DOI:10.1158/1055-9965.EPI-07-0457

Review

Molecular Epidemiology and Biomarkers in Etiologic Cancer Research: The New in Light of the Old

Paolo Vineis[1,2] and Frederica Perera[3]

[1]Department of Epidemiology and Public Health, Imperial College London, London, United Kingdom; [2]Institute for Scientific Interchange Foundation, Torino, Italy; and [3]Department of Environmental Sciences and Columbia Center for Children's Environmental Health, Mailman School of Public Health, Columbia University, New York, New York

Abstract

The purpose of this review is to evaluate progress in molecular epidemiology over the past 24 years in cancer etiology and prevention to draw lessons for future research incorporating the new generation of biomarkers. Molecular epidemiology was introduced in the study of cancer in the early 1980s, with the expectation that it would help overcome some major limitations of epidemiology and facilitate cancer prevention. The expectation was that biomarkers would improve exposure assessment, document early changes preceding disease, and identify subgroups in the population with greater susceptibility to cancer, thereby increasing the ability of epidemiologic studies to identify causes and elucidate mechanisms in carcinogenesis. The first generation of biomarkers has indeed contributed to our understanding of risk and susceptibility related largely to genotoxic carcinogens. Consequently, interventions and policy changes have been mounted to reduce risk from several important environmental carcinogens. Several new and promising biomarkers are now becoming available for epidemiologic studies, thanks to the development of high-throughput technologies and theoretical advances in biology. These include toxicogenomics, alterations in gene methylation and gene expression, proteomics, and metabonomics, which allow large-scale studies, including discovery-oriented as well as hypothesis-testing investigations. However, most of these newer biomarkers have not been adequately validated, and their role in the causal paradigm is not clear. There is a need for their systematic validation using principles and criteria established over the past several decades in molecular cancer epidemiology. (Cancer Epidemiol Biomarkers Prev 2007;16(10):1954–65)

Introduction

In 1982, Perera and Weinstein (1) proposed ''molecular cancer epidemiology'' as a new paradigm for cancer research that incorporated biomarkers into epidemiologic studies to reveal mechanisms and events occurring along the theoretical continuum between exposure and disease. The biomarkers were categorized as markers of internal dose, markers of biologically effective dose, markers of early response/effect, or markers of susceptibility. The U.S. National Academy of Sciences in 1987 convened a workshop on the use of biomarkers in environmental health research that adopted this paradigm and expanded it to include a fourth category: altered structure and function (Fig. 1; refs. 2-5).

Received 5/21/07; revised 7/13/07; accepted 7/23/07.

Grant support: Compagnia di San Paolo to the Institute for Scientific Interchange Foundation; European Union for the project Environmental Cancer Risk, Nutrition and Individual Susceptibility (P. Vineis); NIH; National Institutes of Environmental Health Sciences grants 5P01ES09600, 5RO1ES08977, 1R01CA127532, and R01CA69094; U.S. Environmental Protection Agency grants R827027 and RD-832141; Bauman Family Foundation; and New York Community Trust.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Frederica P. Perera, Department of Environmental Sciences and Columbia Center for Children's Environmental Health, Mailman School of Public Health, Columbia University, 100 Haven Avenue, Tower 3, #25F, New York, NY 10032. Phone: 212-304-7280; Fax: 212-544-1943. E-mail: fpp1@columbia.edu

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-0457

The original model—that of a continuum of molecular/genetic alterations leading to cancer that can be accessed using biomarkers—remains fundamentally valid. Research in molecular epidemiology in the last 20 years has followed this general model and many key research findings are supportive (6). As summarized in Table 1, evidence supporting the paradigm of cancer as a continuum involving measurable molecular/genetic events includes: (*a*) studies showing a correlation between external measurements of exposure and biomarkers of biologically effective dose or early biological response/effect, such as carcinogen-DNA or carcinogen-hemoglobin adducts, in relation to exposure to polycyclic aromatic hydrocarbons (PAHs; ref. 7), acrylamide (8), styrene (9), or 1,3-butadiene (9); (*b*) studies showing overall correlations between DNA or protein adduct levels and environmental exposures to carcinogens via smoking, the workplace, or the ambient air, with significant interindividual variation in adduct levels (10); (*c*) studies confirming the ability of certain carcinogen-DNA adducts (11-13) and chromosome aberrations (14) to predict cancer; and (*d*) studies confirming the role of certain genetic variants [single nucleotide polymorphisms (SNPs)] in modulating the risk of cancer, particularly in subjects exposed to carcinogens (15).

This review provides the platform for an assessment of the potential of new epigenetic biomarkers and ''omic'' technologies in cancer prevention and recommendations

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

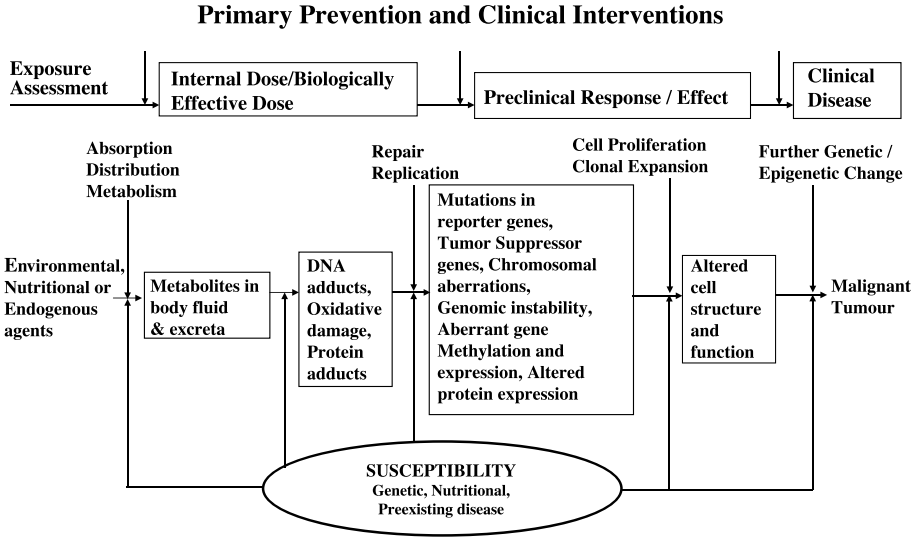

Figure 1. Updated model for molecular epidemiology. Adapted from Perera and Weinstein (1), National Research Council (2), Schulte et al. (3), and Harris (4).

for next steps in molecular epidemiologic research. In the next years, the use of new technologies/markers, such as proteomics, metabonomics, and epigenomics, will become highly relevant, particularly for longitudinal epidemiologic longitudinal studies of chronic disease. When combined with the best of the earlier validated biomarkers of dose, effect, and susceptibility, such new markers have the potential to add considerably to knowledge about the mechanistic pathways that relate pathogenic exposures to disease onset and also to serve as informative early markers of disease risk. However, research is needed to establish the validity and applicability of these new biomarkers/technologies.

In the following paragraphs, we consider the lessons we can draw from the three related areas of biomarker use: etiologic discovery, carcinogen identification and evaluation, and primary prevention. Rather than an encyclopedic review, we have chosen to present several paradigmatic examples of each area. Among promising biomarkers and technologies not included here are those related to inflammation and obesity (15, 16), genome-wide scans (16), and tumor markers (17).

The Use of Biomarkers in Etiologic Cancer Research: What Have We Learned? The Examples of Benzene and Tobacco Smoke/PAHs

Benzene. The example of benzene and hematologic malignancies is paradigmatic because it combines markers of several different types that belong to the carcinogenesis pathway shown in Fig. 1. Some of the most interesting results have come from investigations using early preclinical response markers. Prospective

Table 1. Discoveries that support the original model of molecular epidemiology

Marker linked to exposure or disease	Exposure	Reference
Internal dose		
Urinary metabolites (NNK, NNN)	Nitrosocompounds in tobacco	Hecht et al. (41, 89)
Biologically effective dose		
DNA adducts	PAHs, aromatic compounds	Tang et al. (11)
Albumin adducts	AFB_1	Groopman et al. (90)
		Santella et al. (91)
Hemoglobin adducts	Acrylamide	Hagmar et al. (8)
	Styrene	Vodicka et al. (9)
	1,3-Butadiene	Albertini et al. (92)
Preclinical effect	Exposure and/or cancer	
Chromosome aberrations	Lung	Bonassi et al. (14)
	Leukemia	Smith et al. (85)
	Benzene	Holecková et al. (93)
HPRT	PAHs	Perera et al. (94)
	1,3-Butadiene	Ammenheuser et al. (95)
Glycophorin A	PAHs	Lee et al. (96)
Gene expression	Cisplatin	Gwosdz et al. (97)
Genetic susceptibility		
Phenotypic markers	e.g., DNA repair capacity in head and neck cancer	Berwick and Vineis (98); Cheng et al. (99)
SNPs		
NAT2, GSTM	Bladder	Garcia-Closas (100)
CYP1A1	Lung	Vineis et al. (101)

Abbreviations: NNK, 4-(methylnitrosamino)-I-(3-pyridyl)-1-butanone; NNN, *N*′-nitrosonornicotine; HPRT, hypoxanthine phosphoribosyltransferase.

