Environmental Genomics Workshop

June 28-29, 2005

National Institute of Environmental Health Sciences

National Institutes of Health

Research Triangle Park, North Carolina

The National Institute of Environmental Health Sciences (NIEHS) held a workshop on June 28-29, 2005 to establish priorities and future directions for NIEHS in environmental genomics. The workshop was attended by a diverse group of scientists from the extramural community and by representatives from the NIEHS intramural and extramural programs[1]. In her opening remarks, Dr. Gwen Collman (Chief, NIEHS Susceptibility and Population Health Branch) briefly reviewed the Environmental Genome Project efforts. She asked workshop participants to assess current methods in environmental genomics, to identify deficiencies in available research tools for environmental genomics, and to define new opportunities that will allow NIEHS to make a unique contribution to this rapidly evolving field. Dr. David Schwartz (Director, NIEHS) reiterated these points, briefly reviewed existing NIEHS programs in environmental genomics, and presented a framework for NIEHS' future in this area. The CMGCC (Comparative Mouse Genomics Centers Consortium) and TRC (Toxicogenomics Research Consortium) programs, the molecular epidemiology planning grants, and the ELSI (Ethical, Legal, and Social Implications) programs, as well as the intramural core facilities and intramural and extramural research projects were briefly described. Dr. Schwartz also described the ongoing resequencing efforts of the Center for Rodent Genetics (CRG). Specifically, the CRG has contracted for the resequencing of 15 strains of mice by Perlegen. The resequencing of environmentally relevant human genes was also mentioned. These NIEHS programs have generated numerous resources for the scientific community.

Dr. Schwartz emphasized at the onset of the workshop that NIEHS is at a turning point in its environmental genomics efforts. NIEHS is now moving from developing infrastructure to answering questions about environmental health (EH). Specifically, Dr. Schwartz asked the workshop participants to focus on NIEHS' contribution to understanding complex (chronic) diseases that would have the largest impact on society and public health overall. It was Dr. Schwartz's opinion that NIEHS should give greater emphasis to disease endpoints than in the past, while establishing a unique identity for NIEHS within NIH. Dr. Schwartz also expressed an interest in having NIEHS participate in issues related to global environmental health.

Dr. Schwartz solicited input from workshop participants in guiding NIEHS' future efforts, asking them to bring their collective wisdom, knowledge, and experience to bear on the task at hand: to develop a new strategic vision for environmental genomics at NIEHS. To move the field of environmental health science forward, the key areas in EH that must be addressed are exposure assessment, training, and bioinformatics infrastructure. Dr. Schwartz mentioned that the recommendations of this workshop will be considered by the NIEHS Strategic Planning Group this fall when they develop a framework outlining the Institute's goals over the next five years.

Session I: Gene Discovery

The first workshop session on Gene Discovery was moderated by Dr. Jeffrey Murray (University of Iowa). The charge of this session was to assess current approaches and identify novel methodologies for gene discovery as it pertains to understanding environmentally related disease. Dr. Murray asked participants how the power of the present databases (EGP, HapMap, ENCODE, etc.) and resources could be best utilized to do multiple analyses to identify genetic and environmental factors for human diseases. Both candidate gene and whole genome approaches were discussed as valid in environmental studies. A general discussion of the importance of developing bioinformatics and biostatistical infrastructure and the utilization of cross-strain/cross-species comparisons for gene discovery ensued. The development of pathway analysis tools informed by comparative genomics to understand toxicological mechanisms was suggested. Workshop participants emphasized the overall value of utilizing model organism systems in parallel with human studies. Since there are many “genome projects” in existence today, participants agreed that NIEHS should focus EGP-related efforts on studying how environmental exposures can be utilized to understand individual susceptibility to disease.

Session II: Gene Function

The second workshop session on Gene Function was moderated by Dr. Doug Bell (NIEHS). Dr. Bell pointed out that genomics-based technologies provide tools for understanding the mechanisms by which exposures lead to disease and why the risk of disease varies from individual to individual. This information can form the basis of developing models and policies for risk assessment and risk management. Historically, NIEHS has made important contributions in these areas. Dr. Bell suggested that future NIEHS goals in advancing the field of risk assessment are to identify susceptible human populations with greater precision and more fully understand the mechanisms of toxic responses as well as the function of human disease-susceptibility genes. Workshop participants discussed current issues in exposure assessment and risk assessment during Session II. The participants recommended the development of improved DNA biomarkers of exposure. Numerous investigators in this workshop recommended the incorporation of genetics into toxicological and environmental health studies. The importance of studying age-related susceptibility to exposure was stressed during this session. It was also suggested that NIEHS could make a unique contribution to the understanding of the role of epigenetics in disease susceptibility.

