

Strategies for validation and testing of DNA methylation biomarkers

Christa Noehammer¹, Walter Pulverer¹, Melanie R. Hassler², Klemens Vierlinger¹, Manuela Hofner¹, Matthias Wielscher¹, Triantafillos Liloglou³, David McCarthy⁴, Taylor J. Jensen⁵, Anders Nygren⁶, Henning Gohlke⁶, Geert Trooskens⁷, Maarten Braspenning⁸, Wim Van Criekinge⁷, Gerda Egger², Andreas Weinhaeusel^{1†}

¹AIT Austrian Institute of Technology, Health & Environment Department, Molecular Diagnostics, Vienna, Austria

²Clinical Institute of Pathology, Medical University of Vienna, Vienna, Austria

³University of Liverpool, Department of Molecular & Clinical Cancer Medicine, Liverpool L3 9TA, UK

⁴RiboMed Biotechnologies, Inc., Carlsbad, CA 92010, USA

⁵Sequenom Center for Molecular Medicine, San Diego, CA, 92121, USA

⁶Sequenom, Inc, San Diego, CA, 92121, USA

⁷Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, B-9000, Ghent, Belgium

⁸ NXT-Dx, Fr. Rooseveltlaan 349/B.43, 9000 Gent, Belgium.

† Author for Correspondence: AIT Austrian Institute of Technology, Health & Environment Department, Molecular Diagnostics, Muthgasse 11/2, 1190 Vienna, Austria. Tel. ++43-50550-7622, Fax ++43-50550-4450, andreas.weinhaeusel@ait.ac.at

ABSTRACT (120 words)

DNA methylation is a stable covalent epigenetic modification of primarily CpG dinucleotides that has recently gained considerable attention for its use as biomarker in different clinical settings, including disease diagnosis, prognosis and therapeutic response prediction. Although the advent of genome-wide DNA methylation profiling in primary disease tissue has provided a manifold resource for biomarker development, only a tiny fraction of DNA methylation-based assays have reached clinical testing. Here, we provide a critical overview of different analytical methods that are suitable for biomarker validation, including general study design considerations, which might help to streamline epigenetic marker development. Furthermore, we highlight some of the recent marker validation studies and established markers that are currently commercially available for assisting in clinical management of different cancers.

KEYWORDS (5-10)

DNA methylation, biomarker, assay validation, circulating free DNA, Bisulfite-deamination, qPCR, deep sequencing, pyrosequencing, MALDI

BODY of ARTICLE

DNA methylation testing has become a major approach in biomarker development. Several different concepts of DNA methylation testing for candidate marker confirmation and validation have been developed over the recent years. With respect to diagnostics of human disease, most efforts focus on the testing of 5-methylcytosine (5mC) within CpG dinucleotides. Although non-CpG methylation has been confirmed in stem cells in the last 5 years [1] and elucidation of biological function of modified 5mC as well as 5-hydroxymethylcytosine (5hmC) has started, the major work in the methylation biomarker field is concentrating on 5mC in the CpG context. Today genome-wide technologies are mostly used for discovery of methylation biomarkers. For confirmation of findings from the initial screenings, as well as for validation of markers in large patient cohorts, a number of different methylation testing methods are available that rely on one of the following 3 basic principles: (1) bisulfite deamination – where unmethylated cytosine is converted to uracil and methylated cytosine is resistant to conversion, (2) methylation sensitive restriction enzymes which cut DNA depending on the presence or absence of 5mC and (3) affinity based methods, using proteins for fractionation of methylated versus unmethylated DNA. Based on these 3 principles, we focus in this review on methods enabling (1) quantitative analyses, which is a prerequisite for successful biomarker analyses, (2) parallel testing of multiple markers, as well as on (3) methods which have the capability for high throughput analyses. Along these lines we have compiled an overview of methods fulfilling to our best knowledge these prerequisites for optimal confirmation and validation of methylation marker panels and ideally enable the parallel testing of even hundreds of candidate markers.

										setup- cost (96 targets@ 80% success rate)* incl design	analyses cost / 96 targets x 384 samples - "selfmade - assay"			
Method	Commercial assays (research panels) available	Bisulfite-Treatment	Input ng gDNA per assay	single CpG resolution	Multiplexing	high throughput	equipment	dynamic range (orders)	quantification	person hours (minimal)	**material (incl Oli-gos)	person hours **	material (excl. oligos - using a 384 well qPCR machine)	internal standard
Bisulfite-PCR	MethylLight (Qiagen)	yes	10-100ng	-	moderate	+	qPCR	3(4?)	calibration curve	160	22000	272	20000	SssI-DNA & UM DNA
MSRE-PCR	SABiosciences/Qiagen	no	10-100ng	-	high* (48x preamplification)	+++	qPCR	3	calibration curve	160	3200	40 -272	(6000 ***)- 20000	undigested DNA
Pyro-sequencing	PyroMark CpG Assays (Qiagen)	yes	10-100ng	++	no	++	Pyromark Q96ID/MD, Q24	2	direct (%)	260	5000	384-768	70000	SssI-DNA & UM DNA
Deep Sequencing	MethylSeq (Raindance)	yes		+++ (and single amplicon)	high** (pooled PCR products)	+	NGS	4	direct (read numbers)	depends on NGS system	depends on NGS system	depends on NGS system	depends on NGS system	SssI-DNA & UM DNA
MALDI-based DNA methylation detection	Epityper (Sequenom)	yes	10ng	+	no	+++	MALDI-TOF	3	direct (%)	120	89150	48	21120	-
MBD	various **	no	1-100ng	-	moderate	++	various*	3	Ratio (ME/total)	140	4500	272	15000	SssI-DNA & UM DNA*
	** MethylMagnet - RiboMed; MethylQuest - Millipore; MethylCap - Diagenode; MethylMiner - Life Technologies; MethylCollector - Active Motif				*- multiplexed preamplification and single qPCR readout (not guilty for SABiosciences kit); **- single PCRs - pooled sequencing upon barcoding		*-LC-MS with Absorption, qPCR			* according MIQE	** oligos calc @ 0.2€/base; probe calc @200€; 0.5€/Rxn	** 3x 384-well PCR/day (8.5h)	***HTqPCR -Biomark	*-parallel samples

Table 1: Overview and characteristics of methods for DNA methylation biomarkers testing and validation

Methods and strategies for DNA methylation testing & validation.

Bisulfite based quantitative PCR testing

Bisulfite treatment of genomic DNA provides a suitable option to distinguish methylated from unmethylated cytosine residues by different downstream applications [2]. Several quantitative assays have been developed in order to interrogate the methylation status of selected loci. Generally, two different assay setups can be distinguished; either PCR reactions are performed to amplify bisulfite converted DNA irrespective of methylation status (MIP) (no CpGs in primer sequence) or PCR reactions are methylation specific (MSP) (CpGs within primer sequence).

The first established quantitative MSP assay was MethyLight, which employs MSP combined with methylation-sensitive probing (TaqMan®) [3]. This highly sensitive assay is capable of detecting methylated DNA in a 10000-fold excess of unmethylated DNA, which makes it suitable for the analysis of complex heterogeneous material such as clinical samples. Several quality control reactions controlling for sample integrity and quantity, sample recovery after bisulfite conversion as well as bisulfite conversion efficiency are advisable for assay setup [4]. In order to specifically amplify methylated DNA primers and probe should contain 1-5 CpGs, which can be designed using MethMarker, a platform for the design and optimization of gene-specific DNA methylation assays [5]. Four PCR reactions using two different samples are needed to determine the methylation level of a selected region; the bisulfite converted DNA of the sample of interest and bisulfite converted 100% methylated DNA as a reference (*in vitro* MSssI treated) are amplified using methylation-specific primers and probe for the gene of interest plus a methylation independent, bisulfite conversion specific set of primers and probe for a reference locus such as the repetitive ALU-C4 locus to control for DNA input. The methylation level is then calculated using the PMR (percentage of methylated ratio) value, which gives a relative measure of DNA methylation, based on a standard curve of a dilution of MSssI treated completely methylated DNA [4]. MethyLight assays can be multiplexed using different fluorescent labeled probes allowing for high-throughput applications [6]. A higher grade of sensitivity and accurate quantitation can be obtained by employing digital MethyLight, which allows for the detection of single molecule DNA methylation [7].

This methodology is found in the literature frequently as quantitative MSP (qMSP) and seems to be a method of preference when testing targets in clinical specimens such as bronchial washings [8] and serum [9] and biopsy material [10].

A more economic variation of MethyLight represents sensitive melting analysis after real time methylation specific PCR (SMART-MSP), which relies on probe-free MSP using DNA intercalating fluorescent dyes combined with high resolution melting (HRM) analysis, which allows for the identification of false positives [11]. Evaluation of melting curves subsequent to PCR amplification provides information relating to the specificity of the reaction. Incomplete bisulfite conversion and false priming yields shifted melting curves compared to fully methylated samples and provides an additional quality control. Heterogeneous DNA methylation can also be detected based on altered melting curves providing an advantage compared to conventional MethyLight, which only detects highly methylated sequences. As for MethyLight, the methylation level of a gene of interest is determined based on a methylation-independent internal control and a calibrator sample of 100% methylated DNA. Dilution of the methylated standard down to 0.1% could be reproducibly detected, thus resulting in sensitivity comparable to MethyLight.

Two further assays with high analytical sensitivity include HeavyMethyl and MS-HRM, which are both based on methylation independent priming of bisulfite converted DNA [12,13]. HeavyMethyl uses

blocking oligos, binding to unmethylated DNA, which overlaps with primer binding sequences, thus allowing for amplification of methylated sequences only. Quantification of methylated DNA is obtained by fluorescent probing and by using PMR calculations analog to the MethyLight approach. HeavyMethyl assays were reported to reach a relative sensitivity of 1:8000, detecting a few copies of methylated DNA in 400ng of non-methylated background DNA [12]. The HeavyMethyl assay is used for the commercially available kits “Epi proColon” and “Epi proLung” by Epigenomics AG to detect DNA methylation in colon and lung cancer, respectively [14,15].

Methylation specific high resolution melting (MS-HRM) analysis, which also uses MIP to amplify bisulfite converted DNA, quantifies methylation levels based on melting profiles of amplicons of bisulfite converted DNA. Highly methylated sequences contain a larger number of CpGs in their sequence and therefore have higher melting temperatures compared to less methylated samples, which contain more TpGs and thus shifted melting profiles. Methylation levels are estimated based on melting curves of PCR products of standards with known methylation status and methylation levels as low as 0.1% can be detected [13]. Drawbacks of this method might be difficulties in interpreting melting curves of heterogeneous methylated samples.

In sum, bisulfite-based qPCR is highly sensitive and suitable for high throughput analyses and some assays are already in use for clinical testing.

Quantitative DNA Methylation analysis with methylation sensitive restriction enzymes (MSRE)

One of the first technologies for the investigation of DNA methylation made use of restriction enzymes (REs) [16]. Today DNA methylation analysis typically involves the use of MSREs that cut only unmethylated DNA but not methylated DNA (e.g. *Acil*, *HpaII*). The combination with quantitative PCR (qPCR) based detection enables a reliable and simple detection of DNA methylation targeting native DNA sequences. Consequently, only methylated DNA is specifically amplified during PCR [17]. In contrast to MSREs, a very limited number of REs is available that cut only methylated DNA, like *GlaI*, *McrBC* and *SgeI* [18,19].

