1st NCI Epidemiology Leadership Workshop
Tobacco, Diet, and Genes
Summary Minutes
September 19-21, 2004

The NCI's 1st Epidemiology Leadership workshop, focusing on tobacco, diet, and genes, was convened at 5:00 p.m. on September 19, 2004, in Chicago, Illinois. The first evening began with a poster session, introductions, and an overview of the activities of the Epidemiology and Genetics Research Program (EGRP), a component of the Division of Cancer Control and Population Sciences (DCCPS), National Cancer Institute (NCI). The evening ended with a keynote address. The second day featured a poster session, a state-of-the-science plenary session, and concurrent Working Group meetings. The third day concluded the meeting with Working Group reports, a second keynote address, and a panel discussion on the future of consortia.

Sunday, September 19

Edward Trapido, Sc.D., and Robert Croyle, Ph.D.
(DCCPS)
Workshop Objectives

Dr. Trapido, EGRP Associate Director, explained that the purpose of the workshop was to: (1) identify barriers and gaps in cancer epidemiology related to the topics of tobacco, diet, and genes and facilitate the conduct of research in these areas; and (2) highlight the research achievements of EGRP grantees. Questions to be addressed during the workshop included:

- What is the role of epidemiology in working with new technologies?
- How can changing U.S. demographics be used as a cancer epidemiology laboratory?
- How will cancer epidemiologists adapt to global access to data and changing requirements for data sharing and confidentiality?
- What will the impact be of changing health care on epidemiologic research?
- What kinds of shared resources are needed?

Dr. Croyle, DCCPS Director, said that this workshop is one in a series of activities to be undertaken as part of a review of the EGRP, as has occurred in other parts of DCCPS. It is an important opportunity to identify the epidemiologic questions in tobacco, diet, and genetic research that need support and facilitation and what is needed to overcome barriers. It also is a key opportunity to visit where the field of cancer epidemiology has been and where it is going. He charged the participants to look broadly and think about growing the science 5, 10, and 15 years from now while being realistic about funding. He asked that they consider how cancer epidemiology can be integrated into and inform other areas of biomedical research, such as genomics and proteomics. Participants also were asked to identify needs and scientific questions that will drive implementation of scientific findings to benefit public health.

Susan Curry, Ph.D.
(Institute for Health Research and Policy, School of Public Health, University of Illinois at Chicago)
First Keynote Address

Dr. Curry summarized the policy recommendations from *Fulfilling the Potential of Cancer Prevention and Early Detection*, a 2003 report by the National Cancer Policy Board (NCPB) of the Institute of Medicine.[1] The report begins with an imaginary newspaper headline: "*Interventions succeed in preventing 100,000 cancer cases and 60,000 cancer deaths each year* …." Such a headline is within reach because prevention interventions are available now that can reduce the future burden of cancer and reduce the risk for other chronic diseases. The report calls for a national strategy to ensure that the promise of cancer prevention and early detection is realized.

The report examines the effectiveness of cancer screening methods and interventions to alter lifestyle habits, approaches to enhancing the potential benefits of proven interventions, the problem of adopting new technology when the science is uncertain, professional education and training needs, Federal and state programs that support cancer prevention and early detection, and research trends and opportunities.

The NCPB report makes 12 recommendations, but Dr. Curry considers that only 4 initiatives are needed to realize the current potential for reducing the burden of cancer through changes in health behavior and participation in screening for early detection:

- Increase access to and public demand for state-of-the-art interventions.
- Make prevention a "standard of care" in health care delivery.
- Apply successful population-level policy strategies from tobacco use (e.g., smoking regulations and taxes) to other behaviors.
- Increase investment in research to evaluate policy initiatives and study dissemination of proven programs.

The current potential to reduce the cancer burden through prevention and early detection is built on a solid foundation of epidemiological research. There are three important lessons from this work. First, when pulling science through to policy, data become "evidence," and they must be convincing. We may need to examine the standards for convincing evidence in light of changes in research trends. If the standards of consistent findings across a large number of well-designed studies, a demonstrated dose-response relationship, biologically plausible mechanisms, and supportive laboratory evidence are inadequate as evidence from new research directions, then the field needs to articulate new standards. For example, with an almost infinite number of gene-environment and gene-gene interactions to examine as determinants of biological and/or behavioral risk factors for cancer, we need clear guidelines for determining "what is important."

Second, models that estimate potential reductions in cancer incidence and mortality have confined themselves to strength of association and biological latency. The models do not take into account social or political latency to implementation. When we do take these factors into account, it is humbling to see huge lags between discovery and development and more disappointing chasms between development and delivery. One estimate is that it takes 17 years for a clinical innovation to move from research into

[1] Curry S, Byers T, Hewitt M. *Fulfilling the potential of cancer prevention and early detection*. Washington, DC: Institute of Medicine; 2003. Available from: http://www.iom.edu/reports.asp.

practice. Perhaps one reason for this is that the people and systems that are responsible for delivery—through health care, community initiatives, and policy changes—are not the people engaged in discovery and development. If we recognize that optimizing the impact of our discoveries is in the hands of others, then an important question is: "What is the pull-through? Who has to know what? What resources will be needed? And what infrastructure is in place that can be leveraged to get new innovations into practice sooner rather than later?"

Third, we cannot underestimate the importance of acting on what we know now. Policy and practice cannot and should not wait for the "magic answer."

One-third of cancer deaths are preventable, and even if not all at-risk people will stop smoking or lose weight or become physically active, reduction of an estimated 60,000 deaths per year is achievable. The message is simple: that the ultimate goal is to use science to improve public health. It is up to health leaders to minimize the gap between what we know and what we do.

Monday, September 20

Laurence Kolonel, M.D., Ph.D.

(Cancer Research Center, University of Hawaii)

Perspectives on Diet, Genes, and Cancer

Dr. Kolonel provided a perspective on the status of diet, genes, and cancer in epidemiologic research. Several early studies in Hawaii demonstrated that cancer incidence varies by ethnic group and that cancer risks change dramatically in migrants from Japan to Hawaii, with further changes in their offspring. These findings support the conclusion that environmental and lifestyle factors are the major determinants of cancer risk.

Historically, epidemiologists have addressed the relationship between nutrition and cancer using a reductionist approach, initially emphasizing high-fat foods, then progressively examining animal versus vegetable fat sources, total intake of fat, saturated versus unsaturated fat, polyunsaturated versus monounsaturated fat, omega-6 versus omega-3 polyunsaturated fat, and long-chain omega-3 versus short-chain omega-3 fat, ultimately focusing on individual fatty acids. Over the past 30 years, the major feature of this research has been inconsistency. Proposed reasons for this inconsistency include measurement error, the multifactorial nature of causation, and individual variations in susceptibility. To this, perhaps, should be added the inappropriateness of a reductionist approach to research in this area.

The influence of diet on carcinogenesis may be less specific (with regard to constituents) than has been assumed. Given the fact that humans have evolved (i.e., are adapted) to a harsh environment in which the consumption of certain types of foods is necessary to sustain the organism, it seems unlikely that the species is so fragile that ingesting too much of one particular fatty acid or too little of one particular carotenoid is the "cause" of any site-specific cancer. Indeed, there appear to be common mechanisms of carcinogenesis (e.g., oxidative stress), and nutrition-related cancer more likely reflects a macro-level metabolic imbalance (such as between agents that promote and agents that reduce oxidation).

Based on the weak associations usually found in dietary studies of cancer and on such observations as the "excessive" risk of colon cancer in Japanese migrants to Hawaii, researchers have concluded that genetic susceptibility factors must account for at least a component of individual and group differences in risk for specific cancers. However, 10 years of genotyping research have—as with dietary research—produced inconsistent results. Although measurement error is a less likely explanation in this instance than in dietary assessment studies, other factors may be implicated, including lack of adequate power to examine

interactions, inability to distinguish between markers and true causal variants, population stratification and ethnic admixture, and the complexity of metabolic pathways. As with dietary exposures, holistic rather than reductionist research strategies may be useful in approaching questions concerning genetic susceptibility.

Overall, it seems clear that a diet to reduce cancer risk must entail a proper balance between plant and animal food sources and between the intake and expenditure of energy. Though we have not yet completely elucidated the specific elements of such a diet, the current recommended dietary guidelines will probably prove to be correct in the end.

Neil Caporaso, M.D.

(Division of Cancer Epidemiology and Genetics [DCEG], NCI)

Perspectives on Tobacco, Genes, and Cancer

Dr. Caporaso provided perspective on epidemiologic research on tobacco, genes, and cancer. In recent years, there has been a reduction in per capita tobacco consumption in the United States and a correlated drop in lung cancer incidence. However, tobacco is still a major cause of cancer mortality worldwide, as well as of mortality due to heart disease and other conditions.

Conventional treatment of disseminated cancer is not sufficiently efficacious. Real progress in reducing the burden of cancer requires a focus on molecular targets and pathways that are critical for development and maintenance of cancer phenotypes. Genetics can be discussed in terms of populations, families, and somatic mutations.

To study genetic risk, it is necessary to focus on the gene-environment relationship and how that relationship is associated with disease. Many studies of susceptibility genes in lung cancer are incomplete and face serious challenges, such as numerous false positives and negatives. Larger and better-designed studies are needed that focus on well-characterized genes in logical groups. It may be necessary to develop a whole-genome approach to lung cancer research. This future cancer research will focus on molecular targets and pathways utilizing and standardizing the newest technologies. One example of a standard method for selecting candidate genes includes answering the following questions: (1) Does the gene have a plausible function? (2) Does it contain function-altering polymorphism? (3) How prevalent is this polymorphism in a defined population? In addition, proteomics will advance cancer research by characterizing protein differences between cancer cases and controls. Creating protein profiles that correspond with stages of tumor development will not only increase the rate of cancer diagnosis, but also provide novel drug targets to stop the spread of disease.

Given the complex scientific questions involved in lung cancer research and the public health burden, larger integrated studies and research consortia are needed. A large, interdisciplinary lung cancer case-control study has been initiated that integrates molecular epidemiology (focusing on behavior, exposure, genetics, and disease) and translational medicine (focusing on outcomes).

