Is There a Link Between Genome-Wide Hypomethylation in Blood and Cancer Risk?

Kevin Brennan and James M. Flanagan

Abstract

Cancer cells display widespread genetic and epigenetic abnormalities, but the contribution to disease risk, particularly in normal tissue before disease, is not yet established. Genome-wide hypomethylation occurs frequently in tumors and may facilitate chromosome instability, aberrant transcription and transposable elements reactivation. Several epidemiologic case–control studies have reported genomic hypomethylation in peripheral blood of cancer patients, suggesting a systemic effect of hypomethylation on disease predisposition, which may be exploited for biomarker development. However, more recent studies have failed to reproduce this. Here, we report a meta-analysis, indicating a consistent inverse association between genomic 5-methylcytosine levels and cancer risk [95% confidence interval (CI), 1.2–6.1], but no overall risk association for studies using surrogates for genomic methylation, including methylation at the LINE-1 repetitive element (95% CI, 0.8–1.7). However, studies have been highly heterogeneous in terms of experimental design, assay type, and analytical methods. We discuss the limitations of the current approaches, including the low interindividual variability of surrogate assays such as LINE1 and the importance of using prospective studies to investigate DNA methylation in disease risk. Insights into genomic location of hypomethylation, from recent whole genome, high-resolution methylome maps, will help address this interesting and clinically important question.

Introduction

Epigenetics is the mitotically heritable control of gene expression and chromatin structure through covalent modification by DNA methylation, posttranslational modifications of histone proteins, and control of gene expression by noncoding RNAs (1, 2). DNA methylation is the addition of methyl groups to cytosines within CpG dinucleotides, to form 5-methylcytosine (5meC) and is catalyzed by the DNA methyltransferases (DNMT; refs. 2, 3). DNA methylation is dependent on the 1-carbon metabolism pathway, which requires dietary supply of folate, homocysteine and choline, and other micronutrients (4, 5). The role of DNA methylation as a negative regulator of transcription initiation at CpG islands (CGI), regions of high CpG density found in approximately 60% of genes, is well established, and is required for silencing of developmental and tissue specific genes (2, 6, 7). DNA Methylation facilitates transcriptional repression independently through preventing binding of DNA polymerases and transcription factors, or through interaction with methyl binding domain proteins and other epigenetic mechanisms such as repressive histone marks (8, 9). The functional roles of DNA methylation elsewhere in the genome is less well understood; however, DNA methylation is thought to confer genomic stability and integrity, and high methylation at repetitive elements is proposed to protect against expression of transposable elements (TE) and endogenous retroviruses (10, 11). Epigenetic regulation is often studied in the context of environmental and population health, as widespread DNA methylation patterns are known to be affected by environmental, lifestyle, and demographic factors that affect complex disease risk, such as diet, carcinogen exposure, reproductive factors, and age (12, 13). Furthermore, this “plasticity” or susceptibility to chemical alteration of the epigenome provides promise for therapeutic intervention where abnormal epigenetic patterns occur in disease or in relation to disease risk (14). A current research topic is the potential of DNA methylation variability in normal tissues such as peripheral blood, for development of cancer predisposition biomarkers (15, 16). This review and meta-analysis aims to synthesise all of the current data and to answer the question of whether there is a link between genome-wide DNA hypomethylation and cancer risk.
DNA Methylation and Cancer

DNA methylation dysregulation occurs in almost all cancers, and is becoming established as a hallmark of cancer (1, 11, 17). Most research has been concerned with local hypermethylation of promoter CpG islands, as transcriptional repression because of promoter hypermethylation of tumor suppressor genes, such as MLH1 and BRCA1 is implicated in driving several cancers (7, 8). However, the first described cancer-associated epigenetic phenomenon was the lower net percentage of 5meC in tumor tissue compared with equivalent normal tissue, known as “genome-wide” or “global” hypomethylation, that occurs frequently in all cancer types (18, 19). Despite considerable research, it remains unclear whether promoter hypermethylation and genome-wide hypomethylation are mechanistically linked, or are independent processes. (20, 21). Whole-genome DNA methylation (methylome) analysis has so far revealed than an overall increase in methylation variability (hypervariability) may be more prevalent than discrete changes in DNA methylation levels in cancer, suggesting a widespread loss of epigenetic control (9, 17, 22). Recent research has shown that most cancer-associated DNA hypermethylation occurs at “CpG shores,” regions flanking CpG islands, rather than within CpG islands (17, 23). Furthermore, genome-wide hypomethylation is largely confined to large genomic “hypomethylation blocks” (9, 17), that tend to occur in regions that display intermediate methylation levels in normal tissue, termed partially methylated domains (PMD; refs. 17, 22), and regions of low CpG density (23). That hypomethylation is specifically associated with repetitive elements has been called into question, with only a modest enrichment of repeats in hypomethylation blocks (17, 24). Such apparent compartmentalization of hypomethylation may reflect higher order changes in chromatin structure (9). Whereas genome-wide hypomethylation has long been implicated in loss of transcriptional repression, recent evidence suggests that hypomethylation at PMDs or gene bodies may be often associated with gene repression through formation of repressive chromatin (24). Therefore, understanding of cancer epigenetics is rapidly advancing with the advent of whole-methylome sequencing, revealing a greater complexity and organization to the epigenome than was previously appreciated.