Complete digestion of the DNA is an essential step prior to amplification by PCR as even smallest amounts of uncleaved DNA are detectable and lead to false-positive results. For that reason an appropriate assay design is mandatory. To ensure complete cleavage of the DNA, a minimal number of at least 2-3 cut sites is recommended within the target sequence. The combination of different MSREs is also recommended as it increases the number of possible restriction sites. Such multi digests further compensate for incomplete digestion, which might be caused by the use of just one MSRE. Additionally a combination of enrichment methods (e.g. by MBD-proteins) and MSRE based methylation detection may increase the sensitivity of qPCR results [20].

Many open source tools for primer design are available online (e.g. Primer3 [21]). However, the design for MSRE (q)PCR assays is more tricky, as there is no design tool available that considers the cut sites per se. Therefore, it is necessary to check for the number of cut sites present in a defined PCR amplicon, preferentially using a genome browser (e.g. UCSC genome browser [22]).

To qualify assays and to exclude experimental bias in every analyzed data set, primers should be tested prior to use as suggested in MIQE guidelines by Bustin SA et al. A serial dilution covering ideally 3 log units per analyzed primer pair is usually sufficient to create a 5 point standard curve for every primer as suggested by the MIQE guidelines [23]. The PCR efficiency, slope, intercepting point with

y axis and the correlation coefficient may be given for every PCR-assay to allow an estimation of the influence of the assay performance on the qPCR data as already implemented into an updated delta Ct method by Pfaffl [24] resulting in a PCR-efficiency corrected delta Ct method. In terms of MSRE specific controls PCR values may be corrected for input DNA amounts by subtraction from methylation specific controls, which may be imprinted loci or even more straightforward genomic DNA fragments without MSRE restriction sites [25,26]. Comparable to qMSP every sample may also be divided into a mock and a MSRE reaction followed by qPCR, which also allows an assessment of the portion of methylated DNA for a specific locus as shown by Pulverer et al. [27]

Assays for methylation analysis often target sequences of high GC content, therefore optimization of the assay with different PCR enhancers like DMSO (reduction of secondary structure, facilitates amplification of GC rich templates) or TMAC (prevents unspecific priming) is often advisable to increase assay performance. It is also highly recommended to control the DNA for complete cleavage after the digestion step by assays targeting genes with known methylation status [26].

Depending on the number of investigated targets, the digestion reaction should start with 100 ng of genomic DNA, an amount which can easily be upscaled. About 5-10 ng of digested DNA should be applied to the final qPCR reaction. It has been demonstrated that 1% of methylated DNA can be distinguished from a complete lack of methylation with a limit of detection of 19.58 pg methylated DNA [27]. Serial dilutions of methylated DNA in unmethylated DNA yielded recovery rates between 99%-155%. The higher the content of methylated DNA the more precisely was the recovery rate. Thus, MSRE coupled qPCR assays are suitable for the determination of methylated DNA in heterogeneous samples containing both methylated and unmethylated fractions [25,27,28].

Overall MSRE qPCR assays are an alternative to the DNA degrading bisulfite based methods and allow accurate methylation testing.

Pyrosequencing-based methylation analysis (PMA)

Pyrosequencing is a highly versatile methodology offering significant advantages in DNA methylation analysis [29]. The main advantages of PMA are (a) the ability to quantitatively interrogate multiple CpGs, (b) include multiple bisulfite controls, (c) being a fast and medium cost method. Limitations are largely due to the low temperature (28 °C) of the reaction, which enables the formation of many secondary structures; this is why careful assay design and template optimisation are very important for successful assays.

A brief description of the method is given in Figure 1. Every CpG position is interrogated by a sequential injection of C and T (or G and A if reverse sequencing primer is used). The methylation percentage is determined by the relative incorporation of the two nucleotides at the site. A representative pyrogram is shown in Figure 2.

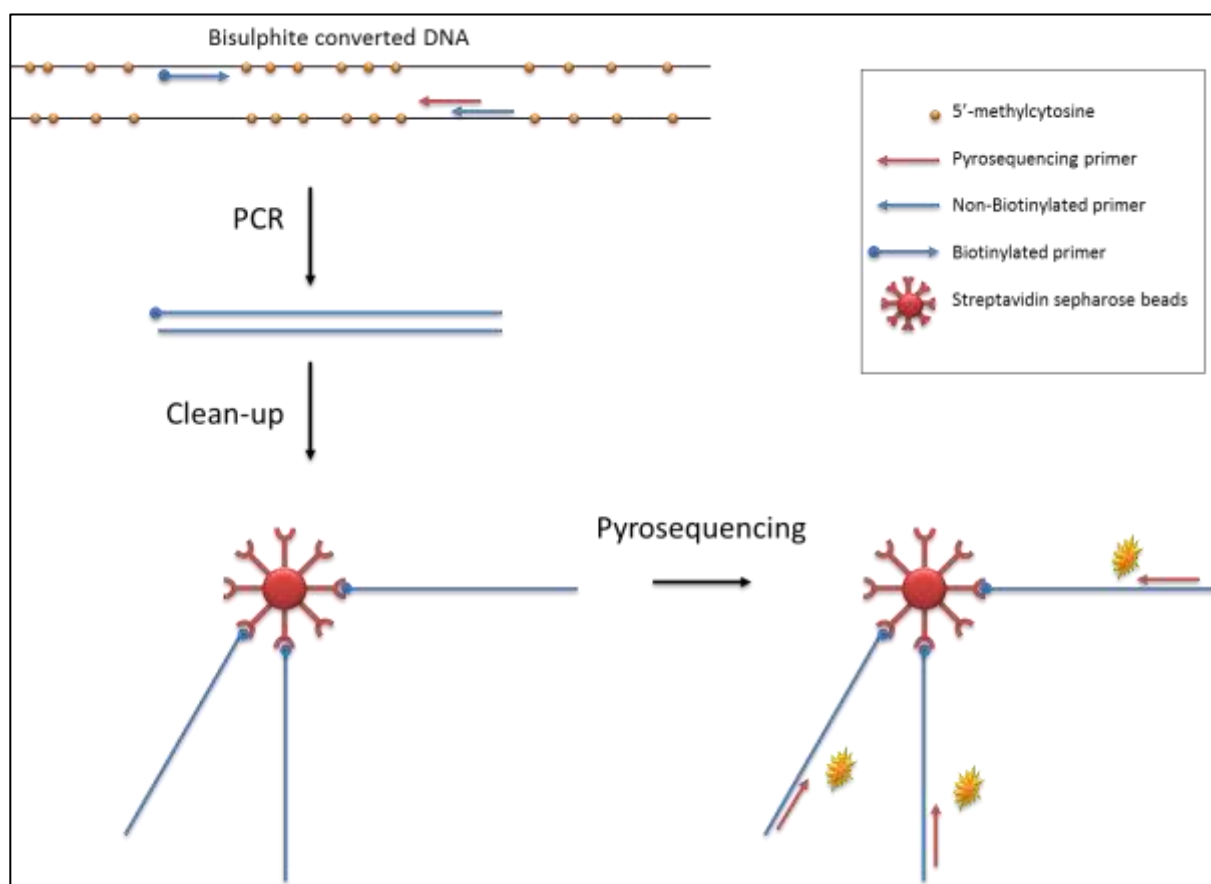


Figure 1: Schematic description of PMA work flow. The CpG bearing sequences of interest are amplified by PCR, after bisulfite conversion, utilising flanking CpG-free regions. One of the PCR primers is biotinylated at its 5' end allowing for a fast (~10 min) post-PCR clean-up involving streptavidin sepharose beads. Following clean up, the single (biotinylated) strand immobilised on the beads is added in the reaction plate containing the annealing buffer and sequencing primer and the plate inserted in the equipment for the reaction to commence. The remaining reagents are dispensed in a pre-determined manner by the equipment through a cartridge.

Assay design demands unique software (Pyromark Assay Design 2.0, Qiagen). It generally requires some experience to achieve optimal designs, as the automated primer selection function rarely results in good designs for DNA methylation, while it works well for the SNP assays, for which it was originally designed. In addition, the T_m calculation in this software is not optimal (to the author's experience) thus additional software (e.g. Primer Express) should be used for this purpose. Here are some brief guidelines for the design:

1. The read length is normally 60-70 nt, although >100 nt is achievable in optimal designs. Thus the amplicon length should be kept fairly short; definitely below 300 bp and optimally up to 150 bp.
2. The primers normally cannot include CpGs and should be avoided if possible. Successful designs may utilise primers with a mismatched nucleotide (G/A) or a mixed nucleotide (C/T, also known as "wobble") at such a C position within the three to four 5' positions of a 22mer for example. However, appropriate optimisation and validation have to be undertaken to eliminate or minimise the potential bias for methylated or unmethylated target copies.
3. Optimally, PCR primer length should be between 18 and 24, with T_m of no less than 48°C (calculation for 200 nM oligo, 50 mM K+). T_m should also not exceed 65°C, but this is extremely un-

likely to be met when targeting bisulfite converted DNA. Typically for PCR, primers should not differ more than 2°C in T_m.

4. Although homopolymers should be generally avoided within PCR primers, this rule will be frequently compromised in bisulfite DNA related designs. Still, if it cannot be avoided, homopolymers should be limited at the 5' of the primer.

Once somebody gets experience, the success rate of designed assays can be over 95%.

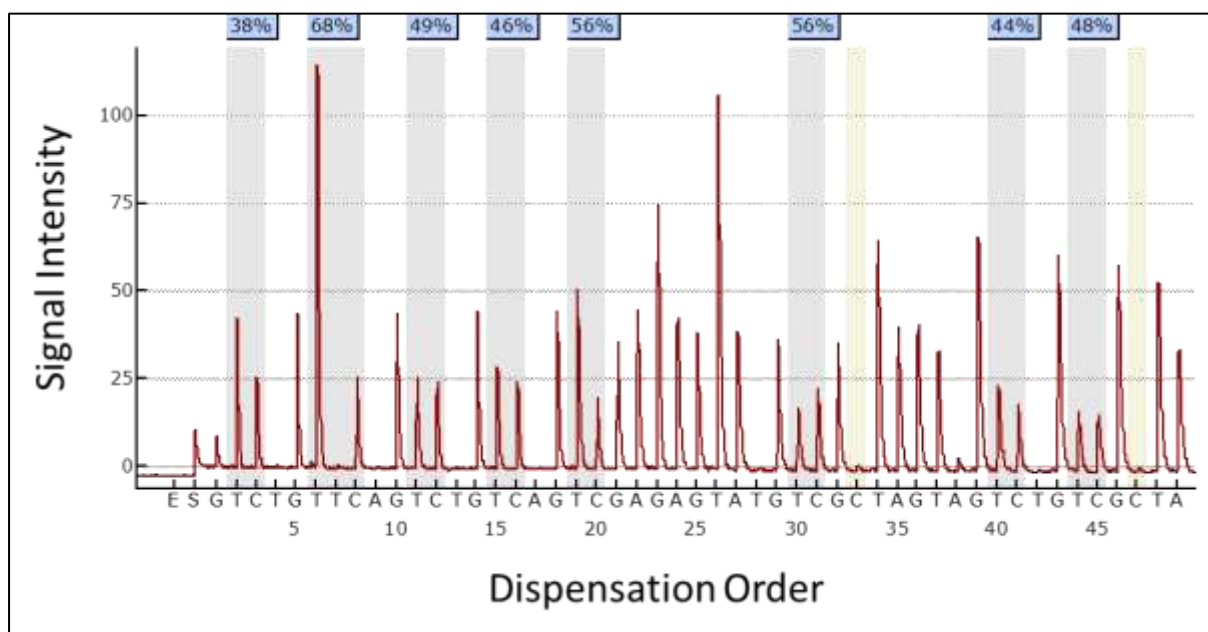


Figure 2: Example of pyrogram from PMA run. Interrogated CpGs are indicated by the grey lanes. The percentages in the boxes above demonstrate the degree of methylation detected. The color of the box reflects the quality control result (blue= pass, yellow=check, red=fail). Percent methylation (%) is calculated by the software as $[C/(C+T)]$ for each CpG dinucleotide. The C dispensations at positions 33 and 47 (yellow lanes, corresponding to C not within a CpG dinucleotide) are for bisulfite conversion quality control purposes.

