Epigenetic Research in Cancer Epidemiology: Trends, Opportunities, and Challenges

Mukesh Verma1, Scott Rogers1, Rao L. Divi1, Sheri D. Schully1, Stefanie Nelson1, L. Joseph Su1, Sharon A. Ross2, Susan Pitch3, Deborah M. Winn1, and Muin J. Khoury1,4

Abstract

Epigenetics is emerging as an important field in cancer epidemiology that promises to provide insights into gene regulation and facilitate cancer control throughout the cancer care continuum. Increasingly, investigators are incorporating epigenetic analysis into the studies of etiology and outcomes. To understand current progress and trends in the inclusion of epigenetics in cancer epidemiology, we evaluated the published literature and the National Cancer Institute (NCI)–supported research grant awards in this field to identify trends in epigenetics research. We present a summary of the epidemiologic studies in NCI's grant portfolio (from January 2005 through December 2012) and in the scientific literature published during the same period, irrespective of support from the NCI. Blood cells and tumor tissue were the most commonly used biospecimens in these studies, although buccal cells, cervical cells, sputum, and stool samples were also used. DNA methylation profiling was the focus of the majority of studies, but several studies also measured microRNA profiles. We illustrate here the current status of epidemiologic studies that are evaluating epigenetic changes in large populations. The incorporation of epigenomic assessments in cancer epidemiology studies has and is likely to continue to provide important insights into the field of cancer research. Cancer Epidemiol Biomarkers Prev; 23(2); 223–33. ©2013 AACR.

Introduction

Cancer is both a genetic and epigenetic disease (1). Epigenetics is defined as heritable changes in gene expression that are not the result of changes in the DNA sequence. Four major epigenetic mechanisms are DNA methylation, histone modification, nucleosome remodeling, and microRNAs (miRNAs), which have been shown to play a role in a number of complex diseases, including cancer (2, 3). DNA methylation in particular has been extensively assessed in breast, colon, esophageal, lung, pancreas, ovary, prostate, and other cancers (4–12). Through their effects on genomic stability and gene expression, epigenetic changes influence carcinogenesis from initiation through progression, throughout a person's lifespan, and, in some cases, across generations (13). Epigenetic events that are relevant to cancer risk are believed to occur early in cancer development, thus, may serve as potential “first hits” for tumorigenesis. Epigenetic marks reflect both an individual’s genetic background and exposure to different environmental factors, and thus may be useful for understanding the impact of environmental exposures in carcinogenesis (14). Because epigenetic changes occur before or during early tumor development, they can be modulated by diet, drugs, and other external factors such as infectious agents; epigenetic profiling may provide clues to mitigate an individual’s risk of cancer (15–17). Mill and Heijmans recently proposed that improved understanding of the mechanism of cancer progression can be understood by studying epigenetics in populations as a part of an integrated functional genomic study (18). Epigenetic changes in comparison with genetic ones are reversible and are acquired in a gradual manner and this feature provides a huge potential for cancer prevention strategies. In addition, therapies targeting epigenetic mechanisms have been shown to modify or inhibit gene expression and some have shown modest effects in clinical research settings.

To understand the current state of the field of epigenetics in cancer epidemiology, we evaluated the research project grant (RPG) awards funded by the National Cancer Institute (NCI) and the published literature in PubMed for trends in epigenetic research in cancer epidemiology across the cancer control continuum in studies conducted in human populations. This report presents a summary of our findings, particularly in the context of studying risk and cancer-relevant exposures, including nutrition and...
infectious agents, as well as practical matters such as the type of cancers being studied, and the methods and techniques that are both emerging and commonly used. Overall, we sought to present an overview of the progress in the inclusion of epigenetics in cancer epidemiology studies and to identify scientific questions related to epigenetics that cancer epidemiology can address.

Materials and Methods

Criteria and terms used for identifying cancer epigenetics and epidemiology grants and publications (search strategy and analysis)

NCI-supported RPGs related to epigenetic epidemiology funded from January 1, 2005, to December 31, 2012, were included in the portfolio analysis, and the scientific terms used in analyzing grants in different categories are shown in Table 1. The portfolio was analyzed using NCI’s Portfolio Management Application software version 13.4. Search and selection criteria used for the grant proposal to be classified as “epigenetic epidemiology” study were as follows: “One OR more terms from column 1 of Table 1” AND “one OR more terms from column 2 of Table 1 AND ‘Human.’” In addition, the criteria for inclusion of a project in the analysis were as follows: (i) the focus of the project is cancer; (ii) study involves human subjects; (iii) focus of at least one of the specific aims in the project is cancer epigenetics; and (iv) had at least 100 cases. We excluded studies that focused solely on polymorphisms in genes encoding DNA methyltransferases or miRNAs. After applying these criteria and exclusions, 79 RPGs were identified for further analysis. The authors of this report coded the grant abstracts for and analyzed the data by study design, organ site, biospecimen type used, exposure evaluated (if applicable), and method/technology used for epigenetic analysis.

