratio.txt
# genome.fa.fai can be prepared by `samtools faidx genome.fa`
```

or for bigWig without strand information

```
python batmeth2_to_bigwig.py -sort genome.fa.fai prefix.methratio.txt
```

Run *BatMeth2* to see information on usage.

Tip: For feature requests or bug reports please open an issue on github.

1.9 PlotMeth

BatMeth2: An Integrated Package for Bisulfite DNA Methylation Data Analysis with Indel-sensitive Mapping.

- *python library*
- *bt2profile*
- *bt2basicplot*
- *bt2chrprofile*
- *bt2heatmap*

1.9.1 python library

install library required

```
pip install numpy
pip install pandas
pip install matplotlib
pip install seaborn
```

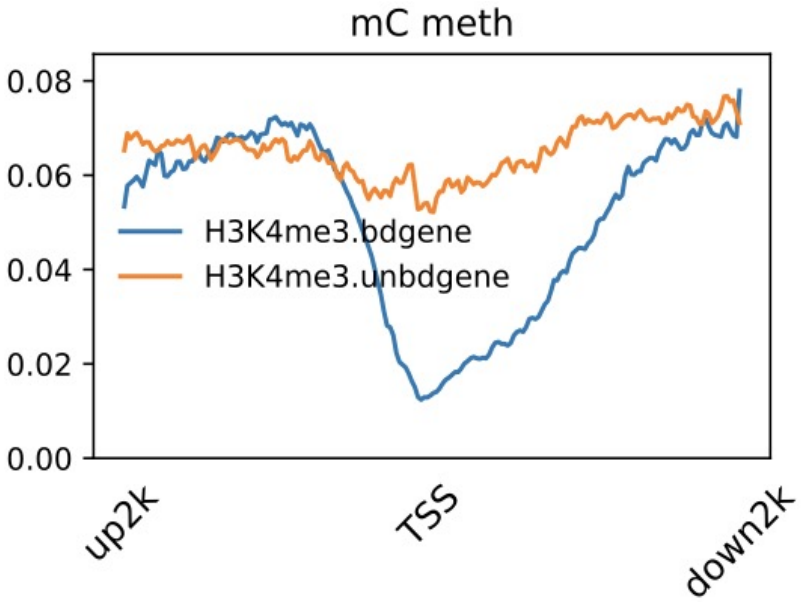
1.9.2 bt2profile

Plot DNA methlation profile across gene/ TE/ predefined bed region, such as peak or dmr region. The input DNA methylation level matrix is produced by *Calulate mC across predefined regions*.

The *.TSSprofile.txt *.centerprofile.txt and *.AverMethylevel.txt are calulated by *Calulate mC across predefined regions*.

```
$ BatMeth2 methyGff -o H3K4me3.bdgene H3K4me3.unbdgene \
    -G genome.fa -m methratio.txt \
    -b H3K4me3.bdgene.bed H3K4me3.unbdgene.bed -B

$ bt2profile.py -f H3K4me3.bdgene.TSSprofile.txt \
    H3K4me3.unbdgene.TSSprofile.txt \
    -l H3K4me3.bdgene H3K4me3.unbdgene \
    --outFileName H3K4me3.output.meth.pdf \
    -s 1 1 -xl up2k TSS down2k --context C
```



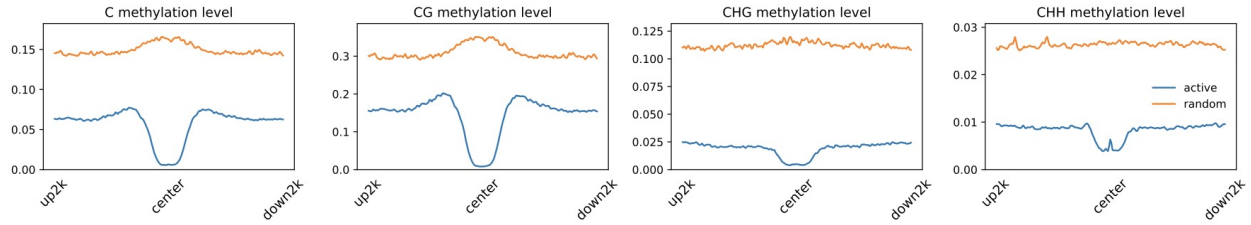

```
$ BatMeth2 methyGff -o active random \
    -G genome.fa -m methratio.txt \
    -b active.bed random.bed -B

