Molecular Epidemiology of Cancer

Yen-Ching Chen, ScD; David J. Hunter, MD, ScD

ABSTRACT Epidemiology is very successful in identifying environmental and lifestyle factors that increase or reduce risk of specific cancers, leading to cancer prevention strategies. However, the etiology of many types of cancer is still poorly understood, despite extensive use of questionnaires and interview-based approaches in conventional epidemiologic studies. The integration of molecular techniques into epidemiology studies may provide new insights and has been referred to as molecular epidemiology. For instance, our ability to make connections between lifestyle and cancer risk is limited by difficulty in accurately measuring exposure to many carcinogens—newer molecular markers of exposure may provide better information. The completion of the Human Genome Project gives us knowledge of the genetic variations that presumably underlie the fact that a family history of cancer is a risk factor for most cancer types. Some of this excess risk has been explained over the last decade by identification of mutations in genes that give rise to a very high familial risk. Molecular epidemiologists are searching for genes that may give rise to much smaller increases in individual risk, but account for much of the residual risk associated with family history. These genes may also interact with environment and lifestyle factors such that cancer risk is not equally elevated in all persons exposed to an environmental factor (but not genetically susceptible), or all gene carriers (but not exposed to the environmental factor). Molecular markers may help to differentiate tumors with the same histologic appearance into different etiologic subtypes. Finally, response to treatment may be determined by molecular subtypes of the tumor, or inherited variation in drug metabolism. Examples will be given of how use of molecular techniques is informative in epidemiological studies of cancer and is predicted to lead to improvements in cancer incidence, early detection, and mortality. (CA Cancer J Clin 2005;55:45–54.) © American Cancer Society, Inc., 2005.

INTRODUCTION

Molecular epidemiology refers to the use of molecular biology techniques in epidemiologic research. A PubMed search on the term “molecular epidemiology” in July 2004 found more than 3,500 citations; the vast majority of these describe the use of molecular techniques to identify subtypes of infectious organisms and thus make inferences about their transmission and pathogenesis. The term was first popularized in the context of infectious diseases, and in the early 1980s it was applied to chronic disease research. Schulte defined the term as “the incorporation of molecular, cellular, and other biologic measurements into epidemiologic research”; and inspection of the citations in the above search shows that most of the applications of molecular epidemiology to chronic disease are in cancer research.

A well-established paradigm for molecular epidemiology contrasts it with “traditional” epidemiology (Figure 1). In this comparison, “traditional” epidemiology is concerned with correlating exposures with cancer outcomes, and everything between the cause (exposure) and the outcome (a cancer) is treated as a “black box.” This formulation is somewhat unfair to “traditional” epidemiologists, who have usually been very interested in the biologic basis of their observed associations. Indeed, the Bradford-Hill criteria for assigning causality to an association featured biologic plausibility as an important criterion. However, it is by and large fair to characterize the interaction of epidemiology with biology as mostly passive—explanations were sought in the work of molecular biologists, toxicologists, and others for mechanisms by which a specific exposure might cause a specific cancer type. Applications to cancer were also hindered by our limited understanding of the molecular pathways and mechanisms underlying cancer causation, resulting in limited opportunities for rational integration of molecular analysis into cancer epidemiology studies. Furthermore, many “traditional” epide-
miologic studies are performed using databases, mailed questionnaires, or telephone interviews, providing little opportunity for obtaining the biologic samples necessary for molecular analyses.

In molecular epidemiology, the epidemiologist is much more of a participant in the assessment of the biologic basis for an association, by using biologic measurements to assess exposure, internal dose, biologically effective dose, early biologic effect, altered structure/function, invasive cancer diagnosis, tumor metastasis and prognosis. In this way, the epidemiologist may help open up the “black box” by examining the events intermediate between exposure and disease occurrence or progression.

The continuum of events represented in Figure 1 may be measured by “traditional” means, eg, a questionnaire measurement of exposure, or a histologic assessment of invasive cancer. Increasingly, however, the tools of molecular biology are used to provide more accurate and specific measurements.

