CRITICAL REVIEW OF GENE-ENVIRONMENT INTERACTIONS IN PSYCHIATRY:
EVIDENCE CONSISTENT WITH THE POSSIBILITY OF PUBLICATION BIAS, LOW
POWER, MULTIPLE TESTING, AND TYPE I ERRORS IN THE LITERATURE.

by

LARAMIE ELIZABETH DUNCAN

B.S. University of Georgia, 2003
M.A. University of Colorado, 2006

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the joint degree of
Doctor of Philosophy
Department of Psychology and Neuroscience

2010
This thesis entitled:
Critical Review of Gene-Environment Interactions in Psychiatry: Evidence Consistent with the Possibility of Publication Bias, Low Power, Multiple Testing, and Type I Errors in the Literature.
Written by Laramie Elizabeth Duncan
has been approved for the Department of Psychology and Neuroscience

Erik G. Willcutt

Matthew C. Keller

Soo H. Rhee

Matthew B. McQueen

Sona Dimidjian

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.
Duncan, Laramie Elizabeth (joint Ph.D. in Clinical Psychology and Neuroscience)

Critical Review of Gene-Environment Interactions in Psychiatry: Evidence Consistent with the Possibility of Publication Bias, Low Power, Multiple Testing, and Type I Errors in the Literature.

Thesis directed by Professor Matthew C. Keller and Erik Willcutt

**Introduction:** Measured gene-by-environment interaction studies have typically been conducted in a candidate GxE (cGxE) fashion, analogous to the candidate gene association studies that were used to search for genetic main effects. Such cGxE research in psychiatry has received widespread attention and acclaim, yet cGxE findings are also controversial. We were interested in determining whether cGxE findings were robust and might help to explain some of the missing heritability in psychiatric genetics or if, in aggregate, cGxE findings were consistent with the existence of publication bias, low power, multiple testing, and type I errors in the cGxE literature in psychiatry.

**Method:** We applied modified meta-analytic procedures to all published studies (to our knowledge) from the first decade of cGxE research in psychiatry, 2000-2009, collapsing across reported interactions in order to identify prevailing trends.

**Results:** Most novel cGxE studies were significant (96%), but only a minority of replication attempts were significant (32%). These findings are consistent with the existence of publication bias among novel GxE studies. There may also be publication bias among replication attempts because significant replication attempts had smaller sample sizes, on average, than null replication attempts. Furthermore, rates of positive replications, observed sample sizes, and power calculations suggested that studies were underpowered. Additionally, patterns of
expanding and branching hypotheses have been reported across time, and could be partially due to multiple testing and publication bias. Finally, through simulations we show that low power biases the observed form of interactions (i.e., ‘crossover’ versus ‘non-crossover’).

**Conclusion:** These results are consistent with the hypothesis that published studies provide a biased representation of all cGxE tests that have been conducted and also suggest that many reported positive findings may be type I errors.
Acknowledgments

This thesis could not have been completed without the expert and wonderful advice given by Dr. Matthew Keller and other faculty members: Drs. Soo Rhee, Matt McQueen, Gary McClelland, John Hewitt, Erik Willcutt, and Sona Dimidjian. Nor would it have been possible without the supportive intellectual community at the Institute for Behavioral Genetics in Boulder, Colorado; thank you.

Thanks also to generous support from the National Institutes of Health, the National Institutes of Mental Health, and the National Institute of Child Health and Human Development via support to LED (T32 grant # HD007289) and MCK (grant MH085812).
Contents

Chapter 1: Introduction........................................................................................................1

1.1 Reasons to expect GxEs in psychiatric genetics ........................................1

1.2 Reasons for caution regarding cGxE studies in psychiatry......................4

1.3 Motivation for thesis.............................................................................................6

Chapter 2: Studies Included.........................................................................................8

Chapter 3: Detecting publication bias in the cGxE literature in psychiatry.......10

3.1 Putative publication bias among novel cGxE studies.............................10

3.2 Putative publication bias among replication attempts.........................12

3.3 Implications.........................................................................................................14

Chapter 4: Power to detect GxEs of various effect sizes.................................16

4.1 Estimating the effect size of GxEs.................................................................16

4.2 Implications of power calculations and observed sample sizes..........17

4.3 Rate of type I errors (false positives) in the literature.........................19

Chapter 5: Potential ‘signatures’ of multiple testing, type I errors, low power, and publication bias in the cGxE literature in psychiatry.........20

5.1 Expanding and branching hypotheses in the cGxE literature.............21

5.2 Form of interactions: reversals of effect (i.e. ‘crossovers’) dominate the cGxE literature in psychiatry........................................25

5.2.1 Implications of reversal interactions.......................................................28

5.2.2 Alternative explanations for reversal interactions in the literature.................................................................29

5.2.3 Synthesis of information regarding the form of published interactions.........................................................30
Chapter 6: Conclusions

6.1 Reasons why a preponderance of type I errors in the cGxE literature in psychiatry may not be surprising

6.2 Limitations

6.3 Recommendations for cGxE studies

6.4 Future directions

6.5 Final synopsis

Bibliography

Appendix
Figures

1. Evidence consistent with hypothesis of publication bias among novel cGxE studies ..........11
2. Evidence consistent with the hypothesis of publication bias among replication attempts of cGxE studies ..................14
3. Power calculations for various GxE effect sizes (top graph) and histogram of observed sample sizes (bottom graph) .................17
4A. Expanding hypotheses ..............................................................22
4B. Branching hypotheses ..............................................................23
5. Reversals of effects in observed and simulated data .....................26
6. Demonstration of different patterns of reversals in hypothetical interactions ..............................................................27
Chapter 1
Introduction

Numerous gene-environment interactions (GxEs) have been reported in psychiatry, appearing in high-profile journals such as Science, PNAS, and JAMA (Binder, et al., 2008; Caspi, et al., 2002; Caspi, et al., 2003; Kaufman, et al., 2004). They appear to explain some of the ‘missing heritability’ that has been elusive in psychiatric genetics despite extensive linkage, candidate gene, and genome-wide association studies (GWASs) (Manolio, et al., 2009). But critics wonder if these results are too good to be true. Roughly a decade ago, similar papers that described direct effects of candidate genes (i.e., association studies) were published in Science (Lesch, et al., 1996), Nature Genetics (Lesch, et al., 1996), and other prominent journals. Subsequently, most of the associations failed to replicate (Colhoun, McKeigue, & Smith, 2003; Kluger, Siegfried, & Ebstein, 2002; Munafo, et al., 2003; Schinka, Letsch, & Crawford, 2002), and based on the observed rate of failures to replicate, Colhoun estimated that 95% of positive association results were actually type I errors (Colhoun, et al., 2003). Does the same fate await GxE studies? This thesis examines studies from the first decade of candidate GxE (cGxE) research in psychiatry (2000-2009) and finds evidence consistent with the existence of publication bias, low power, multiple testing, and type I errors among positive cGxE findings in psychiatry.

1.1 Reasons to expect GxEs in psychiatric genetics

Gene-environment interactions (GxEs) occur when the effect of the environment depends on one’s genotype or, equivalently, when the effect of one’s genotype depends on the
environment. There are strong theoretical reasons to believe that GxEs exist (Moffitt, Caspi, & Rutter, 2005; Rutter, Thapar, & Pickles, 2009). For one, GxE effects can be conceptualized as main effects of genes if the phenotype is ‘response to the environment.’ For example, ‘change in depression in response to stressful life events’ would be a ‘response to the environment’ phenotype. Given that genetic influences on behavior are ubiquitous e.g., (Eaves, Last, Young, & Martin, 1978), it stands to reason that ‘response to the environment’ phenotypes would also be heritable. Thus, the existence of GxEs can be inferred.

More importantly however, many lines of evidence support the existence of GxEs. In animals, the Siamese cat provides a validated example of a GxE in which genetic background and temperature interact to produce the distinctive appearance of the cat (Imes, Geary, Grahn, & Lyons, 2006; Lyons, Imes, Rah, & Grahn, 2005). A temperature-dependent enzyme causes the face, tail, paws, and other extremities of the cat (i.e. the colder parts) to be darker colored than the body (Imes, et al., 2006; Lyons, et al., 2005). In humans, a well-known GxE involves phenylketonuria, PKU, a condition which only develops in individuals who consume a typical diet and have risk genotypes; the disorder is treatable among individuals with risk genotypes if a low phenylalanine diet is consumed (Pietz, et al., 1999). For these and other examples, specific genetic risk loci and environmental variables involved in the interaction have been identified (Imes, et al., 2006; Lyons, et al., 2005; Pietz, et al., 1999).

