Genetic Epidemiology and the Future of Disease Prevention and Public Health

Muin J. Khoury

"As he chats with the young mother, the doctor flicks a cotton swab into the mouth of her infant son, collecting a small sample of mucus from inside his cheek. In the back room of his office, he inserts the sample into a machine which extracts DNA from the mucus cells and compares it with the genetic material on a dime-size chip. Minutes later, a computer printer begins to spit out a list of the infant's genes. Fortunately, all but a few genes are labeled 'normal.' It is those few that the doctor discusses as he explains the results to the mother. "Your son's genetic inheritance is generally good", he says "but he is somewhat predisposed to skin lesions. So starting right away, he should be protected against excessive exposure to the sun." And the doctor warns, "he may well be susceptible to cardiovascular disease later in life. To lessen this risk, after about age 2, he should begin a lifelong low-fat high fiber diet"" (1, p. 24).

The preceding is a futuristic vision of the practice of preventive medicine in the 21st century from *Time* magazine. Although this scenario may seem like science fiction and is loaded with significant ethical, legal, and social complications, the tremendous progress in the Human Genome Project which is rapidly unraveling our estimated 50,000 to 100,000 genes (2, 3) may turn this vision to reality. The futuristic predictions by the popular press are often echoed by the scientific community (4). As of mid 1996, more than 8,000 human genes have been identified and cataloged as part of the Online Mendelian Inheritance in Man (OMIM), with more than 5,000 mapped to specific gene loci (5). The field of medical genetics is rapidly expanding beyond the boundaries of single gene disorders that traditionally have been associated with individually uncommon specific genetic syndromes, inborn errors of metabolism, and birth defects to the realm of common chronic diseases including cancer, coronary artery disease, Alzheimer's disease, diabetes, and almost all diseases of major public health impact (6). Some of these diseases are represented in this volume of *Epidemiologic Reviews*. Recent examples of gene discoveries in diseases of significant public health impact include the gene for hereditary hemochromatosis, the most common autosomal recessive disorder (affecting 3 to 8 per 1,000 persons), which leads to significant morbidity and mortality from iron overload (7); the breast cancer genes BRCA1 and BRCA2 (8); and a chemokine receptor gene variant (CKR5), present in about 1 percent of the population, found to confer protection against HIV infection and progression to AIDS (9).

Clearly, more research is needed to map and sequence disease genes, and policies are needed to ensure the appropriate and ethical use of predictive genetic testing in the population (10), especially in the absence of acceptable and effective medical intervention (e.g., Huntington's disease (11) and Alzheimer's disease (12)). Potential misuse of genetic information in the workplace or by insurers have to be dealt with. Nevertheless, epidemiology will assume a greater role in assessing disease risks and defining genotype-phenotype correlations in different populations for the thousands of genes that will be identified. Genetic epidemiology will become an integral component of the epidemiology paradigm.

Interest in the field of genetic epidemiology has mushroomed over the last decade as evidenced by an increase in publications (13) and in training programs in genetic epidemiology, many of which are connected to schools of public health (14). I predict that the future holds tremendous promise for people interested in careers in genetic epidemiology, and that genetic epidemiology will have a significant impact on the future of public health.

GENETIC EPIDEMIOLOGY AND THE FUTURE OF PUBLIC HEALTH

Despite the advances in molecular genetics and the implications of their use in disease prevention, it is not
entirely clear when and how genetic tests and services can be applied to preventing common human diseases of complex etiology. In a recent commentary, The Centers for Disease Control and Prevention (CDC) Genetics Working Group recently reviewed the impact of genetic technology on the practice of public health using the Institute of Medicine's model of the future of public health (15); this model defines the core functions of public health: 1) assessment of the health and health needs of the population, 2) public health policy development, and 3) assurance and evaluation of public health programs (16). Although the different visions of the future of public health are influenced by a combination of medical, technologic, social, political, and fiscal developments (17-19), new gene discoveries, which occur on an almost daily basis, require that the public health community take a leadership role in translating the results of these discoveries into effective and appropriate strategies to prevent disease and disability in the general population by targeting environmental, behavioral, and medical interventions to each person's genetic susceptibility. As summarized in Table 1, genetic epidemiology will play a central role in providing data that will be useful for evaluating disease risks, developing sound health policies, evaluating genetic testing, and providing valuable information for medical and public health professionals and the public. This discussion focuses on three areas: 1) population-based assessment, 2) policy development and the concepts of disease prevention, and 3) dissemination information about human genetics.