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

studies have shown that, at the population level, chromosome aberrations are able to predict the onset of cancer, including hematologic malignancies. Combined analyses of data from Nordic and Italian prospective cohort studies involving 3,541 subjects found that chromosomal aberrations were significant predictors of cancer (14). In the Italian cohort, cancer predictivity of high chromosomal aberrations was greatest for hematologic malignancies, with a standardized mortality ratio of 5.49 (95% confidence interval, 1.49-140.5; ref. 18).

Specific chromosomal aberrations have been observed in both preleukemia and leukemia patients exposed to benzene as well as in otherwise healthy benzene-exposed workers (19). By use of fluorescent *in situ* hybridization and PCR, Zhang et al. found that high occupational benzene exposure increased the frequencies of aberrations in chromosomes 5, 7, 9, 8, and 11, aberrations that are frequently seen in acute myeloid leukemias and in preleukemic myelodysplastic syndrome.

In the same studies on Chinese workers, protein expression patterns were detected by surface-enhanced laser desorption/ionization-time of flight, a technique widely used in proteomic research (see below). Surface-enhanced laser desorption/ionization-time-of-flight analysis of exposed and unexposed subjects revealed that lowered expression of PF4 and CTAP-III proteins is a potential biomarker of the early biological effects of benzene and may play a role in the immunosuppressive effects of benzene (20).

Finally, Lan et al. (21) investigated 20 candidate susceptibility genes in the same Chinese cohort. After accounting for multiple comparisons, SNPs in five genes were associated with a statistically significant decrease in total WBC counts among exposed workers [*IL-1A* (−889C>T), *IL-4* (−1098T>G), *IL-10* (−819T>C), *IL-12A* (8685G>A), and *VCAM1* (−1591T>C)]. This report provides evidence that SNPs in genes that regulate hematopoiesis influence benzene-induced hematotoxicity. (We note that an association between certain SNPs and cancer in one study does not necessarily translate to other populations and exposure scenarios and must be confirmed).

Tobacco Smoke/PAHs and Lung Cancer. Most of the molecular epidemiologic research on lung cancer has targeted tobacco smoke and PAHs as model carcinogens. PAHs are found in tobacco smoke, in outdoor air from fossil fuel combustion, in indoor air from cooking, heating, and smoking, and in the diet (22). PAHs are animal carcinogens, including transplacentally, and human carcinogens (23, 24). PAHs, such as benzo(*a*)pyrene, form adducts with DNA, a mechanism considered to be a critical early event in PAH-induced tumorigenesis (25, 26).

Since 1982 (27), several molecular epidemiologic case-control studies have found associations between PAH-DNA or related aromatic-DNA adducts measured in WBCs and lung cancer (11, 28-30). Whereas the earlier studies were subject to the limitations of retrospective studies, two later investigations nested within prospective cohort studies found adducts to be predictive of lung cancer within specific exposure groups (11). A meta-analysis of aromatic/PAH-DNA adducts and lung cancer concluded that current smokers with high levels of adducts have an increased risk of lung cancer (13).

Several lines of evidence support the causal role of PAH/aromatic adducts in lung cancer: (*a*) significant correlations have been observed between adducts in peripheral blood and lung tissue (29, 31); (*b*) in lung tumors of smokers, the patter of mutations in the *p53* tumor suppressor gene (which is mutated in 40-50% of lung tumors) is consistent with the types of DNA adducts and mutations induced experimentally by benzo(*a*)pyrene (32, 33); and (*c*) similar to the example of benzene, certain SNPs or genes involved in the metabolism or detoxification of PAHs or in the repair of PAH DNA adducts have been implicated as effect modifiers in lung carcinogenesis.

More recently, as will be discussed in ''Categories of Epigenetic and Omic Biomarkers,'' epigenetic mechanisms have emerged as important in lung cancer related to tobacco smoking (34-36).

Conclusion. In summary, studies using biomarkers of biologically effective dose (adducts in the case of tobacco smoke/PAHs) or preclinical effect (chromosomal aberrations and proteomics in the case of benzene; *p53* for tobacco smoke/PAHs) and individual susceptibility (SNPs) have been valuable in elucidating the steps that link benzene to the onset of leukemia and tobacco smoke/PAHs to the onset of lung cancer.

The Use of Biomarkers/Molecular Epidemiology in Cancer Risk Assessment and Prevention

Biomarkers in Hazard Evaluation: The Example of the IARC Monographs. Biomonitoring and molecular epidemiologic studies have provided mechanistic data on carcinogens that have been used in risk assessment and in some cases regulation of those carcinogens. The Monographs on the Evaluation of Carcinogenic Risks to Humans have been published by the IARC since 1971 as a guide to regulatory and public health agencies in their decision making. The Monographs report the assessment made by groups of experts of the weight of evidence in humans and experimental animals based on detailed guidelines (see the IARC Web site).[4] The evidence in humans or animals is graded using the categories of ''sufficient,'' ''limited,'' ''inadequate,'' and ''evidence of lack of carcinogenicity.'' Then, animal and human evidence is combined for the overall evaluation that consists of classifying the agents into group 1 (carcinogenic to humans), group 2A (probably carcinogenic to humans), group 2B (possibly carcinogenic to humans), group 3 (not classifiable), and group 4 (probably not carcinogenic to humans). Since 1997 (Monograph 54), mechanistic evidence, including biomarker data in humans or animals, has been used to ''upgrade'' or ''downgrade'' the classification of the agents.[5]

Table 2 shows the chemicals or exposures for which mechanistic data contributed substantially to the final evaluation of carcinogenicity and Appendix A explains the underlying rationale. The application of mechanistic

[4] http://www.iarc.fr

[5] Mechanistic data are not defined precisely, but the IARC Monograph Preamble states that ''mechanistic and other evidence judged to be relevant to an evaluation of carcinogenicity and of sufficient importance to affect the overall evaluation... this may include data on preneoplastic lesions, tumour pathology, genetic and related effects, structure-activity relationships, metabolism and toxicokinetics, physicochemical parameters and analogous biological agents.''

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

Table 2. Chemicals or agents for which evidence from mechanistically relevant biomarkers was used in the assessment of carcinogenic hazards to humans within the IARC Monographs program

Mechanistic evidence used to upgrade hazards from 2A (probably carcinogenic to humans) to 1 (carcinogenic to humans)		
Ethylene oxide		
Neutrons		
2,3,7,8-TCDD		
Mechanistic evidence used to upgrade hazards from 2B (possibly carcinogenic to humans) to 2A (probably carcinogenic to humans)		
Acrylamide	Dibenz(*a*,*h*)anthracene	MOCA
Adriamycin	Diethyl sulfate	Methyl methanesulfonate
Azacitidine	Dimethylcarbamoyl chloride	MNNG
Benz(*a*)anthracene	1,2-Dimethylhydrazine	*N*-methyl-*N*-nitrosourea
Benzidine-based dyes	Dimethyl sulfate	*N*-nitrosodiethylamine
Benzo(*a*)pyrene	Epichlorohydrin	*N*-nitrosodimethylamine
Captafol	Ethylene dibromide	Procarbazine hydrochloride
Chloramphenicol	*N*-ethyl-*N*-nitrosourea	Styrene-7,8-oxide
CCNU	Etoposide	Teniposide
Chlorozotocin	Glycidol	Tris(2,3-dibromopropyl) phosphate
Cisplatin	IQ	UV radiation A, B, C
Clonorchis sinensis	5-Methoxypsoralen	Vinyl bromide
Mechanistic evidence used to upgrade hazards from 3 (not classifiable as to carcinogenicity to humans) to 2B (possibly carcinogenic to humans)		
Aziridine		
Bleomycins		
1,2-Epoxybutane		
Gasoline		
Mechanistic evidence used to downgrade hazards from 2B (possibly carcinogenic to humans) to 3 (not classifiable as to carcinogenicity to humans)		
Amitrole		
Atrazine		
Di(2-ethylhexyl)phthalate		
Ethylenethiourea		
D-Limonene		
Melanin		
Saccharin		
Sulfamethazine		

Abbreviations: MOCA, 4,4′-methylenebis-2-chloroaniline; 2,3,7,8-TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; MNNG, *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine; CCNU, 1-(2-chloroethyl)-3-cyclohexyl-L-nitrosourea.

evidence from biomarkers and molecular epidemiology to carcinogen hazard identification is justified by the fact that many animal carcinogens have not been adequately studied epidemiologically. This situation is not likely to be due to chance because there is usually little motivation to fund studies in human populations exposed to specific chemicals, the statistical power of studies that are conducted tends to be low, and exposure assessment is difficult. As a consequence, these studies are few and difficult to interpret at best. Molecular evidence in humans can reasonably be integrated with the experimental data to fill the gap.

