Opportunities and Challenges for Selected Emerging Technologies in Cancer Epidemiology: Mitochondrial, Epigenomic, Metabolomic, and Telomerase Profiling

Mukesh Verma (1)*, Muin J. Khoury (1), John P.A. Ioannidis (2)

(1) Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

(2) Stanford Prevention Research Center, Department of Medicine, and Department of Health Research and Policy, Stanford University School of Medicine; and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA 94305

Running Title: Emerging Technologies in Cancer Epidemiology

Keywords: Epigenetics, methylation, mitochondria, risk assessment, telomerase,

Abbreviations: miRNA, microRNA; mtDNA, mitochondrial DNA; RTL, Relative Telomerase Length; TERT, Telomerase reverse transcriptase

Word Count: 3868

Number of Figures: 0

Number of Tables: 4

*Corresponding Author:

Mukesh Verma, Ph.D.
Chief, Methods and Technologies Branch, Epidemiology and Genomics Research

Program, Division of Cancer Control and Population Sciences, National Cancer Institute,
National Institutes of Health. 6130 Executive Boulevard, Room 5100, Bethesda, MD
20892-7324

Phone: 301-594-7344
Fax: 301-435-6609
E-mail: vermam@mail.nih.gov
Abstract

Remarkable progress has been made in the last decade in new methods for biological measurements using sophisticated technologies that go beyond the established genome, proteome, and gene expression platforms. These methods and technologies create opportunities to enhance cancer epidemiologic studies. In this article, we describe several emerging technologies and evaluate their potential in epidemiologic studies. We review the background, assays, methods, and challenges, and offer examples of the use of mitochondrial DNA and copy number assessments, epigenomic profiling (including methylation, histone modification, microRNAs (miRNAs), and chromatin condensation), metabolite profiling (metabolomics), and telomere measurements. We map the volume of literature referring to each one of these measurement tools and the extent to which efforts have been made at knowledge integration (e.g. systematic reviews and meta-analyses). We also clarify strengths and weaknesses of the existing platforms and the range of type of samples that can be tested with each of them. These measurement tools can be used in identifying at-risk populations and providing novel markers of survival and treatment response. Rigorous analytical and validation standards, transparent availability of massive data, and integration in large-scale evidence are essential in fulfilling the potential of these technologies.
Introduction

Tremendous progress has been made recently in the development and use of sophisticated technologies for enhancing biological measurements beyond the classic platforms of genomics, proteomics, and gene expression profiling. The advent of these tools offers unique opportunities and challenges for their use in human studies, and cancer epidemiology may benefit from incorporating such measurements. In this review, we assess the landscape of this emerging literature, and discuss several of these methods. We specifically address mitochondrial DNA and copy number assessments, epigenomic profiling (including assessments of methylation patterns, histone modification, microRNAs (miRNAs), and chromatin condensation), metabolite profiling (metabolomics), and telomere measurements. For each measurement platform we offer a background introduction, describe the main assays and methods, and list the main remaining challenges. Finally, we overview the use of these methods in the cancer epidemiology literature, the types of samples they can be used on, and their overall strengths and weaknesses.

Overview of the literature landscape

Table 1 shows the advent of these measurement platforms in the overall literature, and also focused on cancer, human studies, and specific types of designs. As shown, the volume of publications is still relatively limited compared with the massive literature on genomics/genetics and gene expression profiling, but many of these measurements already have as large literatures as proteomics with several tens of thousands of papers
overall, and several thousands of papers focused on cancer in particular. Methylation and
telomere-related papers have an especially strong cancer focus, with approximately 40% of
the literature focusing on cancer (as compared with 13% of the overall PubMed).
Moreover, 78-85% of the cancer literature on all these platforms is on humans. Their use in
traditional epidemiologic studies is still relatively limited, accounting for a small fraction
of this rapidly expanding literature, with only methylation-related epidemiologic studies
exceeding 1,000. Many systematic reviews have also started being published, but meta-analyses
remain uncommon, with only a few dozen being available. Most of these meta-analyses
focus on single markers, and they almost ubiquitously depend on published summary
data. This raises concerns about the breadth of coverage of the evidence and the reliability of
inferences.

