NIH Public Access
Author Manuscript
Genet Epidemiol. Author manuscript; available in PMC 2012 August 09.

Published in final edited form as:
Genet Epidemiol. 2011 May ; 35(4): 217–225. doi:10.1002/gepi.20571.

Gene-Environment Interplay in Common Complex Diseases: Forging an Integrative Model—Recommendations From an NIH Workshop

Ebony B. Bookman[1,*], Kimberly McAllister[2], Elizabeth Gillanders[3], Kay Wanke[4], David Balshaw[2], Joni Rutter[5], Jill Reedy[3], Daniel Gnaughressy[2], Tanya Agurs-Collins[3], Dina Paltoo[6], Audie Atienza[3], Laura Bierut[7], Peter Kraft[8], M. Daniele Fallin[9], Frederica Perera[10], Eric Turkheimer[11], Jason Boardman[12], Mary L. Marazita[13], Stephen M. Rappaport[14], Eric Boerwinkle[15], Stephen J. Suomi[16], Neil E. Caporaso[17], Irva Hertz-Picciotto[18], Kristen C. Jacobson[19], William L. Lowe[20], Lynn R. Goldman[9], Priya Duggal[9], Megan R. Gunnar[21], Teri A. Manolio[1], Eric D. Green[1], Deborah H. Olster[4], and Linda S. Birnbaum[2] for the NIH G × E Interplay Workshop participants

[1]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland [2]National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina [3]National Cancer Institute, National Institutes of Health, Bethesda, Maryland [4]Office of Behavioral and Social Sciences Research, National Institutes of Health, Bethesda, Maryland [5]National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland [6]National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland [7]Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri [8]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts [9]Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland [10]Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York [11]Department of Psychology, University of Virginia, Charlottesville, Virginia [12]Department of Sociology and Institute of Behavioral Science, University of Colorado at Boulder, Boulder, Colorado [13]Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh School of Dental Medicine, Pittsburgh, Pennsylvania [14]Department of Environmental Health Sciences, School of Public Health, University of California at Berkeley, Berkeley, California [15]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas [16]Laboratory of Comparative Ethology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland [17]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland [18]Department of Public Health Sciences, University of California at Davis, Davis, California [19]Department of Psychiatry, University of Chicago, Chicago, Illinois [20]Division of Endocrinology, Department of Medicine, Northwestern University Feinberg School of Medicine,

© 2011 Wiley-Liss, Inc.

*Correspondence to: Ebony B. Bookman, Office of Population Genomics, Office of the Director, National Human Genome Research Institute, 530 Davis Drive Room 3130 MSC K3-02, Morrisville, NC 27560. ebony.bookman@nih.gov

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript
AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Chicago, Illinois [21]Institute of Child Development, University of Minnesota, Minneapolis, Minnesota

Abstract

Although it is recognized that many common complex diseases are a result of multiple genetic and environmental risk factors, studies of gene-environment interaction remain a challenge and have had limited success to date. Given the current state-of-the-science, NIH sought input on ways to accelerate investigations of gene-environment interplay in health and disease by inviting experts from a variety of disciplines to give advice about the future direction of gene-environment interaction studies. Participants of the NIH Gene-Environment Interplay Workshop agreed that there is a need for continued emphasis on studies of the interplay between genetic and environmental factors in disease and that studies need to be designed around a multifaceted approach to reflect differences in diseases, exposure attributes, and pertinent stages of human development. The participants indicated that both targeted and agnostic approaches have strengths and weaknesses for evaluating main effects of genetic and environmental factors and their interactions. The unique perspectives represented at the workshop allowed the exploration of diverse study designs and analytical strategies, and conveyed the need for an interdisciplinary approach including, data sharing, and data harmonization to fully explore gene-environment interactions. Further, participants also emphasized the continued need for high-quality measures of environmental exposures and new genomic technologies in ongoing and new studies.

Keywords

gene-environment interaction; epidemiology; study design; genetics; environment

INTRODUCTION

Susceptibility to the majority of human diseases and disorders is complex and multifactorial, involving both genetic and environmental factors. Complex diseases and disorders, such as cardiovascular disease, cancer, diabetes, and psychiatric diseases and disorders, also generally have high prevalence and therefore place the greatest burden on society. The development of high density genotyping platforms has allowed investigators to screen hundreds of thousands of genetic variants to test for associations with disease. To date, Genome Wide Association Studies (GWAS) have identified over 800 statistically significant ($P \leq 5 \times 10^{-8}$) findings in various diseases and conditions [Hindorff et al., 2010]. In contrast to variants identified for Mendelian disorders, most variants identified through GWAS have small to modest effects, thus substantial heritability remains to be explained for virtually all complex diseases [Thomas, 2010a,b;Wright and Christiani, 2010]. This unexplained genetic variation may be due to low frequency alleles or other types of variation not captured by current GWAS techniques and/or to underdeveloped data analysis methods for detecting complex interaction [Manolio et al., 2009; Mechanic et al., 2008; Mukherjee et al., 2009b]. The detection of complex interaction such as gene-environment interaction is important as it can not only allow the opportunity to discover novel genes whose effects are modified by

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

different environments, but it also provides insight into biological pathways that may not be detected outside of interaction analyses [Thomas, 2010a,b]. Few published GWAS have addressed gene-environment interaction, a critical aspect of the architecture of complex disease [Kazma et al., 2010; Rutter et al., 2006].