The Use of Biomarkers in Exposure Intervention, Chemoprevention, and Dietary Preventive Trials

Biomarkers of DNA Damage in Preventive Trials. Primary prevention encompasses a spectrum of measures that include reduction and avoidance of exposure, inhibition of carcinogen activation after it has entered the body, blocking interactions with the genome, and suppression of the propagation of premalignant changes. Many preventive trials have used DNA damage as an intermediate biomarker or end point. Examples of molecular epidemiologic studies that have documented the benefits of reduction of exposure include studies of smokers enrolled in smoking cessation programs. In one study, blood samples were drawn from 400 participants before they began the program and then at multiple time points after smoking cessation. Forty of the 400 smokers were successful at quitting (compliance was verified by assaying cotinine levels in blood). Levels of the biomarkers, PAH-DNA, and 4-aminobiphenyl-hemoglobin adducts were significantly reduced by 8 weeks after quitting, reflecting cessation (37). Similarly, following a reduction in air concentrations of PAHs in a Finnish iron foundry, both PAH-DNA and aromatic-DNA adduct levels in workers' blood samples declined significantly (8). In a follow-up to the smokers' cessation study, a randomized clinical trial of vitamins E and C in smokers found that, among treated women, but not treated men, this was a significant decline in PAH-DNA adducts compared with controls, suggesting a benefit of antioxidants in reducing procarcinogenic DNA damage in female smokers and that hormonal factors may be important in modulating antioxidant effects.

Several dietary or vitamin supplementation randomized studies have used oxidative damage markers as intermediate outcomes. Free radicals, which are produced naturally in the body, can cause oxidative damage of DNA, lipids, proteins, and other cell constituents, contributing to the onset of cancers and other chronic diseases (38). Oxidative damage to DNA plays a major role in carcinogenesis, and all living cells have defense mechanisms in place to counter this damage. The simplest mechanism involves foods and nutrients with antioxidant properties, which work by intercepting free radicals and preventing cellular damage (38, 39). To establish the potential chemopreventive properties of

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

antioxidants, investigators have used markers such as 8-hydroxy-2′-deoxyguanosine and the comet assay as intermediate markers in interventions (38). Moller and Loft (39, 40) have reviewed these intervention studies and concluded that most had extremely low statistical power (sample size usually ≤20), that the interventions led to modest changes in 8-hydroxy-2′-deoxyguanosine (~10%), and that single doses of antioxidants seemed to be more effective than multiple doses, whereas the type of antioxidant was not crucial.

Other research, not reviewed here, has used urinary metabolites of tobacco-specific carcinogens as intermediate biomarkers in chemoprevention studies [see Hecht et al. (41) for example]. In conclusion, promising markers are available, such as DNA adducts, oxidative damage, and urinary metabolites, but they await application in large-scale and well-designed trials.

Enzyme Induction as a Biomarker/Target in Preventive Trials. A wide range of chemicals that variously interact with cellular functions has been studied experimentally in the attempt to block the processes that lead to carcinogenesis (42, 43). Examples of anti-inflammatory compounds having *in vitro* chemopreventive activity are piroxicam, sulindac, aspirin, celecoxib, and curcumin. The selective estrogen receptor modulators, tamoxifen and raloxifene, are beneficial in the prevention of estrogen-dependent tumors. Retinoids and vitamin A derivatives, such as targretin and fenretinide, have also been investigated in the prevention of tumors as have compounds containing sulfur, such as sulforaphane and oltipraz and the steroid dehydroepiandosterone. The biomarkers used to determine the chemopreventive activity of these compounds are quite often activities of enzymes (42, 43).

Probably, the most widely studied among these compounds is oltipraz. Oltipraz has been introduced in high-risk populations based on its activity in inducing phase II enzymes (i.e., the detoxification-induction paradigm). However, results in smokers are not very encouraging. In one study, to determine if oltipraz could induce glutathione *S*-transferases and thereby reduce adduct levels of tobacco smoke constituents in the lungs and other target organs, chronic smokers were enrolled in one of three arms: 400 or 200 mg/wk oral oltipraz or placebo. There was no significant difference between treatment and placebo groups in PAH-DNA adducts in lung epithelial cells measured by immunoperoxidase staining. Likewise, no significant differences were found in PAH or benzo(*a*)pyrene-7,8-diol-9,10-epoxide adducts measured in blood, oral lining cells, or bladder lining cells (44).

Oltipraz has been used more extensively in the prevention of hepatocarcinoma from the grain/cereal contaminant aflatoxin, again based on the detoxification-induction paradigm (41). Oltipraz is a potent inducer of phase II enzymes involved in the detoxication of aflatoxin. Another agent, chlorophyllin, impedes the bioavailability of aflatoxin by forming molecular complexes and enhancing their elimination in the fecal stream (41). Several recent randomized clinical trials with oltipraz and chlorophyllin have been conducted in individuals exposed to dietary aflatoxin and at high risk for development of liver cancer. Both chemopreventive agents modulated levels of aflatoxin-protein adducts in the treated subjects (41). However, these were small-scale studies and there is no evidence yet that the administration of oltipraz can be a reasonable strategy for cancer prevention as an alternative to the improvement of food quality. A simple and effective way of preventing aflatoxin contamination of food has been shown in an intervention trial in Africa (43). In this example, rather than acting through the expensive administration of chemopreventive agents in a very deprived context, the authors modified the practice of storing food, thus preventing degradation and contamination from molds producing aflatoxin.

New Epigenetic and Omic Biomarkers

Categories of Epigenetic and Omic Biomarkers. Several new and exciting biomarkers are becoming available for epidemiologic studies, thanks to the development of high-throughput technologies and theoretical advances in biology. However, most of these markers have not yet been validated, and their role in the causal paradigm is not clear. An exhaustive review is not possible here, and we refer the reader to other critical reviews, particularly for gene expression and toxicogenomics (the study of the complex interaction between the genome and chemicals in the environment or drugs, cells as they relate to disease causation; refs. 34, 35, 43, 45-47). Here, we describe some achievements and some methodologic issues raised by the emerging fields of epigenetics and omics.

Epigenetics and Promoter Methylation. Epigenetic mechanisms of carcinogenesis (i.e., mechanisms that do not depend on structural changes in DNA but on functional regulation such as DNA methylation) are increasingly identified as key steps in the pathway from exposure to cancer. DNA methylation is an important epigenetic determinant of gene expression because it determines the process by which the instructions in genes are converted to mRNA, directing protein synthesis (36). DNA methylation [i.e., the covalent addition of methyl groups (CH_3) to cytosine that precedes a guanosine in the DNA sequence (the CpG dinucleotide)] occurs naturally and is thought to have a role in suppressing gene expression. CpG dinucleotides are enriched in the promoting regions of genes (CpG islands), and for this reason, methylation is thought to be involved in gene expression. Hypermethylation of promoter regions is associated with gene transcriptional silencing and is a common mechanism for the inactivation of tumor suppressor genes in human cancer (48). DNA methylation is heritable (i.e., it can pass from one generation of cells to the next).

Several genes are commonly the target of promoter hypermethylation, for example in lung cancer, including the *p16* gene (*p16I NK4a/CDK N2A*), *DAPK, RAR-β, RASSF1*, and O^6*MGMT* (a DNA repair gene; ref. 49). Global hypomethylation has also been reported (50). Both current and former smoking have been associated with aberrant *p16, DAPK, RASSF1A*, and *RAR-β* methylation (49). A recent investigation found that two alternative pathways could be detected in the biopsies of smoking and nonsmoking lung cancer patients, one involving methylation and *K-ras* mutations and the other *EGRF* mutations in the absence of gene methyl-

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

ation (51). In a prospective study, promoter hypermethylation of multiple genes (including those mentioned above) in the sputum was able to predict lung cancer onset with sensitivity and specificity of 64% (52). Notably, aberrant promoter methylation can be detected in the plasma of lung cancer patients (53), and a recent article by Russo et al. (54) described high frequencies of *ECAD* and *DAPK* methylation in lymphocytes of smokers versus nonsmokers. The capacity of some airborne particulate carcinogens (including those from tobacco smoke) to induce hypermethylation in the regulatory regions of tumor suppressor genes has also been investigated in animal studies (55). Overall, the animal models support involvement of promoter methylation and other epigenetic mechanisms in carcinogen-induced lung carcinogenesis (52, 56). Acetylation is another mechanism that is key in epigenetic pathways, although it has been studied less extensively than methylation in cancer epidemiology (57).

