Validation of DNA methylation profiling in formalin-fixed paraffin-embedded colon and breast tissue using the Infinium HumanMethylation450 Microarray

H.R. Slaats
Maastricht University
h.slaats@student.maastrichtuniversity.nl

Abstract

DNA methylation is not only crucial for normal physiology but it is also involved in a myriad of diseases, which has raised wide interest in the development of methods for genome-wide DNA methylation profiling. Illumina's Infinium 450K platform is regarded as the most attractive, powerful, and cost-effective technique currently available for generating quantitative DNA methylation profiles in clinical tissue samples. Formalin-fixation and paraffin-embedding (FFPE) represents the standard method to preserve these clinical tissue samples. However, formalin-fixation introduces severe DNA damage, which restricts current DNA methylation analysis. The present study responds to this substantial limitation by examining the usability of Illumina's DNA restoration protocol to allow for routinely analysis of formalin-fixed tissue specimens with the Infinium HumanMethylation450 technology. It is hypothesized that the HumanMethylation450 microarray in conjugation with Illumina's Infinium HD FFPE DNA Restoration solution can be successfully employed as a discovery assay. To test this hypothesis, differentially methylated CpG sites between FFPE colon tissues and breast cancer specimens have been identified by 450K methylation microarray analysis, followed by validation with nested methylation specific PCR (MSP). This study also statistically assessed the performance of fresh frozen (FF), FFPE DNA-unrestored and FFPE DNA-restored colon tissue samples on the Infinium HumanMethylation450 microarray. The present study showed good correlation between matched FF and FFPE DNA-restored tissue specimens. The presence of a significant bias in methylation assignment introduced by formalin-fixation

was excluded. MSP-based validation of the microarray's discovery potential showed unexpected and rather questionable data. The current results suggest that – compared to the existing input titration standards – a considerable lower amount of input DNA may be successfully used on the Infinium platform. No significant correlation was detected between the length of FFPE storage and sample variability. It is concluded that the Infinium HumanMethylation450 microarray can be successfully used on FFPE sample material in conjugation with Illumina's Infinium HD FFPE DNA restoration solution, while the microarray's discovery potential remains to be elucidated.

Keywords

DNA methylation, HumanMethylation450 Microarray, Formalin-Fixed Paraffin-Embedded, Infinium HD FFPE restoration solution.

Introduction

DNA methylation is a biochemical process that attaches chemical tags called methyl-groups to cytosine DNA nucleotides. These methyl-groups switch genes off or on by affecting interactions between DNA and the cell's gene-reading machinery (1, 2). DNA methylation occurs predominantly on cytosines within cytosine-guanine (CpG) dinucleotides, which are frequently concentrated in large clusters called CpG islands (3). Normal cells are subjected to a well-balanced DNA methylation, which is essential for maintaining cellular function (4). DNA methylation is, however, also involved in a myriad of cancers and other diseases, which has raised wide interest in the development of methods for genome-wide DNA methylation profiling (5)Wingdings fonts also offer arrows in careful gradations of weight and different directions and styles. For variety and impact as bullets, asterisks, and ornaments, Windings also offers a varied set of geometric circles, squares, polygons, targets, and stars.. In 2011, Illumina introduced the Infinium HumanMethylation450 microarray (Infinium Methylation 450K; Illumina, Inc. CA, USA), which makes it possible to assess the methylation status of 485.512 cytosines distributed across the entire human genome. Each Infinium probe reports a β-value, which represents the ratio between methylated probe intensity and overall probe intensity, ranging between 0 (least methylated) and 1 (most methylated). The 450K methylation microarray is regarded as the most attractive, powerful, and cost-effective technique currently available for generating genome-wide DNA methylation profiles in clinical tissue samples (6, 7). Formalin-fixation and paraffin-embedding (FFPE) represents the standard method to preserve these clinical tissue samples, with over 400 million FFPE tissue samples archived in biobanks worldwide (8). However, formalin-fixation introduces severe DNA damage, which poses

Validation of DNA methylation profiling in formalin-fixed paraffin-embedded colon
and breast tissue using the Infinium HumanMethylation450 Microarray
H.R. Slaats