Causes and Consequences of Genome-Wide Hypomethylation

The cause of genome-wide hypomethylation remains poorly understood, and several factors have been implicated (11, 21). Dietary insufficiency of folate and other 1-carbon metabolism-dependent micronutrients, or germline or somatic mutations in members of the pathway can induce hypomethylation in tumor, and normal tissues including blood (4, 14, 25, 26). Exogenous exposures, such as carcinogen exposure may influence methylation by inducing DNA damage or by affecting DNMT enzyme activity (4, 11, 27, 28). Furthermore, the efficiency of DNA methylation may be affected by age (12, 29). Finally, neoplasia may induce “field-effect” hypomethylation in cancer and surrounding histologically normal tissues, through sequestration of DNA methylation enzymes and substrates by rapidly proliferating, dividing cancer cells, resulting in failure of DNA methylation maintenance in surrounding tissue (11, 30–35).

Several tumorigenic events may result from genome-wide hypomethylation. These include chromosomal instability, which may facilitate gene-dosage alterations, mutations, genetic recombination, large deletions, or translocations (2, 11, 36). Hypomethylation may facilitate altered expression of oncogenes and/or tumor suppressors, as well as aberrant transcription of noncoding RNAs via transcriptional read-through subsequent to a loss of repression at repetitive DNA (37–39). Expression of repetitive elements is a feature of many cancers, with largely unknown consequences (37, 40). Whether genome-wide hypomethylation represents an early or causative tumorigenic event or a passive consequence of cancer remains unknown (reviewed elsewhere (21)). The detection of hypomethylation in cancer precursor lesions and normal adjacent tissue, and the induction of cancers in animal models with experimentally induced hypomethylation suggest a causative role (30, 41–43). However, the occurrence of many cancers without apparent hypomethylation, and the progression of hypomethylation with cancer stage suggest a passive role (20, 21).

Transposable Element Hypomethylation

Because of the conventional view that genome-wide hypomethylation is associated with repetitive elements, surrogate assays for genome-wide methylation have been developed specifically targeting consensus sequences within repetitive elements, including LINEs (Long Interspersed Nucleotide Element 1, LINE1), SINEs (Alu), and Satellite repeats (Sat2; refs. 44, 45). LINE1 (hereafter referred to as “L1”) is the only autonomous [capable of independent retrotransposition (transposition via an RNA intermediate)], and most highly expressed TE in the human genome, comprising approximately 17% of the human genome, with more than 500,000 copies (11, 46, 47). A full length L1 element is approximately 6 kb long with a bidirectional, noncanonical promoter, and 2 open reading frames coding for an endonuclease and retrotransposition machinery proteins (11, 46, 47). L1 transcription is largely regulated by DNA methylation of the 5’ promoter; however, most L1 elements are truncated and cannot be transcribed (37, 38, 40). Approximately 100 L1 elements are functionally capable of retrotransposition; however, only a few contribute to the vast majority of retrotransposition events (46). Alu elements, of which there are multiple families, are the most common TE in the human genome, with approximately 1.1 million copies, comprising roughly 11% of the genome (11, 47). Alu elements are nonautonomous and require the L1 transposition machinery for transposition (11). Satellite repeats (including Sat2) are short tandemly repeated noncoding DNA, frequently in centromeric and heterochromatic regions of chromosome 1 (48). Both L1
and Alu elements are heavily methylated in normal somatic tissue, however, hypomethylation of both, especially L1, is often detectable in tumors (20, 30, 49). Whereas transcription of TEs are required for their transposition, transposition-independent consequences of transcription may have functional consequences, as an estimated 7% of the human transcriptome is derived from gene ectopic expression from the L1 antisense promoter through RNA interference (39) and through driving host gene ectopic expression from the L1 antisense promoter (50).

DNA Methylation Variability and Cancer Risk

We and others have hypothesized that the epigenetic variability may contribute to risk of cancer development (14, 15, 51–53). "Epigenetic epidemiology" refers to the investigation of epigenetic patterns associated with disease risk, and such research holds great potential for cancer prevention and cancer risk-biomarker development, especially where abnormal epigenetic patterns may be detectable in easily accessible tissues such as blood, which is suitable for population screening (15, 16).

Methylation variability may contribute to, or be predictive of, risk of tumorigenesis in other tissues for several reasons. Many innate cancer risk factors, such as age, anthropometric factors, and genetic factors affect epigenetic patterns in blood (54–56). For example, rare epimutations in MLH1 and MSH2, which consist of local DNA hypermethylations events detectable in all tissues and conferring high risk of colorectal cancer, have been shown to occur because of genetic polymorphisms (57, 58). Blood DNA methylation is also affected by exposure to environmental and lifestyle cancer risk factors, such as smoking, alcohol, and other carcinogens (59, 60), therefore, blood methylation may provide useful biomarkers for acquired or environmentally induced cancer risk (12, 13). Finally, inflammation predisposes to risk of many cancers (61, 62), and affects DNA methylation (63), suggesting that blood methylation may reflect immune effects on cancer risk.

Promising for this avenue of research is the identification of a variable methylated region within the ATM gene associated with breast cancer risk (64). Furthermore, hypermethylation of BRCA1 is reportedly associated with increased prevalence of BRCA1-methylated breast tumors (65, 66), suggesting a functional link between blood detectable DNA methylation and disease histology (67–69). High-throughput discovery studies have identified cancer-associated DNA methylation signatures in blood of patients with breast (69), ovarian (68), bladder (67), and head and neck cancers (70), providing potential cancer diagnostic biomarkers.

