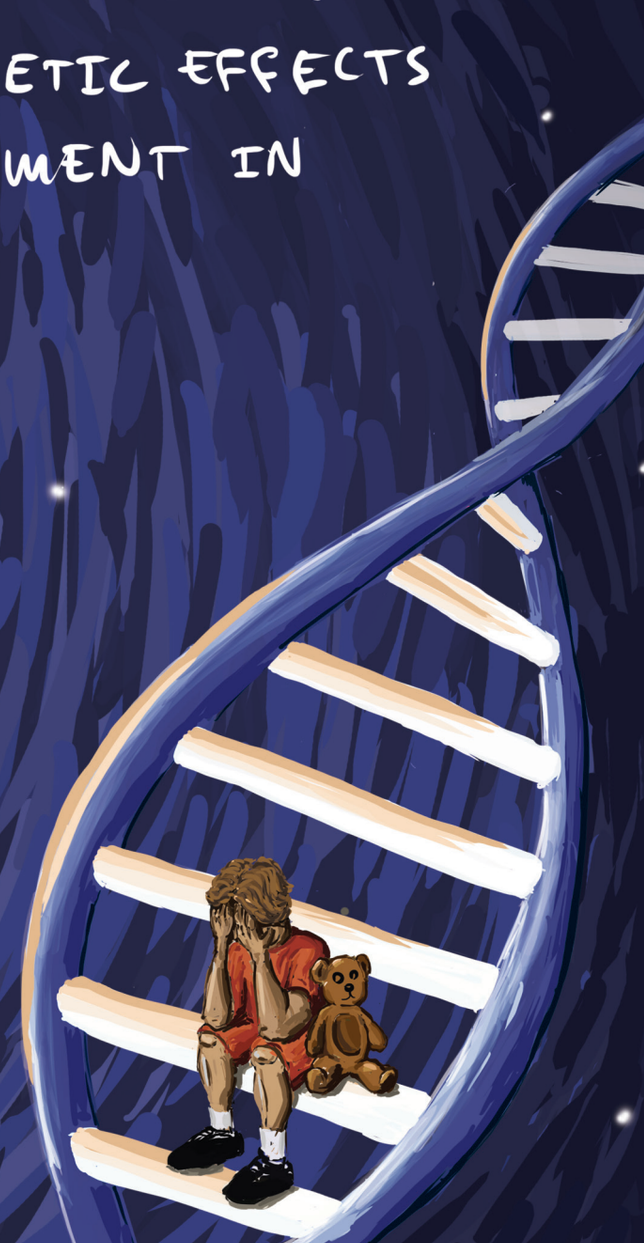
WOUTER J. PEYROT – THE COMPLEX LINK BETWEEN GENETIC EFFECTS AND ENVIRONMENT IN DEPRESSION

THE COMPLEX LINK BETWEEN GENETIC EFFECTS AND ENVIRONMENT IN DEPRESSION

WOUTER JOHANNES PEYROT



VU UNIVERSITY AMSTERDAM
DEPARTMENT OF PSYCHIATRY

VRIJE UNIVERSITEIT

The complex link between genetic effects and environment in depression

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op vrijdag 20 januari 2017 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Wouter Johannes Peyrot

geboren te Amsterdam

promotoren:	prof.dr. B.W.J.H. Penninx
	prof.dr. D.I. Boomsma
copromotor:	prof.dr. N.R. Wray

TABLE OF CONTENT

Chapter 1
General introduction

Major Depressive Disorder: prevalent, heterogeneous, and disabling

The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV and DSM-5) classifies major depressive disorder (MDD) in individuals with depressive symptoms for at least two weeks nearly every day, which consist of at least one of the two core symptoms of depressed mood or decreased interest (anhedonia), added to a total of five symptoms when also considering the seven secondary symptoms of weight or appetite change, change in sleep, psychomotoric change, fatigue or loss of energy, feelings of guilt or worthlessness, decreased ability to concentrate, and thoughts of dead or suicide. MDD is considered a common disorder as it affects around 15% of people sometime in life,[1,2] a number far exceeding the approximate 1% of people diagnosed with psychiatric disorders such as schizophrenia (SCZ), bipolar disorder (BIP) or autism (ASD).[1] The first onset of MDD can occur at all stages in life, but most often MDD presents between the age of 20 and 40 to have a chronic course of more than 24 months in approximately 20% of patients.[3,4] Notably, MDD affects women twice as often as men.[5,6] As a consequence of the wide diagnostic criteria requiring only five of a total of nine symptoms, the clinical presentation of MDD varies considerable from one patient to the other. The DSM IV and DSM 5 acknowledges this heterogeneity by defining subtypes of MDD, such as: *melancholic depression* characterized by i.a. loss of pleasure in all activities (severe anhedonia), lack of response to positive stimuli, excessive weight loss, early-morning waking and a clear day-pattern of symptoms; *atypical depression* characterized by i.a. weight gain and increased sleep; *catatonic depression* characterized by i.a. almost full inability to speak or move; and a *depression with psychotic features* with delusions concerning guilt, punishment, disease or financial debt possibly accompanied by auditory hallucinations from a devaluating nature. Some milder forms of MDD are self-limiting and require no other interventions than psycho-education about symptoms and lifestyle advice, whereas more severe forms of MDD requires therapies ranging from cognitive behavioral therapy or antidepressant therapy for moderate episodes, tricyclic antidepressants with lithium for severe episodes, up to electroconvulsive therapy for episodes with severe motoric and psychotic features as well as for therapy resistant severe episodes.[7] Despite the range of therapeutic strategies available, not all MDD can be treated, leading to live long symptoms in some. Indeed, the suicide rate amongst MDD patients in the USA has been estimated at approximately 3.4% more often in male (7%) than in female (1%) patients.[8] People suffering from MDD are often unable to participate in working and social life, and the World

Health Organization has predicted that by 2030 MDD will be globally leading in disease burden.[9]

The etiology of Major Depressive Disorder remains largely unknown

The etiology and pathophysiology of MDD is largely unknown, in particular when compared to other medical conditions such as, for example, diabetes mellitus, which pathophysiology has been pinpointed to failure of the islets of Langerhans in the pancreas to produce insulin (type I) or by peripheral insulin resistance (type II). Nevertheless, despite the largely unknown etiology of MDD, many associates have been identified leading to hypotheses about MDD's pathophysiology. For example, MDD is known to be associated to lower educational attainment,[10] stressful life-events, childhood trauma,[11,12] and personality characteristics, but also to medical conditions such as diabetes mellitus[13] and cardiovascular disease,[14] and neurobiological measurements such as hypothalamic-pituitary-adrenal axis indicators,[15] hippocampal volume loss,[16] and inflammation.[17,18] Childhood trauma, often defined as trauma before the age of 16, is one of the most notable risk factors, with an OR for MDD between 2 and 3, that also increases risk amongst MDD patients to suffer from psychotic features, to attempt suicide and to achieve poorer treatment outcome.[12] In a step towards understanding MDD's etiology, Kendler et al. have suggested a developmental model where three broad pathways interact; internalizing factors (genetic risk factors, neuroticism, low self-esteem, early-onset anxiety, and past history of major depression), externalizing factors (genetic risk factors, conduct disorder, and substance misuse) and adversities.[19,20] From a purely biological perspective, MDD has been hypothesized to arise from synaptic deficiency of monoamines (serotonin, dopamine, noradrenalin) given the effectiveness of synaptic monoamine increasing medication such as selective serotonin reuptake inhibitors, but this hypothesis is now considered too simplistic.[21,22] It is clear that MDD is associated to many social, psychological and biological factors, but its etiology still remains largely unknown, as do the reasons for its association with these factors. Some associations may be causal, others may be a consequence of MDD, and yet others may be due to a shared etiology, which may include shared genetic risk.

Relevance of genetic research

Genetic research might reveal itself as powerful catalyzer of research in MDD's etiology in the years to come. Firstly, the nature of the association between MDD risk and genetic variants is unique, because the direction of causality is certain:

genetic risk variants impact MDD and not visa versa. Notably, this property is already lost in the next physiological level, because the association between MDD and the expression of a gene may be attributable to gene-expression impacting MDD but also to MDD impacting gene-expression via e.g. increased stress. Secondly, genome-wide association studies (GWAS) provide the opportunity for hypothesis-free testing of all possible pathophysiological pathways potentially inspiring novel therapies. Thirdly, genetic research can help to understand why individuals differ in their vulnerability for MDD by, for example, assessing the proportion of variation in MDD risk in the population attributable to genetic effects expressed as the so called heritability (h^2). Fourth, theory about the distribution of genetic variants and mating patterns can help to understand why psychiatric disorders still exist in the population despite their unfavorable effects on reproductive fitness, which may be of minor interest for MDD (impacted by very little natural selection) but can be much more relevant for disorders such as schizophrenia (SCZ) and autism (ASD) associated with a clear reduced fecundity.[23] Fifth, genetic variants that influence multiple traits can potentially help to understand, at least in part, some of the many comorbidities associated with MDD. Importantly, the points of relevance described here have not yet been fulfilled as psychiatric genetic research has met major challenges, but suggest nevertheless the exciting potential of psychiatric genetic research in general.

Genetic research in MDD

Genetic research in MDD had been quickly evolving in the years up to 2011, when this PhD project commenced. Until around 2009, genetic research on Major Depressive Disorder (MDD) had mainly focused on twin and family studies, linkage studies, and candidate gene studies, but has since increasingly concentrated on genome-wide association studies (GWAS) with case-control data. Population based twin and family studies pointed to an heritability for MDD in the general population of around $h^2 \approx 0.35$ by considering MDD risk in relation to the expected genetic similarity between family members without requiring information on genotypes.[24–26] Notably, increased heritability estimates have been reported for hospitalized depression (0.48-0.75)[27] and for lifetime diagnosis based on repeated assessments in women (0.66).[28] Linkage studies provided the first attempts to find specific genetic regions associated to MDD in large pedigrees or sib-pair studies by testing for linkage between genetic loci and disease status. However, linkage studies can only detect genetic regions with large effect on MDD risk, and did not lead to consistent findings.[29] In candidate gene studies, a single or couple of genetic variants were being tested based on a

priori hypotheses about gene function by comparing individuals with MDD to healthy unrelated controls, and although these studies did point to some potentially associated genes[30] these results showed little consistency.[31] Around 2005, new hope arose as -following sequencing of the first human genome costing 3 billion US dollar and published in 2001-[32] techniques for genotyping had progressed to provide the opportunity to genotype at relatively low cost over 500,000 genome-wide single nucleotide polymorphisms (SNP), which resulted in published MDD GWAS results from 2009 onwards.[33,31,34–37] Nevertheless, the largest GWAS up to 2011, comprising 5763 MDD cases and 6901 healthy controls, found no genome-wide significantly associated SNP, which was disappointing to many.[31]

Datasets used in this thesis

This thesis analyzes both empirical data and simulated data. The empirical data come from the Netherlands Study of Depression and Anxiety (NESDA),[38] Netherlands Twin Registry (NTR),[39] and Psychiatric Genomics Consortium (PGC).[40] NESDA is an ongoing longitudinal cohort study of MDD and anxiety disorders whose nearly 3,000 subjects were recruited from mental health care settings, general practitioners, and the general population in the period from 2004 to 2007. NTR has been collecting data on Dutch twin families since 1991 and comprises data on nearly 90,000 adult individuals. NESDA and NTR collaborate in their genetic research on MDD where NESDA provides most of the cases and NTR most of the controls that were all genotyped together.[29] The PGC is an ever-growing international collaboration combining genotype data from cohorts from multiple countries (the USA, Australia, Germany, Denmark, Sweden, the UK, and NESDA and NTR from the Netherlands).

Part A: Genetic effects and environment in depression

Why were no genome-wide significant loci for MDD found by 2011?

A sample of 5763 cases and 6901 controls for a GWAS on MDD was considered very large at the time, and the lack of significantly associated loci in MDD, as well as the dearth of findings for other traits, inspired development of novel methods and reconsideration of the expected genetic architecture (number of risk loci, their frequency, their effect sizes, and the way in which they act together). First, it should, however, be noted that testing loci in around 1,000,000 independent genomic regions requires control of false positive findings by setting a stringent

level of significance at $5 * 10^{-8}$ $(0.05/1{,}000{,}000)$. Naturally, large sample sizes are needed to balance this high multiple testing burden. Nevertheless, despite this stringent significance threshold, the MDD GWAS was still expected to detect 20% of risk loci with an odds ratio (OR) of 1.2 or larger, and having found none implicated effective loci for MDD would likely have much smaller effects.[31] Because the loci tested were still expected to explain (a large part of) the heritability estimated at 35% from family studies, it was concluded that MDD is most likely affected by many loci with small effect pointing to a polygenic genetic architecture. This polygenic architecture was further confirmed by methods developed just before 2011 to test for the overall effect of all loci at once. Polygenic risk scores were constructed in a target sample by counting the number of risk alleles based on GWAS results from an independent discovery sample, and significantly predicted a small proportion of variation in MDD.[41] These findings were in line with results from another method, genomic-relatedness-matrix restricted maximum likelihood (GREML), which compares concordance in disease status within pairs of individuals to their genetic relatedness based on SNP data, to find that a considerable proportion of variation in MDD was explained by genotyped SNPs, referred to as the SNP-heritability.[42] In addition, So et al developed a method to assess the SNP-heritability from z-statistics from GWAS results,[43] which showed similar estimates of a SNP-heritability of around 0.3 for MDD[42] in data from the Netherlands Study of Depression and Anxiety (NESDA)[38] and Netherlands Twin Registry (NTR).[39] Following these considerations about the genetic architecture, it was assumed in 2011 (when this PhD project commenced) that significantly associated SNPs would be found for MDD with increasing sample size. These considerations were, notably, not unique for MDD and also applicable for other traits, such as schizophrenia (SCZ) and height, where additional significant SNPs were also anticipated with increasing sample size.

Challenges unique for genetic research MDD

In addition to the challenge for most traits introduced by the high polygenicity and small SNP effect sizes, the GWAS on MDD also met more unique challenges as it should be noted that other traits had already detected significant loci by 2011, such as the seven loci found for SCZ.[44] Although the SCZ GWAS contained more cases (12,945) than the MDD GWAS (5763), the difference in results was also considered to be attributable to the different disease-characteristics of MDD compared to SCZ. First of all, family studies pointed to a lower heritability for MDD (0.35) compared to SCZ (0.8),[1] which already suggested that loci would be harder to find for MDD than for SCZ (assuming roughly the same number of

effective loci for both traits implying smaller average effect sizes for MDD).[31] However, other factors were also considered relevant, such as the high prevalence of MDD and its diagnostic heterogeneity.

To understand the impact of the high lifetime prevalence of MDD (15% compared to 1% for SCZ) on the power to detect associated loci, it is helpful to consider the liability-threshold model, which assumes that MDD and SCZ are underpinned by an unobserved disease-liability resulting from both genetic and environmental effects (typically assumed normally distributed), and that individuals are affected when they exceed a liability threshold (defined by the population prevalence). Under this model, individuals with MDD have less extreme MDD-liability values than individuals with SCZ have SCZ-liability values, because MDD has a lower disease-threshold following from its higher prevalence. A GWAS on MDD can, thus, be compared to a GWAS on height comparing individuals with average height of e.g. 170 (controls) to individuals with height of 180 (cases), whereas SCZ can be thought of as comparing individuals with height of 172 to individuals with height of 210. This illustrates why GWAS on MDD have less power than GWAS on SCZ, and it has indeed been suggested that GWAS on MDD would require four times the number of cases than GWAS on SCZ.[31,45] Second, it can be seen that the average MDD-liability in MDD controls (individuals without MDD) is lower than the SCZ-liability in SCZ controls, which suggests that screening of controls is more important for MDD than for SCZ; a difference further exaggerated by the later onset of MDD and thus larger uncertainty of disease-status in controls. The importance of screening controls in relation to the disease prevalence is addressed in **Chapter 6** of this thesis with respect to the SNP heritability.

Another factor often discussed in relation to the lack of significant GWAS findings for MDD is its heterogeneity and nosological uncertainty. First of all, MDD isn't based on etiology of disease but on clustering of psychological and physical symptoms, which is in line with the other psychiatric disorders, but not with somatic disorders such as, for example, diabetes mellitus. Moreover, compared to other psychiatric disorders, MDD has a relatively wide range of diagnostic criteria leading to large heterogeneity in symptoms: of the nine MDD criteria a minimum of only 5 are required (including at least one of two core symptoms) resulting in 227 possible combinations of symptoms to meet MDD diagnosis.[46] The number of possible combinations increases even further as some of the criteria are loosely defined to include two opposing symptoms, such as either gain or loss in both the weight-criterion and sleep-criterion, and feelings of either psychomotor agitation or retardation. One could hypothesize that different

combinations of DSM-IV criteria (leading to a MDD diagnosis) might be attributable to different pathological pathways, which could partly explain why GWAS's on MDD lack power to detect associated loci. It has, furthermore, been hypothesized that etiology of MDD could differ across different environmental conditions (irrespective of the combination of diagnostic criteria); a phenomenon referred to as gene-by-environment interaction.

Gene-by-environment interaction in MDD with candidate genes

Research on gene-by-environment interaction (GxE) tests whether genetic effects are moderated by environmental conditions resulting in a combined impact of environmental and genetic effects different from the sum (or product) of their individual effects. If GxE-effects were to exist in MDD, they would form an additional challenge for GWAS to detect genetic effects when environmental conditions are not appropriately accounted for. Studies on GxE-effects in MDD are, therefore, relevant to inform optimal GWAS design, but also to gain insight in potential different pathophysiological pathways across environmental strata, or environmental potentiating of genetic effects. In principal, GxE-studies can be conducted for all genetic variants across numerous environmental conditions, and many different genes and environmental factors have indeed been studied in the candidate-gene era. The most illustrative example in this aspect is the 2003 paper of Caspi and colleagues, in which childhood trauma was found to increase the impact on MDD of the length polymorphisms in the serotonin-transporter-linked polymorphic region (5-HTTLPR).[47] This interaction-effect was considered a scientific break-through at the time that fitted well with the hypothesized relevance of the serotonin transporter, which is the target of antidepressant medication inhibiting serotonin reuptake in synapses. However, the initial finding was followed by numerous conflicting replication efforts, and even meta-analyses lead to opposing conclusions.[48–50] It had been argued that the conflicting findings in replication efforts were attributable to differences in study design,[51] but the 2011 study of Fergusson et al followed a very similar design to the original paper to find no evidence for the interaction-effect reported by Caspi et al.[52] Taken all together, it seems unlikely that the original finding is generalizable to other cohorts. In addition, a critical review suggested that the GxE-literature from the candidate-gene era suffered from publication bias, because 96% of novel GxE-studies yielded significant results compared to only 27% of replication studies, and because smaller replication studies reported more significant results than the larger samples (with the threshold for significance set at 0.05 typical for candidate gene studies).[53] In **Chapter 2** of this thesis, a well-described large

sample containing individuals from NESDA and NTR adds to the discussion by testing for interaction between childhood trauma and other environmental conditions with 5-HTTLPR, while also considering the single nucleotide polymorphism, rs25531, that has been found to moderate the function of 5-HTTLPR.[54]

Gene-by-environment interaction in MDD with genome-wide information

Although research on main genetic effects had evolved from candidate-genes to a hypothesis-free GWAS approach, research on GxE-effects has not yet followed this progress. As discussed above, the GWAS on main genetic effects lacked power with a total of 5763 cases and 6901 controls collected from many contributing cohorts, and detailed information on environmental conditions was available for only some of the cohorts contributing to the overall GWAS sample. As a consequence, power was lacking to test single-SNP GxE-effects for all SNPs as this would require a genome-wide significance threshold of $5 * 10^{-8}$. Nevertheless, GWAS samples had also been applied for polygenic risk score (PRS) analysis to capture the effect of all genotyped SNPs at once by utilizing SNP-effect estimates from an independent discovery sample. Notably, PRS analyses require a significance threshold of only approximately 0.05 (because only 1 test in target set), given the availability of independent discovery results. Furthermore, contrary to single SNPs (or candidate genes), the PRS had a repeatedly confirmed effect on MDD making it a more feasible instrument to test for GxE than single loci (assuming it is unlikely that single loci have completely opposing effects across different environmental conditions).[55] The interpretation of GxE results with PRS (PRSxE) is more complex than the interpretation of GxE for candidate genes, but the relevance of PRSxE is still found in its potential to point to environmental conditions with increased genetic effects informing optimal GWAS design, and in its possible contribution to obtain insight in MDD's heterogeneity. In **Chapter 4** of this thesis, the impact of PRS on MDD is compared between individuals exposed and individuals not exposed to childhood trauma in NESDA. This single-cohort finding is, subsequently, tested in **Chapter 5** with data from the international collaboration of the Psychiatric Genomics Consortium (PGC),[40] which allows combining data of several cohorts to optimize sample size.

MDD and educational attainment

In addition to aiming to detect causal genetic risk variants, genetic research can also contribute by testing whether shared genetic effects can help to explain comorbidities or phenotypic association. An association of particular interest is

the increased risk for MDD in individuals with lower educational attainment (EA), which has been confirmed in various western countries with a three percent decrease in MDD risk per additional year of education estimated in a meta-analysis of 37 studies.[10] This association might reflect an impact of lower EA on increased MDD risk (via e.g. less effective coping strategies), an impact of MDD on one's possibilities to obtain his or her full educational potential, or a third factor impacting both EA and MDD risk. Such a third factor could, for example, comprise of certain personality characteristics and of shared genetic effects. Analyses for the causes of the association between MDD risk and lower EA are relevant to further understand the etiology of MDD, but also as this might inform future preventions programs to reduce this deleterious link. In **Chapter 3** of this thesis different methods are, therefore, applied to test if the association between MDD risk and lower EA could be attributable to shared effects of genotyped SNPs.

Part B: Methodological aspects of study design, SNP-heritability, power and assortative mating

Different types of GWAS design

The GWAS cohorts that contribute to the Psychiatric Genomics Consortium (PGC) have recruited cases and controls with different strategies, which is likely to impact results from association testing but also estimates of the SNP heritability (the proportion of population variance in disorder risk attributable to genome-wide genotyped SNPs). Some of the MDD cohorts have ascertained cases from clinical settings and others from population; while most cohorts have carefully screened controls some do not.[56] Nevertheless, all MDD cohorts have recruited unrelated controls, which contrasts some of the ADHD and autism cohorts that apply proband-parents trio data. Trio data of affected probands and their parents is essential to detect de novo mutations, perform imprinting studies, and obtain accurately phased haplotypes, but also provides pseudocontrols constructed from the non-transmitted parental alleles for association testing. Pseudocontrols have regularly been applied in candidate gene studies to protect against population stratification, but have also been taken forward for GWAS studies where other methods are also available to protect against stratification with GWAS data such as genomic principal components[57] or mixed model association analysis.[58] In **Chapter 6** of this thesis, the GWAS-properties of the trio design and use of unscreened controls are addressed by deriving the expected SNP-heritability and power to detect a risk variant, while also considering that some of

the trio cohorts overrepresented multiplex family (with more than one affected proband), and while taking into account that assortative mating has been found to occur for most psychiatric traits.[59]

Assortative mating

Interestingly, a number of studies have found evidence for assortative mating for psychiatric traits (a population spouse-correlation in risk for psychiatric disease). Depending on the mechanisms leading to assortment, there are different consequences of assortative mating. Under phenotypic assortment, assortative mating impacts on the genetic architecture of traits and on genetic tests. In **Chapter 6** of this thesis, the consequence are derived of assortative mating for SNP heritability estimates and power to detect single risk variants, and in **Chapter 7** boundaries are defined for the genetic consequences of assortative mating for psychiatric traits in terms of the population disease prevalence and heritability in the next generation.

Aims of this thesis

This thesis aims to study the complex link between genetic effects and environment in depression in real data, and to explore boundaries for some of the consequences of GWAS study design and assortative mating from a theoretical perspective. **Chapter 2** contributes to the debate on the possible moderating effect of 5-HTTLPR on the link between childhood trauma and depression by testing this GxE-effect in NESDA and NTR. **Chapter 3** contributes to the research on the many phenotypic associates of MDD by testing whether the deleterious link between lower education attainment and increased MDD risk can be explained by genome-wide genotyped SNPs. **Chapter 4** adds a hypothesis to the literature on heterogeneity of MDD's genetic effects by testing for interaction between polygenic risk scores and childhood trauma in depression. **Chapter 5** places the findings from Chapter 4 in a broader context by analyzing childhood trauma and polygenic risk for MDD in the large international Psychiatric Genomics Consortium. **Chapter 6** aims to serve decisions for GWAS study design by addressing the consequences of the trio design and unscreened controls on estimates of SNP-heritability and power to detect genetic risk variants. **Chapter 7** contributes to the literature on assortative mating by exploring boundaries for the genetic consequences of assortative mating with respect to population prevalence and heritability in the next generation.

REFERENCES

1. Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat. Rev. Genet. *13*, 537–551.

2. Graaf, R. de, Have, M. ten, Gool, C. van, and Dorsselaer, S. van (2012). Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. Soc. Psychiatry Psychiatr. Epidemiol. *47*, 203–213.

3. Eaton, W.W., Anthony, J.C., Gallo, J., Cai, G., Tien, A., Romanoski, A., Lyketsos, C., and Chen, L.S. (1997). Natural history of Diagnostic Interview Schedule/DSM-IV major depression. The Baltimore Epidemiologic Catchment Area follow-up. Arch. Gen. Psychiatry *54*, 993–999.

4. Spijker, J., de Graaf, R., Bijl, R. V, Beekman, A.T.F., Ormel, J., and Nolen, W.A. (2002). Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br. J. Psychiatry *181*, 208–213.

5. Kuehner, C. (2003). Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. Acta Psychiatr. Scand. *108*, 163–174.

6. Pedersen, C.B., Mors, O., Bertelsen, A., Waltoft, B.L., Agerbo, E., McGrath, J.J., Mortensen, P.B., and Eaton, W.W. (2014). A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. JAMA Psychiatry *71*, 573–581.

7. Spijker J, Bockting CLH, Meeuwissen JAC, Vliet IM van, Emmelkamp PMG, Hermens MLM, B.A., and Van (2013). Multidisciplinaire Richtijn Depressie (3e revisie). Trimbos-Instituut.

8. Blair-West, G.W., and Mellsop, G.W. (2001). Major depression: does a gender-based down-rating of suicide risk challenge its diagnostic validity? Aust. N. Z. J. Psychiatry *35*, 322–328.

9. World Health Organisation (2008). The global burden of disease: 2004 update. WHO Press.

10. Lorant, V., Deliège, D., Eaton, W., Robert, A., and Ansseau, M. (2003). Socioeconomic inequalities in depression: a meta-analysis. Am. J. Epidemiol. *157*, 98–112.

11. Hovens, J.G.F.M., Wiersma, J.E., Giltay, E.J., van Oppen, P., Spinhoven, P., Penninx, B.W.J.H., and Zitman, F.G. (2010). Childhood life events and childhood trauma in adult patients with depressive, anxiety and comorbid disorders vs. controls. Acta Psychiatr. Scand. *122*, 66–74.

12. Teicher, M.H., and Samson, J. a (2013). Childhood maltreatment and psychopathology: A case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. Am. J. Psychiatry *170*, 1114–1133.
13. Rotella, F., and Mannucci, E. (2013). Diabetes mellitus as a risk factor for depression. A meta-analysis of longitudinal studies. Diabetes Res. Clin. Pract. *99*, 98–104.
14. Goldstein, B.I., Carnethon, M.R., Matthews, K.A., McIntyre, R.S., Miller, G.E., Raghuveer, G., Stoney, C.M., Wasiak, H., and McCrindle, B.W. (2015). Major Depressive Disorder and Bipolar Disorder Predispose Youth to Accelerated Atherosclerosis and Early Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation *132*, 965–986.
15. Vreeburg, S.A., Hoogendijk, W.J.G., van Pelt, J., DeRijk, R.H., Verhagen, J.C.M., van Dyck, R., Smit, J.H., Zitman, F.G., and Penninx, B.W.J.H. (2009). Major Depressive Disorder and Hypothalamic-Pituitary-Adrenal Axis Activity. Arch. Gen. Psychiatry *66*, 617.
16. Arnone, D., McKie, S., Elliott, R., Juhasz, G., Thomas, E.J., Downey, D., Williams, S., Deakin, J.F.W., and Anderson, I.M. (2013). State-dependent changes in hippocampal grey matter in depression. Mol. Psychiatry *18*, 1265–1272.
17. Howren, M.B., Lamkin, D.M., and Suls, J. (2009). Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. Psychosom. Med. *71*, 171–186.
18. Vogelzangs, N., Duivis, H.E., Beekman, A.T.F., Kluft, C., Neuteboom, J., Hoogendijk, W., Smit, J.H., de Jonge, P., and Penninx, B.W.J.H. (2012). Association of depressive disorders, depression characteristics and antidepressant medication with inflammation. Transl. Psychiatry *2*, e79.
19. Kendler, K.S., Gardner, C.O., and Prescott, C.A. (2006). Toward a comprehensive developmental model for major depression in men. Am. J. Psychiatry *163*, 115–124.
20. Kendler, K.S., Gardner, C.O., and Prescott, C.A. (2002). Toward a comprehensive developmental model for major depression in women. Am. J. Psychiatry *159*, 1133–1145.
21. Krishnan, V., and Nestler, E.J. (2008). The molecular neurobiology of depression. Nature *455*, 894–902.
22. Hirschfeld, R.M. (2000). History and evolution of the monoamine hypothesis of depression. J. Clin. Psychiatry *61 Suppl 6*, 4–6.
23. Power, R.A., Kyaga, S., Uher, R., MacCabe, J.H., Långström, N., Landen, M., McGuffin, P., Lewis, C.M., Lichtenstein, P., and Svensson, A.C. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia

nervosa, or substance abuse vs their unaffected siblings. JAMA Psychiatry *70*, 22–30.
24. Polderman, T., and Benyamin, B. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. *47*, 702–709.
25. Kendler, K.S., Gatz, M., Gardner, C.O., and Pedersen, N.L. (2006). A Swedish national twin study of lifetime major depression. Am. J. Psychiatry *163*, 109–114.
26. Sullivan, P.F., Neale, M.C., and Kendler, K.S. (2000). Genetic Epidemiology of Major Depression: Review and Meta-Analysis. Am. J. Psychiatry *157*, 1552–1562.
27. McGuffin, P., Katz, R., Watkins, S., and Rutherford, J. (1996). A hospital-based twin register of the heritability of DSM-IV unipolar depression. Arch. Gen. Psychiatry *53*, 129–136.
28. Foley, D.L., Neale, M.C., and Kendler, K.S. (1998). Reliability of a lifetime history of major depression: implications for heritability and co-morbidity. Psychol. Med. *28*, 857–870.
29. Boomsma, D.I., Willemsen, G., Sullivan, P.F., Heutink, P., Meijer, P., Sondervan, D., Kluft, C., Smit, G., Nolen, W. a, Zitman, F.G., et al. (2008). Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur. J. Hum. Genet. *16*, 335–342.
30. López-León, S., Janssens, A.C.J.W., González-Zuloeta Ladd, A.M., Del-Favero, J., Claes, S.J., Oostra, B.A., and van Duijn, C.M. (2008). Meta-analyses of genetic studies on major depressive disorder. Mol. Psychiatry *13*, 772–785.
31. Wray, N.R., Pergadia, M.L., Blackwood, D.H.R., Penninx, B.W.J.H., Gordon, S.D., Nyholt, D.R., Ripke, S., MacIntyre, D.J., McGhee, K. a, Maclean, a W., et al. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. Mol. Psychiatry *17*, 36–48.
32. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.
33. Sullivan, P.F., Geus, E.J.C. De, Willemsen, G., James, M.R., Smit, J.H., Zandbelt, T., Arolt, V., Baune, B.T., Blackwood, D., Cichon, S., et al. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. Mol. Psychiatry *14*, 359–375.
34. Muglia, P., Tozzi, F., Galwey, N.W., Francks, C., Upmanyu, R., Kong, X.Q., Antoniades, a, Domenici, E., Perry, J., Rothen, S., et al. (2010). Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. Mol. Psychiatry *15*, 589–601.
35. Shyn, S.I., Shi, J., Kraft, J.B., Potash, J.B., Knowles, J. a, Weissman, M.M.,

Garriock, H. a, Yokoyama, J.S., McGrath, P.J., Peters, E.J., et al. (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. Mol. Psychiatry *16*, 202–215.

36. Shi, J., Potash, J.B., Knowles, J.A., Weissman, M.M., Coryell, W., Scheftner, W.A., Lawson, W.B., DePaulo, J.R., Gejman, P. V, Sanders, A.R., et al. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. Mol. Psychiatry *16*, 193–201.

37. Lewis, C.M., Ng, M.Y., Butler, A.W., Cohen-Woods, S., Uher, R., Pirlo, K., Weale, M.E., Schosser, A., Paredes, U.M., Rivera, M., et al. (2010). Genome-wide association study of major recurrent depression in the U.K. population. Am. J. Psychiatry *167*, 949–957.

38. Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W.J., Assendelft, W.J.J., et al. (2008). The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int. J. Methods Psychiatr. Res. *17*, 121–140.

39. Boomsma, D.I., de Geus, E.J.C., Vink, J.M., Stubbe, J.H., Distel, M. a, Hottenga, J.-J., Posthuma, D., van Beijsterveldt, T.C.E.M., Hudziak, J.J., Bartels, M., et al. (2006). Netherlands Twin Register: from twins to twin families. Twin Res. Hum. Genet. *9*, 849–857.

40. Sullivan, P.F. (2010). The psychiatric GWAS consortium: big science comes to psychiatry. Neuron *68*, 182–186.

41. Demirkan, A., Penninx, B.W.J.H., Hek, K., Wray, N.R., Amin, N., Aulchenko, Y.S., Dyck van, R., Geus de, E.J.C., Hofman, A., Uitterlinden, a G., et al. (2011). Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol. Psychiatry *16*, 773–783.

42. Lubke, G.H., Hottenga, J.J., Walters, R., Laurin, C., Geus, E.J.C. De, Willemsen, G., Smit, J.H., Middeldorp, C.M., Penninx, B.W.J.H., Vink, J.M., et al. (2012). Estimating the Genetic Variance of Major Depressive Disorder Due to All Single Nucleotide Polymorphisms. Biol. Psychiatry *72*, 707–709.

43. So, H.-C., Li, M., and Sham, P.C. (2011). Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genet. Epidemiol. *35*, 447–456.

44. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P.H., Mors, O., Mortensen, P.B., et al. (2009). Common variants conferring risk of schizophrenia. Nature *460*, 744–747.

45. Yang, J., Wray, N.R., and Visscher, P.M. (2010). Comparing apples and oranges: equating the power of case-control and quantitative trait association

studies. Genet. Epidemiol. *34*, 254–257.
46. Zimmerman, M., Ellison, W., Young, D., Chelminski, I., and Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? Compr. Psychiatry *56*, 29–34.
47. Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science *301*, 386–389.
48. Munafò, M.R., Durrant, C., Lewis, G., and Flint, J. (2009). Gene X environment interactions at the serotonin transporter locus. Biol. Psychiatry *65*, 211–219.
49. Risch, N., Herrell, R., Lehner, T., Liang, K.-Y., Eaves, L., Hoh, J., Griem, A., Kovacs, M., Ott, J., and Merikangas, K.R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. JAMA *301*, 2462–2471.
50. Karg, K., Burmeister, M., Shedden, K., and Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. Arch. Gen. Psychiatry *68*, 444–454.
51. Caspi, A., Hariri, A.R., Holmes, A., Uher, R., and 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. Am. J. Psychiatry *167*, 509–527.
52. Fergusson, D.M., Horwood, L.J., Miller, A.L., and Kennedy, M. a (2011). Life stress, 5-HTTLPR and mental disorder: findings from a 30-year longitudinal study. Br. J. Psychiatry *198*, 129–135.
53. Duncan, L.E., and Keller, M.C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am. J. Psychiatry *168*, 1041–1049.
54. Wendland, J.R., Martin, B.J., Kruse, M.R., Lesch, K.-P., and Murphy, D.L. (2006). Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR and rs25531. Mol. Psychiatry *11*, 224–226.
55. Kendler, K.S., and Gardner, C.O. (2010). Interpretation of interactions: guide for the perplexed. Br. J. Psychiatry *197*, 170–171.
56. Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H.R., Boomsma, D.I., Cichon, S., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.
57. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in

genome-wide association studies. Nat. Genet. *38*, 904–909.
58. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. *46*, 100–106.
59. Nordsletten, A.E., Larsson, H., Crowley, J.J., Almqvist, C., Lichtenstein, P., and Mataix-Cols, D. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. JAMA Psychiatry.

Chapter 2

No gene-by-environment interaction with 5-HTTLPR in a large Dutch sample

Wouter J. Peyrot
Christel M. Middeldorp
Rick Jansen
Johannes H. Smit
Eco J. C. de Geus
Jouke-Jan Hottenga
Gonneke Willemsen
Jacqueline M. Vink
Susanne Virding
Isabel Barragan
Magnus Ingelman-Sundberg
Sarah C. Sim
Dorret I. Boomsma
Brenda W. J. H. Penninx

Published in *The Journal of Affective Disorders 2013, 146:91-99*

Paper at publisher's website:
http://www.jad-journal.com/article/S0165-0327(12)00615-5/fulltext

ABSTRACT

Background: There is ongoing interest in the possible interaction of the serotonin-transporter-linked polymorphic region (5-HTTLPR) with environmental factors in determining Major Depressive Disorder (MDD). The current study contributes to this research area by comprehensively examining the interaction-effects and direct-effects of 5-HTTLPR and four environmental factors on MDD prevalence and course in a well-characterized longitudinal sample.

Methods: The sample consisted of 1625 patients with a CIDI-confirmed diagnosis of MDD and 1698 screened controls from the Netherlands. Four MDD outcomes were studied as dependent variables: one main MDD prevalence-outcome (lifetime MDD), two more severe MDD prevalence-outcomes (suicidal and chronic MDD), and one MDD course outcome (chronic versus non-chronic MDD). Because SNP rs25531 modifies the effect of 5-HTTLPR, haplotypes of 5-HTTLPR and rs25531 were measured. For the four MDD outcome measures, we examined the direct effects of 5-HTTLPR/rs25531-haplotypes, four environmental factors (stressful life-events, sexual abuse, low educational attainment, and childhood trauma) and their interaction in logistic regression models.

Results: The environmental factors had large and consistent effects on all four MDD outcomes, including course of MDD. The 5-HTTLPR/rs25531-haplotype had a suggestive effect on course of MDD, but not on presence of MDD. Gene-by-environment interaction was significant (<0.05) for one of the sixteen tests performed, which is not more than expected by chance.

Limitations: Environmental factors were not assessed before the onset of MDD.

Conclusions: Environmental factors had a strong impact on the presence and course of MDD, but no evidence for gene-by-environment interaction was found.

INTRODUCTION

Since the first findings of Caspi and colleagues in 2003,[1] there has been ongoing interest in a possible interaction between the serotonin-transporter-linked polymorphic region (5-HTTLPR), which contains a repeat length polymorphism, and environmental factors in Major Depressive Disorder (MDD). Caspi et al. showed that individuals with at least one short allele of 5-HTTLPR experienced more depressive symptoms, diagnosable depression and suicidality following a stressful life-event. Many studies aimed to replicate these findings with contradictory results, and two meta-analyses published in 2009 by Munafò et al.[2] and Risch et al.[3] combining data of 5 and respectively 14 studies showed no evidence for this gene-by-environment interaction. However, an ensuing meta-analysis in 2011 by Karg and colleagues used a different definition of stressful life events and a meta-analysis method that allowed inclusion of more studies (56 studies containing a total of 40749 subjects).[4] This meta-analysis supported Caspi's finding.

A major cause proposed for the conflicting results in studies on gene-by-environment interaction involving 5-HTTLPR lays in the different nature and measurement of the environmental factors that are considered.[5] Studies are not always comparable, because environmental factor-measurements differ both in content and in timing to the onset of MDD. For example, Karg et al. studied several environmental factors in their meta-analysis and found stronger gene-by-environment interaction for childhood trauma than for stressful life-events.[4] Low educational attainment is another environmental factor that is strongly associated with MDD,[6] but also to socioeconomic status,[7] for which some studies found interaction effects involving 5-HTTLPR[8,9] and some did not.[10]

Another possible explanation for the inconsistent findings in studies on gene-by-environment interaction involves differences in the depression measures used. Although some studies used DSM- based diagnosis of MDD, others employed continuous scales of self-reported (often milder) depression symptoms. In the meta-analysis of Karg et al., studies with self-reported depression showed less evidence for gene-by-environment interaction than studies with interview assessed depression.[4] Possibly, this was because self-reported depression measures are often state-measures neglecting remitted depression symptoms, whereas DSM-based depression measures might mark the more severe and clinically relevant depressed patients.

A final explanation for the inconsistent findings in studies on gene-by-environment interaction with 5-HTTLPR may be the measurement of the functional variants in the 5-HTTLPR. Functional characterisation of the 5-HTTLPR

has evolved and it has been shown that long alleles of 5-HTTLPR that form a haplotype with the G allele of rs25531 are functionally equivalent to short alleles, which are less expressed.[11,12] Risch et al. and Karg et al. made no mention of rs25531 in their meta-analysis and few studies took rs25531 into account. However, 5-HTTLPR-L/rs25531-G haplotypes have a frequency of 6.5% and,[12] therefore, it seems crucial to take rs25531 into account when aiming to study the functional length of 5-HTTLPR.

The present study contributes to the ongoing debate by comprehensively examining four environmental factors (stressful life-events, sexual abuse, educational attainment and childhood trauma) and their interaction-effects with the functional length of 5-HTTLPR in a large and well-characterized study from the Netherlands. We studied patients with an interview-assessed and DSM-IV based diagnosis of MDD compared to carefully screened controls and, in addition to two more severe prevalence-outcomes (chronic MDD and suicidal MDD) to further increase the contrast between patients and controls. Since the 5-HTTLPR polymorphism has not only been linked to the onset of depression, but also to its chronicity,[13] we additionally examined gene-by-environment interaction in the course of MDD.

METHODS

Subjects

Data from 1727 unrelated MDD patients and 1792 healthy controls from the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Registry (NTR) were analyzed. The NESDA study is an ongoing longitudinal cohort study of MDD and anxiety disorders and its subjects were recruited from mental health care settings, general practitioners, and the general population in the period from 2004 to 2007.[14] The NTR study has been collecting data on Dutch twin families since 1991 and comprises data on nearly 22,000 subjects who have been assessed longitudinally for depressive symptoms (multiple instruments), anxiety and neuroticism.[15] Subjects from the NTR were included in this study based on longitudinal data up to 2005. Both studies (NESDA and NTR) were approved by the institutional Review Board and all their participants provided written informed consent. Most patients were from NESDA (1598 versus 129 from NTR) and patients were included in the current study when they were between 18 and 77 years of age and had a lifetime DSM-IV diagnosis of MDD. MDD diagnosis was assessed by specially trained clinical staff in a face-to-face interview using the Composite International Diagnostic Interview (CIDI, version 2.1). Persons who were not fluent in Dutch and those with a primary diagnosis of a psychotic

disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use dependence were excluded at NESDA study baseline. Most controls were from the NTR (1640 versus 152 from NESDA) and were included when they had no lifetime diagnosis of MDD, did not take any medication that may have been prescribed to treat MDD, had no known first-degree relatives with MDD and a low factor score based on a multivariate analysis of depressive complaints, anxiety, neuroticism, and somatic anxiety.[16] The 152 controls from NESDA had no lifetime diagnosis of MDD or anxiety disorder, as assessed by CIDI, no other major psychiatric disorder and scored low (<4) on the Inventory of Depressive Symptoms scale.[18] MDD patients and controls that were included had North-European ancestry, were matched for age and gender and were unrelated.

