American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2017.

Vol. 186, No. 7
DOI: 10.1093/aje/kwx227
Advance Access publication:
July 15, 2017

Special Article

Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases

Kimberly McAllister, Leah E. Mechanic*, Christopher Amos, Hugues Aschard, Ian A. Blair, Nilanjan Chatterjee, David Conti, W. James Gauderman, Li Hsu, Carolyn M. Hutter, Marta M. Jankowska, Jacqueline Kerr, Peter Kraft, Stephen B. Montgomery, Bhramar Mukherjee, George J. Papanicolaou, Chirag J. Patel, Marylyn D. Ritchie, Beate R. Ritz, Duncan C. Thomas, Peng Wei, and John S. Witte, on behalf of workshop participants

* Correspondence to Dr. Leah E. Mechanic, Genomic Epidemiology Branch, Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, 9609 Medical Center Drive, Room 4E104, MSC 9763, Bethesda, MD 20892 (e-mail: mechanil@mail.nih.gov).

Initially submitted November 7, 2016; accepted for publication March 16, 2017.

Recently, many new approaches, study designs, and statistical and analytical methods have emerged for studying gene-environment interactions (G×Es) in large-scale studies of human populations. There are opportunities in this field, particularly with respect to the incorporation of -omics and next-generation sequencing data and continual improvement in measures of environmental exposures implicated in complex disease outcomes. In a workshop called "Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases," held October 17–18, 2014, by the National Institute of Environmental Health Sciences and the National Cancer Institute in conjunction with the annual American Society of Human Genetics meeting, participants explored new approaches and tools that have been developed in recent years for G×E discovery. This paper highlights current and critical issues and themes in G×E research that need additional consideration, including the improved data analytical methods, environmental exposure assessment, and incorporation of functional data and annotations.

environmental exposure; gene-environment interaction; genome-wide association study

Abbreviations: GWAS, genome-wide association study; G×E, gene-environment interaction.

Genetic and environmental factors are thought to contribute to the etiology of most complex diseases. Through genome-wide association studies (GWAS), thousands of common loci associated with complex diseases have been identified (1–3). Researchers have been motivated to discover and describe how the interplay of these factors influences disease risk and outcomes. There are several reasons to study gene-environment interaction (G×E): providing insights into the biology of disease (e.g., developing new models for disease etiology based on observed G×E findings), building better prognostic models (e.g., using genotype to inform treatment and prognosis), identifying possible high-penetrance subgroups (e.g., increased genotype-specific risk in premenopausal women), or identifying genetic subgroups with higher exposure-specific disease risk for prevention efforts (e.g., increased environmental-specific risk for individuals with a particular genotype) (4–7). Furthermore, in the search for novel genes via GWAS, the modifying effects of environmental risk factors are not often taken into account; therefore, leveraging G×E may result in discovery of additional disease susceptibility loci (5, 8, 9). Despite interest in G×E, there are few agreed-upon successes where the effect of exposure differs across genotypes (and vice versa). Numerous reasons have been suggested as contributors to the small number of successes, including the inherently low power of tests for G×E, the complexity of measuring environmental exposures, the difficulty of incorporating the temporality of environmental exposures, measurement error, limited range of genetic and/or environmental variation, scale dependence in the definition of statistical interaction, and lack of data on the biological consequences of most genetic variants (10–13).

The past few years have seen an emergence of new approaches, study designs, and statistical and analytical methods for exploring G×Es in large-scale studies of human populations. Further, new opportunities in this field continue to be developed with respect to the incorporation of -omics and next-generation sequencing data and improvements in measures of environmental exposures implicated in complex disease outcomes. Therefore, on October 17–18, 2014, the National Institute of Environmental Health Sciences and National Cancer Institute held a workshop at the 64th Annual Meeting of the American Society of Human Genetics—"Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases"—to explore these new approaches and tools for G×E discovery. Based on the discussions, we prepared 4 articles that provide an update on: 1) the state of the science in analytical methods (14); 2) opportunities for incorporation of biological knowledge into G×E analyses (15); 3) advances in environmental exposure assessment in human populations (16); and 4) lessons learned from G×E successes (17). In addition, this article develops some overarching themes and sets the stage for this series. Because environmental factors may be modifiable, defining subpopulations of individuals that are most susceptible to environmental factors through G×E analysis may provide targets to improve public health. This idea is consistent with the goal for President Obama's recently launched Precision Medicine Initiative at the National Institutes of Health: to better understand how individual variability contributes to differences in response to treatment or prevention (18, 19).

ANALYTICAL METHODS

G×E studies require much larger sample sizes than studies targeting either genetic or environmental main effects alone (20). Further, when performing G×E studies on a genome-wide scale, sometimes referred to as genome-wide interaction studies, sample size requirements are substantially further inflated to account for the multiple comparisons performed (5, 21). Therefore, a goal of the development of G×E methods has been to improve power to detect associations. As detailed in Gauderman et al. (14), many different methods have been explored in the context of case-control studies as alternatives to traditional G×E tests, including case-only studies (22), empirical Bayes (23), Bayesian model averaging (24), joint tests (9, 25, 26), case-parent approaches (27–29), and 2-step approaches (21, 25, 30–35). Other approaches include set-based methods, which combine multiple variants or G×Es and which may be particularly appropriate for studies of rare variants (36–40). In addition, several methods have been developed to analyze G×E for quantitative outcomes (41–48).