Candidate GxE (cGxE) studies, in which a risk locus is hypothesized and all variables in the interaction are measured, attempt to characterize specific interactions (Moffitt, et al., 2005) such as those mentioned in the preceding paragraph, and cGxE studies involving psychiatric phenotypes are the subject of this thesis. We are aware of 97 cGxE studies conducted in the first decade of such research in psychiatry - a majority of which reported positive results. The most
often studied cGxE in all of psychiatry, first examined by Caspi and colleagues, involves a putative interaction between a serotonin transporter related polymorphism (5-HTTLPR) and stressful life events in predicting depression (‘Depression ~ 5-HTTLPR x Stressful life events’) (Caspi, et al., 2003). This interaction and closely related variants have been studied over 30 times, with mixed results (Duncan, Willcutt, & Keller, 20--) . To date, two meta-analyses have failed to find support for the original interaction (Munafo, Durrant, Lewis, & Flint, 2009; Risch, et al., 2009), suggesting that the original report was a type I error or that the effect is not robust. Only one other interaction has been subjected to meta-analysis in cGxE research in psychiatry, and it also involved an interaction initially reported by Caspi and colleagues: ‘Antisocial problems ~ MAOA x Maltreatment’ (Caspi, et al., 2002; Kim-Cohen, et al., 2006). While the results of the meta-analysis (co-authored by Caspi and colleagues) were favorable, we believe that they should be regarded with some caution for three reasons. First, the original study (Caspi, et al., 2002) was included in the meta-analysis (Kim-Cohen, et al., 2006). Second, a study that received widespread criticism was included in the meta-analysis (Foley and Riley 2007; Joober, Sengupta et al. 2007; Thapar, Harold et al. 2007)(Kim-Cohen, et al., 2006). Third, the selection of which results to use (from new data reported in conjunction with the meta-analysis) (Kim-Cohen, et al., 2006) was debatable because significant results were selected instead of non-significant results, even though the interaction yielding non-significant results was arguably more similar to the original interaction. While the results of the meta-analysis were significant, an independent meta-analysis is warranted. A number of other interactions have been examined in multiple studies (Duncan, et al., 20--), and they serve as excellent candidates for replication attempts and eventually meta-analyses, which will provide empirical evidence regarding whether or not cGxE findings in psychiatry are robust.
While no specific cGxEs in psychiatry have been unequivocally supported, results from latent variable analyses suggest that interactions involving psychiatric phenotypes exist. Latent variable analyses (e.g. twin studies) estimate genetic risk based on genetic relatedness, as opposed to cGxE studies, which measure specific polymorphisms directly. Furthermore, by measuring genetic effects in aggregate, such studies may be able to detect cumulative genetic effects, composed of many small effects, which might be difficult to detect individually. Latent variable studies have shown that people with low genetic risk are less likely to develop depression in response to stressful life events (Kendler, et al., 1995) and that people with high genetic risk are more susceptible to developing conduct problems in response to maltreatment (Jaffee, et al., 2005). Studies such as these provide a compelling rationale to search for specific cGxEs in psychiatry. Furthermore, latent variable studies indicate that both genes and the environment are important etiological factors in psychiatric disorders. Heritability estimates for depression, bipolar disorder, schizophrenia, eating disorders, and ADHD range from approximately 40 to 90% (Burmeister, McInnis, & Zollner, 2008; Craddock, O'Donovan, & Owen, 2005; Faraone, 2000; Sullivan, Kendler, & Neale, 2003; Sullivan, Neale, & Kendler, 2000). Such heritability estimates bolstered the rationale to hunt for direct effects of polymorphisms, and they are often included in the rationale for testing specific cGxEs e.g., (Becker, El-Faddagh, Schmidt, Esser, & Laucht, 2008; Moffitt, Caspi, & Rutter, 2006).

1.2 Reasons for caution regarding cGxE studies in psychiatry

Despite the reasons supporting the search for cGxEs in psychiatry, there are also reasons to be skeptical of such findings. One criticism of cGxE research stems from the fact that
extensive searches for genetic main\textsuperscript{1} effects in psychiatry, using linkage, candidate gene, and GWA studies, have yet to explain much of the heritability of psychiatric disorders (Burmeister, et al., 2008; Colhoun, et al., 2003; Collins, 2010; Manolio, et al., 2009). Given that power to detect interactions is lower than power to detect main effects (assuming equal effect sizes) (McClelland & Judd, 1993), detecting robust GxEs may be even more difficult than detecting genetic main effects. The fundamental problem described here is low power.

Another problem for cGxE studies is the inherently high risk of type I errors, (i.e., ‘false positives’ that indicate the existence of an interaction when there is no real interaction). The high risk of type I errors is due to the virtually unlimited number of GxE hypotheses that could be tested, coupled with the standard practice of using $p < 0.05$ as the threshold for statistical significance. The number of testable GxE hypotheses is the \textit{product} of the number of genotypic, phenotypic, and environmental variables available for study; thus, the magnitude of the multiple testing problem is demonstrated by the fact that there are over a million genetic polymorphisms available for study by researchers who are interested in tens of psychiatric disorders and tens of environmental variables in cGxE studies (Duncan, et al., 20--). Therefore 100 million is a conservative estimate for the number of testable GxE hypotheses (1 million * 10 * 10). Even with a more modest example, which is likely to be more representative of variables available to individual researchers, of two genetic polymorphisms, two phenotypes, and five environmental variables, 20 tests can be conducted. On average, one test will be ‘significant’ (i.e. $p < 0.05$) if corrections are not made for multiple testing. Furthermore, given that interaction hypotheses can be tested in many different ways using alternative measures of variables and alternative ways of coding variables (i.e. a continuous variable can be dichotomized, additive genetic effects can be

\textsuperscript{1} Throughout this paper ‘main’ effect refers to the effect of a genetic or environmental variable in a regression model that does not include an interaction term.
recoded to reflect recessive, dominant, or other effects, etc.), many more statistical tests are possible. As Sullivan (2007) articulated with respect to candidate gene association studies, the variety of options for testing alternative hypotheses makes finding ‘significant’ results highly likely, and the problem is even greater in cGxE studies because of the additional variable.

Given the magnitude of this multiple testing problem, the concern is that a tremendous number of GxE hypotheses will be tested, but only those that are ‘significant’ (without correction for multiple testing) will make their way to publication. This is the phenomenon of publication bias, and it produces a distorted representation of findings in a particular area of study. It also increases the field-wise type I error rate (Ioannidis, 2005). Thus, multiple testing, publication bias, and type I errors are a major concern for cGxE research.

Another concern about cGxE research, which pertains more broadly to the detection of all interactions, is the possibility that an ‘interaction’ may actually be an artifact due to distributional properties of variables or scaling effects (Eaves, 2006). Such artifactual interactions can be detected at alarmingly high rates, as high as 100 percent of the time, given certain conditions (Eaves, 2006). While these conditions are not likely (indeed, they cannot be true for all interactions because numerous null reports of cGxEs are available), they illustrate the point that statistical interactions can be misleading. Furthermore, as noted by Eaves, an artifact can also be replicated, perhaps even more easily than a real interaction, and thus consistent replication attempts may not be evidence of true interactions (Eaves, 2006).