Assessment of environmental factors

Lifetime and recent stressful life-events. The number of various stressful life-events encountered in lifetime and those encountered in past year (recent stressful life-events) were assessed rather comparably in both studies. In NESDA, stressful life-events were assessed with the Brugha List of Threatening Experiences.[17] This assessment took place at the same day as the CIDI interview. Questionnaires from the NTR[19,20] were matched resulting in the following six stressful life-events encountered during lifetime up to assessment (combined prevalence in patients and controls between brackets): severe disease or victim of physical violence of self (35.2%); severe disease or victim of physical violence of close relative (68.9%); death of close relative (85.7%); forced dismissal from job (20.7%); ending of enduring intimate relationship (39.9%); and being robbed (33.8%). The number of different life-events encountered in this study ranged from 0 to 6.

Sexual abuse. The occurrence of lifetime sexual abuse was assessed in a slightly different way across studies. NESDA-subjects were asked if they were ever touched or forced to touch someone in a sexual way against their will. NTR-subjects were asked if they had ever been victim of a sexual misdeed, which was specified as being raped or assaulted. Because different wording across studies could have resulted in different prevalences, cohort-status (NESDA or NTR) was added as an additional covariate to all analyses focusing on sexual abuse. Although results in cohort-status unadjusted and cohort-status adjusted analyses differed slightly in estimated effect sizes, overall conclusions on the importance of sexual abuse and their interaction with 5HTTLR-gene were very comparable in unadjusted and adjusted analyses.

Educational attainment. Educational attainment was defined as the years required to obtain the highest diploma attained.

Childhood trauma. Childhood trauma was measured in NESDA with the instrument of the Netherlands Mental health Survey and Incidence Study.[21] Subjects were asked for emotional neglect, psychological abuse, physical abuse and sexual abuse. The definition of emotional neglect included lack of parental attention or support and ignorance of one's problems and experiences. Psychological abuse was defined as being verbally abused, undeserved punishment, subordinated to siblings and being blackmailed. Physical abuse was defined as being kicked or hit with hands or an object, beaten up or physical abuse in any other way. Sexual abuse was defined as being sexually approached against your will, meaning being touched or having to touch someone in a sexual way. Participants answered 'yes' or 'no' to each of the four forms of childhood trauma and were asked to give an indication about the frequency on a five-point scale, '1' once, '2' sometimes, '3' regular, '4' often and '5' very often. In the analyses, the frequencies were categorized into three groups (0: absent, 1: once or sometimes, 2: regular, often and very often). The number of different traumas encountered were combined with their frequencies, resulting in a sum score ranging from 0 to 8, as has been defined before.[22–24]

MDD outcomes

Three MDD prevalence outcomes and one MDD course outcome were examined as dependent variables. The first MDD prevalence-outcome compared all MDD patients (defined by a lifetime DSM-IV based diagnosis) to healthy controls. Two additional MDD prevalence-outcomes were examined to further increase the contrast between patients and controls and compared suicidal MDD patients to healthy controls and chronic MDD patients to healthy controls. Suicidal MDD was defined as having ever attempted to commit suicide as assessed in the CIDI interview. Chronic MDD was defined as having a MDD diagnosis and symptom duration of more than two years. Symptom duration was obtained for NESDA only using baseline and 2-year follow-up data of those individuals with an MDD diagnosis one year prior to baseline (n = 997). Symptom duration was assessed using the Life Chart Interview (LCI), which uses a calendar approach to assess the percentage of time that symptoms were present during the four years prior to and the two years following baseline. Computations with LCI data were described in more detail by Penninx et al.[25] In the analyses on recent stressful life-events MDD patients with a past year diagnosis of MDD were included only, changing the

main MDD and suicidal MDD measure, but not the chronic MDD measure, as chronicity could only be assessed for subjects with past year MDD.

In addition to comparing patients to healthy controls, chronic MDD patients were also compared to non-chronic MDD patients. This case-only analysis yielded the opportunity to examine gene-by-environment interaction on the course of MDD.

5-HTTLPR and rs25531

Sample collection procedures and DNA isolation were harmonized between NESDA and NTR as previously described.[26] The 5-HTTLPR/rs25531 haplotypes were assessed by the PCR protocol described by Wendland et al.[12] at the Karolinska Institute in Stockholm (Sweden). In short, genomic DNA amounting to at least 10 ng was used for PCR amplification of the long (L) and short (S) promoter repeats (forward primer: 5'-TCCTCCGCTTTGGCGCCTCTTCC-3', reverse primer: 5'-TGGGGGTTGCAGGGGAGATCCTG-3'). Half of the reaction product was digested with HpaII FastDigest (FD0514, Fermentas) at 37°C for 10 minutes whereas half was left undigested. HpaII digests amplicons carrying the rs25531-G genotype, but leaves rs25531-A undigested. The digested and undigested amplicons were separated on a 3 % Ultrapure agarose (Invitrogen) gel at 160 V for approximately 1 hour. Due to the methylation sensitivity of HpaII, 160 samples were also digested using the methylation-insensitive MspI FastDigest (FD0544, Fermentas) and no discrepancies regarding digestion and interpretation of genotypes were discovered. Based on the length difference between the S and L amplicons, and the resulting digested amplicons caused by the presence of rs25531-G, the 5-HTTPR/rs25531 haplotypes could be resolved.

Quality control of genotypes was performed with additionally genotyped trios (30 trios) and duplicates (18 duplicates) . There were no Mendelian errors or mismatches of duplicates. In addition, the 5-HTTLPR/rs25531 haplotypes were in Hardy Weinberg Equilibrium ($p = 0.9$). The 5-HTTLPR/rs25531 haplotypes were used to define the functionality of the length polymorphism, with the 5-HTTLPR-long/rs25531-G haplotypes classified as short 5-HTTLPR alleles. The number of functional short alleles (0, 1 or 2) from all unrelated subjects were used to test for direct gene effects, and gene-by-environment interaction on the presence and course of MDD.

Statistical Analyses

MDD patients were compared to healthy controls with respect to age, gender, and the four environmental factors using t-test and chi-square statistics.

Associations between the number of functional short 5-HTTLPR alleles and environmental factors were examined using linear regression and linear-by-linear chi-square statistics to test for gene-environment correlation as this can influence tests of gene-by-environment interaction and lead to spurious results.[27]

The impact of the number of functional short 5-HTTLPR alleles (0, 1, or 2), environmental factors and gene-by-environment interaction was examined for the three MDD prevalence-outcomes and MDD course-outcome. First, the direct effects of the number of functional short 5-HTTLPR alleles and the direct effects of the four environmental factors were assessed and, subsequently, their interaction effects were assessed. In the analyses of interaction effects, the main effects of the number of functional short 5-HTTLPR alleles and the concerning environmental factor were included. All analyses were conducted using logistic regression with age and gender as covariates. Analyses focusing on sexual abuse additionally included the subject's cohort-status (NESDA or NTR). All together sixteen interaction effects (four environmental factors times four MDD outcomes) were examined and we, therefore, had to correct for multiple comparison. However, the MDD prevalence outcomes were correlated and Bonferroni-correction would have been too stringent. Therefore, the threshold for significance was set at 0.05 with the number of tests taken into account in the interpretation of the results.

Although the interaction model (MDD outcome = $b_0 + b_1 \cdot$5-HTTLPR + $b_2 \cdot$E + $b_3 \cdot$5-HTTLPR x E + $b_4 \cdot$gender + $b_5 \cdot$age) provides tests for 5-HTTLPR (b_1) and the environmental factor (b_2), these coefficients do not represent the direct effects on the MDD outcome, but rather the effect of 5-HTTLPR when E = 0 and the effect of E when 5-HTTLPR = 0 respectively.[28] Therefore, we examined the direct effects of 5-HTTLPR, the direct effects of the environmental factors and their interaction-effects in separate models as described above.

In addition to the straightforward test for multiplicative interaction, additional tests were performed for interaction as a departure from additivity using the procedure described by Knol et al.[29] In this procedure we used the outcome of a logistic regression model to estimate the Relative Excess Risk due to Interaction (RERI). A RERI smaller than or bigger than zero indicates evidence for additive interaction. The 95 percent confidence intervals of the RERI were estimated using bootstrap simulations. In this way we tested for additive interaction for the five main environmental factors (lifetime stressful life-events, recent stressful life-events, sexual abuse, educational attainment and childhood trauma) with 5-HTTLPR for the four main outcome measures (all MDD, suicidal MDD, chronic MDD and chronic versus non-chronic MDD).

Finally, in order to compare the impact of different environmental factors on MDD prevalence and course, the variation explained by the four environmental with age and gender as covariates, as estimated by Nagelkerke pseudo R-Squares, were compared. Nagelkerke R-Squares are used in logistic regression to approximate the R-Square known from linear regression. Analyses were conducted in SPSS and R.[30]

Table 1. Sample characteristics by MDD status

	MDD patients N = 1625	**Healthy controls N = 1698**	**p-value***
Sociodemographic variables			
Age, mean years ± SD	42.62 ± 12.56	43.91 ± 14.45	0.006
Women, %	69.66	61.66	<0.001
Environmental factors			
Lifetime stressful life-event, number ± SD	3.30 ± 1.24	2.29 ± 1.19	<0.001
N with data available	*1593*	*1411*	
Recent stressful life-event, number ± SD	0.60 ± 0.83	0.45 ± 0.68	<0.001
N with data available[1]	*929*	*1411*	
Lifetime sexual abuse, %	33.27	6.06	<0.001
N with data available	*1596*	*1419*	
Educational attainment, years ± SD	11.98 ± 3.25	12.62 ± 3.30	<0.001
N with data available	*1597*	*1467*	
Childhood trauma, number ± SD	2.00 ± 2.24	0.35 ± 1.00	<0.001
N with data available	*1502*	*143*	
Types of MDD			
Number suicidal MDD, N (%)	123 (75.69%)		
Number chronic MDD, N (%)[2]	396 (24.37%)		

*p-values based on t-test (continuous or > 5 ordered categories) and chi-square statistics (dichotomous variables).

[1]patients are subjects with recent MDD only

[2]out of 837 MDD patients with a recent MDD diagnosis and data on chronicity.

Table 2. Correlation of 5-HTTLPR* to environmental factors

		5-HTTLPR*				
Environmental stressor		**0**	**1**	**2**	**Beta**	**p-trend**
Lifetime stressful life-events	Mean value (SE) N	2.82 (0.05) 731	2.87 (0.03) 1547	2.72 (0.05) 726	-0.03	0.16
Recent stressful life-events	Mean value (SE) N	0.51 (0.03) 729	0.51 (0.02) 1544	0.47 (0.03) 725	-0.02	0.29
Sexual abuse	Prevalence in % N	17.76 732	21.48 1555	21.02 728		0.12
Educational attainment	Mean value (SE) N	12.14 (0.12) 746	12.32 (0.08) 1575	12.38 (0.12) 743	0.03	0.16
Childhood trauma	Mean value (SE) N	1.66 (0.11) 402	1.96 (0.08) 859	1.83 (0.11) 384	0.03	0.27

*p-trends are based on linear regression (continuous or > 5 ordered categories) and linear-by-linear chi-square tests (dichotomous variables). The effect size (beta) is displayed for linear regression results.

RESULTS

Sample characteristics

A total of 1625 MDD patients and 1698 healthy controls had valid data on their number of functional short 5-HTTLPR alleles, i.e. 5-HTTLPR/rs25531 haplotypes. Data on environmental factors were missing for 319 (stressful life-events), 308 (sexual abuse), 259 subjects (educational attainment). Data on childhood trauma were only available for NESDA participants, leaving 1502 patients and 143 controls for these analyses.

The demographics of healthy controls (n = 1698) and MDD patients (n = 1625) are given in Table 1. Controls were slightly older (43.9 versus 42.6 years, p = 0.006) and less often female (61.7% versus 69.7%, p <0.001). Stressful life-events, sexual abuse, and childhood trauma were significantly more frequent and educational attainment was significantly lower in MDD patients compared to healthy controls . Out of all 1625 MDD patients, 123 (7.6%) had suicidal MDD. Out of all 837 patients with a current diagnosis of MDD one year prior to NESDA baseline and sufficient LCI data, 396 (47%) fulfilled criteria for chronic MDD.

Table 2 reports the associations between environmental factors with number of functional short 5-HTTLPR alleles. The occurrence of environmental factors was not significantly associated with the number of functional short 5-HTTLPR alleles, indicating the absence of gene-environment correlation.

Effects of 5-HTTLPR, environmental factors and their interaction on MDD prevalence measures

Table 3 displays the number of patients and controls and the effect of the number of functional short 5-HTTLPR alleles on the three MDD prevalence-outcomes (MDD, suicidal MDD and chronic MDD). In none of the analyses, the number of functional short 5-HTTLPR alleles had any effect on MDD outcome measures.

Effects of environmental factors on the three MDD prevalence-outcomes were strong. The effects of the number of lifetime stressful life-events (with ORs of 2.01, 2.59 and 2.11 per stressful life event; all p < 0.001) , recent stressful life-events (with ORs of 1.29 , 1.36 and 1.29 per stressful life event; all p < 0.001) and childhood trauma (with ORs of 1.98 , 2.47 and 2.15 per increased score on the overall index; all p < 0.001) were strong and rather comparable across the MDD prevalence outcomes. In contrast, high educational attainment was protective for the three MDD outcomes with ORs of 0.94 , 0.84 and 0.86 respectively (all p < 0.001). For sexual abuse, the OR was smallest for MDD versus controls (OR = 2.79), higher for chronic MDD (OR = 3.17), and very high for suicidal MDD (OR = 6.46) (all p < 0.001). When we examined the effect of sexual abuse separately within

Table 3. Effects of 5-HTTLPR, environmental factors and their interaction on MDD, suicidal MDD and chronic MDD versus healthy controls

Lifetime stressful life-events (E)

	N		5-HTTLPR		E		5-HTTLPR x E	
MDD outcome measure	**Patient**	**Control**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
MDD versus healthy control	1593	1411	0.96 (0.86-1.06)	0.39	2.01 (1.88-2.16)	<0.001	1.00 (0.91-1.10)	1.00
Suicidal MDD versus healthy control	121	1411	1.04 (0.80-1.36)	0.75	2.59 (2.20-3.07)	<0.001	1.32 (1.03-1.69)	0.03
Chronic MDD versus healthy control	395	1411	1.03 (0.88-1.21)	0.70	2.11 (1.91-2.33)	<0.001	0.97 (0.84-1.12)	0.63

Recent stressful life-events (E)

	N		5-HTTLPR		E		5-HTTLPR x E	
MDD outcome measure	**Patient**	**Control**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
MDD versus healthy control	929	1411	0.91 (0.81-1.02)	0.11	1.29 (1.15-1.44)	<0.001	1.05 (0.89-1.23)	0.59
Suicidal MDD versus healthy control	93	1411	0.88 (0.65-1.18)	0.40	1.36 (1.03-1.77)	0.03	1.16 (0.79-1.73)	0.45
Chronic MDD versus healthy control	395	1411	1.03 (0.88-1.21)	0.70	1.29 (1.10-1.49)	<0.001	1.06 (0.86-1.31)	0.61

Sexual abuse (E)

	N		5-HTTLPR		E		5-HTTLPR x E	
MDD outcome measure	**Patient**	**Control**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
MDD versus healthy control	1596	1419	0.99 (0.82-1.20)	0.94	2.79 (1.90-4.13)	<0.001	1.61 (0.92-2.79)	0.09
Suicidal MDD versus healthy control	121	1419	1.07 (0.77-1.47)	0.70	6.46 (3.61-11.80)	<0.001	2.04 (0.94-4.51)	0.07
Chronic MDD versus healthy control	396	1419	1.06 (0.81-1.39)	0.68	3.17 (1.96-5.28)	<0.001	1.31 (0.67-2.61)	0.43

Table 3. (continued)

Educational attainment (E)

	N		5-HTTLPR		E		5-HTTLPR x E	
MDD outcome measure	**Patient**	**Control**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
MDD versus healthy control	1597	1467	0.96 (0.87-1.07)	0.45	0.94 (0.92-0.96)	<0.001	1.00 (0.97-1.03)	0.80
Suicidal MDD versus healthy control	121	1467	1.05 (0.81-1.36)	0.74	0.84 (0.79-0.90)	<0.001	1.03 (0.95-1.12)	0.50
Chronic MDD versus healthy control	396	1467	1.04 (0.89-1.22)	0,64	0.86 (0.83-0.89)	<0.001	0.98 (0.93-1.03)	0.50

Childhood trauma (E)

	N		5-HTTLPR		E		5-HTTLPR x E	
MDD outcome measure	**Patient**	**Control**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
MDD versus healthy control	1502	143	0.97 (0.76-1.25)	0.83	1.98 (1.66-2.44)	<0.001	1.03 (0.78-1.37)	0.83
Suicidal MDD versus healthy control	115	143	1.04 (0.74-1.46)	0.81	2.47 (1.98-3.23)	<0.001	1.19 (0.85-1.68)	0.30
Chronic MDD versus healthy control	396	143	1.06 (0.81-1.39)	0.67	2.15 (1.78-2.67)	<0.001	1.09 (0.81-1.44)	0.57

The direct effects of the number of functional short 5-HTTLPR alleles (0, 1 or 2), four various environmental factors (E) and their interaction effects (5-HTTLPR x E) on the three MDD prevalence-outcomes. For every combination of E and MDD outcome measure, the direct effect of 5-HTTLPR, direct effect of E, and their interaction effect were estimated in separate logistic regression models including age, gender and, for the analysis focusing on sexual abuse, also cohort-status. The models estimating 5-HTTLPR x E included 5-HTTLPR and E in addition. Note that for recent stressful life-events only subjects with recent MDD were included.

NESDA and NTR – to prevent a potential impact of measurement differences across studies – strongly significant effects of sexual abuse on the prevalence of MDD were confirmed (in NESDA: OR = 2.23 with 95% CI = 1.41-3.52; and in NTR: OR = 4.01 with 95% CI = 2.20-7.33) indicating that measurement differences are not responsible for the described association.

The impact of the environmental factors on MDD versus controls were compared to each other in two steps. First, Nagelkerke R-Squares were compared in the sample who had data on lifetime stressful life-events, sexual abuse and educational attainment available (1587 MDD patients and 1370 controls). Lifetime stressful life-events explained more variation (R^2 = 0.210) than sexual abuse (R^2 = 0.158), and both explained more variation than educational attainment (R^2 = 0.026). In addition, including all these three environmental factors in one model, all effects remained significant with $p < 0.001$, which indicates that they had independent effects on MDD. Second, to compare childhood trauma to the three other environmental factors, the Nagelkerke R-Squares were also computed on subjects with data on all four environmental factors available (NESDA only with 1501 MDD patients and 143 controls). In this comparison, childhood trauma explained more variation (R^2 = 0.145) than the other three environmental factors (R^2 = 0.046. R^2 = 0.029 and R^2 = 0.054 respectively). When all four environmental factors were examined in one model, sexual abuse had no significant effect, but the other three environmental factors did. Including sexual abuse subsequently into a model with one of the three other environmental factors, indicated that the original impact of sexual abuse was included in the effect of childhood trauma. This seems logical, because the childhood trauma measure included sexual abuse before the age of sixteen, which is also included in lifetime sexual abuse variable.

Gene-by-environment interaction with the number of functional short 5-HTTLPR alleles was tested for 5 environmental factors and 3 outcomes. Of these 15 tests, one interaction-effect had a p-value smaller than 0.05, namely, the interaction-effect with lifetime stressful life-event on suicidal MDD versus healthy controls (OR = 1.32, p = 0.03). The direction of this interaction effect is in line with the finding of Caspi: extra copies of the short allele and a higher number of stressful life-events contributing more to the risk on MDD than their additional risk. In the tests for interaction as departure from additivity, all 95% confidence intervals included 0, indicating no evidence for additive interaction.

Table 4. Effects of 5-HTTLPR, environmental factors and their interaction on chronic MDD versus non-chronic MDD

	N		5-HTTLPR		E		5-HTTLPR x E	
Environmental factor	**Chronic**	**Non-chronic**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
Lifetime stressful life-events (E)	395	441	1.24 (1.01-1.51)	0.04	1.12 (1.00-1.26)	0.04	1.03 (0.88-1.20)	0.74
Recent stressful life-events (E)	395	441	1.24 (1.02-1.51)	0.04	1.01 (0.86-1.20)	0.87	1.04 (0.82-1.32)	0.74
Sexual abuse (E)	396	441	1.24 (1.02-1.52)	0.03	1.69 (1.25-2.29)	<0.001	1.20 (0.78-1.84)	0.41
Educational attainment (E)	396	441	1.24 (1.02-1.52)	0.03	0.91 (0.87-0.95)	<0.001	0.95 (0.89-1.02)	0.15
Childhood trauma (E)	396	440	1.25 (1.02-1.53)	0.03	1.15 (1.08-1.23)	<0.001	1.03 (0.94-1.13)	0.52

The direct effects of the number of functional short 5-HTTLPR alleles (0, 1 or 2), four various environmental factors (E) and their interaction effects (5-HTTLPR x E) on course of MDD. For every E, the direct effect of 5-HTTLPR, direct effect of E, and their interaction effect were estimated in separate logistic regression models including age, gender and, for the analysis focusing on sexual abuse, also cohort-status. The models estimating 5-HTTLPR x E included 5-HTTLPR and E in addition.

Effects of 5-HTTLPR, environmental factors and their interaction on the course of MDD

Within patients, impact of environmental factors on the course of MDD was evaluated (Table 4). The effect of the number of functional short 5-HTTLPR alleles on the course of MDD was 1.24 (p = 0.03) indicating that the functional short allele increases the risk of a chronic course of MDD. The environmental factors had direct effects on the course of MDD: number of lifetime stressful life-events (OR = 1.12, p = 0.04), sexual abuse (OR = 1.69, p < 0.001) and childhood trauma (OR = 1.15, p < 0.001) increased the risk of chronic MDD versus non-chronic MDD and educational attainment was protective (OR = 0.91, p < 0.001). Only the recent stressful life-events had no significant effect on the course of MDD (OR = 1.01, p= 0.87). Most variation was explained by childhood trauma and educational attainment, with Nagelkerke R-Squares of 0.046 and 0.049 respectively, and variation explained by lifetime stressful life-events (0.022) and sexual abuse (0.033) was considerably less. When these four environmental factors were included in a model together a Nagelkerke R-Square of 0.081 was found. No significant gene-by-environment interaction-effect on the course of MDD was found for any of the five environmental factors. In the tests for interaction as departure from additivity, no evidence was found for additive interaction.

As suicidality also selects a more severe subgroup of MDD patients, suicidal MDD patients were compared to non-suicidal MDD patients in post-hoc analyses, analogous to comparing chronic versus non-chronic MDD patients. These analyses showed that all environmental factors had a significant effect on the risk of suicide attempt for MDD patients in the expected direction: lifetime life-events (OR = 1.41; 95%CI 1.21-1.65; p < 0.001), sexual abuse (OR = 2.69; 95%CI 1.82-4.00; p < 0.001) and childhood trauma (OR = 1.34; 95%CI 1.24-1.45; p < 0.001) increased the risk on suicide attempts, whereas higher educational attainment reduced this risk (OR = 0.90; 95%CI 0.85-0.96; p < 0.001). The functional length of 5-HTTLPR had no effect on the risk of suicide attempts in MDD patients. Of the four environmental factors tested, only lifetime life-events showed significant interaction with 5-HTTLPR (OR = 1.41; 95%CI 1.12-1.76; p = 0.002).

Imputation of 5-HTTLPR

We checked whether it was possible to increase sample size by imputing 5-HTTLPR using the haplotype proxy of surrounding SNPs from Vinkhuyzen et al.[31] This haplotype proxy consists of rs2129785-T and rs11867581-A and tags the short allele of 5-HTTLPR with an r^2 of 0.775. The frequencies of 5-

HTTLPR/rs2129785/rs11867581-haplotypes given by Vinkhuyzen et al. were used to simulate a reference sample of 2823 subjects (the number of subjects in the discovery set of Vinkhuyzen et al). With this simulated reference sample and the available SNP data, Beagle- software was used to impute 5-HTTLPR.[32] 5-HTTLPR was imputed with an r^2 of 0.72 and comparison of imputed to genotyped values of 5-HTTLPR showed an rather low accuracy of 87%. This limited accuracy does not justify imputation of 5-HTTLPR on an extended sample of NESDA and NTR. Moreover, Vinkhuyzen et al found that rs25531 could not be tagged by SNPs giving rise to an even greater inaccuracy of imputation with respect to the functional length of 5-HTTLPR.

DISCUSSION

Since first reported by Caspi and colleagues in 2003,[1] there has been a fierce controversy about the reproducibility of the interaction between 5-HTTLPR and environmental factors in MDD. This controversy can only be resolved by empirical data. To that aim, we tested the effects of four environmental factors (stressful life-events, sexual abuse, educational attainment, and childhood trauma), the number of functional short 5-HTTLPR alleles and their interaction on MDD prevalence and course in a large sample from the Netherlands. We found that the environmental factors had large and consistent direct effects on both prevalence and course of MDD. Additionally, our results suggested that the environmental impact is stronger for the more severe outcomes (suicidal MDD patients versus controls and chronic MDD patients versus controls) than for the main outcome (all MDD patients versus controls). Comparison of Nagelkerke R-Squares showed that of all environmental factors, childhood trauma explained most variation in both the prevalence and course of MDD. We did not find a direct gene-effect of the number of functional short 5-HTTLPR alleles on the prevalence of MDD, but we did find some effect (p=0.03) on course of MDD, with short alleles contributing to chronic course. MDD course is a very different concept and outcome than the MDD prevalence outcome. Consequently, it may be too conservative to further adjust the course analyses for the number of associations tested for the prevalence outcomes. Nevertheless, it is clear that our finding for a direct effect on course of MDD deserves confirmation in future longitudinal studies. Out of all 16 tests conducted, we found one gene-by-environment interaction-effect with an uncorrected p-value smaller than 0.05, namely, in the test with stressful life-events for the suicidal MDD outcome (p = 0.03). We argued the Bonferroni-correction to be too stringent, because the different MDD

outcomes are correlated, but this single p-value of 0.03 is not more than expected by chance.

Our finding that the environmental factors have a large impact on MDD is in line with previous literature.[1,6] The absence of a direct effect of 5-HTTLPR on MDD prevalence-outcomes in our study is in line with the meta-analysis by Risch et al.,[3] but not with the meta-analysis by Clarke et al.,[33] where a direct effect of 5-HTTLPR on MDD was found. Because the direct gene-effect found by Clarke et al. was very small, the lack of a gene-effect in our study could also be caused by the limited sample size. The effect of environmental factors on course of MDD, and childhood trauma in particular, is in line with a recent meta-analysis.[34] The suggestive effect of 5-HTTLPR on course of MDD found in our study is interesting, but more studies are needed, as the effect of 5-HTTLPR on course of MDD has not been examined in many studies. One other study found no such effect,[13] whereas another study did find an effect of 5-HTTLPR on the course of MDD,[35] but in the opposite direction with long alleles contributing to chronic course. We found no evidence for gene-by-environment interaction, which is in line with the meta-analysis of Risch et al. and Munafò et al.,[2,4] but it is not in line with the meta-analysis of Karg et al.[4]

Some studies found significant interaction-effects in tests different to the one published by Caspi et al.[1] For example, Uddin et al. reported gene-by-environment interaction for males only[9] and Brummett et al. reported gene-by-environment-by-gender interaction.[8] We performed these tests on our sample to be comprehensive, but none of the interaction effects in male-only and female-only analyses nor results for gene-by-environment-by-gender interaction were significant, indicating that sex-specific results are not further providing evidence for the presence of gene-by-environment interaction. Uher et al. found that gene-by-environment interaction is stronger for childhood trauma in persistent depression,[36] but in our study we found no interaction effect for childhood trauma in chronic MDD. Some have argued that biological interaction should better be tested as departure from additivity than as departure from multiplicativity, as in biological interaction both causes are needed for a disease to develop.[29,36] Therefore, we additionally tested for interaction as departure from additivity, but these analyses also showed no evidence for interaction.

Our study is well sized with around 1600 patients and 1400 controls for the main MDD prevalence outcome. For comparison, only three of the 56 studies included in the meta-analysis of Karg et al. had larger sample sizes.[4] Nevertheless, only interaction-effect sizes smaller than 0.87 and larger than 1.17 could be detected with a power of 0.8 for stressful life-events in our study, as computed

with Quanto-software.[37] Thus, our study lacked power for possible smaller interaction effects. Therefore, and because we had additional subjects with genome-wide SNP data but without measurements of 5-HTLLPR, we checked whether it was possible to increase sample size by imputing 5-HTTLPR using the haplotype proxy of surrounding SNPs from Vinkhuyzen et al.[31] However, limited accuracy of this imputation in our sample was found. Moreover, Vinkhuyzen et al. found that rs25531 could not be tagged by SNPs giving rise to an even greater inaccuracy of imputation with respect to the functional length of 5-HTTLPR. Therefore, we restricted our analyses to the subjects for which 5-HTTLPR/rs25531-haplotypes were genotyped.

The question remains whether we would have been able to detect interaction-effects when we would have had a larger sample size. However, of the 56 studies included in the meta-analysis of Karg et al.,[4] eight contained more than 1000 subjects and of these only two found positive results for gene-by-environment interaction. This phenomenon, that larger studies produce more negative findings than smaller studies, was explained by Duncan et al. as an indication of publication bias amongst smaller studies.[5]

In our study we performed analyses with 5-HTTLPR/rs25531-haplotypes (denoted as functional 5-HTTLPR alleles) instead of plain 5-HTTLPR alleles, because long alleles of 5-HTTLPR are functionally equivalent to short alleles when they form a haplotype with the G-allele of rs25531.[38] However, only few studies took rs25531 into account and, therefore, we also conducted the interaction analysis for the number of plain short 5-HTTLPR alleles. These analyses showed two interaction effects with uncorrected p-values smaller that 0.05 out of the sixteen tests performed: for sexual abuse in the main MDD prevalence outcome and for educational attainment in the MDD course outcome. Thus, analyses with plain 5-HTTLPR yielded one more interaction-effect with a p-values smaller than 0.05 than the analyses with functional 5-HTTLPR. However, because two tests out of sixteen with a p-value smaller than 0.05 is still not very convincing, we suspect this finding to be due to chance.

The role of genetic factors in the relation between educational attainment and MDD has been studied by Lopez-Leon et al.,[39] who found that shared genetic factors play a role in the co-occurrence of lower socioeconomic status and symptoms of depression. Educational attainment is an easy measure to test, has a strong direct effect on MDD,[6] and has often been used as proxy for socioeconomic status, for which some evidence of gene-by-environment interaction was found.[8,9] To the best of our knowledge we were the first to test educational attainment (as measurement for environmental stress) for gene-by-

environment interaction with 5-HTTLPR. We found no correlation between 5-HTTLPR and educational attainment, suggesting that the gene does not contribute to the gene-by-environment correlation reported by Lopez-Leon et al. and that the interaction could be meaningfully tested. Nevertheless, low educational attainment, like the other sources of stressful life events did not support a gene-by-environment interaction for 5-HTTLPR.

Our study has several strengths. First, we analyzed a large sample of MDD patients and controls. As stated before, only three of the 56 studies included in the meta-analysis of Karg et al. had larger sample sizes.[4] Second, we used DSM-IV based diagnoses of MDD which ensured we studied the clinically relevant MDD patients. Moreover, our controls were screened for lifetime MDD diagnosis and a low probability of developing MDD later on in life. Third, we studied four different environmental factors making it less likely that we missed any gene-by-environment interaction in our sample. Fourth, we not only studied MDD patients compared to controls, but also two subgroups of more severe MDD patients in order to increase the contrast. Finally, we studied 5-HTTLPR/rs25531-haplotypes instead of plain 5-HTTLPR which is in line with the latest insights into the function of 5-HTTLPR.

There also are some limitations, including the assessment of stressful life-events only after the onset of MDD. In addition, some MDD patients had a lifetime, but no current diagnosis of MDD and because we assessed the lifetime occurrence of stressful life-events and sexual abuse, some reciprocal causation may have occurred. Reciprocal causation for life-events and MDD was reported by Middeldorp et al. and describes the phenomenon that life-events do not only increases the chance of developing MDD, but that MDD also increases the chance of encountering stressful life-events.[40] This reciprocal causation might have influenced the results for stressful life-events and sexual abuse, but not for childhood trauma and educational attainment and we found no gene-by environment interaction for these environmental factors either. We used two different studies with not entirely equal time and measurement phrames, and the controls in both studies were not screened for other psychiatric disorders (NTR) or for family history of MDD (NESDA), which all may have had a potential influence on our results. However, as we used stringent other selection criteria as described and only used the most comparable environment instruments in analyses, it seems unlikely that this has had a significant impact. An additional limitation is that a current depressive mood could influence the recall of environmental factors.

Our study has several potential clinical implications. It shows that environmental factors increase the risk on MDD, as well as the risk on chronic MDD for already depressed patients. In addition, our study suggests that childhood trauma contributes more to the risk on MDD than stressful life-events, sexual abuse or education attainment. For the course of MDD, childhood trauma and educational attainment contribute comparable risks on developing a chronic course of MDD and more than stressful life-events or sexual abuse. In sum, our study shows the strong effects of environmental factors on both the prevalence and course of MDD. It also shows the larger impact of childhood trauma compared to stressful life-events, sexual abuse and educational attainment. However, no evidence for gene-by-environment interaction effects for the 5-HTTLPR gene was found.

REFERENCES

1. Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science *301*, 386–389.

2. Munafò, M.R., Durrant, C., Lewis, G., and Flint, J. (2009). Gene X environment interactions at the serotonin transporter locus. Biol. Psychiatry *65*, 211–219.

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

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

5. Duncan, L.E., and Keller, M.C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am. J. Psychiatry *168*, 1041–1049.

6. Graaf, R. de, Have, M. ten, Gool, C. van, and Dorsselaer, S. van (2012). Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. Soc. Psychiatry Psychiatr. Epidemiol. *47*, 203–213.

7. Bjelland, I., Krokstad, S., Mykletun, A., Dahl, A. a, Tell, G.S., and Tambs, K. (2008). Does a higher educational level protect against anxiety and depression? The HUNT study. Soc. Sci. Med. *66*, 1334–1345.

8. Brummett, B.H., Boyle, S.H., Siegler, I.C., Kuhn, C.M., Ashley-Koch, A.,

Jonassaint, C.R., Züchner, S., Collins, A., and Williams, R.B. (2008). Effects of environmental stress and gender on associations among symptoms of depression and the serotonin transporter gene linked polymorphic region (5-HTTLPR). Behav. Genet. *38*, 34–43.

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

10. Uddin, M., de los Santos, R., Bakshis, E., Cheng, C., and Aiello, A.E. (2011). Building conditions, 5-HTTLPR genotype, and depressive symptoms in adolescent males and females. J. Adolesc. Health *49*, 379–385.

11. Hu, X.-Z., Lipsky, R.H., Zhu, G., Akhtar, L. a, Taubman, J., Greenberg, B.D., Xu, K., Arnold, P.D., Richter, M. a, Kennedy, J.L., et al. (2006). Serotonin transporter promoter gain-of-function genotypes are linked to obsessive-compulsive disorder. Am. J. Hum. Genet. *78*, 815–826.

12. Wendland, J.R., Martin, B.J., Kruse, M.R., Lesch, K.-P., and Murphy, D.L. (2006). Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR and rs25531. Mol. Psychiatry *11*, 224–226.

13. Zannas, A.S., McQuoid, D.R., Steffens, D.C., Chrousos, G.P., and Taylor, W.D. (2012). Stressful life events, perceived stress, and 12-month course of geriatric depression: direct effects and moderation by the 5-HTTLPR and COMT Val158Met polymorphisms. Stress *15*, 425–434.

14. Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W.J., Assendelft, W.J.J., et al. (2008). The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int. J. Methods Psychiatr. Res. *17*, 121–140.

15. Boomsma, D.I., de Geus, E.J.C., Vink, J.M., Stubbe, J.H., Distel, M. a, Hottenga, J.-J., Posthuma, D., van Beijsterveldt, T.C.E.M., Hudziak, J.J., Bartels, M., et al. (2006). Netherlands Twin Register: from twins to twin families. Twin Res. Hum. Genet. *9*, 849–857.

16. Boomsma, D.I., Beem, A.L., van den Berg, M., Dolan, C. V, Koopmans, J.R., Vink, J.M., de Geus, E.J., and Slagboom, P.E. (2000). Netherlands twin family study of anxious depression (NETSAD). Twin Res. *3*, 323–334.

17. Brugha, T., Bebbington, P., Tennant, C., and Hurry, J. (1985). The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. Psychol. Med. *15*, 189–194.

18. Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., and Trivedi, M.H. (1996). The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol. Med. *26*, 477–486.

19. Middeldorp, C.M., Cath, D.C., Vink, J.M., and Boomsma, D.I. (2005). Twin and genetic effects on life events. Twin Res. Hum. Genet. *8*, 224–231.
20. Middeldorp, C.M., de Geus, E.J.C., Willemsen, G., Hottenga, J.-J., Slagboom, P.E., and Boomsma, D.I. (2010). The serotonin transporter gene length polymorphism (5-HTTLPR) and life events: no evidence for an interaction effect on neuroticism and anxious depressive symptoms. Twin Res. Hum. Genet. *13*, 544–549.
21. Graaf, R. de, Bijl, R. V, Smit, F., Vollebergh, W.M., and Spijker, J. (2002). Risk factors for 12-month comorbidity of mood, anxiety, and substance use disorders: findings from the Netherlands Mental Health Survey and Incidence Study. Am. J. Psychiatry *159*, 620–629.
22. Hovens, J.G.F.M., Wiersma, J.E., Giltay, E.J., van Oppen, P., Spinhoven, P., Penninx, B.W.J.H., and Zitman, F.G. (2010). Childhood life events and childhood trauma in adult patients with depressive, anxiety and comorbid disorders vs. controls. Acta Psychiatr. Scand. *122*, 66–74.
23. Wiersma, J.E., Hovens, J.G.F.M., van Oppen, P., Giltay, E.J., van Schaik, D.J.F., Beekman, A.T.F., and Penninx, B.W.J.H. (2009). The importance of childhood trauma and childhood life events for chronicity of depression in adults. J. Clin. Psychiatry *70*, 983–989.
24. Hovens, J.G.F.M., Giltay, E.J., Wiersma, J.E., Spinhoven, P., Penninx, B.W.J.H., and Zitman, F.G. (2012). Impact of childhood life events and trauma on the course of depressive and anxiety disorders. Acta Psychiatr. Scand. *126*, 198–207.
25. Penninx, B.W.J.H., Nolen, W. a, Lamers, F., Zitman, F.G., Smit, J.H., Spinhoven, P., Cuijpers, P., de Jong, P.J., van Marwijk, H.W.J., van der Meer, K., et al. (2011). Two-year course of depressive and anxiety disorders: results from the Netherlands Study of Depression and Anxiety (NESDA). J. Affect. Disord. *133*, 76–85.
26. Boomsma, D.I., Willemsen, G., Sullivan, P.F., Heutink, P., Meijer, P., Sondervan, D., Kluft, C., Smit, G., Nolen, W. a, Zitman, F.G., et al. (2008). Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur. J. Hum. Genet. *16*, 335–342.
27. Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. Twin Res. *5*, 554–571.
28. Jaccard, J., and R, T. (2003). Interaction effects in multiple regression. In Sage University Paper Series On Quantitative Applications in the Social Sciences, (Thousand Oaks, CA: Sage),.
29. Knol, M.J., van der Tweel, I., Grobbee, D.E., Numans, M.E., and Geerlings, M.I.

(2007). Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int. J. Epidemiol. *36*, 1111–1118.
30. R Core Team (2015). R: A Language and Environment for Statistical Computing.
31. Vinkhuyzen, a a E., Dumenil, T., Ryan, L., Gordon, S.D., Henders, a K., Madden, P. a F., Heath, a C., Montgomery, G.W., Martin, N.G., and Wray, N.R. (2011). Identification of tag haplotypes for 5HTTLPR for different genome-wide SNP platforms. Mol. Psychiatry *16*, 1073–1075.
32. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.
33. Clarke, H., Flint, J., Attwood, a S., and Munafò, M.R. (2010). Association of the 5- HTTLPR genotype and unipolar depression: a meta-analysis. Psychol. Med. *40*, 1767–1778.
34. Nanni, V., Uher, R., and Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. Am. J. Psychiatry *169*, 141–151.
35. Myung, W., Lim, S.-W., Kim, J., Lee, Y., Song, J., Chang, K., and Kim, D.K. (2010). Serotonin transporter gene polymorphisms and chronic illness of depression. J. Korean Med. Sci. *25*, 1824–1827.
36. Uher, R., Caspi, A., Houts, R., Sugden, K., Williams, B., Poulton, R., and Moffitt, T.E. (2011). Serotonin transporter gene moderates childhood maltreatment's effects on persistent but not single-episode depression: replications and implications for resolving inconsistent results. J. Affect. Disord. *135*, 56–65.
37. Gauderman, W.J. (2002). Sample size requirements for matched case-control studies of gene-environment interaction. Stat. Med. *21*, 35–50.
38. Shiroma, P.R., Geda, Y.E., and Mrazek, D.A. (2010). Pharmacogenomic implications of variants of monoaminergic-related genes in geriatric psychiatry. Pharmacogenomics *11*, 1305–1330.
39. López-León, S., Choy, W., and Aulchenko, Y. (2009). Genetic factors influence the clustering of depression among individuals with lower socioeconomic status. PLoS One *4*, 1–5.
40. Middeldorp, C.M., Cath, D.C., Beem, a L., Willemsen, G., and Boomsma, D.I. (2008). Life events, anxious depression and personality: a prospective and genetic study. Psychol. Med. *38*, 1557–1565.

Chapter 3

Association between depression and lower educational attainment not explained by shared genetic effects in ~25,000 individuals

Wouter J. Peyrot, Sang H. Lee, Yuri Milaneschi, Abdel Abdellaoui, Enda M. Byrne, Tonu Esko, Eco J.C. de Geus, Gibran Hemani, Jouke-Jan Hottenga, Stefan Kloiber, Douglas F. Levinson, Susanne Lucae, Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Nick G. Martin, Sarah E. Medland, Andres Metspalu, Lili Milani, Markus M. Noethen, James B. Potash, Marcella Rietschel, Cornelius A. Rietveld, Stephan Ripke, Jianxin Shi, Social Science Genetic Association Consortium, Gonneke Willemsen, Zhihong Zhu, Dorret I. Boomsma, Naomi R. Wray, Brenda W.J.H. Penninx

Published in *Molecular Psychiatry 2015, 20:735-743*

Paper at publisher's website:
http://www.nature.com/mp/journal/v20/n6/full/mp201550a.html
doi:10.1038/mp.2015.50

ABSTRACT

An association between lower educational attainment (EA) and an increased risk for depression has been confirmed in various western countries. This study examines whether pleiotropic genetic effects contribute to this association. Therefore, data were analyzed from a total of 9,662 Major Depressive Disorder (MDD) cases and 14,949 controls (with no lifetime MDD diagnosis) from the Psychiatric Genomics Consortium with additional Dutch and Estonian data. The association of EA and MDD was assessed with logistic regression in 15,138 individuals indicating a significantly negative association in our sample with an odds ratio for MDD 0.78 [0.75-0.82] per standard deviation increase in EA. With data of 884,105 autosomal common SNPs, three methods were applied to test for pleiotropy between MDD and EA: (i) genetic profile risk scores (GPRS) derived from training data for EA (independent meta-analysis on 120,000 subjects) and MDD (using a ten-fold leave-one-out procedure in the current sample) (ii) bivariate Genomic-Relationship-Matrix Restricted Maximum Likelihood (GREML), and (iii) SNP effect concordance analysis (SECA). With these methods we found (i) that the EA-GPRS did not predict MDD status, and MDD-GPRS did not predict EA, (ii) a weak negative genetic correlation with bivariate GREML analyses, but this correlation was not consistently significant, (iii) no evidence for concordance of MDD and EA SNP effects with SECA analysis. To conclude, our study confirms an association of lower EA and MDD risk, but this association was not due to measurable pleiotropic genetic effects, which suggests that environmental factors could be involved such as, for example, socioeconomic status.