Metabonomics. Metabonomics is the study of the complete set of low-molecular weight metabolites present in a cell or organism at any time. With high-throughput techniques (nuclear magnetic resonance spectroscopy and liquid chromatography-mass spectrometry), it is possible to measure a large number of metabolites simultaneously and to define individual metabolic profiles that can be used to predict the onset of common diseases (58). Use of data processing and chemometric models has already allowed the characterization of certain disease states and metabolic disorders (58). Although several cross-sectional metabonomic studies of various cancers have been undertaken (59, 60), no longitudinal study has yet been carried out. Few validation studies of metabonomics have been published. One investigation has analyzed repeat samples from dietary studies (61). As part of a major phenotyping investigation, the authors used high-resolution ^{1}H nuclear magnetic resonance spectroscopy to characterize 24-h urine specimens obtained from population samples in Japan (n = 259), Chicago, Illinois (n = 315), and China (n = 278). They investigated analytic reproducibility, urine specimen storage procedures, interinstrument variability, and split specimen detection. The multivariate analytic reproducibility of the nuclear magnetic resonance screening platform was >98%, and most classification errors were due to heterogeneity in handling of urine specimens. In addition, cross-population differences in urinary metabolites could be related to genetic, dietary, and gut microbial factors.

Proteomics. Proteomics is the analysis of the total protein output encoded by the genome. Proteomic research to date has mainly involved proteomic pattern profiling of tissue and body fluids by mass spectrometry with sophisticated bioinformatic tools to identify proteins within the complex proteomic profile that discriminate between normal, benign, or disease states (52, 56). Proteomic approaches have been used with some success for the molecular classification of tumors (62) and to develop a discriminatory pattern that distinguishes normal sera from that of ovarian, prostate, breast, lung cancer, and, more recently, cutaneous T-cell lymphoma (56, 63-70). To a lesser extent, proteomics has been applied to study physiologic or pathologic changes associated with external ''environmental'' exposures. Proteomic approaches have identified, for example, changes in proteins associated with oxidative stress (71). The investigation of proteomic patterns could be a powerful tool both for the identification of intermediate changes that relate environmental exposures to disease onset and as an early marker of cancer. However, methodologic issues need to be resolved before application in prospective studies. In a critique of early articles, Diamandis (72) identified several methodologic problems. These included the lack of reproducibility of analytic methods; the lack of reproducibility of proteomic patterns in different series of patients and by different laboratories; unresolved effects of different protocols for sample collection, processing, and storage; possible selection effects in cases and controls (bias, confounding) partly because of the opportunistic sampling that characterized the early studies; the possible effect of drugs/other treatments; and inappropriate or nonreproducible data analysis. Many of these concerns apply to other epigenetic and omic biomarker studies as well and have been successfully addressed in more recent studies.

In conclusion, for the epigenetic and omic technologies, systematic validation studies are urgently needed.

Incorporation into Etiologic Studies of New Intermediate Biomarkers Using Omic Technologies. It is timely to discuss the challenge of incorporating epigenetic and omic biomarkers into the molecular epidemiology of cancer because, in contrast to traditional methods, high-throughput omic and epigenetic technologies allow massive investigations not based on a priori hypotheses. Such intermediate markers (either reflecting early preclinical effects or early disease) could be used for etiologic purposes (to investigate the causes and mechanisms of disease onset) or for clinical purposes (early diagnosis, prognosis, and follow-up). We will refer exclusively to the former, but many of our considerations apply to the clinical purposes as well (see refs. 73, 74 for a review of biomarker-based tools for cancer screening, diagnosis, and treatment).

As we have seen, the term intermediate marker has been used most often in the context of chemoprevention. Here, we have used the term in a very broad sense to encompass all measurable markers (in body fluids or in cells) that lie within the putative causal pathway linking an exposure to the onset of disease. If we refer to the classic scheme (Fig. 1), intermediate markers can play a role in each of the steps: they can be related to exposure (e.g., metabonomics), they can be related to early changes in the causal pathways leading to disease (such as promoter methylation, gene mutations, or telomere length), or they can express epiphenomena of preclinical disease (e.g., mutations in plasma DNA as a consequence of tumor apoptosis). It is very important that the biological significance of a marker is made explicit beforehand because false expectations can arise as a consequence of an erroneous interpretation of the role of a biomarker. For example, some markers (those on the right side of the scheme) have clinical relevance or can be useful for screening, others cannot. In the study of etiology, there are at least two important roles that intermediate markers can play: (*a*) to increase sensitivity and/or an earlier detectability in comparison with other markers and (*b*) to show biological plausibility for an exposure proposed as etiologically relevant.

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

Past experience with more traditional biomarkers is relevant for several reasons: it has confirmed the main theories on carcinogenesis, with evidence coming from disparate fields such as epidemiology and molecular biology; it has shown that very rarely does a single biomarker allow an exhaustive understanding of the carcinogenesis process and integration of biomarkers is necessary; and it has highlighted the many facets of biomarker validation. At the same time, the current era is different from the past and deserves special attention for the following reasons: (*a*) omic and new epigenetic methods tend to be discovery-oriented rather than oriented to testing specific hypotheses, (*b*) the main feature of current technologies is the ability to do massive testing of markers (i.e., thousands of markers at a time, potentially in thousands of subjects), and (*c*) such new intermediate markers introduce increased potential for confounding. So, although our ability to measure new intermediate markers has considerably increased, making the current phase potentially very exciting, methodologic challenges have expanded more than proportionally. In fact, much uncertainty surrounds the validity and applicability of new technologies [see Ransohoff (75) for example].

Another important difference between the newer and the earlier biomarkers is that the traditional cancer paradigm was very much centered around DNA and mutations, whereas recent research has uncovered several additional intermediate steps between genotype and phenotype and the importance of gene expression/modulation for carcinogenesis. Therefore, we expect that new biomarkers will be less centered around mutations, although combinations of both types of biomarkers could be informative because pathways are not mutually exclusive.

Validating Promising Intermediate Markers. The early story of proteomics is an example of the risks implicit in the premature use of a technology that has not been sufficiently validated. In particular, a concept that is often unclear is the difference between technical and field validation. Technical validation has to do with intrinsic measurement error and analytic sensitivity. Field (or epidemiologic) validation is related to how a certain marker behaves in the population, depending on biological variability within the population (3).

Biomarker validation requires several steps. A marker may be extremely powerful in increasing our understanding of the natural history and pathogenesis of a disease but

