Molecular Epidemiology: On the Path to Prevention?

Frederica P. Perera

Cancer prevention has been the stated goal of molecular cancer epidemiology for the past 17 years. In this review, progress toward that goal is evaluated by using as examples well-studied environmental exposures—i.e., tobacco smoke, polycyclic aromatic hydrocarbons, aflatoxin B1, benzene, and hepatitis B virus—and their roles in lung, breast, and liver cancers and leukemia. The contributions of molecular epidemiology discussed here include providing evidence that environmental agents pose carcinogenic risks, helping establish the causal roles of environmental factors in cancer, identifying environment–susceptibility interactions and populations at greatest risk, and developing new intervention strategies. Molecular epidemiologic and other data indicate that assessment of carcinogenic risks should address both the range of risk across the population and the risk to subgroups who may be at high risk because of genetic or acquired susceptibilities, including young children. However, for the most part, research results have not yet been effectively translated into risk assessments and preventive health policies. An infrastructure linking scientists, policy makers, and other constituencies is needed to facilitate this process.

To extend our knowledge, the second generation of molecular epidemiologic research should include large-scale, collaborative studies incorporating validated biomarkers and automated technologies. An incentive to make the necessary investment is the recognition that prevention of only 20% of cancer in the United States would result in 200,000 fewer new cases diagnosed each year and an annual savings of $21.4 billion in direct costs alone. (J Natl Cancer Inst 2000; 92:602–12)

Although it stands on the shoulders of seroepidemiology and genetic epidemiology, molecular cancer epidemiology as a recognized discipline is quite new, having been formalized only in the 1980s (1–7). In 1982, Weinstein and I offended some purists in epidemiology and molecular biology alike by proposing a novel research approach that we termed “molecular cancer epidemiology” in which “advanced laboratory methods are used in combination with analytic epidemiology to identify at the biochemical or molecular level specific exogenous and/or host factors that play a role in human cancer causation” (1). To give structure to molecular epidemiology, we delineated four categories of biomarkers: internal dose, biologically effective dose, response, and susceptibility. Our hope was that, by introducing biomarkers into epidemiology, researchers “should be able to predict human risks more precisely than hitherto possible” (1). My colleagues and I then tested whether carcinogen-specific DNA damage, as a marker of biologically effective dose, could be detected in human blood and lung tissue in vivo and reported in 1982 that polycyclic aromatic hydrocarbon (PAH)–DNA aduct levels were indeed present in biologic samples collected within a pilot lung cancer case–control study (8). This was the first of many reports of the presence of carcinogen–DNA adducts in vivo in a human population. [There had previously been a single report (9) of methylated purine in human liver as a result of an intentional poisoning.]

During the past 17 years, molecular cancer epidemiology has been variously perceived as a new field, an advanced form of clinical epidemiology, or a subdiscipline of public health epidemiology (10,11); however, the label has stuck, and many schools of public health and research institutes now have programs in molecular epidemiology. A formal Molecular Epidemiology Group is affiliated with the American Association for Cancer Research, and several international associations have been established.

The stated goal of molecular cancer epidemiology is the prevention of cancer (1), a disease that in the United States claims over half a million lives annually, with more than 1 million new cases diagnosed each year and attendant direct annual costs of $107 billion (12). This review will trace the progress of this relatively new, still-developing field toward prevention. Is it advancing steadily or only “slouching” toward that goal? What has it accomplished thus far? What is its future? What resources are needed to overcome obstacles along the path to cancer prevention?

The cancers used as examples here—lung, breast, and liver cancers and leukemia—exact a terrible toll worldwide. Many lines of evidence indicate, even more clearly than in 1982, that the great majority of these and other cancers are, in principle, preventable because the factors that determine their incidence are largely exogenous or environmental (13–15). These factors include exposures related to lifestyle and occupation and pollutants in air, water, and the food supply. Genetic factors are clearly important in terms of influencing individual susceptibility to carcinogens; in some rare forms of human cancer, hereditary factors may play a decisive role. There is, however, increasing recognition that controlling external factors presents the greatest opportunity for primary cancer prevention and its immediate benefits. This awareness has lent greater urgency to the search for more powerful tools in the form of early-warning systems to identify causal environmental agents and flag risks well before the malignant process is entrenched. In this review, research on causal roles of specific environmental exposures—tobacco smoke, PAHs, aflatoxin B1 (AFB1), benzene, and hepatitis B virus (HBV)—in the cancers cited above will provide illustrations of the part molecular epidemiology can play in identifying and reducing risk.
For this critical review, it has been necessary to select only a few of the many possible examples of molecular epidemiologic studies. The field is strewn with studies either that failed to use validated biomarkers or whose designs did not adequately consider the biology of the end point. The studies selected for inclusion have avoided these major pitfalls, although the research discussed includes pilot or transitional studies that, by their nature, had less than optimal designs. The author hopes that this discussion will prompt other critical reviews of molecular epidemiology’s progress, out of which will come course-corrections for the field.