The large number of available methods, as well as novel software tools to support the application of these methods (31, 49, 50), creates opportunities to better study G×Es in genome-wide settings. Researchers may therefore wonder which method to use for their studies. Several previous simulation studies suggest that none of these G×E methods is universally the most powerful approach (31, 32, 51–54). Therefore, decisions about the most appropriate approach depend on several considerations, including the hypotheses to be tested, likely genetic architecture, study design attributes, and characteristics of the population being studied. Investigators should be cautious about applying multiple methods to their data without an a priori basis for choosing among the results—simply picking those with the most "significant" findings to report would clearly be a biased strategy that could contribute to spurious associations and to what has been referred to as a "vibration of effects" (55, 56). Some of the new methods, however, provide flexible frameworks for combining multiple tests with an appropriate permutation procedure to evaluate the significance of the overall results (31, 32). The collection of methods allows investigators to address specific scientific questions and offers new opportunities for studies of G×E in large populations.

FUNCTIONAL VALIDATION AND DISCOVERY

Despite the recent success of GWAS at identifying risk loci, variants identified are by design not usually the causal variants, defined as the functional genetic variant that influences risk of disease and explains the association. Currently, the underlying biological mechanism contributing to disease risk is only known for a small proportion of these loci. Therefore, more research to functionally characterize risk loci is now being performed, providing opportunities by which G×E analyses may offer new insights into disease development (57). An understanding of the biological consequences of particular genetic differences could lead to specific mechanistic hypotheses, identifying relevant exposures to test and specifying relevant statistical models. As described in Ritchie et al. (15), these approaches include using functional annotations for discovery and validation, studying molecular phenotypes (e.g., epigenetics or gene expression) to improve G×E discovery, and leveraging in vitro and in vivo models for these studies.

Several large public databases (such as the Encyclopedia of DNA Elements (ENCODE), Epigenomics Roadmap, Genotype-Tissue Expression (GTEx), and the Cancer Genome Atlas (TCGA)) have facilitated the functional annotation and interpretation of many genomic regions, which can be used to prioritize candidate G×E markers (32). Many disease-associated single-nucleotide polymorphisms identified in GWASs appear to be located in noncoding or regulatory regions, which are often affected by environmental exposures (58–60). The Encyclopedia of DNA Elements and Roadmap Epigenomics programs have helped to define many of the regulatory regions, and new tools developed by these programs and others now allow functional annotation information—such as the genomic location of histone modification states, methylation patterns, transcription-factor binding sites, and DNAse hypersensitivity sites or other higher-order chromosomal structural information—to be overlaid with GWAS results and could be integrated into G×E analyses (61–65). Projects like Genotype-Tissue Expression have greatly increased the compendium of putative biological functions of genetic variants. However, neither Genotype-Tissue Expression nor large-scale epigenomics projects provide information on effects of genetic and genomic functions across a range of environmental conditions. To explore genetic effects in response to environment, in vitro studies have now perturbed cells and recorded responses to various drugs, infections, and other exposures. Through the use of intermediate molecular phenotypes such as gene expression, these efforts have demonstrated success in illustrating how an exposure may influence gene function,

suggesting potential candidate genes or variants for G×E studies (66–70).

In addition to data resources, the use of population-based mouse resources (such as the Collaborative Cross, Diversity Outbred, and Hybrid Mouse Diversity Panel) and other appropriate mouse models have also been leveraged to assist in the discovery or replication of G×Es. These population-based, variant-enriched mouse resources have been designed to mimic the genetic diversity of human populations and can be used to replicate or inform G×E hypotheses by using carefully controlled exposures in mouse studies. Several recent examples have exemplified the power of these resources to map genetic variants related to susceptibility to environmental exposures (71, 72). Although both in vitro and model systems have led to potential mechanistic insights, linking of these to human populations remains challenging.

There are many approaches for incorporating biological knowledge to improve analytical methods (73) for G×E in both the discovery and the validation phase. Incorporating functional annotation data and a priori biological information (such as knowledge on biological pathways and metabolomics or gene-expression data collected on individuals) to inform methods for analyzing G×E data has aided in the discovery of new G×E findings in recent years (74). For example, Bayesian variable selection (75, 76), the Algorithm for Learning Pathway Structure (77), and the PEAK algorithm (74) are all methods that incorporate external biological information and properties of the data set itself to increase power over agnostic approaches to detect interactions. Another approach is to use 2-stage modeling where functional annotations are used to prioritize variants (78, 79) for G×E studies. As one example, Biofilter was designed to build biologically plausible models of gene-gene and G×Es to test for associations based on biological features using biological knowledge from the public domain (78, 80). These types of filtering approaches are also being explored to prioritize environmental exposures by using databases such as the Comparative Toxicogenomics Database, which links exposures to genes (81). However, challenges still exist in linking environmental exposures into currently available ontological knowledge resources, although some investigators are beginning to navigate these challenges (82). Furthermore, all these databases and functional annotations depend on the quality and extent of existing biological knowledge (73).

ENVIRONMENTAL EXPOSURES

The complex realities of environmental exposures have long made measurement of exposures substantially more complicated than inherited genetic measurements (e.g., genotypes) and single nucleotide variants in particular. The technologies and approaches to incorporate exposures into human population studies have therefore lagged behind genomics capabilities (11, 83). Assessing exposure impact must take into context not just the variety of exposures themselves (physical, often complex chemical mixtures, biological, and psychosocial) but also the source and place of exposure, the timing during a person's life trajectory, the route of contact (skin, lung, diet), metabolism/excretion, and distribution in target tissues. All of these factors may affect the ultimate disease risk associated with environmental exposures. In addition, in the classic environmental exposure paradigm, studies may focus on measurements to capture internal versus external exposure, early markers of disease, or an ultimate biological response, which further adds to the complexity of exploring the impact of environmental exposures.

