

## Meeting Report

# Diet, Nutrition, and Genetic Susceptibility

Rashmi Sinha<sup>1</sup> and John D. Potter

Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland 20892-7374 [R. S.], and Fred Hutchinson Cancer Research Center, Seattle, Washington 98104-2092 [J. D. P.]

### Introduction

On January 22–23, 1997, a workshop was held in Washington, D.C., under the auspices of the Division of Cancer Epidemiology and Genetics, National Cancer Institute (Bethesda, MD), and the Fred Hutchinson Cancer Research Center (Seattle, WA). The purpose of the workshop was to provide a forum for the discussion of the interplay between diet, genes, and cancer risk. This is just part of the burgeoning area in epidemiology that is studying gene-environment interaction more generally; however, it was felt by the organizers that a workshop on genes and diet was warranted, because there is considerable research activity in this area, and because it offered the opportunity to define new directions. It was not intended to review the field in detail, but rather to consider, from a variety of disciplinary perspectives, the future possibilities. Participants were asked to: (a) address issues surrounding the incorporation of genetic markers into epidemiological studies of diet and disease and (b) identify strengths and weaknesses in current research and gaps in understanding and propose new research strategies, tools, and hypotheses.

### The Workshop

The workshop was divided into four sessions followed by some summary observations. Following opening remarks by Joseph Fraumeni and a general introduction, "Diet and Genetic Susceptibility," by Rashmi Sinha, the first session, "Variation in Metabolic Profiles as a Source of Susceptibility to Cancer," laid out some basic principles for the study of genetic susceptibility (Neil Caporaso) and of nutrition (Christine Swanson) in relation to cancer. Examples were then detailed from various other exposure areas of how the interplay of environment and genes can modify disease risk. The papers covered occupation (Nathaniel Rothman), tobacco (Jack Taylor), and apoprotein genetics and Alzheimer disease, as well as some research pitfalls when a gene is the risk factor (Gail Jarvik).

The second session, "Interaction of Diet/Nutrition and Genetic Susceptibility," chaired by Walter Willett, considered known dietary risk factors for a variety of cancers and the evidence that metabolic and receptor polymorphisms might modify risk. Allan Hildesheim discussed nitrosamines, cytochrome P4502E1, and nasopharyngeal cancer; Katherine

McGlynn, folate, alcohol, methylenetetrahydrofolate reductase, and colon cancer; Rashmi Sinha, heterocyclic amines, cytochrome P4501A2, *N*-acetyltransferase, and colorectal cancer; Douglas Bell, the variation in NAT1 and gastrointestinal cancers; Lea Harty, alcohol, alcohol dehydrogenase, and oral cancer; and Loïc LeMarchand, the possible genetic/environmental explanation for the recently acquired high risk of colorectal cancer in Japanese Americans. These presentations resulted in a prolonged discussion among the formal discussants (Robert Hoover, Nicholas Lang, Steven Tannenbaum, and Regina Ziegler), the presenters, and the other workshop members. The themes that emerged from all the presentations and the discussion sessions are summarized below.

The third session, chaired by Allan Conney, "Emerging Hypotheses," explored some newer and more speculative areas. This included a discussion of hormone metabolism in breast and prostate cancers (Brian Henderson) and a molecular genetic perspective on DNA repair and cell-cycle control and emerging evidence that these processes may be influenced by diet (Bernard Weinstein). Genes and chemical carcinogenesis and how animal models cast light on diet and cancer were discussed by Fred Kadlubar. Alisa Goldstein shifted the focus to explore the intriguing question of whether high-penetrance, low-prevalence susceptibility genes may interact with diet. John Potter drew attention to the possibility that yet another wrinkle (bulge?) in the diet/gene/cancer picture was the possible role played by genes that influence obesity. The formal discussants were Colin Campbell, Gloria Petersen, Nancy Potischman, Elio Riboli, and Thomas Sellers. Both the presentations and ensuing discussions highlighted the complexity confronting researchers in this emerging, interdisciplinary field of diet and genetic susceptibility.

The final session, "How to Advance the Field," provided perspectives from both the population sciences (chair, Robert Hoover) and laboratory sciences (chair, Fred Kadlubar). The workshop heard from Sholom Wacholder and Patricia Hartge on some of the statistical pitfalls and opportunities, from Monserrat Garcia-Closas and Nathaniel Rothman on misclassification, from Stephanie London on subgroup analysis and publication bias, and from Kenneth Buetow on genetic approaches to the analysis of complex traits. Lionel Poirier presented perspectives on resolving differences between animal and human data. Steve Hecht discussed the complexity of exposures and biological pathways. Johanna Lampe spoke about the way in which diet can influence metabolic phenotype. Finally, Helmut Bartsch presented data on some exciting new DNA-adduct biomarkers derived from lipid peroxides. The formal discussants (Christine Ambrosone, Douglas Bell, Neil Caporaso, David Eaton, Richard Hayes, David Hunter, Arthur Schatzkin, Mark Schiffman, Peter Shields, and Rashmi Sinha) and the participants had a lively discussion, sobered somewhat by evidence that sample sizes required for studying the interplay between genes and diet appear to be very large indeed.

