Chapter 17

Approaches for the Analysis and Interpretation of Whole Genome Bisulfite Sequencing Data

Tim Stuart, Sam Buckberry, and Ryan Lister

Abstract

DNA methylation is a covalent modification of DNA that plays important roles in processes such as the regulation of gene expression, transcription factor binding, and suppression of transposable elements. The use of whole genome bisulfite sequencing (WGBS) enables the genome-wide identification and quantification of DNA methylation patterns at single-base resolution and is the gold standard for analysis of DNA methylation. Computational analysis of WGBS data can be particularly challenging, as many computationally intensive steps are required. Here, we outline a step-by-step approach for the analysis and interpretation of WGBS data. First, sequencing reads must be trimmed, quality checked, and aligned to the genome. Second, DNA methylation levels are estimated at each cytosine position using the aligned sequence reads of the bisulfite treated DNA. Third, regions of differential cytosine methylation between samples can be identified. Finally, these data need to be visualized and interpreted in the context of the biological question at hand.

Key words DNA methylation, Whole genome bisulfite sequencing, Bioinformatics, Genomics

1 Introduction

The methylation of cytosine bases in DNA (DNA methylation) is a covalent DNA modification able to be faithfully maintained across genome replication and cell divisions and provides a nongenetic mechanism for the transmission of information over generations. While DNA methylation is a stable epigenetic modification, its deposition throughout the genome can also be dynamic, changing through development and in response to environmental stimuli [1]. For this reason, the analysis of genome-wide patterns of DNA methylation is of great interest for studies aiming to understand the molecular basis underlying developmental pathways and gene regulation.

Measurement of DNA methylation in the genome can be achieved in a number of ways. A global average level of DNA methylation can be estimated through liquid chromatography-mass spectrometry after hydrolysis of the DNA to mononucleotides or

Albert Jeltsch and Marianne G. Rots (eds.), *Epigenome Editing: Methods and Protocols*, Methods in Molecular Biology, vol. 1767, https://doi.org/10.1007/978-1-4939-7774-1_17,

mononucleosides, but this provides no contextual information as to where DNA methylation marks are located in the genome [2]. Antibodies, or other proteins that can bind 5-methylcytosine, have been used successfully to pull down methylated fragments of DNA, which can then be sequenced and mapped back to a reference genome [3]. However, such methods are unable to identify which cytosines in a pulled-down DNA fragment are methylated and so are unable to capture the DNA sequence context surrounding the methylated base. As the cytosine base context surrounding the methylated site has been shown to be highly important, with different molecular machinery used for the deposition and maintenance of DNA methylation at different cytosine base contexts [4], this inability of pulldown-based methods to detect the DNA methylation sequence context is a major limitation.

The treatment of DNA with sodium bisulfite deaminates cytosine bases to uracil, but does not deaminate 5-methylcytosine, providing a mechanism for discrimination between the two at single-base resolution [5]. By following sodium bisulfite treatment with PCR, uracil bases are replaced with thymine by the DNA polymerase. One common method for measuring DNA methylation after bisulfite conversion is with microarrays, which involve hybridizing sodium bisulfite-treated DNA fragments to probes to quantify methylation levels at specific loci throughout the genome. While these arrays can cover hundreds of thousands of cytosines, they only capture information on a small percentage of cytosines genome-wide and are typically only available for mouse and human genomes.

A more comprehensive approach is the sequencing of sodium bisulfite-treated and PCR-amplified DNA fragments. Once sequenced, the DNA methylation pattern of the original DNA molecule can be inferred by the presence of thymine base calls at cytosine positions. Whole genome bisulfite sequencing (WGBS) allows the most comprehensive analysis of DNA methylation patterns genome-wide and is the current gold standard for DNA methylation analysis. For more targeted studies, amplicon-based methods that use PCR to amplify the number of reads from a particular region can be used instead.

Since the first use of WGBS for measuring genome-wide patterns of DNA methylation [6, 7], many computational tools have been developed facilitating the analysis of WGBS data. Generally, the analysis of WGBS data involves several main steps. First, sequencing reads must be preprocessed. Second, reads are mapped to a reference genome, allowing for differences between the reads and reference sequence due to bisulfite conversion. This can be done by aligning to a three-base genome, where all cytosines are replaced with thymines. Third, DNA methylation levels genome-wide must be quantified based on the reads mapped to each cytosine base. Finally, further analysis specific to the biological question

of interest must be performed, typically involving the identification of regions of differential DNA methylation between samples or regions of the genome. These steps are outlined in detail below.