Mitochondrial DNA

Background

Mitochondria play an important role in cellular energy metabolism, free-radical
generation, and apoptosis. During neoplastic transformation the mitochondrial genome
may be damaged with accumulation of somatic mutations in the mitochondrial DNA
(mtDNA). These mutations could represent a means for tracking tumor progression.
Mitochondria contain their own genome (16.5 kb), along with transcription, translation,
and protein assembly machinery and maintain genomic independence from the nucleus
(1, 2). Both germline and somatic alterations in mtDNA have been observed in cancer and
other diseases (3-6). For example, the polymorphism G10398A within the NADH
dehydrogenase (ND3) subunit of Complex I has been probed for association with breast
cancer, neurodegenerative diseases, Alzheimer’s disease, Friedreich’s ataxia, longevity, and amyotrophic lateral sclerosis (7). Somatic mitochondrial mutations have been detected in different tumor types, including in breast, colon, esophageal, endometrial, head and neck, liver, kidney, lung, oral, ovarian, prostate, and thyroid cancers, leukemia and melanoma (3–10). Most somatic mutations are homoplasmic in nature (i.e., all mitochondria carry the same mutations), with mutant mtDNA becoming dominant in tumor cells. Furthermore, the number of copies of mtDNA per cell can vary in normal and disease states (8). The mitochondrial genome lacks introns and is organized in 21 major haplogroups named after the letters of the alphabet (4, 9-12). Some haplogroups have been associated with specific cancers in specific populations (3, 4). Tools for characterizing and measuring mtDNA characteristics (including MitoChip) are available and are sufficiently high-throughput for assessing large numbers of epidemiological samples (13, 14). Numerous epidemiologic studies have been conducted using mitochondria information to examine cancer risk factors, natural history, screening markers, response to therapy and/or long term outcomes.

Assays and methods

Tissues, blood cells, exfoliated cells, and biofluids are a good source of mtDNA. To measure alterations in mtDNA (deletions, single nucleotide polymorphisms [SNPs], mutations, copy number), total DNA is usually isolated, followed by polymerase chain reaction (PCR) and nucleotide sequencing. The entire mitochondrial genome is amplified first in two long-range PCR reactions, followed by sequencing. Using MitoChip, mtDNA fragments are amplified and prepared for array hybridization according to the Affymetrix
protocol for the GeneChip Customseq array (15). Investigators also have used restriction fragment length polymorphism (RFLP) analysis for mtDNA variations in tissue samples (16).

For haplogroup analysis, a hierarchical system combines multiplex PCR amplification, multiple single-base primer extensions, and capillary-based electrophoretic separation (17). The output of the GeneChip DNA analysis generates a report of the individual and total numbers of SNPs. Sequence variations are verified against reference mtDNA. Typically samples with call rates <95% are discarded. mtDNA molecules and the virtual number of mitochondria per cell are calculated with reference to a nuclear housekeeping gene (18). Laser capture microdissection can be used to separate different cell types, e.g. epithelial cells from stroma in for ovarian cancer (19, 20). A transparent thermoplastic film is attached to the tissue on the histopathology slide and cells are localized by microscopy. Different cell types are identified and targeted through the microscope with a 15–30 μm carbon dioxide laser beam pulse. The strong focal adhesion allows selective procurement of targeted cells suitable for mtDNA isolation and characterization.

**Challenges**

Determining an accurate mtDNA copy number is difficult, because in some situations mtDNA becomes integrated into the nuclear genome at nonspecific sites (8, 21-23). Another challenge is the simultaneous characterization of nuclear and mtDNA in cases and controls. Although technically possible, such studies have not yet been conducted within large epidemiologic studies. Selection of sample source is another problem. When
mutations in blood DNA were compared with mutations in breast cancer tissue from the same patient, the mutations did not match. This suggests that blood might not be the most appropriate biospecimen (24).

