Published in final edited form as:

Expert Opin Med Diagn. 2011 September 1; 5(5): 375–379. doi:10.1517/17530059.2011.590129.

The Promise and the Problems of Epigenetics Biomarkers in Cancer

Attila T. Lorincz

Wolfson Institute of Preventive Medicine, Queen Mary University of London, London UK

Abstract

Epigenetics plays an important role in tissue differentiation and phenotypic changes are associated with extensive modifications in epigenetic patterns such as DNA methylation (DNAm), histone methylation and acetylation marks, and micro RNAs. From a diagnostic perspective, DNAm is one of the more tractable epigenetic changes; differentially affecting a large number of genes. Variations can be measured accurately, on any given set of CG sites, by sequencing bisulfite converted DNA. The promise of DNAm is that biomarkers can be found in every kind of gene from rare unstable cell-cycle enzymes to highly expressed structural proteins. Almost any kind of biological specimen is amenable and the changes can be measured in tissue biopsies, scrapes, aspirates, urine, blood and other fluids. To date, hundreds of differentially methylated genes have been identified in cancer that can potentially function as biomarkers. The most common kind of change, studied, is CpG island methylation of tumor suppressor genes that modifies transcription. Despite a great many candidates, few of the DNAm markers have been adequately validated for routine clinical use. Important current limitations of epigenetic biomarkers are that assays are diverse; gene lists are large; comparative data are few, and disagreements in published papers are frequent.

Keywords

Epigenetics; Epigenomics; DNA methylation; Cancer biomarker; Diagnostics

1. Introduction

Cancer is a result of uncontrolled abnormal cell proliferation and differentiation with a number of different etiological components. Influences from the environment interact over time with intrinsic factors and inherited genotype to eventually exceed some critical error threshold. Cumulative coding changes combine with perturbations in cellular machinery to produce abnormal cascades of RNA and protein expression and eventually long-term alterations in tissue homeostasis within differentiation compartments. These changes may originate in relatively undifferentiated reservoirs such as stem cells or in more differentiated cells that de-differentiate in the process of carcinogenesis. Changes in tissue phenotype can become self-perpetuating and lead to a permissive context of proliferation, invasion and metastasis. Today it is an impossible task to elucidate all changes that may occur in a given cancer patient; however, a key role of diagnostics is to reveal those alterations that lead to meaningful improvements in diagnosis, prognosis and therapeutic risk predictions.

Correspondence to: Attila T. Lorincz. Wolfson Institute of Preventive Medicine, Queen Mary University of London, London EC1M 6BQ, UK. a.lorincz@qmul.ac.uk; Phone +44 (0)20 7882 3540 .

Declaration of interest.

A Lorincz is supported by Cancer Research UK programme grant C569/A10404

Epigenetics relates to a modifying reversible control system superimposed on the genetic code which carries profound long-term effects on allele expression [1, 2]. With respect to DNAm and transcription it is reasonable to state that genes highly methylated in the promoter region may be on but more likely off and genes with little or no methylation in promoters may be on or off depending on the positions and quantitative levels of methylation in combination with other factors. Although there are multiple aspects of epigenetics this *Editorial* is presented in the context of DNA methylation of specific gene sets. Global changes in hypomethylation and hypermethylation patterns are also a diagnostic possibility for cancer screening but these approaches are probably too crude for use in personalized medicine. The linkage of histone epigenetic marks such as the expressive H3K9Ac or the repressive H3K27me3 to specific genes requires complex chromosomal immunoprecipitation techniques [1-3] and this approach is still firmly in the basic research realm. It is quite clear that there are diverse methylation levels and locations that affect tissue-specific expression of alleles, potentially unmasking recessive mutations. Perhaps surprisingly the correlation between DNAm patterns and gene transcription as assessed in microarray studies in human tissues is low [4] indicating that these alternative approaches will identify many different biomarkers. A hazard for readers interested in practical clinical diagnostic applications is that the epigenetic literature contains many generalized simplifications and assumptions from expression rules based on cell lines and animal models that are only weakly relevant to interpreting the role of DNAm in human tissues.