2 Materials

2.1 Software

The programs listed in Table 1 are all freely available and required for the analysis of WGBS data [8–16].

The programs listed in Table 2 are all freely available and are recommended for the analysis of WGBS data.

2.2 Hardware

A Linux computer or cloud computing resource such as Amazon Web Services or Google Genomics is required.

2.3 Data

1. Genome sequence and annotation: https://support.illumina.com/sequencing/sequencing_software/igenome.html, http://hgdownload.cse.ucsc.edu/downloads.html.
2. Lambda phage genome sequence: https://www.ncbi.nlm.nih.gov/nuccore/215104?report=fasta.
3. Published datasets relevant to your experiment: https://www.ncbi.nlm.nih.gov/geo/, https://www.ncbi.nlm.nih.gov/sra.

3 Methods

3.1 Preparation

3.3.1 Install Software

Install the required software listed in Table 1, and optionally install the software listed in Table 2. Instructions for the installation and testing of this software can be found at the website for each program.

3.1.1 Trim Reads

Adapter bases and low-quality bases should be removed from reads before alignment. To correctly trim the adapter sequence, it is essential to know the sequence of the adapters used in the experiment. Currently, the most common adapter is the Illumina TruSeq adapter with the following sequence: AGATCGGAAGAGCAC ACGTCTGAACTCCAGTCAC.

If you are unsure of the correct adapter sequence, running FastQC (*see* **step 5**) can often correctly identify the adapter sequence used.

To trim adapters and low-quality bases from the reads, use cutadapt (*see* **Note 1**).

For single-end reads:

```
cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC \
         -m 50 -q 10,10 reads.fastq  \
         > trimmed_reads.fastq
```

Table 1
Essential software for WGBS data analysis

Name	Link
Python	https://www.python.org/download/releases/2.7.2/
Java	http://www.oracle.com/technetwork/java/javase/downloads/index.html
Bcl2fastq	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
FastQC [8]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Cutadapt [9]	http://cutadapt.readthedocs.io/en/stable/guide.html
Seqtk	https://github.com/lh3/seqtk
Bowtie2 [10]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Picard	https://broadinstitute.github.io/picard/
SAMtools [11]	http://www.htslib.org/
BSseeker2 [12]	http://pellegrini.mcdb.ucla.edu/BS_Seeker2/
Pysam	http://pysam.readthedocs.io/en/latest/
DSS [13]	http://bioconductor.org/packages/release/bioc/html/DSS.html
GNU coreutils	https://www.gnu.org/software/coreutils/coreutils.html
R	https://www.r-project.org/

Table 2
Optional software for WGBS data analysis

Name	Link
Virtualenv	https://virtualenv.pypa.io/en/stable/
Virtualenvwrapper	https://virtualenvwrapper.readthedocs.io/en/latest/
IGV	http://software.broadinstitute.org/software/igv/
BEDTools [14]	http://bedtools.readthedocs.io/en/latest/
deepTools [15]	http://deeptools.readthedocs.io/en/latest/
ViewBS	https://github.com/xie186/ViewBS
SRA toolkit	https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/
Docker	https://www.docker.com/
MultiQC [16]	http://multiqc.info/
pigz	https://zlib.net/pigz/

For paired-end reads:

```
cutadapt -a AGATCGGAAGA -A CTCTTCCGATC -q 10,10 \
         -o trimmed_read1.fastq -p trimmed_read2.fastq \
         read1.fastq read2.fastq
```

3.1.2 Compress Reads

As DNA sequencing data can be very large, FASTQ files should be gzip compressed at this stage to save disk space (*see* **Note 2**).

For single-end reads:

```
gzip trimmed_reads.fastq
```

For paired-end reads:

```
gzip trimmed_read1.fastq trimmed_read2.fastq
```

3.1.3 Sequencing Quality Control

To assess the quality of your sequencing reads, run FastQC on each FASTQ file. This will generate a html report that summarizes each quality check performed.

