PAPER

Check for updates

In vitro diagnosis of DNA methylation biomarkers with digital PCR in breast tumors†

Xingye Cui,‡[a,b] Lei Cao,‡[a,b] Yalin Huang,[a,b] Dan Bai,[b,c] Shan Huang,[d] Min Lin,[a,b] Qingzhen Yang,[a,b] Tian Jian Lu,[e] Feng Xu (ID) [a,b] and Fei Li (ID) *[a,b]

Liquid biopsy of cancers using DNA methylation biomarkers has received significant interest, where the quantification of multiple biomarkers is generally needed for improving the sensitivity and specificity of cancer diagnosis. However, the inefficiency of the traditional quantitative polymerase chain reaction (qPCR)-based MethyLight assay for detecting the extremely low concentration of methylated DNA fragments in body fluids limits its clinical applications. Here, we developed an ultrasensitive microwell chip digital polymerase chain reaction (dPCR)-based MethyLight assay. Using the synthesized samples, the developed MethyLight assay can achieve 10^3–10^4-fold lower limit of detection and 1–16-fold lower limit of quantification than the traditional MethyLight assay. Four hypermethylated alleles (*RARβ2*, *BRCA1*, *GSTP1* and *RASSF1A*) related to breast cancer in twenty-three clinical samples were tested using the microwell chip dPCR-based MethyLight assay. The results showed that the dPCR assay achieves ~2 times enhancement in the cancer detection rate over the traditional quantitative PCR. Furthermore, the dPCR can detect the healthy and benign samples, which are undetectable using the traditional MethyLight assay. In multiple gene analysis, we achieved the highest detection rate of 93.3% (in the "OR" format of *RARβ2* and *GSTP1*). Lastly, the estimated cut-off values in the dPCR assay were: <1, ~1 to 100 and >100 (copies per μL) referring to the healthy, benign and malignant breast cancers, respectively. Therefore, the developed microwell chip dPCR-based MethyLight assay could provide a powerful tool for cancer biopsy diagnosis and disease monitoring.

Introduction

Cancer has become a major public health problem in the world, where timely and accurate diagnosis of cancer could bring favorable clinical outcomes.[1] Abundant biomarkers have been explored for the diagnosis of cancer.[1,2] For instance, clinical tests of prostate-specific antigen[3] by established immunoassays and breast cancer 1/2, early onset (*BRCA1/2*) gene mutations using polymerase chain reaction (PCR) have been used to evaluate the risks of prostate cancer and breast cancer, respectively.[4,5] Although widely adopted, clinical results are still unsatisfactory due to the limited accuracy of early cancer diagnosis based on these biomarkers.[6] Recently, the epigenetic alteration of the DNA methylation status (*e.g.*, allele specific hypermethylation in the gene promoter regions and global hypomethylation of the genome) has been found to be specifically associated with carcinogenesis in certain cancer types, such as lung cancer, prostate cancer and breast cancer.[7–9] Herein, allele specific hypermethylation that regulates tumor suppressors and DNA-repair genes (*e.g.*, MutL homolog 1, colon cancer, nonpolyposis type 2 (*E. coli*) (*hMLH1*), *BRCA1*, adenomatosis polyposis coli (*APC*)) can lead to human cancer malignancy at an early stage.[10–12] For example, the aberrantly methylated glutathione *S*-transferase pi 1 (*GSTP1*) and retinoic acid receptor, beta (*RARβ2*) genes are associated with prostate cancer and metastatic breast cancer.[13,14] Considering that methylated DNA fragments are present in accessible body fluids (*e.g.*, blood, plasma, urine, and saliva),[10,13,15,16] methylation biomarker-based cancer diagnosis has recently attracted significantly increasing interest.[17]

a The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: feili@mail.xjtu.edu.cn; Fax: +86-29-82667486; Tel: +86-29-82667486
b Bioinspired Engineering and Biomechanics Center (BEBC), Xi'an Jiaotong University, Xi'an 710049, China
c Department of Biochemistry and Molecular Biology, School of Medicine, Xi'an Jiaotong University, Xi'an 710061, China
d Department of Radiotherapy, Second Affiliated Hospital, Xi'an Jiaotong University, Xi'an 710004, China
e MOE Key Laboratory for Multifunctional Materials and Structures, Xi'an Jiaotong University, Xi'an 710049, China
†Electronic supplementary information (ESI) available: Sequences employed in this work. See DOI: 10.1039/c8an00205c
‡These authors contributed equally to the work in this paper.

Sodium bisulfite conversion or methylation-sensitive-/dependent restriction enzymes (MSREs/MDREs) based methods are the most commonly used techniques for DNA methylation analysis.[18] Among them, the methylation specific PCR (MSP) and the MethyLight assay are the established methods for the routine monitoring of allele specific methylation status.[19,20] However, there are several challenges associated with these methods, such as low quantitative ability, low precision, need for external calibrations, vulnerability to PCR variations or inhibitors and large sample consumption, which limit their widespread clinical applications.[19,21] Besides, the concentrations of target molecules are extremely low in body fluids (i.e., the DNA concentration is $\sim$5–20 ng ml^{-1} in the plasma of a healthy individual), which has been the main obstacle for cancer diagnosis based on methylated DNA fragments.[22]

Recently, digital PCR (dPCR) has emerged as a powerful ultrasensitive tool for DNA methylation analysis due to its ultrahigh sensitivity and improved assay accuracy based on the system's robustness and absolute quantification capability without needing external references.[23,24] For instance, by employing the methy-BEAMing (beads, emulsion, amplification and magnetics) dPCR, $\sim$1 methylated DNA molecule in 5000 unmethylated ones in the plasma of colorectal cancer patients has been successfully detected, which is much more sensitive and specific compared to the conventional Bs-pyrosequencing and methylation specific PCR (MSP) methods.[25] Later, a droplet dPCR-based MethyLight assay has been developed with the capability of detecting as few as 19 and 38 copies per 20 μL of methylated Enah/Vasp-like (EVL) and neurotrophic receptor tyrosine kinase 3 (NTRK3) genes, respectively, which are ten-fold lower than the conventional quantitative PCR (qPCR)-based MethyLight assay (379 copies per 20 μL for both genes).[26] However, the carcinogenesis and cancer development often involve the alteration of multiple genes' methylation status rather than a single allele's, where the analysis of the methylation of multi-genes can significantly improve the detection sensitivity and specificity compared to a single gene.[27] Although the qPCR-based MethyLight assay has been established for the analysis of prostate cancer,[28] such a technique suffers from several limitations including primer design, fluorescent dye selection, optimization of reaction conditions and need for a reference gene to quantification. As mentioned above, absolute quantification with a dPCR system can alleviate these limitations. However, the absolute quantification of circulating DNA methylation biomarkers in human blood circulation with chip-based dPCR has not been exploited yet.

