aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2015, Vol. 10(1) 37–59

DOI: 10.1177/1745691614556682
pps.sagepub.com
SAGE

Candidate Gene–Environment Interaction Research: Reflections and Recommendations

Danielle M. Dick[1], Arpana Agrawal[2], Matthew C. Keller[3], Amy Adkins[1], Fazil Aliev[1], Scott Monroe[4], John K. Hewitt[2], Kenneth S. Kendler[1], and Kenneth J. Sher[5]

[1]Department of Psychiatry, Virginia Commonwealth University; [2]Department of Psychiatry, Washington University in St. Louis; [3]Institute for Behavioral Genetics, University of Colorado Boulder; [4]Department of Psychology, University of Notre Dame; and [5]Department of Psychological Sciences, University of Missouri

Abstract

Studying how genetic predispositions come together with environmental factors to contribute to complex behavioral outcomes has great potential for advancing the understanding of the development of psychopathology. It represents a clear theoretical advance over studying these factors in isolation. However, research at the intersection of multiple fields creates many challenges. We review several reasons why the rapidly expanding candidate gene–environment interaction (cG×E) literature should be considered with a degree of caution. We discuss lessons learned about candidate gene main effects from the evolving genetics literature and how these inform the study of cG×E. We review the importance of the measurement of the gene and environment of interest in cG×E studies. We discuss statistical concerns with modeling cG×E that are frequently overlooked. Furthermore, we review other challenges that have likely contributed to the cG×E literature being difficult to interpret, including low power and publication bias. Many of these issues are similar to other concerns about research integrity (e.g., high false-positive rates) that have received increasing attention in the social sciences. We provide recommendations for rigorous research practices for cG×E studies that we believe will advance its potential to contribute more robustly to the understanding of complex behavioral phenotypes.

Keywords

genetics, candidate genes, G×E, gene–environment interaction

There have been radical shifts as to belief about whether human behavior is more strongly determined by genes or by environment over the course of scientific history. Researchers in different fields seemingly advocated for the importance of one over the other (the so-called nature vs. nurture debate), with some camps studying genetic influence and others studying environmental factors. It is now widely accepted that both genetic and environmental influences are important, and characterizing how these influences come together to impact outcome, that is, the study of gene–environment interaction (G×E), has become an important area of study across multiple disciplines. That said, few research topics have generated more controversy and less clarity than the study of candidate gene–environment interaction (cG×E) in complex behavioral outcomes. Following the publication of cG×E studies in high-profile scientific journals (Caspi et al., 2002, 2003), in the last decade researchers have witnessed an explosion of interest in this area. There has been an exponential increase in the number of cG×E studies published, with researchers from diverse backgrounds routinely incorporating cG×E components into their studies. However, there has been growing skepticism about the replicability of many of these

Corresponding Author:
Danielle M. Dick, Department of Psychiatry, Virginia Commonwealth University, 1200 East Broad St., P.O. Box 980126, Richmond, VA 23298-0126
E-mail: ddick@vcu.edu

findings (e.g., Risch et al., 2009) and increasing concern about the quality of this rapidly expanding literature.

This concern led the National Institute on Alcohol Abuse and Alcoholism to sponsor a workshop in January 2013 that brought together a small group of researchers to discuss these challenges and to provide recommendations for how to move the field forward. Those discussions formed the foundation for this article, in which we review a number of reasons why the existing cG×E literature should be considered with a degree of caution. This is not to imply that true discoveries are absent in the literature. However, there are reasons to be concerned about the methods used and the conclusions drawn from many cG×E studies. Drawing from accumulating findings in psychiatric **genomics**,[1] we consider potential pitfalls and logical inconsistencies with some of the extant cG×E literature. We discuss ways of refining the development of cG×E hypotheses, conducting statistically rigorous analyses, and interpreting findings within the broader context of genetics research—all directions that we believe hold promise for advancing the potential of cG×E studies to contribute more robustly to the understanding of complex behavioral phenotypes.

History

The idea that genetic or biological predispositions are likely to interact with environmental factors to contribute to psychiatric and substance use disorders has been entertained for quite some time (Whytt, 1765). Long before it was feasible and cost-efficient to measure specific genes, researchers of twin studies documented that the importance of overall genetic influences (i.e., **heritability**) could vary considerably as a function of measured environmental factors (Kendler & Eaves, 1986). For instance, Kendler and colleagues found that people at highest genetic risk for depression (i.e., individuals with an identical twin with a history of depression) were significantly more likely than individuals not carrying a genetic predisposition to have an onset of the disorder in the presence of exposure to a severe stressful life event, suggesting that genetic factors influence the risk for major depression in part by altering individual sensitivity to the depression-inducing effects of stressful life events (e.g., Kendler et al., 1995). With methodological advances that allowed twin researchers to model how genetic influences change as a function of the environment (T. M. Button et al., 2009), studying G×E became a popular area of research in behavior genetics (T. M. Button, Lau, Maughan, & Eley, 2008; Dick, Bernard, et al., 2009; Dick, Pagan, Holliday, et al., 2007; Dick, Pagan, Viken, et al., 2007; Dick, Rose, Viken, Kaprio, & Koskenvuo, 2001; Dick, Viken, et al., 2007; Harden, Hill, Turkheimer, & Emery, 2008; Purcell, 2002; R. J. Rose, Dick, Viken, & Kaprio, 2001; South & Krueger, 2008).

The accumulating body of research has demonstrated that the importance of genetic influences can vary dramatically as a function of environmental context; alternatively phrased, the importance of environmental influences can vary dramatically as a function of genetic factors. For example, it has been demonstrated that genetic influences on adolescent substance use and externalizing behavior are far stronger under conditions of low parental monitoring (Dick, Pagan, Viken, et al., 2007; Dick, Viken, et al., 2007); high peer deviance (T. M. Button et al., 2009; Dick, Pagan, Holliday, et al., 2007; Dick, Pagan, Viken, et al., 2007; Harden et al., 2008); and state, school, and neighborhood conditions that provide reduced social monitoring and enhanced opportunity to use (Boardman, 2009; Dick, Bernard, et al., 2009; Dick et al., 2001; R. J. Rose et al., 2001). However, this body of G×E research did not gain widespread recognition outside the field of twin research. It was not until the influential *Science* publication by Caspi et al. (2003), attributing part of the genetic sensitivity to the depressogenic effects of stressful life events to variations in a specific DNA sequence (a **polymorphism** in the serotonin-transporter-linked polymorphic region [5-HTTLPR]), that G×E research became a widely recognized area of study outside the field of behavior genetics. However, an important distinction arose between the G×E work conducted in the field of behavior genetics and the widespread adoption of G×E by other fields, particularly the social sciences. Historically, the research conducted by behavior geneticists focused on "latent" genetic influences. This means that the importance of genetic factors is estimated statistically by phenotypic similarity across individuals with different degrees of genetic and environmental sharing, with methodologies such as family, twin, and adoption studies (Bergeman & Plomin, 1989). Using this method, researchers estimate the overall importance of genetic effects on a phenotype, that is, the total contribution of all genes influencing the phenotype. G×E in this context means that the overall importance of genetic variance differs across environments. In contrast, most G×E researchers in fields outside behavior genetics have studied measured candidate genes. In these studies, researchers test whether the association of a specific genetic variant with a given outcome varies across different environments. We refer to these studies as cG×E, and they are the focus of this review.

The publication of several high-profile cG×E studies (e.g., *MAOA* × Maltreatment in Antisociality; Caspi et al., 2002), as well as the technological advances in genetics that made genotyping accessible and cost-efficient, likely contributed to the dramatic increase in cG×E research. Regardless of the validity and reproducibility of those initial efforts (which continues to be debated; Brown & Harris, 2008; Clarke, Flint, Attwood, & Munafo, 2010;

Culverhouse et al., 2013; Karg, Burmeister, Shedden, & Sen, 2011; Munafo, Durrant, Lewis, & Flint, 2009; Risch et al., 2009), they left their mark on the field by creating widespread recognition of the potential importance of the interplay between genetic and environmental factors in developmental pathways underlying the etiology of behavioral outcomes in ways that the latent G×E work of behavior geneticists had failed to do. In the excitement surrounding the initial cG×E findings, and spurred by funding initiatives that encouraged research in this area, investigators from disparate backgrounds incorporated measured genotypes into their studies. In the wake of historical tension surrounding the relative importance of genetic versus environmental effects (the so-called nature vs. nurture debate), G×E provided a conciliatory framework that could facilitate a synthesis of scientific fields that had historically been at odds with one another.

However, early enthusiasm for cG×E findings has waned as the number of failures to replicate original findings mounted. In many ways, the progression of cG×E studies has closely paralleled the trajectory of studies of the main effects of candidate genes, with early enthusiasm and adoption of genotyping candidate genes giving way to a literature plagued by small studies, failures to replicate, and a proliferation of novel findings with effect sizes that appeared at odds with what was subsequently found with well-powered studies (Neiswanger, Kaplan, & Hill, 1995). However, the study of genetic main effects has advanced dramatically since the early days of candidate gene research. We believe that what has been learned about the genetics of complex behavior from studies of genetic main effects yields insights into previous cG×E studies and ways to improve such research in the future. We begin by providing a broad review of developments in the field of psychiatric genetics over the past decade and discuss how this knowledge can inform studies of cG×E.

Early statistical genetics: Linkage and candidate gene studies

The field of statistical genetics, focused on finding genes that contribute to behavioral outcomes and disorders, has undergone rapid advances over the past decade. As researchers' knowledge about genetics has progressed, so too have the methodologies favored for gene identification (see Fig. 1 for an overview of common gene-finding strategies). Linkage and candidate gene studies were early gene identification strategies, believed to have complementary strengths.

Using linkage studies, researchers agnostically scanned the mapped genome by looking for chromosomal regions that were shared among affected family members (suggesting there was a gene in that region that contributed to the disorder). The advantage of linkage was that it did not require any a priori knowledge of the underlying biology of the outcome, in theory making it possible to discover new genes involved in the outcome that could expand the understanding of the biology of the disorder. Linkage studies were used to successfully identify many genes that contributed to **Mendelian disorders**, in which a single gene following a straightforward inheritance pattern with a major impact on outcome was present (Gusella et al., 1983; Murray et al., 1982; Tsui et al., 1985). However, linkage methods were less successful when applied to complex behavioral outcomes in which many genes are likely to be involved, each having just a small effect on the behavior, along with the environment.

In contrast to the hypothesis-free linkage approach, researchers using candidate gene studies focused on genes, and variants within those genes, that were hypothesized to have biological relevance to the outcome of interest. In this way, candidate gene studies had the advantage of being more precise than linkage studies in that they had the potential to pinpoint specific genes or genetic variants, rather than just specific chromosomal regions. However, they relied on the investigator to correctly "guess" what genes were biologically relevant to the outcome. Despite thousands of candidate gene publications on behavioral phenotypes, the approach remains controversial, and very few candidate gene findings are widely accepted within the genetics community. Over time, it has become clear that the genetic architecture of behavioral traits is highly complex, that effect sizes of genetic polymorphisms are likely to be small (as discussed later), and that scientists' ability to predict a priori which genes are likely to be relevant to a behavioral outcome has been very poor (Bosker et al., 2011; Colhoun, McKeigue, & Davey Smith, 2003; Need et al., 2009; Sullivan et al., 2008).

Later statistical genetics: Genome-wide association studies (GWAS)

As the cost of genotyping dropped during the 2000s, it became possible to conduct association tests, as are done in candidate gene studies, but across the entire genome. Such **GWAS** are hypothesis-free as with linkage studies, but they have much higher power to detect small effects of common variants. In GWAS, hundreds of thousands to millions of **genetic markers** known as single nucleotide polymorphisms (**SNPs**) are genotyped across the genome in an attempt to identify **common variants** that are associated with a particular outcome (disorder, behavior, etc.), suggesting that a particular genetic variant (or one very nearby) contributes to the outcome. In a sense, GWAS combine the advantages of linkage studies