In recent years, however, exciting new opportunities have become available for environmental exposure assessment. The potential importance of examining the totality of internal and external exposures, referred to as the "exposome," has been recognized (83, 84). Several recent commentaries have described considerations for measurements of the exposome (85–88). Innovative technologies—including activity monitors, improved sensors, global positioning systems, and geographic information systems—enable new and more detailed exposure measurements, although issues of the timing of exposure measures persist and should be considered. Moreover, development of biological response markers for assessment of exposure (such as changes in gene expression, transcriptomic signatures, and DNA methylation profiles) has been useful for G×E discovery (89–92). Another opportunity is the exploration of environmental exposures in a more agnostic discovery-based fashion, similar to GWASs. These studies, termed environment-wide association studies (EWASs), have led to discoveries of environmental factors associated with disease (93–96).

Key challenges and considerations remain associated with assessing environmental exposures in G×E studies (16), including how to select the most appropriate study designs, incorporate high throughput -omic measures (e.g., metagenome, metabolome) and sensor technologies into human population-based studies, assess long-term exposure, integrate a variety of divergent external exposure and internal response data, and further advance statistical approaches to handle the dynamic nature of exposure data. We are now at the early stages of exploring which novel exposure assessment technologies can be appropriately applied to larger population studies most effectively. To this end, some 2-stage study designs have been investigated (26, 97–100). Given the extreme cost of incorporating some sophisticated environmental measures into large-scale studies of human populations, the question of what can be accomplished with dense (i.e., repeated measures of a marker or measurement of multiple analytes using an -omic platform) environmental measures on a subsample and extrapolating to a larger sample size—and whether simulations can demonstrate that this approach increases power to detect G×E—is currently being explored (51, 97, 101).

Several analytical methods have been developed for the unique considerations of exposure assessment. New statistical methods can adjust much better for exposure misclassification (which has been shown to lead to inflated type I errors and substantially reduced power). These approaches should allow for obtaining greater power with smaller sample sizes. In addition, novel statistical methods have been developed to detect gene × longitudinal exposure interactions by taking into account long-term time-varying exposures (102). Importantly, as researchers begin to combine exposure data to obtain the larger sample sizes required for G×E research across studies, they have to address the possibility that exposures were measured using different approaches or have very different distributions in and between populations, such that exposure misclassification could produce spurious associations (14). There is also the challenge

of exposure-related population stratification for studies relating to G×Es (103). Meanwhile, multiple measures can sometimes increase power for detecting associations. For example, in a recent study, continuous monitoring was shown to reduce the sample size required in a clinical trial context (104).

G×E EXAMPLES FROM HUMAN POPULATION STUDIES

By examining G×E successes, it may be possible to improve the design of G×E studies for the future. Examples of G×E successes range from Mendelian-like traits (e.g., phenylketonuria) to complex diseases (*NAT2* variants, smoking, and bladder cancer) and response to therapies (*HLA-B**1502 variant and carbamazepine-induced Stevens-Johnson syndrome) (17). In addition, several recent studies examined the use of polygenic risk scores, generated from common genetic variation, to assess the impact of environmental factors on individuals with low versus higher genetic risk (105–108). In Ritz et al. (17), highlighting some of the most successful G×Es identified to date, several common themes have emerged, including the strength of focusing on metabolic pathways for a specific exposure; the utility of studying unique, highly, or diversely exposed populations; the necessity of using high-quality exposure assessment methods; the need for large sample sizes; and the utility of model systems to demonstrate genetic function when replication is challenging in population-based studies. These suggest important avenues for undertaking successful future research in G×E.

THEMES AND FUTURE DIRECTIONS

Inclusion of diverse populations may facilitate G×E research by improving power for discovery of causal genetic variants and environmental factors associated with disease. Transethnic differences in the distribution of linkage disequilibrium can be leveraged to improve fine mapping to identify potential causal alleles (109–112). Combining admixture mapping with conventional GWASs may also facilitate discovery of novel loci (113). Using this latter approach, novel loci were identified associated with total immunoglobulin E levels (114) and asthma (115). Last, using geographically diverse populations might expand the distribution of the environmental exposure and thus increase power to detect interactions (13). Performing genetic studies on populations of diverse ancestry might improve our understanding of disease mechanisms, and such studies are required to ensure all populations benefit equally from this research (116).

Replication is an essential component of genetic association studies, and the requirement for independent replication contributed to the success of GWASs (117, 118). However, replication and meta-analysis become challenging as G×E studies become sophisticated in analytical methods, exposure assessment, and incorporation of functional information. Differences in the underlying distribution of environmental exposures, patterns of linkage disequilibrium, and genetic modifiers can reduce the power to detect the same level of interaction in independent studies. Moreover, an appropriate human replication study might not (yet) exist in studies of a rare disease, genetic variant, or environmental exposure; where exposures are unique to particular populations; or where the initial finding was obtained within a large consortium comprising all known studies of a specific outcome (12). As illustrated in Ritz et al. (17), describing G×E successes and incorporation of biological knowledge, in some situations functional studies could serve to provide support for initial G×E observations in absence of a suitable replication population. Moreover, as the field considers gene- and pathway-based approaches to study G×E, replication may become further complicated as different combinations of genes in different data sets may be observed in the interaction. Some have argued that replication requirements might be met if the underlying biological pathway is the same even if replication was not observed with the individual single-nucleotide polymorphism or gene (15). More consideration is needed of standards for replication, definitions of replication, and alternative approaches for replication and verification of G×E results.

