Methpat Manual

Nicholas C. Wong and Bernard J. Pope
Version 0.2

July 9, 2014

Contents

1 Requirements

- Whole genome bisulfite short read sequencing data (WGBS), or Reduced Representation Bisulfite Sequencing (RRBS), or Bisulfite Amplicon Sequencing
- Bismark v0.12.2 and it's dependencies (bwa v0.7.5a-r405)
- Perl 2.7 or above and pip
- Methpat and it's dependencies https://github.com/bjpop/methpat
- For visualisation of html files, current versions of Firefox or Chrome

2 Installation instructions

Goto GitHub site for further detail and instructions. https://github.com/bjpop/methpat
or
Download the current master from git hub with the following command as Administrator (or Root).

```
pip install git+https://github.com/bjpop/methpat.git
```

3 Usage

Methpat was designed to take output files from Bismark and summarise DNA methylation patterns into a text file format that is amenable to downstream statistical analysis. The premise has been that the allelic DNA methylation patterns are as important biologically as the level of DNA methylation measured.

```
usage: methpat [-h] [--dump_reads FILE] [--count_thresh THRESH] --amplicons
               AMPLICONS_FILE --logfile FILENAME --html FILENAME
               [--webassets {package,local,online}]
               BISMARK_FILE

positional arguments:
BISMARK_FILE          input bismark file

optional arguments:
  -h, --help            show this help message and exit
  --dump_reads FILE     dump the read methylation information to FILE
  --count_thresh THRESH only display methylation patterns with at least THRESH
                        number of matching reads
```

```
--amplicons AMPLICONS_FILE filename of tab-delimited BED format text file containing
 amplicon/Region of Interest coordinates
--logfile FILENAME     log progress in FILENAME
--html FILENAME        save visualisation in html FILENAME
--webassets {package, local, online} define location of D3.js libraries.
```

Package defines the required library files are in a common directory and related to Methpat. Local defines the required library files are manually placed into the same directory as the html file. Online defines the required library files are found on the internet (requires internet connection).

3.1 Amplicons File Definintion

The amplicons or regions of interest are defined by this tab-delimited text file. The order of regions listed in this file will correspond to the order of display in the html file generated. The file requires 7 columns which define:

1. chromosome
2. start coordinate of amplicon/region of interest
3. end coordinate of amplicon/region of interest
4. name of amplicon/region of interest
5. size of amplicon/region of interest
6. length of forward PCR primer of amplicon/region of interest, number of bp to ignore from the start
7. length of reverse PCR primer of amplicon/region of interest, number of bp to ignore from the end

See DemoData for a working example.

4 Worked Example

Run Bismark as per instructions to prepare the reference genome as outlined in bismark manual.

Run Bismark to align raw read files (fastq) as below with no mismatches:

```
bismark -n 0 -q -o <nameOfOutPutDirectory> --bam --ambiguous
--chunkmbs 10204 --non_directional /path/to/reference/genome <fastq.file>
```

Run bismark_methylation_extractor to create an output text file containing read DNA methylation pattern information:

```
bismark_methylation_extractor -s <bismarkAlignedFileName.bam>
```

This creates a series of text files prefixed with the methylation orientation and reference genome strand data was extracted from. For the purposes here, the files prefixed with CpG are then processed with methpat.

In the directory containing the bismark output files run:

```
for i in `ls CpG*`
do j=`echo ${i} | sed 's;CpG\_OT\_;;g' | sed 's;_L001_R1_001fastq_bismark.txt;;g'`
methpat --dump_reads ${j}.dump --count_thresh 1 --amplions mitoAmplicons.txt
--logfile ${j}.log --html ${j}.html --webassets "local" ${i} > ${j}.methpat.txt
done
```

This generates a methpat.txt file which summarises the read count and DNA methylation pattern information for all amplicons/regions of interest listed in the amplicons file. A html file is also generated for visualisation in Firefox or Chrome browsers.

5 Visualisation

Once open in a browser (see Figure 1), there are several options that could be changed for the visualisation of the data.