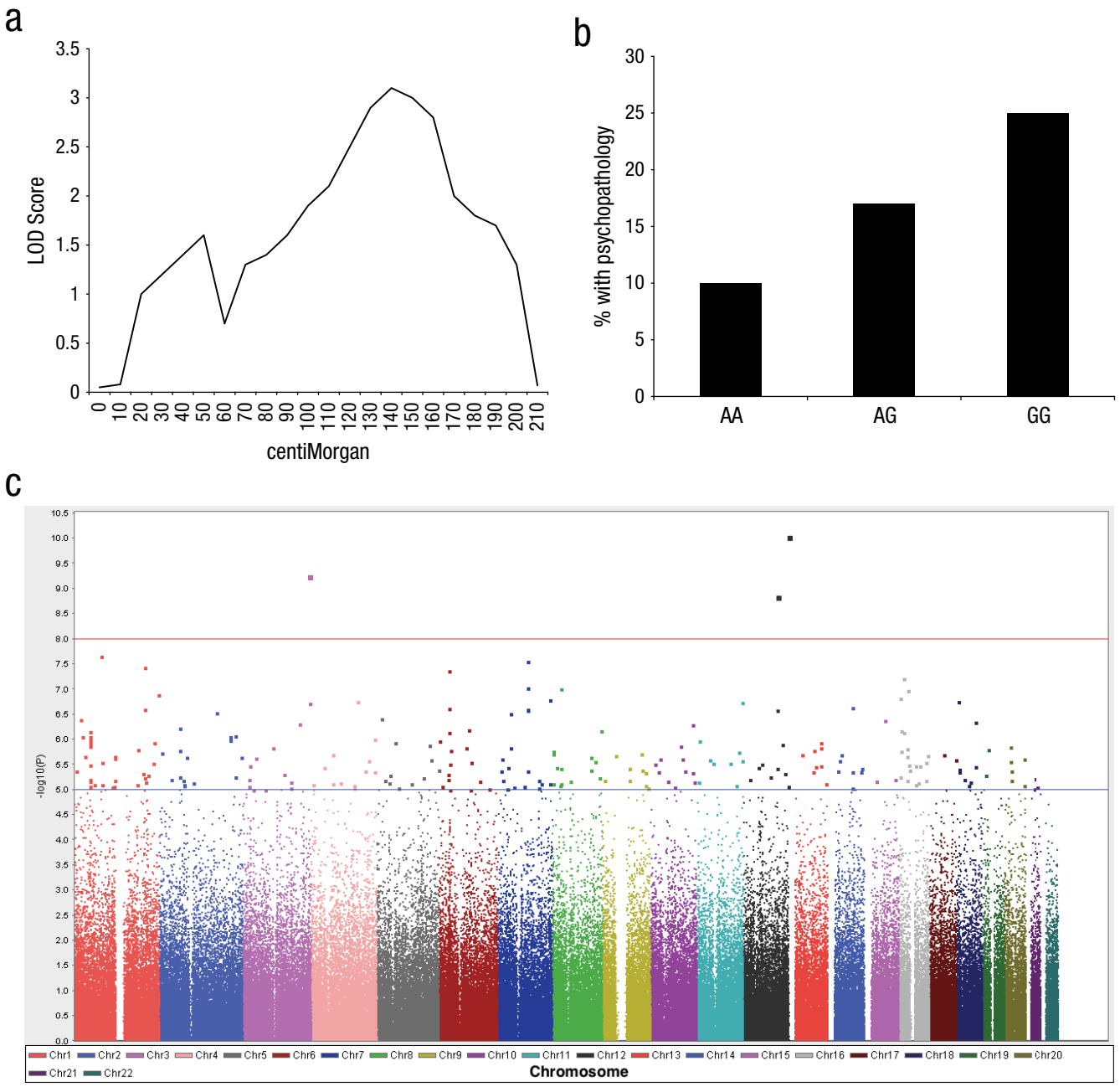

Fig. 1. Overview of common gene-finding strategies. LOD = logarithm of the odds.

(agnostic screening across the genome) and candidate gene studies (more precise localization of the gene/genetic variant).

GWAS were made possible by the rapid discovery of many more genetic variants through the International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/), which had the goal of developing a public resource to catalogue normally occurring genome-wide variation in SNPs by annotating the sample genomes of small groups of ethnically homogenous individuals (e.g., Caucasians of European descent are represented by Utah residents with ancestry from northern and western Europe collected in 1980) to create what is known as a **reference panel**. In addition to providing a database for identification and comparison of the polymorphic nature of SNPs, HapMap also allows for the examination of the extent to which neighboring SNPs are correlated with each other via a **population genetics** process called **linkage disequilibrium** (LD). When two or more SNPs are in high LD (e.g., correlations > 0.8), if one of the SNPs is genotyped, then the genotype at the others can be probabilistically inferred via **imputation**. The ability to infer surrounding genotypes on the basis of LD patterns means that researchers can now cost-efficiently scan the genome with a much smaller subset of

markers than previously needed (e.g., in the range of 370,000–600,000) and impute the remaining commonly occurring markers across the genome. Whereas early attempts to impute variation were restricted to common SNPs (>5% **minor allele frequency**), the landmark 1000 Genomes Project (http://www.1000genomes.org/) used a similar strategy to identify less common SNPs (≤1%). Researchers' growing knowledge of genetic variants across the genome, coupled with exponential decreases in costs of high-density GWAS arrays (>1 million SNPs at <$100 per participant currently), now allows investigators to interrogate more than 10 million polymorphisms in relation to outcomes.

Two primary points have become apparent over the last several years from GWAS: (a) The effect sizes associated with individual genetic variants are very small, usually with odds ratios on the order of 1.1, and (b) researchers' ability to select a priori which genes are viable candidates for psychiatric and substance use disorders has been poor (Kendler, 2013; Sullivan, Daly, & O'Donovan, 2012). There are rare exceptions, such as the role of alcohol dehydrogenase genes in alcohol dependence (Shen et al., 1997). Researchers now realize that early candidate gene studies, as well as early atheoretical systematic gene-finding efforts (such as linkage studies), were underpowered to detect genes with the small effect sizes that more recent studies suggest are likely to be realistic. Although GWAS were more successful for some conditions (Crohn's, diabetes, macular degeneration; Manolio & Collins, 2009), like linkage studies, early GWAS were largely unsuccessful in the area of psychiatric and substance use disorders (Kendler, 2013). Few SNPs were detected that met **genome-wide levels of significance**. As increasingly large sample sizes have been procured, researchers now know that most early GWAS were simply underpowered (Visscher, Brown, McCarthy, & Yang, 2012). Amassing sample sizes through large consortia on the order of tens of thousands or more has revealed that the number of significant findings increases as the sample sizes increase.

The discoveries from these large GWAS are robust and replicable, and for certain phenotypes, the amount of variance explained in total from genome-wide significant SNPs is becoming nontrivial (e.g., 60% for Type 1 diabetes, 10% for height; Visscher et al., 2012). In the area of psychiatric genetics, studies of the genetic basis of schizophrenia are currently enjoying the most success, in which sample sizes on the order of >13,800 cases and >18,000 controls have now been accumulated, leading to the detection of >3,500 loci in 12 genomic regions that contribute to the disorder (Ripke et al., 2013). It is noteworthy, however, that even with such impressive sample sizes, Ripke et al. (2013) acknowledged in their landmark study (which is undergoing a further growth in sample size) the lack of power to detect **genotype relative risks** < 1.1.

Reasons to Be Concerned About the Published cG×E Literature

The emerging genetics literature suggests that the small sample sizes (e.g., n < 1,000), typical of candidate gene studies to date, are likely to be grossly underpowered for detecting genetic influences with small effect sizes. (The reason that many, perhaps most, researchers using candidate gene studies have reported positive associations despite such lack of power is discussed later.) Some of the confusion about expected effect sizes and the sample sizes needed for cG×E studies may surround differences in the conceptualization of G×E effects across different fields (see Fig. 2 for an illustration of this). There are two ways that cG×E studies can be conceptualized: (a) as a genotype moderating the association between an environmental factor and an outcome (i.e., increasing exposure to major stressful life events is more strongly associated with an increased risk for depression in the presence of the short allele of 5-HTTLPR—often the default conceptualization for psychologists) or (b) as an environment moderating the association between a genotype and an outcome (i.e., the association between the short allele of 5-HTTLPR and depression is strongest in individuals experiencing major stressful life events—often the default conceptualization for geneticists). Statistically, these are equivalent and indistinguishable, but they can lead to different interpretations of the same data and different expectations about the likelihood of detecting a G×E effect. Conceptualizing G×E as a genetic effect (on an environment–behavior association) may lead one to assume small effect sizes on the basis of the growing GWAS literature demonstrating that genetic effects on complex outcomes generally have very small effect sizes. Conceptualizing G×E as an environmental effect on a gene–behavior association may lead one to assume larger effects sizes. Nevertheless, under either interpretation, the effect size of the candidate gene is critical. Recognizing the modest main effect sizes produced by individual candidate gene polymorphisms, it may be overly optimistic to presume that the effect sizes associated with cG×E will be systematically larger. This is clearly a matter of some debate, as if certain types of cross-over interactions were prevalent, it would be possible. Nevertheless, the lessons we have learned from the history of gene finding underscore the need for researchers to be cognizant of the strong possibility that they are dealing with small effect sizes (or to provide strong justification for why larger effect sizes are expected) and to demonstrate that their samples are adequately powered.

In addition to what GWAS have taught researchers about genetic effect sizes, GWAS have also been informative as to the likelihood that a hypothesized candidate gene will be associated with the hypothesized outcome. Robust and replicable GWAS signals (e.g., *CACNA1C* for

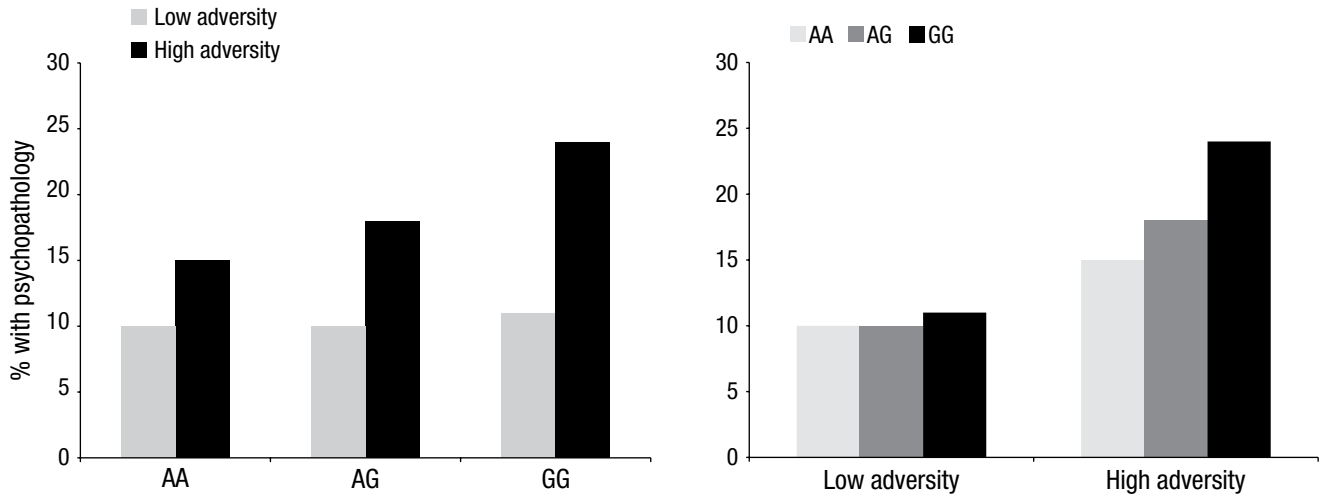

Fig. 2. Alternative portrayals of a Candidate Genotype (in this example, a single nucleotide polymorphism) × Environment (in this example, high vs. low adversity) interaction. The left panel shows that the strongest impact of high adversity on psychopathology is found for individuals with the GG genotype. Alternatively, the right panel shows that the strongest association of the genotype with psychopathology is among those experiencing high adversity. These are alternate presentations of the same interaction.

schizophrenia, Ripke et al., 2013, and bipolar disorder, Ruderfer et al., 2014) have tended to be distinct from those that were routinely hypothesized in candidate gene studies (e.g., *COMT, MAOA*; Craddock, Dave, & Greening, 2001), casting considerable doubt regarding the burden of a priori evidence for these selections. With rare exceptions (e.g., rs16969968 in the *CHRNA5–CHRNA3–CHRNB4* cluster, associated at $p < 10–70$ across several GWAS, was initially posited as a candidate gene; Tobacco and Genetics Consortium, 2010), widely studied candidate genes were not found to be significant when studied systematically across the backdrop of the genome in well-powered studies (Bosker et al., 2011; Collins et al., 2012; Lasky-Su et al., 2008). In fact, some of the findings to emerge from GWAS in other areas, such as Crohn's disease, have suggested new pathways that were not previously suspected to play a role in the disorder and that drastically changed the presumed understanding of the underlying biology of the disorder (Manolio & Collins, 2009). Past experience would suggest that best guesses for candidate genes affecting environmental sensitivity (i.e., cG×E) are unlikely to fare better than they have for other phenotypes investigated to date in GWAS. The combination of low prior likelihood of a given candidate being correct, compounded by likely small effect sizes and low power for any given truly associated variant, suggests that the false discovery rate—the proportion of "discoveries" in candidate gene main effect and cG×E studies that are actually false—may be unacceptably high (Duncan & Keller, 2011).

In addition to the potential for low power and low **prior probabilities** associated with the study of candidate genes (Munafo, 2009), it is also likely that there is insufficient correction for and underreporting of multiple testing. Publication bias is also probable, whereby authors are more likely to submit, and editors are more likely to accept, cG×E findings that are statistically significant. A recent report notes that problems contributing to and a consequence of such bias are rampant in the cognitive sciences (Ioannidis, Munafo, Fusar-Poll, Nosek, & David, 2014), and there is similar evidence in the cG×E literature (Duncan & Keller, 2011). Specifically, the vast majority of first reports of a given cG×E finding were positive, whereas a much lower proportion of attempted replications were positive. Ioannidis et al. (2014) referred to this as the *Proteus phenomenon*, whereby the rapid publication of positive findings might temporarily create a halo in which negative findings might be more readily entertained by journal editors for a short period of time. Furthermore, and inconsistent with expectations based on statistical power, Duncan and Keller (2011) found that the larger the cG×E sample, the less likely it was to be significant. This trend would not be expected if cG×E findings were valid. These observations are consistent with widespread publication bias in the cG×E literature. Publication bias can arise when authors, editors, and reviewers believe that positive findings are more worthy of publication than are negative or null results. There are many uninteresting ways for an empirical test of a hypothesis to fail to support it—the hypothesis was implausible to start with, the power of the test is inadequate, the operationalization of the variables is not valid, and so forth. As a consequence, the greater interest in positive results is understandable, especially when they bring an insightful increment to, or even transformation of, researchers' understanding. Yet, interest in positive findings is clearly misplaced if they are false, as they can (mis)guide research efforts and funding priorities.