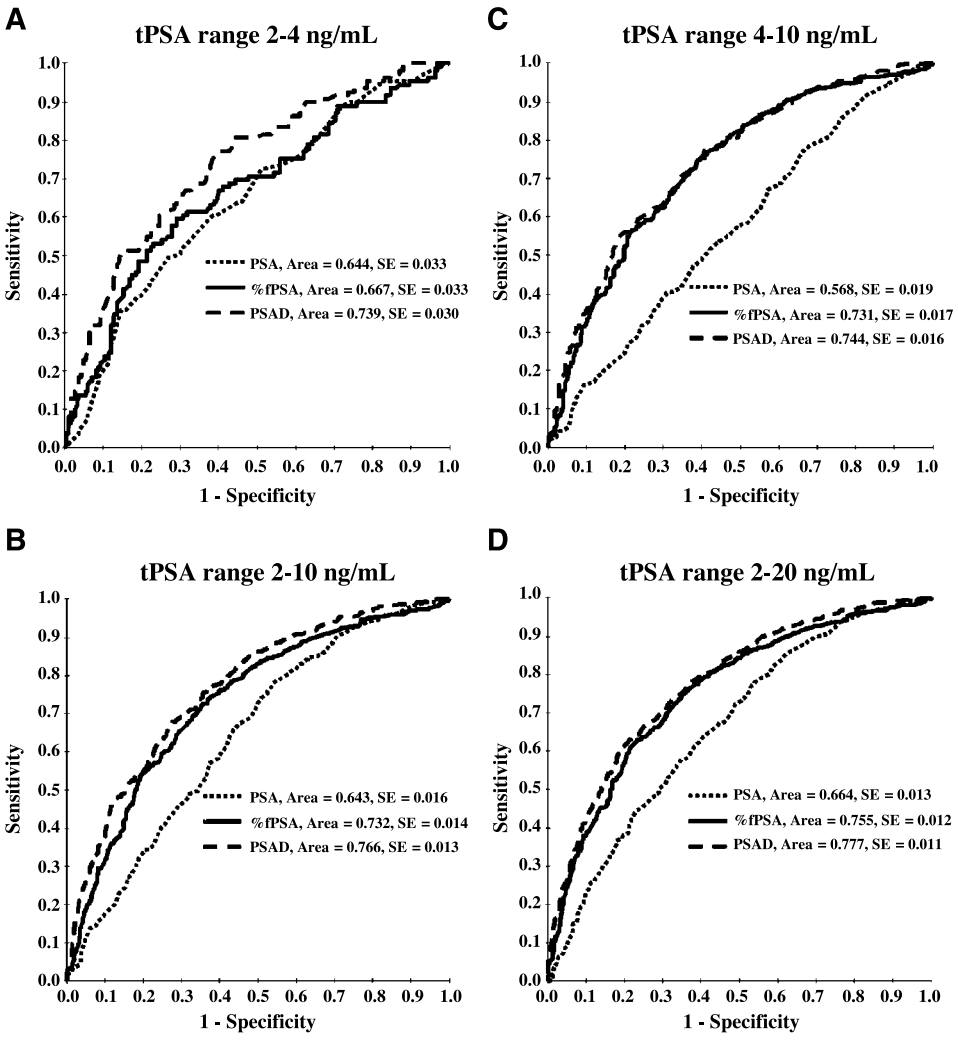

Figure 2. Receiver operating characteristic curve analysis of total PSA (*tPSA*), percent free PSA (*%fPSA*), and PSA density (*PSAD*). Data are shown for 307 patients with total PSA values in the range of 2 to 4 ng/mL (**A**), 1,282 patients with total PSA values in the range of 2 to 10 ng/mL (**B**), 975 patients with total PSA values in the range of 4 to 10 ng/mL (**C**), and 1,809 patients with total PSA values in the range of 2 to 20 ng/mL (**D**). Adapted from Stephan et al. (102).

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

may still do very poorly as a predictor for preventive or clinical purposes. One of the most important goals of validation is to characterize the ability of the marker to predict disease and, in intervention studies, reflect the modification of the natural course of disease.

Figure 2 shows one of the main summary measures of the contribution of a biomarker to the prediction of disease onset, the receiver operating characteristic curve. The figure refers to the ability of prostate-specific antigen (PSA) levels (a serum tumor marker) to predict the presence of prostate cancer. The receiver operating characteristic curve is a measure of the overall capability of the marker to predict the disease, which is a function of its sensitivity and specificity. An ''area under the curve'' of 1 or close to 1 indicates perfect prediction, whereas an area close to 0.5 indicates random association between the marker measurement and the probability of disease onset. In the figure, the maximum area under the curve is 0.77, not a particularly good performance. Unfortunately, we do not have yet similar examples of predictive ability for newer omic biomarkers. Although the receiver operating characteristic curve is established for PSA and other traditional markers, it is still unknown for markers that could be potentially much more effective. Although large trials have yet to be completed, it seems that PSA is not very useful for population screening of prostate cancer because it does not distinguish between the rare cancers that progress to frank malignancy and lead to metastases and death and the much more frequent cancers that would remain clinically inapparent. It is expected that gene expression microarrays or proteomics could be more predictive than PSA, but candidate biomarkers have yet to be identified and receiver operating characteristic curves developed.

A major aim of biomarker validation is to characterize biomarker variability. The main components of biomarker variability that affect the design and interpretation of epidemiologic studies are (*a*) biological variability related to the subject [i.e., variability between subjects (intersubject) and within subjects (intrasubject)]; (*b*) variability due to measurement error, including intralaboratory and interlaboratory variability; and (*c*) random error. Methodologic issues should be discussed within the context of specific biomarker categories. When we design and analyze an epidemiologic study using biomarkers, we want to minimize total intragroup variability to identify intergroup differences (e.g., between exposed and unexposed or between diseased and healthy subjects), if they exist. Total intragroup variation is the weighted sum of intersubject, intrasubject, sampling, and laboratory variation, with weights that are inversely correlated to the numbers of subjects, number of measurements per subject, and analytic replicates used in the study design, respectively. Obviously, if we do not have detailed information, we cannot adjust for intragroup variation. Therefore, in epidemiologic studies using biomarkers, it is important to collect, whenever possible, (*a*) repeat samples (day-to-day, month-to-month, or year-to-year variation may be relevant depending on the marker), (*b*) potentially relevant information on subject characteristics that may influence intersubject variation, (*c*) information on conditions under which samples have been collected and laboratory analyses have been conducted (batch, assay, and specific procedures). To know more about how the variability in laboratory measurements influences study design decisions, see Rundle et al. (76).

Understanding Whether an Intermediate Biomarker Belongs to the Causal Pathway. One of the main challenges with intermediate biomarkers is to understand whether they belong to the causal pathway between exposure and disease, whether they are simply a side effect of exposure or disease, or whether their measurement is confounded by some other exposure. For example, it is likely that certain mutations are genuine intermediate markers in the causal pathway, whereas others are a consequence of the disease, such as genomic instability that arises in cancer cells.

An example of the uncertain status of a marker is the association between folic acid and colorectal cancer: the inverse association observed between the two, in fact, could be due to confounding by other dietary factors and not a causal association. A way to show that folic acid can contribute to the disease process independently of confounding by other risk factors would be to show that the levels of folate are associated with different genotypes for genes involved in folate metabolism, such as *MTHFR*, and that such genotypes also predict the disease. ''Mendelian randomization'' has been suggested in fact as a way to overcome confounding (i.e., by exploiting the random allocation of alleles from parents to the offspring). The association between a gene variant and a disease is not subject to the confounding by behavioral or socioeconomic factors that has led to misleading findings in many conventional observational epidemiologic studies. In the case of folic acid, because the different alleles of the *MTHFR* gene are independent of confounding factors (e.g., dietary habits), and are assorted randomly from one generation to the next, the finding that a polymorphism in MTHFR was associated with cancer would provide indirect proof of a genuine involvement of folate in the etiology of the disease (77).

Complementary Study Designs: The ''Meet in the Middle'' Concept. The U.S. National Institute for Environmental Health Sciences has published a strategic plan that raises many of the points above and, in particular, stresses the importance of integrating environmental exposures within the study of the natural history of disease. In describing the role of different biomarkers and the need for their validation, the National Institute for Environmental Health Sciences document also launches a plan for the development of new markers for exposure assessment. The underlying philosophy is expressed thus: ''The study of how an environmental agent affects molecular targets, cellular function, tissue function, and organism survival will need to be related up and down a continuum of biological complexity that ultimately informs us about the etiology, pathogenesis, and distribution of disease. Scientific contributions from epidemiology, toxicology, molecular and cellular biology, bioinformatics, clinical medicine, and many other fields will need to be coordinated and integrated.''[6]