**Epigenomics**

**Background**

Epigenetics may affect gene expression without changing the nucleotide sequence. The four major components of epigenetic machinery include DNA methylation, histone modification, microRNA (miRNA) expression and processing, and chromatin condensation (25, 26). Methylation and histone markers have been used in studies trying to determine the etiology of breast, colon, esophageal, gastric, liver, lung, pancreas, ovary, prostate, renal, and other cancers (25-31).

miRNA profiling has been used in cases and controls in some epidemiologic studies (e.g. disease survival in lung cancer and therapy outcome in bladder cancer) (32-35). High-throughput miRNA quantification technologies such as the miRNA microarray (36-41), bead-based flow cytometry (42), and real-time (RT)-PCR-based Taqman miRNA assay (43, 44) can be used for miRNA profiling.

Epigenetic biomarkers may offer advantages over other types of biomarkers because they reflect a person’s genetic background plus environmental exposures. Most epigenetic events occur early in cancer development and thus can be used for early detection. Epigenetic alterations also respond to environmental changes, and technologies are
available to measure these changes (45, 46). Altered epigenomic profiling can be seen in response to toxins and environmental pollutants (47-50). Different environmental exposures may affect different components of the epigenetic machinery. For example, exposure to metal carcinogens such as nickel, chromate, arsenite, and cadmium has increased recently because of occupational exposures, the massive growth of manufacturing activities, increased consumption of nonferrous materials, and disposal of waste products (51). These metals are potentially weak carcinogens: although they do not damage DNA directly (as does radiation), they may exert carcinogenic effects by epigenetic mechanisms, especially after chronic exposure (50).

Epigenetic alterations can be reversed by chemicals and can activate gene expression. Thus multiple potential uses have been proposed for epigenetic biomarkers in cancer intervention and treatment (25, 26, 29, 30, 42, 52-64). Observational, experimental, and clinical studies in different diseases, especially cancer, have shown that nutrients may influence epigenetic regulation, e.g. folic acid can supply methyl groups (57, 59, 65-68). Ingredients in some natural foods show properties similar to the inhibitors of histone acetylation.

Epidemiologic studies have been conducted in bladder (30), breast (69, 70), cervical (71), colon (72), gastric (26, 73, 74), head and neck (55, 75), liver (25, 52, 76), and renal (77-79) cancers using methylation profiling and/or polymorphisms in genes involved in initiating or maintaining methylation (53, 54, 78, 80, 81). These studies have suggested
associations between methylation markers and cancer development that need further validation. In most studies blood rather than tissue was used for analysis.

**Assays and methods**

Both tissues and biofluids have been used for epigenetic analysis. MethyLight technology, pyrosequencing, and chromatin immunoprecipitation-on-chip (ChIP-on-chip) can measure epigenetic alterations in cancer (82, 83). For methylation profiling, quantitative methylation-specific polymerase chain reaction (QMSP) assays are performed, followed by pyrosequencing (84). All assays use sodium bisulfite followed by alkali treatment (85). Bisulfite reacts with unmethylated cytosines and converts them to uracil. Methylated cytosines and other bases are not affected by bisulfite treatment. In the PCR reaction, all converted cytosines behave like uracils. MethyLight is the most common method used to determine the methylation profile in real-time (82, 86-88). MethyLight is a high-throughput, quantitative methylation assay that utilizes fluorescence-based, real-time PCR technology and requires no manipulation after the PCR reaction. It can detect a methylation allele among 1,000 unmethylated alleles.

The most common method for miRNA profiling in cancer samples is the GeneChip microarray technology developed by Affymetrix. For histone profiling, monoclonal antibodies against specific histone modifications are used for chromatin immunoprecipitation (89, 90). Another popular epigenetics technique is the ChIP assay followed by next-generation sequencing (ChIP-seq) analysis, which can detect genome-wide histone modifications and methylation (91).
Challenges

Unlike the genome, which is the same for all types of cells, the epigenome is dynamic and changes with cell type and age. Therefore, the epigenome should be evaluated several times to follow cancer-associated alterations. The biggest challenge is the choice of sample (tissue vs. blood). Blood, which is collected in most epidemiologic studies, may not be an adequate sample, because epigenetic profiles and alterations of blood cells do not match those of tissue. Use of blood cells is also problematic because blood is a mixture of cells with different half-lives, ranging from 6 hours for neutrophils to months and years for macrophages and memory cells. Epigenetic changes are dynamic and continuously evolve during cancer development. Epigenetic changes are tissue-specific and cell type-specific. The research question itself determines the most appropriate tissue to be selected for epigenetic analyses.