Advantages of consortium-based research include large sample size, the ability to create subgroups, increased statistical power, standardization, rapid confirmation, pooled resources, and access to advanced technology. Disadvantages include slow progress, the burden of managing resources and support, territorial issues when acknowledging success or failure, and some weaknesses in the ability to conduct meta-analyses. Despite the disadvantages, the consortium approach is still an attractive design when assessing multiple outcomes in epidemiologic research.

Stephen Chanock, M.D.
(Center for Cancer Research (CCR), NCI)
Utilizing Human Variation in the Human Genome

Dr. Chanock said that studying genetic variation in the human genome by using single-nucleotide polymorphisms (SNPs), the most common genetic variants in the genome, represents an important tool to dissect the extent of genetic contribution to specific cancers. It is likely that there will be differences in the contribution of genetics according to cancer types. Moreover, specific variants could be informative in determining the risk for cancer and its outcomes, including both toxicity and efficacy of drug therapy. To explore the genetic contribution of common variants to cancer, it is necessary to carefully annotate the extent of genetic variation across a gene or region by defining the common SNPs and their co-inheritance on haplotypes.

Currently, individual genotyping is expensive. Particular care must be extended to annotating the nearby SNPs, which could alter the validity of an SNP assay. Two parallel efforts have been advancing that are producing critical maps of key SNPs to be applied in case-control and cohort studies. The first is the generation of an authoritative map of SNPs across the genome, known as the International HapMap (http://www.hapmap.org), which genotypes SNPs at a high density across the genome in reference groups from Africa, Northern European, and Pacific Rim backgrounds. The second approach is based on resequence analysis of genes or regions around known SNPs of high interest, such as the SNP500Cancer (http://snp500cancer.nci.nih.gov), which has more than 7,500 sequenced SNPs in its database.

Studies designed to identify common variants in cancer susceptibility can be performed in the case-control or cohort setting. For the foreseeable future, candidate genes—either singly or in sets—usually drawn from interesting biological pathways will be applied to studies. In this regard, candidate genes must be chosen wisely, and populations must be selected with care. However, some candidate genes are difficult to study because there is sequence homology with other genetic segments, precluding the ability to amplify unique regions. Accordingly, the study design and hypothesis must be carefully considered, and positive results require validation in follow-up studies and, perhaps, different populations.

For genetic variation to be useful as an early intervention or predictor, well-designed studies will need to identify haplotypes, followed by careful analysis including genetic and laboratory-based investigation, which will finally characterize the “causal” variants. In turn, these will be considered for screening or public health approaches to early detection and personalized medicine.

Virginia Hartmuller, Ph.D., R.D., F.A.D.A.
(EGRP, DCCPS, NCI)
Opportunities at NIH in Diet and Energy Balance

Dr. Hartmuller said that many research grant opportunities are available in the area of diet, energy balance, and cancer. Details, including funding mechanisms, eligibility, and submission deadlines, can be found by searching under the Program Announcement (PA)/Program Announcement Reviewed (PAR) or Request for Applications (RFA) numbers on the NIH Office of Extramural Research Web site at http://grants1.nih.gov/grants/guide/index.html. Following is a list of these opportunities by title:

- Small Grants Program for Cancer Epidemiology (PAR 03-010)
- Improving Diet and Physical Activity (PAR 03-009)
- Exfoliated Cells, Bioactive Food Components, and Cancer (PA 04-114)
- Transdisciplinary Research on Energetics and Cancer (TREC) (RFA CA-05-010)

- TREC Coordinating Center (RFA CA-05-011)
- Studies of Energy Balance and Cancer in Humans (PA 04-124)
- Obesity and the Built Environment (RFA ES 04-003)
- Diet, Composition, and Energy Balance (PA 04-033)

Further information on NIH diet and energy-balance activities is available at the following NCI Web sites:

- NCI Energy Balance: Weight and Obesity, Physical Activity, and Diet: http://www.cancer.gov/cancertopics/energybalance
- NIH Obesity Research: http://obesityresearch.nih.gov
- NIH Obesity Research Funding Opportunities: http://obesityresearch.nih.gov/funding/funding.htm

Deborah Winn, Ph.D.

(EGRP, DCCPS, NCI)

Bioinformatics for Cancer Epidemiologic Research

Dr. Winn described the DCCPS Health Informatics Initiative, which is a collaboration among DCEG, DCCPS, NCI Center for Bioinformatics (NCICB), and the extramural community. The scope of the initiative includes addressing informatics needs related to surveillance systems, behavioral research, intervention research, cancer outcomes, quality of care, and survivorship. Its primary goal is to transform research in cancer control and population sciences by using the benefits of health informatics to:

- Improve the efficiency and reduce the difficulties of conducting research.
- Increase cross-disciplinary and collaborative research, especially through the use of standards and common platforms.
- Create opportunities for data and other linkages.
- Increase capacity for research and the use of research to understand mechanisms of cancer causation, monitor the burden of cancer, develop interventions to prevent cancer, detect it earlier, reduce its consequences, and inform policy decisions.

Trends influencing the need for a bioinformatics initiative include: the NIH data-sharing policy; increasing collaboration among scientists who may "speak different languages"; growth of consortia; the need for "crosswalks" across existing studies; epidemiology studies involving data from many content domains (e.g., genotypes, biochemistry, clinical records, questionnaire data); increased pressure to conduct research faster, better, and cheaper; the need to translate from biology to epidemiology to prevention and back again; NIH *Roadmap* components related to bioinformatics and computational biology (http://nihroadmap.nih.gov); and strong NCI bioinformatics development efforts in other areas (e.g., International HapMap Project).

Activities within the initiative include:

- An Enterprise Vocabulary System (EVS) for population sciences and cancer control concepts.
- A population sciences and cancer control "workspace" within the NCICB cancer Biomedical Informatics Grid (caBIG) project.
- Models and applications (e.g., a risk-prediction workshop).

The EVS is a set of services and resources that addresses NCI's needs for controlled vocabulary and common data elements. The NCI Thesaurus and Metathesaurus, which is a biomedical thesaurus created

specifically to meet the needs of the Institute, is produced by the EVS project. The Common Data Elements (CDE) Dictionary contains a repository of detailed information pertaining to each standard element used to collect essential data related to cancer research. The CDE facilitates development of questionnaires and forms and can be used to compare data elements from different studies.

The benefits of common data elements and establishment of a caBIG workspace include:

- Building a repository of information about the data collected by studies.
- Facilitating the creation of questionnaires and forms.
- Comparing data elements from different studies.
- Crosswalking between studies if they have been added to the EVS.
- Incorporating common data elements and thesauruses from other disciplines.
- Ensuring that other disciplines can benefit from current knowledge about appropriate ways to collect information related to population sciences and cancer control.

The EGRP-supported Cancer Family Registries (CFR) Informatics Center is being established to make the informatics practices of the registries consistent with NCICB standards. The Center will convert all existing CFR questionnaires and forms (e.g., family history information) into common data elements. DCCPS invited interested attendees to provide questionnaires that they would like to share with others and asked for extramural input and expertise in a variety of capacities.

Useful NCICB Web sites include:

- NCICB home page: http://ncicb.nci.nih.gov.
- Cancer data standards repository: http://ncicb.nci.nih.gov/core/caDSR.
- Common data elements browser: http://cdebrowser.nci.nih.gov/CDEBrowser/.
- NCI Metathesaurus Web page: http://ncimeta.nci.nih.gov/indexMetaphrase.html.

Mukesh Verma, Ph.D.
(EGRP, DCCPS, NCI)
Opportunities in Epigenetics

Dr. Verma explained that one goal of EGRP's Analytic Epidemiology Research Branch (AERB) is to stimulate population-based research on the application of epigenetic markers to cancer epidemiology. Genetic and biochemical markers are currently in use, but the potential of epigenetic and proteomic markers has yet to be fully explored. Epidemiological biomarkers are used in cohort studies to follow disease prevalence to identify populations at high risk. Use of biomarkers in population-based studies raises questions of reliability, sensitivity, specificity, reproducibility, and scalable capacity for automation—issues that are central when selecting assays to determine factors associated with carcinogenesis.

Epigenetics is the study of mitotically heritable changes not caused by DNA sequence alterations. Epigenetic control of gene expression is essential for normal cell behavior such as cell cycle control, DNA damage repair, apoptosis, and imprinting, but misdirected control can result in the silencing of tumor-suppressor genes or the activation of proto-oncogenes. Epigenetic control is regulated by environmental, hormonal, and genetic factors. Epigenetic changes can be reversed, which provides an opportunity for chemical intervention.

Hypo- and hypermethylation of gene regulatory regions and the deacetylation and acetylation of histones

in chromatin regulatory domains are important for normal cell growth and development. Any abnormality in these processes may result in tumor development. The expression of key regulatory genes in tumor suppression, metastasis, DNA repair, and angiogenesis pathways have been shown to be regulated by hypermethylation. Inhibitors of methylation and deacetylation such as Zebularine and Trichostatin are used clinically to help control the epigenetic effects, and these drugs may be useful for chemoprevention.

Several potential epigenetic markers identified in different cancers have proven useful for screening and determining disease progression. For example, the extent of promoter DNA methylation has been shown to affect the probability of survival in esophageal cancer and can determine class of tumor in acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). Assays have been developed for determining methylation state from biofluids and exfoliated cells using automated, high-throughput technologies. In conclusion, epigenetic markers can be used for cancer detection and risk assessment to identify populations at high risk of developing cancer.

Daniela Seminara, Ph.D., M.P.H.

(EGRP, DCCPS, NCI)

Epidemiology, Consortia, and "Large-Scale Science": Challenges and Opportunities

Dr. Seminara said that the challenges confronting large collaborative science initiatives in epidemiology relate to the breadth of study goals, the need for definable deliverables and endpoints, and the long-range strategic planning and timeframe. Initial infrastructure costs are high, and the large financial and intellectual investment requires management and oversight by the funders as well as by multiple investigators. Large projects also require extensive bioinformatics support. However, the exchange of concepts, approaches, and intellectual discourse produces results beyond those possible with single-discipline studies. Large science projects may be seen as a fundamental requirement for obtaining answers to complex scientific questions within the population science framework of epidemiological studies.