INTRODUCTION

An association between lower educational attainment (EA) and increased risk for Major Depressive Disorder (MDD) has been confirmed in various Western countries. A meta-analysis of 37 studies from mainly western countries found a 3 per cent decrease in log odds ratio for depression per additional year of education.[1] Research of the World Mental Health Survey Initiative also found that those with high educational levels are generally at lower risk for depression in high-income countries, although Japan showed an inverted association.[2] The international Consortium of Psychiatric Epidemiology found a negative correlation in the United States and the Netherlands,[3] which was confirmed in a recent study in the Netherlands.[4]

The association of lower EA and increased MDD risk could result from multiple, not necessarily independent, effects; including causal, environmental or pleiotropic genetic effects. Lower EA could lead to an increased MDD risk (social causation), for example via stress associated with lower socioeconomic status, or via less effective coping strategies or unhealthier lifestyles among those with lower EA.[5,6] However, lower EA could also be the result of MDD vulnerability, for example when the onset of MDD is at an early age before educational goals would have been achieved. Alternatively, a third factor could be in play impacting on both, such as personality characteristics or less developed cognitive abilities, causing lower EA and increased risk for MDD. Such a third factor could also consist of pleotropic genetic effects (or linkage disequilibrium between effective variants) resulting in genetic correlation (the part of the phenotypic correlation caused by shared additive genetic effects), because EA[7] and MDD[8–10] both have a confirmed genetic basis.

It is relevant to understand the mechanisms of the association between lower EA and MDD, because this can have important implications for prevention strategies of MDD and its consequences. When lower EA would increase MDD risk, the responsible mechanisms should be studied and subsequently addressed, for example by providing psycho-education about these mechanisms to those with lower EA. However, when shared genetic effects would link EA and MDD no responsible mechanisms can be addressed, and prevention would be restricted to general advice to prevent MDD.

The possible impact of pleiotropic genetic effects on lower EA and increased MDD risk has not received much study. We are aware of three such studies, of which two find a substantial negative genetic correlation between EA and cross-sectional measures of depressive symptoms obtained via self-report questionnaires.[11,12] One study used DSM-IV based diagnosis of MDD with a twin

design and generally supported the social causation model and found only a small genetic correlation.[5] To the best of our knowledge, no study combined DSM-IV based diagnosis and genome-wide SNP data to test for pleiotropic genetic effects between lower EA and MDD risk.

The current study was conducted to test for pleiotropic genetic effects between lower EA and MDD diagnoses in a large sample of ~25,000 subjects from the Psychiatric Genomics Consortium[13] with additional Estonian and Dutch data. We applied the following SNP-based methods: genetic profile risk score (GPRS) analyses, bivariate Genomic-Relationship-Matrix Restricted Maximum Likelihood (GREML) analysis, and SNP effect concordance analysis (SECA).

METHODS

Subjects

Genotype and phenotype data of ten cohort studies were combined: eight cohorts[14–21] included in the Psychiatric Genomics Consortium (PGC)[13] plus two additional cohorts. The first additional cohort was from the Netherlands and combined additional independent data from the Netherlands Study of Depression and Anxiety[22] and the Netherlands Twin Registry[23] (NESDA/NTR-2). The second additional cohort was a population-based sample from Estonia (EGCUT).[24] The numbers of cases and controls per cohort are displayed in Table 1.

MDD cases and controls

All cases (N=9,662) had a DSM-IV or ICD-10 based diagnosis of MDD in lifetime according to a structured diagnostic instrument. Most controls (N=14,949) were randomly selected from the population and screened for a lifetime history of MDD. A more detailed description of the PGC-cohorts was given previously[13] and is summarized in Table S1. For the NESDA/NTR-2 cohort, MDD-cases were diagnosed with the DSM-IV based CIDI interview (CIDI, version 2.1), and controls scored low on various mental health screening questionnaires (NTR)[25] or had no diagnosis of a psychiatric disorder in their lifetime (NESDA). For the EGCUT cohort, MDD-cases were identified using International Classification of Diseases (ICD-10) codes F32 (*depressive disorder*) and/or F33 (*recurrent depressive disorder*), and MDD-controls excluded all subjects with a lifetime ICD10 psychiatric diagnosis (category F).[24]

Educational attainment

Educational attainment (EA) was assessed in seven of the ten contributing cohorts (EGCUT, GenRED, GSK, NESDA/NTR-1, NESDA/NTR-2, QIMR, and STAR-D).

For NESDA/NTR-1 and NESDA/NTR-2, EA was defined as the years of education required for the highest diploma attained following the Dutch educational system. For QIMR and EGCUT, EA was defined as the US years of education required for the highest diploma attained following the international ISCED classification.[7] For GSK, EA was defined as the number of years that school was attended. For STAR*D, EA was expressed in years of education. For GenRED, EA was assessed in controls only as the highest diploma attained and ranged from 1 to 5 labeling the following educational levels: lower than high school (1), high school (2), some college (3), bachelor degree (4), higher than bachelor degree (5).

The EA measure was corrected per cohort for year of birth and sex, in line with the recent meta-analysis from the Social Science Genetic Association Consortium.[7] Thereby, the standardized residuals were obtained after regression of EA on sex, year of birth (YOB), YOB^2, YOB^3, and the interaction of sex with YOB, YOB^2, and YOB^3. For STAR*D and GSK, YOB was not available and substituted with age. In all cohorts, EA was defined in individuals over 25 years of age only, so that they had time to achieve their educational potential. The distribution of EA z-scores is displayed in Figure S1.

Genotyping, quality control, and imputation

Genotyping, quality control, and imputation were performed in line with previous publications and are described in detail in the Supplemental Materials. In short, quality controlled SNPs with a MAF > 0.01 from the HapMap3 reference panel[26] were imputed and yielded information on 884,105. With these SNPs the Genomic-Relationship-Matrix was estimated and unrelated subjects selected (with maximum pairwise genetic relationships 0.05, which is approximately equivalent to second cousins), using the GCTA software.[27] All of the subsequent genetic analyses were corrected for possible confounding cohort- and genotyping effects by including a categorical covariate labeling the ten cohorts, and within cohorts the different genotyping batches, where applicable (i.e. three batches within NESDA/NTR-2, two batches within EGCUT, and two batches within QIMR). Ancestry-informative principal components were based on the Genomic-Relationship-Matrix and estimated with the GCTA software.[27]

Genetic Profile Risk Scores (GPRS)

Preparation of the genetic profile risk scores based on EA discovery results (EA-GPRS) and MDD discovery results (MDD-GPRS) is described in detail in the Supplemental Materials. In short, the procedure from Purcell et al[28] implemented in Plink[29] was applied. The independent EA discovery results were from the recent

meta-analyses on US years of schooling from the Social Science Genetics Association Consortium (SSGAC)[7] containing around 120,000 subjects. EA-GPRS analyses were not conducted for BMH, GenRED, and STAR*D, because no independent discovery results were available. To obtain the MDD discovery results was slightly more elaborate, because no large MDD cohort exists that is independent from PGC. Therefore, a ten-fold leave-one-cohort-out approach was followed, and the discovery results were thus based on around 8,000 cases and 12,000 controls.

The GPRS were based on the same set of independent SNPs. First, the SNPs were selected with results available for all of the discovery sets. Second, this set of SNPs was pruned to a set of 76,516 independent SNPs with a maximum pairwise r^2 of 0.25 based on a sliding window of 200 SNPs with steps of 5 SNPs.[29] The EA-GPRS and MDD-GPRS were then estimated based on all SNPs up to p-value thresholds (P_T) in the respective discovery results of 0.001, 0.01, 0.1, and 1 respectively. Consequently, all GPRS with $P_T = 1$ were based on the exact same SNPs, but GPRS with different P_T were based on different sets of SNPs depending on the respective discovery results (see Table S2). The GPRS were standardized per cohort to a mean of 0 and standard deviation of 1 to aid interpretability of results.

Statistical analyses

The association of EA to MDD risk (phenotypic correlation) was assessed with logistic regression within EGCUT, GSK, NESDA/NTR-1 and 2, QIMR, and STAR*D separately, and in the combined sample correcting for covariates labeling the cohorts.

Genetic Profile Risk Score analyses

In the first method to test for pleiotropic genetic effects we estimated the *across-trait* effects of EA-GPRS on MDD and, vice versa, the effects of MDD-GPRS on EA. For comparison, we also estimated the *within-trait* effects of EA-GPRS on EA and MDD-GPRS on MDD. The effects of GPRS on EA and MDD were assessed with linear and logistic regression respectively. For the full sample, the effects were assessed for the GPRS based on P_T of 0.001, 0.01, 0.1, and 1; for the individual cohorts, the effects were only assessed for the GPRS based on $P_T = 1$.

The proportions of variation explained in EA and MDD were estimated as additional measures of the impact of GPRS. For EA, this proportion was derived as the R^2 of the linear regression model including the covariates and the polygenic risk score, minus the R^2 of the model including the covariates only. For MDD,

Nagelkerke's pseudo R^2 were derived and corrected for the covariates by substituting the null (or intercept) model in Nagelkerke's equation for the model including the covariates (adjusted equation in Supplemental Materials). Lee at al indicated that Nagelkerke's pseudo R^2 can be biased by ascertainment, when the proportion of cases in the study sample differs from the population disease frequency.[30] Therefore, they proposed an R^2 measure that is robust against ascertainment bias and interpretable on the liability scale. This liability R^2 was obtained by rescaling Nagelkerke's R^2 for an MDD population prevalence of K=0.2 (see Supplemental Materials).[30]

Bivariate Genomic-Relationship-Matrix Restricted Maximum Likelihood (GREML)

The GREML mixed linear model method was used (i) to assess the proportion of variation in EA and MDD explained by genome-wide common SNPs (SNP-h^2) and (ii) to assess the pleiotropic genetic effects between MDD and EA (genetic correlation), as implemented in GCTA.[27,31,32] The MDD SNP-h^2 was expressed on the liability scale for a population prevalence of K=0.2 by converting the SNP-h^2 on the observed scale (controls 0; cases 1) with equation (23) from Lee et al.[33] Bivariate GREML estimates of the genetic correlation are approximately the same on the liability scale as on the observed scale,[32] which implies that (i) its value does not depend on population disease prevalence K and (ii) that the genetic correlation between the binary MDD status and continuous EA measure could be estimated. The genetic correlation was, first, estimated with EA information from both cases and controls. This estimate could, however, potentially be confounded by case ascertainment (which may not be education independent). Therefore, the genetic correlation was estimated a second time with EA information from controls only and MDD status from both cases and controls. The GPRS- and GREML-analyses were corrected for sex, the first 10 (GPRS) or 20 (GREML) principal components and covariates labeling the cohorts and genotype batches. The necessity to correct for the principal components is indicated by a significant correlation between some of the GPRS with some of the principal components (Table S3).

SNP effect concordance analysis (SECA)

In SNP effect concordance analysis (SECA; http://neurogenetics.qimrberghofer.edu.au/SECA)[34] association results are analyzed, rather than individual genotyped data, to test for concordance between two traits with respect to the SNP effects significance as well as their directions. We applied SECA on the EA meta-analyses results from the Social Science

Genetics Association Consortium (SSGAC)[7] and MDD association results on our own sample.

RESULTS

The overall sample consisted of 9,662 patients with MDD in lifetime and 14,949 controls with a mean age of 46.2 (SD 15.6) and 59.4% female; information on EA was available for 5,373 cases and 9,765 controls (Table 1). In all cohorts with EA information available for both cases and controls, the phenotypic associations between EA and MDD was negative, with an overall odds ratio of 0.78 (95%CI: 0.75-0.82, p=2.2e-31) per standard deviation increase in EA (Figure 1). This negative association was consistent for MDD cases with known age of onset > 30. The strongest association was found in GSK with an OR of 0.45 (95%CI: 0.40-0.50). When GSK was left out of the analyses, the overall association remained significant with an OR of 0.88 (95%CI: 0.84-0.92). The association was comparable in male and female (Figure S2).

GPRS analyses

The GPRS had *within-trait* predictive effects as expected. The MDD-GPRS predicted MDD with most predictive power for the polygenic risk score including all SNPs (P_T=1), with an odds ratio of 1.13 (p=1.7e-16) and an R^2 of 0.4% on the liability scale (Table 2A). The EA-GPRS predicted EA also in the expected direction, again with most predictive power for GPRS including all SNPs, with a beta of 0.11 (p=2.7e-37) and an R^2 of 1.2% (Table 2A). However, we found no significant *across-trait* prediction: the MDD-GPRS did not predict EA (beta=-0.01 p=6.7e-2) and the EA-GPRS did not predict MDD (OR=0.99 p=5.9e-1, Table 2A). Secondary analyses, performed within all cohorts separately, indicated that the within-trait predictive effects were consistent in all cohorts, and that the lack of across-trait predictive power was also consistent for all cohorts (Table 2B). In addition, no correlation was found between the MDD-GPRS and the EA-GPRS themselves (P_T=1; correlation coefficient of 0.006, p=0.413). In additional analyses, across-trait predictive effects on MDD were tested for GPRS based on the SSGAC EA outcome tagging College completion (College-GPRS).[7] College completion distinguishes more in the extreme end of the EA distribution, and has a confirmed genetic basis.[7] However, no predictive effects of the College-GPRS on MDD were found (OR=0.99, p=0.74 for P_T=1; Table S4).

Table 1. Sample characteristics

	N with MDD		Age			N with EA		Mean EA z-score	
Study (Abbreviation)	**Case**	**Control**	**Mean**	**SD**	**% female**	**Case**	**Control**	**Case**	**Control**
Bonn/Mannheim (B)	925	1282	46.9	13.3	55.7	0	0	-	-
EGCUT (E)	508	5345	48.8	20.1	53.2	446	4569	-0.06	0.01
GenRED (GE)	976	1215	47.0	16.2	56.1	0	1170	-	0.00
GSK (GS)	866	863	51.9	13.5	66.9	866	862	-0.36	0.36
MPIP (M)	337	533	48.1	13.9	54.5	0	0	-	-
NESDA/NTR-1 (N1)	1560	1123	44.5	13.2	64.3	1382	875	-0.08	0.12
NESDA/NTR-2 (N2)	236	1201	40.5	14.7	62.4	211	759	-0.28	0.09
QIMR (Q)	1432	1686	43.4	10.9	61.0	1258	1402	-0.02	0.01
RADIANT (R)	1605	1573	44.3	12.5	66.6	0	0	-	-
STAR*D (S)	1217	128	43.5	14.0	56.9	1210	128	-0.03	0.30
Overall	9662	14949	46.2	15.6	59.4	5373	9765	-0.10	0.06

The number of cases with a diagnosis of MDD in lifetime, controls without MDD, mean age and its standard deviation, and the percentage of female is displayed for the ten cohorts separately and for the overall sample. In addition, the number of cases and controls with information on EA available and their mean EA z-score are displayed. The EA z-scores were defined as the standardized residuals of the regression of EA on sex, year of birth (YOB), YOB^2, YOB^3, and the interaction of sex with YOB, YOB^2, and YOB^3. Abbreviations: EA, educational attainment; EGCUT, Estonian Genome Center of the University of Tartu; GEnRED, Genetics of Recurrent Early-Onset Depression; GSK, GlaxoSmithKline; MDD, major depressive disorder; MPIP, Max Planck Institute of Psychiatry; NESDA, Netherlands Study of Depression and Anxiety; NTR, Netherlands Twin Registry; QIMR, Queensland Institute of Medical Research; STAR*D, Sequenced Treatment Alternatives to Relieve Depression; YOB, year of birth.

Table 2. The effect of genomic risk profile scores (GRPS), based on depression (MDD-GRPS) and educational attainment (EA-GRPS) discovery results, on MDD and EA in the overall sample and the separate studies

	Effect on MDD									
	N		Effect		R-squared (%)		Effect on EA			
	Case	Control	OR	P-value	NK	Liability	N	Beta	P-value	R2(%)
A. Results in overall sample										
Effect of MDD-GRPS										
p < 0.001	9662	14949	1.03	0.021	0.03	0.03	15985	-0.01	0.188	0.01
p < 0.01	9662	14949	1.04	0.010	0.04	0.04	15985	-0.02	0.034	0.03
p < 0.1	9662	14949	1.11	1.0e-11	0.28	0.29	15985	-0.02	0.046	0.02
p < 1	9662	14949	1.13	1.7e-16	0.41	0.42	15985	-0.01	0.067	0.02
Effect of EA-GRPS										
p < 0.001	6544	12324	1.02	0.167	0.02	0.02	13477	0.05	1.2e-09	0.27
p < 0.01	6544	12324	1.00	0.842	0.00	0.00	13477	0.09	2.4e-23	0.73
p < 0.1	6544	12324	0.98	0.245	0.01	0.01	13477	0.10	2.1e-31	1.00
p < 1	6544	12324	0.99	0.594	0.00	0.00	13477	0.11	2.7e-37	1.20
B. Results in separate studies										
Effect of MDD-GRPS (all threshold p<1; letters represent separate studies)										
B	925	1282	1.14	0.003	0.54	0.54	-	-	-	-
E	508	5345	1.10	0.040	0.16	0.30	5015	-0.01	0.402	0.01
GE	976	1215	1.07	0.111	0.16	0.16	1170	-0.01	0.647	0.02
GS	866	863	1.09	0.097	0.22	0.21	1728	-0.04	0.111	0.14
M	337	533	1.05	0.461	0.08	0.09	-	-	-	-
N1	1560	1123	1.24	6.1e-08	1.50	1.50	2257	-0.02	0.373	0.04

Table 2. (continued)

	N		Effect		R-squared (%)		Effect on EA			
	Case	Control	OR	P-value	NK	Liability	N	Beta	P-value	R2(%)
N2	236	1201	1.19	0.019	0.70	0.98	1817	-0.02	0.330	0.05
Q	1432	1686	1.09	0.017	0.25	0.24	2660	-0.01	0.776	0.00
R	1605	1573	1.15	9.7e-05	0.65	0.63	-	-	-	-
S	1217	128	1.17	0.098	0.45	0.79	1338	0.03	0.345	0.07
Effect of EA-GRPS (all threshold p<1; letters represent separate studies)										
E	508	5345	0.95	0.257	0.05	0.09	5015	0.08	1.2e-08	0.64
GS	866	863	0.96	0.358	0.07	0.06	1728	0.07	0.004	0.48
M	337	533	0.96	0.524	0.06	0.06	-	-	-	-
N1	1560	1123	0.94	0.151	0.10	0.10	2257	0.17	1.1e-15	2.79
N2	236	1201	0.95	0.512	0.05	0.08	1817	0.16	1.0e-11	2.52
Q	1432	1686	1.00	0.994	0.00	0.00	2660	0.12	1.7e-09	1.36
R	1605	1573	1.04	0.241	0.06	0.06	-	-	-	-

The impact of the genetic profile risk scores (GPRS), based on EA discovery results (EA-GPRS) and MDD discovery results (MDD-GPRS), on target MDD and on target EA were estimated with respectively logistic and linear regression, while including sex, the first 10 principal components and covariates labeling the cohorts and genotype batches. The impact on MDD was, in addition, estimated as Nagelkerke's R^2 and the R^2 on the liability scale; the impact on EA as the standard R^2 of linear regression. On the overall sample, the effects were estimated for GPRS on different sets of SNPs with different thresholds of significance in the discovery set (p<0.001; p<0.01; p<0.1; p<1) (Panel A). The impact on MDD and EA in the separate cohorts was, subsequently, estimated for the polygenic risk scores based on all SNPs (threshold p<1) (Panel B). The number of individuals included in the analyses is displayed: note that individuals from B, GE, and S are excluded from the analyses with polygenic risk scores based on EA discovery results, because these cohorts were (partly) included in the discovery phase. B=Bonn/Mannheim; E=EGCUT; GE=GenRED; GS=GSK; M=MPIP; N1=NESDA/NTR-1; N2=NESDA/NTR-2; Q=QIMR; R=RADIANT; S=STAR*D

GREML analyses

GREML analyses in the overall study sample generated an estimate of MDD SNP-h^2 of 0.173 (SE=0.017, p<1e-16) on the liability scale (K=0.2); this finding was not solely driven by one of the individual cohorts, because the MDD SNP-h^2 was estimated at consistent values when one cohort was left out at the time (Table 3). The MDD SNP-h^2 was larger when expressed on the liability scale (0.173) than on the observed scale (0.126), with a larger SNP-h^2 for larger values of disease frequency (as expected from equation (23) from Lee et al[33]; Table S5). The EA SNP-h^2 was estimated at 0.124 (SE=0.019, p=2.8e-11) when EA information in both cases and controls was taken into account (Table 3A), and at 0.144 (SE=0.030, p=1.5e-6) when EA information of controls only was utilized (Table 3B). Again, these estimates were not solely driven by one of the individual cohorts (Table 3). The genetic correlation between MDD and EA was estimated at -0.253 (SE=0.087, p=0.004) when EA information of both cases and controls was taken into account (Table 3A). Since a correlation between genetic and environmental factors is likely to be partitioned into the genetic variance and covariance components, we explored the robustness of this estimate by limiting EA to be measured only in controls. When taking into account EA of controls only and MDD status from cases and controls, the genetic correlation dropped considerably and was no longer significantly different from 0 with an estimate of -0.110 (SE=0.105, p=0.298; Table 3B). In post-hoc analyses we tested if EA moderated the polygenic effects on MDD, but found no such evidence with neither GPRS- nor GREML-analyses (Supplemental Materials).

SNP effect concordance analysis

SECA showed no evidence for genetic correlation. The primary SECA test divided the SNPs in 144 subsets based on significance of association with MDD and EA smaller than respectively 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0. Not a single of these subsets contained a larger number of SNPs than expected by chance, i.e. no concordance was found with respect to the MDD and EA SNP effect *significances*. When comparing the directions of SNP effects, only four of the 144 subsets showed nominally correlated directions of effect, which is not more than expected by chance (permuted empirical p-value 0.244), indicating no concordance with respect to the MDD and EA SNP effect *directions*.

DISCUSSION

This study tested the existence of pleiotropic genetic effects (genetic correlation) between major depressive disorder (MDD) and lower educational attainment (EA)

on individual genotype data from a large sample of approximately 25,000 subjects from Western countries. To start, a strong negative phenotypic association was found with an OR for MDD of 0.78 per SD increase in EA, which is in line with findings from a meta-analysis of 37 studies from mainly western countries by Lorant et al.[1] Our first test for genetic correlation was negative with no across- trait predictive power of the GPRS: EA-GPRS did not predict MDD, and MDD-GPRS did not predict EA. In the second test for genetic correlation, GREML analyses did not show consistent evidence for genetic correlation. The third test, SNP effect concordance analysis (SECA), also showed no evidence for concordance of EA and MDD SNP effects with respect to their significance or direction.

The GPRS in our study had within-trait predictive power in line with previous findings,[7,13] and were based on an independent EA discovery sample from the SSGAC[7] of approximately 120,000 subjects and independent MDD leave-one-cohort-out discovery samples of approximately 8,000 cases and 12,000 controls. These numbers seem adequate, but the discovery sets would ideally have been even larger, because most predictive power was still found for the GPRS including all SNPs (P_T=1) indicating that true effect SNPs were associated in the discovery sample with p-values close to 1.[28] Nevertheless, Dudbridge power calculations suggested that the EA-GPRS were well powered to predict MDD when the genetic correlation would have been around -0.2 (Figure S3).[35] Our GPRS results, therefore, indicate that a large genetic correlation between EA and MDD is unlikely, but could not exclude a small genetic correlation of around -0.1.

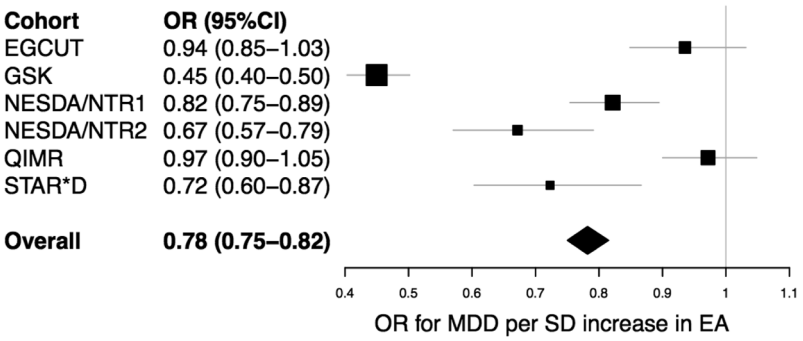


Figure 1. Forest plot of the phenotypic association between EA and MDD. The OR for MDD per SD increase in EA is displayed for the individual cohorts, as well as for the overall sample. The ORs were estimated with logistic regression of MDD on the corrected EA z-scores, which were defined as the standardized residuals of the regression of EA on sex, year of birth (YOB), YOB^2, YOB^3 and their interaction with sex.

Table 3. GREML estimates of the proportion of variation explained by common SNPs in major depressive disorder (MDD SNP-h^2) and educational attainment (EA SNP-h^2), and the genetic correlation between MDD and EA; estimated with EA information from both cases and controls (upper rows) and from controls only (lower rows).

	MDD SNP-h^2					EA SNP-h^2				Genetic correlation		
Cohorts included	**Case**	**Control**	**Est**	**se**	**p-value**	**N**	**Est**	**se**	**p-value**	**Est**	**se**	**p-value**
A. EA in both cases and controls												
All	9662	14949	0.173	0.017	<1e-16	15985	0.124	0.019	2.8e-11	-0.253	0.087	0.004
All excluding B	8737	13667	0.155	0.019	1.1e-16	15985	0.124	0.019	3.1e-11	-0.249	0.096	0.009
All excluding E	9154	9604	0.207	0.022	<1e-16	10970	0.133	0.026	4.5e-07	-0.265	0.102	0.009
All excluding GE	8686	13734	0.183	0.019	<1e-16	14815	0.130	0.020	9.0e-11	-0.345	0.091	1.4e-04
All excluding GS	8796	14086	0.172	0.019	<1e-16	14257	0.120	0.021	7.6e-09	-0.231	0.099	0.020
All excluding M	9325	14416	0.177	0.018	<1e-16	15985	0.124	0.019	3.9e-11	-0.208	0.088	0.018
All excluding N1	8102	13826	0.150	0.020	2.1e-14	13728	0.120	0.022	3.0e-08	-0.239	0.109	0.028
All excluding N2	9426	13748	0.174	0.018	<1e-16	14168	0.118	0.021	1.4e-08	-0.294	0.097	0.002
All excluding Q	8230	13263	0.199	0.020	<1e-16	13325	0.133	0.022	1.6e-09	-0.237	0.091	0.009
All excluding R	8057	13376	0.161	0.020	1.0e-15	15985	0.124	0.019	3.6e-11	-0.206	0.097	0.033
All excluding S	8445	14821	0.187	0.019	<1e-16	14647	0.126	0.020	6.3e-10	-0.294	0.092	0.001

Table 3. (continued)

	MDD SNP-h^2					EA SNP-h^2				Genetic correlation		
Cohorts included	**Case**	**Control**	**Est**	**se**	**p-value**	**N**	**Est**	**se**	**p-value**	**Est**	**se**	**p-value**
B. EA in controls only												
All	9662	14949	0.173	0.017	<1e-16	9765	0.144	0.030	1.5e-06	-0.110	0.105	0.298
All excluding B	8737	13667	0.156	0.019	1.1e-16	9765	0.144	0.030	1.6e-06	-0.113	0.115	0.327
All excluding E	9154	9604	0.208	0.022	<1e-16	5196	0.177	0.054	0.001	0.004	0.130	0.972
All excluding GE	8686	13734	0.183	0.019	<1e-16	8595	0.159	0.034	2.6e-06	-0.187	0.110	0.087
All excluding GS	8796	14086	0.172	0.019	<1e-16	8903	0.133	0.033	4.4e-05	-0.120	0.119	0.310
All excluding M	9325	14416	0.178	0.018	<1e-16	9765	0.144	0.030	1.4e-06	-0.065	0.105	0.537
All excluding N1	8102	13826	0.151	0.020	1.7e-14	8890	0.144	0.033	9.4e-06	-0.112	0.124	0.369
All excluding N2	9426	13748	0.174	0.018	<1e-16	9006	0.132	0.032	4.0e-05	-0.176	0.117	0.133
All excluding Q	8230	13263	0.200	0.020	<1e-16	8363	0.143	0.035	3.6e-05	-0.148	0.114	0.196
All excluding R	8057	13376	0.161	0.020	1.0e-15	9765	0.144	0.030	1.4e-06	-0.059	0.116	0.610
All excluding S	8445	14821	0.188	0.019	<1e-16	9637	0.140	0.030	3.3e-06	-0.084	0.107	0.434

Bivariate- GREML was performed to estimate the proportion of variation explained in MDD (MDD SNP-h^2) and EA (EA SNP-h^2) by genome-wide common SNPs, as well as the genetic correlation between MDD and EA. First, EA information of both cases and controls was analysed (first line upper Panel), but the correlation thus found should be interpreted with caution, because it could be biased (see Methods). Therefore, analyses were repeated taking EA information of controls only into account (first line lower Panel). In order to estimate the robustness of these findings, the analyses were repeated leaving one cohort at the time (additional lines in upper and lower Panel). The MDD SNP-h^2 was expressed on the liability scale assuming a population prevalence K of 20% (see Table S5 for comparison of the MDD SNP-h^2 for different values of K with the SNP-h^2 on the observed scale). MDD was available for all ten cohorts; EA was only available for cohorts E, GE, GS, N1, N2, Q, and S. The analyses were corrected for sex, the first 20 principal components and a categorical covariate labelling the cohorts and genotype batches. B=Bonn/Mannheim; E=EGCUT; GE=GenRED; GS=GSK; M=MPIP; N1=NESDA/NTR-1; N2=NESDA/NTR-2; Q=QIMR; R=RADIANT; S=STAR*D

We performed GREML analyses to estimate the MDD SNP-h^2, EA SNP-h^2 and genetic correlation. The MDD SNP-h^2 found (0.17) was considerably smaller than the one previously found by Lubke et al (0.32),[10] which could well be due to the actual differences in SNP-h^2 across cohorts; the sample of Lubke was included in the current study as NESDA/NTR-1 and indeed had the largest contribution to the overall SNP-h^2 of all cohorts (Table 3). The EA SNP-h^2 (0.14 in controls only) was of the same magnitude (less than 2 SE difference) as the SNP-h^2 found by Rietveld et al (0.2).[7] The GREML estimate of the genetic correlation was somewhat complicated to interpret. A significant negative genetic correlation was found (-0.25, p=0.004) when EA information of both cases and controls was taken into account, but we fear this finding could be biased particularly in the context of genotype and environment correlation. In fact, when taking EA information of only controls into account, the estimate of genetic correlation dropped considerably and was no longer significant (-0.11, p=0.30). However, we note that this estimate was conservative as it reduced variation in EA, and we note the negative point estimate and high standard error showing that this analyses was underpowered to draw definitive inference. Taken all together, the GREML analyses could be in line with a small genetic correlation of around -0.1. In addition to the two methods based on individual level genotype data, we also performed analyses on GWAS summary statistics with the recently published SECA method[34] and found no evidence for genetic correlation.

To the best of our knowledge only three previous studies tested for a genetic correlation between MDD and EA. López-León et al used a family based approach in 2,383 subjects to find a negative genetic correlation of -0.65 and -0.50 between EA and self-reports of depressive symptoms based on respectively the Center for Epidemiologic Studies Depression Scale (CES-D) and the Hospital Anxiety and Depression Scale (HADS-D).[11] Boardman et al also used cross-sectional CES-D assessments and found a genetic correlation of -0.7 with GREML-analyses.[12] Mezuk et al used a twin design with depression assessed with the DSM-IV based Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders (SCID-I), and their study generally supported social causation as cause for the link between lower EA and increased MDD risk, and found only a small genetic correlation of -0.22.[5] The studies of López-León et al and Boardman et al contrast our finding of no, or at most a small, genetic correlation, but this could be because they tested symptom reports of depressive state at a specific point in time, whereas our study tested the presence of a more clinical construct: DSM-IV or ICD-10 based lifetime diagnosis of MDD. Indeed, our results appear in line with the findings from Mezuk et al who also used DSM-IV based diagnoses of

MDD. Furthermore, we found that the association between lower EA and MDD remained when cases with an age of first MDD onset > 30 were taken into account exclusively. This indicates it is unlikely that MDD directly causes a lowering of EA, as it can be assumed that one reaches his or her education potential before the age of 30, which is in line with the suggested social causation by Mezuk et al.[5]

The finding that there is no, or at most a small, genetic correlation between lower EA and MDD is relevant, because this implies that non-genetic factors play an important role, and that underlying mechanisms may possibly be accessible to interventions. For example, when the social causation model would be studied in more detail, this could potentially lead to underlying clues on how lower socioeconomic status could contribute to vulnerability for MDD, or alternatively how higher socioeconomic status may buffer against vulnerability for MDD. For instance, lower socioeconomic status has shown to be associated to less healthy life styles (less physical exercise, more smoking, higher BMI, and more use of alcohol),[36,37] less adequate medical treatment seeking behavior,[38] less knowledge about MDD,[39] and higher vulnerability to experience stressful life events.[40] These factors could all contribute to increased MDD risk. However, future research should be conducted to elucidate the most important underlying mechanism as these may hint to either public or personal actions to best prevent MDD amongst individuals with lower EA. Yet another mechanism underlying the link between lower EA and MDD could possibly be found in a third factor other than genetic effects, such as a certain personality characteristic or less developed cognitive abilities, that causes both lower EA and increased MDD risk.

Our study has several strengths, but also some limitations. First, our study is one of the first and largest studies to test for pleiotropy between lower EA and MDD, and we used individual level genotype data. In addition, we used clinically relevant DSM-IV and ICD-10 based diagnoses of MDD. Furthermore, we applied three distinct methods that essentially lead to the same conclusion. A limitation of our study is that the discovery samples of the polygenic risk score analyses were not optimally sized with maximum predictive power of the GPRS including all SNPs (P_T=1). However, this is a limitation of most current genetic studies, and we feel our discovery samples were adequately powered given the availability of relevant genetic cohorts up to date. Furthermore, the genetic basis of MDD is strong enough to study pleiotropy, as has been indicated in previous work from the Psychiatric Genomics Consortium that indicate a genetic correlation between MDD-schizophrenia (0.43±0.06), and MDD-bipolar disorder (0.47±0.06) with both GREML-[41] and GPRS-analyses.[42] Another limitation is that we could have missed

pleiotropic effects amongst rare SNPs with a MAF < 0.01. This limitation could be addressed with a family or twin study, but it would be surprising when SNPs with MAF < 0.01 would have large pleiotropic effects while SNPs with MAF > 0.01 show no such evidence.

To conclude, we did confirm a negative phenotypic association between MDD and EA, but found no evidence that this association is due to genetic factors, which indicates that a large genetic correlation between lower EA and MDD is unlikely, but does not exclude a small genetic correlation of around -0.1. Understanding of the possible pathways between lower EA and MDD risk requires further research including twin analyses for an additional estimate of the upper bound of the genetic correlation. Nevertheless, we believe that the finding of the absence of large pleiotropic genetic effects underlying the established correlation of lower EA with increased MDD risk may be relevant, as it points to non-genetic mechanisms that may be accessible to interventions aimed at breaking this deleterious link.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

APPENDIX 1

The following people are not listed as co-authors on this manuscript, but did contributed to the Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium,[13] from which individual level genotype and phenotype data was used. We are grateful for their contribution to the Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. The views presented in the present paper may not reflect the opinions of the individuals listed below. We thank: Lewis CM, Hamilton SP, Weissman MM, Breen G, Blackwood DH, Cichon S, Heath AC, Holsboer F, Madden PA, McGuffin P, Muglia P, Pergadia ML, Lin D, Müller-Myhsok B, Steinberg S, Grabe HJ, Lichtenstein P, Magnusson P, Perlis RH, Preisig M, Smoller JW, Stefansson K, Uher R, Kutalik Z, Tansey KE, Teumer A, Viktorin A, Barnes MR, Bettecken T, Binder EB, Breuer R, Castro VM, Churchill SE, Coryell WH, Craddock N, Craig IW, Czamara D, Degenhardt F, Farmer AE, Fava M, Frank J, Gainer VS, Gallagher PJ, Gordon SD, Goryachev S, Gross M, Guipponi M, Henders AK, Herms S, Hickie IB, Hoefels S, Hoogendijk W, Iosifescu DV, Ising M, Jones I, Jones L, Jung-Ying T, Knowles JA, Kohane IS, Kohli MA, Korszun A, Landen M, Lawson WB, Lewis G, Macintyre D, Maier W, Mattheisen M, McGrath PJ, McIntosh A, McLean A, Middeldorp CM, Middleton L, Montgomery GM, Murphy SN, Nauck M, Nolen WA, Nyholt DR, O'Donovan M, Oskarsson H, Pedersen N,

Scheftner WA, Schulz A, Schulze TG, Shyn SI, Sigurdsson E, Slager SL, Smit JH, Stefansson H, Steffens M, Thorgeirsson T, Tozzi F, Treutlein J, Uhr M, van den Oord EJ, Van Grootheest G, Völzke H, Weilburg JB, Willemsen G, Zitman FG, Neale B, Daly M, Sullivan PF.

APPENDIX 2

The following people who are not listed as co-authors on this manuscript contributed to the original GWAS meta-analysis on educational attainment,[7] on which the present paper is based. Data access has been granted under section 4 of the Data Sharing Agreement of the Social Science Genetic Association Consortium (SSGAC). The views presented in the present paper may not reflect the opinions of the individuals listed below. The SSGAC is grateful to the authors of Rietveld et al[7] for providing the meta-analysis data. We thank: Arpana Agrawal, Eva Albrecht, Behrooz Z. Alizadeh, Jüri Allik, Najaf Amin, John R. Attia, Stefania Bandinelli, John Barnard, François Bastardot, Sebastian E. Baumeister, Jonathan Beauchamp, Daniel J. Benjamin, Kelly S. Benke, David A. Bennett, Klaus Berger, Lawrence F. Bielak, Laura J. Bierut, Jeffrey A. Boatman, Patricia A. Boyle, Ute Bültmann, Harry Campbell, David Cesarini, Christopher F. Chabris, Lynn Cherkas, Mina K. Chung, Dalton Conley, Francesco Cucca, George Davey-Smith, Gail Davies, Mariza de Andrade, Philip L. De Jager, Christiaan de Leeuw, Jan-Emmanuel De Neve, Ian J. Deary, George V. Dedoussis, Panos Deloukas, Jaime Derringer, Maria Dimitriou, Gudny Eiriksdottir, Niina Eklund, Martin F. Elderson, Johan G. Eriksson, Daniel S. Evans, David M. Evans, Jessica D. Faul, Rudolf Fehrmann, Luigi Ferrucci, Krista Fischer, Lude Franke, Melissa E. Garcia, Christian Gieger, Håkon K. Gjessing, Patrick J.F. Groenen, Henrik Grönberg, Vilmundur Gudnason, Sara Hägg, Per Hall, Jennifer R. Harris, Juliette M. Harris, Tamara B. Harris, Nicholas D. Hastie, Caroline Hayward, Andrew C. Heath, Dena G. Hernandez, Wolgang Hoffmann, Adriaan Hofman, Albert Hofman, Rolf Holle, Elizabeth G. Holliday, Christina Holzapfel, William G. Iacono, Carla A. Ibrahim-Verbaas, Thomas Illig, Erik Ingelsson, Bo Jacobsson, Marjo-Riitta Järvelin, Min A. Jhun, Magnus Johannesson, Peter K. Joshi, Astanand Jugessur, Marika Kaakinen, Mika Kähönen, Stavroula Kanoni, Jaakkko Kaprio, Sharon L.R. Kardia, Juha Karjalainen, Robert M. Kirkpatrick, Philipp D. Koellinger, Ivana Kolcic, Matthew Kowgier, Kati Kristiansson, Robert F. Krueger, Zóltan Kutalik, Jari Lahti, David Laibson, Antti Latvala, Lenore J. Launer, Debbie A. Lawlor, Terho Lethimäki, Jingmei Li, Paul Lichtenstein, Peter K. Lichtner, David C. Liewald, Peng Lin, Penelope A. Lind, Yongmei Liu, Kurt Lohman, Marisa Loitfelder, Pamela A. Madden, Patrick K.E. Magnusson, Tomi E. Mäkinen, Pedro Marques Vidal, Nicolas W. Martin, Marco Masala, Matt McGue, George

McMahon, Osorio Meirelles, Michelle N. Meyer, Andreas Mielck, Lili Milani, Michael B. Miller, Grant W. Montgomery, Sutapa Mukherjee, Ronny Myhre, Marja-Liisa Nuotio, Dale R. Nyholt, Christopher J. Oldmeadow, Ben A. Oostra, Lyle J. Palmer, Aarno Palotie, Markus Perola, Katja E. Petrovic, Patricia A. Peyser, Ozren Polašek, Danielle Posthuma, Martin Preisig, Lydia Quaye, Katri Räikkönen, Olli T. Raitakari, Anu Realo, Eva Reinmaa, John P. Rice, Susan M. Ring, Samuli Ripatti, Fernando Rivadeneira, Thais S. Rizzi, Igor Rudan, Aldo Rustichini, Veikko Salomaa, Antti-Pekka Sarin, David Schlessinger, Helena Schmidt, Reinhold Schmidt, Rodney J. Scott, Konstantin Shakhbazov, Albert V. Smith, Jennifer A. Smith, Harold Snieder, Beate St Pourcain, John M. Starr, Jae Hoon Sul, Ida Surakka, Rauli Svento, Toshiko Tanaka, Antonio Terracciano, Alexander Teumer, A. Roy Thurik, Henning Tiemeier, Nicholas J. Timpson, André G. Uitterlinden, Matthijs J.H.M. van der Loos, Cornelia M. van Duijn, Frank J.A. van Rooij, David R. Van Wagoner, Erkki Vartiainen, Jorma Viikari, Peter M. Visscher, Veronique Vitart, Peter K. Vollenweider, Henry Völzke, Judith M. Vonk, Gérard Waeber, David R. Weir, Jürgen Wellmann, Harm-Jan Westra, H.-Erich Wichmann, Elisabeth Widen, James F. Wilson, Alan F. Wright, Jian Yang, Lei Yu, Wei Zhao.

REFERENCES

1. Lorant, V., Deliège, D., Eaton, W., Robert, A., and Ansseau, M. (2003). Socioeconomic inequalities in depression: a meta-analysis. Am. J. Epidemiol. *157*, 98–112.

2. Bromet, E., Andrade, L.H., Hwang, I., Sampson, N. a, Alonso, J., de Girolamo, G., de Graaf, R., Demyttenaere, K., Hu, C., Iwata, N., et al. (2011). Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. *9*, 90–106.

3. Andrade, L., Caraveo-Anduaga, J.J., Berglund, P., Bijl, R. V, De Graaf, R., Vollebergh, W., Dragomirecka, E., Kohn, R., Keller, M., Kessler, R.C., et al. (2003). The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. Int. J. Methods Psychiatr. Res. *12*, 3–21.

4. Graaf, R. de, Have, M. ten, Gool, C. van, and Dorsselaer, S. van (2012). Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. Soc. Psychiatry Psychiatr. Epidemiol. *47*, 203–213.

5. Mezuk, B., Myers, J.M., and Kendler, K.S. (2013). Integrating social science and behavioral genetics: testing the origin of socioeconomic disparities in depression using a genetically informed design. Am. J. Public Health *103*, S145–S151.