In this work, to detect extremely low concentrations of methylated DNA fragments for cancer diagnosis, we developed a microwell chip dPCR-based MethyLight assay. The assay consists of plasma DNA extraction (Fig. 1(a)), bisulfite treatment of DNA samples (Fig. 1(b)), digital PCR amplification that uses the hydrolysis probe method (Fig. 1(c) (I–III)) and the data analysis process (Fig. 1(c) (IV)). The performance of the assay was evaluated as compared to the conventional
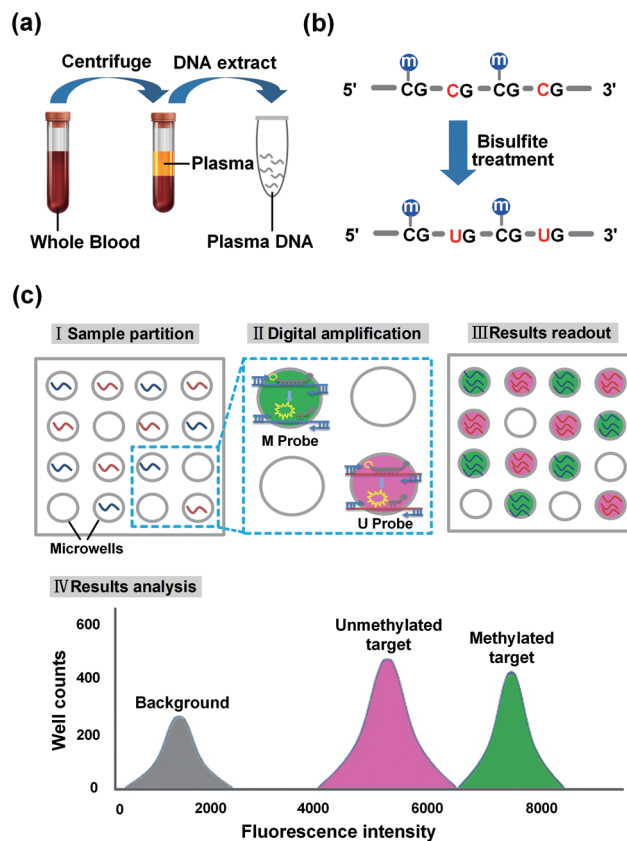

Fig. 1 Scheme of the dPCR-based MethyLight assay. (a) Circulating DNA was extracted from the plasma. (b) After bisulfite treatment, methylated cytosines in the target DNA remained unchanged, while the unmethylated cytosines turned into uracils, making the sequence change. (c) Hydrolysis probe-based digital PCR quantification of methylated templates (FAM™ fluorescein labeled, green color) and unmethylated templates (VIC™ fluorescein labeled, red color).

qPCR-based MethyLight assay. Typically, the concentrations of biomarkers, which are critical for cancer detection and staging, can be reflected by the limit of detection (LOD) and the limit of quantification (LOQ) of the assay. In this work, the LOD is defined as the lowest detectable concentration of samples, which discriminates the healthy and the cancer samples. The LOQ refers to the lowest detectable ratio of methylated molecules in total molecules at a given concentration ($P < 0.05$ vs. control), which is the base for cancer staging. In standard DNA sample analysis, the dPCR assay shows prevalent 10^3–10^4-fold lower LOD and 1–16-fold lower LOQ than the results obtained using the conventional qPCR assay. In the detection of clinical plasma samples, the results of the multiplexed analysis of hypermethylated alleles (RARβ2, BRCA1, GSTP1 and ras association domain family member 1 (RASSF1A)) in breast cancer show that the dPCR assay exhibits one-fold enhancement in the cancer detection rate over the qPCR assay. More importantly, the developed microwell chip-based dPCR assay can discriminate benign cases from malignant breast cancers unambiguously, while qPCR can only detect the malignant cases.

Materials and methods

Standard DNA samples

Sets of sequences of *GCK* gene and *PDX1* gene in powder form were purchased from Sangon Biotech (Shanghai) and used as the standard DNA samples. The standard methylated and unmethylated DNA sequences of *GCK* gene (NM_033507) were designed with two cytosine residues (representing two methylation sites in methylated DNA molecules) in the TaqMan probe binding region, instead of two thymine residues in the unmethylated counterparts. Sequences of *PDX1* gene (NM_000209) were designed with four cytosines in the TaqMan probe binding region. The detailed information of the sequences can be found in the ESI, Table S1.†

Clinical sample collection, process and target hypermethylated alleles

Blood samples of fifteen female patients primarily diagnosed with breast tumor or cancer were collected following the approved protocols of the local IRB committee from the Second Affiliated Hospital of Xi'an Jiaotong University (China). No clinical treatment was applied to the patients before blood collection. Samples collected from young male and female volunteers were used as negative and healthy controls. For real blood samples, informed consents were obtained from all the human subjects in this study. Specifically, 1.5 mL peripheral blood was collected from each patient (blood collection tubes containing EDTA, BD Vacutainer®). The blood samples were centrifuged at 14 500 rpm for 15 min, after which ~600 μL supernatant (plasma) for each sample was obtained for DNA extraction. DNA extraction was done by using a QIAamp DNA mini kit (QIAGEN, Germany) and eluted in 20 μL TE buffer. Then, the extracts were modified using an EpiTect Fast DNA Bisulfite Kit (QIAGEN, Germany) following the protocol provided by the manufacturer. The product was eluted in 20 μL TE buffer and stored at −20 °C until use. To test the ability of the MethyLight assay, four hypermethylated alleles (*RARβ2*, *BRCA1*, *GSTP1* and *RASSF1A*) that are associated with breast cancer genesis and development were selected as targets for the following assays.

qPCR assays

All qPCR analyses were performed on a real time PCR system (Model 7500 Fast, Thermo Fisher Scientific®). The information of the primer and probe sequence is shown in the ESI, Table S1†. The reaction volume was set as 20 μL and each reaction contains 2 μL DNA sample, 10 μL Premix Ex TaqTM (2×) (Takara®), 0.4 μL (10 nM) of each primer (forward primer/reverse primer), 0.8 μL (10 nM) TaqMan probe, 0.2 μL ROX Reference Dye II (50×) (Takara®) and 6.2 μL ddH$_2$O. The thermal cycling conditions were as follows: initial denaturation step at 95 °C for 5 min; 40 cycles of 95 °C for 30 s annealing at the temperature (refers to Tm, shown in the ESI, Table S1†) for 30 s, then 72 °C for 30 s; and a final extension step at 72 °C for 8 min. The fractional cycle number (C_q value) obtained from the software of the PCR system (V2.0.6, Thermo Fisher

Scientific®) was calculated based on the exponential nature of PCR kinetics. The lower C_q value means more copies of initial DNA templates. The data were analyzed using the software and the C_q value was determined with the single threshold method. The unpaired *t*-test was applied to determine the significance between the two groups. The *P*-value <0.05 was accepted with a significant difference.