In addition, we searched the published literature on epigenetic epidemiology in PubMed from January 1, 2005, to December 31, 2012, using the following search criteria: (epigenesis, genetic[mh] OR epigenomics[mh] OR “DNA methylation” OR methylation[ti] OR ’histone modification” OR CIMP[tiab] OR microRNAs[mh] OR “CpG Islands/genetics”[mh] OR methylation[ti]) AND neoplasms[mh] AND (epidemiologic studies[mh] OR risk[mh] OR ’population-based’[tiab] OR ’odds ratio’[tiab] OR hazard[tiab] OR cohort[tiab]) AND Humans[Mesh]. The following elimination criteria were applied: (i) studies with less than 100 cancer cases; (ii) studies published before 2005; (iii) experimental studies in animals; (iv) review articles and articles reporting meta-analyses; (v) letters, commentaries, editorials, and news articles; and (vi) studies that are not clearly epidemiologic. This search yielded 486 publications that are relevant for analysis. As with the grant analysis outlined above, authors listed above coded the publications on and analyzed the data by study design, organ site, biospecimen type used, exposure evaluated (if applicable), and method/technology used for epigenetic analysis. To check all manual coding, each publication and grant was coded by two different authors and any discrepancies were resolved.

Results

The type of exposures/modifiers proposed to be examined in NCI-supported grants and those examined in publications is shown in Fig. 1. The grant proposals supported by the NCI most often studied the influence of nutrition, drugs/treatment, or infectious agents. Other grants focused on energy balance, exogenous hormones, chemical exposures (e.g., pesticides), and physical exposures (e.g., radiation). In the literature, the majority of publications explored the effect of nutrition, smoking, drugs and treatments, and infectious agents on epigenetic processes.

The results from the literature analysis revealed that the number of publications addressing epigenetic changes in cancers of the colon were the largest, followed by breast and lung (Fig. 2). The NCI-supported grant portfolio analysis indicated that in the field of epigenetics, breast cancer was the most frequently studied cancer type (Fig. 2). The next two most frequently investigated cancers were colorectal and lung cancer, in that order. Other organ sites examined for epigenetic changes included pancreas, ovary, liver, gastric, and head and neck cancers.

The portfolio analysis found that 67 of 79 grants planned to analyze methylation levels, mostly at selected individual loci (data not shown). Some of these grants also proposed to assess CpG island methylator phenotype (CIMP) status, promoter methylation, and microsatellite instability. Investigators proposed to assess histone
modifications along with methylation levels in cancers of the bladder, cervix, and prostate, and also in myelodysplastic syndrome. Noncoding RNAs, particularly miRNAs, were the focus of seven studies, with 5 studies planned to assess methylation at specific loci along with miRNAs.

Concerning biospecimen types (Fig. 3) used for epigenetic analysis, tumor tissues (40 grants) and blood (35 grants) were the most predominant specimen types collected, although a few grants proposed using buccal cells, paraffin-embedded tissues, plasma, sputum, mouthwash, urine, DNA from blood spot, stool, toenail, and saliva for epigenetic analysis. Thirteen grant proposals used both blood and tumor tissue specimens for comparison analysis, whereas 22 grant proposals used normal tissue specimens for comparison with tumor specific epigenetic markers. The portfolio analysis also found that 53 of 79 studies used tumor tissue for epigenetic analysis, whereas 13
studies planned to analyze epigenetic changes simultaneously in both leukocyte and tumor DNA. Table 2 shows the types of biospecimens used for epigenetic analysis in published cancer epidemiology studies. Similar to the trend in grant proposals, both tissue samples and blood were the frequently used biospecimens for epigenetic analysis in published studies (19). The published literature also suggested that investigators used buccal cells, sputum, urine, cervical swab, and pancreatic fluid for epigenetic analysis in epidemiologic studies (Fig. 3), albeit less frequently. Several investigators sought to determine whether blood cells could be used as a surrogate tissue for examining epigenetic profiles in tumor tissue (19, 20).