Histone profiling uses ChIP assays that employ antibodies against post-translational modifications of histones (92-94). Obtaining high-quality monoclonal antibodies for use against cancer-associated histone modifications is challenging, because monoclonal antibodies show batch effects (92). A central resource of large amounts of high-affinity, high-quality monoclonal antibodies is needed.

Proteins that bind to the methylated regions have been characterized, along with methylation patterns. These proteins are identified by methylated DNA immunoprecipitation (methyl DIP), which involves the hybridization of
immunoprecipitated methylated DNA to microarrays or deep sequencing of the DNA in the immunoprecipitated DNA complex (95). Improvements are required, however, to adapt this process for large-scale use in addressing such problems as low resolution when using microarrays, difficulty in obtaining sufficient coverage when deep sequencing is used, and high false-discovery rates.

Taking precautions while collecting and storing samples for miRNA analysis can be challenging in epidemiologic studies. Ideally tissue samples are snap-frozen and stored at −70°C (96, 97). Fixed tissues can be problematic for miRNA analyses if proper protocols are not applied (98, 99). In miRNA analysis, different control RNAs are run simultaneously. During miRNA profiling, primers to the internal controls should be included to avoid false-positive results (100).

**Metabolomics**

**Background**

The metabolome measures directly the output of biological pathways and thus may be more representative of the functional state of cells than other “omics” measures. Metabolomics is the study of low molecular weight molecules or metabolites produced within cells and biological systems. Metabolomic profiling may help discover new disease-risk, screening, diagnostic, and prognostic biomarkers. This technology also provides novel insights into disease mechanisms (101-103). The metabolome reflects cellular activity at the functional level and, hence, can be used to discern mechanistic information during normal and disease states (104-107). In clinical samples (serum,
urine), metabolites are more stable than proteins or RNA. The number of epidemiologic studies that use metabolomic profiling is still small compared with other technologies (Table 1), but applications are developing quickly (103, 104, 108) and validation studies are expected in the near future.

**Assays and Methods**

Metabolomic profiling is performed in blood or urine. Metabolomics involves two major technologies—mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR)—that can measure hundreds to thousands of unique chemical entities (101). The advantages of NMR include comprehensive generation of metabolite profiles by a single nondestructive method, full automation with high-throughput capacity, a well-established mathematical and statistical tool box, and very high analytical reproducibility (104). Disadvantages of NMR are its relative insensitivity in detecting metabolites with concentrations in the micromole range and above; and dependence on the quality of sample collection and handling, and on the available metadata. MS-based metabolomics typically consist of three basic components: (i) the “front end” fractionation of complex mixtures, (ii) mass spectral data acquisition, and (iii) metabolite identification and characterization by database searching. Advantages of MS include that the technique is highly sensitive and can detect metabolites with picomole concentrations, it requires small biospecimen volumes, separation by chromatography enables metabolites to be individually identified and quantified, and high-throughput automation is feasible (109, 110). Disadvantages of MS include expensive consumables, relatively lower analytical reproducibility, poor representation of highly polar metabolites when using standard
chromatography protocols, and more complex software and algorithms required for routine data analysis (111, 112).

**Challenges**

Special attention must be paid to optimize protocols for maximizing the reproducibility, sensitivity, and quantitative reliability of metabolomics analysis. Furthermore, multivariate statistical modeling approaches are needed for better visualization and analysis of data. False-positive results can make interpretation difficult unless multiplicity is properly accounted for. Advancements in automatic sample preparation and handling, robotic sample delivery systems, automatic data processing, and multivariate statistical approaches can help streamline and standardize the process, but there are a number of different platforms (113-120) and familiarity is required for their proper use.