Microwell chip-based dPCR assays

The microwell chip-based dPCR assay was referenced by the operational workflow provided by Thermo Fisher Scientific®. The reaction mixture consists of the 3D mixture (2×), and all primers and probes that are the same as those used in the qPCR assays. Briefly, each reaction mixture was 20 μL, containing 2 μL DNA sample, 10 μL 3D mixture (2×), 0.5 μL (10 nM) of each primer, 1 μL (10 nM) of TaqMan probe and 8 μL ddH$_2$O. Then, as per the dPCR protocol, 14.5 μL PCR mixture was loaded to the 20K™ microwell chip (Thermo Fisher Scientific®) using the QuantStudio 3D digital PCR Chip Loader. All dPCR amplifications were performed on the ProFlex™ 2× Flat PCR system using standard conditions: 95 °C for 5 min; 50 cycles of 95 °C for 30 s and annealing temperature for 1 min (ESI, Table S1†). Each assay was performed at least in triplicate to determine the measurement errors. Thereafter, data readout and analysis were processed using the QuantStudio™ 3D Analysis Suite™ Software (V1.0). The default confidence level (%) is 95% and the default desired precision (%) is 10%, where the calculated precision (%) for the data group is defined as the size of the confidence interval for distinguishing between two sample concentrations at a given confidence level (1). To assess the specificity of the dPCR assay, initial experiments were conducted using 100% methylated DNA and 100% unmethylated DNA. The NTC contained all PCR components except for the DNA template, which has been used for determining the assay threshold.

LOD determination

The LOD is defined as the lowest detectable concentration of the samples. To determine the LOD of the qPCR assay, the standard DNA sample was first prepared as 50 ng μL^{-1} and then was double gradient diluted from 20 pg μL^{-1} to 0.3×10^{-3} pg μL^{-1}. For the LOD of the dPCR assay, the DNA sample was double gradient diluted from $0.1–0.01 \times 10^{-3}$ pg μL^{-1}. Each assay was conducted at least in three independent runs with four replicates in a run. The linear regression analyses were performed to obtain the correlation coefficients (R^2) and the LOD.

LOQ determination

The LOQ refers to the lowest detectable ratio of methylated molecules in unmethylated total molecules at a given concentration (*P* < 0.05 *vs.* control). In qPCR, the input DNA was 20 ng per sample. For each subgroup, it was prepared with methylation percentages of 100%, 50%, 25%, 12.5%, 6.25%, 3.125%, 1.56%, 0.78%, 0.39% and 0% per 20 μL reaction volume. Then, they were tested at least four times in a run and at least three independent runs in qPCR, and at least two inde-

pendent chips in dPCR. The significance between two groups was determined through the unpaired t-test and P-value <0.05 was accepted with a significant difference.

Data analysis

In qPCR, the C_q values of a range of DNA were linearly regressed against the log transformation of the starting quantity using the Origin software (V9.2.272). In dPCR, the number of positive wells for the FAM fluorescence channel, the calculated template concentration for each chip under the conditions of 95% Poisson confidence intervals and default precision (<10%) were obtained and analyzed using the QuantStudio™ 3D Analysis Suite™ Software (V1.0). For the LOD and LOQ assays, statistical analysis was performed using the GraphPad Prism Software (V5.04). The unpaired t-test was applied to determine the significance between two groups, where P-value <0.05 was accepted as a significant difference. For the tests using clinical samples, we analysed all cases in a "double-blind" way to eliminate subjective bias and personal preferences of experimenters or participants.

Data availability

All data generated or analysed during this study are included in this published article (and its ESI† files).

Results and discussion

Negative control experiment

To confirm the feasibility of the PCR conditions and specificity of the designed primers and probes (see the ESI, Tables S1 and S2†) on detecting methylated targets, preliminary experiments were conducted by the conventional methylation specific PCR (MSP) and qPCR-based MethyLight assays using sets of standard DNA samples. No amplified signal (typically C_q (cycle of quantification) <35) from both the unmethylated DNA samples and the No template control (NTC) (data not shown) was observed, indicating that the established methods and materials are feasible for the following LOD and LOQ assays.

Detection limit of qPCR

To determine the LOD of the conventional qPCR-based MethyLight assay, we plotted C_q values against the log transformation of the final DNA concentration (since there is 2 μL DNA sample in 20 μL PCR reaction volume, the prepared DNA concentration should be divided by 10) for standard DNA samples of gene NM_000209 (Fig. 2(a)) and gene NM_033507 (Fig. 2(b)). We observed that the C_q values of gene NM_000209 firstly experience a plateau for DNA concentration from 3.05×10^{-3} pg μL^{-1} to 9.76×10^{-2} pg μL^{-1}, then linearly decrease with the DNA concentration ascending from 0.195 pg μL^{-1} to 20 pg μL^{-1} ($R^2 = 0.99$, inset of Fig. 2(a)). Therefore, the LOD of gene NM_000209 is determined as 0.195 pg μL^{-1} in the qPCR assay. For gene NM_033507, the linear range of the detection is from 0.78 pg μL^{-1} to 20.0 pg μL^{-1} ($R^2 = 0.99$, inset of
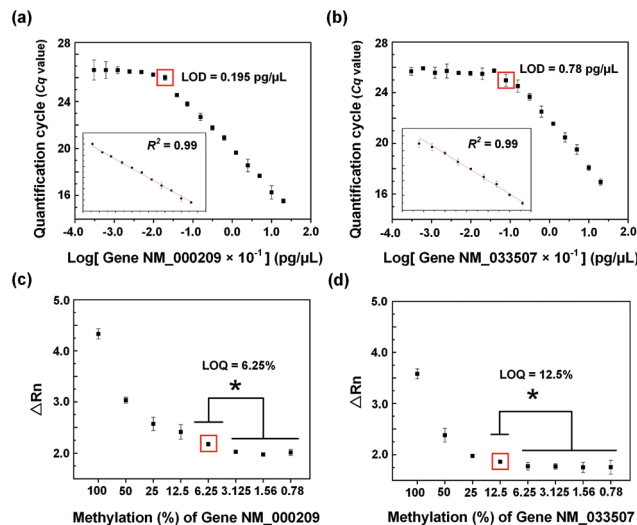

Fig. 2 LOD and LOQ for the qPCR-based MethyLight assay using standard DNA samples. (a) The LOD for gene NM_000209 is 0.195 pg μl^{-1}. (b) The LOD for gene NM_033507 is 0.78 pg μl^{-1}. (c) and (d) The LOQ was 6.25% for gene NM_000209 and 12.5% for gene NM_033507, respectively ($p < 0.05$) (Δr_n means normalized reporter signal).

Fig. 2(b)) and the LOD of gene NM_033507 is 0.78 pg μL^{-1}, which is four times the LOD of gene NM_000209. The differences in the linear range and LOD between these two genes may be attributed to the variations of DNA sequences and discrepant amplification efficiency of the qPCR.