**Contributions of Molecular Epidemiology**

Let us consider the extent to which molecular epidemiology has furthered prevention of cancer by 1) providing evidence that specific environmental agents pose human carcinogenic hazards, 2) establishing their causal role, 3) identifying subsets of the population who are at special risk, and 4) using this information to suggest or to develop new and more effective strategies to reduce risk.

**Providing New Evidence That Environmental Agents Pose Carcinogenic Risks: the Role of Tobacco Smoke and PAHs in Cancer**

**Lung cancer.** Tobacco smoke was identified in 1949 as a potent human lung carcinogen and still ranks at the very top of the list of environmental carcinogens (16). It contains 55 known carcinogens, including PAHs (e.g., benzo[a]pyrene [BaP]), 4-aminobiphenyl (4-ABP), and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), a nitrosamine derived from nicotine (17,18). [For a review of the evidence of the carcinogenicity of tobacco smoke, see (18.)] PAHs are also found in outdoor air from automobile exhaust and emissions from 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 (19,20).

It has long been known from experimental research that many carcinogens, including the PAHs, exert their effects by binding to DNA and forming adducts that may lead to mutation and, ultimately, to cancer (21). Thus, using adducts as biomarkers has the theoretical advantage that they reflect chemical-specific genetic damage that is mechanistically relevant to carcinogenesis (21,22).

In 1982, PAH–DNA adducts were first detected in human subjects in vivo, specifically in white blood cells (WBCs) and lung tissue from lung cancer patients, most of whom were smokers. Samples from some of the nonsmoking cancer patients and control subjects in the same study showed the same molecular fingerprints, presumably as a result of passive smoking and other exposures to PAHs (8). Subsequent studies in healthy, exposed populations (active smokers, coke-oven and foundry workers, and residents of Poland, the Czech Republic, and the People’s Republic of China exposed to air pollution from coal burning) have found increased concentrations of PAH–DNA adduct levels in blood and other tissues, with no apparent threshold for DNA binding (23–26). Most studies of PAH–DNA adduct levels in human subjects have observed substantial interindividual variability among persons with similar exposure, about 30- to 70-fold for those measured in WBCs by immunoassay (26,27). These findings are consistent with traditional epidemiologic data showing elevated risk of lung cancer in PAH-exposed populations.

Although not all studies have been positive, since 1982 more evidence has been developed that PAH–DNA adducts in WBCs or lung tissue may be risk markers for lung cancer (28–30). In one study (29), higher PAH–DNA adduct levels were found in WBCs from 119 case subjects (compared with 98 control subjects), after adjustment for amount of smoking, dietary PAH exposure, and other potential confounders. While the difference was greatest for subjects who smoked, the same pattern of higher adduct levels in case subjects than in control subjects was seen in former smokers and nonsmokers. Unmeasured variability in smoking, diet, or indoor or outdoor PAH concentrations may partially explain these findings; however, they are also consistent with evidence that some individuals are predisposed to genetic damage from PAHs and, thereby, to lung cancer (30–32).

They support the theory that cumulative damage resulting from the overwhelming of DNA repair processes by genotoxic chemicals that bind to DNA is a major cause of cancer (33). The study of interactions between environment and cancer susceptibility has become one of the most active areas of prevention-related research. (See below for further discussion.)

Supporting molecular evidence that PAHs play an important role in lung cancer comes from the observations that the p53 tumor suppressor gene is mutated in 40%–50% of lung tumors and that the pattern of mutations is consistent with the types of DNA adducts and mutations induced experimentally by BaP (34,35).

Caution is necessary in interpreting results from some studies of PAH–DNA adduct levels and cancer risk. Case–control studies alone are, by their retrospective nature, unable to definitively establish causality, and the predictive value of adducts has not been established in prospective or nested case–control studies. 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 tissue such as blood are a valid surrogate for those in target tissue (7). The relationship between adduct concentration in blood and that in target tissue must be established for individual carcinogens. With respect to PAH–DNA, an experimental study (36) has shown ubiquitous binding of BaP metabolites to DNA and protein. Using an immunoassay for PAH–DNA, Tang et al. (29) found a correlation between DNA adducts in WBCs and lung tumor tissue from the same case subjects (r = .34; P = .05). Similarly, Wiencke et al. (37) used the 32P-postlabeling assay to demonstrate a correlation between aromatic-DNA adducts in WBCs and non-tumor lung tissue (r = .77; P<.001).

**Breast cancer.** There is conflicting but suggestive epidemiologic evidence that active tobacco smoking and passive tobacco smoking contribute to the incidence of breast cancer (38–40). Like other constituents of tobacco smoke, such as 4-ABP and some heterocyclic amines, a number of PAHs are potent mammary carcinogens in bioassays (41,42). In addition to being genotoxic and mutagenic as a result of DNA adduct formation, PAHs alter estrogen metabolism and binding to the estrogen receptor. In vitro studies show that human breast epithelial tissue is capable of metabolizing PAHs to their ultimate mutagenic/carcinogenic forms (42).