Session III: Disease Susceptibility and Risk

The third workshop session on Disease Susceptibility and Risk was moderated by Dr. Martyn Smith (University of California). In his opening remarks, Dr. Smith suggested that the next phases in the "omics" revolution will be shaped by an increased use of nanotechnology, proteomics, and metabolomics. Dr. Smith stated that an emphasis of “omics” applications and systems biology approaches will be necessary in molecular epidemiology. To implement and apply these technologies effectively to the study of gene-environment interactions, Dr. Smith encouraged biomedical scientists to recruit and interact extensively with engineers. The Superfund Basic Research Program was mentioned as a good model for getting engineers involved and utilizing the power of interdisciplinary research. Dr. Smith encouraged the development of consortia to study exposed populations as well. The development of new initiatives to study genome-wide

association studies for environmentally induced diseases and new studies on the impact of environmental exposures on disease progression was discussed during this session. The participants recommended the development of “omics” technology and biomarkers of exposure and effect with a focus on disease outcome. Numerous scientists attending this workshop encouraged NIEHS to utilize existing large-scale longitudinal cohort studies for environmental exposure studies and to focus on disease outcomes and prognostic factors. There was also a general discussion relating to the possible creation of an NIEHS repository of samples from environmental health studies. Participants stressed the importance of connecting epidemiologists with scientists who measure environmental exposures.

Panel Discussion/Prioritizing Final Recommendations

Dr. William Suk (NIEHS, Director, Center for Risk and Integrated Sciences) moderated the final panel discussion focused on prioritizing the recommendations made throughout the workshop. The panelists for this session were Dr. John Groopman (Johns Hopkins University), Dr. Mary-Claire King (University of Washington), Dr. Frank Gilliland (University of Southern California) and Dr. Ken Paigen (The Jackson Laboratory). Dr. Suk charged the workshop participants to generate a "short list' of the highest priority goals or initiatives that were discussed throughout the workshop for advancing the field of environmental health. The following list of top priorities was generated:

- **Develop novel biomarkers of exposures and tools for measuring exposure that better predict health and risk for disease**
- **Develop high-throughput relatively predictive lower-organism model systems**
- **Promote the application of comparative genomics to toxicology**
- **Promote the integration of genetics and toxicology as they pertain to environmental health**
- **Focus on understanding mechanisms of environmentally induced diseases**
- **Modernize environmental medicine in light of the lack of occupational exposure in the U.S. population**
- **Understand the role of epigenetics in disease susceptibility and define the epigenome**
- **Leverage/partner with other organizations to identify and study existing cohorts for EH studies; find a well-characterized genetic cohort and apply environmental exposure aspects to it**
- **Improve tools and resources for large scale analysis of gene-environment interactions**

Major Meeting Discussion Points

The following topics were discussed and repeatedly emphasized throughout the meeting: comparative genomics/organism model systems, discovery mechanisms for genes and/or environmental factors, systems biology and “omics” development, computational tools (bioinformatics), animal modeling, epigenetics, clinical applications of environmental health research, population/epidemiological studies, and environmental

exposures/biomarker development. A summary of the extensive discussions made with respect to these general topics areas in all sessions of this meeting are as follows:

Comparative Genomics/Organism Model Systems

Pathway analysis (e.g., focused analysis of gene-gene interactions) could play an important role in our understanding of toxicological responses. Dr. Rick Young (Whitehead Institute) suggested an in-depth study of known pathways involved in environmental exposure responses. He also suggested studying pathways and networks in relevant cells and tissues. In addition, Dr. Young encouraged the sequencing of the whole genomes of more organisms. Yeast, *C. elegans*, and Drosophila genetics systems have been well worked out and should be better utilized. The robustness of the yeast model system has increased as the number of high quality large-scale yeast databases has increased (e.g., the yeast transcriptome, proteome, gene-regulatory interactome, protein interactome). It was stated that there is good annotation in the yeast system with all of the genes having been knocked out and yeast expression data for exposures is also available. All transcriptome and chromatin information as well as many of the protein-protein interactions are also known for yeast. Dr. Young stated that it is important to look at the entire picture and put all the datasets together to get the best idea of what's going on in an organism. Dr. Young recommended that the technology developed for yeast genetics is applied to vertebrate systems.

Elements conserved between species are likely to be important for biological networks. Dr. Jonathan Freedman (Duke University) stated that the *C. elegans* system as a model organism is almost as complete as the yeast system. *C. elegans* will be useful in linking human diseases with toxicological studies and for detecting subtle mutations and polymorphisms as well as studying the activation of proto-oncogenes. However, it will be a challenge to extrapolate results from the yeast or worm model systems to humans, in part because orthologous phenotypes for these systems are not always obvious. Mice or yeast bearing "humanized" genes or gene regions can be very valuable for gene-function studies as well. For homologous genes, it was suggested that one can study the mammalian gene complement in model systems and for non-complementary genes, one can humanize or introduce the human gene into the model organism to study its function. NIEHS was encouraged to use phylogenetic comparisons between mammals and lower invertebrates or vertebrates to study the role of regulatory elements for their potential impact on diseases.

It was suggested that NIEHS may want to build databases utilizing cell lines in which allelic differences are correlated with phenotypes and have biological plausibility. Several participants stated that it will be important to be able to define "phenotype" and its application to disease in both animals and *in vitro* systems. Dr. Young also encouraged the use of cross-species comparisons to validate and confirm pathways deduced in one model system to other systems.

Needs/Gaps

- There needs to be more success stories of going from comparative genomics and animal modeling to clinical research.

- There is a need for toxicologists to connect with physicians and a need to develop better parallel efforts between animal models and human studies.