Many exciting opportunities exist for studies of G×E. There is the emerging recognition that developmental exposures may lead to disease throughout life, and efforts have focused on beginning to address how much of the environmental exposure risk for many disease outcomes may be attributable to in utero exposures or other particularly vulnerable windows of susceptibility (childhood, adolescence, etc.). Successful integration of large volumes of diverse data types (including data from geographic information systems, sensors, metabolomics, and other -omics) will create opportunities to generate unique insights. Epigenetics tools open up new opportunities to directly link environmental exposure to the genome and generate new exposure biomarkers (e.g., methylation of cancer-specific genes associated with dietary folate and alcohol in colorectal cancer (119) or smoking exposure in lung cancer (120), as discussed in reviews of opportunities in epigenetic epidemiology (121, 122)). Moreover, epigenomics, as well as other -omic technologies, may elucidate mechanisms by which exposures contribute to disease. The role of the microbiome as a key environmental risk factor for many complex disease phenotypes is starting to be appreciated and extensively studied. In addition, molecular phenotype data creates opportunities to examine disease subtypes or more precisely classify disease. This may eventually reduce heterogeneity in studies and improve the power to study G×E associations, assuming molecular characterizations are performed with the correct cell type, tissue, or appropriate surrogate tissue for the hypothesis being tested.

Additional areas of research may allow further advances in G×E discovery and replication. The field needs to determine how to best leverage experimental studies in animals or human cell lines to aid in discovering and functionally validating G×Es. Moreover, it is unclear how to best leverage existing family and twin-based studies for examining G×E. In incorporating functional information into G×E studies, questions remain about the appropriate balance between using prior or external information and the characteristics of the data set being studied when building analytical models and appropriate methods for linking environmental exposures information into available biological knowledge databases, which are usually focused on genes and pathways. In addition, because many G×E findings to date have modest effect sizes or have not been extensively replicated (11, 123), it is worthwhile to explore the general question of when to make the effort of attempting to identify these complex types of interactions. Even with modest effect sizes, if a G×E finding is sufficiently replicated in human populations and supported by other experimental data, this information could provide

insights into possible disease mechanisms. Finally, given the reduced power to detect G×E combinations with present methods, approaches that examine higher-order interactions should be taken on cautiously.

Despite many recent advances in analytical methods for G×E discovery, and some validation in recent years, additional statistical methods are needed for studies of copy number and rare genetic variation, survival traits, analysis of trios, and meta-analysis and pooling in large consortiums. In addition, many of the assumptions about expected G×E findings are based on results from genetic simulation studies, but these expectations have not always directly correlated to G×E observations in real population studies. Therefore, the question remains whether simulation studies have been designed with realistic assumptions about the underlying genetic architecture of the traits and whether better simulation approaches are needed (124).

Extended collaboration and data sharing will also advance G×E research. Large epidemiologic consortiums, such as the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, with longitudinal measures of environmental exposures have been heavily leveraged in recent years as a way to examine repeated environmental exposures over time and attempt to incorporate cumulative and time-varying exposures into assessments of complex disease risk (125). There is also a need for further collaboration to allow validation of biomarkers in larger cohorts. Meta-analysis and pooling methodology and efforts will likely need to be advanced to have the power to detect G×E in rarer diseases. Standards are needed to describe the adequate criteria for identifying, reproducing, and reporting a G×E finding; a place to publish negative findings would allow researchers to avoid repeating failed experiments (11). There is also a need for greater integration and education with other fields to better design studies of G×E. Specifically, toxicology expertise will be needed to allow validation in experimental models of G×E discoveries. Last, the sharing of environmental and epidemiologic data has lagged behind genomic data sharing. Some have suggested that an environmental data-sharing policy mirroring the National Institutes of Health genomic data-sharing policy could advance this effort in the environmental health-science fields. However, there are unique sensitivities and ethical issues related to the sharing of environmental data that must be considered, including participant confidentiality and privacy issues (because environmental exposure data with global-positioning-systems information can allow specific identification of the sources of exposure) and legal/regulatory matters (e.g., regulatory reporting, remediation, and reform).

Researchers are exploring ways to apply G×E findings to risk-prediction studies as a possibility for targeted screening or intervention. Questions remain about the optimal approaches for risk-prediction models, including how to integrate biomarkers and external exposures and how best to model the joint effects of genetic markers, biomarkers, and lifestyle and environmental exposures (126). Although most statistical methods for detecting G×E focus on identifying departures from a multiplicative relative-risk model, the absence of multiplicative interactions will typically imply the presence of additive interaction (i.e., when there are marginal genetic and environmental effects). Additive interactions may have public health implications because they suggest that the difference in absolute risks between exposed and unexposed groups differs across genetically defined subgroups (105–108, 126). If an exposure causes disease, then an intervention to remove the exposure will prevent more cases in a genetically sensitive population than an in an equivalently sized genetically insensitive population. Important challenges that remain include determining whether the exposure in fact causes disease, developing effective interventions to change exposures, and evaluating whether targeted or population-level interventions optimize the risk-benefit trade-off. As with main effects, where it is well understood that observational findings of associations across individuals do not necessarily imply that an intervention to change exposure will change any individual's outcome, so an additive interaction does not necessarily imply that a genetically targeted intervention would be a more effective strategy for prevention. Modern methods of causal inference (127, 128) may be useful for estimating the causal difference in disease rates between genetically targeted and population-wide exposure interventions. Finally, the lessons and approaches for research into how the combination of genes and environment contribute to disease relates broadly to the studies of precision medicine and precision prevention. These types of studies may lead to insights for targeting prevention, intervention, or treatment in the future.