Successful pyrosequencing is heavily dependent on the quality and quantity of the PCR product. Too little product will result in high noise-to-signal ratio, while too much will end up in peak tailing and possible loss of its quantitative efficiency. Typically, PCR optimisation for pyrosequencing involves testing dilutions of the biotinylated primer into non-biotinylated, ranging from 1:1 to 1:2. The use of lower amounts of biotinylated primer reduces the competition of the unused excess primer to the PCR product and therefore enhances the signal. The thermal profile has of course to be optimized, using the highest possible temperature that does not compromise yield. The PCR product has to be checked by agarose electrophoresis and should be free of artifacts, primer dimers etc. No other optimisation, beyond PCR amplicon quality/quantity, is required for the pyrosequencing reaction itself.

The performance and linearity of the assay has to be shown ahead of screening by the inclusion of a standard curve of artificially methylated (normally by *SssI* methyltransferase) into unmethylated DNA. It has to be noted that DNA methylation is tissue specific and therefore normal tissue DNA cannot *de facto* be assumed to be unmethylated for every gene/sequence. Synthetic (whole genome amplified) DNA is preferable for this negative control (unmethylated) purpose.

Data analysis is one of the major advantages of pyrosequencing: it is automated, taking no more than 1 minute. No data pre-processing is required.

As mentioned above, assuming assay design is sound, pyrosequencing performance depends solely on the quality and quantity of the PCR product. DNA inputs between 10 ng and 100 ng in the PCR reaction are expected to provide similar results in the pyrosequencing reaction. The nature of the method does not allow for multiplexing, however one can potentially run 96 different assays per plate; of course this requires 96 different PCR products. As a typical post PCR method utilizing non-allele-specific primers, pyrosequencing can detect reliably DNA methylation down to 5%. Lower levels will be associated with very high variability.

The above mentioned characteristics make PMA ideal for screening sequences in primary disease tissue, either in single target approaches or, frequently, in technical and biological validation of DNA methylation microarray results. To date, it is considered as the gold standard technique for this purpose and has contributed to a very large number of studies [8,30-32].

Targeted deep-amplicon bisulfite sequencing (TDBS)

Bisulfite sequencing applies different sequencing methods on bisulfite-treated genomic DNA in order to determine the methylation status of CpG dinucleotides. Initial sequencing methods used sub-cloning to combine positional information with allelic information. Hereby typically 10-100 clones or alleles were sequenced. With the availability of Next Generation Sequencing technology, which provides a clonal read out, deep sequencing upon bisulfite conversion has become a very effective approach for analyzing methylation patterns. For analyses of clinical samples heterogeneity of sample due to biology as well as due to preprocessing (e.g. formalin fixation) has forced investigators to increase the sequencing depth which would have been almost impossible using the classical cloning approach but is easily achieved using today's massive parallel sequencing options [33]. The massively parallel sequencing can easily and directly sequence >100 bisulfite PCR products in a single sequencing-run without sub-cloning. This technology shows also high robustness, and superiority with respect to multiplexing of e.g. 25 gene-related CpG rich regions from more than 40 individual samples in a single sequencing run [34]. Today's limitation of TDBS is rather the capacity for setting up of PCR assays for multiple candidate loci, whereas analyzing different patient samples is easily achieved when a pool of amplicons from different individuals is ligated with barcoded sequencing adapters during TDBS sample preparation.

For design of assays recommendations already given for bisulfite specific PCR amplification in previous sections should be considered. As for pyrosequencing the primer-sequences have to be designed spanning the region of interest and avoid CpGs within the primers. Due to lower complexity of bisulfite treated DNA, it is more difficult to find suitable primers. However for specific target amplification software like MethPrimer [35] design program, which is based on the popular Primer3 [21] program, supports both MSP and bisulfite sequencing primer (BSP) design. Another example of primer design software for bisulfite converted DNA is BiSearch [36]. The algorithm is not based on other primer design software but starts from an own implementation of the nearest-neighbor method to calculate the melting temperature of the DNA strands. An important feature of BiSearch is the 'specificity check' i.e., the algorithm uses a simple search method to find other targets of the primers in the bi-

sulfite treated reference genome. MSPprimer [37] generates possible primers based on a sliding window. The specificity check of the primers is based on the specificity-determining subsequence (SDSS) theory [38]. The SDSS is the smallest 3' sub- sequence of a primer for which the fraction of the template associated with that subsequence exceeds a given threshold

[39]. For precise quantification of methylation ratios, deep amplicon sequencing requires a methylation independent amplification of the targeted regions. In practice, methylation independent assays are often hard to achieve due to PCR bias, which favors the amplification of unmethylated sequences. Optimization with positive (methylated DNA) and negative (unmethylated DNA) control samples is therefore highly recommended. These controls are available as commercial kits or can be generated from a DNA source [40].

Sequenced bisulfite-treated DNA is mostly devoid of cytosines, which means they will not map without mis-matches to a standard reference genome. Most algorithms [41] unmethylate in silico the remaining cytosines before mapping to the unmethylated reference amplicon, and reconvert the methylated sites to cytosines after mapping.

Comparable to a normal MSP, a limited amount of input bisulfite DNA (10-100ng) is required for deep amplicon sequencing. Multiple assays per sample can be done in multiplex, allowing for testing of multiple markers in multiple samples.

Quantitative DNA methylation analysis with MALDI-TOF mass spectrometry

Region-specific quantitative DNA methylation analysis can be performed using base-specific cleavage coupled to Matrix Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry [42-44]. This methodology (EpiTYPER analysis) enables high-throughput assessment of DNA methylation within target regions, enabling the quantitative assessment of methylation levels for the majority of CpG dinucleotides across a region of interest [45-47].

Proper assay design is a prerequisite for robust performance and consistent interpretation of DNA methylation levels. Two distinct parameters govern assay design: the source, quality, and quantity of the DNA and factors linked to the assay itself. The quantity of amplifiable DNA present during the PCR reaction directly contributes to the quantitative accuracy of an assay [48]. The process of treating DNA with sodium bisulfite results in the degradation of DNA, reducing the amount of effective starting material proportional to the length of the downstream amplicon/assay [49]. When combined with long amplicons or coupled to chemically, biologically, or otherwise fragmented input material, the impact of this treatment can ultimately result in the dramatic reduction of the number of amplifiable DNA molecules. For full length genomic DNA, current bisulfite treatment methods should result in a sufficient number of amplifiable copies even at amplicon lengths reaching 500-600 bp; however, for fragmented DNA including circulating cf (cell free) DNA or FFPE (formalin fixed paraffin embedded) material, unpublished data suggest that amplicon sizes should be reduced to 100-200bp.

EpiTYPER assay design is augmented by the use of the EpiDesigner assay design tool (<http://epidesigner.com>). This tool facilitates the process including assay selection, oligonucleotide ordering and downstream assay processing. Primers are designed to hybridize to regions devoid of CpG dinucleotides but containing cytosine residues. These parameters enhance amplification of bisulfite converted DNA strands while mitigating methylation-dependent amplification, allowing quantification of approximately 82% of CpG dinucleotides located in CpG islands using a single cleavage

reaction and yielding assay success rates of >90% [42].

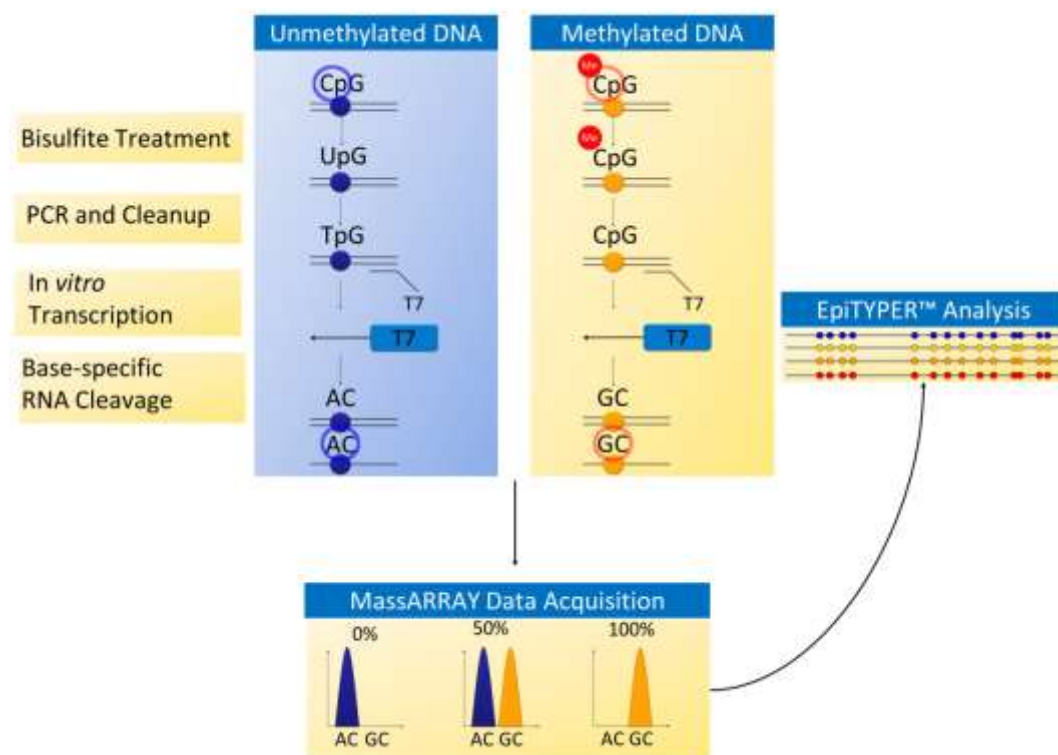


Figure 3: Schematic describing DNA methylation analysis using MassARRAY (EpiTYPER analysis). DNA is first treated with sodium bisulfite, resulting in a methylation-dependent sequence change. Target regions are subsequently amplified, products subjected to *in vitro* transcription and base-specific cleavage, using RNase A and data acquired on the MassARRAY MALDI-TOF mass spectrometer. Data are then processed and interpreted using EpiTYPER software. Colors represent DNA methylation levels (blue=unmethylated, red=methylated). Intermediate colors highlight that the DNA methylation measurement is quantitative, resulting in the ability to discern intermediate DNA methylation levels.

