Impact of gDNA Integrity on the Outcome of DNA Methylation Studies

Application Note

Nucleic Acid Analysis

Authors

Emily Putnam, Keith Booher, and
Xueguang Sun
Zymo Research Corporation,
Irvine, CA, USA

ZYMO RESEARCH

The Beauty of Science is to Make Things Simple®

THE
Epigenetics
COMPANY™

Introduction

A growing area of epigenetics research focuses on changes in gene expression that occur without altering the primary sequence, exclusively based on DNA, RNA, or protein modifications. Epigenetic modifications have been shown to change based on nutrition and environmental factors, and have been implicated in human diseases such as cancer and aging. In some cases, epigenetic changes are heritable. Several regulators involved in epigenetics are known, including:

- DNA methylation and hydroxymethylation

- Histone modification and chromatin remodeling

- Small and large regulatory RNAs

- RNA methylation

DNA methylation is the most widely studied modification involved in epigenetics. In mammalian cells, DNA methylation mainly involves the transfer of a methyl group from S-adenosyl methionine to the carbon 5 position of a cytosine residue to produce 5-methylcytosine. DNA methylation is mainly implicated in the repression of transcriptional activity. DNA methylation studies are performed to discover new disease markers, drug targets, and to further explain gene expression data, for example, understanding phenotypes that classical genetic studies are unable to resolve.

This Application Note focuses on 5-methylcytosine (5-mC) DNA methylation and its detection. By far, the most commonly used method for DNA methylation analysis is bisulfite sequencing due to its high resolution detection when combined with sequencing. Advances in next generation sequencing made it possible to perform bisulfite sequencing at a genome-wide scale. The DNA methylation analysis based on reduced representation bisulfite sequencing, a high-throughput technique, allows analysis on a single nucleotide level[1]. This method is based on the enrichment of genome regions with high CpG content (sites within the genome where a cytosine is next to a guanine) using a combination of restriction enzymes and bisulfite sequencing, and is applicable to any species with a reference genome.

Agilent Technologies

In addition to the genome-wide bisulfite sequencing, targeted studies are performed to validate genome-wide discoveries with a larger group of samples[2]. The next generation targeted bisulfite sequencing approach allows high throughput, and provides single-base pair resolution.

It is known that the quality of the DNA starting material is crucial for next generation sequencing[3]. The experiments described in this Application Note were carried out to verify that the gDNA integrity of the starting material has an impact on DNA methylation analysis based on next generation bisulfite sequencing. Therefore, the initial gDNA quality, as indicated by a numerical measure of DNA integrity (DIN)[4], was determined with the Agilent 2200 TapeStation system and the Agilent Genomic DNA ScreenTape assay. DIN was correlated to the next generation bisulfite sequencing results, to determine if the initial DNA integrity had an effect on the quality of the DNA methylation studies.

Experimental

Material

Restriction enzymes, TaqαI and MspI, were purchased from New England Biolabs (Ipswich, MA, USA). FFPE DNA MiniPrep, Quick-DNA Universal, DNA Clean & Concentrator-5, ZR-96 DNA Clean & Concentrator, Zymoclean Gel DNA Recovery, EZ DNA Methylation-Lightning kits, and Rosefinch software were from Zymo Research Corp. (Irvine, CA, USA).

NuSieve 3:1 Agarose Lonza was obtained from (Basel, Switzerland). The HiSeq system and the MiSeq Reagent Micro Kit v2 were from Illumina, Inc. (San Diego, CA, USA). The Access Array System was from Fluidigm, Inc. (South San Fransisco, CA, USA). The Agilent 2200 TapeStation system (p/n G2965AA) with Agilent TapeStation Analysis software, Agilent Genomic DNA ScreenTape (p/n 5067-5365), and Agilent Genomic DNA Reagents (p/n 5067-5366), D1000 ScreenTape (p/n 5067- 5582) and D1000 Reagents (p/n 5067-5583) were obtained from Agilent Technologies, Inc. (Santa Clara, CA, USA).