47

great hindrance to genome-wide DNA methylation analysis and, thereby, restricts insight into numerous disease mechanisms (9). The destructive interaction between formalin and DNA emphasizes the necessity for careful evaluation of methylation microarrays before they can be confidently used on formalin-fixed sample material. To restore degraded FFPE DNA in preparation for the Infinium assay, Illumina has introduced the Infinium HD FFPE DNA restoration solution (Illumina, Inc. CA, USA). The present study aims at validating the usage of the HumanMethylation450 microarray on FFPE sample material in conjugation with Illumina's DNA repair protocol. Based on Illumina's claims concerning the 450K methylation platform, it is hypothesized that the HumanMethylation450 microarray in conjugation with the Infinium HD FFPE DNA restoration solution can be successfully employed as discovery assay.

Material and methods

This study assessed the performance of matched fresh frozen (FF), FFPE DNA-unrestored and FFPE DNA-restored non-neoplastic colon tissue samples on the Infinium HumanMethylation450 microarray. The present study also identified differentially methylated CpG-probes between FFPE non-neoplastic colon tissues and breast cancer specimens by 450K methylation microarray analysis, followed by cross-validation with nested methylation specific PCR (MSP). Nested MSP is a molecular technique that involves methylation-dependent DNA amplification to rapidly determine the methylation status of multiple CpG-loci within a CpG-island.

Results

Correlation between FF and FFPE Sample Methylation

The colon tissue dataset consisted of subgroups, with every subgroup containing a FF, FFPE DNA-unrestored and FFPE DNA-restored sample from the same individual. As FF tissue samples constitute the best material for isolating relative large amounts of high-quality DNA, each FF sample functioned as high-quality standard (10). To examine the concordance between the high-quality FF samples and matched FFPE tissue specimens, a DNA fingerprinting of the samples was performed using the 65 single nucleotide polymorphisms (SNPs) included on the HumanMethylation450 BeadChip. SNPs are DNA sequence variations that are individual-specific and, therefore, expected to be similar between all matched samples. The DNA fingerprinting analysis assessed whether the SNP methylation statuses were consistent between matched colon samples. As shown for colon sample 08.06923 (figure 1A), the SNP methylation profile between matched FF and

FFPE DNA-restored samples fell along the identity line in the scatter dot plot. The identity line is the *x=y* line that is used as reference: when the two data sets are equal to each other, the corresponding scatters fall along the identity line. Hence, matched FF and FFPE DNA-restored samples showed highly similar SNP methylation. However, the SNP methylation analysis between matched FF and FFPE DNA-unrestored colon samples revealed a lower degree of similarity by showing less clustering along the identity line (figure 1B). Thus, the SNP methylation profile of FF samples – which contain high-quality DNA – showed higher resemblance to matched FFPE DNA-restored samples compared with matched FFPE DNA-unrestored samples.
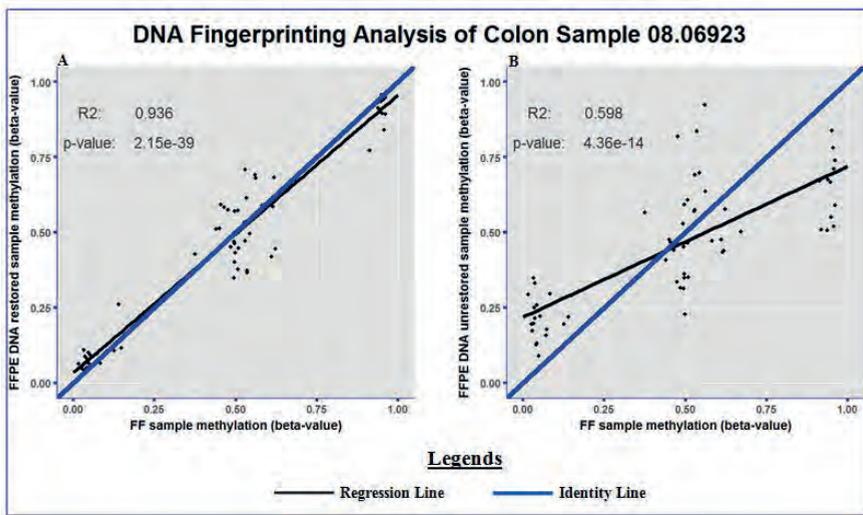

Figure 1. Representative example of the SNP methylation status between matched colon samples. The methylation statuses of the 65 SNPs included on the 450K methylation microarray are shown between matched FF and FFPE DNA-restored (A) and matched FF and FFPE DNA-unrestored (B) colon samples. The R^2 and p-value of the regression line are shown in the top left corner of each corresponding graph.