ACKNOWLEDGMENTS

Author affiliations: Genes, Environment, and Health Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina (Kimberly McAllister); Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland (Leah E. Mechanic); Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire (Christopher Amos); Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Hugues Aschard, Peter Kraft); Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France (Hugues Aschard); Center of Excellence in Environmental Toxicology and Penn SRP Center, Perelman School of Medicine, University of Pennsylvania Philadelphia, Pennsylvania (Ian A. Blair); Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania Philadelphia, Pennsylvania (Ian A. Blair); Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Nilanjan Chatterjee); Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, Maryland (Nilanjan Chatterjee); Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California (David Conti, W. James Gauderman, Duncan C. Thomas); Biostatistics and Biomathematics Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington (Li Hsu); Division of Genome Sciences, National Human

Genome Research Institute, Bethesda, Maryland (Carolyn M. Hutter); California Institute for Telecommunications and Information Technology, Qualcomm Institute, University of California San Diego, La Jolla, California (Marta M. Jankowska); Department of Family Medicine and Public Health, School of Medicine, University of California San Diego, La Jolla, California (Jacqueline Kerr); Department of Genetics, Stanford University School of Medicine, Stanford, California (Stephen B. Montgomery); Department of Pathology, Stanford University School of Medicine, Stanford, California (Stephen B. Montgomery); Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan (Bhramar Mukherjee); Division of Cardiovascular Sciences, Prevention and Population Sciences Program, National Heart, Lung, and Blood Institute, Bethesda, Maryland (George J. Papanicolaou); Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts (Chirag J. Patel); Department of Biochemistry and Molecular Biology, Center for Systems Genomics, Pennsylvania State University, University Park, Pennsylvania (Marylyn D. Ritchie); Biomedical and Translational Informatics, Geisinger Health System, Danville, Pennsylvania (Marylyn D. Ritchie); Department of Epidemiology, Fielding School of Public Health, University of California Los Angeles, Los Angeles, California (Beate R. Ritz); Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas (Peng Wei); and Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California (John S. Witte).

The National Institute of Environmental Health Sciences provided funds to support "Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases." Research reported in this publication was supported by the National Cancer Institute (grants U19CA203654 and U01CA196386 to C.A.; R01CA140561, R01CA201407, and P01CA196569 to D.C.; P01CA196569 and R01CA201407 to W.J.G.; R01CA189532, R01CA195789, and P01CA53996 to L.H.; R21CA169535 and R01CA179977 to J.K.; R01CA169122 to P.W; R01CA169122; and R01CA201358 to J.S.W.), National Heart, Lung, and Blood Institute (grants R01HL116720 and R21HL126032 to P.W.), National Human Genome Research Institute (grant R21HG007687 to H.A.), the National Institute of Environmental Health Sciences (grants P30ES07048, and R21ES024844 to WJG; R21ES020811 to B.M.; R00ES023504 and R21ES025052 to C.J.P.) of the National Institutes of Health, and the National Science Foundation (grant NSF DMS 1406712 to B.M.). S.B.M. is supported by the National Institutes of Health (grants R01HG008150, R01MH101814, U01HG007436, and U01HG009080). This work is funded, in part, under a grant with the Pennsylvania Department of Health (grant SAP 4100070267) to M.D.R.

We thank the participants in the workshop "Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases."

The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

1. Hindorff LA, Gillanders EM, Manolio TA. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. *Carcinogenesis*. 2011;32(7):945–954.
2. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106(23):9362–9367.
3. Stadler ZK, Thom P, Robson ME, et al. Genome-wide association studies of cancer. *J Clin Oncol*. 2010;28(27): 4255–4267.
4. Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet*. 2005;6(4):287–298.
5. Thomas D. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010;11(4):259–272.
6. Thompson WD. Effect modification and the limits of biological inference from epidemiologic data. *J Clin Epidemiol*. 1991;44(3):221–232.
7. Le Marchand L, Wilkens LR. Design considerations for genomic association studies: importance of gene-environment interactions. *Cancer Epidemiol Biomarkers Prev*. 2008;17(2): 263–267.
8. Boffetta P, Winn DM, Ioannidis JP, et al. Recommendations and proposed guidelines for assessing the cumulative evidence on joint effects of genes and environments on cancer occurrence in humans. *Int J Epidemiol*. 2012;41(3):686–704.
9. Kraft P, Yen YC, Stram DO, et al. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007;63(2):111–119.
10. Bookman EB, McAllister K, Gillanders E, et al. Gene-environment interplay in common complex diseases: forging an integrative model-recommendations from an NIH workshop. *Genet Epidemiol*. 2011;35(4):217–225.
11. Hutter CM, Mechanic LE, Chatterjee N, et al. Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genet Epidemiol*. 2013; 37(7):643–657.
12. Mechanic LE, Chen HS, Amos CI, et al. Next generation analytic tools for large scale genetic epidemiology studies of complex diseases. *Genet Epidemiol*. 2012;36(1):22–35.
13. Kraft P, Aschard H. Finding the missing gene-environment interactions. *Eur J Epidemiol*. 2015;30(5):353–355.
14. Gauderman WJ, Mukheerjee B, Aschard H, et al. Update on the state of the science for analytical methods for gene-environment interactions. *Am J Epidemiol*. 2017;186(7): 762–770.
15. Ritchie MD, Davis JR, Aschard H, et al. Incorporation of biological knowledge into the study of gene-environment interactions. *Am J Epidemiol*. 2017;186(7):771–777.
16. Patel CJ, Kerr J, Thomas DC, et al. Opportunities and challenges for environmental exposure assessment in population-based studies [published online ahead of print July 14, 2017]. *Cancer Epidemiol Biomarkers Prev*. (doi:10.1016/j.cmpb.2003.08.003).
17. Ritz BR, Chatterjee N, Garcia-Closas M, et al. Lessons learned from past gene-environment interaction successes. *Am J Epidemiol*. 2017;186(7):778–786.
18. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372(9):793–795.

