Summary of the NIH Genes, Environment, and Health Initiative (GEI) Workshop on Gene-Environment Interplay in Common Complex Diseases: Forging an Integrative Model

Given the current state-of-the-science and the progress made in GEI since its inception, NIH sought input from the scientific community on optimal ways to move forward with investigations of gene-environment (GxE) interplay in health and disease. Experts from a variety of disciplines including genetic epidemiology, social science, environmental science, and statistical genetics were invited to give advice about the future direction of gene-environment interaction studies.

On January 25, 2010, scientists representing a wide array of disciplines met to evaluate the state-of-the-science and recommend ways to move forward. The workshop began with speakers giving their perspective on progress made and future needs from the point of view of genetic researchers and environmental researchers. Topics discussed were genome-wide association studies, statistical genetics, epigenetics, molecular exposures, developmental psychology, and social science.

The workshop speakers made a number of important points about the state-of-the-science and needs for the investigation of gene-environment interaction. For example, the need to incorporate more exposure measures into genetic studies was an overarching theme of the talks. It was noted that single nucleotide polymorphisms (SNPs) were not independent predictors of disease, therefore SNPs should not be studied independently. There is a need to look at heritability and environmental interactions rather than just allele by environment interaction. The difficulty of randomization of social exposures was also mentioned as a problem in the study of gene-environment interaction. For instance, it is hard to randomize parental divorce status and socio-economic status of childhood neighborhoods. Also, incorporation of multiple levels of the environment and longitudinal environmental measures are needed in gene-environment interaction studies as the environment changes with time. It was also stressed that large sample sizes are needed to detect effects and genetic and environmental effect modifiers should be leveraged to increase statistical power. Speakers noted that study design should be informed by gene-environment hypotheses. Also, strong statistical evidence for association and precise replication are required up front for the study of gene-environment interaction. It was noted that negative results are also a finding and should be incorporated into our hypotheses. The need for integration of environment, genetics and epigenetics in the same samples was stressed as there is growing evidence of environmental susceptibility of epigenetic marks. In addition, epigenetics can act as a mediator of genetic effects. Speakers also noted throughout their presentations that there is a need for new statistical methods and measurement tools.

Next participants of the workshop were broken out into breakout groups and charged with answering an overarching question: “Given the current state-of-the-science and the progress made in the GEI program, what are optimal ways to move forward with investigations of gene-environment interplay in health and disease?” As they discussed this question, they were specifically asked for their input on theory and study design, methods and data analysis, and phenotypes, endophenotypes, and other variables. To prompt participants, the 12 questions below were asked to act as a catalyst for discussion.

Q1: What problems can best be solved by “discovery-driven” approaches? What problems are best addressed by “hypothesis driven” approaches and what theoretical approaches would be most helpful in guiding hypothesis generation?

Q2: Should G-E studies be targeted, that is, focused on a particular gene, exposure, phenotype, or disease? Or should studies be broad, designed to encompass as many factors as possible?
Q3: To what extent can existing studies be adapted to investigate G-E interplay? Which questions will require the development of new cohorts?
Q4: Are there research designs that allow us to investigate the complexity (on G and E sides) without infinitely large sample sizes? Conversely, how do we design studies to avoid major pitfalls?
Q5: What analytic strategies might be most useful at this point in investigating G-E interplay? Can multiple strategies be combined in a single "proof-of-principle" study?
Q6: How do we integrate more complex environmental measures into our models? How do we approach incorporating different non-discrete environmental variables? What statistical/computational methods are needed to integrate these disparate data streams?
Q7: What level of mechanistic understanding is needed to verify G or E 'hits' before follow up in GxE studies?
Q8: What statistical tools and resources are needed?
Q9: What are the characteristics of end-points or variables that are "ready to go" for GxE studies? Are there specific diseases, traits, biological phenotypes, or environmental exposures that currently meet these characteristics?
Q10: Should we focus on complex phenotypes, or search for associations to the underlying mechanisms or intermediate/endophenotypes?
Q11: How do we integrate variables in G-E studies, many of which are interdependent, that incorporate a comprehensive view of "environment"?
Q12: What are the best strategies to measure environmental variables and exposures in large cohorts? What is needed to incorporate next-generation tools to scale up to large epidemiological studies?

