TECHNICAL NOTE

High-throughput Infinium™ methylation array QC using DRAGEN™ Array Methylation QC software

illumina®

Introduction

For more than a decade, accurate and scalable Infinium methylation BeadChips have facilitated groundbreaking discoveries on the role of epigenetic mechanisms in human health and disease.[1] Illumina offers multiple BeadChip options for epigenetic studies, including the Infinium Methylation Screening Array-48 and the Infinium MethylationEPIC BeadChip. The Infinium Methylation Screening Array-48 features broad coverage of known methylation associations for human traits, disease phenotypes, exposures, aging, and large-scale epigenetic analyses,[2] while the Infinium MethylationEPIC BeadChip offers genome-wide coverage of known methylation sites.[3]

To measure the proportions of methylated and unmethylated cytosine bases using Infinium methylation BeadChips, the input DNA must first undergo bisulfite conversion where unmethylated cytosine bases are converted to thymine and methylated bases are retained as cytosine. The bisulfite-treated DNA is amplified and loaded onto an Infinium methylation array for the Infinium methylation assay and scanned on the iScan™ System.

Quality control (QC) analysis takes place after scanning and makes sure that the BeadChip data are measuring actual methylation differences and are not the result of experimental issues. The cloud-based DRAGEN Array Methylation QC software delivers high-throughput, quantitative reporting of control metrics for the array. DRAGEN Array Methylation QC software offers multiple features to ensure data quality and accelerate data workflows, including:

- Adjustable thresholds to determine pass/fail status
- Informative data summary plots for a quick visual check of each analysis batch
- Automated determination of detection p-value, beta-values, and m-values from each methylation sample
- Convenient deployment on BaseSpace™ Sequence Hub, a cloud-based software with an intuitive interface for easy analysis kickoff

This technical note examines sample QC methods employed by DRAGEN Array Methylation QC software to determine data quality. Where applicable, these methods will be compared to the methods in GenomeStudio™ Methylation Module (Table 1).

Table 1: DRAGEN Array Methylation QC vs. GenomeStudio Methylation Module features

	DRAGEN Array Methylation QC	GenomeStudio Methylation Module
Deployment	Cloud-based Graphical user interface	Installation on local hardware Graphical user interface
QC capabilities	Algorithm-based analysis of 21 quantitative control metrics with adjustable thresholds Data summary plots Proportion of CG probes passing with an adjustable p-value threshold	Control plots (requires manual visual evaluation) Number of probes passing
Background normalization method	Noob	Background subtraction
Dye bias correction method	Nonlinear	Linear
Detection p-value method	Negative controls for null, log transformed signal	Negative controls for null

Quantitative QC metrics

DRAGEN Array Methylation QC software features automated analysis of control probes included in Infinium methylation BeadChip content. These control probes are used to compute 21 metrics that help users assess individual aspects of BeadChip processing and sample performance. Within the results, the software provides information to identify issues with specific workflow steps, such as bisulfite conversion, staining efficiency, and extension (Table 2). The sample-dependent controls evaluate sample quality and performance, while the sample-independent controls evaluate the quality of specific steps in the process flow.

The software can automatically identify samples with failing performance metrics. Failing samples tend to have low DNA input quantity and low beta-value correlation with the associated higher input replicates (Figure 1). The automated analysis is more efficient than existing BeadChip workflows that begin with a manual QC analysis, including the GenomeStudio Methylation Module workflow. The automation also removes variability associated with subjective human analysis.

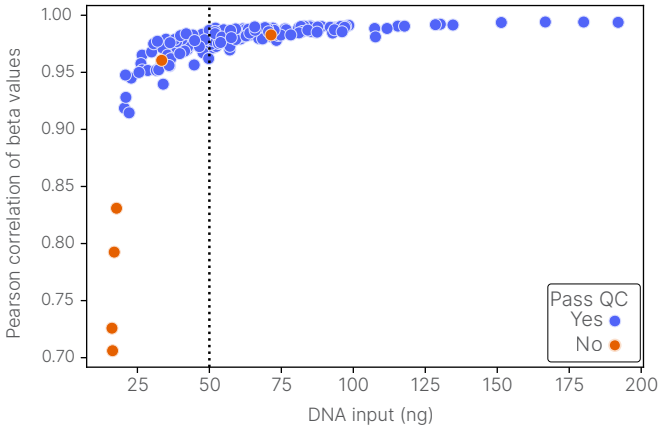


Figure 1: DRAGEN Array Methylation QC automatically identifies samples with poor performance—The X-axis indicates the smaller DNA input of a sample-replicate pair analyzed on the Infinium Methylation Screening Array, which has a required minimum DNA input of 50 ng. The Y-axis shows the Pearson correlation of the beta values for the paired replicate samples. Seven samples out of 239 pairs had one replicate failing QC. All samples failing were the lower DNA input sample in a replicate pair. A p-value threshold of 0.05 was applied before computing correlations. The summary data plot shown was generated for illustrative purposes using third-party software and is not part of DRAGEN Methylation QC analysis.