6. Turner, R.J., and Lloyd, D.A. (1999). The stress process and the social

distribution of depression. J. Health Soc. Behav. *40*, 374–404.
7. Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science (80-.). *340*, 1467–1471.
8. Dongen van, J., Slagboom, P.E., Draisma, H.H.M., Martin, N.G., and Boomsma, D.I. (2012). The continuing value of twin studies in the omics era. Nat. Rev. Genet. *13*, 640–653.
9. Demirkan, A., Penninx, B.W.J.H., Hek, K., Wray, N.R., Amin, N., Aulchenko, Y.S., Dyck van, R., Geus de, E.J.C., Hofman, A., Uitterlinden, a G., et al. (2011). Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol. Psychiatry *16*, 773–783.
10. Lubke, G.H., Hottenga, J.J., Walters, R., Laurin, C., Geus, E.J.C. De, Willemsen, G., Smit, J.H., Middeldorp, C.M., Penninx, B.W.J.H., Vink, J.M., et al. (2012). Estimating the Genetic Variance of Major Depressive Disorder Due to All Single Nucleotide Polymorphisms. Biol. Psychiatry *72*, 707–709.
11. López-León, S., Choy, W., and Aulchenko, Y. (2009). Genetic factors influence the clustering of depression among individuals with lower socioeconomic status. PLoS One *4*, 1–5.
12. Boardman, J.D., Domingue, B.W., and Daw, J. (2014). What can genes tell us about the relationship between education and health? Soc. Sci. Med.
13. Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H.R., Boomsma, D.I., Cichon, S., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.
14. Rietschel, M., Mattheisen, M., Frank, J., Treutlein, J., Degenhardt, F., Breuer, R., Steffens, M., Mier, D., Esslinger, C., Walter, H., et al. (2010). Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. Biol. Psychiatry *68*, 578–585.
15. Sullivan, P.F., Geus, E.J.C. De, Willemsen, G., James, M.R., Smit, J.H., Zandbelt, T., Arolt, V., Baune, B.T., Blackwood, D., Cichon, S., et al. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. Mol. Psychiatry *14*, 359–375.
16. Shi, J., Potash, J.B., Knowles, J.A., Weissman, M.M., Coryell, W., Scheftner, W.A., Lawson, W.B., DePaulo, J.R., Gejman, P. V, Sanders, A.R., et al. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. Mol. Psychiatry *16*, 193–201.
17. Muglia, P., Tozzi, F., Galwey, N.W., Francks, C., Upmanyu, R., Kong, X.Q.,

Antoniades, a, Domenici, E., Perry, J., Rothen, S., et al. (2010). Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. Mol. Psychiatry *15*, 589–601.

18. Ising, M., Lucae, S., Binder, E.B., Bettecken, T., Uhr, M., Ripke, S., Kohli, M. a, Hennings, J.M., Horstmann, S., Kloiber, S., et al. (2009). A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. Arch. Gen. Psychiatry *66*, 966–975.

19. Wray, N.R., Pergadia, M.L., Blackwood, D.H.R., Penninx, B.W.J.H., Gordon, S.D., Nyholt, D.R., Ripke, S., MacIntyre, D.J., McGhee, K. a, Maclean, a W., et al. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. Mol. Psychiatry *17*, 36–48.

20. Lewis, C.M., Ng, M.Y., Butler, A.W., Cohen-Woods, S., Uher, R., Pirlo, K., Weale, M.E., Schosser, A., Paredes, U.M., Rivera, M., et al. (2010). Genome-wide association study of major recurrent depression in the U.K. population. Am. J. Psychiatry *167*, 949–957.

21. Shyn, S.I., Shi, J., Kraft, J.B., Potash, J.B., Knowles, J. a, Weissman, M.M., Garriock, H. a, Yokoyama, J.S., McGrath, P.J., Peters, E.J., et al. (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. Mol. Psychiatry *16*, 202–215.

22. Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W.J., Assendelft, W.J.J., et al. (2008). The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int. J. Methods Psychiatr. Res. *17*, 121–140.

23. Boomsma, D.I., de Geus, E.J.C., Vink, J.M., Stubbe, J.H., Distel, M. a, Hottenga, J.-J., Posthuma, D., van Beijsterveldt, T.C.E.M., Hudziak, J.J., Bartels, M., et al. (2006). Netherlands Twin Register: from twins to twin families. Twin Res. Hum. Genet. *9*, 849–857.

24. Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., Perola, M., Ng, P.C., Mägi, R., Milani, L., et al. (2014). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int. J. Epidemiol. *Epub*, 1–11.

25. Boomsma, D., Beem, A., van den Berg, M., Dolan, C., Koopmans, J., Vink, J., de Geus, E., and Slagboom, P. (2012). Netherlands twin family study of anxious depression (NETSAD). Twin Res. *3*, 323–334.

26. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P.I.W., Deloukas, P., Gabriel, S.B., et al. (2010). Integrating common and rare genetic variation in diverse human populations.

Nature *467*, 52–58.
27. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
28. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.
29. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
30. Lee, S.H., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2012). A Better Coefficient of Determination for Genetic Profile Analysis. Genet. Epidemiol. *36*, 214–224.
31. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of heritability for human height. Nat. Genet. *42*, 565–569.
32. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M., and Wray, N.R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics *28*, 2540–2542.
33. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am. J. Hum. Genet. *88*, 294–305.
34. Nyholt, D.R. (2014). SECA: SNP effect concordance analysis using genome-wide association summary results. Bioinformatics *30*, 2086–2088.
35. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. PLoS Genet. *9*, e1003348.
36. Pampel, F.C., Krueger, P.M., and Denney, J.T. (2010). Socioeconomic Disparities in Health Behaviors. Annu. Rev. Sociol. *36*, 349–370.
37. Cutler, D.M., and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. J. Health Econ. *29*, 1–28.
38. Fletcher, J.M., and Frisvold, D.E. (2009). Higher Education and Health Investments: Does More Schooling Affect Preventive Health Care Use? J. Hum. Cap. *3*, 144–176.
39. Johnston, D.W., Lordan, G., Shields, M.A., and Suziedelyte, A. (2015). Education and health knowledge: Evidence from UK compulsory schooling reform. Soc. Sci. Med. *127*, 92–100.

40. Lantz, P.M., House, J.S., Mero, R.P., and Williams, D.R. (2005). Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. J. Health Soc. Behav. *46*, 274–288.
41. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S. V, Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.
42. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379.

Supplement of Chapter 3. Association between depression and lower educational attainment not explained by shared genetic effects in ~25,000 individuals

Genotyping, quality control, and imputation

Genotyping, quality control, and imputation of the PGC-cohorts have been described in detail previously.[1] In short, the PGC-cohorts were all genotyped following their own protocol and on separate platforms (Supplementary Table 1). During quality control SNPs were removed with missingness ≥ 0.02, case-control difference in SNP missingness ≥ 0.02, difference in allele frequency to HapMap3 ≥ 0.15, or Hardy-Weinberg equilibrium (HWE) p-value < 1e-6; subjects were removed with missingness ≥ 0.02 or diverging ancestry. The PGC-cohorts were then imputed against the CEU+TSI HapMap3 reference panel,[2] and post-imputation QC selected SNPs with imputation $r^2 > 0.6$ and MAF > 0.01.

The NESDA/NTR-2 cohort was genotyped on the Affymetrix 6.0 Human Genome-Wide SNP Array in three separate batches. SNPs with a significant different allele frequency between any two of the three batches (Chi-square p-value < 0.01) were excluded to correct for batch effects. Subsequent quality control was performed four times and based on information from the separate batches and the combined NESDA/NTR-2 sample and removed SNPs with MAF < 0.1, missingness > 0.05, HWE p-value < 0.05, Mendelian error rate ≥ 0.01. Based on the combined sample only, additional SNPs were removed with a significant difference in missingness between cases and controls (p < 0.05), and with a difference in allele frequency to HapMap3 ≥ 0.15. Subjects were removed, based on information of the combined sample, with SNP missingness > 0.01 or Mendelian error rate ≥ 0.01. The NESDA/NTR-2 cohort was then imputed against the CEU+TSI HapMap3 reference panel,[2] and with post-imputation QC SNPs were selected with imputation INFO > 0.8 and MAF > 0.01.

The EGCUT- cohort was genotyped partly on Illumina Human 370 CNV-duo chip (unrelated N=1514)[3] and partly on Illumina Human Omni Express (unrelated N=5188).[4] Both of these parts of the EGCUT-cohort were processed separately. With quality control SNPs were removed with MAF ≤ 0.01 or HWE p-value < 1e-6. The SNPs were then imputed against the 1000 Genomes reference panel, and with post-imputation QC SNPs were selected with imputation INFO > 0.8 and MAF > 0.01. The HapMap3 SNPs were selected and lifted from hg19 to hg18 in order to align with the other cohorts.

The genotype data of the PGC-cohorts, NESDA/NTR-2 cohort, and EGCUT cohort were merged and yielded information on 884,105 overlapping SNPs.

Genetic Profile Risk Scores (GPRS)

The GPRS were based on EA discovery results (EA-GPRS) and MDD discovery results (MDD-GPRS) following the procedure described by Purcell et al[5] and implemented in Plink.[6] In order to test GPRS in a target cohort, it is essential for the target cohort to be independent from the discovery cohort.[5] The EA discovery results were from the recent meta-analyses on US years of schooling from the Social Science Genetics Association Consortium (SSGAC),[7] which contains overlapping individuals with BMH, EGCUT, GenRED, NESDA/NTR-1, NESDA/NTR-2, QIMR, and STAR-D. We had available the SSGAC results separately excluding (i) EGCUT, (ii) NESDA/NTR-1 and 2, and (iii) QIMR. These three sets of EA discovery results were based on around 120,000 subjects, and applied to estimate the EA-GPRS in the respective cohorts. The EA discovery results excluding NESDA/NTR-1 and 2 were in addition applied to estimate risk scores in GSK, MPIP, and RADIANT. EA-GPRS were not estimated for BMH, GenRED, and STAR*D, because no independent discovery results were available.

To obtain the MDD discovery results was slightly more elaborate, because no large MDD cohort exists that is independent from PGC. Therefore, a ten-fold leave-one-cohort-out approach was followed, in which every cohort was once left out as target cohort, while the nine other cohorts would serve as MDD discovery set. In these discovery sets genome wide association studies were performed correcting for sex, covariates labelling the cohorts and genotype batches, and ten principal components. In this manner, independent MDD discovery results were obtained for all of the ten cohorts, and the discovery results were thus bases on around 8,000 cases and 12,000 controls depending on the size of the cohort left out (see Table 1 of main manuscript).

Nagelkerke's and the liability R^2

For MDD, Nagelkerke's R^2 were derived and corrected for the covariates by substituting the null (or intercept) model in Nagelkerke's equation for the model including the covariates. The corrected Nagelkerke's R^2 were thus estimated as

$$R^2_{NK} = \left\{1 - \left(\frac{Likelihood\ model\ covariates\ only}{Likelyhood\ full\ model}\right)^{2/N}\right\} / \{1 - (Likelihood\ model\ covariates\ only)^{2/N}\}$$

Lee at all indicated that Nagelkerke's R^2 can be biased by ascertainment, when the proportion of cases in the study sample differs from the population disease frequency.[8] Therefore, they proposed a R^2 measure that is robust against

ascertainment bias and interpretable on the liability scale. This R^2 was obtained by rescaling Nagelkerke's R^2 for a MDD population prevalence of K=0.2 following equation (15) from Lee et al.[8] Suppose that Nagelkerke's R^2_{NK} has been estimated on a sample with a proportion of cases P for a disease with population frequency K. From Table 1 from Lee et al it follows that $R^2_{Cox\ \&\ Snell} = \in R^2_{NK}$, with $\in = 1 - P^{2P}(1-P)^{2(1-P)}$. On page 216 Lee et al show that $R^2_{Cox\ \&\ Snell}$ is approximately equal to the $R^2_{O_{cc}}$ on the observed scale in a linear model. Subsequently, equation (15) from Lee et al can be applied

$R^2_{Liability} = CR^2_{O_{cc}} / \left(1 + C\theta R^2_{O_{cc}}\right)$, where

$C = \frac{K(1-K)}{z^2}\frac{K(1-K)}{P(1-P)}$ and $\theta = m\frac{P-K}{1-K}\left(m\frac{P-K}{1-K} - t\right)$, with

t: the threshold of the standard normal distribution above which a proportion of P is found, and

m: the selection intensity, which equals z/K where z equals the height of the standard normal distribution at t

Gene-by-environment interaction

In post-hoc analyses we tested if the polygenic effects on MDD were moderated by EA. First, we tested if the effect of MDD-GPRS (P_T=1) on MDD was moderated by EA (continuous measure), but found no such evidence with an interaction effect of OR=1.02 (p=0.26). Second, we applied GREML analyses. Therefore, EA was partitioned in two equal parts per study: low EA and high EA. Subsequently, the MDD SNP-h^2 on the observed scale was estimated at 0.11 (SE=0.04; p=0.008) in low EA and 0.08 (SE=0.04; p=0.047) in high EA. The genetic correlation between MDD in low EA and MDD in high EA was estimated at 0.87 (SE=0.40; p=0.755 for H_0: r_G=1), showing no evidence for a difference of the genetic effects on MDD in low and in high EA.

Table S1. Brief summary of contributing cohort studies

Cohort	Cases	Controls	Genotyping platform
Psychiatric Genomics Consortium (PGC MDD1) - cohorts			
Bonn/Mannheim (BMH)	Inpatients	Population-based, non-screened	Illumina 610K
GenRED (GE)	Volunteers with early onset/ recurrent MDD	Population-based, screened	Affymetrix 6.0
GSK (GS)	Recurrent MDD from clincal centre	Population-based and clinical, screened	Illumina 550K
MPIP (M)	Inpatients	Population-based, screened	Illumina 317K
NESDA/NTR-1 (N1)	Outpatients & population-based	Population-based, screened	Perligen 600K
QIMR (Q)	Population-based from Australia	Population-based, screened	Illumina 317 & Illumina 610
RADIANT (R)	Cases with MDD from clincal centres	Population-based and volunteer, screened	Illumina 610K
STAR*D (S)	Cases from clinical trial	Population-based, screened	Affymetrix 500K/5.0
Other cohorts			
NESDA/NTR-2 (N2)	Outpatients & population-based	Population-based, screened	Affymetrix 6.0
EGCUT (E)	Population-based volunteers	Population-based volunteers, screened	Illumina 370 & Illumina HOE

Information of the eight PGC cohorts is displayed. This Table closely resembles Supplementary Table 2 from the Psychiatric Genomics Consortium; see the publication from the Psychiatric Genomics Consortium for more details.

Table S2. Number of SNPs included in polygenic risk scores

	B	N2	E	N1	GE	GS	M	Q	R	S
Discovery MDD with threshold										
p < 0.001	91	96	88	94	104	104	104	93	92	106
p < 0.01	902	894	878	843	908	918	904	922	870	892
p < 0.1	8225	8292	8273	8103	8369	8345	8401	8321	8118	8303
p < 1	76516	76516	76516	76516	76516	76516	76516	76516	76516	76516
Discovery EA years with threshold										
p < 0.001	-	179	203	179	-	179	179	213	179	-
p < 0.01	-	1290	1304	1290	-	1290	1290	1359	1290	-
p < 0.1	-	9327	9538	9327	-	9327	9327	9578	9327	-
p < 1	-	76516	76516	76516	-	76516	76516	76516	76516	-

The Genetic Profile Risk Scores (GPRS) were based on a set of 76,516 independent SNPs with results available from all discovery sets. The GPRS based on all SNPs (threshold of discovery p- value <1) are thus based on the exact same SNPs. The GPRS based on the lower thresholds (0.001, 0.01, and 0.1) are based on different set of SNPs based on significance in the respective discovery results. EA-GPRS could not be estimated for B, GE and S, because we had no independent discovery results available for these cohorts. B=Bonn/Mannheim; E=EGCUT; GE=GenRED; GS=GSK; M=MPIP; N1=NESDA/NTR-1; N2=NESDA/NTR-2; Q=QIMR; R=RADIANT; S=STAR*D

Table S3. Correlation between polygenic risk scores and the first ten principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Discovery MDD with threshold										
p < 0.001	0.01*	0.01	-0.01	-0.00	-0.01	<0.01	0.01*	-0.02**	0.02*	-0.02**
p < 0.01	-0.00	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01	0.01	-0.00
p < 0.1	-0.02**	0.01	-0.01	-0.01	<0.01	-0.01	-0.02***	-0.00	-0.01	-0.00
p < 1	-0.01	0.01	-0.02***	-0.00	0.01	-0.01	-0.02**	-0.01	-0.00	<0.01
Discovery EA years with threshold										
p < 0.001	-0.00	-0.02**	0.01*	-0.01	-0.01	0.01	-0.01	-0.00	-0.01*	0.02*
p < 0.01	-0.01*	-0.00	0.02*	-0.01	-0.01	<0.01	-0.01	-0.01*	<0.01	0.02*
p < 0.1	-0.01	-0.01	0.04***	-0.01	0.01	0.01	-0.01	-0.01*	<0.01	0.01
p < 1	-0.02**	-0.03***	0.04***	-0.02*	0.02***	0.01	-0.01	-0.01	-0.00	<0.01

* p<0.05 ** p<0.01 *** p<0.001. Some of the GPRS were correlated to some of the first ten principal components. This underlines the necessity to correct the analyses for population stratification.

Table S4. The effect of genomic risk profile scores based on College-completion discovery results (College-GRPS) on MDD and EA (in years of education in the overall sample.

	Effect on MDD									
	N		Effect		R-squared (%)		Effect on EA			
	Case	Control	OR	P-value	NK	Liability	N	Beta	P-value	R^2(%)
Effect of College-GRPS										
p < 0.001	6544	12324	0.99	0.683	0.00	0.00	13477	0.05	6.2e-10	0.28
p < 0.01	6544	12324	0.99	0.609	0.00	0.00	13477	0.07	4.6e-16	0.49
p < 0.1	6544	12324	0.99	0.564	0.00	0.00	13477	0.08	3.1e-20	0.63
p < 1	6544	12324	0.99	0.737	0.00	0.00	13477	0.10	1.7e-29	0.94

The main EA-GPRS analyses were based on discovery results from years of education. The SSGAC, however, also conducted a meta-analyses on the binary measure of College completion.7 This measure is also of interest, because it distinguishes in the extreme end of the EA-distribution. Therefore, additional analyses were performed with GPRS based on College completion discovery results (College-GPRS). The College-GPRS had a slightly smaller impact on EA than main EA-GPRS (main Table 2), and had no impact on MDD.

Table S5. Comparing GREML estimates of the MDD SNP-h^2 on the observed scale to estimates on the liability scale

	N		MDD SNP-h^2 observed scale			MDD SNP-h^2 liability scale				
Cohorts	**Case**	**Control**	**Estimate**	**SE**	**P-value**	**K=0.10**	**K=0.15**	**K=0.20**	**K=0.25**	**K=0.30**
All	9662	14949	0.126	0.013	<1e-16	0.139	0.158	0.173	0.184	0.193
All excluding B	8737	13667	0.113	0.014	1.1e-16	0.125	0.142	0.155	0.165	0.173
All excluding E	9154	9604	0.159	0.016	<1e-16	0.167	0.190	0.207	0.221	0.232
All excluding GE	8686	13734	0.133	0.014	<1e-16	0.147	0.167	0.183	0.195	0.204
All excluding GS	8796	14086	0.124	0.014	<1e-16	0.138	0.157	0.172	0.183	0.192
All excluding M	9325	14416	0.130	0.013	<1e-16	0.143	0.162	0.177	0.189	0.198
All excluding N1	8102	13826	0.107	0.014	2.1e-14	0.121	0.137	0.150	0.160	0.168
All excluding N2	9426	13748	0.129	0.013	<1e-16	0.140	0.160	0.174	0.186	0.195
All excluding Q	8230	13263	0.144	0.014	<1e-16	0.160	0.182	0.199	0.212	0.223
All excluding R	8057	13376	0.116	0.014	1.0e-15	0.129	0.147	0.161	0.171	0.180
All excluding S	8445	14821	0.133	0.013	<1e-16	0.151	0.171	0.187	0.200	0.209

The MDD SNP-h^2 on the observed scale was compared to the SNP-h^2s on the liability scale for different disease population frequencies K. In our sample the MDD SNP-h^2 was larger expressed on the liability than expressed on the observed scale. Furthermore, the MDD SNP-h^2 increased further when assuming a larger disease frequency K. B=Bonn/Mannheim; E=EGCUT; GE=GenRED; GS=GSK; M=MPIP; N1=NESDA/NTR-1; N2=NESDA/NTR-2; Q=QIMR; R=RADIANT; S=STAR*D

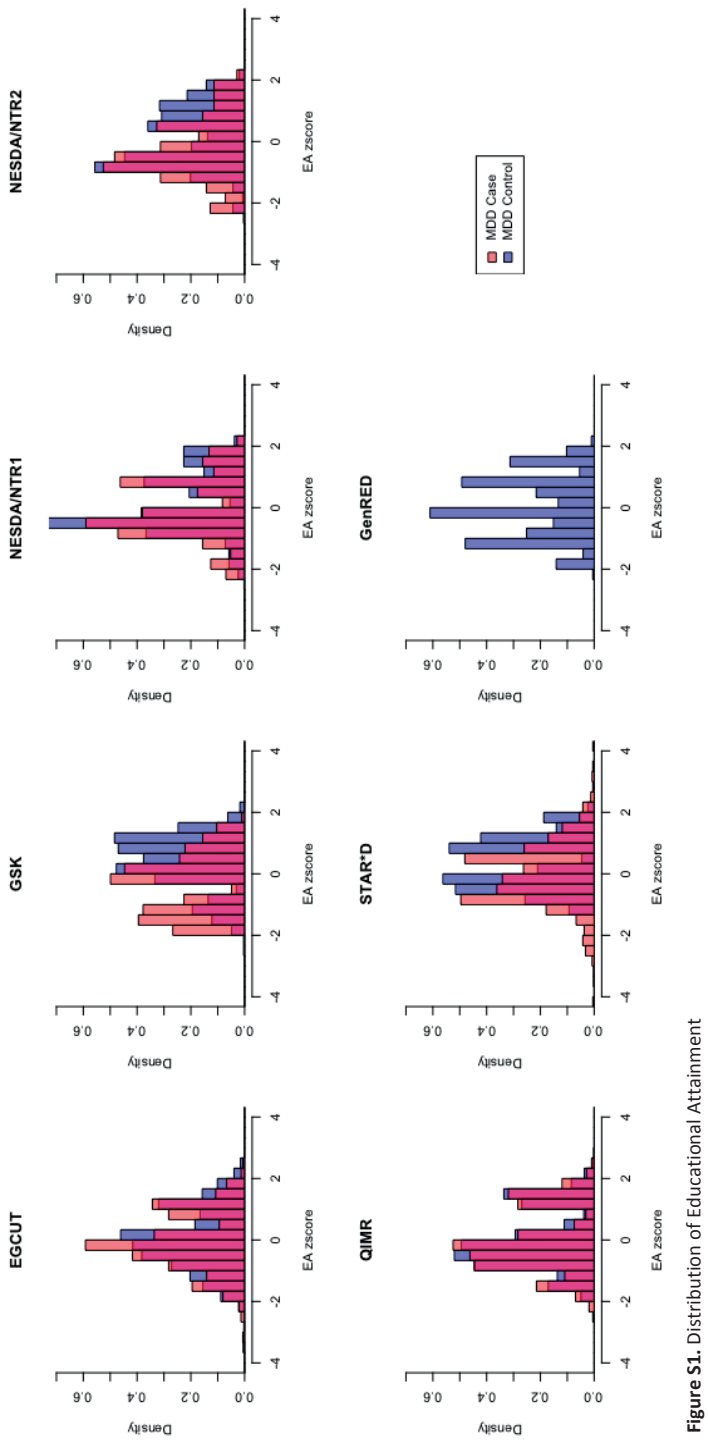


Figure S1. Distribution of Educational Attainment

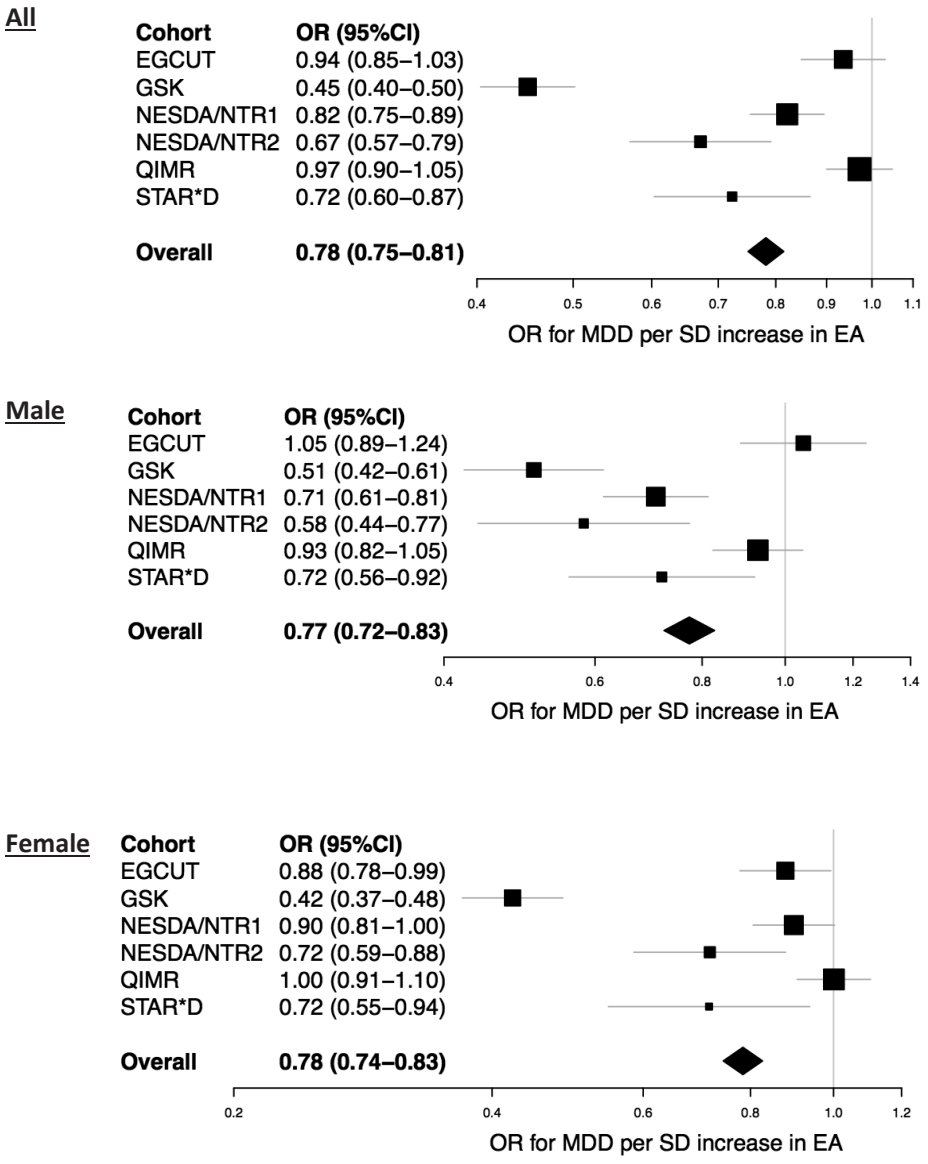


Figure S2. Association between EA and MDD in male and female

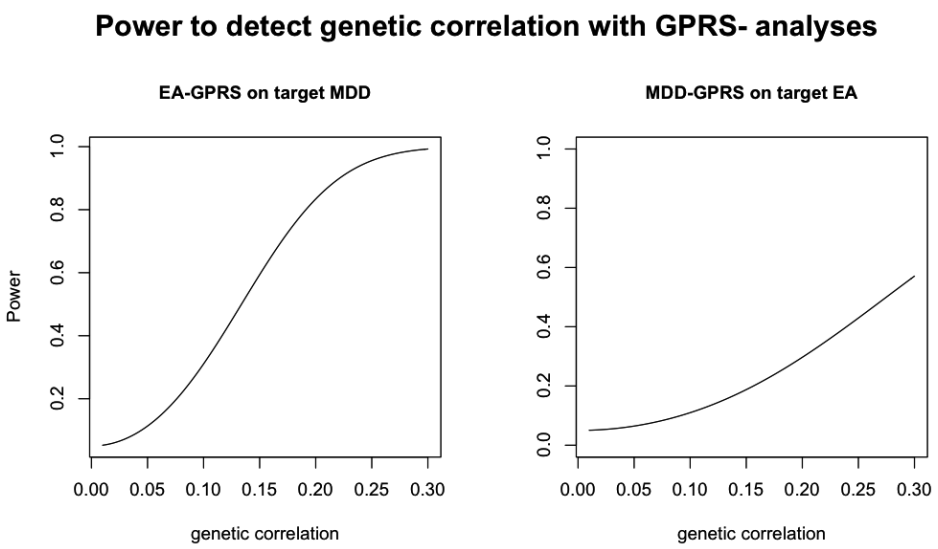


Figure S3. Dudbridge power calculations for GPRS-analyses. Dudbridge's method (available at https://sites.google.com/site/fdudbridge/software/) was applied to estimate the power to detect genetic correlation with GPRS- analyses for $\alpha = 0.05$.[9] This power can be derived from to the SNP-h2 estimated with Dudbridge's method, as well as with the SNP-h2 following from GREML analyses.[10] Based on our results in main Table 2, Dudbridge's method estimated the MDD SNP-h2 at 0.13 and the EA SNP-h2 at 0.09 (applying the R function "estimateVg2FromP"), which estimates used to calculate the power as displayed (applying the R function "polygenescore"). The power to detect genetic correlation was larger for the EA-GPRS predicting MDD than for the MDD-GPRS predicting EA, because of the difference in discovery sample size (~120,000 and ~20,000 respectively). Furthermore, our GPRS- analyses seemed well powered to detect a genetic correlation of -0.2, but could not exclude a smaller genetic correlation of around -0.1. Note that Dudbridge estimates of the SNP-h^2s were lower than the GREML-estimates from this study (main Table 3) and previous studies;[7,11] these power analyses will therefore likely represent a lower bound of the power to detect genetic correlation with GPRS-analyses.

REFERENCES

1. Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H.R., Boomsma, D.I., Cichon, S., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.

2. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P.I.W., Deloukas, P., Gabriel, S.B., et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.

3. Nelis, M., Esko, T., Mägi, R., Zimprich, F., Zimprich, A., Toncheva, D., Karachanak, S., Piskácková, T., Balascák, I., Peltonen, L., et al. (2009). Genetic structure of Europeans: a view from the North-East. PLoS One *4*, e5472.
4. Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., Perola, M., Ng, P.C., Mägi, R., Milani, L., et al. (2014). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int. J. Epidemiol. *Epub*, 1–11.
5. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.
6. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
7. Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science (80-.). *340*, 1467–1471.
8. Lee, S.H., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2012). A Better Coefficient of Determination for Genetic Profile Analysis. Genet. Epidemiol. *36*, 214–224.
9. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. PLoS Genet. *9*, e1003348.
10. Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A.E., Dudbridge, F., and Middeldorp, C.M. (2014). Research Review: Polygenic methods and their application to psychiatric traits. J. Child Psychol. Psychiatry *55*, 1068–1087.
11. Lubke, G.H., Hottenga, J.J., Walters, R., Laurin, C., Geus, E.J.C. De, Willemsen, G., Smit, J.H., Middeldorp, C.M., Penninx, B.W.J.H., Vink, J.M., et al. (2012). Estimating the Genetic Variance of Major Depressive Disorder Due to All Single Nucleotide Polymorphisms. Biol. Psychiatry *72*, 707–709.

Chapter 4

Increased polygenic effects on depression in individuals exposed to childhood trauma in the Netherlands Study of Depression and Anxiety

Wouter J. Peyrot
Yuri Milaneschi
Abdel Abdellaoui
Patrick F. Sullivan
Jouke J. Hottenga
Dorret I. Boomsma
Brenda W.J.H. Penninx

Published in *The British Journal of Psychiatry 2014, 205:113-119*

ABSTRACT

Background: Research on gene-by-environment interaction in major depressive disorder (MDD) has thus far primarily focused on candidate genes, while genetic effects are known to be polygenic.

Aims: To test whether the effect of polygenic risk scores on MDD is moderated by childhood trauma.

Methods: The study sample consisted of 1645 participants with a DSM-IV based diagnosis of MDD and 340 screened controls from the Netherlands. Chronic or remitted episodes (severe MDD) were present in 956 participants. The occurrence of childhood trauma was assessed with the Childhood Trauma Interview and the polygenic risk scores were based on genome-wide meta-analysis results from the Psychiatric Genomics Consortium.

Results: The polygenic risk scores and childhood trauma independently affected MDD risk, and evidence was found for interaction as departure from both multiplicativity and additivity, indicating that the effect of polygenic risk scores on depression is increased in the presence of childhood trauma. The interaction effects were similar in predicting all MDD risk and severe MDD risk and explained a comparable proportion of variation in MDD risk as the polygenic risk scores themselves.

Conclusions: The interaction effect found between polygenic risk scores and childhood trauma implies that (1) studies on direct genetic effect on MDD gain power by focusing on individuals exposed to childhood trauma, and that (2) individuals with both high polygenic risk scores and exposure to childhood trauma are particularly at risk for developing MDD.

INTRODUCTION

Research on gene-by-environment (GxE) interaction in major depressive disorder (MDD) aims to understand the heterogeneity of environmental and genetic risk factors, but has thus far primarily focused on candidate genes with inconclusive findings.[1,2] On the one hand, research on GxE could select individuals with increased vulnerability for environmental factors based on their genetic make-up. Alternatively, research on GxE could select environmental conditions that lead to increased expression of genetic effects. Insights into GxE interaction is therefore of general importance for psychiatric research and contributes to the understanding of MDD's complex etiology.

Thus far, GxE interaction has primarily been tested for candidate genes such as the serotonin transporter gene (5-HTTLPR), for which opposing results were found in very similar single studies,[3,4] as well as in meta-analyses.[5,6] Several environmental factors have been analyzed in this respect, and among the most important factors is childhood trauma, which has a strong impact on MDD risk.[7–10] Nevertheless, although some consistent evidence for interaction between childhood trauma and 5-HTTLPR was found, these GxE findings remain controversial.[1] The progress from a candidate-gene to a hypothesis-free genome-wide approach is hampered by lack of statistical power and inconsistent assessment of environmental stressors across GWAS cohorts.

Research on main genetic effects, on the other hand, has indicated that the risk of MDD is not merely increased by the effect of one or a few single nucleotide polymorphisms (SNPs), but by polygenic variation.[11,12] One of the methods applied to point at these polygenic effects, first introduced for schizophrenia,[13] uses polygenic risk scores and was later applied to MDD.[11] The polygenic risk scores are obtained after carrying out a genome-wide association study in a discovery sample and then taking SNPs up to a certain threshold of significance, or even all SNPs, to predict MDD in an independent target sample. The contributions of these large numbers of SNPs are weighted by their effect size in a GWA or meta-analysis. The effect of polygenic risk scores on MDD was repeatedly confirmed and explains up to 1-2 percent of variation.[11,14,15]

Even though this has not yet been studied for MDD, it is likely that causal genetic variants for MDD are located throughout all of the genome, as has been found for other complex traits such as height and body mass index.[16] Also, SNPs contributing to pleiotropy between schizophrenia and bipolar disorder are found dispersed throughout the genome.[17] The finding that the risk of MDD is increased by polygenic variation suggests that research of interaction effects should also focus on polygenic information. With an expected abundance of causal variants

for MDD, environmental conditions that increase genetic effects are more likely to be found when polygenic information is taken into account. When environmental conditions that increase genetic effects are found, individuals exposed to these conditions can be selected for future research to study the impact of single loci on MDD with increased power. Nevertheless, to the best of our knowledge, no research on GxE interaction in MDD has focused on polygenic information thus far.

The current study focused on polygenic information to test for gene-by-environment interaction in MDD, and examined whether polygenic risk scores interact with the presence of childhood trauma in a large and well-characterized sample from the Netherlands. This sample consists of participants with DSM-IV based diagnosis of MDD and screened controls, with the presence of childhood trauma assessed in face-to-face interviews. The meta-analysis from the Psychiatric Genomics Consortium (PGC)[15] excluding our sample (leaving 7544 cases and 7754 controls) was used as discovery sample to construct the polygenic risk scores.

METHODS

Subjects

The sample consisted of participants from the Netherlands Study of Depression and Anxiety (NESDA), which is an ongoing longitudinal cohort study of depressive and anxiety disorders, with participants recruited from mental health care settings, general practices and the general population in the period from 2004 to 2007.[18] Participants with MDD in their lifetime (N=1645) were diagnosed in a face-to-face interview with a trained clinical staff-member following the DSM-IV based Composite International Diagnostic Interview (CIDI, version 2.1). Over half of these participants (N=956) suffered from severe MDD with remitted (more than one) episodes and/or chronic (longer than two years of) complaints, as assessed with the life-chart, a calendar-approach to calculate the percentage of time symptoms were present during four years prior to baseline and two years following baseline.[19] Controls (N=340) were screened in a similar face-to-face CIDI interview and had no diagnosis of a depressive, dysthymic, anxiety, or other psychiatric disorder in lifetime. Participants were from North-European ancestry and were excluded when they were not fluent in speaking Dutch or when they suffered from another primary diagnosis, such as a psychotic, obsessive compulsive, bipolar or severe substance use disorder. The NESDA study was approved by the institutional Review Board and all participants provided written informed consent.

Childhood Trauma

Childhood trauma was assessed in a face-to-face interview with a trained clinical staff-member following the Childhood Trauma Interview (CTI) from the Netherlands Mental Health Survey and Incidence Study.[7] The CTI assesses the domains of emotional neglect, psychological abuse, physical abuse and sexual abuse before the age of sixteen, and yields a score ranging from 0 to 8 by adding the frequencies of occurrence (0- absent ; 1- once or sometimes; 2- regularly, often or very often). In the CTI the four domains are assessed by a first question asking whether the traumatic event occurred (yes or no), and a subsequent second question asking how often the event occurred. In the first question the traumatic events were specified as follows: emotional neglect as the lack of parental attention or support and ignorance of one's problems and experiences; psychological abuse as being verbally abused, undeserved punishment, subordinated to siblings and being blackmailed; physical abuse as being kicked or hit with hands or an object, beaten up or physical abuse in any other way; and sexual abuse as being sexually approached against one's will, meaning being touched or having to touch someone in a sexual way. The CTI is a well-established instrument, which measurements of childhood trauma show a strong impact on depressive and anxiety disorders[7,20] as well as on structural and functional brain abnormalities.[21,22] The CTI also shows strong content validity when compared to the Childhood Trauma Questionnaire (CTQ),[23] with Spearman's rho correlation of 0.69 ($p<0.001$) in a subset of NESDA with both the CTI and CTQ assessed at different time points.

Genotyping and Quality Control

Methods for blood sampling and DNA extraction were described previously.[24] The manufacturer's protocol was followed to genotype the autosomal SNPs on the Affymetrix 6.0 Human Genome-Wide SNP Array. With quality control, SNPs were excluded that: had probes that mapped badly against NCBI Build 37/UCSC hg19; had a minor allele frequency smaller than 1%; had a missing rate greater than 5%; deviated from Hardy-Weinberg equilibrium with a p-value smaller than 0.001, thus leaving 498,592 SNPs to analyze. Participants were excluded when: they showed a Contrast QC < 0.4 (CQC, a quality metric from Affymetrix representing how well allele intensities separate into clusters); fell outside of the main cluster of a PCs reflecting a batch effect;[25] had a missing rate greater than 5%; had excess genome-wide heterozygosity or inbreeding levels ($F < -0.10$ or > 0.10); had genotypes with inconsistencies regarding reported gender; or had non-European/non-Dutch ancestry as indicated with principal component analysis.[25]

Polygenic Risk Scores

The polygenic risk scores were created based on the results from a large meta-analysis from the Psychiatric Genomics Consortium (PGC)[15] excluding participants from the Dutch GWAS cohort[26] that included NESDA participants (thus yielding 7544 cases and 7754 controls in the discovery set). Risk scores were obtained following the method described by Purcell and colleagues[13] with the PLINK-software.[27] From the meta-analysis, SNPs were selected that had an imputation INFO score > 0.9 and MAF > 0.02, and low linkage disequilibrium (LD) to each other ($r^2 < 0.25$ within 500kb window, filtering for significance; PLINK-command --clump--p1 1 --clump-p2 1 --clump-r2 0.25 --clump-kb 500). The meta-analysis results of SNPs up to eight p-value thresholds (0.001; 0.01; 0.05; 0.1; 0.2; 0.3; 0.4; and 0.5) were selected to compute the polygenic risk scores in our sample; the numbers of SNPs thus included were 150, 1209, 5028, 8905, 16081, 22355, 28018, and 32870 respectively. The polygenic risk scores were standardized to a mean of zero and standard deviation of one to aid interpretation of results.

Statistical analyses

Participants with MDD were compared to controls with respect to age, gender, and their childhood trauma score (range 0-8) with t-tests for continuous and chi-square-tests for binary variables. The effect of polygenic risk scores on the childhood trauma score, i.e. gene-environment correlation, was tested with linear regression, because such an effect could potentially bias tests for interaction.[28] Two binary MDD outcomes were analyzed as dependent variables: all participants with MDD versus controls (all MDD risk), and participants with severe MDD versus controls (severe MDD risk). The direct effects of polygenic risk scores (model 1) and the childhood trauma score (model 2) on MDD risks were assessed in separate logistic regression models. Subsequently, tests for interaction were performed with logistic regression to test for interaction as departure from multiplicativity (model 3) and, secondly, with analyses of relative excess risks due to interaction (RERI, model 4) to test for interaction as departure from additivity. The RERIs were computed with the method described by Knol and colleagues, as $RERI=e^{\beta CT+\beta PRS+\beta PRSxCT}-e^{\beta CT}-e^{\beta PRS}+1$.[29] The RERI's 95% confidence intervals were computed with bootstrapping with 10,000 iterations. The difference between interaction as departure from additivity and interaction as departure of multiplicativity is that the first represents a situation where the combined effect is larger than the sum of the individual effects of the polygenic risk score and childhood trauma, whereas the latter represents a situation where the combined effect is larger than the product of the individual effects. It has been argued that

interaction as departure from additivity is more in line with biological interaction.[29]

Nagelkerke's R^2 were estimated to assess what proportion of variation in all MDD risk was explained by the polygenic risk scores (PRS) and childhood trauma (CT) independently, as well as their interaction CTxPRS. Therefore, several R^2 estimates were compared: between the model with only covariates and the model additionally including CT (R^2 of CT); between the model with covariates and CT and the model additionally including PRS (R^2 of PRS); and between the model with covariates, CT and PRS and the model additionally including PRSxCT (R^2 of CTxPRS). Nagelkerke's R^2 may, however, be biased by a sample's ascertainment when a disproportional number of cases is selected from the population.[30] Therefore, we also computed an alternative R^2 measure for the PRSs, which was recently proposed by Lee and colleagues. This R^2 measure is based on the liability scale, directly comparable to the heritability, and robust against ascertainment bias. [30] Lee's R^2 estimates in our sample were based on a Dutch lifetime prevalence of MDD of 18.7%.[31]

All analyses were corrected for age, gender, and three ancestry-informative principal components to take possible population stratification into account, and the tests for interaction (models 3 and 4) included polygenic risk scores and the childhood trauma score as additional covariates. Effects were considered significant when p-values were <0.05 or when RERIs 95% confidence intervals did not contain zero. All analyses were performed in R.[32]

Table 1. Effect of polygenic risk scores (PRS) on childhood trauma

PRS thresholds	Beta	P-value
p<0.001	<0.01	0.991
p<0.01	-0.01	0.769
p<0.05	0.02	0.733
p<0.1	0.01	0.847
p<0.2	-0.01	0.883
p<0.3	-0.02	0.754
p<0.4	-0.01	0.904
p<0.5	0.01	0.907

Effects of polygenic risk scores on childhood trauma (i.e. gene-environment correlation) were estimated with linear regression including three principal components, age and gender as covariates.