My colleagues and I first detected aromatic carcinogen–DNA adducts in breast tumor and non-tumor tissue from a small number of breast cancer patients and control subjects who were
undergoing reduction mammoplasty (43). The method used in this study, 32P-postlabeling with the P1 nuclease extraction procedure, detects a variety of aromatic adducts, including PAH–DNA adducts (44). A subsequent study (45), which used the same method, assayed adduct levels in normal tumor-adjacent tissues from 87 breast cancer patients and normal tissues from 29 mammoplasty patients and found case subjects to have significantly higher adduct levels than control subjects. (Throughout this review, the term “significant” is used only to refer to statistically significant results.) Neither study controlled for potential confounding variables. A recent case–control study of 119 breast cancer patients and 108 control subjects with benign breast disease, all from the same source population, measured PAH–DNA adduct levels in breast tissue by immunohistochemistry. The method used an antibody that recognizes (+)-antibenzo[a]pyrene diol-epoxide (BPDE)–DNA and structurally related PAH–DNA adducts (46,47). After controlling for known risk factors for breast cancer and PAH exposure via tobacco smoking and dietary sources, elevated levels of PAH–DNA adducts in breast tissue were associated with a significantly increased risk of breast cancer (Rundle A, Tang DL, Hibsboosh H, Estabrook A, Schnabel F, Perera FP: manuscript submitted for publication). Because PAH–DNA adduct levels in breast tissue reflect both PAH exposure and individual susceptibility to PAH-induced genetic damage, this finding suggests that individual variations in metabolic pathways and DNA repair mechanisms play an important role in breast cancer risk.

As in lung cancer, the pattern of p53 mutations in breast tumors is consistent with a role of PAHs; they are predominantly G→T transversions, which are induced experimentally by BaP (34,48). PAHs are, however, but one of many classes of carcinogens (including aromatic amines and nitrosamines) in tobacco smoke, polluted air, and other environmental sources, and the evidence regarding PAH does not rule out a causal role for these other environmental carcinogens.

Conclusion. Molecular epidemiologic research on lung cancer confirmed what was already known about tobacco smoking and lung cancer, so it provided no new data on causality. Rather, the research provided valuable insights into the mechanisms by which tobacco smoke constituents exert their effects. On the other hand, while molecular epidemiologic studies have not proved that PAHs and other aromatic compounds cause human lung and breast cancers, they have provided new evidence supporting that hypothesis. To further test the PAH–lung cancer hypothesis, my colleagues and I (49) have undertaken a nested case–control study of lung cancer within a prospective cohort study of more than 22,000 physicians. A large-scale, population-based case–control study on Long Island, NY, is also in progress (50).

Providing Definitive Evidence of Causality in a Prospective Cohort or Nested Control Design

Prospective studies or case–control studies nested within them are able to establish the predictive value of biomarkers by avoiding the temporal problem posed by retrospective studies: that the marker may reflect the disease rather than the risk factor(s). But because of their considerable cost and because the biologic samples collected therein are so precious and limited in number and quantity, few biomarkers have been definitively established as predictors of cancer. Biomarkers that have been prospectively associated with the cancers addressed in this review include metabolites of AFB1 and AFB2 adducts, which have been studied in conjunction with liver cancer, and chromosomal aberrations, which are biomarkers for many cancers, including leukemia where they may be indicative of exposure to benzene.

Biomarkers related to AFB1 and liver cancer. During the past 30 years, research in experimental animals and humans has confirmed that the food-borne mutagen AFB1 is a human hepatocarcinogen, acting synergistically with HBV. AFB1 is a fungal metabolite present in grains and cereals as a result of improper storage. Humans appear to be as sensitive as the most sensitive animal species tested to AFB1-induced DNA damage (51). Research has indicated that a number of biomarkers of the internal or biologically effective dose of AFB1 (AFB1 metabolites, AFB1–albumin adducts, and AFB1–N7-guanine adducts in urine) and HBV surface antigen seropositivity are risk markers for liver cancer on a population level. In 1992, Ross et al. (52) reported a prospective study of 18,244 men in Shanghai, the People’s Republic of China, among whom there were 22 incident cases of liver cancer. Analysis of urine samples banked 1–4 years prior to diagnosis from the case subjects and matched control subjects gave relative risks (RRs) of 2.4 (95% confidence interval [CI] = 1.0–5.9) for any of the AFB1 metabolites and 4.9 (95% CI = 1.5–16.3) for detectable AFB1–N7-guanine adducts. There was a strong interaction between the serologic marker of HBV infection and the AFB1 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 cohort showed that the presence of any urinary AFB1 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 AFB1–N7-guanine adducts. A synergistic interaction between the presence of urinary AFB1 biomarkers and HBV seropositivity resulted in a 59.4-fold (95% CI = 16.6–212.0) elevation in HCC risk (53). The implication for prevention is that both reduction in dietary levels of AFB1 and widespread HBV vaccination are needed, since the benefits of the latter will not be manifest for many years (52). (These biomarkers have subsequently been used as outcome measures in an intervention trial with the antischistosomal drug oltipraz. See below for further discussion.)