Diseases that cause the largest burden on society and are, in turn, a tremendous public health concern, include complex diseases such as cancer, cardiovascular disease, asthma, psychiatric disorders, and diabetes [Kazma et al., 2011]. The investigation of gene-environment interplay will provide a deeper understanding of the etiology of complex disease and our ability to prevent and treat disease because it will allow us to identify genetic vulnerabilities that are ultimately realized—or possibly prevented—when combined with particular environmental exposures. Many Phase III clinical trials fail due to serious side effects in a subset of patients. The identification of drug-gene interaction effects would greatly benefit drug development and therapy, and thus public health. The evaluation of gene-environment interplay will also facilitate the detection of novel pathways and thus identify new targets for drug therapy. Identification of genetic factors that interact with environmental factors will pinpoint pathways through which the environmental exposures may act and therefore identify specific environmental exposures that require further investigation [Ordovas and Tai, 2008]. Further, the identification of gene–environment interactions can provide guidance for personalized interventions on specific environments based on a person's genetic background. Moreover, studies of this type increase both: (1) the chances to discover factors that contribute to disease through an interaction effect with no marginal effect; and (2) our ability to identify environmental effects that act mainly in genetically susceptible groups [Thomas, 2010]. The inclusion of environmental measures into genetic epidemiology studies will not only help with interpretation by evaluating a more complete picture of complex disease, but it may also increase the power of the study to elucidate the underpinnings of health and disease. The study of gene-environment interplay is also important because modification of the environment, e.g. diet, lifestyle, toxic exposures, is likely to remain more feasible than modification of the genome for the foreseeable future for prevention and intervention measures to improve public health.

Most current genetic or environmental epidemiology studies are powered to detect main effects and are not designed to detect variables that interact with other genetic and/or environmental factors. Moreover, while genotyping is relatively straightforward, environmental exposures have been more difficult to sufficiently measure on a large scale due to the non-static and manifold nature of exposures. The study of gene-environment interplay will identify biological mechanisms and pathways that may be important for predicting disease risk, and will also be necessary to evaluate the benefit of prevention and treatment strategies that differ by genotype [Amato et al., 2010; Thomas, 2010]. In fact, when interactions are in opposite directions in different exposure groups, i.e. different kinds of exposures or different levels of exposures, main effects of genetic or environmental factors may not be identified [Caspi et al., 2010; Murcray et al., 2009]. Moreover, genes that only influence disease occurrence in the presence of an environmental factor will likely be missed in the marginal analysis unless the effect size of the interaction is large. For example, in a study by Andreasen et al. [2008], no association was observed between INSIG2 rs7566605 and obesity; however, when the level of physical activity was taken into account,

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

physically inactive homozygous carriers of the risk C-allele had significantly higher BMI. This example highlights the inherently interdisciplinary nature of gene-environment studies of health. Understanding why the C-allele at SNP rs7566605 may increase the risk of obesity when combined with physical inactivity requires detailed information about the basic functional properties of this gene, such as how its expression is controlled and how this expression may be affected by environmental factors. In addition, environmental risk factors such as physical inactivity need to be understood from a broad social environmental context within studies of gene-environment interplay if targeted intervention or prevention strategies are the ultimate goals.

Although gene-environment interactions relevant to human disease have been identified, opportunities and strategies to investigate gene-environment interactions have been difficult and limited to date, most of the studies that investigate gene-environment interaction typically have focused on a single exposure or candidate gene. However, it is fairly routine for environmental epidemiology studies to include candidate gene effects so there are some identified interactions between specific environmental and genetic susceptibilities. For example, results from one study showed interaction between *N*-acetyltransferase 2 (NAT2) gene variants and smoking leading to bladder cancer. Compared to *NAT2* rapid/intermediate acetylators, *NAT2* slow acetylators had an increased overall risk of bladder cancer [Odds Ratio (OR) = 1.4] that was stronger for cigarette smokers than for never smokers (OR = 1.8; *P*-interaction = 0.008) [Garcia-Closas et al., 2005; Thomas, 2010a,b]. Another gene-environment study includes an interaction between monoamine oxidase-A (MAOA) variants and contextual adversity as predictive of antisocial behavior. Individuals possessing the low-MAOA activity allele are predisposed to become antisocial when they experience adverse experiences in childhood. This interaction of a functional MAOA gene polymorphism (MAOA-uVNTR) and childhood adversity has been replicated repeatedly [Belksky et al., 2009; Kim-Cohen et al., 2006; Widom and Brzustowicz, 2006]. Despite these examples, new discoveries of gene-environment interaction remain a challenge.

In 2006, the National Institutes of Health (NIH) prioritized the investigation of complex factors that contribute to health and disease by establishing the *Genes, Environment and Health Initiative* (GEI; http://www.gei.nih.gov). The primary goals of GEI were two fold: (1) identify novel associations of genetic variants with human disease and disorders; and (2) develop new exposure biology tools to improve measures of environmental factors potentially important in complex disease. Environment was broadly defined to include airborne chemical and biological agents, dietary intake, physical activity, addictive substances, and psychosocial stress. In addition, GEI set out to develop and validate new methods for identifying the interaction between environmental exposures and genetic factors that contribute to human disease. Although the GEI has discovered many new genetic susceptibility loci and developed multiple exposure assessment tools and biomarkers of exposure, the effort to date has been focused on establishing a foundation and has targeted either genetic or environmental contributors to disease separately; a comprehensive approach to understanding gene-environment interactions that underlie major diseases still faces numerous challenges, including inadequate tools and methods for incorporating gene and environment measures in the analysis; limited availability or lack of standardization of

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

exposure measures; inappropriate study designs; and underpowered studies for capturing gene-environment interactions.

