

Epigenetic Biomarkers of Breast Cancer Risk: Across the Breast Cancer Prevention Continuum

Mary Beth Terry, Jasmine A. McDonald, Hui Chen Wu, Sybil Eng and Regina M. Santella

Abstract Epigenetic biomarkers, such as DNA methylation, can increase cancer risk through altering gene expression. The Cancer Genome Atlas (TCGA) Network has demonstrated breast cancer-specific DNA methylation signatures. DNA methylation signatures measured at the time of diagnosis may prove important for treatment options and in predicting disease-free and overall survival (tertiary prevention). DNA methylation measurement in cell free DNA may also be useful in improving early detection by measuring tumor DNA released into the blood (secondary prevention). Most evidence evaluating the use of DNA methylation markers in tertiary and secondary prevention efforts for breast cancer comes from studies that are cross-sectional or retrospective with limited corresponding epidemiologic data, raising concerns about temporality. Few prospective studies exist that are large enough to address whether DNA methylation markers add to the prediction

M. B. Terry (✉) · J. A. McDonald · S. Eng
Department of Epidemiology, Mailman School of Public Health, Columbia University,
New York, NY, USA
e-mail: mt146@columbia.edu

J. A. McDonald
e-mail: jam2319@columbia.edu

S. Eng
e-mail: sme12@columbia.edu

M. B. Terry · R. M. Santella
Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY,
USA
e-mail: mt146@columbia.edu

R. M. Santella
e-mail: rps1@columbia.edu

H. C. Wu · R. M. Santella
Department of Environmental Health Sciences, Mailman School of Public Health,
Columbia University, New York, NY, USA
e-mail: hw2057@columbia.edu

of tertiary and secondary outcomes over and beyond standard clinical measures. Determining the role of epigenetic biomarkers in primary prevention can help in identifying modifiable pathways for targeting interventions and reducing disease incidence. The potential is great for DNA methylation markers to improve cancer outcomes across the prevention continuum. Large, prospective epidemiological studies will provide essential evidence of the overall utility of adding these markers to primary prevention efforts, screening, and clinical care.

Keywords Biomarker • Breast cancer • DNA methylation • Plasma • Prevention • Prognosis • Recurrence • Serum • Survival • Breast tissues

Introduction

Breast cancer mortality rates have steadily decreased since 1990; however, breast cancer remains the second leading cause of cancer deaths in women in the United States [1]. Breast cancer is the most common cancer in women in the United States, and the incidence is increasing dramatically in very young women under age 40 years [2]. Women at higher risk of breast cancer due to family history and/or specific genetic alternations have an earlier age of onset than women at average risk and screening mammography is less sensitive in younger women [3]. Early detection of breast cancer increases treatment options, including surgical resection and therapeutic interventions [4]. Thus, finding markers that can help detect cancer early, particularly in younger women, that complement and/or improve existing methods will help in reducing incidence and mortality from breast cancer.

Biomarkers can be a useful tool for monitoring disease risk and prognosis. For example, in cardiovascular disease, blood pressure and blood markers such as lipid levels are measured routinely throughout adulthood. These markers prove particularly useful when combined with other cardiovascular disease risk factors in predicting risk through models that can readily be employed in the community and clinic. Breast cancer risk assessment models provide estimates of the absolute risk of breast cancer within a fixed time horizon (e.g., 5 or 10 years) or for the remaining lifetime of a woman. For example, women with a 5-year risk of 1.67% or higher are classified as “high-risk” and are eligible for taking tamoxifen or raloxifene to reduce breast cancer risk based on the FDA guidelines. The Gail model is the most frequently used risk prediction tool in United States clinics; however, the model is not recommended for high-risk women such as those with a strong family history of breast cancer [5, 6]. Breast cancer risk assessment methods, just like cardiovascular disease models, may benefit from the addition of biomarker and intermediate marker information. However, at present, there are no existing validated plasma/serum biomarkers for breast cancer. Only a few biomarkers (such as estrogen receptor) have utility for diagnosis and prognosis (reviewed in [7]). Thus, there is a great need for sensitive biomarkers to detect early neoplastic changes and to facilitate the detection of breast cancer at an early treatable stage.