Quantification limit of qPCR

To obtain the LOQ of the qPCR-based MethyLight assay, 20 ng DNA was set as the total DNA input, considering the linear range and LOD of the qPCR assay (Fig. 2(a) and (b)). We plotted the normalized reporter signal (marked as ΔR_n) against the percentage of DNA methylation for gene NM_000209 and gene NM_033507 (Fig. 2(c) and (d)). We observed that ΔR_n decreases with decreasing percentages of DNA methylation from 100% to 6.25% for gene NM_000209 (Fig. 2(c)) and from 100% to 12.5% for gene NM_033507 (Fig. 2(d)). There is no significant difference for ΔR_n among groups with a methylation percentage lower than 6.25% for gene NM_000209 and lower than 12.5% for gene NM_033507 ($P > 0.05$), respectively. Thus, the LOQs are determined to be 6.25% and 12.5% for gene NM_000209 and gene NM_033507, respectively.

Detection limit of dPCR

To gain the LOD of the dPCR assay, serials of two-fold diluted standard DNA samples were prepared to minimize errors that may be caused by ten-fold sample dilution (Fig. 3(a) and (b)). We observed that the copy numbers measured from our dPCR assay decrease with decreasing DNA concentrations from 25×10^{-3} pg μL^{-1} to NTC. Based on the default precision value (<10%), the dPCR assay can detect as low as 0.195×10^{-3} pg μL^{-1} (~23.9 copies per μL) and 0.09×10^{-3} pg μL^{-1} (~14.9 copies
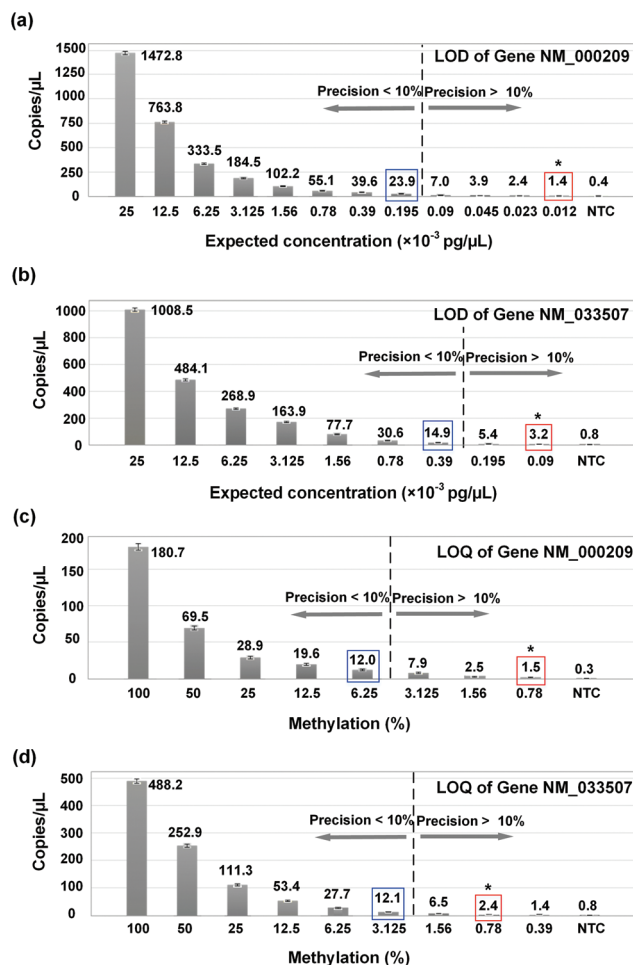
Fig. 3 LOD and LOQ for the dPCR-based MethyLight assay using standard DNA samples. (a) The LOD for gene NM_33209 is 0.195×10^{-3} pg μl^{-1} (precision = 8.2%). (b) The LOD for gene NM_33507 is 0.39×10^{-3} pg μl^{-1} (precision = 10.1%), ($p < 0.05$ vs. NTC). (c) The LOQ for gene NM_33209 is 6.25%. (d) The LOQ for gene NM_33507 is 3.13% ($p < 0.05$ vs. NTC).

per μL) for gene NM_000209 and for gene NM_033507 (blue boxes in Fig. 3(a) and (b)), respectively. Particularly, by using statistical analysis without considering the default precision, the LOD is determined to be 0.012×10^{-3} pg μL^{-1} (~1.4 copies per μL, precision = 39.3%) and 0.09×10^{-3} pg μL^{-1} (~3.2 copies per μL, precision = 20.9%) for gene NM_000209 and NM_033507 (red boxes in Fig. 3(a) and (b)), respectively ($P < 0.05$ vs. NTC). These results indicate that there are 10^3–10^4-fold improvements for the dPCR-based assay over the LODs yielded by using the qPCR assay.

Quantification limit of dPCR

Based on the LOD results, for the measurement of the dPCR assay's LOQ, we used 3.1×10^{-3} pg and 12.5×10^{-3} pg as the total input DNA amount for gene NM_000209 and NM_033507, respectively. The measured concentration of the target DNA decreases with decreasing percentage of methylated DNA from 100% to 0% (Fig. 3(c) and (d)). Based on the

default precision value (<10%), the LOQ of the dPCR assay is determined to be 6.25% (~12 copies per μL methylated templates) for gene NM_000209 and 3.125% (~12.1 copies per μL methylated templates) for gene NM_033507 (blue box in Fig. 3(c) and (d)), respectively. Similarly, by using statistical analysis rather than considering the default precision, groups with methylation percentage below 0.78% show no significant difference between each other ($P > 0.05$ vs. NTC). Therefore, the LOQ is 0.78% (~1.5 copies per μL methylated templates, precision = 30.3%) for gene NM_000209 and 0.78% (~2.4 copies per μL methylated templates, precision = 22.4%) for gene NM_033507 (red box in Fig. 3(c) and (d)), respectively. Therefore, there is ~1–16 times lower LOQ for the dPCR-based assay than the qPCR assay.

Analysis of clinical samples by qPCR and dPCR

Since the dPCR assay has shown great improvements in terms of both LOD and LOQ, to further illuminate its application in cancer diagnosis, we analyzed hypermethylated alleles in the clinical plasma samples of breast cancers. First, twenty-three clinical samples (peripheral blood) were collected and divided into four groups: the healthy male controls ($n = 4$), the healthy female controls ($n = 4$), the benign tumors ($n = 7$) and the malignant breast cancers ($n = 8$). Then, we confirmed the benign tumors and the malignant breast cancers using clinical microscopy and physiological methods. Furthermore, we optimized the qPCR- and dPCR-based MethyLight assays for selected alleles in the analysis of all clinical plasma samples. The detailed sample information and detection results are listed in Table 1.

To explore the ability of multigene diagnosis with the dPCR assay, we proposed two logic strategies to analyze the four alleles (RARβ2, BRCA1, GSTP1 and RASSF1A) in cancer detection, i.e., the "OR" and "AND" format assays. In the "OR" format assay, a positive case is defined if as long as one gene in each group of gene combinations is detected as positive, which is usually applied to the early diagnosis of cancers.[29,30] In the "AND" format assay, a positive case is defined if all the genes in each group of gene combinations are detected as positive, which is of great importance for cancer prognosis and cancer metastasis prediction. Based on these, we tried to set up cut-off values for each hypermethylated allele based on the detection results of the cases.