1.3 Motivation for thesis

In a companion paper that reviews the first decade of cGxE research in psychiatry (2000-2009), we summarize the replication status of all interactions, including the fact that only two
interactions have been studied enough times to warrant meta-analyses (Caspi, et al., 2002; Caspi, et al., 2003; Duncan, et al., 20--; Kim-Cohen, et al., 2006; Munafo, et al., 2009; Risch, et al., 2009). Therefore, strict meta-analyses provide a limited window through which to view all of cGxE research in psychiatry (Duncan, et al., 20--; Rutter, et al., 2009). Consequently, we collapse across interactions in this paper and employ modified meta-analytic procedures, as allowed by the data. Our goal is to determine if the literature, in aggregate, supports the validity of cGxE findings or whether it is consistent with the existence of publication bias, low power, type I errors, and multiple testing in the cGxE literature in psychiatry.
Chapter 2
Studies Included

We attempted to find all cGxE studies conducted in the first decade of cGxE research in psychiatry, 2000-2009 (inclusive). Studies were identified using Medline, Pubmed, and Google Scholar using combinations of the following search terms: gene*, environment*, interact*, psych*, and moderate*. In addition, bibliographies were hand-searched for cGxE studies. In order to be included, the phenotype in a cGxE study had to be a DSM-IV diagnosis (Association, 1994) or a closely-related construct (e.g., neuroticism). Only observational, as opposed to experimental, studies were included; thus all pharmacogenetic studies were excluded. Studies were only included if there was variation across participants for phenotypic, genetic, and environmental variables. When follow-up studies of previously published reports were available, we included only the follow-up studies. In total, 97 studies, encompassing 101 samples, met inclusion criteria. Details about these studies are available in (Duncan, et al., 20--), and a chart of included studies is available in the appendix. To the best of our knowledge, these are all of the cGxE studies published in the first decade of cGxE research in psychiatry (Aguilera, et al., 2009; Altink, et al., 2008; Amstadter, et al., 2009; Bakermans-Kranenburg & van Ijzendoorn, 2006; Bau, Almeida, & Hutz, 2000; Becker, et al., 2008; Bet, et al., 2009; Binder, et al., 2008; Blomeyer, et al., 2008; Bradley, et al., 2008; Brookes, et al., 2008; Brummett, et al., 2008; Caspi, et al., 2002; Caspi, et al., 2005; Caspi, et al., 2003; Cervilla, et al., 2007; Chipman, et al., 2007; Chorbov, et al., 2007; Chotai, Serretti, Lattuada, Lorenzi, & Lilli, 2003; Cicchetti, Rogosch, & Sturje-Apple, 2007; Covault, et al., 2007; Dick, et al., 2006; DiLalla, Elam, & Smolen, 2009; Ducci, et al., 2008; Eley, et al., 2004; Foley, et al., 2004; Fox, et
Chapter 3
Detecting publication bias in the cGxE literature in psychiatry

Publication bias, the tendency to publish significant findings instead of null results\(^2\), is problematic because it produces a distorted representation of findings in an area of study. Publication bias can occur if authors decide not to submit null results or if journal editors opt not to publish them. For example, if twenty samples were tested for the presence of an ‘interaction’ that was not real, on average, one of them would yield a p-value < .05. Due to publication bias, the ‘significant’ result (i.e. the one with p<.05) might be the only one to be published. Examined alone, the significant finding might appear compelling, whereas the finding would appear compatible with chance if all twenty tests had been reported together. This illustrates how publication bias can lead to inappropriate interpretations about the likelihood that a finding could have occurred by chance alone.

3.1 Putative publication bias among novel cGxE studies
An indirect way to determine whether or not publication bias has occurred in the cGxE literature is to examine the rate of null results among novel cGxE studies compared to the rate of null results among replication attempts\(^3\). Publication bias could work as follows: null, novel results might be deemed uninteresting or difficult to interpret because of power, therefore they might not be published. In contrast, following a significant report of a cGxE, both null and

\(^2\) Other forms of publication bias are possible, but we only consider positive publication bias (i.e. disproportionate publication of positive results) in this report.

\(^3\) For all studies, we identified interactions mentioned or alluded to in the abstract. For replication attempts, results were considered to be null if the originally reported interaction was not replicated.
significant results for replication attempts might be deemed to be worthy of publication. Therefore, the rate of null results would be lower among novel reports than replication attempts. This is exactly what we found in the cGxE literature in psychiatry. As shown in figure 1, a mere 4% of novel studies were null versus 68% of replication attempts (\( \chi^2 = 35.6, \text{df} = 1, p = 2.4 \times 10^{-9} \)). The p-value reported here should be interpreted with caution because many of the replication attempts were not independent (e.g., the ‘Depression ~ 5-HTTLPR x Stressful life events’ interaction was tested multiple times, violating the assumption of independence). Therefore we re-ran the analysis, excluding all but the first published replication attempt for each interaction. Despite the reduction in replication attempts available for analysis and the attendant loss of power, the results were still highly significant (\( p = 3.3 \times 10^{-5} \)). These results suggest that publication bias may exist among novel GxE studies.

Figure 1. Evidence consistent with hypothesis of publication bias among novel cGxE studies
3.2 Putative publication bias among replication attempts

The preceding analyses relied on the assumption that replication attempts are a better representation of reality than novel reports, but it is also possible that publication bias exists among replication attempts. If so, our results would be conservative because 68% would be an underestimate of the true percentage of replication attempts that are null. To evaluate this possibility, we focused on replication attempts of previously reported, significant interactions. There were 39 such replication attempts, which corresponded to 12 originally-reported interactions. We reasoned that two relationships might be observed between sample size and the significance (significant versus null) of replication attempts. First, if the cGxE effects are real, then significant (p< .05) replication attempts should have larger sample sizes, on average, than null replication attempts given their greater power (Cohen, 1992)\(^4\). Alternatively, if the originally reported interaction was a type I error and there was publication bias among replication attempts, then significant replication attempts might have smaller sample sizes, on average, than null replication attempts. This would be the case if a cGxE finding was a type I error, and many research groups attempted to replicate the finding. ‘Replications’ would be detected at a rate of approximately alpha (i.e. 5%) and due to publication bias, small studies might be preferentially published if they had significant results. Meanwhile, large replication attempts might be published irrespective of results (yielding a higher fraction of null results among larger than smaller studies).

Figure 2 shows evidence consistent with the hypothesis of publication bias among replication attempts. The mean sample size of significant replication attempts was 217, whereas

\(^4\) This assumes that the interaction is real, has a large enough effect size to be detected, and that there are no systematic differences between smaller and larger samples. Potential confounding differences are discussed in the interpretation of these results.
the mean sample size of null replication attempts was 706 ($W=49, p=.001$). The nonparametric Wilcoxon rank-sum test was used because sample sizes were not normally distributed. However, parametric tests, with and without removal of outliers, were also significant. The same relationship that was observed in this aggregate analysis was also observed for each of the four interactions that allowed individual analysis, although none of the individual analyses reached statistical significance. Thus, the observation of smaller sample sizes among significant replication attempts is not confined to one interaction. If publication bias is the correct explanation for these results, than more than 68% of replication attempts were actually null. As the true percentage of replication attempts that do not achieve significant p-values ($p < 0.05$) approaches 95%, it becomes more and more plausible that original findings were type I errors.

However, we note that publication bias may not be the correct explanation for the fact that significant replication attempts had smaller sample sizes than null replication attempts. It is also possible that there are systematic differences between larger and smaller studies. For example, quality of measurement may vary with sample size. If smaller studies had less measurement error, then they might afford more power than larger studies. Differences in study populations could also be to blame. These and other explanations cannot be ruled out.
**Figure 2. Evidence consistent with the hypothesis of publication bias among replication attempts of cGxE studies**

Null replication attempts had larger sample sizes than significant replication attempts

3.3 Implications

We contend that publication bias provides a parsimonious explanation for our results. If true, many more tests of cGxE hypotheses have been conducted than reported in the literature. In essence, publication bias may be acting as a filter that catches the 5% of cGxE hypotheses that, by chance alone, reach statistical significance. To the extent that this is true, the field-wise type I error rate will be elevated and cGxE hypotheses will appear more credible than they actually are. As suggested by Sullivan regarding novel reports from association studies, we recommend that *novel* reports of cGxEs be treated with “exceptional caution” (Sullivan, 2007). If publication bias exists among replication attempts, as our results suggest that it may, then
replication attempts deserve cautious interpretation as well, and efforts to determine if cGxE findings are robust will be hampered.
Chapter 4

Power to detect GxEs of various effect sizes

Power to detect GxEs depends on sample size, the effect size of the interaction, ascertainment strategy, and the distribution of genetic and environmental variables. Figure 3 depicts power to detect GxEs across the range of sample sizes from published cGxE studies for three effect sizes: ‘tiny’ (accounting for 0.1% of phenotypic variation), ‘small’ (1%), and ‘large’ (10%). A histogram of sample sizes from published cGxE studies is given in the bottom portion of Figure 3, displaying the positively skewed distribution of sample sizes with a median of 345 but a range of 30 - 4,175. For these power calculations we assumed unselected samples and binary distributions for the genetic and environmental variables (with p=0.5).