Despite early promise, the full potential of metabolomics cannot be fully realized at the present time. Challenges include the limited availability of high-quality metabolite reference standards and of facilities that provide high-quality metabolomics services. To characterize unknown metabolites, standard, well-characterized metabolites are spiked with the clinical samples. The idea is to develop both isotopically labeled (i.e., $^{15}$N, $^{13}$C, or $^2$H) and unlabeled metabolite standards for use with MS and/or NMR, respectively. Compounds need to be synthesized in GLP laboratories with ISO 9000 certification and purified either by chromatographic methods or crystallization to > 95% purity. Classes of metabolites that require reference standards for metabolite identification include but are
not limited to glycolytic and other energy intermediates, amino acid metabolism, lipids (phospholipids, glycerolipids, sphingolipids, glycolipids, oxylipins), acylcarnitines and acylglycines, secondary drug metabolites, secondary food metabolites, and fatty acids. The lack of widely-used robust automation tools and techniques in MS-based platforms remains a major limiting factor in high-throughput discovery and in transitioning such platforms to clinical chemistry laboratories (121)

**Telomerase**

**Background**

Telomeres, the ends of chromosomes, are specialized nucleoprotein structures that consist of guanine (G)-rich repetitive DNA sequences complexed with proteins (122-124). Telomeres are required for maintenance, proper replication, and segregation of chromosomes. Without telomerase caps, human chromosomes undergo end-to-end fusion, forming dicentric and multicentric chromosomes that break during mitosis, leading to the activation of DNA damage checkpoints and initiation of the $p53$ pathway with growth arrest and cell death (125). Somatic cell telomeres shorten by 50–200 bp with each cell division, leading to replicative senescence and irreversible growth arrest. Telomere length is maintained by the protein telomerase, which adds TTAGGGG repeats at the ends of chromosomes (126). Telomerase encompasses a catalytic subunit with telomerase reverse transcriptase (TERT) activity, a telomerase RNA component (TERC) that acts as a template for DNA synthesis, and the protein dyskerin (Dkc1), which binds and stabilizes TERC. Telomerase protects the chromosome ends from unscheduled DNA repair and degradation. Both the length of the telomere repeats and the integrity of
Telomere-binding proteins are important for telomere protection. Telomere shortening below a certain threshold length and/or alterations in the functionality of telomere-binding proteins can result in loss of telomere protection, leading eventually to apoptosis (127). Telomere dysfunction has been hypothesized to promote the acquisition of genetic lesions essential to cancer progression. Several epidemiologic studies have examined the average relative telomere length (RTL) as a potential biomarker for predisposition to bladder, colon, head and neck, lung, renal, and skin cancers (126, 128, 129). Biospecimen collection response rates are greater for buccal cells than for blood samples. PCR-based assays have been developed to measure telomerase activity in epidemiologic samples (130). In addition, the area around the TERT gene has been hypothesized to be a cancer polymorphism “hot spot” in different cancers (131-134).

**Assays and Methods**

DNA from any type of cells is suitable for telomerase assays and can be isolated as described in reference (130). The PCR-based assay includes controls for inter-plate and intra-plate variability of threshold cycle values. RTL is calculated as the ratio of telomere repeat copy number to single-gene copy number in samples, compared with the reference DNA sample. Telomere length also can be determined by quantitative fluorescent in situ hybridization (TQ-FISH) (135, 136) where paraffin-embedded tissues are hybridized with fluorescence-tagged telomere probes.

**Challenges**
When studying the association between disease risk and telomere length, it is critical to determine the telomere length accurately. Discrepancies have been reported between telomere length-based studies and telomerase activity-based studies. In contrast to the belief that reduced telomere length reflects a risk of cancer, contradictory results were obtained by different investigators (134, 137-139). Nonsignificant RTL shortening was observed in a breast cancer nested case-control study (130, 138). Study limitations that affect all epidemiologic observational studies, such as subject selection procedures, confounding, measurement errors, analysis, or selective reporting, might explain discrepancies.

