American Journal of Epidemiology
Vol. 186, No. 7
DOI: 10.1093/aje/kwx228
Advance Access publication:
July 15, 2017

Special Article

Update on the State of the Science for Analytical Methods for Gene-Environment Interactions

W. James Gauderman*, Bhramar Mukherjee, Hugues Aschard, Li Hsu, Juan Pablo Lewinger, Chirag J. Patel, John S. Witte, Christopher Amos, Caroline G. Tai, David Conti, Dara G. Torgerson, Seunggeun Lee, and Nilanjan Chatterjee

* Correspondence to Dr. W. James Gauderman, Division of Biostatistics, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, 2001 North Soto Street, 202-K, Los Angeles, CA 90032 (e-mail: jimg@usc.edu).

Initially submitted November 7, 2016; accepted for publication April 25, 2017.

The analysis of gene-environment interaction (G×E) may hold the key for further understanding the etiology of many complex traits. The current availability of high-volume genetic data, the wide range in types of environmental data that can be measured, and the formation of consortiums of multiple studies provide new opportunities to identify G×E but also new analytical challenges. In this article, we summarize several statistical approaches that can be used to test for G×E in a genome-wide association study. These include traditional models of G×E in a case-control or quantitative trait study as well as alternative approaches that can provide substantially greater power. The latest methods for analyzing G×E with gene sets and with data in a consortium setting are summarized, as are issues that arise due to the complexity of environmental data. We provide some speculation on why detecting G×E in a genome-wide association study has thus far been difficult. We conclude with a description of software programs that can be used to implement most of the methods described in the paper.

exposure; gene-environment interaction; GWAS; power; software; statistical models

Abbreviations: BMI, body mass index; BT, burden-type; df, degree(s) of freedom; EB, empirical Bayes; G-E, gene-environment; GWAS, genome-wide association study; G×E, gene-environment interaction; SNP, single nucleotide polymorphism; VC, variance component.

Gene-environment interaction (G×E) can be defined broadly as the interplay between a gene and an environmental factor as they affect some trait. For example, epidemiologists may be interested in studying how genetic susceptibility might predispose subgroups of the population to enhanced effects of an environmental exposure. Alternatively, geneticists may be interested in studying how exposure to an environmental fact may stimulate the expression of a gene and lead to disease.

In the first article in this series, McAllister et al. (1) provide a broad overview of the state of the science related to G×E. In this article, we summarize the latest statistical methods available for analysis of G×E. We address many of the practical questions that statisticians face as they try to uncover G×Es for complex traits, and we summarize some of the latest approaches to address these issues (Table 1).

MODELS OF G×E

Basic models

Consider a study of a disease outcome (D), environmental factor (E), and genetic factor (G), with data also collected on a set of potential confounders (**C**). Exposure E can represent an exogenous environmental variable (e.g., air pollution), personal exposure (e.g., smoking), or other personal characteristic (e.g., sex). Although a single genetic locus may be of interest, most studies now genotype a large number (e.g., 1 million) of single nucleotide polymorphisms (SNPs) on each study subject. Each locus may be coded as G = 0, 1, or 2 for the number of minor alleles, or dominant, recessive, or codominant coding can be used. Additional untyped SNP genotypes are now routinely imputed

Table 1. Challenges and Potential Solutions in the Analysis of Gene-Environment Interactions, Based on Issues Raised at the Workshop "Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases," United States, October 2014

Challenges	Old Approach	Solutions/New Approach	Section Heading
Interaction can be dependent on scale	Only multiplicative scale considered	Consider evaluating interaction on both additive and multiplicative scales	Models of G×E
SNP-based analyses can lack power	Single-step analysis subject to multiple comparisons burden due to large number of SNPs considered at once	Conduct more efficient 2-step tests	Detecting interactions in a GWAS
	Single-variant approach agnostic to biological information	Conduct gene-based/set-based tests	G×E with gene sets
	Individual studies report results independently	Conduct meta-analysis across studies/cohorts	G×E analysis in a consortium setting
	Only homogeneous populations considered, typically of European decent	Consider admixture analysis, if appropriate	G×E analysis in a consortium setting
Exposure measurement can be inconsistent and imperfect	Individual studies independently determine method of exposure measurement	Work towards common core of exposures and definitions	The complexity of exposure
	Employ easiest measurement method for largest study sample possible	Prioritize improving precision of measurements	Why haven't many G×Es been identified?
Software is not available to conduct efficient G×E analysis	Individual analysts tweak existing software to generate limited G×E results	Implement new software designed for high-volume G×E analyses using novel methods	Available software for analysis of G×E

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

based on the 1,000 Genomes (2) or Haplotype Reference Consortium (3) panels.