Subsequent studies in Taiwan of incident HCC case subjects and matched control subjects whose urinary AFB1 metabolites, AFB1–albumin, and AFB1–DNA adduct levels were measured in stored urine samples gave results consistent with the prior results from the People’s Republic of China (54). In HBV-infected men, the risk of HCC was increased with detectable AFB1–albumin and AFB1–DNA adduct levels (RR = 10.0; 95% CI = 1.6–60.9) (55).

The molecular evidence of a causal role for AFB1 is strengthened by other molecular data. In HCCs worldwide, a correlation has been observed between dietary AFB1 exposure and a characteristic mutation in the p53 tumor suppressor gene, a G→T transversion at codon 249 (AGG to AGT) (56).

Chromosomal aberrations and various cancers. Unlike carcinogen–DNA adducts, chromosomal aberrations are a nonchemical-specific marker. Combined analyses of data from Nordic and Italian prospective cohort studies involving 3541 subjects found that chromosomal aberrations were significant predictors of cancer. In the Nordic cohort, among subjects with
high frequencies of chromosomal aberrations, the standardized mortality ratio (SMR) was 1.53 (95% CI = 1.13–2.05), compared with 2.01 (95% CI = 1.35–2.89) in the Italian cohort (57). In the Italian cohort, cancer predictivity of high chromosomal aberrations was greater for hematologic malignancies than for all cancers: An SMR of 5.49 (95% CI = 1.49–140.5) was calculated for lymphatic and hematologic malignancies (58). Chromosomal aberrations could theoretically be useful as biomarkers to identify individuals within high-risk populations (e.g., chemical industry workers) who could benefit from greater surveillance or chemoprevention. They do not, however, provide clues as to the specific exposure(s) responsible for increased risk of cancer.

One remedy to the lack of exposure specificity of chromosomal aberrations is the analysis of exposure-specific patterns of the aberrations. For example, benzene is a model chemical leukemogen. Specific chromosomal aberrations have been observed in both preleukemia and leukemia patients exposed to benzene as well as in otherwise healthy benzene-exposed workers (57). By use of fluorescent in situ hybridization and the polymerase chain reaction, Smith and Zhang (59) 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.

Acute lymphocytic leukemia accounts for almost 25% of all childhood cancers (60). While more studies are needed, a number have reported associations between parental exposure to benzene and childhood leukemia, pointing to the importance of parental genotoxic exposures (59). 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 (61,62). Providing direct evidence of a prenatal initiation of pediatric leukemias, the MLL–AF4 gene fusion sequence has been detected in neonatal blood spots of leukemia patients subsequently diagnosed at ages 5 months to 2 years (63,64). Translocations involving 11q23 have been associated with treatment with chemotherapy agents that are strong inhibitors of topoisomerase II, an enzyme involved in DNA replication (63). A structural feature common to many topoisomerase II-inhibiting drugs and other chemicals (such as the benzene metabolite 1,4-benzoquinone) is the quinone moiety.

**Conclusion.** While it is likely that many biomarkers shown to be promising in cross-sectional or case–control studies will not survive the test of predictivity within a prospective study, those that do should have great utility in identifying at-risk populations and individuals and in serving as outcome markers in interventions, substituting for clinical manifestations of overt disease. Similarly, validated risk biomarkers will be useful as endpoint markers in risk assessments of environmental carcinogens, allowing regulatory and educational interventions in a more timely manner than is possible when tumor incidence or mortality is the sole outcome.

Of all the biomarkers related to environmental exposures and their biologic effects, chromosomal aberrations are the best validated as predictors of risk. AFB1 metabolites, AFB1–albumin, and AFB1–DNA adducts have also been shown to be biomarkers of cancer risk in case–control studies nested within prospective studies.

**Documenting Environment–Susceptibility Interactions and Identifying Populations at Greatest Risk**

A goal of prevention research should be to understand exposure–susceptibility interactions while adhering 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 (23,65–67). Although results from research on interactions have often been inconclusive and even conflicting, molecular epidemiologic studies indicate that some subgroups may have heightened susceptibility to environmental exposures.

The categories of susceptibility factors that can modulate environmental risks—i.e., genetic predisposition, ethnicity, age, gender, and health and nutritional impairment—have been reviewed in detail elsewhere (23,65,66). With respect to the cancers and exposures discussed in this review, molecular epidemiologic studies have reported a number of interactions between exposures to tobacco smoke, PAH, AFB1, or benzene and various susceptibility factors. These findings illustrate the complexities of interactions between environmental carcinogens and genetic and nongenetic susceptibility factors. While susceptibility of the young has been clearly demonstrated for a number of carcinogens, in most cases, the available data must be confirmed before they can be translated into specific risk assessment and prevention measures. The preliminary nature of much of the available data challenges the research community to further delineate the multiple complex interactions that determine individual cancer risk.

