bisReadMapper_v1 USER MANUAL

This software only works on Linux operating systems (64 bits), and have been tested on:
Ubuntu 10.04 LTS
Perl v5.10.1
SOAP 2.21release (Only works on 64 bits systems!)
Samtools 0.1.8

Download required softwares:
1) SOAP - http://soap.genomics.org.cn/down/soap2.21release.tar.gz
After downloading, extract the files with:
```
tar -xvzf soap2.21release.tar.gz
```

2) Samtools - http://sourceforge.net/projects/samtools/files/samtools/
After downloading, extract the files with:

```
bzip2 -d samtools-0.1.8.tar.bz2
tar -xvf samtools-0.1.8.tar
```

Make sure that these dependencies are installed for samtools to compile:
```
sudo apt-get install zlib1g-dev libncurses5-dev
```

Then navigate into the folder and compile:
```
cd samtools-0.1.18
make
```

3) Check that Perl version 5 or later is installed
```
perl --version
```

Download genome references:
1) full reference genome can be downloaded from UCSC Genome Browser. Note: chromosome names should be the standard names.

For hg18:
```
wget ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/chromosomes/chr*.fa.gz
```
For hg19:
```
wget ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/chr*.fa.gz
```

Move unwanted chromosome files from the directory before concatenating files to get standard chromosomes only.
i.e. `mkdir nonStandardChromosomes`
i.e. `mv chr*_*.fa.gz > nonStandardChromosomes/`

To get a single genome file, concatenate all chromosome files into one with less:
i.e. `less chr*.fa.gz > HsGenome18.fa`

For SNP calling, d*ownload the latest reference dbSNP file compatible with your specific reference genome:*

For hg18:
```
wget ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/snp130.txt.gz
```
For hg19:
```
wget ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/snp132.txt.gz
```

To decompress dbSNP file:
i.e. `gzip -d snp130.txt.gz`

Required steps:
1) run genomePrep on single genome file (program will write to the current directory) . This program have two

optional inputs: context=[all, cg], and convert=[yes,no]. The default is to perform calls only on CpGs (context=cg, and convert=yes) and to perform in silico bisulfite conversion (generates *.bis.fwd and *.bis.rev files.) If convert=no is used, the converted files will not be generated.

i.e: /path/to/`genomePrep.pl genome.fasta`
Perform in silico bisulfite conversion and call only CpGs.

i.e: /path/to/`genomePrep.pl genome.fasta context=all convert=no`
Do not perform in silico bisulfite conversion and call all C (when you already have a converted genome.)

2) run 2bwt-builder (Go to http://soap.genomics.org.cn/soapaligner.html for more instructions) on the *.bis.fwd and *.bis.rev files.

i.e: /path/to/soap/`2bwt-builder genome.fasta.bis.fwd`
i.e: /path/to/soap/`2bwt-builder genome.fasta.bis.rev`

3) run samtools faidx on the *.bis.fwd and *.bis.rev files

i.e: /path/to/samtools/`samtools faidx genome.fasta.bis.fwd`
i.e: /path/to/samtools/`samtools faidx genome.fasta.bis.rev`

4) prepare params.txt file for each sample (See example file: sampleParams.txt)

reads= provide full path to reads file(s), for pair-end, list two reads files separated by a comma (,) and no space in between

name= (optional) provide the name of the sample to be used for labeling output files [default=Sample]

allC= (optional) to call methylation on all C position (1) or only CG will be called (0) [default=0]

length= (optional) provide the read length for calculating allowable mismatches (1 per 40 for single-end and 1 per 30 for paired-end) [default=80]

refDir= provide path to reference genome index directory
Note: there should only be one bisulfite reference per directory
Required contents for refDir: *.bis.fwd, *.bis.rev, SOAP *.bis.fwd.index, SOAP *.bis.rev.index, *.chrSizes, *.cpositions.txt

snp= (optional) provide full path to dbSNP file
Note: No SNP will be called if this file is not provided

soapDir= provide path to the SOAP software

samtoolsDir= provide path to the Samtools software

alignMode= specify whether to align reads as paired (P) or single-end (S)
Note: must provide two reads files for setting P

qualBase= Illumina base quality scale, Phred33 (33) or Phred64 (64)

numCPU= (optional) specify the number of processor cores available for mapping [default:1]

trim5= Number of base to trim from the 5' end of reads [default:0]

trim3= Number of base to trim from the 3' end of reads [default:0]

4) run bisReadMapper.pl (program will output to the current directory)

i.e: /path/to/bisReadMapper_v1.2/`bisReadMapper.pl sampleParams.txt > sample.log`