Biomarkers used for molecular epidemiology in cancer research can be categorized into different classes: markers of exposure (eg, presence of a cancer-causing virus),3 markers of dose (eg, amount of the cancer-causing virus),4 markers of internal dose (eg, DNA adducts formed after the metabolic activation of certain carcinogens such as aflatoxin),5,6 markers of biologically effective dose (eg, somatic p53 mutations that may both indicate exposure to a specific carcinogen and be one of the genetic “hits” in the multihit model of carcinogenesis),7 markers of altered structure/function (eg, chromosomal aberrations),8,9 markers of susceptibility (eg, metabolic polymorphisms in genes that are involved in carcinogen metabolism or detoxification),10 markers of cancer subtype (eg, estrogen and progesterone receptors in breast cancer), and markers of prognosis (eg, metabolic polymorphisms in drug-metabolism gene) (Table 1). Some of these examples are discussed in more detail below.

**MARKERS OF EXPOSURE**

Measuring exposures (environmental and lifestyle factors potentially related to cancer) can be very challenging, particularly if the relevant exposures occurred in the distant past. If an exposure is measured with substantial error, then the statistical relation between the exposure and cancer occurrence will be greatly attenuated. Since we are often interested in common, but relatively weak, causes of cancer (eg, exposures that may double or halve the risk of cancer, as opposed to a factor like cigarette smoking that increases the risk of lung cancer 20-fold or more), the ability to estimate exposures accurately can be the difference between establishing a causal link and an inconclusive study. Thus, for a strong risk factor like smoking, the association is attenuated, but still observable, if misclassification of the exposure exists. However, the association between disease and a weak risk factor may be so attenuated as to be statistically indistinguishable from chance results. The history of studies of human papillomavirus (HPV) and cervical cancer provides an example of how the availability of increasingly accurate molecular techniques has strengthened the relation between exposure and disease. In this example discussed below, the relevant measurement concerns the presence or absence of a causal infection; however, the exposure may have occurred in the distant past and cannot be measured, or accurately recalled, at the time of cancer diagnosis. A second example of aflatoxin and liver cancer illustrates the use of the “molecular fingerprint” from past exposure to infer its existence.

**HPV and Cervical Cancer**

More than 100 genotypes of HPV have been identified, although only a subset of these have carcinogenic potential.15 HPV selectively infects the epithelium of skin and mucous membranes. Specific HPV types are associated with squamous cell carcinoma, adenocarcinoma, and dysplasias of the cervix, penis, anus, vagina, and vulva.16 Using current technologies, HPV DNA can be detected in 95% to 100% of cervical cancer specimens,17 and has been called a “necessary cause” of cervical cancer.18,19

**HPV Detection Techniques**

Initially, nonamplifying techniques (eg, Southern blot, in situ hybridization, and dot-blot) used...
radio-labeled nucleic acid probes to detect HPV infection in cervical samples. The disadvantages of these direct-probe approaches include low sensitivity and the need for relatively large amount of purified DNA. More sensitive amplification methods include Hybrid Capture (HC) and polymerase chain reaction (PCR). The Hybrid Capture II (Digene, Inc.) is a commercial HPV detection kit, used for detection of all high-risk oncogenic and most of low-risk nononcogenic HPV genotypes. The genotype-specific RNA probes bind to single-stranded HPV DNA and the hybrids are “captured” and detected using antibodies and chemiluminescent detection.

The various PCR approaches for HPV detection can detect either a group of HPV types or a single type of HPV. After PCR is performed by using consensus primers, specific HPV genotypes can be determined by restriction fragment length polymorphisms, linear probe assays, direct sequencing, or genotype-specific PCR primers.

The strength of association between HPV and cervical cancer depends on the sensitivity of the HPV detection technique. Schiffman and Schatkin compared two studies utilizing different methods for HPV detection in which the less sensitive Southern blot approach resulted in a lower estimate of the relative risk and attributable fraction than the more sensitive PCR approach. Similarly, Bosch, et al. compared three HPV DNA testing techniques (Southern hybridization, PCR, and Virapap) in relation to the odds ratios (OR) and attributable risk fractions for 926 female cases and controls with cervical scrapes in Spain. Virapap was a commercial dot blot screening test kit and could be used for seven HPV types. For cervical cancer outcome, ORs obtained by using Southern hybridization (OR, 16.3; 95% confidence interval (CI), 7.7–34.4) and PCR technique (OR, 24.3; 95% CI, 14.4–41.0) were about 2.5 and 4 times higher than OR obtained by utilizing Virapap (OR, 6.3; 95% CI, 3.4–11.6), respectively. The risk of cervical cancer attributable to HPV infection was substantially higher for Southern hybridization and PCR techniques (attributable risk fraction, 33% and 67%, respectively) as compared to the Southern blot technique (attributable risk fraction, 21%).
pared with Virapap (attributable risk fraction, 23%). Subsequent studies using increasingly sensitive methods have estimated the proportion of cases of cervical cancer with evidence of HPV infection to be greater than 99%. These data show the importance of using the most sensitive molecular analysis in molecular epidemiology studies if accurate estimates of relative risk and attributable risk fractions are to be made.