$ bt2profile.py -f active.centerprofile.txt \
    random.centerprofile.txt \
```

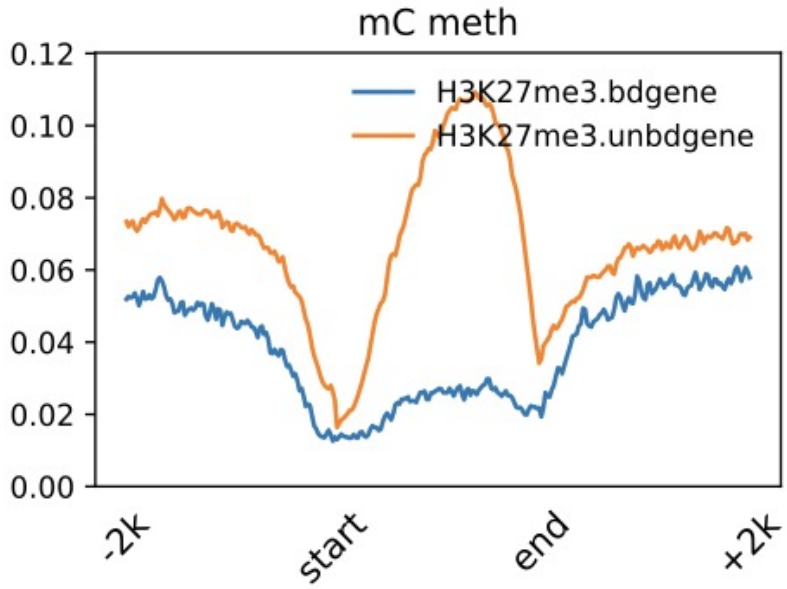
(continues on next page)

(continued from previous page)

```
    -l active random \
    --outFileName active_random.output.meth.pdf \
    -s 1 1 -xl up2k center down2k
```

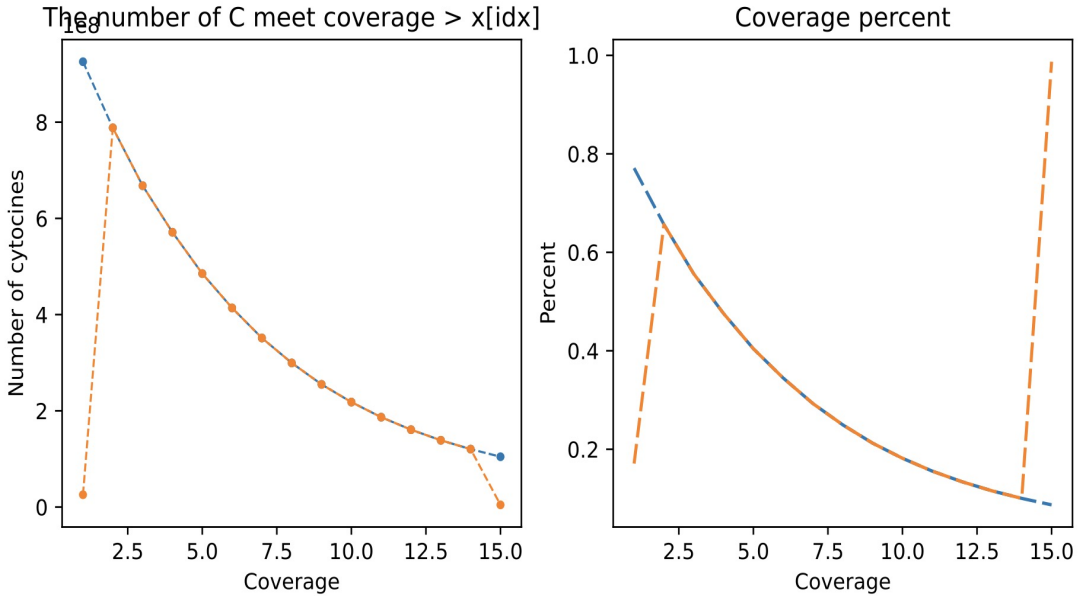