These problems are not unique to the study of cG×E (Ioannidis et al., 2014; Spellman, 2012). Concern about unacceptably high false-positive rates in the social sciences has garnered growing attention over the past several years. In a provocative article by Ioannidis (2005) entitled "Why Most Published Research Findings Are False," several conditions are outlined that contribute to why a novel research finding ultimately may be in error. These conditions include smaller studies; smaller effect sizes; greater number and lesser preselection of tested relationships; greater flexibility in designs, definitions, outcomes, and analytic models; greater interests and prejudices surrounding the area of research; and situations in which there are more scientists in a field involved in chase of statistical significance. We believe that all of these conditions are likely to contribute to findings in the study of cG×E.

A more recent article compellingly demonstrated how flexibility in data collection, analysis, and reporting can dramatically increase false-positive rates (Simmons, Nelson, & Simonsohn, 2011). This recognition has led to a growing movement in the social sciences to adopt new practices to promote research integrity (Cumming, 2014), including prespecification of studies and hypotheses, avoidance of selection and other questionable data-analytic practices, complete reporting of analyses and variables, and encouragement of replication. A "new statistics" method has been proposed that includes recommended statistical practices, such as estimation based on effect sizes, confidence intervals, and meta-analyses (Cumming, 2014). These are practices that have already become more widespread in the genetics field (Agrawal et al., 2012; Boraska et al., 2014; Ripke et al., 2013; Steinberg et al., 2014; Stephens et al., 2013; Thompson et al., 2014), in which a plague of inconsistent and nonreplicable genetic main effects has led to the adoption of more rigorous statistical practices, which have proven successful in advancing the field (Corvin, Craddock, & Sullivan, 2010; Sullivan et al., 2012; van Assen, van Aert, Nuijten, & Wicherts, 2014). Following these guidelines would go a long way toward improving the quality and trustworthiness of the cG×E literature as well.

In the following sections, we focus on practical problems as they relate to cG×E. We focus this discussion on two major components of cG×E research—the core ingredients of the interaction (i.e., how G [gene] and E [environment] are measured) and the recipe for combining them (i.e., statistical problems with modeling their interaction).

The ingredients of cG×E: The choice of genetic and environmental variables

In the enthusiasm surrounding the study of cG×E, many investigators have expanded their studies to include measures of G or E, when that was not the original focus of the study. This expansion into new areas has happened in both directions: Researchers who have focused on carefully characterizing environmental effects have expanded their studies to include measured genes, and researchers who have focused on gene finding have expanded their studies to include measures of the environment. In theory, this expansion of cross-disciplinary science is a positive development; however, an unfortunate corollary has been that the added component does not always represent the state of the science in the other respective field.

The choice of G. For example, the vast majority of the G that has been incorporated into cG×E studies consists of a handful of "usual suspect" candidate genes (Munafo, 2006; e.g., *SLC6A4* [also known as *5-HTT, MAOA, DRD2, COMT*]), as discussed earlier. Often a single genetic marker is genotyped to represent the gene. Genotyping a single marker in a gene does not reflect the state of the science in genetics, which has moved toward more comprehensive approaches to gene finding. It may be appropriate to genotype a single genetic variant when that variant has a known **functional** impact on the gene (i.e., it produces an observable alteration in the manner in which the gene encodes the protein product). However, above and beyond this simple annotation of function, the impact of a candidate polymorphism is very challenging to establish. At its simplest, even when modeling a single variant, the characterization of that variant (or its mode of inheritance) can profoundly impact detection of cG×E. For instance, which **allele** is assigned as a **risk allele** and how many copies of this allele are required to quantify the diathesis need to be determined. How to model a genotype becomes a particular concern with smaller samples, in which **homozygotes** (individuals who carry two copies of a given allele) of the minor allele are often combined with **heterozygotes** (individuals who carry one copy of a given allele) in the interest of statistical power, potentially obfuscating the complexity underlying the genetic model. Consider the consequences if one applied this practice to other known genetic outcomes: For example, researchers know that two copies of one of several of the *CFTR* **mutations** are required for the diagnosis of cystic fibrosis (a disorder with a **recessive** mode of genetic inheritance) and that this is etiologically distinct from the one copy of the *HTT* trinucelotide expansion required for a diagnosis of Huntington's disease (a **dominant** disorder). Imagine the confusion if a prenatal genetic counselor ascribed the same degree of vulnerability to a fetus that tests positive for one copy (a **carrier**, unaffected with disease) versus two copies (will manifest the disease) of the *CFTR* mutation. Although the action of genetic variants on complex traits is not **monogenic**, the technical specification of their purported mode of action is still significant.

Large-scale, gene-finding efforts have moved to systematic screens of the genome (e.g., Sklar et al., 2011; Treutlein et al., 2009; Wray et al., 2012), pathway and network analysis (Weng et al., 2011), and integration with model organism genetics (Zhao et al., 2012) as new avenues for increasing the probability of identifying relevant genes. Further, when a gene of interest is being studied, genetic variants across the gene are usually genotyped to capture the many locations across that gene that could be involved[2] in altering gene regulation or function and producing different effects on behavior. For example, many early studies genotyped the Taq1A1 polymorphism in *DRD2* as a hypothesized biological candidate for a variety of outcomes related to reward deficiency, for which dopamine transmission was thought to play a role (Blum et al., 1990; Meyers et al., 2013; Young, Lawford, Nutting, & Noble, 2004). Subsequently, it was discovered that the polymorphism was actually located in the neighboring gene *ANKK1* (Neville, Johnstone, & Walton, 2004). Researchers using more systematic studies of genetic variation across the region containing *ANKK1* and *DRD2* have found evidence that multiple genes may be involved (Gelernter et al., 2006; Yang et al., 2007, 2008). Differences in the way that individual genes, gene networks, and whole genomes are systematically studied has led to increasing distance between the gene-finding world and studies of cG×E, which still focus largely on the usual suspect candidate polymorphisms. Clearly, better integration of these research areas is necessary.

The choice of E. An equally important issue concerns the choice of E. The challenges within this area are illustrated in the literature examining life stress as an environmental risk factor for depression. In attempting to replicate the original cG×E results of Caspi et al. (2003) pertaining to life events, many investigators have incorporated a wide variety of ad hoc measures of life stress (Monroe & Reid, 2008). Almost any form of adversity or challenge, at any time in a person's life, has been used as an alternative index of "stress." For example, replication studies on this cG×E topic have included participant life stress occurring over a range of time, from 1 month through a lifetime. Other researchers have adopted life event scales known to possess serious measurement deficiencies (Monroe, 2008) or simply have used unique measures never used previously. In one review of this literature, researchers reported psychometric properties of the stress measures in only 5 of 18 studies (Monroe & Reid, 2008). Finally, even when roughly common omnibus measures of life events were adopted, how the life events were combined (over time, severity, and type) for the final index of stress varied greatly across investigations (Uher & McGuffin, 2010).

The use of measures of the environment with proven reliability, empirical precedent, and theoretical plausibility is critical for advancing the field. The elasticity and looseness of the environmental construct can present serious problems for measurement and for the likelihood of detecting a cG×E (Monroe & Reid, 2008). In the example of stress, it is highly likely that different types of stress are relevant for different types of cG×Es for particular kinds of disorder or disease. For example, chronic stress over years may be most relevant for conditions that develop over protracted periods of time (e.g., coronary heart disease), whereas acute, aversive life events may be most relevant for conditions that often appear to come on rather quickly (e.g., a major depressive episode). Additionally, the developmental timing of the stressor can be critical as the social and biological impact can be expected to vary as a function of stage of development. More specific theoretical dimensions associated with stress should accompany such temporal distinctions (e.g., loss or humiliation life events vs. life events conveying threat and danger vs. other types of adaptive challenges). In a very real sense, there should be candidate "stressors" proposed for particular conditions based on a theoretical understanding of the plausible underlying mechanisms. Viewed from an alternative perspective, no one would expect to find a "true" cG×E if the wrong gene was assessed for the particular outcome. In a similar manner, any potentially valid cG×E will go undetected if the wrong form of environment is assessed or if the right form of the environment is assessed poorly (Monroe & Reid, 2008).

Problems with the recipe: Statistical concerns in cG×E research

There are a number of statistical considerations that influence the detection and interpretation of cG×E that have not received widespread attention in the literature.

We highlight some of the most critical issues next, and many of our concerns are consistent with those of others who have recently written on this topic (e.g., Roisman et al., 2012; Zammit, Lewis, Dalman, & Allebeck, 2010; Zammit, Owen, & Lewis, 2010).

The importance of scale. First, evidence for interactions can depend on choice of scale as well as choice of statistical model. Despite one's conviction in the presence or absence of cG×E, interactions are statistical phenomena and only have meaning in the context of a specific statistical and measurement model. However, many behavioral measures have no "true scale." One might, for instance, argue that a construct such as height has a true scale and, moreover, has a *ratio* scale with a meaningful zero point and equal intervals between data

points (Stevens, 1946): A board of 2 feet is twice as long as a board that is 1 foot long. However, many, if not most, of the constructs in the behavioral sciences do not have meaningful zero points and, thus, are scaled somewhat arbitrarily. Quantitative scales without meaningful zero points can vary further as to whether they have meaningful intervals between measurement points or simply reflect differences in relative magnitude. This is true of measures of both behavioral outcomes (e.g., a depression score) and environments (e.g., family function, peer deviance, neighborhood disintegration). The scale of measurement matters profoundly in interaction research because evidence for an interaction can change solely depending on arbitrary choice of scale (Eaves, 2006). For example, predictors that combine multiplicatively to influence the outcome variable will combine additively if the outcome is log-transformed. In such situations, the significance of the interaction term depends on how the outcome is scaled, which, in most behavioral research, is arbitrary.

The selection of model. As with choice of transformation, choice of how to model interactions can profoundly affect evidence for them. In particular, there has long been debate in epidemiology regarding the relative utility of risk differences versus risk ratios. That is, if the rate of illness is 10 and 50 per 10,000 in groups unexposed and exposed to some risk factor, are researchers more interested in the risk difference (40 new cases per 10,000) or the risk ratio of 5? From a practical perspective (e.g., public heath impact, focus for possible prevention, advice to patients), the risk-difference approach has much to recommend it. However, the risk-ratio approach is the more dominant in large part because it is easily implemented statistically in logistic regression. This distinction is critical because it defines the baseline model from which researchers assess interactions. In a risk-difference framework, an interaction reflects a deviation from a model in which risk factors add together. In a risk-ratio framework, an interaction reflects a deviation from a model in which risk effects multiply. Accordingly, the usage of logistic regression to study cG×E for binary outcomes of interest (such as presence/absence of a disorder) fundamentally changes the nature of the relationship between two variables: Multiplication on the original scale of a variable conforms to addition on the logarithmic scale. Thus, an "interaction" on the original scale can "disappear" or even be of opposite sign on the logarithmic scale, and vice-versa. We refer the reader to other sources (Kendler & Gardner, 2010; Zammit, Lewis, et al., 2010) for further discussion of these important issues.

The use of cross-product terms. Statistical tests of cG×E effects often rely on the modeling of a **cross-product term** in a regression-type model. Valid detection of true interactions in these models requires that factors that could produce spurious interactions be ruled out. For example, when predictors are correlated and quadratic terms are not modeled, the cross-product term can carry the variance of the unmodeled quadratic term and generate spurious interactions (Lubinski & Humphreys, 1990). Moreover, failure to include quadratics can also result in false-negative findings of interactions or the reversal of sign of true interactions (Ganzach, 1997). More generally, if the underlying relationship between G or E and an outcome is nonlinear (e.g., a spline or higher order polynomial), misspecification of the analysis by failing to include a term to model the nonlinearity can generate a significant interaction term in the absence of a true interaction.

The use of a cross-product term can be particularly problematic for modeling three-level, categorical genotypes. In the standard practice of assuming an **additive** genetic model, the use of a cross-product term will force the slope difference to be the same among all genotypic groups (e.g., the difference in slope between people carrying 0 vs. 1 copies of the risk allele is constrained to be the same as the slope difference between individuals who carry 1 vs. 2 copies of the risk allele). It also forces the lines for the three genotypic groups to all cross at the same point when an interaction exists. Accordingly, an interaction will only be accurately represented by the cross-product term when these conditions are met, and there is no a priori reason to assume that these constraints are sensible. This means that the regression lines implied by the use of a cross-product term may not accurately reflect the interaction present in the data. Figure 3 illustrates the problem and demonstrates how a reparameterization of the regression equation with parameters additional to the single cross-product term can correct it (as further delineated in Aliev, Latendresse, Bacanu, Neale, & Dick, 2014).