Methylation Signal Intensity Analysis

The 450K methylation array is a two-color assay that employs two different assay technologies to quantitatively measure methylation levels. The raw methylated and unmethylated signal intensities are represented as red and green color channels, which correspond to the unmethylated and methylated state respectively. To assess the 450K methylation microarray for a bias in methylation assignment introduced by formalin-

Validation of DNA methylation profiling in formalin-fixed paraffin-embedded colon
and breast tissue using the Infinium HumanMethylation450 Microarray
H.R. Slaats

49

fixation, the red/green signal intensity ratio was calculated for FF, FFPE DNA-unrestored and FFPE DNA-restored colon samples (figure 2). The red/green-ratio's did not significantly differ between the three fixation methods, with even the poor-quality DNA derived from FFPE DNA-unrestored colon samples not reaching significance.
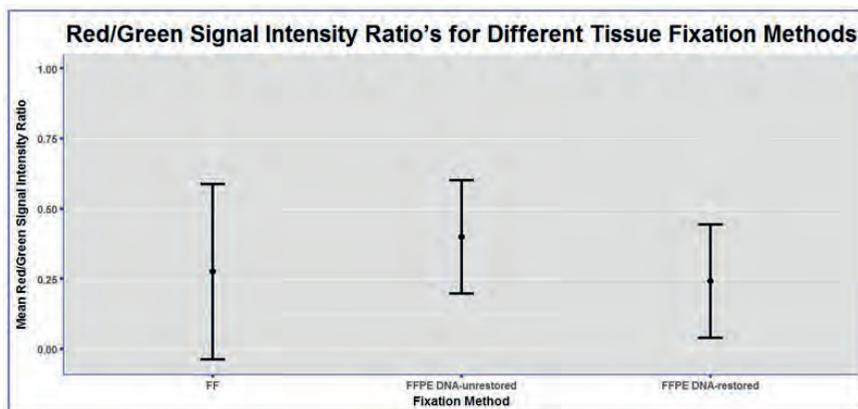

Figure 2. Red/green signal intensity analysis. The mean (± SD) ratio between the red (unmethylated) and green (methylated) signal intensities has been visualized for FF, FFPE DNA-unrestored and FFPE DNA-restored colon samples.

Cross-Validation of the 450K Results with Methylation Specific PCR

To validate the discovery potential of the 450K methylation microarray, the 450K results were cross-validated with nested MSP. A selection of differentially methylated CpG-probes between formalin-fixed breast cancer specimens and non-cancerous colon tissues was used to validate Illumina's 450K methylation microarray. CpG-probes with an average β-value lower than 0.2 in one dataset and higher than 0.5 in the other were considered differentially methylated. These stringent conditions were applied to ensure sufficient specificity, as nested MSP is highly sensitive for the detection of low methylation levels.

Table 1. MSP-based methylation status of nine panels of differentially methylated CpG-probes for FFPE breast and colon samples. The examined panels of differentially methylated CpG-probes and the genes in which these panels are located are depicted in the first column. The percentages reflect the relative amount of methylated samples in the corresponding breast or colon dataset characterized by nested MSP. CpG-probe panels are colored dark gray if the results were inconsistent with the 450K array data and light gray if the results were consistent with the 450K array data.