Most women with evidence of HPV infection do not develop cervical cancer and most infections resolve within 1 to 2 years. Thus, the current focus of molecular epidemiology studies is to find the determinants of initial infection, subsequent persistence of oncogenic HPV infection, and progression of early cervical abnormalities to invasive cancer, using large prospective studies of women with repeated measurements of HPV, cervical cytology, and potential cofactors for infection, progression, and invasion. It has been suggested that the amount of virus present in cervical tissues, as estimated by quantitative PCR, may distinguish HPV carriers at low and high risk of development of cervical cancer.

Interaction Between p53, Rb and HPV in Cervical Carcinomas

Further evidence for both the causality of the HPV-cervical cancer association, and the mechanism for the association, comes from studies of the biological interactions of HPV proteins with cellular proteins that are critical in controlling abnormal cellular proliferation. Integration of the genome of cancer-causing HPV subtypes into the host genome leads to expression of the viral oncogenic proteins E6 and E7. The E6 protein binds to and inactivates p53, and the E7 protein binds to and degrades the Rb protein. Inactivation and degradation of these two tumor-suppressor proteins directly contributes to genetic instability and loss of normal cell cycle controls, thus providing a mechanism for the compelling epidemiological

| Class of Biomarkers | Example | Agent/Exposure | Health Effect | Reference(s) |
|---------------------|---------|----------------|---------------|--------------|
| Dose marker         | Quantify human papillomavirus viral load by using real-time polymerase chain reaction | Human papillomavirus | Cervical cancer | 3 |
| Viral infection     | Cervical human papillomavirus presence | Human papillomavirus | Cancers in cervix, penis, anus, vagina, and vulva | 4 |
| Adducts             | Aflatoxin adducts in serum or urine | Aflatoxin | Hepatocellular | 5,6 |
| Acquired p53 mutation | Codon 249 mutations in p53 in hepatocellular carcinoma | Aflatoxin | Hepatocellular carcinoma | 7 |
| Metabolic polymorphisms | Inherited polymorphisms of NAT2 gene | Cooked meats and heterocyclic aromatic amines | Colorectal cancer | 10 |
| Chromosomal aberrations | Chromosomal aberrations detected in peripheral blood lymphocytes | Genotoxic carcinogens (eg, polyaromatic hydrocarbon, pesticides, butadiene, disiocyanate, and styrene) | A variety of cancer types | 8,11 |
| Diagnosis           | Risk factors for breast cancer are different according to estrogen and progesterone receptor status of the tumors | Environmental and lifestyle factors | Breast cancer | 12,13 |
| Treatment response  | Inherited difference in metabolic activation of chemotherapy agents | Thiopurine drugs/thiopurine S-Methyltransferase | Myelosuppression | 14 |
association between specific HPV subtypes and cervical cancer. Although smoking has been associated with risk of cervical cancer in epidemiologic studies, there has been concern that this association was confounded by smokers having more or different sexual partners and thus being more likely to be exposed to HPV. In studies in which HPV testing was used to identify HPV, smoking increased the risk of subsequent detection of cervical intraepithelial neoplasia. By limiting the study population to women with HPV, these findings strengthen the evidence that smoking causes cervical cancer, rather than merely reflecting an association of smoking with the probability of acquiring HPV infection.

SOMATIC MUTATIONS AND CANCERS

Ultraviolet-induced p53 Mutation and Skin Cancer

Cyclobutane pyrimidine dimers and pyrimidine-pyrimidone(6–4) photoproducts are two major forms of DNA damage after ultraviolet-B radiation. Among cyclobutane pyrimidine dimers, thymine-cytosine and cytosine-cytosine dimers are the most mutagenic. Thymine-cytosine → thymine-thymine and cytosine-cytosine → thymine-thymine mutations are often found in the p53 sequence of ultraviolet-induced skin cancer cells and have been observed in animal studies. Also, pyrimidine-pyrimidone(6–4) photoproducts are formed in DNA after ultraviolet-B radiation and may lead to DNA mutations. Although conventional epidemiology has long established the link between skin exposure and skin cancer, the specificity of the relation between ultraviolet-B and the p53 mutation permits epidemiologic data to identify the specific wavelengths (290 to 320 nm) in sunlight that are carcinogenic.

