THE JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY
ACAMH
Journal of Child Psychology and Psychiatry 52:12 (2011), pp 1223–1238 doi:10.1111/j.1469-7610.2011.02466.x

Research Review: Gene–environment interaction research in youth depression – a systematic review with recommendations for future research

Erin C. Dunn,[1,2] Monica Uddin,[3] S.V. Subramanian,[1] Jordan W. Smoller,[4] Sandro Galea,[5] and Karestan C. Koenen[1,2,5]

[1]Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA; [2]Center on the Developing Child, Harvard University, Cambridge, MA; [3]Center for Molecular Medicine and Genetics and Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI; [4]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; [5]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA

Background: Depression is a major public health problem among youth, currently estimated to affect as many as 9% of US children and adolescents. The recognition that both genes (nature) and environments (nurture) are important for understanding the etiology of depression has led to a rapid growth in research exploring gene–environment interactions (GxE). However, there has been no systematic review of GxE in youth depression to date. **Methods:** The goal of this article was to systematically review evidence on the contribution of GxE to the risk of child and adolescent depression. Through a search of PubMed and PsycINFO databases to 1 April 2010, we identified 20 candidate gene–environment interaction studies focused on depression in youth (up to age 26) and compared each study in terms of the following characteristics: research design and sample studied; measure of depression and environment used; genes explored; and GxE findings in relation to these factors. **Results:** In total, 80% of studies (n = 16) found at least one significant GxE association. However, there was wide variation in methods and analyses adopted across studies, especially with respect to environmental measures used and tests conducted to estimate GxE. This heterogeneity made it difficult to compare findings and evaluate the strength of the evidence for GxE. **Conclusions:** The existing body of GxE research on depression in youth contains studies that are conceptually and methodologically quite different, which contributes to mixed findings and makes it difficult to assess the current state of the evidence. To decrease this heterogeneity, we offer 20 recommendations that are focused on: (a) reporting GxE research; (b) testing and reporting GxE effects; (c) conceptualizing, measuring and analyzing depression; (d) conceptualizing, measuring and analyzing environment; (e) increasing power to test for GxE; and (f) improving the quality of genetic data used. Although targeted to GxE research on depression, these recommendations can be adopted by GxE researchers focusing on other mental health outcomes. **Keywords:** Depression, children, adolescents, youth, gene, environment, interaction.

Introduction

This article presents a systematic review of the evidence for genotype by environment interaction (GxE) in youth depression, which currently affects as many as 9% of US youth (Avenevoli, Knight, Kessler, & Merikangas, 2008). The etiology of depression is complex, resulting from both genetic and environmental factors. Twin studies show that the heritability of youth-onset depression ranges from 30% to 80%, with the remaining variance explained by environmental factors (Rice, Harold, & Thapar, 2002). Although this suggests that depression is moderately to highly heritable, neither candidate gene nor genome-wide association studies have identified robust associations between specific genes and depression (Lopez-Leon et al., 2008; Shaikh et al., 2008; Shyn et al., 2009; Sullivan et al., 2009). By contrast, environmental risk factors for depression are well-documented and include poverty (Brooks-Gunn & Duncan, 1997; McLeod & Shanahan, 1996), negative family relationships and parental divorce (Gilman, Kawachi, Fitzmaurice, & Buka, 2003; Repetti, Taylor, & Seeman, 2002), and child maltreatment (Chapman et al., 2004; Widom, DuMont, & Czaja, 2007). However, only a minority of youth exposed to these environments develop depression, raising questions about individual differences in genetic vulnerability (or sensitivity) to adverse environments. GxE research addresses such questions by examining whether individuals with specific alleles (i.e. alternative forms of DNA sequence at a specific locus) or genotypes (i.e. the combination of alleles that an individual carries at a specific locus) are more or less sensitive to the effects of their environments (Brown & Harris, 2008;

Conflict of interest statement: No conflicts declared.

Khoury, Davis, Gwinn, Lindegren, & Yoon, 2005; Moffitt, Caspi, & Rutter, 2006; Uher & McGuffin, 2008).

The goal of the current article was to systematically identify and summarize studies that tested for GxE in relation to depression among children and adolescents. We focused specifically on the characteristics of these studies with respect to both their methods and findings. We did so to inventory the published literature with respect to their research designs, compliment existing meta-analyses (which have emphasized statistical findings, rather than methodological issues), and ultimately provide substantive conclusions that could guide future research in this area. We also sought to discern the level of heterogeneity in GxE findings that existed across studies specifically among youth, given that previous reviews, which included studies of multiple age ranges, concluded that there were mixed findings for GxE (Munafo, Durrant, Lewis, & Flint, 2009; Uher & McGuffin, 2008, 2010). We focused on GxE research among youth, defined as age 26 and below, as this age span represents a 'sensitive period' characterized by rapid social, psychological, and biological changes and marks the time when depression often first emerges (Rudolph, 2009; Steinberg & Morris, 2001). By focusing specifically on depression in youth, rather than depression in adulthood, we hoped to gain greater etiological insights into the development of this disorder. Previous reviews have also not attended to these developmental periods (Brown & Harris, 2008; Karg, Burmeister, Shedden, & Sen, 2011; Monroe & Reid, 2008; Munafo et al., 2009; Risch et al., 2009; Uher & McGuffin, 2008, 2010). For example, out of the 28 studies cited in a recent meta-analysis (Risch et al., 2009), only about one third (10 studies) focused on youth.

Methods

Search procedures

We systematically identified articles published as of 1 April 2010 through PubMed and PsycINFO search engines. We used a combination of database-specific index terms (e.g. 'depression', 'genetics', 'environment', 'social environment') and individual terms located in the title or abstract (e.g. 'gene', 'environment', 'interaction', 'moderation', 'modification', 'stress', 'abuse', 'depress*'). We applied limits to searches in both databases to eliminate articles: (a) focused on individuals older than age 26, (b) not written in English, (c) focused on non-human animals, (d) not published in a peer-reviewed journal, or (e) based on reviews or meta-analyses. We also searched for articles by examining the references pages of review articles, meta-analyses and other empirical articles published since 2005. Through these searches, we located 278 studies.

We applied the following criteria to these 278 articles to identify the articles for our review: (a) measured major depressive disorder or depressive symptoms as a unique outcome, (b) included participants age 26 or younger, and (c) focused on a specific candidate GxE interaction (as opposed to a twin or adoption study). We examined depression because it has received the most attention in the adult literature and popular press. In addition, we believed by narrowing our phenotype we would be able to make a more comprehensive assessment of the state of the literature. We used age 26 as a liberal cut-point to capture the developmental periods of childhood, adolescence and young adulthood. This cut-point enabled us to be more inclusive and thus incorporate as many articles relevant to youth as possible. After applying these criteria, 20 studies remained, which were published between 1 July 2003 and 1 April 2010.

Results

For the purposes of describing the current state of GxE research, we summarize the research design and study samples, measurement of depression, genotype, environment and main study findings of these 20 studies (see Table 1).

Research design and study samples

Research design. The research design used to test for GxE varied across studies. Slightly more than half (n = 11) of the studies included in this review used data from an ongoing longitudinal study (Caspi et al., 2003; Chipman et al., 2007; Chorbov et al., 2007; Gibb, Benas, Grassia, & McGeary, 2009; Gibb, Uhrlass, Grassia, Benas, & McGeary, 2009; Guo & Tillman, 2009; Hammen, Brennan, Keenan-Miller, Hazel, & Najman, 2010; Nilsson, Sjoberg, Leppert, Oreland, & Damberg, 2009; Sjoberg et al., 2006; Uddin et al., 2010; Vaske, Makarios, Boisvert, Beaver, & Wright, 2009). Although in these studies data may have been collected longitudinally, the association between environmental exposure(s) and outcome was not always prospective. For example, out of 11 studies, five assessed some or all exposures included in the test for GxE concurrently with depression (Chipman et al., 2007; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009; Guo & Tillman, 2009; Hammen et al., 2010) and one, which was analyzed as a case–control study (Chorbov et al., 2007) did not have a clear exposure–disease association. Of the remaining nine, five collected data cross-sectionally (Aslund et al., 2009; Benjet, Thompson, & Gotlib, 2010; Chipman, Jorm, Tan, & Easteal, 2010; Eley et al., 2004; Haeffel et al., 2008), and four collected some or all exposure data (i.e. exposure to child maltreatment) prior to depression assessment (Cicchetti, Rogosch, & Sturge-Apple, 2007; Cicchetti, Rogosch, Sturge-Apple, & Toth, 2010; Kaufman et al., 2004, 2006), with two not reporting assessing the environmental exposure or outcome repeatedly (Kaufman et al., 2004, 2006) and two assessing only the environmental exposure repeatedly (Cicchetti et al., 2007, 2010).