The importance of covariates. Failure to properly control for potential confounds can also be problematic in cG×E research. In nonexperimental research, researchers typically enter potential confounding variables (e.g., gender, ethnicity, socioeconomic status, genotype quality) into regression equations to control for their effects. However, this approach controls only for the additive effects of covariates; it does nothing to control for the potential confounding effects that these covariates might have on the interaction itself (Keller, 2014; Yzerbyt, Muller, & Judd, 2004). To properly control for confounders in cG×E research, investigators must also evaluate all relevant Gene × Covariate and Environment × Covariate interaction terms. To date, virtually no cG×E researchers have appropriately controlled for all covariate interactions

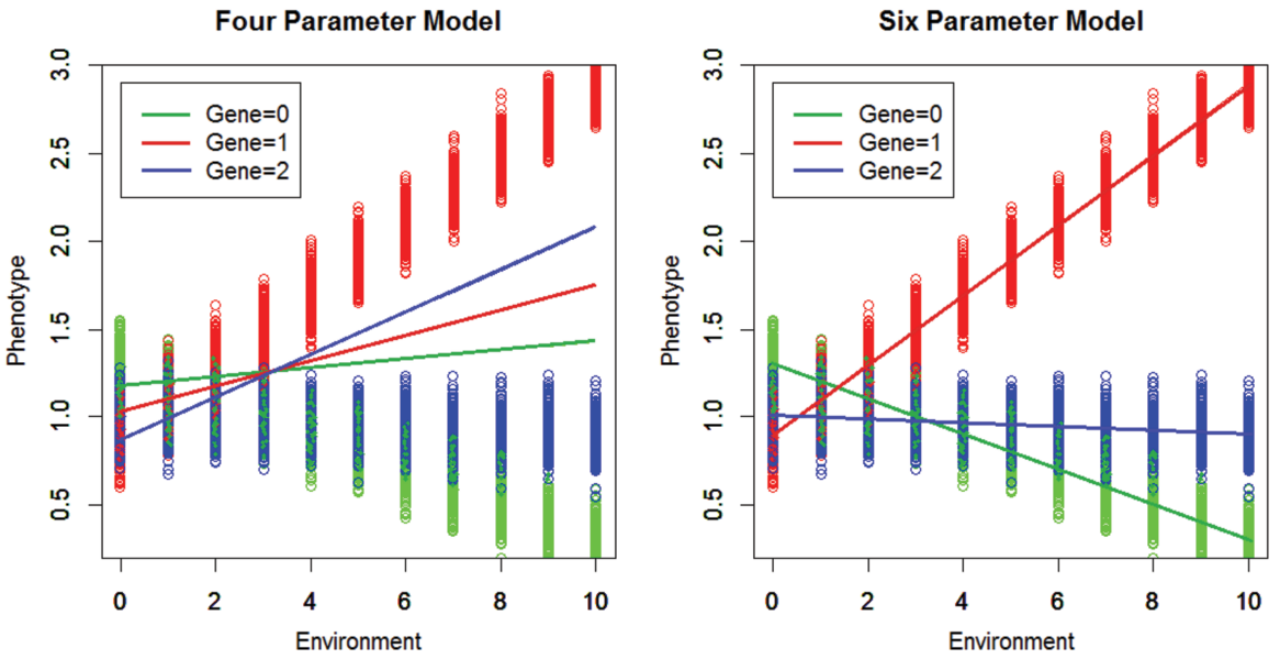

Fig. 3. The figure presents simulated phenotypic data for three genotypic groups (Gene = 0, 1, 2, indicating groups of individuals who carry 0, 1, or 2 copies of a particular allele), each shown in a different color. The four-parameter model corresponds to the case in which the interaction term is modeled by a cross-product term only. Although a significant interaction is detected, the corresponding linear regression lines do not match the data points, and the slopes are incorrectly ordered from 0 to 1 to 2 on the basis of the constraints imposed by the use of the cross-product term to model the interaction. Thus, although the model would produce a significant interaction, the regression lines implied by the model inaccurately represent the data and would be misleading as to the nature of the interaction. The data can be accurately reproduced by an extended parameterization of the regression model (six-parameter model) as detailed in Aliev et al.'s (2014) study.

(Keller, 2014). This failure to include covariates is particularly concerning in mixed-ethnicity samples, in which **stratification** can not only produce spurious genetic main effect association to be detected (Price, Zaitlen, Reich, & Patterson, 2010) but can also cause Ethnicity × Environment interactions to appear as spurious Gene × Environment interactions. This is because the frequency of alleles naturally varies across ethnic populations and, in the presence of a coincidental excess of affected individuals belonging to one ethnic group, spurious associations and interactions with polymorphisms of no functional consequence, except a degree of natural ethnic variation, may emerge.

Power to detect and characterize different types of interactions. Yet another concern is the low power to detect most plausible forms of interactions in the first place (McClelland & Judd, 1993) in observational field studies as compared with experimental studies in which independent variables can be efficiently manipulated. Under many conditions, theoretically meaningful interactions are likely to be quite small, accounting for 1% of the outcome variance, and the power to detect most plausible interactions will be quite limited without a large *N*.

Further, even if an interaction is detected, discerning the true pattern of an interaction from observed results is even more tenuous. In recent years, there has been great interest in determining the form of the observed interaction in cG×E research because the interpretation of disordinal (i.e., cross-over) interactions theoretically differs from ordinal interactions. Specifically, cross-over interactions lend themselves to a differential susceptibility interpretation in which a given "risk" or "malleable" allele is associated with both poorer outcomes in a "bad" environment but better outcomes in a "good" environment; ordinal interactions lend themselves to a diathesis-stress interpretation in which it is the combination of a risk-conferring allele and a bad environment that exacerbates the likelihood of manifesting the outcome (e.g., Belsky, Bakermans-Kranenburg, & van Ijzendoorn, 2007; Belsky et al., 2009). However, simulations demonstrate that, conditional on a Type I error, the form of an ostensibly "significant" interaction is usually of a cross-over (i.e., disordinal) nature, especially when samples sizes are small (Sher & Steinley, 2013). Boardman et al. (2014) recently made a similar observation in reference to the emerging genome-wide, gene-by-environment approach (Cornelis et al., 2012; Mukherjee, Ahn, Gruber, & Chatterjee, 2012; Thomas, Lewinger, Murcray, &

Gauderman, 2012) by demonstrating that when many interaction tests are performed, the most significant *p* values will come from disordinal interactions even when such interactions are generated from random data. Boardman et al. (2014) noted that these findings "conform to the differential susceptibility model but will not tell us anything meaningful about the way in which environments systematically moderate genetic factors . . . because they will likely be a statistical artifact" (p. 123).

Moreover, even true ordinal interactions that are statistically significant can appear to be of a cross-over type (Sher & Steinley, 2013) because of random error. All linear interactions imply a cross-over at some point, even if outside the range of observed values. This is not a trivial issue because a typical practice is to plot values +1 standard deviation and −1 standard deviation above and below the mean of the moderator (Aiken & West, 1991). However, −1 standard deviation can represent values that rarely or never exist in nature for skewed predictors. We note that some authors (Roisman et al., 2012) have recently recommended extending Aiken and West's (1991) guidelines to ±2 standard deviations to provide 95% coverage of the observed values. Given the highly skewed nature of many environmental exposures, attention to the underlying distribution of all study constructs is necessary so as not to generate misleading regression plots covering regions of sparse or imaginary data.

Recently, techniques for estimating the standard error of the cross-over point have been proposed that could, in principle, allow stronger inferences about the actual form of the interaction (Widaman et al., 2012). Alternatively, establishing significant regions around each regression line with standing approaches (e.g., Johnson & Neyman, 1936) to characterize where two slopes overlap and where they do not could also be used to increase confidence that an ostensible cross-over shows a desired degree of statistical differentiation from an ordinal interaction. Such approaches, in principle, could provide greater confidence in believing a true cross-over has been detected. Consistent with the other points made earlier, such approaches depend on being confident that the interaction is not an artifact of scaling, is not caused by (unmodeled) nonlinearity, and is not a Type I error.

cG×E versus gene–environment correlation (rGE). The term rGE refers to instances in which exposure to environment is nonrandom and correlated with genetic vulnerability,[3] whether through passive, active, or evocative processes.[4] For instance, in classical behavior genetics, rGE is represented by genetic factors that influence the outcome (e.g., alcohol and tobacco use) and the environment (e.g., peer relationships; Harden et al., 2008) as indexed by a genetic correlation. Similarly, in measured gene studies, presence of rGE is indexed by variations in genotype or allelic frequency as a function of the environmental exposure. For instance, Salvatore et al. (2014) reported an association between a polygenic score for alcohol problems and peer deviance, indicating that individuals who are at genetic risk for alcohol problems are also more likely to have deviant peer groups. It is likely that for many outcomes both rGE and cG×E may be important; however, the presence of rGE can complicate the interpretation of cG×E. Researchers using behavioral genetic and twin models implement several statistical approaches to account for and even explicitly model rGE (e.g., Eaves & Erkanil, 2003; Purcell, 2002; van der Sluis, Posthuma, & Dolan, 2012). In measured gene studies, the first step in testing for potential rGE involves estimation of a correlation between genotype and environment. In the absence of such a correlation, as has been noted for 5-HTTLPR and stressful life events, cG×E testing can proceed without concern. If the correlation is solely attributable to outliers in the environmental measure, removal or Winsorization may eliminate rGE (e.g., Bogdan, Williamson, & Hariri, 2012). A modest correlation between genotype and environment may require more careful consideration. For instance, in revisiting the interaction between a polymorphism in the *MAOA* gene and exposure to childhood physical abuse in the development of antisocial behaviors (Caspi et al., 2002), Kim-Cohen et al. (2006) examined whether the *MAOA* genotype was correlated with not only exposure to abuse (i.e., evocative rGE) but also to maternal antisocial behavior (i.e., passive rGE), with the latter being a key correlate of transmission of risk for antisociality and for increased likelihood of exposure to abuse. There was evidence for the latter, whereby maternal antisocial behavior was correlated with offspring exposure to abuse; however, the effect of the interaction persisted even after accounting for this effect. Alternatively, Salvatore et al. (2014) accounted for rGE by residualizing both their polygenic score and their environmental measures (parental knowledge and peer deviance) for each other prior to testing for cG×E in the etiology of alcohol use problems. The interaction between parental knowledge and the polygenic scores remained significant; however, the interaction with peer deviance was no longer significant, indicating the possibility of both rGE and cG×E for the former but rGE alone for the latter. Therefore, although both mechanisms of gene–environment interplay (rGE and G×E) may be at work, testing for rGE is necessary before conclusions regarding G×E are made. When rGE is presented, methods to account for it should be implemented. In some instances, relevant data may not be available (e.g., availability of parental phenotypes to test for passive rGE), or the correlation may be complex and mediated by other unmeasured factors. In such instances, the possibility that rGE

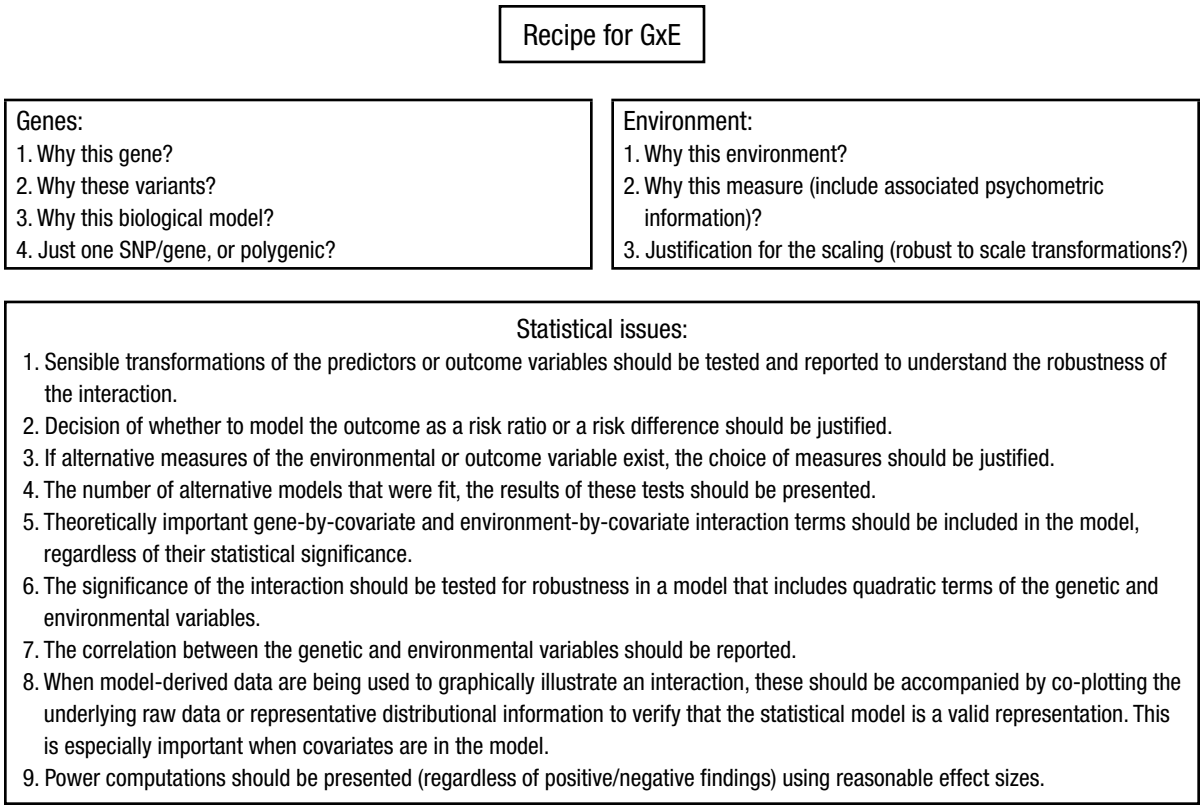

Fig. 4. Recommendations for rigorous gene–environment interaction (G×E) research practices. SNP = single nucleotide polymorphism.

may contribute to the relationship between genotype and environment should be acknowledged.