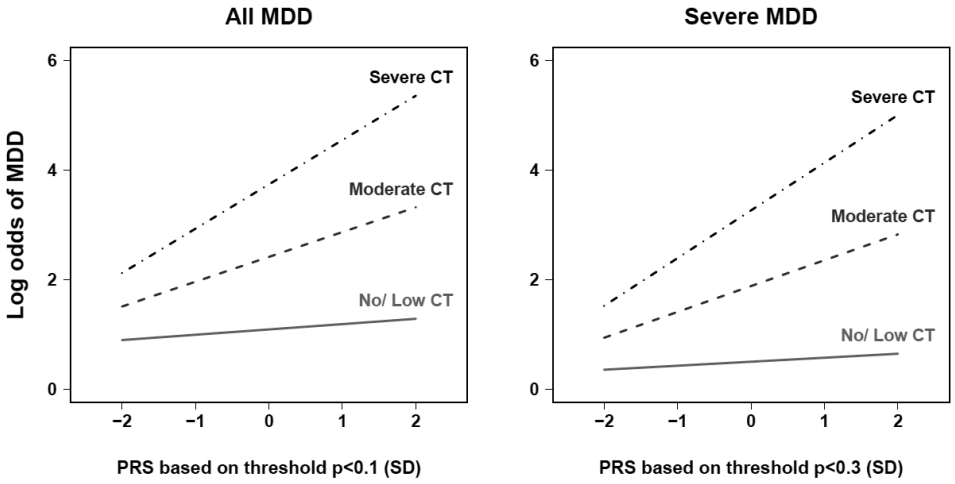


Figure 1. Interaction between childhood trauma (CT) and the polygenic risk score (PRS). The interaction-effects as departure of multiplicativity in predicting risk on all MDD and risk on severe MDD are visualized by displaying the direct effects of the polygenic risk scores (PRS) based on threshold p<0.1 and p<0.3 respectively for three childhood trauma levels, with childhood trauma (CT)-scores of 0-1; 2-3; and 4-8 respectively.

RESULTS

Participants with MDD (N=1645) had a mean age comparable to that of the 340 healthy controls (42.2 [SD 12.5] and 43.3 [SD 14.5] respectively, p=0.172), and were slightly more often female (68% and 57% respectively, p<0.001). The mean childhood trauma score was 1.75 (SD 2.17, range 0-8), and mean scores of the four childhood trauma domains (range 0-2) were 0.76 (0.95) for emotional neglect (EN), 0.50 (0.84) for psychological abuse (PsA), 0.22 (0.57) for physical abuse (PhA), and 0.24 (0.52) for sexual abuse (SA). The scores of the domains were all correlated with each other with Pearson correlation coefficients of 0.61 for EN-PsA, 0.40 for EN-PhA, 0.24 for EN-SA, 0.55 for PsA-PhA, 0.23 for PsA-SA, and 0.26 for PhA-SA (all p<0.001). Childhood trauma occurred more often in participants with MDD than in healthy controls with mean childhood trauma main scores of 1.99 (SD 2.24) and 0.56 (SD 1.29) respectively (p<0.001). None of the polygenic risk scores had an effect on childhood trauma with beta-estimates around zero and all p-values well over 0.05, thus excluding gene-environment correlation and its potential bias on interaction tests (Table 1). The polygenic risk scores significantly predicted MDD risk (model 1), with slightly larger but comparable effects in predicting severe MDD risk compared to predicting all MDD

risk (Table 2). The polygenic risk scores based on five of the eight studied thresholds were predictive in all MDD risk (thresholds 0.05; 0.1; 0.2; 0.3; 0.4) and the polygenic risk scores based on six thresholds were predictive in severe MDD risk (thresholds 0.05; 0.1; 0.2; 0.3; 0.4; 0.5). The polygenic risk scores based on threshold $p<0.05$ had the largest effect on all MDD risk, with an OR of 1.22 per standard deviation increase of the polygenic risk score ($p=0.001$). The presence of childhood trauma also predicted MDD risk (model 2), again with slightly larger but comparable effects in predicting severe MDD risk compared to all MDD risk (with ORs of 1.64 [$p<0.001$] and 1.69 [$p<0.001$] respectively, per childhood trauma score unit increase [range 0-8], Table 2). Evidence was then found for interaction as departure from both multiplicativity (model 3, odds ratios >1) and additivity (model 4, RERIs>0), indicating that the effect of polygenic risk scores on MDD is increased in the presence of childhood trauma (Table 2). The largest interaction effect in predicting all MDD was found for the polygenic risk score based on threshold $p<0.1$ with an OR of 1.15 ($p=0.005$); the largest interaction effect in predicting severe MDD was found for the polygenic risk score based on threshold $p<0.3$ with an OR of 1.16 ($p=0.005$). These two interaction-effects were visualized for their departure of multiplicativity by displaying the direct effects of the polygenic risk scores for three childhood trauma levels, with childhood trauma scores of 0-1; 2-3 and 4-8 respectively (Figure 1). Figure 1 shows that the polygenic risk scores have limited impact in predicting MDD risk in individuals with no/low exposure to childhood trauma, but large impact in individuals with high exposure to childhood trauma. The impact of the four separate childhood trauma domains on the interaction effects were compared by conducting analyses of each domain separately in predicting all MDD risk. The estimates of interaction thus found were in the same direction for all domains (OR>1), but appeared more significant for the domains of emotional neglect and psychological abuse, than for the domains of physical abuse and sexual abuse (Table 3). This difference in significance is possibly due to the lower frequency of occurrence of physical abuse and sexual abuse. Most variation in all MDD risk was explained by childhood trauma (~13%), but the proportions explained by the polygenic risk scores (in addition to the variation explained by childhood trauma) and their interaction-effects (in addition to the variation explained by childhood trauma and the polygenic risk score) were of comparable magnitude (~0.5%, Table 4). Note that Lee's R^2 estimates were comparable to Nagelkerke's R^2 estimates for the polygenic risk scores, which indicates that ascertainment bias did not largely impact our results (Table 4).

Table 2. Interaction between polygenic risk scores (PRS) and childhood trauma (CT) in predicting MDD risk and direct effects of PRSs and CT

	Direct effects				PRS-by-CT interaction			
	PRS (model 1)		CT (model 2)		Multiplicative (model 3)		Additive (model 4)	
PRS thresholds	OR	P-value	OR	P-value	OR	P-value	RERI	95% CI
All MDD (1645 cases and 340 controls)								
p<0.001	1.01	0.808	1.64	<0.001	1.06	0.288	0.08	-0.08:0.25
p<0.01	1.12	0.059	1.64	<0.001	1.09	0.080	0.21	0.04:0.47
p<0.05	1.22	0.001	1.64	<0.001	1.14	0.008	0.37	0.14:0.71
p<0.1	1.18	0.005	1.64	<0.001	1.15	0.005	0.34	0.13:0.64
p<0.2	1.15	0.021	1.64	<0.001	1.12	0.014	0.29	0.10:0.56
p<0.3	1.13	0.037	1.64	<0.001	1.14	0.005	0.30	0.11:0.56
p<0.4	1.13	0.035	1.64	<0.001	1.13	0.010	0.28	0.08:0.55
p<0.5	1.11	0.081	1.64	<0.001	1.12	0.018	0.24	0.04:0.50
Severe MDD (956 cases and 340 controls)								
p<0.001	1.02	0.805	1.69	<0.001	1.07	0.185	0.09	-0.08:0.28
p<0.01	1.11	0.116	1.69	<0.001	1.11	0.054	0.21	0.02:0.46
p<0.05	1.22	0.002	1.69	<0.001	1.14	0.013	0.37	0.14:0.72
p<0.1	1.2	0.005	1.69	<0.001	1.14	0.008	0.36	0.13:0.69
p<0.2	1.17	0.016	1.69	<0.001	1.13	0.017	0.33	0.10:0.67
p<0.3	1.17	0.017	1.69	<0.001	1.16	0.005	0.36	0.13:0.69
p<0.4	1.17	0.016	1.69	<0.001	1.14	0.009	0.34	0.11:0.70
p<0.5	1.15	0.032	1.69	<0.001	1.14	0.014	0.30	0.07:0.63

Direct effects of the polygenic risk scores (PRS), childhood trauma (CT) and their interaction-effects were estimated in four separate logistic regression models. The effects of the PRS (model 1) and CT (model 2) were estimated in models with age, gender and three principal components as covariates. The interaction effects were estimated in a model additionally including PRS and CT as covariates (model 3 and model 4). The RERI's represent tests for interaction as departure from additivity and were computed by $e^{\beta CT+\beta PRS+\beta PRSxCT}-e^{\beta CT}-e^{\beta PRS}+1$.

Table 3. Interaction between polygenic risk scores (PRS) and four childhood trauma (CT) domains in predicting all MDD risk and direct effects of the four CT domains (1645 cases and 340 controls).

	Emotional neglect (EN)				Psychological abuse (PsA)			
	EN		PRS x EN		PsA		PRS x PsA	
PRS thresholds	OR	P-value	OR	P-value	OR	P-value	OR	P-value
p<0.001	2.57	<0.001	1.10	0.307	2.40	<0.001	1.03	0.809
p<0.01	2.57	<0.001	1.15	0.130	2.40	<0.001	1.16	0.216
p<0.05	2.57	<0.001	1.15	0.128	2.40	<0.001	1.37	0.006
p<0.1	2.57	<0.001	1.18	0.069	2.40	<0.001	1.36	0.007
p<0.2	2.57	<0.001	1.21	0.032	2.40	<0.001	1.29	0.025
p<0.3	2.57	<0.001	1.22	0.027	2.40	<0.001	1.36	0.007
p<0.4	2.57	<0.001	1.21	0.035	2.40	<0.001	1.32	0.016
p<0.5	2.57	<0.001	1.19	0.056	2.40	<0.001	1.31	0.018
	Physical abuse (PhA)				Sexual abuse (SA)			
	PhA		PRS x PhA		SA		PRS x SA	
PRS thresholds	OR	P-value	OR	P-value	OR	P-value	OR	P-value
p<0.001	2.90	<0.001	1.24	0.297	2.19	<0.001	1.12	0.503
p<0.01	2.90	<0.001	1.30	0.215	2.19	<0.001	1.05	0.785
p<0.05	2.90	<0.001	1.42	0.081	2.19	<0.001	1.08	0.653
p<0.1	2.90	<0.001	1.36	0.137	2.19	<0.001	1.11	0.508
p<0.2	2.90	<0.001	1.23	0.288	2.19	<0.001	1.10	0.561
p<0.3	2.90	<0.001	1.18	0.381	2.19	<0.001	1.12	0.460
p<0.4	2.90	<0.001	1.11	0.577	2.19	<0.001	1.14	0.393
p<0.5	2.90	<0.001	1.10	0.609	2.19	<0.001	1.16	0.356

Four childhood trauma domains (ranging from 0 to 2) were tested for their direct-effects and interaction-effects with polygenic risk scores on all MDD risk, using logistic regression analyses adjusted for age, gender and three principal components. The main effect of the polygenic risk scores is displayed in Table 2 (model 1).

Table 4. Proportion of variation in risk on all MDD explained by childhood trauma (CT), polygenic risk scores (PRS) and their interaction effect.

	Nagelkerke's R^2 (in percentages)			Lee's R^2
PRS thresholds	**CT**	**PRS**	**CTxPRS**	**PRS**
p<0.001	12.68	<0.00	0.09	0.02
p<0.01	12.68	0.30	0.24	0.41
p<0.05	12.68	0.90	0.53	1.26
p<0.1	12.68	0.66	0.60	0.94
p<0.2	12.68	0.49	0.46	0.64
p<0.3	12.68	0.40	0.59	0.54
p<0.4	12.68	0.40	0.49	0.56
p<0.5	12.68	0.26	0.42	0.39

To estimate proportions of variation in risk on all MDD explained by childhood trauma (CT), polygenic risk scores (PRS) and their interaction (CTxPRS), Nagelkerke's R^2 were compared: between the model with only covariates and the model additionally including CT (R^2 of CT); between the model with covariates and CT and the model additionally including PRS (R^2 of PRS); and between the model with covariates, CT and PRS and the model additionally inclu-ding PRSxCT (R^2 of CTxPRS). Age, gender and three principal components were included as covariates. Lee's proposed estimate of R^2 was computed[30] using a Dutch lifetime prevalence of MDD of 18.7%.[31]

DISCUSSION

This is the first study that focuses on polygenic risk scores to test for gene-by-environment interaction in major depressive disorder (MDD). Within our sample we found increased effects of polygenic risk scores on MDD in the presence of childhood trauma, with evidence for interaction as departure from both multiplicativity and additivity. These interaction-effects were comparable in predicting all MDD risk and severe (chronic or recurrent) MDD risk, although effects were slightly larger in the latter. The interaction-effects were driven by all of the four domains included in the childhood trauma measure (emotional neglect, psychological abuse, physical abuse and sexual abuse). We found that the proportion of variation in all MDD risk explained by the interaction effects was comparable to the proportion explained by the polygenic risk scores (both ~0.5%).

Thus far, polygenic information has not been taken into account in research on gene-by-environment interaction in MDD, but there has been ongoing research for interaction with candidate genes. The motivation for research on gene-by-environment interaction in MDD is found in its contribution to understanding the complex etiology of MDD,[33] and its possibility to select environmental conditions with increased genetic effects. Nevertheless, research

on candidate genes has led to rather contradictory results: in research on the well-known serotonin transporter gene (5-HTTLPR) even meta-analyses differ in their conclusions[5,6,34] with concerns about publications bias.[1] However, because genetic effects on MDD are polygenic in nature,[11,12] we argued that gene-by-environment interaction should be tested with polygenic information.

The interaction effect thus found within our sample between polygenic risk scores and childhood trauma in MDD has two implications. The first implication is that polygenic risk scores have increased effects in the presence of childhood trauma, as illustrated in Figure 1, which indicates that research on direct genetic effects potentially gains power by focusing on individuals exposed to childhood trauma. Therefore, if numbers would allow it would be very useful to perform a genome-wide association study within, for example, the collaborative Psychiatric Genomics Consortium[15] in individuals who experienced childhood trauma. Because interaction-effects are symmetrical, we could, however, also have illustrated that childhood trauma has more impact in individuals with high polygenic risk scores. Thus, the second implication is that individuals with high polygenic risk scores are more vulnerable for the effects of childhood trauma, which has potential clinical relevance, for example in profiling of MDD, but also in possible future prevention programs. When replicated in independent samples, the interaction effect found might add a modest but important piece to the complex puzzle of MDD's etiology.

The direct effects of the polygenic risk scores and childhood trauma in predicting MDD risk in our sample are in line with previous findings. The proportion of variation in MDD explained by the polygenic risk scores (R^2 ~0.5%) was in line with previous findings of Demirkan and colleagues[11] and the Psychiatric GWAS Consortium.[15] Although Nagelkerke's R^2 could have suffered from ascertainment bias because of the large proportion of participants with MDD in our sample, its estimates were of the same magnitude as Lee's estimates of R^2, indicating that ascertainment did not largely affect our results.[30] The choice of the SNP p-value cut-off in the discovery sample tends to be arbitrary, which is why we presented results for eight different cut-offs in this study, and results were comparable for cut-offs larger than 0.05. In general, we anticipate that lower cut-offs are preferable over higher cut-offs when the discovery sample size increases and SNP effects can be found with more certainty. The impact of childhood trauma in predicting MDD risk in our sample is also in line with previous findings, for example those of MacMillan[9] and De Graaf.[7] Furthermore, evidence for interaction was found as departure from both multiplicativity and

additivity, the latter of which has been argued to be more in line with biological interaction.[29,35]

The impact of polygenic risk scores on MDD could have been studied in several environmental conditions, but we hypothesized that the presence of childhood trauma is a likely candidate. The presence of childhood trauma showed most consistent results in previous research on interaction with candidate genes,[5] and it is a severe form of stress with a large and life-long impact, resulting in a large main effect on MDD prevalence.[8,10] Furthermore, childhood trauma generally occurs before the onset of MDD (in our sample 84.7% of the MDD subjects had their first episode after age 16), thereby largely excluding the potential source of bias from reciprocal causation, i.e. when MDD results in environmental stress.[36] In our study, childhood trauma was assessed with the Childhood Trauma Interview (CTI), which is a well-established instrument that has shown to predict onset of depressive and anxiety disorders[7,20] as well as an enduring impact on structural and functional brain abnormalities.[21,22] Our finding that childhood trauma increases the effects of polygenic risk scores on MDD fits with a recent review of Teicher and Samson, which indicates that MDD patients with childhood trauma have more severe mood, neurovegetative and endogenous symptoms, and more comorbidities and psychotic features than MDD patients without childhood trauma.[10]

The approach applied in this study, to test for gene-by-environment interaction with polygenic risk scores, has both advantages and disadvantages. This is the first study to apply this approach to MDD, but Meyers and colleagues have applied it to smoking behaviour before. They observed interaction effects on smoking behaviour between polygenic risk scores for smoking and the number of traumatic events experienced as well as for polygenic risk scores and neighborhood social cohesion (effective n=399).[37] An advantage of the polygenic risk scores -approach is that polygenic risk scores are based on genome-wide SNP data, but result in a one-dimensional summary measure, with corresponding requirements of significance ($p<0.05$). Consequently, the sample size of the target sample can be much smaller than in GWAS studies testing SNPs independently. A disadvantage is, however, that particular aspects of the multi-dimensional polygenic information is lost, which could lead to biased results, for example when certain SNPs show increased effects on MDD in the presence of childhood trauma while other SNPs show decreased effects on MDD in the presence of childhood trauma. If this hypothetical situation would occur, both interaction-effects would be leveled out in tests with the one-dimensional polygenic risk scores summary measure. Nevertheless, at the present time sample sizes are

insufficient to examine the impact of many SNPs in GxE studies and, therefore, studying polygenic risk scores seems an elegant approach for testing the general hypothesis of the existence of gene-by-environment interaction.

Our study has several strengths. First, it was based on DSM-IV based diagnoses of MDD, which ensures we studied participants with clinically relevant MDD. Second, controls were carefully screened for any lifetime psychiatric diagnosis. Third, childhood trauma was assessed in a face-to-face interview by specially trained clinical staff. Fourth, polygenic risk scores were based on a large and independent discovery sample, which adds to the accuracy of the polygenic risk scores. However, there are also some limitations, including potential recall bias of childhood trauma by the mood of participants with MDD. The number of controls in our sample was rather limited, but we carefully checked for ascertainment bias and found none. Even though controls were carefully screened for MDD, they could potentially develop MDD later in life, especially because MDD has a high prevalence of approximately 20 percent.

To conclude, we show that the effect of polygenic risk scores on MDD is increased in the presence of childhood trauma in our sample. Our finding implicates that power in research on direct genetic effects is larger in the presence of childhood trauma, but it also implicates that subjects with high polygenic risk scores form a potential group for MDD prevention, because of their increased vulnerability for childhood trauma. Future research should be conducted to replicate our finding, especially in the light of the inconclusive findings in research on interaction in MDD thus far. In addition, future research could also be designed to test interaction with polygenic information applying different techniques. A possible technique to apply could be genome-wide complex trait analyses (GCTA) to test for interaction with the genetic relationship matrix.[38] The present study was underpowered to conduct such analyses,[39] but future efforts combining data from several independent GWAS cohorts could potentially reach power to test for interaction with GCTA. Further research is required, but our results suggest that gene-by-environment interaction plays a considerable role in the polygenic effects on MDD.

REFERENCES

1. Duncan, L.E., and Keller, M.C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am. J. Psychiatry *168*, 1041–1049.

2. Klengel, T., and Binder, E.B. (2013). Gene-environment interactions in major depressive disorder. Can. J. Psychiatry. *58*, 76–83.

3. Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science *301*, 386–389.
4. Fergusson, D.M., Horwood, L.J., Miller, A.L., and Kennedy, M. a (2011). Life stress, 5-HTTLPR and mental disorder: findings from a 30-year longitudinal study. Br. J. Psychiatry *198*, 129–135.
5. Karg, K., Burmeister, M., Shedden, K., and Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. Arch. Gen. Psychiatry *68*, 444–454.
6. Risch, N., Herrell, R., Lehner, T., Liang, K.-Y., Eaves, L., Hoh, J., Griem, A., Kovacs, M., Ott, J., and Merikangas, K.R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. JAMA *301*, 2462–2471.
7. Graaf, R. de, Bijl, R. V, Smit, F., Vollebergh, W.M., and Spijker, J. (2002). Risk factors for 12-month comorbidity of mood, anxiety, and substance use disorders: findings from the Netherlands Mental Health Survey and Incidence Study. Am. J. Psychiatry *159*, 620–629.
8. Hovens, J.G.F.M., Wiersma, J.E., Giltay, E.J., van Oppen, P., Spinhoven, P., Penninx, B.W.J.H., and Zitman, F.G. (2010). Childhood life events and childhood trauma in adult patients with depressive, anxiety and comorbid disorders vs. controls. Acta Psychiatr. Scand. *122*, 66–74.
9. MacMillan, H.L., Fleming, J.E., Streiner, D.L., Lin, E., Boyle, M.H., Jamieson, E., Duku, E.K., Walsh, C. a, Wong, M.Y., and Beardslee, W.R. (2001). Childhood abuse and lifetime psychopathology in a community sample. Am. J. Psychiatry *158*, 1878–1883.
10. Teicher, M.H., and Samson, J. a (2013). Childhood maltreatment and psychopathology: A case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. Am. J. Psychiatry *170*, 1114–1133.
11. Demirkan, A., Penninx, B.W.J.H., Hek, K., Wray, N.R., Amin, N., Aulchenko, Y.S., Dyck van, R., Geus de, E.J.C., Hofman, A., Uitterlinden, a G., et al. (2011). Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol. Psychiatry *16*, 773–783.
12. Lubke, G.H., Hottenga, J.J., Walters, R., Laurin, C., Geus, E.J.C. De, Willemsen, G., Smit, J.H., Middeldorp, C.M., Penninx, B.W.J.H., Vink, J.M., et al. (2012). Estimating the Genetic Variance of Major Depressive Disorder Due to All Single Nucleotide Polymorphisms. Biol. Psychiatry *72*, 707–709.
13. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C.,

Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.

14. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379.

15. Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H.R., Boomsma, D.I., Cichon, S., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.

16. Yang, J., Manolio, T. a, Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. *43*, 519–525.

17. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S. V, Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.

18. Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W.J., Assendelft, W.J.J., et al. (2008). The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int. J. Methods Psychiatr. Res. *17*, 121–140.

19. Penninx, B.W.J.H., Nolen, W. a, Lamers, F., Zitman, F.G., Smit, J.H., Spinhoven, P., Cuijpers, P., de Jong, P.J., van Marwijk, H.W.J., van der Meer, K., et al. (2011). Two-year course of depressive and anxiety disorders: results from the Netherlands Study of Depression and Anxiety (NESDA). J. Affect. Disord. *133*, 76–85.

20. Peyrot, W.J., Middeldorp, C.M., Jansen, R., Smit, J.H., de Geus, E.J.C., Hottenga, J.-J., Willemsen, G., Vink, J.M., Virding, S., Barragan, I., et al. (2013). Strong effects of environmental factors on prevalence and course of major depressive disorder are not moderated by 5-HTTLPR polymorphisms in a large Dutch sample. J. Affect. Disord. *146*, 91–99.

21. van Harmelen, A.-L., van Tol, M.-J., Demenescu, L.R., van der Wee, N.J. a, Veltman, D.J., Aleman, A., van Buchem, M. a, Spinhoven, P., Penninx, B.W.J.H., and Elzinga, B.M. (2013). Enhanced amygdala reactivity to emotional faces in adults reporting childhood emotional maltreatment. Soc. Cogn. Affect. Neurosci. *8*, 362–369.

22. van Harmelen, A.-L., van Tol, M.-J., van der Wee, N.J. a, Veltman, D.J., Aleman, A., Spinhoven, P., van Buchem, M. a, Zitman, F.G., Penninx, B.W.J.H., and Elzinga,

B.M. (2010). Reduced medial prefrontal cortex volume in adults reporting childhood emotional maltreatment. Biol. Psychiatry *68*, 832–838.
23. Bernstein, D.P., Stein, J. a, Newcomb, M.D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., Desmond, D., et al. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. Child Abuse Negl. *27*, 169–190.
24. Boomsma, D.I., Willemsen, G., Sullivan, P.F., Heutink, P., Meijer, P., Sondervan, D., Kluft, C., Smit, G., Nolen, W. a, Zitman, F.G., et al. (2008). Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur. J. Hum. Genet. *16*, 335–342.
25. Abdellaoui, A., Hottenga, J.-J., de Knijff, P., Nivard, M.G., Xiao, X., Scheet, P., Brooks, A., Ehli, E. a, Hu, Y., Davies, G.E., et al. (2013). Population structure, migration, and diversifying selection in the Netherlands. Eur. J. Hum. Genet. *21*, 1277–1285.
26. Sullivan, P.F., Geus, E.J.C. De, Willemsen, G., James, M.R., Smit, J.H., Zandbelt, T., Arolt, V., Baune, B.T., Blackwood, D., Cichon, S., et al. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. Mol. Psychiatry *14*, 359–375.
27. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
28. Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. Twin Res. *5*, 554–571.
29. Knol, M.J., van der Tweel, I., Grobbee, D.E., Numans, M.E., and Geerlings, M.I. (2007). Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int. J. Epidemiol. *36*, 1111–1118.
30. Lee, S.H., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2012). A Better Coefficient of Determination for Genetic Profile Analysis. Genet. Epidemiol. *36*, 214–224.
31. Graaf, R. de, Have, M. ten, Gool, C. van, and Dorsselaer, S. van (2012). Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. Soc. Psychiatry Psychiatr. Epidemiol. *47*, 203–213.
32. R Core Team (2015). R: A Language and Environment for Statistical Computing.
33. Manolio, T. a, Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L. a, Hunter,

D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.
34. Munafò, M.R., Durrant, C., Lewis, G., and Flint, J. (2009). Gene X environment interactions at the serotonin transporter locus. Biol. Psychiatry *65*, 211–219.
35. Uher, R., Caspi, A., Houts, R., Sugden, K., Williams, B., Poulton, R., and Moffitt, T.E. (2011). Serotonin transporter gene moderates childhood maltreatment's effects on persistent but not single-episode depression: replications and implications for resolving inconsistent results. J. Affect. Disord. *135*, 56–65.
36. Middeldorp, C.M., Cath, D.C., Beem, a L., Willemsen, G., and Boomsma, D.I. (2008). Life events, anxious depression and personality: a prospective and genetic study. Psychol. Med. *38*, 1557–1565.
37. Meyers, J.L., Cerdá, M., Galea, S., Keyes, K.M., Aiello, a E., Uddin, M., Wildman, D.E., and Koenen, K.C. (2013). Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit Neighborhood Health Study. Transl. Psychiatry *3*, e290.
38. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
39. Visscher, P.M., Hemani, G., Vinkhuyzen, A.E., Chen, G.-B., Lee, S.H., Wray, N.R., Goddard, M.E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet. *10*, e1004269.

Chapter 6

Disease and polygenic architecture: avoid trio-design and appropriately account for unscreened controls for common disease

Wouter J. Peyrot
Dorret I. Boomsma
Brenda W.J.H. Penninx
Naomi R. Wray

Published in *The American Journal of Human Genetics 2016, 98:382-391*

Paper at publisher's website:
http://www.cell.com/ajhg/fulltext/S0002-9297(15)00512-1
DOI: http://dx.doi.org/10.1016/j.ajhg.2015.12.017

ABSTRACT

Genome-wide association studies (GWAS) are an optimal design for discovery of disease risk loci for diseases whose underlying genetic architecture includes many common causal loci of small effect (a polygenic architecture). We consider two designs, which deserve careful consideration if the true underlying genetic architecture of the trait is polygenic: parent-offspring trios and unscreened controls. We assess these designs in terms of quantification of the total contribution of genome-wide genetic markers to disease risk (SNP-heritability) and power to detect an associated risk allele. First, we show that trio-designs should be avoided when: i) the disease has a lifetime risk > 1%; ii) trio probands are ascertained from families with more than one affected sibling under which scenario the SNP-heritability can drop by over 50%, and power can drop as much as from 0.9 to 0.15 for a sample of 20,000 subjects; iii) assortative mating occurs (spouse correlation of the underlying liability to the disorder) which decreases the SNP-heritability but not the power to detect a single locus in the trio design. Some studies use unscreened rather than screened controls as these can be easier to collect; we show that the estimated SNP heritability should then be scaled by dividing by $(1 - K * u)^2$ for disorders with population prevalence *K* and proportion of unscreened controls *u*. When omitting to scale appropriately, the SNP-heritability of, for example, major depressive disorder (*K* = 0.15) would be underestimated by 28% when none of the controls are screened.

INTRODUCTION

Optimal experimental design of genetic studies of disease for discovery of associated loci depends on the underlying genetic architecture of the trait. Although the true genetic architecture of the trait is usually not known, different experimental designs aim at exposing causal loci of differing population frequencies. For example, the optimal experimental design to detect *de novo* mutations is a trio design in which affected probands and their parents are genotyped.[1] In contrast, genome-wide association studies (GWAS) are an optimal design for a genetic architecture that includes many common causal loci of small effect (a polygenic architecture). Here, we consider two designs of GWAS, which we show deserve careful consideration: designs based on parent-offspring trios and on unscreened controls. We assess these designs in terms of quantification of the total contribution to disease risk of genome-wide genetic markers, via estimation of so-called SNP-heritability,[2] and the power to detect an associated risk allele.

Our study is motivated by experiences with GWAS designs for psychiatric disorders, but our results are parameterized based on baseline disease risk and heritability, and are, therefore, applicable to the full range of diseases and disorders with a polygenic genetic architecture of underlying risk. For psychiatric disorders, GWAS have had variable success in detecting genome-wide significant common single nucleotide polymorphisms (SNPs). On the one hand, 108 significant loci were recently found for schizophrenia (SCZ [MIM 181500]) in a sample comprising 36,989 cases,[3] whereas only 2 loci were found in one study on Major Depressive Disorder (MDD [MIM 608516])[4] but none in another,[5] no loci for attention-deficit/ hyperactivity disorder (ADHD [MIM 143465]),[6] and only single-study genome-wide significant loci for autism spectrum disorder (ASD [MIM 209850]).[7–9] Sample size is pivotal in explaining this discrepancy, since much smaller numbers of cases were included for MDD (5303 and 9240 respectively), ADHD (2960), and ASD (2705, 1984, and 1553 respectively) than for SCZ. Other contributing factors have, nevertheless, been proposed, such as the impact of *de novo* mutations in ASD[10,11] (although these are only expected to explain a small proportion variation),[12] lower family-based heritability of MDD (~0.4 versus ~0.8 for SCZ, ASD and ADHD, assuming a similar genetic architecture between disorders)[13], higher prevalence and greater heterogeneity of MDD.[14] Here, we show that the trio design, which is regularly applied in ASD and ADHD, and use of unscreened controls deserves careful consideration in the context of an underlying polygenic architecture, which is an important consideration for design of future studies which strive to increase sample size.[15]

The impact of trio-design and the use of unscreened controls on the SNP-heritability have, to the best of our knowledge, not yet been described, probably because the methods for estimation of SNP-heritability were only developed in recent years.[16,17] The impact on the power to detect a single locus has, on the other hand, been studied in the pre-GWAS era of candidate genes,[18–21] but we could find no clear-cut comparison of the power to detect an associated risk allele with trio-studies versus screened control studies, and we will therefore also give an overview of these differences. We investigate the trio-design and the use of unscreened controls by analytical derivation followed by simulation studies to validate theory. Assortative mating (correlation in liability between spouses) is included in our trio design analyses, because this has been reported for a range of psychiatric disorders.[22–25] For example, a spouse-correlation on the Social Responsiveness Scale (a quantitative measure of autistic traits) of 0.29 has been reported in a population sample[23] and of 0.26 in parents of ASD probands.[22] For ADHD a spouse correlation of 0.11 on the ADHD-index in population samples has been reported.[25] In trio designs genotypes of proband cases are compared to genotypes of pseudocontrols (the non-transmitted parental alleles).

SNP-HERITABILITY CALCULATIONS

The SNP-heritability estimates the total proportion of variance tagged by common SNPs from genome-wide association study.[2,16] If samples with GWAS data are population samples, then the variance estimated on the observed scale ($\hat{h}_o^2$) is expressed with the Robertson's transformation on the liability scale ($\hat{h}_l^2$) as[26]

$$\hat{h}_l^2 = \hat{h}_o^2 \frac{K(1-K)}{z^2}. \qquad \text{[Equation 1]}$$

Quantification on the liability scale is most interpretable as it allows direct comparisons to estimates of heritability from family data that are reported on this scale, and to estimates of variance explained by individual genome-wide significant loci. However, usually GWAS samples are oversampled for cases compared to population samples and the transformation of proportion of variance attributable to SNPs estimated from case-control data ($\hat{h}_{occ}^2$) must also account for the proportion of cases in the sample P by[2,27]

$$\hat{h}_l^2 = \hat{h}_{occ}^2 \frac{K^2(1-K)^2}{P(1-P)z^2}, \qquad \text{[Equation 2]}$$

which reduces to Equation 1 when the sample is a population sample and $P = K$. However, these transformations assume that controls are screened. To account for controls being unscreened, we define F as the proportion of falsely classified controls, $F = \frac{N_{false\ controls}}{N_{false\ controls}+N_{true\ controls}} = \frac{N_{false\ controls}}{N_{controls}}$. We closely followed the derivations of Golan at al (paragraphs 1.2 and 1.3 of their Supplemental Materials)[27] to derive an updated equation (Table S1) validated by simulation (Table S2)

$$\hat{h}_l^2 = \hat{h}_{occ}^2 \frac{K^2(1-K)^2}{P(1-P)(1-F)^2 z^2}, \qquad \text{[Equation 3]}$$

which reduces to Equation 2 when $F = 0$ and controls are screened. If a proportion u of the controls are a random sample from the population then one can assume that $F \approx Ku$. Therefore, if it is unknown if controls are screened or not, the potential underestimation when all controls are unscreened ($u = 1$) of the SNP-heritability $\hat{h}_l^2$ estimated from the standard Equation 2 can be assessed as $\hat{h}_l^2(1-K)^2$ and thus depends on baseline risk K. In trio designs where probands are ascertained randomly, the pseudocontrols are equivalent to unscreened controls under a polygenic model (Figure S1).

For the trio-design, the SNP-heritability was derived for a disease parameterized with normally distributed phenotypic (l) and genetic (G) liabilities with means $E(l) = E(G) = 0$ and variances $V_l = 1$ and $V_G = h_l^2$, the true heritability on the liability scale in the parental generation.[28] Under the liability-threshold model, individuals are deemed affected when their liability l is larger than threshold T such that $P(l > T|l \sim N(0,1)) = K$. Parental assortative mating was taken into account by parameterizing a spouse liability correlation of ρ_l and genetic correlation of $\rho_G = h_l^2\rho_l$.[28] The $E(G)$ of proband cases and pseudocontrols were derived by considering the variance-covariance matrix of l and G of individuals that could contribute to a trio design (proband, sibling, mother, father, pseudocontrol). To account for the affected proband, the variance-covariance matrix of random families was conditioned on the proband being affected by accounting for the reduction in variance as result of the Bulmer effect[29] in related individuals described by Tallis.[30] To account for a second affected sibling, the variance-covariance matrix was further conditioned on the sibling also being affected. Details of these derivations are provided in the Supplemental Methods, and were validated with a simulation study in R (Table S3 & Table S4).[31]

Figure 1 Panel A displays the SNP-heritability assessed from unscreened controls, which is equivalent to estimates from pseudocontrols from random families with at least one affected proband (dotted line Figure 1 Panel A), and screened controls (solid lines Figure 1). While the standard transformation (Equation 2) applied to derive estimates of SNP-heritability on the liability scale ($\hat{h}_l^2$) is expected to give unbiased estimates of the true SNP-heritability when cases are randomly ascertained and controls are screened (Figure 1, Panel A solid line), the transformation underestimates h_l^2 by a factor $(1-K)^2$ when diseases are common (high K) and controls are unscreened or are pseudocontrols (Figure 1 Panel A dashed line). The estimated heritability from the Equation 2 transformation $\hat{h}_l^2$ severely underestimates h_l^2 when data result from a trio design with probands ascertained from multiplex families (Figure 1 Panel B dotted line), for example, $\hat{h}_l^2 = 0.31$ for $K = 0.05$ and $h_l^2 = 0.5$, since the mean liability of pseudocontrols is greater than the average in the population and so the contrast in genetic values between cases and pseudocontrols is less than between cases and screened controls (Table 1 Panel B), which is not fully compensated by the fact that cases from multiplex families have higher mean liability than randomly selected cases (Table 1 Panel A). In contrast, when cases are selected from multiplex families and controls are screened controls the estimated SNP-heritability based on the standard transformation is an overestimate of h_l^2 (for example, $\hat{h}_l^2 = 0.75$ for $K = 0.05$ and $h_l^2 = 0.5$). When controls are unscreened, the SNP-heritability is found in between the SNP-heritability from screened and pseudocontrols (dashed lines Figure 1), when SNP-heritabilities are estimated using equation 2. In the context of assortative mating, a trio design comparison of probands to pseudocontrols yield decreased $\hat{h}_l^2$ (Figure 1 Panel C; Table 1 Panel C, spouse correlation $\rho_l = 0.3$). Again, comparing the probands to screened controls (from the offspring generation) does in fact overestimate the heritability in the parent generation h_l^2; this is, however, a well-known consequence of assortative mating and is not restricted to the trio-design ($V_{G,offspring} = V_{G,parents} + \frac{1}{2}\rho_{G,parents}V_{G,parents}$).[29] The most pronounced difference between screened controls and pseudocontrols is found for probands with an additional affected sibling in the context of parental assortative mating (Figure 1 Panel D; Table 1 Panel D).

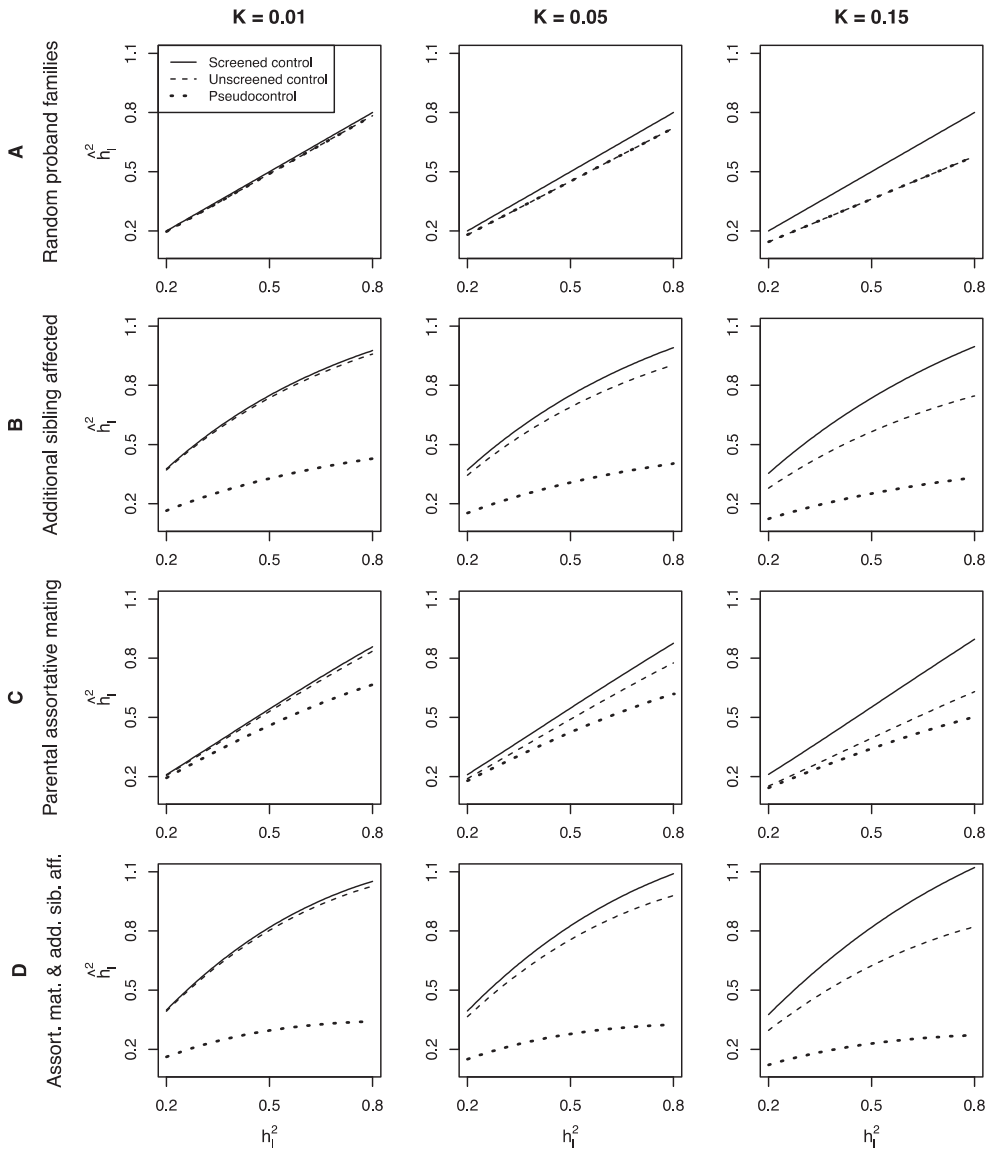


Figure 1. Relationship between the True SNP Heritability and Its Estimates Based on the Standard Transformation with Equation 2 from Trio Data, Screened Controls, and Unscreened Controls. The SNP-heritability $\hat{h}_l^2$ that would be estimated based on the standard liability transformation equation (Equation 2) for GWAS studies using pseudocontrols (dotted lines), unscreened controls (dashed lines) and screened controls (solid lines) compared to the true parental SNP-heritability h_l^2 for designs based on randomly ascertained proband families (Panel A), families with an additional affected sibling (Panel B), in the context of parental assortative mating with a correlation on the liability scale of $\rho_l = 0.3$ (Panel C), and families with an additional affected sibling in the context of parental assortative mating (Panel D) for disorders with lifetime risk $K =$

$0.01, K = 0.05$ and $K = 0.15$. The pseudocontrols of random proband families are equivalent to unscreened controls (dashed and dotted lines Panel A overlap), and the slope of these lines are defined by $(1-K)^2$, i.e. the underestimation of $\hat{h}_l^2$ when mistakenly applying Equation 2 rather than Equation 3 to transform the heritability on the observed scale to the liability scale when none of the controls are screened.

Table 1. Mean genetic liabilities and SNP-heritability estimated from the standard transformation with Equation 2 from GWAS using trio-design, screened controls, or unscreened controls for actual parental heritability 0.5

		Mean genetic liability (E(G))						
			Control			$\hat{h}_l^2$ assessed from proband		
K	h_l^2 parents	Case	Screened	Unscreened	Pseudo	Screened	Unscreened	Pseudo
A. Random proband families								
0.01	0.5	1.333	-0.013	0.000	0.000	0.500	0.490	0.490
0.05	0.5	1.031	-0.054	0.000	0.000	0.500	0.451	0.451
0.15	0.5	0.777	-0.137	0.000	0.000	0.500	0.361	0.361
B. Additional sibling affected								
0.01	0.5	1.634	-0.013	0.000	0.543	0.749	0.736	0.328
0.05	0.5	1.275	-0.054	0.000	0.424	0.750	0.690	0.307
0.15	0.5	0.972	-0.137	0.000	0.323	0.735	0.565	0.251
C. Parental assortative mating								
0.01	0.5	1.386	-0.016	0.000	0.097	0.542	0.530	0.459
0.05	0.5	1.075	-0.060	0.000	0.075	0.547	0.490	0.424
0.15	0.5	0.812	-0.148	0.000	0.057	0.552	0.395	0.341
D. Additional sibling affected and parental assortative mating								
0.01	0.5	1.706	-0.016	0.000	0.670	0.818	0.803	0.296
0.05	0.5	1.335	-0.060	0.000	0.525	0.826	0.756	0.278
0.15	0.5	1.021	-0.148	0.000	0.402	0.818	0.624	0.230

The mean genetic liabilities *E(G)* are displayed for proband cases, unrelated screened controls, unrelated unscreened controls, and their pseudocontrols as well as the SNP-heritability $\hat{h}_l^2$ estimated from Equation 2 from comparing cases to these three sets of controls, for different parameterization of baseline disease risk *K* and a fixed underlying heritability of $h_l^2=0.5$. The proband cases are parameterized in line with Figure 1 to be selected from random proband families (Panel A), families with an additional affected sibling (Panel B), families in the context of parental assortative mating (Panel C), and families with an additional affected sibling in the context of assortative mating (Panel D) respectively.