[6] http://www.niehs.nih.gov/external/plan2006/

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

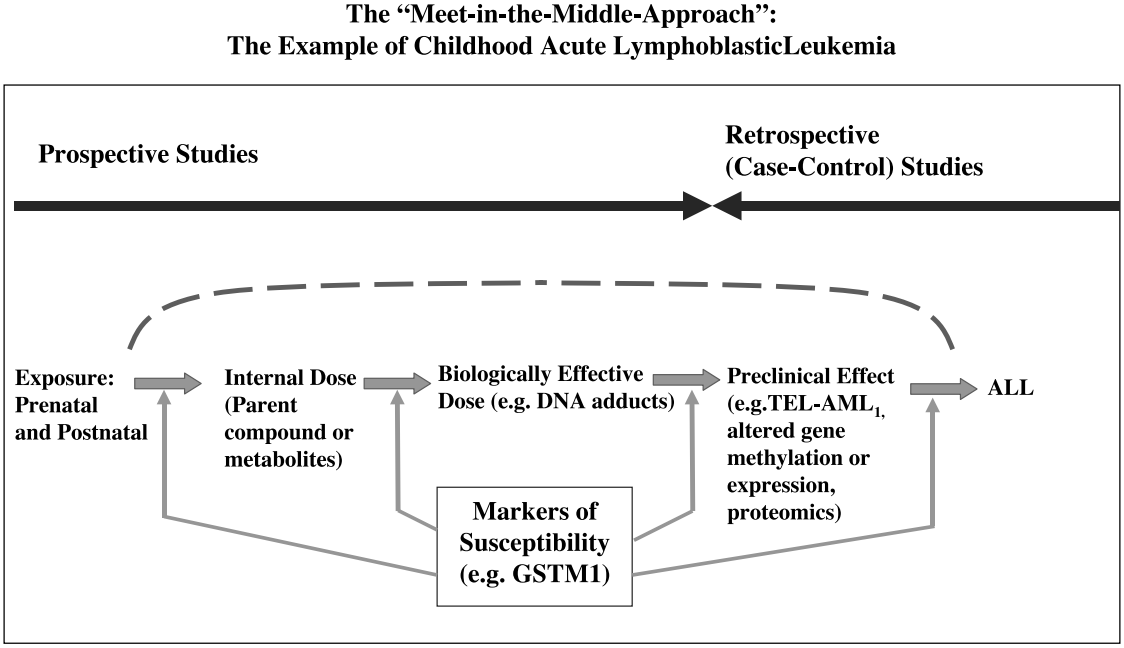

Figure 3. In this theoretical example, a biomarker of preclinical effect (TEL-AML1), DNA methylation, and gene or protein expression is linked to a specific environmental exposure within a prospective cohort study. The same biomarker is found to be present in a subgroup of acute lymphocytic leukemia (*ALL*) cases in a retrospective (case-control) study. The association between the identified exposure and biomarker is then retrospectively evaluated within the case-control study using available questionnaire, biomarker, or other data. If confirmed as significant risk factor, the exposure can be targeted for interventions.

We propose a related approach that emphasizes the complementarity between different types of markers and study designs: the use of complementary studies to validate new biomarkers such as gene or protein expression in the absence of large-scale prospective cohort studies or where cancer is rare. An example is childhood leukemia. Although few risk factors have been confirmed (78), several prospective molecular epidemiologic studies are suggesting an etiologic role for *in utero* and childhood exposure to environmental pollutants (79-81). Such prospective studies have not only obtained monitoring and biomarker data on exposure but also measured biomarkers, such as the internal dose of the toxicant, carcinogen-DNA adducts, gene mutations, chromosomal aberrations, and/or gene methylation or expression. Their small sample size (<1,000 subjects in each study) does not permit assessment of childhood acute lymphocytic leukemia as an outcome. However, other retrospective studies have analyzed biomarkers in specimens from childhood cancer cases and controls and have also gathered questionnaire and other data on environmental exposures. These include the United Kingdom Children's Cancer Study group (82), Children's Oncology Group, and the Northern California Childhood Leukemia Study (83). Interesting findings include a signal mutation (TEL-AML1) found in 25% of childhood acute lymphocytic leukemia and found to be present in neonatal bloodspots of children who subsequently developed acute lymphocytic leukemia (84). The TEL-AML1 fusion is acquired prenatally and constitutes the ''first hit'' in childhood leukemia. In addition, Smith et al. (85) have reported striking differences in gene methylation patterns between different cytogenetic subgroups of childhood leukemia, suggesting that epigenetic events are important in development of some forms of the disease and that different etiologic agents may be involved. They and others are exploring the use of gene methylation or proteomics to subclassify childhood leukemia. An ongoing study within the Columbia Center for Children's Environmental Health cohort in New York City is assessing the associations between prenatal exposures (e.g., PAHs and pesticides) and TEL-AML1 gene fusion in cord blood samples.[7] If links are established between specific *in utero* exposures and this proleukemic biomarker, case-control studies could then retrospectively evaluate the associations between TEL-AML1 and the same exposures. A finding that preclinical biomarkers shown to be related to particular exposures in prospective studies are also elevated in certain subclasses of leukemia would strengthen causal links between specific exposures and disease, eventually allowing targeted efforts to reduce those exposures. Referring to the molecular epidemiology paradigm, as shown in Fig. 3, prospective studies working from left to right (from exposure to preclinical response) would directly complement retrospective studies backtracking from right to left (from clinical disease to preclinical response to exposure). This ''meet in the middle approach'' has potential to open new avenues for prevention by identifying the specific environmental factor(s) responsible.

[7] M. Orjuela, unpublished data.

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

Conclusions

In conclusion, the previous generation of biomarkers has contributed greatly to our understanding of risk and susceptibility related largely to genotoxic carcinogens. Consequently, interventions and policy changes have been mounted to reduce risk from several important environmental carcinogens. More recently, developed biomarkers have considerable potential in molecular epidemiology because they reflect another equally important mechanism of carcinogenicity: epigenetic alterations that affect the expression of genes and proteins. These biomarkers can be measured by high-throughput methods, allowing large-scale studies that are discovery-oriented and that can be followed by hypothesis-testing studies. Several large collaborative studies and consortia to coordinate research efforts and enhance statistical power by increasing sample size have already been formed to study specific biomarkers and cancers [see Boffetta et al. (86) for example]. However, there is an urgent need for systematic validation of epigenetic and omic markers using the principles established during the last decades with the earlier generation of biomarkers. Once validated, the newer biomarkers can be combined with the more traditional biomarkers in hypothesis-testing studies and interventions to reduce cancer risk.

Appendix A. Use of mechanistic data in the IARC Monographs

More than 900 agents have been evaluated since the inception of the Monograph Program in 91 volumes: 108 were in group 1, 64 in group 2A, and 240 in group 2B. Table 2 shows all the agents for which upgrading or downgrading occurred based on mechanistic evidence, including molecular epidemiologic or biomarker data. Most of the category changes involved agents (36 in number) for which the evidence was sufficient in animals but inadequate in humans. These would have been classified as 2B but mechanistic evidence justified upgrading them to 2A. Three agents were upgraded from 2A to 1. An example of the latter upgrading based on human mechanistic evidence is 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, included in group 1, in spite of limited evidence in humans. The rationale was the following: (*a*) 2,3,7,8-tetrachlorodibenzo-*p*-dioxin is a multisite carcinogen in experimental animals that has been shown by several lines of evidence to act through a mechanism involving the aryl hydrocarbon receptor, (*b*) this receptor is highly conserved in an evolutionary sense and functions the same way in humans as in experimental animals, and (*c*) tissue concentrations (internal dose of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin) are similar in heavily exposed human populations in which an increased overall cancer risk was observed and in rats exposed to carcinogenic dosage regimens in bioassays (87).

An example of upgrading from 2B to 2A is the chemotherapeutic agent cisplatin, which was done based on observed genotoxicity in treated patients, including sister chromatid exchanges in peripheral blood lymphocytes, antigenicity against cisplatin-DNA adducts, chromosomal aberrations, and micronuclei (88). Similar effects were observed *in vitro* and in experimental animals. Eight agents were downgraded from 2B to 3. However, none of these evaluations involved human biomarker data.