4.1 Estimating the effect size of GxEs

In order for power calculations to be useful, however, we need estimates of the true effect sizes of GxEs. Given that genetic main effects explain less than one percent of phenotypic variance each for psychiatric disorders (Plomin & Davis, 2009), we reasoned that the effect sizes of GxEs might be comparable, hence our use of 1% and 0.1% as estimates of effects sizes for GxEs. Furthermore, GxEs can be conceptualized as main effects of genes if the phenotype is ‘response to the environment’. Viewed this way, genetic effects on ‘response to the environment’ phenotypes may be comparable to other genetic effects in psychiatric genetics. Thus, an upper bound of 1% for the effect size of GxEs might be appropriate. However, true effect sizes of specific GxEs are not known so we also included an effect size ten times larger than any known genetic main effect, a GxE accounting for 10% of phenotypic variance.
4.2 Implications of power calculations and observed sample sizes

The implications of these power calculations, using observed sample sizes, are sobering. For example, assuming an effect size of 1%, two thirds of studies were underpowered. Moreover, assuming an effect size of 0.1%, then all studies were underpowered. With power estimates this low, the arbitrary choice of alpha level becomes the most important factor in
determining power; thus power would be substantially lower with lower alpha levels. In summary, it is very likely that most published cGxE studies were underpowered, perhaps severely so.

One of the problems with low power is that it necessarily increases the fraction of type I errors among positive reports (Green, et al., 2008). Detection of type I errors can be more likely than detection of true effects if power is sufficiently low and/or if the fraction of tested hypotheses that are actually true is sufficiently low (Ioannidis, 2005). In psychiatric genetics, we have reason to believe that both power to detect GxEs and the fraction of hypotheses that are actually true may be low. First, as illustrated above, power to detect GxEs may have been very low in most published studies. Second, candidate gene association studies often failed to replicate, and in retrospect, it was estimated that 95% of association findings were type I errors (Colhoun, et al., 2003). The same may prove to be true of candidate GxE studies, although replication attempts are necessary to test this empirically. Nevertheless, the possibility that most or all positive reports of cGxEs in psychiatry may be type I errors must be acknowledged.

Finally, we considered the possibility that GxEs actually had extremely large effects sizes. In order for the median sample size to yield adequate power (80%), the GxE would need to account for approximately 27% of phenotypic variance. The problem with this explanation is that it is inconsistent with the observed data: only 32% of replication attempts were significant as compared to 80% of replication attempts expected be significant with 80% power. Instead, the results of our power calculations further support the hypothesis of publication bias in the cGxE literature because publication bias could explanation the high rate of positive reports in the GxE literature (i.e. 96% of novel GxE studies), despite low power (power = alpha because hypotheses are actually incorrect). While these results are important because they apply across all cGxE
studies in psychiatry, they may not be applicable to any particular GxE. Furthermore, as discussed in the conclusion section, systematic differences between novel studies and replication attempts could also account for these results.

4.3 Rate of type I errors (false positives) in the literature

Given findings regarding power and publication bias, it is important to consider the prevalence of type I errors in the cGxE literature in psychiatry. A rough but conservative estimate is that 65% of novel positive reports should actually be null to the extent that the following assumptions are correct: We used the observed percentage of positive replication attempts (32%) as a proxy for power\(^5\). Given 32% power, we calculated the number of tests (134) necessary to achieve the observed number of statistically significant, novel GxEs (N=43). This implies that 3.1 (=134/43) statistical tests were conducted per positive report; thus Bonferroni correction (Bonferroni 1936) yields a significance threshold of \(p < .016\). Sixty-five percent of observed \(p\)-values for novel GxEs did not survive this correction for multiple testing.

While this estimate may be useful as a ‘ballpark’ figure, it is extremely rough and should be interpreted with a great deal of caution; there is reason to believe that it is an underestimate (because of publication bias among replication attempts), but it may also be an overestimate if idiosyncratic factors account for the low rate of positive replication attempts.

\(^5\) However, 32% is likely to be an overestimate of power because of putative publication bias among replication attempts.
Chapter 5

Potential ‘signatures’ of multiple testing, type I errors, low power, and publication bias in the cGxE literature in psychiatry.

In the previous two sections we presented evidence and analyses consistent with the presence of publication bias in the cGxE literature and low power to detect GxEs. These problems imply a high rate of type I errors in the literature. If true, we reasoned that ‘signatures’ of these problems might exist across the literature. In this section we discuss three possible signatures of these problems. However we suggest caution in interpreting putative signatures because alternative explanations are possible. Indeed, they are certainly true to some extent, as noted in the following sections.

Imagine that a researcher tests a previously reported hypothesis and fails to find support for it. Publishing null results is one option, but another option is to test alternative hypotheses involving minor alterations to the original hypothesis. An incomplete list of possible alterations, all of which were reported in the literature, is as follows: 1) genetic variables can be recoded to reflect additive, dominant, recessive, and other effects. Phenotypic and environmental variables can be measured continuously and then dichotomized or turned into ordinal variables; 2) additional polymorphisms in and around the gene of interest and haplotypes can be added to models, as can covariates and additional phenotypic, genetic and environmental variables; 3) alternative phenotypic, genotypic, and environmental variables can be tested; for example, neuroticism can be tested instead of depression; 4) finally, effects can be tested in sub-samples instead of full samples. Notably, each of these alterations results in a multiplicative increase in the number of statistical tests conducted (if all hypotheses are tested systematically),
demonstrating the magnitude of the multiple testing problem and the virtual inevitability of finding a p-value < .05. An analogous point about association studies was made by Sullivan (2007).

5.1 Expanding and branching hypotheses in the cGxE literature

If researchers responded to finding null results for an original hypothesis, such as the ‘Depression ∼ 5-HTTLPR x Stressful life events’ interaction by adding variables and substituting alternative variables for those in the original interaction, then we would expect to see patterns of expanded and branched hypotheses, respectively, as depicted in Figure 4. Figure 4A shows a series of expanding hypotheses in which additional variables are added to previously-reported hypotheses in subsequent studies. This particular series begins with a direct association hypothesis, ends with a 4-way interaction, and encompasses the most often studied cGxE hypothesis in all of psychiatry: ‘Depression ∼ 5-HTTLPR x Stressful life events’ (Caspi, et al., 2003; Duncan, et al., 20--; Kaufman, et al., 2006; Kim, et al., 2007; J. A. Lasky-Su, Faraone, Glatt, & Tsuang, 2005; Wichers, et al., 2008). Notably, meta-analyses have failed to support the direct association hypothesis and the two-way (5-HTTLPR x Stressful life events) interaction hypothesis (J. A. Lasky-Su, et al., 2005; Munafo, et al., 2009; Risch, et al., 2009), findings consistent with the possibility that these original reports were type I errors. The three- and four-way interactions have not yet been subjected to meta-analyses, so we do not know if they will prove to be robust. However, as the number of terms in an interaction increases, power to detect such interactions (with equivalent effect sizes) decreases; so power to detect 3-way interactions is lower than power to detect 2-way interactions, and so on (McClelland & Judd, 1993).
Figure 4A. Expanding hypotheses

Hypotheses can expand via the addition of new variables.

- Direct association: Depression $\sim$ 5-HTTLPR
- 2-way interaction: Depression $\sim$ 5-HTTLPR $\times$ Stressful life events
- 3-way interaction: Depression $\sim$ 5-HTTLPR $\times$ BDNF $\times$ Stressful life events
- 4-way interaction: Depression $\sim$ 5-HTTLPR $\times$ BDNF $\times$ Stressful life events $\times$ Social support

Phenotypic variables
Genotypic variables
Environmental variables
Hypotheses can branch via the substitution of alternative variables
5-HTTLPR=serotonin transporter-linked polymorphic region, ADHD=attention deficit hyperactivity disorder, PTSD=post-traumatic stress disorder, 5-HTR=serotonin receptor, BDNF=brain-derived neurotrophic factor, CRHR1=corticotropin releasing hormone receptor
Forty-nine branched hypotheses are depicted in Figure 4B in which alternative phenotypic, genotypic, or environmental variables are substituted for the corresponding variable in the original ‘Depression ~ 5-HTTLPR x Stressful life events’ interaction. A majority, but not all of these hypotheses, were reported as significant findings. The origin of these branched and expanded hypotheses has implications for the progress of scientific knowledge. If positive reports of branched and expanded hypotheses were ‘discovered’ via multiple testing and failure to correct for statistical tests, then they are likely to slow scientific progress, confuse consumers of the literature, and waste research dollars on unpromising replication attempts. However, branched hypotheses may also be constructive additions to the literature. Carefully planned and targeted branched hypotheses may be tested when precise replication attempts are precluded by the limitations of extant datasets. There may also be theoretical reasons to test particular branched hypotheses, including efforts to determine the boundary conditions of an originally reported hypothesis. For example, Kendler reported testing an ‘Anxiety ~ 5-HTTLPR x Stressful life events’ interaction in order to assess the specificity of the original interaction involving depression, and this is a useful contribution to the literature (Kendler, et al., 2005). In addition, researchers may attempt to replicate previously reported branched hypotheses. Thus, branching hypotheses may be beneficial, but continuously branching and expanding hypotheses, in the absence of robust effects, may serve as a signature of multiple testing, low power and/or type I errors, and publication bias in the cGxE literature. Future replication attempts and meta-analyses will be needed to determine if branched and expanded hypotheses replicate or not.
5.2 Form of interactions: Reversals of effect (i.e. ‘crossovers’) dominate the cGxE literature in psychiatry