Genome-wide DNA methylation analysis

Human brain tissue from the same donor was used for the genome-wide methylation analysis. Genomic DNA was extracted from fresh frozen tissue using the Quick-DNA Universal kit, and was used as a control for the experiments. Four tissue samples were formalin-fixed, paraffin embedded (FFPE) using varying lengths of paraffin treatment times (30 to 120 minutes). Following FFPE treatment, gDNA from the four samples was extracted using the ZR FFPE DNA MiniPrep kit. A 200 ng amount of extracted gDNA was sequentially digested using 60 units of TaqαI and 30 units of MspI. Next, the DNA fragments were ligated to pre-annealed adapters containing 5′-methyl-cytosine instead of cytosine. The adaptor-ligated fragments of 150 to 250 bp and 250 to 350 bp were recovered from a 2.5 % agarose gel. The fragments were then bisulfite-treated. Preparative-scale PCR was performed, and the resulting products were purified for HiSeq sequencing.

Targeted DNA methylation analysis

Genomic DNA was extracted from 282 mammalian samples, and bisulfite converted. Then, targeted amplification was performed using unbiased, bisulfite-specific primers to cover CpG sites in the specified regions of interest. Primers were created with a dedicated bisulfite-converted DNA-specific primer design tool (proprietary Zymo Research software). Multiplex amplification of all samples using region of interest specific primer pairs and the Access Array system was performed. The amplification conditions were chosen for PCR amplicons ideally ranging from 100 bp to 300 bp. The resulting amplicons were pooled for harvesting and subsequent bar-coding. Samples were purified, then prepared for massively parallel sequencing and paired-end sequencing. Sequence reads were identified using standard base-calling software, then analyzed using proprietary Zymo Research software. CpG site coverage was calculated as a percentage of CpG site covered by sequencing reads versus total number of CpG sites for the designed assays.

DNA analysis with the 2200 TapeStation system

DNA analysis was performed using the 2200 TapeStation system in combination with the Genomic DNA or D1000 ScreenTape assay, according to the manufacturer's instructions[5,6].

Results and Discussion

Genome-wide DNA methylation analysis

Human brain tissue from the same donor was used for the genome-wide DNA methylation analysis. Genomic DNA was extracted from fresh frozen tissue and was used as a control for the following experiments. In addition, four tissue samples were subjected to FFPE using varying lengths of paraffin treatment times to generate varying levels of DNA degradation. The gDNA integrity of this starting material was determined using the Genomic DNA ScreenTape assay and the 2200 TapeStation system. Figure 1 shows the obtained gel image, the DIN values indicating the gDNA integrity, and the DNA concentrations. The numerical assessment ranged from 1–10. A high DIN value indicates highly intact gDNA, and a low DIN degraded gDNA[4].

DNA extraction from FFPE samples has proven to be especially challenging[7]. The most common issues are cross-linking and fragmentation of DNA during tissue processing and storage, affecting the yield and quality of gDNA obtained. Figure 1 demonstrates that, as expected, the gDNA yield and quality obtained for FFPE samples was lower, compared to the fresh frozen sample. The obtained DIN values for the FFPE samples ranged from 4.0 to 7.1. The gDNA extracted from the fresh frozen sample had a relative high gDNA quality with DIN 8.3, and the highest DNA concentration.

Following this initial quality control step, the five gDNA samples were subjected to the workflow for genome-wide DNA methylation analysis based on reduced representation bisulfite sequencing, as outlined in Figure 2.