<table>
<thead>
<tr>
<th>Biospecimens</th>
<th>Examples of types of cancer studied (all involved methylation marks unless noted in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>Bladder cancer (43), breast cancer (44, 45), cervical cancer (46), colon cancer (47), esophageal cancer (miRNA profile; ref. 48), gastric cancer (49), head and neck cancer (50), leukemia (51), liver cancer (52), liver cancer (miRNA profile; ref. 53), lung cancer (54), prostate cancer (55), and renal cancer (56)</td>
</tr>
<tr>
<td>Duodenal secretion</td>
<td></td>
</tr>
<tr>
<td>Exfoliated cells from oral rinse</td>
<td></td>
</tr>
<tr>
<td>Formalin-fixed, paraffin-embedded tissues</td>
<td></td>
</tr>
<tr>
<td>Pancreatic secretion</td>
<td>Pancreatic cancer (57)</td>
</tr>
<tr>
<td>Salivary rinse</td>
<td>Lung cancer (63)</td>
</tr>
<tr>
<td>Sputum</td>
<td>Lung cancer (64, 65)</td>
</tr>
<tr>
<td>Tumor tissue</td>
<td>Bladder cancer (66), breast cancer (68), colon cancer (69, 70), esophageal cancer (71), gastric cancer (72), glioblastoma (73), head and neck cancer (74), laryngeal and hypolaryngeal cancer (75), liver cancer (76), lung cancer (59, 77, 78), neuroblastoma (79), oral cancer (80), ovarian cancer (81, 82), pancreatic cancer (57, 83), prostate cancer (59, 84), rectal cancer (85), and renal cancer (59, 86)</td>
</tr>
<tr>
<td>Urine</td>
<td>Bladder cancer (87), prostate cancer (88, 89)</td>
</tr>
<tr>
<td>Uterine/cervix swab</td>
<td>Cervical cancer (46)</td>
</tr>
</tbody>
</table>
Case–control study design was the most predominantly used study design type both in RPGs and in publications (Fig. 4). Investigators also used case-only cohort and nested case–control studies for epigenetic studies, but to a lesser extent. Eight grant proposals used either a mixed-methods approach or different approaches to address cancer etiology questions.

Several methodologies are currently used to generate a global view of DNA methylation, miRNA expression, and histone modifications in various cell types. As noted earlier, most of the projects examined both global and tumor-specific DNA methylation changes (11, 20, 21). Investigators used several different types of technologies for methylation analyses (Fig. 5), and the most commonly used were methylation-specific PCR, Methyl Light, and pyrosequencing-based technologies (22). Ulrich and colleagues described a method that screens a broad scale of factors in relation to DNA methylation (23). In grants supported by the NCI, Methyl Light and pyrosequencing were most commonly used technologies for methylation profiling; whereas methylation-specific PCR, Methyl Light, was the routinely used technology in the published
studies (Fig. 5). Both grant awards and publications used mostly PCR-based miRNA analysis technologies. However, a recent trend is to use microarray or sequencing-based technologies because of their high-throughput and broad dynamic range. Pyrosequencing of methylation regions was the preferred method in publications in which global methylation was studied, see for example ref. 21. Chromatin immunoprecipitation methods using antibodies for specific histone modifications, such as ChIP-PCR, ChIP-on-ChIP (Chip on microarray Chip), and ChIP-SAGE (ChIP combined with serial analysis of gene expression), were routinely used for understanding histone modifications and associated DNA sequences in both grant awards and published articles, whereas more recent publications used ChiP-Seq.

Overall, the major hypotheses being addressed in both the grant portfolio and the publications include (i) whether diet or environmental exposure is associated with specific epigenetic marks, patterns, and/or miRNA expression, and whether epigenetic factors were related to cancer risk or survival from cancer; (ii) whether an epigenetic profile in blood or other accessible biospecimens is related to the epigenetic profile observed in the tumor; and (iii) whether an epigenetic pattern detected in tumor tissue differs from that of the normal tissue surrounding it.

In the publications studies, there are numerous examples of epidemiologic studies that reported the association of methylation and risk of different cancers. Several examples will illustrate this. In a case–control study nested in the prospective Shanghai Women’s Health Study, 192 gastric cancer cases and 384 matched controls were used to analyze methylation of Alu and long interspersed nucleotide elements (LINE), and results indicated that hypomethylation of these repeat sequences was associated with increased risk of cancer (24). However, breast cancer was not associated with Alu and LINE hypomethylation, as reported in another group of participants from a cohort in New York (25). A case–control study design demonstrated an association of HPV16 DNA hypermethylation with high-grade intraepithelial cervical cancer (26). Characterization of methylated promoters of other oncogenic papilloma viruses, HPV18, HPV31, and HPV41, indicated that methylation of these viruses was a general phenomenon and diagnostic assays, based on these results, could be developed (27). A case–control study showed a correlation between promoter methylation and testicular tumors (28). Constitutional BRCA1 promoter methylation determined in peripheral blood was shown to strongly predispose toward the development of tumors with features that resemble BRCA1-mutated tumors (29). Thus, screening of peripheral blood for BRCA1 methylation may predict risk for breast cancer.