2. The Promise of Epigenetic Biomarkers

Ongoing efforts to improve prognosis and prediction in solid cancers has focused on ever larger sets of markers used in combinations because individual immunochemistry or RNA expression markers have reached limits of diminishing return. Since it is a huge effort to validate each new marker and have it assessed by regulatory authorities for approval in routine practice the broad dissemination of most new biomarkers fails. It seems better to try for as much new informational content as possible in each round of validation to justify the efforts involved. Newer validations are now mostly conducted in cooperation with diagnostic and pharmaceutical companies searching for new markets. These may relate to routine diagnostics, to companion diagnostics for personalized therapies, or for guiding drugs with less than acceptable risk or efficacy profiles to individuals where the benefits outweigh the risks. Along these lines the recent work on RNA-based biomarkers has focused on large sets of transcript signatures increasingly identified by means of microarrays. These larger gene sets are typified in the breast cancer arena by complex tests such as Oncotype DX, Mammaprint, and the core intrinsic subtype sets, for example PAM-50 [5, 6]. Similar work involving the prognostic value of 31 cell cycle proliferation (CCP) transcripts in another major cancer, namely of the prostate, has recently been published [7].

DNAm are a relatively new and exciting complement to these evolving RNA expression panel approaches and increasingly DNAm biomarker sets will come from: 1) the careful validation of existing candidate DNAm genes in novel diagnostic combinations and 2) from new discoveries based on DNA methylation microarrays and other global approaches such as deep DNAm sequencing. At least 10% of expressed genes are controlled by methylation and up to 60% of genes have local CpG islands, thus it is a reasonable assumption that many and perhaps all cellular pathways have at least one gene controlled by methylation. If some of these methylated genes are regulatory then entire pathways may be turned on or off by the methylation of a few key genes. This may be a particularly efficient way of reducing overall RNA noise in the transcriptome in differentiated tissues for the longer term as compared to maintaining particular states by a multitude of unstable intermediary transcription factors and their modifiers (phosphorylation, miRNA etc). It is realistic to consider differential DNAm as affecting a set of genomic master switches

controlling many aspects of tissue proliferation and differentiation. The great majority of genes are present in the genome at comparable copy number and amenable to search and discovery as biomarkers. This is not the case for RNA and protein biomarkers because genes expressed at low levels or as unstable transcripts are difficult or not possible to detect in microarrays. Similarly protein biomarkers are usually limited in a practical diagnostic sense to assays based on antibodies or mass spectroscopy that can detect the higher level expression of the proteins.

By a conservative assumption of 25,000 genes in the human genome, with 10% methylated, there are potentially 2,500 candidate methylation biomarkers if one assumes only one biomarker function per gene. In addition methylation of non-coding regions is known to affect the cancer genome so the total number of relevant biomarkers may be much higher. To date only a small percentage (less than 10%) of these putative markers have been investigated. New advances in methylome screening technologies make possible the discovery of many more sequences of interest, such approaches have been summarized recently by Laird [8]. Once these biomarkers are in hand the challenge will be the validation of specific clinical applications, as discussed below. Table 1 shows strengths and weaknesses of the main current DNA methylation assays and Table 2 lists some of the more important diagnostic DNAm genes in prostate cancer studied by our epigenetics team. Many recent studies, including from our laboratory focusing on prostate and breast cancer, have revealed numerous interesting CpG island DNAm genes and gene combinations with significant potential in diagnostic and prognostic applications [9, 10]. In prostate cancer a large set of genes are differentially methylated to high levels in cancer as compared to non-malignant prostate. For example the quantitative methylation levels of any of *RARB*, *GSTpi*, *HIN1*, *APC*, *BCL2*, *CCND2*, *CDH13*, *EGFR5*, *NKX2-5*, and *RASSF1a* are capable of separating prostate cancer from non-cancer with a sensitivity and specificity of greater than 90%. It appears that there are many such diagnostic genes from which one could choose a small set to be validated for optimum performance. Several of these genes also have field effects and possibly could be used to indicate false-negative initial diagnoses in men requiring re-biopsy or to screen for cancer in bodily fluids. Similar situations are found in breast, colon, bladder and other cancers, thus DNAm may have a role in a wide variety of diagnostic applications. Our work on prostate and breast cancer has also revealed DNAm genes promising as independent prognostic markers for cancer recurrence [10, unpublished data]. An appealing aspect of DNAm markers as compared to RNA expression markers is that meaningful diagnostic and prognostic information may be attainable with relatively small sets of well defined DNAm genes (3 to 7 genes) as compared to RNA panels that typically have 20 to several hundred genes. For routine diagnostic applications fewer genes are preferred because these are amenable to more robust and less costly assays.