A third potential signature of publication bias and low power and/or type I errors may exist in the observed *form* of published interactions. Figure 5 displays the observed proportion of interactions with particular forms\(^6\). We use the terms ‘reversal of genetic effect’ and ‘reversal of environmental effect’ to describe the form of interactions because these terms allow for more nuanced descriptions of the form of interactions than the terms ‘crossover’ and ‘non-crossover’ do (see Figure 6). As shown in Figure 5, only 8% of observed interactions involved no reversals (i.e., ‘non-reversal interactions’). Ninety percent of interactions involved reversals of genetic effect and nearly half of interactions (45%) involved reversals of environmental effect. We submit these observations to the literature for others to consider.
Figure 5. Reversals of effects in observed and simulated data

For this analysis we examined only novel reports of interactions because we didn’t want the form of any particular interaction to be overrepresented in this sample (i.e. we attempted to gather a representative sample of GxEs in psychiatry). For the same reason, we examined only the first graphed interaction from each novel study.
Figure 6. Demonstration of different patterns of reversals in hypothetical interactions

Hypothetical interactions. Boxes A and B demonstrate how the terms ‘crossover’ and ‘non-crossover’ are ambiguous. A crossover interaction is observed in A, but no crossover is apparent in B. Importantly, the only difference between A and B is the arbitrary choice of which variable (environmental or genetic) is plotted on the x-axis. Reversal language provides a more nuanced description of the form of the interaction graphed in A and B. This interaction involves a reversal of genetic effect (as evidenced by crossed lines in B) and no reversal of environmental effect (in B, severe maltreatment is always the risk environment). An interaction involving reversal of environmental but not genetic effect is also possible (not shown), as are interactions involving no reversals of effect (C) and reversal of both effects (D). In total this yield four possible forms of interactions.
5.2.1 Implications of reversal interactions

If these findings are true, then they are quite intriguing because true reversal of effect interactions allow for the possibility that large GxEs could exist in the absence of ‘main’ effects of environment and/or genotype. In fact, they are the only type of interaction that can be detected if there are no genetic or environmental main effects (Rogers, 2002). Furthermore, researchers should be aware that the effect size of non-reversal interactions is mathematically constrained to be 50% or less of the combined effect size of the genetic and environmental main effects (Rogers, 2002). Notably, all published studies without main effects could only detect crossover interactions. Furthermore, those with modest main effects were severely underpowered to detect non-crossover interactions. Thus, if researchers believe, as we do, that non-reversal interactions are generally more plausible than reversal interactions, then they should search for GxEs involving genetic and environmental variables known to have robust main effects on the phenotype of interest, a suggestion that has been offered before (Risch, et al., 2009).

We expect non-reversal interactions to be more likely than reversal interactions because of the seemingly implausible (to us) scenarios that reversal interactions imply. Contrary to the expectation that a given genotype will be the ‘risk’ genotype across different levels of the environment, reversal of genetic effect implies that the risk status of a given genotype will be reversed in one environment compared to another. Similarly, one level of the environment will not be the ‘risk’ environment for all people; rather the risk environment will be reversed depending on genotype. Using the example of the ‘Depression ~ 5-HTTLPR x Stressful life events’ interaction, a reversal of environmental effect implies that high levels of stressful life
events will increase the likelihood of depression for some people but be \textit{protective} against depression for people with certain genotypes.

Such reversals of effect seem implausible for two reasons. For one, it is surprising that stressful life events could be \textit{protective} against depression (Kendler, Karkowski, & Prescott, 1999). Additionally, such interactions imply that polymorphisms have effects strong enough to reverse the effect of the environment, a prospect that seems inconsistent with the typically modest effects of individual polymorphisms (Plomin & Davis, 2009). In contrast, non-reversal interactions seem plausible because they imply that genetic variables have subtle influences on environmental effects, and vice versa. In other words, non-reversal interactions imply that the magnitude (but not the direction) of genetic effects is influenced by environment. Similarly, the magnitude (but not the direction) of environmental effects is influenced by genotype. Indeed, it seems implausible that genetic and environmental effects are completely impervious to the effects of one another. We provide these explanations merely as a springboard for discussion, and reiterate that we do not know the true form of interactions involving psychiatric phenotypes. Indeed, a systematic report of interactions in mice found many reversals and also found that interaction effects tended to be larger than main effects (Valdar, et al., 2006), which is consistent with the presence of reversal interactions. Thus it may be the case that reversals are common.

5.2.2 Alternative explanations for reversal interactions in the literature

Alternatively, the observed profile of results may be inaccurate. As shown in Figure 5B via simulations, low power biases the observed form of interactions; increasing power by increasing \(N\) changes the observed profile of reversals, even though the form of the true interaction does not change. For the simulation results presented in Figure 5B we conducted 1,000 runs for each of three scenarios that differed only in terms of sample size (\(N = 345, 2,000,\)
and 10,000). For each sample size, the profile of reversals is derived from significant interactions only, mimicking the effect of publication bias. For these simulations, we identified a combination of genetic, environmental, and GxE effect sizes (accounting for 0.5%, 1.8%, and 0.1% of phenotypic variance each) that produced a profile of reversals similar to the observed profile of reversals in Figure 5A when power was low (i.e., when N=345). However, the true form of an interaction is dictated by the relative magnitude of genetic, environmental, and GxE effects, and this particular combination of effect sizes yields a non-crossover interaction (as depicted in Figure 5B). Therefore, detection of any other form of interaction is inaccurate. As shown in the first simulated bar, the observed form of the interaction is inaccurate most of the time. In the right two columns, we demonstrate that identification of the true form of the interaction improves with increasing power. Thus, low power will bias the form of observed interactions.

Another possibility for the observed profile of reversals is that findings are actually type I errors. We created a profile of reversals identical to the observed profile (data not shown) in a variety of simulations involving no real GxE effect. Thus, another possibility is that the observed profile of reversals is entirely misleading because positive results are actually type I errors.

5.2.3 Synthesis of information regarding the form of published interactions

The observed profile of reversals may reflect the true form of GxEs in psychiatry. However, through simulations we demonstrated two other possibilities: 1) the observed profile of interactions may be inaccurate because low power biases the observed form of interactions; 2) the observed profile of reversals could be completely inaccurate if it is due to the publication of type I errors (data not shown). Thus, the observed form of an interaction should be interpreted
cautiously, and highly powered studies of robust interactions will be needed to determine the true form of interactions. Finally, we apply Rogers’ (2002) findings about mathematical constraints on the effect size of non-reversal interactions to GxE studies, noting that non-reversal interactions cannot be found in the absence of main effects. Additionally, samples of thousands rather than hundreds of subjects will be necessary for adequate (> 80%) power to identify non-reversal interactions when main effects are modest (i.e. combined genetic and environmental main effects account for 10% or less of phenotypic variance).
Chapter 6

Conclusion

We examined the cGxE literature in psychiatry and found multiple lines of evidence consistent with the possibility that many positive reports of cGxEs are type I errors. In addition, we showed that the observed form of an interaction (assuming that it is real) could often be incorrect because low power biases the observed form of interactions. This finding could be important for cGxE research in psychiatry because many published studies were likely underpowered. Additionally, we noted patterns of expanding and branching hypotheses in the literature, and suggest that such patterns will serve as signatures of type I errors if robust interactions are not eventually identified.

6.1 Reasons why a preponderance of type I errors in the cGxE literature in psychiatry may not be surprising.

The above-mentioned conclusions may sound surprising at first, but we argue that recent advances in knowledge about the genetic architecture of psychiatric disorders make them less surprising than they would have been even two years ago. For one, GxEs can be conceptualized as main effects if the phenotype is ‘response to the environment’. Viewed this way, genetic effects on ‘response to the environment’ phenotypes may be comparable in size to other genetic effects in psychiatry, accounting for less than one percent of phenotypic variance each (Plomin & Davis, 2009). If so, then cGxE studies are likely to mimic association studies with initially positive reports, followed by mixed and predominately
null results, and ultimately few or no robust findings (Dahlman, et al., 2002; Wacholder, Chanock, Garcia-Closas, El ghormli, & Rothman, 2004).