THE GEI WORKSHOP: GENE-ENVIRONMENT INTERPLAY IN COMMON COMPLEX DISEASES—FORGING AN INTEGRATIVE MODEL

On January 25, 2010, over 150 scientists representing a wide array of scientific fields met to evaluate the state-of-the-science in the study of gene-environment interplay in complex disease and made recommendations on needs and next steps for the field. Speakers discussed approaches utilized for identifying genetic and/or environmental risk factors for complex disease, including genetic epidemiology, statistical genetics, environmental epidemiology, epigenetics, molecular exposures, developmental psychology, and social sciences. Recommendations were made for needs related to the investigation of gene-environment interaction from their unique perspectives and fields of study. The need to incorporate more exposure measures into genetic studies, data sharing, and data harmonization were some overarching themes of the presentations.

Laura Bierut, Washington University, noted that the genetic risks for complex diseases that have been detected in GWAS are modest, raising the possibility that rare variants are important and/or that environmental factors may also be at work [Cornelis et al., 2010]. This requires that we pursue large-scale studies, with sample sizes in the tens of thousands, which would best be done through consortia in which phenotypes and environmental exposures are harmonized.

Peter Kraft, Harvard School of Public Health, warned that there are limits to what statistical tools can tell us about disease etiology. He encouraged meta-analyses of existing data as well as stratification by exposure, which could reveal effects in subsets of the population that are obscured when looking at a total population sample.

Frederica Perera, Columbia University, highlighted the relative ease of access to the genome compared with the challenges of measuring environmental exposures and responses in assessing gene-environment interaction studies, suggesting that measures of environmental exposures and responses should be further developed.

Eric Turkheimer, University of Virginia, suggested that the study of gene-environment interplay can benefit from lessons learned in the behavioral and social sciences. He noted that many of the problems faced by social scientists involve small and non-additive effects of large numbers of potential causes that cannot be experimentally controlled. He pointed out that social scientists have learned not to depend entirely on statistical significance, but also to consider non-significant results as an indication of replicable causal effects.

Jason Boardman, University of Colorado at Boulder, also suggested that social scientific studies provide a useful framework for gene-environment studies because of the long tradition of theorizing, measuring, and modeling the environmental influences on complex phenotypes. Social scientists broadly construe the environment as multilevel (e.g. families, neighborhoods, schools, or workplaces), multidimensional (e.g. normative or institutional) and longitudinal (e.g. historical changes and intra-individual change). He proposed that

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

hypotheses should be built upon evidence from heritability by environment (H × E) studies. For example, one could quantify the heritability of complex phenotypes, test for variation in heritability across environments, test gene-environment interaction theories about the source of this variation, and then demonstrate this association within this framework. Since genetic and environmental factors clearly interact to produce complex diseases, genetic variants should not be studied in isolation from the social and physical environment. Dr. Boardman also stressed the incorporation of cross-sectional and/or longitudinal environmental measures and large sample sizes.

The need for integration of environment, genetics, and epigenetics in the same study was emphasized by Dani Fallin of Johns Hopkins University as there is growing evidence of epigenetic changes due to environmental exposures. Thus, epigenetics can act as a mediator of environmental exposure through genetic regulation. Developmental windows of susceptibility and tissue-specificity of epigenetic effects will add to the complexity of these studies. All speakers noted that study design should be informed by gene-environment hypotheses and negative results should be published to inform future hypotheses. Speakers also agreed that there is a need for development of new statistical methods and measurement tools.

Following the presentations, workshop participants formed breakout groups and were charged with answering the overarching question: "Given the current state-of-the-science and the progress made in the GEI program and other gene-environment studies to date, what are optimal ways to move forward with investigations of gene-environment interplay in health and disease?" Investigators were asked for their input in three content areas: (1) theory and study design; (2) methods and data analysis, and (3) phenotypes, endophenotypes, and other variables. To prompt participants, 12 questions (Table I) were asked to serve as a catalyst for discussion.

RECOMMENDATIONS

Participants reconvened to hear the recommendations from each breakout group and discuss overall recommendations and strategies to move forward. The recommendations below capture the main points and suggestions from the participants of the workshop, but do not necessarily represent a consensus view of all participants.

