Molecular Epidemiology and the Genetics of Environmental Cancer

Peter G. Shields, MD, Curtis C. Harris, MD

Environmental, occupational, and recreational exposures to carcinogens contribute to cancer risk in humans. Cancer formation is a multistage process involving tumor initiation, promotion, conversion, and progression. Carcinogens can affect any of these stages through genetic and epigenetic mechanisms. The association of a suspected carcinogenic exposure and cancer risk can be studied in populations with classic epidemiologic techniques. However, these techniques are not applicable to the assessment of risk in individuals. Molecular epidemiology, in contrast, is a field that integrates molecular biology, in vitro and in vivo laboratory models, biochemistry, and epidemiology to infer individual cancer risk. Carcinogen-macromolecular adduct levels, and somatic cell mutations can be measured to determine the biologically effective dose of a carcinogen. Molecular epidemiology also explores host cancer susceptibilities, such as carcinogen metabolic activation, DNA repair, endogenous mutation rates, and inheritance of mutated tumor suppressor genes. Substantial interindividual variation for each of these biologic end points has been shown and, therefore, highlights the need for assessing cancer risk on an individual basis. Given the pace of the last decade, it is feasible that the next 10 years will allow molecular epidemiologists to develop a cancer-risk profile for an individual that includes assessment of a number of factors. This will help focus preventive strategies and strengthen quantitative risk assessments.

THE STUDY of environmental carcinogenesis finds its roots in the epidemiology of life-style-related and occupation-ally related cancers.1 In 1759, John Hill reported that tobacco snuff could cause oral cavity cancers. This was followed by a publication from Percival Pott who concluded that working as a chimney sweep, exposure to soot, and poor hygiene led to scrotal cancer. Other associations followed, such as lung cancer and asbestos, arsenic, or radium; bladder cancer and aromatic amines; bone sarcoma and radium; skin cancer and ionizing radiation; and leukemia and radiation or benzene. Equally important to the individual's cancer risks is related to endogenous rates of mutation, genetic instability, alterations in proto-oncogenes (naturally occurring genes that, when altered, contribute as oncogenes to cancer formation) and tumor suppressor genes (naturally occurring genes that are involved in controlling normal cell growth, differentiation, and development). This article will review the current understanding of the multistage process of carcinogenesis, its role in molecular epidemiology, and the application of molecular epidemiology to assessing carcinogenic risk. (Due to the limitations of citing multiple references herein, review articles are frequently cited and individual studies are offered as examples of the available literature. The reference list is not intended to be all-inclusive.)

MULTISTAGE CARCINOGENESIS

Carcinogenesis is a multistage process of normal growth, differentiation, and development gone awry. It is driven by spontaneous and carcinogen-induced genetic and epigenetic events. Figure 1 presents a simplified scheme. Tumor initiation involves the direct effects of carcinogenic agents on DNA, mutations, and altered gene expression. The attendant defects are involved in tumor promotion, whereby cells have selective reproductive and clonal expansion capabilities through altered growth, resistance to cytotoxicity, and dysregulation of terminal differentiation.1 Tumor promotion further involves an "initiated" cellular clone that may also be affected by growth factors that control signal transduction. During this process, progressive phenotypic changes and genomic instability occur (aneuploidy, mutations, or gene amplification).1 These genetic changes enhance the probability of initiated cells transforming into a malignant clone, the odds of which are increased during repeated rounds of cell replication.1 During tumor progression, angiogenesis allows for a tumor to grow beyond 1 or 2 mm in size.1 Ultimately,
tumor cells can disseminate through vessels, invading distant tissues and establishing metastatic colonies. The role of proto-oncogenes and their activation have become increasingly apparent in the multistage model of carcinogenesis. These genes are important to the regulatory mechanisms of growth, cell cycle, and terminal differentiation. Activation of proto-oncogenes enhances the probability of neoplastic transformation, which can either be an early or late event. Carcinogens can cause mutations in proto-oncogene DNA sequences, or they can act as tumor promoters enhancing the activity of oncogene protein products.

