Biochemical and Biophysical Research Communications 455 (2014) 70–83

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Review

Epigenetic epidemiology of cancer

Timothy M. Barrow [a,b,c], Karin B. Michels [a,d,e,*]

[a] Institute for Prevention and Tumor Epidemiology, Freiburg Medical Center, University of Freiburg, 79106, Germany
[b] German Consortium for Translational Cancer Research (DKTK), Heidelberg, Germany
[c] German Cancer Research Center (DKFZ), Heidelberg, Germany
[d] Obstetrics and Gynecology Epidemiology Center, Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[e] Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 4 May 2014
Available online 11 August 2014

Keywords:
Epigenetics
Epidemiology
Cancer
DNA methylation
Biomarkers

ABSTRACT

Epigenetic epidemiology includes the study of variation in epigenetic traits and the risk of disease in populations. Its application to the field of cancer has provided insight into how lifestyle and environmental factors influence the epigenome and how epigenetic events may be involved in carcinogenesis. Furthermore, it has the potential to bring benefit to patients through the identification of diagnostic markers that enable the early detection of disease and prognostic markers that can inform upon appropriate treatment strategies. However, there are a number of challenges associated with the conduct of such studies, and with the identification of biomarkers that can be applied to the clinical setting. In this review, we delineate the challenges faced in the design of epigenetic epidemiology studies in cancer, including the suitability of blood as a surrogate tissue and the capture of genome-wide DNA methylation. We describe how epigenetic epidemiology has brought insight into risk factors associated with lung, breast, colorectal and bladder cancer and review relevant research. We discuss recent findings on the identification of epigenetic diagnostic and prognostic biomarkers for these cancers.

Contents

Abbreviations: 5-FU, 5-flurouracil; ACT, Adriamycin, Cytoxan and Taxol; ChIP-seq, chromatin immunoprecipitation sequencing; CIMP, CpG island methylator phenotype; CRC, colorectal cancer; EPIC, European Prospective Investigation into Cancer and Nutrition; EWAS, epigenome-wide association studies; HPLC, high performance liquid chromatography; HR, hazard ratio; LUMA, luminometric methylation assay; MBD-seq, methyl-CpG binding domain protein sequencing; MeDIP-seq, methylated DNA immunoprecipitation sequencing; NSCLC, non-small cell lung cancers; OR, odds ratio; RRBS, reduced representation bisulfite sequencing.

* Corresponding author at: Obstetrics and Gynecology Epidemiology Center, Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. Fax: +1 617 732 4899.

E-mail addresses: timothy.barrow@uniklinik-freiburg.de (T.M. Barrow), kmichels@research.bwh.harvard.edu (K.B. Michels).

http://dx.doi.org/10.1016/j.bbrc.2014.08.002
0006-291X/

1. Introduction

Epigenetic epidemiology includes the study of variation in epigenetic traits and the risk of disease in populations. The merging of these two fields can facilitate insight into which epigenetic marks are associated with cancer, whether some of these marks explain the link between certain exposures and cancer, and how these epigenetic marks can be utilised as biomarkers. Epigenetic epidemiology can therefore serve to promote primary cancer prevention by identifying risk factors and their method of action, secondary prevention by establishing markers of early disease, and tertiary prevention by establishing markers of disease progression and drug resistance. Biomarkers can be identified by the adoption of appropriate study designs, in conjunction with a solid understanding of how any of the three cornerstones of epigenetics may be involved in carcinogenesis: DNA methylation; chromatin and histone modifications; and non-coding RNAs.

In addition to the identification of markers of drug resistance, insight into epigenetic dysregulation of the genome also provides new bases for therapies. For example, azacitidine is a demethylating agent that acts through inhibition of DNA methyltransferases and has been approved for use with myelodysplastic syndromes. It is currently in Phase I and Phase II clinical trials for use with diffuse large B-cell lymphomas, non-small cell lung cancers (NSCLC), breast cancer, pancreatic cancer, and oesophageal cancer. Histone deacetylase inhibitors, such as vorinostat and romidepsin, have been approved for use with the treatment of T-cell lymphomas. While no microRNA-based therapies have yet been approved for clinical usage, Phase I clinical trials are underway investigating the use of an miR-34 mimic, MRX34, with liver cancer and lymphoma patients. It is therefore evident that better understanding the epigenetic basis of cancer is enabling the development of a range of new therapeutic options.

In this review, we will describe how epidemiology studies have related epigenetic variation with environmental factors and have identified diagnostic and prognostic biomarkers that can be applied in the clinical setting. We will describe the challenges in study design, and we will review progress that has been made in identifying biomarkers of disease risk, and especially the efforts in developing non-invasive means of screening patients. We will focus exclusively on work using primary human tissues, and we will pay particular attention to large-scale and prospective studies due to their relative strength in identifying biomarkers and the epigenetic dysregulation associated with cancer.

