diagenode
A Hologic Company

APPLICATION NOTE

Diagenode Premium RRBS technology: Cost-effective DNA methylation mapping with superior CpG resolution and coverage

Anne-Clémence Veillard (1), Paul Datlinger (2), Miklos Laczik (1), Sharon Squazzo (3) and Christoph Bock (2)

(1) Diagenode, Liège, Belgium. **(2)** CeMM - Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. **(3)** Diagenode Inc., Denville, USA.

Correspondence should be addressed to: anne-clemence.veillard@diagenode.com

Abstract

DNA methylation is an important epigenetic mark with broad relevance in development and disease. Reduced representation bisulfite sequencing (RRBS) enables genome-scale DNA methylation analysis with all the advantages of sequencing technology (accuracy, robustness, flexibility) at a fraction of the cost of whole genome bisulfite sequencing. Diagenode Premium RRBS kit makes this powerful technology readily available to any interested laboratory. Its optimized protocol provides genomic coverage for 3.5 to 4 million CpG dinucleotides in the human genome, which is 5-10 fold higher than the coverage of commercially available DNA methylation microarrays. The method covers not only CpG islands and promoter regions, but also a broad sampling of other functional elements including enhancers, CpG island shores and non-coding RNAs. Finally, its optimized workflow with multiplexing prior to bisulfite conversion allows for convenient processing of up to 96 samples in a single experiment, thus enabling epigenome-wide association studies in large cohorts and in any vertebrate species using an optimized and easy-to-use kit.

Introduction

DNA methylation provides an essential mechanism of epigenetic gene regulation in embryonic development, with an important role for lineage commitment, genomic imprinting and chromosome stability. DNA methylation patterns are deregulated in many diseases and most strongly in cancer.

Several methods exist for mapping DNA methylation patterns throughout the genome. DNA methylation microarrays such as the Infinium assay make it easy to study large cohorts, but they provide low genome-wide coverage, are susceptible to batch effects and available only for the human genome. Whole genome bisulfite sequencing (WGBS) provides accurate and robust coverage for most CpGs in the genome, but the associated sequencing costs are limiting its application to small sample sizes, making it hard to draw biologically relevant conclusions. The RRBS technology combines the strengths of both assays, providing an accurate and cost-effective assay that has all the advantages of bisulfite sequencing, including its robustness and flexible use in any species.

RRBS is based on the observation that DNA methylation in vertebrates occurs mainly at CpG dinucleotides[1], and it uses the restriction enzyme MspI (target site: C^CGG) to generate genomic fragments starting and ending with a CpG dinucleotide independently of their DNA methylation status. Subsequent size selection and bisulfite sequencing provides high coverage for CpG islands and promoter regions, and also subset of other genomic elements such as enhancers, CpG island shores and non-coding RNAs. RRBS has seen a recent rise in popularity as researchers switch from microarray-based assays to next generation sequencing.

Diagenode Premium RRBS kit overcomes many limitations of the original RRBS protocol, which was time-consuming and

tedious, and it unlocks the advantages of genome-scale bisulfite sequencing for any interested laboratory. Key points of the optimized RRBS protocol include: (i) superior coverage of 3.5 to 4 million CpGs; (ii) support for allele-specific measurements; (iii) robustness toward batch effects; (iv) applicable to any vertebrate species; (v) compatible with low-input samples; and (vi) feasible for formalin-fixed, paraffin-embedded (FFPE) samples. Due to its versatility and high reproducibility, RRBS is the technology of choice for accurate genome-scale DNA methylation analysis[2, 3].

Workflow

Diagenode Premium RRBS technology offers a complete solution optimized for DNA methylation mapping in vertebrate species. The kit contains reagents for enzymatic digestion, library preparation, bisulfite conversion and amplification and is available in two sizes: 24 and 96 reactions. Depending on the species, up to 24 samples can be combined and sequenced together, which reduces handling time, reagent consumption and cost per sample. For human, we recommend sequencing up to 6 samples per lane on the Illumina HiSeq 2000/2500 platform or up to 12 samples per lane on the HiSeq 3000/4000 and on the NextSeq high output configuration.

