BS-SNPer User Guide

1. Introduction

BS-SNPer is an ultrafast and memory-efficient package, a program for BS-Seq variation detection from alignments in standard BAM/SAM format (Li, et al., 2009). The purpose of this program is to call SNP using bisulfite sequencing data.

2. System requirement

BS-SNPer works on Unix (Linux, Ubuntu, MacOSX, etc) based systems.

Hardware requirements

> One computing node equipped with at least 8 GB Memory

Software requirements

> Linux 64-bit version
>
> GCC 4.6.0 or higher
>
> Perl 5.16.3 or higher

3. Getting started

Install

Download BS-Snper from https://github.com/hellbelly/BS-Snper. After extracting the downloaded package, execute the command ./BS-Snper.sh. Make sure the executable file rrbsSnp is generated.

Run

You can run BS-Snper in Linux or MAC OS, using the command like:

```
perl BS-Snper.pl --interval hg19.len --fa hg19.fa --input merge.sort.bam --output SNP.candidate --methoutput Meth.out --minhetfreq 0.1 --minhomfreq 0.85 --minquali 15 --mincover 10 --maxcover 1000 --minread2 2 --errorate 0.02 --mapvalue 20 >SNP.out2 2>SNP.log
```

4. Input file

Any alignments in standard BAM/SAM format.

5. Output files

Meanings of the fields of the output SNP information are:

1. CHROM: Chromosome

2. POS: Coordinate

3. ID: Just like vcf format, but in this version, we didn't open this function. The user can get the information according to the DBSNP dataset easily.

4. Ref: reference base(s): Each base must be one of A,C,G,T (case insensitive).

5. ALT: alternate base(s).

6. QUAL: quality: Phred-scaled quality score.

7. FILTER: filter status: PASS if this position has passed all filters, i.e. a call is made at this position.

8. GENOTYPE: genotype of this position.

9. FREQUENCY: allele frequency

10. Number_of_watson: the number of A,T,C,G covered in Watson strand.

11. Number_of_crick: the number of A,T,C,G covered in Crick strand.

12. Mean_Quality_of_Watson: the mean base quality of A,T,C,G covered in Watson strand.

13. Mean_Quality_of_Crick: the mean base quality of A,T,C,G covered in Crick strand.

PS: Please refer to the format of VCF from 1st to 7th column.

6.   Contact information

If you have any problem please do not hesitate to contact us!

gaoshengjie@genomics.org.cn