Epigenetic modifications (e.g., DNA methylation) refer to heritable and modifiable markers that regulate gene expression without changing the underlying DNA sequence. DNA methylation may play an important role in tumorigenesis by silencing tumor suppressor genes [8–12]. Emerging evidence suggests that aberrant DNA methylation can begin very early in breast tumor progression [13] and can be detected in body fluids [14]. Similarities between methylation patterns found in primary tumor specimens and in blood plasma indicate the potential utility of blood-based molecular detection of breast cancer [15–18]. Emerging evidence has shown that DNA methylation of select genes measured in plasma results in sensitivities >90% for detecting breast cancer [15, 19]. These results suggest that DNA methylation has promise for screening. As we review, however, the evidence base is far from complete with many small studies and of a cross-sectional design that limit any inferences about temporality. Where there are gaps, we suggest study designs and the types of evidence that may prove useful in addressing these gaps.

Breast cancer is a heterogeneous disease with very different therapeutic responses and outcomes. Gene expression profiles have been used for breast cancer classification and have served as prognostic and therapeutic predictors. However, there are still major challenges in accurate early prediction of breast cancer incidence, detection and prognosis. Given that DNA methylation changes are plausibly critical components of the molecular mechanisms involved in breast cancer, distinct DNA methylation profiles may help improve the accuracy of prediction of incidence, detection and prognosis. The number of genes identified as being aberrantly methylated in breast cancer is rapidly growing (reviewed in [20]). These genes encompass multiple pathways leading to malignancy, including the six alterations proposed by Hanahan and Weinberg required to transform a healthy cell into a cancer cell: unlimited replication potential, self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death, sustained angiogenesis, and tissue invasion and metastasis [21].

In this chapter, we review the methods used to assay DNA methylation in human studies and the evidence to date from clinical and epidemiological studies on DNA methylation and breast cancer. We focus our review on describing the most common measurement techniques used to ascertain DNA methylation in human studies and then evaluate the evidence base for DNA methylation to enhance tertiary prevention (reduction of morbidity after diagnosis and improving overall survival), secondary prevention through early detection of disease, and primary prevention as a risk marker to reduce overall breast cancer incidence.

DNA Methylation, Definitions and Measurement Methods

Epigenetics is defined as changes in gene expression in the absence of changes in DNA sequence. Levels of DNA methylation, histone modifications and microRNA expression are the three main epigenetic drivers of altered gene expression. As the evidence base is largest for DNA methylation biomarkers, here, we concentrate

on studies of DNA methylation, specifically 5-methylcytosine (5mC), which results from the addition of a methyl group to the 5' position of cytosine primarily in CpG sequences. DNA methylation is essential in development and cell differentiation, silencing of transposable elements, genomic imprinting and X-chromosome inactivation. In cancer, it is well established that tumors have lower levels of total 5mC than adjacent tissues (reviewed in [22]). This hypomethylation is primarily in repetitive elements which make up the majority of our DNA and leads to their re-activation, increased illegitimate recombination, and genomic instability. This loss of methylation is an early event in carcinogenesis. Gene-specific hypomethylation can also occur and results in the re-expression of affected genes. Gene-specific hypermethylation, particularly in CpG island promoters, is the more common and well-studied event and is associated with gene inactivation. Thus, we now know that inactivation of tumor suppressor genes is not only the result of mutation but also of DNA methylation. In breast cancer, as discussed below, a large number of genes have been identified as having hypermethylated CpG island promoters and include those involved in DNA repair, cell-cycle regulation, apoptosis, chromatin remodeling, cell signaling, transcription and tumor cell invasion.

In addition to 5mC, which is present at levels of about 4 % of the cytosines, 5-hydroxymethylcytosine (5hmC) is present but at much lower levels. This is the result of Tet enzyme oxidation of 5mC [23]. This family of enzymes can further oxidize 5hmC to 5-formylcytosine and 5-carboxylcytosine, both of which are substrates for thymidine–DNA glycosylase, a DNA repair enzyme. This pathway of oxidation and base removal and repair is believed to be a mechanism for removal of the methyl group from cytosine.

A large number of methods have been developed for analysis of DNA methylation including evaluation of total 5mC; levels of methylation in repetitive elements that are a large fraction of the human genome as an indirect measure of global methylation; and levels in specific genes, primarily in CpG-rich promoter regions, but also in gene bodies and regions more distant from genes. While a large number of methods have been developed for the analysis of DNA methylation (reviewed in [24–28]), a much more limited range of assays has been applied to human health studies. These methods as well as their strengths and limitations are given in Table 1. Early studies digested DNA to nucleosides and analyzed levels of 5-methyldeoxycytidine (5mdC) by high performance liquid chromatography (HPLC) or used antibodies to bind to 5mC to obtain qualitative data. More recently, liquid chromatography-mass spectrometry (LC/MS) that allows the use of an internal standard for highly accurate and sensitive quantitation has been used [29]. This has also facilitated the quantitation of 5hmdC; however, this method as well as HPLC generally requires 1 µg of DNA [30]. Another method takes advantage of the ability of the *SssI* prokaryotic methylase enzyme to indiscriminately methylate all unmethylated CpG sequences using [³H]S-adenosylmethionine as the methyl group donor [31]. Therefore, the ability of DNA to incorporate [³H] methyl groups *in vitro* is inversely related to endogenous DNA methylation. Another method that looks at general levels of DNA methylation is the luminometric methylation assay (LUMA) which specifically analyzes 5mC in C^mCGG regions. It takes advantage of a pair of