FFPE

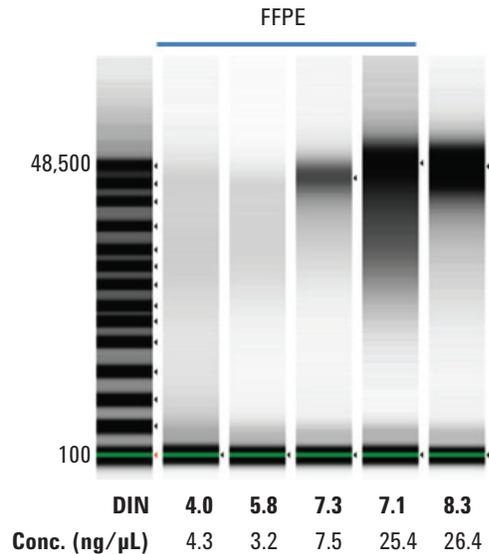|					
48,500					
100					
DIN	**4.0**	**5.8**	**7.3**	**7.1**	**8.3**
Conc. (ng/µL)	4.3	3.2	7.5	25.4	26.4

Figure 1. Analysis of gDNA extracted from four FFPE samples (lanes 1 to 4) and one fresh frozen sample (last lane) from human brain tissue. gDNA integrity (DIN) and concentration were determined using the Agilent 2200 TapeStation system and the Agilent Genomic DNA ScreenTape assay.

gDNA preparation

Restriction enzyme digest

Adapter ligation

Size selection
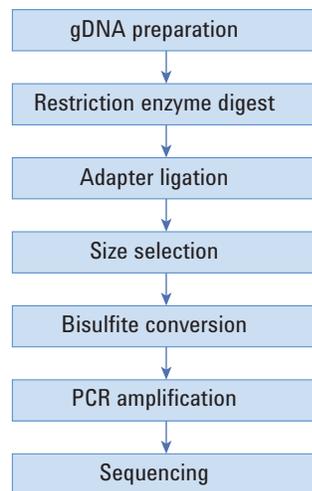
Bisulfite conversion

PCR amplification

Sequencing

Figure 2. Genome-wide DNA methylation analysis workflow based on reduced representation bisulfite sequencing.

The gDNA samples were digested using a combination of TaqαI (T|CGA) and MspI (C|CGG). Neither restriction enzyme is sensitive to CpG methylation. Fragments of different sizes with CpG at the end were generated. Next, methylated sequence adapters were ligated to the DNA fragments. Then, a size selection of fragments was performed to enrich the informative sequences. The following bisulfite treatment of the size-selected DNA fragments resulted in the conversion of unmethylated cytosine residues to uracil. 5-Methylcytosine is protected from conversion. The bisulfite converted DNA was then amplified using PCR, and a barcode sequence was added for sample identification post-sequencing. Following the PCR amplification step, the obtained samples were analyzed using the 2200 TapeStation system and the D1000 ScreenTape assay.

In all PCR reactions, DNA of approximately 200 bp was detected with different concentrations. Figure 3 demonstrates that the DNA library yield after PCR amplification varied depending on the DIN value of the initial gDNA sample. A low DNA integrity of the starting material resulted in a low DNA yield after amplification.

Finally, the obtained DNA libraries from the five samples were subjected to next generation sequencing. Due to the bisulfite conversion, the methylated cytosine residues are read as cytosine. The nonmethylated cytosines are converted to uracil, and read as thymine. Special software was used for alignment and data analysis. The alignment to the reference genome allowed the software to identify base pairs within the genome that were methylated. To demonstrate the effect of DIN integrity on the quality of the sequencing results, the number of covered CpG sites was determined and compared to the DNA integrity of the initial gDNA samples (Figure 4).

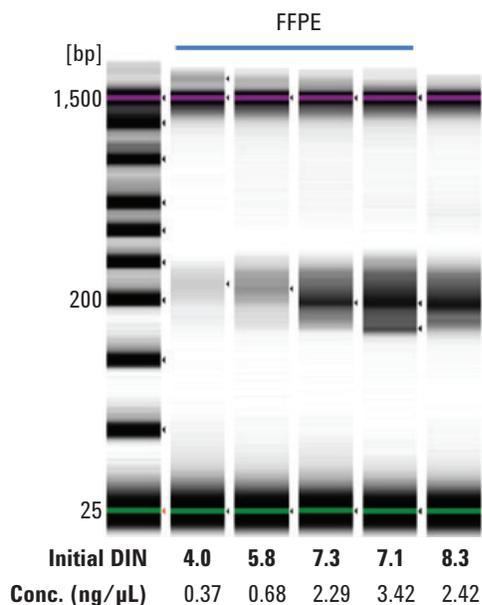| Initial DIN	4.0	5.8	7.3	7.1	8.3
Conc. (ng/µL)	0.37	0.68	2.29	3.42	2.42

