CHAPTER 19.

Cancer

Frederica P. Perera and Paolo Vineis

Summary

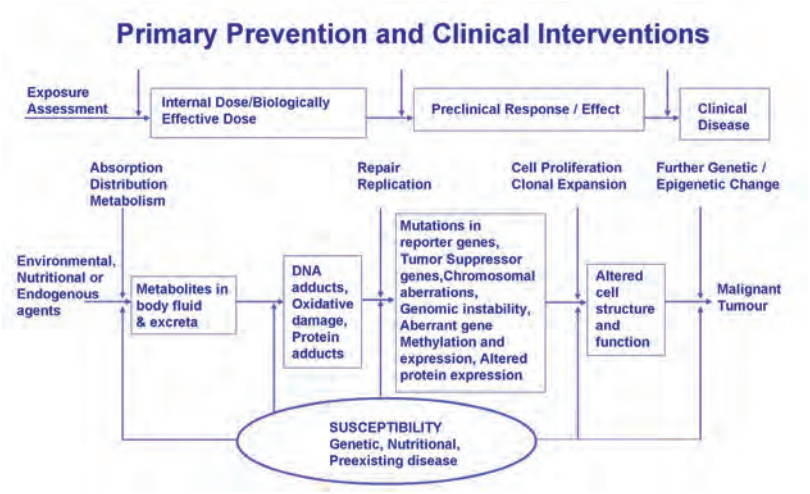
Molecular epidemiology was introduced in the study of cancer in the early 1980s, with the expectation that it would help overcome some important limitations of epidemiology and facilitate cancer prevention. The first generation of biomarkers has indeed contributed to our understanding of mechanisms, risk and susceptibility as they relate largely to genotoxic carcinogens, resulting in interventions and policy changes to reduce risk from several important environmental carcinogens. New and promising biomarkers are now becoming available for epidemiological studies, including alterations in gene methylation and gene expression, proteomics and metabolomics. However, most of these newer biomarkers have not been adequately validated, and their role in the causal paradigm is not clear. Systematic validation of these newer biomarkers is urgently needed and can take advantage of the principles and criteria established over the past several decades from experience with the first generation of biomarkers.

Prevention of only 20% of cancers in the United States alone would result in 300 000 fewer new cases annually, avoidance of incalculable suffering, and a savings in direct financial costs of over US$20 billion each year (1). Molecular epidemiology can play a valuable role in achieving this goal.

Introduction

In 1982, “molecular cancer epidemiology” was proposed 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. Four categories of biomarkers were described: internal dose, biologically effective dose, early response/effect and susceptibility (2). In 1987, the United States National Academy of Sciences (NAS) convened a workshop on the use of biomarkers in environmental health research that adopted this concept and expanded it to include a fifth category: altered structure and function. Figure 19.1 summarizes the general paradigm proposed in 1982 and expanded in 1987 (3). The fundamental concept

Figure 19.1. Updated model for molecular epidemiology (figure compiled from (2,3,177))



of a continuum of molecular/genetic alterations leading to cancer that can be accessed using biomarkers remains valid.

Most of the focus thus far has been on biomarkers of genotoxicity. The field is now expanding rapidly to include high-throughput methods to detect alterations in the expression of genes, rather than structural changes. In this chapter, examples are provided of the accomplishments in molecular cancer epidemiology: studies that have provided evidence of causality and mechanisms, documented environment–susceptibility interactions and identified at-risk populations. The promise and challenge of new "omic" and epigenetic biomarkers (4–9), including their translational potential and need for validation, are then discussed. A discussion follows of the strengths, limitations and lessons learned from molecular epidemiologic research to date, and future directions for this field. Rather than an encyclopaedic review, presented are several paradigmatic examples of each area. Among promising biomarkers and technologies not included here are those related to inflammation and obesity (10,11), genome-wide scans (11), and tumour markers (12).

Context and public health significance

The context of this chapter on molecular cancer epidemiology is the need to prevent cancer, a disease that in the United States alone claims over half a million lives annually, with more than 1.5 million new cases diagnosed each year and attendant direct annual costs of US$107 billion (1). Many lines of evidence indicate, even more clearly than in 1982, that the great majority of cancers are in principle preventable, because the factors that determine their incidence are largely exogenous or environmental (5–8). These include exposures related to lifestyle (diet and smoking), occupation, and pollutants in the air, water and food supply. Genetic factors are largely important in terms of influencing individual susceptibility to carcinogens; only in some rare forms of human cancer do hereditary genetic factors play a decisive role. This awareness has lent greater urgency to the search for more powerful early-warning systems to identify causal environmental agents and flag risks well before the malignant process is entrenched.

Contributions of molecular epidemiology

The following sections refer to studies that have employed various study designs, the strengths and limitations of which are discussed in Chapters 14–18.

Providing evidence on causality and mechanisms: Examples

Polycyclic aromatic hydrocarbons/tobacco smoke and lung cancer

Most of the molecular epidemiologic research on lung cancer has targeted tobacco smoke as a model carcinogen. Polycyclic aromatic hydrocarbons (PAHs) such as benzo[a]pyrene (B[a]P) are one of 55 known carcinogens in tobacco smoke, are among the most studied, and often serve as a representative tobacco smoke carcinogen (13,14). Other tobacco carcinogens include 4-aminobiphenyl (4-ABP) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) (14–16). PAHs are also found in outdoor air from fossil fuel combustion via automobile exhaust, emissions from coal-fired power plants, and other industrial sources; in indoor air from tobacco smoking, cooking and heating; and in the diet from consumption of smoked or grilled food (17,18). By several routes of exposure in adult animals, PAHs cause tumours including lung, liver and skin tumours (19) (see also (20) for review). PAHs are also transplacental carcinogens experimentally (21,22). PAHs such as B[a]P form adducts with DNA, a mechanism considered to be a

critical early event in PAH-induced tumorigenesis, since adducts can lead to mutations and ultimately to cancer. As biomarkers, carcinogen-DNA adducts have the advantage of reflecting chemical-specific genetic damage that is mechanistically relevant to carcinogenesis (23,24).

In 1982, PAH-DNA adducts were detected in human subjects *in vivo*, specifically in white blood cells (WBCs) and lung tissue from lung cancer patients, most of whom were smokers (25). Using more sensitive laboratory methods to measure adducts, subsequent studies in healthy exposed populations (i.e. active smokers, coke-oven and foundry workers, and residents of Poland, the Czech Republic and China who were exposed to air pollution from coal burning) have found increased concentrations of PAH-DNA adduct levels in blood and other tissues compared to unexposed individuals, with no apparent threshold for DNA binding (26–29). These findings are consistent with traditional epidemiologic data showing elevated risk of lung cancer in PAH-exposed populations (see (20) for a review). Substantial interindividual variability has been observed in adduct levels among persons with similar exposure; about 30- to 70-fold for adducts in WBCs (29,30).

Although not all studies have been positive, since 1982 considerable evidence has mounted that PAH-DNA adducts in WBCs or lung tissue are risk markers for lung cancer (31–33). In one case–control study, higher PAH-DNA adduct levels were found in WBCs from 119 case subjects (compared with 98 control subjects), after adjusting for smoking, dietary PAH exposure and other potential confounders (32).

Caution is necessary in interpreting results from studies of DNA adduct levels and cancer risk. As discussed in Chapter 14, by their retrospective nature, case–control studies alone are unable to definitively establish causality. In addition, because the carcinogenic impact of adducts depends on the tissue and genes affected, one cannot assume *a priori* that adduct levels measured in blood are a valid surrogate for those in target tissue (34). The relationship between adduct concentration in blood and target tissue must be established for individual carcinogens. With respect to PAH-DNA, an experimental study (35) has shown ubiquitous binding of B[a]P metabolites to DNA and protein. Two other studies have found significant correlations between DNA adducts in WBCs and lung tissue from the same case subjects (35,36).

More recently, in a case–control study nested within the prospective Physicians' Health Study of over 14 000 men, it was evaluated whether DNA damage in blood samples collected at enrolment significantly predicted risk, consistent with the hypothesis that cases have greater biological susceptibility to PAHs and other aromatic tobacco carcinogens (37). The subjects in this nested case–control study were 89 cases of primary lung cancer and 173 controls, matched on smoking, age and duration of follow-up. Aromatic DNA adducts were measured in WBCs by the nuclease P1-enhanced ^{32}P-postlabelling method that primarily detects smoking-related adducts. Healthy current smokers who had elevated levels of aromatic DNA adducts in WBCs were approximately three times more likely to be diagnosed with lung cancer 1–13 years later than were current smokers with lower adduct concentrations (odds ratio (OR) = 2.98; 95% CI = 1.05–8.42; $P = 0.04$). The same relationship was not seen among former smokers and never smokers. The findings suggested that individuals who become cases have greater biological susceptibility to tobacco carcinogens, a biological difference that seems to manifest most clearly while exposure is still ongoing.

A second nested case–control study on lung cancer (newly diagnosed after recruitment) within the European Prospective Study Into Cancer and Nutrition (EPIC) cohort measured aromatic PAH-DNA adducts as markers of the biologically effective dose of PAHs, and mutations in the *ras* and *p53* genes in plasma DNA as markers of early preclinical effects. Cases included subjects with newly diagnosed lung cancer (n = 115) accrued after a median follow-up of seven years among the EPIC former smokers and never smokers. Unlike the prior nested case–control study, no current smokers were included. Adducts were associated with the subsequent risk of lung cancer among never smokers (OR = 4.04; 95% CI = 1.06–15.42) and among the younger age groups.

A meta-analysis of aromatic PAH-DNA adducts and lung cancer (38) concluded that current smokers with high levels of adducts have an increased risk of lung cancer, supporting a causal role of aromatic compounds in the etiology of lung cancer. While unmeasured variability in smoking, diet, or indoor/outdoor PAH concentrations may partially explain the finding of higher adduct levels in individuals with lung cancer, the results are also consistent with other evidence that some individuals are predisposed to genetic damage from PAHs and thereby to lung cancer (33,39,40). Taken together, the results of many studies support the theory that cumulative damage resulting from genotoxic chemicals that bind to DNA is a major cause of cancer (40).

Supporting molecular evidence that PAHs play an important role in lung cancer comes from observations that the *p53* tumour suppressor gene is mutated in 40–50% of lung tumours, and that the pattern of mutations in those tumours is consistent with the types of DNA adducts and mutations induced experimentally by B[a]P (41,42). Smokers with lung cancer show a pattern of mutations in *p53* that is different (with some exceptions) from that of non-smokers (43). Moreover, as discussed above, certain single nucleotide polymorphisms (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.

In addition to genetic damage and gene mutations, epigenetic mechanisms are now emerging as important in lung cancer related to tobacco smoking (discussed in a later section).

In summary, studies using biomarkers of biologically effective dose, early preclinical effect/response, and individual susceptibility (SNPs) have been valuable in elucidating the steps that link tobacco smoke/PAH exposure to the onset of lung cancer.

AFB_1, HBV and liver cancer

During the past 30 years, research in experimental animals and humans has confirmed that the foodborne mutagen aflatoxin B1 (AFB_1) is a human hepatocarcinogen acting synergistically with the hepatitis B virus (HBV). AFB_1 is a fungal metabolite present in grains and cereals due to improper storage (44). Research has indicated that several biomarkers of the internal or biologically effective dose of AFB_1 (AFB_1 metabolites, AFB_1-albumin adducts, and AFB_1-N^7-guanine adducts in urine) and HBV surface antigen seropositivity are risk markers for liver cancer on a population level. In 1992, a prospective study in Shanghai, China found that among 18 244 men there were 22 incident cases of liver cancer (45). Analysis of urine samples banked 1–4 years before diagnosis from the case subjects and matched control subjects gave relative risks (RRs) of 2.4 (95% CI = 1.0–5.9) for any of the AFB_1 metabolites, and 4.9 (95% CI = 1.5–16.3) for detectable AFB_1-N^7-guanine adducts. There was a strong interaction between the serologic marker of HBV infection and the AFB_1 markers. Among individuals with chronic hepatitis infection who were also aflatoxin-positive, the RR was 60 (95% CI = 6.4–561.8). A subsequent follow-up study of 55 hepatocellular carcinoma (HCC) case subjects and 267 control subjects from the same cohort showed that the presence of any urinary AFB_1 biomarker significantly predicted liver cancer (RR = 5.0; 95% CI = 2.1–11.8) with an RR of 9.1 (95% CI = 2.9–29.2) for the presence of AFB_1-N^7-guanine adducts. A synergistic interaction between the presence of urinary AFB_1 biomarkers and HBV seropositivity resulted in a 59-fold (95% CI = 16.6–212.0) elevation in HCC risk (46). The implication for prevention is that both reduction in dietary levels of AFB_1 and wide-scale HBV vaccination are needed, since the benefits of the latter will not be manifest for many years (45). These biomarkers have subsequently been used as outcome measures in an intervention trial with the antischistosomal drug oltipraz (see further discussion below).

In Taiwan, China, subsequent studies of incident HCC case subjects and matched controls whose levels of AFB_1 metabolites, AFB_1-albumin and AFB_1-DNA adducts were measured in stored urine samples gave results consistent with the prior results from the PRC prospective study (47). In HBV-infected men with detectable AFB_1-albumin and AFB_1-DNA adduct levels, the risk of HCC was increased by 10-fold (RR = 10.0; 95% CI = 1.6–60.9) (48).

Other molecular data on the causal and mechanistic role of AFB_1 involve the *p53* gene. Early studies suggested a characteristic mutation spectrum in the human *p53* gene in HCC in South Africa and China, where it was observed that about 50% of the patients had a relatively rare mutation, a G to T transversion at codon 249 (49). This mutation was not previously observed in patients living in areas where food contamination by aflatoxins is not common; furthermore, the same mutation could be induced experimentally by AFB_1 *in vitro.* More recently, however, cells were incubated with AFB_1 and the types of DNA adducts induced in *p53* were studied (42). It was observed that adducts were mainly in sites different from codon 249, the one that the 'fingerprint' theory based on human data had implicated. In addition, the expected adducts in codon 249 were rapidly repaired (50% in seven hours). Therefore, the argument that aflatoxin exerts its carcinogenic activity by leaving a signature in a specific codon, and via a specific mechanism in *p53*, was considerably weakened. The apparent association of *p53*-specific mutations with aflatoxin now appears to be due to the selective advantage of mutated cells after exposure to HBV, rather than a causal event in the pathogenic process.

This example illustrates problems encountered in the use of human cancer gene fingerprints as

definitive links between an exposure and a specific cancer. These difficulties include:

- the multifactorial nature of human cancers that hampers their attribution to single carcinogenic agents and/or the identification of a pathogenetic pathway common to several cancers;
- the high genetic instability of cancer cells that may increase the frequency of mutations in certain cancer genes regardless of exposure factors;
- the importance of DNA repair mechanisms and of the corresponding degree of population variation;
- tissue selection bias that may affect the results, although its extent is difficult to establish;
- the simultaneous presence of clinical (e.g. treatment) and biological factors (e.g. stage, grading) related to the exposure and to the frequency of mutations that may confound its association;
- the need for consideration of temporal sequences in the activation/deactivation of cancer genes;
- the fact that several different carcinogens may induce the same *p53* mutation, and attribution to one of those carcinogens requires careful consideration of all relevant exposures.

For these reasons, the original hypothesis, that cancer fingerprints could be identified and used to recognize exposure-related tumours, has not been fully confirmed.