Aflatoxin, Hepatitis Viruses, and Hepatocellular Carcinoma

Aflatoxin is a naturally occurring mycotoxin found in contaminated foods such as peanuts, some other nuts, corn, soy sauce, and fermented soybeans. Aflatoxin exposure, hepatitis B or hepatitis C infection, and alcohol intake are major risk factors for hepatocellular carcinoma. Aflatoxin-nucleic acid adducts in urine and serum are intermediate biomarkers for assessing the biological effective dose of aflatoxin. Quantification of aflatoxin-serum albumin adducts has also been validated in several experimental and epidemiologic studies, making it a useful screening tool for large-scale studies. The finding that a high proportion of liver tumors diagnosed in aflatoxin-exposed areas have the characteristic G:C somatic mutation in codon 249 of the p53 gene adds to the associations between serum levels of aflatoxin and liver cancer risk. Aflatoxin B1 induces typical G:C to T:A transversions at the third base in codon 249 of p53. This molecular marker of aflatoxin-induced mutagenesis helped to establish that aflatoxin is a human liver carcinogen, responsible for a high proportion of the tumors in geographic areas with high dietary exposure to aflatoxin.

MARKERS OF SUSCEPTIBILITY

At each step in the continuum in Figure 1, between-person differences in the absorption, activation, and detoxification of carcinogens, or the response to DNA damage caused by some carcinogens, may mediate the relation between exposure and cancer outcome. The most common markers of susceptibility are mutations in specific genes that confer increased or decreased risk. Familial clustering of specific cancer sites has long been recognized, and family history of a specific cancer is associated with increased risk of that cancer for most cancer types. Molecular analysis now permits some of this familial clustering to be traced to these mutations (for examples in breast and colorectal cancer, see below). A spectrum exists ranging from “high penetrance” susceptibility genes, ie, those with a high probability of cancer in mutation carriers, to “low penetrance” genes, in which cancer is diagnosed in a much lower proportion of carriers or only those exposed to specific environmental or life-
style factors develop the disease. High penetrance mutations are usually identified by studying families at high risk of specific cancers or multiple cancer syndromes. Linkage analysis is used to identify chromosomal segments that occur more commonly in family members diagnosed with cancer than in those free of cancer. Specific mutated genes are then sought in these segments. Low penetrance gene variants do not give rise to familial clustering and are sought by comparing the prevalence of the variant in a series of unrelated cases compared with controls, often called “association studies.”

**Germline Mutations and Cancer**

Germline mutations are inherited from a parent and are present in each cell in the body. Up to 5% of breast cancer and up to 15% of colorectal cancer are attributable to these high penetrance hereditary germline mutations, with an even larger percentage associated with low penetrance mutations.

**Breast Cancer**

About 5% of breast cancer is attributable to rare high-penetrance mutations in a small number of specific genes (e.g., BRCA1, BRCA2, ATM, PTEN, and TP53); mutations in BRCA1 and BRCA2 account for up to 50% of hereditary and familial breast cancer. Each gene shows some differences in the familial pattern of cancer: BRCA1 is more strongly associated with ovarian cancer, BRCA2 with male breast cancer, ATM with radiosensitivity, PTEN with Cowden’s syndrome (including hamartomas of the skin and other organs), and p53 with Li-Fraumeni syndrome (including sarcomas and a variety of other tumors). These genes were originally identified in high-risk families in which information from the pedigrees was used to identify the chromosomal localization of the mutated genes. Follow-up studies of carriers of mutations in these genes are examining whether other genes are involved in modifying the age at onset or lifetime incidence of the disease.

Twin and family history studies have shown that the proportion of breast cancer due to inherited susceptibility is substantial, and only part of this is accounted for by these high penetrance genes. Variants in genes that give rise to lower relative risks may be much more common, and thus account for higher attributable risks than for high penetrance genes. These gene variants do not give rise to classic pedigrees, and are studied in conventional molecular epidemiology studies in conjunction with known lifestyle and environmental causes. Because of the lower relative risks being studied, large case-control studies are needed. Most of the current work examines genetic variants in candidate genes such as steroid-hormone metabolism genes. Future approaches will include whole-genome single nucleotide polymorphism searches, in which the whole genome of cases and controls is queried for polymorphisms that are more or less common in cases.

