**Whole-genome bisulfite sequencing data analysis reveals differentially methylated regions on a dataset of epilepsy in animal models**

W. Souza1, B. S. Carvalho2, D. B. Dogini1, I. Lopes-Cendes1

1Department of Medical Genetics, School of Medical Sciences, 2Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing; University of Campinas – UNICAMP and the Brazilian Institute of Neuroscience and Neurotechnology, Campinas, SP, BRAZIL

**Introduction:** Whole-genome bisulfite sequencing (WGBS) became the gold standard for mapping DNA methylation profile of vertebrate species. In this methodology, DNA molecules are treated with sodium bisulfite that chemically convert cytosine residues into uracil maintaining 5-methylcitosine unchanged. During polymerase chain reaction uracil residues are transcribed as thymine. Bisulfite-treated DNA are quantified by high-throughput sequencing technologies [1]. There are software tools available to process each analysis step, such as filtering and aligning bisulfite-treated reads [2]. However, there is a lack of bioinformatics protocols for WGBS data analysis. Differentially methylated regions (DMR) are sections of the genome where we observe significant differences in methylation levels when comparing treated to control individuals. We developed a data analysis protocol to process WGBS data and identify DMRs. We applied this protocol to study the genetic landscape of epilepsy on pilocarpine-induced animal model.

**Materials and Methods:** The dataset consists of 2 samples of control animals and 2 samples of pilocarpine-treated animal model of epilepsy. We generated these data using Illumina TruSeq DNA Methylation kit and Illumina HiSeq 2500 sequencer. Our protocol for WGBS data analysis is comprised of several steps: A) quality assessment via Rqc Bioconductor package; B) trimming and adapter clipping through Tim Galore!; C) mapping to the *Rattus* *novergicus* reference genome (v5) using Bismark [3] (duplicate and multi-mapped reads were removed); D) determination of the DNA methylation frequency as the number of reads that aligned on dinucleotides CpG and; E) WGBS data smoothing and DMR identification performed with the bsseq software [4].

**Results:** Quality assessment of bisulfite sequencing dataset showed a potential bias on raw sequencing data: we expect a low percentage of cytosine calls due to bisulfite conversion; however, we observed an increase of cytosines towards the end of the reads. After trimming adapter sequences, we noticed that the proportion of cytosine calls was constant across all cycles. Comparison between control group and pilocarpine-treated groups showed differences on DNA methylation profiles. We identified eighteen DMRs on these WGBS data.

**Discussion:** Due to the lack of alternatives, developing this protocol for bisulfite sequencing data analysis was essential to identify differences on DNA methylation profiles between control and pilocarpine-treated animal models. The identified DMRs are indicators for future studies, as they can be associated to gene expression regulation, playing a fundamental role on integrative genomic analyses.

**Conclusion:** We developed a WGBS data analysis protocol. We used this protocol to analyze dataset of control animals and pilocarpine-treated animals. We found differences on DNA methylation profiles.

**References:** [1] Bock C et al., Nat Biotechnol. 28(10):1106–14, 2010; [2] Bock C., Nat Rev Genet. 13(10):705–19, 2012; [3] Krueger F, Andrews SR., 27(11):1571–2, 2011; [4] Hansen KD et al., Genome Biol. 13(10):R83, 2012.

Supported by: CEPID-FAPESP