Recommendations

- NIEHS should promote the exploitation of model systems in EH studies, with an associated effort to improve methods for extrapolating from model systems to human populations.

Discovery of Gene, Gene-Environment, Gene-Gene Interactions:

Novel approaches to mapping genes for complex disorders and the utilization of genetics and genomics in EH studies were discussed. Dr. Jan Vijg (University of Texas Health Science Center) stated that future NIEHS initiatives need to be framed in terms of haplotypes and groups of genes rather than individual SNPs. One needs to apply large genome concepts (large deletions, etc.), genome architecture, and copy number of genes as well. Dr. King emphasized the use of genome architecture (large-scale inversions, deletions, duplications, etc.) to identify critical genes that could be screened in larger populations. Identifying the genes or alleles can be a first step to identifying the specific environmental assault that drives penetrance or alters the onset of the disease. Dr. King recommended the continued exploitation of genomics as a tool in the eventual identification of environmental factors for complex diseases. Dr. Demetrius Maraganore (Mayo Clinic) suggested that NIEHS should focus on candidate gene studies but with a disease approach (e.g., have an initiative for whole-genome association studies for asthma, etc.).

Dr. Young, Dr. Stephen Chanock (National Cancer Institute), and others stressed the importance in continuing to improve sequencing capacity and reduce sequencing cost. NIEHS should support cost effective whole genome sequencing technology development to bring costs down to $1000 or less. Dr. Young also encouraged efforts to determine the sequence of regulatory and intragenic DNA regions along with complete chromatin structure and histone code with a focus on chromatin modeling. He suggested developing comprehensive maps of gene-gene, gene-protein, and protein-protein interaction networks in appropriate model systems as well as the complete mapping of the regulatory structure of an organ (e.g., the liver) and the mapping of all proteins and protein-protein interactions in particular cell types of interest (hepatocytes, for example, as has been done in yeast for some cell types). Once mapping is complete, one could then examine how environmental exposures affect these processes. It was mentioned that metabolism studies with genetics may be of interest to NIEHS as well.

Standardization and optimization of tools for DNA sequencing and genotyping efforts is needed as well. Dr. Ivan Rusyn (University of North Carolina, Chapel Hill), Dr. Mary-Claire King (University of Washington), and others indicated that there is a great need for improved, powerful, and cost-effective genotyping technologies. Dr. King and others suggested that oligonucleotide pools might be useful to genotype the same SNPs across studies.

The improvement of tools and resources for large scale analysis of gene-environment interactions will be very important. Several participants stressed the need for better approaches to compare combinations of SNPs with environmental factors. New approaches to map genes include mapping loci of susceptibility in cell lines (e.g., Vivian

Cheung studies) were mentioned. NIEHS was asked to consider including QTL analysis studies in mice followed by association studies in humans as a useful approach to identify complex disease genes.

Needs/Gaps

- Additional bioinformatics tools and infrastructure are needed to analyze the necessarily large datasets required for multifactorial (e.g., gene by gene by environmental factor A; gene by environmental factor A by environmental factor B) studies.
- Accurate phenotyping methods/measures (imaging, physiology, biochemistry), human environmental monitoring for exposures, as well as large N's (human sample sizes) will be necessary to study gene-environment interactions.
- Cost effective models for functional analysis of human SNPs are needed

Recommendations

- NIEHS should promote new initiatives in DNA sequencing and genotyping
- NIEHS should partner with other organizations (i.e., Illumina) for advances in genotyping/sequencing technology
- NIEHS should stimulate genome-wide association studies in EH
- NIEHS should support a HapMap equivalent for appropriate mouse strains for use in QTL analyses.

Systems Biology/"Omics":

Many workshop participants indicated that continued development of genomics-based technology is important. Dr. King and others emphasized that dissemination of new technology to NIH grantees and/or the public is as important as developing the technology itself. There was a general discussion related to developing systems biology agendas and biological tools for data interpretation. Many participants stated the importance of developing better standards/controls and establishing and promoting criteria for the stabilization of novel platforms, including proteomics and metabolomics. There was consideration about how to best integrate the platforms (microarray, proteomic, metabolomic, etc.) when developing these technologies as well. It was noted that metabolomics is non-invasive and may fit in well when using cohorts. It was stated that proteomics could possibly be done on formaldehyde-fixed tissues as well. There is a need to develop new algorithms for metabolomics and proteomics as well as the application of those already developed more broadly. One can interpret these "omics" data using reverse engineering approaches as well. The suggestion was made to utilize NTP resources to help facilitate these efforts.