Appropriate infrastructure requirements for large science projects include large, well-ascertained populations and the integration of cutting-edge genomic/molecular technologies, as well as complete exposure, outcome, clinical, and molecular assessment. Consensus building is also critical, and NCI continues to express its interest and support for large interdisciplinary epidemiology studies.

EGRP fosters the development of consortia by identifying and assessing research priorities to provide resources, coordination, and communication among participating groups and with other consortia. In cooperation with investigators, EGRP can facilitate the research implementation and translational process to evaluate performance, develop milestones, and establish "best practices." EGRP also works closely with the principal investigators of cooperative agreements and contracts to ensure that study participants receive credit for their efforts, that skilled grantees manage projects, and that all NIH funders provide oversight.

Of particular interest for the support of consortia is a novel mechanism recently approved by NIH for support of interdisciplinary research. Grant applications for this new consortium model should each focus on a specific and synergistic aspect of the team effort. The number and types of mechanisms for each consortium vary, but each must include a Project Leadership and Management Grant headed by an eminent scientist who can articulate the vision and science of the team effort. Consortia Grants (RFAs) may be issued to solicit additional grant applications to allow researchers to join a consortium. Satellite Grants may be issued to enable researchers to become affiliated with, rather than integrated into, a consortium. All grantees would be expected to understand that continued support depends on the progress

of each grant and the achievement of consortium goals and that oversight would be by a cross-Institute extramural team.

EGRP-supported epidemiology consortia are structured to allow considerable design flexibility. One example is the Breast and Colon Cancer Family Registries (CFRs). The CFRs have the largest and best-characterized biospecimen collection of breast and colon cancer population-based families in the world, which, combined with associated risk factor data, supports high-throughput molecular technologies and gene-environment studies. The consortia are driven by an interdisciplinary research agenda that can adapt to advances in the science. Proposals for collaborative protocols are strongly encouraged from national and international groups with appropriate expertise.

A second example is the Cohort Consortium formed by NCI (EGRP/DCEG) to study gene-gene and gene-environment interactions in the etiology of cancer. General cohort studies worldwide with more than 10,000 subjects, blood samples, and questionnaire data on important cancer risk factors were invited to participate. Many of these cohorts are currently involved in collaborative studies of gene-environment interactions in breast, prostate, and pancreatic cancer, and other "spinoff" studies are anticipated.

The challenges facing future consortia are extensive and cover diverse issues such as informed consent, support for large data sets, centralized biorepositories, authorship, and appropriate criteria for evaluation and measures of productivity. EGRP will continue to work with NCI, the extramural community, and investigators to find solutions to these challenges.

Graham Colditz, M.D., Dr.P.H.

(Brigham and Women's Hospital)

Team Science and Tenure: Is There a Conflict?

Dr. Colditz said that "team science" or participation in consortia is becoming common and necessary as funding budgets shrink and large data sets are required to answer critical questions. Consortia-type activities are already a common feature of the epidemiology programs at NCI—approximately 10 percent of principal investigators with projects in the current portfolio already participate in them. These opportunities include the Specialized Programs of Research Excellence (SPOREs), Early Detection Research Network (EDRN), Interdisciplinary Research Centers (IR Centers), Transdisciplinary Tobacco Use Research Centers (TTURCs), and consortia. However, there are no guidelines defining the role of each team player within a consortium. Although this "division of credit" is hard for all researchers, it is most difficult for junior investigators who are focused on obtaining tenure. Under the current system, it is difficult to evaluate the impact of an investigator's work or identify his or her individual contribution. Therefore, NCI must work with the extramural research community to develop guidelines that will clearly delineate the contribution of each investigator.

As a leader in the field of epidemiology, EGRP can provide solutions by developing mechanisms to identify the contribution of each investigator engaged in team science. The Program can achieve this in part by using the consortia Web sites and by mandating that individual contributions be footnoted in journal articles. It can support junior investigators by educating academic leaders (deans, provosts, and department chairs) on the purpose, structure, and major contributions made through team science. In this way, EGRP can influence the way team science participation impacts tenure decisions. In addition, it can work to improve the peer-review process for consortia and persuade journals to allow long lists of authors. It should work to maintain funding of the basic components that feed into consortia activities over time. "Best practices" for consortia are needed, and as they are developed, EGRP can serve as a

repository for them. The challenge of consortia is to frame the evaluation of large endeavors and support the academic careers of their contributors.

Jon Kerner, Ph.D.
(DCCPS)
Translating Data into Action (Optional Lunch Presentation and Discussion)

Dr. Kerner discussed the importance of moving scientific findings into health practice and described the two commonly understood approaches used to disseminate research findings into the community: passive diffusion and active dissemination. Both processes involve the communication of information from one population to another, but dissemination is an active process and diffusion is not. One goal of active dissemination of health information is to improve health outcomes in underserved and disease-burdened populations. The NCI Translating Research Into Improved Outcomes (TRIO) program focuses on three priorities: using surveillance data to identify needs, track progress, and motivate action in the community; developing tools to assist accessing and promoting evidence-based interventions; and enlisting the support and interest of regional and local partners to address infrastructure barriers to implementing evidence-based interventions. Thus, the dissemination of scientific findings is critical to improving health.

Three models of dissemination were described. First, the Knowledge Synthesis Model, which assesses the appropriateness of individual discoveries in the context of systematic reviews of research evidence, was described. Second, Grant Support Models, which provide grant and contract support for dissemination research and research dissemination, were discussed. For example, should there be supplements to cancer epidemiology grants or contracts to facilitate the diffusion and dissemination of epidemiologic research findings? Third, a Partnership Model, which brings together researchers, advocates, and direct-service personnel to work on the shared goal of integrating science with service, was described. An example is Cancer Control PLANET (Plan, Link, Act, Network with Evidence-based Tools), which is a Web portal to access tools for comprehensive cancer control planning, implementation, and evaluation (http://cancercontrolplanet.cancer.gov). The Web portal was developed by NCI in collaboration with the American Cancer Society (ACS), Agency for Healthcare Research and Quality (AHRQ), Centers for Disease Control and Prevention (CDC), and Substance Abuse and Mental Health Services Administration (SAMHSA). Models for systems changes to achieve effective diffusion and dissemination of scientific findings and research products were also discussed.

Dr. Kerner challenged participants to think about findings or products from cancer epidemiology that may be ready for diffusion and dissemination in order to advance research and benefit public health. He suggested a framework for epidemiology and surveillance dissemination activities that takes into consideration funding partners (e.g., other components of the National Institutes of Health [NIH] and Department of Health and Human Services [HHS], state health departments); intermediaries (e.g., professional organizations and societies, nonprofit organizations); and target audiences (e.g., scientists, public health community, general public).

Participants were invited to share their ideas about diffusion and dissemination in the area of cancer epidemiology with leaders of EGRP's KTT: Dr. Deborah Winn and Ms. Linda Anderson.

Working Group Sessions

Four Working Groups met concurrently to focus on Susceptibility in Tobacco Carcinogenesis: Genotypes Versus Phenotypes; Challenges to Diet/Energy Balance Epidemiology Research; Haplotypes Versus Genotypes; and Design Issues and Strategies in the Study of Rare Cancers. The Working Groups were

charged with three tasks: (1) identify and prioritize the scientific research opportunities needed to advance cancer epidemiology, (2) define the scientific resources needed to address these opportunities, and (3) compare and contrast these priorities with the current NCI research portfolio in cancer epidemiology. They also were asked to identify gaps in science, scientific barriers, and potential solutions as they relate to tobacco, diet, and genetic epidemiology research. The Appendix contains summaries of their discussions.

Tuesday, September 21

Working Group Reports

On the second day of the meeting, the Chairs of the four Working Groups reported before all meeting participants on the results of their respective breakout sessions. (See Appendix for lists of members of the Working Groups.)

Challenges to Diet/Energy Balance Epidemiology Research

Chair: Lawrence Kushi, Sc.D. (Kaiser Permanente)
Co-Chairs: Rachel Ballard-Barbash, M.D., M.P.H. (DCCPS, NCI)
Virginia Hartmuller, Ph.D., R.D. (EGRP, DCCPS, NCI)

On behalf of the Working Group, Dr. Kushi presented an overview of the issues discussed. Five priority areas were discussed.

1. Integration of Diet/Weight and Physical Activity

The Working Group said that although most people recognize that obesity and physical activity are linked to the risk of developing cancer, the independent effects of diet, physical activity, and obesity are unknown and thus represent an area of research interest. Current gaps in the science include a need for small-scale intervention studies for biomarkers and large, community-level intervention studies to address social factors. Investigators need better ways to assess changes in physical activity and diet as well as better tools to measure changes in activity that could be incorporated into intervention studies. Also needed are a greater understanding of energy balance at different stages in the life cycle and an understanding of the role of diet and energy balance in rare cancers.

Methodological issues and barriers identified include the need for better tools to measure change in physical activity and diet and tools that allow simultaneous data collection on diet and physical activity.

Regarding the direction of future science related to energy balance, opportunities that the Working Group considers promising include community intervention and obesity initiatives, as well as diet, physical activity, and biospecimen measures at various stages in the life cycle.

2. Biomarkers of Diet

The Working Group noted that sensitive biomarkers are needed to determine the interaction between exposures of nutritional status and their effect on gene expression and cancer risk. Questions raised included: (1) whether serum and tissue biomarkers are adequate, (2) whether it is possible to develop biomarkers for dietary patterns, and (3) how data can be interpreted where biomarkers are available. The Working Group said that biomarkers that incorporate metabolism are also needed.

3. Dietary Assessment

Dietary assessment was identified as a problem area. The standard method in cancer epidemiology, the Food Frequency Questionnaire (FFQ), usually a self-report, is associated with a recognized measurement error, and there is need for and interest in capturing more information on foods, nutrients, and dietary patterns. A critical gap exists in retrospective assessment to help determine the importance of dietary exposures at different stages in the life cycle (e.g., childhood and adolescence). Different assessment tools for different life stages are needed as an approach to diet in the context of cancer diagnosis. This includes, for example, an understanding of the most appropriate time or times to measure diet for cancer etiology and treatment and whether a self-report or an interview is the better way to obtain these data.