Table 1 Gene–environment interaction (GxE) studies examining depression or depressive symptoms in youth (n = 20)

Author (year)	Study sample and research design[a]	Measurement of outcome	Genetic variant	Measurement of environment	GxE interaction detected and scale tested
Aslund et al. (2009)	Cross-sectional; community-based; n = 1,482; Swedish youth; ages 17–18	DSRS	*5-HTTLPR*	Exposure to violence	Yes overall for symptoms (additive scale) and diagnosis (multiplicative scale); when stratified by sex, observed interaction with both outcomes in girls only
Benjet et al. (2010)	Cross-sectional; community-based; n = 78 females; mean age 12.2	Brief CDI	*5-HTTLPR*	Relational peer victimization	Yes (additive scale)
Caspi et al. (2003)	Longitudinal; n = 847; assessed at age 26	DIS and survey	*5-HTTLPR*	Stressful life events	Yes (additive and multiplicative scale)
Chipman et al. (2007)	Two community-based Australian samples: 1. Cross-sectional; n = 2,095; 20–24 years old	GDAS	*5-HTTLPR*	Stressful life events; childhood adversities	No (multiplicative scale)
	2. Longitudinal; n = 584; age 15–16 and n = 655; ages 17–18	SMFQ	*5-HTTLPR*	Family stressors; persistent family adversity	No for family stressors when cohort aged 15–16 or 17–18 (multiplicative scale); Yes (multiplicative scale) with family adversity when cohort was aged 17–18, but not when 15–16
Chipman et al. (2010)	Cross-sectional; community-based; n = 2,099; ages 20–24	GDAS	*HTR1A*	Childhood adversity; stressful life events	No (multiplicative scale) when adjusted for multiple comparisons
Chorbov et al. (2007)	Case–control; n = 227; twins; ages 13–23	Diagnosis of a lifetime DSM–IV MDD	*5-HTTLPR*	Traumatic life events	No (multiplicative scale) for s/l genotype or number or presence of s alleles; yes (multiplicative) for number of L_A alleles
Cicchetti et al. (2007)	Cross-sectional; n = 339; mean age 16.7; with/without history of maltreatment	DISC	*MAOA* and 5-*HTTLPR*	Child maltreatment	Yes (additive scale) for *MAOA*; no for *5-HTTLPR*
Cicchetti et al. (2010)	Cross-sectional; n = 858; mean age 9.19; with/without history of maltreatment	CDI	*5-HTTLPR*	Child maltreatment	No (additive scale)
Eley et al. (2004)	Cross-sectional; n = 377; UK adolescents; ages 10–20	SMFQ	*5-HTTLPR*, *5HT2A*, *5HT2C*, *MAOA*, *TPH1*	High- versus low-risk family environment	Trend (multiplicative scale) with *5-HTTLPR*, *5HT2A* and *TPH1* overall; among females, significant interaction with *5-HTTLPR*; among males, significant interaction with *TPH1*; no for *5HT2C*, *MAOA*
Gibb, Benas, et al. (2009)	Longitudinal; n = 74; mothers and their children; mean age 9.96	CDI	*5-HTTLPR*	Maternal depressive symptoms; attentional bias for facial displays of emotion	Yes (additive scale) for maternal depression; 3-way interaction with maternal depression and attentional bias

Table 1 (Continued)

Author (year)	Study sample and research design[a]	Measurement of outcome	Genetic variant	Measurement of environment	GxE interaction detected and scale tested
Gibb, Uhrlass, et al. (2009)	Longitudinal; *n* = 100; mothers and their children; mean age 9.97	CDI	*5-HTTLPR*	Maternal expressed emotion; inferential styles for the causes of negative events, and styles for consequences and self-characteristics; maternal depression	Yes (additive scale) for all interactions tested except found trend when excluded youth with history of depression; no (additive scale) with inferential styles
Guo and Tillman (2009)	Longitudinal; *n* = 2,286; twins/siblings; ages 13–25 (Add Health)	CES–D	*DRD2* and *DRD4*	Developmental period (e.g. adolescence vs. young adulthood)	No (additive scale)
Haeffel et al. (2008)	Cross-sectional; *n* = 176; court-ordered into juvenile detention; mean age 16.2	BDI and K-SADS	*SLC6A3*	Maternal rejection	No for *rs6347* and *rs2652511* with K-SADS (multiplicative scale); yes for *rs40184* for BDI (additive scale) and borderline for K-SADS (multiplicative scale); trend with *rs40184* and BDI (additive scale)
Hammen et al. (2010)	Longitudinal; *n* = 346; mean age 23.7	BDI–II	*5-HTTLPR*	Chronic family stress/ disorder; acute stress	No with acute stress (additive scale); yes with chronic stress, gender and genotype (additive scale)
Kaufman et al. (2004)	Cross-sectional; *n* = 101; mean age 10; with/without history of maltreatment	MFQ	*5-HTTLPR*	Exposure to maltreatment; social support	Yes (additive scale) with maltreatment; no 2-way interaction with genotype and social support; 3-way interaction with maltreatment, genotype and social support
Kaufman et al. (2006)	Cross-sectional; *n* = 196; mean age 9.3; with/without history of maltreatment	MFQ	*5-HTTLPR* and *BDNF*	Exposure to maltreatment; social support	Trend (additive scale) with maltreatment; significant 3-way interaction with *BDNF*, *5-HTTLPR* and maltreatment; significant 4-way interaction with *BDNF*, *5-HTTLPR*, maltreatment and social support
Nilsson et al. (2009)	Longitudinal; *n* = 180; Swedish youth; ages 19–22	DSRS	*AP-2B*	Type of residence; parent marital status; psychosocial risk	Yes (additive scale) for type of residence overall (marginal in boys and significant in girls); yes for marital status overall and separately for boys and girls; yes for psychosocial risk overall
Sjoberg et al. (2006)	Longitudinal; *n* = 180; Swedish youth; ages 19–22	DSRS	*5-HTTLPR*	Type of residence; parent marital status; traumatic conflicts within the family; psychosocial risk	Yes (additive scale) among boys, for type of residence, separated families, and psychosocial index; among girls, significant interaction for traumatic conflicts and psychosocial index; Trend for other variables
Uddin et al. (2010)	Longitudinal; *n* = 1,084; twins/siblings; ages 12–20; mean age 16 (Add Health)	CES–D	*5-HTTLPR*	County-level deprivation	Yes (additive scale) for males; not for females

Table 1 (Continued)

Author (year)	Study sample and research design[a]	Measurement of outcome	Genetic variant	Measurement of environment	GxE interaction detected and scale tested
Vaske et al. (2009)	Longitudinal; *n* = 2,380; twins/siblings (Add Health)	CES–D	*DRD2*	Exposure to violence	No (additive scale) for full sample or when stratified by sex; yes among females; effects seen for black but not white females

BDI, Beck Depression Inventory; CDI, Children's Depression Inventory; CES–D, Center for Epidemiological Studies of Depression Scale; DIS, Diagnostic Interview Schedule; DISC, Diagnostic Interview Schedule for Children; DSRS, Depression Self-Rating Scale (of the DSM–IV); GDAS, Goldberg Depression and Anxiety Scales; K-SADS, Schedule for Affective Disorders and Schizophrenia; MDD, Major Depressive Disorder; MFQ, Mood and Feelings Questionnaire; SMFQ, Short Mood and Feelings Questionnaire.
Additive scale, refers to when risks add in their effects (e.g. linear regression); multiplicative scale, refers to when risks multiply in their effects (e.g. logistic regression).
[a]Refers to how data for the overall project were collected.

Sample. At least 12 unique samples were examined. Although some drew at least part of their sample from a high risk or clinical population (Cicchetti et al., 2007, 2010; Haeffel et al., 2008; Kaufman et al., 2004, 2006), the majority (*n* = 15) studied community-based samples. Studies varied in their sample size, with the smallest sample including 74 (Gibb, Benas, et al., 2009) and the largest 2,380 (Vaske et al., 2009).

Race/ethnicity. Studies varied with respect to the amount of racial/ethnic diversity in the sample. Five studies exclusively focused on participants who were White (Aslund et al., 2009; Caspi et al., 2003; Chipman et al., 2007, 2010; Chorbov et al., 2007), five included a majority of participants who were White (Benjet et al., 2010; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009; Haeffel et al., 2008; Hammen et al., 2010), and the remaining either included more diverse samples (Cicchetti et al., 2007, 2010; Guo & Tillman, 2009; Kaufman et al., 2004, 2006; Uddin et al., 2010; Vaske et al., 2009) or did not report this information (Eley et al., 2004; Nilsson et al., 2009; Sjoberg et al., 2006).

Sex. Studies were balanced with respect to sex, although two limited their sample to males (Haeffel et al., 2008) or females (Benjet et al., 2010).

Age. Studies varied in the age of the youth included, with 20% (*n* = 4) focusing on youth ages 12 and below (Benjet et al., 2010; Cicchetti et al., 2010; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009), 4% (*n* = 1) on youth ages 13–18 (Aslund et al., 2009), 25% (*n* = 5) on youth ages 19 and above (Caspi et al., 2003; Chipman et al., 2010; Hammen et al., 2010; Nilsson et al., 2009; Sjoberg et al., 2006), and roughly 40% (*n* = 8) on a wide age range (Chipman et al., 2007; Chorbov et al., 2007; Eley et al., 2004; Guo & Tillman, 2009; Kaufman et al., 2004, 2006; Uddin et al., 2010; Vaske et al., 2009). Two studies provided only the mean age of their sample, although both referred to studying 'adolescents' (Cicchetti et al., 2007; Haeffel et al., 2008). A total of 14 studies included youth in adolescence (between ages 10–24).

Measurement of depression (i.e. phenotype)

Outcome. The majority (*n* = 15) assessed depressive symptoms (Benjet et al., 2010; Chipman et al., 2007, 2010; Cicchetti et al., 2010; Eley et al., 2004; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009; Guo & Tillman, 2009; Hammen et al., 2010; Kaufman et al., 2004, 2006; Nilsson et al., 2009; Sjoberg et al., 2006; Uddin et al., 2010; Vaske et al., 2009). Of the remaining five, one assessed a depression diagnosis (Chorbov et al., 2007) and four assessed both a depression diagnosis and depressive symptoms (Aslund et al., 2009; Caspi et al., 2003; Cicchetti et al., 2007; Haeffel et al., 2008).

Symptom measures. Seven different measures were used to capture depressive symptoms, with the most commonly used measures being a brief or complete version of the Children's Depression Inventory (Benjet et al., 2010; Cicchetti et al., 2010; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009), a brief or complete version of the Mood and Feelings Questionnaire (Chipman et al., 2007; Eley et al., 2004; Kaufman et al., 2006, 2004) and a modified version of the Center for Epidemiological Studies of Depression Scale (Guo & Tillman, 2009; Sjoberg et al., 2006; Uddin et al., 2010; Vaske et al., 2009). Three collapsed symptoms into binary high- versus low-depressed categories (Chipman et al., 2007, 2010; Eley et al., 2004).