Tumor suppressor genes also play an important role in carcinogenesis. Like proto-oncogenes, these genes regulate cell growth and terminal differentiation. However, they have the opposite effect by limiting growth and stimulating terminal differentiation (Table 1). If inactivated, these genes increase the probability of neoplastic transformation and is a dominant function. Tumor suppressor genes are recessively inherited and generally require loss of both genomic alleles. This is exemplified by inheritance of a predisposition to retinoblastoma and/or osteosarcoma, where patients possess a hereditary loss of the Rb1 gene on chromosome 13. In the familial form, loss of one allele is inherited and the other is lost through later mutation. Loss of suppressor and antitumor metastasis genes can be an early or late event involving several steps, including angiogenesis and metastasis.

Tumor suppressor genes can be activated by mutagenic mechanisms. For example, the p53 tumor suppressor gene, located on chromosome 17, is the most commonly altered suppressor gene among all tumors studied so far. Germ line mutations of p53 have been found in familial cancer syndromes (Li-Fraumeni syndrome). This gene normally encodes for a phosphoprotein involved in control of cell proliferation. Single-base substitutions can result in loss of function or production of proteins that either interfere with normal function or otherwise directly enhance neoplastic transformation.

The method by which carcinogenesis affects DNA and produces mutational spectra is varied. Chemical carcinogens interact directly with DNA by covalently binding to nucleotides and forming adducts (Fig 2). If present at the time of DNA synthesis, these adducts can cause mutations. For example, adducts are present in the diet and are linked to esophageal and gastric carcinoma, resulting in base substitutions due to mispairing at sites where adducts are formed. Further, site specificity may be dependent on the type of nitrosocompound affecting specific oncogene or tumor suppressor gene function. Among the best studied is the rat family of proto-oncogenes. The ras protein products are involved in signal transduction pathways initiated by growth factors and hormones at cell membrane receptors. In several experimental systems, activation is associated with tumor formation, angio genesis, and metastasis. Mutation of ras proto-oncogenes has been observed in human cancers, including lung, colon, pancreas, melanoma, leukemia, and thyroid. Base substitutions at codons 12, 13, and 61 on exposure to such agents as nitrosamines, polycyclic aromatic hydrocarbons, and radiation causes activation in animal models.

Some mutations may reflect specific carcinogen exposures and may exhibit target organ specificity. For example, p53 mutations at codon 249, frequently observed in hepatocellular carcinoma from China or southern Africa, are consistent with the type of damage caused by aflatoxin B, exposure, a common dietary carcinogen linked to this tumor. In contrast, several types of p53 mutations have been observed in lung cancer, which is consistent with a multiple carcinogen exposure from tobacco smoke.

Some carcinogens can damage DNA directly, eg, bis(chloromethyl)ether, but most require metabolic activation by cytochrome P-450 metabolic enzymes. The primary role of cytochrome P-450 enzymes is detoxification. During metabolism, however, functional groups can be added or exposed that result in the formation of reactive electrophilic intermediates that bind covalently...
Hydrogen peroxide, superoxide anions, and other agents, and asbestos all cause oxidative damage. Oxidative DNA damage can be prevented chemically (vitamin E, glutathione, uric acid) or enzymatically (superoxide dismutase, catalase, peroxidase).

On a molecular basis, cells also possess the ability to repair DNA damage. An extensively studied repair enzyme is O\textsubscript{2-}-alkylguanine-DNA-alkyltransferase. This enzyme repairs damage from alkylating agents such as nitrosamines.

It is a suicide protein in that it transfers the alkyl group to itself and becomes inactivated. Cytotoxicity and tumor cell resistance are negatively correlated with the levels of this enzyme, and levels vary within organs and among people.

In almost every step of the multistage process of carcinogenesis, person-to-person differences in cancer susceptibility can be acquired or inherited. Inheritance of the ability to metabolize debrisoquin sulfate, an antihypertensive medication, is correlated with lung cancer risk. Another example is the aryl hydrocarbon hydroxylase enzyme involved in the metabolism of polycyclic aromatic hydrocarbons. Activity varies in lung tissue but can also be inducible on exposure to agents such as tobacco smoke.

Induction can even vary among individuals and is notably higher in patients with lung cancer than in non-cancer controls. Levels of aryl hydrocarbon hydroxylase activity have also been correlated with prognosis in lung cancer.