References

1. Perera FP, Weinstein IB. Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. J Chronic Dis 1982;35:581–600.
2. National Research Council. Biological markers in environmental health research. Environ Health Perspect 1987;74:3–9.
3. Schulte PA, Rothman N, Schottenfeld D. Design considerations in molecular epidemiology. In: Schulte PA and Perera FP, editors. Molecular epidemiology: principles and practices. San Diego (CA): Academic Press, Inc.; 1993. p. 159–98.
4. Harris CC. p53 tumor suppressor gene: at the crossroads of molecular carcinogenesis, molecular epidemiology, and cancer risk assessment. Environ Health Perspect 1996;104:435–9.
5. National Research Council. Regulating pesticides in food: the Delaney paradox. Washington (DC): National Academy of Sciences, National Academy Press; 1987.
6. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004;10:789–99.
7. Hemminki K, Dickey CP, Karlsson S, et al. Aromatic-DNA adducts in foundry workers in relation to exposure, lifestyle and CYP1A1 and glutathione transferase M1 genotype. Carcinogenesis 1997;18:345–50.
8. Hagmar L, Wirfalt E, Paulsson B, Tornqvist M. Differences in hemoglobin adduct levels of acrylamide in the general population with respect to dietary intake, smoking habits and gender. Mutat Res 2005;580:157–65.
9. Vodicka P, Koskinen M, Stetina R, et al. The role of various biomarkers in the evaluation of styrene genotoxicity. Cancer Detect Prev 2003;27:275–84.
10. Perera FP. Molecular epidemiology: on the path to prevention? J Natl Cancer Inst 2000;92:602–12.
11. Tang D, Phillips DH, Stampfer M, et al. Association between carcinogen-DNA adducts in white blood cells and lung cancer risk in the physicians health study. Can Res 2001;61:6708–12.
12. Peluso M, Munnia A, Hoek G, et al. DNA adducts and lung cancer risk: a prospective study. Cancer Res 2005;65:8042–8.
13. Veglia F, Matullo G, Vineis P. Bulky DNA adducts and risk of cancer: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2003;12:157–60.
14. Bonassi S, Znaor A, Norppa H, Hagmar L. Chromosomal aberrations and risk of cancer in humans: an epidemiologic perspective. Cytogenet Genome Res 2004;104:376–82.
15. Vineis P. Individual susceptibility to carcinogens. Oncogene 2004;23:6477–83.
16. Suuriniemi M, Agalliu I, Schaid DJ, et al. Confirmation of a positive association between prostate cancer risk and a locus at chromosome 8q24. Cancer Epidemiol Biomarkers Prev 2007;16:809–14.
17. Zhang ZF, Cordon-Cardo C, Rothman N, Freedman AN, Taylor JA. Methodological issues in the use of tumour markers in cancer epidemiology. IARC Scientific Publications No. 142. Lyon: International Agency for Research on Cancer; 1997. p. 201–13.
18. Hagmar L, Bonassi S, Stromberg U, et al. Chromosomal aberrations in lymphocytes predict human cancer: a report from the European Study Group on Cytogenetic Biomarkers and Health (EDCH). Cancer Res 1998;58:4117–21.
19. Zhang L, Yang W, Hubbard AE, Smith MT. Nonrandom aneuploidy of chromosomes 1, 5, 6, 7, 8, 9, 11, 12, and 21 induced by the benzene metabolites hydroquinone and benzenetriol. Environ Mol Mutagen 2005;45:388–96.
20. Vermeulen R, Lan Q, Zhang L, et al. Decreased levels of CXC-chemokines in serum of benzene-exposed workers identified by array-based proteomics. Proc Natl Acad Sci U S A 2005;102:17041–6.
21. Lan Q, Zhang L, Shen M, et al. Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. Cancer Res 2005;65:9574–81.
22. International Agency for Research on Cancer. Polynuclear aromatic compounds. Part 1. Chemical, environmental, and experimental data. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Lyon (France): IARC; 1983. p. 1–453
23. Bostrom CE, Gerde P, Hanberg A, et al. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. Environ Health Perspect 2002;110:451–88.

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

24. Anderson LM, Jones AB, Miller MS, et al. Metabolism of transplacental carcinogens. Perinatal and multigeneration carcinogenesis. Lyon (France): IARC; 1989.
25. Miller EC, Miller JA. Mechanisms of chemical carcinogenesis. Cancer 1981;47:1055–64.
26. Perera F. Environment and cancer: who are susceptible? Science 1997;278:1068–73.
27. Perera FP, Poirier MC, Yuspa SH, et al. A pilot project in molecular cancer epidemiology: determination of benzo[*a*]pyrene adducts in animal and human tissues by immunoassay. Carcinogenesis 1982;3:1405–10.
28. Kriek E, van Schooten FJ, Hillebrand MJX, et al. DNA adducts as a measure of lung cancer risk in human exposed to polycyclic aromatic hydrocarbons. Environ Health Perspect 1993;99:71–5.
29. Tang D, Santella RM, Blackwood A, et al. A molecular epidemiological case-control study of lung cancer. Cancer Epidemiol Biomarkers Prev 1995;4:341–6.
30. Bartsch H, Rojas M, Alexandrov K, Risch A. Impact of adduct determination on the assessment of cancer susceptibility. Recent Results Cancer Res 1998;154:86–96.
31. Wiencke JK, Thurston SW, Kelsey KT, et al. Early age at smoking initiation and tobacco carcinogen DNA damage in the lung. J Natl Cancer Inst 1999;91:614–9.
32. Greenblatt MS, Bennett WP, Hollstein M, Harris CC. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res 1994;55:4855–78.
33. Denissenko M, Pao A, Tang MS, Pfiefer G. Preferential formation of benzo[*a*]pyrene adducts at lung cancer mutational hotspots in p53. Science 1996;274:430–2.
34. Potter JD. Epidemiology, cancer genetics and microarrays: making correct inferences, using appropriate designs. Trends Genet 2003;19:690–5.
35. Potter JD. At the interfaces of epidemiology, genetics and genomics. Nat Rev Genet 2001;2:142–7.
36. Kim YI. Folate and DNA methylation: a mechanistic link between folate deficiency and colorectal cancer? Cancer Epidemiol Biomarkers Prev 2004;13:511–9.
37. Mooney LA, Santella RM, Covey L, et al. Decline on DNA damage and other biomarkers in peripheral blood following smoking cessation. Cancer Epidemiol Biomarkers Prev 1995;4:627–34.
38. Evans MD, Dizdaroglu M, Cooke MS. Oxidative DNA damage and disease: induction, repair and significance. Mutat Res 2004;567:1–61.
39. Moller P, Loft S. Oxidative DNA damage in human white blood cells in dietary antioxidant intervention studies. Am J Clin Nutr 2002;76:303–10.
40. Moller P, Loft S. Dietary antioxidants and beneficial effect on oxidatively damaged DNA. Free Radic Biol Med 2006;41:388–415.
41. Hecht SS. Tobacco and cancer: approaches using carcinogen biomarkers and chemoprevention. Ann N Y Acad Sci 1997;833:91–111.
42. Levi MS, Borne RF, Williamson JS. A review of cancer chemopreventive agents. Curr Med Chem 2001;8:1349–62.
43. Crowell JA. The chemopreventive agent development research program in the Division of Cancer Prevention of the US National Cancer Institute: an overview. Eur J Cancer 2005;41:1889–910.
44. Kelley MJ, Glaser EM, Herndon JE 2nd, et al. Safety and efficacy of weekly oral oltipraz in chronic smokers. Cancer Epidemiol Biomarkers Prev 2005;14:892–9.
45. Waters MD, Fostel JM. Toxicogenomics and systems toxicology: aims and prospects. Nat Rev Genet 2004;5:936–48.
46. Webb PM, Merritt MA, Boyle GM, Green AC. Microarrays and epidemiology: not the beginning of the end but the end of the beginning. Cancer Epidemiol Biomarkers Prev 2007;16:637–8.
47. Garcia-Closas M, Vermeulen R, Sherman ME, Moore LE, Smith MT, Rothman N. Application of biomarkers in cancer epidemiology. In: Schottenfeld D, Fraumeni JF Jr, editors. Cancer epidemiology and prevention. New York: Oxford University Press; 2006.
48. Robertson KD, Wolffe AP. DNA methylation in health and disease. Nat Rev Genet 2000;1:11–9.
49. Alberg AJ, Brock MV, Samet JM. Epidemiology of lung cancer: looking to the future. J Clin Oncol 2005;23:3175–85.
50. Brena RM, Huang TH, Plass C. Quantitative assessment of DNA methylation: potential applications for disease diagnosis, classification, and prognosis in clinical settings. J Mol Med 2006;84:365–77.
51. Toyooka S, Tokumo M, Shigematsu H, et al. Mutational and epigenetic evidence for independent pathways for lung adenocarcinomas arising in smokers and never smokers. Cancer Res 2006;66:1371–5.
52. Belinsky SA, Klinge DM, Dekker JD, et al. Gene promoter methylation in plasma and sputum increases with lung cancer risk. Clin Cancer Res 2005;11:6505–11.
53. Bearzatto A, Conte D, Frattini M, et al. p16(INK4A) hypermethylation detected by fluorescent methylation-specific PCR in plasmas from non-small cell lung cancer. Clin Cancer Res 2002;8:3782–7.
54. Russo AL, Thiagalingam A, Pan H, et al. Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer. Clin Cancer Res 2005;11:2466–70.
55. Vineis P, Husgafvel-Pursiainen K. Air pollution and cancer: biomarker studies in human populations. Carcinogenesis 2005;26:1846–55.
56. Honorio S, Agathanggelou A, Schuermann M, et al. Detection of RASSF1A aberrant promoter hypermethylation in sputum from chronic smokers and ductal carcinoma *in situ* from breast cancer patients. Oncogene 2003;22:147–50.
57. Shen L, Issa JP. Epigenetics in colorectal cancer. Curr Opin Gastroenterol 2002;18:68–73.
58. Lindon JC, Holmes E, Bollard ME, Stanley EG, Nicholson JK. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. Biomarkers 2004;9:1–31.
59. Odunsi K, Wollman RM, Ambrosone CB, et al. Detection of epithelial ovarian cancer using ^{1}H-NMR-based metabonomics. Int J Cancer 2005;113:782–8.
60. Yang J, Xu G, Zheng Y, et al. Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. J Chromatogr B Analyt Technol Biomed Life Sci 2004;813:59–65.
61. Dumas ME, Maibaum EC, Teague C, et al. Assessment of analytical reproducibility of ^{1}H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. Anal Chem 2006;78:2199–208.
62. Kikuchi T, Carbone DP. Proteomics analysis in lung cancer: challenges and opportunities. Respirology 2007;12:22–8.
63. Conrads TP, Zhou M, Petricoin EF III, Liotta L, Veenstra TD. Cancer diagnosis using proteomic patterns. Expert Rev Mol Diagn 2003;3:411–20, Review.
64. Conrads TP, Fusaro VA, Ross S, et al. High-resolution serum proteomic features for ovarian cancer detection. Endocr Relat Cancer 2004;11:163–78.
65. Oh JM, Brichory F, Puravs E, et al. A database of protein expression in lung cancer. Proteomics 2001;1:1303–19.
66. Zhang Z, Bast RC, Jr., Yu Y, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004;64:5882–90.
67. Granville CA, Dennis PA. An overview of lung cancer genomics and proteomics. Am J Respir Cell Mol Biol 2005;32:169–76.
68. Esteller M. The necessity of a human epigenome project. Carcinogenesis 2006;27:1121–5.
69. Chen G, Gharib TG, Thomas DG, et al. Proteomic analysis of eIF-5A in lung adenocarcinomas. Proteomics 2003;3:496–504.
70. Yanagisawa K, Shyr Y, Xu BJ, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet 2003;362:433–9.
71. Dalle-Donne I, Scaloni A, Giustarini D, et al. Proteins as biomarkers of oxidative/nitrosative stress in diseases: the contribution of redox proteomics. Mass Spectrom Rev 2005;24:55–99.
72. Diamandis EP. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst 2004;96:353–6.
73. Gruenewald TL, Seeman TE, Ryff CD, Karlamangla AS, Singer BH. Combinations of biomarkers predictive of later life mortality. Proc Natl Acad Sci U S A 2006;103:14158–63.
74. Ambrosone CB, Rebbeck TR, Morgan GJ, et al. New developments in the epidemiology of cancer prognosis: traditional and molecular predictors of treatment response and survival. Cancer Epidemiol Biomarkers Prev 2006;15:2042–6.
75. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer 2005;5:142–9.
76. Rundle AG, Vineis P, Ahsan H. Design options for molecular epidemiology research within cohort studies. Cancer Epidemiol Biomarkers Prev 2005;14:1899–907.
77. Sharp L, Little J. Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. Am J Epidemiol 2004;159:423–43.
78. Buffler PA, Kwan ML, Reynolds P, Urayama KY. Environmental and genetic risk factors for childhood leukemia: appraising the evidence. Cancer Invest 2005;23:60–75.
79. Perera F, Tang D, Whyatt R, Lederman SA, Jedrychowski W. DNA damage from polycyclic aromatic hydrocarbons measured by benzo[*a*]pyrene-DNA adducts in mothers and newborns from Northern Manhattan, the World Trade Center Area, Poland, and China. Cancer Epidemiol Biomarkers Prev 2005;14:709–14.
80. Bocskay KA, Tang D, Orjuela MA, Xinhua L, Warburton DP, Perera FP.