In summary, with this caveat in mind, studies using biomarkers of biologically effective dose, early preclinical response/effect and individual susceptibility (SNPs) have been valuable in elucidating the steps that link AFB_1 and HBV exposure to the risk of liver cancer.

Benzene and leukaemia

Benzene exposure occurs in the workplace and in the ambient environment largely because it is a component of gasoline (50). Another major source of public exposure to benzene is cigarette smoking. The example of benzene and haematological malignancies is paradigmatic, as it involves a single type of malignancy and combines biomarkers of several different classes that belong to the carcinogenic pathway shown in Figure 19.1. The various exposure markers include unmetabolized benzene in urine (UBz) and all major urinary metabolites (phenol (PH), *E,E*-muconic acid (MA), hydroquinone (HQ), and catechol (CA)), as well as the minor metabolite, S-phenylmercapturic acid (SPMA), all of which have been investigated among Chinese workers exposed to benzene (51). However, the most interesting results have come from investigations on early response/effect markers, specifically chromosomal aberrations.

Classical studies have shown that prospective data on chromosome aberrations are able to predict the onset of haematological malignancies. Combined analyses of data from Nordic and Italian prospective cohort studies, involving 3541 subjects, found that chromosomal aberrations were significant predictors of cancer (52). In the Nordic cohort, among subjects with high frequencies of chromosomal aberrations, the OR for all cancer deaths was 2.35 (95% CI = 1.31–4.23), compared with 2.66 (95% CI = 1.26–5.62) in the Italian cohort (53). In the Italian cohort, cancer predictivity of high chromosomal aberrations was greater for haematologic malignancies, with a standardized mortality ratio (SMR) of 5.49 (95% CI = 1.49–140.5) (54).

Specific chromosomal aberrations have been observed in both preleukemia and leukaemia patients exposed to benzene, as well as in otherwise healthy benzene-exposed workers (55). By use of fluorescent *in situ* hybridization (FISH) and the polymerase chain reaction (PCR), it was 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 mass spectrometry (SELDI-TOF MS). SELDI-TOF analysis of exposed and unexposed subjects revealed that lowered expression of PF4 and CTAP-III proteins is a potential biomarker of benzene's early biologic effects and may play a role in the immunosuppressive effects of benzene (56).

Finally, 20 candidate susceptibility genes were investigated in the same Chinese cohort (57). 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 finding provides evidence that SNPs in genes that regulate haematopoiesis modify benzene-induced haematotoxicity. However, as is clarified later, much research on genetic variants and gene-environment interactions shows inconsistencies, and causal assessment is delicate, particularly when replication is lacking.

Molecular epidemiologic studies have also been conducted on acute

lymphocytic leukaemia (ALL) in children, a disease that accounts for almost 25% of all childhood cancers. While more studies are needed, several have reported associations between parental or environmental exposure to benzene, or benzene-emitting sources, and childhood leukaemia, underscoring the potential importance of transplacental benzene exposures (58,59).

In summary, studies using biomarkers of internal dose, biologically effective dose, early preclinical effect/response and individual susceptibility (SNPs) have been valuable in elucidating the steps that link benzene exposure to the onset of leukaemia and other haematologic changes.

Nutritional factors and cancer

In the field of nutritional epidemiology, the investigation of biomarkers has shed some light on the role of obesity and metabolic syndrome in cancer. A high body mass index (BMI) has long been known to be associated with an increased risk of cancer at several sites, as the European Prospective Investigation into Cancer and Nutrition (EPIC) and other investigations have recently confirmed (60–63). The metabolic syndrome related to obesity is also suspected of a causal relationship with cancer (64). The metabolic syndrome is a constellation of central adiposity, impaired fasting glucose, elevated blood pressure and dyslipidemia (high triglyceride and low HDL cholesterol). The association of cancer with obesity and the metabolic syndrome has been unclear on biological grounds. Recently, however, several investigations have unveiled the role played by hormones and other intermediate markers related to key metabolic pathways. In particular, circulating insulin-like growth factor binding protein 1 (IGFBP-1), leptin, C-peptide and insulin are factors modified by obesity and have been associated with cancer. Higher circulating insulin levels may modulate cell proliferation and apoptosis, either directly or indirectly, by increasing the bioactivity of IGF-I, and decreasing the bioactivity of some of its binding proteins. Caloric restriction is a powerful way to reduce the occurrence of cancers, in particular lymphomas, induced by carcinogenic chemicals in *TP-53* deficient mice (65).

The evidence overall, however, is still incomplete. In a case–control study nested within the EPIC cohort involving 10 western European countries, serum C-peptide concentration was positively associated with an increased colorectal cancer risk for the highest versus the lowest quintile (OR = 1.56, 95% CI = 1.16–2.09, p for trend < 0.01). When stratified by anatomical site, the cancer risk was stronger in the colon (OR = 1.67, 95% CI = 1.14–2.46, p for trend < 0.01) than in the rectum (OR = 1.42, 95% CI = 0.90–2.25, p for trend = 0.35). No clear colorectal cancer risk associations were observed for IGFBP-1 or IGFBP-2. This large prospective study confirms that hyperinsulinemia, as determined by C-peptide levels, is associated with an increased colorectal cancer risk (66). In a nested case–control study in the prospective Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial, which examined associations between IGF-1 and IGFBP-3 and risk of prostate cancer, a total of 727 incident prostate cancer cases and 887 matched controls were selected for a similar analysis. There was no clear overall association between IGF-1, IGFBP-3 and IGF-1:IGFBP-3 molar ratio (IGFmr) and prostate cancer risk; however, IGFmr was associated with risk in obese men (BMI > 30, p for trend = 0.04), with a greater than two-fold increased risk in the highest IGFmr quartile (OR = 2.34, 95% CI = 1.10–5.01). Risk was specifically increased for aggressive disease in obese men (OR = 2.80, 95% CI = 1.11–7.08) (67). However, in the EPIC study only a weak association was found between these factors (IGFmr not analysed) and prostate cancer (68).

Another associated line of research refers to the role of inflammation and immunity in obesity. The posited mechanism would imply immune impairment that accompanies obesity, and possibly a gene-environment interaction with leptin and other genes implicated in obesity (69).

While the relationships among the different factors involved in the relationship between cancer and obesity and the metabolic syndrome, as well as the precise causal pathways, are still far from clear (70), this is a promising area of research.

Arsenic and urothelial cancer

As in several studies mentioned above, tumour markers have been used to help identify causal environmental exposures in bladder cancer. A recent study of the differential expression of molecular markers in tissues of arsenic-related urothelial cancers (AsUC) (n = 33), non-AsUC (n = 20), and normal bladder urothelia from patients with benign diseases (n = 4) were examined for multiple selected molecular markers responsible for various cellular functions, including *Bcl-2, p53*, and *c-Fos* (71). The expression of *Bcl-2* and *c-Fos* in AsUC was significantly higher than in

non-AsUC (P = 0.004 and P = 0.02, respectively), suggesting different carcinogenic pathways in the two etiologic groups. Such studies of the etiological heterogeneity of tumours at the molecular level may provide great insight into the mechanisms and causal pathways to carcinogenesis, which may lead to appropriate preventive strategies to reduce the incidence of cancer related to specific exposures (72).

Documenting environment-susceptibility interactions and identifying populations at greatest risk

To be effective, prevention strategies must target the most susceptible populations. This requires research to identify genetic and other susceptibility factors. Such research on exposure-susceptibility interactions must adhere to sound ethical principles, both in the conduct of research and in the communication of results and conclusions, in such a way as to discourage their inadvertent or intentional misuse (26,73–75). Although results from research on interactions have often been inconclusive and even conflicting, molecular epidemiologic studies indicate that some subgroups and individuals may have heightened susceptibility to environmental exposures. The categories of susceptibility factors that can modulate environmental risks, such as genetic predisposition, ethnicity, age, gender, and health and nutritional impairment, have been reviewed in detail elsewhere (26,73,74). With respect to the cancers and exposures discussed in this review, molecular epidemiologic studies have reported interactions between exposures to tobacco smoke, PAHs, AFB_1 or benzene and various susceptibility factors. These findings illustrate the complexities of interactions between environmental carcinogens and both genetic and non-genetic susceptibility factors. Susceptibility of the young has also been clearly demonstrated for several carcinogens.

Genetic susceptibility

Genes vary in their penetrance (the frequency, under given environmental conditions, with which a specific genotype is expressed by those individuals that possess it). Highly penetrant mutations in genes that are directly involved in carcinogenesis and confer a high risk of cancer in carriers represent the tail of a distribution of individual susceptibility to carcinogenesis (76). Less penetrant susceptibility may be conferred by common variants (SNPs) in genes that mediate the metabolism of carcinogens or DNA repair (77). For example, polymorphisms in certain cytochrome P450 (CYP) enzymes increase the oxidative metabolism of diverse endogenous and exogenous chemicals to their carcinogenic intermediates, while genetic variants in phase II (detoxifying) enzymes, such as glutathione S-transferase (*GST*), N-acetyltransferase (*NAT*), and epoxide hydrolase (*EH*) detoxify certain carcinogenic metabolites. Polymorphisms in DNA repair genes such as *XPD* or *XRCC1* can modulate risks from agents that directly or indirectly damage the DNA.

Rare and highly penetrant mutations in cancer genes may exert their effects without interacting with external exposures (usually by directly interfering with basic mechanisms of cell replication and differentiation), but gene–environment interactions are intrinsic to the mode of action of common, low-penetrance polymorphisms. The penetrance of a mutation is determined by other endogenous factors, including the importance of the function of the protein encoded by the gene (e.g. in key regulatory aspects of the cell cycle, as in the case of the *BRCA1* gene), the functional importance of the mutation (e.g. a total loss of function due to a truncating deletion versus a mild loss of function due to a point mutation), the existence of alternative pathways that can substitute for the loss of function, and interactions with other genes.

Most genes act in a sequence or in cascades. This is typical, for example, of metabolic and DNA repair genes. Genotyping according to pathways is likely to be much more rewarding then genotyping for single SNPs, in terms of both biological plausibility and statistical power (see discussion on the role of DNA repair genes (78)).

A large number of SNPs have been studied in molecular epidemiological investigations in recent decades, thanks to the development of quick and relatively inexpensive genetic techniques. However, only a few clear associations with cancer risk have been detected with reasonable certainty (i.e. consistently across different populations). Even with these SNPs most closely linked to cancer risk, caution is needed in extrapolating from one population and exposure scenario to another.

An example of a SNP consistently implicated in cancer is the methylenetetrahydrofolate reductase (*MTHFR*) gene, which plays an important role in the folate metabolism pathway (40,78). This enzyme provides the methyl group required for *de novo* methionine synthesis, and indirectly, for DNA methylation; therefore, it controls DNA stability and mutagenesis (79–81). Common *MTHFR*

polymorphisms (C677T and A1298C) have been associated with reduced enzyme activity *in vitro* which, in the case of C677T, affects the metabolism of folate, consequently increasing homocysteine levels and (theoretically) the risk of colon cancer (82). According to a systematic review, in most studies *MTHFR* 677TT (10 studies, >4000 cases) and 1298CC (four studies, >1500 cases) were associated with moderately reduced colorectal cancer risk. In four of five genotype-diet interaction studies, 677TT subjects who had higher folate levels (or a high-methyl diet) had the lowest cancer risk (82).

An interaction of *MTHFR* SNPs with alcohol intake has also been reported, with high alcohol consumption levels decreasing DNA methylation, probably by hindering folate absorption, metabolism and excretion (83). Alcohol is thought to increase risk of cancer through its antagonist effects on folate. A study of health professionals examined folate, alcohol, *MTHFR* and alcohol dehydrogenase 3 (*ADH3*) polymorphisms in relation to risk of colorectal adenomas in 379 cases and 726 controls (84). *MTHFR* genotypes were not found to be appreciably related to risk of adenoma, but men who were TT homozygotes and who consumed 30+ g/day of alcohol had an OR of 3.52 (95% CI = 1.41–8.78) relative to drinkers of ≤ 5 g/day with the CC/CT genotypes (84).

Studies investigating the folate-*MTHFR*-cancer risk relationship have largely shown inverse associations of breast cancer risk with folate intake in all genotype groups, particularly among subjects with the 677TT genotype (85,86). Although the evidence is not conclusive, *MTHFR* provides a good example of how inherited gene variants can modify the cancer risk associated with dietary and other exposures.

With respect to lung cancer, various studies have implicated genetic polymorphisms involved in PAH metabolism (e.g. *CYP1A1* or *GST*) and DNA repair (e.g. *XRCC1*) as effect modifiers capable of increasing risk from PAHs (87–91). Increased risk of hepatocellular carcinoma has been associated with the *GSTM1* null/*GSTT1* null genotype in conjunction with smoking and drinking (92). The *GSTM1* null genotype, the low-activity epoxide hydrolase genotype, and a genetic polymorphism in *CYP2E1* also appear to confer greater risk of liver cancer (93,94).

Regarding leukaemia, the hepatic cytochrome P450 2E1 enzyme plays a key role in the activation of benzene to its ultimate haematotoxic and genotoxic benzoquinone metabolites (95). The NAD(P)H:quinone oxidoreductase (*NQO1*) and two subclasses of GSTs (M1 and T1) are involved in the detoxification of the ultimate benzoquinones and their reactive benzene oxide intermediates, respectively (50,95). A case–control study of occupational benzene poisoning in Shanghai showed that individuals homozygous for the *NQO1*^{609} C→A mutation were at a 7.6-fold (95% CI = 1.8–31.2) greater risk of poisoning (96). Benzene poisoning was linked to risk of preleukemia and leukaemia. Theoretically, individuals with high activities of cytochrome P450 2E1 and homozygous mutations in the *NQO1*, *GSTT1* and *GSTM1* genes would be at highest risk of benzene haematotoxicity (50), but this inference has not been demonstrated conclusively. As noted earlier, polymorphisms in several IL and VCAM genes have been implicated in benzene haematoxicity (57).

DNA repair capacity is a particularly important source of variability in susceptibility to cancer. In addition to rare syndromes that involve faulty repair and genetic instability (e.g. ataxia-teleangectasia, Fanconi anaemia, Bloom syndrome, and xeroderma pigmentosum) (97), individuals commonly vary in their capability to repair DNA damage, at least in part due to genetics. The role of SNPs in three DNA repair genes (*XRCC1-Arg399Gln, exon 10; XRCC3-Thr241Met, exon 7;* and *XPD-Lys751Gln, exon 23*), and their combination in modulating the levels of DNA adducts in a population sample of healthy individuals has been investigated (98). The ^{32}P-postlabelling assay was used to measure aromatic DNA-adduct levels in WBCs from peripheral blood. A dose–response relationship between the number of at-risk alleles and the levels of adducts (P = 0.004) was observed, suggesting that the combination of multiple variant alleles may be more important than single SNPs in modulating cancer risk; hence the importance of focusing on gene pathways in the study of gene-environment interactions.

In addition to SNPs or polymorphisms in DNA repair genes, phenotypic tests have been widely used in recent years to measure DNA repair. The mutagen sensitivity assay, based on DNA damage (usually chromosome breaks) induced with chemical (bleomycin) or physical mutagens (radiation), unscheduled DNA synthesis, ^{3}H-thymidine incorporation, or count of pyrimidine dimers are examples of tests by which DNA repair is inferred from the different frequency of DNA damage induced in cancer cases and controls, without direct evidence of repair. Other phenotypic tests (e.g. the

plasmid cat gene test, the ADPRT modulation test, or immunoassays based on antigenicity of thymidine) are based on some direct evidence of repair (77).