In summary, although the statistical approaches for modeling interactions are well established, having confidence in the statistical validity of an interaction requires due diligence on the part of the investigator. These include attention to scaling issues, characterizing the underlying linearity of the relationships under investigation and determining whether nonlinear models are necessary, controlling for relevant confounders including various forms of rGE, and ensuring plotted results are not unduly influenced by the constraints imposed for rendering an easy-to-interpret graph. Perhaps the greatest challenge is to minimize the likelihood that an observed interaction is not a Type I error given that various data sets have a large number of candidate Gs and candidate Es; there is considerable flexibility in approach to analysis; and under most plausible conditions, power to detect G×E is likely to be low. Many of the issues described earlier (as well as some others) are described by Roisman et al. (2012), who provided a list of thoughtful guidelines for addressing various issues, such as characterizing whether an obtained interaction is of a cross-over type (and to the extent that the magnitude of the cross-over is meaningful), the problem of nonlinearity, and Type I errors.

Recommendations

Although the list of challenges associated with characterizing cG×E is long, many of these can be addressed by adopting a handful of rigorous research practices. Later in the article, we delineate a series of recommendations that we believe will help advance the study of G×E and ensure that the literature provides meaningful progress for science. We focus at greatest length on the issues pertaining to the G in G×E, under the assumption that this will be most unfamiliar to social scientists. Many of the other concerns, reviewed in less detail later in the article, are not specific to the study of cG×E but complement the broader discussion (Cumming, 2014; Ioannidis, 2005; Lakens & Evers, 2014; Simmons et al., 2011) in the social sciences about how to produce robust, replicable findings that advance science. The recommendations described in this article are reviewed in Figure 4.

Selection of genes

Given the prevailing skepticism surrounding candidate gene research, the burden of proof for the selection of a candidate gene is high. Such a rationale should be convincingly articulated in a manner specific to the

phenotype and environment under study. The crux of the argument for selection of a particular gene to study lies in the statistical prior probabilities for the gene, that is, on the basis of prior evidence and the quality of the source of that evidence, how likely is it that this is a robust candidate. There is nothing inherently wrong with studying candidate genes, though the very idea of candidate gene research has fallen out of favor because of the historical issues with studying hypothesized biological candidates that have not held up in more systematic well-powered studies, as reviewed earlier. There are notable cases in which hypothesized biological candidates have shown robust associations with outcome. For instance, rs1229984 in the *ADH1B* gene was one of the earliest candidate gene variants proposed in the etiology of alcoholism, particularly in Asians (Agrawal & Bierut, 2012).[5] The variant, which is rather rare in European American populations, has recently been identified in adequately powered GWAS as well (Bierut et al., 2012). Similarly, early candidate gene and experimental studies implicated a SNP in a nicotinic receptor (rs16969968 in *CHRNA5*) as a risk factor for tobacco smoking and nicotine dependence, but the role of this variant in cigarette smoking was not widely accepted until it was identified in multiple meta-analyses of cigarettes smoked per day, the largest of which had a sample size exceeding 70,000 (J. Z. Liu et al., 2010; Thorgeirsson et al., 2010; Tobacco and Genetics Consortium, 2010).

Knowing that the prior probabilities for a candidate gene selected on hypothesized biological rationale is low, there are a variety of other methods for selecting candidate genes for study that should produce more robust and reliable findings, including focusing on genes with either large main effects or stronger a priori evidence. For the latter, methods of gene selection that have a greater likelihood of producing meaningful cG×E results include focusing on candidates suggested by well-powered GWAS or meta-analyses, or by model organism work, ideally with replication. Identification of genotypes of substantial main effect is challenging and limited. For example, the ε2/ε3/ε4 polymorphism at the Apolipoprotein E locus is a known and important risk factor for Alzheimer's and coronary heart disease (Corder et al., 1993; Ward et al., 2009). Such large-effect polymorphisms seem to be particularly compelling candidates for cG×E research because the genetic effect is already known; only an environmental modification of this effect is required for evidence for an interaction.

A systematic strategy of gene identification followed by efforts to characterize moderation of the effect associated with that gene can be found in the example of *GABRA2* and moderation by parental monitoring. *GABRA2* was originally identified by the Collaborative Study on the Genetics of Alcoholism, the largest gene identification project in the area of alcohol dependence (Begleiter et al., 1995), by systematically interrogating gamma-aminobutyric acid receptor genes with evidence for involvement in ethanol-related responses (Harris, 1999) that were located in a region of linkage identified with both clinical alcohol dependence phenotypes and electrophysiological endophenotypes (Ghosh et al., 2003; Porjesz et al., 2002; Reich et al., 1998). Association between *GABRA2* and alcohol dependence was subsequently reported (Edenberg et al., 2004) and replicated by multiple independent research groups around the world using a variety of research designs (Covault, Gelernter, Hesselbrock, Nellissery, & Kranzler, 2004; Enoch, Schwartz, Albaugh, Virkkunen, & Goldman, 2006; Fehr et al., 2006; Lappalainen et al., 2005; Olfson & Bierut, 2012; Soyka et al., 2008). Although there have been failures to replicate (Covault, Gelernter, Jensen, Anton, & Kranzler, 2008; Drgon, D'Addario, & Uhl, 2006; Matthews, Hoffman, Zezza, Stiffler, & Hill, 2007), a recent meta-analysis has confirmed the evidence for association (Li et al., 2014). In addition, translational researchers have found the role of *GABRA2* in rodent drinking (Dixon, Walker, King, & Stephens, 2012; J. Liu et al., 2011) and in the brain's response to alcohol-related (Kareken et al., 2010) and monetary reward cues (Villafuerte et al., 2012). Although *GABRA2* SNPs have not been identified via GWAS, they typically have the lowest *p* values of candidate polymorphisms extracted from GWAS data (Olfson & Bierut, 2012). Interaction between *GABRA2* and parental monitoring was tested on the basis of the twin literature suggesting that parental monitoring moderates the relative importance of overall genetic effects (as inferred on the basis of comparisons of twins, not with measured genotypes) on substance use outcomes in adolescence (Dick, Pagan, Viken, et al., 2007; Dick, Viken, et al., 2007); genetic effects assumed greater importance under conditions of lower parental monitoring, presumably because adolescents with lower monitoring have more access to the substance and opportunity to express their genetic predisposition. An interaction between *GABRA2* and parental monitoring was tested in an independent sample; a stronger association between the gene and trajectories of externalizing behavior was found under conditions of lower parental monitoring, as was hypothesized on the basis of the twin findings (Dick, Latendresse, et al., 2009). This series of studies illustrates how different literatures and study designs (linkage and association studies; twin studies; and longitudinal, developmental studies) were integrated to characterize a cG×E effect. In this case, the candidate gene under study was selected on the basis of a series of converging pieces of evidence that indicated involvement in alcohol and externalizing outcomes before it was studied in the context of cG×E.

In short, not all candidate genes are created equal, and there is not a single pathway to determining their viability in a cG×E study. The burden is on the researcher to

provide a compelling argument for the study of a particular candidate. Furthermore, we suggest that the bar should be much higher than the field has insisted on to date.

Selection of genetic variants

In addition to strong justification for the selection of the gene under study, a second area that should be justified is the genotyping strategy: If only a select number of polymorphisms in a gene are being genotyped, how and why were those selected? There are several methods for selecting polymorphisms—one can utilize SNP content from a preexisting GWAS array or genotype custom content individually or en masse (e.g., as offered by the Illumina Golden Gate technology; Hodgkinson et al., 2008). Prices for GWAS arrays, especially those designed to include custom content, have dropped considerably, and often it is far more expensive and less cost-efficient to genotype a small number of polymorphisms than to genotype in large scale. In some cases, the technology required to genotype a particular variant, especially one that is not a SNP (e.g., a variable number of **tandem repeats**), is specialized and may entail unique requirements, as most commercial arrays and custom genotyping platforms may not include this. Nonetheless, if only a few SNPs can be genotyped, *tagging* a gene may be preferable to simply pursuing the usual suspects. Tagging refers to identifying all variation, regardless of function, that captures variation across the gene. This includes important regulatory regions, such as **promoters** and **enhancers**, that are increasingly recognized as key contributors (Zannas & Binder, 2014). Reliance on simple annotation of "function" (i.e., typically a **nonsynonymous exonic** variant, meaning that the location is known to be in a part of the gene that produces an alteration in the gene's protein product) is short-sighted, as modern annotations available via the identification of **epigenetic** marks along the genome suggest that even **intronic** variants can have a profound impact on genomic action (e.g., Ziller et al., 2013). An exciting and upcoming possibility is a highly cost-efficient chip being designed by the Psychiatric Genomics Consortium that will include custom common and rare variation and will be based on SNPs nominated by expert consensus and validated via meta-analytic methods.

Modeling genes

Another consideration is that how genotype is coded implies a biological model, as reviewed earlier, and, thus, the model should be specified and justified (or explicitly stated as exploratory, with appropriate corrections for multiple testing). Genotypes should not be collapsed purely to increase power, and if they are, then effects should be described across all genotype groups for completeness. With increasing recognition that individual genetic polymorphisms on their own are likely to have very small effects, there should be justification provided if single polymorphisms are being studied in isolation. Further, with growing knowledge about gene networks and integrated functional pathways, there is opportunity to think more broadly about the potential for incorporating gene networks or pathways into tests of G×E, allowing one to go beyond focusing on a single gene and instead focus on sets of genes that interact biologically. This brings its own set of complications, as decisions must be made about the nature of genetic effects across the pathway or network: Is a mutation in any of the genes in the network sufficient? Are all variations within the network expected to have an equal effect on outcome? Are mutations across multiple genes in the network acting cumulatively to affect outcome? Similar questions can be asked about multiple variants within any given gene of interest. Should **polygenic risk scores** that capture risk across the genome be used? The answers to these questions depend on the investigator's theory behind how the environment is operating: Does the investigator believe that all genes involved in outcome should be moderated by that environment in a parallel fashion or only subsets of the relevant genes? There is no straightforward answer to these questions, but what is clear is that deeper thought must be given to these issues to move the study of cG×E forward. Justification for the choices made in any given G×E study should be included in the publication. Because it is challenging to keep up with advances in the field of genetics, we suggest that this is an area in which collaborations between geneticists and psychologists can be particularly fruitful, as connection to the latest findings from statistical and psychiatric genetics about the rapidly evolving knowledge of the underlying genetic basis for a given outcome of interest can ensure that the genetics being integrated into psychological science represents the latest advances from the field of genetics. The annual meetings of the Behavior Genetics Association (www.bga.org) and the International Society for Psychiatric Genetics (www.ispg.net) are opportunities to learn about the latest advances in psychiatric and behavioral genetics and to potentially develop collaborations with scientists working directly in those areas. In addition, Text Box 1 of this review (in the Supplemental Material) lists a variety of genetics resources that may be useful to investigators in the social sciences that are interested in adding an informed genotyping component to their study.

Selection of the environment

Investigators should provide theoretical rationale for the selection of the environmental factor under study. Why is

there reason to believe that this aspect of the environment will have a moderating effect? Does the investigator believe this is an environmental factor with a time-limited or enduring effect? Justification for the scale of measurement of the environment should be provided as well as reporting of results on biologically defensible transformations of the independent and dependent variables. The environmental measure should reflect the "state of the art" method for measuring the particular feature of the environment. If it does not, justification for why this particular measure should be relevant/adequate should be included as well as traditional supportive psychometric information.

Accurate reporting of multiple testing

Researchers should explicitly detail how many total polymorphisms were available to them, how many were tested, and what types of guarding against inflation of Type I errors were made. In an identical vein, researchers should explicitly detail how many environmental variables, and how many different methods of operationalizing these environmental variables, were considered, as well as transformations of such, and selection of models. These types of procedures for genes are now routine in GWAS but are rare in cG×E studies. Such explicit disclosures will help increase confidence in positive findings that survived proper multiple testing corrections.

Power and small samples

Investigators should demonstrate that the sample being used has adequate power to detect an interaction effect for the variables under study. Power computations should be presented regardless of the nature of the finding. In other words, researchers that produce studies with positive findings should be encouraged to present power computations as well (K. S. Button et al., 2013). Computations should be specific to the statistical technique, distributions of the variables under investigation, and the hypothesized form of the interaction. Power analyses should assume realistic cG×E effect sizes. GWAS suggest that main effects are typically of a small magnitude, most accounting for less than 1% of the variance in a psychiatric phenotype. If the investigator has reason to believe that a larger effect size is likely for their study, this should be clearly spelled out and justified. For instance, one might hypothesize that genes exert stronger effects on endophenotypes (e.g., neuroimaging outcomes) that are, arguably, more proximal to their action (though see Flint & Munafo, 2007; Munafo & Flint, 2009). Or, it is possible that the use of a phenotypic measure thought to be of much greater reliability and validity than existing clinical phenotypes could enhance power. Finally, investigators should use meta-analysis and/or integrative data analysis (e.g., Hussong, Curran, & Bauer, 2013) to combine data across multiple studies so that small-scale research can contribute to more definitive results.

A question that sometimes arises is what can and should be done with studies whose small sample size is necessitated by the prohibitive costs associated with measurement of the phenotype, for example, with behavioral or neuroimaging studies, multiwave longitudinal studies, or studies of special populations. Regardless of the nature of the variables (e.g., self-report vs. neuroimaging), a small sample size nearly always implies reduced power. In fact, K. S. Button and colleagues (2013) reported that the median statistical power across 49 meta-analyses of neuroscience studies was typically less than 20%, with the estimate plummeting to 8% when examining neuroimaging studies alone—and that is without estimation of G or G×E (K. S. Button et al., 2013). Some have argued that genotypic effect sizes associated with endophenotypes (such as neuroimaging outcomes) are likely to be higher because the outcome is more proximal to the biological substrate (E. J. Rose & Donohoe, 2013); yet, there is no demonstrable evidence that researchers understand the genetic underpinnings of threat-related amygdala reactivity and habituation any better than they do the etiology of depression and anxiety. In other words, endophenotypes may be as polygenic as self-report measures (Flint & Munafo, 2007).