**Colorectal Cancer**

There are four hereditary colorectal cancer syndromes: familial adenomatous polyposis, hereditary nonpolyposis colorectal cancer (HNPCC), Peutz-Jeghers syndrome, and juvenile polyposis. HNPCC accounts for up to 3% of all colorectal carcinomas, and is associated with mutations in one or more of several mismatch repair (MMR) genes. Two human homologues of two bacterial MMR proteins, hMSH2 and hMLH1, were discovered in the early 1990s by studying families with a high burden of colorectal cancer. Subsequent analyses of a larger set of families showed that germline mutations in these two MMR genes are responsible for 70% to 90% of all HNPCC cases.

It is possible that a large proportion of colorectal cancer cases result from inherited medium- or low-penetrance alleles, in which environmental factors may interact with the inherited factors. For example, MSH6 is a low-penetrance gene compared with MLH1 and MSH2 mutations; mutations in MSH6 are associated with a later age of onset of the
disease and account for about 10% of all germline MMR gene mutations in HNPCC patients.\textsuperscript{47} Future studies of medium- and low-penetrance alleles, which are more difficult to identify and assess, will be needed to explain the balance of colorectal cancer cases associated with a family history.

\section*{Susceptibility Genes: Carcinogen Metabolism and Detoxification Genes}

Almost all exogenous carcinogens require activation by metabolic enzymes, and detoxification enzymes frequently exist to deactivate carcinogens or their intermediate metabolites. Inherited polymorphisms in these enzymes may alter their rate of activation or detoxification, thus increasing or decreasing the carcinogenic potential of the environmental exposures they act on.

The high temperature cooking of animal protein creates carcinogens such as heterocyclic amines and polycyclic aromatic hydrocarbons.\textsuperscript{48} Heterocyclic amines are metabolized by a number of enzymes, including N-acetyltransferase 2 (NAT2); single nucleotide polymorphisms in the coding sequence for NAT2 give rise to different forms of the NAT2 protein and these divide Caucasian populations into slow acetylators (about 55% of the population) and rapid acetylators (about 45%). Thus, a theoretical basis exists for about half of the Caucasian population to metabolize heterocyclic amines differently from the other half, and have different susceptibility to genotoxic damage after exposure to heterocyclic amines in cooked meat. Almost all studies have shown that when analyzed without consideration of meat consumption, there is very little difference in risk of colorectal cancer between slow and rapid acetylators. Several studies have shown that colorectal cancer cases are more likely to be rapid acetylators if they consume diets high in red meat intake.\textsuperscript{10} In some, but not all, studies that assessed exposure to heterocyclic amines by using information on meat intake, and preferred methods of cooking the meat (eg, grilling versus broiling versus roasting),\textsuperscript{49} rapid acetylators consuming higher predicted heterocyclic amine intake have again been at somewhat higher risk of colorectal cancer. Studies such as these can help identify the specific human carcinogens in complex mixtures, like diet, and may ultimately lead

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Proportion of Breast Cancer Attributable to Known and Unknown Germline Genetic Mutations.}
\end{figure}
to preventive advice for populations (eg, changes in meat cooking practices) or individuals (eg, those with susceptibility genotypes for common environmental exposures).

As our knowledge of the molecular mechanisms of carcinogenesis has expanded rapidly, additional gene pathways have become the focus of molecular epidemiology research. These include pathways such as DNA repair, cell cycle control, immune response, and the inflammatory response. Genes in these pathways are candidates in the search for between-person genetic variants that alter cancer risk or response to the environment.

MARKERS OF ALTERED STRUCTURE/FUNCTION

The first genotoxicity biomarker shown to be associated with cancer risk was based on chromosomal aberrations in peripheral blood lymphocytes. High levels of chromosomal aberrations were associated with increased total cancer incidence in cohorts in Nordic countries and increased total cancer mortality in an Italian cohort. These results support an approximately two-fold cancer risk among subjects with high frequencies of chromosomal aberrations. Factors related to genetic susceptibility are known to affect the frequency of chromosomal aberrations in peripheral blood lymphocytes. Thus, chromosomal aberrations may serve as an integrated index of susceptibility, as well as DNA damaging exposure in populations. Other markers of altered structure/function are being actively sought to identify high-risk individuals.