Much could be learned from researchers in other disciplines, such as microarray technology, about developing statistical methods for analysis of foods and dietary patterns from existing data. Measuring exposures in childhood and the use of alternative or additional diet tools, such as supermarket purchase patterns, should be considered. Also needed are studies in cancer patients during treatment and survivorship, as disease and treatment could alter nutrient metabolism.

4. Genetic Susceptibility

The best approach to studying genetic susceptibility and diet interactions—for example, SNP or haplotype interaction with dietary factors—is not yet known. What is clear, however, is that large sample sizes will be necessary to study the effects of weak diet and gene effects. The variation in genetic associations is wide, and many studies are underpowered, requiring the application of advanced statistical methods. The Working Group members felt it important to determine how linkage between gene functionality and phenotypic observations will influence interpretation of data. Similarly, it is important to determine when genetic measurements should be used to examine a hypothesis related to a biological mechanism.

Any major emphasis on genetics will need to involve large consortia and multidisciplinary teams looking at macro- as well as micro-level factors. Underscoring these efforts is a matter of clinical and public health relevance given the possible small effects and large amount of funding required to support these projects.

5. Population Coverage

A major issue facing population coverage is ensuring continued efforts to conduct studies in low- and high-risk populations—including international studies, if this is where the questions can best be addressed. It also will be necessary to determine how best to incorporate non-U.S. foods, such as those consumed by recent immigrant populations, into nutrient composition research databases**.** Trans-NIH-funded opportunities will be important.

Discussion also focused on NIH-funded cohorts. The Working Group members felt it important to continue funding ongoing cohorts rather than forming new, large cohorts to study diet-gene interactions. Other important areas of discussion included use of different designs for investigation of diet-gene interactions, cancer prognosis and survival related to diet issues, dissemination of findings, and innovative technologies.

Next steps should include polling meeting attendees via e-mail about their interest in continuing their involvement, as well as the logistics of meeting again at national meetings or participating in conference calls. Participants were asked to identify key topics for evaluation by smaller working groups and subgroups, with the possibility of holding workshops to learn from experts in other fields. The importance of coordinating efforts across appropriate Institutes and Cancer Centers with an interest in diet, obesity, and energy balance was emphasized.

Haplotypes Versus Genotypes

Chair: Stephen Gruber, M.D., Ph.D. (University of Michigan Medical School)
Co-Chairs: Thomas Sellers, Ph.D., M.P.H. (H. Lee Moffitt Cancer Center and Research Institute)
Daniela Seminara, Ph.D., M.P.H. (EGRP, DCCPS, NCI)

The Working Group reported that although haplotype and genotype analysis (specifically, the International HapMap Project) is progressing, there is concern that not enough genetic information is being included from minority populations. Resequencing will take many years to obtain fine-level structures, and coordinated, possibly centralized, resequencing efforts have been considered for the use of haplotypes in gene-environment studies. The optimal analytical design is not clear because of the inherent cost of genotyping, and there are also advantages and disadvantages to pooling versus single-gene haplotype strategies. Both are useful and provide different insights. Guidelines on the state of the science were suggested to help educate colleagues and study sections on this issue, and it was recommended that communication between genomicists and epidemiologists be improved by holding education sessions at professional meetings.

Collecting biosamples to study genotypes as well as haplotypes should be encouraged, recognizing that considerable costs are involved in doing this. The Working Group discussed whether biosamples are needed in all case-control studies and whether new case-control studies should be started. One solution might be to require biospecimen collection in PAs.

Principal investigators also should be educated about the informed consent process and Health Insurance Portability and Accountability Act (HIPAA) regulations to ensure that they are allowed to perform genetic analysis on samples at time of collection and reanalysis and secondary analysis in the future. The Working Group also felt that it is “too soon for dogma” and that educational guidelines and resources should encourage multiple and creative haplotype-based approaches.

A cross-Institute NIH *Roadmap* approach for resequencing SNPs could be considered, as well as the provision of better links to available informatics resources. It was noted that most tools for estimating haplotypes have been developed by statistical geneticists and genetic epidemiologists. Encouraging communication between geneticists and cancer epidemiologists can bridge fields that have developed in parallel and is likely to cross-fertilize both fields of investigation. There is now an opportunity to move the field forward to incorporate greater detail and higher resolution by promoting methodologic development.

NCI needs to take a larger role to encourage collaborative data reanalysis and secondary analysis to compare methods. Policies could be implemented to allow data sharing. Journals also could be encouraged to leave study results online for extended periods and to publish negative results to improve access and availability. An alternative might be to develop strategies to provide researchers with long-term data storage and access. Prioritizing where in the genome researchers should focus their efforts to identify genetic variations (i.e., within coding or regulatory, noncoding regions) may require guidelines.

It is important that genetic variation in underrepresented populations in the United States be studied. Furthermore, tools that are relevant to the analysis of genomic variation need to be developed in parallel with methods for the analysis of epigenetic profiles of cancers secondary to methylation or other somatic regulatory control.

Design Issues and Strategies in the Study of Rare Cancers

Chair: Nathaniel Rothman, M.D., M.P.H., M.H.S. (DCEG, NCI)
Co-Chairs: Sholom Wacholder, Ph.D. (NCI)
Isis Mikhail, M.D., M.P.H., Dr.P.H. (EGRP, DCCPS, NCI)

The goal of this workshop was to solicit input from NCI investigators on the need to study rare tumors. This workshop focused on adult tumors only. A tumor is deemed rare if the incidence is less than 15 per 100,000, approximately equivalent to 40,000 new cases per year in the United States. Childhood tumors were not considered because there is an active epidemiology effort by the Children's Oncology Group that enrolls over 90 percent of children with cancer.

The group discussed why rare tumors should be studied. First, there is an opportunity to find the "low-hanging fruit" in the first investigations of rare tumors, in contrast to additional studies of more common but also far more extensively studied tumors. The payoff might be particularly large for genetic factors if, as suggested by N. Risch from twin studies, lower-incidence tumors tend to have a greater genetic component. In the past, extremely rare tumors with single etiologies have been identified, such as retinoblastoma (RB), angiosarcoma, and clear cell carcinoma (CA) of the vagina. The resulting insights have also helped provide an understanding of the pathogenesis of tumors with a more complex etiology. Three examples are: Al Knudson's work on RB, Fred Li's work on adrenocortical CA, and studies of vinyl chloride and angiosarcoma. Moreover, there are public health reasons for studying rare tumors. While individually low, the aggregated incidence and resultant mortality of all rare tumors combined comprise an important fraction of the total cancer burden. Thus, ignoring rare cancers would close off opportunities to reduce the disease burden.

Individual tumor sites have some characteristics that may make them particularly appropriate to study. Some rare tumors are rapidly lethal (e.g., cancers of the pancreas and esophagus) or have rising incidence rates (e.g., esophageal cancer). From a population perspective, rare tumors can have disproportionately higher rates among some ethnic groups, as seen in the high rates of nasopharyngeal carcinoma among the Chinese population or high rates of male breast cancer among Zambian men in Africa (15%) versus lower rates among men in the United States (0.1%). As a consequence, the Working Group felt that neglecting the study of rare tumors had ethical implications at both the population and individual patient levels.

The group agreed that the first step should be the review of Surveillance, Epidemiology, and End Results (SEER) Program descriptive data to show overall incidence rates, their temporal trends and geographic variation, and racial or ethnic variation. This would help identify tumors that would provide the best opportunity for study and give a sense of feasibility. Other steps include identifying opportunities for using specimens and data in existing cohorts. The Group felt a need to determine where there are ongoing treatment trials to which a case-control study can be added, as is commonly done in pediatric cancers. Finally, the distinct, albeit costly, advantages of *de novo* efforts were considered at length.

Three study designs were discussed in terms of their utility in studying rare tumors:

- *Cohort Studies.* Pooling data and, potentially, biologic samples from existing cohorts would be cost-effective and could be done relatively quickly. However, the degree of completeness of case ascertainment for rare tumors is unclear and would need to be considered. Given the limited number of cases, this study design would allow the identification of moderate to strong risk factors for the rarer tumors.
- *Clinical Trials.* Clinical trials ascertain cases and collect biospecimens already and, thus, can be a helpful setting for a case-control study, as shown for the pediatric cancers. There are a number of methodological issues that need to be considered, including those factors that determine which cases

are enrolled into such trials. However, as the key challenge in the study of rare tumors is sample size, efforts should be made to explore taking advantage of trials.

- *De Novo Designs.* New standalone case-control studies are ideal but can be expensive, especially as they are likely to require multiple centers. SEER-based studies could provide an infrastructure to carry out population-based studies. A key advantage of hospital-based studies would be the ability to collect pretreatment blood samples and tumor tissue and to process samples in ways that allow state-of-the-art analysis. These studies could use molecular tumor assays for defining disease subgroups and in studies of gene expression and proteomics and could be integrated with studies of prognosis and treatment. Studying multiple tumors simultaneously in a case-control study offers economy of scale from, perhaps, a single core questionnaire and biospecimen collection protocol. A study that is hospital-based, particularly at major cancer centers, seems particularly sensible, since it would allow a potentially relatively rapid accrual of cases. Hospital or clinic controls would be appropriate. The disadvantages are not as severe as believed, and the Working Group felt that researchers cannot afford to be overly fastidious, particularly for initial studies. In fact, strong apparent risk factors will be quite unlikely to be artifacts of well-designed and well-conducted hospital-based case-control studies.

One strategy is to capitalize on existing infrastructure is the NIH-funded General Clinical Research Centers (GCRCs). A local principal investigator could be in charge; involvement of an orphan disease program would provide focus; and GCRC facilities could be used to collect blood samples for vigorous phenotype/genotype assessment. To initiate these studies, supplemental funds could be provided to Cancer Centers to explore feasibility. Once significant pilot data are collected, R01 grant applications could be submitted to expand the research.

