CpG Methylation Analysis from Targeted Sequencing of Bisulfite Converted DNA

Jason Warner1, Steve Kotsopoulos1, Michael Samuel1, Jochen Hampe2, Alexander Herrmann2, Kathryn Stemshorn2, Karol Szafrański2, Peter Nuernberg2, Darren Link1, Jeff Olson1, RainDance Technologies, Inc., Lexington, MA, 2University of Kiel, Kiel, Germany.

OVERVIEW

RainDance Technologies (RDT) has extended the capabilities of the RDT 1000 to target genomic regions of bisulfite converted DNA. This approach used in conjunction with high-throughput sequencing enables researchers to measure the methylation status of targeted regions of the genome with complete sequence coverage, specificity, and uniformity.

Recent advances in high-throughput sequencing technologies have enabled the analysis of differential methylation patterns at a genome-wide scale. These genome-wide approaches have enabled the discovery of epigenetic variations associated with disease progression, including cancer. To validate these candidate regions with a sufficient number of samples, a targeted approach is required to maximize the breadth and the depth of coverage for these regions.

Targeted bisulfite sequencing data provides a unique set of challenges. The majority of bisulfite-converted DNA is devoid of cytosines, which creates difficulties when mapping the short sequencing reads from Next-Generation Sequencing platforms. Due to the nature of a bisulfite converted genome, the sense and antisense strands are no longer complementary following bisulfite conversion, which further increases the size and complexity of the sequence space.

RainDance has developed a Targeted Bisulfite Sequencing Analysis Pipeline designed to address these challenges by standardizing the collection, storage, and assembly of bisulfite-converted DNA. Our approach utilizes any standard FASTQ format that is pre-processed using a custom RainDance Perl script. The processed FASTQ file is inputted into the CLC Genomics Workbench from CLC Bio for sequence alignment to produce a standard SAM assembly. The SAM file is then further processed with another custom RainDance Perl script to produce the final data output.

The RainDance Targeted Bisulfite Sequencing Analysis Pipeline was demonstrated to enable the quantitative analysis of CpG methylation regions with high completeness (C1 of 97.3%; AviC of 31.0%), specificity (76.8% of mapped reads were on target) and uniformity (base coverage at 0.2x of AviC was 75.9%) of alignment of sequencing reads.

References


Conclusions

Our strategy for analysis of targeted bisulfite sequencing data maximized the specificity and accuracy of the alignment of sequencing reads to enable quantitative analysis of CpG methylation. The RDT Targeted Bisulfite Sequencing Workflow and Analysis pipeline yielded very high completeness (C1 of 97.3%; AviC of 31.0%), specificity (76.8% of mapped reads were on target), and uniformity (base coverage at 0.2x of AviC was 75.9%). The statistical power of the mi-crofluidic-based PCR format will enable highly quantitative measurement of CpG methylation using targeted resequencing of bisulfite-converted template DNA.