In another study, analysis of tumor tissues indicated that insulin growth factor II (IGF-II) promoter methylation was associated with ovarian cancer progression (30). In one nested case–control study, hypermethylation of runt-related transcription factor 3 (RUNX3) was associated with advanced gastric lesions, which were susceptible to Helicobacter pylori infection, a risk factor for gastric cancer (31). Inactivation of PTEN due to hypermethylation was associated with cervical cancer (32). Studies in Colon Cancer Family Registry indicated that relatives of colorectal cancer patients with hypermethylation of MLH1 gene might be at increased risk of colorectal and gastric cancer and possibly ovarian and liver cancer (33).

Less commonly observed in the publications analysis are the effects of epigenetic factors on cancer survival, with very few examples illustrating this. Formalin-fixed tumor tissues from patients with non–small cell lung cancer were analyzed for hypermethylation of selected genes (p16, MGMT, DAPK, RASSF1, CDH1, LET7-3-a, and PTEN), and results indicated that p16 hypermethylation was associated with a worse outcome in patients with age 60 years or younger but not in older than 60 years (12).

Serum DNA shed from tumors was used to evaluate methylation of specific genes associated with breast cancer in a case–control study (34). Delgado-Cruzata and colleagues demonstrated that global DNA methylation levels measured in white blood cells may be a potential biomarker of breast cancer risk even within families at higher risk of cancer (21).

Discussion

Through our analyses of the NCI grant portfolio and the published literature on the use of epigenetic markers in epidemiologic research, we identified a number of challenges and opportunities for the field (Table 3). There is tremendous potential to integrate the knowledge of epigenetics research with genetic characteristics, environmental predisposition, and lifestyle factors in cancer epidemiologic studies. The field of cancer epidemiology research should leverage the use of epigenetic information when used appropriately to advance translational research and to enhance practice in cancer prevention, detection, diagnosis, and treatment.

Breast, lung, colon, and prostate cancers were the most studied organ sites in grants and publications, whereas a few studies investigated other cancer sites. This result was anticipated because these are among the most common cancers and epidemiologists would likely have access to higher numbers of samples from these cancer types. The cancer site in which NCI-supported research is yielding many promising epidemiologic insights is colon cancer, particularly the identification of CIMP, a molecular subtype of colon cancer. Traditional histopathologic examination cannot distinguish CIMP+ with CIMP− individuals (35, 36). A prospective study of duration of smoking cessation and colorectal cancer risk was based on CIMP analysis and results indicated a protective effect of smoking cessation on a DNA methylation–related carcinogenesis pathway leading to CIMP+ phenotype (37). In the Netherlands Cohort Study, body size and physical activity were associated with risk of colorectal cancer in CIMP+ and CIMP− individuals (38). Central adiposity increased and high physical activity decreased the risk of
Table 3. Trends, opportunities, and challenges in the cancer epigenetics and epidemiology field