Other general suggestions included creating a common rare tumor protocol; collecting baseline information on all rare tumors using pooling and data sharing; banking samples; conducting international studies comparing higher rates in other countries with focus on U.S. population (e.g., liver cancer in Asian populations in Asia vs. Asian-American immigrants in California); open-minded, innovative use of controls, such as using “reasonable but not perfect” controls (e.g., relatives and friends of patients and routinely screened persons) or no controls at all; and, because of cancer rarity, looking for high relative risk, as low relative risk seems not to be cost-effective from a research portfolio perspective. Studies using data in innovative ways include the Cancer Research Network (an HMO funded by NCI) and Kaiser Permanente data studies. Studies pooling common tumors from preexisting studies could use data from the Dutch study (n=80,000), the Greek-Italian study, the Iowa Study, and others, identifying the number of cases from these studies.

Susceptibility in Tobacco Carcinogenesis: Genotypes Versus Phenotypes

Chair: Peter Shields, M.D. (Lombardi Cancer Center, Georgetown University)
Co-Chairs: Stephen Marcus, Ph.D. (NCI)
Xifeng Wu, M.D. (The University of Texas M.D. Anderson Cancer Center)

The Working Group felt it important to define phenotypes in terms of looking at the molecular, microscopic, organ, or whole-body level in order to better understand genotypes and haplotypes. These phenotypes, which are often used interchangeably with intermediate or early-detection markers, include phenotypes of behavior (patterns of smokers, choice of cigarettes), exposure markers (metabolites and metabolic profiles, expression profiles), and biologically effective dosage (DNA adducts).

Other phenotypes include markers of harm that integrate susceptibility and exposure, such as mutations and methylation profiles in morphologically normal or premalignant cells; markers of susceptibility that affect DNA repair capacity; tumor phenotypes (because mutations or genetic instability within tumors

themselves may be important); and clinical outcomes and cancer prognosis. Other tobacco-related diseases or markers (chronic obstructive pulmonary disease, lipid profiles, or white blood cell counts) could be considered phenotypes, as could symptoms, such as coughing or shortness of breath at an early age, and comorbid traits, such as drinking and depression.

Using genotypes to determine cancer risk has several advantages: SNPs are widely available; they can be predictive of how the host will respond over a lifetime; and their analysis is inexpensive. Disadvantages include an often lower statistical power and the fact that there are nonreplication issues. Phenotypes represent complex genotypes and so have a conceptually better predictive ability, but phenotype analysis is expensive, is less useful in the field and clinic, and involves challenging methodology. Phenotypes also incur nonreplication issues, but phenotype analysis has the advantage of looking at all genes within a pathway without *a priori* knowledge of what they are. Therefore, it is easier to examine complex genotypic traits and to understand which pathways are important; the odds of finding a moderately penetrant gene are increased, and it is also possible to obtain risk-factor data and mechanistic information. It is also useful to be able to identify and validate genotypes to identify SNPs that are highly predictive of a phenotype, including "extreme outlier" phenotypes such as a subject becoming sick after his or her first cigarette.

Several phenotypic markers have demonstrated good statistical power and consistency in small cohorts and case-control studies performed on different cancer types and at multiple institutions. This is particularly true for DNA adducts and mutagen sensitivity assays. These studies also provide information about the host's response to ongoing and multiple-source tobacco exposure. The response can be quantitated and studied and, for some phenotypic markers, can be studied in the target tissue (e.g., sputum, urine).

The Working Group then developed a list of tobacco-related issues that require further study—most importantly, which populations are most susceptible to tobacco use and why, which tobacco-related cancers need to be prioritized for further study, and the need for tobacco control efforts to provide information to the susceptible populations.

Smoking-associated behavioral research requires further investigation of "potential reduction exposure products" (PREPs) and tobacco products other than cigarettes. To facilitate the comparison of smoking-related risks, the development of a standardized methodology of risk assessment will identify populations at high risk of tobacco use initiation. Risk assessment tools also are needed to identify populations at risk for tobacco-related cancers. These cancer risk assessment tools will identify tobacco-related cancer risk in family studies and in the 40 million-plus former smokers.

Through basic science research, our understanding of the biology of lung cancer will increase. This will illuminate the causal relationship between exposure to tobacco smoke and lung cancer. Our understanding of the etiology and progression of the disease will be increased by additional research focusing on the exposure-histology relationships, why certain lung lesions regress, and the resulting clinical outcomes of tobacco-related cancers. To accomplish this, array biomarkers must improve and use of surrogate tissue for target tissues must be validated.

As mentioned by the Haplotypes Versus Genotypes Working Group, improved methodologies are needed to confidently establish a relationship between genetic polymorphisms and tobacco-related phenotypes. These methodologies need to be less costly, require less tissue, be quantitative, have high throughput, and identify extreme phenotypes. Lastly, an SNP prioritization scheme for tobacco-related cancers needs to be developed.

There are still many unresolved environmental tobacco smoke issues. Studies showing functional outcomes, markers of harm, carcinogen exposure, expression profiles, and susceptibility factors are needed to assess the physiological effects of environmental smoke. Tackling these priorities requires development of a phenotype panel for scanning genotypes, improved cohort designs, and mechanisms for systematically following the cohort controls. Finally, strong partnerships with health care providers and improved interview measures will facilitate cross-study analyses.

Catherine DeAngelis, M.D., M.P.H.

(Editor-in-Chief, *Journal of the American Medical Association*)

Second Keynote Address: Making a Difference With Epidemiology Research

Dr. DeAngelis posed two crucial questions: Is the future of medicine going to be centered on genetics and genomics? And, if so, will this diminish the importance of epidemiological research?

Epidemiology identifies the impact of disease on populations and, thus, is essential in deciding which research questions require the greatest attention. The results of epidemiological studies guide decisions regarding intervention development and public health policy.

Medical journals play an important role in epidemiology research by carefully reviewing studies, prioritizing them by their findings, publishing the best articles, assisting authors in presenting data and information, and disseminating this information to health professionals, investigators, policy makers, news media, and the public.

Biomedical journals offer many opportunities to publish research findings. There are more than 15,000 biomedical journals worldwide, 4,800 of which are indexed in Medline. Medline contains more than 15 million citations going back to the 1950s, and more than 2,000 are added every day.

Publication of epidemiological data in medical journals can have an enormous impact on public policy, public opinion, and health care practices. For example, in 2002, evidence from randomized trials demonstrated adverse cardiovascular disease events and other risks with hormone therapy in the form of oral estrogen combined with progestin. In the months following the publication of these results, hormone therapy prescriptions declined significantly.

Epidemiological research is making a significant impact by providing evidence of the increasing prevalence of obesity and confirming the value of dietary guidelines in reducing mortality. Between 1966 and 1971, only 5 epidemiological papers on nutrition in medicine were listed in Medline; between 1986 and 1990, the number of similar papers was 34; and by 2001–2004, that number had increased to 231. To some extent, this pattern reflects increasing funding on this topic, but it also reflects the increasing interest on the part of the scientific community and the public that led to that increase in funding.

Epidemiology has been slower in contributing to knowledge about the role of tobacco in the U.S. death rate. Between 1966 and 1990, only 8 epidemiological papers on tobacco in medicine were listed in Medline; between 1986 and 1990, the number of similar papers was 30; and by 2001–2004, that number had increased to 52. Findings published in 2000 demonstrated that estimates of deaths caused by smoking are not altered by adjusting for demographic or behavioral factors—contrary to claims made by the tobacco industry.

A review of causes of death in the United States between 1990 and 2000 indicates that tobacco, poor diet, and lack of physical activity were the leading factors. The trends in publication of epidemiological papers on these topics correlates with the increased public interest and public health emphasis placed upon them.

Robert Hiatt, M.D., Ph.D.
(University of California, San Francisco Comprehensive Cancer Center)
The Future of Consortia

Dr. Hiatt described factors that are both pushing and pulling the field of epidemiology toward establishing more and larger consortia. The field is pulled in the direction of consortia by scientific opportunities that can best be addressed by multiple institutions and multiple disciplines. It also is pushed in the same direction by the lack of competitiveness of small studies and their limited ability to answer important questions, as well as the demands placed on investigators by new NIH data-sharing guidelines.

The transdisciplinary approach to science is critical to conducting research through consortia. Transdisciplinary science involves collaborations in which exchanging information, altering discipline-specific approaches, sharing resources, and integrating different disciplines achieve a common scientific goal.

Epidemiologists as a group like to work with colleagues from their own and other disciplines, and they are good at it because their training encourages a "big picture" worldview. Epidemiologists will be able to successfully work in consortia if there is an opportunity to advance scientific discovery and its application, their careers do not suffer, their universities and places of employment support participation in consortia, and appropriate financial support is available.

Consortia can advance science by undertaking research not possible in individual institutions and by addressing questions not likely to be considered by individual disciplines. Fears that participation in large-scale collaborations will harm scientific careers (e.g., by limiting opportunities for first-authored publications and investigator-initiated research grants) may not be realistic. The experience of the NCI-supported CFRs has been that there have been opportunities for the advancement of junior investigators. Efforts are needed to ensure that career advancement is not hindered by participation in consortia.

The question of institutional support is more problematic. The typical "silo mentality" within many academic institutions results from a number of factors, such as competition for funding, the complexity of indirect cost allocations, and the tendency to give credit for contributions to departments rather than to scientists. While many academic institutions have a real commitment to transdisciplinary science, others that claim to have transdisciplinary programs have not removed barriers to the success of those programs.

Consortial enterprises require larger amounts of financial commitment over a longer period than individual studies. They are difficult to support with current infrastructure and funding mechanisms, and NIH needs to invent new ways of addressing this problem.

The stakeholders in transdisciplinary science include universities; other, quasi-academic institutions; government; foundations; and industry. The "rules of the road" for sharing responsibility and credit for research among multiple stakeholders have not yet been ironed out.