19. Khoury MJ, Gwinn ML, Glasgow RE, et al. A population approach to precision medicine. *Am J Prev Med*. 2012;42(6):639–645.
20. Aschard H. A perspective on interaction effects in genetic association studies. *Genet Epidemiol*. 2016;40(8):678–688.
21. Murcray CE, Lewinger JP, Conti DV, et al. Sample size requirements to detect gene-environment interactions in genome-wide association studies. *Genet Epidemiol*. 2011;35(3):201–210.
22. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med*. 1994;13(2):153–162.
23. Mukherjee B, Chatterjee N. Exploiting gene-environment independence for analysis of case-control studies: an empirical Bayes-type shrinkage estimator to trade-off between bias and efficiency. *Biometrics*. 2008;64(3):685–694.
24. Li D, Conti DV. Detecting gene-environment interactions using a combined case-only and case-control approach. *Am J Epidemiol*. 2009;169(4):497–504.
25. Dai JY, Logsdon BA, Huang Y, et al. Simultaneously testing for marginal genetic association and gene-environment interaction. *Am J Epidemiol*. 2012;176(2):164–173.
26. Han SS, Rosenberg PS, Ghosh A, et al. An exposure-weighted score test for genetic associations integrating environmental risk factors. *Biometrics*. 2015;71(3):596–605.
27. Kistner EO, Shi M, Weinberg CR. Using cases and parents to study multiplicative gene-by-environment interaction. *Am J Epidemiol*. 2009;170(3):393–400.
28. Umbach DM, Weinberg CR. Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med*. 1997;16(15):1731–1743.
29. Weinberg CR, Umbach DM. A hybrid design for studying genetic influences on risk of diseases with onset early in life. *Am J Hum Genet*. 2005;77(4):627–636.
30. Dai JY, Kooperberg C, Leblanc M, et al. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*. 2012;99(4):929–944.
31. Gauderman WJ, Zhang P, Morrison JL, et al. Finding novel genes by testing G × E interactions in a genome-wide association study. *Genet Epidemiol*. 2013;37(6):603–613.
32. Hsu L, Shuo J, Dai Y, et al. Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genet Epidemiol*. 2012;36(3):183–194.
33. Kooperberg C, Leblanc M. Increasing the power of identifying gene × gene interactions in genome-wide association studies. *Genet Epidemiol*. 2008;32(3):255–263.
34. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*. 2009;169(2):219–226.
35. Gauderman WJ, Thomas DC, Murcray CE, et al. Efficient genome-wide association testing of gene-environment interaction in case-parent trios. *Am J Epidemiol*. 2010;172(1):116–122.
36. Chen H, Meigs JB, Dupuis J. Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered*. 2014;78(2):81–90.
37. Jiao S, Hsu L, Bézieau S, et al. SBERIA: set-based gene-environment interaction test for rare and common variants in complex diseases. *Genet Epidemiol*. 2013;37(5):452–464.
38. Lin X, Lee S, Christiani DC, et al. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics*. 2013;14(4):667–681.
39. Lin X, Lee S, Wu MC, et al. Test for rare variants by environment interactions in sequencing association studies. *Biometrics*. 2016;72(1):156–164.
40. Tzeng JY, Zhang D, Pongpanich M, et al. Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. *Am J Hum Genet*. 2011;89(2):277–288.
41. Aschard H, Zaitlen N, Tamimi RM, et al. A nonparametric test to detect quantitative trait loci where the phenotypic distribution differs by genotypes. *Genet Epidemiol*. 2013;37(4):323–333.
42. Brown AA, Buil A, Viñuela A, et al. Genetic interactions affecting human gene expression identified by variance association mapping. *Elife*. 2014;3:e01381.
43. Levene H. Robust tests for equality of variances. In: Olkin I, ed. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Palo Alto, CA: Stanford University Press; 1960:278–292.
44. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984;40(4):1079–1087.
45. Paré G, Cook NR, Ridker PM, et al. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet*. 2010;6(6):e1000981.
46. Wang G, Yang E, Brinkmeyer-Langford CL, et al. Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. *Genetics*. 2014;196(2):413–425.
47. Yang J, Loos RJ, Powell JE, et al. *FTO* genotype is associated with phenotypic variability of body mass index. *Nature*. 2012;490(7419):267–272.
48. Zhang P, Lewinger JP, Conti D, et al. Detecting gene-environment interactions for a Quantitative Trait in a Genome-Wide Association Study. *Genet Epidemiol*. 2016;40(5):394–403.
49. Bhattacharjee S, Chatterjee N, Han S, et al. *An R Package for Analysis of Case-Control Studies in Genetic Epidemiology*. R package version 3.10.0. Bethesda, MD; 2012. http://bioconductor.org/packages/release/bioc/html/CGEN.html.
50. Su YR, Di C, Hsu L, et al. A unified powerful set-based test for sequencing data analysis of GxE interactions. *Biostatistics*. 2017;18(1):119–131.
51. Boonstra PS, Mukherjee B, Gruber SB, et al. Tests for gene-environment interactions and joint effects with exposure misclassification. *Am J Epidemiol*. 2016;183(3):237–247.
52. Cornelis MC, Tchetgen EJ, Liang L, et al. Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. *Am J Epidemiol*. 2012;175(3):191–202.
53. Mukherjee B, Ahn J, Gruber SB, et al. Testing gene-environment interaction in large-scale case-control association studies: possible choices and comparisons. *Am J Epidemiol*. 2012;175(3):177–190.
54. Thomas DC, Lewinger JP, Murcray CE, et al. Invited commentary: GE-Whiz! Ratcheting gene-environment studies up to the whole genome and the whole exposome. *Am J Epidemiol*. 2012;175(3):203–207.
55. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5):640–648.
56. Patel CJ, Burford B, Ioannidis JP. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J Clin Epidemiol*. 2015;68(9):1046–1058.