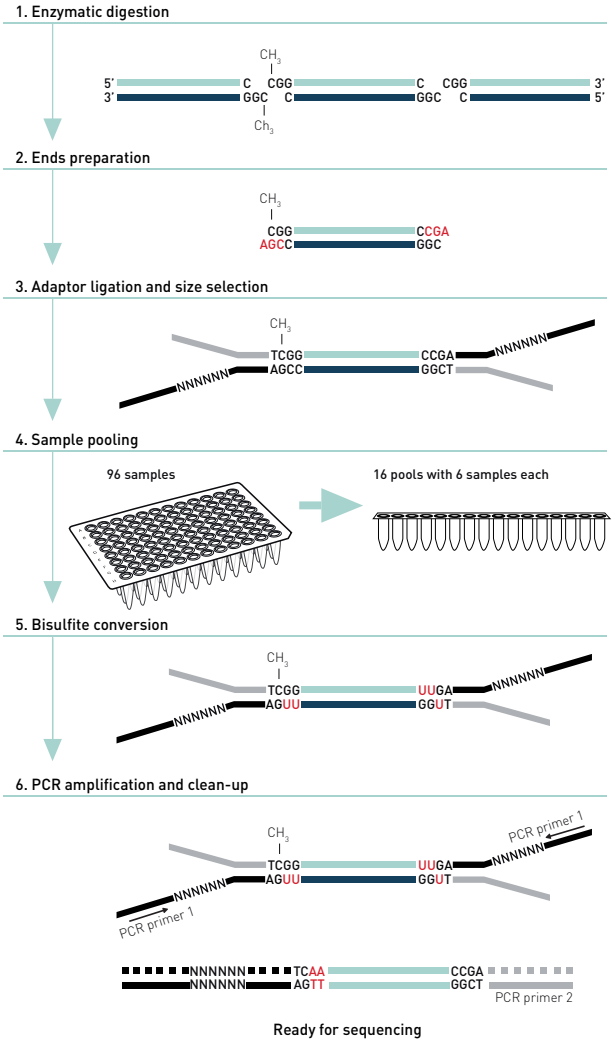


Figure 1: *Reduced Representation Bisulfite Sequencing (RRBS) Workflow*

The protocol described in Figure 1 starts from 100 ng of DNA and has been validated on FFPE samples without any modifications. In the first step, genomic DNA is digested with the restriction enzyme MspI, followed by a single-tube library preparation workflow completely devoid of clean-up steps, which minimizes material loss during library preparation. The kit includes methylated and unmethylated spike-in controls to precisely monitor bisulfite conversion efficiencies. Carefully optimized bead-based size selection retains even the smallest library fragments while efficiently removing adaptor dimers, resulting in an increased genomic coverage.

One distinguishing feature of Diagenode's protocol is that the samples are pooled early on, which reduces handling time and cost per sample. To maximize the pooling precision, custom software was developed and is available on the Diagenode website: https://www.diagenode.com/documents/rrbs-pooling-aid. This software assists the researcher in defining pools of samples based on their performance throughout the protocol and the compatibility of their barcode indices. Each kit contains 24 different barcodes, which allows the pooling of 6 to 24 samples. The pooled samples are then bisulfite converted using optimized conditions to decrease DNA degradation while keeping highly efficient conversion of unmethylated cytosines. To avoid potential bias during the enrichment PCR, the minimum number of amplification cycles is determined based on qPCR results. Diagenode MethylTaq Plus enzyme was specifically developed to amplify bisulfite-converted DNA with high accuracy and efficiency, which further reduces the number of PCR cycles needed.

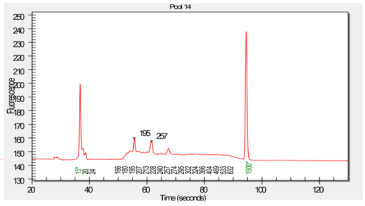
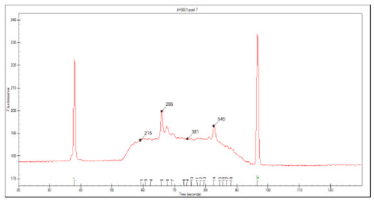
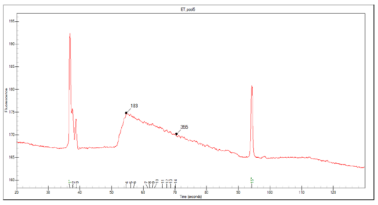
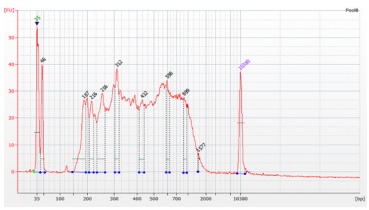
A variety of bioinformatics tools were specifically developed for the analysis of RRBS sequencing data[4]. The typical analysis routine starts by trimming reads for the removal of adaptor sequences, sequencing artifacts or low-quality bases, which can be done using TrimGalore or Trimmomatic. Subsequent alignment and DNA methylation calling is performed using bisulfite aligners such as Bismark[6] or BSMAP/RRBSMAP[7]. The resulting DNA methylation data can be visualized with genome browsers such as the UCSC Genome Browser[8] or the Integrative Genomics Viewer[9] (Fig. 2A-B). Further analysis is possible with software tools such as RnBeads[10], which provides a comprehensive pipeline for analyzing and interpreting DNA methylation data.

Sequencing data and results obtained with Diagenode Premium RRBS kit

Diagenode Premium RRBS kit generated DNA methylation profiles of high coverage and quality for all vertebrate species tested, as exemplified in Table 1 and Figure 2.

Table 1 shows typical data obtained when using human fresh-frozen and FFPE samples, rat, dog, and zebrafish samples. The genomic coverage depends on the species and exceeded 3.9 million for the human fresh-frozen sample, and the bisulfite conversion rate consistently exceeded 99% in all species. A mean coverage of 10-20 per CpG should be aimed for, although some applications with small DNA methylation differences or extensive tissue heterogeneity will profit from deeper sequencing. The expected library profiles depend on the distribution of MspI restriction sites in the genome and are specific to each species. Size distribution profiles of the final RRBS library facilitate excellent quality control prior to sequencing.