The Cohort Consortium presents a useful example of a transdisciplinary research effort. The scientific rationale for this project is based on the need for large cohorts, economies of scale, need for accelerated research, and establishment of collaborative networks for investigators. This Consortium brought exciting opportunities—along with challenges, such as how to handle biospecimens, publications, and intellectual property. There also have been concerns about what investigators would gain from participation and the nature of NCI's role in the Consortium.

Case-control consortia have been built on a somewhat different model. Whereas the Cohort Consortium was created when NCI pulled groups together to address new questions, case-control consortia arose when investigators with common interests asked NCI for support and guidance.

There are multiple ways to approach the development and organization of consortia based on the appropriate epidemiological design (e.g., cohort studies, case-control studies, family registries). Regardless of study design, however, the principles for consortia development can be generalized to guide their establishment in response to important research questions. Multiple mechanisms can be used to fund consortia, depending on the scientific questions being addressed and resources required. New mechanisms also may need to be created.

The speaker's view is that intellectual leadership and direct management of consortia are best placed in the hands of extramural investigators because they have the most at stake. This is congruent with sound NIH principles of review and support. NCI intramural scientists and extramural advisors are valued partners who will be critical to generating the appropriate research mechanisms and infrastructural support for consortia.

Transdisciplinary research and consortia are likely to be more common in the future. There is a critical need for revision in the academic reward system for intellectual input to scientific infrastructure and collaborative enterprises by participants in these efforts. The value of consortia that are led by epidemiologists, or in which they participate, must be made clear to funders, government, and the public. Society will demand increased attention to translation and dissemination of the results of research performed by consortia.

Michael Thun, M.D., M.S.
(American Cancer Society)
Insights From the Breast and Prostate Cancer Cohort Consortium

Dr. Thun provided background information on the rationale for the Cohort Consortium and on the progress of its first proof-of-principle study. The Cohort Consortium was initiated by NCI in 1999 to establish a network of epidemiologists with existing large cohort studies to study gene-environment interactions and cancer, as well as to foster interactions among epidemiologists, population geneticists, and statisticians to integrate rapid advances in genomic research into large-scale epidemiological studies. (More than 20 cohorts are represented in the Cohort Consortium.)

The work of the Consortium began with a proof-of-principle study focusing on breast and prostate cancers; it is funded by EGRP through May 2007. The study is a collaboration involving six large cohorts (with multiple subcohorts) and two genome centers. It involves a comprehensive survey of candidate genes involved in steroid hormone metabolism, the IGF pathway, and their related receptors. Investigators are conducting pooled, nested case-control studies of putative functional SNPs (identified by resequencing) and haplotype tag SNPs associated with breast and prostate cancers. Other specific aims are to assess the association of genetic variants in a subset of cohorts with plasma hormones and IGF levels (and whether the associations vary by plasma level of hormones); examine gene-environment interactions with known lifestyle and anthropometric risk factors; and demonstrate feasibility and benefit of this large-scale, multicenter, multidisciplinary collaboration.

Based on lessons learned so far through the Breast and Prostate Cancer Cohort Consortium (BPC3) Study, some of the benefits of participation in the Cohort Consortium are that the work is scientifically interesting and stimulating; findings are potentially more informative than results from single, underpowered cohorts; and the decision not to participate may have costs (i.e., if epidemiologists cannot "hang together," they may "hang separately," especially in terms of the study of gene-gene and gene-environment interactions).

Some challenges associated with participation in the BPC3 Study are that the team process is inherently more cumbersome than independent research; larger cohorts that have already developed separate studies may jeopardize their existing funding or scientific investment in the cohort; younger and nontenured investigators may be "invisible" in team science, thereby harming career advancement; and trust and respect, although honorable and desirable, require structures that ensure fairness and maximize opportunities.

The process of operating consortia is complex. Plans and procedures must reflect reality and be clearly defined in writing. Responsibilities and roles for protocol development must be delineated. One way to divide work is by topic (e.g., pathways, genetic variants, haplotype construction, genotyping, phenotype analyses, analytical approach, cohort descriptions). The rationale for a consortium is strongest when a prospective design is essential to the research project, when other approaches have been unable to provide adequate sample size in a reasonable timeframe, and when costs can be reduced by using existing cohorts and collectively purchasing laboratory equipment.

Lessons learned from the BPC3 Study include:

- Most of the work is done in small Working Groups.
- Large conference calls are useful mainly for updates and communication across Working Groups.
- Busy schedules and long-distance communication complicate decision making.
- The proposal will be strengthened through genuine involvement of population geneticists.
- Decisions must be made concerning:
  - Whether genotyping should be centralized or multiple centers will use their own genotyping facilities.
  - Whether researchers will publish pooled analyses as soon as they are completed or publication will be delayed until individual centers have published.

Publication guidelines followed by the Breast and Prostate Cancer Cohort Consortium are that criteria for authorship are defined broadly (to be inclusive rather than exclusive); publications must fulfill the criteria of the International Committee of Medical Journal Editors; all authors and their organizations are listed at end of each paper; all who qualify as authors are listed in PubMed; and the byline under each article title states "From the Breast and Prostate Cancer Cohort Consortium," with a footnote directing readers to the author list.

At its conclusion, the BPC3 Study will be successful if it demonstrates the feasibility of studying gene-environment interactions by systematically collecting and pooling data from existing cohort studies; promotes collaboration involving population geneticists, biostatisticians, and epidemiologists at all phases of the study; and shows that "the whole is greater than the sum of the parts." The overall goal is to show that definitive results can be found more quickly and accurately by consortia than by individual, sequential studies.

The cohort consortium approach has enormous potential. Realizing this potential will require adapting individual and institutional behavior to this new research environment.

Patricia Hartge, Sc.D.
(DCEG, NCI)
Case-Control Consortia: Ways to Move Forward

Dr. Hartge explained how case-control consortia can advance cancer epidemiology. Consortia (both cohort and case-control) can yield important scientific gains more definitively, faster, and at lower costs than multiple sequential studies. Typical consortium goals include conducting specific collaborative studies with significant public health importance, fostering collaborative links, discussing common problems, and recommending solutions. Many consortia operate multiple laboratories rather than establish centralized facilities. A partial list of case-control consortia that have been initiated includes projects focusing on cancers of the bladder, brain, breast, lung, esophagus, and oral cavity, as well as lymphomas, and multiple myeloma.

The 4-year-old International Consortium of Investigators Working on Non-Hodgkin's Lymphoma Epidemiologic Studies (InterLymph Consortium) is an example of the case-control consortium model. This international consortium involves more than 20 case-control studies. There are multisite "mini-consortia" within the Consortium, such as the IARC EpiLymph Study. Consortium meetings are open to lymphoma researchers who are not involved in these case-control studies. The InterLymph Consortium began by focusing on studies for which biospecimens were available because of the group's interest in focusing on genetics.

A Genetic Polymorphism Working Group conducted the InterLymph Consortium's proof-of-principle study to demonstrate that a group of case-control studies could work together to share data, come to an agreement on how to analyze the data, and collaborate on developing a publication. This group identified 13 SNPs that seemed worthy of investigation. Data from eight international studies were pooled, and although some of the studies had very limited resources to devote to it, the effort was successful. EGRP provided support to bring the groups together, and groups with more extensive resources assisted those with fewer resources.

Within the InterLymph Consortium, several similar data-pooling projects have been initiated, and several others are in the planning phase. Eventually, the Consortium will result in approximately 20 such projects. One major benefit of pooling results of several studies is the elimination of false leads that, in the absence of collaborative analysis, might result in launching studies that are unlikely to produce useful findings.

The InterLymph Consortium experience also has shown that marginal costs are not high and that focusing on cancers with 1 to 3 percent lifetime risk is a good guideline.

Consortia are grappling with several scientific issues, such as what emphasis to place on genes versus environment, which genes to study, how to manage subgroups, and which analytical approaches to use. Common logistical issues include study design and management, leadership, support, communications, and evaluation. Procedures and resources developed by the first consortia (such as Web portals) can be used as models to save new consortia from having to "reinvent the wheel."

The growth of epidemiological consortia has training and career development implications for cancer epidemiology investigators who are not participating in consortia. Mechanisms should be developed to allow new investigators to become involved in consortium-based research in addition to working independently on questions that do not lend themselves to collaborative investigation. Data-sharing requirements may work to the benefit of new investigators who are looking for promising research questions and need to develop relationships with established researchers. However, consortia will not

provide opportunities for all outside investigators desiring to do secondary analyses, primarily because collecting new data and/or biospecimens may not be feasible.

Finally, EGRP should consider developing guidelines and technical assistance for consortia on organizing publication committees to assist junior investigators and create standards for handling publications that are based on the work of multiple institutions.

Appendix: Working Group Members

Challenges to Diet/Energy Balance Epidemiology Research

Chair: Lawrence Kushi, Sc.D. (Kaiser Permanente)
Co-Chairs: Rachel Ballard-Barbash, M.D., M.P.H. (DCCPS, NCI)
Virginia Hartmuller, Ph.D., R.D. (EGRP, DCCPS, NCI)