Dr. David Balshaw (NIEHS) discussed several ongoing trans-NIH engineering "omics" and informatics initiatives. BECON (the NIH Bioengineering Consortium) was established in 1997 and is currently administered under the leadership of the NEBIB (the National Institute of Biomedical Imaging and Bioengineering). BECON has released a series of program announcements to foster the application of engineering concepts to biomedical research and includes a special emphasis on nanotechnology and team

science. BECON has taken the lead in developing NIH's policy on accepting multiple-PIs on investigator initiated awards. BISTIC (Biomedical Information Science and Technology Initiative Consortium) focuses on making optimal use of computer science and technology to address medical research. BISTIC initiatives have fostered innovative applications of informatics and computational techniques to biomedical science and has established the foundation for the NIH Roadmap National Centers of Biomedical Computing. Dr. Chanock pointed out that NCI has also developed a Cancer Bioinformatics Grid (ca/BIG). Dr. Raymond Tennant (NIEHS) stated that Ca/BIG is linked to the CEBS (Chemical Effects on Biological Systems) database [developed by the NIEHS National Center for Toxicology (NCT)] as well. However, several participants stated that there is little or no consideration for environmental exposures in any of these initiatives. The NIH Roadmap is also stimulating the maturation of systems biology in particular through the activities of the Technology Centers for Networks and Pathways, the Metabolomics Technology Development initiative, the National Centers for Biomedical Computing, the nanomedicine development centers, and the Interdisciplinary Research Centers activities.

Needs/Gaps

- To develop predictive markers, there is a need for additional knowledge of the pathways and networks altered by exposure.

Recommendations

- NIEHS should develop tools, databases, and technologies for elucidating the global biological response to exposures and make them available to the public
- NIEHS should push the agenda for metabolomic profiling and develop metabolomics- and proteomics-based biomarkers of early effects
- NIEHS should participate more fully in trans-NIH initiatives in "omics"
- NIEHS should develop a proteome database which includes tissue-proteome responses to chemical exposures as predictive markers

Computational/Bioinformatics/Biostatistics

Workshop participants expressed serious concern over the lack of bioinformatics tools and support for their research. As a result, there was strong consensus over the immediate need for new bioinformatics initiatives and resources, as well as additional bioinformatics expertise. The lack of support for currently available bioinformatics tools appeared to be of special concern. This problem was thought to reflect a structural deficiency within the bioinformatics community, that is, the lack of a "service component" for highly complex and sophisticated informatics tools. Dr. Tim Rebbeck (University of Pennsylvania) suggested that there are some novel approaches and methods that have been developed but have not been applied which could be used to study gene-gene and gene-environment interactions. Dr. Rebbeck stated that networking of bioinformatics scientists is necessary. Participants pointed out that there was a need for common datasets as well. NIEHS was encouraged to sponsor a workshop focused on the utilization of public datasets for bioinformatics development similar to the workshops NIH has previously held for genetic databases. It was stressed that investigators need the

ability to analyze complex data sets at a reasonable charge as a service as well.

Many workshop participants including Dr. Rusyn and Dr. Maraganore suggested that NIEHS should promote the rapid development and dissemination of improved bioinformatics tools for toxicogenomics-based datasets (e.g., transcriptomic, proteomic, metabolomic, and epidemiology-based studies of exposed humans or animals). In this context, Dr. Tennant emphasized that the CEBS database, which will be a full-scale toxicogenomics database, will be available to the public in 2005. At that time, complete exposure datasets for approximately 20 chemicals will be available. This availability of common data sets should both enhance existing resources as well as stimulate development of additional bioinformatics resources in the EH community and could serve as a model for integrated databases. It was suggested that NIEHS work on algorithm and platform integrations to make CEBS more usable with interfaces everyone can use. It was stated that CEBS is missing a service component. Dr. Tennant suggested that for the next iteration of CEBS, it needs: 1) more complete databases for meta-analyses, 2) real tools that people will want to use, and 3) better facilitation of communication and posting of information. Dr. Tennant stated that after NIEHS opens the CEBS database up to the community they will bring bioinformaticians and biologists together for a workshop to discuss how well CEBS works and how to improve its functionality further.

Dr. Maraganore stated that we lack the ability to address the level of complexity necessary for identifying gene-environment interactions computationally and we don't know enough about exposures. Dr. Maraganore gave a specific example of a whole genome association study for Parkinson's disease as an example of the complexity and difficulties in identifying genes and environmental factors for complex diseases. The suggestion was made to develop chat rooms to link computational/computer scientists with biologists. It was also pointed out that proprietary software is a problem in advancing science.

There was considerable discussion about engaging an industry partner, such as GlaxoSmithKline (GSK), to promote rapid development of new bioinformatics tools and resources. Dr. Allen Roses (GSK) described a powerful GSK bioinformatics tool called BioBucketBlast and expressed willingness to demonstrate this tool to any interested scientists and to discuss partnering opportunities with NIEHS/NIH.

There was a general discussion about current trans-NIH initiatives in Bioinformatics. Although there are no formal RoadMap initiatives, ca/BIG (NCI) is one existing initiative relating to bioinformatics. The participants discussed whether a formal initiative should be led by National Library of Medicine or NIEHS or as a joint effort of many NIH Institutes. It was stated that it was unlikely that the National Library of Medicine would do the complex analyses that are necessary.