TABLE 1. The Impact of genetic epidemiology on the future of public health

1. Will provide data on the public health impact of human genes and their interaction with preventable risk factors on disease morbidity, mortality, and disability in various populations
2. Will provide data to guide health policy guidelines on the appropriate use of genetic testing in disease prevention and public health programs
3. Will provide data to evaluate the impact of population-based prevention programs that reduce morbidity and disability associated with disease genes
4. Will provide data on the laboratory quality of genetic testing
5. Will become increasingly needed in core training programs in epidemiology and public health
6. Will provide core quantitative disease genetic risk information in integrated and online genetics information systems used by medical and public health professionals and the public

Genetic epidemiology and population-based assessment

As the branch of epidemiology that studies the role of genetic factors and their interactions with environmental factors in the occurrence of disease in various populations (20), genetic epidemiology will become a fundamental component of most epidemiologic studies of risk factors for all human diseases. The concept of gene-environment interaction is becoming a central theme in epidemiologic studies that assess causes of human disease in populations (20, 21). Almost all human disease results from the interaction between our genetic inheritance and the environment, broadly defined to include infectious, chemical, physical, social, and lifestyle factors. This concept is exemplified by the impact of phenylalanine in the diet on the occurrence of mental retardation among individuals who are deficient in phenylalanine hydroxylase enzyme (22). Another example is the impact of dietary intake of iron on the occurrence of multiple organ failure from iron overload among individuals with the hereditary hemochromatosis genotype (23). Rothman once affirmed that it is easy to show that any disease is 100 percent genetic and 100 percent environmental as well (24), an affirmation which highlights the premise of gene-environment interaction in human disease.

Although the technologic and methodological tools to measure allelic variation and to demonstrate the presence of genetic components in families and populations continue to evolve, the ability to integrate genetic studies in clinical and epidemiologic research may be quite simple, as described in the scenario from Time magazine (1). I can foresee a time when the classic epidemiologic paradigm of the two-by-two table correlating disease status (e.g., in a case-control study), with the presence or absence of an exposure or risk factor, will be replaced by a two-by-four table in which the underlying genotype at a particular locus or at groups of loci will be measured and evidence of interaction with the risk factor will routinely be sought in almost every epidemiologic study (table 2). Indeed, the case-control method is ideal to measure the impact of genetic variation on disease risk because inherited genetic make-up does not change over time, is not affected by disease status, and thus should be measured using any available tissue sample containing DNA (25). Case-control analyses should take into account exposure dose-response; interaction of alleles at the same and different loci, as well as interaction among nongenetic factors; and adjustment for potential confounding factors. The failure to consider the genetic component of any disease-risk factor association will dilute the impact of the risk factor in the
genotypes and exposures in the population, thereby reducing the ability to detect effects of genotypes and exposures in the population (26, 27).

Thus, genetic epidemiology will provide the scientific foundation to measure the magnitude of disease risk associated with different alleles both in the presence and absence of certain key nongenetic risk factors for the disease (Khoury et al. Centers for Disease Control and Prevention, Atlanta, Georgia, unpublished manuscript). Under different models for gene-environment interaction, stratification by underlying genotypes can markedly improve the predictive value of disease risk factors in order to better target prevention efforts (28).