Table 1 Methods commonly used for analysis of DNA methylation in epidemiologic studies

Assay	Method	Advantages	Disadvantages
5methylC by HPLC or LC/MS	Enzymatic digestion of DNA followed by analysis	Absolute, specific quantification of total levels of 5mC; can also measure 5hmC	Expensive equipment especially for LC/MS; MS need internal standard
[³ H] Methyl acceptance assay	Enzymatic radioactive labeling of non-methylated CpG sites	Global measure of methylation	Uses radioactivity, variable with batch of enzyme and S-adenosylmethionine; requires highly accurate DNA quantitation
Luminometric methylation assay (LUMA)	Methylation sensitive restriction digestion followed by pyrosequences	Global measure of methylation at CCGG sites	Limited CpG sites evaluated; requires high quality DNA
<i>Bisulfite treatment-based assays</i>			
Combined bisulfite restriction analysis (COBRA)	Restriction digestion of PCR amplified regions	Provides semi-quantitative data at specific regions	Analysis limited to specific restriction target sites; gel analysis limits high throughput
Methylation- specific PCR (MSP)	Separate primers for methylated and non-methylated DNA followed by gel analysis	Requires limited equipment; sensitive to 0.1 % methylated alleles	Gel analysis limits high throughput; not quantitative; only one region analyzed per assay
Fluorescence-based real time methylation-specific PCR	MSP but with fluorescence detection	Suitable for high throughput; highly quantitative; sensitive to 1/10 ⁵	Requires more expensive instrumentation than PCR; only measure DNA fully methylated on sites covered by primers; only one region analyzed per assay
MethylLight	Adds Taqman probe to real time PCR	Suitable for high throughput; highly quantitative	Requires more expensive instrumentation than PCR; only measure DNA fully methylated on sites covered by primers; sites covered by probe must also be fully methylated; only one region analyzed per assay

Table 1 (continued)

Assay	Method	Advantages	Disadvantages
Pyrosequencing	Sequencing by synthesis after amplification with non-methylation-specific primers; sequencing probe also does not contain CpG sites	Relative level of methylation at each CpG analyzed; control for efficiency of bisulfite conversion	Requires expensive instrumentation; only one region analyzed per assay; sometimes impossible to design appropriate primers and probe for specific region
Illumina Infinium 27k and 450k CpG HumanMethylation BeadChip arrays	Two types of assays; type 1 uses two probes per CpG (methylated and unmethylated); type 2 uses degenerate probes and two colors	Information from across the genome; easily interpretable beta values for methylation level	Expensive instrumentation and arrays; only interrogates sites on the array; two chemistries, specific probes and SNPs require careful data analysis
Next generation sequencing	Varies by platform	High resolution analysis of each cytosine in the genome; also obtain genetic information	Expensive; large fraction of C converted to T (reduced sequence complexity) complicates sequence alignment

isoschizomer restriction enzymes that cut DNA differentially based on methylation status. Sequencing of the product allows determination of methylation but only in CCGG sequences [32].

A major advance in analysis of DNA methylation resulted from the demonstration that treatment of DNA with sodium bisulfite resulted in deamination (the removal of an amine group) of unmethylated cytosines converting them to uracil while leaving 5mC intact. Since uracil pairs with adenine, polymerase chain reaction (PCR) primers can be designed with either an A or a G opposite the position of the C in CpG sequences. Cs in non CpG sequences, since generally not methylated, will be converted to U and an A will be used in the PCR primer. Upon PCR, the U is amplified as a T. Thus, Cs in unmethylated CpG sites are converted to Ts while methylated CpG sites remain as Cs. In methylation specific PCR (MSP), two sets of primers are designed specifically for the modified DNA strand encompassing several CpG sites, one assumes a C and the other a T in Cs in CpG sites. PCR is then followed by gel analysis for qualitative determination of whether methylated and/or unmethylated DNA is present [33].