57. Freedman ML, Monteiro AN, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*. 2011;43(6):513–518.
58. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190–1195.
59. John S, Sabo PJ, Thurman RE, et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet*. 2011;43(3):264–268.
60. Nicolae DL, Gamazon E, Zhang W, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6(4):e1000888.
61. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790–1797.
62. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473(7345):43–49.
63. Guo Y, Conti DV, Wang K. Enlight: web-based integration of GWAS results with biological annotations. *Bioinformatics*. 2015;31(2):275–276.
64. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40(Database issue):D930–D934.
65. Yao L, Tak YG, Berman BP, et al. Functional annotation of colon cancer risk SNPs. *Nat Commun*. 2014;5:5114.
66. Barreiro LB, Tailleux L, Pai AA, et al. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc Natl Acad Sci USA*. 2012;109(4):1204–1209.
67. Fairfax BP, Humburg P, Makino S, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*. 2014;343(6175): 1246949.
68. Grundberg E, Adoue V, Kwan T, et al. Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet*. 2011;7(1):e1001279.
69. Qiu W, Rogers AJ, Damask A, et al. Pharmacogenomics: novel loci identification via integrating gene differential analysis and eQTL analysis. *Hum Mol Genet*. 2014;23(18): 5017–5024.
70. Wei P, Yang Y, Guo X, et al. Identification of an association of TNFAIP3 polymorphisms with matrix metalloproteinase expression in fibroblasts in an integrative study of systemic sclerosis-associated genetic and environmental factors. *Arthritis Rheumatol*. 2016;68(3): 749–760.
71. French JE, Gatti DM, Morgan DL, et al. Diversity outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity. *Environ Health Perspect*. 2015;123(3):237–245.
72. Rasmussen AL, Okumura A, Ferris MT, et al. Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science*. 2014;346(6212):987–991.
73. Ritchie MD, Holzinger ER, Li R, et al. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet*. 2015;16(2):85–97.
74. Baurley JW, Conti DV. A scalable, knowledge-based analysis framework for genetic association studies. *BMC Bioinformatics*. 2013;14:312.
75. Quintana MA, Conti DV. Integrative variable selection via Bayesian model uncertainty. *Stat Med*. 2013;32(28): 4938–4953.
76. Quintana MA, Schumacher FR, Casey G, et al. Incorporating prior biologic information for high-dimensional rare variant association studies. *Hum Hered*. 2012;74(3–4):184–195.
77. Baurley JW, Conti DV, Gauderman WJ, et al. Discovery of complex pathways from observational data. *Stat Med*. 2010; 29(19):1998–2011.
78. Pendergrass SA, Frase A, Wallace J, et al. Genomic analyses with biofilter 2.0: knowledge driven filtering, annotation, and model development. *BioData Min*. 2013;6(1):25.
79. Sun X, Lu Q, Mukherjee S, et al. Analysis pipeline for the epistasis search—statistical versus biological filtering. *Front Genet*. 2014;5:106.
80. Bush WS, Dudek SM, Ritchie MD, et al. Biofilter: a knowledge-integration system for the multi-locus analysis of genome-wide association studies. *Pac Symp Biocomput*. 2019:368–379.
81. Davis AP, Grondin CJ, Lennon-Hopkins K, et al. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res*. 2015; 43(Database issue):D914–D920.
82. Audouze K, Brunak S, Grandjean P. A computational approach to chemical etiologies of diabetes. *Sci Rep*. 2013;3: 2712.
83. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*. 2005;14(8):1847–1850.
84. Wild CP. The exposome: from concept to utility. *Int J Epidemiol*. 2012;41(1):24–32.
85. Cui Y, Balshaw DM, Kwok R, et al. The exposome: embracing the complexity for discovery in environmental health. *Environ Health Perspect*. 2016;124(8):A137–A140.
86. Dennis KK, Auerbach SS, Balshaw DM, et al. The importance of the biological impact of exposure to the concept of the exposome. *Environ Health Perspect*. 2016; 124(10):1504–1510.
87. Dennis KK, Marder E, Balshaw DM, et al. Biomonitoring in the era of the exposome. *Environ Health Perspect*. 2016; 125(4):502–510.
88. Turner MC, Nieuwenhuijsen M, Anderson K, et al. Assessing the exposome with external measures: commentary on the state of the science and research recommendations. *Annu Rev Public Health*. 2017;38:215–239.
89. Gibson G. The environmental contribution to gene expression profiles. *Nat Rev Genet*. 2008;9(8):575–581.
90. van Breda SG, Wilms LC, Gaj S, et al. The exposome concept in a human nutrigenomics study: evaluating the impact of exposure to a complex mixture of phytochemicals using transcriptomics signatures. *Mutagenesis*. 2015;30(6): 723–731.
91. Shaw JG, Vaughan A, Dent AG, et al. Biomarkers of progression of chronic obstructive pulmonary disease (COPD). *J Thorac Dis*. 2014;6(11):1532–1547.
92. Alexander N, Wankerl M, Hennig J, et al. DNA methylation profiles within the serotonin transporter gene moderate the association of 5-*HTTLPR* and cortisol stress reactivity. *Transl Psychiatry*. 2014;4:e443.
93. Patel CJ, Chen R, Kodama K, et al. Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. *Hum Genet*. 2013; 132(5):495–508.
94. Patel CJ, Bhattacharya J, Butte AJ. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One*. 2010;5(5):e10746.