**Genetic epidemiology and disease prevention**

How can genetic epidemiology change the future of public health and disease prevention? After all, genetic risk factors are unmodifiable, and, therefore, it can be argued that disease prevention and health promotion should target only preventable risk factors such as environmental chemical exposures, infectious agents, and diet through behavior modification, environmental intervention, and clinical preventive services. In addition, the word “prevention” leads us to ethical debates about eugenic implications of prevention through prenatal diagnosis and pregnancy terminations of fetuses at risk for various diseases and disabilities. Juenst recently articulated the difference between the concepts of “phenotypic prevention,” which emphasizes the efforts to prevent clinical manifestations of a genetic condition, and the concept of “genotypic prevention,” which emphasizes the efforts to prevent transmission of particular genotypes to the next generation (e.g., through reproductive counseling and prenatal diagnosis) (29). Stone and Stewart (30) recently pointed out the ethical dilemmas involved in applying the new genetic technologies for population-based testing if sound screening principles are overtaken by technical ability, professional zeal, and consumer demand. I have recently suggested (31) that in order for genetics and public health to overlap successfully, public health professionals should keep their efforts targeted on preventing disease, disability, and mortality among people who inherit specific genotypes using targeted risk factor, medical, behavioral, and environmental intervention. This is clearly the phenylketonuria and hemochromatosis model for prevention and not the reproductive genotypic model for prevention. This model of prevention will allow society to begin to address some of the ethical issues involved in the applications of genetic technologies and gene testing. Thus, I view the future of genetics in public health as dealing primarily with conditions in which there are available medical and environmental interventions that interact with susceptibility genotypes. As for those conditions in which specific genotypes have been associated with increased disease risk but for which no effective intervention is available (e.g., breast cancer, Alzheimer’s disease), genetic epidemiologists will provide medical and public health practitioners with both the quantitative data on allele frequencies in different populations and the role of interacting cofactors with the goal of finding disease interventions. These data will be crucial in evaluating the impact of targeted intervention according to individual genetic susceptibilities. The *Time* magazine scenario described earlier will not become a reality unless we have good risk estimates from genetic epidemiologic studies that can guide health policy decisions on the appropriate use of genetic tests and services to better target the traditional disease prevention and health promotion activities of public health.

**Genetic epidemiology and the dissemination of medical and public health information**

Finally, I predict that genetic epidemiology will make great contributions to the dissemination of genetic risk information to medical and public health professionals and the public. The progress of the Human Genome Project has led to an explosion of genetic information, ranging from information about DNA and protein sequences to allelic variation found at many gene loci. As the number of genes discovered increases over the next few years, the amount of basic, clinical, and epidemiologic data on these genes and their associations with specific diseases will increase rapidly. Even for single-gene disorders, many allelic variants have been found. In mid 1996, OMIM listed 87 allelic variations at the cystic fibrosis transmembrane conductance regulator gene (CFTR), and up to 500 mutations may be identified (5). The resulting
genotype-phenotype correlations and disease predictions turn out to be complicated even for simple gene conditions. It is not difficult to envision the amount of genetic-epidemiologic information that may become available on 50,000 genes, each with many allelic variants. With increasing population-based genetic-epidemiologic studies on the prevalence of alleles in different populations, the association of alleles with disease risks, and the role of gene-environment interaction, a plethora of information on the disease-positive and disease-negative predictive values for allelic variants will be produced. These quantitative evaluations will be available for diagnostic purposes, presymptomatic testing, and even for predictive purposes. In addition, evolving methods of testing will produce information on the validity and reliability of testing methods in terms of analytic sensitivity and specificity and of the predictive values for each test.

Because of continuous changes in the management of patients and families with single gene and multifactorial disorders, including genetic counseling and possibly gene therapy, data need to be made available on the impact of early intervention, and on the cost-effectiveness of intervention efforts for society as a whole (32). Health policy makers need to decide whether to target prevention strategies to the whole population or to the subsets of the population at highest risk for disease on the basis of genetic susceptibility (33). The significant public concerns about the ethical, legal, and social implications of the Human

