White Paper

Agilent
Trusted Answers

Quantifying Cancer-Specific DNA Methylation Using a Novel Metric

Agilent Avida Methylation Index (MI)

Introduction

The assessment of epigenetic modifications, particularly DNA methylation changes, is a powerful approach for identifying disruptions in physiological processes such as cancer and aging. One key strategy involves analyzing differentially methylated regions (DMRs), which are genomic regions that exhibit methylation differences between diseased and normal samples. In cancer research, profiling DMRs from tumor and non-tumor samples can offer a sensitive and precise method for discovering biomarkers, elucidating mechanisms of epigenetic dysregulation, and distinguishing cancer subtypes.

Traditional tools for analyzing DNA methylation data typically rely on average methylation metrics across DNA molecules within a biological sample. These metrics are derived from: (1) methylation fractions at individual CpG sites, calculated by dividing the number of methylated cytosines by the total number of molecules covering the CpG site; and (2) regional methylation density, represented by the average or a variation thereof across all CpG methylation fractions within a defined genomic region of interest. To capture more nuanced methylation patterns, methylation haplotype load (MHL) quantifies methylation across a mixture of DNA molecules, incorporating weights based on the length of methylated haplotype blocks within a DMR.[1]

Relying on average-based measurements presents challenges when analyzing cell-free DNA (cfDNA) from liquid biopsies, particularly due to the low abundance of cfDNA originating from affected tissues such as tumors. To address this limitation, we introduce a novel DNA methylation detection and analysis method that moves beyond traditional averaging approaches. Agilent Avida technology offers a next-generation sequencing (NGS) approach designed to detect differential methylation changes from targeted, cancer-relevant genomic regions. We describe the strategy to target DMRs and quantify methylation status using an innovative, Agilent proprietary metric, Avida Methylation Index (MI). The Avida MI is a model-based score that reflects the methylation profile of known cancer-associated DMRs, enabling robust and scalable analysis for translational cancer research applications.

Methods

The design of an Avida panel targeting DMRs based on data mining from public databases, targeted methyl sequencing on a set of tumor and normal samples, and the development of a predefined model on the training set to enable calculation of the Avida MI metric for a sample are described here.

Biomarker discovery: Defining cancer-relevant DMRs for use with tissue and plasma samples

To determine DMRs that could be used to profile cancer samples derived from tissue or plasma, *in silico* and wet lab experiments were performed (Figure 1). We searched for potential cancer biomarkers by data mining public databases containing methylation data from arrays and sequencing. The selected DMR candidates were evaluated with confirmatory methyl sequencing experiments using a number of tissue samples (normal and tumor) and cfDNA from pooled healthy plasma samples. The final target region covered approximately 3,400 cancer-relevant DMRs and is available as a catalog panel, Avida Methyl 3400 DMR Cancer panel (p/n 5280-0059), for use with the Avida Methyl Reagent kit (p/n G9429A) in the Avida Methyl workflow.

Methylation data analysis pipeline

The raw sequencing data in FASTQ format was used as input for the methylation analysis pipeline. Briefly, the pipeline steps include:

1. Sequencing reads are aligned to the reference genomes, taking into consideration C to T and G to A base conversions.
2. The aligned reads are filtered and trimmed based on quality metrics.
3. Filtered reads are analyzed to remove duplicates based on strand and genomic location.
4. Quality control (QC) metrics, such as raw coverage, deduplicated coverage, and alignment percentage, are collected. The QC metrics can be used as sample pass or fail criteria.
5. CpG and CHX (where H is a non-G base and X can be any one of the four bases) methylation information is extracted from the deduplicated reads and methylation profiling of CpGs in the targeted regions is conducted. A filtering technique to exclude noise introduced by the C>T conversion process is used to enhance the robustness of the data used for downstream analysis.
6. Proprietary computational steps are implemented to reduce methylation noise and artifacts in the sample.
7. The methylation profile for each molecule of a sample is used as input into a "pre-defined" computational model to calculate the Avida MI for the sample.