Table 1: *Examples of RRBS data generated with Diagenode's Premium RRBS Kit using samples from four different species.*

Species	No. of reads aligned (Alignment rate)	No. of unique CpGs covered	Bisulfite Conversion Rate	Mean coverage per CpG	Library profile
Human	21,787,346 (74%)	3,913,287	99.4%	14	
Human FFPE	27,290,117 (79 %)	2,525,053	99.7%	28	
Rat	27,499,313 (86 %)	1,663,104	99.0%	38	
Dog	11,377,427 (62 %)	3,315,166	99.4%	8	
Zebrafish	29,035,877 (83 %)	1,685,466	99.3%	48	

Illustrating typical results of the RRBS kit, Figure 2 shows differential DNA methylation around intergenic CpG islands in IGF2 and the unmethylated promoter region of GAPDH in two cell lines. The results were highly reproducible between replicates.

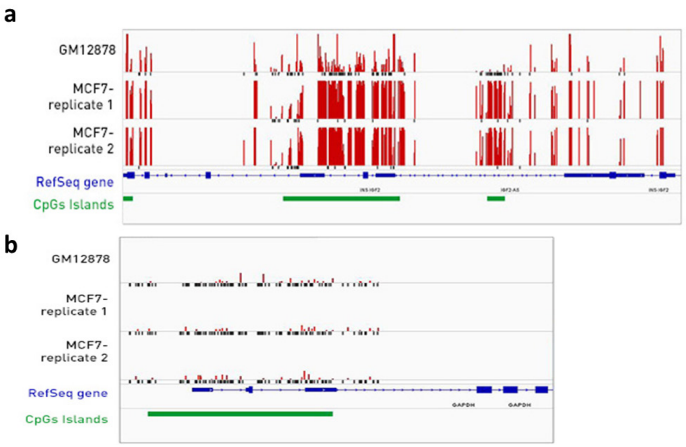


Figure 2: *DNA methylation profiles obtained with Diagenode Premium RRBS kit. Two human cell lines were analyzed: GM12878 and MCF7. The MCF7 cell line comprised two biological replicates. Each peak represents the DNA methylation percentage at one CpG. The methylat-ed CpGs are shown in red and the unmethylated CpGs in grey. Panel a displays the IGF2 (insulin-like growth factor 2) gene and panel b displays a CpG island in the GAPDH (glyceraldehyde-3-phosphate dehydrogenase) gene.*

Conclusion

Diagenode Premium RRBS technology is a highly-multiplexed solution for studying DNA methylation cost-efficiently in vertebrate samples including FFPE, clinical and low-input material. The kit contains all the reagents needed to transform genomic DNA into sequencing-ready RRBS libraries with a superior coverage of 4 million CpGs.

Link to the supplier's website

Please note that the RRBS technology is also available as a service at Diagenode. For more information on Diagenode kits and services for epigenetic studies please visit:
www.diagenode.com

References

1. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R (2005). **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.** *Nucleic Acids Res 33, 5868-5877.*

2. Gu, H. et al. **Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution.** *Nat. Methods 7, 133–136 (2010).*

3. Bock, C. et al. **Quantitative comparison of genome-wide DNA methylation mapping technologies.** *Nat. Biotechnol. 28, 1106–1114 (2010).*

4. Bock, C. **Analysing and interpreting DNA methylation data.** *Nat. Rev. Genet. 13, 705–719 (2012).*

5. Martin, M. **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet.journal 17, 10 (2011).*

6. Krueger, F. & Andrews, S. R. **Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.** *Bioinforma. Oxf. Engl. 27, 1571–1572 (2011).*

7. Xi Y, Bock C, Müller F, Sun D, Meissner A, Li W (2012). **RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing.** Bioinformatics 28, 430-432

8. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AF, Speir ML, Zweig AS, Haussler D, Kuhn RM, Kent WJ (2015). **The UCSC Genome Browser database: 2015 update.** *Nucleic Acids Res 43, D670-681.*

9. Robinson, J. T. et al. **Integrative genomics viewer.** *Nat. Biotechnol. 29, 24–26 (2011).*

10. Assenov Y, Müller F, Lutsik P, Walter J, Lengauer T, Bock C (2014). **Comprehensive analysis of DNA methylation data with RnBeads.** *Nat Methods 11, 1138-1140.*

Europe Diagenode sa / LIEGE SCIENCE PARK // Rue du Bois Saint-Jean, 3 // 4102 Seraing (Ougrée) // Belgium // Phone: (+32) 4 364 20 50 // E-mail: info@diagenode.com
USA Diagenode Inc. / 400 Morris Avenue, Suite 101 // Denville, NJ 07834 // USA // Phone: (+1) 862 209-4680 // Mail: info.na@diagenode.com