Replication is key for findings based on small samples, but, as Button and colleagues have noted, the winner's curse often inflates initial results, and unless the replication sample is substantially better powered, the ceiling placed on average sample size likely perpetuates false-positive findings until enough samples have been amassed to conduct a meta-analysis (K. S. Button et al., 2013). Meta-analysis and integrative data analysis (Hussong, Curran, & Bauer, 2013) are particularly attractive techniques to combine data across multiple studies so that small scale research can contribute to more definitive results. In instances in which that is impossible—perhaps because the sample is rare and unique, because the outcome under study is highly novel, or because the environmental factors have been measured with superior assessments that other studies do not include—we recommend that researchers acknowledge their limitations and present a candid account of power in the study, even if it implies that their findings might be spurious. Readers of this literature should carefully consider the caveats of such small sized studies so as not to perpetuate a series of cG×E emerging from and replicated in small samples.

Manipulation/presentation of data

The statistical properties that surround the detection of the interaction should be specified and discussed. Is the

interaction significant on an additive or multiplicative scale? Investigators should provide evidence that detected interaction effects are robust to transformation of scale and should provide a strong defense of the scaling used. If it is not robust to transformation, the implications of this should be discussed. Common artifacts associated with nonlinearity should also be evaluated and ruled out. Similarly, if the interaction is detected under some conditions or with some measures of the environment but not others, this should be reported and incorporated into the theory behind what the findings contribute to the literature. Perhaps most important, the raw data along with 95% confidence intervals should always be plotted and reported in the publication, not just the model-derived regression lines, which can be misleading, especially when some combinations of the independent variable have small sample sizes and when control covariates are included in the model. Because apparent cG×E effects can be due to stratification, genotyping artifacts, and even gender (Keller, 2014), careful attention must be paid to the variables under study to bolster confidence that the moderation effect is due to the specific gene/environment under study. For example, if White individuals are more sensitive to environmental trauma, leading to posttraumatic stress disorder, in a mixed ethnicity sample examining environmental Trauma × Gene interactions, any SNP that differentiates White individuals from Black individuals will also show an apparent cG×E. In this case, the gene is not the true moderator—race is; the gene was merely correlated with race. To increase confidence in cG×E findings and to eliminate alternative explanations for them, researchers need to include all relevant Gene × Covariate and Environment × Covariate interaction terms in their models. Researchers should consider including quadratic transformations of the environmental term as well as covariates, as these may change the interpretation of the interaction depending on multi-collinearity. Finally, the assumptions that are imposed by modeling interactions with cross-product terms in linear regression or by the use of logistic regression (as discussed earlier) should be acknowledged and, when appropriate, justified.

Addressing publication bias

In circumstances in which the proportion of false findings is high, as we suspect is the case for candidate gene main effect and cG×E studies, the issue of publication bias should be considered by both authors and editors. Authors, journal editors, and reviewers should understand that novel positive findings may often be false, and so they should set higher standards of evidence, including stricter standards of significance testing, full accounting for all sources of multiple testing, and direct independent replication prior to publication. Equally, investigators should strive to submit, and editors to publish, adequately powered, independent direct replication attempts, irrespective of the outcome—positive, negative, or null. Meta- or mega-analyses that combine data across multiple studies while accounting for heterogeneity allow smaller sized studies to contribute to hypotheses necessitating larger samples, such as cG×E. Such collaborative research should be encouraged. These kinds of recommendations, aimed at reducing the role of publication and other biases, are being adopted by journals that have an interest in cG×E; some recent examples are *Behavior Genetics* (Hewitt, 2012), the *Journal of Abnormal Child Psychology* (Johnston, Lahey, & Matthys, 2013), and *Drug and Alcohol Dependence* (Munafo & Gage, 2013). We would recommend that other journal editors whose journals deal with substance abuse, psychiatry, and related fields adopt similar policies.

Conclusions

Studying how genetic predispositions and environmental circumstances come together to contribute to complex behavioral outcomes has great potential for advancing the understanding of the development of psychopathology. It represents a clear theoretical advance over studying these factors in isolation. However, research at the intersection of multiple fields creates many challenges. Studying cG×E requires appropriate understanding of genetic mechanisms, appropriate measurement of the environment, as well as a conceptual framework for integrating the two with respect to a specific outcome of interest and, critically, of the statistical principles that underpin cG×E studies. It is not likely, or expected, that every investigator conducting cG×E research will have the requisite expertise in all of these areas; accordingly, we encourage cG×E studies that are collaborative efforts involving individuals with common interests but diverse expertise.

The National Institutes of Health has initiatives aimed at addressing challenges associated with genotypic research, many of which are also relevant to the study of cG×E. For example, in recognition of the fact that gene identification requires large numbers, but one of the challenges that is often encountered as researchers attempt to pool their samples is the use of different measures across studies, the PhenX initiative was launched (www.phenX.org; Hamilton et al., 2011). PhenX brought together panels of experts across a variety of research areas to come up with recommended consensus measures (including both outcome and environmental measures) for inclusion in genetics studies to encourage the use of common measures to facilitate cross-study comparisons and analyses. There are limitations to this approach: Brief, low-burden measures were preferentially selected to encourage more widespread uptake, which may result in less precise or comprehensive

assessments in the case of some constructs. However, it represents a step toward facilitating collaborative efforts in genetics research. The use of standardized measures across studies could also help advance the cG×E field, with greater emphasis placed on replication and combined analyses across research groups to enhance sample size and corresponding power. The Office of Behavioral and Social Sciences Research at the National Institutes of Health also has resources on its website (http://obssr.od.nih.gov/) to help social scientists with the incorporation of genetic information into their studies. Many of these issues are not specific to the study of psychopathology. In a recent article on G×Es in cancer epidemiology, coming out of a National Cancer Institute think tank, Hutter, Mechanic, Chatterjee, Kraft, and Gillanders (2013) described similar challenges.

As investigators who have explored cG×E hypotheses ourselves, with some of our own work not meeting the standards delineated in this review, we were compelled to ask "how can we do better?" We hope we have outlined some such strategies. Through greater awareness of the challenges in conducting cG×E research, resources available to aid in conducting high-quality cG×E studies, and proactive efforts to move cG×E studies in this direction, it is our hope that this growing area of research will eventually reach its potential to deeply inform to the understanding of complex behavioral outcomes.

Author Contributions

All authors contributed to writing sections of the article and to reviewing drafts.

Acknowledgments

This article developed out of discussions from the workshop "Challenges and Opportunities in G×E Research: Creating Consensus Recommendations for Guidelines on G×E Research," sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in January 2013. In addition to the authors who contributed to the writing of this article, other workshop participants included Lindon Eaves (Virginia Commonwealth University), Andrew Heath (Washington University in St. Louis), and Eric Turkheimer (University of Virginia). We are deeply grateful for the leadership of Marcia S. Scott (chair of the NIAAA Gene and Environment Team), Dionne C. Godette, and Mariela C. Shirley—who are affiliated with the Division of Epidemiology and Prevention Research, NIAAA (Mariela C. Shirley is now at the Office of Research on Women's Health), National Institutes of Health—in organizing support for the workshop and for their helpful comments on drafts of this article.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

Support for the authors involved in writing this article includes the following: K02AA018755, R01DA070312, and U10AA008401 to Danielle M. Dick; R01AA018333 to Danielle M. Dick and Kenneth S. Kendler; K01MH085812 and R01MH100141 to Matthew C. Keller; K02DA32573, R21AA21235, and R01DA23668 to Arpana Agrawal; DA011015 to John K. Hewitt; and K05 AA017242 to Kenneth J. Sher. We are also grateful to Erica Spotts, Health Scientist Administrator at the Office of Behavioral and Social Science Research (OBSSR), for her assistance in securing funding (R01 AA018333-04S1 to Danielle M. Dick) to create a website to facilitate understanding of the concepts related to conducting G×EcG×E research, as delineated in this article. This information will be hosted through the OBSSR website (anticipated 2015 release).

Supplemental Material

Additional supporting information may be found at http://pps.sagepub.com/content/by/supplemental-data

Notes

1. Words in bold font in the text are defined in the Supplemental Material.
2. This is often called LD-tagging and indicates that you can use information about the LD or correlation structure across variants within and across genes to know how many genetic variants you need to genotype to capture most of the variable locations in the gene that could be associated with outcome (see Text Box 1 in the Supplemental Material for software).
3. Mendelian randomization is a related and novel conceptualization of genotype representing environmental exposure propensity and might explain cG×E in the presence of rGE. We refer the reader to Smith (2010) to learn more.
4. Passive rGE is more common during childhood and adolescence and refers to individuals being differentially exposed to an environment without their own initiative, most likely because aspects of their environment are provided by their parents with whom they also share genetic variance. For instance, antisocial parents may pass on a genetic liability to antisociality and expose their children to an abusive environment. Evocative rGE refers to environments that an individual elicits/evokes from others on the basis of his or her genetic predisposition. For instance, an antisocial adolescent may evoke harsher parenting. Finally, as an individual matures into adulthood, the importance of active rGE increases. Here, an individual actively selects his or her environment, or niche, on the basis of characteristics of his or her genetic predisposition. For example, antisocial youths may select into high-risk neighborhoods or affiliate with deviant peers.
5. The *ADH1B* polymorphism (rs1229984) has been implicated in Asian populations to afford protection against the development of alcoholism, putatively via flushing (facial reddening due to accelerated conversion of ethanol to acetaldehyde); rs1229984 was not originally detected in GWAS on European American populations. The minor allele frequency of this variant in European Americans is low (<5%), and commercial GWAS arrays rely on common variation. Accordingly, the SNP was neither genotyped on these arrays nor could it be reliably

imputed. However, genotyping this polymorphism resulted in genome-wide significant association signals (Bierut et al., 2012), and with the advent of more recent arrays that target this variant better, there is now evidence in GWAS of an association between rs1229984 and alcoholism at $p = 1.2 \times 10^{-31}$ (Gelernter et al., 2014). In this instance, GWAS did not initially guide identification of a logical and validated candidate gene, highlighting that all techniques have their strengths and weaknesses.

References

Agrawal, A., & Bierut, L. (2012). Identifying genetic variation for alcohol dependence. *Alcohol Research*, *34*, 274–281.

Agrawal, A., Freedman, N. D., Cheng, Y. C., Lin, P., Shaffer, J. R., Sun, Q., . . . Bierut, L. J. (2012). Measuring alcohol consumption for genomic meta-analyses of alcohol intake: Opportunities and challenges. *American Journal of Clinical Nutrition*, *95*, 539–547. doi:10.3945/ajcn.111.015545

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Aliev, F., Latendresse, S. J., Bacanu, S., Neale, M., & Dick, D. M. (2014). Testing for measured gene–environment interaction: Problems with the use of cross-product terms and a regression model reparameterization solution. *Behavior Genetics*, *44*, 165–181. doi:10.1007/s10519-014-9642-1

Begleiter, H., Reich, T., Hesselbrock, V., Porjesz, B., Li, T. K., Schuckit, M., . . . Rice, J. (1995). The collaborative study on the genetics of alcoholism. *Alcohol Health & Research World*, *19*, 228–236.

Belsky, J., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current Directions in Psychological Science, 16*, 300–304.

Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry, 14*, 746–754.

Bergeman, C. S., & Plomin, R. (1989). Genotype-environment interaction. In M. H. Bornstein & J. S. Bruner (Eds.), *Interaction in human development* (pp. 157–171). Hillsdale, NJ: Erlbaum.

Bierut, L., Goate, A., Breslau, N., Johnson, E., Bertelsen, S., Fox, L., . . . Edenberg, H. (2012). ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Molecular Psychiatry*, *17*, 445–450. doi:10.1038/mp.2011.124

Blum, K., Noble, E. P., Sheridan, P. J., Montgomery, A., Ritchie, T., Jagadeeswaran, P., . . . Cohn, J. B. (1990). Allelic association of human dopamine D2 receptor gene in alcoholism. *Journal of the American Medical Association*, *263*, 2055–2060.

Boardman, J. D. (2009). State-level moderation of genetic tendencies to smoke. *American Journal of Public Health*, *99*, 480–486.

Boardman, J. D., Domingue, B. W., Blalock, C. L., Haberstick, B. C., Harris, K. M., & McQueen, M. B. (2014). Is the gene–environment interaction paradigm relevant to genome-wide studies? The case of education and body mass index. *Demography*, *51*, 119–139. doi:10.1007/s13524-013-0259-4

Bogdan, R., Williamson, D., & Hariri, A. R. (2012). Mineralocorticoid receptor Iso/Val (rs5522) genotype moderates the association between previous childhood emotional neglect and amygdala reactivity. *American Journal of Psychiatry*, *169*, 515–522. doi:10.1176/appi.ajp.2011.11060855

Boraska, V., Franklin, C. S., Floyd, J. A., Thornton, L. M., Huckins, L. M., Southam, L., . . . Builk, C. M. (2014). A genome-wide association study of anorexia nervosa. *Molecular Psychiatry*, *19*, 1085–1094. doi:10.1038/mp.2013.187

Bosker, F. J., Hartman, C. A., Nolte, I. M., Prins, B. P., Terpstra, P., Posthuma, D., . . . Nolen, W. A. (2011). Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Molecular Psychiatry*, *16*, 516–532. doi:10.1038/mp.2010.38

Brown, G. W., & Harris, T. O. (2008). Depression and the serotonin transporter 5-HTTLPR polymorphism: A review and a hypothesis concerning gene–environment interaction. *Journal of Affective Disorders*, *111*, 1–12. doi:10.1016/j.jad.2008.04.009

Button, K. S., Ioannidis, J. P., Morkrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376. doi:10.1038/nrn3475

Button, T. M., Lau, J. Y., Maughan, B., & Eley, T. C. (2008). Parental punitive discipline, negative life events and gene–environment interplay in the development of externalizing behavior. *Psychological Medicine*, *38*, 29–39.

