BS-seq Methods

Pablo Cingolani

1 Methods

1.1 Sequencing Analysis

In total we sequenced **NUMBER-OF-LANES** lanes using Illumina Genome Analyzer GAIIx with a Paired End Cluster Generation Kit. Obtaining **NUMBER-OF-READS**, **76** bp, pair-ended reads. Sequence extraction and base calling was done using Illumina Pipeline v1.6 software. Image analysis and basecalling was done using *SCS* (Sequencing Control Software).

1.2 Mapping

Read mapping and downstream analysis was done using our in-house system (MoreBs version 1.0) which uses either BWA[5][6] or Bowtie[3] for read alignment. Both alignment programs are based on Burrows-Wheeler transform and create *SAM*[7] output format. In this case BWA was selected as main mapping method in order to have better alignment near insertions and deletions. MoreBs parameters were the following **INSERT PARAMETERS HERE**.

1.3 Methylation Assessment

MoreBs perform methylation calls by invoking *Samtools* to create *BAM* files, *sorted BAM* files and *mpileup*. In this last step methylation calls are produced using an MAQ[8] probabilistic model. It must be noted that BAQ[4] model is explicitly disabled by MoreBs, since some of the its assumptions do not apply for methylation calls. Finally *BcfTools* package is invoked to produce methylation calls in *VCF4.0*[2] format.

As a final step, MoreBs performs statistics on methylation as well as ranks of hypo-methylated and hyper-methylated genes by means of Fisher exact test. Multiple testing is corrected using False Discovery Rate methodology[1].

Some statistics were carried out using custom programs in R programming language (www.r-project.org).

2 References

References

[1] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[2] Broad Institute. VCF (Variant Call Format) version 4.0. *http://www.1000genomes.org/wiki/Analysis/vcf4.0*, 2010.

[3] B. Langmead, C. Trapnell, M. Pop, and S.L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3):R25, 2009.

[4] H. Li. Mathematical Notes on SAMtools Algorithms. *http://lh3lh3.users.sourceforge.net/download/samtools.pdf*, 2010.

[5] H. Li and R. Durbin. Fast and accurate short-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(5), 2009.

[6] H. Li and R. Durbin. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5):589, 2010.

[7] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078, 2009.

[8] H. Li, J. Ruan, and R. Durbin. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research*, 18(11):1851, 2008.