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

Chromosomal aberrations in cord blood are associated with prenatal exposure to carcinogenic polycyclic aromatic hydrocarbons. Cancer Epidemiol Biomarker Prev 2005;14:506–11.
81. Eskenazi B, Gladstone EA, Berkowitz GS, et al. Methodologic and logistic issues in conducting longitudinal birth cohort studies: lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. Environ Health Perspect 2005;113:1419–29.
82. Ablett S, Pinkerton CR. Recruiting children into cancer trials—role of the United Kingdom Children's Cancer Study Group (UKCCSG). Br J Cancer 2003;88:1661–5.
83. Jensen CD, Block G, Buffler P, Ma X, Selvin S, Month S. Maternal dietary risk factors in childhood acute lymphoblastic leukemia (United States). Cancer Causes Control 2004;15:559–70.
84. Wiemels JL, Greaves M. Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. Cancer Res 1999;59:4075–82.
85. Smith MT, McHale CM, Wiemels JL, et al. Molecular biomarkers for the study of childhood leukemia. Toxicol Appl Pharmacol 2005;206:237–45.
86. Boffetta P, Armstrong B, Linet M, Kasten C, Cozen W, Hartge P. Consortia in cancer epidemiology: lessons from InterLymph. Cancer Epidemiol Biomarkers Prev 2007;16:197–9.
87. IARC. Polychlorinated dibenzo-para-dioxins and polychlorinated dibenzofurans. IARC monographs on the evaluation of carcinogenic risks to humans, vol. 69. Lyon (France): International Agency for Research on Cancer; 1997.
88. IARC. Some antiviral and antineoplastic drugs, and other pharmaceutical agents. IARC monograph on the carcinogenic risks to humans, vol. 76. Lyon (France): International Agency for Research on Cancer; 2000.
89. Hecht SS, Carmella SG, Murphy SE, Akerkar S, Brunnemann KD, Hoffmann D. A tobacco-specific lung carcinogen in the urine of men exposed to cigarette smoke. N Engl J Med 1993;329:1543–6.
90. Groopman JD, Kensler TW, Links JM. Molecular epidemiology and human risk monitoring. Toxicol Lett 1995;82:763–9.
91. Santella RM, Zhang YJ, Chen CJ, et al. Immunohistochemical detection of aflatoxin B1-DNA adducts and hepatitis B virus antigens in hepatocellular carcinoma and nontumorous liver tissue. Environ Health Perspect 1993;99:199–202.
92. Albertini RJ, Sram RJ, Vacek PM, et al. Biomarkers in Czech workers exposed to 1,3-butadiene: a transitional epidemiologic study. Res Rep Health Eff Inst 2003;116:1–141; discussion 143–62.
93. Holecková B, Piesová E, Sivikova K, Dianovský J. Chromosomal aberrations in humans induced by benzene. Ann Agric Environ Med 2004;11:175–9.
94. Perera FP, Mooney LA, Stampfer M, et al. Associations between carcinogen-DNA damage, glutathione *S*-transferase genotypes, and risk of lung cancer in the prospective Physicians' Health Cohort Study. Carcinogenesis 2002;23:1641–6.
95. Ammenheuser MM, Bechtold WE, Abdel-Rahman SZ, Rosenblatt JI, Hastings-Smith DA, Ward JB, Jr. Assessment of 1,3-butadiene exposure in polymer production workers using HPRT mutations in lymphocytes as a biomarker. Environ Health Perspect 2001;109:1249–55.
96. Lee KH, Lee J, Ha M, et al. Influence of polymorphism of GSTM1 gene on association between glycophorin a mutant frequency and urinary PAH metabolites in incineration workers. J Toxicol Environ Health A 2002;65:355–63.
97. Gwosdz C, Balz V, Scheckenbach K, Bier H. p53, p63 and p73 expression in squamous cell carcinomas of the head and neck and their response to cisplatin exposure. In: Bier H, editor. Adv Otorhinolaryngol 2005;62:58–71.
98. Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J Natl Cancer Inst 2000;93:874–97.
99. Cheng L, Eicher SA, Guo Z, Hong WK, Spitz MR, Wei Q. Reduced DNA repair capacity in head and neck cancer patients. Cancer Epidemiol Biomarkers Prev 1998;7:465–8.
100. Garcia-Closas M, Malats N, Silverman D, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet 2005;366:649–59.
101. Vineis P, Veglia F, Benhamou S, et al. CYP1A1 T3801 C polymorphism and lung cancer: A pooled analysis of 2,451 cases and 3,358 controls. Int J Cancer 2003;104:650–57.
102. Stephan C, Stroebel G, Heinau M, et al. The ratio of prostate-specific antigen (PSA) to prostate volume (PSA density) as a parameter to improve the detection of prostate carcinoma in PSA values in the range of <4 ng/mL. Cancer 2005;104:993–1003.

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research

DOI:10.1158/1055-9965.EPI-07-0457

Correction

In an article (1) in the October 2007 issue, there was an error in the references. Reference 2 should replace reference 5 in every in-text citation.

2. National Research Council. Biological markers in environmental health research. Environ Health Perspect 1987;74:3–9.

Should replace.

5. National Research Council. Regulating pesticides in food: the Delaney paradox. Washington (DC): National Academy of Sciences, National Academy Press; 1987.

Reference

1. Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: The new in light of the old. Cancer Epidemiol Biomarkers Prev 2007;16:1954–65.

Copyright © 2007 American Association for Cancer Research.
doi:10.1158/1055-9965.EPI-16-12-COR

Downloaded from cebp.aacrjournals.org on November 2, 2011
Copyright © 2007 American Association for Cancer Research