This basic bisulfite treatment methodology has also been applied to real time fluorescence PCR eliminating the need to run gels, as well as to microarray analysis, sequencing and other types of assays. There are a number of variations of the real time assays, some using a combination of methylated and unmethylated primers with cyber green for quantification of amplified DNA and others using a control gene [21–23]. A specific variation of real time PCR, the MethyLight assay uses Taq-Man probes for quantification [34]. The fluorescence-based PCR assays are much more sensitive than MSP, but also allow high throughput since they can be run on 96- or 384-well plates. All the PCR methods that use methylation specific primers/probes detect only those DNA strands that are fully methylated for the CpG sites that are interrogated by the primers or probe; they cannot discriminate between 5mdC and 5hmdC. While small quantities of DNA are required for each PCR reaction, bisulfite modification is generally carried out on a minimum of 250 ng of DNA. All bisulfite-based assays also are dependent on the complete conversion of C to T for accurate data. In addition, differential PCR efficiency with methylated and non-methylated primers can impact results.

Bisulfite sequencing has been used extensively in epidemiologic studies for analysis of methylation. For both analysis of specific genes as well as repetitive elements (e.g., LINE-1, Alu), pyrosequencing has been the method of choice due to its relatively low cost [35]. In contrast to real time PCR, primers do not contain CpGs so that both methylated and unmethylated DNA will be amplified. The sequencing probe that sits adjacent to the region of interest also does not contain CpG sites. Synthesis of the DNA strand from the 5' to 3' direction is carried out one base at a time by incubation with the appropriate triphosphate (dNTP) based on known DNA sequence. Each incorporation event is accompanied by the release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. ATP sulfurylase converts the PPi to ATP in the presence of adenosine 5'phosphosulfate and this ATP drives an enzymatic reaction that generates light. When sequencing through positions that might contain either a C or a T, both G and A dNTPs are sequen-

tially added, which allows calculation of average level of methylation of each CpG site in the region sequenced, which is generally <300 base pairs in length. Allele-specific methylation data, or methylation along a single strand of DNA, can only be obtained if PCR products are cloned prior to sequencing, but this is not feasible in epidemiologic studies. Pyrosequencing is also not accurate at very low or high levels of methylation. The sensitivity limitation for pyrosequencing is ~5%. Next generation bisulfite sequencing is the most comprehensive method of analysis as it allows determination of methylation of multiple regions at the same time or even across the genome. Different platforms utilize different technologies, but all provide large amounts of data even with relatively small amounts of DNA. However, there are cost limits in the utilization of these platforms in epidemiological studies.

Bisulfite treated DNA has also been analyzed using Illumina Infinium HumanMethylation BeadChips that evaluate methylation of >27,000 or >450,000 CpGs. Two types of chemistries are used on the 450 K chips that lead to some differences in data, but both provide beta values or percent of methylation at each site. The low cost per data point and ease of data interpretation have made these arrays commonly used in epidemiologic studies. While results are frequently referred to as genome-wide analysis data, they are limited to the specific CpG sites on the chip. However, the 450 K array covers 99% of RefSeq genes, with an average of 17 CpG sites per gene region distributed across the promoter, 5'UTR, first exon, gene body, and the 3'UTR [36]. The 450 K array covers 96% of CpG islands, with additional coverage in island shores and the regions flanking them. One challenge with methylation studies is knowing which region of the DNA to analyze [37]. Most gene-specific methylation studies evaluate promoter regions upstream and downstream of transcription start sites. While these regions are clearly important, we now know that that intra-genic CpG sites as well as CpG shores may also be important (reviewed in [38, 39]). While levels of gene expression are often of primary interest, the relation between methylation levels and gene expression is rarely evaluated.

DNA methylation is dynamically changing over the lifecourse, but most studies only have samples from one time point. Here, we describe how DNA methylation markers may be useful in improving prognosis and overall survival (tertiary prevention), early detection (secondary prevention) and primary prevention. The importance of DNA methylation markers across all stages of the prevention continuum is strengthened by the recent data from The Cancer Genome Atlas (TCGA) on DNA methylation of over 800 breast tumors using Illumina Infinium HumanMethylation BeadChips. The data have dramatically expanded the number of genes identified as aberrantly methylated in breast cancer [40]. Knowing whether these aberrantly methylated genes in the tumor tissue are influenced by modifiable factors across the lifecourse, and/or affect early detection and tumor growth, and/or response to treatment and overall survival will have major implications for primary, secondary, and tertiary prevention efforts. In TCGA, unsupervised clustering analysis of the methylation array data identified five distinct DNA tumor groups. Group 3 showed a hypermethylation phenotype, was significantly enriched in the *luminal-B* mRNA subtype, and was under-represented for *PIK3CA*, *MAP3K1* and *MAP2K4* mutations. Group 5 showed the lowest levels of DNA methylation, overlapped with the