The multiple stages of carcinogenesis are best exemplified by a model of human colorectal tumorigenesis, as recently described by Vogelstein and others (summarized in a study by Fearon and Vogelstein\textsuperscript{14}). In the early stages, and apparently more common in patients with familial polyposis, loss or inactivation of a candidate tumor suppressor gene, MCC located on chromosome 5q, is associated with cellular hyperproliferation. Later, a phenotypic appearance of adenoma occurs that is accompanied by hypomethylation with genomic instability. More advanced tumors involve oncogenes and tumor suppressor genes not observed in early adenomas. These include Ki-ras mutations on chromosome 12; mutation of p53 tumor suppressor genes on chromosome 17; and a deletion of DCC, the putative tumor suppressor gene on chromosome 18q and that may be involved in cell-to-cell adhesion and possibly metastasis. Other allelic losses can occur in other chromosomes. Thus, it appears that at least six genetic events occur in the development of colorectal carcinoma. The combined events are more important than the actual order in which they occur.

**CARCINOGEN EXPOSURE ASSESSMENT**

Genetic damage from carcinogens results from a number of determinants, including exposure, absorption, metabolism, and DNA repair. Each of these can be affected by host factors. The amount of carcinogen that reaches the target macromolecule (e.g., DNA) in a sensitive organ is the biologically effective dose. Measurements of DNA and protein adducts in humans have been the focus of research in numerous laboratories. Central to these studies, however, is the application of sensitive and specific assays that are required to detect femtomole and attomole levels of adducts in microgram amounts of DNA. Current methods are challenged because of the complexity and multitude of possible exposures in human tissues. Previous research has demonstrated the need to develop assays using authentically synthesized standards and micropreparative techniques. Moreover, corroborative techniques are needed for validation.

A variety of assays are available to identify carcinogen DNA adducts in human tissues. These include the phosphorosirrus postlabeling assay for adducts in microgram amounts of DNA. The choice of tissue to be studied (e.g., blood, lung, buccal mucosa, or urine) can be critical for determination of exposure assessments because of the multiple factors that determine the degree of addition.

Exposures to polycyclic aromatic hydrocarbons (PAH) compounds are associated with an increased risk of lung and skin cancer. Industrial pollution, fossil fuels, and tobacco smoke account for the major environmental sources. Dietary exposures also occur due to overcooked or charcoal-broiled meats. Carcinogenic PAHs are metabolized by cytochrome P-450 monoxygenases, and reactive epoxide intermediates readily form adducts in DNA. Adduct levels have been correlated with exposure in coke-oven workers, tobacco consumption, and urban vs rural residences, but decreases during vacation from occupational sources have been observed. Seasonal variation in adduct levels has also been observed. Some studies have found correlations with tobacco consumption, but this may be due to other sources of exposure. For example, dietary exposures to PAHs can contrib-
Nitrosamines and other nitrosocompounds are potential human carcinogens. These compounds readily alkylate DNA and form adducts but most require metabolic activation. Exposure can occur through endogenous formation of nitrosamines or directly from dietary sources, cosmetics, drugs, household commodities, and tobacco smoke. Endogenous formation occurs in the stomach from the reaction of nitrosatable amines and nitrate, which are converted to nitrates by bacteria. Endogenous nitrosamine formation can be estimated by administration of proline, a nontoxic precursor that reacts with nitrates and is excreted in the urine. Excretion of nitrosoproline varies among individuals and is associated with risk of stomach and esophageal cancer. Whether this is related to endogenous host factors or coexisting dietary exposures remains unknown. It has also been shown that coadministration of vitamin C can reduce the rate of endogenous nitrosation.

The reaction of nitrosamine metabolites and DNA results in several types of adducts, only some of which are promutagenic. These adducts can be measured in human tissues (Fig 3) and have been correlated with cancer incidence. Using a radioimmunoassay for O'-ethyldeoxyguanosine, adduct levels were found to be higher in persons with esophageal cancer in China. In a separate study, levels of O'-ethylthymine were found to be higher in liver samples from persons with liver and other cancers.

Among the best studied potential dietary carcinogens are aflatoxins produced by Aspergillus flavus and Aspergillus parasiticus. These molds are contaminants of corn, peanuts, sorghum, rice, and other agricultural products. Levels of aflatoxin vary between regions. Aflatoxin consumption has been linked to primary hepatocellular carcinoma in areas where liver cancer is common, eg, the Orient and Africa.