In contrast to genotype-based studies of DNA repair, for which the evidence is still largely inconsistent (78,99), most studies using phenotypic tests from which DNA repair is inferred have shown highly statistically significant results (100). When odds ratios were available, they were between 2.8 and 10.3, suggesting a strong association. However, the results are limited by potential confounding (i.e. the possibility that some exposure or characteristic of the patient is associated with DNA repair and is a risk factor for cancer, thus creating a spurious relationship between DNA repair and the disease). Repair enzymes can be induced in several ways, such as by stresses that damage DNA (e.g. pro-oxidative stress that could result from several endogenous and exogenous exposures). For example, in human studies, several tests of DNA repair were affected by characteristics such as age, sunlight, dietary habits, exposure to pro-oxidants, and cancer therapies (100). While age and therapies were usually controlled for in most studies, dietary habits might have acted as confounders, since both the intake and the plasma level of carotenoids and other antioxidants have been shown to be lower in cancer patients compared to healthy controls. The extent of such potential confounding is unknown, but could be substantial.

Another major limitation of many tests is that DNA repair is only indirectly inferred from DNA damage. To draw firm conclusions about a cause-effect relationship, more information about the biological meaning of tests is needed—for instance, whether they actually reflect DNA repair or a general or specific impairment of the DNA repair machinery.

Many investigations of gene–environment interactions (GEI) are underway in different parts of the world. Some ongoing studies are extremely large (e.g. EPIC, United Kingdom Biobank); all of them employ similar methods for genotyping (Taqman and high-throughput methods, such as Illumina). However, the quality of exposure assessment (e.g. diet, air pollution) is extremely variable. Ideally, understanding GEI requires determining, with equal resolution, both environmental exposures (e.g. to pesticides, air pollutants, ETS or dietary constituents) and genetic variants that are postulated to modulate the effects of the environmental exposures. However, there is an asymmetry between the two in that genotyping is much more accurate than most of methods used to measure environmental exposures. This implies a lower degree of genetic classification error, which in turn means an easier identification of associations between genes and disease than with environmental exposures and disease.

Suppose that classification error is expressed by the correlation coefficient between each exposure "assessor" and a reference standard ($r = 1$ means no error, $r = 0.9$ means a 10% classification error). For different expected relative risks that associate exposure with disease, one can compute the relative risks under different conditions of classification error. For example, a classification error of 10% implies the drop of a relative risk of 2.5 to 2.3 (i.e. little change). With an extreme classification error of 90%, however, even a relative risk of 2.5 becomes 1.1 (i.e. undetectable with common epidemiological methods). The lesson is that false-negative results are much more likely when analysing the role of environmental exposures than genetic variables (while in the latter case false-positives may be the main problem). In addition, very large numbers of subjects are needed if one wants to study interaction, for example, between a frequent exposure (prevalence 25%) and a frequent genotype (prevalence 50%). Presume that classification error is 20% for the environmental exposure (sensitivity = 80%) (in actuality, classification error for most exposures is likely to be much larger). Classification error could be around 7% for genotyping (sensitivity 93%). This is realistic, since genotyping techniques are currently validated and extremely accurate. The consequence of this situation is that approximately 1800 cases would be needed to observe main effects of genes if no classification error occurs, 2700 if exposure is incorrectly classified 20% of the time, and 3200 if the genotype is also mistaken 7% of the time. To study the effect of interactions, four times more subjects than those estimated above would be needed.

False-positives seem to be a common problem in genetic research, often due to small numbers and statistical instability. As pointed out by Ioannidis (the "Proteus phenomenon"), gene–disease associations that seem to be strong at first appear to be much weaker when larger studies are conducted (101). Publication bias contributes to this problem. For this reason, initiatives like the Venice criteria have been launched to provide sound systematic reviews of the genetic evidence (102).

In conclusion, the evidence concerning the role of low-penetrant genes in cancer is contradictory and

difficult to interpret. Most observed associations between cancer and low-penetrant gene variants have been weak or very weak (with 20–50% increases in cancer risk). This, in fact, is inherent in the concept of low penetrance. However, the penetrance of a gene variant depends on interaction with external exposures, the internal environment, or other genes. Thus, the strength of association is a relative, not absolute, concept and requires the study of interactions. Nonetheless, interactions themselves are obviously difficult to investigate, as the study of a two-way interaction requires a sample size four times larger than the study of a main effect; therefore, little is known about the nature and strength of gene–environment interactions.

Genome-wide association studies (GWAS) and new methodological issues

Technical developments, with platforms such as Illumina or the Affymetrix microchips, offer the possibility of analysing up to 550 000 or even 1 million gene variants in one run. This revolution is giving rise to an unprecedented wave of new potential discoveries, as is testified by several papers in *Nature, Science*, and *Nature Genetics* in 2007, such as the Wellcome Trust Case-Control Study Consortium (103). Regarding cancer, a successful story is represented by the identification of chromosome 8q24 as the probable locus of a genetic risk factor for prostate cancer. Family-based linkage studies, association studies, and studies of tumours had already highlighted human chromosome 8q as a genomic region of interest for prostate cancer susceptibility loci. Recently, a locus at 8q24, characterized by both a SNP and a microsatellite marker, was shown to be associated with prostate cancer risk in Icelandic, Swedish and US samples (104). These data suggest that the locus on chromosome 8q24 harbours a genetic variant associated with prostate cancer, and that the microsatellite marker is a stronger risk factor for aggressive prostate cancers defined by poorly differentiated tumour morphology. Evidence has now been provided that colon cancer might also be associated with the same region. Using a multistage genetic association approach comprising 7480 affected individuals and 7779 controls, researchers have also identified markers in chromosomal region 8q24 associated with colorectal cancer (105). This example is interesting for two reasons: reverse genetics (the possibility that etiologic pathways for cancers that elude epidemiological research can be discovered starting from genetic susceptibility) and pleiotropy (the ability of certain gene variants to increase/modulate the risk for quite different diseases). (See Chapter 6 and (106,107) for a summary of recent GWAS findings.)

Apart from the 8q24 success story, many other contributions to the potential understanding of cancer and other diseases have come from GWAS. Exfoliation glaucoma is a striking example for which a potent signal has been identified, but this is an exception (in addition to being a non-cancer example). A cancer example is the KITLG gene and testicular carcinoma (see (108)). A summary of the locuses associated with cancer and other diseases after GWAS is available in the so-called GWAS catalogue of the National Human Genome Research Institute (http://www.genome.gov/GWAStudies).

However, genome-wide scans are clearly open to an even greater risk of false-positive findings related to multiple comparisons. Also, the interaction with external exposures is usually ignored. Design issues, including the investigation of traits that show strong familial aggregation, the selection of clinically homogeneous populations, and selection of cases that have a family history, are emerging as very influential on the success of genome-wide studies (109). (See also Chapter 6 for discussion of methodologic issues.)

Ethnicity, gender and nutritional factors

Ethnicity also appears to affect cancer risk. For example, higher rates of various smoking-related cancers in blacks may be partially explained by the finding that, in black smokers, urinary concentrations of NNK metabolites and serum concentrations of cotinine, a nicotine metabolite, exceeded those in white smokers (110). However, the effect of unmeasured differences in the exposure levels of the subjects cannot be ruled out.

Although studies have been inconsistent, there is evidence that women may be inherently more susceptible than men, on a dose-for-dose basis, to certain lung carcinogens. Several epidemiologic studies indicate that women smokers are 1.7- to 3-fold more likely to develop lung cancer than are male smokers with the same exposure (111,112). The level of PAH-DNA adducts and the frequency of G:C→T:A transversions in *p53* were elevated in lung tumours from female smokers compared with those from male smokers (112–114). Adduct levels in non-tumour lung tissue were also higher in women than in men, with a higher ratio of adduct levels to pack-years in women (115). The greater expression of the *CYP1A1* gene found in lung tissue

of female smokers suggests a possible mechanism for this gender difference. In addition, a case–control study of lung cancer found that the effect of the *GSTM1* null genotype on lung cancer risk was significant among women, but not among men (116).

Nutritional deficits resulting in low levels of antioxidants can also heighten susceptibility to lung and other carcinogens by increasing DNA damage and subsequent mutation and carcinogenesis by oxygen radicals, PAHs and other chemical carcinogens (117). Heavy smokers with low plasma levels of micronutrients, such as retinol and the antioxidant α-tocopherol, appear to have reduced protection against carcinogen-induced DNA damage (118). In several studies, these effects were seen only in smokers with the *GSTM1* null genotype, illustrating the importance of interactions between multiple susceptibility factors (119,120). Sensitivity to mutagens, as measured by bleomycin-induced chromatid breaks, was also increased in cultured lymphocytes of healthy individuals with low plasma levels of antioxidants (121).

A special case of susceptibility: The fetus and young child

Compared with exposures occurring in adult life, exposures *in utero* and in the early years can disproportionately increase the risks of childhood cancer and many types of cancer later in life (122–124). Experimental and epidemiologic data indicate that because of differential exposure or physiologic immaturity, fetuses, infants and children experience greater risks than adults from a variety of environmental toxicants, including PAHs, nitrosamines, pesticides, tobacco smoke, air pollution and radiation. The underlying mechanisms may include increased exposure to toxicants, greater absorption or retention of toxicants, reduced detoxification and DNA repair, the higher rate of cell proliferation during early stages of development, and the fact that cancers initiated in the womb and in the early years have the opportunity to develop over many decades (for a review, see (125)) (126–129).

New evidence has emerged in recent years on the role played by *in utero* exposures in relation to the development of cancer in childhood and adult life. Fetuses and newborns seem to be particularly susceptible to diverse carcinogens (126–129). In a series of studies, PAH-DNA adducts were evaluated in mother-newborn pairs in central Europe, the USA and China (130,131). Consistently, levels of adducts in newborn cord blood were the same or higher than in the mothers' blood, although estimated transplacental exposure based on experimental studies was on the order of one tenth of maternal exposures. These observations across a gradient of exposure and in four different ethnic groups suggested that the fetus may be 10-fold more susceptible to DNA damage than the mother, and that *in utero* exposure to PAHs may disproportionately increase carcinogenic risk. Underscoring the potential risk of transplacental exposure to carcinogens, PAH/aromatic DNA adducts in cord blood were positively associated with hypoxanthine-guanine phosphoribosyltransferase (HPRT) mutant frequency in newborns (25). These studies provided molecular evidence of links between transplacental exposure to common air pollutants and somatic mutations indicative of increased cancer risk. In another study, airborne PAHs, measured by personal air monitoring during pregnancy, were significantly associated with stable aberration frequencies in cord blood (132). However, the epidemiologic evidence is still inconclusive on the role of transplacental exposure to PAHs and air pollution and childhood cancer (133).

Other investigators have reported that prenatal or postnatal exposure to tobacco smoke or its constituents were associated with increased frequencies of DNA and haemoglobin adducts, as well as chromosomal aberrations in newborns or children (134,135). A significant association between paternal smoking (without maternal smoking) and death from childhood cancer was found (136). A significant difference in the HPRT mutational spectrum was reported between newborns of mothers exposed to environmental tobacco smoke (ETS) and newborns of unexposed mothers. Their results suggested that V(D)J recombinase mutations, which are associated with leukaemia and lymphomas, are induced by ETS exposure (137). A meta-analysis of 11 studies of childhood exposure to maternal and paternal ETS found a very small excess risk for childhood cancer (RR = 1.10; CI = 1.03–1.19) (138). In addition, early-life exposure to ETS is suspected of playing a causal role in adult cancer. Three studies found that childhood exposure to ETS increased the risk of lung cancer in adults (139–141).

There is direct chromosomal evidence of a link between *in utero* exposures and cancer in infancy and childhood. Approximately 75% of infant acute leukemias have a reciprocal translocation between chromosome 11q23 and one of several partner chromosomes, including chromosome 4, which creates a fusion of the MLL gene at 11q23 and the AF4 gene at 4q21. Providing direct evidence of prenatal

initiation of infant leukemias, the MLL-AF4 gene fusion sequence has been detected in neonatal blood spots of leukaemia patients subsequently diagnosed at ages five months to two years (141). Similarly, a signal mutation (TEL-AML1) observed in 25% of childhood acute lymphocytic leukaemia (ALL) cases was found to be present in neonatal bloodspots of children who subsequently developed ALL (142). The interpretation is that the TEL-AML1 fusion is acquired prenatally and constitutes the "first hit" in childhood leukaemia.

Adolescence and young adulthood are also viewed as sensitive stages of life because of greater proliferative activity in epithelial cells of certain tissues, as seen in radiation-induced breast cancer (143). Initiation of smoking at an early age confers a higher risk of lung, bladder and possibly breast cancer (144). Breast cancer risk associated with the *NAT2* slow acetylator genotype was higher in women who began smoking under the age of 16 years (145). In addition, aromatic DNA adduct levels were highest in lung tissue of former smokers who had smoked during adolescence, suggesting either that smoking at a young age induces more persistent adducts or that young smokers are more susceptible to DNA adduct formation (146).

In conclusion, molecular epidemiology has provided valuable data on the existence of complex interactions between environmental exposures and susceptibility factors, and has spurred researchers to investigate further differences in susceptibility among subsets of the population. Neither experimental nor conventional epidemiologic research alone could have done this. Although more research is needed before risk assessors can routinely develop quantitative estimates of the risks to sensitive subsets posed by specific environmental agents, the information obtained thus far has relevance to risk assessment and prevention. For example, government agencies are already beginning to require that regulatory policies explicitly protect children as a susceptible group.

The promise and challenge of new "omic" and epigenetic biomarkers

Types of new biomarkers

Several new and exciting biomarkers are becoming available for epidemiological studies thanks to the development of high-throughput technologies and theoretical advancements in biology. However, most of these markers have not yet been adequately validated, and their role in the causal paradigm is not clear. An exhaustive review is not possible here, and the reader is referred to Chapter 5 and other critical reviews, in particular for gene expression and toxicogenomics (147–149).

Toxicogenomics

Toxicogenomics refers to the study of the complex interaction between the cells' genome and chemicals in the environment or drugs, as they relate to disease. One method for genome-wide analysis, comparative genomic hybridization (CGH), provides a molecular cytogenetic approach for genome-wide scanning of differences in DNA sequence copy number (150). This technique has been attracting widespread interest among cancer researchers, as evidenced by the rapidly expanding database of CGH publications that already covers about 1500 tumours, and is beginning to reveal genetic abnormalities characteristic of certain tumour types or stages of tumour progression. In theory, such genomic differences could be exploited to gain insights into the risk factors involved (150).

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, since it determines the process by which the instructions in genes are converted to mRNA, directing protein synthesis (81). DNA methylation, that is, the covalent addition of methyl groups (CH3) to cytosine that precedes a guanine in the DNA sequence (the CpG dinucleotide), occurs naturally and plays a role in suppressing gene expression. CpG dinucleotides are enriched in the promoting regions of genes (CpG islands). Hypermethylation of promoter regions is associated with gene transcriptional silencing, and is a common mechanism for the inactivation of tumour suppressor genes in human cancer (151). DNA methylation is heritable; it passes from one generation of cells to the next.