The most extensively studied susceptibility factors are the relatively common genetic polymorphisms that determine the metabolic fate of carcinogens and the level of DNA damage they exert. Polymorphisms in certain cytochrome P450 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, detoxify certain carcinogenic metabolites. Genetically determined variation in DNA repair modulates risks from agents that directly or indirectly damage the DNA. There is also evidence that risks from tobacco smoke, PAH, AFB1, and benzene vary with ethnicity, gender, age, and nutritional status. (See discussion and references below.)

**Lung cancer.** Two CYP1A1 variants, the MspI polymorphism and the closely linked exon 7 polymorphism that results in an altered protein having valine in place of isoleucine (Ile→Val), have been associated with increased risk of lung cancer among smokers in Japanese and a few Caucasian populations (68,69). In Caucasian populations, in whom the frequency of the polymorphisms is lower than in the Japanese, studies of CYP1A1 polymorphisms by themselves have been largely nonpositive; however, the WBCs of U.S. smokers with the exon 7 mutation had higher levels of PAH–DNA adducts than those of U.S. smokers without the variant (70). PAH–DNA adduct levels were also elevated in the placenta and cord blood of newborns with the CYP1A1 MspI gene polymorphism, suggesting that it increases risk from transplacental PAH exposure (71). In lung tissue of adults, adduct levels have been associated with CYP1A1 gene expression and CYP1A1 enzyme activity (72). Lung tumors of Japanese smokers with the susceptible CYP1A1 genotype were more likely to have p53 mutations than those of Japanese patients without the genotype (73). Variations
in genes coding for other cytochrome P450 enzymes may also modulate lung cancer risk, although the evidence is less compelling (74).

The M1 variant of GST (GSTM1) detoxifies reactive intermediates of PAHs and other carcinogens. The GSTM1 locus is entirely absent in approximately 50% of Caucasians, and this deletion has been associated quite consistently with increased risk of lung cancer (75). The frequencies of PAH–DNA adducts and p53 mutations were higher in lung biopsy samples from subjects with the GSTM1 null genotype (76,77).

The importance of combined and interactive effects of multiple gene variants is illustrated by the observation in Japanese populations that individuals with both the CYP1A1 exon 7 (Val/Val) and GSTM1 null genotypes have an eightfold (95% CI = 1.74–38.47) higher frequency of p53 gene mutations (73) and a sixfold (95% CI = 2.28–13.3) greater risk of lung cancer compared with persons with neither genotype (68,78). Studies of the role of CYP1A1 and GST in lung cancer risk in Caucasians have yielded inconsistent results, possibly reflecting ethnic differences in gene prevalence or linkage. Consistent with the Japanese studies, however, a case–control study of a largely Caucasian population showed a doubling of lung cancer risk (95% CI = 1.0–3.4) in patients heterozygous for the CYP1A1Msp1 and the GSTM1 null genotypes (79). In a European study (31), BPDE–DNA adduct levels in bronchial tissue and WBCs were elevated in smokers with both high CYP1A1 activity and GSTM1 inactivity.

Genetic polymorphisms in genes for DNA repair enzymes may also influence lung cancer risk (80–83). Moreover, DNA repair capacity varies among individuals as a result of environmental challenges and physiologic factors (84,85). Compared with healthy control subjects, patients with lung cancer were almost sixfold (odds ratio [OR] = 5.7; 95% CI = 2.1–15.7) more likely to have reduced ability to repair DNA damage that was induced by BPDE, the carcinogenic diol-epoxide metabolite of BaP (84,86).

Ethnicity also appears to affect lung cancer risk. The 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 (87), although the effect of unmeasured differences in the exposure levels of the subjects cannot be ruled out.

Women appear to be inherently more susceptible than men, on a dose-for-dose basis, to certain lung carcinogens. A number of epidemiologic studies indicate that women smokers are 1.7-fold to threefold more likely to develop lung cancer than male smokers with the same exposure (49,88). The level of PAH–DNA adducts and the frequency of G: C→T: A transversions in p53 were elevated in lung tumors from female smokers compared with those from male smokers (49,76,89). Adduct levels in non-tumor lung tissue were also higher in women than in men, and women had a higher ratio of adduct levels to pack-years (90). 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 (91) found that the effect of the GSTM1 null genotype on lung cancer risk was significant among women but not among men.

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 (92). 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 (93). In two studies (70,94), these effects were seen only in smokers with the GSTM1 null genotype, illustrating the importance of interactions between multiple susceptibility factors. 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 (95).

Breast cancer. There have been fewer studies on the role of genetic polymorphisms in breast cancer, and no clear patterns have emerged. An association between the CYP1A1 Msp1 polymorphism and breast cancer risk has been observed in African-American women (96). Elevated risk of breast cancer has also been noted for smoking at an early age in conjunction with the CYP1A1 Msp1 T→C transition at nucleotide 6235 and the CYP1A1 exon 7 A→G transition at nucleotide 4889: 2.5% and 2.2% of the cases of breast cancer, respectively, are attributable to these factors (97). Breast cancer risk was slightly elevated with the CYP1A1 exon 7 (Ile→Val) polymorphism (OR = 1.61; 95% CI = 0.94–2.75) and was highest for those women who smoked up to 29 pack-years (OR = 5.22; 95% CI = 1.16–23.56) (98).