Table 2: Controls on Infinium methylation BeadChips analyzed by DRAGEN Array Methylation QC

Control category	Details	No. of metrics	Sample dependent
Restoration	Indicates whether FFPE restoration was successful	1	No
Staining	Efficiency of staining step	2	No
Extension	Extension efficiency of A,T,C, and G nucleotides	2	No
Hybridization	Efficiency of hybridization using synthetic targets instead of amplified sample DNA	2	No
Target removal	Efficiency of stripping step after extension reaction	2	No
Bisulfite conversion	Efficiency of bisulfite conversion (separate metrics for Infinium type I and type II probes)[4]	6	Yes
Specificity	Identify nonspecific extension (separate metrics for Infinium type I and type II probes)	4	Yes
Nonpolymorphic	Overall assay performance, from amplification to detection, involving query of a particular base in a nonpolymorphic region of the genome	2	Yes

Updated methylation QC pipeline

DRAGEN Array Methylation QC software produces dependable pass/fail detection of probes on Infinium methylation BeadChips due to improved background normalization, dye bias correction, and detection p-value calculation.

Background normalization

DRAGEN Array Methylation QC software uses normal-exponential convolution on out-of-band probes (Noob) for background detection. The Noob method models out-of-band signal from thousands of Infinium I probes across the BeadChip to achieve a more accurate background calculation than the background subtraction method used in the GenomeStudio Methylation Module that models a limited number of negative control probes. Noob has been shown to have lower probe-level variation across replicates than the background subtraction method, supporting more accurate signal estimates.[5,6]

Dye bias correction

Infinium methylation arrays use signals from Cy3 (green) and Cy5 (red) dyes to measure methylation levels at CpG sites. These dyes perform differently in the assay and a dye bias step correction must be performed before methylation levels can be accurately determined. DRAGEN Array Methylation QC uses nonlinear dye bias correction for differences in red/green channel signal that is an improvement over the linear dye bias correction used in the GenomeStudio Methylation Module.[7]

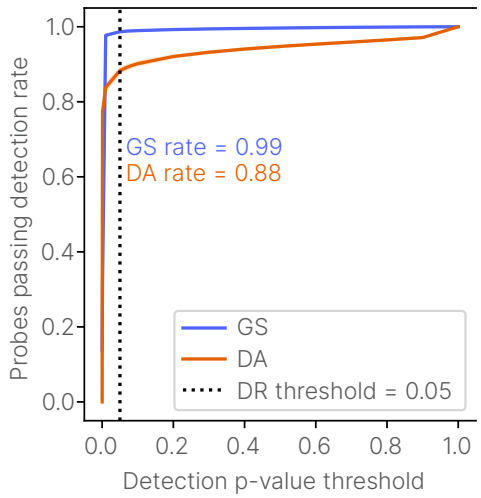
Detection p-value threshold

The detection p-value threshold indicates the probability that the signal is background noise. A cutoff of $p < 0.05$ is usually recommended for detection p-values. For calculating detection p-values, DRAGEN Array Methylation QC models the null distribution by fitting a normal distribution to the log of the signal intensity of the negative control probes.

DRAGEN Array Methylation QC controls false positive rate

DRAGEN Array Methylation QC software calculates background normalization, dye bias correction, and detection p-value differently in comparison to the GenomeStudio Methylation Module, leading to differences in detection p-values. DRAGEN Array Methylation QC analysis results in lower probes passing detection p-value rates (Figure 2), with the resulting data being higher quality with lower false positive results.

A. Infinium Methylation Screening Array-48



B. Infinium MethylationEPIC v2.0

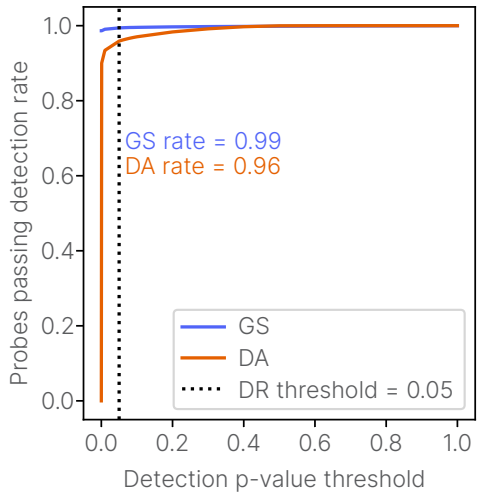


Figure 2: Probes passing detection rates using DRAGEN Array Methylation QC (DA) software vs. GenomeStudio Methylation Module (GS)— Probes passing detection p-value rate (DR) threshold is set at the recommended $p \leq 0.05$ for this analysis. Analysis is shown for DNA isolated from HeLa cell cultures and run on (A) Infinium Methylation Screening Array-48 and (B) Infinium MethylationEPIC v2.0.