For single-end reads:

```
fastqc trimmed_reads.fastq.gz
```

For paired-end reads:

```
fastqc trimmed_read1.fastq.gz trimmed_read2.fastq.ga
```

Inspect the html report produced by running FastQC. The most important aspects of this report are the base quality scores and overrepresented sequences. Examples of good and bad FastQC reports can be found at the FastQC website: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Overrepresented sequences may indicate inadequate read trimming. Poor quality scores indicate a poor sequencing run which may be due to numerous reasons.

3.1.4 Genome Preparation

Download the required reference genome for the organism used in the experiment (*see* **Note 3**).

If lambda DNA spike-in was used prior to bisulfite conversion, add the lambda genome as an additional chromosome to the genome by concatenating the FASTA files.

```
cat genome.fa lambda.fa > genome_lambda.fa
```

3.1.5 Build the Genome Index

For read alignment, we recommend using BS-Seeker2 [12]. To build the genome index, you simply need to supply the genome FASTA file and the aligner you plan to use. We recommend using Bowtie2 [10].

```
python bs_seeker2-build.py -f genome_lambda.fa
--aligner=bowtie2
```

3.2 Alignment

3.2.1 Align Reads

Next, reads must be aligned to the reference genome (*see* **Note 4**).

For single-end reads:

```
python bs_seeker2-align.py -g genome_lambda.fa \
                           --aligner=bowtie2 \
                           -u unmapped.fa \
                           -o mapped.bam \
                           --bt2-p <number of cores> \
                           -i trimmed_reads.fastq.gz
```

For paired-end reads:

```
python bs_seeker2-align.py -g genome_lambda.fa \
                           --aligner=bowtie2 \
                           -u unmapped.fa \
                           -o mapped.bam \
                           --bt2-p <number of cores> \
                           -1 trimmed_read1.fastq.gz \
                           -2 trimmed_read2.fastq.gz
```

3.2.2 Sort BAM File

After alignment, the mapped reads should be sorted by position (*see* **Note 5**):

```
samtools sort -@ <number of cores> -T temp -O bam
mapped.bam sorted
```

3.2.3 Remove PCR Duplicates

PCR duplicates can be computationally identified following alignment to the reference genome with relatively high accuracy for genome coverages typical for WGBS experiments. Use the MarkDuplicates tools in picard to identify and remove PCR duplicates from the aligned reads:

```
java -jar picard.jar MarkDuplicates I=sorted.bam \
                                    O=filtered.bam \
                                    M=duplicate_stats.txt \
                                    REMOVE_DUPLICATES=true \
                                    AS=true
```

3.2.4 Remove Intermediate Files

At this stage, the intermediate files should be removed to avoid unnecessary use of disk space:

```
rm mapped.bam sorted.bam
```

3.3 Quantifying DNA Methylation

3.3.1 Call DNA Methylation

Next, call DNA methylation (*see* **Note 6**):

```
python bs_seeker2-call_methylation.py -i filtered.bam --sorted \
                                      -o <sample name> \
                                      --db <path to genome db>
```

3.4 Post-Alignment Quality Control

3.4.1 Assess Methylation Bias in Read Position

DNA methylation levels should be unrelated to the position of the methylated cytosine in a sequencing read. Any methylation bias along the length of a read therefore indicates inadequate adapter trimming prior to alignment. An important quality check is to assess this methylation bias in the mapped reads. As the methylation

information is stored under the XM read tag by BS-Seeker2, the methylation bias along reads can easily be assessed (*see* **Note 7**).

3.4.2 Assess Alignment Statistics

The percentage of mapped reads and multi-mapped reads are reported by BS-Seeker2 and should be recorded. A low percentage of mapped reads, or high percentage of multi-mapped reads, may indicate a problem (*see* **Note 8**).

The percentage of duplicate reads removed by picard is also reported in the duplicate removal step (stored in the duplicate_stats.txt file from **step 7**) and should be recorded. A high percentage of PCR duplicates may indicate that too many PCR cycles were used, and alterations to the wet-lab protocol may be needed in future.

3.4.3 Assess Non-conversion Rate

The bisulfite non-conversion rate can be estimated for each cytosine base context by counting the number of C base reads mapped to a known unmethylated sequence, typically the lambda phage genome if it was spiked into the library prior to bisulfite conversion. As the lambda genome is totally unmethylated, each C read indicates a non-converted base. By calculating the total percentage of lambda cytosines with C reads, this gives an estimate of the bisulfite non-conversion rate. A high non-conversion rate (>2%) indicates that the bisulfite conversion performed poorly and new data may need to be collected.