### TABLE 3. Examples of online genetics information systems

<table>
<thead>
<tr>
<th>System</th>
<th>World wide web address</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Mendelian Inheritance in Man (OMIM)</td>
<td><a href="http://www.ncbi.nlm.nih.gov/OMIM">www.ncbi.nlm.nih.gov/OMIM</a></td>
<td>Catalog of human genes and genetic disorders started by V. McKusick; contains summary chronologic information and links to MEDLINE</td>
</tr>
<tr>
<td>Genome Data Base (GDB)</td>
<td>gdbwww.ncbi.nlm.nih.gov</td>
<td>Official repository for genomic mapping data at Johns Hopkins University, Baltimore, MD; contains links to OMIM and enzyme databases</td>
</tr>
<tr>
<td>Human Gene Mutation Database (HGMD)</td>
<td><a href="http://www.cf.ac.uk/uwcm/mg/hgmdc.html">www.cf.ac.uk/uwcm/mg/hgmdc.html</a></td>
<td>Catalog of known gene mutations associated with inherited diseases developed at the Institute of Medical Genetics, Cardiff, Wales</td>
</tr>
<tr>
<td>GENLINE</td>
<td>medhelp.netusa.net/www/agag.htm</td>
<td>Expert-authored electronic textbook under development by the University of Washington, Seattle, WA; contains disease profiles including diagnosis, management, and counseling</td>
</tr>
<tr>
<td>Alliance for Genetic Support Groups (AGSG)</td>
<td>medhelp.netusa.net/www/agag.htm</td>
<td>Nonprofit organization, a resource for consumers and professionals looking for genetic support groups and genetic services; extensive list of disease-specific support groups</td>
</tr>
<tr>
<td>National Center for Genome Resources (NCGR)</td>
<td>nccr.org/hcgr/hcgr.html</td>
<td>Nonprofit organization in Santa Fe, NM, providing information on ethics, gene therapy, social issues, public policy, and education</td>
</tr>
<tr>
<td>Information for Genetic Professionals</td>
<td><a href="http://www.kumc.edu/GEC/prof/gene/prof.html">www.kumc.edu/GEC/prof/gene/prof.html</a></td>
<td>General information for genetic counselors, clinical geneticists, with links to research and teaching developed at University of Kansas Medical Center, Kansas City, KS</td>
</tr>
</tbody>
</table>
Genome Project (34–36), and increasing commercial pressures to make genetic tests available for predictive purposes even in the absence of effective and acceptable intervention (30), make it even more crucial for the public and health care providers to have current information on health policy statements from government groups, consensus conferences, and professional groups about the appropriateness and the implications of genetic testing.

In addition, public health officials need to have population- and state-specific data on the prevalence of the alleles in question and on the public health impact on disease, disability, and death in populations and subpopulations and various ethnic groups. Such information will be extremely helpful in planning, implementing and evaluating public health policy for clinical and public health programs that use genetic tests and services.

Public health information systems are beginning to emerge as technology improves and demand for such information increases. Thacker and Stroup (37) have recently described a comprehensive system for public health surveillance in the United States based on a network of data systems. Baker et al. (38) have recently described the launching of the CDC’s Information Network for Public Health Professionals (INPHO). INPHO was established in 1992 to strengthen the public health infrastructure in the United States through computer linkage, information access, and data exchange among federal, state, local, and private organizations.

With advances in biomedical informatics (39), and the increasing popularity of the World Wide Web (www) at the expense of traditional publications (40), the last few years have witnessed the proliferation of many online genetic information systems. Table 3 lists some examples of currently available online information systems, including those for gene-mapping information (human genome database), gene sequence information (GenBank), and gene mutation information (human gene mutation database). OMIM provides detailed abstracted information from the literature on all genes and associated phenotypes in an expanding text format that includes copious references. GENLINE is being developed by geneticists at the University of Washington as an expert-authored electronic publication containing up-to-date genetic disease profiles, including information on diagnosis, management, genetic counseling and testing, and support groups. The same group has worked on a companion online directory for all laboratories that provide genetic testing. Some examples of additional sources of online general genetics information include the National Center for Genome Resources, and many World Wide Web pages from universities, such as the one from the University of Kansas Medical Center, and support organizations such as the Alliance for Genetic Support Groups (table 3). None of the online systems, however, provide population-specific data on disease risks, allele frequencies, prevention-effectiveness analyses, or laboratory quality data.

Khoury et al. (Centers for Disease Control and Prevention, Atlanta, Georgia, unpublished manuscript) have recently described the rationale and need for developing an integrated genetics information system for use in medicine and public health, and have described its knowledge-base content. This is a proposed online system that provides up-to-date information on the role of human genes in disease risk in different populations, and can be used to 1) integrate clinical and research information, 2) plan and evaluate population-based disease prevention strategies, and 3) provide educational materials for health professionals and the public. Such a knowledge base should be peer reviewed and continuously updated.

An integrated genetics information system could evolve as a confederation of information sources and as a partnership of federal agencies, state health departments, academic institutions, support groups, and industry. An important component of this system will be genetic-epidemiologic information on human genes. I predict that genetic epidemiologists will provide an important leadership role needed for such an effort. The contributions of genetic epidemiology to an integrated genetics information system will facilitate the translation of the Human Genome Project into the practice of medicine and public health in the 21st century.

REFERENCES
9. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele...


