

README_ILLUMINA

Featured Updated Sep 9, 2013 by drcy...@gmail.com

Installation

Building Indexes.

Mapping

Output

Sample Run

Installation

To install the package, do the following.

untar the package:

tar -zxvf batmethxxxx

change to installation directory batmethxxxx

cd batmethxxxx

Configure and Install

./configure

make

make copy

Building Indexes.

All executables of this package can be found in batmethxxxx/bin folder after make copy.

Two sets of indexes needs to be built. They are built by calling build_index as follows.

build_index genome.fa GTOA

build_index genome.fa CTOT

1) will turn all G's to A's and build an index with the base name genome.fa-GtoA

2) will turn all C's to T's and build an index with the base name genome.fa-CtoT

To build the required indexes and delete unwanted files run the script ***build_all genome.fa***

Mapping

To map the data set METH to the genome with base name GENOME do,

./batmeth -q GENOME -i INPUT METH -n <number of mismatches> -o <Output Prefix>

./split <FINAL RESULT> GENOME <number of mismatches> y <all...output...files...from...batmeth...>

Output

<Final_result> will contain all uniquely mapped hits (after PCR duplicated removal) in the following format:

Line 1: Header

Line 2: <read_ID> <Original Read>

Line 3: <read-orientation> <chrNum> <strand> <chrPos_0> <mismatch> <read_length>

<methylation status>

<read-orientation> is an integer: [1-4]. 1 & 3 has the read and reverse-complement of the read mapped to a C->T converted reference respectively. 2 & 4 will have its read and reverse-complement of the read mapped to a G->A converted reference respectively.

<methylation status> is a string which contains per-base methylation-related information on the read. This string consists of {=,M,U,A,C,G,T}. = is for a match between the read and the reference; A/C/G/T is the genomic base to which the read has a corresponding mismatch with. M means that this positional base in the read is methylated; otherwise, it will be an U.

Please ignore the other fields which are appended to the end of the lines.

Sample Run

```
[linux]$ .batmeth -g genome.fa -i query.fa -n 2 -p 4 -O 1
```

```
[linux]$ .split output.out genome.fa 2 y query.fa.0 query.fa.1 query.fa.2 query.fa.3
```