The most widely studied putative epigenetic risk marker for cancer is genome-wide or "global" DNA methylation in blood. The studies investigating this, their findings, and the research approaches used, will be discussed further.

Current Methods for Investigation of Genome-Wide Hypomethylation

A challenge in investigating genome-wide DNA methylation in population studies is the lack of cost-effective, high-throughput assays that have both wide genome coverage and high resolution (45). Early studies used methods that measure genomic 5meC content using 5meC-specific antibodies (71), methyl-acceptance assays (72) and high-performance liquid chromatography (HPLC) (73). Such methods, however, give no information about the spatial arrangement or genomic location of DNA methylation and require large amounts of input DNA, making them unsuitable for population studies using precious patient samples. Widespread DNA methylation patterns can be measured using methylation sensitive restriction enzymes with relatively high resolution; however, methylation analysis is biased toward regions of high CpG density, as the enzyme restriction sites occur at sequences such as CCGG, and CGCG for the HpaII and HhaI restriction enzymes, respectively (45, 74). Most DNA methylation assays are based on bisulphite conversion of DNA, where incubation of DNA with sodium bisulphite causes deamination of unmethylated, but not methylated, cytosines to uracil (45). Sequencing of bisulphite converted DNA represents the "gold standard" for DNA methylation analysis (16, 75); however, real-time PCR, restriction enzyme based [combined bisulphite restriction analysis (COBRA; ref. 76)], and microarray-based methods (75) are also used. Most popular and convenient for population studies is the measurement of methylation at the repetitive elements such as L1, Alu, and Sat2, which are distributed at high frequency throughout the human genome (11, 44). These are considered "surrogate" measures of genome-wide DNA methylation as their methylation is thought to reflect genome-wide methylation levels (77). However, the efficacy of these surrogate assays for genome-wide methylation has been questioned (refs. 14, 78; discussed below), and a reinterpretation of the results reported using these assays is warranted.

Comparison of Published Reports of Genome-Wide Methylation in Blood and Cancer Risk