**Comments and Conclusions**

Table 2 summarizes some strengths and weaknesses for each of the methods discussed above. Not all samples are suitable for these methods and technologies. A list of biospecimens and the appropriate technology for analyzing samples is provided in Table 3. Selected examples where technologies described in this article are applied for different epidemiologic studies are given in Table 4.

We have described the advent of several new biological measurement methods that may be of use in cancer epidemiology and beyond. We make some final comments here about the evolution of this evidence.

First, while we discussed each platform in isolation, it is possible that information obtained from multiple markers and multiple platforms may be most informative in some
circumstances. Detecting multiple markers in cancer epidemiology has been suggested from time to time (140-143). For example, El-Tayeh et al. (141) suggested evaluating alpha-fetoprotein (AFP), alpha-L-fucosidase (AFU), transforming growth factors alpha and beta (TGF-α and TGF-β), and interleukin-8 (IL-8) simultaneously to enhance the sensitivity and specificity of hepatocellular carcinoma. Large-scale assessment at multiple times of the genome, proteome, transcriptome, and metabolome has been recently described (144), and as platforms become less expensive, such combined assessments may become feasible in larger samples of patients. Selecting between complexity and parsimony remains a prominent challenge.

Second, for most of the platforms that we described, most of the ongoing research is discovery-oriented and replication efforts are still at their infancy. Not surprisingly, no meta-analysis to-date is available on any mtDNA topic and only few have been performed on epigenetic or telomere markers. This poses challenges in interpreting the reliability of the published results. Validation efforts should include not only cross-validation or bootstrapping on the same samples and datasets, but also external validation in independent diverse datasets, preferably also by different teams of investigators (145-148). Reporting of these complex studies is also not standardized and would benefit from adoption of relevant reporting guidelines (149-151).

Third, handling complex omics and related data collected in cancer epidemiology presents another challenge. The vast amount of data and biases that are introduced create a need for fast and effective computer analysis programs and for transparent large-scale data repositories. Most studies using the discussed platforms are done by single teams,
but there is an increasing interest in larger coalitions of teams and consortia. Public availability of raw data, protocols, and analysis codes for these complex investigations could go a long way towards improving the transparency, reliability, and reproducibility of this research (145, 152).

In summary, progress continues to be made in emerging technologies in the cancer epigenetics and epidemiology fields, and some of the technologies are ready to be used in larger scale while others need improvements in analytical validity, high-throughput performance and sensitivity of detection. In the coming years, we expect that these emerging technologies may be used for different epidemiologic studies to contribute to a more comprehensive understanding of cancer risk factors, understand natural history and evaluate screening markers, and understand responses to therapy and/or evaluate longer term outcomes. Epidemiologic studies may also inform future randomized controlled trials to explore clinical utility for different applications in practice.

**Acknowledgements**

We are thankful to Andrew Freedman, Elizabeth Gillanders, Somdat Mahabir, Britt Reid, Sheri Schully, and Daniela Seminara for reading the manuscript and providing useful comments.
Table 1. Landscape of the literature on emerging types of measurements

<table>
<thead>
<tr>
<th>Measurements</th>
<th>All</th>
<th>Cancer</th>
<th>Human</th>
<th>Case control</th>
<th>SR</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Cancer</td>
<td>Human</td>
<td>Case control</td>
<td>SR</td>
<td>MA</td>
</tr>
<tr>
<td>Emerging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrial DNA</td>
<td>49990</td>
<td>5294</td>
<td>4211</td>
<td>134</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Methylation</td>
<td>63132</td>
<td>20745</td>
<td>16604</td>
<td>1274</td>
<td>137</td>
<td>25</td>
</tr>
<tr>
<td>Histone modification</td>
<td>18002</td>
<td>5862</td>
<td>4643</td>
<td>74</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>miRNA</td>
<td>19886</td>
<td>7844</td>
<td>6089</td>
<td>405</td>
<td>77</td>
<td>19</td>
</tr>
<tr>
<td>Chromatin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>condensation</td>
<td>5649</td>
<td>1300</td>
<td>995</td>
<td>10</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>7533</td>
<td>948</td>
<td>673</td>
<td>86</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Telomerase</td>
<td>25074</td>
<td>9765</td>
<td>8332</td>
<td>326</td>
<td>43</td>
<td>12</td>
</tr>
<tr>
<td>Classic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genome OR genomic*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR genetic</td>
<td>1529835</td>
<td>278971</td>
<td>224763</td>
<td>16488</td>
<td>2654</td>
<td>1080</td>
</tr>
<tr>
<td>Proteom*</td>
<td>54664</td>
<td>10108</td>
<td>8046</td>
<td>454</td>
<td>149</td>
<td>7</td>
</tr>
<tr>
<td>Gene expression</td>
<td>883585</td>
<td>201768</td>
<td>159270</td>
<td>5975</td>
<td>968</td>
<td>193</td>
</tr>
</tbody>
</table>