3.4.4 Assess Genome-Wide Methylation Average in each Context

The genome-wide average methylation level is reported by BS-Seeker2 (see the BS-Seeker2 log file) and should be similar to what is expected for the sequenced organism or close relatives.

3.5 Differential DNA Methylation

When comparing multiple samples, as is almost always done in WGBS experiments, the first step should be the identification of cytosines that are differentially methylated between the samples. We recommend using DSS for differential methylation detection in WGBS experiments [13, 17] (*see* **Note 9**).

3.5.1 Preprocessing DNA Methylation Data

Input to DSS needs to be a text file in the format:

```
<chromosome><position><methylated_count>
<unmethylated_count>
```

If you wish to detect differential DNA methylation in one cytosine context in particular (e.g., the CG context), extract CG positions from the BS-Seeker2 CGmap file (*see* **Note 10**). If you wish to find differential DNA methylation in all cytosine contexts, the CGmap file should be split into three files: CG, CHG, and CHH. To extract only the CG positions from the BS-Seeker2 CGmap file, run:

```
awk 'BEGIN {FS=OFS="\t"} {if ($4 == "CG") print $1,
$3, $7, $8-$7}' sample.CGmap > cg_positions.tsv
```

3.5.2 Find Differentially Methylated Positions

Use DSS to identify differentially methylated positions. Within R, run the following commands:

```
> library(DSS)
> column_names <- c("chr", "pos", "N", "X")
> condition1 <- read.table('condition1_cg.tsv',
header=column_names)
> condition2 <- read.table('condition2_cg.tsv',
header=column_names)
> experiment <- makeBSseqData(list(condition1,
condition2),
+ c('C1', 'C1'))
> dmlTest <- DMLtest(experiment, group1=c('C1'),
group2=c('C2'))
```

The `dmlTest` object contains *p*-values for each cytosine position and can be filtered to give a set of significantly differentially methylated positions:

```
> dml <- callDML(dmlTest, delta=0.1,
p.threshold=0.001)
```

The parameters `delta` and `p.threshold` can be altered to change the significance cutoff, altering the percentage change in methylation and *p*-value cutoff, respectively.

3.5.3 Find Differentially Methylated Regions

Following from the above R code, the `dmlTest` object can be used to detect larger regions of differential methylation:

```
> dmrs <- callDMR(dmlTest, delta=0.1, minCG=3
+             p.threshold=0.001, minlen=50,
+             dist.merge=100)
```

Many different arguments can be passed to the callDMR function in order to change the minimum requirements for a DMR to be identified. The argument `delta` can be altered to change the minimum change in methylation required to call a DMR, and `minCG` can be altered to change the minimum number of CG sites allowed in a DMR. The p-value cutoff can be set using the `p.threshold` argument. The minimum DMR size reported can be controlled using the `minlen` and `dist.merge` arguments.

These DMR positions can be saved as a BED (Browser Extensible Data) format file, for later analysis steps.

```
> write.table(dmrs, "cg_dmrs.bed", sep="\t",
+ row.names=FALSE, quote=FALSE)
```

3.6 Interpretation

3.6.1 Data Visualization

An important step when analyzing any genomic data is data visualization. Genome browsers are well suited to this task, as different data types can be stacked by the addition of different browser tracks, such as gene annotations, DMR positions, ChIP-seq data, and RNA-seq data. We recommend using the IGV or UCSC genome browser to visualize DNA methylation (Fig. 1).