Recommendations

<u>Theory and Study Design</u>

There were many recommendations on theory and study design. Targeted as well as broad approaches are needed depending on the research question. Participants agreed that a variety of approaches are needed to capture data on the genome, epigenome, etiology of disease, and environmental exposures. Both broad and targeted approaches should aim to be multi-level and multi-dimensional whenever possible. Targeted studies (hypothesis driven), that focus, for example, on phenotype, disease, or environmental exposure, are best for detecting situations in which genes are moderating the environment. Broad (discovery-driven) approaches are more appropriate in situations where the environment is modifying genetic effects. If well designed, such studies can provide a better range of the environmental exposure, phenotypes, and phenotype groups. Multi-level and multi-dimensional designs are more easily incorporated; however, this approach is far more costly than a targeted approach. Discovery-based research should be encouraged in both environmental research as well as genomics. Whatever approach is used, the goal should be to inform policy decisions, address health disparities, and improve public health.

It was noted that gene-environment interaction studies using existing cohorts can leverage existing investments and many of the cohorts have great sophistication of phenotypic measures, high depth of follow-up information, and stored biological specimens. Disadvantages of using existing cohorts are that they often lack sufficient or appropriate ethnic representation or age groups, lack variables that span disease domain, and exposures may be lacking or not relevant

to the new question. Also, the informed consent may not cover new uses of data and/or samples or allow for data sharing.

New cohorts are required for studies requiring very large sample sizes with measurements spanning multiple disease domains; studies requiring modern environmental exposures and individual monitoring; studies requiring certain specimens (e.g., epigenetic studies cannot use amplified DNA); studies requiring time critical measures; and studies requiring certain minority groups including youth and disadvantaged populations. Participants noted that studies should have policies and procedures for collection and storage of biological specimens and for access to materials and data by researchers. In addition, investigators should develop plans for re-contacting individuals found to have rare variants for additional experimental studies.

Methods and Analysis

No single proof-of-principle study is perfect or adequate; therefore, participants felt that various study designs and approaches were needed—that is, integrative designs that include biomarkers, pathways to disease, and outcome measures as well as more focused studies. Behavioral and social factors should be factored into these designs.

Participants felt that gene-environment interaction studies require more analytical methods and approaches, thus continued support is needed for methods development. Also, it was felt that the field suffers from a shortage of people trained in computational biology. Participants encouraged support of a clearinghouse for standardized approaches and tools as it was felt that it would benefit the research community.

Phenotypes, Endophenotypes, and Variables

Participants agreed that for large-scale studies, phenotypes should be low burden, have validated measurement accuracy and reliability, be easily accessible and related to the disease being studied, and have low cost. It was noted that there is a tradeoff between the need for low-cost, harmonized phenotypic measures that allow for the pooling of data from large studies and the need to discover new intermediate phenotypes and biomarkers that more precisely characterize both the exposure and the disease risk.

Conclusion

In conclusion, there are many common pathways leading to health or disease and there are both genetic and environmental effects along that pathway. There is a need for a heterogeneous approach to gene-environment interaction studies in order to reflect differences in diseases and the stages of human development. Different models and sample sizes will be needed based on what is known. For example, developmental stages, environmental changes, gene expression, and disease latency may require longitudinal studies starting early in life. Studies that are more hypothesis generating may be more appropriate when less is understood about a given disease (e.g., autism versus some forms of cancer). Environmental measurement technologies should be incorporated into ongoing and new gene-environment interaction studies. There should also be an emphasis on data sharing and harmonization.

For more information about the workshop, please contact Ebony Bookman, OPG at ebony.bookman@nih.gov.