The evidence with regard to the effect of GSTM1 alone is conflicting; however, the risk of breast cancer has been found to increase as the number of putative high-risk GST genotypes increased (P for trend <.001). The OR was 3.77 (95% CI = 1.10–12.88) for a combined genotype of GSTM1 null, GSTT1 null, and either GSTP1 valine heterozygosity or GSTP1 valine homozygosity (99). As in lung cancer, the GSTs in breast cancer may act through the modulation of DNA damage caused by PAHs and other substrates. For example, in the previously cited case–control study of breast cancer, PAH–DNA adduct levels in breast tumor tissue were significantly increased in women with the GSTM1 null genotype (Rundle A, Tang DL, Hibshoosh H, Estabrook A, Schnabel F, Perera FP, et al.: manuscript submitted for publication).

NAT2 deactivates carcinogenic aromatic amines via N-acetylation. Fifty percent to 60% of Caucasians and 30%–40% of African-Americans have the slow acetylator genotype (100). Among postmenopausal women, smoking increased breast cancer risk only among the women with the slow acetylator NAT2 genotype (41). In addition to the CYP1A1 and GST polymorphisms discussed above, genetic variation in estrogen and other receptors that are instrumental in the toxicokinetics of carcinogens can strongly influence breast cancer risk (101,102).

The effects of metabolic polymorphisms on risk appear to vary with ethnicity (101). In addition, the observation that the pattern of p53 mutations in breast tumors varies between black and white women and also between Japanese and Western women suggests that these groups differ in their environmental exposure to carcinogens or in their susceptibility to those exposures (35,103).

For reviews of other suspected environmental breast carcinogens and their interactions with susceptibility factors, see (104–107).

Liver cancer and leukemia. Increased risk of AFB1-induced HCC has been associated with the GSTM1 null/GSTT1 null
genotype in conjunction with smoking and drinking, with an apparent interaction between the genotype and low levels of certain carotenoids (108). The GSTM1 null genotype and the low-activity epoxide hydrolase genotype also appear to confer greater risk of liver cancer (109). Furthermore, there is some evidence that a genetic polymorphism in CYP2E1 increases susceptibility to this disease (110).

With respect to leukemia, the hepatic cytochrome P450 2E1 enzyme plays a key role in the activation of benzene to its ultimate hematotox- genicity (111). NAD(P)H: quinone oxidoreductase (NQO1) [NAD(P)H = reduced nicotinamide adenine dinucleotide or reduced nico- tinamide adenine dinucleotide phosphate] and two subclasses of GSTs (M1 and T1) are, in contrast, involved in the detoxification of the ultimate benzoquinones and their reactive benzene oxide intermediates, respectively (59,111). While the exact role of the GSTs is unclear, a case–control study of occupational benzene poisoning in Shanghai (112) showed that individuals homozygous for the NQO1 G609 C→A mutation were at a 7.6-fold (95% CI = 1.8–31.2) greater risk of poisoning. Benzene poisoning was linked to risk of preleukemia and leukemia. 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 hematotox- icity (59), but this inference has not been demonstrated.

A special case: susceptibility of the fetus and young child to all cancers. Compared with exposures occurring in adult life, exposures in utero and in the early years can disproportionately increase the risks of many types of cancer later in life (113–115). Experimental and epidemiologic data indicate that, because of differential exposure or physiologic immaturity, 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, or the fact that cancers initiated in the womb and in the early years have the opportunity to develop over many decades.

Molecular epidemiologic studies indicate that babies in the womb and young children exposed to tobacco smoke or PAHs experience a higher internal dose of toxicants or greater genetic damage than adults who are similarly exposed. In cord blood of newborns at delivery, levels of cotinine and PAH–DNA adducts were higher (by 70% and 30%, respectively) than in the mothers’ blood, also sampled at delivery (116). Although the difference in adduct levels was not statistically significant, this finding suggests that the fetus has at least a 10-fold higher sensitivity to genetic damage than the mother, since PAH exposure to the fetus is estimated to be about 10% of that to the mother.

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 (117). Initiation of smoking at an early age confers a higher risk of lung, bladder, and (possibly) breast cancers (118). Breast cancer risk associated with the NAT2 slow acetyl- lator genotype was higher in women who began smoking under the age of 16 years (41). 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 (38).