Promoter methylation is a mechanism that regulates gene expression and is believed to play a crucial role in lung carcinogenesis. Several genes are commonly the target of promoter hypermethylation in lung cancer, including the *p16* gene (*$p16^{INK4a}$/CDKN2A*), *DAPK, RAR-β, RASSF1* and *O^6MGMT* (a DNA-repair gene) (152). Global hypomethylation has also been

observed (153). Both current and former smoking have been associated with aberrant *p16*, *DAPK, RASSF1A* and *RAR-β* methylation (152). Recently, investigators have found that two alternative pathways can be detected in the biopsies of smoking and non-smoking lung cancer patients: one involving methylation and *K-ras* mutations, and the other *EGRF* mutations in the absence of gene methylation (154). 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% (155). Notably, aberrant promoter methylation can be detected in the plasma of lung cancer patients (156); high frequencies of *ECAD* and *DAPK* methylation have been reported in lymphocytes of smokers versus non-smokers (157). The capacity of some airborne particulate carcinogens to induce hypermethylation in the regulatory regions of tumour suppressor genes has also been demonstrated in animal studies (158). Overall, the animal models support involvement of promoter methylation and other epigenetic mechanisms in carcinogen-induced lung carcinogenesis (159,160).

Acetylation is another key mechanism in epigenetic pathways, although it has been studied less extensively than methylation in cancer epidemiology (161).

Metabolomics

The study of the complete set of low-molecular weight metabolites present in a cell or organism at any time is metabolomics, sometimes referred to as metabolomics. With high-throughput techniques (NMR spectroscopy and LC-MS) 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 (162). Use of data processing and chemometric models has already allowed the characterization of disease states and metabolic disorders (163). While several cross-sectional metabonomic studies investigating various cancers have been undertaken (164,165), no longitudinal study has yet been carried out, and few validation studies have been published. In one investigation of repeat samples from dietary studies (166), high-resolution ^{1}H NMR spectroscopy was used to characterize 24-hour urine specimens obtained from population samples in Japan (n = 259), Chicago, USA (n = 315), and China (n = 278). The authors investigated analytical reproducibility, urine specimen storage procedures, interinstrument variability, and split specimen detection. The multivariate analytical reproducibility of the NMR 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

The study of an organism's entire complement of proteins is known as proteomics. Proteomics has been used for the investigation of several types of cancer (167–170) and of physiological or pathological changes associated with external exposures. Proteomic studies have identified, for example, changes in proteins associated with oxidative stress (171). 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, methodological issues need to be resolved before application in prospective studies. In a critique of early papers, Diamandis (2004) identified several methodologic problems: the lack of reproducibility in analytical 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 and processing, freeze–thaw, and duration of 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 non-reproducible data analysis. Many of these concerns apply to other epigenetic and "omic" biomarkers and have been addressed in subsequent proteomic studies. In conclusion, for all the "omic" technologies, validation studies are urgently needed.

Incorporating new intermediate epigenetic or "omic" biomarkers into etiologic studies

Epigenetic and "omic" technologies can provide intermediate markers (either reflecting exposure/effective dose, early effects, or preclinical disease) for etiologic purposes (to investigate the causes and mechanisms of disease onset) or for clinical purposes (early diagnosis, prognosis, follow-up). This chapter refers to etiologic purposes, but many of the considerations apply to clinical purposes as well (see (172) and (173) for a review of biomarker-based tools for cancer screening, diagnosis and treatment). While past experience with earlier biomarkers

is relevant, the current era is different and poses new challenges for the following reasons: "omic" and new epigenetic methods tend to be discovery-oriented, rather than oriented to testing specific hypotheses; the main feature of current technologies is the ability to perform massive testing of markers (i.e. thousands of markers at a time), potentially in thousands of subjects; and 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, methodological challenges have expanded more than proportionally. In fact, much uncertainty surrounds the validity and applicability of new technologies (see (174, 175)).

Feasibility is also an issue. For example, it is currently prohibitively expensive and labour-intensive to perform expression array analysis for every subject in large studies. An alternative is to select a small subset of matched pairs of exposed and unexposed subjects (or subjects with and without preneoplastic lesions) and discover differentially exposed genes. Once several target genes are identified, real-time PCR analysis can be used to quantify expression of selected genes in all subjects (176).

Another important difference between the earlier and newer biomarkers is that the traditional cancer paradigm was very much centred around DNA damage and mutations, while recent research has uncovered several additional intermediate steps between genotype and phenotype, and has highlighted the importance of gene expression/modulation in carcinogenesis. Therefore, combinations of both types of biomarkers are expected to be informative, since pathways are not mutually exclusive.

Several critical steps in the putative causal pathway linking exposure to the onset of cancer can be explored with intermediate markers. Referring to the classical scheme (Cf. Figure 19.1), intermediate markers can play a role in each of the following steps: they can be related to exposure (e.g. metabolomics); related to early effects or changes in the causal pathways leading to disease (like promoter methylation, gene mutations, or changes in telomere length); or they can express epiphenomena of pre-clinical disease (e.g. mutations present in plasma DNA as a consequence of tumour cell apoptosis). It is very important that the biological significance of a marker be made explicit beforehand, because false expectations can arise as a consequence of an erroneous interpretation of a biomarker's role. For example, some markers (those on the right side of the scheme) have clinical relevance or can be useful for screening, others cannot.

Validating promising intermediate markers

A concept that is often unclear is the difference between technical and field validation. Technical validation has to do with intrinsic measurement error and analytical sensitivity. Field (or epidemiological) validation is related to how a certain marker behaves in the population, depending on biological variability within the population (177).

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 may still perform 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.

One of the main summary measures of the contribution of a biomarker to the prediction of disease onset is the receiver operating characteristic (ROC) curve. The ROC curve is a measure of the overall capability of the marker to predict the disease, which is a function of sensitivity and specificity. An area under the curve (AUC) of 1 or close to 1 indicates perfect prediction, while an area close to 0.5 indicates random association between the marker measurement and the probability of disease onset. The maximum AUC for the prostate serum antigen (PSA) test (a serum tumour marker to predict the presence of prostate cancer) is only 0.77 (178). It is possible that gene expression microarrays or proteomics could perform better than the PSA test, but no candidate biomarker has yet been identified.

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: biologic variability related to the subject (i.e. variability between subjects (intersubject) and within subjects (intrasubject)); variability due to measurement error, including intralaboratory and interlaboratory variability; and random error. Methodological issues should be discussed within the context of specific biomarker categories. When epidemiologic studies employing biomarkers are designed and analysed, the goal is 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 analytical replicates used in the study design, respectively. Obviously, if detailed information is not available, intragroup variation cannot be adjusted for. Therefore, in epidemiologic studies employing biomarkers it is important to collect, whenever possible: repeat samples (day-to-day, month-to-month, or year-to-year variation may be relevant depending on the marker); information on subject characteristics that may influence intersubject variation; and information on conditions under which samples have been collected and laboratory analyses have been conducted (batch, assay, specific procedures). (For more about how the variability in laboratory measurements influences study design decisions, see (179).)

To increase power and improve the quality of studies, consortia like the NCI Cohort Consortium have recently been created. While these have been set up mainly to share questionnaire or GWAS data, consortia can also be extremely helpful for biomarker research including omics and biomarker validation. One recent example is the series of papers that examined the association between Vitamin D and several cancer sites (180).

Design issues

Study design issues identified with earlier biomarkers, such as mutation, oxidative damage, and adducts are particularly relevant for the use of newer intermediate markers, such as proteomic changes. Only prospective studies allow for a proper temporal evaluation of the role of intermediate biomarkers. The use of the cross-sectional design in the analysis of *p53* mutations has been an invaluable tool in the investigation of liver carcinogenesis. However, the cross-sectional design of the early studies did not allow researchers to exclude the possibility that mutations were a consequence of cell selection rather than of the original causal agent, such as aflatoxins (181). In other words, what was observed was the spectrum of mutations in liver cancers as the consequence of a long and complex process involving the effect of carcinogens, DNA repair, and the selection of cells carrying specific mutations conferring a selective advantage to cells. Therefore, in principle, prospective studies are better for the understanding of time relationships between exposure, intermediate biomarkers, and disease although they have the limitation of usually being based on a single spot biological sample, which does not allow the measurement of intraindividual variation.

Randomized clinical trials (RCTs) with biological samples have been repeatedly performed. For example, trials have used dietary changes as the intervention and oxidative damage or DNA adducts as the outcome. Though RCTs are probably the best design to conclude causality in epidemiology, they also have limitations, particularly the short half-life of most biomarkers and most interventions, compared to the long-term exposures that are needed to cause chronic diseases like cancer.

Another issue with respect to some biomarkers is that it is often difficult to understand whether the marker is intermediate in the pathway leading from exposure to disease, or it is just a consequence of exposure with no role in disease onset, or even an epiphenomenon of disease with no relationship to exposure. For example, micronuclei seem to originate from exposure to clastogens, but can lead to cell death and therefore are likely not to be intermediate in the causal pathway.

Types of bias that are common in other epidemiological studies may become dramatic when biological samples are collected and biomarkers are measured. For example, in a study on pancreas cancer, out of more than 1000 eligible patients, the investigators were only able to extract DNA from 46 biopsies (182). The patients with a biopsy available were more frequently white and the tumour size was on average 179 mm, versus 570 mm among the patients whose biopsy was not made available. This discrepancy is likely to introduce bias if one tries to correlate the prevalence of somatic mutations with exposure characteristics, such as occupation. As further example, in a case–case study, patients with pancreatic cancer seen at two general hospitals were retrospectively identified (183). Their clinical records were abstracted and paraffin-embedded samples retrieved from pathology records. DNA was amplified and mutations in codon 12 of the *K-ras* gene were detected. Results on the mutations were obtained for 51 of the 149 cases (34.2%). Mutation data were over five times more likely to be available from one of the hospitals. In particular, subjects with mutations were more likely to have received a treatment with curative intent (OR = 11.56; 95% CI = 2.88–46.36).

In addition, special forms of confounding may affect molecular epidemiology. An example is the levels of plasma DNA in cancer patients and controls, in the context

of a multicentre cohort study. Researchers found that, although the level of plasma DNA seemed to predict the onset of cancer, it was also strongly associated with the recruitment centre. This was due to modalities of blood collection and storage, since a longer time elapsing between blood drawing and storage in liquid nitrogen was associated with higher DNA levels due to greater white blood cell death. Thus, the association between cancer and plasma DNA levels could be confounded by centre, since cancer rates also differed by centre in this multicentre study (184).

Translation of research into preventive programmes

Assessing risk

Risk assessment for low-level exposures

One of the main challenges for epidemiology in recent decades has been the need to characterize and quantify the effects of low-level exposures to carcinogens. Such exposures are widespread (e.g. traffic-related air pollution and ETS) but extremely difficult to study with conventional epidemiological tools. There has been a heated debate on the shape of dose–response relationships in carcinogenesis (i.e. on the extrapolation from high- to low-levels of dose), an issue of great public health significance. Epidemiological studies are often underpowered to study the carcinogenic effects of very low levels of exposure. Using biomarkers, molecular epidemiology can mitigate the problem that very large numbers of subjects are needed to detect small effects on cancer risk, by providing individual estimates of dose and intermediate markers of procarcinogenic damage that can be used *in lieu* of cancer as an outcome.

To illustrate these points, results are described from a series of analyses that have been carried out by investigators in the EPIC study on the effects of low-level exposure to ETS and air pollution on lung cancer. ETS and air pollution share several characteristics: they are widespread exposures in both developed and developing countries, they have chemical components in common, and they are associated with increased risks of lung cancer and other diseases (185,186). The lung cancer relative increase is around 20–30%, approximately of the same magnitude for both ETS and air pollution at the typical exposure levels in Western countries (158,187). In EPIC, relative risks were found in the order of 1.4–1.5 for exposure to ETS in adulthood and the risk of lung cancer, based on the prospective investigation of about 120 000 subjects with information on ETS and 117 newly diagnosed lung cancers in non-smokers (140). Biomarkers were used in several different ways. First, cotinine was used to validate the questionnaire information on ETS exposure, demonstrating a strong association with self-reported exposure ($P < 0.001$). Second, biomarkers of genetic susceptibility strengthened the epidemiological association between low-level exposures to carcinogens and cancer. The risk associated with ETS was higher in subjects with three or more at-risk alleles for genes involved in carcinogen metabolism (*GSTM1, GSTM3, GSTP1, GSTT1, CYP1A1, CYP1B1, NAT2, MnSOD, MPO,* and *NQO1*), with an odds ratio of 2.86 compared to 1.33 in those with less than three alleles (140). These results have implications for risk assessment in that they show a modest, but significant, increase in cancer risk at low levels of exposure to environmental carcinogens, and demonstrate that genetic factors can substantially increase risk to certain subsets of the population.

Developing dose–response models for assessing the risk of carcinogens

A major issue relevant to risk assessment is whether to view carcinogenesis as a linear process, involving the accumulation of several additive events, or as a nonlinear process. Molecular data on carcinogenic mechanisms have been instrumental in developing and validating different statistical models for carcinogen risk assessment. At the time of the initial development of the molecular epidemiology paradigm (2,3), the dominant model of carcinogenesis was the "multistage" model proposed by Armitage and Doll that postulated the existence of about six stages in cancer development (188). Armitage and Doll's model was consistent with the paradigm, implying several steps between exposure and cancer, and an important role for duration of exposure to carcinogenic stimuli. Steps were postulated to be heritable from one cell to the progeny, and critical genetic changes were hypothesized to be irreversible. In fact, after the multistage model was originally proposed, Vogelstein demonstrated, on the basis of molecular pathology, that the development of colon cancer was likely to require six mutations or chromosome aberrations (189).

In addition, Knudson had suggested, based on its age distribution, that retinoblastoma (Rb) in children was likely to be due to two mutations, one inherited and one acquired (190). The Knudson "two-hit" model for retinoblastoma was supported by the discovery of the

first tumour suppressor gene (*Rb*) that in fact requires two mutations, one inherited and one acquired, to give rise to the tumour (191). Thus, both models were examples of the success of combining epidemiological observations, mathematical models, and molecular or biomarker evidence.

Yet another statistical model, based directly on molecular evidence, derives from the identification of hereditary syndromes that predispose to colon cancer (hereditary non-polyposis colon cancer) through mutations in the mismatch repair genes (77). The corresponding model postulates that the rate of mutations is too low to explain the incidence of cancer in human populations; therefore, a "mutator phenotype" (the inherited or acquired ability to develop frequent mutations, such as through a defect in DNA repair machinery) would be necessary (192,193). A cascade of mutations, originated by the inability to repair DNA damage, would better explain the high frequency of colon cancers in some families, than the simple accumulation of spontaneous or acquired mutations. The same could be true for "sporadic" cancers.

Several other models have been proposed in recent years that accommodate and reflect new molecular information on carcinogenic mechanisms. One recent model (194) is based on the concept that carcinogenesis is a Darwinian process in which transformed cells acquire a selective advantage over normal cells. The term "selectogen" has been proposed for carcinogens that act by increasing the ability of mutated cells to acquire selective advantage (in given environments) over normal cells. A biomarker that has been used to explore such a Darwinian concept of carcinogenesis is mutation in the *HPRT* reporter gene. The X-chromosomal gene for *HPRT* serves as a simple reporter gene (i.e. it indicates the induction of mutations) and is now finding use in studies of *in vivo* selection for mutations arising in either somatic or germinal cells (195). This line of research, however, is still in its infancy.

All these apparently diverse models are in fact generally compatible and consistent with molecular data. The picture that is emerging is that environmental stimuli can increase genomic instability (in addition to inherited variants of instability), which in turn leads to chromosome aberrations or mutations that increase the selective advantage of cells in stressful environments, and induces the carcinogenic process. However, the problem with mathematical models is that often they are compatible with different biological interpretations and do not easily accommodate certain aspects of carcinogenesis, such as epigenetics. The incorporation of non-genetic biomarkers into risk assessment models is still in a very early stage.