For a case-control study, logistic regression is typically used to model G×E, with the form (model 1):

$$\text{Logit}(\Pr(D = 1 \mid G, E, \mathbf{C}) = \beta_0 + \beta_G G + \beta_E E + \beta_{G\times E} G\times E + \boldsymbol{\beta}_{\mathbf{C}}\mathbf{C}. \quad (1)$$

Compared to an unexposed noncarrier (E = 0, G = 0), $OR_G = \exp(\beta_G)$ measures the "main effect" of G (E = 0, G = 1) and $OR_E = \exp(\beta_E)$ measures the main effect of E (E = 1, G = 0). The corresponding odds ratio when E = 1 and G = 1 is $\exp(\beta_G)\exp(\beta_E)\exp(\beta_{G\times E})$, and so the "interaction odds ratio" is $OR_{G\times E} = \exp(\beta_{G\times E})$ and measures the departure from the multiplicative effects of the corresponding main effects. Note that the choice of coding system (e.g., E = 1 for obese, E = 0 for nonobese) is arbitrary and has no impact on the significance of the interaction test. However, the coding system is important for interpretation of β_G, β_E, and $\beta_{G\times E}$, in that the magnitude and/or direction of these effects can change when shifting the coding of G and/or E (4).

The G×E model (model 1) builds on the typical model (model 2) that does not consider G×E:

$$\text{Logit}(\Pr(D = 1 \mid G, \mathbf{C})) = \mu_0 + \mu_G G + \boldsymbol{\mu}_{\mathbf{C}}\mathbf{C}, \quad (2)$$

where E may be included in **C**. Here $OR_G = \exp(\mu_G)$ measures the "marginal odds ratio" of G, interpreted as averaging (or marginalizing) over the exposure-specific effects of G.

For the simplest situation of a binary exposure, binary G, and no covariates, Web Figure 1 (available at https://academic.oup.com/aje) shows how cell counts from two 2 × 2 tables can be used to compute the interaction and marginal effects described above. In the context of a cohort study, one could replace models 1 and 2 with log-linear models to estimate relative risks RR_G, RR_E, and $RR_{G\times E}$ (5), or with proportional hazards models to estimate hazard rate ratios for time-to-disease data (6). For a quantitative outcome, linear regression is typically used (see below).

Interpretations of interaction parameters

Interpretation of G×E depends on the underlying scale on which G×E effects are modeled. The classical result of Prentice and Pyke (7) ensures that estimates of (β_G, β_E, $\beta_{G\times E}$, $\beta_{\mathbf{C}}$) obtained from model 1 are valid and efficient under case-control sampling. Most commonly, interaction in epidemiology refers to the departure from multiplicative effects described above.

Another form of interaction that is not commonly assessed is the departure from additivity on the absolute risk scale. The additive effect is defined as $G\times E_{ADD} = RR_G \times RR_E \times RR_{G\times E} - RR_G - RR_E + 1$ for a relative risk model, with analogous form for an odds ratio model. Departure from additivity implies that absolute risk reduction associated with removal of one risk factor depends on the levels of another and vice versa. As such, the model has direct relevance for evaluation of public health impact of risk-factor intervention (8). Furthermore, many mechanistic forms of interactions, such as under the sufficient-component cause model (8, 9) and various modern extensions (10), have been shown to yield superadditive effects ($G\times E_{ADD} > 0$).

It is useful to understand the relationships between different models for interactions (Figure 1). A supermultiplicative G×E ($RR_{G\times E} > 1$) automatically implies the effects of G and E are superadditive. Conversely, a superadditive G×E can correspond to either sub- or super-multiplicative effects of

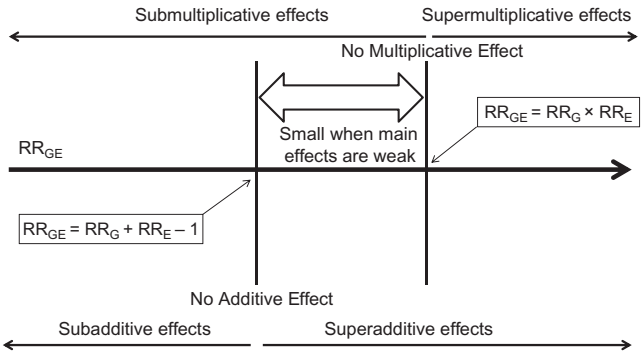

Figure 1. Comparison of additive and multiplicative models for the joint effects of gene (G) and environment (E). Assuming binary G and binary E, the joint-effect relative risk $RR_{GE} = RR_G \times RR_E \times RR_{G\times E}$—relative risk for G = 1 and E = 1 compared with G = 0 and E = 0—is shown under the additive and multiplicative models. Under the 2 different models, the RR_{GE} is determined by underlying main-effect relative risks (RR_G and RR_E), that is, relative-risk associated with one factor (e.g., G = 1 vs G = 0) while the other factor is fixed at baseline (e.g., E = 0) in 2 different functional forms. Throughout, it is assumed that G = 1 and E = 1 correspond to higher-risk categories than G = 0 and E = 0 ($RR_G > 1$ and $RR_E > 1$).

G and E. In absence of a main effect of G and/or E, the additive and multiplicative models coincide. Thus for G and/or E with weak main effects, which is often the case for common SNPs, models may be hard to distinguish without large sample size. Recent studies conducted to explore genome-wide association study (GWAS) results suggest that the multiplicative model often provides reasonable approximation of G×E joint effects on disease risks (11–14). However, although the multiplicative model may be the "accepted" analysis approach, superadditive effects have been reported (15), demonstrating that investigation of interactions both scales can be informative for understanding joint G and E effects.