```
$ bt2profile.py -f H3K27me3.bdgene.AverMethylevel.txt \
    H3K27me3.unbdgene.AverMethylevel.txt \
    -l H3K27me3.bdgene H3K27me3.unbdgene \
    --outFileName H3K27me3.output.meth.pdf \
    -s 1 1 1 -xl up2k TSS TES down2k
```

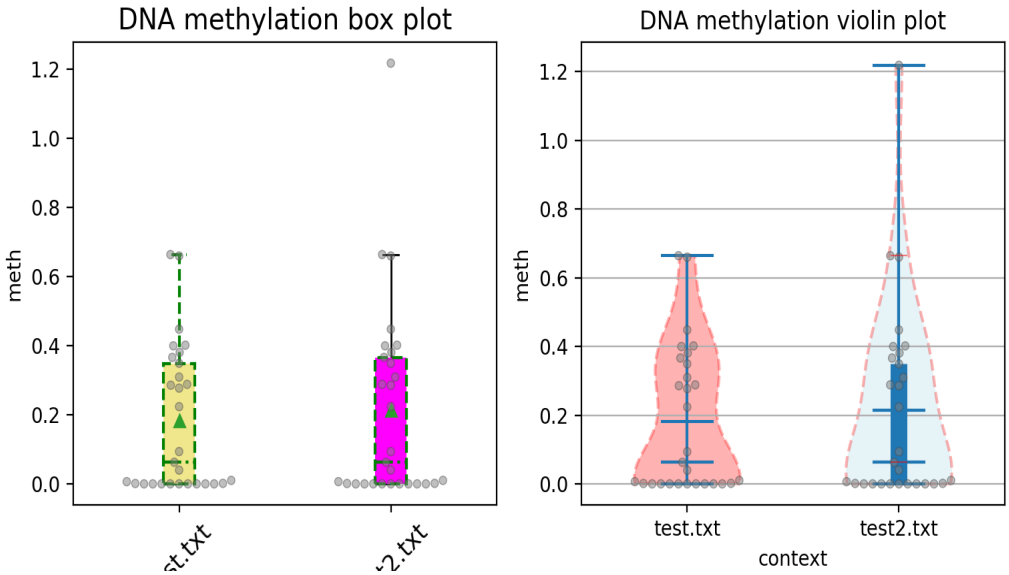


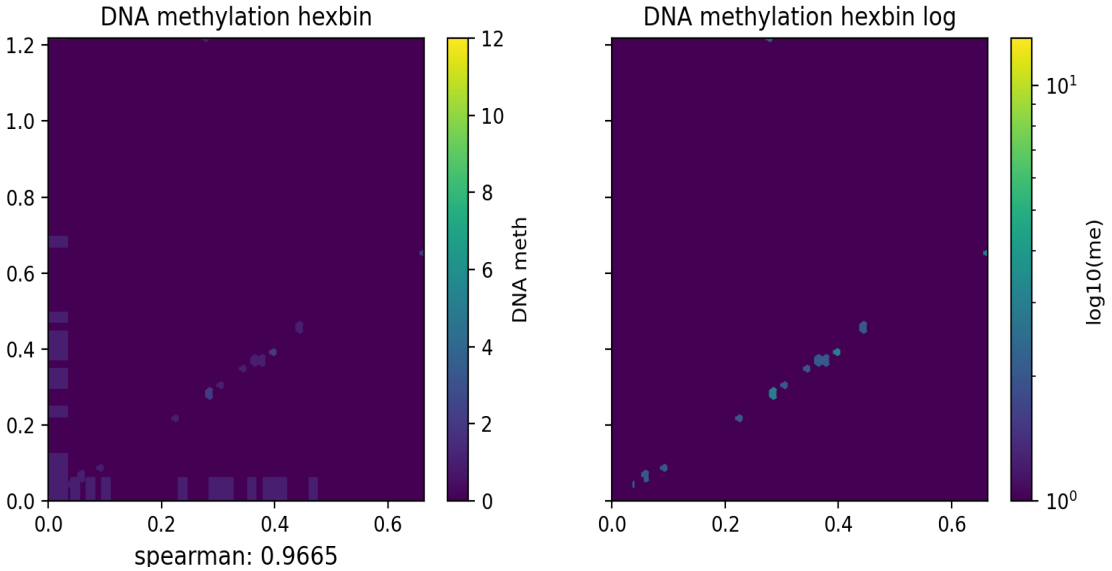
1.9.3 bt2basicplot

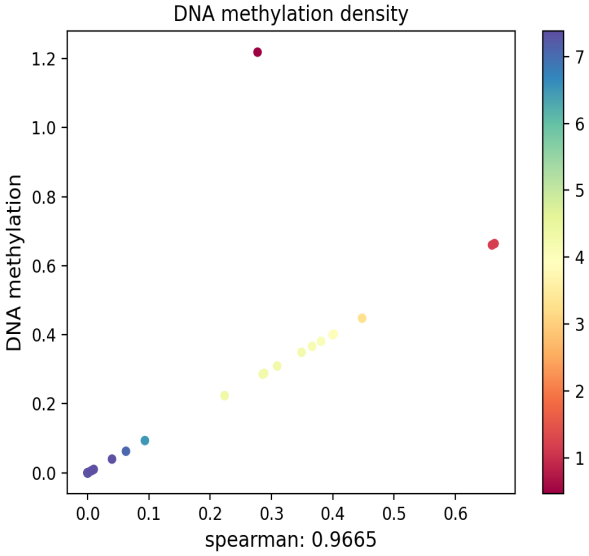
```
$ python3 bt2basicplot.py -c coverfile.txt coverfile2.txt -o tt.pdf
```

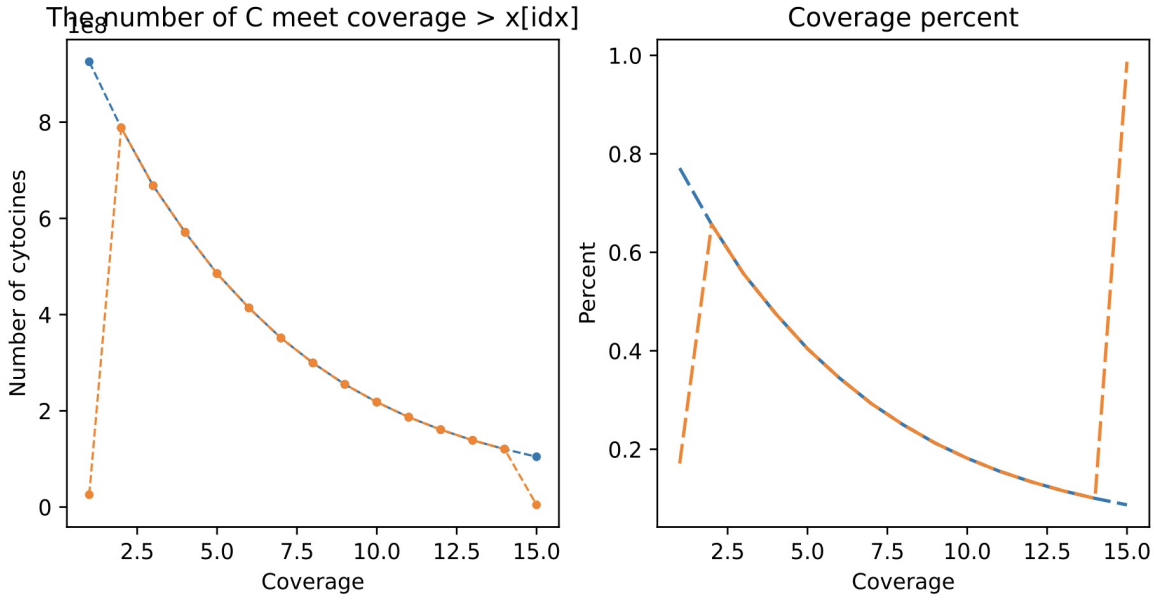