95. Hall MA, Dudek SM, Goodloe R, et al. Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank. *Pac Symp Biocomput*. 2014:200–211.
96. McGinnis DP, Brownstein JS, Patel CJ. Environment-wide association study of blood pressure in the National Health and Nutrition Examination Survey (1999–2012). *Sci Rep*. 2016;6: 30373.
97. Ahn J, Mukherjee B, Gruber SB, et al. Bayesian semiparametric analysis for two-phase studies of gene-environment interaction. *Ann Appl Stat*. 2013;7(1):543–569.
98. Breslow NE, Chatterjee N. Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *J R Stat Soc Ser C Appl Stat*. 1999;48(4):457–468.
99. Chatterjee N, Chen YH. Maximum likelihood inference on a mixed conditionally and marginally specified regression model for genetic epidemiologic studies with two-phase sampling. *J R Stat Soc Series B Stat Methodol*. 2007;69(2): 123–142.
100. Wacholder S, Weinberg CR. Flexible maximum likelihood methods for assessing joint effects in case-control studies with complex sampling. *Biometrics*. 1994;50(2):350–357.
101. Stenzel SL, Ahn J, Boonstra PS, et al. The impact of exposure-biased sampling designs on detection of gene-environment interactions in case-control studies with potential exposure misclassification. *Eur J Epidemiol*. 2015; 30(5):413–423.
102. Wei P, Tang H, Li D. Functional logistic regression approach to detecting gene by longitudinal environmental exposure interaction in a case-control study. *Genet Epidemiol*. 2014; 38(7):638–651.
103. Shi M, Umbach DM, Weinberg CR. Family-based gene-by-environment interaction studies: revelations and remedies. *Epidemiology*. 2011;22(3):400–407.
104. Dodge HH, Zhu J, Mattek NC, et al. Use of high-frequency in-home monitoring data may reduce sample sizes needed in clinical trials. *PLoS One*. 2015;10(9):e0138095.
105. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst*. 2014;106(11):dju305.
106. Garcia-Closas M, Rothman N, Figueroa JD, et al. Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer Res*. 2013;73(7): 2211–2220.
107. Joshi AD, Lindstrom S, Husing A, et al. Additive interactions between susceptibility single-nucleotide polymorphisms identified in genome-wide association studies and breast cancer risk factors in the Breast and Prostate Cancer Cohort Consortium. *Am J Epidemiol*. 2014;180(10):1018–1027.
108. Maas P, Barrdahl M, Joshi AD, et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol*. 2016;2(10): 1295–1302.
109. Franceschini N, van Rooij FJ, Prins BP, et al. Discovery and fine mapping of serum protein loci through transethnic meta-analysis. *Am J Hum Genet*. 2012;91(4):744–753.
110. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
111. Liu CT, Buchkovich ML, Winkler TW, et al. Multi-ethnic fine-mapping of 14 central adiposity loci. *Hum Mol Genet*. 2014;23(17):4738–4744.
112. Wu Y, Waite LL, Jackson AU, et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet*. 2013;9(3):e1003379.
113. Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. *Nat Rev Genet*. 2011;12(8): 523–528.
114. Pino-Yanes M, Gignoux CR, Galanter JM, et al. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. *J Allergy Clin Immunol*. 2015;135(6):1502–1510.
115. Galanter JM, Gignoux CR, Torgerson DG, et al. Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-Environments and Admixture in Latino Americans study. *J Allergy Clin Immunol*. 2014;134(2):295–305.
116. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature*. 2011;475(7355):163–165.
117. NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, et al. Replicating genotype-phenotype associations. *Nature*. 2007;447(7145):655–660.
118. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. *Stat Sci*. 2009;24(4):561–573.
119. van Engeland M, Weijenberg MP, Roemen GM, et al. Effects of dietary folate and alcohol intake on promoter methylation in sporadic colorectal cancer: the Netherlands Cohort Study on diet and cancer. *Cancer Res*. 2003;63(12):3133–3137.
120. Zochbauer-Muller S, Lam S, Toyooka S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer*. 2003;107(4): 612–616.
121. Cortessis VK, Thomas DC, Levine AJ, et al. Environmental epigenetics: prospects for studying epigenetic mediation of exposure–response relationships. *Hum Genet*. 2012;131(10): 1565–1589.
122. Bakulski KM, Fallin MD. Epigenetic epidemiology: promises for public health research. *Environ Mol Mutagen*. 2014;55(3): 171–183.
123. Simonds NI, Ghazarian AA, Pimentel CB, et al. Review of the gene-environment interaction literature in cancer: what do we know? *Genet Epidemiol*. 2016;40(5):356–365.
124. Chen HS, Hutter CM, Mechanic LE, et al. Genetic simulation tools for post-genome wide association studies of complex diseases. *Genet Epidemiol*. 2015;39(1):11–19.
125. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2(1):73–80.
126. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*. 2016;17(7):392–406.
127. VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. *Epidemiology*. 2007;18(3):329–339.
128. VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology*. 2009;20(1):6–13.