Sample methylation profiling and assay methylation noise removal

As an output of the methylation analysis pipeline, each unique read contains CpG methylation information from a unique DNA molecule captured by the assay. Two metrics were collected from each read: (1) N: the total number of CpGs in the read, and (2) M: the number of methylated CpGs in the read. Based on these metrics, a third metric was calculated as f = M/N, the fraction of CpGs that are methylated in the read. The data pairs (N, f) were collected for each of the unique molecules on all DMRs in the assay. A scatter plot showing f (y-axis) vs N (x-axis) was generated for a DMR, with each unique read in the DMR shown as a dot in the plot (see Figure 3).

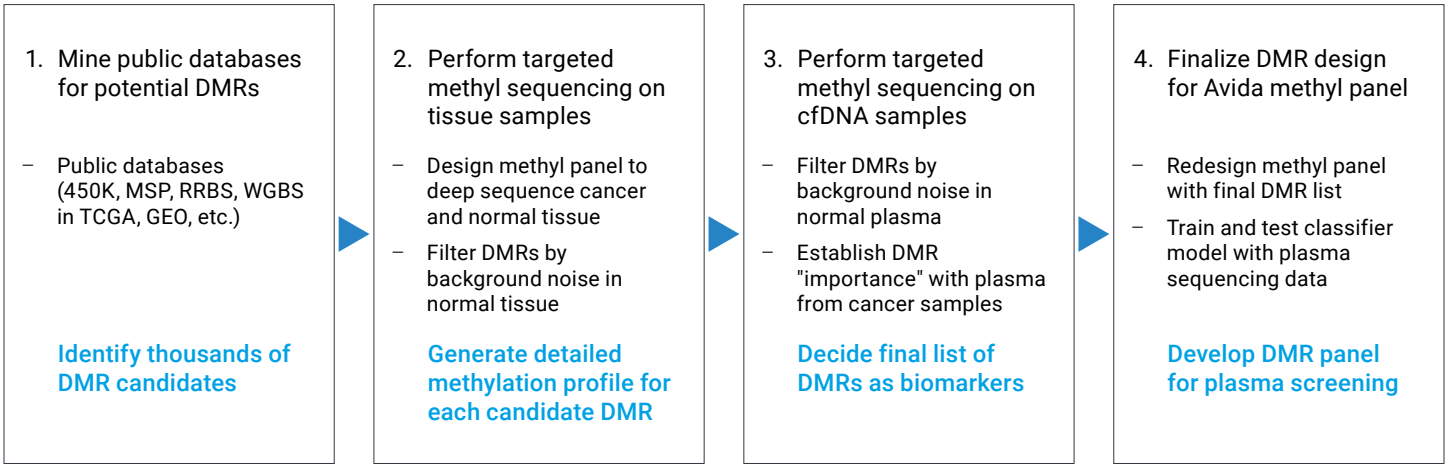

Figure 1. Schematic to identify potential cancer methylation biomarkers that are compatible with tissue and plasma samples.

Reference material serial dilution testing

To demonstrate the usage and performance of the Avida MI, a serial dilution of hypermethylated reference DNA was spiked into hypomethylated baseline cfDNA at 0.003%, 0.03%, 0.3%, and 3%, with 0% as the baseline. Samples were prepared with the Avida Methyl Reagent kit and targets were captured with the Avida Methyl 3400 DMR Cancer panel following the protocol (Avida Methyl Reagent Kits Protocol). Ten nanograms of cfDNA input was used for each sample. Three replicates were generated for each concentration.

The libraries were sequenced, and the analysis steps described above were applied to the FASTQ data for each sample.

Results

Assessing differential methylation across DMRs

Clear differences in methylation levels were observed between representative sets of tumor (gDNA from tissue) and normal (gDNA from tissue, cfDNA from plasma) samples, all enriched for DMRs using a subset of the Avida Methyl 3400 DMR Cancer panel[2,3] (Figure 2).

Assessing differential methylation within a single DMR

The methylation profiles within a single DMR, based on a scatter plot of the methylation fraction for each CpG site in the DMR, show a clear difference in the methylation levels between normal and cancer tissue samples (Figure 3). Normal tissue samples show many hypomethylated molecules, along with a range of partially methylated molecules (Figure 3A). By contrast, cancer tissue samples show a high number of hypermethylated molecules (Figure 3B).