SR: systematic review; MA: meta-analysis.

Based on PubMed searches performed in November 2012; the last three columns are limited to studies that are already retrieved by Cancer and Human filters; the Human column is limited to studies that are already retrieved by the Cancer filter. Searches for
the different measurements are: “mitochondrial DNA”, “methylation OR hypermethylation”, “histone AND (modification OR methylation OR acetylation)”, “miRNA OR microRNA”, "chromatin condensation", “metabolomic* OR metabolome OR metabonomic***”, “telomere***”, “genome OR genomic* OR genetic”, “proteome***", “gene expression”. 
Table 2. Comparison of selected emerging methods and technologies for use in cancer epidemiology

<table>
<thead>
<tr>
<th>Method/Technology</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>mtDNA polymorphism</td>
<td>Whole genome can be sequenced for a large number of samples because of the small size of the mt genome (16.5 kb)</td>
<td>mtDNA sometimes can become integrated into the nuclear genome, and identifying the integrated mt genome in the nuclear genome is tedious</td>
</tr>
<tr>
<td>mtDNA copy number</td>
<td>May provides additional information for identifying risk- and survival-associated biomarkers</td>
<td>Experiments should be done very carefully because the number of copies varies during the disease development</td>
</tr>
<tr>
<td>Methylation profiling</td>
<td>Provides a mechanism for studying gene activation/inactivation without a change in the genome</td>
<td>Careful selection of the method is key to avoiding false-negative and false-positive results; tissue</td>
</tr>
<tr>
<td>Method</td>
<td>Description</td>
<td>Limitation</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>miRNA profiling</td>
<td>Requires a small amount of sample and provides additional information to understand epigenetically mediated gene regulation; information can be used in targeted intervention studies</td>
<td>Tissue specificity can be a concern</td>
</tr>
<tr>
<td>Nuclear magnetic resonance (NMR)</td>
<td>Quantitation of analytes is accurate with full automation and high-throughput capacity and high reproducibility</td>
<td>Identifying products can be challenging because of its insensitivity in detecting metabolites with concentrations in micromole range</td>
</tr>
<tr>
<td>Mass spectroscopy (MS)</td>
<td>Extremely sensitive, can detect analytes at picomole range; requires small biospecimen volumes</td>
<td>Requires expensive consumables; poor representation of highly polar metabolites</td>
</tr>
<tr>
<td>Metabolite profiling</td>
<td>Can be done easily in patient biofluids</td>
<td>Standards for all metabolites are not</td>
</tr>
<tr>
<td>Telomerase activity and telomere size variation</td>
<td>Suitable for paraffin-embedded tissue samples</td>
<td>Since the length of telomere changes with age, subject selection is very critical</td>
</tr>
</tbody>
</table>

mt: mitochondrial
Table 3. Sample types appropriate for selected methods and technologies that can be used in cancer risk assessment, detection, prognosis, and survival