The meeting concluded with Colin Campbell, Lawrence Kolonel, Steven Tannenbaum, Alfred Knudson, and Robert Hoover making summary observations on future research di-

Received 6/2/97; accepted 6/3/97.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> To whom requests for reprints should be addressed, at Nutritional Epidemiology Branch, National Cancer Institute, Executive Plaza North, Room 430, 6130 Executive Boulevard, Rockville, MD 20892-7374.

reactions in a session chaired by Barbara Hulka. John Potter summarized the themes that had emerged from the presentations and especially the discussions.

### The Workshop Themes

Coherent themes and recommendations covered five major areas: technical issues, research approaches, underlying biology, resources, and the broader societal perspective.

**Technical Issues.** The major technical issues discussed relate to study design, analytic methods (both laboratory and population), and the need for quality control and data management expertise.

In relation to study design, the importance of well-designed epidemiological studies with high response rates (including for any biomarker component), adequate sample sizes, and accurate exposure measurement was emphasized. There was a strong view (although not everyone agreed) that we should be moving away from case-control to cohort designs, not because of the genetic questions but because these designs are better able to reduce recall and selection bias and their influence on dietary (and other self-reported) data. However, it was noted that there are different advantages in using case-control study designs; these are relatively quick and can explore new hypotheses requiring data that may not have been collected in existing cohort studies. Furthermore, the sample size requirements to examine the interplay between genes and diet are similar for both cohort and case-control studies.

Studying various ethnic groups provides opportunities for examining gene/diet interactions. Studies involving several ethnic groups have increased heterogeneity of both genetic background and diet, thus increasing power. If certain gene-diet interactions are confirmed and the importance of specific carcinogens varies across ethnic groups based on gene frequency, this will have important implications for prevention. Appropriate attention always needs to be paid to differences by ethnicity and risk status, because specific ethnic groups and high-risk populations are likely to vary both by genes and behavior.

There was a strong perceived need for close attention to maintaining statistical power, especially given the focus on interactions and specific subgroups. It is plausible that the number of subjects in studies investigating gene-diet interaction need to be in the thousands. The important determining factors are prevalence of the genetic polymorphisms and dietary exposure, degree of misclassification of each, as well as the strength of effect.

As we develop this area of research, the improvement of analytical methods is central. Identified areas include the need for specially tailored or refined food frequency questionnaires that are focused on a specific set of related dietary exposures (this is in addition to, not instead of, the broader instruments). Comments on the need for better biological measures of exposure recurred regularly. It was noted, again more than once, however, that biological methods of measuring exposure are themselves limited (for example, high intraindividual variation results in a short-term marker that is not appropriate for measuring usual dietary exposure) and that, for some exposures (for example, age at menarche, lifetime smoking history, or adolescent diet), a biological measure is not easy to imagine. The likelihood of new high-speed genetic methods (for example, the emerging chip technology) is much more certain and more likely to add enormously to our ability to collect genetic data, although not, in itself, to our ability to manage and interpret such data.

That leads to the third area of need in technical develop-

ment, appropriate quality control and data management. Here, the need is for scientists who are well versed in conducting cross-disciplinary studies [such education programs are being developed at the Fred Hutchinson Cancer Research Center (Seattle, WA) and the Division of Cancer Epidemiology and Genetics, National Cancer Institute (Bethesda, MD)] and for laboratory techniques for handling large numbers of samples and computing methods for analyzing multiple potential interactions in large datasets.

**Research Methods: An Ensemble of Approaches to Gene/Diet Studies.** As the meeting progressed, the need for an ensemble of approaches, not only for diet and genes, but essentially for any question involving both environmental data and genotyping, became clearer. The needed models and study designs extend from molecules to populations and include cell models, including simple organisms, knockout cells, and affected germline cells; studies of tissues; animal studies, both knock-out mice (and rats?) and standard carcinogenesis experiments; small intensive human studies, both observational and those involving metabolic experiments; and population-based studies, including studies of families, other high-risk groups, and studies of the general population.

Against this background, there was a strong emphasis on the employment of the hypothesis-driven nature of good epidemiology as well as exploiting the phenomenological methods that characterize current molecular genetic research. There was a strong preference for hypothesis-driven (what are the good questions?) rather than technology-driven (what is possible?) approaches.