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

Developing New and More Effective Prevention Strategies

Primary prevention: reduction or elimination of exposures causally related to cancer. Primary prevention encompasses a spectrum of measures that include 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 (7). Examples of molecular epidemiologic studies that have documented the benefits of reduction of exposure include a study of 400 smokers enrolled in a smoking cessation program. Blood samples were drawn from the participants before they began the program and then at multiple time points after smoking cessation. Forty of the 400 smokers were successful at quitting. (Compliance was verified by assaying cotinine levels in blood.) Levels of other biomarkers, PAH–DNA and 4-ABP–hemoglobin adducts, refected cessation; by 8 weeks after quitting, their concentrations were signif- icantly reduced (119). 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 (120).

Primary prevention measures for AFB1-related liver cancer include the reduction of mold growth in harvested crops and the modulation of the metabolism of AFB1 to enhance detoxification (121). A study in rats (122) showed that oltipraz significantly lowered AFB1–albumin adduct levels and tumor incidence. Al- bumin adducts were significantly associated with risk of HCC overall but not within experimental treatment groups, suggesting that adduct levels could be useful in monitoring intervention on a population level but would have limited utility in identifying individuals destined to develop HCC (122). A randomized, phase IIa, placebo-controlled intervention trial (123) in Qidong, the People’s Republic of China, then used biomarkers to assess whether oltipraz can reduce the risk of liver cancer. Measurement of levels of urinary AFB1 metabolites, used as end points of efficacy, showed that intermittent, high-dose oltipraz inhib- ited phase I activation of AFB1, while sustained, low-dose olti- praz increased phase II conjugation and detoxification of the carcinogen (121,124). A phase IIb trial with oltipraz is ongoing.

Based in part on data supporting a role of antioxidants in genetic damage in smokers (70,94), a randomized clinical trial of vitamins C and E (93) is measuring PAH–DNA adduct levels as an intermediate marker of the efficacy of chemoprevention. Similarly, the role of retinoids in the chemoprevention of lung cancer is being tested by use of intermediate end-point markers.
In addition, biomarker assessment may be useful in primary prevention by providing motivational feedback to persons contemplating or enrolled in smoking cessation programs or undergoing lifestyle changes, such as dietary modification (33,127,128). Biomarkers can also be used to monitor the benefits of regulations and other policies aimed at reducing exposure of workers and the general population to pollutants. The significant decline registered in blood lead levels nationwide following banning of lead-soldered cans and removal of lead from gasoline is a good example of this type of monitoring and testifies to the success of the regulations (129).

Conclusion. Research demonstrates that biomarkers can measure the efficacy of exposure reduction and that clinical trials using biomarker levels as end points have the potential for making big gains in terms of immediate prevention. Both types of efficacy monitoring are more efficient and provide more immediate results than conventional approaches that rely solely on clinical end points, such as a diagnosis of cancer.

**Progress in Risk Assessment and Public Health Policy**

Molecular epidemiology has advanced prevention by identifying carcinogenic hazards, providing definitive etiologic data, furthering our understanding of individual susceptibility to environmental carcinogens, and facilitating clinical trials. Molecular epidemiology has not yet led to broad policy changes to prevent or to reduce exposure to carcinogens, but it has pointed the way.

What policy changes are indicated? The data on genetic susceptibility do not support a policy of large-scale individual screening, since many different common polymorphisms are likely to be involved in individual risk of cancer (33). The data on the interactions between environment and susceptibility have underscored the importance of preventing prolonged exposures to genotoxic carcinogens, even at low levels, since they can result in DNA damage that begins at a very early age, even in the womb, and accumulates over a lifetime (34,130). In principle, the same concern about prolonged exposure beginning at an early age applies to carcinogens that cause aberrant gene expression, cell cycle control, differentiation, and cancer progression (131).

Effectively linking molecular epidemiology to risk assessment and health policy formulation will, in most cases, require additional research to confirm and to further elucidate many of the reported interactions between specific environmental exposures and susceptibility factors (e.g., gene–environment, gene–gene–environment, and gene–nutrition–environment interactions). But the current data, taken together, have bearing on fundamental principles of risk assessment. For example, up to the present time, risk assessments by agencies such as the U.S. Environmental Protection Agency (EPA) have assumed that the population is biologically homogeneous in response to carcinogens. This default assumption can lead to underestimates of risk to the population and to sensitive subgroups, leading to standards and policies that are not adequately health protective or equitable (132,133).

The theoretical importance of focusing intervention strategies (regulations, public education programs, health surveillance, behavior modification, and chemoprevention programs) on the subgroups at greatest risk as a result of genetic or acquired susceptibility (134,135) is illustrated in Fig. 1 (based on (136)). The figure illustrates that, while the distribution of susceptibility/risk is symmetrical on a log scale, it is right-skewed on a linear scale so that, 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 one of median-sensitive individuals. (Although the numbers shown in Fig. 1 are the upper 95% confidence limit of risk with respect to uncertainty, the estimated increase in risk for individuals with 99th-percentile sensitivity is similar if the arithmetic mean estimates of risk with respect to uncertainty are 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.