To demonstrate the improvements in DRAGEN Array Methylation QC analysis, we examined the rate of Y-chromosome probes passing detection using two different sample types at two different input DNA amounts from four female subjects. Using data run on the Infinium Methylation Screening Array-48, we calculated the Y-chromosome false positive rate as the proportion of probes passing a detection threshold p-value = 0.05. The empirical false positive rates for DRAGEN Array Methylation QC and GenomeStudio Methylation Module were compared with the expected false positive rate (0.05). In this analysis, the DRAGEN Array Methylation QC results are closer to the expected false positive rate of 0.05 for all samples compared to the GenomeStudio Methylation Module, which produces a higher false positive rate (Figure 3, Table 3). This experiment demonstrates that DRAGEN Array Methylation QC produces a lower false positive rate, resulting in higher-quality data for downstream analysis.

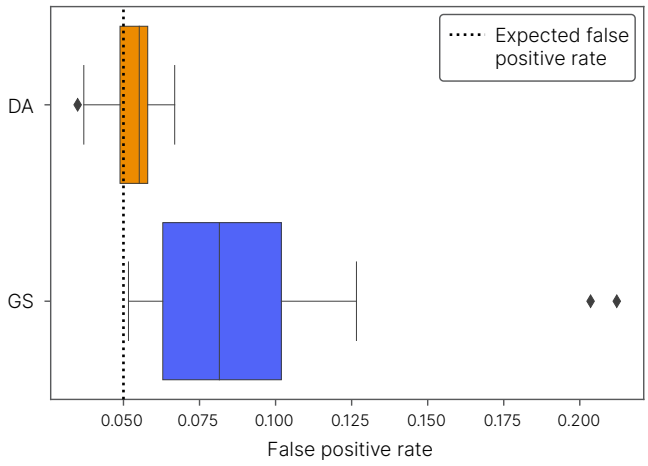


Figure 3: DRAGEN Array Methylation QC controls Y-chromosome false positive rates for female samples—Y-chromosome methylation false positive rates (FPR) at threshold p-value ≥ 0.05 with female samples demonstrate that DRAGEN Array (DA) Methylation QC generates data closer to the expected FPR range than QC analysis with GenomeStudio (GS) Methylation Module and the same data. Analysis was performed using the Infinium Methylation Screening Array-48 assays.

Table 3: Average Y-chromosome false positive rate across four female samples[a,b]

	DRAGEN Array Methylation QC		GenomeStudio Methylation Module	
DNA input	250 ng		250 ng	
Sample	Blood	Saliva	Blood	Saliva
Average FPR[c]	0.058	0.063	0.098	0.126
DNA input	50 ng		50 ng	
Sample	Blood	Saliva	Blood	Saliva
Average FPR[c]	0.052	0.053	0.092	0.133

FPR; false positive rate

a. Two DNA concentrations and two tissue types were used from four female subjects for a total of 16 samples
b. Assays were performed on the Infinium Methylation Screening Array-48
c. The expected FPR is 0.05

Summary

DRAGEN Array Methylation QC is a cloud-based software recommended for the analysis of Infinium methylation arrays. In comparison to the local-based GenomeStudio Methylation Module, automated analysis using DRAGEN Array Methylation QC allows labs to conduct higher throughput methylation QC analysis and more consistently and effectively assess data quality. DRAGEN Array Methylation QC analysis of Infinium methylation control probes generates easy-to-interpret results assessing sample and BeadChip processing procedures, and provides raw data that can be used for troubleshooting when needed. As demonstrated in this technical note, the optimized QC algorithms of DRAGEN Array Methylation QC result in better data sets for improved downstream analysis.

Learn more

Analysis of methylation array data

DRAGEN Array software

References

1. Wei S, Tao J, Xu J, et al. Ten Years of EWAS. *Adv Sci* (Weinh). 2021;8(20):e2100727. doi:10.1002/advs.202100727
2. Illumina. Infinium Methylation Screening Array. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-methylation-screening-array-data-sheet-m-gl-01893/infinium-methylation-screening-array-data-sheet-m-gl-01893.pdf. Published 2024. Accessed March 20, 2024.
3. Illumina. Infinium MethylationEPIC v2.0 BeadChip. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-methylation-epic-data-sheet-m-gl-01156/infinium-methylation-epic-data-sheet-m-gl-01156.pdf. Published 2022. Accessed March 20, 2024.
4. Illumina. Illumina Methylation BeadChips achieve breadth of coverage using two Infinium™ Chemistries. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-coverage-tech-note-m-gl-00922/infinium-coverage-tech-note-m-gl-00922.pdf. Published 2022. Accessed April 8, 2024.
5. Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res*. 2013;41(7):e90. doi:10.1093/nar/gkt090.
6. Fortin JP, Triche TJ Jr, Hansen KD. Preprocessing, normalization, and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2017;33(4):558-560. doi:10.1093/bioinformatics/btw691
7. Zhou W, Triche TJ Jr, Laird PW, Shen H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res*. 2018;46(20):e123. doi:10.1093/nar/gky691

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-02644 v1.0