Twenty-three publications have reported population based cancer case–control studies investigating blood genomic DNA methylation in relation to cancer risk (Table 1). It is important to note that the comparability of these studies is limited by several differences between each, including cancer type, assay, study design (prospective/retrospective), length of time between blood-draw and diagnosis, sample size, sex (male/female/mixed), ethnicity, cancer treatment exposure, and analytical/statistical methods. Furthermore, populations at different cancer risk are included, for instance, one report included elderly men at high cancer risk (79), whereas another included Asian women at lower risk of breast cancer (80). Some reports included more than 1 "study," because of use of different assays, different patient populations, or different study designs, so altogether there were 34 individual studies of genome-wide methylation.
Table 1. Study details abstracted from all reports included in meta-analysis and all relevant studies excluded.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Author</th>
<th>Publication date</th>
<th>Journal</th>
<th>Cancer type</th>
<th>Measure-ment Assay</th>
<th>Categoricalsplit</th>
<th>Sex</th>
<th>Country/continent</th>
<th>n (case/ ctrl)</th>
<th>Prospective/retrospective</th>
<th>Newtly diagnosed</th>
<th>Pre/ postchemo-therapy</th>
<th>Cases(description)</th>
<th>Controls(description)</th>
<th>Unaffected</th>
<th>Unaffected on screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>Pufulete</td>
<td>5/2/2003</td>
<td>GASTROENTEROLOGY</td>
<td>Colorectal</td>
<td>5meC Methyl acceptance Tertiles</td>
<td>Mixed</td>
<td>Pufulete</td>
<td>No</td>
<td>UK</td>
<td>28/76</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Cancer screening</td>
<td>Unaffected</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>Pufulete</td>
<td>5/2/2003</td>
<td>GASTROENTEROLOGY</td>
<td>Colorectal adenoma</td>
<td>5meC Methyl acceptance Tertiles</td>
<td>Mixed</td>
<td>Pufulete</td>
<td>No</td>
<td>Spain</td>
<td>35/57</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Cancer screening</td>
<td>Unaffected on screening</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>Lim</td>
<td>1/1/2008</td>
<td>GASTROENTEROLOGY</td>
<td>Colorectal</td>
<td>5meC HPLC and MS Tertiles</td>
<td>Female</td>
<td>Lim</td>
<td>No</td>
<td>US</td>
<td>115/115</td>
<td>Retrospective</td>
<td>Prospective</td>
<td>Pretreatment</td>
<td>Cancer screening</td>
<td>Unaffected</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>Moore</td>
<td>4/9/2008</td>
<td>LANCET ONCOL</td>
<td>Bladder</td>
<td>5meC HPCE and hpaII Quartiles</td>
<td>Mixed</td>
<td>Moore</td>
<td>No</td>
<td>Spain</td>
<td>775/397</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Unreported</td>
<td>Hospital based</td>
<td>Hospital based</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>Choi</td>
<td>7/7/2009</td>
<td>CARCINOGENESIS</td>
<td>Breast</td>
<td>5meC Mass spectrometry Tertiles</td>
<td>Female</td>
<td>Choi</td>
<td>No</td>
<td>U.S.</td>
<td>176/173</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Hospital-based Patients (unrelated conditions)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Hou</td>
<td>10/15/2010</td>
<td>INT J CANCER</td>
<td>Gastric</td>
<td>Alu Pyrose-quencing Tertiles</td>
<td>Mixed</td>
<td>Hou</td>
<td>Caucasian Poland</td>
<td>302/419</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Mixed</td>
<td>Unaffected at follow up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>Zhu</td>
<td>3/22/2011</td>
<td>CANCER CAUSECONTROL</td>
<td>All (combined)</td>
<td>Alu Pyrose-quencing Quartiles</td>
<td>Male</td>
<td>Zhu</td>
<td>Boston</td>
<td>206/384</td>
<td>Retrospective</td>
<td>No</td>
<td>Posttreatment</td>
<td>Baseline screening of prospective population cohort</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>Rusiecki</td>
<td>4/3/2011</td>
<td>EPIGENOMICS</td>
<td>Gastric</td>
<td>Alu Pyrose-quencing Tertiles</td>
<td>Mixed</td>
<td>Rusiecki</td>
<td>Oman</td>
<td>110/104</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Mixed</td>
<td>Hospital based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>Xu</td>
<td>2/27/2012</td>
<td>FASEB J</td>
<td>Breast</td>
<td>LUMA CCGG/C3 Quintiles</td>
<td>Female</td>
<td>Xu</td>
<td>No</td>
<td>US</td>
<td>1055/1101</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Posttreatment</td>
<td>Cancer registry, first primary breast cancer Randomly selected (digit dialing)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on the following page)
<table>
<thead>
<tr>
<th>Reference Author</th>
<th>Publication date</th>
<th>Journal</th>
<th>Cancer type</th>
<th>Measurement Assay</th>
<th>Categorical split</th>
<th>First author</th>
<th>Reported restricted ethnicity</th>
<th>Country/continent</th>
<th>n (case/ctrl)</th>
<th>Prospective/retrospective</th>
<th>Newly diagnosed</th>
<th>Pro/ postchemo-therapy</th>
<th>Cases (description)</th>
<th>Controls (description)</th>
</tr>
</thead>
<tbody>
<tr>
<td>87 Di</td>
<td>10/12/2011</td>
<td>J ZHEJIANG UNIV-SC B</td>
<td>Hepatocellular</td>
<td>LINE1 Pyrosequencing</td>
<td>Quartiles Mixed</td>
<td>Di</td>
<td>No</td>
<td>China</td>
<td>315/356</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Hospital based</td>
<td>Unaffected at case selection</td>
</tr>
<tr>
<td>97 Gao</td>
<td>1/3/2012</td>
<td>BRIT J CANCER</td>
<td>Gastric</td>
<td>LINE1 Pyrosequencing</td>
<td>Quartiles Female</td>
<td>Gao</td>
<td>No</td>
<td>Shanghai</td>
<td>192/382</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Pretreatment</td>
<td>Incident cases from prospective female cohort</td>
<td></td>
</tr>
<tr>
<td>74 Xu</td>
<td>2/2/2012</td>
<td>FASEB J</td>
<td>Breast</td>
<td>LINE1 Pyrosequencing</td>
<td>Quartiles Female</td>
<td>Xu</td>
<td>No</td>
<td>US</td>
<td>106/4/1100</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Posttreatment</td>
<td>Cancer registry, first primary breast cancer</td>
<td></td>
</tr>
<tr>
<td>80 Cash</td>
<td>3/1/2012</td>
<td>INT J CANCER</td>
<td>Bladder</td>
<td>LINE1 Pyrosequencing</td>
<td>Tertiles Mixed</td>
<td>Cash</td>
<td>Han</td>
<td>China</td>
<td>510/528</td>
<td>Retrospective</td>
<td>No</td>
<td>Posttreatment</td>
<td>Cancer registry, randomly selected (digit dialing)</td>
<td></td>
</tr>
<tr>
<td>15 Brennan</td>
<td>5/1/2012</td>
<td>CANCER RES</td>
<td>Breast</td>
<td>LINE1 Pyrosequencing</td>
<td>Quartiles Female</td>
<td>Brennan</td>
<td>No</td>
<td>China</td>
<td>252/255</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Pretreatment</td>
<td>Unaffected on follow up</td>
<td></td>
</tr>
<tr>
<td>15 Brennan</td>
<td>5/1/2012</td>
<td>CANCER RES</td>
<td>Breast</td>
<td>LINE1 Pyrosequencing</td>
<td>Quartiles Female</td>
<td>Brennan</td>
<td>No</td>
<td>Europe</td>
<td>263/257</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Pretreatment</td>
<td>Incident cases from prospective female cohort</td>
<td></td>
</tr>
<tr>
<td>89 Wu</td>
<td>5/13/2012</td>
<td>CARCINOGENESIS</td>
<td>Hepatocellular</td>
<td>LINE1 Pyrosequencing</td>
<td>Quartiles Mixed</td>
<td>Wu</td>
<td>No</td>
<td>Taiwan</td>
<td>305/1254</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Pretreatment</td>
<td>Incident cases from prospective female cohort</td>
<td></td>
</tr>
<tr>
<td>47 Wu</td>
<td>6/7/2012</td>
<td>CARCINOGENESIS</td>
<td>Breast</td>
<td>LINE1 Pyrosequencing</td>
<td>Quartiles Female</td>
<td>Wu</td>
<td>No</td>
<td>US/Australia</td>
<td>279/940</td>
<td>Retrospective</td>
<td>No</td>
<td>Posttreatment</td>
<td>High-risk families</td>
<td></td>
</tr>
<tr>
<td>76 Hsiung</td>
<td>1/6/2007</td>
<td>CANCER EPIDEM BIOMAR</td>
<td>Head and neck</td>
<td>URE1 COBRA^</td>
<td>Tertiles Mixed</td>
<td>Hsiung</td>
<td>No</td>
<td>Boston</td>
<td>278/536</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Posttreatment</td>
<td>Hospital based</td>
<td></td>
</tr>
<tr>
<td>89 Wu</td>
<td>5/13/2012</td>
<td>CARCINOGENESIS</td>
<td>Bladder</td>
<td>Sat2 MethyLight</td>
<td>Quartiles Mixed</td>
<td>Wu</td>
<td>No</td>
<td>Taiwan</td>
<td>305/1254</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Pretreatment</td>
<td>Incident cases from prospective female cohort</td>
<td></td>
</tr>
<tr>
<td>47 Wu</td>
<td>6/7/2012</td>
<td>CARCINOGENESIS</td>
<td>Breast</td>
<td>Sat2 MethyLight</td>
<td>Quartiles Female</td>
<td>Wu</td>
<td>No</td>
<td>Australia</td>
<td>266/333</td>
<td>Retrospective</td>
<td>No</td>
<td>Posttreatment</td>
<td>High-risk families</td>
<td></td>
</tr>
<tr>
<td>26 Cho</td>
<td>12/30/2010</td>
<td>ANTICANCER RES</td>
<td>Breast</td>
<td>Alu MethyLight</td>
<td>NA Female</td>
<td>Cho</td>
<td>No</td>
<td>Turkey</td>
<td>4090</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Healthy volunteers</td>
<td></td>
</tr>
<tr>
<td>69 Widschwendter</td>
<td>7/16/2008</td>
<td>PLOS ONE</td>
<td>Breast</td>
<td>Alu MethyLight</td>
<td>NA Female</td>
<td>Widschwendter</td>
<td>No</td>
<td>Germany</td>
<td>189/80</td>
<td>Retrospective</td>
<td>Prospective</td>
<td>Pretreatment</td>
<td>Hospital based</td>
<td></td>
</tr>
<tr>
<td>73 Choi</td>
<td>7/7/2009</td>
<td>CARCINOGENESIS</td>
<td>Breast</td>
<td>LINE1 Pyrosequencing</td>
<td>NA Female</td>
<td>Choi</td>
<td>No</td>
<td>US</td>
<td>1918</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Hospital based</td>
<td></td>
</tr>
<tr>
<td>26 Cho</td>
<td>12/30/2010</td>
<td>ANTICANCER RES</td>
<td>Breast</td>
<td>LINE1 MethyLight</td>
<td>NA Female</td>
<td>Cho</td>
<td>No</td>
<td>Turkey</td>
<td>4090</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Healthy volunteers</td>
<td></td>
</tr>
<tr>
<td>98 Pyathilake</td>
<td>5/27/2011</td>
<td>NUTRITION</td>
<td>Bladder</td>
<td>LINE1 Pyrosequencing</td>
<td>NA Female</td>
<td>Pyathilake</td>
<td>No</td>
<td>US</td>
<td>139/273</td>
<td>Prospective</td>
<td>Yes</td>
<td>Pretreatment</td>
<td>Incident CIN cases from prospective female cohort</td>
<td></td>
</tr>
<tr>
<td>83 Patchsung</td>
<td>5/15/2012</td>
<td>PLOS ONE</td>
<td>Bladder</td>
<td>COBRA</td>
<td>NA Mixed</td>
<td>Patchsung</td>
<td>No</td>
<td>Thailand</td>
<td>6145</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Unreported</td>
<td>Healthy elderly volunteers</td>
<td></td>
</tr>
</tbody>
</table>