An exciting new discovery is differential DNA methylation of human papillomavirus (HPV). HPV 16 has approximately 113 scattered CG sites but no CpG island, many of the CGs are differentially methylated and increased methylation of the L1, L2 and E2 ORFs are associated with cervical carcinogenesis [11]. Other viruses such as Hepatitis B Virus, Epstein Barr Virus, and Human T Lymphotropic Virus also exhibit increased methylation of certain genomic regions in cancer. If these observations on the regulatory impact of differential methylation of scattered CG sites are generalized to other viruses and to the human genome it will be an important opportunity to discover new kinds of DNAm biomarkers for cancers and chronic diseases.

3. The Problems of Epigenetic Biomarkers

There is a great need for standardization and clinical validation of specific applications. This situation is in-part related to the use of diverse assays with various technical limitations that

may give non-comparable results in different studies (Table 1). DNAm assays are relatively complex and require great care in primer design and optimization. Rigorous quality control must be implemented to identify problems such as poor DNA quality, inhibitors, and contamination. For assays that use bisulfite the efficiency of C conversion may vary and must be carefully checked in all specimens by inclusion of appropriate controls

Many of the more popular DNAm assays such as Methylation Specific PCR (MSP) are either non-quantitative or produce semi-quantitative data with weakly defined cutoffs based on complicated ratios of the test and control genes. Such assays are fine in a research setting but are prone to reproducibility problems in routine clinical testing laboratories [1,9]. The terms hypermethylation and hypomethylation are often used in epigenetics research settings but these are unhelpful and potentially misleading for routine diagnostic applications because they are imprecise. Some DNAm assays employ restriction enzymes that either cut or do not cut at certain methylated C positions. These assays are limited to areas that contain the relevant enzyme sites, they can be affected by specimens of poor quality that contain inhibitors or degraded DNA and are not applicable to formalin fixed tissues. Assays based on methylated DNA immunoprecipitation (MeDIP) suffer from a requirement for a relatively large number of adjacent methylated sites and have little to no ability to provide fine quantitative resolution of methylation levels at specific CG sites. Similarly high-density (HD) DNAm microarrays have not been adequately validated and in particular there is a generalized lack of data on inter-laboratory reproducibility on clinical specimens, with comparative data presented usually for only one or a few genes. HD arrays are costly for routine testing and more useful for new biomarker discovery in different tissues. Increasingly researchers are adopting an absolute quantitative approach that gives accurate percentage methylated by employing PCR pyrosequencing, high throughput whole genome bisulfite sequencing, DNA melting analyses, epiTYPER, and other accurate approaches [10-12].

Adding to problems produced by diverse assays is the very large number of biomarkers to be validated and the fact that epigenomics has had less research support than other biomarker approaches. A consequence of diverse assays and diverse markers is that while there are many published epigenetics papers few compare the same markers using the same assays in similar clinical specimens. This leads to candidate markers whose real performance is unknown even though some of their details may have been published several times by research groups focusing on a particular favored small set of markers.

False discovery is a general problem with new biomarkers and these issues are also of great concern in the DNAm field; however, some false discovery within reason is preferable to a failure of true discovery and it is the job of subsequent validation studies to reveal biomarkers with real value. Our research team undertook a broad laboratory review of candidate DNAm genes in an attempt to bring a level of comparability and eventual validation to some of the diverse data in the literature. In both prostate and breast cancer specimens we studied 28 to 30 candidate genes using a highly accurate PCR-PSQ assay [10]. We found that levels of methylation were important and informative in both a diagnostic and prognostic aspect and built multi-gene risk models that can categorize patients at risk for cancer and for cancer recurrence. We mostly confirmed and expanded on existing data for a majority of the DNAm genes. There were also quite a number of candidate genes not confirmed as biomarkers by our studies, for example *CDH1*, a reported DNAm gene that has caused confusion in the literature because of widely disparate reports on both the level of methylation and the regions of the gene that were methylated [1, 10]. This situation indicates the continuing need to expand efforts to compare results on specific DNAm genes and to carefully define the gene regions studied using standardized and accurate assays.

