Methylation Haplotype Analysis User Guide

Methylation Haplotype Analysis starts from the BAM files. BS-seq or RRBS sequencing data are recommend to be aligned by BisReadMapper [[1](#_ENREF_1)] or Bismark [[2](#_ENREF_2)].

Quality trim (Trim-galore), adaptor trim (Trim-galore, Cutadapt or Cleanadaptors), pair-end sequencing stitch (COPE) and any other preliminary processing should be conducted according to the sequencing strategy of the authors. Our custom Perl script could extract haplotype information (haploinfo) and another Perl script could provide the calculation of methylation haplotype load (mhl). Please be remember the bismark alignment has totally different SAM format and flag labelling system. Please be sure to give bismark or BisReadMaper options when BAM2Hapinfo were processing. Some other script such as LD (R2) calculation is also provided which take haploinfo file as the input. All the script will be updated continuously. Any question or suggestion, please don't hesitate to email Dr. Kun Zhang < kzhang@bioeng.ucsd.edu >.

1, Bam files to methylation haploinfo files

perl ~/bin/mergedBam2hapInfo.pl

USAGE: mergedBam2hapInfo.pl target\_list\_file merged\_bam <bisReadMapper|bismark> ChrosmeZizeFile CpG\_Position\_File

1, target\_list\_file: bed file to assign the genomic regions to achieve the methylation haplotype.

2, merged\_bam is the BAM file for each sample combined from chr1 to chrY

3, Alignment method: bisReadMapper or bismark. If you use BWA or BOWTIE, please choose bisReadMapper.

4, ChrosmeZizeFile: please download it from UCSC

5, CpG\_Position\_File: CpG position in the genome.

2, haploinfo files to methylation haplotype load matrix

perl hapinfo2mhl.pl Hapinfo\_File\_List.txt > MHL.output.txt

Hapinfo\_File\_List.txt: Is the list of the hapinfo files, each line is one hapinfo files. Please assign absolute directory.

1. Diep, D., et al., *Library-free methylation sequencing with bisulfite padlock probes.* Nat Methods, 2012. **9**(3): p. 270-2.

2. Krueger, F. and S.R. Andrews, *Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.* Bioinformatics, 2011. **27**(11): p. 1571-2.