Subsequent to bisulfite conversion, a region-specific PCR reaction is performed to enrich for the region of interest and incorporate a polymerase recognition sequence. Due to differences in assay specific parameters, optimization of PCR conditions should be performed to ensure robust yield. The amplified product is then subjected to a simultaneous *in vitro* transcription reaction and base specific cleavage using RNase A with the products measured using MALDI-TOF mass spectrometry (Figure 3). This downstream process contributes a minority of the assay variance [48], enabling consistent downstream reaction conditions amenable to automation.

As previously described, the quality of the methylation assay is dependent upon both the assay and the input material. For assay specific controls, template consisting of *in vitro* unmethylated, methylated, and a mixture of the aforementioned sample types can be used to quantitatively assess the methylation values for the assay [42]. For sample qualification, EpiTYPER also provides a method enabling the rapid assessment of the quality of the input material [48]. Ensuring the quality of both the assay and input material enhances the likelihood of robust, reproducible assay performance.

The interpretation of the data from the MALDI-TOF mass spectrometer is enriched through the use of the EpiTYPER software. This software computes the methylation ratio for each interrogated CpG unit and allows for data export. Finally, an additional layer of data processing can be performed through the use of the MassArray bioconductor package [50] or other analytical tools where needed. Each reaction should start with 10ng or more of bisulfite converted DNA. Using this amount, it has been demonstrated that a methylation level of 5% can be distinguished from a complete lack of methylation. The standard deviation of relative DNA methylation levels between 10-90% is approximately 5%; the measurement variance increases outside of this analytical range [42]. The ability to multiplex assays within a single well is sacrificed for the quantitative determination of multiple CpG dinucleotides, a property shared by most direct quantitative methods. While direct multiplexing is limited, the parallel assessment of multiple markers from multiple samples can be rapidly assayed. Indeed, a high-throughput MALDI-TOF is capable of processing 6000 reactions per day [42]. For example, the previously described assessment of 47 targeted regions across 96 individual samples could be performed in a single day [42]. Overall, EpiTYPER analysis provides a high throughput method for quantitative DNA methylation analysis, lending itself to biomarker validation and development.

MBD-affinity capture based quantification

MBD (Methyl Binding Domain) fusion proteins bind specifically to double stranded DNA that is methylated at CpG sites on both strands. They demonstrate a bias for high CpG densities and preferentially extract methylated CpG islands, which are of particular interest for clinical assays (Figure 4) [51-54]. As DNA methylation is likely to be heterogeneous within an affected CpG island, validation of an affinity-based assay should target the CpG island region that originally was identified in biomarker discovery [55]. The specificity of an affinity assay is affected by fragmentation of the DNA prior to fractionation, which unlinks the region of interest from neighboring CpG sites whose methylation status could bias the fractionation. Isolation of intact islands can usually be achieved with restriction endonucleases that recognize sites that contain only A and T.

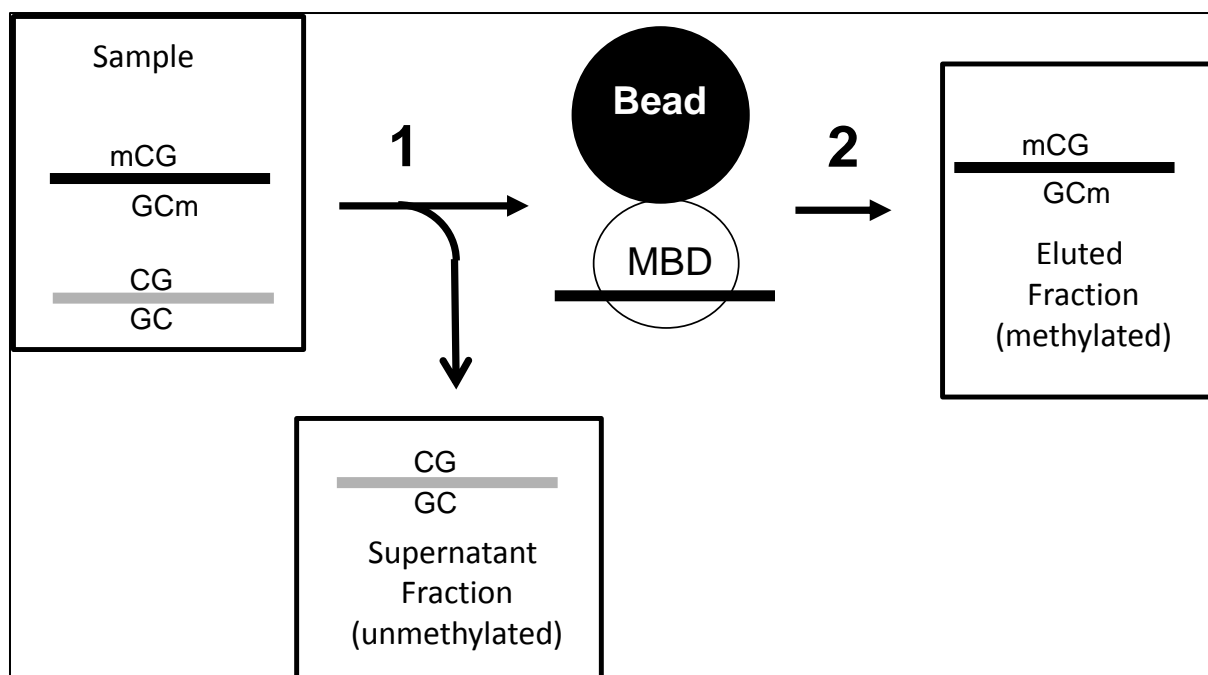


Figure 4: Methylated DNA fractionation with immobilized MBD domain proteins. A sample containing a mixture of methylated and unmethylated fragments is incubated with an MBD-domain protein that is immobilized to magnetic beads. Methylated fragments bind to the bead via the MBD protein (step 1) and are released in an eluted fraction by denaturing the protein or by increasing the NaCl concentration (step 2). The supernatant of the binding reaction contains the unmethylated fragments.

MBD beads are expected to have different binding capacities for different fragments in a genomic DNA preparation because fragments with different CpG content will compete for binding with different efficiencies, which becomes important at very high DNA inputs. A standard MBD-bead volume should be defined by titrating MBD-beads at the maximum DNA input using a sample that is methylated in the target region. The stringency of the binding reaction is an important variable, particularly if the target has few CpGs. Stringency is controlled by NaCl concentration in the binding and wash buffers.

After MBD separation of the methylated and unmethylated target fragments, the extent of methylation of that target is determined by quantifying how much target is present in the eluted vs. bound fractions. PCR bias against methylated DNA can affect PCR efficiency, especially in the early cycles of amplification [56], so T_m-lowering additives should be included at concentrations that minimize the difference in PCR efficiency between unmethylated and artificially methylated samples [6]. Optimizations of denaturation conditions (temperature and duration) can be applied to first 3 cycles, when methylated strands comprise a significant fraction of the template pool, after which cycles can be performed under conditions that have been recommended for the amplification of unmethylated G+C rich DNA [57].

Recommendations on targeting and PCR optimization apply equally to all commercially available MBD-affinity kits. Kits differ in other ways based on the design of the fusion proteins. Those based on glutathione-S-transferase (GST) are dimeric (MethylMagnet, RiboMed; MethylQuest, Millipore; MethylCap, Diagenode), while those with hexa-his affinity tags are monomeric (e.g. MethylMiner, Life Technologies; MethylCollector, Active Motif). Stringency and binding capacity optimizations for one commercial kit are not necessarily transferable to another.

Assay controls confirm the effectiveness of the binding and elution steps. A genomic DNA that is unmethylated for the target and an artificially methylated positive control can be processed in parallel with the test samples. An alternative control for the fractionation step is to assay for an imprinted gene in a normal sample, which should be 50% methylated.

MBD affinity assays have a large dynamic range. Successful fractionation is possible with less than 0.1 ng of genomic DNA, after eliminating nonspecific binding to the bead matrix [54]. Heavily damaged FFPE DNA samples with little amplifiable DNA can be successfully fractionated [58]. The upper limit is controlled by the bead input volume. The sensitivity of the assay is influenced by the downstream detection method. Quantitative PCR is satisfactory for undamaged DNAs but might require excessive numbers of cycles to analyze damaged DNAs. Amplification via CAP (Coupled Absorption-PCR) uses promoter-linked PCR primers followed by Absorption (Abortive Transcription) and is up to 1000 fold more sensitive than qPCR [58]. Adding signal amplification to target amplification makes clinically important FFPE samples more accessible.

General sample considerations

Tissue: native fresh frozen, FFPE, alternative fixatives

Proper preservation of resected tissue is crucial for a broad range of biological studies. The tissue preservation method of choice for DNA methylation analysis is to deep-freeze the tissue. However, deep-freezing requires a lot of laborious equipment, particularly -80°C refrigerators with secured electrical power supply or liquid nitrogen storage. Consequently, the procedure of formalin fixation and paraffin embedding (FFPE), first described in 1893, evolved into the standard method for tissue fixation [59]. Formalin allows morphological preservation of tissue. However, formalin degrades DNA by fragmentation and causes DNA-protein cross links at the expense of molecule integrity [60]. Optimized fixation protocols can help to overcome these problems, e.g. cold fixation at 4°C reduces fragmentation of the DNA [59,61], revealing the need for standardized protocols to diminish molecule degradation.

Many methods have been developed to analyze DNA methylation applicable to FFPE samples, including protocols for pyrosequencing [62], microarray based genomewide methylation studies (http://www.illumina.com/products/infinium_ffpe_dna_restoration_solution.ilmn) and qPCR based methods [3]. Nevertheless, FFPE tissue remains a tricky source for isolation of nucleic acids for proper quantitative methylation analysis.

Another negative aspect has been addressed by Hamilton et al., who found altered methylation patterns within the MGMT gene caused by the formalin fixation compared to fresh frozen tissue [63]. Therefore, there is an increasing demand for alternative fixation approaches. RCL2, an alcohol based fixation reagent, for example does not degrade the DNA and widely preserves tissue morphology [27,64]. Other commercial available fixatives are PAXgene, Allprotect (both from Qiagen, Germany) and RNAlater (Invitrogen, UK). All of those fixatives allow the isolation of high quality DNA comparable to DNA derived from fresh frozen samples. However there is evidence that Allprotect and RNAlater impairs immunohistochemical analyses [65].

Bodyfluids

The potential of DNA methylation markers for clinical diagnostics in body fluids has been shown already in many studies associated with various neoplastic diseases [66,67]. Body fluids can usually be obtained using non-invasive (e.g. saliva, sputum, urine) or minimal invasive (e.g. serum, plasma) methods and their extraction can also be implemented in routine diagnostic tests, representing significant benefits for patients as well as clinicians. On the other hand analyzing DNA methylation markers in body fluids is technically challenging due to several factors. In many body fluids like plasma, serum, cell-free saliva or urine (i) the total amount of DNA is relatively low, (ii) the amount of cell-free DNA (cf DNA) derived from affected cells is minimal in comparison to “normal” DNA present in the fluid and (iii) cf DNA is usually highly fragmented [68-72].