Figure 3. Analysis of DNA samples after PCR amplification from four FFPE samples (lanes 1 to 4) and one fresh frozen sample (last lane) from human brain tissue. The analysis was performed using the Agilent 2200 TapeStation system and the Agilent D1000 ScreenTape assay. The table shows DIN values of the initial gDNA samples, and the determined DNA concentration.
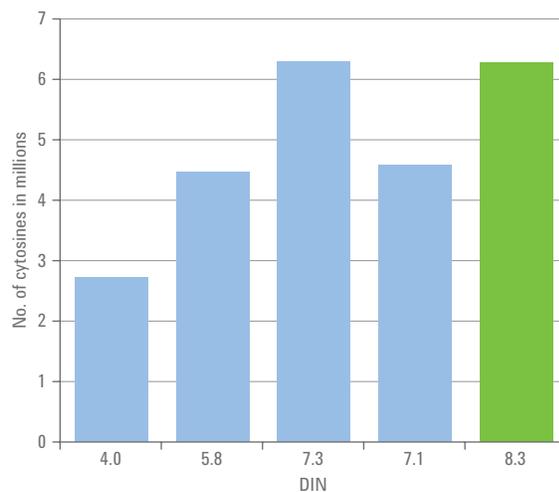

Figure 4. Correlation of the gDNA integrity and the covered CpG sites from four FFPE samples (blue) and one fresh frozen sample (green) from human brain tissue.

Figure 4 demonstrates that the total number of covered CpG sites varied, depending on initial DIN value of the gDNA samples. For the fresh frozen sample, a CpG coverage of slightly above six million was obtained; a similar value was obtained for the FFPE sample with a DIN of 7.3. It can be observed that, with decreasing DNA integrity of the gDNA starting material, the CpG coverage decreases.

Hexbin plots were used for the analysis of the obtained large data sets. Figure 5 compares the methylation levels for overlapping CpG sites for the four FFPE samples with DIN ranging from 4 to 7.3 to the fresh frozen sample with a DIN value of 8.

The number of overlapping CpG sites comparing the gDNA obtained from the FFPE sample with DIN 4 as starting material, with the DNA extracted from the fresh frozen tissue (DIN 8) was low