Needs/Gaps

- There are infrastructural weak points to using whole-genome association studies to identify gene-environment interactions.
- Common, public, integrated bioinformatics datasets are needed
- Availability of a cost effective service center to analyze complex datasets is necessary

Recommendations

- ❖ NIEHS should promote the development of bioinformatics tools and infrastructure
- ❖ NIEHS should encourage bioinformaticians to network
- ❖ NIEHS should develop and provide bioinformatics and biostatistics resources to EH grantees and the scientific community
- ❖ NIEHS should sponsor a workshop on the utilization of public datasets for bioinformatics development similar to the workshops NIH held for genetic databases.
- ❖ NIEHS should partner with industry (i.e., GSK) or other ICs to stimulate further advances in bioinformatics development

Animal Modeling:

The utility of animal model systems to identify genetic variations of interest to particular diseases as well as the limitations of using animal models were discussed. Variability in tumor endpoints was a concern. It was pointed out that the exposure may be the same but the tissue specificity may be very different between species. For example, the same exposure may result in liver tumors in rodents while producing bladder cancer in humans. Dr. Paigen stated that the limited genetic variability in laboratory mice is also a problem since all lab mice are at least partially inbred. Dr. Paigen and Dr. Tennant stressed the importance of the mouse strain used in influencing the phenotype you obtain. Finally, Dr. Tennant suggested that improved annotation will be necessary to more fully understand genetic variation in response to environmental exposures.

Dr. Paigen pointed out that data for 7 or 8 complex diseases (e.g., atherosclerosis, osteoporosis, asthma, type II diabetes, etc.) suggest that the same set of susceptibility genes for these diseases are probably operative in all mammals. With this information one can utilize mammalian model systems to map susceptibility loci in the rodent and then find the human complement through association studies in humans. One could also use this approach to study epistasis, complex networks, or to identify which genetic alleles are working together to produce a given phenotype.

Dr. Paigen advocated for continued emphasis on well-characterized chemical exposures that have known public health significance such as ozone and mercury. Dr. Paigen and Dr. Tennant also felt that it was important to understand mouse-strain specific susceptibility to toxic and carcinogenic compounds. Dr. Paigen proposed an initiative that addresses both of these questions, in which a panel of 35 inbred mouse strains would be screened for susceptibility to several environmental agents. Using powerful mouse genetic tools, this experiment could help identify genetic markers associated with unique strain- or species-specific susceptibility to environmental exposures.

The suggestion was also made to use common environmental agents (ozone, mercury, etc.) to study how much variation in susceptibility there is across mouse strains (e.g., acetaminophen in liver toxicity across mouse strains). Several participants stated that it may be more informative to study chemicals where humans and mice react in the same way. It was stressed that similar studies should be conducted in appropriate human populations simultaneously with the rodent studies. There was a general discussion about whether the National Toxicology Program (NTP) should focus on shorter-term

assays rather than long-term mammalian studies. Dr. Christopher Portier (NIEHS) briefly described the NTP program which has studied over 500 environmental agents in mice and rats. Responses have been assayed, recorded, and tissues collected.

Participants discussed the analysis of phenotypic effects of gene knockdowns and overexpression. The importance of evaluating the effect of gene dosage was also emphasized. In contrast to knockout alleles, which are often lethal and therefore relatively uninformative, overexpressing and underexpressing alleles can be easier to study and more informative for gene function studies. Workshop participants questioned whether NIEHS should also study alternative strategies for knocking out genes such as siRNA.

There was little support among workshop participants for the proposed NIH initiative to generate a complete library of mouse knockouts for all genes. The value of this project was questioned on the basis of poor cost-effectiveness, inappropriate use of resources, and poor experimental design (e.g., functional analysis on a gene-by-gene basis is likely to be much more valuable). It was pointed out that a knock-in or conditional mouse model may give more information than the null mutation. It was stressed that humanization of the mouse for particular genes is also of interest. In general, the workshop participants supported creation of a resource that would generate specific mouse knockout strains on an "as-needed" or "on-demand" basis. It was recommended that NIEHS contribute to technology development in mouse knockout methodologies. The suggestion was made for NIH to do a pilot project to assess utility of a repository of knockout mice, using the Lexicon efforts as an example. Workshop participants encouraged NIEHS to make knockout technology available to all at low cost and focus on knockouts that would be appropriate for the NIEHS mission (related to exposure, environment, etc.).

Recommendations

- NIEHS should screen a panel of 35 inbred mouse strains for susceptibility to chemicals that are significant public health hazards
- NIEHS should support the development of knockout technologies for mouse models
- NIEHS should support mouse knockout resource as an “on-demand” service
- NIEHS should investigate alternatives to mouse knockout approaches to study gene expression and function
- NIEHS should utilize mammalian model systems to pull out complex gene interactions.
- NTP specimens should be made available to the scientific community for further analyses.
- NIEHS should promote simultaneous comparative genomics approaches between animal models and humans
- NIEHS should promote a combination of *in silico* approaches and human epidemiology to model human genes, SNPs, or haplotypes in mice after appropriate prescreening assays.

Epigenetics:

Dr. Randy Jirtle (Duke University) advocated for additional research on the impact of environmental exposure during fetal development. For some critically important mammalian genes that regulate growth, there is a critical window during fetal development in which environmental exposure alters genomic imprinting. These epigenetic changes can have significant phenotypic consequences including increased cancer susceptibility in the adult.