Scientists and policy makers now widely agree that the new molecular epidemiologic and other data invalidate the assumption of population homogeneity (132,135,137,138) and suggest that assessment of risk from pesticides and the many industrial chemicals in air, water, and household products should address both the range of risk across the population and risks to specific differentially susceptible populations. Such groups might in-

---

**Fig. 1.** The theoretical distribution of cancer susceptibility and risk across a population that is heterogeneous with respect to sensitivity to a hypothetical nonthreshold carcinogen [based on (136)]. The x-axis represents the percentile of sensitivity; the y-axis represents the number of individuals. 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. Panel [a] shows the distribution on a log scale. Panel [b] shows the same distribution on a linear scale.
clude children, genetically susceptible subgroups, one or more ethnic group, or one gender, for which the demographics should be presented and considered in the risk assessment.

Even when currently available data support differential risk assessment, progress has been slow. For example, the Food Quality Protection Act of 1996 requires that, in setting standards for pesticides, the EPA must consider the special sensitivity of children to these agents, but few standards have been set to implement this policy with respect to the most hazardous carcinogenic pesticides, such as atrazine. Policies and regulations that would truly protect those who are likely to bear a disproportionate risk would reduce both the incidence and burden of cancer and would promote environmental justice (133).

Considering differential susceptibility in risk assessment is easiest when comprehensive data exist for a specific carcinogen, but that is the exceptional case. When this kind of information is unavailable, existing data on related compounds and susceptible subgroups could be used to develop science-based, default standards that reflect the current understanding of individual variability. The default standards could be applied generically or to various classes of carcinogens.

**OBSTACLES TO PROGRESS AND FUTURE NEEDS**

A number of obstacles impede the progress of molecular epidemiology toward prevention. The inadequacy of data on environmental hazards severely limits our ability to focus on priorities for prevention, particularly for children’s health (138). The need for basic carcinogenicity testing of the 70,000 chemicals in the EPA’s inventory is urgent: “Until more information is available, it is difficult to assess the possible role of these chemicals in childhood cancer and to take steps to reduce exposure to children” (138).

A major investment in infrastructure is needed to support the second generation of molecular epidemiologic research: large-scale interdisciplinary, collaborative studies to determine the predictive power of biomarkers and to measure interactions between genes and environment and between nongenetic susceptibility factors and environment. The studies must be of sufficient size and complexity to focus on multiple genotype-phenotype and other susceptibility combinations, taking advantage of recent advances in genomics, DNA microarray technology (DNA chips), and informatics (139). Recently launched nationwide efforts to assemble and to characterize a large panel of polymorphic genes through the Human Genome and Environmental Genome Projects should greatly facilitate this research (140). Funding agencies like the National Institutes of Health will need to step forward to support larger collaborative efforts that require sustained funding over many years. In particular, more studies of children as a large vulnerable subgroup should be a priority (141).

The lack of resources for interdisciplinary research training in molecular epidemiology is another obstacle for the field. An investment is needed to counter the increasing specialization of graduate and postgraduate training in U.S. universities (121).

A further handicap to progress toward prevention is the lack of an infrastructure to ensure rapid translation of molecular epidemiologic research findings into risk assessment and public health protection. Such an infrastructure would include mechanisms for sustained funding of translational, interdisciplinary research in risk assessment and for more communication among disciplines. This formal structure is needed to promote a dialogue between scientists, policy makers, and stakeholders (community and public-interest organizations, grass-roots groups, industry, and other constituencies).

**CONCLUSION**

The examples of the association of tobacco smoke and PAHs with lung and breast cancers, AFB1 with liver cancer, and benzene with leukemia indicate that molecular epidemiology has already 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 provide new and more effective strategies to reduce risk. What is now needed is timely translation of existing data 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. Understanding risks to children is a research priority (137).

Prevention of only 20% of cancer in the United States alone would result in 200,000 fewer new cases diagnosed annually and a savings in attendant financial costs of $21.4 billion (estimates for 1999) (13). Based on the progress so far, it is reasonable to expect that investment in molecular epidemiologic research should pay off within the next 5 years with cancer policies conducive to prevention.

**REFERENCES**

(22) Perera FP, Environment and cancer: who are susceptible? Science 1997;278:1068–73.


Hattis D. Human interindividual variability in susceptibility to toxic effects: from annoying detail to a central determinant of risk. Toxicology 1996;115:1–5.


NOTES

Supported by Public Health Service Research grants 5R01CA53772, 5P03CA13696-23, and 5R01CA60904 (National Cancer Institute) and grants 1R01ES06722, 1P50ES09600-01, 1P30ES09089-01, and 1RO1ES08977-01 (National Institute of Environmental Health Sciences), National Institutes of Health, Department of Health and Human Services; by U.S. Army grant DAMD17-94-J-4251; by Department of Energy grant DE-FG02-93 ER61719; and by awards from the Glady’s and Roland Harriman Foundation, the Bauman Family Foundation, the Robert Wood Johnson Foundation, the W. Alton Jones Foundation, New York Community Trust, and the Irving A. Hansen Memorial Foundation.

Manuscript received August 5, 1999; revised January 13, 2000; accepted January 28, 2000.