Positive rates for detecting breast cancer related genes in "OR" format assays with qPCR and dPCR

We first analyzed the results from the "OR" format assays, in terms of the positive rate defined as the proportion of the detected positive cases of 15 clinically confirmed cases and the benign positive rate defined as the proportion of the detected benign positive cases of seven clinically confirmed cases (Fig. 4). We observed that the dPCR assay shows a higher total positive rate, which is nearly double compared to the qPCR assay, no matter whether in single, dual, triple or quadruple gene combinations. Interestingly, as for the benign detection rate, none has been detected by qPCR, but most of the benign cases can be identified by the dPCR assay. In addition, the posi-

Table 1 Clinical diagnosis and demographic characteristics of samples among the investigated patterns of *BRCA1*, *RASSF1A*, *GSTP1*, and *RARβ2* methylation by qPCR and dPCR. "—": no amplification; "√": positive case; "*": cut-off values (data beyond the value was defined as positive, and the opposite was negative)

| | | | qPCR (C_q value) | | | | dPCR (copies μL^{-1}) | | | |
Samples	Age	Groups (clinical pathological diagnosis) (M: male; F: female)	BRCA1	RASSF1A	GSTP1	RARβ2	BRCA1	RASSF1A	GSTP1	RARβ2
C1#	25	Negative control (M)	—	—	—	—	0	0	0	0.25
C2#	31	Negative control (M)	—	—	—	—	0.23	0.80	0.24	0.24
C3#	29	Negative control (M)	—	—	—	—	0	0	0.77	0
C4#	24	Negative control (M)	—	—	—	—	0.16	0	0.23	0
C5#	24	Healthy control (F)	—	—	—	—	0.51	0.80	0.15	0.82
C6#	21	Healthy control (F)	—	—	—	—	0.75	0.31	0.24	0.68
C7#	24	Healthy control (F)	—	—	—	—	0.99*	0.88	0.25	0.08
C8#	22	Healthy control (F)	—	—	—	—	0.46	0.97*	1.06*	0.97*
P1#	30	Benign (breast hyperplasia) (F)	—	—	—	—	0.72	0.64	0.38	0.78
P2#	63	Benign (adenoma fibrosum) (F)	—	—	—	—	0.52	0.32	13.62	0.22
P3#	32	Benign (adenoma fibrosum) (F)	—	—	—	—	0.70	0.15	86.47	28.42
P4#	52	Benign (adenoma fibrosum) (F)	—	—	—	—	2.53	1.07	1.83	1.37
P5#	45	Benign (adenoma fibrosum) (F)	—	—	—	—	0.47	1.46	2.43	1.36
P6#	32	Benign (adenoma fibrosum) (F)	—	—	—	—	5.38	53.04	3.13	6.05
P7#	26	Benign (adenoma fibrosum) (F)	—	—	—	—	1.56	1.51	4.72	1.49
P8#	66	Malignant (breast cancer) (F)	—	√ (28.88)	—	√ (32.53)	0.15	245.74	179.36	139.04
P9#	28	Malignant (breast cancer) (F)	—	—	—	√ (32.46)	0.32	2.15	0.15	134.97
P10#	48	Malignant (breast cancer) (F)	√ (21.30)	—	√ (31.82)	—	125.72	0.22	164.00	1.37
P11#	30	Malignant (breast cancer) (F)	—	√ (29.01)	√ (32.46)	√ (30.73)	0.45	256.07	428.77	165.00
P12#	68	Malignant (breast cancer) (F)	—	—	√ (30.73)	√ (33.34)	1.06	254.11	240.05	123.52
P13#	57	Malignant (breast cancer) (F)	—	—	—	—	38.64	0.92	127.60	72.22
P14#	54	Malignant (breast cancer) (F)	√ (31.58)	—	√ (33.34)	—	69.45	0.15	162.13	0.00
P15#	51	Malignant (breast cancer) (F)	—	√ (15.61)	√ (17.94)	√ (19.91)	1.47	2574.60	2148.10	874.12
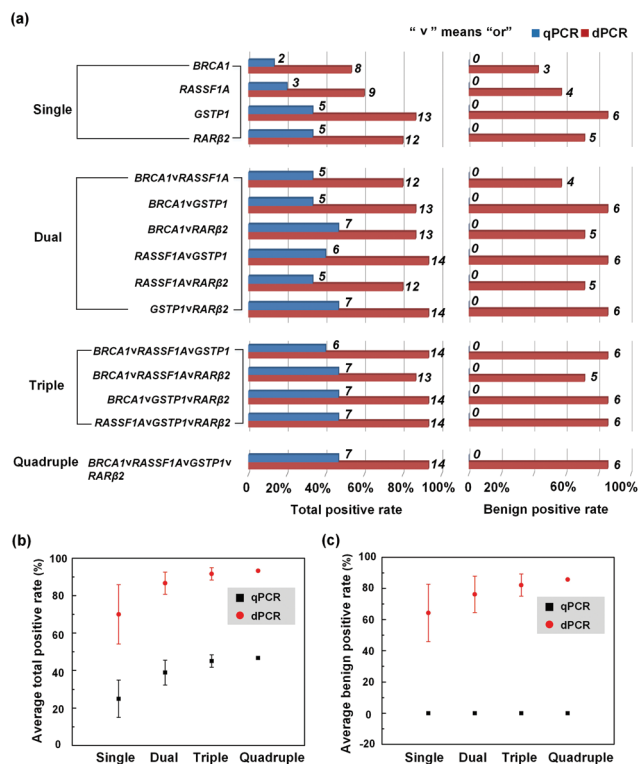
Fig. 4 Positive rates for detecting breast cancer related genes in "or" format gene combinations with qPCR and dPCR. (a) Total positive rates and benign positive rates for detecting *BRCA1*, *RASSF1A*, *GSTP1* and *RARβ2* in "or" format gene combinations with qPCR and dPCR (use confirmed 7 benign samples and 8 malignant samples). (b) The average total positive rates with four kinds of "or" format gene combinations. (c) The average benign positive rates with four kinds of "or" format gene combinations.

tive rate (both the total positive rate and benign positive rate) increases with increasing number of genes in combinations from single to quadruple, either for the qPCR or dPCR assay (Fig. 4(a)). These results indicate that the dPCR assay has over 100% performance enhancement in the cancer detection rate compared to the qPCR assay. For further quantitative comparison, we quantified the average total positive rate (average value of total positive rates from gene combinations that have the same number of genes) and average benign positive rate (average value of benign positive rates from gene combinations that have the same number of genes) against the amount of genes on analysis (Fig. 4(b) and 3(c)). It indicates that with more genes on analysis, that is, from single to quadruple, the average total positive rate and average benign positive rate constantly increase by ~20% for both qPCR and dPCR assays. In Fig. 4(c), the dPCR assay's average benign positive rate reaches ~65% to ~85% corresponding to the single and quadruple situation in the "OR" format assays. Not surprisingly, the average benign positive rates of qPCR in any combinations are 0%.