Mechanistic interpretation of statistical interaction is difficult because of its dependence on scale. However, certain forms of scale-invariant interactions can provide important mechanistic insights. In a "pure interaction" model when the effect of G is present only in the presence/absence of E (corresponding to $\beta_G = \beta_E = 0$ in model 1), interaction would be evident irrespective of scale. Genetic markers of acetylation in the N-acetyltransferase 2 gene (*NAT2*) have been associated with bladder cancer only among smokers (16). Another form of invariant interaction is the qualitative interaction where the presence of one risk factor may reverse the effect of another.

DETECTING INTERACTIONS IN A GWAS

Basic tests: G×E and the 2-df test

In context of model 1, detection of multiplicative interaction is based simply on the test of the null hypothesis that $\beta_{G\times E} = 0$, and a Wald, Score, or likelihood ratio test may be applied. The same type of test is used if conditional logistic, log-linear, or Cox regression are used for analysis. Standard GWAS screens for marginal G effects are based on the test of H_0: $\mu_G = 0$ from model 2. A 2-degrees-of-freedom (df) procedure based on model 1 (17) or a combination of models 1 and 2 (18) can be used to test the joint null H_0: $\beta_G = \beta_{G\times E} = 0$. For many models, 2-df tests have better power to detect genes than either the marginal G or 1-df G×E test alone (17).

Regardless of the specific test, one should carefully consider potential confounders **C**, G×C, and E×C interactions, because G×E effects can themselves be confounded by other interactions (19). Some potential confounders seem like obvious choices for inclusion in the model—for example, principal components to adjust for ancestry. Others require more judgment, such as whether or not to adjust for body mass index (BMI) in a study of gene-diet interaction. While BMI may seem like an obvious confounder, it may also be a mediator of dietary effects, and including BMI, G × BMI, and diet × BMI as adjustments may reduce the ability to detect G × diet signals.

In a GWAS, both G and E are each typically coded using a single trend variable, yielding a 1-df test of interaction. However, the effect of E can often be specified as categorical, ordinal, or continuous form depending on the nature of the underlying measurements. It may be desirable to code a complex exposure using a flexible model to avoid bias in the test for interaction due to model misspecification (20). This will translate to a multi-df test of interaction, which can reduce power and should be avoided if a single trend variable can be justified.

Testing for additive interaction can be numerically complex due to constraints required to guarantee risk estimates are bounded between 0 and 1 for all risk-factor combinations. For rare diseases and categorical risk factors, the additive model can be specified in the form of a general logistic regression model where the interaction term is specified by a nonlinear function of the main effects (21).

Case-only and empirical Bayes approaches

Piegorsch et al. (22) showed that for binary (G, E), the parameter $\beta_{G\times E}$ in model 1 can be estimated by using data from cases only (Web Figure 2). To allow for nonbinary G and/or E, and adjustment for confounders, one can adopt a polytomous logistic model for case-only analysis, as in model 3:

$$\text{Logit } P(G = g \mid E, \mathbf{C}, D = 1) = \gamma_g + \gamma_{gE}E + \boldsymbol{\gamma}_{\mathbf{gc}}^{\mathbf{T}}\mathbf{C} \quad g = 1, 2 \tag{3}$$

where each parameter quantifies the effect of a heterozygous (g = 1) or homozygous (g = 2) carrier of a risk allele compared with a homozygous noncarrier (g = 0). Under the assumption of gene-environment (G-E) independence conditional on **C**, the test of H_0: $\gamma_{1E} = \gamma_{2E} = 0$ is a valid test of multiplicative interaction. One often desires a test of trend across G = 0, 1, and 2, which can be accomplished in model 3 by setting $\gamma_{1E} = \gamma_E$ and $\gamma_{2E} = 2\gamma_E$ and testing H_0: $\gamma_E = 0$. A limitation of case-only analysis is that one cannot estimate main effects β_G and β_E, and thus cannot retrieve the subgroup effects of genotype across exposure strata (e.g., $OR_{D|E=1}$ and $OR_{D|E=0}$). Alternative "case-only" approaches that use controls to estimate main effects have been proposed (23, 24).

If the G-E independence assumption is violated, estimates derived from case-only analysis will be biased (25), either toward or away from the null depending on directions of the

G-E association and G×E. To reduce bias but retain some efficiency of the G-E independence assumption, an empirical Bayes (EB) strategy has been proposed (26). In the simple case of a binary G and E, the EB estimator $\hat{\beta}_{EB}$ is constructed using estimates from models 1 and 3:

$$\hat{\beta}_{EB} = \hat{\gamma}_{gE} + K(\hat{\beta}_{GxE} - \hat{\gamma}_{gE})$$
$$\text{where } K = \hat{\theta}^2_{gE}/(\hat{\sigma}^2_{G\times E} + \hat{\theta}^2_{gE}) \quad (4)$$

The intuitive explanation behind the EB estimator is that if there is G-E independence in the population (i.e., $\theta_{gE} = 0$), γ_{gE} and $\beta_{G\times E}$ will be approximately equal, and one should favor the case-only estimate $\hat{\gamma}_{gE}$ for its increased efficiency. On the other hand, if the data provide evidence of G-E dependence ($\hat{\theta}_{gE} > 0$) or if the variance of $\hat{\beta}_{G\times E}$ is small ($\hat{\sigma}^2_{G\times E} \to 0$), larger weight is assigned to $\hat{\beta}_{G\times E}$.