basal-like mRNA subtype, and had a high frequency of *TP53* mutations. Other studies examining the associations between whole-genome DNA methylation and breast cancer classification found that there were distinct methylation patterns by hormone receptor status [41, 42] and by *BRCA* mutation state [43]. Methylation profiling was also shown to reflect the cell type composition of the tumor microenvironment, specifically T lymphocyte infiltration [44]. In addition, methylation patterns in selected genes were significantly associated with disease progression [41, 42] and survival [45]. Thus, DNA methylation markers by enhancing molecular characterization of breast tumors show potential utility in population health prevention and screening and clinical care. Here we review the evidence to evaluate its potential across the cancer prevention continuum starting with improving outcomes after diagnosis and ending with primary prevention.

DNA Methylation Markers and Tertiary Prevention and Role in Prognosis

Extensive data examining DNA methylation in tissue samples at the time of diagnosis exist, however, there are far fewer studies that have prospectively followed breast cancer cases to examine how DNA methylation patterns at the time of diagnosis relate to overall survival and prognosis after breast cancer diagnosis. For example, although there have been several thousand studies that report DNA methylation and breast cancer, when we used the following search strategy in MEDLINE from the earliest available publication to September 2014 (the following search terms included forms of methylation + breast cancer + prognosis or recurrence or survival + serum or plasma in varied combinations) using two separate and independent reviewers, we only found 82 studies of DNA methylation in tissue or plasma at the time of diagnosis that examine DNA methylation and prognosis. Of these 82 studies, we reviewed the subset that specifically followed up patients longitudinally to evaluate whether DNA methylation markers are related to overall prognosis and mortality and that met the following additional criteria: (1) reported on either disease-free survival (DFS) or overall survival (OS) using survival regression methods and (2) had at least 30 events of either relapse or death (Table 2). We used these criteria because we specifically wanted to focus on whether DNA methylation markers predicted DFS or OS, over and beyond the standard clinical prognostic markers. As evidenced by TCGA, many DNA methylation markers map to subtypes of tumors [40]. For clinical utility, it is necessary to know whether new markers add to the prediction of DFS and OS after considering standard clinical metrics like stage, grade, tumor size, and nodal status. To do so, multivariable regression models are needed; such models require large sample sizes to yield precise estimates. For example, in one study that we did not include in Table 2 because it did not meet the criterion for the number of events, the overall unadjusted association of methylation in the *NEUROD1* gene with relapse free survival was 0.8 (relative risk (RR)=0.8, 95% confidence interval (CI)=0.3–1.8) but the adjusted association was over six-fold (relative

Table 2 Summary of studies evaluating the tertiary prevention potential of DNA methylation markers using breast tissue or plasma

	Sample size (N)	Follow-up time	Genes	Outcomes (HR (95% CI))	Recurrence	Overall survival/mortality
<i>Tissue studies</i>						
Huang 2013 [52]	175 primary breast cancer samples	Mean 45.81 months (15–124 months)	<i>PTPRO</i>	–	–	Reference group: Unmethylated 3.27 (0.75–14.21)
Xu 2013 [56]	167 triple-negative breast cancer treated with adjuvant chemotherapy	Median 9 years (0.41–15.1 years)	<i>BRCA1</i>	Reference group unmethylated	–	Reference group unmethylated
	675 non-triple-negative breast cancer treated with adjuvant chemotherapy			0.45 (0.24–0.84) for triple-negative		0.43 (0.19–0.95) for triple-negative
				1.56 (1.16–2.12) for non-triple negative		1.53 (1.05–2.21) for non-triple negative
Hsu 2013 [57]	139 early stage breast cancer samples including 21 triple-negative	Up to 120 months ^a	<i>BRCA1</i>	Reference group unmethylated: 12.19 (2.29–64.75)	–	Reference group unmethylated: 16.38 (1.37–195.45)
Lu 2012 [53]	348 primary breast cancer samples	Median 86 months (8–108 months)	<i>HOTAIR</i>	–	Reference group low methylated: 0.95 (0.52–1.73)	Reference group low methylated: 1.15 (0.58–2.31)

Novel Biomarkers in the Continuum of Breast Cancer

Stearns, V. (Ed.)

2016, XIX, 279 p. 34 illus., 4 illus. in color., Hardcover

ISBN: 978-3-319-22908-9