GxE studies are currently in a ‘candidate GxE’ era, similar to the candidate gene association era in which individual association hypotheses were tested. This is in contrast to the GWAS era, in which hundreds of thousands of association hypotheses are tested simultaneously. Notably, nearly all candidate gene association studies failed to replicate (Colhoun, et al., 2003).

Three factors likely contributed to the lack of success in candidate gene association studies: 1) low power given small effect sizes; 2) inchoate knowledge of the effects of polymorphisms, particularly with respect to disease etiology, which typically precluded the possibility of accurate a priori hypotheses; and 3) the low prior probability that one given polymorphism, out of the millions of possible polymorphisms, had an appreciable effect on a phenotype of interest. Notably, all of these factors are likely to be even more problematic in cGxE studies. First, the effect sizes of GxEs may be comparable to those of genetic main effects (this is an unbiased guess; GxE effects may actually be smaller or larger, on average, than genetic main effects). However, power to detect interactions is lower than power to detect main effects (of equal effect sizes) (McClelland & Judd, 1993). Therefore, the power problem that plagued association studies may be even worse for candidate GxE studies. Second, in general, even less is known about the interaction of specific polymorphisms with specific environmental variables than is known about the main effects of polymorphisms. Thus, candidate GxE hypotheses will tend to be less informed than candidate gene association hypotheses, rendering a priori cGxE hypotheses less accurate than a priori association hypotheses, which were usually incorrect (Colhoun, et al., 2003). Third, there are many more potential cGxE hypotheses than candidate gene association hypotheses. Therefore it is likely that the prior probability that any
one candidate GxE has an appreciable effect on a phenotype of interest is even lower than the prior probability that a particular candidate gene has an effect on a phenotype of interest. Taken together, these factors suggest that identifying robust GxEs may be even more difficult than identifying robust genetic main effects and that most or all positive findings may be type I errors. However, this assertion is based on the above-mentioned arguments and empirical evidence will be needed to definitively determine the promise of cGxE research in psychiatry.

6.2 Limitations

There are several limitations in our study that should be considered when interpreting our findings. First, since we collapsed across interactions for our analyses our results cannot be applied to any particular interaction. In other words, what is generally true of cGxE studies in psychiatry may not be true for individual interactions.

Second, our analyses depended on assumptions about the lack of systematic differences between particular groupings of interactions. An important assumption was that replication attempts did not differ from novel reports in systematic ways. If replication attempts were disproportionately devoted to interactions that were, on average, less replicable (e.g., due to type I errors, variables that were difficult to measure, and/or small effect sizes), then our finding of putative publication bias among novel reports could be exaggerated or incorrect. In addition, our estimate of observed power to detect GxEs (32%, the percentage of positive replication attempts) would be an underestimate of the true power to detect the average reported GxE. We checked to see if replication samples were smaller than novel samples, which could render replication attempts less powerful and reliable. However, replication samples, on average, were larger than
novel samples (615 versus 536, non-significant difference), so we found no support for this particular explanation.

Systematic differences between replication attempts and novel studies would not affect conclusions about putative publication bias among replication attempts because the relevant analyses were restricted to replication attempts. Instead, as noted previously, systematic differences between large and small studies could provide explanations other than publication bias for the observation that significant replication attempts tended to have smaller, rather than larger, sample sizes than null replication attempts. For example, if smaller studies used more accurate measures, then they might be more powerful by virtue of lower measurement error.

Third, our ascertainment strategy could have biased our results. In general, we probably missed some published cGxE studies. We know of no reason to expect that the studies included in this review constitute a biased sample of published cGxE studies, though this possibility cannot be ruled out. Additionally, for our first analysis, we searched for GxEs in the abstracts, but not the body of papers. We did this for two reasons: first, the latter was not feasible because it would have required us to look through every paper ever published that examined a genetic variant – which is obviously too time consuming. Second, we confined our search to the abstracts in an effort to avoid certain biases. For example, we might have worked harder to find GxEs in the body of some papers rather than others, and this could be a potential source of bias. Nevertheless, it is likely that more novel cGxE hypotheses have been reported than we identified. To the extent that this is true, our finding of publication bias among novel GxE studies could be exaggerated or incorrect.

Finally, as with any review or meta-analysis, some subjective decisions were required. For example, studies needed to be classified as novel studies OR replication attempts. Such
decisions can be difficult because, in reality, studies exist on a continuum of similarity to dissimilarity, and any choice is inherently subjective. In all analyses we attempted to be as unbiased as possible. Furthermore, for the analysis that involved the greatest number of subjective decisions (i.e., the analysis of sample sizes in significant versus non-significant replication attempts) we performed a check of potential bias. First, we coded studies, trying to be as unbiased as possible. Then, we coded studies two other ways: 1) biased toward detecting publication bias, and 2) biased against detecting publication bias. Upon re-running the analyses, we found that results were in the same direction for all three methods of coding the data, thought the result did not reach statistical significance when data coding was biased against the possibility of detecting publication bias. Thus, we felt confident that these biases were not the source of our findings.

6.3 Recommendations for cGxE studies

Given our findings we offer three recommendations for future cGxE studies that may improve the reliability of reported findings and aid in discerning true from false GxE effects. First, researchers should consider power when designing studies. Second, replication attempts, even those with small samples and null results, are valuable contributions to the literature and should be afforded such status in journals. In fact, independent replication attempts are arguably more important than novel reports of cGxEs. Last, all statistical tests that were performed should be reported in order to decrease the multiple testing problem and the associated risk of type I errors. Failing to do so clutters the literature with confusing results, slows the progress of science, and leads to wasted or poorly appropriated resources. Additional recommendations are provided in our companion paper (Duncan, et al., 20--).
6.4 Future directions

As has occurred with investigations of main effects of genes, we expect that genome-wide methodologies will soon supplant the candidate gene approach in GxE research (Khoury & Wacholder, 2009; Murcray, Lewinger, & Gauderman, 2009; Sonuga-Barke, et al., 2008). In psychiatry, we know of only one study that included an environmental measure in a GWAS design (excluding pharmacogenetic and other experimental studies, as stated in the methods section) (Sonuga-Barke, et al., 2008). This ‘GWASxE’ study examined the interaction of approximately half a million SNPs with maternal expressed emotion on the phenotype of ADHD (Sonuga-Barke, et al., 2008). Although the authors found no genome-wide significant results, they provided a useful demonstration of the methodology and identified cGxEs for future study (Sonuga-Barke, et al., 2008). Importantly, GWASxE studies could be used to systematically look for GxEs with very large effects, which could be clinically useful. However, it is probably more likely that GWASxE studies will suffer from the same problems of low power and inadequate sample sizes that have plagued GWAS studies. Nevertheless, the use of larger samples will lead to improved power, and GWASxE methodology will undoubtedly uncover some reliable GxEs at some point in the future.

6.5 Final synopsis

Taken together, the results of the present study suggest that early failures to replicate cGxE interactions such as the ‘Depression ~ 5-HTTLPR x Stressful life events’ interaction (Caspi, et al., 2003)(Munafo, et al., 2009; Risch, et al., 2009) may portend more failures of replication in the future. We report our conclusions with some apprehension given concern that
true GxEs may be discounted because of these findings. However, the risk of wasting resources on unpromising hypotheses is also a realistic possibility. In sum, our results provide a rationale for interpreting cGxE findings more cautiously and they are consistent with the criticisms of GxE research offered by others e.g. (Eaves, 2006; Risch, et al., 2009).


Duncan, L. E., Willcutt, E. G., & Keller, M. C. (20--). Empirical review of the first decade of mGxE research in psychiatry.