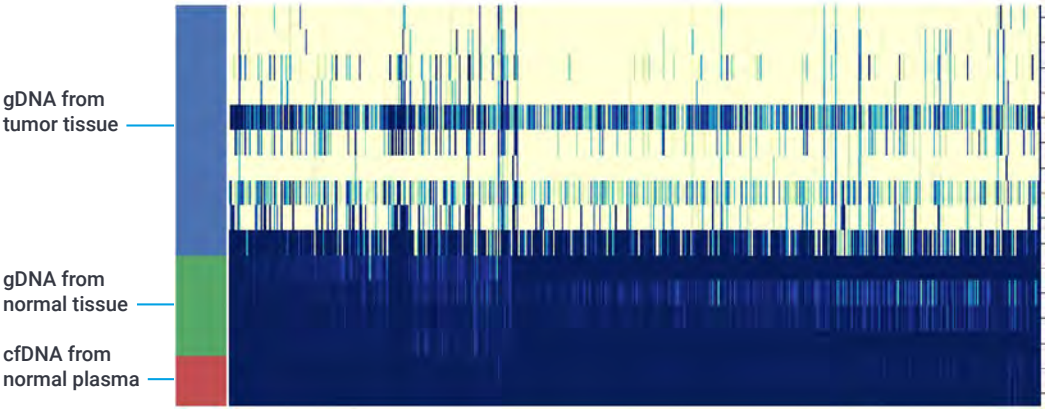

Figure 2. Methylation signal intensity in tissue and cfDNA samples targeted with a subset of the Agilent Avida Methyl 3400 DMR Cancer panel. Libraries were prepared using the Agilent Avida Methyl Reagent kit. The rows of the heatmap correspond to samples and the columns correspond to the targeted DMRs.[2,3] Very low methylation levels are shown as dark blue and high methylation levels for targeted DMR are shown as yellow.

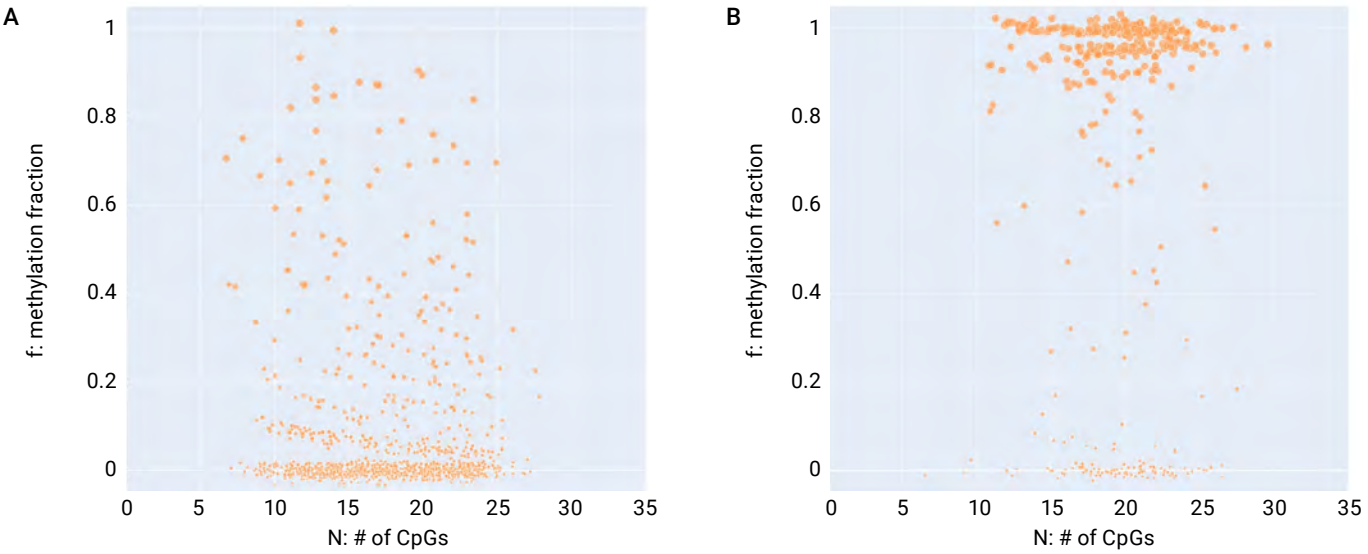

Figure 3. Methylation patterns within a single DMR in gDNA derived from normal and cancer tissue samples. Each dot represents a molecule obtained for a single DMR for each tissue.