MSP-based Validation of 450K Array Data

CpG-Probes	DNA Methylation Status ascertained from array data		DNA Methylation Status Characterized by MSP Pilot Breast Samples + Pilot Colon Samples		DNA Methylation Status Characterized by MSP Pilot Breast Samples + MSP Colon Sample Cohort	
	Breast Samples	Colon Samples	Methylation in Breast Samples (%) n=8	Methylation in Colon Samples (%) n=5	Methylation in Breast Samples (%) n=8	Methylation in Colon Samples (%) n=8
Gene: Neu1 cg20383155, cg18363192 cg27347290	Unmethylated	Methylated	50	100	50	50
Gene: CHAD cg10421979, cg19046826 cg22627841	Unmethylated	Methylated	100	100	100	75
Gene: CMTM2 cg07153665, cg06666025 cg05546863, cg16626067	Methylated	Unmethylated	100	100	100	13
Gene: VWCE cg01698298, cg07402729 cg06282270, cg05190577	Methylated	Unmethylated	100	20	100	0
Gene: AC007618.1 cg05166022, cg02794451 cg15442907, cg13647052	Methylated	Unmethylated	100	100	100	38
Gene: MX2 cg21390801, cg06752054 cg02949186	Methylated	Unmethylated	100	100	100	25
Gene: GFI1 cg24683414, cg17858328 cg26027669	Methylated	Unmethylated	100	100	100	13
Gene: MTCP1NB cg05461361, cg08446187 cg27268405	Methylated	Unmethylated	100	20	100	25
Gene: AC025171.1 cg05551979, cg26542283 cg03723510, cg22055815	Methylated	Unmethylated	100	60	100	0

Legends

■ = Consistent with 450K Array Data   ■ = Inconsistent with 450K Array Data

Based on the 450K array data, it was expected that the selection of differentially methylated CpG-probes located in the genes Neu1 and CHAD were unmethylated in the breast samples and methylated in the colon samples, while the expectations were vice versa for the selection of differentially methylated CpG-probes located in the other genes (table 1). However, upon MSP-based analysis of the original pilot samples (8 breast cancer samples together with 5 non-cancerous colon samples), only four out of nine CpG-probe panels fit these expectations, with the other five CpG-probe panels showing methylation in all samples that were expected to turn out unmethylated. It was, however, striking to see that the number of CpG-probe panels that matched the array-based expectations markedly increased after replacement of the original colon samples with a panel of 8 additional FFPE non-cancerous colon samples, which are further referred to as the MSP colon sample cohort (table 1). Thus, validation based on the original pilot samples did not show results that were in line with the microarray-based expectations, while the MSP colon sample cohort showed methylation assignments that were mostly consistent with the array-based expectations.

Another interesting observation concerns the number of samples that failed to be characterized by MSP. A considerable amount of the specimens from the MSP colon sample

Validation of DNA methylation profiling in formalin-fixed paraffin-embedded colon and breast tissue using the Infinium HumanMethylation450 Microarray
H.R. Slaats

51

cohort failed to produce any results, while none of the other samples were not annotated by MSP (data not shown). This indicates that the colon specimens from the MSP colon sample cohort performed worse on the nested MSP assay compared to the other tissue samples.

Input Titration Analysis

The ultimate success of the 450K methylation assay strongly depends on using an accurately quantified amount of input DNA. Illumina recommends to use a minimal amount of 500ng of DNA for the manual 450K methylation assay. To test whether robust results could also be realized with a lower amount of input DNA, the CpG methylation status was compared between technical colon replicates that were analyzed on the 450K methylation microarray using a variable amount of input DNA. The concordance between the technical replicates was calculated in terms of Euclidean distance, with a decreasing Euclidean distance corresponding to a higher similarity between the replicates. Figure 3 reveals no significant correlation between the amount of input DNA and the similarity between colon replicates. It is, however, important to note that an amount of input DNA of 75ng or lower is associated with a relative high Euclidean distance as well as with a high variability in Euclidean distance, which corresponds to a low measurement accuracy. In contrast, 100ng of input DNA or more seems to perform well as this shows a limited Euclidean distance, indicating a high similarity between the technical colon replicates. Thus, in contrast to Illumina's recommendations of 500ng of input DNA, current analysis between technical colon replicates suggests that the cut-off value for a proper 450K array analysis may lie between the 75 and 100ng of input DNA.
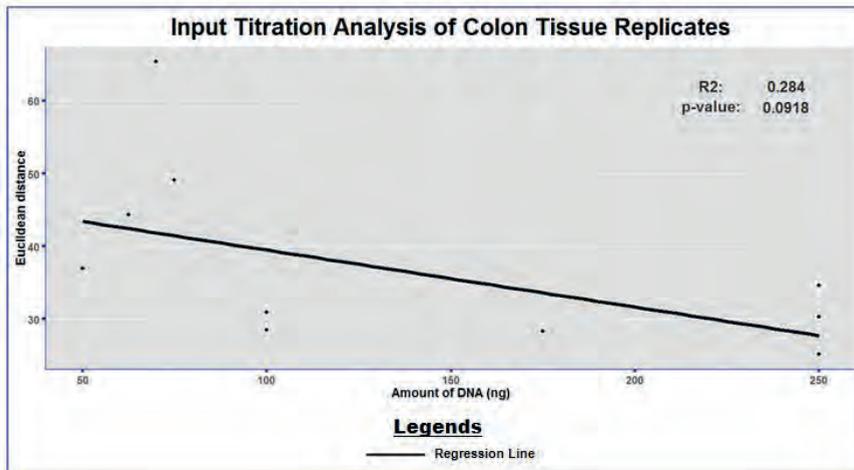