Chen et al. (27) provided a more general EB framework, which is implemented in the CGEN software package (28). Li and Conti (29) developed a Bayes model averaging framework, where the weights in Equation 4 are defined by posterior probabilities. Both EB and Bayes model averaging approaches often provide greater power than a standard case-control test, while providing improved (although not perfect) control of type I error compared with a case-only test in the presence of population G-E association (26) (Web Appendix 1).

Efficient 2-step tests

Several "2-step" approaches have been proposed to improve the efficiency of G×E analysis while controlling type I error, both for disease (30–35) and quantitative (36, 37) traits. All of these methods use the following general approach:

Step 1 screen: For all M (e.g., 1 million) SNPs, compute screening test statistic T_1 and corresponding P value p_1.

Step 2 test: Prioritize SNPs based on p_1 (e.g., conduct step 2 only on m SNPs with $p_1 < 0.05$), and compute G×E test statistic T_2 with corresponding P value p_2. Power is increased by the need to adjust in step 2 for only $m < M$ tests.

A key requirement for validity of any 2-step procedure is that T_1 and T_2 must be independent. In a case-control study, two types of step-1 screening tests have been proposed: 1) test of marginal D versus G (DG) association (33), based on model 2; and 2) test of E versus G (EG) association (35), based on model 3 applied to the combined case-control sample. The step-2 test is based on $\beta_{G\times E}$ from model 1. It has been shown that tests of $\mu_G = 0$ and $\gamma_{gE} = 0$ (in the combined sample) are independent of the test of $\beta_{G\times E} = 0$, and so either is a valid screening statistic (38). In the presence of G×E, one can typically expect nonzero values of both μ_G and γ_{gE}, making either the DG or EG screening statistic useful for identifying those SNPs that are most likely to be involved in a G×E. Three additional 2-step methods have been proposed including H2 (34), Cocktail (32), and EDGE (31), each of which use the DG and EG screening statistics in combination to further improve efficiency. Both standard tests and 2-step methods are implemented in the GxEScan software program (39).

G×E analysis for quantitative measures

Quantitative trait analyses in plant and animal models have clearly identified the importance of G×Es, in some cases having profound impact on phenotypes such as longevity in *Drosphila* (40) or flowering time in *Arabiaopsis* (41). In humans, Winkler et al. (42) identified 15 loci showing evidence for age-dependent genetic effects on BMI, 4 of which were not identified previously. A recent analysis suggests that estimates of heritability for quantitative traits can be substantially underestimated when interaction effects are not modeled (43). For a quantitative outcome Y, a linear model of the form of model 5:

$$Y = \beta_0 + \beta_G G + \beta_E E + \beta_{G\times E} G\times E + \boldsymbol{\beta_C C} + \varepsilon \quad (5)$$

is often adopted. As for any such regression model, failure to satisfy basic model assumptions (e.g., linearity, normality of residuals ε) can lead to inflated type I error or reduced power. As for a disease trait, G×E can induce a marginal G effect, here a difference in mean Y across G. This information can be used efficiently in a 2-df joint test of G and G×E (17, 44–46), or to construct a 2-step procedure that screens on marginal G association (37). It has also been shown that G×E induces a difference in the variance of Y across G (36, 37, 47). This variance-heterogeneity information can also be used to develop valid testing procedures that are more powerful than standard tests of G×E or marginal G effects (37, 47). Standard and 2-step testing procedures are implemented in GxEScan (39).

Which analysis should you choose for GWAS?

There have obviously been many statistical approaches developed for testing G×E. Perhaps the most natural is the standard test of $\beta_{G\times E}$ from model 1 or model 5, as this is a simple extension of the kind of model used for marginal G analysis. However, case-only, EB, 2-df, and 2-step procedures can offer improvements in efficiency and should be considered. As an example, Figure 2 shows that the case-only, 2-df, and 2-step EDGE approaches can require substantially lower sample sizes to achieve 80% power for detecting G×E compared with the standard (model 1) G×E test. While this example is representative, the efficiency of various approaches relative to one another varies depending on the underlying true model (31, 32).

G×E WITH GENE SETS

Set-based methods for G×E have emerged for detecting G×E effects within biologically defined sets, such as variants mapping to a particular pathway. A set-based G×E test is a single global test of interaction between an entire set of variants and the exposure of interest, rather than multiple individual tests, one per variant. The idea behind set-based methods is that accumulating multiple weak signals—possibly undetectable in isolation—across a set of variants may result in a detectable overall G×E signal. For rare variants, set-based methods are indispensable because the power to detect G×E with any single mutation is exceedingly small. Additional issues in rare-variant × E interaction analyses are provided in the online supplement (Web Appendix 2).