Metabolic activation of aflatoxin B varies among individuals, resulting in different levels of adducts. Exposure and adduct levels are inversely correlated with residence in industrialized countries. Exposure also has been linked to a mutation in the p53 tumor suppressor gene in hepatocellular carcinomas.

Aromatic amines have been implicated in bladder carcinogenesis, especially in occupationally exposed cohorts and tobacco smokers. Metabolic activation occurs via multiple cytochrome P-450 and acetyltransferase. Internal dosimetry has generally focused on hemoglobin rather than DNA adducts. Levels are higher in smokers than nonsmokers and decrease with cessation.

Among the most important life-style risk factors in carcinogenesis is tobacco smoke exposure. Due to widespread aggressive advertising and the addictive nature of cigarettes, tobacco smoke has become the major cause of cancer. Of growing concern is the documentation that passive exposure to tobacco smoke also increases the risk of lung cancer. Because of the causal relationship to lung and other cancers, tobacco smoke exposure is useful for developing exposure assessments. Putative adducts have been correlated with consumption by the phosphorus 32-postlabeling assay in human lung but not in peripheral lymphocytes or oral mucosa. Tobacco-specific nitrosamines are potent carcinogens in laboratory animals. Levels are higher in secondary rather than mainstream smoke, raising the question of their role in passive smoke exposure. Hemoglobin adducts are increased in smokers more than nonsmokers, while higher levels are reported in snuff dippers. Inhala-
The Li-Fraumeni syndrome, at risk for tumors at multiple sites, inherit mutant p53 genes. Retinoblastoma, another inheritable tumor, is associated with loss of a suppressor gene on chromosome 13q. These individuals are also at risk of other tumors (soft tissue sarcomas and osteosarcomas) within radiation therapy ports. Host susceptibility can be manifested through inheritable interindividul differences in metabolism, DNA repair, genomic instability, or altered proto-oncogene or suppressor-gene expression.

The study of the cytochrome P-450 CYP2D6 enzyme is among the best examples of inheritable interindividul differences in metabolism. CYP2D6 is responsible for metabolism of several medications, including tricyclic antidepressants, β-blockers, and debrisoquin. Poor metabolizers are at risk of adverse drug reactions. However, in an English cohort, a fourfold higher risk of lung cancer was associated with the extensive metabolic phenotype. This association has been confirmed in a second study of lung cancer, with an odds ratio (OR) of 6.1. The extensive metabolic phenotype has also been shown to have an interactive effect with occupational exposures to asbestos and PAHs. A method for genotyping individuals from a small blood sample using the polymerase chain reaction and amplification of mutants and wild-type alleles targeting specific base mismatches is now available. This method, in conjunction with a restriction fragment length polymorphism analysis, can correctly characterize over 95% of metabolic phenotypes. Thus far, the only carcinogenic substrate identified for CYP2D6 is methyl nitrosamino pyridylbutanone but not other tobacco-specific nitrosamines that have been tested (F. Gonzalez, PhD, oral communication, January 1991). Unless there are other unidentified carcinogenic substrates, an alternative explanation is that the gene is in linkage disequilibrium with another gene related to cancer risk.

The cytochrome P-450 CYP1A1 enzyme metabolically activates PAHs. A restriction fragment length polymorphism using the polymerase chain reaction has been described that is reported to correlate with lung cancer risk (OR of 3) in Japanese cancer patients compared with historical controls. Other ethnic populations will also need to be investigated. For example, a striking difference in allelic frequency between whites and blacks in the United States was found.

An investigation involving the metabolism of isoniazid, commonly used in the treatment of tuberculosis, led to the identification of acetylation phenotypes for the enzyme acetyltansferase. This enzyme, besides metabolizing caffeine, also metabolizes certain carcinogenic aromatic amines. The slow acetylation phenotype has been associated with increased risk of bladder cancer. In contrast, rapid acetylation is correlated with the risk of colon cancer. A correlation of aromatic amine adducts and acetylation phenotype in tobacco smokers highlights the relationship of acetylation and DNA damage.