Positive rates for detecting breast cancer related genes in "AND" format assays with qPCR and dPCR

In the "AND" format assays, when more genes are involved in multiple gene analysis, the positive detection rates of cancer for qPCR and dPCR assays both decrease at first and then become stable and reliable, which can be seen from the decrease of standard deviation (SD) from each data point (Fig. 5). Because of the increasing stability and the reliability of positive detection rates as the gene number increases, this kind of multiple gene combination detection strategy may
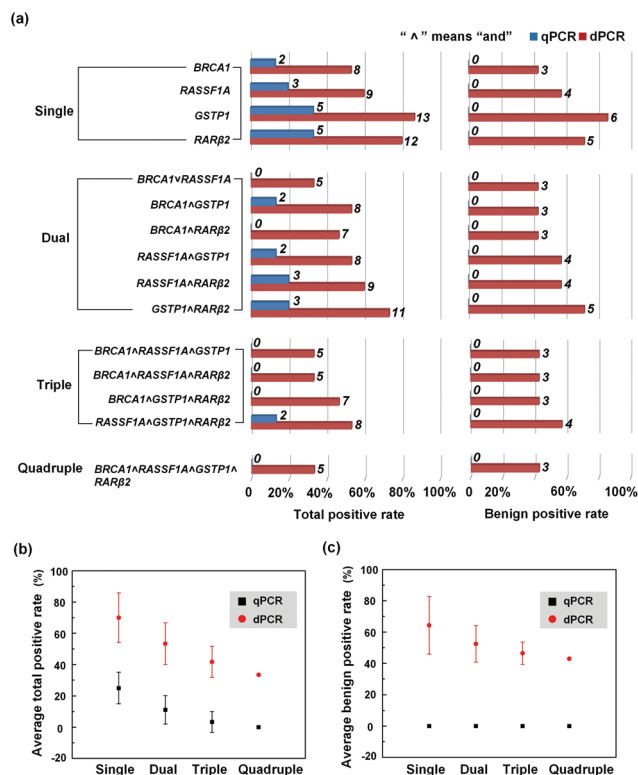
Fig. 5 Positive rates for detecting breast cancer related genes in "and" format gene combinations with qPCR and dPCR. (a) Total positive rates and benign positive rates for detecting *BRCA1*, *RASSF1A*, *GSTP1* and *RARβ2* in "and" format gene combinations with qPCR and dPCR (use confirmed 7 benign samples and 8 malignant samples). (b) The average total positive rates with four kinds of "and" format gene combinations. (c) The average benign positive rates with four kinds of "and" format gene combinations.

Fig. 6 Cut-off values for hypermethylated alleles to classify the healthy, benign and malignant cases with qPCR and dPCR. For qPCR (left *y* axis for black data points): C_q values >35 refers to the healthy and the benign; C_q < = 35 indicates the malignant; for dPCR (right *y* axis for red data points), the data based cut-off values for each allele were plotted.

hold great potential for the prognosis of cancer and tracing the metastases of cancer tissue.[31–33] We observed that the dPCR assay shows higher total positive rate which is nearly triple than the qPCR assay, no matter whether in single, dual, or triple gene combinations (Fig. 5(a) and (b)). Notably, from single gene to quadruple genes analysis, with increasing number of genes, qPCR can detect fewer and fewer combinations (4, 4, 1 and 0 for single, dual, triple and quadruple, respectively) (Fig. 5(a)). As for the benign positive rate, none has been detected by qPCR. But the dPCR assay works outstandingly where approximately half of the benign cases can be detected by the dPCR assay (Fig. 5(a) and (c)). What's more, as the number of gene combinations increases (*i.e.*, detecting more target genes in one single sample), the variation of the detection rate becomes smaller, to some extent, and the detection rates are more stable and credible as reflected by the decrease of SD in each data point (Fig. 5(b) and (c)).

Cut-off values for hypermethylated alleles to classify the healthy, benign and malignant cases with qPCR and dPCR

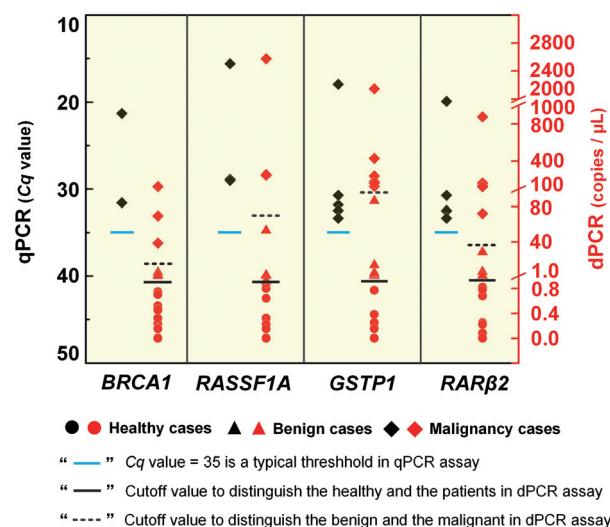To demonstrate the significance of the quantitative results, we correlated the C_q values measured by the qPCR assay to the absolute concentration (copies per μL) obtained by the dPCR assay (Fig. 6). By doing so, we can also set up cut-off values for each hypermethylated allele in the dPCR assay for the clinical diagnosis of breast cancer using plasma samples. The following is a list of the estimated cut-off values for each allele in the order of the healthy, the benign and the malignant cases: for *BRCA1*, the cut-off values are <1 copy per μL, 1–30 copies per μL and >30 copies per μL, respectively; for *RASSF1A*, they are <1 copy per μL, 1–80 copies per μL and >80 copies per μL, respectively; for *GSTP1*, they are <1 copy per μL, 1–100 copies per μL and >100 copies per μL, respectively; and for *RARβ2*, they are <1 copy per μL, 1–40 copies per μL and >40 copies per μL, respectively.

DNA methylation biomarkers, which are essential in the emerging liquid biopsy of cancers, are inaccessible with the existing qPCR methods due to their extremely low abundance in the plasma. Herein, we further developed a microwell chip dPCR-based MethyLight assay for the detection of DNA methylation biomarkers for cancer diagnosis. Digital PCR is based on the limit dilution of the DNA sample into a large number of separate PCR reactions.[34,35] With sufficient dilutions and proper compartment numbers, each reaction compartment will statistically just contain one "1" or no "0" template DNA molecule, while compartments that hold two or more templates would be excluded in subsequent data processing under the guidance of the Poisson distribution.[36,37] Thereafter, the "1" compartments emit fluorescence, while the "0" compartments do not during the hydrolysis probe-based amplification. The number of positive reactions allows template quantification without a standard fluorescence curve. The advantages of the dPCR assay are technically simpler to

perform than the qPCR-based MethyLight assay. And it does not require standard curves to realize the absolute quantification of the targets. However, it's not easy to obtain the ideal quantification results as expected with digital PCR. Any inaccuracy in limiting the dilution step or a trace of nucleic acid contaminant can change the results completely due to the ultrahigh sensitivity of dPCR. But, a proper dilution method, repeated tests and a nucleic acid-free testing environment would lessen or even eliminate the influence. In this work, using the standard DNA samples, we have proven that the LOD and LOQ of the dPCR assay are supersensitive as expected, with 10^3–10^4-fold improvement in LOD and 1–16-fold enhancement in LOQ compared to the conventional qPCR-based MethyLight assay.