Workshop participants expressed significant interest in research on epigenetics, but indicated that they needed additional information on current methods and applications in this field. Thus, it was suggested that NIEHS sponsor a workshop on technical and scientific aspects of epigenetics research encompassing many model organisms in the near future. NIEHS could particularly focus on understanding the role of epigenetics in disease susceptibility and develop a unique niche to help define the epigenome. Dr. Jirtle pointed out that somatic and epigenetic alterations are not detected by whole-genome association studies so there needs to be alternative approaches to identify these changes. One can study individual twins and get information for epigenetics. It was pointed out that those mouse strains (of the 15 that NIEHS is currently resequencing) demonstrating a larger variance in response to environmental exposures may have increased epigenetic lability. A recommendation was made to utilize epigenetic changes as phenotypes and early biomarkers of disease. Several participants asked whether NTP could study whether genotoxic agents alter methylation or make epigenetic changes.

Needs/Gaps

- Alternative approaches to identify somatic and epigenetic alterations

Recommendations

- NIEHS should encourage investigations into the fetal origins of adult disease susceptibility as well as prenatal windows of susceptibility
- NIEHS should sponsor a workshop on epigenetic methods and applications

Clinical Applications to Environmentally Induced Disease:

Several workshop participants suggested that the future in EH lies in bringing EH to the clinic and "to the patient's bedside." This will require greater focus on disease endpoints, translational research, and therapeutic targets and less focus on disease prevention. It was suggested that environmental health might achieve greater clinical relevance if it orients itself more towards prognostic/diagnostic/therapeutic approaches as a means to reduce the burden of environmentally related diseases. Dr. Maraganore pointed out that susceptibility genes do not necessarily make good therapeutic targets. One may have to consider different sets of genes for intervention/prevention of disease than treatment in a population. Dr. Groopman (Johns Hopkins University) and Dr. Maraganore stated that NIEHS should decide whether they want to invest their energies into public health versus clinical medicine, studying susceptibility of populations versus treatment of disease. NIEHS faces a choice between focusing on population endpoints

or individual endpoints, where the former approach reflects a public health orientation and the latter approach reflects a more clinical orientation.

It was suggested that NIEHS broaden its focus from etiologic questions to study what happens after disease as well. Dr. Maraganore stated that outcome-specific factors may determine survival. Dr. Groopman suggested NIEHS should study how environmental exposures affect the outcome and trajectory of disease. He recommended that an RFA be developed to study the impact of environmental exposures on disease outcome. This would likely expand NIEHS' focus from susceptibility factors to factors that modify survival.

Dr. Roses led a general discussion on drug discovery applications that might be applicable to EH and exposure studies. Dr. Roses suggested that one could utilize environmental exposure data to target drugs and treatments therapeutically and make a contribution to this field. Comparative genomics and humanized mouse approaches can be used for drug discovery and treatment. Dr. Roses stated that with ApoE, the mouse KO was humanized for ApoE isoforms and questions were asked regarding the different pathways – are proteins up- or down-regulated? Clinical trials were then proposed for compounds that were successfully screened. Pharmacogenomics and predictive toxicity are very important for drug development. The suggestion to identify gene targets with whole-genome association studies was also made.

It was pointed out that novel approaches may be needed to assess drug susceptibility in the human population as well. For example, Dr. Roses suggested that patients who receive a drug early after its FDA approval be required to participate in a drug susceptibility screening program. For example, the first 500,000 patients to receive the new drug might be required to have their blood banked so that when an adverse effect occurs, the sample is available for genotyping. It was suggested that this would be a very cost-effective method for evaluating and understanding the risk of adverse drug reactions in the human population. There was also a general discussion related to the feasibility of adding a clinical component to the existing environmental health centers that NIEHS is funding.

Recommendations

- ❖ NIEHS should broaden its focus from preventive to prognostic approaches
- ❖ NIEHS should develop an RFA focused on the impact of environmental exposures on disease outcome
- ❖ NIEHS should investigate new approaches to assess susceptibility in the human populations, possibly by working with industry to develop tissue repositories/databases or by adding clinical components to existing environmental health centers.

Population Studies/Epidemiology:

To make population-based studies feasible, NIEHS should partner with outside groups to make use of previously established large cohorts. The question was asked which methodology is best to use for studying EH-related exposures. Dr. Clare Weinberg

(NIEHS) stated that there was no “best” approach. However, NIEHS will get the “biggest bang for their buck” by using existing cohorts.

Several specific cohorts were mentioned at the workshop, including an AARP cohort, the National Children's Study, the Drug-Induced Liver Toxicity Network (DILIN), and a newly proposed AGES (American Gene Environment Study) cohort. NIEHS could make a contribution by measuring exposure in these cohorts. NIEHS should consider further identifying environmental and potential disease links in the Ag health study. One could extensively genotype the individuals in the existing cohort and ask why some people get disease and others do not. Dr. Perry Blackshear (NIEHS) mentioned a newly developed environmental polymorphism registry which will maintain identities for 25 years. This registry will be able to associate SNPs with phenotypes. Other cohorts to consider are the deCODE project in Iceland and Norwegian and Denmark cohorts. Dr. Chanock mentioned the NCI Cohort Consortium as an example of one presently existing cohort that NIEHS might use. A question was asked as to whether the NIEHS can partner with CDC to access the NHANES datasets? Many existing cohorts need partnerships and funding to keep going. Dr. Chanock also mentioned that one could utilize the power of large cohorts to perform co-variant analyses as well.