<table>
<thead>
<tr>
<th>Sample</th>
<th>Method/Technology</th>
<th>Typical Potential Applications in Epidemiologic Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>Methylation analysis (29, 75, 153-155), miRNA analysis (156-159), mtDNA analysis (160), metabolite profiling (161, 162), telomerase assay (134, 163, 164),</td>
<td>Cancer detection, screening, survival, prognosis, risk assessment</td>
</tr>
<tr>
<td>Buccal cells</td>
<td>Methylation analysis (165, 166), mtDNA analysis (167), mtDNA copy number analysis (8)</td>
<td>Cancer detection</td>
</tr>
<tr>
<td>Circulating tumor cells</td>
<td>Telomerase (168)</td>
<td>Cancer detection and prognosis</td>
</tr>
<tr>
<td>Circulating plasma DNA</td>
<td>Methylation analysis (31)</td>
<td>Cancer prognosis</td>
</tr>
<tr>
<td>Exfoliated cells from cervix</td>
<td>Methylation analysis (169, 170), mtDNA analysis (171)</td>
<td>Cancer detection</td>
</tr>
<tr>
<td>Exfoliated cells from urine</td>
<td>Methylation analysis (172, 173), mtDNA analysis (172, 173)</td>
<td>Cancer detection</td>
</tr>
<tr>
<td>Nipple aspirate</td>
<td>Methylation analysis (174, 175), mtDNA analysis (176-178), metabolite profiling (179, 180)</td>
<td>Cancer detection and prognosis</td>
</tr>
<tr>
<td>Serum</td>
<td>miRNA profiling (32)</td>
<td>Cancer prognosis</td>
</tr>
<tr>
<td>Tissue samples</td>
<td>All methods [epigenomics (25, 26, 30, 33, 181-184), telomerase (185), mtDNA (19, 20, 186)]</td>
<td>Cancer detection and prognosis</td>
</tr>
<tr>
<td>Urine</td>
<td>Metabolomic profiling (187, 188)</td>
<td>Cancer detection, identification of lifestyle factors contributing to disease, factors contributing to cancer prognosis</td>
</tr>
</tbody>
</table>
Table 4. Selected examples where mitochondrial, metabolomic, epigenomic, and telomerase profiling was used in screening, risk assessment, and prognosis (e.g. survival, disease aggression or recurrence)

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Examples of Applications in epidemiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial DNA alterations</td>
<td>Germline mtDNA for screening populations for breast cancer (3, 5, 189, 190), breast and esophageal cancer (4), kidney cancer (191), pancreatic cancer (15); to evaluate long term survival in breast cancer (192), somatic mtDNA mutation to identify risk of lung cancer (193, 194)</td>
</tr>
<tr>
<td>Epigenetic profiling</td>
<td>Methylation profiling to evaluate cancer survival in breast cancer (69), head and neck cancer (195), and esophageal squamous cell carcinoma (31); to evaluate disease aggressiveness in bladder cancer (30, 196); to detect disease and disease stratification in head and neck cancer (75) and kidney cancer (184); histone profiling and chromatin condensation in disease recurrence in lung cancer (183), colon cancer (25) and gastric cancer (26); miRNA profiling to evaluate survival in lung cancer (32, 33), miRNA polymorphism to evaluate therapy outcome in bladder and breast cancer (34)</td>
</tr>
<tr>
<td>Metabolomic profiling</td>
<td>Metabolomic profiling in disease detection and stratification in ovarian cancer (102, 103); to evaluate nutritional factors and disease development (197, 198); life-style factors that</td>
</tr>
</tbody>
</table>
contribute to lung cancer development (187); endogenous factors that contribute to endometrial cancer (199); and prostate cancer aggressiveness (200)

**Telomerase profiling**

Association between *TERT* locus polymorphisms and predisposition to cancer (201, 202); *TERT* polymorphism in detecting risk of bladder cancer (203), ovarian cancer (134), and lung cancer (204, 205); tandem repeat minisatellite of telomerase as a risk factor for colorectal cancer (128); telomere length to assess breast cancer treatment outcome (206)

**References**


Opportunities and Challenges for Selected Emerging Technologies in Cancer Epidemiology: Mitochondrial, Epigenomic, Metabolomic, and Telomerase Profiling

Mukesh Verma, Muin J. Khoury and John P.A. Ioannidis

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst December 14, 2012.

Updated version  Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-12-1263

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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, contact the AACR Publications Department at permissions@aacr.org.