POWER CALCULATIONS

The power to detect an associated risk allele in a case-control association test follows from the non-centrality parameter NCP of the X^2 test-statistic. This NCP is expressed in terms of sample size N, proportion of cases in the study v, the allele frequency in cases p_{case}, the allele frequency in controls $p_{control}$, and the mean allele frequency in the sample $\bar{p} = vp_{case} + (1-v)p_{control}$ as

$$NCP = \frac{(p_{case} - p_{control})^2}{\bar{p}(1-\bar{p})\left(\frac{1}{2N*v} + \frac{1}{2N*(1-v)}\right)} \qquad \text{[Equation 4]}$$

and the power as $P\left(x > \sqrt{NCP} + x_T \mid z{\sim}N(0,1)\right)$, where x_T is the z-value quantile-function of the standard normal distribution for the desired significance threshold, here set at $\alpha = 5 * 10^{-8}$ $(x_T = -5.45)$. The power of different experimental designs is reflected in the appropriate expressions of p_{case} and $p_{control}$. We parameterize a disease with a baseline lifetime disease risk K, a di-allelic locus with risk allele frequency $P(B) = p$, non-risk allele frequency $P(b) = q = 1 - p$, relative risk of heterozygotes $RR_{Bb} = P(Disease|Bb)/P(Disease|bb)$, and relative risk of the homozygotes $RR_{BB} = P(Disease|BB)/P(Disease|bb)$.[32,33] When controls are screened, power follows from $p_{case} = k_{bb}RR_{Bb}p\,(1 + p(RR_{Bb} - 1))/K$, where $k_{bb} = P(Disease|\mathrm{bb}) = K/(q^2 + 2pqRR_{Bb} + p^2RR_{BB})$, and $p_{control} = ((1 - k_{bb}RR_{Bb})p(1-p) + (1 - k_{bb}RR_{BB})p^2)/(1-K)$,[33] which agrees with the *Genetic Power Calculator* of Purcell et al.[34] When controls are unscreened, the power of an association study is expressed by Equation 4 with $p_{control} = p$. For the trio-design, power was assessed by substituting in Equation 4 the allele frequency in proband cases and pseudocontrols (the non-transmitted alleles of the parents). When trios are ascertained from families with an additional affected sibling or when there is assortative mating, the risk allele frequency in controls can be derived from combined and conditional genotype frequencies of an individual, its affected sibling and its parents. Under assortative mating expressions are dependent on spouse liability correlation $\rho_{liability}$, which results in the correlation between the parental genotypes as $\rho_{locus} = \rho_{liability}h^2_{locus}$.[28] It follows that assortative mating (for e.g. $\rho_{liability} = 0.3$) has no impact on the power to detect a single locus for loci typical of polygenic architecture that explain less than one percent of variation ($\rho_{locus} = 0.3 * 0.01 = 0.003$).[28] When assuming a small RR_{Bb} typical of complex genetic disease and a multiplicative model on the disease scale ($RR_{BB} = RR^2_{Bb}$, implying additively on the underlying risk scale), the variance attributable

to risk locus can be approximated by $h^2_{locus} \approx 2pq(RR_{Bb} - 1)^2/i^2$ with $i = z/K$ the mean liability of cases, and z the height of the standard normal density function at the threshold corresponding to a baseline disease risk K.[33] The expressions to derive allele frequencies in trios are closed but complex (Supplemental Methods) and were validated by simulation (Table S5).

Figure 2 displays the power to detect an associated risk allele for proband cases from (A) random trios with an affected proband, and (B) multiplex trios with an additional affected sibling, when the risk allele has a frequency of $P(B) = p = 0.2$ for disorders with baseline risk $K = 0.01, 0.05$ and 0.15 in a sample of $N = 10{,}000$ trios (proband cases vs pseudo-controls) against RR_{Bb} given an underlying additive effect ($RR_{BB} = RR^2_{Bb}$) (dotted line). Note that pseudocontrols from random families are equivalent to unscreened controls displayed in Figure 2 for a number of 10,000 unscreened controls (dashed line) and 20,000 unscreened controls (dot-dashed line). The solid line on each graph is the power for 10,000 proband cases compared to 10,000 unrelated screened controls. Figure 2A shows that there is little to be gained in screening controls for diseases of lifetime morbid risk < 1%, but for more common disorders (such as ADHD and MDD) there is an important gain in power, which can also be gained by increasing the number of unscreened controls. When trios come from families with an additional affected sibling, the cases have an increased probability of carrying the risk allele and so when matched with screened controls, there is a gain in power compared to random ascertainment of cases (solid line 2B vs solid line 2A). For example, when $p = 0.2$, $RR_{Bb} = 1.2$, then $p_{proband\ B} = 0.248$ and $p_{proband\ A} = 0.231$ respectively (these frequencies do not depend on K). However, when the association study is of cases from multiplex families compared to pseudocontrols there is little gain in power compared to trios based on randomly selected cases (dotted line 2B vs dotted line 2A), because the pseudocontrols also have increased probability of carrying the risk allele ($p_{pseudocontrol\ B} = 0.215$ and $p_{pseudocontrol\ A} = 0.2$). The maximum power difference between using screened and pseudocontrols depends on RR_{Bb}, K, sample size, and whether probands are ascertained randomly (Table 2 Panel A) or from families with an additional affected sibling (Table 2 Panel B), but is found for a sample comprising 20,000 subjects at $RR_{Bb} = 1.11$ and $K = 0.15$ for probands with additional affected siblings, under which scenario a total sample size of $N = 15{,}945$ is needed when controls are screened vs $N = 44{,}574$ for the pseudocontrol trio design respectively to obtain a power of 0.8. For unscreened controls (equivalent to pseudocontrols from random families), the most pronounced decrease in power in a sample of 20,000 subjects is found for a locus with $RR_{Bb} = 1.14$ in disease

with $K = 0.15$ where unscreened controls yield a power of 0.39 and screened controls of 0.74. As expected, the impact of using screened controls is higher for more common disorders. Allele frequencies in probands, pseudocontrols, and screened controls for all Figure 2 scenarios are presented in Figure S2. Furthermore, the power-differences between pseudocontrol and screened control studies are consistent for other risk allele frequencies e.g., $p = 0.6$ (Figure S3), underlying actual recessive (RR_{Bb}=1; Figure S4) and dominant effects ($RR_{Bb} = RR_{BB}$; Figure S5). In addition, to select only trios with unaffected parents has no impact on power of pseudo-control studies, since although the risk allele frequency in pseudocontrols decreases, the frequency in cases decreases proportionally (Figure S6). When unscreened controls are much easier to obtain then screened controls, the loss of power due to not screening can be balanced by increasing the number of unscreened controls, which is illustrated for different numbers of unscreened controls in Figure S7. Note that Equation 4 defines a limit to the power-gain from increasing the number of unscreened controls, but that when increasing number of unscreened controls from 10,000 to 20,000 the loss of power due to not screening is balanced for all scenarios under consideration here. In Figure 2, the additional x-axis is variance explained by the locus, hence the results generalize to many combinations of p and RR_{Bb} that together explain the same locus variance.[32] While association studies have similar power to detect a locus based on RR_{Bb} regardless of baseline disease risk K, the variance explained by a locus is much larger for high K. Therefore, to detect a risk allele that explains the same proportion of genetic variance, a much larger sample size is needed for larger K (Figure 3).

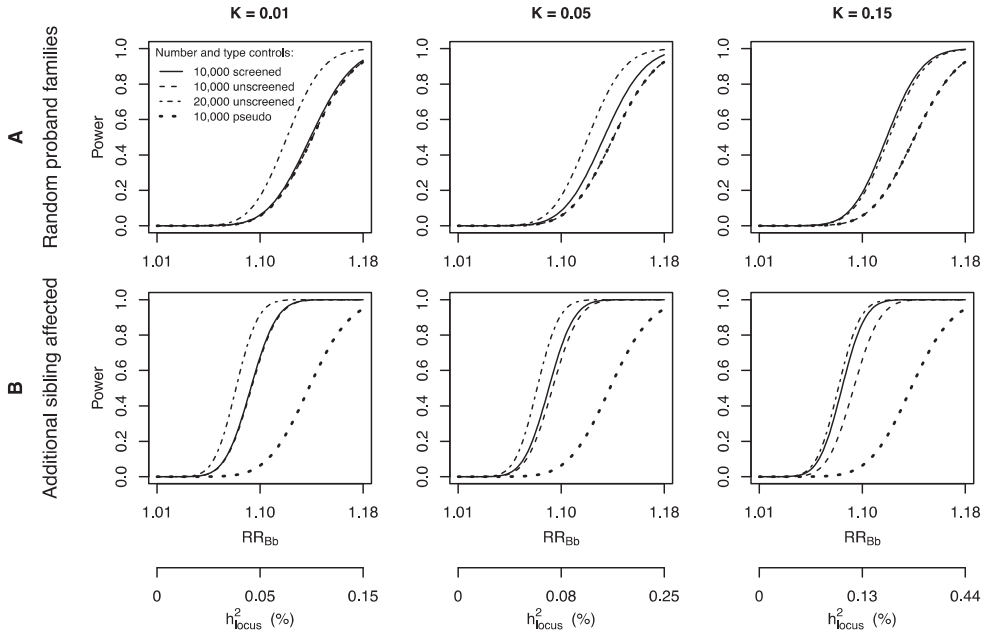


Figure 2. Power to detect a single risk variant in association studies of 10,000 cases that use a trio-design, screened controls, or unscreened controls. Power of association analysis comparing 10,000 probands to 10,000 screened controls (solid line), 10,000 unscreened controls (dashed), 20,000 unscreened controls (dot-dashed), and 10,000 pseudocontrols (dotted) to detect a single associated risk variant for a risk allele with frequency $p = 0.2$, for a baseline disease risk $K = 0.01$, $K = 0.05$ and $K = 0.15$. Power was estimated for risk variants with underlying additive effect ($RR_{BB} = RR_{Bb}^2$) for random ascertainment of probands (Panel A), and probands from families with an additional affected sibling (Panel B). Note that pseudocontrols from random families are equivalent to unscreened controls and that the dotted and dashed line in Panel A overlap. The variation explained on the liability scale was approximated by $h^2_{locus} \approx 2p(1-p)(RR_{Bb}-1)^2/i^2$, where i equals z/K the mean liability of probands, and z the height of the standard normal density function at the threshold corresponding with disease of lifetime risk K.

Table 2. Maximum power difference between trio-design and screened controls studies with 20,000 subjects

		Allele frequencies			Power (N=20,000)		N (power=0.8)	
K	**RR_{Bb}**	**Proband**	**Pseudo**	**Screened**	**Pseudo**	**Screened**	**Pseudo**	**Screened**
A. Proband from random proband families								
0.01	1.147	0.223	0.200	0.200	0.56	0.58	25226	24714
0.05	1.144	0.222	0.200	0.199	0.51	0.63	26327	23712
0.15	1.135	0.221	0.200	0.196	0.39	0.74	29670	21297
B. Proband from families with an additional affected sibling								
0.01	1.115	0.228	0.209	0.200	0.17	0.91	39201	17307
0.05	1.113	0.227	0.209	0.199	0.15	0.92	40533	16923
0.15	1.108	0.226	0.208	0.197	0.11	0.94	44574	15945

The loci with allele frequency p=0.2 from Figure 2 that result in most pronounced decrease in power for pseudocontrol compared to screened control studies for a sample of 10,000 cases and 10,000 controls are displayed in detail. The power difference depends on the baseline disease risk K, its effect size RR_{BP}, and whether the proband cases are from random proband families (A) or families with an additional affected sibling (B) (compare to respectively solid and dotted lines in Figure 2). For these loci, the allele frequencies in proband cases, pseudocontrols and screened controls is displayed, as well as the power given a sample size of N=20,000 (50% cases), and the required sample size to obtain a power of 0.8. Note that pseudo-controls from random families are equivalent to unscreened population controls (A).

DISCUSSION

To summarize our findings, our results generate two important conclusions that trio based samples and unscreened controls for common diseases deserve careful consideration when the underlying genetic architecture is highly polygenic. We have quantified this in two ways, firstly by the underestimation of SNP-heritability through application of the inappropriate transformation equation, and secondly by power calculations of association analysis. We derived a transformation equation for the SNP-heritability that is appropriate for unscreened control samples (Equation 3).

The use of trio designs most commonly occurs for pediatric diseases and disorders in which it is relatively easy to obtain blood samples from parents. Trio designs are needed to detect *de novo* causal mutations,[35] to determine accurately phased haplotypes[35] or to undertake parent-of-origin analyses implied by a hypothesis of parental imprinting.[36] Trio designs have also been considered for detection of gene-environment interaction.[37,38] In the pre-GWAS era trio designs were recommended to protect against potential bias from population

stratification,[1] and although this quality is also sometimes promoted for trio GWAS, with genome-wide SNP data other strategies, such as genomic principal components[39] or mixed model association analysis,[40] appropriately account for population stratification without the need to incur 50% higher costs by genotyping three samples to generate two genomes. While acknowledging the benefits of parent-offspring trios under some experimental paradigms, trio design GWAS have been undertaken without full regard of the implications to power under the genetic architecture implicated by the GWAS paradigm. We draw the following conclusions:

1) If the case probands of trios are ascertained randomly, then the resulting case-pseudocontrol study is equivalent to a case-unscreened control design under a polygenic genetic architecture, and has little impact on the SNP-heritability and power for disorders that are less common, but for more common disorders there is important decrease in SNP-heritability (Figure 1 Panel A) and loss of power (Figure 2 Panel A), inadvertently contributing to the missing heritability problem. For example, in a study on MDD (lifetime risk $K \sim 0.15$)[13,41] where all controls are unscreened, the SNP- heritability (say 0.3) would reduce by a factor of 0.72 (0.72*0.3=0.22) (hence underestimated by 28%) when not accounting for the unscreened controls (i.e. applying Equation 2 rather than Equation 3). For disorders such as MDD, even when controls have been screened it is likely that some controls remain misclassified, as onset can occur throughout the lifetime. Naturally, it should also be noted that when super-controls are used (controls screened to be at the lower end of the liability distribution, for example based on low scores for the personality trait neuroticism in the context of MDD) then SNP-heritability estimates based on the standard transformation equation would be biased upwards. The loss of power due to including unscreened controls can be compensated by increasing the number of controls (Figure 2 & Figure S7), in particular in the context of the continuously decreasing costs for genotyping, but this requires caution when estimating the SNP-heritability, because Equation 3 should then be applied rather than the standard Equation 2.
2) If case probands are ascertained from multiplex families, then the SNP-heritability and power of GWAS are substantially reduced when using pseudocontrols even for less common disorders (see Figure 1 Panel B, and Figure 2 Panel B respectively; modeled on families with two affected siblings). Even in the absence of deliberate ascertainment of multiplex families, studies are likely to be biased by self-ascertainment as parents from multiplex

families may be more concerned with the genetic origins of the disorder. In fact, 43.6% of the 1369 families included in the Autism Genome Project (AGP) had two or more children affected with ASD while counting up to third-degree relatives.[7] However, the proportion of multiplex families is often not reported, as is the case for the family-based studies[42–44] contributing to the last ADHD meta-analysis,[6] which leaves the loss in power due to included multiplex families unknown, but likely. In addition, in a number of families with a first affected child parents will stop to reproduce, so that a second affected child is never observed. Our results are consistent with the simplex versus multiplex and simulation results of Klei et al in analyses of ASD samples.[45]

3) Assortative mating considerably decreases the SNP-heritability assessed from trio-design compared to screened controls also for small K (Figure 1 Panel C), but it does not impact the power to detect a single locus under a polygenic model, because of the small proportions of variation explained by single loci (<1%). Assortative mating is possibly common for psychiatric disorders,[22–25] and needs to be considered when interpreting SNP-heritability in general, and for trio-design in particular. These results and point 2) could explain why lower SNP-based heritabilities were found in the ADHD pseudocontrol samples from the Psychiatric Genomics Consortium compared to case-control samples (see Supplementary Table 5 of Lee et al).[14]

We also take the opportunity to re-emphasize that parameterization of power in terms of genotype relative risk can be misleading since the same RR_{Bb} operating in common disease implies a much higher proportion of variance explained by the locus compared to a locus operating in a less common disease. For example, when the risk allele has frequency $p = 0.2$ and effect size RR_{Bb}= 1.1, the locus explains 0.05%, 0.08% and 0.13% of the variance in phenotypic liability for disorder of frequency $K = 0.01, 0.05,$ and 0.15 respectively. Hence, to detect a locus that explains the same proportion of variance in liability, much larger samples are needed for common disorders (Figure 3). For example, samples of N=4,059 (50% cases 50% screened controls) are needed to detect a locus that explains 0.5% of the variance in liability for a disorder lifetime risk *K* = 0.01 ($RR_{Bb} = 1.39$), compared to samples of N=9,181 when the disorder risk is *K* = 0.15 ($RR_{Bb} = 1.21$). Similar arguments have been used to explain that much larger GWAS samples are needed for MDD compared to schizophrenia.[46]

To the best of our knowledge, the impact of the trio design and use of unscreened controls on the SNP-heritability has not yet been addressed, but our

power analyses build upon a rich literature exploring the characteristics of family-based association studies in the pre-GWAS era. Ferreira et al showed that the trio-based transmission disequilibrium test (TDT) has less power when an additional (non-genotyped) sibling is affected compared to random families with one affected sibling.[18] Li et al,[19] Risch and Teng,[47] and Risch[48] showed that case-control studies are generally more powerful when cases are from families with an additional affected sibling, which is in line with our results (Figure 2 Panel B compared to Panel A). Teng and Risch found that family-based approaches have less power than case-unrelated control strategies for families with multiple affected siblings.[20] Of note, our paper focuses on the pseudocontrol trio design, because this is how the trio design is typically applied in GWAS studies, however the TDT has often been applied for candidate genes and could yield more power for rare disorders as has been indicated by Laird et al.[21] The power to detect a locus with the use of unscreened controls can readily be calculated with the online power calculator of Purcell et al,[34] or the Quanto software from Gauderman.[49] Nevertheless, our study adds also to the current literature on the power to detect a single locus, because we directly compare pseudocontrol-studies to screened control-studies for multiplex families and assortative mating. As expected, there is overall similarity between consequences of design for the power to detect a single risk variant and expected SNP-heritability, but in this study we have formalized these expectations, and also shown that such similarity does not hold when considering assortative mating which impacts on the estimated SNP-heritability but not in power to detect a single risk variant.

To conclude, we advise against the use of trio designs for disorders with a polygenic genetic architecture, such as psychiatric disorders, and we advise careful consideration when using unscreened controls for prevalent disorders, because these designs can results in an underestimated SNP-heritability and decreased power to detect an associated risk allele.

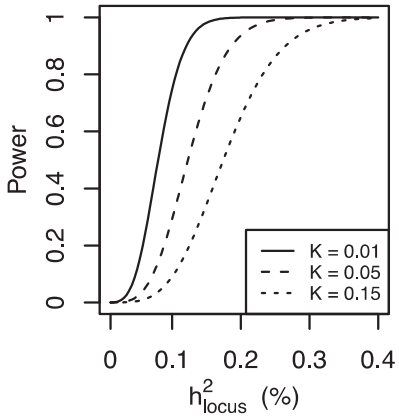


Figure 3. Power to detect an associated locus by the proportion of variation it explains. The power to detect an associated locus depends on the proportion of variation it explains on the liability scale h^2_{locus}, the baseline disease risk K, and is displayed for random case vs screened control. For a locus with the same h^2_{locus} larger sample sizes are required for larger K. h^2_{locus} can be approximated by $2p(1-p)(RR_{Bb}-1)^2/i^2$, where i equals z/K the mean liability of probands, and z the height of the standard normal density function at the threshold corresponding with disease of lifetime risk K. The (complex) relation between allele frequency p, RR_{Bb}, and the non-centrality parameter NCP given h^2_{locus} results in an identical relation between power and h^2_{locus} for varying p.

REFERENCES

1. Spielman, R.S., and Ewens, W.J. (1996). The TDT and other family-based tests for linkage disequilibrium and association. Am. J. Hum. Genet. *59*, 983–989.

2. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am. J. Hum. Genet. *88*, 294–305.

3. Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

4. Cai, N., Bigdeli, T.B., Kretzschmar, W., Li, Y., Liang, J., Song, L., Hu, J., Li, Q., Jin, W., Hu, Z., et al. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature *523*, 588–591.

5. Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H.R., Boomsma, D.I., Cichon, S., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.

6. Neale, B.M., Medland, S.E., Ripke, S., Asherson, P., Franke, B., Lesch, K.-P.,

Faraone, S. V, Nguyen, T.T., Schäfer, H., Holmans, P., et al. (2010). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J. Am. Acad. Child Adolesc. Psychiatry *49*, 884–897.

7. Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. Hum. Mol. Genet. *19*, 4072–4082.

8. Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M.A., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature *459*, 528–533.

9. Weiss, L.A., Arking, D.E., Daly, M.J., and Chakravarti, A. (2009). A genome-wide linkage and association scan reveals novel loci for autism. Nature *461*, 802–808.

10. Ronemus, M., Iossifov, I., Levy, D., and Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. Nat. Rev. Genet. *15*, 133–141.

11. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

12. Gratten, J., Visscher, P.M., Mowry, B.J., and Wray, N.R. (2013). Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. Nat. Genet. *45*, 234–238.

13. Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat. Rev. Genet. *13*, 537–551.

14. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S. V, Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.

15. Sullivan, P.F. (2010). The psychiatric GWAS consortium: big science comes to psychiatry. Neuron *68*, 182–186.

16. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of heritability for human height. Nat. Genet. *42*, 565–569.

17. Chen, G.-B. (2014). Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. Front. Genet. *5*, 107.

18. Ferreira, M.A.R., Sham, P., Daly, M.J., and Purcell, S. (2007). Ascertainment through family history of disease often decreases the power of family-based association studies. Behav. Genet. *37*, 631–636.
19. Li, M., Boehnke, M., and Abecasis, G.R. (2006). Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. Am. J. Hum. Genet. *78*, 778–792.
20. Teng, J., and Risch, N. (1999). The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. II. Individual genotyping. Genome Res. *9*, 234–241.
21. Laird, N.M., and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. Nat. Rev. Genet. *7*, 385–394.
22. Virkud, Y. V, Todd, R.D., Abbacchi, A.M., Zhang, Y., and Constantino, J.N. (2009). Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *150B*, 328–334.
23. Constantino, J.N., and Todd, R.D. (2005). Intergenerational transmission of subthreshold autistic traits in the general population. Biol. Psychiatry *57*, 655–660.
24. Lichtenstein, P., Björk, C., Hultman, C.M., Scolnick, E., Sklar, P., and Sullivan, P.F. (2006). Recurrence risks for schizophrenia in a Swedish national cohort. Psychol. Med. *36*, 1417–1425.
25. Boomsma, D.I., Saviouk, V., Hottenga, J.-J., Distel, M.A., de Moor, M.H.M., Vink, J.M., Geels, L.M., van Beek, J.H.D.A., Bartels, M., de Geus, E.J.C., et al. (2010). Genetic epidemiology of attention deficit hyperactivity disorder (ADHD index) in adults. PLoS One *5*, 1–7.
26. Dempster, E.R., and Lerner, I.M. (1950). Heritability of Threshold Characters. Genetics *35*, 212–236.
27. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: Inferring the contribution of common variants. Proc. Natl. Acad. Sci. U. S. A. *111*, E5272–E5281.
28. Falconer, D., and Mackay, T. (1996). Introduction to quantitative genetics (Essex: Longman).
29. Bulmer, M. (1985). The mathematical theory of quantitative genetics (Oxford: Clarendon press).
30. Tallis, G.M. (1987). Ancestral covariance and the Bulmer effect. Theor. Appl. Genet. *73*, 815–820.
31. R Core Team (2015). R: A Language and Environment for Statistical Computing.
32. Yang, J., Wray, N.R., and Visscher, P.M. (2010). Comparing apples and

oranges: equating the power of case-control and quantitative trait association studies. Genet. Epidemiol. *34*, 254–257.
33. Witte, J.S., Visscher, P.M., and Wray, N.R. (2014). The contribution of genetic variants to disease depends on the ruler. Nat. Rev. Genet. *15*, 765–776.
34. Purcell, S., Cherny, S.S., and Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics *19*, 149–150.
35. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
36. Solter, D. (1988). Differential imprinting and expression of maternal and paternal genomes. Annu. Rev. Genet. *22*, 127–146.
37. Cordell, H.J. (2009). Estimation and testing of gene-environment interactions in family-based association studies. Genomics *93*, 5–9.
38. Gauderman, W.J., Thomas, D.C., Murcray, C.E., Conti, D., Li, D., and Lewinger, J.P. (2010). Efficient genome-wide association testing of gene-environment interaction in case-parent trios. Am. J. Epidemiol. *172*, 116–122.
39. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
40. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. *46*, 100–106.
41. Graaf, R. de, Have, M. ten, Gool, C. van, and Dorsselaer, S. van (2012). Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. Soc. Psychiatry Psychiatr. Epidemiol. *47*, 203–213.
42. Neale, B.M., Lasky-Su, J., Anney, R., Franke, B., Zhou, K., Maller, J.B., Vasquez, A.A., Asherson, P., Chen, W., Banaschewski, T., et al. (2008). Genome-wide association scan of attention deficit hyperactivity disorder. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *147B*, 1337–1344.
43. Mick, E., Todorov, A., Smalley, S., Hu, X., Loo, S., Todd, R.D., Biederman, J., Byrne, D., Dechairo, B., Guiney, A., et al. (2010). Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. J. Am. Acad. Child Adolesc. Psychiatry *49*, 898–905.e3.
44. Elia, J., Gai, X., Xie, H.M., Perin, J.C., Geiger, E., Glessner, J.T., D'arcy, M., deBerardinis, R., Frackelton, E., Kim, C., et al. (2010). Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with

neurodevelopmental genes. Mol. Psychiatry *15*, 637–646.
45. Klei, L., Sanders, S.J., Murtha, M.T., Hus, V., Lowe, J.K., Willsey, a J., Moreno-De-Luca, D., Yu, T.W., Fombonne, E., Geschwind, D., et al. (2012). Common genetic variants, acting additively, are a major source of risk for autism. Mol. Autism *3*, 9.
46. Wray, N.R., Pergadia, M.L., Blackwood, D.H.R., Penninx, B.W.J.H., Gordon, S.D., Nyholt, D.R., Ripke, S., MacIntyre, D.J., McGhee, K. a, Maclean, a W., et al. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. Mol. Psychiatry *17*, 36–48.
47. Risch, N., and Teng, J. (1998). The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. Genome Res. *8*, 1273–1288.
48. Risch, N. (2001). Implications of multilocus inheritance for gene-disease association studies. Theor. Popul. Biol. *60*, 215–220.
49. Gauderman, W.J. (2003). Candidate gene association analysis for a quantitative trait, using parent-offspring trios. Genet. Epidemiol. *25*, 327–338.

Supplement of Chapter 6. Disease and polygenic architecture: avoid trio-design and appropriately account for unscreened controls for common disease

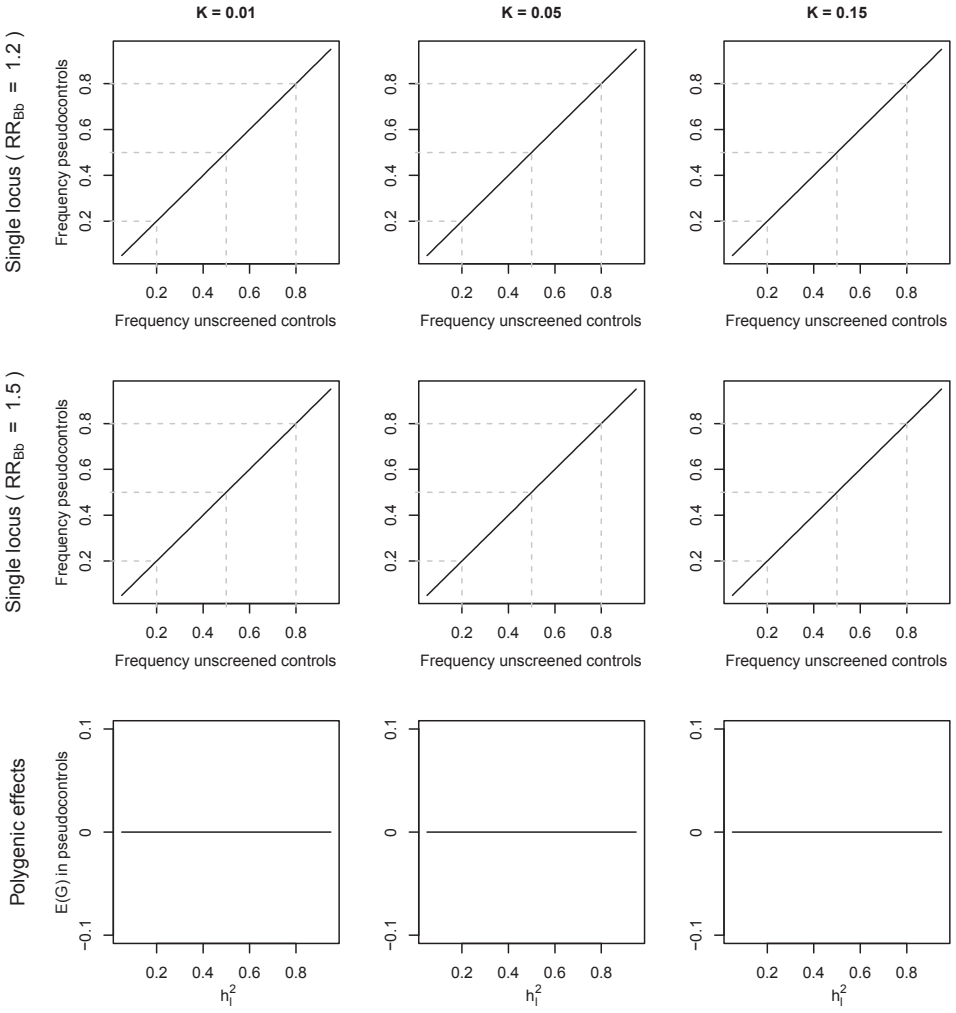


Figure S1. Pseudocontrols of random families with at least one affected proband case are equal to unscreened controls. Pseudocontrols of random families with at least one affected proband case are equal to unscreened controls (i.e. population mean) as displayed for the allele frequency of single loci of different effect-size (first two rows) and the mean genetic liability $E(G)$ (population mean equals 0) for variable heritability h_l^2(bottom row) and different baseline population risk K. The equivalence is exact and follows from the closed formulas provided in the R scripts, but is non-trivial to display in equations, because multiple sequential probabilities were needed to derive at the allele frequency and mean genetic liability in pseudocontrols. The equivalence can be understood intuitively by realizing that the non-transmitted alleles of random proband family are, in fact, part of the population background.

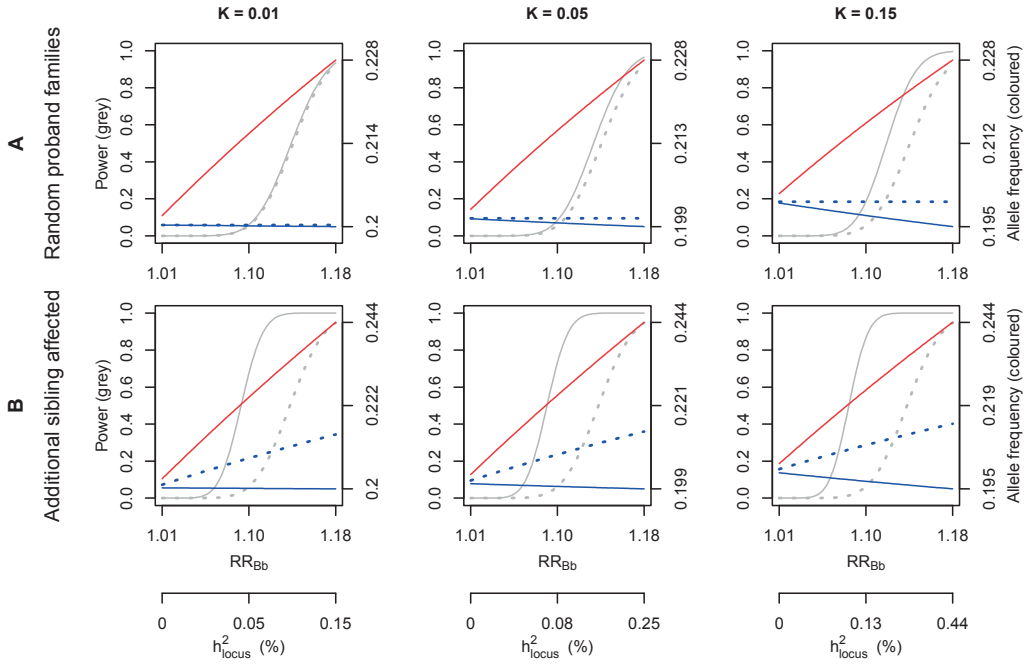


Figure S2. Power to detect a single SNP in trio-design and unscreened control studies, p=0.2. Power to detect a single SNP with risk allele frequency $p = 0.2$ for case vs screened controls (solid grey line) and case vs pseudocontrol (dotted grey line). The allele frequencies of proband cases are displayed as the red solid line, the allele frequency of screened controls as the solid blue line, and the allele frequency of pseudocontrols in the dotted blue line. The allele frequencies of pseudocontrols from proband random families equal unscreened population controls, which is reflected by the horizontal blue dotted lines at 0.2 in Panel A. Note that the grey lines equal the solid and dotted lines in Main Figure 2; the unscreened controls are not displayed in the Supplemental Figures, because they will always have an allele frequency equal to the population frequency.

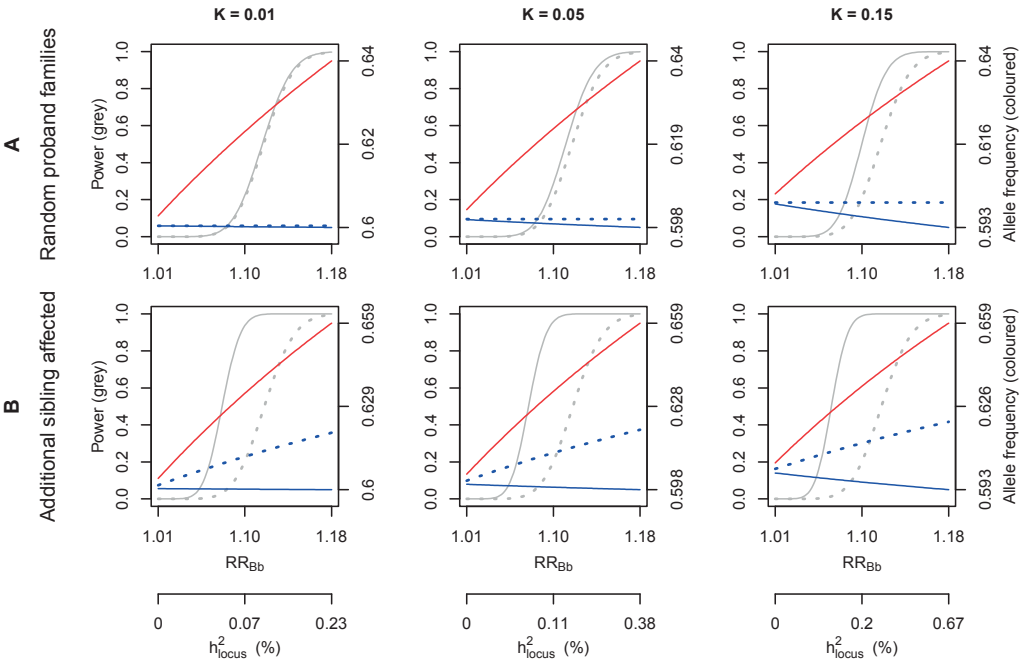


Figure S3. Power to detect a single SNP in trio-design and unscreened control studies, p=0.6. The power is displayed for a risk allele with frequency p=0.6, and results indicate that the conclusions do not depend on the allele frequency (noting that in Figure S2 a locus with p=0.2 was displayed). See the legend of Figure S2 for details.

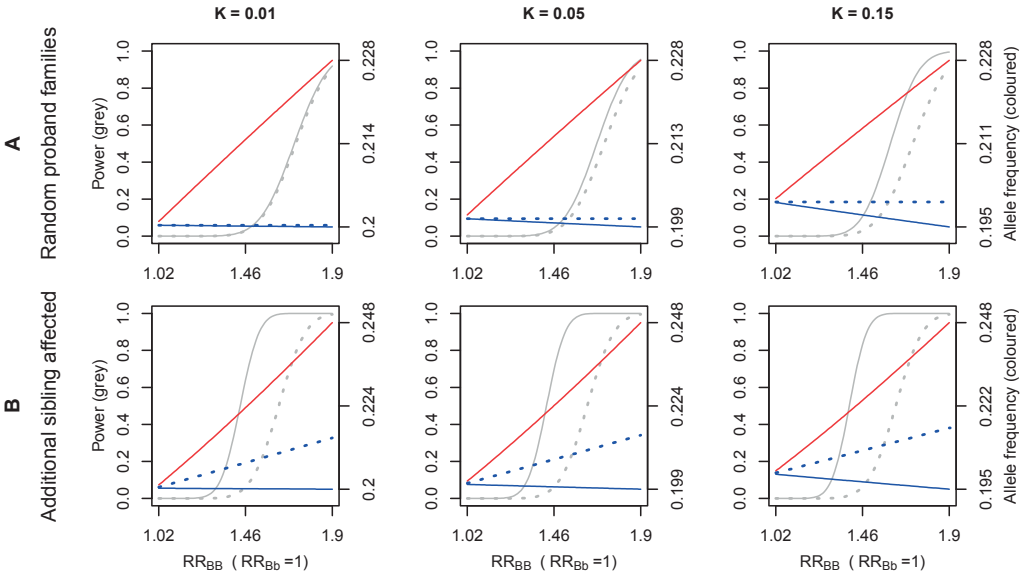


Figure S4. Power in trio design to detect SNP with underlying recessive effect. Power to detect the additive effect a single SNP with risk allele frequency $p = 0.2$ with an underlying recessive effect for case vs screened controls (solid grey line) and case vs pseudocontrol (dotted grey line). The allele frequency of cases is displayed as the red solid line, the allele frequency of screened controls as the solid blue line, and the allele frequency of pseudocontrols in the dotted blue line. Note that the RR_{BB} are being displayed for a larger range than in Figure S2 ($1.9 > 1.18^2 = 1.39$), i.e. an actual recessive allele results in less power given RR_{BB}.

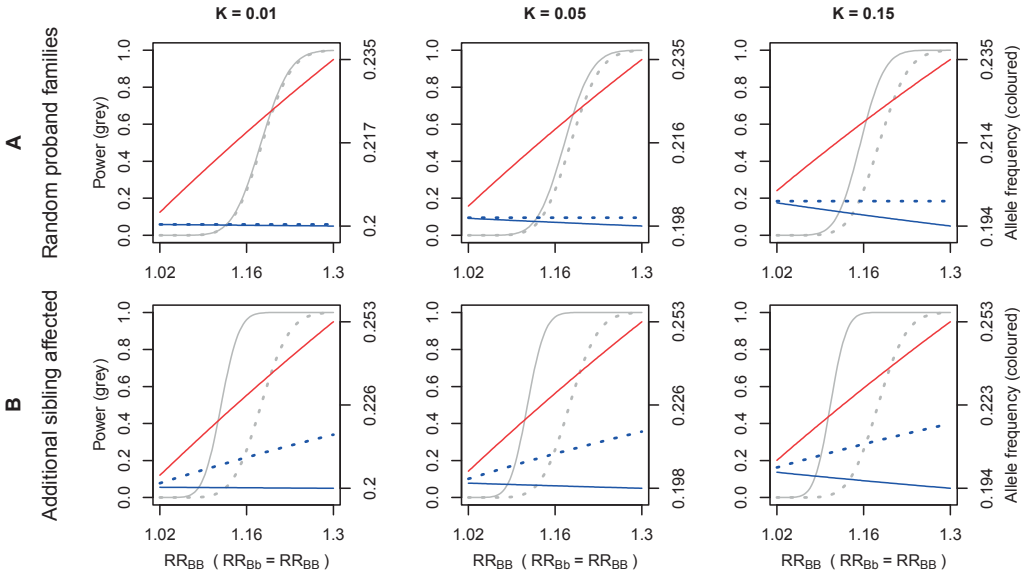


Figure S5. Power in trio design to detect SNP with underlying dominant effect. Power to detect the additive effect a single SNP with risk allele frequency $p = 0.2$ with an actual dominant effect for case vs screened controls (solid grey line) and case vs pseudocontrol (dotted grey line). The allele frequency of cases is displayed as the red solid line, the allele frequency of screened controls as the solid blue line, and the allele frequency of pseudocontrols in the dotted blue line. Note that the RR_{BB} are being displayed for a smaller range than in Figure S2 ($1.3 < 1.18^2 = 1.39$), i.e. a dominant allele results in more power given RR_{BB}.

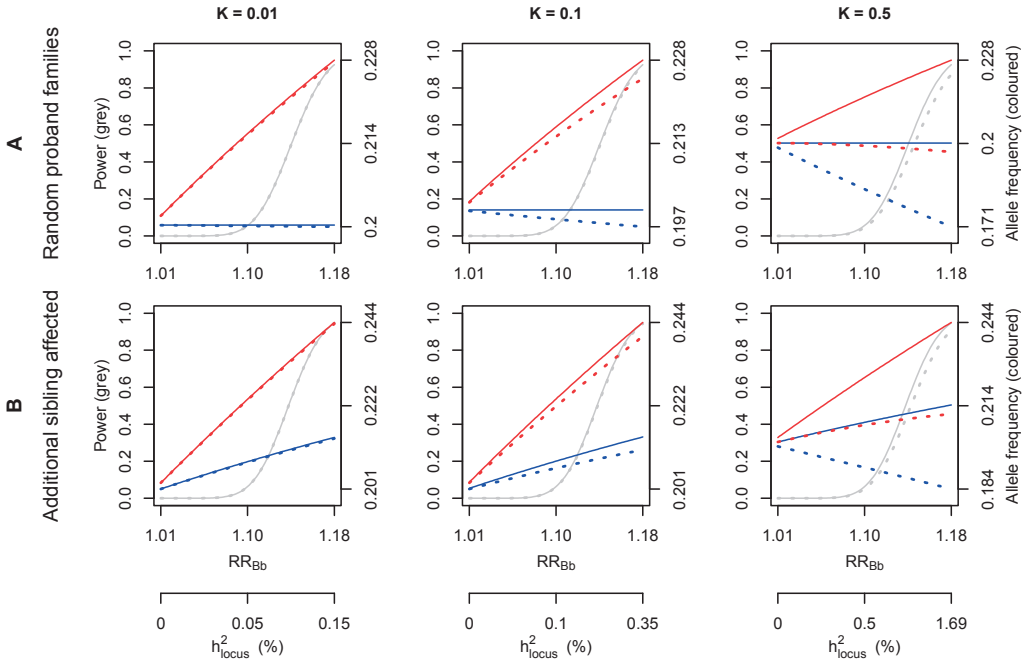


Figure S6. Power to detect SNP in trios with unaffected parents. Power to detect a single SNP with risk allele frequency $p = 0.2$ for cases vs pseudocontrols without conditioning on parents (solid grey line) and case vs pseudocontrol restricted to trios with unaffected parents (dotted grey line). The allele frequency of cases from trios without conditioning on parents is displayed as the red solid line, and the allele frequency of their pseudocontrols as the solid blue line. The allele frequency in cases from trios with unaffected parents is displayed as the red dotted line, and the allele frequency in their pseudocontrols as the dotted blue line. To summarize: solid=no selection on parents; dotted=unaffected parents; grey=power; red=allele frequency case; blue=allele frequency pseudocontrol. Note that the grey lines overlap, i.e. selecting trios with unaffected parents does not increase power in pseudocontrol studies. Furthermore, note that for $K = 0.1$ and $K = 0.5$ the allele frequencies are lower in trios from unaffected parents, but this difference is proportional for cases and pseudocontrol resulting in no power-difference.