<table>
<thead>
<tr>
<th><strong>FIRST AUTHOR</strong></th>
<th><strong>TITLE</strong></th>
<th><strong>YEAR</strong></th>
<th><strong>N sample</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aguilera</td>
<td>Early adversity and 5-HTT/BDNF genes: new evidence of gene–environment interactions on depressive symptoms in a general population</td>
<td>2009</td>
<td>534</td>
</tr>
<tr>
<td>Altink</td>
<td>The dopamine receptor D4 7-repeat allele and prenatal smoking in ADHD-affected children and their unaffected siblings: no gene–environment interaction</td>
<td>2008</td>
<td>946</td>
</tr>
<tr>
<td>Amstadter</td>
<td>Variant in RGS2 moderates posttraumatic stress symptoms following potentially traumatic event exposure</td>
<td>2009</td>
<td>607</td>
</tr>
<tr>
<td>Bakermans-Kranenburg</td>
<td>Gene-environment interaction of the dopamine D4 receptor (DRD4) and observed maternal insensitivity predicting externalizing behavior in preschoolers</td>
<td>2006</td>
<td>47</td>
</tr>
<tr>
<td>Bau</td>
<td>The TaqI A1 allele of the dopamine D2 receptor gene and alcoholism in Brazil: Association and interaction with stress and harm avoidance on severity prediction</td>
<td>2000</td>
<td>229</td>
</tr>
<tr>
<td>Becker</td>
<td>Interaction of dopamine transporter genotype with prenatal smoke exposure on ADHD symptoms</td>
<td>2008</td>
<td>305</td>
</tr>
<tr>
<td>Binder</td>
<td>Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults</td>
<td>2008</td>
<td>676</td>
</tr>
<tr>
<td>Blomeyer</td>
<td>Interaction between CRHR1 gene and stressful life events predicts adolescent heavy alcohol use</td>
<td>2008</td>
<td>280</td>
</tr>
<tr>
<td>Bradley</td>
<td>Influence of Child Abuse on Adult Depression Moderation by the Corticotropin-Releasing Hormone Receptor Gene</td>
<td>2009</td>
<td>422 &amp; 204</td>
</tr>
<tr>
<td>Brookes</td>
<td>Differential dopamine receptor D4 allele association with ADHD dependent of (sic) proband season of birth</td>
<td>2008</td>
<td>1110</td>
</tr>
<tr>
<td>Brummett</td>
<td>Effects of environmental stress and gender on associations among symptoms of depression and the serotonin transporter gene linked polymorphic region (5-HTTLPR)</td>
<td>2008</td>
<td>288 &amp; 142</td>
</tr>
<tr>
<td>Caspi-1</td>
<td>Role of genotype in the cycle of violence in maltreated children</td>
<td>2002</td>
<td>442</td>
</tr>
<tr>
<td>Caspi-3</td>
<td>Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene</td>
<td>2003</td>
<td>847</td>
</tr>
<tr>
<td>Cervilla</td>
<td>The risk for depression conferred by stressful life events is modified by variation at the serotonin transporter 5HTTLPR genotype: evidence from the Spanish PREDICT-Gene cohort</td>
<td>2007</td>
<td>737</td>
</tr>
<tr>
<td>Author</td>
<td>Title</td>
<td>Year 1</td>
<td>Year 2</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------</td>
<td>--------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Chipman</td>
<td>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</td>
<td>2007</td>
<td>2095 &amp; 584</td>
</tr>
<tr>
<td>Chorbov</td>
<td>Relationship of 5-HTTLPR genotypes and depression risk in the presence of trauma in a female twin sample</td>
<td>2007</td>
<td>247</td>
</tr>
<tr>
<td>Chotai</td>
<td>Gene-environment interaction in psychiatric disorders as indicated by season of birth variations in tryptophan hydroxylase (TPH), serotonin transporter (5-HTTLPR) and dopamine receptor (DRD4) gene polymorphisms</td>
<td>2003</td>
<td>1349</td>
</tr>
<tr>
<td>Cicchetti</td>
<td>Interactions of child maltreatment and serotonin transporter and monoamine oxidase A polymorphisms: Depressive symptomatology among adolescents from low socioeconomic status backgrounds</td>
<td>2007</td>
<td>339</td>
</tr>
<tr>
<td>Covault</td>
<td>Interactive effects of the serotonin transporter 5-HTTLPR polymorphism and stressful life events on college student drinking and drug use</td>
<td>2007</td>
<td>302</td>
</tr>
<tr>
<td>Dick</td>
<td>Marital status, alcohol dependence, and GABRA2: Evidence for gene-environment correlation and interaction</td>
<td>2006</td>
<td>1916 &amp; 915</td>
</tr>
<tr>
<td>Ducci</td>
<td>Interaction between a functional MAOA locus and childhood sexual abuse predicts alcoholism and antisocial personality disorder in adult women</td>
<td>2008</td>
<td>187</td>
</tr>
<tr>
<td>Eley</td>
<td>Gene-environment interaction analysis of serotonin system markers with adolescent depression</td>
<td>2006</td>
<td>377</td>
</tr>
<tr>
<td>Foley</td>
<td>Childhood adversity, monoamine oxidase A genotype, and risk for conduct disorder</td>
<td>2004</td>
<td>514</td>
</tr>
<tr>
<td>Fox</td>
<td>Evidence for a gene-environment interaction in predicting behavioral inhibition in middle childhood</td>
<td>2008</td>
<td>73</td>
</tr>
<tr>
<td>Frazzetto</td>
<td>Early trauma and increased risk for physical aggression during adulthood: The moderating role of MAOA genotype</td>
<td>2007</td>
<td>235</td>
</tr>
<tr>
<td>Gacek</td>
<td>Tryptophan hydroxylase 2 gene and alcohol use among college students</td>
<td>2008</td>
<td>351</td>
</tr>
<tr>
<td>Gervai</td>
<td>Infant genotype may moderate sensitivity to maternal affective communications: attachment disorganization, quality of care, and the DRD4 polymorphism</td>
<td>2007</td>
<td>96</td>
</tr>
<tr>
<td>Gibb</td>
<td>Serotonin transporter (5-HTTLPR) genotype, childhood abuse, and suicide attempts in adult psychiatric inpatients</td>
<td>2006</td>
<td>30</td>
</tr>
<tr>
<td>Gillespie</td>
<td>The relationship between stressful life events, the serotonin transporter (5-HTTLPR) genotype and major depression</td>
<td>2004</td>
<td>1091</td>
</tr>
<tr>
<td>Grabe-1</td>
<td>Mental and physical distress is modulated by a polymorphism in the 5-HT transporter gene interacting with social stressors and chronic disease burden</td>
<td>2005</td>
<td>976</td>
</tr>
<tr>
<td>Grabe-2</td>
<td>Serotonin transporter gene (SLC6A4) promoter</td>
<td>2009</td>
<td>3045</td>
</tr>
<tr>
<td>Author</td>
<td>Title</td>
<td>Year</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Gunther</td>
<td>Serotonin transporter gene polymorphism (5-HTTLPR) and anxiety reactivity in daily life: a daily process approach to gene-environment interaction</td>
<td>2007</td>
<td>350</td>
</tr>
<tr>
<td>Haberstick</td>
<td>Monoamine oxidase A (MAOA) genotype and antisocial behaviors in the presence of childhood and adolescent maltreatment</td>
<td>2005</td>
<td>774</td>
</tr>
<tr>
<td>Haeffel</td>
<td>Association between polymorphisms in the dopamine transporter gene and depression</td>
<td>2008</td>
<td>176</td>
</tr>
<tr>
<td>Henquet</td>
<td>COMT Val158Met moderation of cannabis-induced psychosis: a momentary assessment study of 'switching on' hallucinations in the flow of daily life</td>
<td>2009</td>
<td>61</td>
</tr>
<tr>
<td>Huizinga</td>
<td>Childhood maltreatment, subsequent antisocial behavior, and the role of monoamine oxidase A genotype</td>
<td>2006</td>
<td>277</td>
</tr>
<tr>
<td>Jacobs</td>
<td>Stress-related negative affectivity and genetically altered serotonin transporter function: evidence of synergism in shaping risk of depression</td>
<td>2006</td>
<td>374</td>
</tr>
<tr>
<td>Jokela-1</td>
<td>Serotonin receptor 2A gene and the influence of childhood maternal nurturance on adulthood depressive symptoms</td>
<td>2007</td>
<td>1212</td>
</tr>
<tr>
<td>Jokela-2</td>
<td>The influence of urban/rural residency on depressive symptoms is moderated by the serotonin receptor 2A gene</td>
<td>2007</td>
<td>1224</td>
</tr>
<tr>
<td>Jokela-3</td>
<td>The serotonin receptor 2A gene moderates the influence of parental socioeconomic status on adulthood harm avoidance</td>
<td>2007</td>
<td>1246</td>
</tr>
<tr>
<td>Jokela-4</td>
<td>Tryptophan hydroxylase 1 gene (TPH1) moderates the influence of social support on depressive symptoms in adults</td>
<td>2007</td>
<td>341</td>
</tr>
<tr>
<td>Kahn</td>
<td>Role of dopamine transporter genotype and maternal prenatal smoking in childhood hyperactive-impulsive, inattentive, and oppositional behaviors</td>
<td>2003</td>
<td>161</td>
</tr>
<tr>
<td>Kaufman-1</td>
<td>Social supports and serotonin transporter gene moderate depression in maltreated children</td>
<td>2004</td>
<td>101</td>
</tr>
<tr>
<td>Kaufman-2</td>