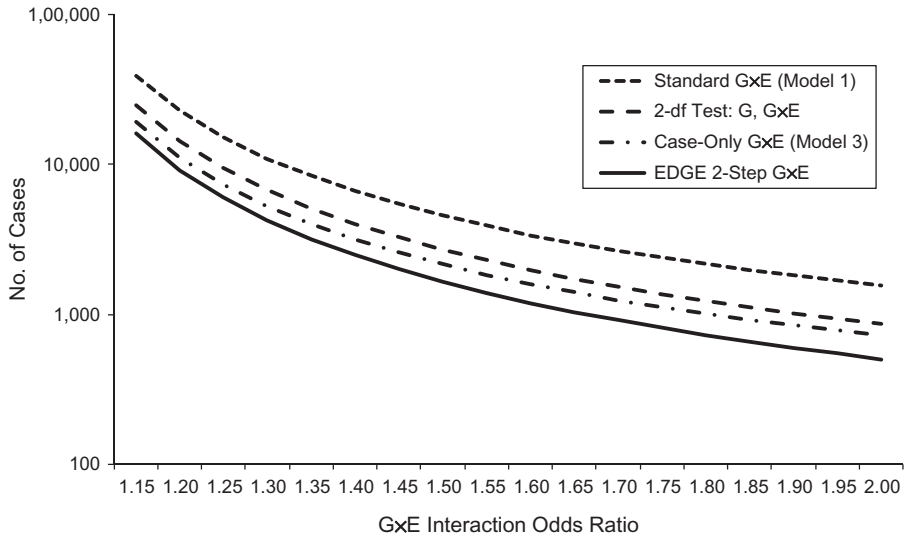

Figure 2. Required sample size (*n*) versus gene-environment interaction (G×E) effect size ($OR_{G\times E}$) to achieve 80% power using 4 different analysis methods, assuming an equal number of controls (except for the case-only analysis). The underlying model assumes that G has minor allele frequency 30% and additive (0–2) genotype coding, E is binary with prevalence 40%, and neither G nor E has a main effect on disease risk ($\beta_G = \beta_E = 0.0$, model 1). The calculations also assume a scan of 1 million single nucleotide polymorphisms and an overall type I error rate of 0.05, yielding significance threshold for a single interaction of single nucleotide polymorphism × E of 5×10^{-8}. For example, when $OR_{G\times E} = 1.5$, a case-control study would require $n = 4{,}557$ cases using a standard test of $\beta_{G\times E} = 0$ from model 1. The same power can be achieved with $n = 2{,}716$ cases by using a 2-df joint test of G and G×E, $n = 2{,}160$ by using a case-only analysis, or with only $n = 1{,}654$ by using the 2-step EDGE (31) approach. QUANTO (83), modified to accommodate the EDGE method, was used for computations.

Methods for testing set-based G×E effects can be broadly classified into 3 categories: burden-type (BT) tests, variance component (VC) tests, and a combination of both. For BT tests, G×E testing uses a "G" defined as a weighted risk score computed on the gene set. The weight can be informed by DG and/or EG screening statistics (48, 49), analogous to 2-step approaches for single SNPs. The VC approach is based on the assumption that G×E effects are random and follow an arbitrary distribution with mean 0 and variance τ^2. Testing G×E can be accomplished using a score test of H_0: $\tau^2 = 0$ (50) or using a regression model of genotypic and phenotypic similarity (51, 52). BT tests perform better when many variants in the set are causal and have effects in the same direction. In contrast, VC tests are more powerful when there is heterogeneity in magnitude and direction of effects. To potentially improve power across a range of underlying scenarios, hybrid methods that combine BT and VC tests have been proposed (53–55). Set-based analogs of the 2-df joint test of G and G×E have also been developed (56, 57). The BT, VC, and hybrid methods are implemented in the MiST-I software program (55).

THE COMPLEXITY OF EXPOSURE

A significant challenge in G×E analysis is the complexity of the environmental data (58). First, environmental exposures are heterogeneous in their type (e.g., continuous or discrete). Second, measurement modalities (e.g., community-level vs. home-level assessment of air pollution) can differ vastly in their measurement-error characteristics (59, 60), which in turn affects bias in estimates and power to detect effects. Third, exposures and their biological effects can vary considerably from the period before conception through adulthood. This has implications in estimation of effect sizes and interpretation. Fourth, exposures are often spatially, temporally, and/or culturally dependent. For example, air pollution levels can vary significantly between rural and urban settings and across decades (61, 62), and racial/ethnic-specific differences in exposure to phthalates (63) and air pollution (64) have been reported. Fifth, multiple environmental exposures can be highly correlated with one another (58, 65–68), making it difficult to identify the independent influence of a single exposure. Additional issues arise related to emerging exposome-type measurements (69).

G×E ANALYSIS IN A CONSORTIUM SETTING

For standard GWAS of marginal genetic effects, achieving sufficient sample size commonly requires merging data from multiple cohorts across a consortium. Due to ethical and data protection constraints, it is usually not possible to share individual level data (to perform a so-called mega-analysis), and the solution has been to perform meta-analysis of cohort-specific analyses. In brief, each study performs the same analysis (e.g., application of model 1), perhaps with some cohort-specific adjustment covariates if necessary.

Recent work has shown that meta-analysis of G×E is asymptotically similar to mega-analysis (70, 71). However, there are some important considerations. First, for a binary E, a consortium may choose to perform stratified analysis. Here the goal is to estimate the marginal G effect separately within the exposed and unexposed, and test for G×E based on heterogeneity across E strata (72–74). Advantages of this stratified approach are that

standard software for marginal G effects can be used, and one obtains stratified estimates and tests of G effects naturally. However, the stratified approach obviously does not extend to continuous E. Additionally, one may be tempted to overinterpret *P* values of G effects within each stratum, rather than being guided by the overall test of G×E that forms the basis for the primary analysis (4).

