Pipeline Overview

The ENCODE Whole-Genome Bisulfite Sequencing (WGBS) pipeline is used for discovering methylation patterns to base granularity. Bisulfite treatment is used to convert cytosines into uracils, but leaves methylated cytosines unchanged. After mapping bisulfite sequencing reads against a Bismark transformed genome, this pipeline then extracts the CpG, CGH, and CHH methylation patterns genome-wide. The WGBS pipeline inputs gzipped DNA-sequencing reads (fastqs) and a Bismark-transformed, Bowtie-indexed genome in a tar.gz archive file. These are processed to generate bam alignment files, which in turn produce:
- Methylation state at CpG (bedMethyl.gz and bigBed)
- The methylation state at CHG (bedMethyl.gz and bigBed)
- The methylation state at CHH (bedMethyl.gz and bigBed)
- Raw signal files of all reads (bigWig)
- SamTools quality metrics, Bismark quality metrics
- Pearson correlation, calculated from the two replicates' methylation states at CpG.

Description of bedMethyl file

The bedMethyl file is a bed9+2 file containing the number of reads and the percent methylation. Each column represents the following:
1. Reference chromosome or scaffold
2. Start position in chromosome
3. End position in chromosome
4. Name of item
5. Score from 0-1000. Capped number of reads
6. Strandedness, plus (+), minus (-), or unknown (.)
7. Start of where display should be thick (start codon)
8. End of where display should be thick (stop codon)
9. Color value (RGB)
10. Coverage, or number of reads
11. Percentage of reads that show methylation at this position in the genome

Genomic References Used in this Pipeline
- Unmodified Genome References and Chromosome Sizes, including lambda genome for generation of comparative statistics (encodeproject.org/ENCSR425FOI)
- Bismark/Bowtie reference (https://www.encodeproject.org/references/ENCSR497EUF/)

References

Krueger, Felix, and Simon R. Andrews. "Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications." *Bioinformatics* 27.11 (2011): 1571-1572.

Tsuji, Junko, and Zhiping Weng. "Evaluation of preprocessing, mapping and postprocessing algorithms for analyzing whole genome bisulfite sequencing

data." *Briefings in bioinformatics* (2015): bbv103.