For each dPCR assay, the detection accuracy can be evaluated by the measured precision value. Typically, the default precision value is equal to 10%, which means 90% confidence level. In our study, the dPCR assay shows greater accuracy and sensitivity. Since the template concentrations of 0.1 pg μL^{-1} and 0.05 pg μL^{-1} are out of the detection range of dPCR (data not shown), the linear range of the dPCR assay is determined to be from 0.025×10^{-3} pg μL^{-1} to 0.012×10^{-3} pg μL^{-1} (gene NM_000209) or 0.09×10^{-3} pg μL^{-1} (gene NM_033507) as presented in Fig. 3(a) and (b), which is fully beyond the LOD of the qPCR assay. Besides, 1–16-fold enhancement of the LOQ achieved in the dPCR assay can be further improved to some extent by increasing the amount of the total input DNA.

Multiple gene analysis is of great significance in cancer diagnosis.[38,39] The analysis and diagnosis of multiple disease related genes can improve the accuracy of diagnosis results. It usually consists of two assay formats. In early cancer diagnosis, the "OR" format assay could improve the detection rate of cancers, while the "AND" format assay could provide directive information on the therapeutic treatment in the management of cancer treatment or in prognosis. In this work, we employed breast cancer as a proof of concept for these two formats of multiple gene analysis. Four hypermethylated alleles (*RARβ2*, *BRCA1*, *GSTP1* and *RASSF1A*) associated with breast cancer genesis and development were quantified by the qPCR assay and the dPCR assay, respectively. We found that when more genes are included in the assay, the dPCR assay shows much more improved detection ability than the qPCR assay, no matter whether the assay format is "OR" or "AND". For example, when detecting quadruple genes, the dPCR assay has one-fold improvement in terms of the total positive rate in the "OR" format and 33.3% improvement in the "AND" format compared to the qPCR-based MethyLight assay (Fig. 4 and 5).

As for the clinical diagnosis of breast cancer through plasma samples, seven benign cases and eight malignant breast cancers have been discriminated unambiguously by our dPCR assay. Based on the results of the dPCR assay, we set up the cut-off values of each gene for discriminating the healthy, the benign and the malignant. However, the qPCR assay can only detect the malignant. Because the concentration of methylated alleles in the malignant cancers is much higher than the threshold value of the qPCR assay (C_q value = 35

refers to qPCR or ~100 copies per µL refers to dPCR), they can be easily detected. In our findings, the concentrations of the four hypermethylated alleles in the benign cases are often below 100 copies per µL. It is noteworthy that for breast cancer diagnosis through DNA methylation biomarkers, our results show that $GSTP1 \lor RAR\beta2$ provides the highest total positive rate (93.3%) among the four selected alleles in the "OR" format of both the dPCR and the qPCR-based MethyLight assays. This gene combination should be considered as preferred biomarkers for the early diagnosis of breast cancer or personalized genetic testing.

Conclusions

In summary, we developed a microwell chip dPCR-based MethyLight assay, and verified its performance in LOD and LOQ of standard samples compared with the qPCR-based MethyLight assay. Based on the chip dPCR, the circulating DNA methylation biomarkers in the human plasma samples of breast tumor patients were absolutely quantified, and the cancer detection rates of dPCR and qPCR were compared to the clinic standards. With the dPCR assay, we have achieved 10^3–10^4-fold lower LOD and 1–16-fold lower LOQ than the conventional qPCR assay for the synthesized DNA samples. The results of the absolute quantification of hypermethylated alleles (*RARβ2*, *BRCA1*, *GSTP1* and *RASSF1A*) in breast cancer samples indicated that the chip dPCR assay exhibits one-fold enhancement in the cancer detection rate over the qPCR assay, due to superior sensitivity. Promisingly, the microwell chip dPCR-based MethyLight assay showed an unambiguous discrimination of benign tumor cases from the malignant ones, while the qPCR assay can only detect the malignant cases. The potential benefit of this work would be that the number of patients who with the disease risk need to turn to aspiration biopsy for a final diagnosis, which is usually painful and time-consuming, will decrease. Thus, dPCR is of importance for early screening diagnosis and timely intervention of the disease. In addition, since a small amount of the blood sample is sufficient for dPCR detection, the way of the dPCR-based liquid biopsy for cancer biomarkers is relatively noninvasive and suitable for long-term monitoring of patients. We envision that the chip dPCR MethyLight assay would be a powerful tool for the *in vitro* diagnosis of cancer patients as well as for post-treatment monitoring.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 H. Schwarzenbach, D. S. Hoon and K. Pantel, *Nat. Rev. Cancer*, 2011, **11**, 426–437.

2 M. S. Pepe, R. Etzioni, Z. Feng, J. D. Potter, M. L. Thompson, M. Thornquist, M. Winget and Y. Yasui, *J. Natl. Cancer Inst.*, 2001, **93**, 1054–1061.

3 D. Psaltis, S. R. Quake and C. Yang, *Nature*, 2006, **442**, 381–386.

4 M. J. Barry and J. H. Hayes, *J. Am. Med. Assoc.*, 2015, **314**, 2073–2074.

5 N. S. Kapoor, L. D. Curcio, C. A. Blakemore, A. K. Bremner, R. E. McFarland, J. G. West and K. C. Banks, *Ann. Surg. Oncol.*, 2015, **22**, 3282–3288.

6 H. G. Welch, D. H. Gorski, P. C. Albertsen, G. L. Andriole, E. D. Crawford, R. L. Grubb III, F. H. Schröder, J. Hugosson and M. J. Roobol, *N. Engl. J. Med.*, 2016, **374**, 594–596.

7 M. Sanchez-Cespedes, M. Esteller, K. Hibi, F. O. Cope, W. H. Westra, S. Piantadosi, J. G. Herman, J. Jen and D. Sidransky, *Clin. Cancer Res.*, 1999, **5**, 2450–2454.

8 R. Grützmann, B. Molnar, C. Pilarsky, J. K. Habermann, P. M. Schlag, H. D. Saeger, S. Miehlke, T. Stolz, F. Model and U. J. Roblick, *PLoS One*, 2008, **3**, e3759.

9 W. A. Palmisano, K. K. Divine, G. Saccomanno, F. D. Gilliland, S. B. Baylin, J. G. Herman and S. A. Belinsky, *Cancer Res.*, 2000, **60**, 5954–5958.