Diagnostic measures. Seven studies used diagnostic measures to capture presence or absence of a depressive disorder or depressive symptoms. Three used the Depression Self-Rating Scale (Aslund et al., 2009; Nilsson et al., 2009; Sjoberg et al., 2006); of these, two used only symptom counts in the analyses (Nilsson et al., 2009; Sjoberg et al., 2006) and one used both symptom counts and a binary indicator (i.e. depressed/not depressed; Aslund et al., 2009). Two used the Diagnostic Interview Schedule (Caspi et al., 2003; Cicchetti et al., 2007); in one case, the authors used only symptom counts in the analysis (Cicchetti et al., 2007) and in the other, the authors used both symptom counts and a binary indicator of a depression diagnosis (Caspi et al., 2003). One study also used an adapted version of the Diagnostic Interview for Children and Adolescents (Chorbov et al., 2007) and another used the Schedule for Affective Disorders and Schizophrenia for School-Aged Children (Haeffel et al., 2008); in both, the authors used a binary measure for the outcome (depressed/not depressed).

Data collection method. Youth were the primary if not sole respondent. In eight (40%) studies, data on the outcome were collected exclusively through youth self-report (Aslund et al., 2009; Benjet et al., 2010; Chipman et al., 2007, 2010; Cicchetti et al., 2010; Eley et al., 2004; Hammen et al., 2010; Sjoberg et al., 2006). The remaining 12 collected data by a trained interviewer (Guo & Tillman, 2009; Kaufman et al., 2004, 2006; Nilsson et al., 2009; Uddin et al., 2010; Vaske et al., 2009), or clinician (Caspi et al., 2003); in one case, these data were collected through a telephone interview, and the background of the interviewer was unclear (Chorbov et al., 2007). Youth also self-reported information about depression for the first assessment, and subsequent assessments were interviewer-administered (Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009). Some studies used a combination of methods (Cicchetti et al., 2007; Haeffel et al., 2008). In only one case did researchers use depression data reported by both youth and another informant (Caspi et al., 2003).

Psychometric properties. Thirty percent of studies (n = 6) reported information about the psychometric properties of the outcome measure in their sample (Benjet et al., 2010; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009; Guo & Tillman, 2009; Haeffel et al., 2008; Uddin et al., 2010).

Measurement of genotype

Polymorphisms examined. Eleven genetic polymorphisms were investigated. These polymorphisms included variants in genes involved in serotonergic function (*SLC6A4*, *HTR1A*, *5HT$_{2A}$*, *5HT$_{2C}$* and *TPH1*), dopaminergic function (*DRD2*, *DRD4* and *SLC6A3*), monoamine catabolism (*MAOA*), brain-derived neurotropic factor (*BDNF*), and a transcription factor implicated in the differentiation of neural crest cells (*AP-2β*). The most commonly examined polymorphism, in 75% of studies (n = 15), was the *5-HTTLPR* variable number tandem repeat, which consists of the s/s, s/l and l/l genotype. Eight studies (Chipman et al., 2010; Chorbov et al., 2007; Eley et al., 2004; Guo & Tillman, 2009; Haeffel et al., 2008; Kaufman et al., 2006; Nilsson et al., 2009; Vaske et al., 2009) either examined the *5-HTTLPR* polymorphism along with other polymorphisms or focused entirely on other genetic markers.

Genotypes analyzed. Of the 15 studies focusing on *5-HTTLPR*, 13 captured the biallelic variant (s/s, s/l or l/l genotype) and three (Chorbov et al., 2007; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009) captured the triallelic version (Hu et al., 2005): (a) L_GS, L_GL_G, SS; (b) L′S: L_AS, L_AL_G; and (c) L′L: L_AL_A. When analyzing these *5-HTTLPR* genotypes, one study combined the s/s and s/l genotype (Cicchetti et al., 2010).

Measurement of environment

The measurement of environment varied considerably across studies, with respect to both the quality of the measures employed and quantity of environments considered.

Types of exposures assessed. With the exception of two studies (Kaufman et al., 2004, 2006), which examined exposure to both risky and protective environments, studies focused on exposure to risky environmental factors. The types of risk factors assessed were diverse, ranging from acute or discretely occurring stressful life events (typically occurring in the past 6 months), including unemployment, housing, financial and relationship stressors, to more potentially chronic or ongoing exposures to childhood adversity, such as experiences of child

abuse and neglect, low family socioeconomic status, high levels of family stress, maternal depression, risky family structure (i.e. separated parents), and levels of maternal expressed emotion and rejection. Studies also assessed environment as developmental period (Guo & Tillman, 2009), aspects of the child's cognition, including attentional bias for facial displays of emotion and attributional style (Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009), type of residence (Nilsson et al., 2009; Sjoberg et al., 2006), and characteristics of the child's broader social ecology, namely county-level socioeconomic deprivation (Uddin et al., 2010).

Method. Questionnaires were the most commonly employed method for obtaining information about environmental exposures, used as the only source of data collection in 35% (n = 7) of studies (Aslund et al., 2009; Benjet et al., 2010; Chorbov et al., 2007; Eley et al., 2004; Guo & Tillman, 2009; Haeffel et al., 2008; Vaske et al., 2009). The remaining studies relied on interviews (primarily with youth), review of administrative records (e.g. Census-data; maltreatment data), computer-based methods or a combination of approaches.

Classification of exposure. There was considerable variation in how studies treated exposure status for the analysis. In 30% of studies (n = 6), environment was treated as a binary variable (i.e. exposed vs. unexposed; Aslund et al., 2009; Cicchetti et al., 2010; Eley et al., 2004; Nilsson et al., 2009; Sjoberg et al., 2006; Uddin et al., 2010), in 37.5% (n = 8) environment was treated as a ordinal variable (ranging from three to five categories, based on frequency of exposure; Caspi et al., 2003; Chipman et al., 2007, 2010; Chorbov et al., 2007; Cicchetti et al., 2007; Gibb, Uhrlass, et al., 2009; Guo & Tillman, 2009; Hammen et al., 2010), and 25% (n = 5) employed continuous measures or scales (Benjet et al., 2010; Gibb, Benas, et al., 2009; Haeffel et al., 2008), or a combination of approaches (Kaufman et al., 2004, 2006). In one case, it was unclear (Vaske et al., 2009).

Psychometric properties. Half (n = 10) reported information about the psychometric properties of the environment measure, either with an internal consistency reliability coefficient or a measure of interrater agreement (Benjet et al., 2010; Cicchetti et al., 2007, 2010; Gibb, Uhrlass, et al., 2009; Haeffel et al., 2008; Hammen et al., 2010; Kaufman et al., 2004, 2006; Nilsson et al., 2009; Vaske et al., 2009). The same number used an existing measure that had been psychometrically evaluated in another sample.

Main study findings

As described in previous sections, there was considerable heterogeneity in the methods and analyses used across studies to test for GxE. This diversity made it difficult to summarize this research, provide a synthesis of the main findings, and led us to emphasize statistical significance over magnitude of effects. We therefore provide a summary of findings on specific aspects of this research.

GxE findings. Sixteen studies found at least one significant ($p \leq .05$) GxE effect. Among the 15 studies investigating *5-HTTLPR*, 13 found at least some evidence in support of GxE, with the risk allele or genotype varying depending on the analyses conducted. With respect to the 10 other polymorphisms examined, two studies found no evidence of a GxE effect for *HTR1A* (Chipman et al., 2010), or $5HT_{2C}$ (Eley et al., 2004); one found mixed evidence for *DAT1* (Haeffel et al., 2008); one found a trend for $5HT_{2A}$ and *TPH* (Eley et al., 2004); one found evidence for *MAOA* (Cicchetti et al., 2007) and another failed to find significant evidence (Eley et al., 2004); one found evidence for *DRD2* (Vaske et al., 2009) and another failed to find significant evidence (Guo & Tillman, 2009); one found evidence for *AP-2β* (Nilsson et al., 2009); and one found evidence of three-way interaction with environment, *BDNF*, and *5-HTTLPR* (Kaufman et al., 2006).

Main effect of genotype. Eight (40%) found significant main effects or a trend ($p < .08$) for some of these specific genes (Aslund et al., 2009; Caspi et al., 2003; Chorbov et al., 2007; Eley et al., 2004; Guo & Tillman, 2009; Kaufman et al., 2004; Uddin et al., 2010; Vaske et al., 2009). Another eight (40%) did not (Benjet et al., 2010; Chipman et al., 2010; Chorbov et al., 2007; Cicchetti et al., 2007, 2010; Hammen et al., 2010; Kaufman et al., 2006; Nilsson et al., 2009), and the remaining four (20%) did not provide sufficient information to make this determination (Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009; Haeffel et al., 2008; Sjoberg et al., 2006).

Main effect of environment. Slightly more than half (n = 12) found significant main effects or a trend for at least one of the environmental variables (Aslund et al., 2009; Benjet et al., 2010; Caspi et al., 2003; Chorbov et al., 2007; Cicchetti et al., 2007, 2010; Gibb, Uhrlass, et al., 2009; Guo & Tillman, 2009; Kaufman et al., 2004; Nilsson et al., 2009; Uddin et al., 2010; Vaske et al., 2009). Of the remaining eight studies, three did not find significant effects of any environmental measure (Eley et al., 2004; Gibb, Benas, et al., 2009; Hammen et al., 2010), four did not provide sufficient evidence to evaluate this association (Chipman et al., 2010; Haeffel et al., 2008; Kaufman et al., 2006; Sjoberg et al., 2006), and one had mixed results based on its inclusion of two different samples (Chipman et al., 2007).

Effect size. Although half of the studies did not provide enough information to evaluate whether the effect for genotype or environment was larger, the studies that did report this information tended to find that the effects for environment were larger.