Button, T. M., Stallings, M. C., Rhee, S. H., Corley, R. P., Boardman, J. D., & Hewitt, J. (2009). Perceived peer delinquency and the genetic predisposition for substance dependence vulnerability. *Drug & Alcohol Dependence*, *100*, 1–8.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., . . . Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*, 851–854.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., . . . Poulton, R. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*, 386–389.

Clarke, H., Flint, J., Attwood, A. S., & Munafo, M. R. (2010). Association of the 5-HTTLPR genotype and unipolar depression: A meta-analysis. *Psychological Medicine*, *40*, 1767–1778. doi:10.1017/S0033291710000516

Colhoun, H. M., McKeigue, P. M., & Davey Smith, G. (2003). Problems of reporting genetic associations with complex outcomes. *Lancet*, *361*, 865–872.

Collins, A. C., Kim, Y., Sklar, P., International Schizophrenia Consortium, O'Donovan, M., & Sullivan, P. F. (2012). Hypothesis driven candidate genes for schizophrenia compared to genome-wide association results. *Psychological Medicine*, *42*, 607–616.

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., . . . Pericak-Vance, M. A. (1993). Gene does of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, *261*, 921–923.

Cornelis, M. C., Tchetgen, E. J., Liang, L., Qi, L., Chatterjee, N., Hu, F. B., & Kraft, P. (2012). Gene–environment interactions in genome-wide association studies: A comparative study of tests applied to empirical studies of type 2

diabetes. *American Journal of Epidemiology*, *175*, 191–202. doi:10.1093/aje/kwr368

Corvin, A., Craddock, N., & Sullivan, P. F. (2010). Genome-wide association studies: A primer. *Psychological Medicine*, *40*, 1063–1077. doi:10.1017/S0033291709991723

Covault, J., Gelernter, J., Hesselbrock, V., Nellissery, M., & Kranzler, H. R. (2004). Allelic and haplotypic association of GABRA2 with alcohol dependence. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 129B*, 104–109.

Covault, J., Gelernter, J., Jensen, K., Anton, R., & Kranzler, H. R. (2008). Markers in the 5′-region of GABRG1 associate to alcohol dependence and are in linkage disequilibrium with markers in the adjacent GABRA2 gene. *Neuropsychopharmacology*, *33*, 837–848.

Craddock, N., Dave, S., & Greening, J. (2001). Association studies of bipolar disorder. *Bipolar Disorders*, *3*, 284–298.

Culverhouse, R. C., Bowes, L., Breslau, N., Nurnberger, J., Burmeister, M., & Fergusson, D. M. . . . 5-HTTLPR, Stress, and Depression Consortium. (2013). Protocol for a collaborative meta-analysis of 5-HTTLPR, stress, and depression. *BMC Psychiatry*, *13*, 304. doi:10.1186/1471-244X-13-304

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7–29. doi:10.1177/0956797613504966

Dick, D. M., Bernard, M., Aliev, F., Viken, R., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2009). The role of socioregional factors in moderating genetic influences on early adolescent behavior problems and alcohol use. *Alcoholism: Clinical and Experimental Research*, *33*, 1739–1748. doi:10.1111/j.1530-0277.2009.01011.x

Dick, D. M., Latendresse, S. J., Lansford, J. E., Budde, J., Goate, A., Dodge, K. A., . . . Bates, J. (2009). Role of GABRA2 in trajectories of externalizing behavior across development and evidence of moderation by parental monitoring. *Archives in General Psychiatry*, *66*, 649–657. doi:10.1001/archgenpsychiatry.2009.48

Dick, D. M., Pagan, J. L., Holliday, C., Viken, R., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2007). Gender differences in friends' influences on adolescent drinking: A genetic epidemiological study. *Alcoholism: Clinical and Experimental Research*, *31*, 2012–2019. doi:10.1111/j.1530-0277.2007.00523.x

Dick, D. M., Pagan, J. L., Viken, R., Purcell, S., Kaprio, J., Pulkkinen, L., & Rose, R. J. (2007). Changing environmental influences on substance use across development. *Twin Research and Human Genetics*, *10*, 315–326. doi:10.1375/twin.10.2.315

Dick, D. M., Rose, R. J., Viken, R. J., Kaprio, J., & Koskenvuo, M. (2001). Exploring gene–environment interactions: Socioregional moderation of alcohol use. *Journal of Abnormal Psychology*, *110*, 625–632.

Dick, D. M., Viken, R., Purcell, S., Kaprio, J., Pulkkinen, L., & Rose, R. J. (2007). Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. *Journal of Abnormal Psychology*, *116*, 213–218. doi:10.1037/0021-843x.116.1.213

Dixon, C. I., Walker, S. E., King, S. L., & Stephens, D. N. (2012). Deletion of the gabra2 gene results in hypersensitivity to the acute effects of ethanol but does not alter ethanol self-administration. *PLoS ONE*, *7*(10), e47135. doi:10.1371/journal.pone.0047135

Drgon, T., D'Addario, C., & Uhl, G. R. (2006). Linkage disequilibrium, haplotype and association studies of a chromosome 4 GABA receptor gene cluster: Candidate gene variants for addictions. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *141B*, 854–860.

Duncan, L., & Keller, M. C. (2011). A critical review of the first ten years of measured gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*, 1041–1049.

Eaves, L., & Erkanil, A. (2003). Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human development change and G × E interaction. *Behavior Genetics*, *33*, 279–299.

Eaves, L. J. (2006). Genotype × Environment interaction in psychopathology: Fact or artifact? *Twin Research and Human Genetics*, *9*, 1–8.

Edenberg, H. J., Dick, D. M., Xuei, X., Tian, H., Almasy, L., Bauer, L. O., . . . Begleiter, H. (2004). Variations in GABRA2, encoding the α2 subunit of the GABA-A receptor are associated with alcohol dependence and with brain oscillations. *American Journal of Human Genetics*, *74*, 705–714.

Enoch, M. A., Schwartz, L., Albaugh, B., Virkkunen, M., & Goldman, D. (2006). Dimensional anxiety mediates linkage of GABRA2 haplotypes with alcoholism. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *141B*, 599–607.

Fehr, C., Sander, T., Tadic, A., Lenzen, K. P., Anghelescu, I., Klawe, C., . . . Szegedi, A. (2006). Confirmation of association of the GABRA2 gene with alcohol dependence by subtype-specific analysis. *Psychiatric Genetics*, *16*, 9–17.

Flint, J., & Munafo, M. R. (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine*, *37*, 163–180.

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, *2*, 235–247.

Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Koesterer, R., Smith, A. H., . . . Farrer, L. A. (2014). Genome-wide association study of alcohol dependence: Significant findings in African- and European-Americans including novel risk loci. *Molecular Psychiatry*, *19*, 41–49. doi:10.1038/mp.2013.145

Gelernter, J., Yu, Y., Weiss, R., Brady, K., Panhuysen, C., Yang, B. Z., . . . Farrer, L. (2006). Haplotype spanning TTC12 and ANKK1, flanked by the DRD2 and NCAM1 loci, is strongly associated to nicotine dependence in two distinct American populations. *Human Molecular Genetics*, *15*, 3498–3507.

Ghosh, S., Begleiter, H., Porjesz, B., Chorlian, D. B., Edenberg, H. J., Foroud, T., . . . Reich, T. (2003). Linkage mapping of beta 2 EEG waves via non-parametric regression. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *118B*, 66–71.

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., & . . . Martin, J. B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, *306*, 234–238.

Hamilton, C. M., Strader, L. C., Pratt, J. G., Maiese, D., Hendershot, T., Kwok, R. K., . . . Haines, J. (2011). The PhenX toolkit: Get the most from your measures. *American Journal of Epidemiology*, *174*, 253–260. doi:10.1093/aje/kwr193

Harden, K. P., Hill, J. E., Turkheimer, E., & Emery, R. E. (2008). Gene–environment correlation and interaction in peer effects on adolescent alcohol and tobacco use. *Behavior Genetics*, *38*, 339–347.

Harris, R. A. (1999). Ethanol actions on multiple ion channels: Which are important? *Alcoholism: Clinical and Experimental Research*, *23*, 1563–1570.

Hewitt, J. (2012). Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics*, *42*, 1–2.

Hodgkinson, C., Yuan, Q., Xu, K., Shen, P., Heinz, E., Lobos, E., . . . Goldman, D. (2008). Addictions biology: Haplotype-based analysis for 130 candidate genes on a single array. *Alcohol & Alcoholism*, *43*, 505–515. doi:10.1093/alcalc/agn032

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology*, *9*, 61–89.

Hutter, C. M., Mechanic, L. E., Chatterjee, N., Kraft, P., & Gillanders, E. M. (2013). Gene–environment interactions in cancer epidemiology: A National Cancer Institute think tank report. *Genetic Epidemiology*, *37*, 643–657.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Ioannidis, J. P., Munafo, M. R., Fusar-Poll, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, prevention. *Trends in Cognitive Sciences*, *18*, 235–241. doi:10.1016/j.tics.2014.02.010

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs*, *1*, 57–93.

Johnston, C., Lahey, B. B., & Matthys, W. (2013). Editorial policy for candidate gene studies. *Journal of Abnormal Child Psychology*, *41*, 511–514.

Kareken, D. A., Liang, T., Wetherill, L., Dzemidzic, M., Bragulat, V., Cox, C., . . . Foroud, T. (2010). A polymorphism in GABRA2 is associated with the medial frontal response to alcohol cues in an fMRI study. *Alcoholism: Clinical and Experimental Research*, *34*, 2169–2178. doi:10.1111/j.1530-0277.2010.01293.x

Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: Evident of genetic moderation. *Archives of General Psychiatry*, *68*, 444–454.

Keller, M. C. (2014). Gene by environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biological Psychiatry*, *75*, 18–24. doi:10.1016/j.biopsych.2013.09.006

Kendler, K. S. (2013). What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Molecular Psychiatry*, *18*, 1058–1966.

Kendler, K. S., & Eaves, L. J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *The American Journal of Psychiatry*, *143*, 279–289.

Kendler, K. S., & Gardner, C. (2010). Interpretation of interactions: Guide for the perplexed. *British Journal of Psychiatry*, *197*, 170–171.

Kendler, K. S., Kessler, R. C., Walters, E. E., MacLean, C. J., Neale, M. C., Heath, A. C., & Eaves, L. J. (1995). Stressful life events, genetic liability, and onset of an episode of major depression in women. *American Journal of Psychiatry*, *152*, 833–842.

Kim-Cohen, J., Caspi, A., Taylor, A., Williams, B., Newcombe, R., Craig, I. W., & Moffitt, T. E. (2006). MAOA, maltreatment, and gene–environment interaction predicting children's mental health: New evidence and a meta-analysis. *Molecular Psychiatry*, *11*, 903–913.

Lakens, D., & Evers, E. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, *9*, 278–292. doi:10.1177/1745691614528520

Lappalainen, J., Krupitsky, E., Remizov, M., Pchelina, S., Taraskina, A., Zvartau, E., . . . Gelernter, J. (2005). Association between alcoholism and Gamma-Amino Butyric Acid alpha2 receptor subtype in a Russian population. *Alcoholism: Clinical and Experimental Research*, *29*, 493–498.

Lasky-Su, J., Neale, B. M., Franke, B., Anney, R. J., Zhou, K., Maller, J. B., . . . Faraone, S. V. (2008). Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *147B*, 1345–1354.

Li, D., Sulovari, A., Cheng, C., Zhao, H., Kranzler, H. R., & Gelernter, J. (2014). Association of gamma-aminobutyric acid A receptor a2 (GABRA2) with alcohol use disorder. *Neuropsychopharmacology*, *39*, 907–918. doi:10.1038/npp.2013.291

Liu, J., Yang, A. R., Kelly, T., Puche, A., Esoga, C., June, J. L., . . . Aurelian, L. (2011). Binge alcohol drinking is associated with GABAA alpha2-regulated Toll-like receptor 4 (TLR4) expression in the central amygdala. *Proceedings of the National Academy of Sciences, USA*, *108*, 4465–4470. doi:10.1073/pnas.10120108

Liu, J. Z., Tozzi, F., Waterworth, D. M., Pillal, S. G., Middelton, L., Berrettini, W., . . . Marchini, J. (2010). Meta-analysis and imputation refines the association of 12q25 with smoking quality. *Nature Genetics*, *42*, 436–440. doi:10.1038/ng.572

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial visualization and mathematical ability. *Psychological Bulletin*, *107*, 385–393.

Manolio, T., & Collins, F. S. (2009). The HapMap and genome-wide association studies in diagnosis and therapy. *Annual Review of Medicine*, *60*, 443–456.

Matthews, A. G., Hoffman, E. K., Zezza, N., Stiffler, S., & Hill, S. Y. (2007). The role of the GABRA2 polymorphism in multiplex alcohol dependence families with minimal comorbidity: Within-family association and linkage analyses. *Journal of Studies on Alcohol and Drugs*, *68*, 625–633.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390.

Meyers, J. L., Nyman, E., Loukola, A., Rose, R. J., Kaprio, J., & Dick, D. (2013). The association between DRD2/ANKK1 and genetically informed measure of alcohol use and problems. *Addiction Biology*, *10*, 523–536. doi:10.1111/j.1369-1600.2012.00490.x

Monroe, S. M. (2008). Modern approaches to conceptualizing and measuring human life stress. *Annual Review of Clinical Psychology*, *4*, 33–52.