The distribution of exposure will almost certainly vary across studies in a consortium for reasons described above. For a continuous exposure, differences in distribution across cohorts are unlikely to affect the estimate of the G×E effect if the interaction effect is mostly linear (i.e., its direction and magnitude do not differ across the exposure range). In fact, such heterogeneity can lead to more precise estimates and greater power to detect G×E (Web Figure 3). If G×E effects are nonlinear (e.g., increased genotypic risks occur only above some threshold of ozone exposure), meta-analysis becomes more complicated and may have low power. Fundamental differences in how exposures are assessed across studies (e.g., different questionnaire items or satellite versus ground-based measurement of air pollution) may make it impossible to analyze G×E for some E in a consortium setting.

The inclusion of diverse and/or admixed populations in a consortium may increase power to detect G×E. Diverse populations can increase both genomic variation and the range of environmental exposures (75, 76). Diverse populations with varying levels of linkage disequilibrium are beneficial for fine mapping of G×Es to identify truly causal variation. In admixed populations, local patterns of genetic ancestry can be used to perform ancestry × E interaction analyses to increase power of discovery over traditional G×E analysis (77).

WHY HAVEN'T MANY G×ES BEEN IDENTIFIED?

While some G×Es have been reported (46, 78, 79), detecting and replicating them has been a challenge. A key reason may be low statistical power or, equivalently, the need for large sample sizes in order to detect interaction effects of moderate magnitude (Figure 2). Analysis (or reanalysis) in the largest possible samples (perhaps in a consortium setting), using the most efficient methods may lead to the identification of additional G×Es.

Measurement error of G and/or E will reduce the effective sample size and adversely affect our ability to detect G×E. The causal locus G is often not directly measured, and we rely on linkage disequilibrium between G and one or more measured marker loci. While linkage disequilibrium is generally high (e.g., $R^2 > 0.8$), it can vary substantially across the genome and across racial/ethnic groups. Difficulties in accurately measuring E as described above can lead to limited correlation between the observed and true E values. For example, dietary measures assessed from food frequency questionnaires commonly have low correlation with those assayed via a 24-hour dietary record (e.g., $R^2 < 0.5$). Obtaining more precise measures of G (e.g., direct sequencing) and E (e.g., repeat measurements or biomarkers of exposure) may be more cost-effective for improving power of G×E analysis than simply increasing sample size (80).

Of course, we cannot rule out the possibility that a G×E does not exist, or is not that important, for a particular phenotype. For example, common genetic variants are likely to have occurred in the more distant past, and over time selection pressure may have weeded out variants with large G×E effects, leaving little detectable interaction in a current study. As another example, G×E may be important for a specific rare variant and rare exposure, but the amount of disease risk attributable to the G×E will likely be negligible and nearly impossible to detect. Alternatives to omnibus G×E tests, including stratified analyses (81), may be more effective for uncovering the genetic and environmental architecture of complex traits.

AVAILABLE SOFTWARE FOR ANALYSIS OF G×E

As is true in many areas of research, the most efficient methods of statistical analysis may not get wide use unless they are implemented in available software. For smaller-scale analyses (e.g., analysis of single variants or a set of candidate genes), popular statistical software such as SAS (SAS Institute, Inc., Cary, North Carolina) or STATA (StataCorp LP, College Station, Texas) can be used for G×E analyses. However, these programs do not scale well to genome-wide analyses or more complex models. In the sections above, we cited 3 software programs specifically designed for high-volume G×E analyses using novel and efficient methods (28, 39, 82). Additionally, many of the papers we cite include links to software programs that implement the corresponding methods. As we continue to move into a more high-volume, "-omics" driven research environment, it is essential that there be a strong focus on developing efficient software tools that implement evolving approaches.

DISCUSSION

G×E analysis may hold the key to further understanding many complex traits. In recent years, more efficient methods for GWAS scans have been developed. These open the door to the analysis (or reanalysis) of existing resources to learn more about the range of genetic and environmental factors that affect a given trait. Modern methods for assessing exposure provide new opportunities but also new challenges for the detection of G×Es. Consortium-based studies or very large cohorts will likely be required to achieve adequate power for the analysis of G×E. The study of an admixed population, either alone or as part of a consortium, may increase power for detecting G×E and will certainly broaden the public-health relevance of any findings. The evolving availability of new -omics technologies will provide us with rich data resources for discovering G×Es and translating them into predictive/diagnostic models. Methods and software development for the analysis of G×E will need to keep pace in order efficiently use these exciting new data resources.

ACKNOWLEDGMENTS

Author affiliations: Division of Biostatistics, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California (W. James Gauderman, Juan Pablo Lewinger, David Conti); Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan (Bhramar Mukherjee, Seunggeun Lee); Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston,

Massachusetts (Hugues Aschard); Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France (Hugues Aschard); Biostatistics Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington (Li Hsu); Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts (Chirag J. Patel); Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California (John S. Witte, Caroline G. Tai); Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire (Christopher Amos); Department of Medicine, University of California San Francisco, San Francisco, California (Dara G. Torgerson); Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Nilanjan Chatterjee); and Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, Maryland (Nilanjan Chatterjee).