4. Expert Opinion

Epigenetic assays are predominantly still in the research arena but some assays are increasingly used in clinical testing, mostly to check for methylation of a small set of genes, for example hypermethylated vimentin (*VIM*) (ColoSure™, LabCorp) in colorectal cancer, *GSTpi*, *APC* and others in prostate cancer. Another example is the potential use of *CDKN2A* or *CDKN2B* (p16^{INK4A}, p15^{INK4B}) methylation in guiding therapy in myelodysplastic syndrome [13]. The relative lack of clinical applications for DNAm markers is related to inadequate validation and marketing. The situation will change in the coming years as the limitations of current biomarkers drives a shift in focus to aspects of biology that can better explain the variability of individual patients, the effects of past exposures, and impact of the environment. Epigenetic biomarkers are complementary to RNA transcript levels, protein immunochemistry, and single nucleotide polymorphism-based risk assessments. A detailed understanding of epigenetic information flow in a given patient, although complicated, is necessary to fully develop the nascent field of personalized medicine. For patients with complex cancers like prostate or breast, today's nomograms and marker models do not accurately indicate risks of progression and best modes of therapy.

Molecular biomarkers continue to grow in importance with the recognition that they are a complement to pathology diagnoses and clinical information. Epigenomics in particular will produce a vast array of novel biomarkers to help refine cancer risk assessment in the next 5-10 years. The study of cell lines and animal model systems are of limited value in epigenetic diagnostics although they may have some value in assisting the discovery of new markers. Given the level of variability observed it will be necessary to check each particular gene in each kind of human tissue carefully to determine the quantitative level and configuration state of methylation and role in transcription. Biomarkers that assist pathologists in improved diagnosis or prognosis have been actively embraced. New molecular SNP tests allow prediction of drug efficacy, for example in colorectal cancer, where detection of a *KRAS* mutation is a negative predictor for anti-EGFR antibody therapies; this test along with many other examples is rapidly gaining acceptance [14]. When methylation markers are similarly validated they are expected to quickly make their way into routine use by the clinical community. In other cases the markers may have an ability to stratify but nevertheless add nothing to the expert diagnosis. Some pathologists may say that there is little value in these latter biomarkers since cancer specimens are always reviewed by a skilled pathologist. However, gene-based biomarkers that reflect an expert's diagnosis do have value in places where there is a shortage of skilled pathologists. In many parts of the developing world gene-based tests could assist in diagnostic algorithms. The assay systems and costs of such markers are important issues, complex and expensive RNA marker sets will have limited appeal leaving the door open for potentially less expensive epigenetic assays with fewer more informative genes.

While continued discovery of new biomarkers is of research value these are not useful in a clinical setting until they have been extensively validated for specific applications by years of testing in large clinical studies. In a routine diagnostic sense there are already too many DNAm candidate markers and unfortunately few have had any meaningful validation. Thus the greatest need in epigenetic diagnostics today is to formally validate existing candidate biomarkers. Their subsequent configuration into clinical assays, regulatory approvals, and marketing efforts to educate potential users are subsequent key milestones to acceptance for routine use in clinical testing labs. Companies hoping to capitalize in the coming era of personalized medicine need to work closely with academic groups who have existing archived clinical specimen sets that can be used to validate the new epigenetic biomarkers in real world applications.

Bibliography

Papers of special note have been highlighted as either of interest (*) or of considerable interest (**) to readers.