Effects by developmental period. Among the 14 studies that focused on adolescents (defined as ages 10–24), only two (Chipman et al., 2010; Guo & Tillman, 2009) did not find at least some evidence for GxE. Among three studies that focused on children or early adolescents (defined as ages 13 and below), one did not find evidence for GxE (Cicchetti et al., 2010).

Overall, we found more consistent evidence in support of GxE than expected. Out of 20 studies, 16 found at least some evidence to suggest a GxE effect, with 13 of the 15 studies examining the *5-HTTLPR* polymorphism finding GxE. However, within these 13 studies, researchers found mixed results concerning the risk allele or genotype; in most cases, the s allele was risky, though in others it was associated with decreased risk. At least two papers provided evidence of heterosis, whereby the heterozygote genotype (e.g. s/l) conferred a protective effect beyond that observed in either homozygote (e.g. s/s or s/l; Sjoberg et al., 2006; Uddin et al., 2010). The evidence was also mixed concerning whether main effects were detected for both genotype and environment across most of the 11 genes examined. Specifically, less than half of the studies detected a main effect of the genotype and slightly more than half detected a main effect of one or more environmental variables.

The heterogeneity in results was matched by (and probably related to) the heterogeneity in conceptual and methodological approaches used to test for GxE. Studies varied tremendously in the populations sampled, methods used to assess environmental exposures, and ultimately test for GxE. For instance, some studies presented the main effects of either environment or genotype as well as the interaction effects after controlling for sex, age and race/ethnicity, whereas others did not include any covariates. Moreover, some studies also presented main effects of either environment or genotype in the presence of interaction terms and some did not. This heterogeneity made it difficult for us to provide the kind of synthesis we sought to at the outset. It is also a major limitation, as methodological diversity probably creates discrepancies of GxE effects across studies and prevents a deeper understanding of potential GxE interactions around youth depression. We suspect that the use of disparate methods across studies is an artifact of the cross-disciplinary nature of GxE research and reflects differences in both conceptual understanding and methodological conventions adopted across disciplines.

In an effort to build bridges across disciplines and guide GxE researchers in conducting more consistent GxE research in the future, we offer in the following sections suggestions to address some of these challenges (Table 2). These recommendations are centered on: (a) reporting GxE research; (b) testing and reporting GxE effects; (c) conceptualizing, measuring and analyzing depression; (d) conceptualizing, measuring and analyzing environment; (e) increasing power to test for GxE; and (f) improving the quality of genetic data. Our hope is that these recommendations will enable GxE researchers to conduct future research that is better methodologically aligned so that substantive conclusions can one day be drawn about GxE effects on depression among youth.

Reporting GxE research

Few studies fully described their methods and analyses, including how tests for interaction were conducted and the nature of the association between exposure and outcome. This lack of specificity is not unique to GxE research. However, the complexity inherent in testing for interactions, combined with the interdisciplinary nature of this work, require that future studies adopt more explicit reporting standards. Future studies should therefore follow both the STROBE (STrengthening the REporting of OBservational studies in Epidemiology) and STREGA guidelines (Strengthening The Reporting of Genetic Association studies; Little et al., 2009; http://www.strobe-statement.org).

Testing and reporting GxE effects

Treatment of the genotype. Although nearly all studies examining *5-HTTLPR* examined the effect of each genotype (s/s, s/l, and l/l) in the analysis, there was one case where investigators combined across genotype groups (Cicchetti et al., 2010), making it impossible to detect differences that exist between s/s and s/l genotypes. This type of grouping presumes a dominant model of inheritance, whereby having at least one s allele increases risk. Given that the mode of inheritance for these candidate genes is unknown, researchers should, when possible, test an additive model (where genotype is an ordinal variable) and use all data to make comparisons across all genotype groups.

Reporting main and interaction effects. The need for more thorough reporting standards especially applies to conducting tests for interaction. This includes reporting actual regression coefficients for all parameters included in a regression model and explicitly noting what variables were included, including not just covariates. Whether or not main effects are tested and reported without the interaction term will depend on the investigators' hypotheses. However, when main effects are reported in the

Table 2 Recommendations for conducting and reporting gene–environment interaction (GxE) research on depression in youth

Reporting GxE research	1. Adopt more rigorous reporting standards (a) Follow STROBE guidelines for reporting observational studies (b) Follow STREGA guidelines for reporting genetic studies
Testing and reporting GxE effects	
Treatment of the genotype	2. Make comparisons across all genotype groups (e.g. separately test for the effects of s/s and s/l genotypes)
Reporting main and interaction effects	3. Adopt more thorough reporting standards (a) Report all parameters included in the regression model 4. Test for main effects with and without the interaction term present depending on hypotheses (b) When the interaction term is present, correctly interpret main effects (c) Present all data to enable interpretation of interaction effects
Conducting formal tests for interaction	5. Test for GxE using traditional methods (i.e. cross-product terms) 6. Incorporate new methods (i.e. test GxE at different values of E) 7. Report descriptive statistics to enable readers to understand the distribution of genotype and exposures and estimate GxE
Treatment of covariates	8. Include all relevant covariates in the analysis, including sex, age and race/ethnicity
Reporting gene–environment correlation	9. Test and report gene–environment correlation
Conceptualizing, measuring and analyzing depression	10. Use a measure of depressive symptoms as the outcome and note the scale (i.e. additive or multiplicative) used to test GxE 11. When diagnostic information is available, use this data to validate results obtained from tests of GxE on symptoms 12. Investigate other aspects of depression, including symptom clusters, age at onset versus course and comorbidity
Conceptualizing, measuring and analyzing environment	
Focus on timing	13. Use more rigorous research designs (i.e. longitudinal, experimental, quasi-experimental)
Examine frequency and duration of exposure	14. Refer to existing reviews on how to conceptualize, measure and analyze data on life stress and depression
Measurement and modeling of environmental exposures	15. Use rigorous methods to reliably and validly assess environment and causal associations of 'E' to mental health outcomes 16. Use more specific methods, rather than indexes or counts, to statistically model aspects of the environment
Incorporate a multilevel approach	17. Examine how the broader social environment (i.e. schools, neighborhoods) modify genetic effects on depression
Examine a wider array of proximal environments	18. Investigate how other proximal environments, such as families and peers, interact with genotypes to influence depression risk
Increasing power to test for GxE	19. Test GxE in larger samples and use more valid and reliable measures of environmental exposures and depressive outcomes
Improving the quality of genetic data	20. Follow STREGA guidelines for reporting genetic studies and test for Hardy–Weinberg equilibrium

presence of an interaction term, they should be correctly interpreted (see Appendix S1). Given our observations, we caution against interpreting any of the genotype or environment main effects reported in the previous studies, unless the authors explicitly described the parameters included in the regression model. Moreover, as the formal tests for interaction were often only presented for one interpretation, that is, either the environment or the genotype was the modifier, we recommend researchers report specific values for each variable included and provide actual beta coefficients so that readers can interpret interaction effects as they wish. Doing so can help lessen concerns that only the most robust GxE finding was reported.

Conducting formal tests of interaction. Future research should report basic descriptive information that may be suggestive of GxE, including a basic data table (e.g. genotype by exposure by outcome or 3×2) that summarizes the distribution of the environmental exposure and genotype according to the outcome. This recommendation is based on the finding that a considerable number of studies were missing basic univariate analyses on environmental exposures and depression outcomes. This trend in reporting perhaps showcases the overemphasis on statistical significance in current thinking on GxE, rather than on magnitude of effects and meaningful observations that are consistent with theory and existing research (Caspi, Hariri, Holmes, Uher, &

Moffitt, 2010). Epidemiologists have advocated for the use of a 'counterfactual approach' to examine GxE, which does not emphasize statistical models and instead focuses on the joint contribution of two combined risk factors or causal effects (Greenland & Rothman, 1998; Institute of Medicine Board on Health Sciences Policy, 2006). By presenting more descriptive data, researchers will be better able to discern any patterning that may exist in the GxE effect even without knowing the results of a formal statistical test for such an interaction (Kraft & Hunter, 2009).

Treatment of covariates. Three covariates – sex, age or developmental period, and race/ethnicity – should be included more explicitly in future GxE research. These factors are important for understanding the etiology of youth depression, the patterning of environmental exposures, and may relate to differences in genotype frequency. Although journal article space often prevents researchers from including this information, attempts should be made wherever possible to test and present the results of analyses according to these three factors (if not in the published article, then as online supplementary material).

Sex: Not all studies controlled for or stratified their results by sex, although the higher prevalence of depression in females is one of the most consistent findings in psychiatric epidemiology. Without controlling for or stratifying tests of GxE by sex (or conducting three-way interactions), it is unknown whether GxE effects may manifest differently for boys and girls and results may be biased. Indeed, several studies reviewed here found different GxE effects for males compared to females (see, e.g. Eley et al., 2004; Uddin et al., 2010).

Age or developmental period: Many studies did not examine whether GxE effects differed across development. Exploration of the salience of age for GxE is important for several reasons. First, some environmental exposures are age-specific. For instance, the risk of exposure to certain types of maltreatment, such as sexual violence, increases over time and sharply peaks for females around early adolescence (U.S. Department of Health and Human Services & Administration on Children, 2008). Second, neuroscience research suggests that there may be sensitive periods for these environmental exposures, whereby their effects on brain structures and functioning are more pronounced during one period in development than another (Gunnar & Quevedo, 2007; Lupien, McEwen, Gunnar, & Heim, 2009). Third, the risk of depression increases with age (Kessler et al., 2005; Rudolph, 2009). Thus, without accounting for or explicitly exploring how age or developmental period influences GxE effects, research may be biased and the field misses an opportunity to better understand how GxE effects may differ across development.

