Challenges to Interpretation of Gene-Environment Research

Presentation by:
Paul A. Schulte
Director, Education and Information Division
National Institute for Occupational Safety and Health
pas4@cdc.gov

Abstract:

Epidemiologists increasingly are presented with powerful new tools to piece together the roles of environmental/occupational and genetic factors in assessing disease risks. High throughput tools, such as microarrays and 2D gels, present new opportunities, but their utility for assessing disease risks in populations is still some time from being realized.

Making sense from population studies incorporating genes and environment has a rich history, but one fraught with problems. Despite notable examples to the contrary, the nature-nurture debate has given way to the fundamental realization that both genes and environment play a role in disease. Unfortunately, however, the degree of increased risk in exposed genetically susceptible individuals is highly dependent on the relationship between genes and exposure in terms of their effect on disease risk; and we seldom have had information about this. Depending on the type of model for gene-environment interaction, risks can vary dramatically. At least six different types of gene-environment interaction have been described. These have been generally considered with the dichotomous conditions, a single susceptibility genotype (present or absent), and a single environmental factor (present or absent). For example, in a paper by Khoury and Wagener, the risk to an exposed person with a rare (1% prevalence) susceptibility genotype ranges from 8% (under a multiplicative model of interaction) to 95.6% (when we assume that the exposure has no effect in susceptibles, but the genotype raises risk in unexposed as well as exposed persons).

Various study designs have been identified for the detection of gene-environment interaction. These include: case-only designs, case-control study designs using unrelated controls, case-control designs using related controls, case-control designs using relatives of cases and population- or hospital-based controls, twin-study designs, family-study designs, and combined segregation and linkage analyses.

There is currently a debate whether bias from population stratification (the mixture of individuals from heterogeneous genetic backgrounds) undermines the credibility of epidemiologic studies designed to estimate the association between genotype and risk of disease. However, Wacholder et al. found only a small bias from stratification in a well-designed case-control study of genetic factors that ignored ethnicity among non-Hispanic, U.S. Caucasians of European origin. When important confounding caused by population stratification does occur, it should be controllable by the usual design and analytical features employed by epidemiologists.

Population studies are necessary if gene-environment interaction is to be assessed because of the large sample sizes required. For example, in the simple case of three binary variables (for genotype, exposure, and disease) there are eight combinations that can reflect five different types of relationships among the three variables: (1) complete independence, (2) partial independence, (3) conditional independence, (4) no interaction, and (5) interaction. Even larger numbers will be needed if more than one genetic and environmental factor are studied. Statistical models to assess interaction have been developed, but rarely applied, beyond the 2x2x2 design because sample size in the cells become small and hence, limit the power. Assessment of interaction can also be affected by exposure and genotype misclassification. Modest error in exposure assessment may result in substantial increases in sample size requirements and this problem is compounded by even small errors in genotype assessment. These challenges have led some investigators to call for a new paradigm to describe interaction that involves some form of network and chaos theory.

Meanwhile, new high throughput technologies will present a number of issues that amplify these earlier challenges. First will be the large amount of data. This will be a challenge in terms of validity, reduction, summarization, analysis, and interpretation. Many early studies have shown different sets of genes perturbed for similar exposures. This may be due to a lack of standardization of approach to guide analysis and interpretation. Studies where expression of genes are measured under different conditions need rules for determining what is an outlier (a potentially important deviation in expression). There is a need to determine whether variability of experience for each gene can be measured in the same or different scale. Replication is necessary to determine whether tested genes have the same natural scales of measurement. Even if the scale of measurement is appropriate for different genes, the question arises whether a useful method has been used for detecting outliers. Different approaches to detecting outliers include principal components analysis and cluster analysis which are being widely used or considered. Depending on how data are transformed can lead to different outcomes of these analyses. There are other statistical questions that will arise in these new and uncharted areas; thus, new problems need exploration by many methods to gain insight into the data and the relationships that might be found.

Other fundamental questions will need to be answered before interpretations of the deluge of data will begin to be possible. These include the ability to empirically define "housekeeping" genes, to identify reproducible artifacts, and to distinguish homoeostatic from pathologic perturbations.

In the short-term, epidemiologists will not so much use the products of high throughput technologies but be involved in the development and validation of them. This includes the development of large scale data bases, the need to determine prevalence and baseline for microarray data and the range of variability, the approaches for addressing the issues of multiple comparisons, and the development of new ways of defining diseases. Microarray data may provide new definitions of cases and controls for epidemiological studies. Epidemiologists will also play a role in the incorporation of microarray data in risk assessments and intervention research. All of these efforts that involve human subjects will require attention to the ethical, legal, and social issues that have been raised in the molecular epidemiology literature, and to the increased questions of interpretation that will undoubtedly arise.

WELCOME

The objective of the International Council of Chemical Associations (ICCA) Genomics Workshop is to review the state-of-the-science in the application of genomic technologies in toxicology, ecotoxicology, and molecular epidemiology and the importance of these new developments in understanding the potential effects of chemicals on humans and the environment. Social, legal, ethical issues and their influence on the direction and application of research will also be highlighted. An anticipated outcome of the workshop will be the identification of options for the global chemical industry to consider as it develops a research strategy and other plans for incorporating these scientific advances into product stewardship.

This workshop is sponsored by ICCA, which is a council of leading trade associations representing chemical manufacturers worldwide, and is recognized as a non-governmental organization by the United Nations. ICCA's purpose is to work on policy issues of international significance, and to promote and coordinate voluntary chemical industry initiatives, including Responsible Care® and the Long-range Research Initiative (LRI). Through the ICCA, the American Chemistry Council, European Chemical Industry Council (CEFIC), and the Japan Chemical Industry Association (JCIA) have coordinated research projects and research strategies in key areas.

This workshop was designed and organized by the ICCA Genomics Workshop Steering Committee, members of which include:

Richard Phillips, ExxonMobil Biomedical Sciences, Inc. (Chair)
Edward Carney, The Dow Chemical Company
Rory Conolly, CIIT Centers for Health Research
Chris Corton, CIIT Centers for Health Research
William Greenlee, CIIT Centers for Health Research
Carol Henry, American Chemistry Council
Koichi Igarashi, Japan Chemical Industry Association
Robert Leboeuf, The Proctor & Gamble Company
Cheryl Morton, American Chemistry Council
Lewis Smith, Syngenta CTL
Ray Tennant, National Institute of Environmental Health Sciences
Jane Teta, Union Carbide Corporation
Donald Versteeg, The Proctor & Gamble Company

Logistical and meeting support is being provided by ICF Consulting and the American Chemistry Council's Member Services.

Thank you for your participation and efforts in this workshop.

ICCA Genomics Workshop

March 7 and 8, 2001

Orlando, Florida