2. The suitability of epigenetic biomarkers

Epigenetic traits have the potential to serve as excellent diagnostic and prognostic markers of cancer. In addition to the stability of DNA methylation and the resistance of microRNAs to RNase-degradation, aberrant epigenetic events are frequently observed in early-stage cancers and in adenomas [1–7]. Significantly increased stochastic variation in DNA methylation has been observed in cervical cells of normal morphology in patients who went on to develop cervical cancer [8], and there is evidence that epigenetic-based tests may offer superior sensitivity to cytology-based ones in the early diagnosis of disease [6].

While gene-specific epigenetic modifications have been reported to occur with great frequency, such as hypermethylation of the *RASSF1A* promoter in >30% of lung [2], bladder [9] and breast [10–12] tumours, it is unlikely that a test based upon a single gene will suffice to identify a large proportion of early-stage cancers. Subsequently, many studies have aimed to establish panels of genes whose synergy offer the greatest sensitivity [12–14]. An alternative approach that has been employed is to utilise the global changes that are commonly observed in tumours. There is evidence that methylation boundaries are disrupted in cancer, such as those between CpG islands and shores, and that the cancer genome contains large regions of hypomethylated blocks [15].

Summation of total DNA genome methylation may therefore be able to serve as a comparatively simple diagnostic marker. While high performance liquid chromatography (HPLC) and mass spectrometry are considered the 'gold standard' for estimating global methylation levels, they are not readily applicable to the clinical setting. Repetitive elements, such as LINE-1 and *Alu*, have been proposed as surrogate markers of global DNA methylation [16], while the pyrosequencing-based luminometric methylation assay (LUMA) has been developed as a simpler direct approach to estimating global DNA methylation [17]. However, while pyrosequencing enables the potential interrogation of DNA methylation at any CpG site, LUMA specifically assesses methylation at CCGG motifs. LUMA is performed by restriction digests of genomic DNA using *Eco*RI in conjunction with one of *Hpa*II (methylation-insensitive) or *Msp*I (methylation-sensitive) that cleave DNA at CCGG motifs, and DNA methylation can then be quantified at these sites by the ratio of the peaks from the *Hpa*II and *Msp*I digests, as determined by pyrosequencing. However, results by these two methods often do not correlate well [18,19]. In a comparison of these approaches, Lisanti and colleagues [20] reported that LINE-1 methylation corresponded best to results by HPLC (r^2 0.96), with *Alu* displaying a weaker correlation (r^2 0.78), while LUMA performed very poorly (r^2 0.04), although LINE-1 was found to overestimate hypomethylation and to underestimate hypermethylation. LINE-1 methylation is also advantageous over LUMA in displaying less variation between samples taken at different time points [21]. Other studies have reported weaker correlations (r^2 0.51–0.70) between the methylation of LINE-1 and *Alu* elements and global DNA methylation measured by HPLC [22]. While the use of repetitive element methylation is preferable to LUMA, their use as surrogate markers for global DNA methylation remains controversial. Furthermore, a clear limitation of these surrogate markers of global methylation is their lack of specificity for different cancer types. However, LINE-1 methylation is in itself of interest, both in terms of its role in carcinogenesis and its potential use as a biomarker, as will be discussed later in this review.

3. Epigenome-wide association studies

It has become increasingly common for epigenetic epidemiology studies to use microarray or next-generation sequencing technology in order to assess epigenetic variation on a substantially larger scale, rather than using a surrogate marker of global methylation, such as LINE-1 (Fig. 1). Referred to as epigenome-wide association studies (EWAS), these can offer a cost-effective means of interrogating large numbers of loci without requiring infeasible quantities of starting material, and can bring insight by identifying novel genes and pathways implicated in disease. The design and analysis of EWAS has been comprehensively reviewed by Michels et al. [23].

The most well-established of these approaches is the use of Illumina's Infinium microarray technology, which can assess DNA methylation at single-nucleotide resolution. These microarrays offer a robust and cost-effective approach to epigenome-wide screening, and their use in epigenetic epidemiologic studies is increasingly common (Fig. 1), but data analysis is complex and requires careful consideration. Whole-genome bisulfite sequencing is prohibitively expensive, and therefore sequencing-based approaches require stratification to enable their application to epidemiological studies. Methyl-CpG binding domain protein sequencing (MBD-seq) and immunoprecipitation-based approaches, such as methylated DNA immunoprecipitation sequencing (MeDIP-seq) and chromatin immunoprecipitation sequencing (ChIP-seq), require substantial