Detoxification of carcinogenic reactive intermediates also can vary among individuals. Glutathione transferase activity is inherited in an autosomal dominant fashion. It is responsible for detoxifying metabolites of aromatic and other compounds. Specific mRN class isoenzymes that have high activity toward trans-stilbene oxide have been correlated with lung cancer risk. An OR of 3.25 for the absence of such activity in adenocarcinoma patients was reported. In a separate study, the absence of this transferase activity was correlated with the risk of stomach and colon cancer.

Genetic differences in proto-oncogene sequences can also be predictive of risk. Using Southern blot analysis and restriction fragment length polymorphism, it has been possible to demonstrate that rare alleles of Ha-ras are detectable primarily in cancer patients but not in controls. These alleles are the result of reiterated DNA sequences arranged in tandem array. These regions may provide an unstable site where recombination or amplification occurs. An association of rare alleles with lung cancer risk was tested in a case-control study, and the association was found to be strongest for black persons. A total of 16 studies have tested the association with lung and other cancers, including bladder, breast, leukemia, brain, and colorectal, but have found conflicting results.

This may be
due to differences in characterization of rare alleles or study designs.

Analysis of the L-msc proto-oncogene also demonstrates a polymorphic restriction enzyme site. The site-presence allele has been detected more frequently in males with soft tissue sarcoma and also in patients with gastric and lung cancer. The association also suggests a worse prognosis at the time of diagnosis for patients with renal cell cancer. A recent analysis of persons with lung cancer failed to reveal the predictive value of this polymorphism.

Interestingly, racial differences were again found, whereby the homozygote site-present genotypes were detected more frequently in American blacks.

**PERSPECTIVE**

The assessment of cancer risk in an individual and on a molecular level is currently not possible because no single genetic factor is sufficiently predictive. It is likely that further elucidation of the multiple stages of cancer and their interactions will need to be identified. Carcinogen exposure and exposure in target tissues, mutational spectra of proto-oncogenes and tumor suppressor genes, and DNA repair all play a role. Given the pace of the last decade, however, it is feasible that the next 10 years will allow molecular epidemiologists to develop a profile for an individual that includes assessment of multiple DNA damage (adducts and mutations), genotypic typing of metabolism, DNA repair, and susceptibilities based on germline and altered proto-oncogenes and tumor suppressor genes. The development of these techniques, however, is incumbent on their trials in well-designed case-control and prospective studies. The ability to perform polymerase chain reactions in formalin-fixed, paraffin-embedded tissue sections and banked serum samples will allow molecular epidemiologists to study archival material.

The current status of quantitative risk assessment, which concerns risks of populations rather than individuals, is hampered by inadequate human data. Assumptions about risk and safety factors required because of regulatory perspectives are not based on sufficient human scientific data. The reliance on animal data leaves open the questions of interspecies and high-to-low dose extrapolations. However, the status of quantitative risk assessment also should improve as the genetics of cancer and its interactions with species are better elucidated. Assessments of cancer risk in individuals have potential impact on society and its ethical considerations, eg, worker selection and potential discrimination. Scientifically, the impact of cancer risk assessments in individuals can best benefit those persons identified as being at risk and the focus of cancer prevention strategies, including education and chemopreventive therapies.

We wish to thank Ainsley Weston, PhD, for his critical review of the manuscript and Bob Julia for his skillful editorial assistance.

**References**

6. Preston-Martin S, Pike MC, Ross RK, Jones JA, Hruby VJ. Genetic factors and risk assessment, which concerns risks of populations rather than individuals, is hampered by inadequate human data. Assumptions about risk and safety factors required because of regulatory perspectives are not based on sufficient human scientific data. The reliance on animal data leaves open the questions of interspecies and high-to-low dose extrapolations. However, the status of quantitative risk assessment also should improve as the genetics of cancer and its interactions with species are better elucidated. Assessments of cancer risk in individuals have potential impact on society and its ethical considerations, eg, worker selection and potential discrimination. Scientifically, the impact of cancer risk assessments in individuals can best benefit those persons identified as being at risk and the focus of cancer prevention strategies, including education and chemopreventive therapies.

We wish to thank Ainsley Weston, PhD, for his critical review of the manuscript and Bob Julia for his skillful editorial assistance.

**Molecular Epidemiology—Shields & Harris**

688 JAMA, August 7, 1991—Vol 266, No. 5