Jacques G. Baillargeon, Ph.D., University of Texas Health Science Center
Kathy B. Baumgartner, Ph.D., University of New Mexico
Gladys Block, Ph.D., University of California, Berkeley
Tim E. Byers, M.D., Ph.D., University of Colorado Health Sciences Center
Ralph B. D'Agostina, Jr., Ph.D., Wake Forest University School of Medicine
Qi Dai, M.D., Ph.D., Vanderbilt University Medical Center
James S. Felton, Ph.D., Lawrence Livermore National Laboratory
Gary E. Fraser, M.B., Ch.B., Ph.D., M.P.H., DipStats, FPACP, FACC, Loma Linda University
Jo Freudenheim, Ph.D., University at Buffalo
Susan M. Gapstur, Ph.D., M.P.H., Northwestern University Medical School
Marc T. Goodman, Ph.D., M.P.H., University of Hawaii
Gabriel P. Haas, M.D., SUNY Upstate Medical University
Steven M. Haffner, M.D., University of Texas Health Science Center at San Antonio
Susan Hankinson, R.N., Sc.D., Brigham and Women's Hospital
Kathy Helzlsouer, M.D., M.H.S., Johns Hopkins University
Elizabeth A. Holly, Ph.D., M.P.H., University of California, San Francisco
Rudolf Kaaks, Ph.D., International Agency for Research on Cancer
Ikuko Kato, M.D., Ph.D., Wayne State University
Larry Kushi, Sc.D., Kaiser Permanente
Johanna W. Lampe, Ph.D., Fred Hutchinson Cancer Research Center
Jing Ma, M.D., Ph.D., Brigham and Women's Hospital
Gertraud Maskarinec, M.D., Ph.D., Cancer Research Center of Hawaii
Anne M. McTiernan, M.D., Ph.D., Fred Hutchinson Cancer Research Center
Maureen A. Murtaugh, Ph.D., R.D., University of Utah
Marian Neuhouser, Ph.D., R.D., Fred Hutchinson Cancer Research Center
Sara H. Olson, Ph.D., Memorial Sloan-Kettering Cancer Center
Manuela A. Orjuela, M.D., Sc.M., Columbia University School of Public Health
Dorothy R. Pathak, Ph.D., M.S., Michigan State University
Nancy Potischman, Ph.D., National Cancer Institute
Lynn A. Rosenberg, Sc.D., Boston University Slone Epidemiology Center
Jessie S. Satia, Ph.D., University of North Carolina at Chapel Hill
Arthur Schatzkin, M.D., Dr.P.H., National Cancer Institute
Richard A. Scribner, M.D., M.P.H., Louisiana State University

Challenges to Diet/Energy Balance Epidemiology Research (continued)

Xiao-Ou Shu, M.D., Ph.D., Vanderbilt University Medical Center
Martha L. Slattery, Ph.D., University of Utah
Victoria L. Stevens, Ph.D., American Cancer Society
Sara S. Strom, Ph.D., The University of Texas M.D. Anderson Cancer Center
Janice E. Stuff, Ph.D., R.D., Baylor College of Medicine
Michael J. Thun, M.D., M.S., American Cancer Society
Richard Troiano, Ph.D., National Cancer Institute
Marilyn Tseng, A.M., Ph.D., Fox Chase Cancer Center
Carmina Valle, M.P.H., National Cancer Institute
Dee W. West, Ph.D., Northern California Cancer Center
Walter Willett, M.D., Harvard School of Public Health
Herbert H. Yu, M.D., Ph.D., Yale University School of Medicine
Jian-Min Yuan M.D., Ph.D., University of Southern California

Haplotypes Versus Genotypes

Chair: Stephen Gruber, M.D., Ph.D. (University of Michigan Medical School)
Co-Chairs: Thomas Sellers, Ph.D., M.P.H. (H. Lee Moffitt Cancer Center and Research Institute)
Daniela Seminara, Ph.D., M.P.H. (EGRP, DCCPS, NCI)

Stephen Chanock, M.D., National Cancer Institute
Graham A. Colditz, M.D., Ph.D., Brigham and Women's Hospital
Daniel W. Cramer, M.D., Sc.D., Brigham and Women's Hospital
Eric J. Duell, Ph.D., Dartmouth Medical School
Patricia Hartge, Sc.D., National Cancer Institute
Karl T. Kelsey, Ph.D., Harvard School of Public Health
Polly Newcomb, Ph.D., M.P.H., Fred Hutchinson Cancer Research Center
Julie R. Palmer, Sc.D., Boston University
Jong Y. Park, Ph.D., M.P.H., M.S., H. Lee Moffitt Cancer Center and Research Institute
Ann G. Schwartz, Ph.D., M.P.H., Karmanos Cancer Institute/Wayne State University
Roy Shore, Ph.D., Dr.P.H., New York University Medical School
Margaret R. Spitz, M.D., M.P.H., The University of Texas M.D. Anderson Cancer Center
John K. Wiencke, Ph.D., University of California, San Francisco

Design Issues and Strategies in the Study of Rare Cancers

Chair: Nathaniel Rothman, M.D., M.P.H., M.H.S. (DCEG, NCI)
Co-Chairs: Sholom Wacholder, Ph.D. (DCEG, NCI)
Isis Mikhail, M.D., M.P.H., Dr.P.H. (EGRP, DCCPS, NCI)

Robert A. Branch, M.D., University of Pittsburgh School of Medicine
Graham A. Colditz, M.D., Dr.P.H., Brigham and Women's Hospital
R. William Field, Ph.D., University of Iowa College of Public Health
Marsha L. Frazier, Ph.D., The University of Texas M.D. Anderson Cancer Center

Anna R. Giuliano, Ph.D., H. Lee Moffit Cancer Center and Research Institute
Sally L. Glaser, Ph.D., Northern California Cancer Center
Donghui Li, Ph.D., Brigham and Women's Hospital
Shelia K. Zahm, Sc.D., National Cancer Institute

Susceptibility in Tobacco Carcinogenesis: Genotypes Versus Phenotypes

Chair: Peter Shields, M.D. (Lombardi Cancer Center, Georgetown University)
Co-Chairs: Stephen Marcus, Ph.D. (DCCPS, NCI)
Xifeng Wu, M.D. (The University of Texas M.D. Anderson Cancer Center)

Marshall W. Anderson, Ph.D., University of Cincinnati
J. Fernando Arena, M.D., Ph.D., National Cancer Institute
Paul J. Brennan, Ph.D., International Agency for Research on Cancer
Neil Caporaso, M.D., National Cancer Institute
David C. Christiani, M.D., Harvard School of Public Health
R. William Field, Ph.D., University of Iowa College of Public Health
Barbara Guest, M.S.W., M.P.H., National Cancer Institute
Stephen S. Hecht, Ph.D., University of Minnesota Cancer Center
Philip Lazarus, Ph.D., Penn State University College of Medicine
Mark Parascandola, Ph.D., National Cancer Institute
Joe Patel, Ph.D., National Cancer Institute
John P. Richie, Ph.D., Penn State Cancer Center
Scott Rogers, Ph.D., National Cancer Institute
Margaret R. Spitz, M.D., M.P.H., The University of Texas M.D. Anderson Cancer Center
Vaurice Starks, B.S., National Cancer Institute
John K. Wiencke, Ph.D., University of California, San Francisco
Deborah Winn, Ph.D., National Cancer Institute

Susceptibility in Tobacco Carcinogenesis: Genotypes Versus Phenotypes

Chair: Peter Shields, M.D. (Lombardi Cancer Center, Georgetown University)
Co-Chairs: Stephen Marcus, Ph.D. (DCCPS, NCI)
Xifeng Wu, M.D. (The University of Texas M.D. Anderson Cancer Center)

The Working Group initially focused on phenotypes related to smoking and tobacco. Members agreed that phenotype characterization provides researchers with an important tool to study tobacco-related cancers. Phenotypes offer predictive capabilities that may lead to the identification of genotypes, biomarkers, comorbidities, and etiologies associated with cancer. They provide insight into complex genotypic traits by identifying pathways central to multiple genotypes. Phenotype classification also facilitates finding genes or SNPs of varying penetrance. The importance of learning from "extreme phenotypes"—such as those diagnosed with lung cancer after smoking their first cigarette—was recognized. By focusing on the genotypes of such individuals, much can be learned about susceptibility for lung cancer. Furthermore, phenotypes provide information about an individual's response to exposures. Using this information, researchers can examine and identify exposure-response relationships related to dose, specific agents, and population response profiles. The Working Group concluded that identifying tobacco-related phenotypes and broadly defining them is important.

Biomarkers of harm provide information on mutations in cells, tissue activities related to exposure, and methylation profiles in blood and epithelial cells. Tools to identify these biomarkers can include cytogenetic studies and various imaging techniques. Biomarkers of susceptibility can provide vital information on individuals or populations vulnerable to tobacco-related cancers. Researchers have examined DNA repair mechanisms and used functional studies from various population groups to identify them. Tumor phenotypes also generate specific genetic and behavior profiles. Other diseases and conditions, such as chronic obstructive pulmonary disease, heart disease, and shortness of breath at an early age, represent specific phenotypic changes that function as physical markers of tobacco use. Finally, comorbidities, such as alcoholism and depression, often are associated with smoking and tobacco use. Thus, there are many biomarkers as well as behavioral markers for smoking and tobacco use.

Three gaps in tobacco and cancer research that require further study were identified: (1) identifying which populations are most susceptible to tobacco use and why; (2) determining which tobacco-related cancers need to be studied; and (3) choosing the types of tobacco control efforts needed to reach susceptible populations.

Another gap in our ability to determine the impact of tobacco on people's health is a lack of new risk-assessment tools. In addition, to assess the smoking-related risks associated with potential reduction exposure products (PREPs) and compare their use with those for cigarettes requires development of standardized risk-assessment methodologies. These methods need to take into account populations at high risk for initiation of tobacco use and to identify populations at risk for tobacco-related cancers. Together, these cancer risk-assessment tools can be used to identify tobacco-related cancer risk in family studies and among the 40 million-plus former U.S. smokers.

Through basic science research, our understanding of the biology of lung cancer will increase. This will illuminate the causal mechanisms between exposure to tobacco smoke and lung cancer. Additional research focusing on exposure-histology relationships, understanding why certain lung lesions regress, and the resulting clinical outcomes of tobacco-related cancers will increase our knowledge of the etiology and progression of the disease. To accomplish this, our array of biomarkers must improve, and our use of surrogate tissue for target tissues must be validated.

Improved methodologies are needed to confidently establish a relationship between genetic variations/polymorphisms and tobacco-related phenotypes. These methodologies need to be less costly, require less tissue, be quantitative, have high throughput, and identify extreme phenotypes. Lastly, an SNP prioritization scheme for tobacco-related cancers needs to be developed. Overcoming these five barriers will advance our understanding, diagnosis, and treatment of tobacco-related cancers.

There are still many unresolved environmental tobacco smoke issues. Studies showing functional outcomes, markers of harm, carcinogen exposure, expression profiles, and susceptibility factors are all needed to assess the physiological affects of environmental smoke. Tackling these priorities requires development of a phenotype panel for scanning genotypes, improved cohort designs, and mechanisms to systematically follow the cohort controls. Lastly, strong partnerships with health care providers and improved interview measures will facilitate cross-study analyses.