A variety of approaches are needed to capture data on environmental exposures, individual genomes, and epigenomes, each of which is likely to contribute to the etiology of disease. Targeted as well as broad approaches are needed in studies of gene-environment interplay, depending on the research question, and studies should include multiple genes and environmental exposures whenever possible. Whichever approach is used, the downstream goals should include ways to help inform policy decisions, address health disparities, and improve public health. Targeted studies (hypothesis driven) that focus, for example, on a specific phenotype, disease, or environmental exposure, are best for investigating situations in which there are known disease associations with particular genes and/or environmental factors. In these situations case-control (including nested-case control) study designs and family study designs may be particularly valuable. These approaches are more cost effective

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

than discovery-driven designs, but focus upon particular hypotheses, and therefore, can miss important effects that lie outside the study design. Broad (discovery-driven, rather than hypothesis-driven) agnostic approaches may be more appropriate in situations where it is unknown how the environment is modifying genetic effects or vice versa. Although more costly than a targeted approach, genome-wide studies can provide a more thorough characterization of the potential interrelationships of environmental exposures, phenotypes, and phenotype groups. Discovery-based research that incorporates both genomics and environmental exposures should be encouraged in both environmental research as well as genomics.

A benefit of using existing population studies for the study of gene-environment interplay is the ability to leverage existing investments, as many completed and ongoing studies have sophisticated phenotypic and exposure measures, follow-up information, and stored biological specimens. Existing cohorts and intervention studies can sometimes be supplemented to collect new data, whether genetic or environmental, to enable the study of gene-environment interplay.

Some environmental factors can be assessed at time points that predate the onset of the disease. Examples include: data from cohort studies' baseline, special exposures documented in databases (e.g. toxins released from industrial sites recorded in emissions inventories), or medical exposures that are extracted from patient records. One could also conduct secondary analyses using archived data and biospecimens which are rich resources for studies of gene-environment interplay. Specimen banks for newborns' bloodspots can serve as a resource in which prenatal exposures can be measured retrospectively. Exposures that accumulate slowly over time, are intermittent, or have a short half-life are particularly challenging, but some biomarkers can provide a record of exposures during previous time periods. The problem of assessing past exposure may be attenuated when the lag between etiologically relevant periods of exposure and onset of disease is short, i.e. when the condition is characterized by acute onset. In addition, measurement tools and resources being developed for gene-environment studies to improve precision or enhance potential for future data harmonization can be leveraged in ongoing studies. These include measurement technology in the form of more sophisticated biomarkers and environmental sensors being developed and validated in the GEI Exposure Biology Program (GEI EBP; www.gei.nih.gov/exposurebiology/). The NHGRI-funded Consensus Measures for Phenotypes and Exposures (PhenX; https://www.phenx.org/), which provides standardized, validated measurement tools for high-priority phenotypes and exposures for GWAS and Phenotype Finder IN Data Resources (PFINDR; http://www.nhlbi.nih.gov/resources/pfindr.htm), a tool to support cross-study data discovery among NHLBI genomic studies, are other valuable resources for the scientific community [Stover et al., 2010].

Existing studies or cohorts may also present some disadvantages for gene-environment interaction studies, as most studies were not originally designed to identify this complexity. Existing studies often lack exposure data from the relevant time of risk, often do not have variables that span disease domains and/or do not provide detailed exposure information relevant to hypotheses, and informed consents may not allow for data sharing or new uses of data or specimens, all of which are important in leveraging these resources. In addition,

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

existing cohorts often do not include populations with sufficient or appropriate representation of the diverse racial, ethnic, age groups, and social backgrounds which may be necessary to detect population or group-specific gene-environment interactions.

New cohorts and study designs are necessary for detecting the multiple genetic and environmental factors that lead to human disease. Desirable characteristics of new cohort studies for the study of gene-environment interaction generally include: large sample size; diverse demographic representation; a broad range of genetic backgrounds and environmental exposures with early and recent exposure data; a broad array of clinical and laboratory measures with regular follow up over long periods of time; high-quality endpoint ascertainment and documentation; and measurements that are appropriate for the cohort(s) being studied. Furthermore, attention to the selection of participants in case-control studies is needed in order to enhance environmental variation as the limited range of exposures and/or limited numbers of subjects in critical exposure groups are impediments to assessing the exposure effects alone, and the problem may be magnified in gene-environment interaction studies. Studies should also have policies and procedures in place for collection and storage of biological specimens and open access of materials and data to other researchers; researchers should develop plans for re-contacting individuals for additional experimental studies, or for follow-up clinical care. There is also a great need for studies in underserved populations (e.g. American Indians, African Americans, immigrants, low-income individuals, the elderly and children) that often bear a disproportionate burden of certain diseases and/or exposures.

It is important to measure environmental exposures at appropriate time points, because many genes are only expressed during specific developmental periods, and some exposures may have greater impact during specific developmental stages. Potential sensitive periods for environmental exposure include but are not limited to time of conception, gestation, infancy, and puberty. Due to the sporadic or cumulative nature, and/or "sensitive timing" of environmental exposures over a lifetime, a large, longitudinal population study will be needed in order to identify some gene-environment interactions.

Much larger sample sizes are needed for detection of interactions than for main effects [Thomas, 2010]. Thousands of cases and controls are needed to detect interaction relative risks of about 1.5 for a candidate gene study or tens of thousands for a GWAS. Power for detecting interactions would be further diminished by measurement error in either exposure or genotype and can have unpredictable effects on the direction of an interaction, particularly if one or both is differentially misclassified. Many designs can be appropriate for studying a complex disease, but the sample size requirements for gene-environment studies will depend on many factors, including the prevalence and dose of the environmental variable(s) of interest, allele frequencies, effect sizes, outcome(s), effect modifiers, and/or covariates. Certain study designs can reduce the needed sample size by altering those parameters. For instance, by enriching a cohort with subjects having a rare variant, researchers can increase the prevalence of that variant within their study population. Likewise, increasing the prevalence of an exposure or effect modifier of interest through over-sampling will alter the required sample size. Similarly, the use of animal studies may allow increased levels of the exposure so that the exposure effect size increases dramatically,

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

thereby reducing the needed sample size, although this approach has to be tempered by a need for realistic "doses" of environmental exposure.