(Continued on the following page)
Table 1. Study details abstracted from all reports included in meta-analysis and all relevant studies excluded (Cont’d)

| Reference Author | Country/continent | Prospective/retrospective | Case–control studies (case/ctrl) | Reported restricted ethnicity | New/old therapy | L1, Alu, or C3 repetitive elements | Categorical split | Country/continent of sample collection | Cancer type | Publication date | Journal | Measure-ment Assay | Firstauthor | New/old therapy | Reported restricted ethnicity | New/old therapy | L1, Alu, or C3 repetitive elements | Categorical split | Country/continent of sample collection | Cancer type | Publication date | Journal | Measure-ment Assay |
|------------------|-------------------|--------------------------|----------------------------------|------------------------------|--------------------------|----------------|------------------------|-----------------|-----------------------------|-----------------|------------------|------------------|----------------|------------------------|-----------------|-----------------------------|-----------------|------------------|------------------|----------------|----------------|
variable, with smaller studies showing the largest ORs, it appears evident that genomic 5mC content is reduced in blood of cancer patients, suggesting potential for biomarker development. Interestingly, these studies represent the earliest published reports investigating blood genomic methylation in relation to cancer risk, and no similar study has attempted to replicate these findings since. Given that most studies using surrogate assays are null for cancer risk, it may be worthwhile returning to 5mC measures in an effort to reproduce this association. Interestingly, the only study (84) investigating widespread genomic methylation, using the restriction enzyme and bisulphite-pyrosequencing-based lumino-metric methylation assay (LUMA), identified a strong protective effect of genomic DNA hypomethylation on breast cancer risk. This inconsistent finding was likely influenced by the use of restriction enzymes, which measure methylation mainly at CGIs, which tend to be unmethylated, meaning that the only detectable change would be an increase in methylation (84).