MARKERS OF DIAGNOSIS

Conventional diagnosis of cancer types has relied on anatomic localization and histopathologic appearance. It has been long speculated, however, that these classifications ignore heterogeneity among tumors of similar histopathologic appearance that may mean that there are two or more different types of cancer that we currently treat as a single entity. A limited number of studies have assessed whether risk factors for breast cancer, for instance, are different according to the estrogen and progesterone receptor status of the tumors; these studies have concluded that differences exist, although more studies of this phenomenon are needed to identify the differences that are consistent across multiple studies. For instance, the relative risk associated with combined hormone replacement therapy (estrogen plus progestin) was 1.67 (95% CI, 1.33–2.10) for estrogen receptor-positive, progesterone receptor-positive breast cancers, compared with 1.21 (95% CI, 0.87–1.68) for estrogen receptor-negative, progesterone receptor-negative breast cancer. The development of gene expression microarrays with thousands of genes has permitted the characterization of tumors previously treated as a single entity into multiple entities based on gene expression patterns. For hematologic malignancies, such as diffuse B-cell lymphoma, gene expression patterns have prognostic significance. Further work based on large-scale assays have reduced the number of genes involved to a smaller number. For instance, six genes (LMO2, BCL6, FN1, CCND2, SCYA3, and BCL2) were selected out of a much large number for prediction of survival in diffuse large B–cell lymphoma based on the gene expression results from microarrays. The progression from microarray to a small number of specific RNAs or proteins that could be tested by using real-time PCR or immunohistochemistry assay will permit treatment decisions to be tailored to the tumor subtype using techniques more familiar to clinical laboratories. It may be no accident that the hematologic malignancies are among those for which epidemiologic studies have provided the fewest clearly established risk factors—if we are studying multiple etiologic entities as if they are a single entity, then the statistical signal for environmental and lifestyle factors associated with just a single subtype will be greatly attenuated. Unfortunately, for molecular epidemiologists interested in etiology rather than prognosis, much of the work in this field has been done on samples with either no or minimal demographic information, and essentially zero environmental or lifestyle information. The challenge for molecular epidemiology is to take advantage of the emerging knowledge about tumor subgroups and design studies that permit simultaneous analysis of both risk factors and tumor subclassification.
MARKERS OF TREATMENT RESPONSE

One of the limitations in cancer chemotherapy is the recognition that some patients experience toxicity from cancer treatment that may be fatal, undermine quality of life, or limit the dose of drugs able to be given. Much of the variation in drug tolerance is thought to result from inherited differences in drug metabolism. Some people respond very favorably to certain drugs, while others experience treatment failure or toxic reactions. Just as most carcinogens require metabolic activation, most drugs also require metabolic activation to their active form. There are inherited differences in the metabolism of drugs, just as there are inherited differences in metabolism of carcinogens. Perhaps the best elucidated example of inherited differences in cancer drug response is the pharmacogenetics of the thiopurine drugs mercaptopurine and azathioprine. Fewer than 1% of persons given these drugs experience life-threatening drug-related myelosuppression; it has been shown that almost all of these people are homozygous for low activity alleles of a critical enzyme that metabolizes these drugs, thiopurine S-methyltransferase.50 Evidence suggests that carriers of one low activity allele require different doses of these drugs than carriers of two copies of the high activity alleles, to minimize toxicity and maximize response. Thus, prescreening for these metabolic enzyme polymorphisms may not only avoid unnecessary deaths due to acute toxicity, but may also help optimize drug dosage.

Progress has been made in identifying somatic alterations in tumors that may predict response to therapy. For example, about 10% of nonsmall-cell lung cancer patients had a rapid and strong response to gefitinib, a tyrosine kinase inhibitor. It has recently been shown that most of these tumors had activating mutations in the epidermal growth factor receptor (EGFR) gene and thus screening lung tumors for these mutations in EGFR may identify patients who can expect a good response to this drug.51

CONCLUSIONS

The integration of molecular biologic techniques into epidemiologic studies offers a panoply of opportunities to better understand the causes of cancer, the natural history of the different types of cancer, and the determinants of survival once a cancer is diagnosed. A hallmark of molecular epidemiology studies is the availability of tissues in which biological measurements can be made along the spectrum from exposure to death. Another hallmark is the underlying idea of heterogeneity—of exposures, susceptibility, and tumor types. This predicts that very large studies will be needed to permit stratification by these different elements of heterogeneity. On the other hand, there may be opportunities to make more robust conclusions in smaller studies, if we can use molecular tools to prestratify by susceptibility or outcome and use more accurate measurements of exposure, thus reducing attenuation of exposure-cancer associations. Ultimately, the ability to maximize the utility of these new molecular methods to better understand the epidemiology of cancers will depend on close collaboration of epidemiologists with molecular biologists, and the training of a new generation of molecular epidemiologists for whom this collaboration is not exceptional, but normal.