```
$ python3 bt2basicplot.py -f prefix1.gene.cg.txt prefix2.gene.cg.txt \
    -c coverfile.txt coverfile2.txt -o tt.pdf
```




DNA methylation hexbin
DNA methylation hexbin log
DNA meth
log10(me)
spearman: 0.9665

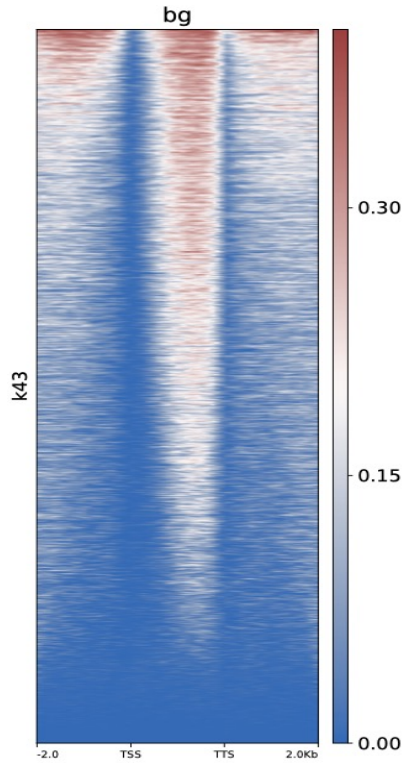

DNA methylation density
DNA methylation
spearman: 0.9665



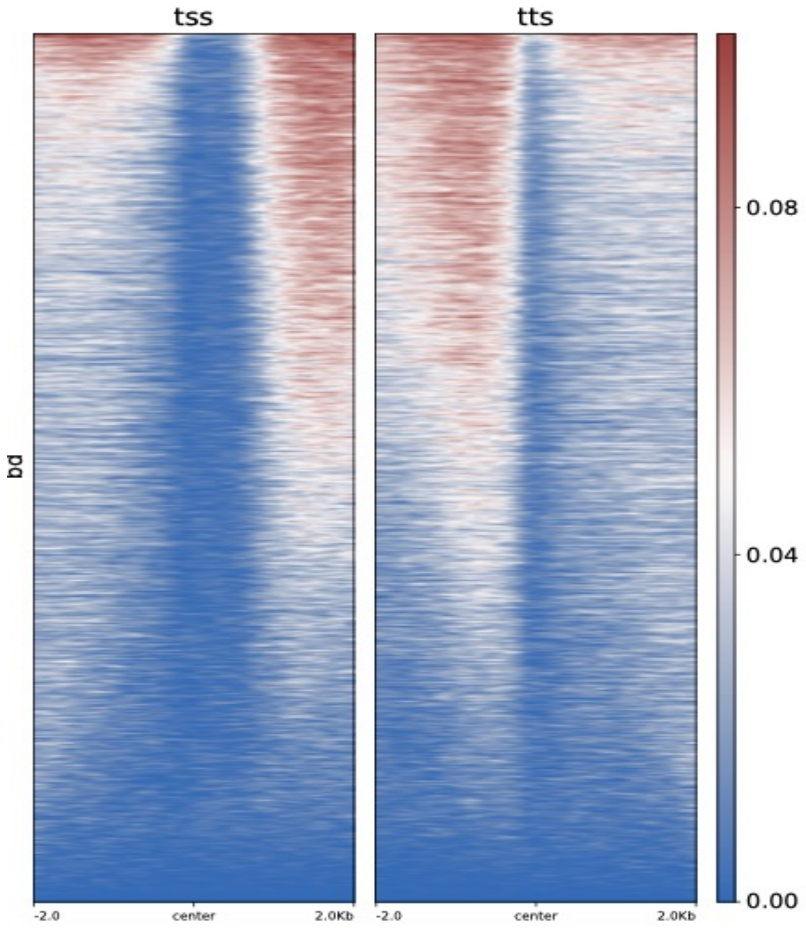
1.9.4 bt2chrprofile

1.9.5 bt2heatmap