There is no overall association between methylation at L1 and cancer risk according to this analysis. Whereas early studies suggested an association of blood L1 hypomethylation with cancer, later studies have failed to reproduce this. Furthermore, the direction of effect on cancer prevalence of L1 hypomethylation was inconsistent between significant studies (85), suggesting that this association may have occurred because of chance. One consideration is that heterogeneity between cancer types may affect the overall findings from this meta-analysis. However, in 3 breast cancer studies there is no association of L1 methylation with either cancer incidence (64) or prevalence (73, 84). Of 5 prospective studies, only 1 (79) identified an association between low L1 methylation and cancer incidence, however, small sample size, combining of multiple cancer types, and selection of a very high-risk population (elderly males) may be confounding factors for this study. Four more recent, larger studies have failed to identify any such association in gastric (90), breast (64), or hepatocellular...
Factors Affecting Epigenetic Epidemiologic Studies

Study design

Many retrospective studies investigating blood DNA methylation in cancer patients inappropriately use the term “risk” to describe associations between methylation and disease. Association with disease risk of blood DNA methylation variability can only be determined using prospective studies, with blood samples collected several years before disease development, because of the possibility of “reverse causality,” that is, the alteration of blood DNA methylation by presence of active disease (15, 68, 89). Whereas the effects of active cancer on blood DNA methylation are unknown, proliferation of lymphocytes, or depletion of required substrates or enzymes may affect blood DNA methylation (91). In addition, blood methylation may be affected by the presence of circulating tumor cells (92), though the contribution to overall blood methylation of this small fraction of tumor cells may not be detectable. An important factor for prospective studies is latency, that is, the duration between sample collection and cancer diagnosis (15), as methylation may be affected by as yet undetected cancer in samples collected shortly before diagnosis. Furthermore, the degree to which latency affects methylation-risk association may help determine whether methylation variability confers long-term or transient risk, or whether methylation is likely to be an early tumorigenic event. Retrospective studies are useful for investigating the relationship between DNA methylation variability and cancer prevalence, and may be used for development of diagnostic cancer biomarkers (16, 68). A potential confounding factor in some retrospective studies is the possible effect of cancer treatment, as several commonly used neoadjuvant chemotherapeutic drugs are known affect inhibit activity of the folate-mediated 1-carbon metabolism pathway (93, 94). Whereas this treatment-induced inhibition of DNA methylation could potentially explain the occurrence of genome-wide hypomethylation in posttreatment blood samples, 4 studies identifying significant associations between genomic hypomethylation and cancer prevalence, including 1 L1 study (87), and 3 5meC studies (31, 72, 73) used pretreatment sample only. Nonetheless, the exposure of cases, but not controls to cancer treatment may represent a confounding factor. Sample size is an important factor for case–control studies, and a confounding factor when comparing different population-based case control studies. Studies of small sample size (n < 40 cases; refs. 72, 79), may have little power to detect subtle methylation variability, perhaps leading to a bias toward nonsignificance, or to chance detection (16). Most of the studies relevant to this report were of small sample size. Conversely, larger sample sizes may lead to detection of statistically significant results with a very small overall effect size (95), which can be misleading. The largest study to date included 1,000 cases/ control pairs, at alpha = 0.05 this study would have 80% power to detect a difference of 0.12 standard deviations. For the L1 assay, with a standard deviation of 1.8% across the population, this would equal a difference in means of 0.06%, which is well below the technical variability in this assay [0.5%–5% (78, 88, 89, 96–99)] and may be biologically meaningless.

Sample selection

Various patient characteristics are potential modifiers of both cancer risk and DNA methylation. Such factors may represent intermediate factors in the link between DNA methylation and disease risk, but may also represent potential confounding factors within the study. For example, where DNA methylation is affected by smoking, an association between DNA methylation and cancer risk may be apparent because of uneven numbers of smokers in the case and control groups. To reduce the effect of such potential confounders, case–control pairs should be matched on potential confounding patient characteristics, as well as technical factors such as experimental batch (16). Furthermore, statistical adjustment for these factors should be applied during analysis. Age is a strong modifier of DNA methylation for some genes (12, 100), and is the biggest risk factor for most cancers, therefore, adjustment for age is essential (12, 100). L1 methylation is significantly lower in females than males (80, 101, 102), probably because of differences in the X and Y chromosomes (103), therefore, population studies investigating genome-wide DNA methylation must be stratified by gender (79, 80, 101, 104). Whereas gender may represent a confounding factor for some of the included studies (72, 85), statistical adjustment may have helped to reduce this bias. Both L1 methylation and cancer risk are modified by ethnicity and environmental carcinogen exposures (55, 80, 99, 105). Hospital-based studies (106) may include controls confounded by conditions unrelated to the disease under investigation. Ideally, controls samples should represent healthy individuals, matched on all potential confounding factors, both technical and biological. Case–control studies nested within prospective cohorts remain the gold standard for this type of analysis (16, 107).