Research reported in this publication was supported by National Institute of Environmental Health Sciences, National Cancer Institute, and National Human Genome Research Institute of the National Institutes of Health under (grants P01CA196569, R01CA201407, P30ES07048, and R21ES024844 to W.J.G.; R21HG007687 to H.A.; R01CA140561, R01CA201407, and P01CA196569 to D.C.; R01CA189532, R01CA195789, and P01CA53996 to L.H.; P01CA196569 and R01CA201407 to J.P.L.; R21ES020811 and NSF DMS 1406712 to B.M.; UL1TR001086, R01LM012012, U19CA203654, and P30CA023108 to C.A.; R00ES023504 and R21ES025052 to C.J.P.; R01CA201358 and CA088164 to J.S.W.; and F31CA200139 to C.G.T.).

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. McAllister K, Mechanic LE, Amos C, et al. Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *Am J Epidemiol*. 2017;186(7):753–761.
2. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56–65.
3. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279–1283.
4. Aschard H. A perspective on interaction effects in genetic association studies. *Genet Epidemiol*. 2016;40(8):678–688.
5. Breslow N, Day N. *Statistical Methods in Cancer Research: II. The Design and Analysis of Cohort Studies*. Geneva, Switzerland: World Health Organization; 1987.
6. Kalbfleisch J, Prentice R. *The Statistical Analysis of Failure Time Data*. 2nd ed. Indianapolis, IN: John Wiley & Sons, Inc.; 2002.
7. Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika*. 1979;66(3):403–411.
8. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health*. 2005;95(suppl 1):S144–S150.
9. Madsen AM, Hodge SE, Ottman R. Causal models for investigating complex disease: I. A primer. *Hum Hered*. 2011;72(1):54–62.
10. VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology*. 2009;20(1):6–13.
11. 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.
12. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the Breast and Prostate Cancer Cohort Consortium. *J Natl Cancer Inst*. 2011;103(16):1252–1263.
13. Barrdahl M, Canzian F, Joshi AD, et al. Post-GWAS gene-environment interplay in breast cancer: results from the Breast and Prostate Cancer Cohort Consortium and a meta-analysis on 79,000 women. *Hum Mol Genet*. 2014;23(19):5260–5270.
14. 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.
15. Song M, Kraft P, Joshi AD, et al. Testing calibration of risk models at extremes of disease risk. *Biostatistics*. 2015;16(1):143–154.
16. Garcia-Closas M, Malats N, Silverman D, et al. *NAT2* slow acetylation, *GSTM1* null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet*. 2005;366(9486):649–659.
17. Kraft P, Yen YC, Stram DO, et al. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007;63(2):111–119.
18. 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.
19. Keller MC. Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry*. 2014;75(1):18–24.
20. Tchetgen Tchetgen EJ, Kraft P. On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. *Epidemiology*. 2011;22(2):257–261.
21. Han SS, Rosenberg PS, Garcia-Closas M, et al. Likelihood ratio test for detecting gene (G)-environment (E) interactions under an additive risk model exploiting G-E independence for case-control data. *Am J Epidemiol*. 2012;176(11):1060–1067.
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. Umbach DM, Weinberg CR. Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med*. 1997;16(15):1731–1743.
24. Chatterjee N, Carroll RJ. Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika*. 2005;92(2):399–418.
25. Albert PS, Ratnasinghe D, Tangrea J, et al. Limitations of the case-only design for identifying gene-environment interactions. *Am J Epidemiol*. 2001;154(8):687–693.
26. 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.