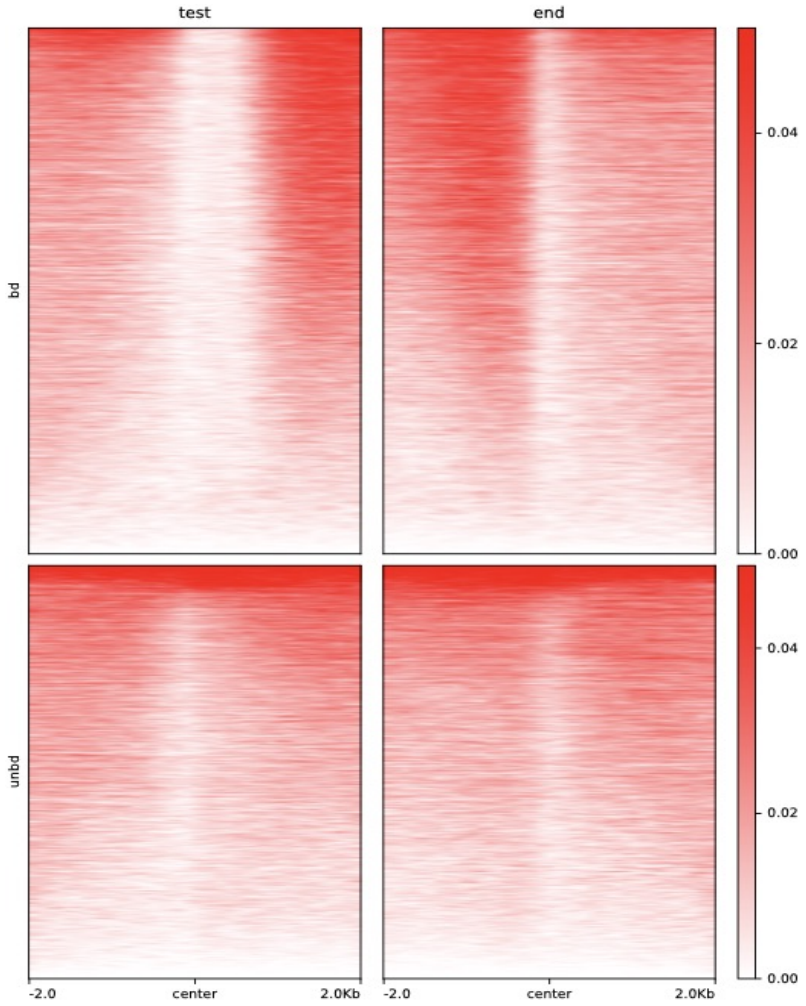
```
$ python bt2heatmap.py -m H3K4me3.bdgene.GENE.cg.txt -l bg \
-o test0.pdf -z k43 -sl TSS -el TTS
```



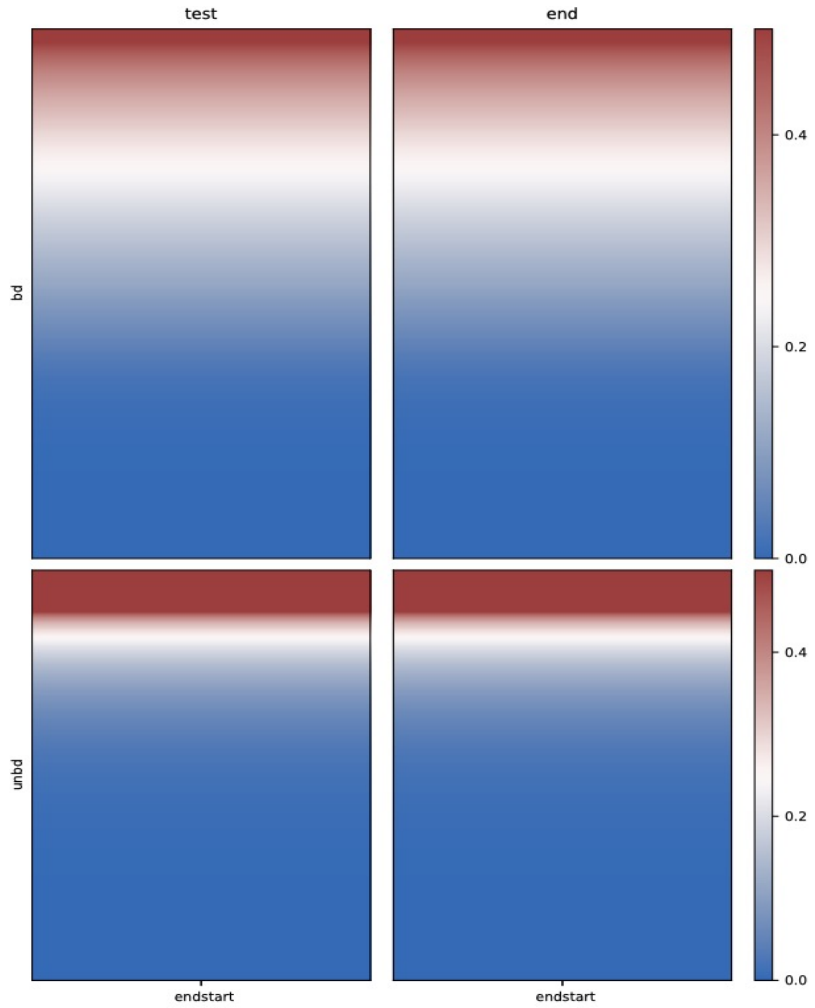
```
$ python bt2heatmap.py -m H3K4me3.bdgene.TSS.cg.txt H3K4me3.bdgene.TTS.cg.txt \
    -l tss tts -o test.pdf --zMax 0.1 --colorMap vlag --centerlabel center -z bd
```



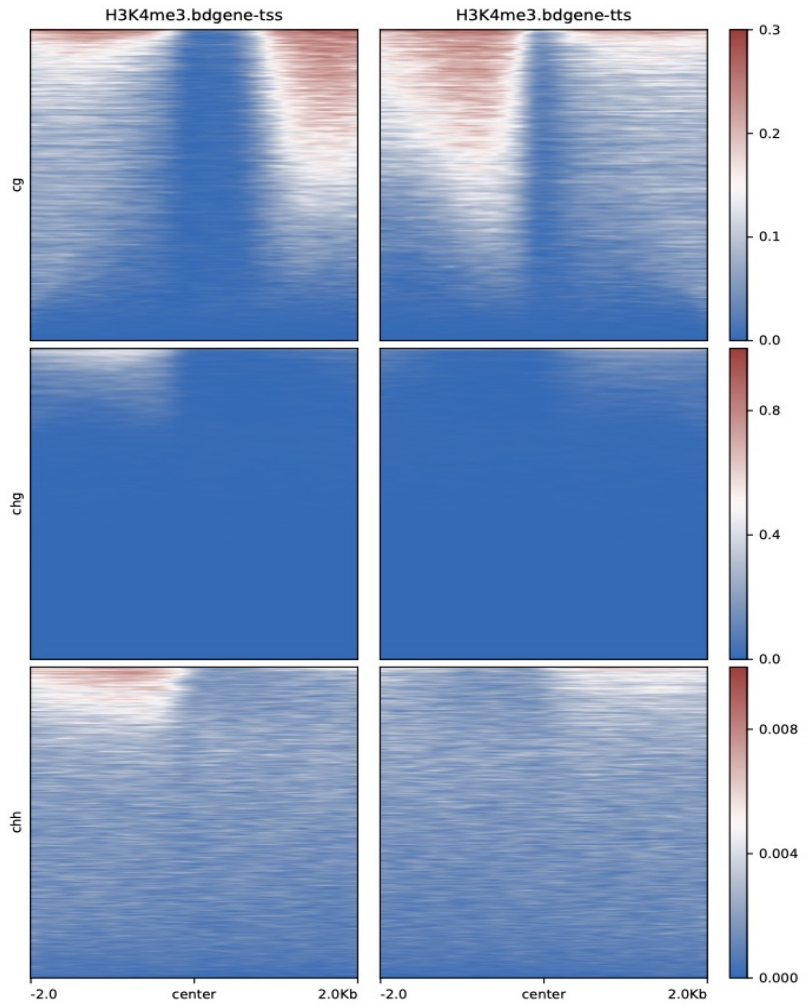