REFERENCES

1. Schulte PA. Molecular Epidemiology: Principles and Practices. San Diego: Academic Press; 1993.
2. Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58:295–300.
3. Josefsson AM, Magnusson PK, Ylitalo N, et al. Viral load of human papilloma virus 16 as a determinant for development of cervical carcinoma in situ: a nested case-control study. Lancet 2000;355:2189–93.
4. Silins I, Wang X, Tadesse A, et al. A population-based study of cervical carcinoma and HPV infection in Latvia. Gynecol Oncol 2004;93:484–92.
5. Wild CP, Jiang YZ, Sabbioni G, et al. Evaluation of methods for quantitation of aflatoxin–albumin adducts and their application to human exposure assessment. Cancer Res 1990;50:245–251.
6. Bennett RA, Essigmann JM, Wogan GN. Excretion of an aflatoxin–guanine adduct in the urine of aflatoxin B1-treated rats. Cancer Res 1981;41:650–654.
7. Olivier M, Hussain SP, Caron de Fromentel C, et al. TP53 mutation spectra and load: a tool for generating hypotheses on the etiology of cancer. IARC Sci Publ 2004:247–270.
8. Hagmar L, Stromberg U, Bonassi S, et al. Impact of types of lymphocyte chromosomal aberrations on human cancer risk: results from Nordic and Italian cohorts. Cancer Res 2004;64:2258–2263.
9. Hagmar L, Stromberg U, Tinnerberg H, Mikoczy Z. Epidemiological evaluation of cytogenetic biomarkers as potential surrogate endpoints for cancer. IARC Sci Publ 2004:207–215.
10. Chen J, Stampfer MJ, Hough HL, et al. A prospective study of N-acetyltransferase geno-
type, red meat intake, and risk of colorectal cancer. Cancer Res 1998;58:3307–3311.
11. Norppa H. Cytogenetic biomarkers and genetic polymorphisms. Toxicol Lett 2004;149: 309–334.
12. Potter JD, Cerhan JR, Sellers TA, et al. Progesterone and estrogen receptors and mammary neoplasia in the Iowa Women’s Health Study: how many kinds of breast cancer are there? Cancer Epidemiol Biomarkers Prev 1995;4: 319–326.
13. Colditz GA, Rosner BA, Chen WY, et al. Risk factors for breast cancer according to estrogen and progesterone receptor status. J Natl Cancer Inst 2004;96:218–228.
14. Weinshilboum R. Inheritance and drug response. N Engl J Med 2003;348:529–537.
15. Cubie A. When is an STD not an STD? HPV and cervical cancer. Microbiol Today 2003;30:58–60.
16. Reichman RC. Human Papillomaviruses. In: Braunwald E, Fauci AS, Kasper DL, et al. Harrison’s Principles of Internal Medicine, 15th ed. New York: McGraw-Hill; 2001:1118–1120.
17. Bosch FX, de Sanjose S. Chapter 1: Human papillomavirus and cervical cancer – burden and assessment of causality. J Natl Cancer Inst Monographs 2003;31:3–13.
18. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12–19.
19. Schiffman MH, Castle P. Epidemiologic assessment of causality. J Natl Cancer Inst Monograph 1994;459:179–193.
20. Davey DD, Zarbo RJ. Human papillomavirus testing–are you ready for a new era in cervical cancer screening? Arch Pathol Lab Med 2003;127:927–929.
21. Hubbard RA. Human papillomavirus testing methods. Arch Pathol Lab Med 2003;127:940–945.
22. Schiffman MH, Schattkin A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. Cancer Res 1994;54:1944–1947.
23. Bosch FX, de Sanjose S, Munoz N. Correspondence Re: M. Schiffman and A. Schattkin, Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. Cancer Res 1994;54:1944–1947. Cancer Res 1994;54:6288–6289.
24. Guerrero E, Daniel RW, Bosch FX, et al. Comparison of ViraPap, Southern hybridization, and polymerase chain reaction methods for human papillomavirus identification in an epidemiological investigation of cervical cancer. J Clin Microbiol 1992;30:2951–2959.
25. Song S, Liem A, Miller JA, Lambert PF. Human papillomavirus types 16 E6 and E7 contribute differently to carcinogenesis. Virology 2000;267:141–150.