Type 1 errors continue to be a concern in the analysis of gene-environment interaction as in other large-scale genomic analyses. Methods to address type 1 errors include the integration of known pathway information to inform analysis; testing for interaction in an environmental subgroup to identify genetic effects only apparent in certain environmental exposures; utilizing case only designs to increase the power to test for interactions; and family-based designs to avoid bias from population stratification [Thomas, 2010]. Other examples of research designs that allow the investigation of complex diseases without prohibitively large sample sizes include: studies of controlled environment(s), intervention studies, and clinical trials. The use of biomarkers as a refinement of outcome or exposure measures and data reduction methods may also decrease the need for large samples. Other exploratory methods such as multiple regression and pattern recognition can also be used for the evaluation of gene-environment interplay.

Studies utilizing cell lines and animal models will be important for elucidation of the function of genetic variants identified by GWAS and for mechanistic studies to understand how environmental factors interact with genetic factors to influence health and disease. Data simulation and animal models can be useful in developing theories about the functionality of a given gene variant but may not be applicable to humans. Environmental manipulation need not be limited to animal models, but can also be conducted with humans, such as human chamber studies (controlled environmental exposures) or special epidemiologic studies (e.g. occupationally exposed individuals). Animal models can also be useful for conducting basic mechanistic studies.

Yet, there remains a need for the development of more sophisticated analytical methods and approaches for gene-environment interaction studies to explore the multiple levels of data in gene-environment interplay. Continued support is needed for bioinformatic and biostatistical tools and methods development. In addition, the field suffers from a shortage of investigators trained in computational biology and statistical methodology capable of developing new methods and analytic tools. Interdisciplinary training programs in computational, statistical, and molecular biology are also needed to train the next generation on global approaches to research. A database for standardized approaches, tools, and GWAS and gene-environment study results would also benefit the research community. Additionally, existing resources for environmental measurement tools and databases [e.g. GEI Exposure Biology program, PhenX, the Pharmacogenomics Knowledge Base (PharmGKB; http://www.pharmgkb.org), the Database of Genotypes and Phenotypes (dbGaP; http://www.ncbi.nlm.nih.gov/gap), etc.] should be leveraged to aid in harmonization of data from multiple studies [Altman, 2007; Mailman et al., 2007; Stover et al., 2010].

Measurement of phenotypes should be precise and accurate, well correlated with the disease, disorder, or trait of interest, easy to measure, low burden, and low cost. However, 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 and use new endophenotypes and

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

biomarkers that more precisely characterize both the exposure and early or subclinical features of the disease. Across studies, there is a need to use more intensive measures on high risk subsets of the study population (e.g. siblings of autistic children). As much information should be captured as possible on subsets, including data on physical and social environments and molecular markers. However, in addition to more intensive and targeted measures, a large number of phenotypic measures that are cheaper and easier to collect could be obtained and harmonized in many studies, thereby allowing for pooling of phenotype results among studies. To further assist phenotype harmonization across studies, phenotypes should be collected, whenever possible, using standardized methods and approaches, e.g. those outlined in PhenX [Stover et al., 2010].

Those conducting phenotype-specific studies should be cautious in assuming that phenotype measurements are the same across studies. More investment is needed in defining and refining the phenotype so that data can be analyzed, harmonized, and interpreted successfully. Previous family studies and linkage studies may be useful to improve phenotyping and determining heritable components, but they must be done carefully, as meta-analyses can provide different results than those of individual studies due to differences in sample ascertainment and phenotype heterogeneity [Manolio et al., 2009].

It is important to focus on complex phenotypes as well as search for associations to the underlying mechanisms or endophenotypes as each approach yields different information. Although the search for intermediate phenotypes or endophenotypes is desirable, measurement of endophenotypes can be as complicated as that of the disease outcome. On the other hand, the search for intermediate phenotypes can lead to the discovery of new molecular phenotypes. Challenge or intervention studies are one approach, when possible, to confirm endophenotypes and gene-environment interactions. This is particularly important when endophenotypes include behaviors that are sensitive to gene-environment correlation. It is important to note that measurement of biomarkers may be complex and needs to account for co-morbidities and the presence of multiple exposures. Also, the importance of the cell type affected in the disease process should not be underestimated. Longitudinal information can help identify time and age-dependent phenotype and biomarker measures and their genetic underpinnings.

DISCUSSION

Although there has been progress in the study of gene-environment interplay, to date relatively little is known about the effects of the interaction of genes and the environment on human disease, let alone how these interactions might change over time. GEI Workshop participants concluded that targeted as well as broad approaches are needed to study gene-environment interactions. Regardless of which approach is used, the multiple interactions that lead to complex disease should be taken into account. A systematic approach to studying gene-environment interaction will increase our understanding of how environmental factors induce disease and toxicity and will provide insight for treatment and prevention [Mukherjee et al., 2009b]. Environmental exposures likely interact with genes through numerous mechanisms; therefore, the study of complex diseases requires a comprehensive model involving multiple genes and multiple environmental exposures and

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

must also consider gene by gene interactions, environment by environment interactions, and other forms of gene-environment interplay [Cordell, 2009; Hodgins-Davis and Townsend, 2009; Thomas, 2010].