Developing new intervention strategies

Primary prevention encompasses a spectrum of measures that includes elimination or avoidance of exposure, prevention of carcinogen activation after it has entered the body, blocking interactions with the genome, and suppressing the propagation of premalignant changes. Several studies have used DNA damage as an intermediate biomarker or endpoint. An example is a study of smokers enrolled in a smoking cessation programme. Levels of biomarkers, PAH–DNA and 4-ABP–haemoglobin adducts, reflecting cessation were significantly reduced by eight weeks after quitting smoking (196). 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 (197).

Other prevention research has used biomarkers as intermediate endpoints in chemoprevention trials. Research studies have shown that isothiocyanates, which occur as conjugates in a wide variety of cruciferous vegetables, are involved in the inhibition of carcinogenesis (14). Isothiocyanates appear to selectively inhibit cytochrome P450 enzymes involved in carcinogen metabolic inactivation; they also induce Phase II enzymes and enhance apoptosis. Phenethyl isothiocyanate is a particularly effective inhibitor of lung tumour induction by the tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and is currently being developed as a chemopreventive agent against lung cancer (110).

Several dietary or vitamin supplementation randomized studies have used DNA adducts or 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 (198). Oxidative damage to DNA plays a major role in carcinogenesis, and all living cells have defence 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 (198,199). To establish the potential chemopreventive properties of

antioxidants, investigators have used markers such as 8-hydroxy-2'-deoxyguanosine (8-OHdG) and the comet assay as intermediate markers in interventions (198). A review of these intervention studies has concluded that most had extremely low statistical power (sample size usually ≤ 20) and that they led to modest changes in 8-OHdG (around 10%) (199,200). In conclusion, promising markers are available for intervention studies, but they await application in large-scale and well designed trials.

A randomized clinical trial of vitamins E and C in smokers found that among women, but not men, there was a significant decline in PAH-DNA adducts in the treatment group (201).

New biomarkers for clinical purposes

The field of biomarkers is rapidly expanding, particularly as far as biomarkers for clinical purposes are concerned. For example, microRNAs, which are very short stretches of RNA with regulatory functions, seem to be extremely promising for the understanding of cancer mechanisms, as well as for developing new therapies (202). In addition, microRNAs are also relevant to chemically-induced cancer (203). More about such new developments can be found in (204).

Policy changes

With regard to public health and environmental policy, molecular epidemiology has not yet led to broad policy changes to prevent or to reduce exposure to carcinogens. However, it has provided impetus for prevention of prolonged exposures to carcinogens, even at low levels, since they can result in DNA damage or epigenetic alterations that begin at a very early age, even *in utero*, and accumulate over a lifetime (41,205,206). In addition, molecular epidemiologic data on interindividual variation in susceptibility refute the default assumption in risk assessment that the population is biologically homogeneous in response to carcinogens. This default assumption can lead to substantial underestimates of risk to the population and to sensitive subgroups, leading to standards and policies that are not adequately health-protective or equitable (129,207).

The theoretical importance of focusing intervention strategies (regulations, public education programmes, health surveillance, behaviour modification, and chemoprevention programmes) on the subgroups at greatest risk as a result of genetic or acquired susceptibility (208,209) is illustrated in Figure 19.2 (210). The figure illustrates that while the distribution of susceptibility/risk is symmetrical on a log scale, it is right-skewed on the linear scale. Thus, for a hypothetical carcinogen with a linear low dose–response curve, the estimated risk would be 38-fold greater for a population of individuals with 99th-percentile sensitivity than for a population of median-sensitive individuals. (This number is the upper 95% confidence limit of risk with respect to uncertainty; the estimated increase in risk is similar if the arithmetic mean estimate of risk with respect to uncertainty is used.) Sensitivity due to genetic and nutritional factors can be compounded in the case of certain groups (e.g. children) who would be expected to have both more exposure and greater age-related susceptibility to certain carcinogens.

Figure 19.2. The theoretical distribution of cancer susceptibility and risk across a population that is heterogeneous with respect to sensitivity to a hypothetical non-threshold carcinogen [based on (200)]. The x-axis represents the percentile of sensitivity; the y-axis, the number of individuals. The numbers in parentheses are the estimated cancer risk for a population of individuals at the indicated percentile of sensitivity. They are derived by use of a Monte Carlo simulation using data on observed human variability in metabolic activation, detoxification and DNA repair, as well as uncertainty in cancer potencies for a set of genetically acting carcinogens. The numbers shown are the upper 95% confidence limit of risk with respect to uncertainty estimates and are similar if the arithmetic mean estimates are used. Panel [a] shows the distribution on a log scale; panel [b] shows the same distribution on a linear scale.

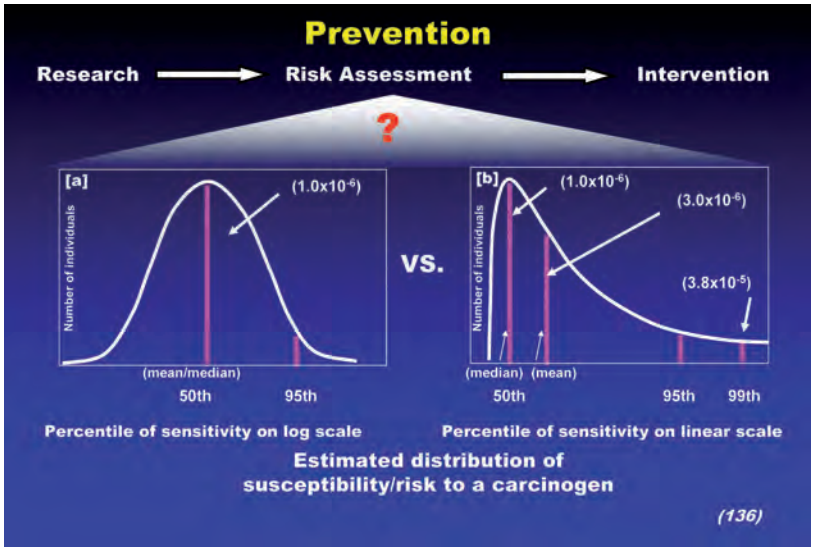


Table 19.1. Discoveries that support the original model of molecular epidemiology*

Marker linked to exposure or disease	Exposure	Reference
Metabolites in body fluids	Nitroso compounds in tobacco	(211,212)
Urinary metabolites (NNK, NNN)		
Exposure/biologically effective dose		
DNA adducts	PAHs, aromatic compounds	(37)
Albumin adducts	AFB_1	(213,214)
Haemoglobin adducts	Acrylamide	(215)
	Styrene	(216)
	1,3-butadiene	(217)
Preclinical effect	**Exposure and/or Cancer**	(52)
Chromosome aberrations	Lung	(218)
	Leukemia	(219)
	Benzene	
HRPT	PAHs	(220)
	1,3-butadiene	(221)
Glycophorin A	PAHs	(222)
Gene expression	Cisplatin	(223)
Genetic susceptibility	e.g. DNA repair capacity in head and neck cancer	(77,224)
Phenotypic markers		
SNPs:		
NAT2, GSTM	Bladder	(225)
CYP1A1	Lung	(226)

*See (2) and (227).

Conclusions and future directions

The examples presented in this chapter show that molecular epidemiology has made extensive progress since the 1980s. It has contributed to prevention by providing new evidence that specific environmental agents pose human carcinogenic hazards, helping to establish their causal role, identifying subsets of the population at special risk, and using this information to develop new and more effective strategies to reduce risk. As a result, some interventions and policy changes have been mounted to reduce risk from several important environmental carcinogens.

As has been seen, recently developed epigenetic and "omic" biomarkers have considerable potential in molecular epidemiology, along with genotoxic markers, because they reflect another equally important mechanism of carcinogenicity: epigenetic alterations that affect the expression of genes and proteins. These can be measured by high-throughput methods, allowing large-scale studies that are discovery-oriented. However, a major challenge is the need for validation of these newer biomarkers so they may be applied in large-scale etiologic and intervention studies. An important development in molecular epidemiology has been the emergence of networks and consortia involving hundreds of researchers and multiple large studies. Examples include the Wellcome Trust Case-Control Consortium, CGEMS (Cancer Genetic Markers of Susceptibility), HuGE (Human Genome Epidemiology Network), ECNIS (Environmental Cancer Risk, Nutrition and Individual Susceptibility), NuGO (The European Nutrigenomics Organization linking genomics, nutrition and health research), and Interlymph in the field of lymphomas. Such initiatives allow coordinated efforts, avoid false-positives and publication bias from several small studies, and contribute to rapid dissemination and replication of new knowledge.

Another challenge and future direction is the timely translation of

data from etiologic and intervention studies into risk assessment and public health policy, as well as focused research to fill gaps in scientific knowledge. Meeting these goals requires an infrastructure to promote a dialogue among scientists, policy-makers and other stakeholders, and a major investment in the second generation of molecular epidemiologic research, including large-scale collaborative studies incorporating validated biomarkers and automated technologies.

Acknowledgements

This work has been made possible by grants from the Compagnia di San Paolo to the ISI Foundation and from the European Union for the project ECNIS (PV), and by the National Institutes of Health, the National Institutes of Environmental Health Sciences (5P01ES09600, 5RO1ES08977, 1R01CA127532, R01CA69094), the US. Environmental Protection Agency (R827027, RD-832141), Bauman Family Foundation, and the New York Community Trust.

References

1. American Cancer Society. Cancer facts and figures 2006. Atlanta (GA): American Cancer Society; 2006.

2. Perera FP, Weinstein IB (1982). Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. *J Chronic Dis,* 35:581–600.doi:10.1016/0021-9681(82)90078-9 PMID:6282919

3. National Academy of Sciences. Regulating pesticides in food: the Delaney Paradox. Washington (DC): National Academy Press; 1987.

4. Weinberg RA (2006). A lost generation. *Cell,* 126:9–10.doi:10.1016/j.cell.2006.06.022 PMID:16839866

5. Chiazze L Jr, Levin DL, Silverman DT. Recent changes in estimated cancer mortality. In: Hiatt HH, Watson JD, Winsten JA, editors. Origins of human cancer. Cold Spring Harbor conferences on cell proliferation, vol 4. Maine: Cold Spring Harbor Laboratory; 1977. p. 33–44.

6. Weinstein IB, Santella RM, Perera FP *et al.* The molecular biology and molecular epidemiology of cancer. The science and practice of cancer prevention and control. New York (NY): Marcel Dekker; 1993.

7. Tomatis L, Aitio A, Wilbourn J, Shukar L (1987). Human carcinogens so far identified. *Jpn J Cancer Res,* 78:887–898. PMID:2513295

8. Willett WC (2002). Balancing life-style and genomics research for disease prevention. *Science*, 296:695–698.doi:10.1126/science.1071055 PMID:11976443

9. U.S. Department of Health, Education and Welfare. Smoking and health. Report of the Advisory Committee to the Surgeon General of the Public Health Service. Washington (DC): Public Health Service; 1964.

10. Vineis P (2004). Individual susceptibility to carcinogens. *Oncogene*, 23:6477–6483. doi:10.1038/sj.onc.1207897 PMID:15322518

11. Suuriniemi M, Agalliu I, Schaid DJ *et al.* (2007). Confirmation of a positive association between prostate cancer risk and a locus at chromosome 8q24. *Cancer Epidemiol Biomarkers Prev,* 16:809–814.doi:10.1158/1055-9965.EPI-06-1049 PMID:17416775

12. Zhang ZF, Cordon-Cardo C, Rothman N *et al.* Methodological issues in the use of tumour markers in cancer epidemiology. In: Toniolo P, Boffetta P, Shuker DEG *et al.*, editors. Application of biomarkers in cancer epidemiology. Lyon: IARC Scientific Publication; 1997(142). p. 201–213.

13. National Research Council. Environmental tobacco smoke: measuring exposures and assessing health effects. Washington (DC): National Academy Press; 1986.

14. Hecht SS (1999). Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst,* 91:1194–1210.doi:10.1093/jnci/91.14.1194 PMID:10413421

15. Wogan GN, Hecht SS, Felton JS *et al.* (2004). Environmental and chemical carcinogenesis. *Semin Cancer Biol,* 14:473–486.doi:10.1016/j.semcancer.2004.06.010 PMID:15489140

16. Bryant MS, Skipper PL, Tannenbaum SR, Maclure M (1987). Hemoglobin adducts of 4-aminobiphenyl in smokers and nonsmokers. *Cancer Res,* 47:602–608. PMID:3791245

17. International Agency for Research on Cancer. Polynuclear aromatic compounds. Part I. Chemical, environmental, and experimental data. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Lyon: IARC Scientific Publication; 1983.

18. U.S. Environmental Protection Agency. Aerometric information retrieval system (AIRS), data for 1985–1990. Research Triangle Park (NC): U.S. Environmental Protection Agency; 1990.

19. Anderson LM, Jones AB, Miller MS, Chauhan DP. Metabolism of transplacental carcinogens. In Napalkov NP, Rice JM, Tomatis L, Yamasaki H, editors. Perinatal and multigeneration carcinogenesis. Lyon: IARC Scientific Publication; 1989(96). p. 155–188.