(n = 660,291) compared to the samples with higher DIN values. The number of overlapping CpG sites increased with increasing DNA integrity. Even though the number of CpG sites overlapping was low, the correlated methylation levels were still fairly high. The correlation was good for all samples (R > 0.9). Figure 5 clearly demonstrates that the initial DNA integrity can significantly impact the success of genome-wide DNA methylation studies.
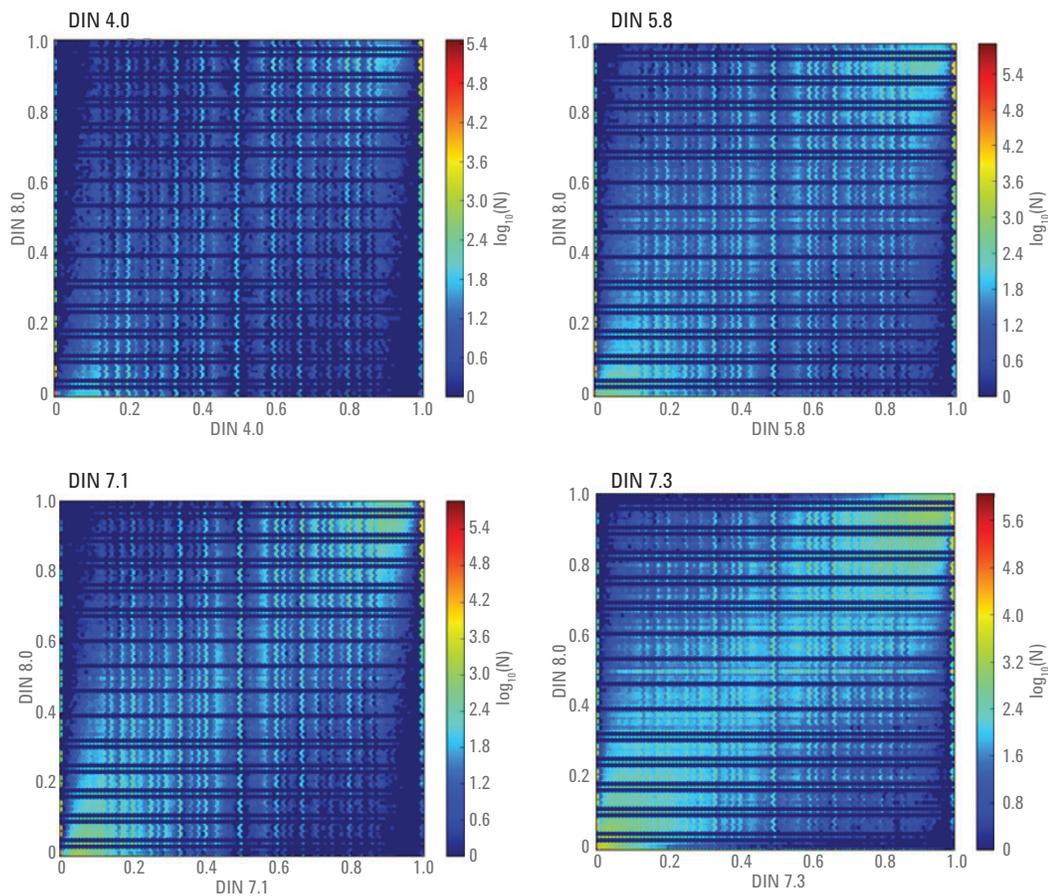

Figure 5. Comparison of the DNA methylation levels for overlapping CpG sites for the four FFPE samples with DIN ranging from 4 to 7.3 to the fresh frozen sample with a DIN value of 8 using Hexbin plots. The correlation between two samples as indicated by the Pearson's R correlation coefficient is displayed at the top. Colors, at the indicated location on the plot, represent point densities, with red indicating points comprised from many values, and blue showing points from only a few values.

Targeted DNA methylation analysis

In addition to the genome-wide bisulfite sequencing, targeted studies are performed to validate genome-wide discoveries with a larger group of samples. Figure 6 shows the targeted bisulfite next generation sequencing workflow.

For this Application Note, gDNA was extracted from 382 mammalian samples, and bisulfite converted. Then, targeted amplification was performed using unbiased, bisulfite-specific primers to cover CpG sites in the specified regions of interest. Due to the DNA degradation induced by the harsh conditions of the bisulfite conversion, multiplex amplification of all samples with conditions for short PCR amplicons ideally ranging from 100 bp to 300 bp was performed. The resulting amplicons were bar-coded using PCR to enable sample identification post-sequencing. Parallel and paired-end sequencing was applied.

The CpG site coverage was calculated as a percentage of CpG site covered by sequencing reads versus the total number of CpG sites for the designed assays. The total number of covered CpG sites was compared to the DNA integrity determined for the initial gDNA starting material. Figure 7 shows the comparison of the determined DIN with the percentage of total CpG sites covered in targeted bisulfite sequencing assays.