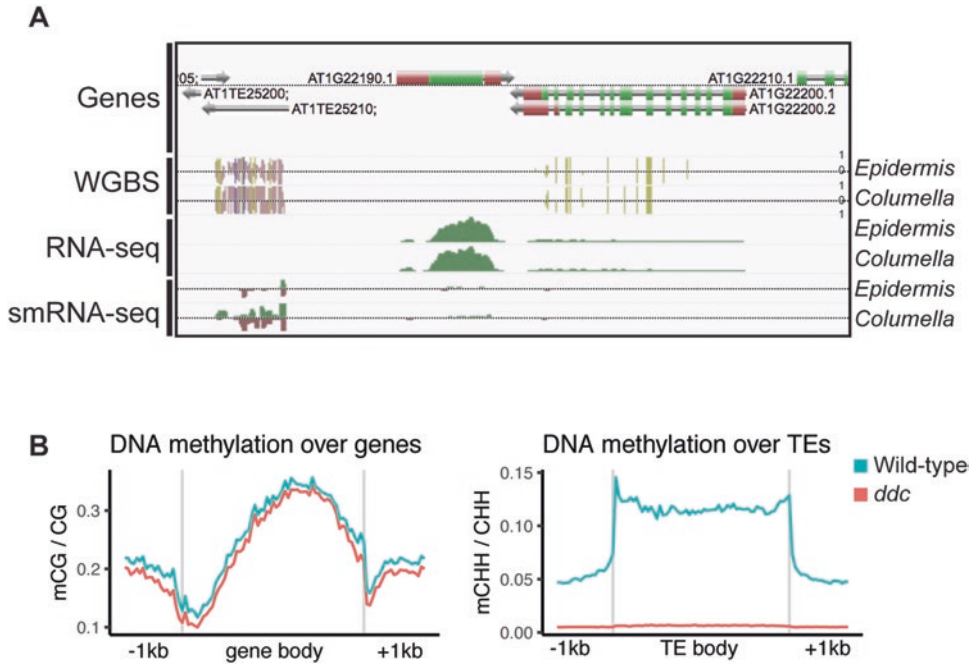

Fig. 1 Example data visualizations for WGBS data. (**a**) A genome browser view showing WGBS, RNA-seq, and small RNA-seq data for two samples side by side. Data is from [19]. (**b**) DNA methylation meta-plots for genes and TEs, showing a comparison between two samples. Plots were prepared using ViewBS (https://github.com/xie186/ViewBS) with data from [20]

To view WGBS data in IGV, load the wig file produced by BS-Seeker2 using the IGV graphical interface. It is also useful to view the sequencing depth of coverage alongside the measured DNA methylation levels. This is done automatically by IGV when you load the bam file. However, you can also create a separate coverage track as a bigwig file using the bamCoverage command in deepTools [15]:

```
bamCoverage -b filtered.bam -o coverage.bw
```

3.6.2 *Interpreting DMRs*

The interpretation of DMRs depends strongly on the biological question at hand, although some general rules can be applied that will be applicable in most cases (*see* **Note 11**). Firstly, the genomic context of the DMR will almost always be of interest. To find which genomic features, such as genes or promoters, are close to identified DMRs, the BEDTools command-line toolset can be used [14]. The GenomicRanges R package provides similar functionality within R [18].

Beyond the simple intersection of DMRs with annotated genomic features, it is generally helpful to integrate other types of data wherever possible. Of particular value is RNA-seq data, ChIP-seq for histone modifications or transcription factors, and

chromatin accessibility data such as ATAC-seq. This can be achieved through additional experiments or by using public datasets made available by previous publications.

4 Notes

1. If your reads come from an Illumina sequencer that used two-color chemistry (the NextSeq or NovaSeq), you will need to specify a different option for quality trimming. Instead of using `-q 10,10`, use `--nextseq-trim 10,10`.
2. Pigz (parallel implementation of gzip) can be used in place of the program gzip for the compression of data in all instances. Pigz can be much faster than gzip and is able to use multiple processing cores.
3. If the sample sequenced was not the same strain as was used to construct the reference genome, the alignment of sequencing reads will be improved by altering the reference to better match the sample sequenced. If genomic variant data for your sample is available, you should substitute SNP bases into the reference to produce a SNP-corrected FASTA file to use as a reference genome.
4. BS-Seeker2 will launch two bowtie2 processes in parallel, so the number of cores specified on the command line is the number of cores made available to each bowtie2 process. For example, if --bt2-p 4 is specified, two bowtie2 processes are launched each using four cores, totaling eight cores being used at the same time.
5. Older versions of samtools use slightly different command line options. If you encounter errors, first check that the options specified are appropriate for the version of samtools you are using. The program sambamba can be used as an alternative to samtools and has slightly better performance.
6. The genome db file is generated when the genome is indexed by BS-Seeker2 and is usually found at `BSseeker2/bs_utils/reference_genomes/<genome_db>`.
7. A python script to find the methylation bias along reads is available here: https://github.com/timoast/ngs-tools/blob/master/mBias.py.
8. The percentage of uniquely mapped reads can vary greatly between experiments but should typically be >50%. Similarly, the percentage of reads with multiple mapping positions (multi-mapped reads) will vary depending on the experiment and the reference genome but should typically be below 20%.