Race/ethnicity: Some studies did not control for race/ethnicity. This issue relates to population stratification, a concept from population genetics that refers to the presence of different allele frequencies among different subpopulations or ancestral groups. In cases where samples are drawn from different population groups, the failure to control for race/ethnicity may lead to biased results. Ideally, research should follow STREGA guidelines (Little et al., 2009) and describe any methods used to assess or address population stratification. At a minimum, studies should control for self-reported race/ethnicity and conduct sensitivity analyses to test whether observed GxE effects vary by race.

Reporting of gene–environment correlation. Given that less than half of the studies (n = 9) reported the results of tests for gene–environment correlation (rGE; Caspi et al., 2003; Chipman et al., 2010; Chorbov et al., 2007; Cicchetti et al., 2007; Gibb, Benas, et al., 2009; Gibb, Uhrlass, et al., 2009; Hammen et al., 2010; Kaufman et al., 2006, 2004), we recommend future research test and report whether rGE is present. rGE refers to the idea that individuals select, modify, and construct their environment (Kendler & Baker, 2007). For GxE, the concern of rGE arises in making causal inferences about the effect of environmental exposures on depression; that is, are genes and environments independent or did genetic factors play some role in determining which environments an individual was exposed to? Tests for rGE can be conducted through simple tests of association between environmental exposures and genotype. However, such tests are limited by the genotypes measured; the absence of association between tested genotypes and the environment does not rule out rGE in general but only reduces the likelihood that it is of particular concern for the specific GxE tested.

If rGE is observed, researchers can conduct stratified analyses, where the risk of depression is estimated separately for each genetic subgroup, rather than using a combined group test of statistical interaction, which presumes that genetic and environmental risks are independent. Otherwise, GxE results may be biased and should be interpreted cautiously (Jaffee & Price, 2007). Prospective or cohort designs, where environmental exposures precede depression onset, are also least likely to be affected by rGE. By contrast, retrospective designs may give rise to rGE, as recall of past events may be influenced by factors under genetic influence (i.e. mood, personality; Jaffee & Price, 2007).

Conceptualizing, measuring and analyzing depression

Future research should use a measure of depressive symptoms (i.e. continuous) rather than diagnosis (i.e. binary) as the outcome, as has already been

done in many GxE studies on youth. Researchers should also explicitly note the scale (i.e. additive or multiplicative) used to detect GxE effects. These recommendations are based on the finding that the way the outcome measure is scaled and whether the GxE effect is tested on the additive (i.e. risks add in their effect, such as linear regression) or multiplicative scale (i.e. risks multiply in their effects, such as logistic regression) influences whether a GxE effect is observed (Greenland & Rothman, 1998; Institute of Medicine Board on Health Sciences Policy, 2006). In fact, changing the scale of the outcome may create interactions that may not have previously existed or eliminate interactions that were once present (Kraft & Hunter, 2009). To that end, analyses based on binary outcomes have been shown through simulations to incorrectly detect a GxE effect when none existed (Eaves, 2006), thus raising concerns about the validity of results based on diagnoses. Using dimensional (rather than categorical) approaches is also warranted given the current emphasis in genetics on understanding how all levels of liability shape complex quantitative traits (i.e. symptoms), rather than on how extremes in liability shape qualitative traits (i.e. disorders; Plomin, Haworth, & Davis, 2009). However, when available, researchers could validate their test of GxE based on symptoms with a test of GxE using diagnoses. Some authors have argued that this approach can help validate results obtained from a previous test of GxE, decreasing the chance of spurious GxE effects (Moffitt et al., 2006). Moreover, future studies can also test for GxE using a broader conceptualization of the depression phenotype (Cross-Disorder Phenotype Group of the Psychiatric GWAS Consortium et al., 2009).

Conceptualizing, measuring and analyzing environment

Focus on timing. There was wide variation in the timing of the exposures assessed across studies, with respect to the temporal relationship between the exposure and outcome (i.e. prospective vs. cross-sectional), ability to make causal inferences (i.e. lag-time between exposure onset and development of depression) and developmental period or stage in the lifecourse considered (i.e. early childhood, childhood, adolescence). For instance, despite being embedded in an ongoing longitudinal study, the exposure and outcome were often measured simultaneously. As noted, little attention was also paid toward understanding the timing of exposures in relation to development. Thus, GxE research would benefit from incorporating more rigorous research designs, including experimental and quasi-experimental approaches.

Examine frequency and duration of exposure. Studies differed in the frequency and duration of each exposure included (e.g. acute or discrete vs. chronic or cumulative occurrence), with some studies examining one-time events occurring close in time to the assessment of depression and others investigating repeated exposures occurring over longer periods of time. These differences not only create challenges in making comparisons but also prevent a deeper understanding of how the frequency, duration and persistence of the exposure influenced detection of a significant GxE interaction (Moffitt et al., 2006; Uher & McGuffin, 2008).

This variation also highlights the different theoretical traditions used to examine the association between life stress and depression (see recent reviews by Cohen, Kessler, & Gordon, 1995; Hammen, 2005; Monroe, 2008; Monroe & Reid, 2008). These approaches vary in the characteristics of the stressor examined and the psychological or biological explanations used to explain how the stressor exerts its effect. For example, in one tradition, researchers argue that depression results from exposure to acute or major, threatening and recent life events (Brown & Harris, 1978, 1989). This approach parallels the notion of 'diathesis-stress', whereby a genetic liability (diathesis) interacts with a negative life experience (stress) to cause depression, with genes exacerbating or buffering the effects of stress (Monroe & Simons, 1991). In another tradition, researchers focus on ongoing and chronic exposures to stress or adversity, such as poverty, child maltreatment and social deprivation, based on the idea that these stressful conditions can accumulate over time, resulting in an increased 'allostatic load' or wear and tear on the body (McEwen & Seeman, 1999). Several authors have outlined the key issues to consider in conceptualizing, measuring and analyzing data about the role of life stress in depression (Cohen et al., 1995; Hammen, 2005; Monroe, 2008). We urge researchers to consult these sources to more carefully capture characteristics of environmental stressors.

Measurement and modeling of environmental exposures. Studies varied considerably with respect to the quality of the measures and approaches used to capture environment or some aspect of stress. Some used reliable and valid scales, whereas others used single items or measures designed by their research team. Some also used self-reported measures, whereas others used interview-based approaches. This is an important distinction because interview-based approaches are more reliable than self-reported checklists (Monroe, 2008). These observations leave the impression that there is considerable 'noise' in the assessment of environment, underscoring the need to incorporate more rigorous methods to reliably and validly assess environmental exposures.

We also found differences in how each study treated environmental exposures in the analyses.

For example, some studies examined counts depicting the total number of events experienced, while others used binary indicators (exposed/unexposed) or scales to capture more complex phenomenon, not limited to stress such as maternal rejection. Greater specificity in capturing features of the environment is needed to better understand how specific exposures, in what combination, and to what degree, are associated with depression.

Moreover, few studies discussed whether the environmental exposures examined were selected because they had environmentally mediated effects on the outcome, in other words that depressive symptoms in youth were caused by environmental features and that this association was not due to genetic factors. This is important because an association between an environment and an outcome may arise due to a third variable, namely common genetic liability. Evidence for environmental mediation (i.e. that environments are associated with depression above and beyond genetic factors) is available for some 'E' variables used in GxE studies (e.g. physical abuse; Jaffee, Caspi, Moffitt, Polo-Tomas, et al., 2004; Jaffee, Caspi, Moffitt, & Taylor, 2004). However, providing robust evidence of environmental mediation for many 'E' variables (e.g. childhood socioeconomic status) is methodologically challenging (Purcell & Koenen, 2005; Rutter, Pickles, Murray, & Eaves, 2001; Turkheimer, D'Onofrio, Maes, & Eaves, 2005; Turkheimer & Waldron, 2000). Requiring that all 'E' variables in GxE studies have demonstrated environmentally mediated effects might unnecessarily limit the exposures that could be considered in this research. For this reason, we argue that GxE studies would be strengthened if the 'E' has been shown to be environmentally mediated; however, we do not argue that environmental mediation is a requirement for 'E' to be included in a GxE study.

Incorporate a multilevel approach. Although the social and physical contexts in which youth develop – their family, school and neighborhood environments – shape their health and risk of depression (Bronfenbrenner, 1979; Gershoff & Aber, 2006; Goodman, Huang, Wade, & Kahn, 2003; Leventhal & Brooks-Gunn, 2000; Mair, Diez Roux, & Galea, 2008), nearly every study reviewed examined proximal or individual-level factors, ignoring the distal social conditions surrounding youth and thus the multilevel nature of disease causation (Diez Roux, 1998; Subramanian, Jones, & Duncan, 2003). Although proximal factors often confer larger risks for disease when compared to distal factors, the ubiquity of exposure to macrosocial environmental variables suggests that their role in determining the population distribution of youth depression may be substantial.

Future GxE research would benefit from incorporating a multilevel approach to conceptualizing and measuring environments for two reasons. First, the social contexts surrounding youth play a pivotal role in determining the resources they can draw from to support their development. For instance, neighborhood and school environments can provide access to assets that affect both whether a child will engage in a behavior that increases their risk of developing depression (e.g. substance abuse) and how they respond to stressors (e.g. positive peer social networks; Berkman & Kawachi, 2000). Moreover, the fact that families are also embedded in these social contexts suggests that they can shape parents' capacity to raise their children (Tendulkar, Buka, Dunn, Subramanian, & Koenen, 2010). Second, distal environments appear to be important for understanding GxE effects. For instance, a study included in this review found that after controlling for individual-level risks, the effect of the *5-HTTLPR* genotype on risk of depression was modified by county-level deprivation (Uddin et al., 2010). Similar GxE effects have been found for urban/rural residence on depression in adults (Jokela, Lehtimaki, & Keltikangas-Jarvinen, 2007; Xu et al., 2009).

