**Methylation Haplotype Analysis User Guide**

File description:

|  |  |
| --- | --- |
| **File** | **Description** |
| bam2hapinfo.pl | Extract methylation haplotype information from aligned BAM files |
| hapinfo2LDR2.pl | Calculate LD R-square on all the genomic intervals for methylation haplotype file. |
| hapinfo2mhb.pl | Infer methylation haplotype blocks from methylation haplotype files |
| hapinfo2mhl.pl | Calculate methylation haplotype load from methylation haplotype files |
| Figure.2. R | R script to prepare Figure 2 simulation |
| GuassianAssignment.R | R script to assign tissue-of-origin of each patients samples based on Z-scores compared with background signals in normal plasma. |
| RAnalysis.R | Other related R scripts and functions for main figures. |

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 additional preprocessing could be conducted before methylation haplotype analysis pipeline. We used one Perl script to extract haplotype information (hapinfo) from bam files, and another Perl script to calculate methylation haplotype load (MHL). Please note that Bismark alignment has totally different SAM format and flag labelling system. Please make sure to specify the Bismark or BisReadMaper option when running *bam2hapinfo.pl*. Additional scripts such as LD (r2) calculation were also provided which take a hapinfo file as the input. All the scripts will be updated continuously. For questions or comments, please contact Dr. Kun Zhang < kzhang@bioeng.ucsd.edu >.

**1, Bam files to methylation hapinfo files**

perl ~/bin/bam2hapinfo.pl

USAGE: bam2hapinfo.pl target\_list\_file merged\_bam <bisReadMapper|bismark> ChrosmeSizeFile 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, ChrosmeSizeFile: please download it from UCSC

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

**2, hapinfo files to methylation haplotype block (MHB)**

perl ~/bin/hapinfo2mhb.pl HapinfoMerge.txt 0.5 > methylationHaplotypeblock.bed

HapinfoMerge.txt: merged hapinfo files and the minimal r2 value to define MHBs.

**3, hapinfo files to LD r matrix by genomic intervals**

Perl ~/bin/hapinfo2LDR2ByBed.pl input.bed input.hapinfo

**3, hapinfo 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 use absolute directory.

**Reference:**

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.