20. Boström CE, Gerde P, Hanberg A *et al.* (2002). Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environ Health Perspect,* 110 Suppl 3;451–488. PMID:12060843

21. Anderson LM, Ruskie S, Carter J *et al.* (1995). Fetal mouse susceptibility to transplacental carcinogenesis: differential influence of Ah receptor phenotype on effects of 3-methylcholanthrene, 12-dimethylbenz[a]anthracene, and benzo[a]pyrene. *Pharmacogenetics,* 5:364–372. doi:10.1097/00008571-199512000-00005 PMID:8747408

22. Yu Z, Loehr CV, Fischer KA *et al.* (2006). In utero exposure of mice to dibenzo[a,l] pyrene produces lymphoma in the offspring: role of the aryl hydrocarbon receptor. *Cancer Res,* 66:755–762.doi:10.1158/0008-5472.CAN-05-3390 PMID:16424006

23. Miller EC, Miller JA (1981). Mechanisms of chemical carcinogenesis. *Cancer*, 47 Suppl;1055–1064.doi:10.1002/1097-0142(19810301)47:5+<1055::AID-CNCR2820471302>3.0.CO;2-3 PMID:7016297

24. Perera FP (1997). Environment and cancer: who are susceptible? *Science*, 278:1068–1073.doi:10.1126/science.278.5340.1068 PMID:9353182

25. Perera FP, Poirier MC, Yuspa SH *et al.* (1982). A pilot project in molecular cancer epidemiology: determination of benzo[a] pyrene–DNA adducts in animal and human tissues by immunoassays. *Carcinogenesis*, 3:1405–1410.doi:10.1093/carcin/3.12.1405 PMID:6295657

26. Perera FP, Santella RM, Brenner D *et al.* (1987). DNA adducts, protein adducts, and sister chromatid exchange in cigarette smokers and nonsmokers. *J Natl Cancer Inst,* 79:449–456. PMID:3114532

27. Mumford JL, Lee X, Lewtas J *et al.* (1993). DNA adducts as biomarkers for assessing exposure to polycyclic aromatic hydrocarbons in tissues from Xuan Wei women with high exposure to coal combustion emissions and high lung cancer mortality. *Environ Health Perspect,* 99:83–87.doi:10.1289/ehp.939983 PMID:8319664

28. Binková B, Lewtas J, Mísková I *et al.* (1996). Biomarker studies in northern Bohemia. *Environ Health Perspect,* 104 Suppl 3;591–597.doi:10.2307/3432828 PMID:8781388

29. Perera FP, Hemminki K, Gryzbowska E *et al.* (1992). Molecular and genetic damage in humans from environmental pollution in Poland. *Nature*, 360:256–258. doi:10.1038/360256a0 PMID:1436106

30. Dickey C, Santella RM, Hattis D *et al.* (1997). Variability in PAH-DNA adduct measurements in peripheral mononuclear cells: implications for quantitative cancer risk assessment. *Risk Anal,* 17:649–656. doi:10.1111/j.1539-6924.1997.tb00905.x PMID:9404054

31. Kriek E, Van Schooten FJ, Hillebrand MJ *et al.* (1993). DNA adducts as a measure of lung cancer risk in humans exposed to polycyclic aromatic hydrocarbons. *Environ Health Perspect,* 99:71–75.doi:10.1289/ehp.939971 PMID:8319662

32. Tang D, Santella RM, Blackwood AM *et al.* (1995). A molecular epidemiological case-control study of lung cancer. *Cancer Epidemiol Biomarkers Prev*, 4:341–346. PMID:7655328

33. Bartsch H, Rojas M, Alexandrov K, Risch A (1998). Impact of adduct determination on the assessment of cancer susceptibility. *Recent Results Cancer Res,* 154:86–96. PMID:10026994

34. Ambrosone CB, Kadlubar FF (1997). Toward an integrated approach to molecular epidemiology. *Am J Epidemiol,* 146:912–918. PMID:9400332

35. Stowers SJ, Anderson MW (1984). Ubiquitous binding of benzo[a]pyrene metabolites to DNA and protein in tissues of the mouse and rabbit. *Chem Biol Interact,* 51:151–166.doi:10.1016/0009-2797(84)90027-9 PMID:6088095

36. Wiencke JK, Thurston SW, Kelsey KT *et al.* (1999). Early age at smoking initiation and tobacco carcinogen DNA damage in the lung. *J Natl Cancer Inst,* 91:614–619.doi:10.1093/jnci/91.7.614 PMID:10203280

37. Tang D, Phillips DH, Stampfer M *et al.* (2001). Association between carcinogen-DNA adducts in white blood cells and lung cancer risk in the physicians health study. *Cancer Res,* 61:6708–6712. PMID:11559540

38. Veglia F, Matullo G, Vineis P (2003). Bulky DNA adducts and risk of cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev,* 12:157–160. PMID:12582026

39. Kellermann G, Shaw CR, Luyten-Kellerman M (1973). Aryl hydrocarbon hydroxylase inducibility and bronchogenic carcinoma. *N Engl J Med,* 289:934–937. doi:10.1056/NEJM197311012891802 PMID:4126515

40. Vineis P, Malats N, Porta M, Real FX (1999). Human cancer, carcinogenic exposures and mutation spectra. *Mutat Res,* 436:185–194. PMID:10095140

41. Greenblatt MS, Bennett WP, Hollstein M, Harris CC (1994). Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res,* 54:4855–4878. PMID:8069852

42. Denissenko MF, Pao A, Tang MS, Pfeifer GP (1996). Preferential formation of benzo[a] pyrene adducts at lung cancer mutational hotspots in P53. *Science*, 274:430–432. doi:10.1126/science.274.5286.430 PMID:8832894

43. Hainaut P, Hollstein M (2000). p53 and human cancer: the first ten thousand mutations. *Adv Cancer Res,* 77:81–137. doi:10.1016/S0065-230X(08)60785-X PMID:10549356

44. Wild CP, Hasegawa R, Barraud L *et al.* (1996). Aflatoxin-albumin adducts: a basis for comparative carcinogenesis between animals and humans. *Cancer Epidemiol Biomarkers Prev,* 5:179–189. PMID:8833618

45. Ross RK, Yuan JM, Yu MC *et al.* (1992). Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet*, 339:943–946.doi:10.1016/0140-6736(92)91528-G PMID:1348796

46. Qian GS, Ross RK, Yu MC *et al.* (1994). A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol Biomarkers Prev,* 3:3–10. PMID:8118382

47. Wang LY, Hatch M, Chen CJ *et al.* (1996). Aflatoxin exposure and risk of hepatocellular carcinoma in Taiwan. *Int J Cancer,* 67:620–625.doi:10.1002/(SICI)1097-0215(19960904)67:5<620::AID-IJC5>3.0.CO;2-W PMID:8782648

48. Yu MW, Lien JP, Chiu YH *et al.* (1997). Effect of aflatoxin metabolism and DNA adduct formation on hepatocellular carcinoma among chronic hepatitis B carriers in Taiwan. *J Hepatol,* 27:320–330.doi:10.1016/S0168-8278(97)80178-X PMID:9288607

49. Hsia CC, Kleiner DE Jr, Axiotis CA *et al.* (1992). Mutations of p53 gene in hepatocellular carcinoma: roles of hepatitis B virus and aflatoxin contamination in the diet. *J Natl Cancer Inst,* 84:1638–1641.doi:10.1093/jnci/84.21.1638 PMID:1279184

50. Smith MT, Zhang L (1998). Biomarkers of leukemia risk: benzene as a model. *Environ Health Perspect,* 106 Suppl 4;937–946. doi:10.2307/3434135 PMID:9703476

51. Kim S, Vermeulen R, Waidyanatha S *et al.* (2006). Modeling human metabolism of benzene following occupational and environmental exposures. *Cancer Epidemiol Biomarkers Prev,* 15:2246–2252. doi:10.1158/1055-9965.EPI-06-0262 PMID:17119053

52. Bonassi S, Znaor A, Norppa H, Hagmar L (2004). Chromosomal aberrations and risk of cancer in humans: an epidemiologic perspective. *Cytogenet Genome Res,* 104:376–382.doi:10.1159/000077519 PMID:15162068

53. Bonassi S, Hagmar L, Strömberg U *et al.*; European Study Group on Cytogenetic Biomarkers and Health (2000). Chromosomal aberrations in lymphocytes predict human cancer independently of exposure to carcinogens. *Cancer Res,* 60:1619–1625. PMID:10749131

54. Hagmar L, Bonassi S, Strömberg U *et al.* (1998). Chromosomal aberrations in lymphocytes predict human cancer: a report from the European Study Group on Cytogenetic Biomarkers and Health (ESCH). *Cancer Res,* 58:4117–4121. PMID:9751622

55. Zhang YJ, Chen Y, Ahsan H *et al.* (2005). Silencing of glutathione S-transferase P1 by promoter hypermethylation and its relationship to environmental chemical carcinogens in hepatocellular carcinoma. *Cancer Lett,* 221:135–143.doi:10.1016/j.canlet.2004.08.028 PMID:15808399

56. Vermeulen R, Lan Q, Zhang L *et al.* (2005). Decreased levels of CXC-chemokines in serum of benzene-exposed workers identified by array-based proteomics. *Proc Natl Acad Sci USA*, 102:17041–17046.doi:10.1073/pnas.0508573102 PMID:16286641

57. Lan Q, Zheng T, Rothman N *et al.* (2006). Cytokine polymorphisms in the Th1/Th2 pathway and susceptibility to non-Hodgkin lymphoma. *Blood*, 107:4101–4108.doi:10.1182/blood-2005-10-4160 PMID:16449530

58. Steffen C, Auclerc MF, Auvrignon A *et al.* (2004). Acute childhood leukaemia and environmental exposure to potential sources of benzene and other hydrocarbons; a case-control study. *Occup Environ Med,* 61:773–778.doi:10.1136/oem.2003.010868 PMID:15317919

59. Feingold L, Savitz DA, John EM (1992). Use of a job-exposure matrix to evaluate parental occupation and childhood cancer. *Cancer Causes Control,* 3:161–169.doi:10.1007/BF00051656 PMID:1562706

60. Lahmann PH, Friedenreich C, Schuit AJ *et al.* (2007). Physical activity and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev,* 16:36–42.doi:10.1158/1055-9965.EPI-06-0582 PMID:17179488

61. Pischon T, Lahmann PH, Boeing H *et al.* (2006). Body size and risk of renal cell carcinoma in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Int J Cancer,* 118:728–738.doi:10.1002/ijc.21398 PMID:16094628

62. Pischon T, Lahmann PH, Boeing H *et al.* (2006). Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst,* 98:920–931.doi:10.1093/jnci/djj246 PMID:16818856

63. Friedenreich C, Cust A, Lahmann PH *et al.* (2007). Anthropometric factors and risk of endometrial cancer: the European prospective investigation into cancer and nutrition. *Cancer Causes Control,* 18:399–413.doi:10.1007/s105 52-006-0113-8 PMID:17297555

64. Russo A, Autelitano M, Bisanti L (2008). Metabolic syndrome and cancer risk. *Eur J Cancer,* 44:293–297.doi:10.1016/j.ejca.2007.11.005 PMID:18055193

65. Hursting SD, Lavigne JA, Berrigan D *et al.* (2004). Diet-gene interactions in p53-deficient mice: insulin-like growth factor-1 as a mechanistic target. *J Nutr,* 134:2482S–2486S. PMID:15333746

66. Jenab M, Riboli E, Cleveland RJ *et al.* (2007). Serum C-peptide, IGFBP-1 and IGFBP-2 and risk of colon and rectal cancers in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer,* 121:368–376.doi:10.1002/ijc.22697 PMID:17372899

67. Weiss JM, Huang WY, Rinaldi S *et al.* (2007). IGF-1 and IGFBP-3: risk of prostate cancer among men in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Int J Cancer,* 121:2267–2273.doi:10.1002/ijc. 22921 PMID:17597108

68. Allen NE, Key TJ, Appleby PN *et al.* (2007). Serum insulin-like growth factor (IGF)-I and IGF-binding protein-3 concentrations and prostate cancer risk: results from the European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev,* 16:1121–1127.doi:10.1158/1055-9965.EPI-06-1062 PMID:17548673

69. Skibola CF, Holly EA, Forrest MS *et al.* (2004). Body mass index, leptin and leptin receptor polymorphisms, and non-hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev*, 13:779–786. PMID:15159310

70. DeLellis Henderson K, Rinaldi S, Kaaks R *et al.* (2007). Lifestyle and dietary correlates of plasma insulin-like growth factor binding protein-1 (IGFBP-1), leptin, and C-peptide: the Multiethnic Cohort. *Nutr Cancer*, 58:136–145. PMID:17640159

71. Hour TC, Pu YS, Lin CC *et al.* (2006). Differential expression of molecular markers in arsenic- and non-arsenic-related urothelial cancer. *Anticancer Res,* 26 1A;375–378. PMID:16475721

72. Zhang ZF, Cordon-Cardo C, Rothman N *et al.* Methodological issues in the use of tumour markers in cancer epidemiology. In: Toniolo P, Boffetta P, Shuker DEG *et al.*, editors. Application of biomarkers in cancer epidemiology.Lyon: IARC Scientific Publication; 1997(142). p. 201–213.

73. Garte S, Zocchetti C, Taioli E. Gene-environment interactions in the application of biomarkers of cancer susceptibility in epidemiology. In: Toniolo P, Boffetta P, Shuker DEG *et al.*, editors. Application of biomarkers in cancer epidemiology. Lyon: IARC Scientific Publication; 1997. p. 251–264.

74. Caporaso NE, Goldstein A. Issues involving biomarkers in the study of the genetics of human cancer. In: Toniolo P, Boffetta P, Shuker DEG *et al.*, editors. Application of biomarkers in cancer epidemiology. Lyon: IARC Scientific Publication; 1997. p. 237–250.

75. Schulte PA, Hunter D, Rothman N. Ethical and social issues in the use of biomarkers in epidemiological studies. In: Toniolo P, Boffetta P, Shuker DEG *et al.*, editors. Application of biomarkers in cancer epidemiology. Lyon: IARC Scientific Publication; 1997. p. 313–318.

76. Vogelstein B, Kinzler KW, editors. The genetic basis of human cancer. New York (NY): McGraw-Hill; 1998.

77. Berwick M, Vineis P (2000). Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst,* 92:874–897.doi:10.1093/jnci/92.11.874 PMID:10841823

78. Matullo G, Dunning AM, Guarrera S *et al.* (2006). DNA repair polymorphisms and cancer risk in non-smokers in a cohort study. *Carcinogenesis*, 27:997–1007.doi:10.1093/carcin/bgi280 PMID:16308313

79. Sharp L, Little J (2004). Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. *Am J Epidemiol,* 159:423–443.doi:10.1093/aje/kwh066 PMID:14977639

80. Eichholzer M, Lüthy J, Moser U, Fowler B (2001). Folate and the risk of colorectal, breast and cervix cancer: the epidemiological evidence. *Swiss Med Wkly,* 131:539–549. PMID:11759174

81. Kim SJ, Kim TW, Lee SY *et al.* (2004). CpG methylation of the ERalpha and ERbeta genes in breast cancer. *Int J Mol Med,* 14:289–293. PMID:15254780

82. Little J, Sharp L, Duthie S, Narayanan S (2003). Colon cancer and genetic variation in folate metabolism: the clinical bottom line. *J Nutr,* 133 Suppl 1;3758S–3766S. PMID:14608111

83. Zhang SM, Willett WC, Selhub J *et al.* (2003). Plasma folate, vitamin B6, vitamin B12, homocysteine, and risk of breast cancer. *J Natl Cancer Inst,* 95:373–380.doi:10.1093/jnci/95.5.373 PMID:12618502

84. Giovannucci E, Chen J, Smith-Warner SA *et al.* (2003). Methylenetetrahydrofolate reductase, alcohol dehydrogenase, diet, and risk of colorectal adenomas. *Cancer Epidemiol Biomarkers Prev,* 12:970–979. PMID:14578131

85. Terry P, Jain M, Miller AB *et al.* (2002). Dietary intake of folic acid and colorectal cancer risk in a cohort of women. *Int J Cancer,* 97:864–867.doi:10.1002/ijc.10138 PMID:11857369

86. Harnack L, Jacobs DR Jr, Nicodemus K *et al.* (2002). Relationship of folate, vitamin B-6, vitamin B-12, and methionine intake to incidence of colorectal cancers. *Nutr Cancer,* 43:152–158.doi:10.1207/S15327914NC432_5 PMID:12588695

87. Zienolddiny S, Campa D, Lind H *et al.* (2006). Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis*, 27:560–567.doi:10.1093/carcin/bgi232 PMID:16195237

88. Leng S, Cheng J, Pan Z *et al.* (2004). Associations between XRCC1 and ERCC2 polymorphisms and DNA damage in peripheral blood lymphocyte among coke oven workers. *Biomarkers*, 9:395–406.doi:10.1080/13547500400015618 PMID:15764301

89. Alexandrov K, Cascorbi I, Rojas M *et al.* (2002). CYP1A1 and GSTM1 genotypes affect benzo[a]pyrene DNA adducts in smokers' lung: comparison with aromatic/hydrophobic adduct formation. *Carcinogenesis*, 23:1969–1977.doi:10.1093/carcin/23.12.1969 PMID:12507920

90. Cascorbi I, Brockmöller J, Roots I (2002). Molecular-epidemiological aspects of carcinogenesis: the role of xenobiotic metabolizing enzymes. *Int J Clin Pharmacol Ther,* 40:562–563. PMID:12503814