Use case: analytical sensitivity of MI tested with reference material

The Avida MI can be used to differentiate samples with very small changes in methylation among samples using spike-ins of hypermethylated reference DNA, as demonstrated in the use case described here.

For comparison, calculating the average methylation level across all CpG sites within the panel, without using MI, failed to reveal subtle differences among various spike-in concentrations of hypermethylated reference DNA tested. This lack of discernibility for concentrations from 0% to 0.3% arises from incomplete methylation conversion or biological noise (Table 1, Figure 4).

By contrast, the Avida MI of the spike-ins are tightly clustered for each spike-in concentration and show clear cluster separation between different concentrations, with an analytical sensitivity down to 0.003% (Table 1, Figure 4).

Table 1. Average methylation levels do not correlate with the reference spike-in concentration, whereas the Avida Methylation Indexes show good correlation.

Spike-in concentration (%)	Technical replicate	Bisulfite conversion rate (%)	Average CpG methylation level (%)	Avida Methylation Index
0.000	1	99.2	0.9	0
	2	99.1	1 .0	0
	3	99.2	0.9	0
0.003	1	99.4	0.7	0.005
	2	99.4	0.7	0.004
	3	99.4	0.7	0.004
0.030	1	99.4	0.7	0.080
	2	99.4	0.7	0.067
	3	99.4	0.6	0.103
0.300	1	99.4	0.8	0.799
	2	99.4	0.8	0.907
	3	99.4	0.8	0.861
3.000	1	99.4	2.3	8.069
	2	99.4	2.3	8.373
	3	99.4	2.3	8.047

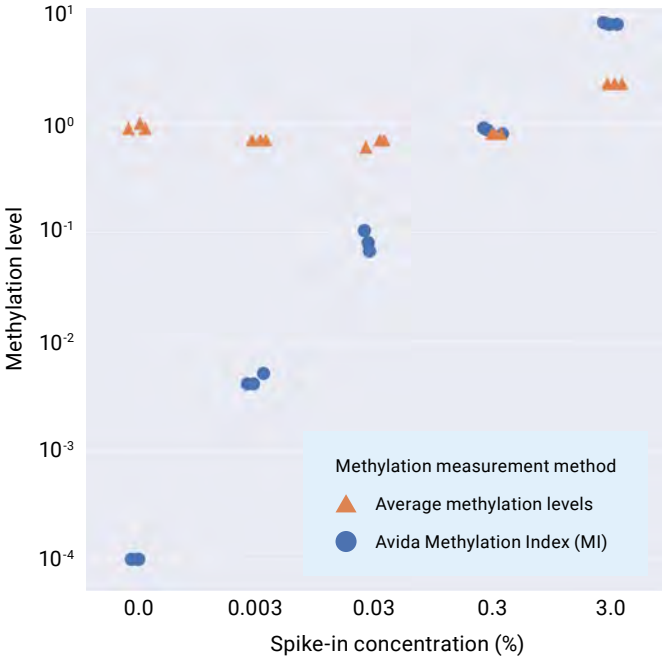

Figure 4. Methylation levels measured for a titration series of hypermethylated reference DNA spike-in samples comparing measurements of average methylation levels to Avida Methylation Indexes. The graph of the average CpG methylation levels (%) (orange triangles), also shown in Table 1, indicated that this averaging method failed to accurately differentiate methylation levels for samples with spike-in concentrations at 0.3% and below. By contrast, the Avida Methylation Indexes (blue dots) showed clear separation and correlation across the entire titration series, and consistency for each spike-in concentration.

Conclusion

Here we outline the biomarker discovery process to develop the Avida Methyl 3400 DMR Cancer panel, its application in generating a predefined analytical model, and the subsequent calculation of the Avida Methylation Index (MI).