Figure 3. Input titration analysis. The Euclidean distance between technical colon replicates has been calculated for an amount of input DNA ranging from 50 to 250ng. The R^2 and p-value of the regression line are shown in the top right corner.

Correlation of Tissue Sample Variability with FFPE Block Age

Besides the type of fixation and the amount of input DNA, FFPE sample block age can contribute to problems in diagnostic testing. Advanced FFPE sample block age corresponds to elevated levels of DNA fragmentation, which may affects the overall success of the Infinium process. To test this, the percentage of differentially methylated CpG-probes between colon and breast sample replicates has been assessed in relation to the FFPE sample block age. Figure 4 reveals no significant correlation between tissue sample variability and FFPE block age, which suggests that the length of FFPE storage has no influence on sample variability.
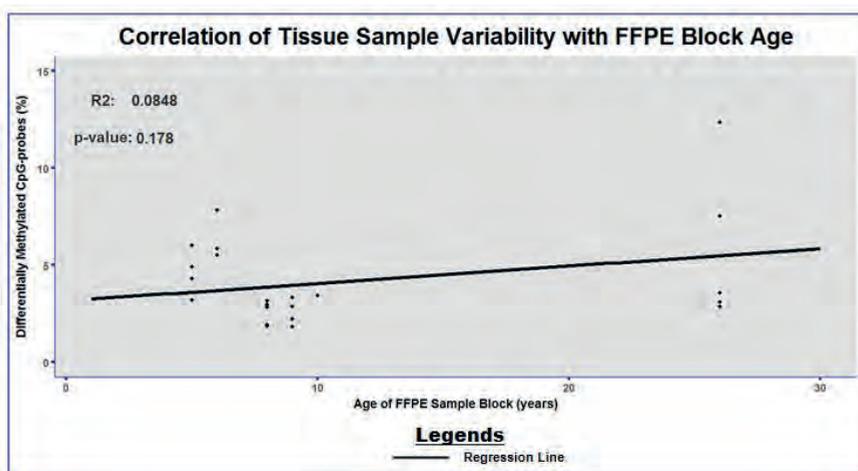

Figure 4. Correlation between the concordance of sample replicates and FFPE block age. The percentage of differentially methylated CpG-probes between colon and breast sample replicates is illustrated in relation to the FFPE sample block age. The R^2 and p-value of the regression line are shown in the top left corner.

Discussion/Conclusion

This study validates DNA methylation profiling in FFPE tissue samples using the Infinium HumanMethylation450 BeadChip array in conjugation with Illumina's HD FFPE DNA restoration protocol. Current DNA fingerprinting analysis reveals that – unlike FFPE DNA-unrestored tissue specimens – FFPE DNA-restored tissue samples show high concordance with FF tissue specimens. This indicates that the Infinium HD FFPE DNA restoration protocol is a competent DNA repair protocol and, therefore, overcomes the limitation of non-degradable input material for the Infinium assay. The present study also

Validation of DNA methylation profiling in formalin-fixed paraffin-embedded colon and breast tissue using the Infinium HumanMethylation450 Microarray
H.R. Slaats

53

shows the absence of a significant difference in red/green signal intensity ratio between FFPE tissue samples and their FF counterparts, which excludes any bias in methylation assignment introduced by formalin-fixation. Thus, both DNA fingerprinting and red/green signal intensity analyses suggest that the 450K platform successfully interrogates the methylation status of CpG-loci when combined with Illumina's DNA restoration procedure. This corresponds to previous validation studies showing highly robust FF and FFPE results on the 450K methylation microarray in conjugation with Illumina's DNA restoration procedure (11, 12).