Researchers can incorporate the social environment in future GxE studies by drawing from ecological theories of child development (Bronfenbrenner, 1979) and using multilevel modeling techniques (Raudenbush & Bryk, 2002; Subramanian et al., 2003). Measurement of the social environment often combines administrative, observational and self-report data. Researchers can use 'geocoding' procedures to link youth's street address to data from publicly available sources (i.e. US Census) that provide cost-effective information about the contexts (i.e. neighborhood, county, or state) surrounding youth. Observational data include systematic observations of social and physical markers of disorder (e.g. graffiti or public intoxication in a neighborhood; Sampson & Raudenbush, 1999). Self-report measures can also be used and when aggregated to the school- or neighborhood level, can capture features of these settings (Dunn & Masyn, 2011; Shinn, 1990); however, brief self-report measures of these environments are lacking. Researchers will face a trade-off between sample size and measurement, with larger samples making it potentially more challenging to collect extensive measures of the environment (as well as phenotype). Regardless of measurement approach, attention should be paid toward measuring the social contexts most influential during a specific period of development (i.e. home and school with younger children; neighborhoods for adolescents).

Examine a wider array of proximal environments. Future research should focus on a broader array of proximal environments. We found limited

attention to protective factors and positive environments, as every study, with the exception of two (Kaufman et al., 2004, 2006), examined negative life events, childhood adversities or some other stressor. Moreover, other types of proximal environments that could be more deeply explored include family and peer relationships and interventions that may play a role in influencing reactions to stress. Ideally, future research should focus on proximal environments that are unlikely influenced by genes. When selecting environments, researchers should pay attention to the salience of age or developmental period, as certain environments, such as peer relationships, may be more salient at different ages.

Increasing power to test for GxE

The results of this review underscore the need for larger and more diverse samples to test for GxE in depression among youth. The studies reviewed here may be too small to detect GxE effects, as the largest (n = 2,380) still fell short of the estimated number needed to adequately test for GxE (Munafo, Durrant, Lewis, & Flint, 2010; Munafo et al., 2009). Existing power estimates to detect GxE are also calculated for optimal conditions and do not take account of other factors that influence power; sample size requirements will vary depending on allele frequency, the magnitude of the GxE interaction, and the strength of the association between exposure and outcome, which is influenced by the reliability and validity of the environment and depression measures (Wong, Day, Luan, Chan, & Wareham, 2003). Future research should try to test for GxE in larger samples, including ongoing cohort studies and population-based epidemiological studies, and by pooling across existing samples. It is also important for power that more valid and reliable measures of environmental exposures and depressive outcomes are used. Where tradeoffs need to be made, gaining more precise measures of environment may be a more worthwhile endeavor than trying to acquire a larger sample (Institute of Medicine Board on Health Sciences Policy, 2006).

Improving the quality of genetic data

Many studies did not report information about the quality of their genetic data. This is problematic as genotyping errors, especially if they are large in magnitude, can significantly bias the results (Little et al., 2009). Therefore, we recommend future GxE research follow STREGA guidelines and provide more specific information about the collection, storage and analysis of DNA. Moreover, not all studies tested for Hardy–Weinberg equilibrium (HWE), which is whether genotype frequencies are consistent with random mating in the studied population. Some also tested for HWE for the entire sample and not in specific subgroups (i.e. stratified by race/ethnicity). Testing for HWE and reporting these results is important as it can provide information about deviation from its assumptions (Hartl & Clark, 2007).

Conclusion

We conducted this literature review in an attempt to understand the state of the science on GxE research in depression among children and adolescents and focus on the methodological approaches used across studies, as these features have not been previously described, including in recent meta-analyses. We also sought to discern the level of heterogeneity in GxE findings that existed across studies, assuming that mixed results would be widespread, as others had concluded (Munafo et al., 2009; Uher & McGuffin, 2008, 2010). The results of this systematic review suggest that the findings of GxE are perhaps not as mixed among youth as believed: most studies reviewed did find some evidence to support GxE, although this finding may reflect publication bias. However, a more salient issue that this review illustrates is the heterogeneity that exists in the methods and conceptual approaches used to conduct studies of GxE. These variations made it difficult to interpret and summarize findings and understand the nature of GxE effects during childhood and adolescence. Research completed by interdisciplinary teams, where there are equally high levels of expertise on genetic and environmental factors, will be important for reducing this heterogeneity and adequately capturing the joint contribution of genetic and environmental factors. We hope that the recommendations reported in Table 2 will provide a useful framework to guide future studies so that more research of better quality can be conducted, compared and replicated, and empirical knowledge of GxE effects among youth can advance in ways that generate new knowledge for prevention and intervention.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Interpreting main effects in the presence of an interaction (PDF).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Acknowledgements

Erin C. Dunn was supported by a predoctoral training grant from the National Institute of Mental Health (MH088074) and a fellowship from the Center on the Developing Child at Harvard University. This work was also supported by National Institute of Health grants DA022720, DA022720-S1 (S.G. and M.U.), MH088283-

01 (M.U.), HL081275 (S.V.S.), MH079799 (J.W.S.), MH078152 and MH082729 (S.G.), and MH070627 and MH078928 (K.C.K.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health. The authors thank Carol Mita for her assistance in conducting the literature search and Felton Earls, MD, Stephanie M. Jones, PhD, Laura Kubzansky, PhD, and Kate McLaughlin, PhD for their insights on conceptualizing 'environment'.

Correspondence to

Erin C. Dunn, Department of Society, Human Development, and Health, Harvard School of Public Health, 677 Huntington Avenue, 7th Floor, Boston, MA 02115, USA; Email: erindunn@hsph.harvard.edu

Key points

- There has been a tremendous growth in research on gene–environment interaction (GxE) within the last decade.
- Most published studies on GxE in depression among children and adolescents find evidence for GxE.
- However, existing research varies in the methods and analyses used to test for GxE; this variation prevents a synthesis of the current state of knowledge of the joint contribution of genetic and environmental influences.
- The recommendations provided in this article will enable more methodologically and conceptually consistent research on GxE in the future.

References

Aslund, C., Leppert, J., Comasco, E., Nordquist, N., Oreland, L., & Nilsson, K.W. (2009). Impact of the interaction between 5HTTLPR polymorphism and maltreatment on adolescent depression: A population-based study. *Behavior Genetics*, *39*, 524–531.

Avenevoli, S., Knight, E., Kessler, R.C., & Merikangas, K.R. (2008). Epidemiology of depression in children and adolescents. In J.R. Abela & B.L. Hankin (Eds.), *Handbook of depression in children and adolescents* (pp. 6–32). New York: Guilford.

Benjet, C., Thompson, R.J., & Gotlib, I.H. (2010). 5-HTTLPR moderates the effect of relational peer victimization on depressive symptoms in adolescent girls. *Journal of Child Psychology and Psychiatry*, *51*, 173–179.

Berkman, L.F., & Kawachi, I. (2000). *Social epidemiology*. New York: Oxford University Press.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Brooks-Gunn, J., & Duncan, G.J. (1997). Effects of poverty on children. *The Future of Children*, *7*, 55–71.

Brown, G.W., & Harris, T.O. (1978). *Social origins of depression: A study of psychiatric disorder in women*. New York: Free Press.

Brown, G.W., & Harris, T.O. (Eds.). (1989). Depression. In *Life events and illness* (pp. 49–93). London: Guilford.

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.

Caspi, A., Hariri, A.R., Holmes, A., Uher, R., & Moffitt, T.E. (2010). Genetic sensitivity to the environment: The case of the serotonin transporter gene and its implications for studying complex diseases and traits. *American Journal of Psychiatry*, *167*, 1–19.

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

Chapman, D.P., Whitfield, C.L., Felitti, V.J., Dube, S.R., Edwards, V.J., & Anda, R.F. (2004). Adverse childhood experiences and the risk of depressive disorders in adulthood. *Journal of Affective Disorders*, *82*, 217–225.

Chipman, P., Jorm, A.F., Prior, M., Sanson, A., Smart, D., Tan, X., & Easteal, S. (2007). No interaction between the serotonin transporter polymorphism (5-HTTLPR) and childhood adversity or recent stressful life events on symptoms of depression: Results from two community surveys. *American Journal of Medical Genetics Part B*, *144B*, 561–565.

Chipman, P., Jorm, A.F., Tan, X.T., & Easteal, S. (2010). No association between the serotonin-1A receptor gene single nucleotide polymorphism rs6295c/and symptoms of anxiety or depression, and no interaction between the polymorphism and environmental stressors of childhood anxiety or recent stressful life events on anxiety or depression. *Psychiatric Genetics*, *20*, 8–13.

Chorbov, V.M., Lobos, E.A., Todorov, A.A., Heath, A.C., Botteron, K.N., & Todd, R.D. (2007). Relationship of 5-HTTLPR genotypes and depression risk in the presence of trauma in a female twin sample. *American Journal of Medical Genetics Part B*, *144B*, 830–833.

Cicchetti, D., Rogosch, F.A., & Sturge-Apple, M.L. (2007). Interactions of child maltreatment and serotonin transporter and monoamine oxidase A polymorphisms: Depressive symptomatology among adolescents from low socioeconomic status backgrounds. *Development and Psychopathology*, *19*, 1161–1180.

Cicchetti, D., Rogosch, F.A., Sturge-Apple, M., & Toth, S.L. (2010). Interaction of child maltreatment and 5-HTT polymorphisms: Suicidal ideation among children from low-SES backgrounds. *Journal of Pediatric Psychology*, *5*(5), 536–546.

Cohen, S., Kessler, R.C., & Gordon, L.U. (Eds.). (1995). *Measuring stress: A guide for health and social sciences*. New York: Oxford University Press.

Cross-Disorder Phenotype Group of the Psychiatric GWAS Consortium, Craddock, N., Kendler, K., Neale, M., Nurnberger, J., Purcell, S., Rietschel, M. … Thapar, A. (2009). Dissecting the phenotype in genome-wide association studies of psychiatric illness. *British Journal of Psychiatry*, *195*, 97–99.