Participants also encouraged NIEHS to consider participating with well-designed cohorts outside the U.S. It was suggested that NIEHS could add the exposure piece to many of these cohorts. Participants suggested that NIEHS focus on at-risk populations around the world. Studying populations with high exposures may be particularly relevant to EH. It was pointed out that U.S. populations are very different from nutritionally compromised populations around the globe. NIEHS could make important contributions by improving the burden of disease worldwide by focusing on more preventable diseases in developing countries. The global transport of disease is an important issue that may be of interest to NIEHS as well. (China and India were suggested as possible countries that NIEHS might consider establishing partnerships in with epidemiologists there).

The importance of proper sample collection (stem cells, DNA, etc.) was discussed. It was stressed that samples should be banked, annotated, and stored properly. Participants encouraged NIEHS to be sensitive to issues relating to DNA storage in NIH repositories, especially for international studies. Extensive phenotyping is very expensive; therefore it was suggested that NIEHS could focus on a subset of cohorts for particular analyses while maintaining specimen repositories for additional analyses if warranted.

The group recommended a long-term RFA that focused on prognostic outcomes with 20-year snapshots. Ideally, the population would be highly phenotyped and monitored for exposures. This population would be followed to see what diseases result. It would be equally important to consider susceptibility associated with race and ethnicity, as well as exposure-resistant populations, considering factors conferring protection as well as risk.

Finally, it was proposed that NIEHS should carefully consider ELSI issues for susceptible populations. Workshop participants recommended that NIEHS develop an initiative on ethical, legal, and social issues relevant to populations that are subject to "unconsented" exposures to environmental toxins. NIEHS should also consider current IRB issues and burdens.

Needs/Gaps

- Proper collection, annotation, and storage of biological samples
- NIEHS needs to identify centers to follow patients longitudinally and use suitable existing cohorts for large scale EH exposure studies

Recommendations

- NIEHS should make use of previously established, well-designed cohorts, including international cohorts of highly exposed populations or adding an exposure component to cohorts already in existence
- NIEHS should consider a long-term prospective study of highly phenotyped individuals to associate disease with exposure
- NIEHS should consider susceptibilities associated with race and ethnicity, as well as resistant populations.
- NIEHS should develop an RFA focused on ELSI issues relevant to "unconsented" exposures
- NIEHS should consider current IRB issues and burdens.
- NIEHS should explore creating an NIEHS repository of EH samples and make samples available for use to the broader scientific community

Environmental Exposures and Exposure Assessment/Biomarker Development:

It was widely acknowledged in this workshop that we cannot measure environmental exposures with the degree of precision that we would like and this is holding back the field. Dr. Groopman stated that we need to measure exposures at quantitative levels below what is possible with current mass spectrometry methods to get a handle on low-dose exposures in the population. We also need to assess sample size requirements for low-dose exposure assessments. The industry does not recognize that there is a market for this type of work, especially to identify small molecular weight components. Finally, it was pointed out that it will be important to attract chemists to this field.

There is a need to continue to focus on analytical biomarkers that can be validated in real samples and to develop tools for validating biomarkers of exposure. Participants suggested the development of an RFA for analytical methodologies. Issues related to analytical methodologies that were discussed include the contamination and destruction of samples as well as background noise level (associated with diet, etc.).

There was a general discussion on the types of environmental exposures that are most relevant to the NIEHS mission. For example, NIEHS could focus on one or more of the following: chemical exposure outside the U.S., novel infectious agents, radiation or radiation oncology, bioterrorism agents, the role of the "built" environment in obesity and diabetes, stress, etc. Dr. Paigen stated that we should not compete with NIAID, so NIEHS should focus on chemical exposures. However, Dr. Groopman pointed out that most exposures are a combination of both chemical and biological. The suggestion was made to pick agents where there is a known human population that is susceptible to begin to study.

It was pointed out that there is a need for the collection of better exposure data (dose or body burden, biological response to that exposure) as well. There was a general discussion of what to measure for accurate exposure assessment. Would you measure

pathogens, antibodies to environmental agents, metabolites of chemicals, or adducts? Dr. Mike Resnick (NIEHS) and Dr. Jack Taylor (NIEHS) discussed using somatic mutational load, radiation damage hotspots, and the integration of exposure data into DNA biomarker data because DNA damage accumulates mutations over time. It was stated that NIEHS might also consider the role of the immune system in response to exposure and that inflammatory responses could be a measure of environmental human outcome over time. There was also a recommendation that *in silico* mapping, QTL analyses, or post-translational modifications/changes related to environmental exposures (measured with proteomics, mass spectrometry, etc.) could be applied to exposure assessment.

It was suggested that NIEHS might want to use cell-based systems (yeast or *C. elegans*) to study exposures. It was pointed out that one needs phenotypic measures of exposure and a way to measure genotoxicity accumulated over a lifetime. (An example was given of a NIDDK database and studying drug-induced liver disease with the liver reactions to drugs). Participants recommended building resources such as genomic DNA, serum, and immortalized lymphocytes to study environmental exposures.

There was a discussion about the Department of Defense's role in developing biosensors. Several scientists asked which biosensors are most relevant to dating exposures in biological systems. [Dr. Dennis Lang (NIEHS) mentioned that NIH money is earmarked for Biodefense but it is not clear how this money will be distributed among the Institutes]. It was pointed out that sparse biological specimens may mean pooling is necessary in developing biosensors.