91. Whyatt RM, Camann DE, Kinney PL *et al.* (2002). Residential pesticide use during pregnancy among a cohort of urban minority women. *Environ Health Perspect,* 110:507–514.doi:10.1289/ehp.02110507 PMID:12003754

92. Yu MW, Chiu YH, Chiang YC *et al.* (1999). Plasma carotenoids, glutathione S-transferase M1 and T1 genetic polymorphisms, and risk of hepatocellular carcinoma: independent and interactive effects. *Am J Epidemiol,* 149:621–629. PMID:10192309

93. McGlynn KA, Rosvold EA, Lustbader ED *et al.* (1995). Susceptibility to hepatocellular carcinoma is associated with genetic variation in the enzymatic detoxification of aflatoxin B1. *Proc Natl Acad Sci USA*, 92:2384–2387. doi:10.1073/pnas.92.6.2384 PMID:7892276

94. Yu MW, Gladek-Yarborough A, Chiamprasert S *et al.* (1995). Cytochrome P450 2E1 and glutathione S-transferase M1 polymorphisms and susceptibility to hepatocellular carcinoma. *Gastroenterology*, 109:1266–1273.doi:10.1016/0016-5085(95)90587-1 PMID:7557094

95. Ross D (1996). Metabolic basis of benzene toxicity. *Eur J Haematol Suppl,* 60:111–118. PMID:8987252

96. Rothman N, Smith MT, Hayes RB *et al.* (1997). Benzene poisoning, a risk factor for hematological malignancy, is associated with the NQO1 609C–>T mutation and rapid fractional excretion of chlorzoxazone. *Cancer Res*, 57:2839–2842. PMID:9230185

97. Vogelstein B, Kinzler KW (2004). Cancer genes and the pathways they control. *Nat Med,* 10:789–799.doi:10.1038/nm1087 PMID:15286780

98. Matullo G, Peluso M, Polidoro S *et al.* (2003). Combination of DNA repair gene single nucleotide polymorphisms and increased levels of DNA adducts in a population-based study. *Cancer Epidemiol Biomarkers Prev,* 12:674–677. PMID:12869411

99. Hung RJ, Hall J, Brennan P, Boffetta P (2005). Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. *Am J Epidemiol,* 162:925–942. doi:10.1093/aje/kwi318 PMID:16221808

100. Berwick M, Vineis P (2005). Measuring DNA repair capacity: small steps. *J Natl Cancer Inst,* 97:84–85.doi:10.1093/jnci/dji038 PMID:15657333

101. Ioannidis JP, Trikalinos TA (2005). Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol,* 58:543–549.doi:10.1016/j.jclinepi.2004.10.019 PMID:15878467

102. Vineis P, Manuguerra M, Kavvoura FK *et al.* (2009). A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility. *J Natl Cancer Inst,* 101:24–36. PMID:19116388

103. Zeggini E, Weedon MN, Lindgren CM *et al.*; Wellcome Trust Case Control Consortium (WTCCC) (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316:1336–1341.doi:10.1126/science.1142364 PMID:17463249

104. Haiman CA, Le Marchand L, Yamamato J *et al.* (2007). A common genetic risk factor for colorectal and prostate cancer. *Nat Genet,* 39:954–956.doi:10.1038/ng2098 PMID:17618282

105. Zanke BW, Greenwood CM, Rangrej J *et al.* (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet,* 39:989–994.doi:10.1038/ng2089 PMID:17618283

106. Chung CC, Magalhaes WC, Gonzalez-Bosquet J, Chanock SJ (2010). Genome-wide association studies in cancer–current and future directions. *Carcinogenesis*, 31:111–120. doi:10.1093/carcin/bgp273 PMID:19906782

107. Hunter DJ, Chanock SJ (2010). Genome-wide association studies and "the art of the soluble". *J Natl Cancer Inst,* 102:836–837. doi:10.1093/jnci/djq197 PMID:20505151

108. Manolio TA (2010). Genomewide association studies and assessment of the risk of disease. *N Engl J Med,* 363:166–176.doi:10.1056/NEJMra0905980 PMID:20647212

109. Amos C. Successful design and conduct of genome-wide association studies. *Hum Mol Genet* 2007;16(Spec No. 2):R220–R225.

110. Hecht SS, Carmella SG, Akerkar S, Richie JP Jr (1994). 4-(Methylnitrosamino)-1-(3,pyridyl)-1-butanol (NNAL) and its glucuronide, metabolites of a tobacco-specific lung carcinogen, in the urine of black and white smokers. *Proc Am Assoc Cancer Res*, 35:286.

111. Zang EA, Wynder EL (1996). Differences in lung cancer risk between men and women: examination of the evidence. *J Natl Cancer Inst,* 88:183–192.doi:10.1093/jnci/88.3-4.183 PMID:8632492

112. Hennekens CH, Buring JE, Manson JE *et al.* (1996). Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med,* 334:1145–1149.doi:10.1056/NEJM199605023341801 PMID:8602179

113. Ryberg D, Hewer A, Phillips DH, Haugen A (1994). Different susceptibility to smoking-induced DNA damage among male and female lung cancer patients. *Cancer Res,* 54:5801–5803. PMID:7954403

114. Kure EH, Ryberg D, Hewer A *et al.* (1996). p53 mutations in lung tumours: relationship to gender and lung DNA adduct levels. *Carcinogenesis*, 17:2201–2205.doi:10.1093/carcin/17.10.2201 PMID:8895489

115. Mollerup S, Ryberg D, Hewer A *et al.* (1999). Sex differences in lung CYP1A1 expression and DNA adduct levels among lung cancer patients. *Cancer Res,* 59:3317–3320. PMID:10416585

116. Tang DL, Rundle A, Warburton D *et al.* (1998). Associations between both genetic and environmental biomarkers and lung cancer: evidence of a greater risk of lung cancer in women smokers. *Carcinogenesis*, 19:1949–1953.doi:10.1093/carcin/19.11.1949 PMID:9855008

117. Block G (1993). Micronutrients and cancer: time for action? *J Natl Cancer Inst*, 85:846–848.doi:10.1093/jnci/85.11.846 PMID:8388060

118. Mooney LA, Perera FP (1996). Application of molecular epidemiology to lung cancer prevention. *J Cell Biochem,* 25 S25;63–68 doi:10.1002/(SICI)1097-4644(1996)25+<63::AID-JCB9>3.0.CO;2-0.

119. Grinberg-Funes RA, Singh VN, Perera FP *et al.* (1994). Polycyclic aromatic hydrocarbon-DNA adducts in smokers and their relationship to micronutrient levels and the glutathione-S-transferase M1 genotype. *Carcinogenesis*, 15:2449–2454.doi:10.1093/carcin/15.11.2449 PMID:7955090

120. Mooney LA, Bell DA, Santella RM *et al.* (1997). Contribution of genetic and nutritional factors to DNA damage in heavy smokers. *Carcinogenesis*, 18:503–509.doi:10.1093/carcin/18.3.503 PMID:9067549

121. Küçük O, Pung A, Franke AA *et al.* (1995). Correlations between mutagen sensitivity and plasma nutrient levels of healthy individuals. *Cancer Epidemiol Biomarkers Prev,* 4:217–221. PMID:7541679

122. National Research Council. Science and judgment in risk assessment. Washington (DC): National Academy Press; 1994.

123. Goldman LR (1995). Children–unique and vulnerable. Environmental risks facing children and recommendations for response. *Environ Health Perspect,* 103 Suppl 6;13–18. doi:10.2307/3432338 PMID:8549460

124. Perera FP (1996). Molecular epidemiology: insights into cancer susceptibility, risk assessment, and prevention. *J Natl Cancer Inst,* 88:496–509. doi:10.1093/jnci/88.8.496 PMID:8606378

125. Perera FP (2000). Molecular epidemiology: on the path to prevention? *J Natl Cancer Inst*, 92:602–612.doi:10.1093/jnci/92.8.602 PMID:10772677

126. Birnbaum LS, Fenton SE (2003). Cancer and developmental exposure to endocrine disruptors. *Environ Health Perspect*, 111:389–394.doi:10.1289/ehp.5686 PMID:12676588

127. Miller RL, Garfinkel R, Horton M *et al.* (2004). Polycyclic aromatic hydrocarbons, environmental tobacco smoke, and respiratory symptoms in an inner-city birth cohort. *Chest,* 126:1071–1078.doi:10.1378/chest.126.4.1071 PMID:15486366

128. Perera FP, Tang D, Tu YH *et al.* (2004). Biomarkers in maternal and newborn blood indicate heightened fetal susceptibility to procarcinogenic DNA damage. *Environ Health Perspect,* 112:1133–1136.doi:10.1289/ehp.6833 PMID:15238289

129. National Research Council. Pesticides in the diets of infants and children. Washington (DC): National Academy Press; 1993.

130. Perera FP, Tang D, Whyatt RM, *et al.* (2004). Comparison of PAH-DNA adducts in four populations of mothers and newborns in the US, Poland and China. *Proc Am Assoc Cancer Res*, 45 (abstract #1975).

131. Whyatt RM, Jedrychowski W, Hemminki K *et al.* (2001). Biomarkers of polycyclic aromatic hydrocarbon-DNA damage and cigarette smoke exposures in paired maternal and newborn blood samples as a measure of differential susceptibility. *Cancer Epidemiol Biomarkers Prev,* 10:581–588. PMID:11401906

132. Bocskay KA, Tang D, Orjuela MA *et al.* (2005). Chromosomal aberrations in cord blood are associated with prenatal exposure to carcinogenic polycyclic aromatic hydrocarbons. *Cancer Epidemiol Biomarkers Prev,* 14:506–511.doi:10.1158/1055-9965.EPI-04-0566 PMID:15734979

133. Buffler PA, Kwan ML, Reynolds P, Urayama KY (2005). Environmental and genetic risk factors for childhood leukemia: appraising the evidence. *Cancer Invest,* 23:60–75. PMID:15779869

134. Neri M, Ugolini D, Bonassi S *et al.* (2006). Children's exposure to environmental pollutants and biomarkers of genetic damage. II. Results of a comprehensive literature search and meta-analysis. *Mutat Res,* 612:14–39.doi:10.1016/j.mrrev.2005.04.003 PMID:16027031

135. Pluth JM, Ramsey MJ, Tucker JD (2000). Role of maternal exposures and newborn genotypes on newborn chromosome aberration frequencies. *Mutat Res,* 465:101–111. PMID:10708975

136. Sorahan T, Prior P, Lancashire RJ *et al.* (1997). Childhood cancer and parental use of tobacco: deaths from 1971 to 1976. *Br J Cancer,* 76:1525–1531.doi:10.1038/bjc.1997.589 PMID:9400953

137. Finette BA, O'Neill JP, Vacek PM, Albertini RJ (1998). Gene mutations with characteristic deletions in cord blood T lymphocytes associated with passive maternal exposure to tobacco smoke. *Nat Med,* 4:1144–1151.doi:10.1038/2640 PMID:9771747

138. Boffetta P (2000). Molecular epidemiology. *J Intern Med,* 248:447–454.doi:10.1046/j.1365-2796.2000.00777.x PMID:11155137

139. Janerich DT, Thompson WD, Varela LR *et al.* (1990). Lung cancer and exposure to tobacco smoke in the household. *N Engl J Med,* 323:632–636.doi:10.1056/NEJM199009063231003 PMID:2385268

140. Vineis P, Airoldi L, Veglia F *et al.* (2005). Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ,* 330:277–280.doi:10.1136/bmj.38327.648472.82 PMID:15681570

141. Gale KB, Ford AM, Repp R *et al.* (1997). Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci USA,* 94:13950–13954.doi:10.1073/pnas.94.25.13950 PMID:9391133

142. Wiemels JL, Greaves M (1999). Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res,* 59:4075–4082. PMID:10463610

143. Shimizu Y, Kato H, Schull WJ (1991). Risk of cancer among atomic bomb survivors. *J Radiat Res (Tokyo),* 32 Suppl 2;54–63.doi:10.1269/jrr.32.SUPPLEMENT2_54 PMID:1823367

144. Hegmann KT, Fraser AM, Keaney RP *et al.* (1993). The effect of age at smoking initiation on lung cancer risk. *Epidemiology*, 4:444–448.doi:10.1097/00001648-199309000-00010 PMID:8399693

145. Moore CJ, Tricomi WA, Gould MN (1986). Interspecies comparison of polycyclic aromatic hydrocarbon metabolism in human and rat mammary epithelial cells. *Cancer Res,* 46:4946–4952. PMID:3093058

146. Palmer JR, Rosenberg L (1993). Cigarette smoking and the risk of breast cancer. *Epidemiol Rev,* 15:145–156. PMID:8405197

147. Potter JD (2003). Epidemiology, cancer genetics and microarrays: making correct inferences, using appropriate designs. *Trends Genet,* 19:690–695.doi:10.1016/j.tig.2003.10.005 PMID:14642749

148. Potter JD (2001). At the interfaces of epidemiology, genetics and genomics. *Nat Rev Genet,* 2:142–147.doi:10.1038/35052575 PMID:11253054

149. Kim YI (2004). Folate and DNA methylation: a mechanistic link between folate deficiency and colorectal cancer? *Cancer Epidemiol Biomarkers Prev,* 13:511–519. PMID:15066913

150. Forozan F, Karhu R, Kononen J *et al.* (1997). Genome screening by comparative genomic hybridization. *Trends Genet,* 13:405–409.doi:10.1016/S0168-9525(97)01244-4 PMID:9351342

151. Robertson KD, Wolffe AP (2000). DNA methylation in health and disease. *Nat Rev Genet,* 1:11–19.doi:10.1038/35049533 PMID:11262868

152. Alberg AJ, Brock MV, Samet JM (2005). Epidemiology of lung cancer: looking to the future. *J Clin Oncol,* 23:3175–3185.doi:10.1200/JCO.2005.10.462 PMID:15886304

153. Brena RM, Huang TH, Plass C (2006). Quantitative assessment of DNA methylation: Potential applications for disease diagnosis, classification, and prognosis in clinical settings. *J Mol Med,* 84:365–377.doi:10.1007/s00109-005-0034-0 PMID:16416310

154. Toyooka S, Tokumo M, Shigematsu H *et al.* (2006). Mutational and epigenetic evidence for independent pathways for lung adenocarcinomas arising in smokers and never smokers. *Cancer Res,* 66:1371–1375.doi:10.1158/0008-5472.CAN-05-2625 PMID:16452191

155. Belinsky SA (2005). Silencing of genes by promoter hypermethylation: key event in rodent and human lung cancer. *Carcinogenesis*, 26:1481–1487.doi:10.1093/carcin/bgi020 PMID:15661809

156. Bearzatto A, Conte D, Frattini M *et al.* (2002). p16(INK4A) Hypermethylation detected by fluorescent methylation-specific PCR in plasmas from non-small cell lung cancer. *Clin Cancer Res,* 8:3782–3787. PMID:12473590

157. Russo AL, Thiagalingam A, Pan H *et al.* (2005). Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer. *Clin Cancer Res*, 11:2466–2470.doi:10.1158/1078-0432.CCR-04-1962 PMID:15814621

158. Vineis P, Husgafvel-Pursiainen K (2005). Air pollution and cancer: biomarker studies in human populations. *Carcinogenesis*, 26:1846–1855.doi:10.1093/carcin/bgi216 PMID:16123121