27. Chen YH, Chatterjee N, Carroll RJ. Shrinkage estimators for robust and efficient inference in haplotype-based case-control studies. *J Am Stat Assoc*. 2009;104(485):220–233.
28. Bhattacharjee S, Chatterjee N, Han S, et al. *CGEN: An R package for analysis of case-control studies in genetic epidemiology, Version 3.8.0.* 2012. http://bioconductor.org/packages/release/bioc/html/CGEN.html.
29. 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.
30. 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.
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, Jiao S, Dai JY, 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, Conti DV, et al. Sample size requirements to detect gene-environment interactions in genome-wide association studies. *Genet Epidemiol*. 2011; 35(3):201–210.
35. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*. 2009;169(2):219–226.
36. 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.
37. 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.
38. 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.
39. Gauderman W, Morrison J, Zhang P. *GxEScan: A Program for Detecting GxE Interactions in a Genome-wide Association Study, Version 0.5.0.* 2016. http://biostats.usc.edu/software.
40. Vieira C, Pasyukova EG, Zeng ZB, et al. Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. *Genetics*. 2000;154(1): 213–227.
41. Juenger TE, Sen S, Stowe KA, et al. Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica*. 2005; 123(1–2):87–105.
42. Winkler TW, Justice AE, Graff M, et al. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet*. 2015;11(10):e1005378.
43. Zuk O, Hechter E, Sunyaev SR, et al. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci USA*. 2012;109(4):1193–1198.
44. Aschard H, Hancock DB, London SJ, et al. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum Hered*. 2010;70(4):292–300.
45. Manning AK, LaValley M, Liu CT, et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet Epidemiol*. 2011;35(1):11–18.
46. Hancock DB, Soler Artigas M, Gharib SA, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet*. 2012;8(12):e1003098.
47. 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.
48. Jiao S, Hsu L, Bezieau 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.
49. Liu Q, Chen LS, Nicolae DL, et al. A unified set-based test with adaptive filtering for gene-environment interaction analyses. *Biometrics*. 2016;72(2):629–638.
50. 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.
51. 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.
52. Zhao G, Marceau R, Zhang D, et al. Assessing gene-environment interactions for common and rare variants with binary traits using gene-trait similarity regression. *Genetics*. 2015;199(3):695–710.
53. 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.
54. Jiao S, Peters U, Berndt S, et al. Powerful set-based gene-environment interaction testing framework for complex diseases. *Genet Epidemiol*. 2015;39(8):609–618.
55. Su YR, Di CZ, Hsu L, et al. A unified powerful set-based test for sequencing data analysis of GxE interactions. *Biostatistics*. 2017;18(1):119–131.
56. Kazma R, Cardin NJ, Witte JS. Does accounting for gene-environment interactions help uncover association between rare variants and complex diseases? *Hum Hered*. 2012; 74(3–4):205–214.
57. Broadaway KA, Duncan R, Conneely KN, et al. Kernel approach for modeling interaction effects in genetic association studies of complex quantitative traits. *Genet Epidemiol*. 2015; 39(5):366–375.
58. Ioannidis JP, Loy EY, Poulton R, et al. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci Transl Med*. 2009;1(7):7ps8.
59. Palmer CD, Lewis ME, Geraghtya CM, et al. Determination of lead, cadmium and mercury in blood for assessment of environmental exposure: a comparison between inductively coupled plasma–mass spectrometry and atomic absorption spectrometry. *Spectrochim Acta Part B At Spectrosc*. 2006; 61(8):980–990.
60. Zeger SL, Thomas D, Dominici F, et al. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect*. 2000; 108(5):419–426.
61. Lurmann F, Avol E, Gilliland F. Emissions reduction policies and recent trends in Southern California's ambient air quality. *J Air Waste Manag Assoc*. 2015;65(3):324–335.
62. Gauderman WJ, Urman R, Avol E, et al. Association of improved air quality with lung development in children. *N Engl J Med*. 2015;372(10):905–913.
63. Silva MJ, Barr DB, Reidy JA, et al. Urinary levels of seven phthalate metabolites in the US population from the National Health and Nutrition Examination Survey (NHANES) 1999–2000. *Environ Health Perspect*. 2004;112(3):331–338.

64. Zou B, Peng F, Wan N, et al. Spatial cluster detection of air pollution exposure inequities across the United States. *PLoS One*. 2014;9(3):e91917.
65. Gauderman WJ, Avol E, Gilliland F, et al. The effect of air pollution on lung development from 10 to 18 years of age. *N Engl J Med*. 2004;351(11):1057–1067.
66. Patel CJ, Ioannidis JP. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J Epidemiol Community Health*. 2014;68(11):1096–1100.
67. Patel CJ, Manrai AK. Development of exposome correlation globes to map out environment-wide associations. *Pac Symp Biocomput*. 2015:231–242.
68. Smith GD, Lawlor DA, Harbord R, et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med*. 2007;4(12):e352.
69. 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).
70. Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol*. 2010;34(1):60–66.
71. Sung YJ, Schwander K, Arnett DK, et al. An empirical comparison of meta-analysis and mega-analysis of individual participant data for identifying gene-environment interactions. *Genet Epidemiol*. 2014;38(4):369–378.
72. Randall JC, Winkler TW, Kutalik Z, et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet*. 2013;9(6):e1003500.
73. Myers RA, Scott NM, Gauderman WJ, et al. Genome-wide interaction studies reveal sex-specific asthma risk alleles. *Hum Mol Genet*. 2014;23(19):5251–5259.
74. Magi R, Lindgren CM, Morris AP. Meta-analysis of sex-specific genome-wide association studies. *Genet Epidemiol*. 2010;34(8):846–853.
75. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
76. Moore SC, Gunter MJ, Daniel CR, et al. Common genetic variants and central adiposity among Asian-Indians. *Obesity (Silver Spring)*. 2012;20(9):1902–1908.
77. Schoeps A, Rudolph A, Seibold P, et al. Identification of new genetic susceptibility loci for breast cancer through consideration of gene-environment interactions. *Genet Epidemiol*. 2014;38(1):84–93.
78. Lubin JH, Kogevinas M, Silverman D, et al. Evidence for an intensity-dependent interaction of *NAT2* acetylation genotype and cigarette smoking in the Spanish Bladder Cancer Study. *Int J Epidemiol*. 2007;36(1):236–241.
79. 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.
80. Wong MY, Day NE, Luan JA, et al. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? *Int J Epidemiol*. 2003;32(1):51–57.
81. Aschard H, Zaitlen N, Lindström S, et al. Variation in predictive ability of common genetic variants by established strata: the example of breast cancer and age. *Epidemiology*. 2015;26(1):51–58.
82. Sun J, Zheng Y, Hsu L. *MiST-I: A program for set-based gene-environment interaction tests.* Version 1.0. 2014. http://research.fhcrc.org/hsu/en/software.html.
83. Gauderman W, Morrison J. *Quanto 1.2: A computer program for power and sample size calculations for genetic-epidemiology studies.* 2009. http://biostats.usc.edu/Quanto.html.