Monroe, S. M., & Reid, M. W. (2008). Gene–environment interactions in depression: Genetic polymorphisms and life stress polyprocedures. *Psychological Science*, *10*, 947–956.

Mukherjee, B., Ahn, J., Gruber, S. B., & Chatterjee, N. (2012). Testing gene–environment interaction in large-scale case-control association studies: Possible choices and comparisons. *American Journal of Epidemiology*, *175*, 177–190. doi:10.1093/aje/kwr367

Munafo, M. R. (2006). Candidate gene studies in the 21st century: Meta-analysis, mediation, moderation. *Genes, Brain, and Behavior, 5*(S1), 3–5.

Munafo, M. R. (2009). Reliability and replicability of genetic association studies. *Addiction*, *104*, 1439–1440. doi:10.1111/j.1360-0443.2009.02662.x

Munafo, M. R., Durrant, C., Lewis, G., & Flint, J. (2009). Gene X environment interactions at the serotonin transporter locus. *Biological Psychiatry*, *65*, 211–219. doi:10.1016/j.biopsych.2008.06.009

Munafo, M. R., & Flint, J. (2009). Replication and heterogeneity in gene x environment interaction studies. *International Journal of Neuropsychopharmacology*, *12*, 727–729.

Munafo, M. R., & Gage, S. H. (2013). Improving the reliability and reporting of genetic association studies. *Drug and Alcohol Dependence*, *132*, 411–413. doi:10.1016/j.drugalcdep.2013.03.023

Murray, J. M., Davies, K. E., Harper, P. S., Meredith, L., Mueller, C. R., & Williamson, R. (1982). Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature*, *300*, 69–71.

Need, A. C., Ge, D., Weale, M. E., Maia, J., Feng, S. H., Heinzen, E. L., . . . Goldstein, D. B. (2009). A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genetics*, *5*(2), e1000373. doi:10.1371/journal.pgen.1000373

Neiswanger, K., Kaplan, B. B., & Hill, S. Y. (1995). What can the DRD2/alcoholism story teach us about association studies in psychiatric genetics? *American Journal of Medical Genetics*, *60*, 272–275.

Neville, M. J., Johnstone, E. C., & Walton, R. T. (2004). Identification and characterization of ANKK1: A novel kinase gene closely linked to DRD2 on chromosome band 11q23.1. *Human Mutation*, *23*, 540–545.

Olfson, E., & Bierut, L. J. (2012). Convergence of genome-wide association and candidate gene studies for alcoholism. *Alcoholism, Clinical and Experimental Research*, *36*, 2086–2094. doi:10.1111/j.1530-0277.2012.01843.x

Porjesz, B., Almasy, L., Edenberg, H. J., Wang, K., Chorlian, D. B., Foroud, T., . . . Begleiter, H. (2002). Linkage disequilibrium between the beta frequency of the human EEG and a GABAA receptor gene locus. *Proceedings of the National Academy of Sciences, USA*, *99*, 3729–3733.

Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews: Genetics*, *11*, 459–463. doi:10.1038/nrg2813

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research*, *5*, 554–571.

Reich, T., Edenberg, H. J., Goate, A., Williams, J. T., Rice, J. P., Van Eerdewegh, P., . . . Begleiter, H. (1998). Genome-wide search for genes affecting the risk for alcohol dependence. *American Journal of Medical Genetics*, *81*, 207–215.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kahler, A. K., Akterin, S., . . . Sullivan, P. F. (2013). Genome-wide association analysis identified 13 new risk loci for schizophrenia. *Nature Genetics*, *45*, 1150–1159.

Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L., Hoh, J., . . . Merikangas, K. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. *Journal of the American Medical Association*, *301*, 2462–2471.

Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis-stress: Recommendations for evaluating interaction effects. *Development and Psychopathology*, *24*, 389–409. doi:10.1017/S0954579412000065

Rose, E. J., & Donohoe, G. (2013). Brain vs behavior: An effect size comparison of neuroimaging and cognitive studies of genetic risk for schizophrenia. *Schizophrenia Bulletin*, *39*, 518–526. doi:10.1093/schbul/sbs056

Rose, R. J., Dick, D. M., Viken, R. J., & Kaprio, J. (2001). Gene–environment interaction in patterns of adolescent drinking: Regional residency moderates longitudinal influences on alcohol use. *Alcoholism: Clinical and Experimental Research*, *25*, 637–643.

Ruderfer, D. M., Fanous, A. H., Ripke, S., McQuillin, A., Amdur, R. L., Schizophrenia Working Group of Psychiatric Genomics Consortium . . . Kendler, K. S. (2014). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Molecular Psychiatry, 19*, 1017–1024.

Salvatore, J., Aliev, F., Edwards, A., Evans, D., Macleod, J., Hickman, M., . . . Dick, D. (2014). Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. *Genes*, *5*, 330–346. doi:10.2290/genes5020330

Shen, Y. C., Fan, J. H., Edenberg, H. J., Li, T. K., Cui, Y. H., Wang, Y. F., . . . Xia, G. Y. (1997). Polymorphism of ADH and ALDH genes among four ethnic groups in China and effects upon the risk for alcoholism. *Alcoholism: Clinical and Experimental Research*, *21*, 1272–1277.

Sher, K., & Steinley, D. S. (2013, June). *Some issues surrounding interactions*. Paper presented at the NIAAA Workshop on Gene–Environment Interactions, Rockville, MD.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Sklar, P., Ripke, S., Scott, L. J., Andreassen, O. A., Cichon, S., Craddock, N., . . . Purcell, S. M. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, *43*, 977–983. doi:10.1038/ng.943

Smith, G. (2010). Mendelian randomization for strengthening causal interference in observational studies: Application to gene x environment interactions. *Perspectives on Psychological Science*, *5*, 527–545. doi:10.1177/1745691610383505

South, S. C., & Krueger, R. F. (2008). Martial quality moderates genetic and environmental influences on the internalizing spectrum. *Journal of Abnormal Psychology*, *117*, 826–837. doi:10.1037/a0013499

Soyka, M., Preuss, U. W., Hesselbrock, V., Zill, P., Koller, G., & Bondy, B. (2008). GABA-A2 receptor subunit gene (GABRA2) polymorphisms and risk for alcohol dependence. *Journal of Psychiatric Research*, *42*, 184–191.

Spellman, B. A. (2012). Introduction to the special section: Data, data, everywhere . . . especially in my file drawer. *Perspectives on Psychological Science*, *7*, 58–59.

Steinberg, S., de Jong, S., Mattheisen, M., Costas, J., Demontis, D., Jamain, S., . . . Stefansson, K. (2014). Common variant 16p11.2 conferring risk of psychosis. *Molecular Psychiatry*, *19*, 108–114. doi:10.1038/mp.2012.157

Stephens, S. H., Hartz, S. M., Hoft, N. R., Saccone, N. L., Corley, R. C., Hewitt, J. K., . . . Ehringer, M. A. (2013). Distinct loci in the CHRNA5/CHRNA3/CHRNB4 gene cluster are associated with onset of regular smoking. *Genetic Epidemiology*, *37*, 846–859. doi:10.1002/gepi.21760

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Sullivan, P. F., Daly, M., & O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: The emerging picture and its implications. *Nature Reviews Genetics*, *13*, 537–551.

Sullivan, P. F., Lin, D., Tzeng, J. Y., van den Oord, E., Perkings, D., Stroup, T. S., . . . Close, S. L. (2008). Genomewide association for schizophrenia in the CATIE study: Results of stage 1. *Molecular Psychiatry*, *13*, 570–584. doi:10.1038/mp.2008.25

Thomas, D. C., Lewinger, J. P., Murcray, C. E., & Gauderman, W. J. (2012). Invited commentary: GE-Whiz! Ratcheting gene–environment studies up to the whole genome and the whole exposome. *American Journal of Epidemiology*, *175*, 203–207; discussion 208-209. doi:10.1093/aje/kwr365

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., & Renteria, M. E. . . . Saguenay Youth Study (SYS) Group. (2014). The ENIGMA Consortium: Large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, *8*, 153–182. doi:10.1007/s11682-013-9269-5

Thorgeirsson, T. E., Gudbjartsson, D. F., Surakka, I., Vink, J. M., Amin, N., Geller, F., . . . Stefansson, K. (2010). Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genetics*, *42*, 448–453. doi:10.1038/ng.573

Tobacco and Genetics Consortium. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics*, *42*, 441–447. doi:10.1038/ng.571

Treutlein, J., Cichon, S., Ridinger, M., Wodarz, N., Soyka, M., Zill, P., . . . Rietschel, M. (2009). Genome-wide association study of alcohol dependence. *Archives of General Psychiatry*, *66*, 773–784.

Tsui, L. C., Buchwald, M., Barker, D., Braman, J. C., Knowlton, R., Schumm, J. W., . . . Donis-Keller, H. (1985). Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science*, *230*, 1054–1057.

Uher, R., & McGuffin, P. (2010). The moderation by the serotonin transporter gene of environmental adversity in depression: 2009 update. *Molecular Psychiatry*, *15*, 18–22.

van Assen, M. A., van Aert, R. C., Nuijten, M. B., & Wicherts, J. M. (2014). Why publishing everything is more effective than selective publishing of statistically significant results. *PLoS ONE*, *9*(1), e84896. doi:10.1371/journal.pone.0084896

van der Sluis, S., Posthuma, D., & Dolan, C. V. (2012). A note on false positives and power in G × E modelling of twin data. *Behavior Genetics*, *42*, 170–186.

Villafuerte, S., Heitzeg, M. M., Foley, S., Yau, W. Y., Majczenko, K., Zubieta, J. K., . . . Burmeister, M. (2012). Impulsiveness and insula activation during reward anticipation are associated with genetic variants in GABRA2 in a family sample enriched for alcoholism. *Molecular Psychiatry*, *17*, 511–519. doi:10.1038/mp.2011.33

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, *90*, 7–24.

Ward, H., Mitrou, P. N., Bowman, R., Luben, R., Wareham, N. J., Khaw, K. T., & Bingham, S. (2009). APOE genotype, lipids, and coronary heart disease risk: A prospective population study. *Archives of Internal Medicine*, *169*, 1424–1429. doi:10.1001/archinternmed.2009.234

Weng, L., Macciardi, F., Subramanian, A., Guffanti, G., Potkin, S. G., Yu, Z., & Xie, X. (2011). SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics*. Advance online publication.

Whytt, R. (1765). *Observations on the nature, causes, and cure of those disorders which are commonly called nervous, hypochondriac, or hysteric*. Edinburgh, Scotland: Becket & Du Hondt.

Widaman, K. F., Helm, J. L., Castro-Schilo, L., Pluess, M., Stallings, M. C., & Belsky, J. (2012). Distinguishing ordinal and disordinal interactions. *Psychological Methods*, *17*, 615–622. doi:10.1037/a0030003

Wray, N. R., Pergadia, M. L., Blackwood, D. H. R., Penninx, B. W. J. H., Gordon, S. D., Nyholt, D. R., . . . Sullivan, P. F. (2012). Genome-wide association study of major depressive disorder: New results, meta-analysis, and lessons learned. *Molecular Psychiatry*, *17*, 36–48.

Yang, B. Z., Kranzler, H. R., Zhao, H., Gruen, J. R., Luo, X., & Gelernter, J. (2007). Association of haplotypic variants in DRD2, ANKK1, TTC12 and NCAM1 to alcohol dependence in independent case control and family samples. *Human Molecular Genetics*, *16*, 2844–2853.

Yang, B. Z., Kranzler, H. R., Zhao, H., Gruen, J. R., Luo, X., & Gelernter, J. (2008). Haplotypic variants in DRD2, ANKK1, TTC12, and NCAM1 are associated with comorbid alcohol and drug dependence. *Alcoholism: Clinical and Experimental Research*, *32*, 2117–2127.

Young, R. M., Lawford, B. R., Nutting, A., & Noble, E. P. (2004). Advances in molecular genetics and the prevention and treatment of substance misuse: Implications of association

studies of the A1 allele of the D2 dopamine receptor gene. *Addictive Behaviors*, *29*, 1275–1294.

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, *40*, 424–431.

Zammit, S., Lewis, G., Dalman, C., & Allebeck, P. (2010). Examining interactions between risk factors for psychosis. *British Journal of Psychiatry*, *197*, 207–211. doi:10.1192/bjp.bp.109.070904

Zammit, S., Owen, M. J., & Lewis, G. (2010). Misconceptions about gene–environment interactions in psychiatry. *Evidenced-Based Mental Health*, *13*, 65–68. doi:10.1136/ebmh.13.3.65

Zannas, A. S., & Binder, E. B. (2014). Gene–environment interactions at the FKBP5 locus: Sensitive periods, mechanisms and pleiotropism. *Genes, Brain and Behavior*, *13*, 25–37. doi:10.1111/gbb.12104

Zhao, Z., Guo, A., van den Oordd, E. J. C. G., Aliev, F., Jia, P., Riley, B. P., . . . Miles, M. F. (2012). Multi-species data integration and gene ranking enriches significant results in an alcoholism genome-wide association study. *BMC Genomics*, *13*(Suppl. 8), S16. doi:10.1186/147-2164-13-S8-S16

Ziller, M. J., Gu, H., Muller, F., Donaghey, J., Tsai, L. T. Y., Kohlbacher, O., . . . Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, *500*, 477–481. doi:10.1038/nature1243