Throughout the meeting, the complexity inherent in the studies of diet and nutrition was emphasized: diet is a complex mixture of compounds, including both carcinogens and anti-carcinogens; looking at one particular compound in isolation does not provide an adequate picture. Furthermore, even though measuring usual diet is laden with problems, it is crucial that we improve dietary-exposure assessment so that the studies examining joint effects with common polymorphisms can provide answers rather than more confusion.

There was a clear consensus that, optimally, laboratory and population scientists should participate jointly in every phase of the study; this has implications for learning each other's languages and methods as well as implications for the development of training programs, as noted above. It was agreed that there is a broader need to encompass the implications, for public health and clinical medicine, of rapidly evolving genomic research methods, moving beyond the laboratory setting to consider how advances and discoveries in basic research affect, and can be translated into, clinical and public health settings.

Finally, it was agreed that there is a need not only for analysis (*e.g.*, how does this gene work?) but also for synthesis, putting the pieces of pathways, molecular changes, and disease progression together into coherent views of pathogenesis.

**Some Biological Questions.** Because the study of genes in population settings is rapidly becoming commonplace, there is a need to think beyond the rather restricted pattern of questions that have characterized the first generation of studies. Some possibilities include: (a) structure/function studies: considering gene function and expression as well as looking for elevated risks associated with particular patterns of mutations; (b) genotype/phenotype studies: studying how closely genotype and phenotype are related for any one gene or gene family (metabolizing enzymes, DNA repair, and others). Such studies should help unmask redundancies in the protection against cancer as

well as help uncover an almost certain high level of complexity in the relevant biology, *e.g.*, induction or repression of gene expression by external agents, multiple pathways with overlapping influences, proteins with multiple functions, and so forth; (c) studies going beyond the role of a single gene in a complex pathway: we have tended to focus on mutations/polymorphisms in known candidate genes for specific enzymes rather than seeking abnormalities across candidate pathways; (d) identification of rate-limiting steps in the complex pathways: these can be measured and exploited in epidemiological studies; (e) studies of the specificity of gene-environmental exposure relationships: because we are at an early stage, we know little, as yet, about these; (f) population studies: these will allow us to detect selective survival of specific genotypes manifested as gene frequencies that change with age, a phenomenon that appears to characterize the apolipoprotein E/Alzheimer's disease story; (g) studies of species specificity: these will allow us to choose appropriate models for future study. We know that some genes, such as cell-cycle and DNA repair genes are highly conserved from simple to complex organisms, but the number of members of any one family and their pleiomorphic effects appear to increase across evolutionary time. Furthermore, even the array of proteins between species, *e.g.*, the P450 enzyme profile in rats and humans, can be markedly different; and (h) studies that move beyond the activating and detoxifying polymorphic enzymes to consideration of steroid, retinoid, and vitamin D receptors; DNA repair enzymes; cell cycle regulators; transcriptional factors; and others.

**Resources.** One of the ideas that echoed across presentations and in the discussion periods was the need to pursue the anomalies. However, detecting whether a particular finding is anomalous is particularly problematic if the universe of findings is not available. Accordingly, there was a strong feeling that publication bias against null studies was a detriment to the development of the field. The need to publish something other

than null associations forces researchers to squeeze findings out of their data, thus adding to the confusion. As a result, only real and imagined "positive" findings tend to be found in the literature. The appropriate response would be the establishment of a database and a clearing house for null studies, studies with sufficient power that, nonetheless, found no association between a specific gene and a particular disease. One important question is what criteria should be used to include studies in this database. The World Wide Web has made such a proposal feasible, but it still needs energy and personnel for its establishment and maintenance.

Another major resource need, on a much grander scale, is for a coherent approach to the characterization of the spectrum of human genetic variability—not just a Genome Project (whose genome?), but a Population Genome Project.

**The Broader Perspective.** In the United States, and perhaps uniquely in the United States, the possibility that the identification of specific genetic variations that influence disease risk might become grounds for loss of health insurance is already a serious worry for many. At present, the primary concern relates to the high-penetrance, high-risk, low-prevalence genes such as *APC*, *BRCA1*, *Huntington*, and so forth. What would be even more deeply troubling would be the characterization of normal human genetic variability (including, for instance, where a particular gene carries a slightly elevated for one disease and perhaps a reduced risk for another) as abnormal. If this kind of thinking took hold, whose genome would we regard as normal? We have seen the extreme results of this kind of bigotry, earlier this century, when the ability to identify differences among us was based only on pseudoscience. However, these dangers should be met firmly, not used as a political excuse for stopping research. We need, now, to ensure that good science with the potential to benefit each of us via improved prevention and treatment is not perverted in the name of profit, racism, or ideology.