Validation of Infinium's discovery potential by using the original pilot samples revealed a considerable amount of CpG-probe panels with methylation assignments that were inconsistent with the array-based expectations, which contrasts the postulated hypothesis. It was, however, interesting to see that – in contrast to the original pilot samples – the methylation assignment of the MSP colon sample cohort was mostly consistent with the array-based expectations. Specimens from the MSP colon sample cohort differed from the other samples with respect to the method that was applied for DNA extraction. The MSP colon sample cohort was subjected to the manual QIAamp DNA Mini Kit (Qiagen), which employs silica-membrane-based nucleic acid purification. The original pilot samples were, however, subjected to the semi-automated Maxwell 16 nucleic acid extraction (Promega) that utilizes paramagnetic particles for cellulose-based binding of nucleic acids. It has been evidenced that Qiagen's DNA extraction technology performs best for applications requiring large PCR products, while the Maxwell extraction has proven to perform superior for applications requiring small PCR products (13). This suggests a more efficient DNA amplification in the original pilot samples as compared to the specimens from the MSP colon sample cohort, since the amplicon size used for the PCR reactions was particularly low. Moreover, a superior DNA amplification in the original pilot samples is in line with the observed lower failure rate among the original pilot samples as compared to the MSP colon sample cohort. Because MSP-based validation of the original pilot samples revealed methylation in all samples that were expected to turn out unmethylated, it is suggested that the more efficient DNA amplification in these samples culminated in over-sensitive methylation detection by nested MSP. Nested MSP has a sensitivity of 1:50.000 (one methylated allele in 50.000 unmethylated alleles), which makes nested MSP an extremely sensitive techniques for analyzing DNA (14). Further research is needed to elucidate the inconsistencies between the Infinium platform and the results obtained by nested MSP and assess the possible role of MSP's high sensitivity herein. This can be achieved by performing a direct MSP instead of a nested MSP, which does not involve a first-stage methylation-independent target enrichment and, therefore, confers less sensitivity.

Another possibility involves a reduction in the amount of amplification cycles used for the PCR reactions, which would also reduce the sensitivity of the MSP assay. An alternative approach applies more stringent conditions used for identifying the differentially methylated CpG-probes. It is, however, important to note that β-value averages rather than individual β-values have been used for identification of the differentially methylated CpG-probes. This implies that the presence of sample outliers need to be taken into account, which explains the low methylation percentages in most of the sample datasets that were expected to turn out unmethylated.

The Infinium platform strongly depends on using an accurately quantified amount of input DNA. In contrast to Illumina's recommendation of using a minimal amount of 500ng of input DNA, current input titration analysis suggests that the cut-off value for a proper 450K array analysis lies much lower, namely between the 75 and 100ng of input DNA. However, current input titration analysis does not allow for a well-substantiated conclusion, as the analysis only includes a limited number of colon replicates. Moreover, Illumina's recommended input DNA quantity is excluded from the analysis. Thus, a more reliable and conclusive input titration analysis requires an increase in sample size as well as expansion of the input quantity range.

Besides sufficient DNA quantity, accurate genome-wide methylation profiling on the 450K platform depends on an adequate DNA quality. Despite the detrimental effect of advanced FFPE block age on DNA integrity, the present study could not detect a significant correlation between tissue sample variability and FFPE block age (15). This suggests that the genomic DNA quality assessment – which constitutes the first step of the methylation microarray protocol – succeeds in excluding poor-quality DNA from further analysis and, thereby, make the Infinium platform independent of FFPE block age. However, this opposes previous research by Dumenil *et al.* who showed a significant correlation between sample variability and FFPE block age, which suggests that FFPE block age negatively influences the 450K array analysis despite employment of the genomic DNA quality assessment (11). These inconsistent results together with the fact that FFPE blocks used in discovery assays are generally of advanced age encourage to merit further research into the impact of FFPE block age on the Infinium assay. Further research should engage on a broad range of FFPE block ages with an evenly distributed sample set, which have not been provided in both the present study and previous research.