```
$ python bt2heatmap.py -m H3K4me3.bdgene.TSS.cg.txt H3K4me3.bdgene.TTS.cg.txt \
    H3K4me3.unbdgene.TSS.cg.txt H3K4me3.unbdgene.TTS.cg.txt \
    -l test end -o test2.pdf --zMax 0.05 --centerlabel center \
    --plotmatrix 2x2 --colorList white,red -z bd unbd
```




```
$ python bt2heatmap.py -f H3K4me3.bdgene.body.cg.txt H3K4me3.bdgene.body.cg.txt \
    H3K4me3.unbdgene.body.cg.txt H3K4me3.unbdgene.body.cg.txt \
    -l test end -o test3.pdf --zMax 0.5 --centerlabel center \
    --plotmatrix 2x2 -z bd unbd
```




```
$ python bt2heatmap.py -m H3K4me3.bdgene.TSS.cg.txt H3K4me3.bdgene.TTS.cg.txt \
    H3K4me3.bdgene.TSS.chg.txt H3K4me3.bdgene.TTS.chg.txt \
    H3K4me3.bdgene.TSS.chh.txt H3K4me3.bdgene.TTS.chh.txt \
    -l H3K4me3.bdgene-tss H3K4me3.bdgene-tts \
    -o H3K4me3.bdgene.TSS_TTS.heatmap.pdf --plotmatrix 3x2 \
    --centerlabel center -z cg chg chh --zMax 0.3 1 0.01
```



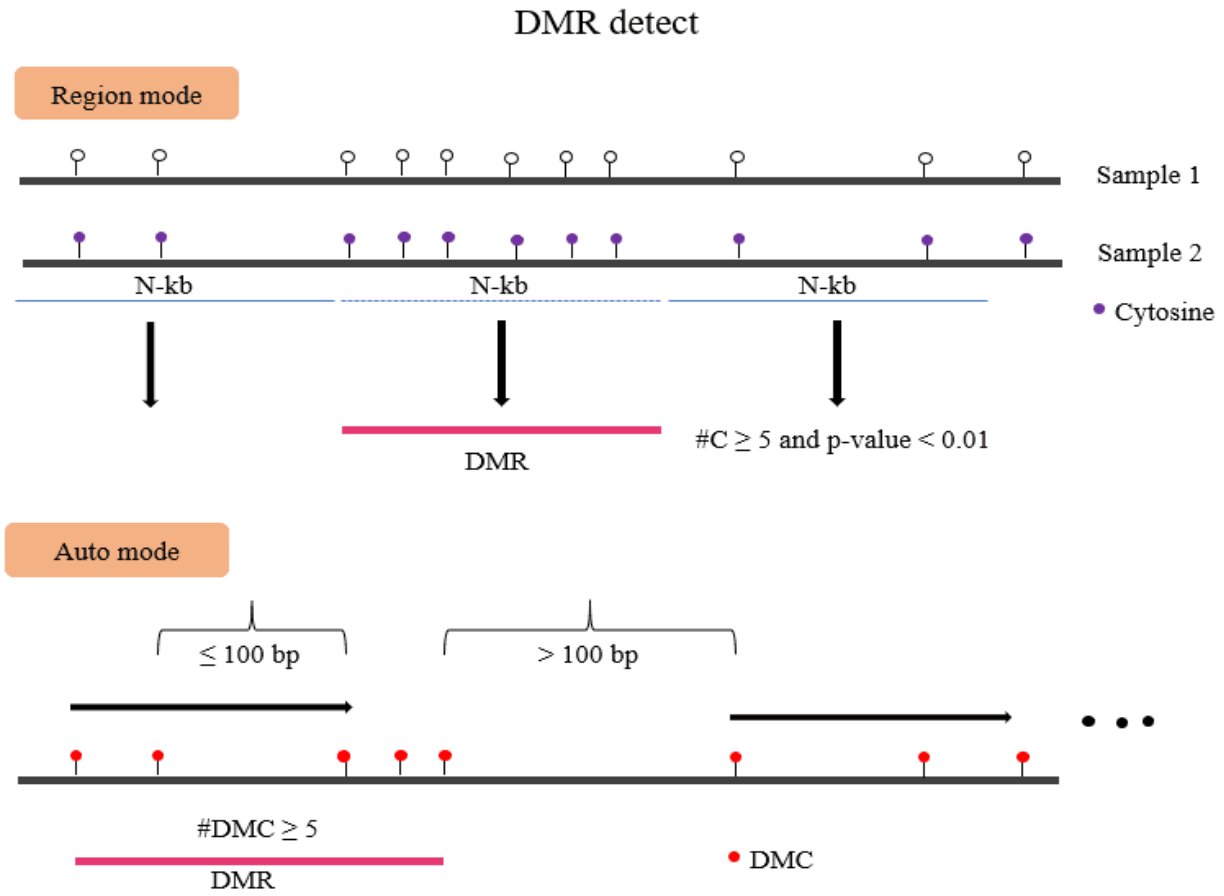
Tip: DNA methylation level distribution on chromosome (bt2chrplot) and DNA methylation level distribution (bt2visul) are currently being tested, and we will update them as soon as possible.

Note: @HZAU.

Tip: For feature requests or bug reports please open an issue on github.

1.10 DiffMeth

1.10.1 BatMeth2 DMC or DMR/DMG



You can get dmc and dmr result with:

```
$ batDMR -g genome.fa -o_dm mutant.output.dmc -o_dmr mutant.output.dmr \
-1 mutant.methratio.txt -2 WT.methratio.txt \
-methdiff 0.2 -minstep 100 -mindmc 5 -pval 0.01
```

obtained hyperhypo dmc/dmr from dmc/dmr results