A variety of new promising approaches can assist in the next generation of studies of complex diseases that incorporate gene-environment interactions. Although GWAS have led to novel findings of new pathways and enlightened our understanding of biological processes, it is often hard to understand the statistical evidence of gene-environment interactions without supporting evidence of biological plausibility [Hindorff et al., 2009]. Thus, incorporation of previously identified biological information into the discovery phase may be a way to reduce false positives and understand results of hypothesis free studies [Hindorff et al., 2009]. Environment Wide Association Studies could be useful in detecting environmental factors associated with some complex disease as some environmental factors can be found with effect sizes comparable to or far exceeding loci identified by GWAS [Patel et al., 2010]. Complementary experimental and observational research designs are also important in testing gene-environment interactions [Murcray et al., 2009]. For example, experiments with animals elucidate biological mechanisms and can validate findings from human observational studies [Murcray et al., 2009]. The use of large-scale genomic platforms in conjunction with a priori planned replication studies would also reduce the number of false positives. For example, replication studies should have a similar range of exposures as the initial study, since studies sampling from different ranges of the full exposure distribution pattern may fail to replicate the original interaction (Fig. 1) [Hindorff et al., 2009; Khoury et al., 2009; Kraft and Hunter, 2010; Moore and Williams, 2009; Shuldiner, 2009]. Selecting subjects based on their environment (e.g. exposure extremes, rare exposures) may help find genetic effects. Stratification by exposure may also reveal effects in subsets of the population that are obscured when looking at a population sample. Misclassification of complex exposures has been a problem in previous studies; therefore, future studies are needed that incorporate biomarkers collected in traditional epidemiological contexts as well as built in new biosensor technology. Systems biology may provide experimental approaches to assess molecular components of the biologic system and assess interactions. However, there remains a large gap between modeling interactions in cellular and biologic processes and the ability to use the information in population studies [Khoury and Wacholder, 2008]. A systems science approach that includes chemical, behavioral, and social environment may help fill that gap and be more useful for studies of population health.

Workshop participants noted the advantages and disadvantages of utilizing existing cohorts and development of new cohort studies. The importance of data sharing and data harmonization was highlighted for both existing and new cohort studies. New cohorts are needed to evaluate the effects of environmental exposures during certain developmental stages. Genetic variants that cause susceptibility to a particular environmental exposure may only do so when the exposure occurs during a developmental window in which the gene is highly expressed as exposure to particular environments during specific stages of development prevents or exacerbates genetic vulnerability to disease [Hindorff et al., 2009; Khoury et al., 2009]. Case-control studies often do not have the power to detect multiplicative interaction due to the small numbers of cases or controls with the same

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

genotype [Mukherjee et al., 2009a]. Also, case-control or family-based designs may collect environmental data at a single point in time, and may subsequently miss periods of risk or vulnerability. If these environmental measures are not reflective of long-term exposures, then the studies are limited not only in their ability to determine exposure-disease associations, but also in their ability to detect gene-environment interactions [Mukherjee et al., 2009a].

Although cohort studies are costly and require long follow up, they allow for exposures and risk factors to be characterized before disease onset [Landrigan et al., 2006; Mukherjee et al., 2009a]. Cohort studies therefore allow investigators to overcome the problem of temporal uncertainty by collecting time-dependent exposure information before disease develops, reducing selection bias, and avoiding the problem of reverse causation [Le Marchand and Wilkens, 2008; Mukherjee et al., 2009a]. Prospective environmental data collection may identify the role of critical windows of susceptibility that likely correspond to the expression of specific genes and gene pathways [Mukherjee et al., 2009a]. Prospective studies also permit evaluation of environmental exposures within individual, family, neighborhood, and/or societal levels [Boardman, 2009; Boardman et al., 2008; Landrigan et al., 2006]. However, cohort studies may be difficult to conduct when the participant burden for intensive or repeated exposure measurement is high, leading to low or biased retention rates. Nested case-control and case-cohort designs may be used for cost-efficiency and may retain the same advantages as cohort studies [Thomas, 2010].

A need for low-cost, precise, validated, and standardized environmental measures was noted repeatedly during the workshop. It is critical to assess environmental exposures precisely and accurately. This can be done by proximal measures of environmental pathogens, multiinformant data, developmental-specific assessments, and by identifying cumulative effects of environmental influences [Van der Zwaluw and Engels, 2009]. For instance, the measurement of biomarkers may allow investigators to understand biological disease processes that account for gene-environment interaction effects by measurement of intermediate metabolites [Thomas, 2010]. Consensus phenotypic measures from resources like PhenX and tools developed by the GEI Exposure Biology Program should be leveraged when designing gene-environment research studies so that data can be harmonized across studies.