In conclusion, the present study suggests that the Infinium HumanMethylation450 platform can be used successfully on FFPE sample material in conjugation with Illumina's Infinium HD FFPE DNA restoration solution and that the resulting data will be concordant to that of corresponding FF DNA samples. The discovery potential of the methylation

Validation of DNA methylation profiling in formalin-fixed paraffin-embedded colon and breast tissue using the Infinium HumanMethylation450 Microarray
H.R. Slaats

55

microarray, however, remains to be elucidated. Since the ultimate success of the 450K platform depends on an accurately quantified amount of input DNA as well as a robust DNA quality, this study also encourages more extensive evaluation of the required input titration and the impact of FFPE block age on the 450K array results.

Role of the student

H.R. Slaats was an undergraduate student in Biomedical Science working under the supervision of Tim C. de Ruijter when the research in this report was performed. The topic was proposed by the supervisor. The nested MSP experiments as well as statistical processing of the 450K array data were done by the student.

References

1. Kass SU, Pruss D, Wolffe AP. How does DNA methylation repress transcription? Trends Genet. 1997 Nov;13(11):444-9. PubMed PMID: 9385841.
2. Razin A, Cedar H. DNA methylation and gene expression. Microbiol Rev. 1991 Sep;55(3):451-8. PubMed PMID: 1943996. Pubmed Central PMCID: 372829.
3. Clark SJ, Harrison J, Frommer M. CpNpG methylation in mammalian cells. Nature genetics. 1995 May;10(1):20-7. PubMed PMID: 7647784.
4. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nature reviews Genetics. 2013 Mar;14(3):204-20. PubMed PMID: 23400093.
5. Portela A, Esteller M. Epigenetic modifications and human disease. Nature biotechnology. 2010 Oct;28(10):1057-68. PubMed PMID: 20944598.
6. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nature reviews Genetics. 2011 Aug;12(8):529-41. PubMed PMID: 21747404. Pubmed Central PMCID: 3508712.
7. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011 Dec;3(6):771-84. PubMed PMID: 22126295.
8. Sah S, Chen L, Houghton J, Kemppainen J, Marko AC, Zeigler R, et al. Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. Genome medicine. 2013;5(8):77. PubMed PMID: 24001039. Pubmed Central PMCID: 3978876.
9. Sikorsky JA, Primerano DA, Fenger TW, Denvir J. DNA damage reduces Taq DNA polymerase fidelity and PCR amplification efficiency. Biochemical and biophysical research communications. 2007 Apr 6;355(2):431-7. PubMed PMID: 17303074. Pubmed Central PMCID: 1945218.
10. Weiss AT, Delcour NM, Meyer A, Klopfleisch R. Efficient and cost-effective extraction of genomic DNA from formalin-fixed and paraffin-embedded tissues. Veterinary pathology. 2011 Jul;48(4):834-8. PubMed PMID: 20817894.
11. Dumenil TD, Wockner LF, Bettington M, McKeone DM, Klein K, Bowdler LM, et al. Genome-wide DNA methylation analysis of formalin-fixed paraffin embedded colorectal cancer tissue. Genes Chromosomes Cancer. 2014 Mar 28. PubMed PMID: 24677610.
12. Moran S, Vizoso M, Martinez-Cardus A, Gomez A, Matias-Guiu X, Chiavenna SM, et al. Validation of DNA methylation profiling in formalin-fixed paraffin-embedded samples using the Infinium HumanMethylation450 Microarray. Epigenetics : official journal of the DNA Methylation Society. 2014 Apr 14;9(6). PubMed PMID: 24732293.

13. Khokhar SK, Mitui M, Leos NK, Rogers BB, Park JY. Evaluation of Maxwell(R) 16 for automated DNA extraction from whole blood and formalin-fixed paraffin embedded (FFPE) tissue. Clinical chemistry and laboratory medicine : CCLM / FESCC. 2012 Feb;50(2):267-72. PubMed PMID: 22022984.

14. Hayat MA. Methods of Cancer Diagnosis, Therapy and Prognosis: Breast Carcinoma: Springer Science + Business Media B.V.; 2008.

15. Craig JM, Vena N, Ramkissoon S, Idbaih A, Fouse SD, Ozek M, et al. DNA fragmentation simulation method (FSM) and fragment size matching improve aCGH performance of FFPE tissues. PloS one. 2012;7(6):e38881. PubMed PMID: 22719973. Pubmed Central PMCID: 3376148.

Validation of DNA methylation profiling in formalin-fixed paraffin-embedded colon and breast tissue using the Infinium HumanMethylation450 Microarray

H.R. Slaats

57