Diez Roux, A.V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health*, *88*, 216–222.

Dunn, E.C., & Masyn, K.E. (2011, May 31). Using multi-level factor analysis and ecometrics analysis to capture features of youth and their experiences of context: An illustration focusing on schools. Paper presented at the Society for Prevention Research Meeting, Washington, DC.

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

Eley, T.C., Sugden, K., Corsico, A., Gregory, A.M., Sham, P., McGuffin, P., Plomin, R., & Craig, I.W. (2004). Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Molecular Psychiatry*, *9*, 908–915.

Gershoff, E.T., & Aber, J.L. (Eds.). (2006). *Neighborhoods and schools: Contexts and consequences for the mental health and risk behaviors of children and youth*. New York: Psychology Press.

Gibb, B.E., Benas, J.S., Grassia, M., & McGeary, J. (2009). Children's attentional biases and 5-HTTLPR genotype: Potential mechanisms linking mother and child depression. *Journal of Clinical Child and Adolescent Psychology*, *38*, 415–426.

Gibb, B.E., Uhrlass, D.J., Grassia, M., Benas, J.S., & McGeary, J. (2009). Children's inferential styles, 5-HTTLPR genotype, and maternal expressed emotion-criticism: An integrated model for the intergenerational transmission of depression. *Journal of Abnormal Psychology*, *118*, 734–745.

Gilman, S.E., Kawachi, I., Fitzmaurice, G.M., & Buka, S.L. (2003). Family disruption in childhood and risk of adult depression. *American Journal of Psychiatry*, *160*, 939–946.

Goodman, E., Huang, B., Wade, T.J., & Kahn, R.S. (2003). A multilevel analysis of the relation of socioeconomic status to adolescent depressive symptoms: Does school context matter? *Journal of Pediatrics*, *143*, 451–456.

Greenland, S., & Rothman, K.J. (Eds.). (1998). Concepts of interaction. In *Modern epidemiology* (2nd edn, pp. 329–342). Philadelphia: Lippincott-Raven.

Gunnar, M., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology*, *58*, 145–173.

Guo, G., & Tillman, K.H. (2009). Trajectories of depressive symptoms, dopamine D2 and D4 receptors, family socioeconomic status and social support in adolescence and young adulthood. *Psychiatric Genetics*, *19*, 14–26.

Haeffel, G.J., Getchell, M., Koposov, R.A., Yrigollen, C.M., DeYoung, C.G., Klinteberg, B. … Grigorenko, E.L. (2008). Association between polymorphisms in the dopamine transporter gene and depression: Evidence for a gene-environment interaction in a sample of juvenile detainees. *Psychological Science*, *19*, 62–69.

Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology*, *1*, 293–319.

Hammen, C., Brennan, P.A., Keenan-Miller, D., Hazel, N.A., & Najman, J.M. (2010). Chronic and acute stress, gender, and serotonin transporter gene-enviornment interactions predicting depression symptoms in youth. *Journal of Child Psychology and Psychiatry*, *51*, 180–187.

Hartl, D.L., & Clark, A.G. (2007). *Principles of population genetics* (4th edn). Sunderland, MA: Sinauer Associates.

Hu, X., Oroszi, G., Chun, J., Smith, T.L., Goldman, D., & Schuckit, M.A. (2005). An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk. *Alcoholism: Clinical and Experimental Research*, *29*, 8–16.

Institute of Medicine Board on Health Sciences Policy (2006). *Study design and analysis for assessment of interactions, genes, behavior, and the social enviornment: Moving beyond the nature nurture debate* (pp. 161–180). Washington, DC: National Academies Press.

Jaffee, S.R., Caspi, A., Moffitt, T.E., Polo-Tomas, M., Price, T.S., & Taylor, A. (2004). The limits of child effects: Evidence for genetically mediated child effects on corporal punishment but not on physical maltreatment. *Developmental Psychology*, *40*, 1047–1058.

Jaffee, S.R., Caspi, A., Moffitt, T.E., & Taylor, A. (2004). Physical maltreatment victim to antisocial child: Evidence of an environmentally mediated process. *Journal of Abnormal Psychology*, *113*, 44–55.

Jaffee, S.R., & Price, T.S. (2007). Gene-environment correlations: A review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*, *12*, 432–442.

Jokela, M., Lehtimaki, T., & Keltikangas-Jarvinen, L. (2007). The influence of urban/rural residency on depressive symptoms is moderated by the serotonin receptor 2A gene. *American Journal of Medical Genetics Part B*, *144*, 918–922.

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

Kaufman, J., Yang, B.Z., Douglas-Palumberi, H., Grasso, D., Lipschitz, D., Houshayar, S., Krystal, J.H., & Gelernter, J. (2006). Brain-derived neurotropic factor 5-HTTLPR gene interactions and environmental modifiers of depression in children. *Biological Psychiatry*, *59*, 673–680.

Kaufman, J., Yang, B.Z., Douglas-Palumberi, H., Houshyar, S., Lipschitz, D., Krystal, J.H., & Gelernter, J. (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 17316–17321.

Kendler, K.S., & Baker, J.H. (2007). Genetic influences on measures of the environment: A systematic review. *Psychological Medicine*, *37*, 615–626.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., & Walters, E.E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602.

Khoury, M.J., Davis, R., Gwinn, M., Lindegren, M.L., & Yoon, P. (2005). Do we need genomic research for the prevention of common diseases with environmental causes? *American Journal of Epidemiology*, *161*, 799–805.

Kraft, P., Hunter, D.J. (2009). The challenge of assessing complex gene-environment and gene-gene interactions. In: Khoury, M.J., Bedrosian, S.R., Gwinn, M., Higgins, J.P.T., Ioannidis, J.P.A., & Little, J., eds. *Human genome epidemiology: Building the evidence for using genetic information to improve health and prevent disease*. (second edition). New York, NY: Oxford University Press, 165–187.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, *126*, 309–337.

Little, J., Higgins, J.P.T., Ioannidis, J.P.A., Moher, D., Gagnon, F., von Elm, E. … Birkett, N. (2009). Strengthening the reporting of genetic association studies (STREGA): An extension of the strengthening of reporting of observational studies in epidemiology (STROBE) statement. *Journal of Clinical Epidemiology*, *62*, 597–608.

Lopez-Leon, S., Janssens, A.C.J.W., Gonzalez-Suloeta Ladd, A.M., Del-Favero, J., Claes, S.J., Oostra, B.A., & van Duijn, C.M. (2008). Meta-analyses of genetic studies on major depressive disorder. *Molecular Psychiatry*, *13*, 772–785.

Lupien, S.J., McEwen, B.S., Gunnar, M.R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour, and cognition. *Nature Reviews Neuroscience*, *10*, 434–445.

Mair, C.F., Diez Roux, A.V., & Galea, S. (2008). Are neighborhood characteristics associated with depressive symptoms? A critical review. *Journal of Epidemiology and Community Health*, *62*, 940–946.

McEwen, B.S., & Seeman, T. (1999). Protective and damaging effects of mediators of stress: Elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences*, *896*, 30–47.

McLeod, J.D., & Shanahan, M.J. (1996). Trajectories of poverty and children's mental health. *Journal of Health and Social Behavior*, *37*, 207–220.

Moffitt, T.E., Caspi, A., & Rutter, M. (2006). Measured gene-environment interactions in psychopathology: Concepts, research strategies, and implications for research, intervention, and public understanding of genetics. *Perspectives on Psychological Science*, *1*, 5–27.

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 research: Genetic polymorphisms and life-stress polyprocedures. *Psychological Science*, *19*, 947–956.

Monroe, S.M., & Simons, A.D. (1991). Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders. *Psychological Bulletin*, *110*, 406–425.

Munafo, M.R., Durrant, C., Lewis, G., & Flint, J. (2009). Gene × environment interactions at the serotonin transporter locus. *Biological Psychiatry*, *65*, 211–219.

Munafo, M.R., Durrant, C., Lewis, G., & Flint, J. (2010). Defining replication: A response to Kaufman and colleagues. *Biological Psychiatry*, *67*, e21–e23.

Nilsson, K.W., Sjoberg, R.L., Leppert, J., Oreland, L., & Damberg, M. (2009). Transcription factor AP-2B genotype and psychosocial adversity in relation to adolescent depressive symptomatology. *Journal of Neural Transmisson*, *116*, 363–370.

Plomin, R., Haworth, C.M.A., & Davis, O.S.P. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*, 872–878.

Purcell, S., & Koenen, K.C. (2005). Environmental mediation and the twin design. *Behavior Genetics*, *35*, 491–498.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis method* (2nd edn). Thousand Oaks, CA: Sage.

Repetti, R.L., Taylor, S.E., & Seeman, T.E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin*, *128*, 330–366.

Rice, F., Harold, G., & Thapar, A. (2002). The genetic aetiology of childhood depression: A review. *Journal of Child Psychology and Psychiatry*, *43*, 65–79.

Risch, N., Herrell, R., Lehner, T., Kung-Yee, L., Eaves, L., Hoh, J. … Merikangas, K.R. (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.

Rudolph, K.D. (Ed.). (2009). *Adolescent depression* (2nd edn). New York: Guilford.

Rutter, M., Pickles, A., Murray, R., & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin*, *127*, 291–324.

Sampson, R.J., & Raudenbush, S.W. (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology*, *105*, 603–651.

Shaikh, S.A., Strauss, J., King, N., Bulgin, N.L., Vetro, A., Kiss, E. … Kennedy, J.L., & International Consortium for Childhood-Onset Mood Disorders (2008). Association study of serotonin system genes in childhood-onset mood disorders. *Psychiatric Genetics*, *18*, 47–52.

Shinn, M. (1990). Mixing and matching: Levels of conceptualization, measurement, and statistical analysis in community research. In P. Tolan, C. Keys, F. Chertok, & L. Jason (Eds.), *Researching community psychology* (pp. 111–126). Washington, DC: American Psychological Association.