1. Li LC, Carroll PR, Dahiya R. Epigenetic changes in prostate cancer: Implication for diagnosis and treatment. *J Natl Cancer Inst.* 2005; 97:103–15. [PubMed: 15657340] * Overview of DNA methylation with an indication of the problems in reproducibility
2. Baylin SB, Belinsky SA, Herman JG, et al. Aberrant methylation of gene promoters in cancer - concepts, misconcepts, and promise. *J Natl Cancer Inst.* 2000; 92:1460–61. [PubMed: 10995795]
3. Cao R, Wang L, Wang H, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science.* 2002; 98:1039–43. [PubMed: 12351676]
4. Killian JK, Bilke S, Davis S, et al. Large-scale profiling of archival lymph nodes reveals pervasive remodeling of the follicular lymphoma methylome. *Cancer Res.* 2009; 69:758–64. [PubMed: 19155300]
5. Sparano JA. TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer.* 2006; 7:347–50. [PubMed: 17092406]
6. Parker JS, Mullins M, Cheang MCU, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* 2009; 27:1160–67. [PubMed: 19204204]
7. Cuzick J, Swanson GP, Fisher G, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. [PubMed: 21310658] Published online February 9, 2011, www.thelancet.com/oncology ** Indicates current state of the art methodology and statistics in new cancer biomarker validation.
8. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. *Nature Rev Genetics.* 2010; 11:191–203. [PubMed: 20125086] ** Comprehensive overview of epigenomic assays.
9. Yegnasubramanian S, Kowalski J, Gonzalgo ML, et al. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer Res.* 2004; 64:1975–86. [PubMed: 15026333]
10. Vasiljevic N, Wu K, Brentnall A, et al. Absolute quantitation of DNA methylation of 28 candidate genes in prostate cancer using pyrosequencing. *Disease Markers.* Published online June 2011. www.iospress.nl/loadtop/load.php?isbn=02780240
11. Brandsma JL, Sun Y, Lizardi PM, et al. Distinct human papillomavirus type 16 methylomes in cervical cells at different stages of premalignancy. *Virology.* 2009; 389:100–07. [PubMed: 19443004]
12. Li M, Chen WD, Papadopoulos N, et al. Sensitive digital quantification of DNA methylation in clinical samples. *Nat Biotechnol.* 2009; 27:858–63. [PubMed: 19684580] * A promising DNA methylation assay.
13. Rodriques EF, Santos-Reboucas CB, Goncalves Pimentel MM, et al. Epigenetic alterations of p15 (INK4B) and p16 (INK4A) genes in pediatric primary myelodysplastic syndrome. *Leuk Lymphoma.* 2010; 51:1997–94. [PubMed: 20658957]
14. Lefferts JA, Tsongalis GJ. Molecular diagnostics: parallels between infectious disease testing and emerging oncology testing. *Expert Opin Med Diagn.* 2010; 4:185–88.

Table 1

A summary of the main DNA methylation methods in use today and their strengths and limitations [1, 8,12].

<ul style="list-style-type: none"> • Bisulfite methods applicable to a variety of clinical specimens and degraded DNA (FFPE, frozens, scrapes, serum, urine, milk, aspirates, etc) <ul style="list-style-type: none"> – qMSP is simple but prone to high variability – PSQ is robust and accurate in range 3% - 97% DNA^{me} but expensive on a per-locus basis – Methyl-BEAMing is complex but may give accurate detection of very low levels of DNA^{me} ~0.1% – Arrays are expensive and better suited to biomarker discovery; however, cost/CG is low • Endonuclease methods need good DNA (FFPE not suitable), prone to false positives • MeDIP biased to higher DNA^{me}, no CG-specific data

FFPE is formalin-fixed paraffin-embedded tissue; qMSP is quantitative methylation specific PCR; PSQ is PCR-based pyrosequencing; DNA^{me} is methylated DNA; BEAMing is a DNA methylation detection method involving beads emulsion amplification and measurement; MeDIP is methylated DNA immunoprecipitation.

Table 2

Some important DNA methylated genes in prostate cancer and associated true positive (TP)* and false positive (FP)* values determined by PCR-pyrosequencing for separating cancer from non-cancer. Adapted from [10] with permission of IOS Press.

Gene	Cut off (% DNAm)	TP* (%)	FP* (%)
<i>RARB</i>	20.8	100	0
<i>GSTP1</i>	7.3	98	0
<i>HIN1</i>	26.5	98	0
<i>APC</i>	26.5	96	0
<i>BCL2</i>	10.4	96	3
<i>CCND2</i>	7.5	92	0
<i>CHD13</i>	27.5	88	0
<i>EGFR5</i>	35.5	96	14
<i>NKX2-5</i>	32.5	88	0
<i>RASSF1A</i>	14.8	92	7
<i>DPYS</i>	37.9	85	0
<i>MDR1</i>	26.8	85	3
<i>PTGS2</i>	13.8	79	0
<i>EDNRB</i>	33.5	77	0
<i>MAL</i>	3.8	85	14
<i>PDLIM4</i>	8	75	3
<i>HLAa</i>	6.11	74	14
<i>TIG1</i>	4.8	65	10
<i>ESR1</i>	38.5	62	7
<i>SLIT2</i>	33.5	56	14
<i>CDKN2A</i>	35.5	40	0
<i>MCAM</i>	1.4	45	18
<i>SFN</i>	96.5	15	3
<i>THRB</i>	27.5	26	21
<i>CDH1</i>	13.2	33	14

* True positives are the percentage of prostate cancers correctly separated from non-cancers using the indicated PSQ cutoffs. False positives are the percentage of non-cancers falsely indicated as cancer using the indicated PSQ cutoff.