```
$ awk -v OFS="\t" 'gsub(/\,/,"\t",$NF)' mutant.output.dmr | \
awk '$(NF-2)>4 && $NF<=1'  > mutant.output.hyper.dmr
$ awk -v OFS="\t" 'gsub(/\,/,"\t",$NF)' mutant.output.dmr | \
awk '!($(NF-2)>4 && $NF<=1)'  > mutant.output.hypo.dmr
$ awk '$NF>0' mutant.output.dmc | awk '{print $1"\t"$2"\t"$2}' \
> mutant.output.hyper.dmc
$ awk '$NF<0' mutant.output.dmc | awk '{print $1"\t"$2"\t"$2}' \
> mutant.output.hypo.dmc
```

1.10.2 Usage

[Main paramaters]	
-o_dm	output file
-o_dmr	dmr output file when use auto detect by dmc
-g|--genome	Genome files
-1	sample1 methy files, sperate by space.
-2	sample2 methy files, sperate by space.
-mindmc	min dmc sites in dmr region. [default : 4]
-minstep	min step in bp [default : 100]
-maxdis	max length of dmr [default : 0]
-pvalue	pvalue cutoff, default: 0.01
-FDR	adjust pvalue cutoff default : 1.0
-methdiff	the cutoff of methylation differention. default: 0.25 [CpG]
-element	caculate predefinded region, input file with id.
-context	Context for DM. [CG/CHG/CHH/ALL]
-L	predefinded regions or loci.
-gz	gzip input file.
-h|--help	

1. Pre-definded regions (Gene/TE/UTR/CDS or other regions)

```
BatMeth2 batDMR -g genome -L -o_dm dm.output.txt -1 [sample1.methC.txt replicates ..] \
-2 [sample2.methC.txt replicates ..]
```

2. Auto define DMR region according the dmc

```
BatMeth2 batDMR -g genome -o_dm dm.output.txt -o_dmr dmr.output.txt -1 [sample1.methC.
↪txt replicates ..] \
-2 [sample2.methC.txt replicates ..]
```

1.10.3 Output file

1. DMC

```
# format
Chrom position starnd context pvalue adjust_pvalue combine_pvalue corrected_pvalue \
cover_sample1 meth_sample1 cover_sample2 cover_sample2 meth.diff
```

2. DMR

```
# format
Chrom start end methlevelInSample1 methlevelInSample2 NdmcInRegion hypermdc,hypodmc
```

Tip: For feature requests or bug reports please open an issue on github.

1.11 Requirements

- *gsl library*
- *zlib library*
- *SAMtools*
- *fastp*

1.11.1 gsl library

The GSL library may need to be installed when the following problems occur during the installation process.

You can download here:

- fatal error: gsl/gsl_matrix_double.h : No such file or directory

`gsl-2.4.tar.gz`

```
./configure --prefix=/disk1/glli/tools/gsl-2.4/
make
make install
```

Add environment variables to ~/.bashrc

```
export C_INCLUDE_PATH=$C_INCLUDE_PATH:~/software/gsl-2.4/include
export CPLUS_INCLUDE_PATH=$CPLUS_INCLUDE_PATH:~/software/gsl-2.4/include
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH::~/software/gsl-2.4/lib
export LIBRARY_PATH=$LIBRARY_PATH::~/software/gsl-2.4/lib
```

And then:

```
$ source ~/.bash
```

1.11.2 zlib library

The GSL library may need to be installed when the following problems occur during the installation process.

- unfound zlib.h

`zlib-1.2.11.zip`

```
./configure --prefix=/disk1/glli/tools/zlib-1.2.11/
make
make install
```

Add environment variables to ~/.bashrc

```
export C_INCLUDE_PATH=$C_INCLUDE_PATH:/disk1/glli/tools/zlib-1.2.11/include
export CPLUS_INCLUDE_PATH=$CPLUS_INCLUDE_PATH:/disk1/glli/tools/zlib-1.2.11/include
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH::/disk1/glli/tools/zlib-1.2.11/lib
export LIBRARY_PATH=$LIBRARY_PATH::/disk1/glli/tools/zlib-1.2.11/lib
```

And then:

```
$ source ~/.bash
```

1.11.3 SAMtools

1.11.4 fastp

fastp, raw reads as input need.

Tip: For feature requests or bug reports please open an issue on github.

While developing BatMeth2, we continuously strive to create software that fulfills the following criteria:

- raw fastq reads quality control and **efficiently align bisulfite sequencing data**
- **calculate DNA methylation level based on sorted BAM file** for single base or chromosome region and genes.
- **new methlation mbw format with index** can calculate DNA methylation level quickly.
- enable **customized down-stream analyses**, espacially with visulization
- generation of **highly customizable images** (change colours, size, labels, file format, etc.)

1.12 Citation

Please cite BatMeth2 as follows:

Zhou Q, Lim J-Q, Sung W-K, Li G: An integrated package for bisulfite DNA methylation data analysis with Indel-sensitive mapping. BMC Bioinformatics 2019, 20:47. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2593-4

Tip: For feature requests or bug reports please open an issue on github.