159. Honorio S, Agathanggelou A, Schuermann M *et al.* (2003). Detection of RASSF1A aberrant promoter hypermethylation in sputum from chronic smokers and ductal carcinoma in situ from breast cancer patients. *Oncogene*, 22:147–150.doi:10.1038/sj.onc.1206057 PMID:12527916

160. Belinsky SA, Liechty KC, Gentry FD *et al.* (2006). Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res,* 66:3338–3344.doi:10.1158/0008-5472.CAN-05-3408 PMID:16540689

161. Shen L, Issa JP (2002). Epigenetics in colorectal cancer. *Curr Opin Gastroenterol,* 18:68–73.doi:10.1097/00001574-200201000-00012 PMID:17031233

162. Lindon JC, Holmes E, Bollard ME *et al.* (2004). Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers*, 9:1–31.doi:10.1080/13547500410001668379 PMID:15204308

163. Lindon JC, Holmes E, Nicholson JK (2006). Metabonomics techniques and applications to pharmaceutical research & development. *Pharm Res,* 23:1075–1088.doi:10.1007/s11095-006-0025-z PMID:16715371

164. Yang SY, Xiao XY, Zhang WG *et al.* (2005). Application of serum SELDI proteomic patterns in diagnosis of lung cancer. *BMC Cancer,* 5:83.doi:10.1186/1471-2407-5-83 PMID:16029516

165. Odunsi K, Wollman RM, Ambrosone CB *et al.* (2005). Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. *Int J Cancer,* 113:782–788.doi:10.1002/ijc.20651 PMID:15499633

166. Dumas ME, Maibaum EC, Teague C *et al.* (2006). Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem,* 78:2199–2208.doi:10.1021/ac0517085 PMID:16579598

167. Esteller M (2006). The necessity of a human epigenome project. *Carcinogenesis*, 27:1121–1125.doi:10.1093/carcin/bgl033 PMID:16699174

168. Granville CA, Dennis PA (2005). An overview of lung cancer genomics and proteomics. *Am J Respir Cell Mol Biol,* 32:169–176.doi:10.1165/rcmb.F290 PMID:15713815

169. Chen J, Hunter DJ, Stampfer MJ *et al.* (2003). Polymorphism in the thymidylate synthase promoter enhancer region modifies the risk and survival of colorectal cancer. *Cancer Epidemiol Biomarkers Prev,* 12:958–962. PMID:14578129

170. Yanagisawa K, Shyr Y, Xu BJ *et al.* (2003). Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet*, 362:433–439.doi:10.1016/S0140-6736(03)14068-8 PMID:12927430

171. Dalle-Donne I, Scaloni A, Giustarini D *et al.* (2005). Proteins as biomarkers of oxidative/nitrosative stress in diseases: the contribution of redox proteomics. *Mass Spectrom Rev,* 24:55–99.doi:10.1002/mas.20006 PMID:15389864

172. Ambrosone CB, Rebbeck TR, Morgan GJ *et al.* (2006). New developments in the epidemiology of cancer prognosis: traditional and molecular predictors of treatment response and survival. *Cancer Epidemiol Biomarkers Prev,* 15:2042–2046. doi:10.1158/1055-9965.EPI-06-0827 PMID:17119026

173. Naas SJ, Moses HL. Cancer biomarkers: the promises and challenges of improving detection and treatment. Washington (DC): National Academies Press; 2007.

174. Ransohoff DF (2005). Lessons from controversy: ovarian cancer screening and serum proteomics. *J Natl Cancer Inst*, 97:315–319.doi:10.1093/jnci/dji054 PMID:15713968

175. García-Closas M, Vermeulen R, Sherman ME *et al.* Application of biomarkers in cancer epidemiology. In Schottenfeld D, Fraumeni JF Jr, editors. Cancer epidemiology and prevention. New York (NY): Oxford University Press; 2006.

176. Forrest MS, Lan Q, Hubbard AE *et al.* (2005). Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. *Environ Health Perspect,* 113:801–807.doi:10.1289/ehp.7635 PMID:15929907

177. Schulte PA, Perera FP, editors. Molecular epidemiology: principles and practices. New York (NY): Academic Press; 1993.

178. Stephan C, Stroebel G, Heinau M *et al.* (2005). 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,* 104:993–1003.doi:10.1002/cncr.21267 PMID:16007682

179. Rundle AG, Vineis P, Ahsan H (2005). Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiol Biomarkers Prev,* 14:1899–1907. doi:10.1158/1055-9965.EPI-04-0860 PMID:16103435

180. Gallicchio L, Helzlsouer KJ, Chow WH *et al.* (2010). Circulating 25-hydroxyvitamin D and the risk of rarer cancers: Design and methods of the Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. *Am J Epidemiol,* 172:10–20.doi:10.1093/aje/kwq116 PMID:20562188

181. Aguilar F, Harris CC, Sun T *et al.* (1994). Geographic variation of p53 mutational profile in nonmalignant human liver. *Science*, 264:1317–1319.doi:10.1126/science.8191284 PMID:8191284

182. Hoppin JA, Tolbert PE, Taylor JA *et al.* (2002). Potential for selection bias with tumor tissue retrieval in molecular epidemiology studies. *Ann Epidemiol,* 12:1–6.doi:10.1016/S1047-2797(01)00250-2 PMID:11750233

183. Porta M, Malats N, Corominas JM *et al.*; Pankras I Project Investigators (2002). Generalizing molecular results arising from incomplete biological samples: expected bias and unexpected findings. *Ann Epidemiol,* 12:7–14.doi:10.1016/S1047-2797(01)00267-8 PMID:11750234

184. Gormally E, Hainaut P, Caboux E *et al.* (2004). Amount of DNA in plasma and cancer risk: a prospective study. *Int J Cancer,* 111:746–749.doi:10.1002/ijc.20327 PMID:15252845

185. U.S. Department of Health and Human Services. The health consequences of involuntary exposure to tobacco smoke: a report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Washington (DC): U.S. Government Printing Office; 2006.

186. National Research Council. Issues in risk assessment. Washington (DC): National Academy Press; 1993.

187. Vineis P, Forastiere F, Hoek G, Lipsett M (2004). Outdoor air pollution and lung cancer: recent epidemiologic evidence. *Int J Cancer,* 111:647–652.doi:10.1002/ijc.20292 PMID:15252832

188. Armitage P, Doll R (1957). A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer,* 11:161–169.doi:10.1038/bjc.1957.22 PMID:13460138

189. Vogelstein B, Fearon ER, Hamilton SR *et al.* (1988). Genetic alterations during colorectal-tumor development. *N Engl J Med,* 319:525–532.doi:10.1056/NEJM198809013190901 PMID:2841597

190. Knudson AG (1996). Hereditary cancer: two hits revisited. *J Cancer Res Clin Oncol,* 122:135–140.doi:10.1007/BF01366952 PMID:8601560

191. Hashimoto T, Takahashi R, Yandell DW *et al.* (1991). Characterization of intragenic deletions in two sporadic germinal mutation cases of retinoblastoma resulting in abnormal gene expression. *Oncogene*, 6:463–469. PMID:2011402

192. Bielas JH, Loeb KR, Rubin BP *et al.* (2006). Human cancers express a mutator phenotype. *Proc Natl Acad Sci USA,* 103:18238–18242.doi:10.1073/pnas.0607057103 PMID:17108085

193. Benedict WF, Banerjee A, Kangalingam KK *et al.* (1984). Increased sister-chromatid exchange in bone-marrow cells of mice exposed to whole cigarette smoke. *Mutat Res,* 136:73–80.doi:10.1016/0165-1218(84)90136-8 PMID:6717473

194. Gatenby RA, Frieden BR (2004). Information dynamics in carcinogenesis and tumor growth. *Mutat Res,* 568:259–273. PMID:15542113

195. Albertini RJ, Ardell SK, Judice SA *et al.* (2000). Hypoxanthine-guanine phosphoribosyltransferase reporter gene mutation for analysis of in vivo clonal amplification in patients with HTLV type 1-associated Myelopathy/Tropical spastic paraparesis. *AIDS Res Hum Retroviruses,* 16:1747–1752. doi:10.1089/088922200501932 54 PMID:11080821

196. Mooney LA, Santella RM, Covey L *et al.* (1995). Decline of DNA damage and other biomarkers in peripheral blood following smoking cessation. *Cancer Epidemiol Biomarkers Prev,* 4:627–634. PMID:8547829

197. Hemminki K, Dickey CP, Karlsson S *et al.* (1997). Aromatic DNA adducts in foundry workers in relation to exposure, life style and CYP1A1 and glutathione transferase M1 genotype. *Carcinogenesis*, 18:345–350. doi:10.1093/carcin/18.2.345 PMID:9054627

198. Evans D, Sheares BJ, Vazquez TL (2004). Educating health professionals to improve quality of care for asthma. *Paediatr Respir Rev*, 5:304–310.doi:10.1016/j.prrv.2004.07.007 PMID:15531255

199. Møller P, Loft S (2002). Oxidative DNA damage in human white blood cells in dietary antioxidant intervention studies. *Am J Clin Nutr,* 76:303–310. PMID:12144999

200. Møller P, Loft S (2006). Dietary antioxidants and beneficial effect on oxidatively damaged DNA. *Free Radic Biol Med,* 41:388–415.doi:10.1016/j.freeradbiomed.2006.04.001 PMID:16843820

201. Mooney LA, Perera FP, Van Bennekum AM *et al.* (2001). Gender differences in autoantibodies to oxidative DNA base damage in cigarette smokers. *Cancer Epidemiol Biomarkers Prev*, 10:641–648. PMID:11401914

202. Nana-Sinkam P, Croce CM (2010). MicroRNAs in diagnosis and prognosis in cancer: what does the future hold? *Pharmacogenomics*, 11:667–669.doi:10.2217/pgs.10.57 PMID:20415558

203. Chen T (2010). The role of MicroRNA in chemical carcinogenesis. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev,* 28:89–124. PMID:20552498

204. García-Closas M, Vermeulen R, Sherman ME *et al.* Application of biomarkers in cancer epidemiology. In Schottenfeld D, Fraumeni JF Jr, editors. Cancer Epidemiology and Prevention. New York (NY): Oxford University Press; 2006. pp 70–88.

205. Perera FP, Whyatt RM, Jedrychowski W *et al.* (1998). Recent developments in molecular epidemiology: A study of the effects of environmental polycyclic aromatic hydrocarbons on birth outcomes in Poland. *Am J Epidemiol,* 147:309–314. PMID:9482506

206. Weinstein IB, Zhou P. Cell cycle control gene defects and human cancer. In Bertino JR, editor. Encyclopedia of cancer. Vol 1. New York (NY): Academic Press; 1997. p. 256–267.

207. Olden K, Poje J (1995). Environmental justice and environmental health. *Bull Soc Occup Environ Health,* 4:3–4.

208. Portier CJ, Bell DA (1998). Genetic susceptibility: significance in risk assessment. *Toxicol Lett,* 102-103:185–189.doi:10.1016/S0378-4274(98)00305-1 PMID:10022252

209. Hattis D (1996). Human interindividual variability in susceptibility to toxic effects: from annoying detail to a central determinant of risk. *Toxicology,* 111:5–14.doi:10.1016/0300-483X(96)03388-4 PMID:8711749

210. Hattis D, Barlow K (1996). Human interindividual variability in cancer risks. Technical and management challenges. *Hum Ecol Risk Assess,* 2:194–220 doi:10.1080/10807039.1996.10387468.

211. Hecht SS, Carmella SG, Murphy SE *et al.* (1993). A tobacco-specific lung carcinogen in the urine of men exposed to cigarette smoke. *N Engl J Med,* 329:1543–1546.doi:10.1056/NEJM199311183292105 PMID:8413477

212. Hecht SS (1997). Tobacco and cancer: approaches using carcinogen biomarkers and chemoprevention. *Ann NY Acad Sci,* 833:91–111.doi:10.1111/j.1749-6632.1997.tb48596.x PMID:9616743

213. Groopman JD, Kensler TW, Links JM (1995). Molecular epidemiology and human risk monitoring. *Toxicol Lett,* 82-83:763–769.doi:10.1016/0378-4274(95)03594-X PMID:8597140

214. Santella RM, Zhang YJ, Chen CJ *et al.* (1993). Immunohistochemical detection of aflatoxin B1-DNA adducts and hepatitis B virus antigens in hepatocellular carcinoma and nontumorous liver tissue. *Environ Health Perspect*, 99:199–202. PMID:8391434

215. Hagmar L, Wirfält E, Paulsson B, Törnqvist M (2005). Differences in hemoglobin adduct levels of acrylamide in the general population with respect to dietary intake, smoking habits and gender. *Mutat Res,* 580:157–165. PMID:15668117

216. Vodicka P, Koskinen M, Stetina R *et al.* (2003). The role of various biomarkers in the evaluation of styrene genotoxicity. *Cancer Detect Prev,* 27:275–284.doi:10.1016/S0361-090X(03)00096-5 PMID:12893075

217. Albertini RJ, Srám RJ, Vacek PM *et al.* (2003). Biomarkers in Czech workers exposed to 1,3-butadiene: a transitional epidemiologic study. *Res Rep Health Eff Inst,* (116):1–141, discussion 143–162. PMID:12931846

218. Smith MT, McHale CM, Wiemels JL *et al.* (2005). Molecular biomarkers for the study of childhood leukemia. *Toxicol Appl Pharmacol,* 206:237–245.doi:10.1016/j.taap.2004.11.026 PMID:15967214

219. Holecková B, Piesová E, Sivikova K (2004). Dianovskỳ J. Chromosomal aberrations in humans induced by benzene. *Ann Agric Environ Med,* 11:175–179. PMID:15627321

220. Perera FP, Mooney LA, Stampfer M *et al.*; Physicians' Health Cohort Study (2002). Associations between carcinogen-DNA damage, glutathione S-transferase genotypes, and risk of lung cancer in the prospective Physicians' Health Cohort Study. *Carcinogenesis,* 23:1641–1646.doi:10.1093/carcin/23.10.1641 PMID:12376472

221. Ammenheuser MM, Bechtold WE, Abdel-Rahman SZ *et al.* (2001). Assessment of 1,3-butadiene exposure in polymer production workers using HPRT mutations in lymphocytes as a biomarker. *Environ Health Perspect,* 109:1249–1255.doi:10.1289/ehp.011091249 PMID:11748032

222. Lee KH, Lee J, Ha M *et al.* (2002). 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,* 65:355–363. doi:10.1080/15287390252808028 PMID:11936216

223. Gwosdz C, Balz V, Scheckenbach K, Bier H (2005). p53, p63 and p73 expression in squamous cell carcinomas of the head and neck and their response to cisplatin exposure. *Adv Otorhinolaryngol,* 62:58–71. PMID:15608418

224. Cheng L, Eicher SA, Guo Z *et al.* (1998). Reduced DNA repair capacity in head and neck cancer patients. *Cancer Epidemiol Biomarkers Prev*, 7:465–468. PMID:9641488

225. García-Closas M, Malats N, Silverman D *et al.* (2005). NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet*, 366:649–659.doi:10.1016/S0140-6736(05)67137-1 PMID:16112301

226. Vineis P, Veglia F, Benhamou S *et al.* (2003). CYP1A1 T3801 C polymorphism and lung cancer: a pooled analysis of 2451 cases and 3358 controls. *Int J Cancer,* 104:650–657.doi:10.1002/ijc.10995 PMID:12594823

227. Committee on Biological Markers of the National Research Council (1987). Biological markers in environmental health research. *Environ Health Perspect,* 74:3–9. PMID:3691432