26. Harris TG, Kulasingam SL, Kiviat NB, et al. Cigarette smoking, oncogenic human papillomavirus, Ki-67 antigen, and cervical intraepithelial neoplasia. Am J Epidemiol 2004;159:834–842.
27. Daya-Grosean L, Dumaz N, Sarasin A. The specificity of p53 mutation spectra in sunlight induced human cancers. J Photochem Photobiol B 1995;28:115–124.
28. Brash DE, Rudolph JA, Simon JA, et al. A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. Proc Natl Acad Sci USA 1991;88:10124–10128.
29. Nakazawa H, English D, Randell PL, et al. UV and skin cancer: specific p53 gene mutation in normal skin as a biologically relevant expression measurement. Proc Natl Acad Sci USA 1994;91:360–364.
30. Ziegler A, Jonason AS, Leffell DJ, et al. Sunburn and p53 in the onset of skin cancer. Nature 1994;372:773–776.
31. Jonason AS, Kimula S, Price GJ, et al. Frequent clones of p53-mutated keratinocytes in normal human skin. Proc Natl Acad Sci USA 1996;93:14025–14029.
32. Dumaz N, van Kranen HJ, de Vries A, et al. The role of UV-B light in skin carcinogenesis through the analysis of p53 mutations in squamous cell carcinomas of hairless mice. Carcinogenesis 1997;18:897–904.
33. Kress S, Sutter C, Strickland PT, et al. Carcinogen-specific mutational pattern in the p53 gene in ultraviolet B radiation-induced squamous cell carcinomas of mouse skin. Cancer Res 1992;52:6400–6403.
34. Ichihashi M, Ueda M, Budiyanto A, et al. UV-induced skin damage. Toxicology 2003;189:21–39.
35. Ettel RA. Mycotoxins. JAMA 2002;287:425–427.
36. International Agency for Research on Cancer. Aflatoxins. IARC Monograph on the Evaluation of Carcinogenic Risks to Humans 2002;82:171–300.
37. Kerr RA. Indoor radon: the deadliest pollutant. Science 1986;240:606–608.
38. Bresac B, Kew M, Wands J, Ozturk M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature 1991;350:429–431.
39. Hsu JC, Metcalf RA, Sun T, et al. Mutation hotspot in the p53 gene in human hepatocellular carcinomas. Nature 1991;350:427–428.
40. Stah F, Hussain SP, Hofseth LJ, et al. TP53 and liver carcinogenesis. Hum Mutat 2003;21:201–216.
41. Bennett IC, Gatts M, Teh BT. The management of familial breast cancer. Breast 2000;9:247–263.
42. Samowitz WS, Curtin K, Lin HH, et al. The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. Gastroenterology 2001;121:830–838.
43. Aaltoinen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 1998;338:1481–1487.
44. Leach FS, Nicolaides NC, Papadopulos N, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 1993;75:1215–1225.
45. Fishel R, Lescoe MK, Rao MR, et al. The human mutator gene homology MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993;75:1027–1038.
46. Nicolaides NC, Papadopoulus N, Liu B, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 1994;371:75–80.
47. Plaschke J, Kruger S, Dietmaier W, et al. Eight novel MSH6 germline mutations in patients with familial and nonfamilial colorectal cancer selected by loss of protein expression in tumor tissue. Hum Mutat 2004;23:285–285.
48. Kize MG, Salmon CP, Pais P, Felton JS. Food heating and the formation of heterocyclic aromatic amine and polycyclic aromatic hydrocarbon mutagens/carcinogens. Adv Exp Med Biol 1999;459:179–193.
49. Felton JS, Knize MG, Salmon CP, et al. Human exposure to heterocyclic amine food mutagens/carcinogens: relevance to breast cancer. Environ Mol Mutagen 2002;39:112–118.
50. Staudt LM. Molecular diagnosis of the hematologic cancers. N Engl J Med 2003;348: 1777–1785.
51. Losos IS, Czerwinski DK, Alizadeh AA, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004;350:1828–1837.
52. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350:2129–2139.