The Avida MI is a highly sensitive, single-value metric derived from a computational pipeline that includes proprietary noise and artifact reduction techniques. It is calculated using a predefined model applied to individual samples, enabling precise quantification of methylation patterns.

By combining the robust differentiation capabilities of the Avida Methyl 3400 DMR Cancer Panel with the Avida MI score, researchers can detect hypermethylated DNA even at extremely low concentrations. This sensitivity makes the Avida MI a potentially valuable tool for translational oncology research, particularly in early detection and molecular profiling.

The descriptions and calculations described here offer an example of how researchers can generate MI distributions across their own sample cohorts. These distributions can serve as the foundation for developing predictive or classification models tailored to specific translational cancer research applications, leveraging the full potential of the Avida Methyl 3400 DMR Cancer panel.

References

1. Moss, J.; Magenheim, J.; Neiman, D.; Zemmour, H.; Loyfer, N.; Korach, A.; Samet, Y.; Konikoff, F. M.; Goren, A.; Glaser, B.; et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **2018**, *9*, 5068
2. Zhao, G. Q.; Bao, Y.; Wang, H.; Vu, P.; Patel, S. P.; Lin, S. Detection of Colorectal Cancer (CRC) through Novel Dual Targeted Sequencing Solution that Enables Methylation and Genomic Alteration Analysis of Cell-Free DNA Without Sample Splitting. Poster presented at the American Society of Clinical Oncology Annual Meeting, Virtual, June 2021.
3. Zhao, G. Q.; Bao, Y.; Wang, H.; Vu, P.; Patel, S. P.; Lin, S. A novel NGS kit solution enables multi-cancer early detection using circulating cell-free DNA based methylation analysis. Poster presented at the American Society of Clinical Oncology Annual Meeting, Chicago, IL, June 2022.
4. Guo, S.; Diep, D.; Plongthongkum, N.; Fung, H. L.; Zhang, K.; Zhang, K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **2017**, *9*, 635.

Appendix

Definitions

Sample – Biological material from which DNA molecules are extracted, such as plasma, blood, tissue (fresh, frozen, fixed), and cultured cells. Sample can also refer to reference material available from commercial vendors, such as from Seracare or Horizon.

DNA molecule – Examples are cell-free DNA (cfDNA) isolated from plasma/serum or genomic DNA (gDNA) isolated from cultured cells, PBMCs (from blood) and tissue. Commercial vendors also offer reference cfDNA and gDNA for research.

cfDNA – Cell-free DNA, also known as circulating free DNA (cfDNA), are degraded DNA fragments that are released into body fluids such as blood plasma, urine, cerebrospinal fluid. The term cfDNA can be used to describe various forms of DNA freely circulating in body fluids, including circulating tumor DNA (ctDNA).

CpG – CpG refers to a specific sequence of nucleotides in DNA where a cytosine nucleotide is followed by a guanine nucleotide along its 5' –> 3' direction. In the process of DNA methylation, a methyl group is added to the cytosine in a CpG site, forming 5-methylcytosine. CpG methylation is a key aspect of epigenetic regulation and is critical for cellular function and identity.

DMR – Differentially methylated regions (DMRs) are genomic regions with different DNA methylation status across different biological samples and regarded as possible functional regions involved in gene transcriptional regulation. The biological samples can be different cells/tissues within the same individual, the same cell/tissue at different times, or cells/tissues from different individuals. DMRs play a crucial role in gene regulation, development, and disease.

MI – The Avida Methylation Index (MI) is a single, numeric metric which is a measurement of the methylation state of a sample that has been enriched for DMRs using the Avida Methyl 3400 DMR Cancer panel.

MHL – Methylation Haplotype Load (MHL) is a metric proposed to quantify the level of coordinated methylation.[4] It is defined in the context of methylation haplotypes, which are stretches of consecutively methylated CpGs. MHL is used to perform tissue-specific methylation analysis at co-methylated CpGs block level.

For more information, please visit
www.agilent.com/

PR7003-469

Published in the USA, November 05, 2025
5994-8766EN

Agilent
Trusted Answers