DNA isolation protocols for body fluids must therefore combine processing of high sample volumes and/or highly sufficient extraction of small-scaled and short-sized DNA fragments, making the choice of the right isolation method a crucial step for successful biomarker detection. A broad range of different kits for nucleic acid isolation from various body fluids is commercially available as well as many different published methods trying to fulfill these demands. **Table 2** summarizes exemplarily DNA isolation results from serum and/or plasma of healthy individuals in several independent studies, revealing large variations in DNA output, based not only on usage of different isolation approaches but also distinct DNA quantification methods, which makes it difficult to define a “gold standard method” for DNA isolation of cf DNA out of body fluids. Furthermore also variables like sampling procedures and sample storage conditions can strongly affect amounts of cf DNA and DNA fragmentation levels, underlining again the high demand of standard operating procedures especially for cf DNA analysis in clinical applications. Messaoudi et al. recently defined various parameters for optimal pre-analytical blood sample handling before cf DNA isolation which could be used as a first step in this direction (Table 4) [73].

Nevertheless highly sensitive DNA methylation detection methods able to identify low copy numbers of methylated DNA and working with limited amounts of available total DNA are essential to enable reliable detection of aberrantly methylated markers [74].

Apart from technical challenges another important factor has to be considered: While aberrant DNA methylation markers detected in urine or sputum are site directed, markers in serum, plasma or saliva can originate from anywhere in the body. Therefore methylation markers identified in these substrates must be specific for a single or only a small group of diseases to enable the clinician to identify the site of malignancy, especially regarding diagnostic approaches.

Table 2: Comparison of DNA isolation results in various studies out of serum and/or plasma derived from healthy individuals

Reference	Serum cfDNA concentration (ng/ml Serum)	Plasma cfDNA concentration (ng/ml Serum)	Isolation method	Quantification method
Xue et al. 2009 [75]	1.67	-	QIAamp DNA Blood Midi Kit (Qiagen)	Real-time PCR (GAPDH gene)
	4.37 (1.49-10.25)	-	THP (Triton/Heat/Phenol protocol)	Real-time PCR (GAPDH gene)
Mori et al. 2005 [76]	263 (+/-51)	-	QIAamp DNA Blood Mini Kit (Qiagen)	PicoGreen quantification assay
Gal et al. 2004 [77]	63 (5-456)	-	QIAamp DNA Blood Mini Kit (Qiagen)	Real-time PCR (β -globin gene)
Wu et al. 2002 [78]	57.1 (+/-30.6)	-	QIAamp 96 DNA Blood Kit (Qiagen)	PicoGreen DNA detection kit
Wielscher et al. 2011 (serum DNA yield varied with sample cohort) [20]	(1) 11.9 (+/-10.9)* (2) 39.7 (+/-32.8)* (3) 12.2 (+/-9.7)* * DNA yields varied depending on at which institution samples were taken	5.8 (+/-5.1)	High pure template preparation kit (Roche)	PicoGreen quantification assay
Board et al. 2008 [79]	24.65 (5-64)	5.07 (2.5-7.5)	QIAamp Viral Spin Kit (Qiagen)	Real-time PCR (AAT gene)
Gautschi et al. 2004 [80]	12.6	1.8	QIAamp DNA Blood Mini Kit (Qiagen)	Real-time PCR (GAPDH gene)
Herrera et al. 2005 [81]	-	10.6 (7.0-14.0)	QIAamp DNA Blood Mini Kit (Qiagen)	Real-time PCR, (β -actin gene)
Jung et al 2004 [82]	-	20	NucleoSpin Blood DNA Purification Kit (MN)	PicoGreen quantification assay
Deligezer al. 2003 [83]	-	44	NucleoSpin Blood DNA Purification Kit (MN)	Spectrophotometry
Stemmer et al 2003 [84]	-	3-22	KingFisher silicate magnetic beads	Picogreen reagent
Chang et al. 2002 [85]	-	7	QIAamp DNA Blood Mini Kit (Qiagen)	PicoGreen™ DNA quantitation kit

Table 3: Guidelines for optimal pre-analytical blood sample handling before cf-DNA isolation

-
- + Use Plasma samples instead of serum to avoid contamination of blood-cell genomic DNA
 - + Use EDTA or cell-free DNA collection tubes for blood sampling to prevent blood-cell lysis
 - + Process blood samples within 4 hours after blood drawing to retain initial DNA concentration and integrity
 - + Include a second high speed centrifugation step after first blood sample centrifugation to remove any remaining cells
 - + Aliquot plasma/serum samples to avoid freeze-thaw cycles, do not freeze/thaw samples more than 2 times to preserve initial DNA integrity
 - + Store plasma/serum samples at -80°C and perform DNA isolation within 9 months after sampling to preserve initial DNA integrity
-

Selecting features for validation studies

For validation studies, it is of paramount importance to choose the markers wisely. This is particularly true, if they are selected from highly multivariate analysis techniques usually employed in DNA methylation screening experiments. Popular choices for genome-wide discovery are WGBS (whole genome bisulfite sequencing) [86] RRBS (reduced representation bisulfite sequencing) [87] and the Infinium BeadChip methylation platform marketed by Illumina (either 27k or 450k). These techniques allow simultaneous measurement of tens of thousands to a few million parameters. A typical screening experiment measures those parameters on 10-50 samples per group. Therefore, a proportion of features will appear differentially methylated purely by chance, leading to a large number of false positive discoveries. For demonstration purposes, we used the method by Pawitan et.al. [88] to estimate the false discovery rate (FDR) and the number of true positives expected to be put forward into validation depending on the size of the discovery set (Assumptions: 10.000 of the 30Mio CpG sites are differentially methylated, the techniques mentioned above measure 30Mio (WGBS), 10Mio (RRBS), 450.000 (450k) and 27.000 (27k) CpG sites). In this setting, if the discovery set consists of only 10 samples per group, more than 80% of the selected features will be false positives (Figure 5). Depending on the discovery technology and the feature numbers it measures, this rate drops rapidly, as the sample size increases, but remains as high as 20% for 30 samples per group in WGBS analysis. There are several strategies to overcome this issue. The obvious one is to use an appropriately sized sample set for discovery. However, there are cases where this is not possible, either due to limited availability of samples or funding to perform these extremely costly experiments. Another strategy would be to reduce the number of analysed features, for example by focusing on genomic regions which are known to play a particularly strong role for the genotype of interest (e.g. CpG island methylation in cancer). The downside of this approach is that it limits the possibility for new discoveries. Alternatively, one could include the genomic location into feature selection: As the influence of DNA methylation is mediated through chromatin structure, it is more likely, that stretches of adjacent differential methylation have a higher influence on gene expression and phenotype than single CpG methylation events. While all of these considerations seem valid, a more simple approach may be employed: modern qPCR technologies also allow to increase the number of features to be selected for validation. Using microfluidic qPCR systems (Fluidigm Biomark, ABI OpenArray, Wafergen) one can validate several hundreds of target features instead of tens. The chance of picking the true positive features, which allow discrimination between clinical entities (healthy versus diseased), increases accordingly. Minimal amounts of DNA are sufficient to validate those hundreds of candidate

markers on hundreds of samples. This is an essential strategy to use limited resources (usually sample numbers and amounts as well as funds) in a very effective and efficient way.

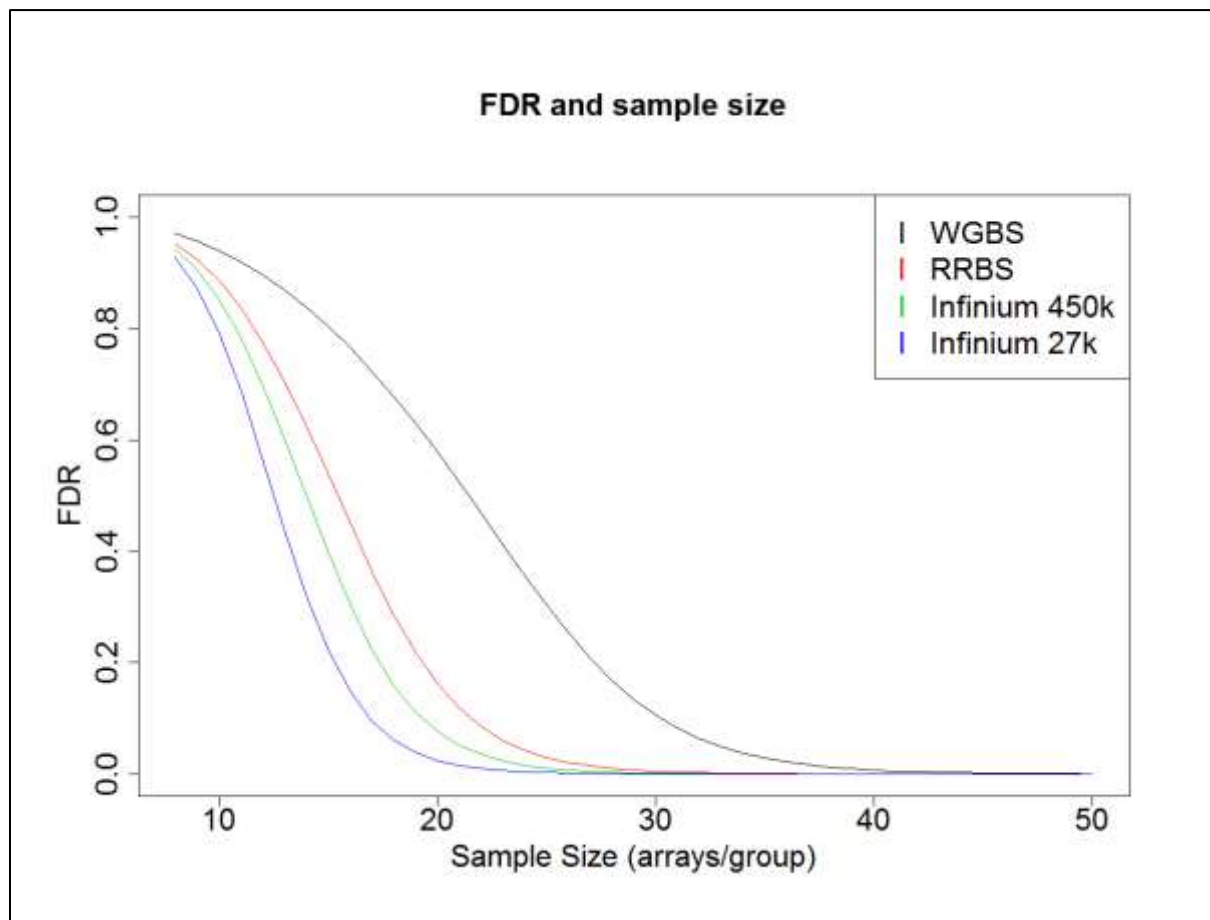


Figure 5: Expected false discovery rates (FDR) of popular genome-wide DNA methylation screening technologies depending on sample sizes.

The first step in designing a (validation) study of association between a biomarker and patient's outcome is to conduct a pilot study in which several measurements per patient are taken. Based on these data, the heterogeneity of the marker within and between individuals can be estimated and used in the process of designing an appropriate study.

Concerning the statistical approach for sample size and power calculations we refer to the specific literature (e.g. [89]). Although no general numbers on sample size can be given due to heterogeneity of markers and samples we however want to give some examples from our own studies such as e.g. lung cancer biomarkers deduced from native tissue (patent number: WO2010086388) obtaining almost perfect classification using 8 markers and a MSREqPCR analysis (studying $n=96$ samples; balanced case-control design). Candidate markers were then reconfirmed using pyrosequencing ($n=174$, 8 markers). To obtain a power of 0.9, sample size calculations using the methylation percentage-values from the markers and 174 samples (comprising DNA samples from tissue of 3 groups of 24 normal lung, 60 adenocarcinomas, and 90 squamous-cell carcinomas) revealed a sample size of 18 biologically distinct samples per group (assuming a Type 1 error of 0.05 and a hypothesized mean 2-fold difference between classes).

