



Library preparation and sequencing

Samples were quantified and quality checked using Qubit 2.0 Fluorometer (Invitrogen, Carlsberg, CA). Based on the kind of biological material (plant or mammals) the digestion of DNA was performed using TaqI or MspI enzymes (NEBs, Ipswich, MA) respectively. The EpiTect Fast Bisulfite Conversion Kit (Qiagen, Hilden, DE) was used to perform the Bisulfite conversion step. RRBS libraries were prepared using the Ovation Ultralow Methyl-Seq system (NuGEN Technologies, San Carlos, CA) following the manufacturer's recommendations. Libraries were quality checked using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and then sequenced on the HiSeq2500 platform (Illumina, San Diego, CA).

Reduced-representation bisulfite sequencing (RRBS-Seq) standard bioinformatic analysis includes:

- Base calling and demultiplexing using standard Illumina pipeline (CASAVA v1.8.2).
- Quality and adapter trimming using Trim Galore.
(https://www.bioinformatics.babraham.ac.uk/projects/trim_galore).
- Bisulfite conversion rate estimation on lambda phage DNA.
- Alignment to the reference genome using Bismark v0.19.0 (Krueger and Andrews 2011) setting the parameters `--bowtie2` and `-N 1`.
- Extraction of methylated cytosines from aligned reads using the *Bismark_methylation_extractor* routine (Krueger and Andrews 2011) by setting the parameter `-ignore=3` to ignore the first three nucleotides in the reads (corresponding to the P1 adapter overhang).
***NOTE:** methylation levels of sites carrying C->T or A->G polymorphisms can be biased. A possible solution is to provide a list of known C->T and A->G polymorphisms of the study sample. In case a list of polymorphisms is not available, inquire about our whole genome sequencing and SNP detection services.*
- Differential methylation analysis using the *calculateDiffMeth* function of methylKit v1.4.1 (Akalin et al. 2012). Differential methylation analysis is performed at base pair resolution and in regions 1000 bp wide.
- Gene annotation of differential methylated positions and regions.

Delivery Files:

- Folder *raw_sequences* containing FASTQ files with raw reads and QC files.
- Folder *alignments* containing alignments files in BAM format and the corresponding index files. BAM files can be visualized using a BAM data visualization tool such as *Tablet* (<https://ics.hutton.ac.uk/tablet/>).
- Folder *metrics_and_statistics* containing for each sample a pdf report summarizing alignment and methylation statistics and a tab-delimited file with comprehensive coverage and methylation statistics in the following format:
<chromosome> <start position> <end position> <methylation percentage> <count methylated> <count unmethylated>
- Folder *differential_methylation* containing differential methylation analysis results. Results of each comparison are included in a separated folder (named *group1_vs_group2*) containing the following tab-delimited files:
 - AllSites.txt: list of all positions covered in all the samples included in the two comparison groups together with the significance of the differential methylation analysis (p-values and q-values) and the methylation percentage change between the two groups. A positive methylation percentage change value means a higher methylation level in group1 with respect to group2; a negative methylation percentage change value means a lower methylation level in group1 with respect to group2. Format: <chromosome> <position> <pvalue> <qvalue> <methylation percentage change>
 - DiffSites_annotated.txt: list of the differentially methylated positions (*i.e.* position in which q-value<0.01 and methylation percentage change higher than 25) with the gene annotation. Format: <chromosome> <position> <pvalue> <qvalue> <methylation percentage change> <closest TSS gene ID> <distance to closest TSS> <closest TSS gene strand> <affected promoter gene ID> <affected exon gene ID> <affected intron gene ID>
 - AllRegions.txt: list of all regions covered in all the samples included in the two comparison groups along with the significance of the differential methylation analysis (p-values and q-values) and the methylation percentage change between the two groups. Format: <chromosome> <start region> <end region> <pvalue> <qvalue> <methylation percentage change>
 - DiffRegions_annotated.txt: list of the differentially methylated regions (*i.e.* regions in which q-value<0.01 and methylation percentage change higher than 25) with the gene annotation. Format: <chromosome> <start region> <end region> <pvalue> <qvalue> <methylation percentage change> <closest TSS gene ID> <distance to closest TSS> <closest TSS gene strand> <affected promoter gene ID> <affected exon gene ID> <affected intron gene ID>

References:

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* [Internet] 13:R87. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23034086>

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* [Internet] 27:1571–1572. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21493656>