Lastly, sufficient power for the detection of gene-environment interaction effects was a repeated theme. Large sample sizes may be needed for ensuring sufficient number of subjects (in key groups) for agnostic studies of gene-environment interaction as these studies require more power than hypothesis driven studies [Murcray et al., 2009]. Gene-environment interaction studies that recruit participants on the basis of their genotype and environmental exposure are better powered to test for genetic variants that increase susceptibility to environmental exposures or environmental exposures that influence genetic susceptibility [Murcray et al., 2009]. However, many environmental and genetic factors are often unknown in observational studies and require more testing which decreases power [Le Marchand and Wilkens, 2008].

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

CONCLUSION

In conclusion, the participants of the workshop, *Gene-Environment Interplay in Common Complex Diseases*: *Forging an Integrative Model*, felt that a comprehensive approach to the study of gene-environment interplay is needed. A diverse set of study designs and sample sizes are appropriate, depending on current knowledge and the specific goals of the study. For example, questions addressing developmental or life stages, environmental changes, changes in gene expression, and disease latency may require longitudinal studies starting early in life. Targeted studies are best for detecting particular gene-environment interactions when disease associations are known, whereas studies that are hypothesis-generating are more appropriate when less is understood about a given disease. Environmental measurement technologies should be developed and incorporated into ongoing and new studies of gene-environment interplay. Finally, an emphasis should be placed on data sharing and harmonization to maximally leverage existing population studies.

Acknowledgments

The authors thank the participants of the GEI Workshop: Gene-Environment Interplay in Common Complex Diseases—Forging an Integrative Model for their thoughtful discussions and input into the recommendations.

REFERENCES

Altman RB. PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. Nat Genet. 2007; 39:426. [PubMed: 17392795]

Amato K, Pinelli M, D'Andrea D, Miele G, Nicodemi M, Raiconi G, Cocozza S. A novel approach to simulate gene-environment interactions in complex disease. BMC Bioinformatics. 2010; 11:8. [PubMed: 20051127]

Andreasen CH, Mogensen MS, Borch-Johnsen K, Sandbæk A, Lauritzen T, Sorensen TI, Hansen L, Almind K, Jorgensen T, Pedersen O, Hansen T. Non-replication of genome-wide based associations between common variants in INSIG2 and PFKP and obesity in studies of 18,014 Danes. PLoS One. 2008; 3:e2872. [PubMed: 18682847]

Belksky J, Jonassaint C, Pluess M, Stanton M, Brummett B, Williams R. Vulnerability genes or plasticity genes? Mol Psychiatry. 2009; 14:746–754. [PubMed: 19455150]

Boardman JD. State-level moderation of genetic tendencies to smoke. Am J Public Health. 2009; 99:480–486. [PubMed: 19150910]

Boardman JD, Jarron M, Onge J, Haberstick BC, Timberlake DS, Hewitt JK. Do schools moderate the genetic determinants of smoking? Behav Genet. 2008; 38:234–246. [PubMed: 18347970]

Caspi A, Hairiri AR, Holmes A, Uher R, Moffitt TE. Genetic sensitivity to the environment: the case of the serotonin transporter gene its implications for studying complex diseases traits. Curr Opin Lipidol. 2010; 21:136–140. [PubMed: 20154612]

Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet. 2009; 10:392–404. [PubMed: 19434077]

Cornelis MC, Agrawal A, Cole JW, Hansel NN, Barnes KC, Beaty TH, Bennett SN, Bierut LJ, Boerwinkle E, Doheny KF, Feenstra B, Feingold E, Fornage M, Haiman CA, Harris EL, Hayes MG, Heit JA, Hu FB, Kang JH, Laurie CC, Liang H, Manolio TA, Marazita ML, Mathias RA, Mirel DB, Paschall J, Pasquale LR, Pugh EW, Rice JP, Udren J, van Dam RM, Wang X, Wiggs JL, Williams K, Yu K. GENEVA Consortium. The Gene, Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. Genet Epidemiol. 2010; 34:364–372. [PubMed: 20091798]

García-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, Tardón A, Serra C, Carrato A, García-Closas R, Lloreta J, Castaño-Vinyals G, Yeager M, Welch R, Chanock S, Chatterjee N, Wacholder S, Samanic C, Torà M, Fernandez F, Real FX, Rothman N. NAT2 slow

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet. 2005; 366:649–659. [PubMed: 16112301]

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA. 2009; 106:9362–9367. [PubMed: 19474294]

Hindorff LA.; Junkins HA.; Manolio, TA. [Accessed on August 6, 2010] A catalog of published genome-wide association studies. Available from: http://www.genome.gov/gwastudies/.

Hodgins-Davis A, Townsend JP. Evolving gene expression: from G to E to G × E. Cell. 2009; 24:649–658

Kazma R, Bonaiti-Pellie C, Norris JM, Génin E. On the use of sibling recurrence risk to select environmental factors liable to interact with genetic risk factors. Eur J Hum Genet. 2010; 18:88–94. [PubMed: 19584901]

Kazma R, Bobron M-C, Génin E. Genetic association and gene-environment interaction: a new method for overcoming the lack of exposure information in controls. Am J Epidemiol. 2011; 173:225–235. [PubMed: 21084555]

Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to genome-environment-wide interaction studies—challenges and opportunities. Am J Epidemiol. 2008; 169:227–230. [PubMed: 19022826]