Challenges to Diet/Energy Balance Epidemiology Research

Chair: Lawrence Kushi, Sc.D. (Kaiser Permanente)
Co-Chairs: Rachel Ballard-Barbash, M.D., M.P.H. (DCCPS, NCI)
Virginia Hartmuller, Ph.D., R.D. (EGRP, DCCPS, NCI)

This Working Group identified many gaps in our knowledge of the relationship between diet and tumor development. The importance of physical activity in tumor development is undeniable; however, the independent effects of diet and obesity on tumor development remain unknown. New interventions need to be developed. Designing clearer methodologies detailing the role that diet, exercise, and obesity play at the social and community levels are needed. Lastly, incorporating these new methodologies into treatment and prevention plans will break down a barrier between basic science research and community involvement.

Another barrier to understanding the relationship between diet and tumor development is the lack of standard biomarkers for diet. This is confounded by a disagreement among scientists as to whether biomarkers should focus on nutritional status or dietary intake. Current dietary assessments include urine and/or serum specimens for biomarker analysis. However, the lack of definitive biomarkers is a barrier to further progress. Identifying these markers and developing diagnostic assays for them will provide a more accurate indication of dietary effects on tumor development. Therefore, several of the large barriers in the area of energy balance and tumor development will be tackled by identifying the independent effects of diet, exercise, and obesity, as well as identification of more specific biomarkers.

A third barrier to assessing the role of diet in tumor development stems from use of the self-reported Food Frequency Questionnaire (FFQ) to assess dietary intake. This method is a technological barrier because it generates many dietary intake inconsistencies that make it difficult to combine data across studies. In addition, the traditional FFQ does not gather information about ethnic foods. This is a methodological gap because the current U.S. immigrant population maintains strong cultural ties and dietary practices similar to those seen in their countries of origin. One solution to this barrier is the creation of a standardized dietary assessment tool. This tool would be strengthened by including not only ethnic food choices, but also the capability to assess dietary intake over the life cycle, during interventions, and during cancer treatment.

Current research efforts are determining the relationship among SNPs, haplotypes, and dietary factors. However, identifying statistically significant diet-gene interactions requires large sample sizes. One solution is to develop consortia that focus on multiple diet-gene interactions. Consortia provide large

datasets that will determine whether these diet-gene interactions are protective or contribute to cancer development.

Haplotypes Versus Genotypes

Chair: Stephen Gruber, M.D., Ph.D. (University of Michigan Medical School)
Co-Chairs: Thomas Sellers, Ph.D., M.P.H. (H. Lee Moffitt Cancer Center and Research Institute)
Daniela Seminara, Ph.D., M.P.H. (EGRP, DCCPS, NCI)

Many methodological questions remain about the use of data on genotypes and haplotypes in epidemiological studies. Two related questions focus on resequencing: When should resequencing be initiated, and should it be centralized? In addition, new methodologies of haplotype analysis are a barrier that must be overcome to advance this field. These new methodologies should detail the type of biospecimen required, the environment in which the biospecimen will be harvested, and the type of study in which the biospecimen will be used. Lastly, this Working Group recommended secondary analysis of data for method comparison. Formulating new methodologies and secondary analyses will provide the solution required to overcome these barriers.

The Working Group asked that the difficulties of including minority groups in studies be addressed as new methodologies for haplotype analysis are created. Including representative numbers of minority populations in studies historically has been a barrier because many U.S. minority populations are dynamic. The Working Group suggested initiating specific studies within minority groups to examine genetic variations that may exist within populations. Together, these changes in the way that minority populations are approached and integrated into studies will increase the accuracy of our analytic models.

As the fields of epidemiology and genetics continue to merge, education and training sessions for collaborating scientists will enhance this interdisciplinary science. These training sessions would familiarize scientists from each discipline with the skills, language, and methodologies of the other fields. Educated exchanges and discussions will increase communication and strengthen collaborations. Lastly, through these education sessions, standardized methodologies for addressing genetic variations within regulatory regions will reach consensus.

Looking toward the future, the Working Group recommended that journals not only archive submissions and supplemental material online, but also publish negative data online.

Design Issues and Strategies in the Study of Rare Cancers

Chair: Nathaniel Rothman, M.D., M.P.H., M.H.S. (DCEG, NCI)
Co-Chairs: Sholom Wacholder, Ph.D. (DCEG, NCI)
Isis Mikhail, M.D., M.P.H., Dr.P.H. (EGRP, DCCPS, NCI)

Overview

The goal of this workshop was to solicit input from NCI investigators on the need to study rare tumors. This workshop focused on adult tumors only. We defined a tumor as rare if the incidence is less than 15/100,000, approximately equivalent to 40,000 new cases per year in the U.S. We considered childhood tumors outside our mandate because there is an active epidemiology effort by the Children's Oncology Group, which enrolls over 90% of children with cancer.

Why should we study rare tumors?

We discussed why rare tumors should be studied. First, there is an opportunity to find the "low-lying fruit" in the first investigations of rare tumors, in contrast to additional studies of more common but also far more extensively studied tumors. The payoff might be particularly large for genetic factors if, as suggested by N. Risch from twin studies, lower-incidence tumors tend to have a greater genetic component. In the past, we have identified some very rare tumors with single etiologies, such as retinoblastoma (RB), angiosarcoma, and clear cell CA of the vagina. The resulting insights have also helped to provide an understanding of the pathogenesis of tumors with a more complex etiology. Three examples are Al Knudson's work on RB, Fred Li's work on adrenocortical CA, and studies of vinyl chloride and angiosarcoma. Moreover, there are public health reasons for studying rare tumors. While individually low, the aggregated total incidence and resultant mortality of all rare tumors combined comprise an important fraction of the total cancer burden. Thus, ignoring rare cancers would close off opportunities to reduce the disease burden.

Individual tumor sites have some characteristics that may make them particularly appropriate to study. Some rare tumors are rapidly lethal (e.g., cancer of the pancreas and esophagus) or have rising incidence rates (e.g., esophageal cancer). From a population perspective, rare tumors can have disproportionately higher rates among some ethnic groups—for example, high rates of nasopharyngeal carcinoma among the Chinese population or high rates of male breast cancer among Zambian men in Africa (15%) versus lower rates among men in the U.S. (0.1%). As a consequence, the Working Group felt that neglecting the study of rare tumors had ethical implications at both the population and individual patient levels.

What are appropriate first steps to the study of the etiology of rare tumors?

The group agreed that the first step should be the review of SEER descriptive data to show incidence rates overall, their temporal trends and geographic variation, and racial or ethnic variation. This would help identify tumors that would provide the best opportunity for study and give a sense of feasibility. Other steps include identifying opportunities for using specimens and data in existing cohorts. The group felt that we need to determine where there are ongoing treatment trials to which a case-control study can be added on, as is commonly done in pediatric cancers. Finally, the distinct, albeit costly, advantages of *de novo* efforts were considered at length.

What study designs would be appropriate?

Three study designs were discussed in terms of their utility in studying rare tumors:

Cohort Studies: Pooling data and potentially biologic samples from existing cohorts would be cost-effective and could be done relatively quickly. However, the degree of completeness of case ascertainment for rare tumors is unclear and would need to be considered. Given the limited number of cases, this study design would allow the identification of moderate to strong risk factors for the rarer tumors.

Clinical Trials: Clinical trials ascertain cases and collect biospecimens already and thus can be a helpful setting for a case-control study, as shown for the pediatric cancers. There are a number of methodological issues that need to be considered, including those factors that

determine which cases enroll into such trials. However, as the key challenge in the study of rare tumors is sample size, efforts should be made to explore taking advantage of trials.

De novo Designs: New standalone case-control studies are ideal but can be expensive, especially as they are likely to require multiple centers. SEER-based studies could provide an infrastructure to carry out population-based studies. A key advantage of hospital-based studies would be the ability to collect pretreatment blood samples and tumor tissue and to process samples in ways that allow state-of-the-art analysis. These studies could use molecular tumor assays for defining disease subgroups and in studies of gene expression and proteomics and could be integrated with studies of prognosis and treatment. Studying multiple tumors simultaneously in a case-control study offers economy of scale from, perhaps, a single core questionnaire and biospecimen collection protocol. A study that is hospital-based, particularly at major Cancer Centers, seems particularly sensible, since it would allow a potentially relatively rapid accrual of cases. Hospital or clinic controls would be appropriate. The disadvantages are not as severe as believed, and the Working Group felt that we cannot afford to be overly fastidious—particularly for initial studies. In fact, strong apparent risk factors will be quite unlikely to be artifacts of well-designed and well-conducted hospital-based case-control studies.

Building infrastructure:

One strategy to capitalize on existing infrastructure is the NIH-funded General Clinical Research Centers (GCRCs). A local PI could be in charge; involvement of an orphan disease program would provide focus; and the GCRC facilities could be used to collect blood samples for vigorous phenotype/genotype assessment. To initiate these studies, supplemental funds could be provided to Cancer Centers to explore feasibility. Once significant pilot data are collected, R01 applications could be submitted to expand the research.

Some general suggestions:

- Create a common rare tumor protocol

- Collect baseline information on all rare tumors: pooling, data sharing
- Bank samples
- Conduct international studies comparing higher rates in other countries with focus on U.S. population: e.g., liver cancer in Asian populations in Asia vs. Asian-American immigrants in California
- Controls: need to keep an open mind
- Innovative studies
    - No need for controls?
    - Get “reasonable but not perfect” controls and tolerate the “critique noise”: e.g., relatives and friends of patient, routinely screened persons, etc. (exposure bias?)
    - Be liberal: Due to cancer rarity, look for high relative risk (RR); a low elevation in RR seems not to be cost effective from a research portfolio perspective

- Cancer Research Network: an HMO funded by NCI
- Kaiser Permanente data
- Pooling of common tumors from pre-existing studies
    - Dutch study: n=80,000
    - Greek-Italian study
    - Iowa Study
    - Other studies
    - # cases identified from these studies

# # #