We then setup a 37-plex MSREqPCR candidate assay and analysed cfDNA from 194 plasma samples (unpublished data, not shown). These data were also used for sample size calculations and a sample size of 41 biological samples per group was estimated to obtain a power of 0.9 (when a Type 1 error of 0.05 and a hypothesized mean 4 fold difference between classes was assumed; sample size doubled to 82 different biological samples per group if a type 1 error of 0,001 was assumed). We conducted test-wise the same 37-plex MSREqPCR assay on a different sample-set of 48 serum samples and calculated a sample size of 64 different biological samples per group from that data for obtaining a power of 0.9 if a type 1 error of 0,05 (0,001) and a 4 fold difference was assumed. Though we used the same methods and protocols for cfDNA isolation and methylation analyses, the sample source effected assay performance and calculated sample size (more than 50% increase in numbers when comparing the samples per group for the “plasma-cohort” (n=41) and “serum-cohort” (n=64) from our example).

Using another data set to confirm methylation markers on DNA from native glioblastoma- and normal brain- tissue derived DNA [27], testing of 8 candidate gene regions by Sequenom’s EpiTyper assays made a sample size of 18 biologically distinct samples per group necessary to obtain a power of 0.9 (assuming a Type 1 error of 0.05 and a hypothesized mean 2-fold difference between classes). This is very similar to the number found in comparison to the pyrosequencing based lung cancer study described above. Although sample-size is very dependent on assay’s performance and standardization, we still believe that the numbers from our examples should be useful for conception of pilot experiments. As can be assumed from comparing numbers from the examples of cfDNA methylation analyses from plasma and serum from different retrospective cohorts and study sites, pilot studies should be performed for distinct sample sources. Especially for cfDNA testing samples from different sources might not be useful for direct comparisons.

Overview of current methylation marker validation studies and established markers

Regarding the clinical implementation of DNA methylation biomarkers, several studies are currently testing the use of early detection biomarkers as well as of prognostic and predictive biomarkers in malignant diseases (for a more comprehensive review on DNA methylation based biomarkers and clinical implementation, see [55,90]. Noteworthy, kits using established markers for early detection of colon, lung and prostate cancer or kits containing DNA methylation markers to predict the recurrence of bladder cancer already exist on the market (Table 4). In colon cancer, promoter methylation of septin 9 (SEPT9) and vimentin (VIM) can be used for early detection of malignant tissue by analysing blood (SEPT9) or stool (VIM) samples of patients [91-94]. Both markers show improved sensitivity and specificity when compared to the fecal occult blood test, which is normally used as a standard non-invasive screening test for colorectal cancer. In lung cancer, it has been shown that promoter methylation of CDKN2A is an early event in the progression of all histological subtypes and can be detected in body fluids from smoking as well as non-smoking lung cancer patients, but no diagnostic kit testing CDKN2A exists up to date [95,96]. Another biomarker already used in clinical settings for distinguishing malignant and benign lung diseases is methylation of SHOX2 [15,97]. SHOX2 methylation is measured in bronchial aspirates, and a sensitivity of 78% and a specificity of 96% have been reported [97]. In prostate cancer research, the latest and probably best-studied methylation marker recently implemented in the clinics is methylation of GSTP1, as it is seen in over 90% of prostate cancer patients, but not in normal prostate or benign prostatic hyperplasia [98]. The promoter methylation status of GSTP1 in urine or plasma can be evaluated as a follow-up in individuals at risk, for ex-

ample after a positive PSA (prostate-specific antigen) test, which would significantly improve specificity and reduce false positive results after PSA screening. In people with hematuria, the DNA methylation status of TWIST-1 and NID-2 is used together with other biomarkers to rule out bladder cancer, and Vimentin and NID-2 methylation form part of a panel of markers assessing the recurrence of bladder cancer [99-101]. DNA methylation markers have also been described as alternative or additional diagnostic tools in the analysis of Papanicolaou (PAP) smears for early cervical cancer detection. Here, the DNA methylation status of four genes (JAM3, EPB41L3, TERT and C13ORF18) has been associated with cervical cancer in high-risk human papillomavirus patients [102].

Regarding prognostic biomarkers, PITX2 methylation in breast and prostate cancer has so far shown promising results as an independent biomarker for evaluating the therapy efficacy or the risk of relapse of affected patients. For example, PITX2 methylation was associated with a poorer outcome of node-negative, estrogen-receptor positive breast cancer patients who did not receive adjuvant therapy, and also with a higher risk of relapse of prostate cancer patients after prostatectomy [103-105]. Other potential prognostic DNA methylation markers such as CDKN2A, CDH13, RASSF1A and APC have been tested in non-small-cell lung carcinoma (NSCLC) and showed association with early recurrence in stage I NSCLC [106]. Additionally, the methylation signature of these genes was predictive for responses to epigenetic therapy, as patients with two or more methylated markers presented with stable disease or objective responses to a demethylating agent and a histone deacetylase inhibitor, whereas patients without methylation in these genes showed no objective responses [107].

Among the group of predictive DNA methylation biomarkers used to predict the response to chemotherapeutic drugs, many belong to the group of DNA repair genes. In line with the finding that absence of repair genes renders tumours more susceptible to alkylating agents, methylation of the MGMT gene, which encodes the DNA repair protein O6-methylguanine DNA methyltransferase, has been shown to be associated with a survival benefit of glioblastoma patients after treatment with the alkylating drug temozolomide [108,109]. Other potential predictive biomarkers are methylation of ESR1 and ARH1 in breast cancer patients, which can predict survival in tamoxifen-treated and non-tamoxifen treated patients, the WRN and UGT1A1 genes in colorectal tumors, which were linked to increased sensitivity to the topoisomerase-1 inhibitor irinotecan, methylation of BRCA1, which can sensitise breast cancer cell lines to treatment with PARP inhibitors, and methylation of the mismatch repair gene MLH1, which has been reported to be associated with cisplatin resistance in ovarian cancer and 5-fluorouracil resistance in colorectal cancer cell lines [110-115].

Biomarker	Applica- tion	Disease	Mate- rial	Sensitivi- ty/Specificity*	Commercial test
SEPT9 [91]	Early de- tection	Colorectal cancer	Blood	70-80%/89-99%	Epi proColonR 2.0 (Epigenomics), ColoVantage™ (Quest Diagnostics), Real Time mS9 (Abbott)
VIM [93]	Early de- tection	Colorectal cancer	Stool	92%/87%	Cologuard™ (Exact Sciences)
SHOX2 [97]	Early de- tection	Lung cancer	Spu- tum	81%/95%	Epi proLungR BL 1.0 (Epigenomics)
GSTP1 [98]	Early de- tection	Prostate cancer	Urine	-	Predictive Biosciences
MGMT [109]	Predictive	Brain cancer	Tumor	-	PredictMDx™ Brain Cancer (MDxHealth)
TWIST2 + NID2 [99]	Predictive	Bladder can- cer	Urine	87.9%/99.9%	CertNDx™ Bladder Cancer Assay Hematuria Assessment (Predictive Biosciences)
VIM + NID2 [101]	Recurrence	Bladder can- cer	Urine	90.5%/95.5%	CertNDx™ Bladder Cancer Assay Recurrence Monitoring (Predictive Biosciences)

Table 4: Commercially available tests based on DNA methylation biomarkers.

CONCLUSION

A variety of methods are nowadays available for efficient analyses of DNA sequences with changed methylation patterns. Usually (human/vertebrate diagnostic) methods focus on the detection of a gain in 5mC methylation, found in the CpG dinucleotide context. This makes it relatively simple to design primers and assays for bisulfite deamination based PCR amplification and DNA methylation testing. Alternatively MSRE and MBD-affinity capture based methods for selective fractionation, amplification and quantification of methylated sequence regions are in use. However, these sequences have a high GC content and thus a high melting temperature, which can cause some difficulties in assay set-up. Methods reviewed here, are all well suited for 1) either parallel analyses of multiples regions of interest, which might be of interest, when e.g. hundred or more methylation sites have to be confirmed upon genome-wide screening, and 2) for efficient high throughput analyses of many clinical samples, which is most often required for biomarker validation. Common to all methods is that they provide a quantitative readout of the methylation-value. qPCR based methods usually provide an overall methylation-measure of the amplified sequence, relative to a calibration curve. Pyrosequencing, deep sequencing and MALDI based testing provide a quantitative single 5mC-site specific readout. Depending on the methylated sequence and sample-type of interest for validation, methods have to be carefully qualified usually with *in vitro* methylated and unmethylated DNA.

An extremely critical issue in DNA methylation testing is the preprocessing of the analysed samples. Along these lines we underline in this article that freshly deep-frozen specimens, which have not undergone any fixation procedure, represent the best starting material for DNA methylation analysis. We further conclude that methylation testing of cell-free DNA in body fluids is quite challenging because of the limited amounts of cf DNA compared to cell-derived DNA and underline to preferentially use deep-frozen plasma instead of serum for methylation testing in blood samples.

We further show in the present review sample size calculations for genome-wide methylation screening studies which, independent from which platform technology is used, reveal that sample numbers of approximately 30 per group still lead to false discovery rates of 20%. Concerning sample-sizes in validation studies we state from own experiences that they are highly dependent on assay's performance and standardization and that especially for cf DNA testing different sources of samples should be avoided since they appear not to be useful for direct comparison.

Last but not least methylation marker validation studies and established diagnostic DNA methylation markers for cancer have been summarized in this review. These might be the forefront of future validation studies for various diseases based on recent methylation screening initiatives.

FUTURE PERSPECTIVE

DNA methylation based biomarker development has increased exponentially over the last decade. Especially in but not solely limited to oncology, DNA methylation based biomarker discovery-studies have been very successful and will be best suited for diagnostic, predictive and prognostic testing. Due to the stability of the DNA as well as the methylation pattern, cfDNA methylation testing as well as a couple of tissue based assays have found their way into the clinics. Although validation of biomarkers is challenging, an appropriate variety of different methods are available, which enable efficient design and qualification of methylation assays for validation studies. We estimate that qPCR based assays will remain the working horse for these studies, for the upcoming few years. Although deep amplicon bisulfite sequencing provides quantitative methylation values and a single C resolution of methylation patterns, PCR amplification is a prerequisite for deep sequencing. For most ap-

proaches methylation quantification using PCR is sufficient and more cost effective. We expect that introduction of third-generation sequencing omitting PCR amplification, and enabling a direct – bisulfite-free read out of DNA methylation (and other types of modification), will unquestionably improve the field of DNA methylation analyses. It is also evident that within the next few years digital PCR testing of methylation analyses will make its way into clinical research and diagnostics. For confirmation of multiple candidate markers using many samples, the high throughput assays will be of certain use.

Within the next 10 years, we expect that DNA methylation-based markers will be validated in suitable cohorts and make their way into clinical routine to facilitate patient screening, monitoring and stratification for therapy decision making.