Khoury MJ, Bertram L, Boffetta P, Butterworth AS, Chanock SJ, Dolan SM, Fortier I, Garcia-Closas M, Gwinn M, Higgins JPT, Janssens ACJW, Ioannidis J, Owen RP, Pagon RA, Rebbeck TR, Rothman N, Bernstein JL, Burton PR, Campbell H, Chokkalingam A, Furberg H, Little J, O'Brien TR, Seminara D, Vineis P, Winn DM, Yu W, Ioannidis JPA. Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. Am J Epidemiol. 2009; 170:269–279. [PubMed: 19498075]

Kim-Cohen J, Caspi A, Taylor A, Williams B, Newcombe R, Craig IW, et al. MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. Mol Psychiatry. 2006; 11:903–913. [PubMed: 16891953]

Kraft, P.; Hunter, D. The challenge of assessing gene-environment and gene-gene interactions. In: Khoury, MJ.; Bedrosian, SR.; Gwinn, M.; Higgins, JPT.; Ioannidis, JPA.; Little, J., editors. Human genome epidemiology. New York: Oxford University Press; 2010.

Landrigan P, Trasco L, Thorpe L, Gwynn C, Lioy P, D'Alston ME, Lipkind HS, Swanson J, Wadhwa PD, Clark EB, Rauh VA, Perera FP, Susser E. The National Children's Study: a 21-year prospective study of 100,000 American children. Pediatrics. 2006; 118:2173–2186. [PubMed: 17079592]

Le Marchand L, Wilkens LR. Design considerations for genomic association studies: importance of gene-environment interactions. Cancer Epidemiol Biomarkers Prev. 2008; 17:263–267. [PubMed: 18268108]

Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007; 39:1181–1186. [PubMed: 17898773]

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–753. [PubMed: 19812666]

Mechanic LE, Luke BT, Goodman JE, Chanock SJ, Harris CC. Polymorphism interaction analysis (PIA): a method for investigating complex gene-gene interactions. BMC Bioinformatics. 2008; 9:146. [PubMed: 18325117]

Moore JH, Williams SM. Epistasis and its implications for personal genetics. Am J Hum Genet. 2009; 85:309–320. [PubMed: 19733727]

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Mukherjee B, Ahn J, Gruber SB, Rennert G, Moreno V, Chatterjee N. Tests for gene-environment interaction from case-control data: a novel study of type 1 error, power and designs. Genet Epidemiol. 2009a; 32:615–626. [PubMed: 18473390]

Mukherjee O, Sanapala KR, Anbazhagana P, Ghosh S. Evaluating epistatic interaction signals in complex traits using quantitative traits. BMC Proc. 2009b; 3:S82. [PubMed: 20018078]

Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. Am J Epidemiol. 2009; 169:219–226. [PubMed: 19022827]

Ordovas JE, Tai ES. Why study gene-environment interactions? Curr Opin Lipidol. 2008; 19:158–167. [PubMed: 18388696]

Patel CJ, Bhattacharya J, Butte AJ. An environment-wide association study (EWAS) on type 2 diabetes mellitus. PLoS. 2010; 5:e10746.

Rutter M, Moffitt TE, Caspi A. Gene-environment interplay and psychopathology: multiple varieties but real effects. J Child Psychol Psychiatry. 2006; 47:226–261. [PubMed: 16492258]

Shuldiner AR. Obesity genes and gene-environment-behavior interactions: recommendations for a way forward. Obesity. 2009; 16:S79–S81. [PubMed: 19037219]

Stover PJ, Harlan WR, Hammond JA, Hendershot T, Hamilton CM. PhenX: a toolkit for interdisciplinary genetics research. Curr Opin Lipidol. 2010; 21:136–140. [PubMed: 20154612]

Thomas D. Gene-environment-wide association studies: emerging approaches. Nat Rev Genet. 2010a; 11:259–272. [PubMed: 20212493]

Thomas D. Methods for investigating gene-environment interactions in cidate pathway genome-wide association studies. Annu Rev Public Health. 2010b; 31:21–36. [PubMed: 20070199]

Van der Zwaluw CS, Engels RCME. Gene-environment interactions alcohol use dependence: current status future challenges. Addiction. 2009; 104:907–914. [PubMed: 19466917]

Widom CS, Brzustowicz LM. MAOA and the "cycle of violence": childhood abuse and neglect, MAOA genotype, and risk for violent and antisocial behavior. Biol Psychiatry. 2006; 60:684–689. [PubMed: 16814261]

Wright RO, Christiani D. Gene-environment interaction children's health development. Curr Opin Pediatr. 2010; 22:197–201. [PubMed: 20090521]

EVALUATION Formatter

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Fig. 1.

(A) The effects of genetic, environment, and gene-environment interaction depend on what part of the environment is sampled. (B) Genetic effects are not phenotypically expressed in the common environments (G = Genetic effect; E = Environmental effect) [Kraft and Hunter, 2010].

AHFormatter EVALUATION

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE 1

GEI workshop breakout questions

GEI workshop breakout questions	
Theory and study design	
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?
Methods and data analysis	
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 G × E studies?
Q8:	What statistical tools and resources are needed?
Phenotypes, endophenotypes and other variables	
Q9:	What are the characteristics of end-points or variables that are "ready to go" for G × E 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?

AH Formatter V6.2 MR6 (Evaluation) http://www.antennahouse.com/