<table>
<thead>
<tr>
<th>Trends</th>
<th>Opportunities</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most of the studies have been conducted using methylation markers.</td>
<td>Integrate epigenetic research with genomics, environmental predisposition, and lifestyle factors.</td>
<td>Follow an individual's epigenomic status, which changes spatiotemporally and compartmentally in tissues, and contributes to variations.</td>
</tr>
<tr>
<td>The majority of the exposures evaluated for their impact on the epigenome were nutrition, smoking, drugs and treatments, and infectious agents.</td>
<td>Incorporate epigenetics into epidemiologic studies of cancer and the environment, which could contribute greatly to our understanding of cancer risk and development.</td>
<td>Improve strategies for epigenetic data analysis and interpretation.</td>
</tr>
<tr>
<td>Most of the studies investigated epigenetic changes at specific individual loci, and very few studies explored changes at multiple loci or interactions among multiple loci.</td>
<td>Determine the stability of epigenetic marks in repeated biospecimen samples from the same people over time.</td>
<td>Conduct large-scale epidemiologic studies to determine whether epigenetic changes detected using blood samples accurately reflect both inherent and acquired epigenetic changes that contribute to cancer risk and impact outcomes.</td>
</tr>
<tr>
<td>Few investigators explored histone modifications along with methylation, nucleosome remodeling, or miRNA expression changes in cancer epidemiology.</td>
<td>Explore the use of epigenomic information to better define cancer subcategories.</td>
<td>Identify new chromatin abnormalities and their association with cancer.</td>
</tr>
<tr>
<td>Most epigenetic studies have been conducted in blood, which may not be an appropriate biospecimen.</td>
<td>Develop improved strategies for epigenetic data analysis and interpretation.</td>
<td>Develop high-throughput technologies for histone modifications and nucleosome remodeling.</td>
</tr>
<tr>
<td></td>
<td>Conduct studies that examine the relationship between epigenetic marks in germline DNA and tumor DNA.</td>
<td>Distinguish between association and causality of epigenetic mark with disease.</td>
</tr>
<tr>
<td></td>
<td>Characterize all the components of the epigenome, which might help to understand the underlying mechanism of cancer risk and identify new biomarkers of cancer initiation and development.</td>
<td>Evaluate relationships between epigenetic marks in germline versus tumor DNA.</td>
</tr>
<tr>
<td></td>
<td>Develop technologies that require small amount of samples compared with the amount currently used which might help to analyze multiple biomarkers in small samples.</td>
<td>Distinguish age-related epigenomic marks with cancer-associated marks.</td>
</tr>
<tr>
<td></td>
<td>Use exposomes with information on well-defined factors (tobacco, diet, occupational exposures, and environmental pollutants) and omics profiling (genomics, transcriptomics, epigenomics, and metabolomics) for evaluating environmental exposure and cancer risk.</td>
<td>Synthesize monoclonal antibodies for histone modifications (currently available antibodies have low dynamic range).</td>
</tr>
<tr>
<td></td>
<td>Understand the role of epigenetics in interaction of cancer-associated infectious agents with host factors.</td>
<td>Develop technologies that use smaller amounts of sample for epigenomic profiling.</td>
</tr>
<tr>
<td></td>
<td>Use resources such as family registries in identifying cancers that tend to cluster in families.</td>
<td>Increase in funds to conduct studies on epigenome-wide association studies.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Determine for how long longitudinal measurements should be taken in individuals at high risk before the disease develops. Two unresolved issues are: (i) whether epigenetic marks are transmitted intact from parent to offspring; and (ii) can we develop an epigenetic transmission test comparable with the transmission disequilibrium test used in genetic epidemiology.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider confounding factors between the epigenome and increasing age and tissue heterogeneity.</td>
</tr>
</tbody>
</table>
Cancer epidemiologic studies are needed to determine whether epigenetic changes detected using blood samples accurately reflect both inherent and acquired epigenetic changes that contribute to cancer risk and impact outcomes (19, 42). In addition, such studies could examine whether these inherent or acquired epigenetic changes reflect the pattern of epigenetic changes in the tumor tissues.

Future research in this area includes developing improved strategies for epigenetic data analysis and interpretation, determining the stability of epigenetic marks in repeated biospecimen samples from the same people over time, and studies that correlate the relationship between epigenetic marks in germline DNA and tumor DNA. Resources that are particularly valuable include studies that have prospectively collected and stored biospecimens, and thus allow for the analysis of epigenetic profiles well before cancer starts developing. The serial sampling of biospecimens at multiple time points across the life course may provide additional value allowing insights into the temporal variation in epigenetic marks over time. Some important research resources that could be explored for such studies include some of the large cancer prevention trials (such as NCI's Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial that obtained repeated biosamples during the trials). Some confounding factors need to be considered in studies of epigenomics and cancer risk, particularly age. Moreover, because multiple types of exposures (e.g., alcohol, tobacco smoking) may modify one's epigenomic profile, these exposures need to be controlled for when examining any one exposure. Studies that correlate epigenetic characteristics in different tissues and body fluids in the same persons would be very useful to validate the use of easily accessible biospecimens such as white blood cells in studies of cancer risk.

In infectious-agent–associated cancers (cervical, liver, gastric cancer, nasopharyngeal carcinoma, and Kaposi sarcoma), epigenetic epidemiology has emerged as another promising area for future research. However, these studies face temporal causality problem. In addition, the reversible and context-dependent nature of epigenetic changes poses challenges to epidemiologic studies. To overcome these challenges, pathogen-associated epigenetic studies should be accompanied by comprehensive longitudinal (multistage and multi-individual) and transgenerational data.

Investment of resources is needed in this area of cancer epigenetics and epidemiology. Epigenetics hold substantial potential for furthering our understanding of the molecular mechanisms of health-related risks due to environmental exposure and individual susceptibility.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.
EPIGENETICS IN CANCER EPIDEMIOLOGY

References