Assay measurement

Repetitive element surrogate assays provide a practical and cost-effective indicator of genomic hypomethylation in tumor tissues and cell lines, where L1 methylation is highly variable (108). However, in blood, L1 methylation displays little variability (109), and it remains unclear how sensitive such assays are to detect subtle DNA methylation variability. Methylation of TEs, particularly L1, is often reported as genome-wide or “global” DNA methylation. However, L1 pyrosequencing measures methylation at only 3 to 4 CpG sites within a pool of L1 elements based on a consensus sequence (44), and is, therefore, not representative of genome-wide methylation (14). The detection of genome-wide hypomethylation by repetitive-element assays in tumor DNA is likely due the
occurrence of a proportion of repetitive elements within hypomethylated domains, rather than to a specific enrichment at repetitive elements (17, 24). Key to the utility of repetitive element methylation as a surrogate for genome-wide methylation is the reported correlation between 5meC measured by HPLC, and methylation of L1 and Alu measured by MethylLight, a methylation-sensitive PCR-based system, in a panel of cell lines and tissues. However, these correlations were moderate ($r = 0.66$ and $0.70$ for L1 and Alu, respectively), and the authors recommended using a composite measure of 2 repetitive elements, ALU-M2 and Sat2-M1 ($r = 0.85$), rather than individual elements, as a surrogate for genome-wide methylation. The only study to measure the correlation between L1 pyrosequencing and 5meC in blood, or indeed any tissue, failed to find any correlation ($r = -0.204$), though the sample size was small ($n = 27$; ref. 73). Whereas L1 methylation measured by pyrosequencing is reduced in cell lines exposed to DNA methylation inhibitors (44), and in many tumors, it is unclear whether L1 methylation reflects subtle genomic methylation variability in blood. This discrepancy may explain the stronger association with cancer risk of blood 5meC than blood L1 methylation.

Two prospective cancer risk studies detected blood hypomethylation at Alu (97) and Sat2 (89), but not at L1 in the same samples, and methylation of L1 and Alu are not correlated (99). This is inconsistent with these assays detecting "genome-wide methylation," but suggests that hypomethylation may be restricted to specific genomic sequences. Widely apparent is the greater variability of repetitive element methylation between different studies than between cases and controls within individual studies (14). L1 methylation varies with ethnicity (55, 105), likely because of L1 genetic variability (14, 90). L1 sequence heterogeneity may also cause cellular and allelic heterogeneity in L1 methylation (110), and may lead to underestimation of methylation levels (77). L1 pyrosequencing, however, displays little technical variation (108), and it will be important to determine whether measures of 5meC are as technically reproducible. L1 elements are differentially methylated at different genomic loci (50, 103, 111), and at different CpG loci within the L1 consensus sequence (98), therefore, assays preferentially amplifying different L1 elements or measuring methylation at different CpG sites may produce inconsistent results.

Finally, methylation variability at individual "functional" repetitive elements may provide greater biomarker potential than methylation of "global" repetitive elements, as hypomethylation of an aberrantly transcribed L1 within the MET oncogene, but not global L1 methylation, was detected in normal bladder in patients with bladder cancer (50).

**Statistical analysis**

Categorical analysis of DNA methylation using OR is the most frequently used measure of disease risk for case control epigenetic epidemiology studies (16). Furthermore, categorical analysis is useful for investigating potentially non-linear relationships between methylation and disease risk and retains more power to detect significant differences compared with linear regression. However, different approaches may introduce bias, for instance, a prominent inconsistency between studies included in this report is the use of different levels of methylation categories, including tertile, quartile, quintiles, and deciles, when a prespecified analysis plan and statistical power should have been described and used for the study. The narrow ranges of DNA methylation reported, especially at L1, would mean that the difference in percentage methylation between categories would be far below the detection sensitivity of the assay, for instance the technical variation for pyrosequencing, one of the most quantitative assays currently available, is around 2% and 3% (64). In our study, we observed an intraclass correlation coefficient for LINE1 in blinded duplicate samples of 0 (95% CI, 0–0.61) that suggests higher within individual variability than between individual variability. A major problem with developing small methylation differences as biomarkers is the overlap between methylation of cases and controls or lack of specificity (14).

**Publication factors**

According to the STROBE-ME (Strengthening the reporting of observational studies in epidemiology-molecular epidemiology) guidelines, all basic statistics of a biomarker measure distribution (mean, median, range, and variance) and details of all other analyses should be reported in studies investigating molecular biomarkers for disease risk (112); however, many reports do not include these, making direct comparison between studies difficult. Publication bias frequently affects meta-analyses of observational studies (113), and there is some evidence that it affects the current analysis.

All of the included reports showing significant association between L1 methylation and cancer included investigations of L1 methylation only, however all reports showing negative results for L1 also included studies showing significant associations between another methylation marker and cancer incidence/prevalence, consistent with a bias toward publication of significant associations. There appears to be a trend toward lower effect size with later publication date, as OR is significantly correlated with publication year among L1 methylation studies (Spearman $r = 0.49$, $P = 0.05$). This appears to be independent of sample size, as the correlations between OR and sample size ($r = -0.19$), and sample size and year of publication ($r = 0.2$), are not significant. A funnel plot including L1 studies only (Supplementary Fig. S1) does not appear to show asymmetry or publication bias; however, the ability of this plot to indicate publication bias may be limited by the small number of studies ($n = 14$), and variable direction of effect in studies showing significant associations. A reporting bias exists whereby categorical analysis of methylation-risk associations tends only to be reported if it reveals statistically significant
results, and if the authors deem categorical analysis appropriate. The inclusion of categorical analysis of our previously reported L1 studies represents an attempt to address this bias; however, this could only be estimated for other studies that reported no association with cancer risk (Supplementary Fig. S2). As expected, inclusion of our studies, and other recent null studies, attenuated the significant association between L1 methylation and cancer risk reported by a recent meta-analysis (86). Inclusion of all relevant reports is critical to the accuracy of meta-analyses (95), therefore, publication and reporting biases may affect our ability to identify the true effect size.