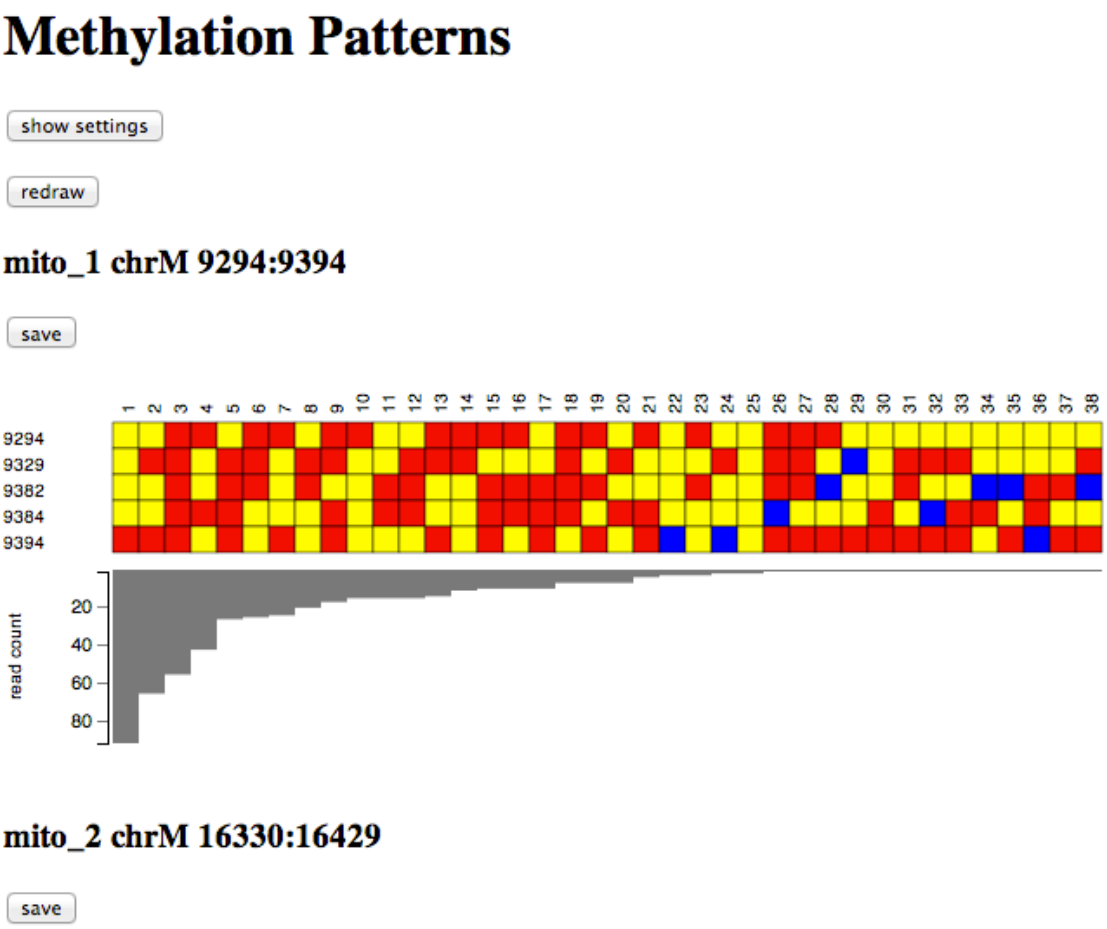

Figure 1: Methpat splash page

5.1 Settings for Visualisation

Various settings can be changed by clicking on the show settings button (Figure 1). By default methpat displays each amplicon with the following defaults:

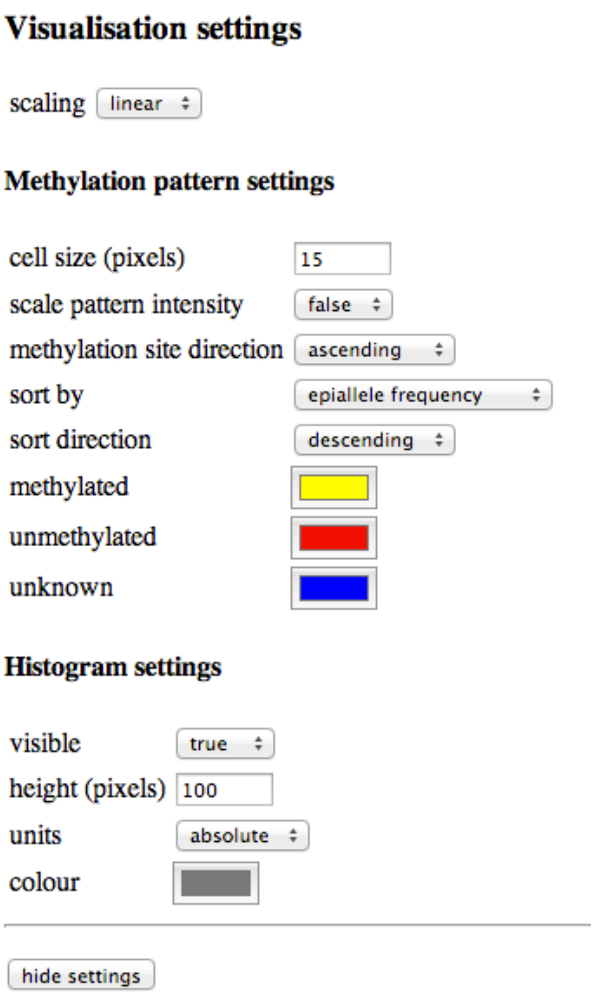

Figure 2: Methpat settings page

- Scaling, Linear
- cell size, 15px
- scale pattern intensity, false
- methylation site direction, ascending
- sort by, epiallele
- sort direction, descending
- methylated colour, yellow

- unmethylated colour, red
- unknown colour, blue
- histogram visibility, true
- histogram height, 100px
- histogram units, absolute
- histogram colour, grey

scaling can be set either as linear or logarithmic (Figure 2).

5.1.1 cell size (pixels)

Define the size of the cells depicting CpG sites.

5.1.2 scale pattern intensity

true or false, add alpha layer over the cells to represent read count abundance.

5.1.3 methylation site direction

ascending or descending, not sure what this is meant to do, only changes plots when sort by is changed to degree of methylation.

5.1.4 sort by

epiallele frequency or degree of methylation.

Epiallele or read frequency for each DNA methylation pattern. Or sort by the degree of DNA methylation across each pattern.

5.1.5 sort direction

ascending or descending.

5.1.6 methylated

Colour selection, in HEX code for Firefox.

5.1.7 unmethylated

Colour selection, in HEX code for Firefox.

5.1.8 unknown

Colour selection, in HEX code for Firefox.

5.1.9 visible

true or false, display histogram or not.

5.1.10 height (pixels)

height of histogram.

5.1.11 units

absolute or percent.
    absolute read count or percentage of total reads.

5.1.12 colour

Colour selection, in HEX code for Firefox.