10 Y. Hashimoto, T. J. Zumwalt and A. Goel, *Epigenomics*, 2016, **8**, 685–703.

11 M. Swellam, M. D. E. Abdelmaksoud, M. Sayed Mahmoud, A. Ramadan, W. Abdel-Moneem and M. M. Hefny, *IUBMB Life*, 2015, **67**, 61–68.

12 Q. Tang, J. Cheng, X. Cao, H. Surowy and B. Burwinkel, *Clin. Epigenet.*, 2016, **8**, 115.

13 L. Maldonado, M. Brait, M. Loyo, L. Sullenberger, K. Wang, S. B. Peskoe, E. Rosenbaum, R. Howard, A. Toubaji and R. Albadine, *J. Urol.*, 2014, **192**, 1542–1548.

14 N. Yamamoto, T. Nakayama, M. Kajita, T. Miyake, T. Iwamoto, S. J. Kim, A. Sakai, H. Ishihara, Y. Tamaki and S. Noguchi, *Breast Cancer Res. Treat.*, 2012, **132**, 165–173.

15 I. Nawaz, X. Qiu, H. Wu, Y. Li, Y. Fan, L.-F. Hu, Q. Zhou and I. Ernberg, *Epigenetics*, 2014, **9**, 1138–1148.

16 M. Szyf, P. Pakneshan and S. A. Rabbani, *Biochem. Pharmacol.*, 2004, **68**, 1187–1197.

17 E. Crowley, F. Di Nicolantonio, F. Loupakis and A. Bardelli, *Nat. Rev. Clin Oncol.*, 2013, **10**, 472–484.

18 N. Redshaw, J. F. Huggett, M. S. Taylor, C. A. Foy and A. S. Devonshire, *BMC Genomics*, 2014, **15**, 1174.

19 C. A. Eads, K. D. Danenberg, K. Kawakami, L. B. Saltz, C. Blake, D. Shibata, P. V. Danenberg and P. W. Laird, *Nucleic Acids Res.*, 2000, **28**, e32–e00.

20 J. G. Herman, J. R. Graff, S. Myöhänen, B. D. Nelkin and S. B. Baylin, *Proc. Natl. Acad. Sci. U. S. A., Early Ed.*, 1996, **93**, 9821–9826.

21 B. J. Hindson, K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, M. Y. Lucero, A. L. Hiddessen and T. C. Legler, *Anal. Chem.*, 2011, **83**, 8604–8610.

22 F. Rothe, J. F. Laes, D. Lambrechts, D. Smeets, D. Vincent, M. Maetens, D. Fumagalli, S. Michiels, S. Drisis and C. Moerman, *Ann. Oncol.*, 2014, **25**, 1959–1965.

23 C. M. Hindson, J. R. Chevillet, H. A. Briggs, E. N. Gallichotte, I. K. Ruf, B. J. Hindson, R. L. Vessella and M. Tewari, *Nat. Methods*, 2013, **10**, 1003–1005.

24 L. Cao, X. Cui, J. Hu, Z. Li, J. R. Choi, Q. Yang, M. Lin, L. Y. Hui and F. Xu, *Biosens. Bioelectron.*, 2016, **90**, 459–474.

25 L. Barault, A. Amatu, F. E. Bleeker, C. Moutinho, C. Falcomatà, V. Fiano, A. Cassingena, G. Siravegna, M. Milione and P. Cassoni, *Ann. Oncol.*, 2015, **26**, 1994–1999.

26 M. Yu, K. T. Carter, K. W. Makar, K. Vickers, C. M. Ulrich, R. E. Schoen, D. Brenner, S. D. Markowitz and W. M. Grady, *Epigenetics*, 2015, **10**, 803–809.

27 G. Nikolaidis, O. Y. Raji, S. Markopoulou, J. R. Gosney, J. Bryan, C. Warburton, M. Walshaw, J. Sheard, J. K. Field and T. Liloglou, *Cancer Res.*, 2012, **72**, 5692–5701.

28 E. Olkhov-Mitsel, D. Zdravic, K. Kron, T. van der Kwast, N. Fleshner and B. Bapat, *Sci. Rep.*, 2014, **4**, 4432.

29 M. O. Hoque, Q. Feng, P. Toure, A. Dem, C. W. Critchlow, S. E. Hawes, T. Wood, C. Jeronimo, E. Rosenbaum, J. Stern, M. Yu, B. Trink, N. B. Kiviat and D. Sidransky, *J. Clin. Oncol.*, 2006, **24**, 4262–4269.

30 L. Barault, A. Amatu, F. E. Bleeker, C. Moutinho, C. Falcomata, V. Fiano, A. Cassingena, G. Siravegna, M. Milione, P. Cassoni, F. De Braud, R. Ruda, R. Soffietti, T. Venesio, A. Bardelli, P. Wesseling, P. de Witt Hamer, F. Pietrantonio, S. Siena, M. Esteller, A. Sartore-Bianchi and F. Di Nicolantonio, *Ann. Oncol.*, 2015, **26**, 1994–1999.

31 S. Roessler, H. L. Jia, A. Budhu, M. Forgues, Q. H. Ye, J. S. Lee, S. S. Thorgeirsson, Z. Sun, Z. Y. Tang, L. X. Qin and X. W. Wang, *Cancer Res.*, 2010, **70**, 10202–10212.

32 S. Asaga, C. Kuo, T. Nguyen, M. Terpenning, A. E. Giuliano and D. S. Hoon, *Clin. Chem.*, 2011, **57**, 84–91.

33 S. J. Dawson, D. W. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S. F. Chin, M. J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas and N. Rosenfeld, *N. Engl. J. Med.*, 2013, **368**, 1199–1209.

34 B. Vogelstein and K. W. Kinzler, *Proc. Natl. Acad. Sci. U. S. A., Early Ed.*, 1999, **96**, 9236–9241.

35 M. Baker, *Nat. Methods*, 2012, **9**, 541–544.

36 V. Marx, *Nat. Methods*, 2014, **11**, 241–245.

37 J. F. Huggett, C. A. Foy, V. Benes, K. Emslie, J. A. Garson, R. Haynes, J. Hellemans, M. Kubista, R. D. Mueller,

T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, C. T. Wittwer and S. A. Bustin, *Clin. Chem.*, 2013, **59**, 892–902.

38 V. Taly, D. Pekin, L. Benhaim, S. K. Kotsopoulos, D. Le Corre, X. Li, I. Atochin, D. R. Link, A. D. Griffiths, K. Pallier, H. Blons, O. Bouche, B. Landi, J. B. Hutchison and P. Laurent-Puig, *Clin. Chem.*, 2013, **59**, 1722–1731.

39 F. Shen, B. Sun, J. E. Kreutz, E. K. Davydova, W. Du, P. L. Reddy, L. J. Joseph and R. F. Ismagilov, *J. Am. Chem. Soc.*, 2011, **133**, 17705–17712.