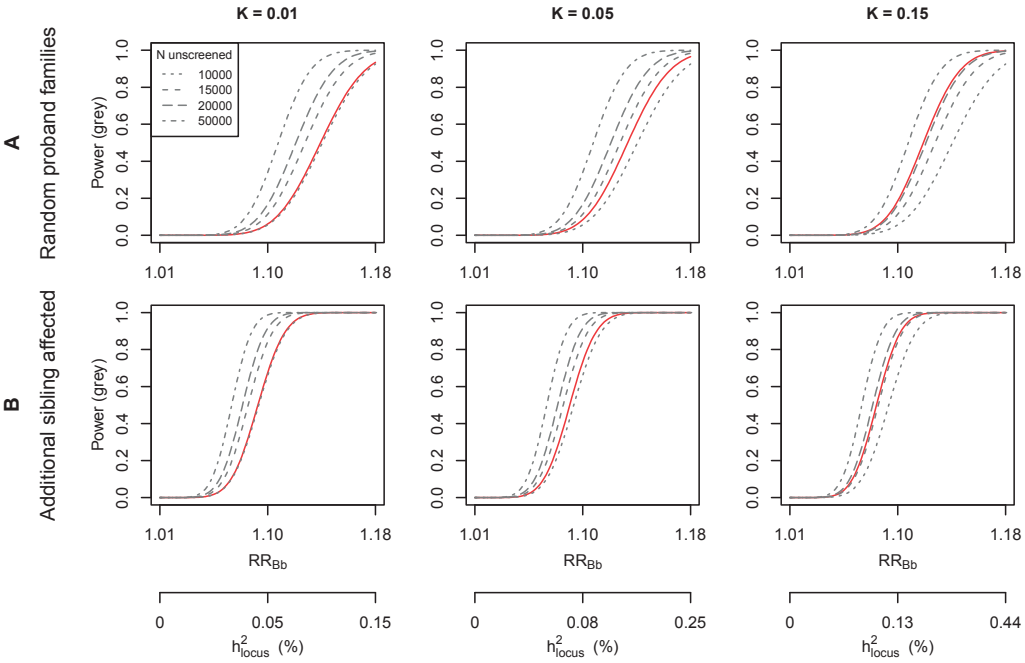


Figure S7. Power to detect a risk variant from screened *vs.* unscreened controls studies. Power to detect a risk variant with risk allele frequency $p = 0.2$ for 10,000 proband cases vs 10,000 screened controls (solid red line) and 10,000 proband cases vs respectively 10,000 unscreened controls (dotted line), 15,000 unscreened controls (short dashed), 20,000 unscreened controls (long dashed), and 50,000 unscreened controls (dot-dashed).

Table S1. Values of the Haseman Elston cross-product accounting for falsely classified controls

$y_{true,i}$	$y_{true,j}$	$y_{assumed,}$	$y_{assumed,j}$	P_{ij}	Z_{ij}
1	1	0	0	$((1-P_{assumed})F)^2$	$\frac{P_{assumed}}{1-P_{assumed}}$
1	1	1	1	$(1-P_{assumed})FP_{assumed}$	-1
1	1	0	0	$P_{assumed}(1-P_{assumed})F$	-1
1	1	1	1	$P^2_{assumed}$	$\frac{1-P_{assumed}}{P_{assumed}}$
1	0	0	0	$P_{assumed}(1-P_{assumed})(1-F)$	-1
1	0	0	0	$(1-P_{assumed})F(1-P_{assumed})(1-F)$	$\frac{P_{assumed}}{1-P_{assumed}}$
0	1	1	1	$(1-P_{assumed})(1-F)P_{assumed}$	-1
0	1	0	0	$(1-P_{assumed})(1-F)(1-P_{assumed})F$	$\frac{P_{assumed}}{1-P_{assumed}}$
0	0	0	0	$((1-P_{assumed})(1-F))^2$	$\frac{P_{assumed}}{1-P_{assumed}}$

To adjust the transformation from the heritability on the observed scale $\hat{h}_o^2$ to the liability scale $\hat{h}_l^2$ for a proportion $F = N_{false\ controls}/N_{all\ controls}$ of falsely classified controls, we closely followed the derivations of Golan et al, which we recommend for further reading (paragraphs 1.2 and 1.3 of their Supplemental Materials).[1] The adjusted expected values of the cross-product Z_{ij} used for Haseman Elston-regression follow from considering the true disease status y_{true} and assumed disease status $y_{assumed}$ with probabilities

$$\mathbb{P}(y_{true} = 1 \,\&\, y_{assumed} = 1) = P_{assumed}$$

$$\mathbb{P}(y_{true} = 1 \,\&\, y_{assumed} = 0) = (1 - P_{assumed})F$$

$$\mathbb{P}(y_{true} = 0 \,\&\, y_{assumed} = 0) = (1 - P_{assumed})(1 - F)$$

The 9 possible pairs, their probabilities $\mathbb{P}_{ij}$ and values of cross-product Z_{ij} are displayed in the Table. The expected values of $\mathbb{E}[Z_{ij}|y_{true,i}, y_{true,j}]$ follow as:

$$\mathbb{E}\left[Z_{ij}\middle|y_{true,i} = y_{true,j} = 1\right] = \frac{\sum \mathbb{P}_{ij|y_{true,i}=y_{true,j}=1} Z_{ij|y_{true,i}=y_{true,j}=1}}{\sum \mathbb{P}_{ij|y_{true,i}=y_{true,j}=1}} = \frac{P_{assumed}(1-P_{assumed})(1-F)^2}{(P_{assumed}+(1-P_{assumed})F)^2}$$

$$\mathbb{E}\left[Z_{ij}\middle|y_{true,i} \neq y_{true,j}\right] = \frac{P_{assumed}(F-1)}{(P_{assumed}+(1-P_{assumed})F)}$$

$$\mathbb{E}\left[Z_{ij}\middle|y_{true,i}=y_{true,j}=0\right]=\frac{P_{assumed}}{1-P_{assumed}}$$

Given these $\mathbb{E}\left[Z_{ij}\middle|y_{true,i},y_{true,j}\right]$ the derivation of Golan et al can be followed with $P_{Golan}=P_{true}=P_{assumed}+(1-P_{assumed})F$ to derive at the transformation of the observed to the liability scale as: $\hat{h}_l^2=\frac{K^2(1-K)^2}{P(1-P)(1-F)^2z^2}\hat{h}_{occ}^2$, where $P=P_{assumed}$.

Table S2. Simulation of falsely classified controls

Simulation parameters				Haseman-Elston regression					
				$\hat{h}_l^2$		$\hat{h}_{occ}^2$(assuming F=0)		$\hat{h}_l^2$ (corrected for F)	
K	h_l^2	P	F	Mean	SE	Mean	SE	Mean	SE
Parameters of Major Depressive Disorder									
0.2	0.4	0.5	0	0.3048	0.0131	0.3983	0.0171	0.3983	0.0171
0.2	0.4	0.5	0.1	0.2467	0.0112	0.3224	0.0146	0.3980	0.0180
0.2	0.4	0.5	0.2	0.1834	0.0095	0.2396	0.0124	0.3744	0.0194
0.2	0.4	0.25	0	0.2288	0.0062	0.3985	0.0107	0.3985	0.0107
0.2	0.4	0.25	0.1	0.1795	0.0088	0.3127	0.0153	0.3861	0.0189
0.2	0.4	0.25	0.2	0.1545	0.0055	0.2691	0.0096	0.4204	0.0150
Parameters of Schizophrenia									
0.01	0.8	0.5	0	1.4699	0.0130	0.8113	0.0072	0.8113	0.0072
0.01	0.8	0.5	0.005	1.4358	0.0116	0.7924	0.0064	0.8004	0.0065
0.01	0.8	0.5	0.01	1.4096	0.0157	0.7780	0.0087	0.7938	0.0089
0.01	0.8	0.25	0	1.0927	0.0055	0.8041	0.0040	0.8041	0.0040
0.01	0.8	0.25	0.005	1.0829	0.0078	0.7969	0.0057	0.8049	0.0058
0.01	0.8	0.25	0.01	1.0737	0.0049	0.7901	0.0036	0.8061	0.0037
Additional parameter settings to further validate the derived equation									
0.2	0.8	0.5	0	0.6282	0.0182	0.8207	0.0238	0.8207	0.0238
0.2	0.8	0.5	0.1	0.4964	0.0117	0.6485	0.0153	0.8006	0.0189
0.2	0.8	0.5	0.2	0.4062	0.0076	0.5307	0.0100	0.8293	0.0156
0.2	0.8	0.25	0	0.4608	0.0077	0.8028	0.0135	0.8028	0.0135
0.2	0.8	0.25	0.1	0.3722	0.0061	0.6484	0.0107	0.8005	0.0132
0.2	0.8	0.25	0.2	0.2956	0.0062	0.5150	0.0109	0.8047	0.0170
0.01	0.4	0.5	0	0.7287	0.0108	0.4022	0.0059	0.4022	0.0059
0.01	0.4	0.5	0.005	0.6993	0.0148	0.3859	0.0082	0.3898	0.0082
0.01	0.4	0.5	0.01	0.7022	0.0132	0.3876	0.0073	0.3954	0.0074
0.01	0.4	0.25	0	0.5395	0.0047	0.3970	0.0035	0.3970	0.0035
0.01	0.4	0.25	0.005	0.5393	0.0076	0.3969	0.0056	0.4009	0.0057
0.01	0.4	0.25	0.01	0.5375	0.0064	0.3956	0.0047	0.4036	0.0048

To validate the Equation 3, $\hat{h}_l^2 = \frac{K^2(1-K)^2}{P(1-P)(1-F)^2 z^2}\hat{h}_{occ}^2$, we performed a simulation study in line with Golan et al (Supplemental Materials paragraph 5.3).[1]

1. MAFs of 10,000 SNPs in full linkage equilibrium were randomly sampled from $U[0.05,0.5]$, and the effect sizes were randomly sampled from $N(0, h_l^2/10{,}000)$.
2. An individual was generated by
    a. Randomly assigning alleles with the probabilities given by the MAFs
    b. Standardizing the allele counts by $(allele\ count - 2 * MAF)/\sqrt{2MAF(1-MAF)}$.

   c. Assessing the genetic liability G as the product of the standardized allele counts with the effects
   d. Assessing the phenotypic liability l as $G + E$ with E randomly drawn from $N(0, 1 - h_l^2)$
   e. Defining disease status $y = 1$ for those with $l > T$ with T the liability threshold corresponding to a proportion of K cases
3. Step 2 was repeated until we obtained $2{,}000$ cases, an additional $F * 2{,}000$ cases which we labeled as controls, and $(1 - F) * 2{,}000$ true controls. The cases and controls were saved in a single ped-file.
4. Plink was used to transform the ped-file to a bim-file,[2] and GCTA[3] to estimate the genetic relationship matrix and to perform cross-product Haseman-Elston regression with the "--HEreg" option yielding $\hat{h}_{occ}^2$.
5. Steps 1-4 were repeated 10 times. The mean of these 10 point-estimates of the SNP-heritability are displays, as well as their standard error (SE) estimated as their standard deviation divided by $\sqrt{10}$.
6. The mean $\hat{h}_o^2$ was, first, transformed to the liability scale assuming $F = 0$ (i.e. with Equation 2, $\hat{h}_l^2 = \frac{K^2(1-K)^2}{P(1-P)z^2}\hat{h}_{occ}^2$), and second, with Equation 3, $\hat{h}_l^2 = \frac{K^2(1-K)^2}{P(1-P)(1-F)^2z^2}\hat{h}_{occ}^2$. Simulation illustrates that Equation 3 appropriately accounts for unscreened controls, because the actual simulated h_l^2 fall within the approximate 95% confidence interval of the mean $\hat{h}_l^2$ from simulation (mean ± 1.96*SE).

Table S3. Analytical derivation of genetic liabilities in trios versus simulation

				Screened controls		Case		Pseudo control		Case | sib aff		Ps contr | sib aff	
Method	***K***	h_l^2	ρ_l	σ^2 ***(G)***	***E(G)***	σ^2 ***(G)***	***E(G)***	σ^2 ***(G)***	***E(G)***	σ^2 ***(G)***	***E(G)***	σ^2 ***(G)***	***E(G)***
Sim	0.001	0.8	0	0.7932	-0.0027	0.2052	2.6945	0.8059	-0.0014	0.2134	2.9642	0.6400	0.9853
Ana	0.001	0.8	0	0.7933	-0.0027	0.2034	2.6937	0.8000	0.0000	0.2133	2.9529	0.6347	0.9788
Sim	0.001	0.8	0.5	0.9450	-0.0058	0.2259	2.8185	0.9360	0.4686	0.2415	3.1014	0.7186	1.4582
Ana	0.001	0.8	0.5	0.9451	-0.0058	0.2250	2.8182	0.9396	0.4697	0.2381	3.0970	0.7162	1.4595
Sim	0.001	0.4	0	0.3982	-0.0013	0.2502	1.3461	0.3991	0.0003	0.2417	1.6929	0.3489	0.5700
Ana	0.001	0.4	0	0.3983	-0.0013	0.2508	1.3468	0.4000	0.0000	0.2384	1.7045	0.3622	0.5674
Sim	0.001	0.4	0.5	0.4377	-0.0017	0.2688	1.4265	0.4392	0.1287	0.2519	1.8069	0.3818	0.7377
Ana	0.001	0.4	0.5	0.4377	-0.0017	0.2668	1.4286	0.4386	0.1299	0.2506	1.8200	0.3896	0.7484
Sim	0.01	0.8	0	0.7596	-0.0216	0.2218	2.1327	0.7996	-0.0004	0.2342	2.3623	0.6462	0.7870
Ana	0.01	0.8	0	0.7595	-0.0215	0.2220	2.1322	0.8000	0.0000	0.2344	2.3578	0.6432	0.7813
Sim	0.01	0.8	0.5	0.8914	-0.0350	0.2488	2.2414	0.9403	0.3723	0.2674	2.4906	0.7281	1.1794
Ana	0.01	0.8	0.5	0.8913	-0.0350	0.2492	2.2423	0.9403	0.3737	0.2642	2.4889	0.7282	1.1733
Sim	0.01	0.4	0	0.3899	-0.0109	0.2552	1.0664	0.4015	-0.0012	0.2451	1.3546	0.3632	0.4459
Ana	0.01	0.4	0	0.3899	-0.0108	0.2555	1.0661	0.4000	0.0000	0.2437	1.3561	0.3637	0.4513
Sim	0.01	0.4	0.5	0.4270	-0.0128	0.2720	1.1315	0.4375	0.1025	0.2571	1.4517	0.3905	0.5990
Ana	0.01	0.4	0.5	0.4271	-0.0129	0.2723	1.1323	0.4386	0.1029	0.2568	1.4509	0.3916	0.5965
Sim	0.1	0.8	0	0.6157	-0.1558	0.2682	1.4039	0.8004	-0.0003	0.2844	1.5857	0.6633	0.5286
Ana	0.1	0.8	0	0.6157	-0.1560	0.2682	1.4040	0.8000	0.0000	0.2818	1.5844	0.6615	0.5261
Sim	0.1	0.8	0.5	0.7104	-0.1982	0.3073	1.4969	0.9420	0.2497	0.3265	1.7023	0.7538	0.8060
Ana	0.1	0.8	0.5	0.7102	-0.1984	0.3071	1.4968	0.9419	0.2495	0.3208	1.6993	0.7530	0.8035
Sim	0.1	0.4	0	0.3539	-0.0780	0.2670	0.7020	0.3998	0.0000	0.2567	0.9043	0.3668	0.3016
Ana	0.1	0.4	0	0.3539	-0.0780	0.2671	0.7020	0.4000	0.0000	0.2562	0.9040	0.3671	0.3009
Sim	0.1	0.4	0.5	0.3851	-0.0873	0.2859	0.7480	0.4392	0.0677	0.2724	0.9727	0.3971	0.4003
Ana	0.1	0.4	0.5	0.3851	-0.0873	0.2858	0.7483	0.4387	0.0680	0.2713	0.9721	0.3961	0.3997

Legend to Table S3.
We validated the analytical estimations (see Supplemental Methods) of the mean genetic liabilities $E(G)$ with a simulation study. The heritability h_l^2, phenotypic correlation between parents ρ_l, the population disease frequency K, and corresponding threshold T were defined as described in the main text. Hereby, the variance-covariance matrix of the genetic liabilities of the parents was defined as

$$\sum(G_m, G_f) = \begin{pmatrix} h_l^2 & \rho_l h_l^2 h_l^2 \\ \rho_l h_l^2 h_l^2 & h_l^2 \end{pmatrix}$$

with $V_G = h_l^2 V_l = h_l^2$. Subsequently, the genetic liabilities of the mothers and fathers were randomly drawn from this bivariate normal distribution. The genetic liabilities of the first and second sibling were independently defined as $G_s = \frac{1}{2}G_m + \frac{1}{2}G_f + G_{residual}$, where $G_{residual}$ represent Mendelian variation and was randomly drawn from the normal distribution with mean 0 and variation $\frac{1}{2}V_G$.[4] The phenotypes l of the siblings were than independently defined as $l_s = G_s + E_s$, with E_s randomly drawn from $N(0,1 - h_l^2)$. To conclude, the genetic liability of the complement $c1$ of the first sibling $s1$ was defined as $G_{c1} = G_m + G_f - G_{s1}$. In this manner, $l_{s1}, G_{s1}, l_{s2}, G_{s2}, G_m, G_f$ and G_{c1} were defined for 10^8 families. We note that the value of $\sigma^2(G_s)$ thus simulated was in line with previous theoretical derivations $V_G + \frac{1}{2}\rho_G V_G$.[4,5] The respective variances, covariances and means were estimated from this simulation study and were in line with the theoretically derived values (see Table S3). Simulations were performed in R.[6]

Table S4. Heuristic prediction of assessed heritability in trios versus simulation

Simulation parameters				$\hat{h}_l^2$ screened control			$\hat{h}_l^2$ pseudocontrol		
				Simulation			Simulation		
K	h_l^2	sib aff	ρ_l	Mean	SE	Pred. $\hat{h}_l^2$	Mean	SE	Pred. $\hat{h}_l^2$
0.3	0.8	Y	0	0.9885	0.0225	0.9864	0.2182	0.0196	0.2331
0.3	0.8	N	0.5	0.9741	0.0155	0.9833	0.3303	0.0139	0.3221
0.3	0.8	Y	0.5	1.2126	0.0113	1.2214	0.1452	0.0129	0.1736
0.1	0.8	Y	0	0.9888	0.0122	0.9957	0.3613	0.0158	0.3682
0.1	0.8	N	0.5	0.9418	0.0152	0.9447	0.5001	0.0129	0.5114
0.1	0.8	Y	0.5	1.2115	0.0105	1.1839	0.2822	0.0107	0.2638
0.01	0.8	Y	0	0.9899	0.0069	0.9764	0.4249	0.0073	0.4287
0.01	0.8	N	0.5	0.8810	0.0096	0.8945	0.6054	0.0067	0.6022
0.01	0.8	Y	0.5	1.1072	0.0045	1.0987	0.3135	0.0057	0.2985
0.3	0.4	Y	0	0.6153	0.0127	0.5913	0.1397	0.0213	0.1491
0.3	0.4	N	0.5	0.4643	0.0162	0.4640	0.2154	0.0180	0.1860
0.3	0.4	Y	0.5	0.6995	0.0210	0.6957	0.1438	0.0132	0.1362
0.1	0.4	Y	0	0.6435	0.0140	0.6340	0.2257	0.0118	0.2391
0.1	0.4	N	0.5	0.4539	0.0086	0.4591	0.3002	0.0104	0.3043
0.1	0.4	Y	0.5	0.7240	0.0117	0.7379	0.1998	0.0083	0.2154
0.01	0.4	Y	0	0.6531	0.0056	0.6445	0.2952	0.0059	0.2824
0.01	0.4	N	0.5	0.4507	0.0075	0.4524	0.3573	0.0043	0.3655
0.01	0.4	Y	0.5	0.7451	0.0057	0.7391	0.2604	0.0093	0.2518

To formally get from the $E(G)$ (Table S3) of cases and controls to the SNP-heritability $\hat{h}_l^2$ that would be assessed is non-trivial, because no normal distribution thresholds exist to define the pseudocontrols or the probands with an additional affected sibling (which form a non-random subset of all cases not defined by a specific threshold). $\hat{h}_l^2$ was therefore heuristically derived and validated with a simulation study of individual level SNP-data. In short, for any baseline disease frequency K, a unique set of T, z, and i can be found such that K equals $P(l > T|l{\sim}N(0,1))$, z the height of the standard normal distribution at T, and $i = z/K$ the mean l of cases, which results in a mean G in cases of ih_l^2. We numerically inverted this equation in R to find an unique equivalent-K matching the difference between $E(G_{case}) - E(G_{(pseudo)control})$. The equivalent- K, corresponding equivalent-z and Equation 3 yields the heritability that would be assessed with Haseman-Elston regression (Pred. $\hat{h}_l^2$), and was validated with simulation study:

1. Following Golan et al,[1] the MAFs of 10,000 SNPs in full linkage disequilibrium were randomly sampled from $U[0.05,0.5]$, and the effect sizes were randomly sampled from $N(0, h_l^2/10{,}000)$.
2. An individual was generated by
    a. Randomly assigning alleles with the probabilities given by the MAFs
    b. Standardizing the allele counts by $(allele\ count - 2 * MAF)/\sqrt{2MAF(1-MAF)}$.

   c. Assessing the genetic liability G as the product of the standardized allele counts with the effects
   d. Assessing the phenotypic liability l as $G + E$ with E randomly drawn from $N(0, 1 - h_l^2)$
   e. Defining disease status $y = 1$ for those with $l > T$ with T the liability threshold corresponding to a proportion of K cases
3. Assortative mating ρ_l was simulated following
   a. The genotypes and phenotypes of 600 men l_{men} and 600 women l_{women} were simulated
   b. A vector V was simulated as $V = \rho_l l_{men} + N(0,1 - \rho_l^2)$ so that $cor(l_{men}, V) = cov(l_{men}, V)/(\sigma_{l_{men}}\sigma_V) = cov(l_{men}, \rho_l l_{men})/(1\sigma_V) = \rho_l/\sqrt{\sigma^2_{\rho_l l_{men}} + 1 - \rho_l^2} = \rho_l$
   c. Subsequently, the l_{women} were ordered in line with V thereby ensuring $cor(l_{men}, l_{women}) = \rho_l$
4. For the 600 pair of spouses, families were generated as follows
   a. Kid-1 got one random allele from the father and one from the mother for all of the 10,000 loci. Subsequently, l and disease status y were generates as described above.
   b. The genetic complement of Kid-1 was formed by the non-transmitted alleles of the parents
   c. Kid-2 was generated as Kid-1
5. Affected proband (Kid-1) were selected as cases. Depending on the type of families simulated, we additionally conditioned on $y_{Kid-2} = 1$.
6. Unaffected Kid-1's were selected as screened controls.
7. Step 2-6 were repeated until 2,000 cases and 2,000 screened controls were collected
8. Cross-product Haseman-Elston regression yielded the $\hat{h}^2_{occ}$ for case vs screened controls and case vs pseudocontrols, which were than transformed to the liability scale with $\hat{h}_l^2 = \hat{h}^2_{occ}\frac{K^2(1-K)^2}{P(1-P)z^2}$
9. Steps 1-8 were repeated 10 times for the different setting of K, h_l^2, and ρ_l. The mean of these 10 point-estimates of the SNP-heritability are displays, as well as their standard error (SE) estimated as their standard deviation divided by $\sqrt{10}$.
10. The heuristically predicted $\hat{h}_l^2$ are within or very close to the *ballpark* 95% confidence interval of the mean $\hat{h}_l^2$ from simulation (mean ± 1.96*SE), which justifies the use of this heuristic approach for Main Figure 1.

Table S5. Analytical derivation of allele frequencies in trios versus simulation

	Genotype relative risk		Random families with at least one affected sibling			Second sibling affected		Second sibling aff. Parents unaffected		Assortative mating parents		
Method	**Bb**	**BB**	**Case**	**Scr control**	**Ps control**	**Case**	**Ps control**	**Case**	**Ps control**	**Case**	**Scr control**	**Ps control**
K=0.01; p=0.2												
Sim	1.00	2.25	0.2381	0.1996	0.1995	0.2723	0.2163	0.2718	0.2155	0.2596	0.1995	0.2052
Ana	1.00	2.25	0.2381	0.1996	0.2000	0.2695	0.2205	0.2688	0.2199	0.2593	0.1994	0.2056
Sim	1.50	2.25	0.2727	0.1993	0.2000	0.3159	0.2316	0.3141	0.2303	0.2865	0.1980	0.2110
Ana	1.50	2.25	0.2727	0.1993	0.2000	0.3171	0.2358	0.3161	0.2349	0.2862	0.1991	0.2109
Sim	2.25	2.25	0.3106	0.1989	0.2002	0.3671	0.2512	0.3660	0.2502	0.3167	0.2012	0.2165
Ana	2.25	2.25	0.3103	0.1989	0.2000	0.3663	0.2475	0.3652	0.2466	0.3169	0.1988	0.2169
K=0.01; p=0.8												
Sim	1.00	2.25	0.8890	0.7991	0.8001	0.9174	0.8424	0.9167	0.8413	0.8909	0.7982	0.8128
Ana	1.00	2.25	0.8889	0.7991	0.8000	0.9179	0.8446	0.9174	0.8437	0.8907	0.7991	0.8131
Sim	1.50	2.25	0.8571	0.7995	0.8004	0.8767	0.8267	0.8763	0.8261	0.8634	0.7992	0.8085
Ana	1.50	2.25	0.8571	0.7994	0.8000	0.8788	0.8283	0.8784	0.8278	0.8637	0.7994	0.8085
Sim	2.25	2.25	0.8181	0.7998	0.7998	0.8233	0.8107	0.8233	0.8104	0.8294	0.8001	0.8029
Ana	2.25	2.25	0.8182	0.7998	0.8000	0.8241	0.8086	0.8239	0.8085	0.8295	0.7997	0.8028
K=0.3; p=0.2												
Sim	1.00	2.25	0.2381	0.1836	0.2000	0.2696	0.2206	0.2415	0.1956	0.2593	0.1730	0.2055
Ana	1.00	2.25	0.2381	0.1837	0.2000	0.2695	0.2205	0.2403	0.1943	0.2593	0.1736	0.2056
Sim	1.50	2.25	0.2727	0.1688	0.2000	0.3171	0.2358	0.2733	0.1980	0.2861	0.1644	0.2109
Ana	1.50	2.25	0.2727	0.1688	0.2000	0.3171	0.2358	0.2732	0.1980	0.2862	0.1628	0.2109
Sim	2.25	2.25	0.3104	0.1527	0.2000	0.3663	0.2475	0.3152	0.2068	0.3169	0.1539	0.2169
Ana	2.25	2.25	0.3103	0.1527	0.2000	0.3663	0.2475	0.3148	0.2060	0.3169	0.1514	0.2169

Table S5. (continued)

	Genotype relative risk		Random families with at least one affected sibling			Second sibling affected		Second sibling aff. Parents unaffected		Assortative mating parents		
Method	**Bp**	**BB**	**Case**	**Scr control**	**Ps control**	**Case**	**Ps control**	**Case**	**Ps control**	**Case**	**Scr control**	**Ps control**
K=0.3; p=0.8												
Sim	1.00	2.25	0.8889	0.7619	0.8000	0.9178	0.8445	0.8953	0.8062	0.8908	0.7609	0.8131
Ana	1.00	2.25	0.8889	0.7619	0.8000	0.9179	0.8446	0.8958	0.8066	0.8907	0.7602	0.8131
Sim	1.50	2.25	0.8571	0.7755	0.8000	0.8787	0.8283	0.8622	0.8055	0.8637	0.7719	0.8085
Ana	1.50	2.25	0.8571	0.7755	0.8000	0.8788	0.8283	0.8621	0.8056	0.8637	0.7726	0.8085
Sim	2.25	2.25	0.8183	0.7922	0.8000	0.8242	0.8086	0.8184	0.8021	0.8294	0.7893	0.8028
Ana	2.25	2.25	0.8182	0.7922	0.8000	0.8241	0.8086	0.8184	0.8026	0.8295	0.7876	0.8028

We checked the analytical estimations (described in Supplemental Methods) of allele frequencies with a simulation study. Genotypes were simulated by first randomly assigning each parent two alleles with frequency $p = P(B)$ of the risk allele B. Then, genotypes of the first and second siblings were defined by assigning them a single random allele from both of their parents. The genotypes of the pseudocontrols were defined as the two alleles of the parents not transmitted to the first sibling. Disease status was randomly assigned to parents, siblings, with a probability of disease per genotype of $P(\text{Disease}|\text{Genotype})$ (see Witte et al for details)[7]. Families with the first sibling affected were selected as proband families with the first sibling serving as the proband case. Assortative mating was simulated as the non-random mating fraction $\alpha = 0.3$ (see Supplemental Methods section 2.4 for details), which correspond to a spouse-correlation at the locus of 0.3 (note that this unrealistic large value is merely to validate theory, because assortative mating will have no impact on allele frequency as for a phenotypic spouse-correlation of 0.3 a locus explaining 1% of variance would have a spouse-correlation of only $0.3 * 0.01 = 0.003$). We simulated 10^8 families and compared allele frequencies in different types of cases, controls, and pseudocontrols to the algebraic estimates. Results displayed in this Table validate the analytical estimations described in the Supplemental Methods that were used to make the relevant Figures and Tables.

SUPPLEMENTAL METHODS

1. Derivation of genetic liabilities in trio design

The mean genetic liabilities (breeding values) $E(G)$ and their variances were subsequently derived for random families (Section 1.1), families with one affected sibling (Section 1.2), and families with two affected siblings (Section 1.3). Therefore, variance-covariance matrices were derived for these family's phenotypic liabilities and genetic liabilities. The mean genetic liability of screened controls in the offspring generation was derived in Section 1.4. The analytical estimates of the mean genetic liabilities and their variances were validated with a simulation study (Table S3). In Table S4, the derived mean genetic liabilities are used to heuristically predict the SNP-based heritability that would be assessed with Haseman Elston-regression, which is again validated with a simulation study.

Consider a complex disease with a population frequency K and heritability h_l^2 in the parental population. Define phenotype l to represent the underlying liability for disease with variance $V_l = 1$ (the choice for V_l is arbitrary, but conveniently set to 1). The variance of genetic liabilities G equals $V_G = V_l h_l^2 = h_l^2$, while the environmental variance equals $V_E = V_l - V_G = 1 - h_l^2$. Assuming that the parents have a phenotypic correlation of $\rho_l \geq 0$, the genetic correlation follows as $\rho_G = h_l^2 \rho_l$ (page 175 of Falconer and Mackay)[8] and the genetic covariance as $\rho_G V_G$.

1.1 Variances and covariances of genetic liabilities in random families

Consider families with a mother (m), father (f), first sibling ($s1$), second sibling ($s2$) and the pseudocontrol of the first sibling (interchangeably referred to as the complement of the first sibling, $c1$). Their genetic liability values are denoted with G_m, G_f, G_{s1}, G_{s2}, respectively. The variance of genetic liabilities in the siblings equals $\sigma^2(G_{s1}) = \sigma^2(G_{s2}) = \sigma^2(G_s) = \sigma^2\left(\frac{1}{2}G_m + \frac{1}{2}G_f\right) + V_{residual}$, where $V_{residual}$ represents Mendelian variation. Bulmer (page 175)[4] proved that $V_{residual} = \frac{1}{2}V_G$, which gives $\sigma^2(G_s) = \sigma^2\left(\frac{1}{2}G_m\right) + \sigma^2\left(\frac{1}{2}G_f\right) + 2\sigma\left(\frac{1}{2}G_m, \frac{1}{2}G_f\right) + \frac{1}{2}V_G = V_G + \frac{1}{2}\rho_G V_G$. In addition, Bulmer showed that the variation of non-genetic effects (E) is not effected by assortative mating, which gives the phenotypic variation of the siblings as $\sigma^2(l_{s1}) = \sigma^2(l_{s2}) = \sigma^2(l_s) = \sigma^2(G_s + E_s) = \sigma^2(G_s) + \sigma^2(E_s) = \sigma^2(G_s) + V_E$. Keeping in mind that $\sigma(G, E) = 0$ per definition, gives $\sigma(l_s, G_s) = \sigma^2(G_s)$, as well as $\sigma(l_{s1}, G_{s2}) = \sigma(l_{s2}, G_{s1}) = \sigma(G_{s1}, G_{s2}) = \sigma\left(\frac{1}{2}G_f + \frac{1}{2}G_m, \frac{1}{2}G_f + \frac{1}{2}G_m\right) = \sigma\left(\frac{1}{2}G_f, \frac{1}{2}G_f\right) + \sigma\left(\frac{1}{2}G_f, \frac{1}{2}G_m\right) +$

$\sigma\left(\frac{1}{2}G_m, \frac{1}{2}A_f\right) + \sigma\left(\frac{1}{2}G_m, \frac{1}{2}G_m\right) = \frac{1}{2}V_G + \frac{1}{2}\rho_G V_G$. The variance of the genetic liabilities in the parents equals $\sigma^2(G_m) = \sigma^2(G_f) = V_G$, and the covariance between fathers and mother equals $\sigma(G_m, G_f) = \rho_G V_G$. The covariance between the siblings and their parents subsequently follows as $\sigma(G_m, l_s) = \sigma(G_f, l_s) = \sigma(G_m, G_s) = \sigma(G_f, G_s) = \sigma\left(G_f, \frac{1}{2}G_m + \frac{1}{2}G_f\right) = \sigma\left(G_f, \frac{1}{2}G_m\right) + \sigma\left(G_f, \frac{1}{2}G_f\right) = \frac{1}{2}V_G + \frac{1}{2}\rho_G V_G$. For the complement of the first sibling, the following covariances are found:

- $\sigma(G_{c1}, l_{s1}) = \sigma(G_{c1}, G_{s1}) = \sigma(G_m + G_f - G_{s1}, G_{s1}) = \sigma(G_m, G_{s1}) + \sigma(G_f, G_{s1}) - \sigma^2(G_{s1}) = V_G + \rho_G V_G - V_G - \frac{1}{2}\rho_G V_G = \frac{1}{2}\rho_G V_G$, and
- $\sigma(G_{c1}, l_{s2}) = \sigma(G_{c1}, G_{s2}) = \sigma(G_m + G_f - G_{s1}, G_{s2}) = (G_m, G_{s2}) + \sigma(G_f, G_{s2}) - \sigma(G_{s1}, G_{s2}) = V_G + \rho_G V_G - \frac{1}{2}V_G - \frac{1}{2}\rho_G V_G = \frac{1}{2}V_G + \frac{1}{2}\rho_G V_G$, and
- $\sigma(G_{c1}, G_m) = \sigma(G_{c1}, G_f) = \sigma(G_m + G_f - G_{s1}, G_f) = \sigma(G_m, G_f) + \sigma^2(G_f) - \sigma(G_{s1}, G_f) = \rho_G V_G + V_G - \frac{1}{2}V_G - \frac{1}{2}\rho_G V_G = \frac{1}{2}V_G + \frac{1}{2}\rho_G$, and finally
- $\sigma^2(G_{c1}) = \sigma^2(G_m + G_f - G_{s1}) = \sigma^2\left(G_m + G_f - \frac{1}{2}G_m - \frac{1}{2}G_f - G_{residual}\right) = \sigma^2\left(\frac{1}{2}G_m, \frac{1}{2}G_f\right) + (-1)^2\sigma^2(G_{residual}) = V_G + \frac{1}{2}\rho_G V_G$

By this, all element were derived of $\sum(l_{s1}, G_{s1}, l_{s2}, G_{s2}, G_m, G_f, G_{c1})$, the 7x7 variance-covariance matrix of random families. The means of $l_{s1}, G_{s1}, l_{s2}, G_{s2}, G_m, G_f$ and G_{c1} all equal zero, noting that assortative mating does not change the mean genetic liability, because $E\left(\frac{1}{2}G_m + \frac{1}{2}G_f + G_{residual}\right) = E\left(\frac{1}{2}G_m\right) + E\left(\frac{1}{2}G_f\right) + E(G_{residual})$, also when $\sigma\left(\frac{1}{2}G_m, \frac{1}{2}G_f\right) > 0$.

1.2 Variances and covariances of genetic liabilities in families with at least one affected sibling

Assortative mating increases the variances of the phenotype l from the parental to the offspring generation with $\frac{1}{2}\rho_G V_G$. The increase in V_l results in a higher disease frequency in the offspring generation, because the liability threshold T remains the same. In order to estimate the reduction in variance in the affected siblings (assume $s1$ to be affected), the offspring population was first described in terms of the standard normal distribution, and than transformed back to the

parental scale. The new disease frequency $K_{offspring}$ follows from $P(x > T \mid x{\sim}N(0, \sqrt{\sigma^2(l_s)}))$, and gives the mean phenotypic value of the affected siblings $s1$ on the standardized liability scale as $i_{offspring} = z_{offspring}/K_{offspring}$, where $z_{offspring}$ is the height of the standard normal distribution $N(0,1)$ at threshold $T_{offspring}$ with $K_{offspring} = P\big(x > T_{offspring} \,\big|\, x{\sim}N(0,1)\big)$. Bulmer showed (page 153)[4] that the reduction of variation in affected siblings on the standardized liability scale equals $k_{offspring} = i_{offspring}(i_{offspring} - T_{offspring})$, and the variance reduction on the parental liability scale thus equals $k = k_{offspring}/\sigma^2(l_s)$. Tallis showed that given normality of G and l in the family members, the new variances and covariances are given by $\sigma(X, Y|s1\ affected) = \sigma(X, Y) - k\sigma(X, l_{s1})\sigma(Y, l_{s1})$, where X and Y represent all pairwise combinations of $l_{s1}, G_{s1}, l_{s2}, G_{s2}, G_m, G_f$ and G_{c1}.[9] By this, all element are defined of $\sum\big(l_{s1}, G_{s1}, l_{s2}, G_{s2}, G_m, G_f, G_{c1} \,\big|\, s1\ affected\big)$, the 7x7 variance-covariance matrix of families with one affected sibling. Given these variances and covariances, the means were derived as follows.

- $E(l_{s1}|s1\ aff) = i_{offspring}\sqrt{\sigma^2(l_s)}$
- $E(G_{s1}|s1\ aff) = \{\sigma^2(G_{s1})/\sigma^2(l_{s1})\} * E(l_{s1}|s1\ aff)$
- $E(l_{s2}|s1\ aff) = \{\sigma(l_{s1}, l_{s2})/\sigma^2(l_{s1})\} * E(l_{s1}|s1\ aff)$
- $E(G_{s2}|s1\ aff) = \{\sigma(G_{s1}, G_{s2})/\sigma^2(G_{s1})\} * E(G_{s1}|s1\ aff)$
- $E(G_m|s1\ aff) = E\big(G_f|s1\ aff\big) = \Big\{(\frac{1}{2}V_G + \frac{1}{2}\rho_G V_G)/\sigma^2(G_s)\Big\} * E(G_{s1}|s1\ aff)$, noting that $\frac{1}{2}V_G + \frac{1}{2}\rho_G V_G$ is the part of $\sigma^2(G_s)$ following from the parents contribution $\frac{1}{2}G_f + \frac{1}{2}G_m$.
- $E(G_{c1}|s1\ aff) = E(G_m|s1\ aff) + E\big(G_f|s1\ aff\big) - E(G_{s1}|s1\ aff)$

1.3 Variances and covariances of genetic liabilities in families with two affected siblings

To derive variances and covariances within families with two affected siblings, we take the estimates of families with one affected sibling as starting point. However, in order to apply Tallis' method to account of reduction in variance when selecting for an affected sibling, G and l need to be normally distributed in all family members. The distribution of l in the first sibling $s1$ is evidentially non-normal, because he is affected. Nevertheless, the distributions of G and l in the other family members are approximately normally distributed, which was illustrated by simulation (not shown) and can be intuitively understood as follows. The first sibling is affected when l_{s1} exceeds the threshold T. However,

because l_{s1} is the sum of G_{s1} and E_{s1} and because G_{s1} and E_{s1} are independent, the violation of normality in $G_{s1|s1\,aff}$ is less than in $l_{s1|s1\,aff}$. In addition, the covariances between $G_{s1|s1\,aff}$ and G and l in the other family members are considerably smaller than 1. Hence, the distribution of G and l in all family members but sibling $s1$ are approximately normally distributed. Furthermore, note that the first and second sibling have equal genetic characteristics when they are both selected to be affected (except for their covariance with the complement, but this characteristic is not needed for this study). The variances and covariances are thus given by

$$\sigma(X, Y \mid s1\ affected\ \&\ s2\ affected) = \sigma(X, Y \mid s1\ affected) - k_2\sigma(X, l_{s2} \mid s1\ affected)\sigma(Y, l_{s2} \mid s1\ affected),$$

where X and Y take all pairwise combinations of l_{s2}, G_{s2}, G_m, G_f and G_{c1}. The variance reduction k_2 is derived analoguously as k. The disease frequency in the second siblings $K_{s2\,|\,s1\,affected}$ follows from $P(x > T \mid x{\sim}N(E(l_{s2}|s1\ aff), \sqrt{\sigma^2(l_{s2}|s1\ affected)}))$, and gives the mean phenotypic value of the affected siblings $s2$ on the standardized liability scale as $i_{s2\,|\,s1\,affected} = z_{s2\,|\,s1\,affected}/K_{s2\,|\,s1\,affected}$, where $z_{s2\,|\,s1\,affected}$ is the height of the standard normal distribution $N(0,1)$ at threshold $T_{s2\,|\,s1\,affected}$ with $K_{s2\,|\,s1\,affected} = P\left(x > T_{s2\,|\,s1\,affected} \,\middle|\, x{\sim}N(0,1)\right)$. The reduction of variation in affected second siblings on the standardized liability scale equals $k_{s2\,|\,s1\,affected} = i_{s2\,|\,s1\,affected}(i_{s2\,|\,s1\,affected} - T_{s2\,|\,s1\,affected})$, and the variance reduction on the parental liability scale thus equals $k_2 = k_{s2\,|\,s1\,affected}/\sigma^2(l_{s2}|s1\ affected)$. This defines $\sum\left(l_{s2}, G_{s2}, G_m, G_f, G_{c1} \,\middle|\, s1\ \&\ s2\ affected\right)$, the 5x5 variance-covariance matrix of families with two affected siblings (leaving out the first sibling $s1$). Given this variance-covariance matrix, the means were derived as:

- $E(l_{s2}|\ s1\ \&\ s2\ aff) = E(l_{s2}|\ s1\ aff) + i_{s2\,|\,s1\,affected}\sqrt{\sigma^2(l_{s2}|\ s1\ affected)}$
- $E(G_{s2}|\ s1\ \&\ s2\ aff) = E(G_{s2}|\ s1\ aff) + \{i_{s2\,|\,s1\,affected}\sqrt{\sigma^2(l_{s2}|\ s1\ affected)}\} * \sigma^2(G_{s2}|\ s1\ affected)/\sigma^2(l_{s2}|\ s1\ affected)$
- $E(G_m|s1\ \&\ s2\ aff) = E\left(G_f\middle|s1\ \&\ s2\ aff\right) = E\left(G_f\middle|s1\ aff\right) + \delta * \{\frac{1}{2}\sigma^2(G_m|s1\ aff) + \frac{1}{2}\sigma(G_m, G_f|s1\ aff)\}/\{\sigma^2(G_{s2}|s1\ aff)\}$, with $\delta =$

$E(G_{s2}|\ s1\ \&\ s2\ aff) - E(G_{s2}|\ s1aff)$, while noting that $\frac{1}{2}\sigma^2(G_m|s1\ aff) + \frac{1}{2}\sigma\big(G_m, G_f|s1\ aff\big) + \frac{1}{2}V_{residual} = \sigma^2(G_{s2}|s1\ aff)$.