For most of the samples, the higher the gDNA integrity, the higher the obtained CpG coverage. However, a few gDNA samples with high DIN resulted in low CpG coverage. This is probably due to additional alterations of the gDNA during extraction from the FFPE tissue. It is known that formalin causes cross-linking of nucleic acids and proteins, increasing the susceptibility of DNA to mechanical stress and decreasing the accessibility to enzymes. In addition, formalin may be oxidized to formic acid, which causes DNA depurination and DNA strand breaks. Therefore, the fixation conditions can significantly impact the quality of the extracted DNA[7].
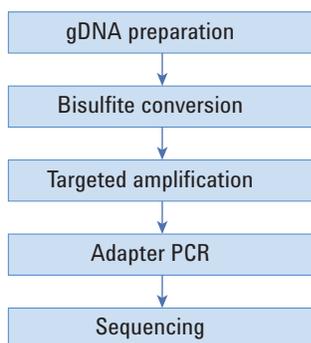

Figure 6. DNA methylation analysis workflow based on targeted bisulfite next generation sequencing.
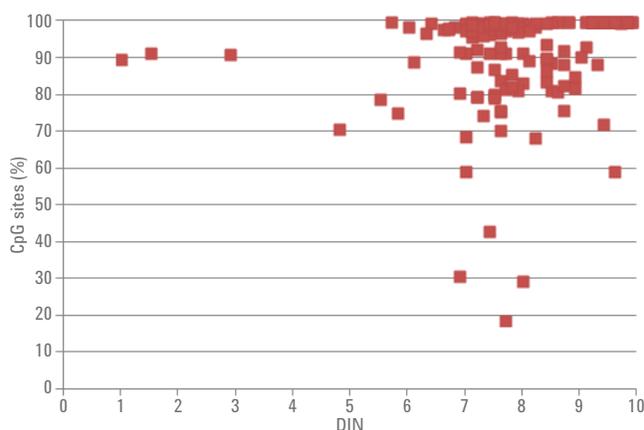

Figure 7. Correlation of the gDNA integrity (DIN) and the covered CpG sites obtained using targeted next generation bisulfite sequencing.

For some gDNA samples with low DIN, a high CpG coverage was achieved by using a higher amount of gDNA starting material, or by increasing the sequencing depth for each of the amplicons.

The gDNA integrity of the starting material can significantly impact the outcome of targeted DNA methylation studies. A high DIN integrity ensures a better assay success with the CpG coverage. Using the Genomic DNA ScreenTape assay and the 2200 TapeStation system for quality control of the gDNA starting material allows optimizing of the experimental conditions to improve the sequencing results. For example, increasing the amount of input DNA or increasing the sequencing depth for each of the amplicons might help to overcome low quality starting material.

Conclusion

The Agilent 2200 TapeStation system and the Agilent Genomic DNA ScreenTape assay provide an optimal tool for quality control of gDNA samples extracted from FFPE tissue for DNA methylation analysis.

The automatically determined DIN correlates with the CpG coverage, a key quality metric. The DIN can be used as quality criteria to determine how to handle individual gDNA samples for downstream workflows, and to ensure successful DNA methylation analysis.

References

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

2. Ashktorab, H; *et al.* DNA methylome profiling identifies novel methylated genes in African American patients with colorectal neoplasia. *Epigenetics* **2014**, *9(4)*, pp 503-512.

3. DNA quality control of formalin-fixed paraffin-embedded and fresh-frozen tissues prior to target-enrichment and next generation sequencing, *Agilent Technologies*, publication number 5991-0483EN, **2012**.

4. DNA Integrity Number (DIN) with the Agilent 2200 TapeStation System & Genomic DNA ScreenTape, *Agilent Technologies*, publication number G5991-5258EN, **2014**.

5. Agilent Genomic DNA ScreenTape System Quick Guide, *Agilent Technologies*, publication number G2964-90040 rev.C, **2014**.

6. Agilent D1000 ScreenTape System Quick Guide *Agilent Technologies*, publication number G2964-90032 Rev. C, **2014**.

7. Tang, W.; *et al.* DNA Extraction from Formalin-Fixed, Paraffin-Embedded Tissue, *Cold Spring Harbor Protocols* **2009**, doi:10.1101/pdb.prot5138.

Agilent Technologies