From a clinical perspective, Dr. Maraganore stated that Americans want to know if they have been exposed and if they are at risk for developing disease. The general difficulty with environmental exposures is that they are low-dose and they occurred long ago. The Achilles' heel of EH is that diseases are caused by exposures that occurred 20 to 30 years ago. Few chemicals are sufficiently persistent in the body to be measured well over time. Participants suggested that perhaps lifetime exposure histories could be assessed from the body burden of fat-soluble compounds (e.g., dioxin). What are the basic tools to assess human exposure over a lifetime? NIEHS needs to create opportunities to answer these questions. Several individuals stated that biological effects may happen at very low levels of continuous environmental exposures and we must be able to measure this. It was pointed out that dating of environmental exposures is still very difficult. The suggestion was also made to use adverse effects as a more persistent and valid indicator of exposure. Dr. Groopman and Dr. Tennant pointed out that the effect of exposure may be more important than exposure per se to study.

Needs/Gaps

- Analytical biomarkers that can be validated in real samples
- Tools for validating biomarkers of exposure
- Biomarkers with a disease focus that integrate exposure over time
- Metabolomics- and proteomics-based biomarkers of early effects
- Novel biomarkers of exposure that predict health and risk for disease

- To anchor biomarkers in mechanistic studies for etiology, prevention/interventions, and prognosis in human and experimental models
- Better, more precise, exposure data (dose or body burden, biological response to that exposure)

Recommendations

- NIEHS should initiate development of an RFA focused on analytical methodologies to detect biomarkers of low-dose exposures
- NIEHS should develop an RFA focused on technology development to make exposure assessment more precise and sensitive
- NIEHS should focus on understanding mechanisms underlying environmentally induced diseases and the translation of phenotypes to exposures.
- NIEHS should bridge the gap between engineering and medicine by focusing on biomarkers/biosensors, biomonitoring (to be able to respond to biothreats by studying the biological monitoring of exposure), and developing signature profiles of biothreats.
- NIEHS should develop signatures of adverse effects in population/animal models.
- NIEHS should investigate age-related susceptibilities to exposures
- NIEHS should develop better approaches to measure combined or multiple exposures and extended exposures.
- NIEHS should develop report cards for the public related to EH.
- NIEHS should utilize Environmental Health Science Centers to provide public access to the tools we have discussed.

Training

Participants stated that environmental health technology will be driven by our young investigators; however, we are not attracting top students into EH. Dr. Groopman pointed out that creative approaches must be used to increase the number of trainees without decreasing the amount of money available per trainee. One approach is to fund physician-scientists or foreign nationals as trainees. He also discussed that NCI uses the R25 (education project) mechanism to generate additional training slots. This is a good but rarely used mechanism for funding trainees. Start-up support for young scientists is particularly important. Dr. King recommended putting 20% of the extramural budget into training.

Many individuals stressed that bioinformaticians are in extremely high demand, so they are both difficult to recruit and to retain. Young bioinformaticians are being offered higher salaries than other young scientists, even though many do not complete postdoctoral training. Thus, the group suggested that new approaches or initiatives may be needed for training and recruiting young faculty in the area of bioinformatics. It was pointed out that more graduate students in biostatistics are also needed. Workshop members suggested that perhaps we should integrate the bioinformatics and biostatistician trainees into teams with other members of the laboratory in multi-

disciplinary projects. There also needs to be cross-disciplinary approaches to integrate toxicology and genetics for the future needs of EH.

There was a general discussion about mature scientists being encouraged to invest in specialized training outside their field of expertise during a sabbatical year or by other appropriate mechanisms. Alternatively, one can recruit scientists who are already bioinformaticians or statisticians and train them in toxicology. However, participants expressed concern over whether we would lose a generation of scientists as we bring more senior scientists up to date.

The importance of educating younger students of potential careers in EH was emphasized. Participants stated that NIEHS should encourage collaborations between scientists in different disciplines within a particular grant. Other mechanisms should be used to reward scientists for time spent in mentoring activities. It was pointed out that the Superfund Basic Research Program is an excellent example of how to integrate training into a program. Dr. Groopman stated that more emphasis should be placed on reaching undergraduates who know about risk assessment but are not familiar with careers and opportunities in EH. The F programs are a useful mechanism allowing postdocs to be a P.I. on a project. This mechanism should be marketed to other disciplines.

Recommendations

- Recruit engineers and chemists to apply their knowledge to interdisciplinary environmental-health research projects
- Young scientists should be encouraged to engage in multidisciplinary projects very early in their careers
- Laboratories should be encouraged to support trainees in several different fields or disciplines
- An alternative to training more young scientists is to "re-train" senior scientists
- The R65 mechanism can be used to encourage senior faculty to mentor young scientists
- Be creative about the mechanisms of support used to facilitate training

Closing Remarks

In closing, Dr. Schwartz thanked all workshop participants for their time and effort. He also welcomed further communication regarding these issues in the days ahead.

[1] See separate attachment for list of participants. Miriam Sander (Page One Editorial Services, Boulder, Colorado) attended the workshop as rapporteur.