Other Considerations for Epigenetic Risk Studies

A concern for all studies investigating blood DNA methylation in relation to cancer in other tissues is the tissue specificity of DNA methylation (91, 100). Several studies have indicated that methylation of repetitive elements is tissue specific, most variable in tumor tissue, and not correlated between tumor and blood (81, 98, 114). Consistently, evidence suggests that genomic hypomethylation in tumor and normal adjacent tissue of bladder and colon cancer was not detectable in blood (30, 50), suggesting that hypomethylation is restricted to the disease-affected tissue. Furthermore, approximately 36% of L1 elements display tissue-specific expression (37). Methylation variability between blood cell types is an important consideration for retrospective studies, as differential methylation may reflect cellular proliferation. A recent microarray-based study found that approximately 39% of CpG sites analyzed were differentially methylated between leukocyte subgroups (91). However, a recent report did not find any association between L1 methylation and blood cell count (115). Another potential caveat for all cancer biomarker studies is tumor subtype heterogeneity, where the biomarker under investigation is a feature of only 1 disease subtype. For example, in tumor tissue, L1 hypomethylation frequently occurs in cancers displaying chromosome instability, but rarely in cancers with microsatellite instability (20, 36, 116). Consistently, L1 hypomethylation is detectable in some colon tumor samples, but not others (20). Therefore, if blood L1 hypomethylation is associated with colorectal cancer risk, it may predict risk of some, but not all cancers.

Conclusions and Future Studies

Epigenetic epidemiology holds great promise for identifying biomarkers of cancer risk and understanding cancer etiology. It is important to investigate methylation variability at individual’s loci in the context of widespread methylation changes associated with cancer; therefore, understanding the contribution of genome-wide DNA methylation to cancer susceptibility and development is needed. The challenges of investigating a genome-wide epigenetic phenomenon in population studies are formidable, because of the requirements of high genome coverage, highly quantitative DNA methylation measurement, and large sample size (16, 45). We conclude from our meta-analysis that genome-wide DNA methylation, as measured by the surrogate L1 methylation assay, is not associated with cancer risk, shown most appropriately by several prospective cohort studies. However, there appears evidence for an association of genomic hypomethylation, measured using more representative assays, such as HPLC or methyl-acceptance assays, with cancer prevalence, though this remains to be validated in prospective studies.

A thorough investigation of the relationship between genome-wide DNA methylation and cancer risk would require whole-genome bisulphite sequencing of large cohorts of prospectively collected blood samples, with careful sample selection, case–control matching, and deep sequencing to achieve the sensitivity required to detect subtle DNA methylation variability (16, 45). Such an approach would provide information about genome-scale methylation changes as well as local methylation variability, thus providing candidate regions for risk biomarker development. An equivalent study using retrospectively collected blood samples may provide candidate diagnostic biomarkers. Whole-genome bisulphite sequencing is extremely costly, however, technological improvement will make this feasible in the coming years (45, 75). Whereas current research using whole-genome sequencing of small numbers of samples cannot provide information about methylation variability across populations, further classification of the regions that consistently undergo methylation changes in cancer may provide candidate regions for a sequence capture approach in population studies. New primer technology will enable the investigation of epigenetic regulation at specific “functional” repetitive elements, such as those L1 elements capable of retrotransposition (47, 50), or driving ectopic expression of neighboring genes (37, 50). Furthermore, availability of the first microarrays specifically covering repetitive elements will provide high-throughput methods of investigating repetitive element methylation on a genome scale (117). The most popular tool for current investigation of DNA methylation variability in population studies is the Illumina 450K methylation beadchip, which measures methylation at individual CpG sites across the entire human genome. These arrays are highly quantitative, high-throughput, and largely unbiased in terms of genomic coverage, making them suitable for biomarker discovery and epigenome-wide association studies (16). Use of these assays will likely provide interesting evidence for the implications of DNA methylation variability in disease susceptibility and development in coming years.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: K. Brennan, J.M. Flanagan
Development of methodology: K. Brennan
Analysis and interpretation of data (e.g., statistical analysis, bioinformatics, computational analysis): K. Brennan, J.M. Flanagan
Writing, review, and/or revision of the manuscript: K. Brennan, J.M. Flanagan
Genome-Wide Hypomethylation and Cancer Risk

References


Grant Support
This work was funded by Breast Cancer Campaign fellowship to J.M. Flanagan. K. Brennan and J.M. Flanagan are funded by Breast Cancer Campaign.

Received July 19, 2012; revised October 17, 2012; accepted October 22, 2012; published OnlineFirst November 7, 2012.

Acknowledgments
The authors thank Prof. Robert Brown for his critical review of this manuscript.

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K. Brennan

Study supervision: J.M. Flanagan

www.aacrjournals.org Cancer Prev Res; 5(12) December 2012 1355
Genome-Wide Hypomethylation and Cancer Risk


Downloaded from cancerpreventionresearch.aacjrournals.org on September 2, 2021. © 2012 American Association for Cancer Research.
Is There a Link Between Genome-Wide Hypomethylation in Blood and Cancer Risk?

Kevin Brennan and James M. Flanagan


Updated version
Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-12-0316

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2012/11/07/1940-6207.CAPR-12-0316.DC1

Cited articles
This article cites 116 articles, 24 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/5/12/1345.full#ref-list-1

Citing articles
This article has been cited by 11 HighWire-hosted articles. Access the articles at:
http://cancerpreventionresearch.aacrjournals.org/content/5/12/1345.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerpreventionresearch.aacrjournals.org/content/5/12/1345.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.