Furthermore, multi-mapping reads are generally not useful for the measurement of DNA methylation, as it cannot be determined which genomic region these reads originated from. For this reason, such reads should be discarded from further analysis. This can have important consequences for the study of DNA methylation in repetitive genomic regions such as transposable elements. The percentage of PCR duplicates should also scale with the number of PCR cycles performed during library preparation, and so investigators should be conscious of keeping the number of PCR cycles to a minimum. Very low mapping efficiency could be indicative of a number of problems with the library or analysis method. Common problems include a high amount of contaminating DNA, leading to reads that are not derived from the reference genome. This can be assessed by mapping a subset of the reads to different genomes to try to determine the origin of the unmapped reads, so that experimental procedures can be altered in the future to avoid similar problems.

9. DSS can be run with or without replicates.
10. As DNA methylation levels at CG dinucleotides are highly correlated between strands, it can be beneficial to aggregate the data for CG position for both strands. This has two benefits: First, it increases the coverage at each position for the DMR calling, and, second, it reduces the number of bases for DMR calling by half which reduces file size and speeds up the analysis. The data for each strand can be aggregated by running the R script available here: https://gist.github.com/SamBuckberry/1fb57d8e6f9927a086516cb536c172f2.
11. Genomic experiments are often discovery-based, meaning that investigators often do not set out with a defined hypothesis to test. This can lead to the problem of forking paths in a data analysis, as comparisons can be made in many different ways, with the explored paths that yield uninteresting results ultimately unreported. Unfortunately, this approach effectively leads to a problem of multiple hypothesis testing and can result in a higher false-discovery rate than should be expected of these types of experiments. There is a simple solution to this, which is to divide the initial data into subsets at an early stage of the analysis. For WGBS data, this could be done with a set of DMRs, withholding say 30% of the DMRs and leaving them untouched. You can then safely explore as many paths as you wish. When you are at the stage of writing up your results, add in the retained DMRs and see what effect it has on your original conclusion. If the effect goes away, it was likely a false positive.

References

1. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14:204–220. https://doi.org/10.1038/nrg3354
2. Friso S, Choi S-W, Dolnikowski GG, Selhub J (2002) A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. Anal Chem 74:4526–4531
3. Weber M, Davies JJ, Wittig D et al (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37:853–862. https://doi.org/10.1038/ng1598
4. Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15:394–408. https://doi.org/10.1038/nrg3683
5. Frommer M, McDonald LE, Millar DS et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89:1827–1831. https://doi.org/10.1073/pnas.89.5.1827
6. Lister R, O'Malley RC, Tonti-Filippini J et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133:523–536. https://doi.org/10.1016/j.cell.2008.03.029
7. Cokus SJ, Feng S, Zhang X et al (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452:215–219. https://doi.org/10.1038/nature06745
8. Andrews S FastQC A Quality control tool for high throughput sequence data. In: bioinformatics.babraham.ac.uk. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 18 Apr 2017
9. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17:10–12. https://doi.org/10.14806/ej.17.1.200
10. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
11. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
12. Guo W, Fiziev P, Yan W et al (2013) BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics 14:774. https://doi.org/10.1186/1471-2164-14-774
13. Feng H, Conneely KN, Wu H (2014) A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic Acids Res 42:e69. https://doi.org/10.1093/nar/gku154
14. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033
15. Ramirez F, Dundar F, Diehl S et al (2014) deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 42:W187–W191. https://doi.org/10.1093/nar/gku365
16. Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354
17. Wu H, Xu T, Feng H et al (2015) Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. Nucleic Acids Res 43(21):e141. https://doi.org/10.1093/nar/gkv715
18. Lawrence M, Huber W, Pagès H et al (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118
19. Kawakatsu T, Stuart T, Valdes M et al (2016) Unique cell-type-specific patterns of DNA methylation in the root meristem. Nat Plants 2(5):16058. https://doi.org/10.1038/nplants.2016.58
20. Stroud H, Greenberg MVC, Feng S et al (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152:352–364. https://doi.org/10.1016/j.cell.2012.10.054