Shyn, S. I., Shi, J., Kraft, J. B., Potash, J. B., Knowles, J. A., Weissman, M. M. … & Hamilton, S. P. (2011). Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Molecular Psychiatry*, *16*(2), 202–215.

Sjoberg, R.L., Nilsson, K.W., Nordquist, N., Ohrvik, J., Leppert, J., Lindstrom, L., & Oreland, L. (2006). Development of depression: Sex and the interaction between environment and a promoter polymorphism of the serotonin transporter gene. *International Journal of Neuropsychopharmaology*, *9*, 443–449.

Steinberg, L., & Morris, A.S. (2001). Adolescent development. *Annual Review of Psychology*, *52*, 83–110.

Subramanian, S.V., Jones, K., & Duncan, C. (2003). Multilevel methods for public health research. In I. Kawachi & L.F. Berkman (Eds.), *Neighborhoods and health* (pp. 65–111). New York: Oxford University Press.

Sullivan, P.F., de Geus, E.J., Willemsen, G., James, M.R., Smit, J.H., Zandbelt, T. … Penninx, B.W. (2009). Genome-wide association for major depressive disorder: A possible role for the presynaptic protein piccolo. *Molecular Psychiatry*, *14*, 359–375.

Tendulkar, S.A., Buka, S., Dunn, E.C., Subramanian, S.V., & Koenen, K.C. (2010). A multilevel investigation of neighborhood effects on parental warmth. *Journal of Community Psychology*, *38*, 557–573.

Turkheimer, E., D'Onofrio, B.M., Maes, H.H., & Eaves, L.J. (2005). Analysis and interpretation of twin studies including measures of the shared environment. *Child Development*, *76*, 1217–1233.

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin*, *126*, 78–108.

Uddin, M., Koenen, K., de los Santos, R., Bakshis, E., Aiello, A., & Galea, S. (2010). Gender differences in the genetic and environmental determinants of adolescent depression. *Depression and Anxiety*, *27*, 658–666.

Uher, R., & McGuffin, P. (2008). The moderation by the serotonin transporter gene of environmental adversity in the aetiology of mental illness: Review and methodological analysis. *Molecular Psychiatry*, *13*, 131–146.

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

U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. (2010). Child Maltreatment 2008. Available from http://www.acf.hhs.gov/programs/cb/stats_research/index.htm#can

Vaske, J., Makarios, M., Boisvert, D., Beaver, K.M., & Wright, J.P. (2009). The interaction of DRD2 and violent victimization on depression: An analysis by gender and race. *Journal of Affective Disorders*, *112*, 120–125.

Widom, C.S., DuMont, K., & Czaja, S.J. (2007). A prospective investigation of major depressive disorder and comorbidity in abused and neglected children grown up. *Archives of General Psychiatry*, *64*, 49–56.

Wong, M.Y., Day, N.E., Luan, J.A., Chan, K.P., & Wareham, N.J. (2003). The detection of gene-environment interaction for continuous traits: Should we deal with measurement error by bigger studies or better measurement? *International Journal of Epidemiology*, *32*, 51–57.

Xu, Y., Li, F., Huang, X., Sun, N., Zhang, F., Liu, P. … Zhang, K. (2009). The norepinephrine transporter gene modulates the relationship between urban/rural residence and major depressive disorder in a Chinese population. *Psychiatry Research*, *168*, 213–217.

Accepted for publication: 12 August 2011
Published online: 29 September 2011

Supporting Information

Interpreting Main Effects in the Presence of an Interaction

We found several studies that interpreted the main effect parameters for the environment, genotype, or both when these variables comprised the lower order terms of a higher order interaction that was also included in the regression model. Although it is possible to interpret these main effects in the presence of an interaction term, the interpretation is not very straight-forward. Given the clear interest in the interaction term (i.e. it is a "higher order" term for the GxE *interaction* comprised of "lower order" terms for genotype and environment), it is important to avoid misinterpreting the other parameters in the model that may lead to misleading conclusions. To facilitate the correct interpretation and reporting of main effects in the presence of an interaction, we provide two hypothetical examples.

Example 1

In the first hypothetical model, we have a regression model with four predictors:

$$y = \beta_0 + \beta_1(\text{genotype}) + \beta_2(\text{environment}) + \beta_3(\text{GxE}) + \beta_4(\text{covariates})$$

In this equation, genotype is an ordinal variable (0=l/l genotype; 1=s/l genotype; and 2=s/s genotype), which reflects an additive model of genetic inheritance, and environment is a binary variable (0=low risk environment; 1=high risk environment). The parameter corresponding to the third predictor (β_3) represents the cross-product of these two terms, forming the genotype-by-environment interaction (GxE). The parameter corresponding to the fourth predictor (β_4) represents a generic term for a set of potential covariates (e.g. sex, age, race/ethnicity).

When all four predictors are included in the model, the interpretation of the first parameter is as follows: β_0, the intercept, is the estimated average depressive symptoms score for youth living in a low risk environment with the l/l genotype who have a score of zero on the

covariates (note: a score of zero on the covariates is often impossible; centering variables prior to the analysis can address the lack of meaningful interpretation of the intercept in these cases). If there was no statistically significant interaction term in the model, then β_1 and β_2 would be interpreted in the usual way as main effects terms: β_1 would refer to the estimated average difference in depressive symptoms for each one unit difference in genotype (l/l vs. s/l vs. s/s), controlling for all other predictors in the model; and β_2 would refer to the estimated average difference in depressive symptoms for each one unit difference in environment (low risk vs. high risk), controlling for all other predictors in the model. However, the presence of the interaction term in the model changes the interpretation of these two coefficients.

If there is a significant interaction term in the model, interpreting β_1 or β_2 in the usual way will not be accurate. The new interpretation of the second and third coefficients is as follows: β_1 is the estimated average difference in depressive symptoms for each one unit difference in genotype (l/l vs. s/l vs. s/s), adjusting for covariates, *for youth living in a low risk environment* (where environment = 0). Likewise, β_2 is the estimated average difference in depressive symptoms associated with a one unit difference in environment (high risk vs. low risk), adjusting for covariates, *for youth with two long alleles* (when genotype = 0). β_3, which is the interaction term, is the parameter of interest in the GxE context. This parameter is interpreted as the estimated difference in the effect of genotype on depressive symptoms (β_1) associated with a one-unit different in environment (β_2) (comparing youth in high vs. low risk environments), adjusting for covariates. Conversely, β_3 can be interpreted as the estimated average difference in the effect of environment on depressive symptoms (β_2) by genotype (meaning, for a one unit difference in genotype (β_1)). Lastly, β_4 represents the expected difference in depressive

symptoms for a one unit difference in the covariates, adjusting for genotype, environment, and the interaction between them.

After treating genotype as an ordinal variable, researchers can then create indicator variables to make planned comparisons across genotype groups. For instance, one could compare specific differences between l/l and the other two genotypes:

$$y = \beta_0 + \beta_1(\text{s/l vs l/l}) + \beta_2(\text{s/s vs l/l}) + \beta_3(\text{environment}) + \beta_4(\text{s/l xE}) + \beta_5(\text{s/s xE})$$

To provide estimates of the main effects for genotype and environment, we would then combine across terms as follows:

		Environment	
Genotype		Low Risk (environment=0)	High Risk (environment = 1)
	l/l (genotype=0)	β_0	$\beta_0+\beta_3$
	s/l (genotype=1)	$\beta_0+\beta_1$	$\beta_0+\beta_1+\beta_3+\beta_4$
	s/s (genotype=2)	$\beta_0+\beta_2$	$\beta_0+\beta_2+\beta_3+\beta_5$

Therefore, as this example illustrates, to accurately report the main effect of environment or genotype in the presence of an interaction term, researchers must provide the parameter estimates for all terms included in the regression model.

Example 2

In a second hypothetical model, we have a regression model with four predictors:

$$y = \beta_0 + \beta_1(\text{genotype}) + \beta_2(\text{environment}) + \beta_3(\text{GxE}) + \beta_4(\text{covariates})$$

In this equation, genotype remains ordinal (0=l/l genotype; 1=s/l genotype; and 2=s/s genotype), however, environment is continuous (values ranging from 0-20, with lower scores indicating lower risk environments). The parameter corresponding to the third predictor (β_3) continues to represent the cross-product of these two terms, forming the GxE, and the fourth predictor (β_3) continues to represent a vector of covariates. When all four predictors are included in the model,

the interpretation of these parameters is as follows: β_0 is the estimated average depressive symptoms score for youth with the l/l genotype who has a score of zero on the environment measure and covariates; β_1 is the estimated average difference in depressive symptoms score for each one unit difference in genotype group (i.e. comparing s/l to l/l; and s/s to s/l) when the environment score equals zero, adjusting for covariates; β_2 is the expected average difference in depressive symptoms score for each one unit difference in environment for youth with the l/l genotype (i.e. when genotype=0), adjusting for covariates; β_3 is the expected average difference in the effect of genotype on depressive symptoms score comparing youth by levels of environments, adjusting for covariates (or conversely, the difference in the effect of environment on depressive symptoms by genotype) and; β_4 is the expected difference in depressive symptoms for each one unit difference in the covariates, adjusting for genotype, environment, and the interaction between them.

Therefore, to report the main effect of environment on depressive symptoms when the interaction term is included in the model, researchers must set the genotype value to zero; it is incorrect to interpret the parameter corresponding to environment without doing so. Similarly, to report the main effect of genotype in the presence of an interaction, the value corresponding to environment must also be set to zero.

Readers interested in learning more about interpreting interactions should consult outside resources (Aiken & West, 1991).

References

Aiken, L.S. & West, S.G. (1991). Multiple Regression: Testing and Interpreting Interactions. New York, NY: Sage Publications.