<td>Brain-derived neurotrophic factor-5-HTTLPR gene interactions and environmental modifiers of depression in children.</td>
<td>2007</td>
<td>196</td>
</tr>
<tr>
<td>Kaufman-3</td>
<td>Genetic and environmental predictors of early alcohol use</td>
<td>2007</td>
<td>127</td>
</tr>
<tr>
<td>Keltikangas-Jarvinen</td>
<td>Nature and nurture in novelty seeking</td>
<td>2004</td>
<td>92</td>
</tr>
<tr>
<td>Kendler</td>
<td>The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression. A replication</td>
<td>2005</td>
<td>549</td>
</tr>
<tr>
<td>Kilpatrick</td>
<td>The serotonin transporter genotype and social support and moderation of posttraumatic stress disorder and depression in hurricane-exposed adults</td>
<td>2007</td>
<td>589</td>
</tr>
<tr>
<td>Kim</td>
<td>Interactions between life stressors and susceptibility genes</td>
<td>2007</td>
<td>732</td>
</tr>
<tr>
<td>Author</td>
<td>Title</td>
<td>Summary</td>
<td>Year</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Kim-Cohen</td>
<td>MAOA, maltreatment, and gene-environment interaction predicting children’s mental health: new evidence and a meta-analysis.</td>
<td>2006</td>
<td>975</td>
</tr>
<tr>
<td>Koenen</td>
<td>Modification of the association between serotonin transporter genotype and risk of posttraumatic stress disorder in adults by county-level social environment</td>
<td>2009</td>
<td>651</td>
</tr>
<tr>
<td>Lahti</td>
<td>Socio-demographic characteristics moderate the association between DRD4 and Novelty seeking</td>
<td>2006</td>
<td>154</td>
</tr>
<tr>
<td>Langley</td>
<td>Testing for gene x environment interaction effects in attention deficit hyperactivity disorder and associated antisocial behavior</td>
<td>2008</td>
<td>266</td>
</tr>
<tr>
<td>Lasky-Su</td>
<td>A study of how socioeconomic status moderates the relationship between SNPs encompassing BDNF and ADHD symptom counts in ADHD Families</td>
<td>2007</td>
<td>345</td>
</tr>
<tr>
<td>Laucht-1</td>
<td>Interacting effects of the dopamine transporter gene and psychosocial adversity on attention-deficit/hyperactivity disorder symptoms among 15-year-olds from a high-risk community sample</td>
<td>2007</td>
<td>305</td>
</tr>
<tr>
<td>Laucht-3</td>
<td>Impact of psychosocial adversity on alcohol intake in young adults: Moderation by the LL genotype of the serotonin transporter polymorphism</td>
<td>2009</td>
<td>309</td>
</tr>
<tr>
<td>Madrid</td>
<td>Stress as a mediating factor in the association between the DRD2 TaqI polymorphism and alcoholism</td>
<td>2001</td>
<td>304</td>
</tr>
<tr>
<td>Neuman</td>
<td>Prenatal smoking exposure and dopaminergic genotypes interact to cause a severe ADHD subtype</td>
<td>2007</td>
<td>770</td>
</tr>
<tr>
<td>Nilsson</td>
<td>Role of monoamine oxidase A genotype and psychosocial factors in male adolescent criminal activity</td>
<td>2007</td>
<td>81</td>
</tr>
<tr>
<td>Nobile-1</td>
<td>Socioeconomic status mediates the genetic contribution of the dopamine receptor D4 and serotonin transporter linked promoter region repeat polymorphisms to externalization in preadolescence</td>
<td>2007</td>
<td>607</td>
</tr>
<tr>
<td>Nobile-1</td>
<td>The influence of family structure the TPH2 G703T and the 5HTTLPR serotonergic genes upon affective problems in children aged 10 14 years</td>
<td>2009</td>
<td>607</td>
</tr>
<tr>
<td>Ozkaragoz</td>
<td>Extraversion. Interaction between D2 dopamine receptor polymorphisms and parental alcoholism.</td>
<td>2000</td>
<td>98</td>
</tr>
<tr>
<td>Perroud</td>
<td>Interaction between BDNF Val66Met and childhood trauma on adult’s violent suicide attempt</td>
<td>2008</td>
<td>813</td>
</tr>
<tr>
<td>Polanczyk</td>
<td>Protective Effect of CRHR1 Gene Variants on the Development of Adult Depression Following Childhood</td>
<td>2009</td>
<td>1037 &amp; 1116</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Title and Highlights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>5-HTTLPR genotype, stressful life events and late-life depression: No evidence of interaction in a French population (2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prom-Wormley</td>
<td>Monoamine oxidase A and childhood adversity as risk factors for conduct disorder in females (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propper</td>
<td>Parenting quality, DRD4, and the prediction of externalizing and internalizing behaviors in early childhood (2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Racine</td>
<td>The possible influence of impulsivity and dietary restraint on associations between serotonin genes and binge eating (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reif</td>
<td>Nature and nurture predispose to violent behavior: serotonergic genes and adverse childhood environment (2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheid</td>
<td>Depressive symptoms in mid-pregnancy, lifetime stressors and the 5-HTTLPR genotype (2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sjoberg-2</td>
<td>Adolescent girls and criminal activity: Role of MAOA-LPR genotype and psychosocial factors (2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sonuga-Barke</td>
<td>Dopamine and serotonin transporter genotypes moderate sensitivity to maternal expressed emotion: the case of conduct and emotional problems in attention deficit/hyperactivity disorder (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stevens</td>
<td>Dopamine transporter gene polymorphism moderates the effects of severe deprivation on ADHD symptoms: developmental continuities in gene–environment interplay (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td>The combined effect of norepinephrine transporter gene and negative life events in major depression of Chinese Han population (2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surtees</td>
<td>Social adversity, the serotonin transporter (5-HTTLPR) polymorphism and major depressive disorder (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor</td>
<td>Early family environment, current adversity, the serotonin transporter promoter polymorphism, and depressive (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Title</td>
<td>Year</td>
<td>Page</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Todd</td>
<td>Gene–environment interactions in the development of combined type ADHD: Evidence for a synapse-based model</td>
<td>2007</td>
<td>770</td>
</tr>
<tr>
<td>van Winkel</td>
<td>Evidence that the COMT Val158Met polymorphism moderates sensitivity to stress in psychosis</td>
<td>2008</td>
<td>56</td>
</tr>
<tr>
<td>Waldman</td>
<td>Gene-environment interactions reexamined: Does mother's marital stability interact with the dopamine receptor D2 gene in the etiology of childhood attention-deficit/hyperactivity disorder?</td>
<td>2007</td>
<td>211</td>
</tr>
<tr>
<td>Wichers</td>
<td>The BDNF Val66Met × 5-HTTLPR × child adversity interaction and depressive symptoms: An attempt at replication</td>
<td>2008</td>
<td>621</td>
</tr>
<tr>
<td>Widom</td>
<td>MAOA and the “Cycle of Violence:” Childhood abuse and neglect, MAOA genotype, and risk for violent and antisocial behavior</td>
<td>2006</td>
<td>409</td>
</tr>
<tr>
<td>Wilhelm</td>
<td>Life events, first depression onset and the serotonin transporter gene</td>
<td>2006</td>
<td>127</td>
</tr>
<tr>
<td>Xu</td>
<td>The norepinephrine transporter gene modulates the relationship between urban/rural residency and major depressive disorder in a Chinese population</td>
<td>2009</td>
<td>835</td>
</tr>
<tr>
<td>Yen</td>
<td>A multilevel analysis of the influence of Apolipoprotein E genotypes on depressive symptoms in late-life moderated by the environment</td>
<td>2008</td>
<td>301</td>
</tr>
<tr>
<td>Young</td>
<td>Interaction between MAO-A genotype and maltreatment in the risk for conduct disorder: Failure to confirm in adolescent patients.</td>
<td>2006</td>
<td>247</td>
</tr>
<tr>
<td>Zalsman</td>
<td>Association of a triallelic serotonin transporter gene promoter region (5-HTTLPR) polymorphism with stressful life events and severity of depression</td>
<td>2006</td>
<td>316</td>
</tr>
<tr>
<td>Zhang</td>
<td>The combined effects of the 5-HTTLPR and 5-HTR1A genes modulates the relationship between negative life events and major depressive disorder in a Chinese population</td>
<td>2009</td>
<td>792</td>
</tr>
</tbody>
</table>