EXECUTIVE SUMMARY

Methods and strategies for DNA methylation testing & validation

- There are several platforms available and in use including qPCR-, sequencing -and MALDI-TOF based methods
- All DNA methylation testing strategies include PCR amplification, the majority of methods rely on bisulfite converted DNA
- Primer design, quality/integrity and amount of input DNA as well as optimization of PCR conditions are critical and important issues in DNA methylation testing & validation

General sample considerations

- Freshly and deeply frozen tissue - and body fluid samples are best to use for DNA methylation analysis
- Cell-free DNA methylation analysis in body fluids is challenging due to the minimal amount of cf DNA compared to cell-derived DNA and the fact that cfDNA is usually highly fragmented
- General recommendations for optimal serum/plasma preprocessing: use plasma instead of serum when possible, use EDTA or cell-free collection tubes, store plasma/serum at -80°C, avoid freeze-thaw cycles

Selecting features for validation studies

- Genome-wide DNA methylation screening typically requires sample sizes of 30 samples per group to still remain with 20% false discovery rates (FDR)
- Sample size calculations for validation studies depend not only on assay performance - and standardization but are also on the source of samples indicating that different sources of sample material should be avoided for direct comparisons

Overview of current methylation marker validation studies and established markers

- Several studies are currently testing the use and clinical implementation of DNA methylation biomarkers including early diagnosis as well as disease prognosis and prediction
- DNA methylation marker kits for early detection of lung-, colon - and prostate cancer are already on the market

Future perspective

- DNA methylation biomarker discovery -and validation studies will further increase and expand to diseases other than cancer
- qPCR based assays will remain investigators' first choice for validation studies
- Third generation sequencing methods, enabling PCR- and bisulfite-free DNA methylation analysis, should revolutionize and improve the field

Acknowledgments:

Work on this review was among others supported by the framework of CTMM, the Center for Translational Molecular Medicine, project DeCoDe (grant 03O-101), the European Community FP7 program (FP7 project no. 202047 "RESOLVE", and FP7 project no. 277849 EURHEALTHAGEING) and the Austrian funding agencies OeNB Anniversary Fund and TECNET.

REFERENCES

1. Lister R, Pelizzola M, Downen RH *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315-322 (2009).
2. Frommer M, McDonald LE, Millar DS *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, 89(5), 1827-1831 (1992).
3. Eads CA, Danenberg KD, Kawakami K *et al.* MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic acids research*, 28(8), E32 (2000).
4. Campan M, Weisenberger DJ, Trinh B, Laird PW. MethyLight. *Methods Mol Biol*, 507, 325-337 (2009).
5. Schuffler P, Mikeska T, Waha A, Lengauer T, Bock C. MethMarker: user-friendly design and optimization of gene-specific DNA methylation assays. *Genome biology*, 10(10), R105 (2009).
6. He Q, Chen HY, Bai EQ *et al.* Development of a multiplex MethyLight assay for the detection of multigene methylation in human colorectal cancer. *Cancer genetics and cytogenetics*, 202(1), 1-10 (2010).
7. Weisenberger DJ, Trinh BN, Campan M *et al.* DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic acids research*, 36(14), 4689-4698 (2008).

8. Nikolaidis G, Raji OY, Markopoulou S *et al.* DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer. *Cancer research*, 72(22), 5692-5701 (2012).
9. Hibi K, Goto T, Shirahata A *et al.* Detection of TFPI2 methylation in the serum of colorectal cancer patients. *Cancer letters*, 311(1), 96-100 (2011).
10. Shames DS, Elkins K, Walter K *et al.* Loss of NAPRT1 Expression by Tumor-specific Promoter Methylation Provides a Novel Predictive Biomarker for NAMPT Inhibitors. *Clinical cancer research : an official journal of the American Association for Cancer Research*, (2013).
11. Kristensen LS, Mikeska T, Krypuy M, Dobrovic A. Sensitive Melting Analysis after Real Time-Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic acids research*, 36(7), e42 (2008).
12. Cottrell SE, Distler J, Goodman NS *et al.* A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic acids research*, 32(1), e10 (2004).
13. Wojdacz TK, Dobrovic A. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic acids research*, 35(6), e41 (2007).
14. Lofton-Day C, Model F, Devos T *et al.* DNA methylation biomarkers for blood-based colorectal cancer screening. *Clinical chemistry*, 54(2), 414-423 (2008).
15. Schmidt B, Liebenberg V, Dietrich D *et al.* SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC cancer*, 10, 600 (2010).
16. Cedar H, Solage A, Glaser G, Razin A. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic acids research*, 6(6), 2125-2132 (1979).
17. Hashimoto K, Kokubun S, Itoi E, Roach HI. Improved quantification of DNA methylation using methylation-sensitive restriction enzymes and real-time PCR. *Epigenetics : official journal of the DNA Methylation Society*, 2(2), 86-91 (2007).
18. Lopez Castel A, Nakamori M, Thornton CA, Pearson CE. Identification of restriction endonucleases sensitive to 5-cytosine methylation at non-CpG sites, including expanded (CAG)_n/(CTG)_n repeats. *Epigenetics : official journal of the DNA Methylation Society*, 6(4), 416-420 (2011).
19. Tarasova GV, Nayakshina TN, Degtyarev SK. Substrate specificity of new methyl-directed DNA endonuclease Glal. *BMC molecular biology*, 9, 7 (2008).
20. Wielscher M, Pulverer W, Peham J *et al.* Methyl-binding domain protein-based DNA isolation from human blood serum combines DNA analyses and serum-autoantibody testing. *BMC clinical pathology*, 11, 11 (2011).
21. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-386 (2000).
22. Kent WJ, Sugnet CW, Furey TS *et al.* The human genome browser at UCSC. *Genome research*, 12(6), 996-1006 (2002).
23. Bustin SA, Benes V, Garson JA *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*, 55(4), 611-622 (2009).
24. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, 29(9), e45 (2001).
25. Pulverer W, Wielscher M, Panzer-Grumayer R *et al.* The stem cell signature of CHH/CHG methylation is not present in 271 cancer associated 5'UTR gene regions. *Biochimie*, 94(11), 2345-2352 (2012).
26. Weinhaeusel A, Thiele S, Hofner M, Hiort O, Noehammer C. PCR-based analysis of differentially methylated regions of GNAS enables convenient diagnostic testing of pseudohypoparathyroidism type 1b. *Clinical chemistry*, 54(9), 1537-1545 (2008).
27. Pulverer W, Hofner M, Preusser M, Dirnberger E, Hainfellner JA, Weinhaeusel A. A simple quantitative diagnostic alternative for MGMT DNA-methylation testing on RCL2 fixed paraffin embedded tumors using restriction coupled qPCR. *Clinical neuropathology*, (2013).
28. Rinner B, Weinhaeusel A, Lohberger B *et al.* Chordoma characterization of significant changes of the DNA methylation pattern. *PLoS One*, 8(3), e56609 (2013).

29. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nature protocols*, 2(9), 2265-2275 (2007).
30. Daskalos A, Logotheti S, Markopoulou S *et al.* Global DNA hypomethylation-induced DeltaNp73 transcriptional activation in non-small cell lung cancer. *Cancer letters*, 300(1), 79-86 (2011).
31. Schache AG, Hall G, Woolgar JA *et al.* Quantitative promoter methylation differentiates carcinoma ex pleomorphic adenoma from pleomorphic salivary adenoma. *British journal of cancer*, 103(12), 1846-1851 (2010).
32. Shaw RJ, Liloglou T, Rogers SN *et al.* Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing. *British journal of cancer*, 94(4), 561-568 (2006).
33. Gries J, Schumacher D, Arand J *et al.* Bi-PROF: bisulfite profiling of target regions using 454 GS FLX Titanium technology. *Epigenetics : official journal of the DNA Methylation Society*, 8(7), 765-771 (2013).
34. Taylor KH, Kramer RS, Davis JW *et al.* Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer research*, 67(18), 8511-8518 (2007).
35. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, 18(11), 1427-1431 (2002).
36. Tusnady GE, Simon I, Varadi A, Aranyi T. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic acids research*, 33(1), e9 (2005).
37. Brandes JC, Carraway H, Herman JG. Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. *Oncogene*, 26(42), 6229-6237 (2007).
38. Miura F, Uematsu C, Sakaki Y, Ito T. A novel strategy to design highly specific PCR primers based on the stability and uniqueness of 3'-end subsequences. *Bioinformatics*, 21(24), 4363-4370 (2005).
39. van Vlodrop IJ, Niessen HE, Derks S *et al.* Analysis of promoter CpG island hypermethylation in cancer: location, location, location! *Clinical cancer research : an official journal of the American Association for Cancer Research*, 17(13), 4225-4231 (2011).
40. Hughes S, Jones JL. The use of multiple displacement amplified DNA as a control for methylation specific PCR, pyrosequencing, bisulfite sequencing and methylation-sensitive restriction enzyme PCR. *BMC molecular biology*, 8, 91 (2007).
41. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC bioinformatics*, 10, 232 (2009).
42. Ehrich M, Nelson MR, Stanssens P *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44), 15785-15790 (2005).
43. Ehrich M, Turner J, Gibbs P *et al.* Cytosine methylation profiling of cancer cell lines. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12), 4844-4849 (2008).
44. van den Boom D, Ehrich M. Mass spectrometric analysis of cytosine methylation by base-specific cleavage and primer extension methods. *Methods Mol Biol*, 507, 207-227 (2009).
45. Baer C, Claus R, Frenzel LP *et al.* Extensive promoter DNA hypermethylation and hypomethylation is associated with aberrant microRNA expression in chronic lymphocytic leukemia. *Cancer research*, 72(15), 3775-3785 (2012).
46. Bullinger L, Ehrich M, Dohner K *et al.* Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood*, 115(3), 636-642 (2010).
47. Novak P, Stampfer MR, Munoz-Rodriguez JL *et al.* Cell-type specific DNA methylation patterns define human breast cellular identity. *PLoS One*, 7(12), e52299 (2012).
48. Ehrich M, Zoll S, Sur S, van den Boom D. A new method for accurate assessment of DNA quality after bisulfite treatment. *Nucleic acids research*, 35(5), e29 (2007).

49. Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic acids research*, 29(13), E65-65 (2001).
50. Reid TFG, J. M. MassArray: Analytical Tools for MassArray Data. *R-Package* (www.bioconductor.org) (2009).
51. Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic acids research*, 31(6), 1765-1774 (2003).
52. Robinson MD, Storzaker C, Statham AL *et al.* Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. *Genome research*, 20(12), 1719-1729 (2010).
53. Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic acids research*, 38(2), 391-399 (2010).
54. Yegnasubramanian S, Lin X, Haffner MC, DeMarzo AM, Nelson WG. Combination of methylated-DNA precipitation and methylation-sensitive restriction enzymes (COMPARE-MS) for the rapid, sensitive and quantitative detection of DNA methylation. *Nucleic acids research*, 34(3), e19 (2006).
55. Mikeska T, Bock C, Do H, Dobrovic A. DNA methylation biomarkers in cancer: progress towards clinical implementation. *Expert review of molecular diagnostics*, 12(5), 473-487 (2012).
56. Kholod N, Boniver J, Delvenne P. A new dimethyl sulfoxide-based method for gene promoter methylation detection. *The Journal of molecular diagnostics : JMD*, 9(5), 574-581 (2007).
57. Aird D, Ross MG, Chen WS *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology*, 12(2), R18 (2011).
58. McCarthy DC, P; Hanna, M. MethylMeter: A Quantitative, Sensitive and Bisulfite-Free Method for Analysis of DNA Methylation. *InTech*, (DNA Methylation), 93-116 (2012).
59. Frankel A. Formalin fixation in the '-omics' era: a primer for the surgeon-scientist. *ANZ journal of surgery*, 82(6), 395-402 (2012).
60. Campos PF, Gilbert TM. DNA extraction from formalin-fixed material. *Methods Mol Biol*, 840, 81-85 (2012).
61. Bussolati G, Annaratone L, Medico E, D'Armento G, Sapino A. Formalin fixation at low temperature better preserves nucleic acid integrity. *PLoS One*, 6(6), e21043 (2011).
62. Doyle B, O'Riain C, Appleton K. Pyrosequencing of DNA extracted from formalin-fixed paraffin-embedded tissue. *Methods Mol Biol*, 724, 181-190 (2011).
63. Hamilton MG, Roldan G, Magliocco A, McIntyre JB, Parney I, Easaw JC. Determination of the methylation status of MGMT in different regions within glioblastoma multiforme. *J.Neurooncol.*, (2010).
64. Denouel A, Boissiere-Michot F, Rochaix P, Bibeau F, Boule N. An alternative fixative to formalin fixation for molecular applications: the RCL2((R))-CS100 approach. *Methods Mol Biol*, 724, 297-307 (2011).
65. Staff S, Kujala P, Karhu R *et al.* Preservation of nucleic acids and tissue morphology in paraffin-embedded clinical samples: comparison of five molecular fixatives. *Journal of clinical pathology*, 66(9), 807-810 (2013).
66. Swarup V, Rajeswari MR. Circulating (cell-free) nucleic acids--a promising, non-invasive tool for early detection of several human diseases. *FEBS letters*, 581(5), 795-799 (2007).
67. O'Driscoll L. Extracellular nucleic acids and their potential as diagnostic, prognostic and predictive biomarkers. *Anticancer research*, 27(3A), 1257-1265 (2007).
68. Jahr S, Hentze H, Englisch S *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer research*, 61(4), 1659-1665 (2001).
69. Holdenrieder S, Burges A, Reich O, Spelsberg FW, Stieber P. DNA integrity in plasma and serum of patients with malignant and benign diseases. *Annals of the New York Academy of Sciences*, 1137, 162-170 (2008).

70. Diehl F, Li M, Dressman D *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(45), 16368-16373 (2005).
71. Chan KC, Zhang J, Hui AB *et al.* Size distributions of maternal and fetal DNA in maternal plasma. *Clinical chemistry*, 50(1), 88-92 (2004).
72. Su YH, Wang M, Brenner DE *et al.* Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. *The Journal of molecular diagnostics : JMD*, 6(2), 101-107 (2004).
73. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. *Clinica chimica acta; international journal of clinical chemistry*, 424, 222-230 (2013).
74. Gormally E, Caboux E, Vineis P, Hainaut P. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. *Mutation research*, 635(2-3), 105-117 (2007).
75. Xue X, Teare MD, Holen I, Zhu YM, Woll PJ. Optimizing the yield and utility of circulating cell-free DNA from plasma and serum. *Clinica chimica acta; international journal of clinical chemistry*, 404(2), 100-104 (2009).
76. Mori T, O'Day SJ, Umetani N *et al.* Predictive utility of circulating methylated DNA in serum of melanoma patients receiving biochemotherapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 23(36), 9351-9358 (2005).
77. Gal S, Fidler C, Lo YM *et al.* Quantitation of circulating DNA in the serum of breast cancer patients by real-time PCR. *British journal of cancer*, 90(6), 1211-1215 (2004).
78. Wu TL, Zhang D, Chia JH, Tsao K, Sun CF, Wu JT. Cell-free DNA: measurement in various carcinomas and establishment of normal reference range. *Clinica chimica acta; international journal of clinical chemistry*, 321(1-2), 77-87 (2002).
79. Board RE, Williams VS, Knight L *et al.* Isolation and extraction of circulating tumor DNA from patients with small cell lung cancer. *Annals of the New York Academy of Sciences*, 1137, 98-107 (2008).
80. Gautschi O, Bigosch C, Huegli B *et al.* Circulating deoxyribonucleic Acid as prognostic marker in non-small-cell lung cancer patients undergoing chemotherapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 22(20), 4157-4164 (2004).
81. Herrera LJ, Raja S, Gooding WE *et al.* Quantitative analysis of circulating plasma DNA as a tumor marker in thoracic malignancies. *Clinical chemistry*, 51(1), 113-118 (2005).
82. Jung K, Stephan C, Lewandowski M *et al.* Increased cell-free DNA in plasma of patients with metastatic spread in prostate cancer. *Cancer letters*, 205(2), 173-180 (2004).
83. Deligezer U, Yaman F, Erten N, Dalay N. Frequent copresence of methylated DNA and fragmented nucleosomal DNA in plasma of lymphoma patients. *Clinica chimica acta; international journal of clinical chemistry*, 335(1-2), 89-94 (2003).
84. Stemmer C, Beau-Faller M, Pencreac'h E *et al.* Use of magnetic beads for plasma cell-free DNA extraction: toward automation of plasma DNA analysis for molecular diagnostics. *Clinical chemistry*, 49(11), 1953-1955 (2003).
85. Chang HW, Lee SM, Goodman SN *et al.* Assessment of plasma DNA levels, allelic imbalance, and CA 125 as diagnostic tests for cancer. *Journal of the National Cancer Institute*, 94(22), 1697-1703 (2002).
86. Habibi E, Brinkman AB, Arand J *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell stem cell*, 13(3), 360-369 (2013).
87. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research*, 33(18), 5868-5877 (2005).
88. Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics*, 21(13), 3017-3024 (2005).

89. Pintilie M, Iakovlev V, Fyles A, Hedley D, Milosevic M, Hill RP. Heterogeneity and power in clinical biomarker studies. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(9), 1517-1521 (2009).
90. How Kit A, Nielsen HM, Tost J. DNA methylation based biomarkers: practical considerations and applications. *Biochimie*, 94(11), 2314-2337 (2012).
91. deVos T, Tetzner R, Model F *et al.* Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clinical chemistry*, 55(7), 1337-1346 (2009).
92. Chen WD, Han ZJ, Skoletsky J *et al.* Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *Journal of the National Cancer Institute*, 97(15), 1124-1132 (2005).
93. Itzkowitz S, Brand R, Jandorf L *et al.* A simplified, noninvasive stool DNA test for colorectal cancer detection. *The American journal of gastroenterology*, 103(11), 2862-2870 (2008).
94. Zou H, Allawi H, Cao X *et al.* Quantification of methylated markers with a multiplex methylation-specific technology. *Clinical chemistry*, 58(2), 375-383 (2012).
95. Tessema M, Yu YY, Stidley CA *et al.* Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers. *Carcinogenesis*, 30(7), 1132-1138 (2009).
96. Palmisano WA, Divine KK, Saccomanno G *et al.* Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer research*, 60(21), 5954-5958 (2000).
97. Kneip C, Schmidt B, Seegebarth A *et al.* SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 6(10), 1632-1638 (2011).
98. Nakayama M, Bennett CJ, Hicks JL *et al.* Hypermethylation of the human glutathione S-transferase-pi gene (GSTP1) CpG island is present in a subset of proliferative inflammatory atrophy lesions but not in normal or hyperplastic epithelium of the prostate: a detailed study using laser-capture microdissection. *The American journal of pathology*, 163(3), 923-933 (2003).
99. Renard I, Joniau S, van Cleynenbreugel B *et al.* Identification and validation of the methylated TWIST1 and NID2 genes through real-time methylation-specific polymerase chain reaction assays for the noninvasive detection of primary bladder cancer in urine samples. *European urology*, 58(1), 96-104 (2010).
100. Costa VL, Henrique R, Danielsen SA *et al.* Three epigenetic biomarkers, GDF15, TMEFF2, and VIM, accurately predict bladder cancer from DNA-based analyses of urine samples. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 16(23), 5842-5851 (2010).
101. Reinert T, Borre M, Christiansen A, Hermann GG, Orntoft TF, Dyrskjot L. Diagnosis of bladder cancer recurrence based on urinary levels of EOMES, HOXA9, POU4F2, TWIST1, VIM, and ZNF154 hypermethylation. *PLoS One*, 7(10), e46297 (2012).
102. Eijsink JJ, Lendvai A, Deregowski V *et al.* A four-gene methylation marker panel as triage test in high-risk human papillomavirus positive patients. *International journal of cancer. Journal international du cancer*, 130(8), 1861-1869 (2012).
103. Nimmrich I, Sieuwerts AM, Meijer-van Gelder ME *et al.* DNA hypermethylation of PITX2 is a marker of poor prognosis in untreated lymph node-negative hormone receptor-positive breast cancer patients. *Breast cancer research and treatment*, 111(3), 429-437 (2008).
104. Weiss G, Cottrell S, Distler J *et al.* DNA methylation of the PITX2 gene promoter region is a strong independent prognostic marker of biochemical recurrence in patients with prostate cancer after radical prostatectomy. *The Journal of urology*, 181(4), 1678-1685 (2009).
105. Banez LL, Sun L, van Leenders GJ *et al.* Multicenter clinical validation of PITX2 methylation as a prostate specific antigen recurrence predictor in patients with post-radical prostatectomy prostate cancer. *The Journal of urology*, 184(1), 149-156 (2010).
106. Brock MV, Hooker CM, Ota-Machida E *et al.* DNA methylation markers and early recurrence in stage I lung cancer. *The New England journal of medicine*, 358(11), 1118-1128 (2008).

107. Juergens RA, Wrangle J, Vendetti FP *et al.* Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. *Cancer discovery*, 1(7), 598-607 (2011).
108. Esteller M, Garcia-Foncillas J, Andion E *et al.* Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *The New England journal of medicine*, 343(19), 1350-1354 (2000).
109. Hegi ME, Diserens AC, Gorlia T *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *The New England journal of medicine*, 352(10), 997-1003 (2005).
110. Widschwendter M, Siegmund KD, Muller HM *et al.* Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer research*, 64(11), 3807-3813 (2004).
111. Agrelo R, Cheng WH, Setien F *et al.* Epigenetic inactivation of the premature aging Werner syndrome gene in human cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8822-8827 (2006).
112. Gagnon JF, Bernard O, Villeneuve L, Tetu B, Guillemette C. Irinotecan inactivation is modulated by epigenetic silencing of UGT1A1 in colon cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12(6), 1850-1858 (2006).
113. Veeck J, Ropero S, Setien F *et al.* BRCA1 CpG island hypermethylation predicts sensitivity to poly(adenosine diphosphate)-ribose polymerase inhibitors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 28(29), e563-564; author reply e565-566 (2010).
114. Zeller C, Dai W, Steele NL *et al.* Candidate DNA methylation drivers of acquired cisplatin resistance in ovarian cancer identified by methylome and expression profiling. *Oncogene*, 31(42), 4567-4576 (2012).
115. Carethers JM, Chauhan DP, Fink D *et al.* Mismatch repair proficiency and in vitro response to 5-fluorouracil. *Gastroenterology*, 117(1), 123-131 (1999).