- $E(G_{c1}|\ s1\ \&\ s2\ aff) = E(G_m|\ s1\ \&\ s2\ aff) + E\big(G_f\big|\ s1\ \&\ s2\ aff\big) - E(G_{s1}|\ s1\ \&\ s2\ aff)$, where $E(G_{s1}|\ s1\ \&\ s2\ aff) = E(G_{s2}|\ s1\ \&\ s2\ aff)$.

1.4 Genetic liabilities of screened controls

Screened controls were selected from the offspring generation, i.e. after one generation of assortative mating. In order to apply the useful properties of the standard normal distribution, the liability scale was inverted to regard controls as 'cases', and later transformed back to the original scale of l in the parental generation. The population frequency of screened controls in the offspring generation is $K_{screened\ controls} = 1 - K_{offspring}$, which gives $i_{screened\ controls}$ and $k_{screened\ controls}$ as described previously in Section 1.2. The variation of genetic liabilities follows as

$\sigma^2(G_{screened\ controls}) = \sigma^2(G_s) - \{k_{screened\ controls}/\sigma^2(l_s)\} * \sigma(l_s, G_s) * \sigma(l_s, G_s)$, and the mean as $E(G_{screened\ controls}) = -1 * \{\sigma^2(G_{s1})/\sigma^2(l_{s1})\} * i_{screened\ controls}\sqrt{\sigma^2(l_s)}$, where the term is multiplied by -1 to transform the mean back to the original parental liability scale of l.

2. Derivation of a single SNP's risk allele frequency in trio design

First, the risk allele frequencies were analytically derived for screened controls, cases, and cases with unaffected parents ('cases' and 'probands' are used interchangeably) (Section 2.1). Second, risk allele frequencies were derived for cases with affected siblings by applying the first set of derived frequencies and by considering IBD-sharing between cases and their siblings (Section 2.2). Third, all acquired estimates were applied to estimate risk allele frequencies in pseudocontrols (Section 2.3). Next we consider the impact of assortative mating (Section 2.4). To conclude, analytical derivations were validated with a simulation study (Table S5).

2.1 Risk allele frequencies in screened controls, cases, and cases with unaffected parents

This Section closely follows the work of Witte et al.[7] Assume the complex disease of interest has a population frequency $P(\mathrm{D}) = K$, and the locus of interest has risk allele B with frequency $P(\mathrm{B}) = p$, and non-risk allele b with frequency $P(\mathrm{b}) = 1 - p = q$. Given Hardy-Weinberg Equilibrium (HWE), the genotype frequencies are $P(\mathrm{bb}) = q^2$, $P(Bb) = 2pq$, and $P(\mathrm{BB}) = p^2$. Under a multiplicative risk model with relative risk of the heterozygote λ, the risk of disease given genotype $P(\mathrm{D}|\mathrm{G})$ can be expressed as $P(\mathrm{D}|\mathrm{bb}) = k_{bb}$, $P(D|Bb) = k_{bb}\lambda$, and $P(\mathrm{D}|\mathrm{BB}) = k_{bb}\lambda^2$, with k_{bb} the disease risk in subjects with genotype bb. The probabilities of genotypes in cases is given by $P(\mathrm{G}|\mathrm{D}) = P(\mathrm{D}|\mathrm{G})P(\mathrm{G})/P(\mathrm{D})$, that is $P(\mathrm{bb}|\mathrm{D}) = k_{bb}q^2/K$, $P(\mathrm{Bb}|\mathrm{D}) = k_{bb}\lambda 2pq/K$, and $P(\mathrm{BB}|\mathrm{D}) = k_{bb}\lambda^2 p^2/K$. Affected individuals, thus, have a risk allele frequency of $p_{case} = P(\mathrm{BB}|\mathrm{D}) + \frac{1}{2}P(\mathrm{Bb}|\mathrm{D})$. Analogously, the probabilities of genotypes in unaffected individuals (i.e., screened controls, sc) are given by $p(\mathrm{bb}|\mathrm{ND}) = (1 - k_{bb})q^2/(1 - K)$, $P(\mathrm{Bb}|\mathrm{ND}) = (1 - k_{bb}\lambda)2pq/(1 - K)$, and $P(\mathrm{BB}|\mathrm{ND}) = (1 - k_{bb}\lambda^2)p^2/(1 - K)$, and they have a risk allele frequency of $p_{sc} = P(\mathrm{BB}|\mathrm{ND}) + \frac{1}{2}P(\mathrm{Bb}|\mathrm{ND})$, and non-risk allele frequency $q_{sc} = 1 - p_{sc}$. The offspring of unaffected parents will have genotype frequencies $P(\mathrm{G} \mid \text{parents unaffected})$ of $P(\mathrm{bb}|\mathrm{pu}) = q_{sc}^2$, $P(\mathrm{Bb}|\mathrm{pu}) = 2p_{sc}q_{sc}$, and $P(\mathrm{BB}|\mathrm{pu}) = p_{sc}^2$, noting that HWE is re-established after one generation. Assuming no correlation between genotype and family environment, the $P(\mathrm{D}|\mathrm{G})$ in offspring of screened controls are equal to $P(\mathrm{D}|\mathrm{G})$ in the baseline population. The probabilities of genotypes in cases (proband) with unaffected parents, therefore, equal $P(\mathrm{bb}|\mathrm{D},\mathrm{pu}) = k_{bb}q_{sc}^2/P(\mathrm{D}|\mathrm{pu})$, $P(\mathrm{Bb}|\mathrm{D},\mathrm{pu}) = k_{bb}\lambda 2p_{sc}q_{sc}/P(\mathrm{D}|\mathrm{pu})$, and $P(\mathrm{BB}|\mathrm{D},\mathrm{pu}) = k_{bb}\lambda^2 p_{sc}^2/P(\mathrm{D}|\mathrm{pu})$, with $P(\mathrm{D}|\mathrm{pu}) = k_{bb}q_{sc}^2 + k_{bb}\lambda 2p_{sc}q_{sc} + k_{bb}\lambda^2 p_{sc}^2$. Note that

all can be expressed in terms of $p, q = 1 - p, K,$ and λ by realizing that $K = \sum_G P(D|G)P(G) = q^2 k_{bb} + 2pqk_{bb}\lambda + p^2 k_{bb}\lambda^2$, and thus $k_{bb} = K/(q^2 + 2pq\lambda + p^2\lambda^2)$. To take account of dominance effect, substitute λ with RR_{Bb} and λ^2 with RR_{BB} in the above.

2.2 Risk allele frequencies in proband with an affected sibling

To estimate the risk allele frequency in cases (proband) with affected siblings, the combined probabilities of genotypes in cases and their siblings is required:

$$\boldsymbol{P}(G_{case}, G_{sib}) = \boldsymbol{P}(G_c, G_s) = \begin{pmatrix} P(bb,bb) & P(bb,Bb) & P(bb,BB) \\ P(Bb,bb) & P(Bb,Bb) & P(Bb,BB) \\ P(BB,bb) & P(BB,Bb) & P(BB,BB) \end{pmatrix}$$

The rows of $\boldsymbol{P}(G_c, G_s)$ thus correspond to the three possible genotypes of cases and the columns to the three possible genotypes of their siblings. $\boldsymbol{P}(G_c, G_s)$ is the sum of four matrices: $\boldsymbol{P}(G_c, G_s|\, IBD = 0)$, $\boldsymbol{P}(G_c, G_s|\, IBD = 1(b))$, $\boldsymbol{P}(G_c, G_s|\, IBD = 1(B))$, and $\boldsymbol{P}(G_c, G_s|\, IBD = 2)$, all weighted by $0.25 = \boldsymbol{P}(IBD = 0) = \boldsymbol{P}(IBD = 1)/2 = \boldsymbol{P}(IBD = 2)$. To illustrate, the three row elements of $\boldsymbol{P}(G_s|\, G_c = Bb, IBD = 1(B))$ follow from basic Mendelian reasoning as $P(G_s = bb|\, G_c = Bb, IBD = 1(B)) = 0 * q_{NT|G_c=Bb}$ (the probability that the IDB-allele is b equals 0; the probability that the non-IBD allele is b depends on its frequency in the non-transmitted alleles from the parents given $G_c = Bb$), $P(G_s = Bb|\, G_c = Bb, IBD = 1(B)) = 1 * q_{NT|G_c=Bb}$, and $P(G_s = BB|\, G_c = Bb, IBD = 1(B)) = 1 * p_{NT|G_c=Bb}$ respectively, where $p_{NT|G_c}$ represents the frequency of B in the non-transmitted alleles from parents given G_c, and $q_{NT|G_c} = 1 - p_{p|G_c}$ the frequency of b. Note that $p_{NT|G_c}$ equals $p_{parents}$ when the parental generation is in HWE, however when the parents are unaffected they are not in HWE and derivation of $p_{NT|G_c}$ is slightly more elaborate (described in Appendix A). When IBD=0, the genotypes G_s depend on the distribution of the non-transmitted genotypes, which is also described in Appendix A. In this manner, the four matrices $\boldsymbol{P}(G_s|\, G_c, IBD)$ are defined as:

$$\boldsymbol{P}(G_s|\, G_c, IBD = 0) = \begin{pmatrix} P(NT = bb|G_c = bb) & P(NT = Bb|G_c = bb) & P(NT = BB|G_c = bb) \\ P(NT = bb|G_c = Bb) & P(NT = Bb|G_c = Bb) & P(NT = BB|G_c = Bb) \\ P(NT = bb|G_c = BB) & P(NT = Bb|G_c = BB) & P(NT = BB|G_c = BB) \end{pmatrix}$$

$$\boldsymbol{P}(G_s|\, G_c, IBD = 1(b)) = \begin{pmatrix} 2q_{NT|G_c=bb} & 2p_{NT|G_c=bb} & 0 \\ q_{NT|G_c=Bb} & p_{NT|G_c=Bb} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{P}(G_s|\, G_c, IBD = 1(B)) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & q_{NT|G_c=Bb} & p_{NT|G_c=Bb} \\ 0 & 2q_{NT|G_c=BB} & 2p_{NT|G_c=BB} \end{pmatrix}$$

$$\boldsymbol{P}(G_s|\, G_c, IBD = 2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

First, the allele frequency in cases with an affected sibling and random parents (in HWE) was derived, where $p_{NT} = p$ irrespective of G_c. Furthermore, define the diagonal matrix with the genotype probabilities in cases, and the diagonal matrix with the probabilities on an affected sibling given the siblings genotype as follows

$\boldsymbol{P}(G_c) = \mathrm{diag}\big(P(G|D)\big) = \mathrm{diag}(P(bb|D), P(Bb|D), P(BB|D))$, and
$\boldsymbol{P}(S = Affected|G_s) = \mathrm{diag}\big(P(D|G)\big) = \mathrm{diag}(P(D|bb), P(D|Bb), P(D|BB))$

Now estimate the combined genotype probabilities of cases and their sibling

$\boldsymbol{P}(\mathrm{G_c}, \mathrm{G_{s=Affected}}|\mathrm{IBD}) = \boldsymbol{P}(\mathrm{G_c}) * \boldsymbol{P}(G_s|\, G_c, IBD) * \boldsymbol{P}(S = Affected|\mathrm{G_s})$, (*Eq 1*)
and
$\boldsymbol{P}\big(G_c, G_{s=Affected}\big) = \sum_{IBD} 0.25 * \boldsymbol{P}\big(\mathrm{G_c}, \mathrm{G}_{s=Affected}\big|\mathrm{IBD}\big)$

Because of the ascertainment on cases the elements of $\boldsymbol{P}(G_c, G_s)$ do not add up to 1. Hence, $\boldsymbol{P}\big(G_{case}, G_{S=Affected}|case, S = Affected\big) = \boldsymbol{P}(G_c, G_s)/\sum \boldsymbol{P}(G_c, G_s)$. The rows of $\boldsymbol{P}\big(G_{case}, G_{S=Affected}|case, S = Affected\big)$ add up to $P(G_c = \mathrm{bb}|case, S = Affected)$, $P(G_c = \mathrm{Bb}|case, S = Affected)$, and $P(G_c = \mathrm{BB}|case, S = Affected)$ respectively. This defines the risk allele frequency in cases with an affected sibling as
$p_{case\,|\,S=Affected} = P(G_c = \mathrm{BB}|case, S = Affected) + \frac{1}{2} P(G_c = \mathrm{Bb}|case, S = Affected)$.

Second, the allele frequency in cases with an affected sibling and unaffected parents was derived analoguously but with p_{NT} depending on G_c (see Appendix A in Section 2.5), and with $\boldsymbol{P}(G_c) = \mathrm{diag}\big(p(G|D, parents\ unaffected)\big)$.

2.3 Risk allele frequencies in pseudocontrols

Pseudo-control (pc) genotypes are the genomic complement genotypes from both parents not transmitted to their offspring. Allele frequencies in pseudocontrols depend on the genotypes of the cases selected, on the genotypes and disease statuses of the siblings and their IBD sharing with the cases. The genotype probabilities in pseudocontrols $P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c}, \mathrm{G_s}\big)$ were estimated as follows and the sum of these $4 * 3 * 3 = 36$ probabilities for a specific $\mathrm{G_{pc}}$ weighted by the probabilities of the genotypes in cases and controls and their IBD-sharing, gives $P(\mathrm{G_{pc}})$.

Define the matrices $\boldsymbol{P}\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, G_c, G_s\big)$ which has rows defined by genotypes of the cases and columns defined by the genotypes of the siblings

$$\begin{pmatrix} P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = bb, \mathrm{G_s} = bb\big) & P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = bb, \mathrm{G_s} = Bb\big) & P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = bb, \mathrm{G_s} = BB\big) \\ P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = Bb, \mathrm{G_s} = bb\big) & P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = Bb, \mathrm{G_s} = Bb\big) & P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = Bb, \mathrm{G_s} = BB\big) \\ P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = BB, \mathrm{G_s} = bb\big) & P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = BB, \mathrm{G_s} = Bb\big) & P\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, \mathrm{G_c} = BB, \mathrm{G_s} = BB\big) \end{pmatrix}$$

Given the parental genotype frequencies $P(G_p = bb)$, $P(G_p = Bb)$ and $P(G_p = BB)$, these $3\ (G_{pc}) * 4\ (IBD) = 12$ matrices follow from basic Mendelian reasoning and are displayed in Appendix B (Section 2.6). With these matrices the values of $P(\mathrm{G_{pc}} = bb)$, $P(\mathrm{G_{pc}} = Bb)$, and $P(\mathrm{G_{pc}} = BB)$ are separately estimated by

$$\boldsymbol{P}(G_{pc}|\mathrm{G_c}, \mathrm{G_s}, case, S = Affected) = \sum_{IBD} 0.25 * \boldsymbol{P}(\mathrm{G_c}, \mathrm{G_{s=Affected}}|\mathrm{IBD}) \circ \boldsymbol{P}\big(\mathrm{G}_{pc}\big|\mathrm{IBD}, G_c, G_s\big)$$

$$\boldsymbol{P}\big(\mathrm{G_{pc}}\big) = \sum \boldsymbol{P}(G_{pc}|\mathrm{G_c}, \mathrm{G_s}, case, S = Affected)$$

Where $\circ$ represent the Hadamard product of two matrices (i.e., when $A = B \circ C$, than $a_{ij} = b_{ij} * c_{ij}$). The probabilities $P(\mathrm{G_{pc}} = bb)$, $P(\mathrm{G_{pc}} = Bb)$, and $P(\mathrm{G_{pc}} = BB)$ do not add up to 1, because they are defined in terms of the full population. Therefore, $P\big(\mathrm{G_{pc}} \mid case, S = Affected\big)$ equal $P(\mathrm{G_{pc}})/\sum_{\mathrm{G_{pc}}} P(\mathrm{G_{pc}})$. This yields

the risk allele frequency in pseudocontrols from cases with affected siblings as $p_{pc\,|\,S=Affected} = P(\mathrm{G_{pc}} = BB) + \frac{1}{2}P(\mathrm{G_{pc}} = Bb)$.

The following variations yield the estimation for the other sets of pseudocontrols. (i) To estimate p_{pc} (without conditioning on affected siblings), replace $\boldsymbol{P}(\mathrm{G_c}, \mathrm{G_{s=Affected}}|\mathrm{IBD})$ by $\boldsymbol{P}(\mathrm{G_c}, \mathrm{G_s}|\mathrm{IBD})$ by substituting the diagonal matrix $\boldsymbol{P}(S = Affected|G_s)$ in the above for the identity matrix $\mathbb{I}$. (ii) To estimate $p_{pc|P=unaffected}$, adjust the parental genotype probabilities accordingly (no longer in HWE) and set $\boldsymbol{P}(G_c) = \mathrm{diag}\big(p(G|D, parents\ unaffected)\big)$. (iii) To estimated $p_{pc|S=Affected\ \&\ P=unaffected}$, combine the substitutions described in (i) and (ii).

2.4 Assortative mating

The impact of assortative mating on a single locus is expressed as the non-random mating fraction α of parents with similar genotypes. The next generation has the following frequencies[8]

$\boldsymbol{P}(G_c = bb|\ assortative\ mating\ parents) = (1-\alpha)q^2 + \alpha(q^2 + \frac{1}{2}pq)$,
$\boldsymbol{P}(G_c = Bb|\ assortative\ mating\ parents) = (1-\alpha)2pq + \alpha pq$, and
$\boldsymbol{P}(G_c = BB|\ assortative\ mating\ parents) = (1-\alpha)p^2 + \alpha(p^2 + \frac{1}{2}pq)$,

when the parental generation is in HWE, and with p the parental frequency of B and q of b. The genotype probabilities of affected siblings are given by $\boldsymbol{P}(\mathrm{G}|\mathrm{D}, a.m.parents) = P(\mathrm{D}|\mathrm{G})P(\mathrm{G}|a.m.\ parents)/P(\mathrm{D})$ analoguous to Section 2.1. Substituting these as $\boldsymbol{P}(\mathrm{G_c})$ in Eq 1 in Section 2.2

$\boldsymbol{P}(\mathrm{G_c}, \mathrm{G_s}|\mathrm{IBD}, a.m.parents) = \boldsymbol{P}(\mathrm{G_c}) * \boldsymbol{P}(G_s|\ G_c, IBD) * \mathbb{I}$,

and following the other steps in Sections 2.1 and 2.2 gives the frequencies of cases and pseudocontrol of parents with assortative mating (not selecting of disease-status of parents or siblings). Note that assortative mating changes the probabilities of the combined genotypes of parents, which is described in Appendix A (Section 2.5).

2.5 Appendix A: allele and genotype frequencies of non-transmitted alleles

When the parents are unaffected, they are not in HWE, in which case the non-transmitted allele and genotype frequencies are dependent on the case's (proband's) genotype G_c. These non-transmitted allele and genotype frequencies are needed to derive the combined probabilities of genotypes in cases and their sibling $\boldsymbol{P}(G_c, G_s)$. (Note that these non-transmitted alleles are not the pseudocontrols of interest.) Suppose the genotypes in the parents have frequencies $P(G_p = bb)$, $P(G_p = Bb)$ and $P(G_p = BB)$. The distribution of the genotypes of pairs of parents with a genotype correlation (non-random mating fraction) α is given by

$$\boldsymbol{P}\left(G_{father}G_{mother}\right) = \begin{pmatrix} P(G_f = bb, G_m = bb) \\ P\left(G_f = b\&b, G_m = Bb\right) \\ P\left(G_f = b\&b, G_m = BB\right) \\ P\left(G_f = B\&b, G_m = bb\right) \\ P\left(G_f = B\&b, G_m = Bb\right) \\ P\left(G_f = B\&b, G_m = BB\right) \\ P\left(G_f = BB, G_m = bb\right) \\ P\left(G_f = BB, G_m = Bb\right) \\ P\left(G_f = BB, G_m = BB\right) \end{pmatrix}$$

$$= \begin{pmatrix} (1-\alpha)P\left(G_p = bb\right)P\left(G_p = bb\right) + \alpha P\left(G_p = bb\right) \\ (1-\alpha)P\left(G_p = bb\right)P(G_p = Bb) \\ (1-\alpha)P\left(G_p = bb\right)P(G_p = BB) \\ (1-\alpha)P\left(G_p = Bb\right)P(G_p = bb) \\ (1-\alpha)P\left(G_p = Bb\right)P\left(G_p = Bb\right) + \alpha P\left(G_p = Bb\right) \\ (1-\alpha)P\left(G_p = Bb\right)P(G_p = BB) \\ (1-\alpha)P\left(G_p = BB\right)P(G_p = bb) \\ (1-\alpha)P\left(G_p = BB\right)P(G_p = Bb) \\ (1-\alpha)P\left(G_p = BB\right)P\left(G_p = BB\right) + \alpha P\left(G_p = BB\right) \end{pmatrix}$$

The distributions of the genotypes of pairs of parents conditional on their offspring G_c are proportional to the pairwise multiplications of the probability of these parental genotypes times the probability of getting offspring with G_c, that is

$$\widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = bb\right) =$$
$$\boldsymbol{P}\left(G_{father}G_{mother}\right)*(1\ \ 0.5\ \ 0\ \ 0.5\ \ 0.25\ \ 0\ \ 0\ \ 0\ \ 0)^T$$
$$\widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = Bb\right)$$
$$= \boldsymbol{P}\left(G_{father}G_{mother}\right) * (0\ \ 0.5\ \ 1\ \ 0.5\ \ 0.5\ \ 0.5\ \ 1\ \ 0.5\ \ 0)^T$$
$$\widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = BB\right)$$
$$= \boldsymbol{P}\left(G_{father}G_{mother}\right) * (0\ \ 0\ \ 0\ \ 0\ \ 0.25\ \ 0.5\ \ 0\ \ 0.5\ \ 1)^T$$

The probabilities of non-transmitted (NT) genotypes are proportional to the sum of the combined parental genotypes resulting in this NT genotype, that is

$$\widetilde{\boldsymbol{P}}(NT = bb|G_c = bb) = (1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = bb\right)$$
$$\widetilde{\boldsymbol{P}}(NT = Bb|G_c = bb) = (0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = bb\right)$$
$$\widetilde{\boldsymbol{P}}(NT = BB|G_c = bb) = (1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = bb\right)$$
$$\widetilde{\boldsymbol{P}}(NT = bb|G_c = Bb) = (0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = Bb\right)$$
$$\widetilde{\boldsymbol{P}}(NT = Bb|G_c = Bb) = (0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = Bb\right)$$
$$\widetilde{\boldsymbol{P}}(NT = BB|G_c = Bb) = (0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = Bb\right)$$
$$\widetilde{\boldsymbol{P}}(NT = bb|G_c = BB) = (0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = BB\right)$$
$$\widetilde{\boldsymbol{P}}(NT = Bb|G_c = BB) = (0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = BB\right)$$
$$\widetilde{\boldsymbol{P}}(NT = BB|G_c = BB) = (0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1) * \widetilde{\boldsymbol{P}}\left(G_{father}G_{mother}|G_c = BB\right)$$

Scaling gives the exact probabilities of the NT genotypes: $P(NT = bb|G_c = bb) = \tilde{P}(NT = bb|G_c = bb)/\left(\tilde{P}(NT = bb|G_c = bb) + \tilde{P}(NT = Bb|G_c = bb) + \tilde{P}(NT = BB|G_c = bb)\right)$ etc. The allele frequencies $p_{NT|G_c}$ follow directly from the NT genotype frequencies.

2.6 Appendix B: pseudocontrol genotypes conditional on IBD, G_c and G_s

Define the matrices $\boldsymbol{P}\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, G_c, G_s\right)$ as

$$\begin{pmatrix} P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = bb, \mathrm{G_s} = bb\right) & P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = bb, \mathrm{G_s} = Bb\right) & P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = bb, \mathrm{G_s} = BB\right) \\ P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = Bb, \mathrm{G_s} = bb\right) & P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = Bb, \mathrm{G_s} = Bb\right) & P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = Bb, \mathrm{G_s} = BB\right) \\ P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = BB, \mathrm{G_s} = bb\right) & P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = BB, \mathrm{G_s} = Bb\right) & P\left(\mathrm{G}_{pc}\middle|\mathrm{IBD}, \mathrm{G_c} = BB, \mathrm{G_s} = BB\right) \end{pmatrix}$$

Given the parental genotype frequencies $P(G_p = bb)$, $P(G_p = Bb)$ and $P(G_p = BB)$, these $3 * 4 = 12$ matrices follow from basic Mendelian reasoning. Note that IBD=0 (between cases and their siblings) indicates that the pseudocontrol shares both alleles with the sibling; IBD=1 indicates that the pseudocontrol shares

the non-IBD allele with the sibling; and IBD=2 indicates that the pseudocontrol and sibling share no alleles. Alleles in the pseudocontrols not shared with the sibling come from the parents with the probabilities derived in Appendix A (Section 2.5). The $\boldsymbol{P}(\mathrm{G}_{pc}|\mathrm{IBD})$ are thus defined as:

$$\boldsymbol{P}(\mathrm{G}_{pc} = bb|\mathrm{IBD} = 0) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{P}(\mathrm{G}_{pc} = bb|\mathrm{IBD} = \mathrm{b}) = \begin{pmatrix} q_{NT|G_c=bb} & 0 & 0 \\ q_{NT|G_c=Bb} & 0 & 0 \\ q_{NT|G_c=BB} & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{P}(\mathrm{G}_{pc} = bb|\mathrm{IBD} = \mathrm{B}) = \begin{pmatrix} q_{NT|G_c=bb} & q_{NT|G_c=bb} & 0 \\ q_{NT|G_c=Bb} & q_{NT|G_c=Bb} & 0 \\ q_{NT|G_c=BB} & q_{NT|G_c=BB} & 0 \end{pmatrix}$$

$$\boldsymbol{P}(\mathrm{G}_{pc} = bb|\mathrm{IBD} = 2) = \begin{pmatrix} P(NT = bb|G_c = bb) & P(NT = bb|G_c = bb) & P(NT = bb|G_c = bb) \\ P(NT = bb|G_c = Bb) & P(NT = bb|G_c = Bb) & P(NT = bb|G_c = Bb) \\ P(NT = bb|G_c = BB) & P(NT = bb|G_c = BB) & P(NT = bb|G_c = BB) \end{pmatrix}$$

$$\boldsymbol{P}(\mathrm{G}_{pc} = Bb|\mathrm{IBD} = 0) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\boldsymbol{P}(\mathrm{G}_{pc} = Bb|\mathrm{IBD} = \mathrm{b}) = \begin{pmatrix} p_{NT|G_c=bb} & q_{NT|G_c=bb} & q_{NT|G_c=bb} \\ p_{NT|G_c=Bb} & q_{NT|G_c=Bb} & q_{NT|G_c=Bb} \\ p_{NT|G_c=BB} & q_{NT|G_c=BB} & q_{NT|G_c=BB} \end{pmatrix}$$

$$\boldsymbol{P}(\mathrm{G}_{pc} = Bb|\mathrm{IBD} = \mathrm{B}) = \begin{pmatrix} p_{NT|G_c=bb} & p_{NT|G_c=bb} & q_{NT|G_c=bb} \\ p_{NT|G_c=Bb} & p_{NT|G_c=Bb} & q_{NT|G_c=Bb} \\ p_{NT|G_c=BB} & p_{NT|G_c=BB} & q_{NT|G_c=BB} \end{pmatrix}$$

$$\boldsymbol{P}(\mathrm{G}_{pc} = Bb|\mathrm{IBD} = 2) = \begin{pmatrix} P(NT = Bb|G_c = bb) & P(NT = Bb|G_c = bb) & P(NT = Bb|G_c = bb) \\ P(NT = Bb|G_c = Bb) & P(NT = Bb|G_c = Bb) & P(NT = Bb|G_c = Bb) \\ P(NT = Bb|G_c = BB) & P(NT = Bb|G_c = BB) & P(NT = Bb|G_c = BB) \end{pmatrix}$$

$$\boldsymbol{P}\left(\mathrm{G}_{pc} = BB \middle| \mathrm{IBD} = 0\right) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\boldsymbol{P}\left(\mathrm{G}_{pc} = BB \middle| \mathrm{IBD} = \mathrm{b}\right) = \begin{pmatrix} 0 & p_{NT|G_c=bb} & p_{NT|G_c=bb} \\ 0 & p_{NT|G_c=Bb} & p_{NT|G_c=Bb} \\ 0 & p_{NT|G_c=BB} & p_{NT|G_c=BB} \end{pmatrix}$$

$$\boldsymbol{P}\left(\mathrm{G}_{pc} = BB \middle| \mathrm{IBD} = \mathrm{B}\right) = \begin{pmatrix} 0 & 0 & p_{NT|G_c=bb} \\ 0 & 0 & p_{NT|G_c=Bb} \\ 0 & 0 & p_{NT|G_c=BB} \end{pmatrix}$$

$$\boldsymbol{P}\left(\mathrm{G}_{pc} = BB \middle| \mathrm{IBD} = 2\right) = \begin{pmatrix} P(NT = BB|G_c = bb) & P(NT = BB|G_c = bb) & P(NT = BB|G_c = bb) \\ P(NT = BB|G_c = Bb) & P(NT = BB|G_c = Bb) & P(NT = BB|G_c = Bb) \\ P(NT = BB|G_c = BB) & P(NT = BB|G_c = BB) & P(NT = BB|G_c = BB) \end{pmatrix}$$

REFERENCES

1. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: Inferring the contribution of common variants. Proc. Natl. Acad. Sci. U. S. A. *111*, E5272–E5281.

2. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

3. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

4. Bulmer, M. (1985). The mathematical theory of quantitative genetics (Oxford: Clarendon press).

5. Lynch, M., and Walsh, B. (1998). Genetics and analysis of quantitative traits. (Sunderland: Sinauer),.

6. R Core Team (2015). R: A Language and Environment for Statistical Computing.

7. Witte, J.S., Visscher, P.M., and Wray, N.R. (2014). The contribution of genetic variants to disease depends on the ruler. Nat. Rev. Genet. *15*, 765–776.

8. Falconer, D., and Mackay, T. (1996). Introduction to quantitative genetics (Essex: Longman).

9. Tallis, G.M. (1987). Ancestral covariance and the Bulmer effect. Theor. Appl. Genet. *73*, 815–820.

Chapter 8

Summary & General discussion

SUMMARY OF THESIS FINDINGS

This thesis aimed to study the complex link between genetic effects and measured environmental risk factors in major depressive disorder (MDD) in empirical data, and to explore boundaries of the consequences of two Genome-Wide Association study (GWAS) designs and assortative mating from a theoretical perspective. Both genes and environment affect MDD risk, but it remains unclear whether both act independently, whether they interact, or whether environmental risk might actually reflect shared genetic effects between MDD risk and e.g. behavioral traits that might increase environmental stress.

The most studied candidate gene in MDD is the serotonin transporter gene. The length polymorphism in the promoter region of this gene (5-HTTLPR) has been hypothesized to influence MDD risk, because an important group of antidepressant drugs acts on the serotonin transporter. A Science paper by Caspi et al suggested that this gene mainly exerts its influence when persons had been exposed to childhood trauma.[1] In **Chapter 2**, the 5-HTTLPR polymorphism was analyzed in 1593 cases and 1411 controls from the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR). In these individuals, four different outcome measures were defined: lifetime MDD, suicidal MDD, chronic MDD, and course of MDD (chronic versus non-chronic). No evidence was found for either direct effects of 5-HTTLPR on these outcome measures or interaction effects between 5-HTTLPR and five environmental risk factors for MDD: lifetime stressful life-events, recent stressful life-events, sexual abuse, childhood trauma, and educational attainment (as proxy for social economic status associated with increased stress).

In **Chapter 3**, the relation between MDD and educational attainment was investigated in approximately 25,000 individuals from the Psychiatric Genomics Consortium (PGC wave 1) with additional Dutch and Estonian data. An increased risk for MDD was confirmed in individuals with lower educational attainment. Subsequently, the possible contribution of shared genetic effects to this link was assessed with three different methods applying data of 884,105 autosomal common single-nucleotide polymorphisms (SNPs). Firstly, polygenic risk scores (PRS) based on GWAS results on education attainment in ~120,000 individuals (EA-PRS) did not affect MDD risk, and PRS based on MDD GWAS results in ~20,000 individuals (MDD-PRS) did not affect EA. Secondly, a non-consistent weak significant negative genetic correlation was found with bivariate genomic-relationship-matrix restricted maximum likelihood (GREML). Thirdly, no concordance was found in either significance or direction of SNP effects across MDD GWAS and EA GWAS results based on SNP effect concordance analysis

(SECA). To conclude, these findings indicate that it is unlikely that shared genetic effects explain a large proportion of the link between MDD risk and lower education attainment, but a small genetic contribution to this deleterious link could not be excluded.

Up to 2014, research on gene-by-environment (GxE) interaction in MDD had mainly focused on candidate genes, such as 5-HTTLPR. However, with the emergence of MDD cohorts with genome wide SNP data, novel methods were developed that allowed to tag genome-wide genetic MDD risk with polygenic risk scores (PRS). In **Chapter 4**, PRS were constructed in 1645 cases and 340 controls from NESDA based on discovery results from the large Psychiatric Genomics Consortium (PGC wave 1). These PRS were found to have an increased impact on MDD risk in individuals exposed to childhood trauma (CT), which suggested gene-by-environment interaction on a genome-wide scale. This interaction-effect was found both as departure from multiplicativity (combined impact of PRS and CT larger than the *product* of the individual effects) and as departure from additivity (combined impact larger than the *sum* of the individual effects), the latter of which has been hypothesized to be more plausible from a biological perspective.

In **Chapter 5**, the interaction between polygenic risk scores (PRS) and childhood trauma (CT) was further tested in seven cohorts from PGC (wave 2) with CT information available in 3,024 cases and 2,741 controls. CT had consistent impact across cohorts, with similar impact in males and females. However, the interaction effects were heterogeneous with a positive interaction effect in NESDA (as in Chapter 4), negative interaction effect in the Radiant-UK study, and no interaction in the other contributing five cohorts, resulting in no overall evidence for interaction between PRS and CT in MDD. The results from Chapter 5 illustrate the heterogeneity of MDD, and suggest that the results of Chapter 4 can best be interpreted as a single cohort phenomenon.

The focus was switched from analyzing empirical data to theoretic work in **Chapter 6.** Two GWAS study designs applied in the Psychiatric Genomics Consortium (PGC) were considered with respect to their power of SNP association analysis and SNP-heritability estimates (proportion of population-variance in disease-risk attributable to genome-wide common SNPs). First, parent-affected-offspring trio data are regularly applied in the subgroups of the PGC analyzing autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD). Trio data are essential to detect *de novo* mutations, but its use may results in reduced power in association analysis and underestimation of the SNP-heritability compared to analyses in case screened-control data. This difference is attributable to (i) potential oversampling of multiplex families (with more than

one affected offspring) and to (ii) assortative mating, which describes the correlation between mating partners in vulnerability for ADHD or ASD. Second, the use of poorly or unscreened controls for common disorders will result in decreased power in association analysis and decreased SNP-heritability estimates. In particular, for MDD with a lifetime prevalence of around 15%, the anticipated underestimation of the SNP-heritability was analytically derived at 28% when none of the controls would be screened. An updated equation was provided to properly scale the SNP-heritability when including unscreened controls (Equation 3 in **Chapter 6**). When aiming to analyze the polygenic effects in psychiatric disorders, it is advisable not to use trio data and to properly scale the SNP-heritability when applying data with unscreened controls.

Research has found significant partner-resemblances for psychiatric disorders, that is, mating partners are more often concordant in psychiatric disorder-status than expected by chance. This phenomenon is often referred to as assortative mating. However, the potential consequences of these partner-resemblances have not been quantified and have been left implicit despite available theory in the quantitative genetics literature. Therefore, in **Chapter 7**, boundaries were quantified for the anticipated consequences for disorder prevalence and heritability under various inevitable assumptions. The consequences are most pronounced when partner-resemblance is attributable to phenotypic assortment (partner-resemblance driven by the psychiatric trait), and are reflected in increased population prevalence and heritability in the offspring generation. From the first generation in which assortative mating takes place, the consequences add generation after generation to reach equilibrium asymptotically over generations. Because of this equilibrium, assortative mating is unlikely to balance the impact of reduced fecundity of psychiatric patients in the long term, as analytically derived in **Chapter 7**. Modeling suggests that the heritability of none of the psychiatric disorders considered is likely to increase with more than 5% from one generation of assortative mating (or 13% over several generations). The population prevalence will increase most for rare disorders with high heritability, such as the prevalence of ASD that might maximally increase 1.5-fold after one generation of assortative mating (or 2.4-fold over several generations). While emphasizing the limitations inherent to the inevitable model assumptions, genetic theory suggests that the consequences of assortative mating are, at most, modest for the heritability, but may be considerable for the population prevalence.

DISCUSSION OF THESIS FINDINGS

GxE interaction research with candidate genes

Over a decade of research on gene-by-environment (GxE) interaction with candidate genes in major depressive disorder (MDD) has led to contradicting findings from which a pattern of non-replication has emerged. **Chapter 2** of this thesis further strengthens the pattern of non-replication of the interaction effect between 5-HTTLPR and childhood trauma presented by Caspi et al.[1] In addition, a critical review from Duncan et al suggested that the GxE literature from the candidate-gene era suffered from publication bias, because 96% of novel GxE studies yielded significant results compared to only 27% of replication studies, and because smaller replication studies reported more significant results than the larger samples.[2] To conclude, it seems unlikely that the Caspi finding is generalizable to other cohorts, and there doesn't appear to be much ground to further study interaction between childhood trauma and 5-HTTLPR in MDD.

GxE with genome-wide information

Chapter 4 of this thesis describes a positive statistical interaction between polygenic risk scores and childhood trauma in the Netherlands Study of Depression and Anxiety.[3] In the context of the critical review of Duncan et al of the literature on GxE with candidate genes,[2] was it worth to publish this novel finding while based on one cohort only? I think it was, as these analyses had several marked differences with the candidate gene literature. First, polygenic risk scores have a significant impact on MDD that is consistent across different cohorts,[4–6] which contrasts the small impact of 5-HTTLPR on MDD that was found in a large meta-analyses showing considerable variety across individual cohorts.[7] Interaction effects can result in only a small main genetic effect, but this requires that the effects in different environmental strata balance exactly. Testing for interaction with polygenic risk scores with confirmed consistent main effects might be more powerful than testing for interaction with candidate genes with small inconsistent effects or no main effect at all.[8] Second, although there were several different polygenic risk scores that could have been tested, for example based on a discovery GWAS on schizophrenia instead of MDD, I felt there was one candidate of most interest: the genome-wide set of polygenic risk scores based on the largest GWAS for MDD at the time.[4] This apparent choice contrasted the many independent candidate genes that were tested contributing to multiple testing and suggested publication bias in the candidate gene literature.[2] Nevertheless, there were of course many environmental risk factors that we could have tested. Exposure to childhood trauma was chosen as environmental

risk factor, because this is one of the strongest and most consistent risk factors with a lifelong impact on MDD risk, which has even been hypothesized to distinguish a neurobiological distinct subtype of MDD.[9] Taken all together, I think that testing for interaction between polygenic risk scores and childhood trauma in MDD was not only novel, but also an obvious step to take at the time, and I think the NESDA finding was justifiably published as a single cohort result. Nevertheless, this single cohort finding confirming polygenic GxE interaction was, subsequently, not replicated by Mullins et al who found evidence for an opposing interaction effect.[10] Analyses combining data of seven cohorts totaling 5,765 individuals showed no overall evidence for interaction in either direction (**Chapter 5** of this thesis).

Interpreting PRSxCT results

How to interpret the positive interaction effect between polygenic risk scores (PRS) and childhood trauma found in **Chapter 4** with NESDA data, contrasting the negative interaction effect found by Mullins et al with the Radiant-UK data?[10] Most importantly, these contrasting findings seem to illustrate the genetic heterogeneity in MDD, which had already been indicated by Lee et al who found that the genetic coheritability between different MDD cohorts showed more variety than between e.g. different schizophrenia cohorts.[11] In **Chapter 5**, an attempt was made to aid further interpretation with a simulation study. Therefore, several different scenarios of underlying genetic architecture of MDD and childhood trauma were simulated, followed by comparison of interaction-effect estimates with results from empirical data, and by comparison of the patterns of mean PRS in exposed controls, unexposed controls, exposed cases and unexposed cases. From simulation, the typical pattern seemed that exposed cases had lower PRS than unexposed cases, and exposed controls had lower PRS than unexposed controls, explained by the fact that those exposed to childhood trauma require less polygenic risk to become affected. Notably, this pattern does not reflect GE-correlation. In the full population, exposed controls are relatively rare as are unexposed cases. Thus, the mean PRS in all unexposed individuals shifts towards the mean in controls and the mean PRS in exposed individuals shifts towards the mean in cases, overall resulting in equal mean PRS in exposed and unexposed individuals in the full population.

It appeared that the NESDA-findings were unlikely to primarily represent different directions of SNP effects in exposed and unexposed individuals, because this would have resulted in a negative interaction-effect attributable to the discovery GWAS that would have primarily tagged effects in unexposed

individuals (the prevalence of CT is approximately 0.25). Rather, the NESDA-findings seemed to best fit the simulated scenario with either decreased environmental variance or increased genetic variance in exposed individuals, explaining the increased effect of PRS in exposed individuals. At first sight, the negative interaction effect in the Radiant-UK study appeared consistent with a different direction of genetic effects in exposed compared to unexposed individuals as just discussed. However, the mean PRS in Radiant showed a distinct pattern with larger PRS in exposed cases than unexposed cases, and larger PRS in exposed controls than in unexposed controls. This pattern seemed to fit simulated data with genetic effects impacting on CT (not ruling out difference of genetic effects, of course). However, this comparison between simulated and empirical data only provides a rough feeling for the genetic architectures that the NESDA-findings and Radiant-findings might be compatible with, because simulation didn't allow for different genetic architectures between discovery and target set, and because simulation was based on several inevitable and partly arbitrary assumptions. In addition, we have to keep in mind that, although interesting from a theoretical perspective, the heterogeneity across the six cohorts studied in **Chapter 5** does not justify too detailed interpretation. Nevertheless, I think more insight is obtained by attempting to understand patterns from PRS interaction-analyses with simulation, than when not attempting at all. Importantly, non-consistent findings in PRSxCT interaction in MDD do not suggest that no SNPxCT interaction effects exist in MDD. For example, if a proportion of 10% of MDD SNPs would be moderated by childhood trauma, their interaction-effects would likely be diluted in PRS-analyses, while interaction in 10% of affective SNPs would be a relevant phenomenon to study.

Methodological challenges in GxE interaction analyses

Several papers have described methodological challenges of GxE analyses in twin studies and candidate gene studies, and these challenges may also apply to tests for GxE analyses with polygenic risk scores. Purcell discussed gene-environment (GE) correlation, which describes the impact of genetic variants on environmental exposure via e.g. personality characteristics, or when considering childhood trauma, via a link of transmitted genetic variants with personality characteristics in the parents.[12] Purcell explained that increased genetic effects in an environmental condition could result from moderated genetic effects (GxE), but also from risk variants being more likely to be present in that environmental condition as a consequence of GE-correlation. Notably, within NESDA and the Radiant-UK study, the PRS based on MDD discovery results were correlated to CT.

Therefore, a simulation study was performed in **Chapter 5** to assess the potential impact of GE-correlation. These simulations indeed confirmed an inflated type I error rate in the context of GE-correlation, but to a modest extent of 0.075 (with alpha set at 0.05) for a strong correlation of 0.3 between G and E. This indicates that the NESDA and Radiant-UK findings are most likely to represent non-spurious single cohort phenomena.

Apart from GE-correlation, Eaves (2005) showed with simulation that even when ruling out gene-environment correlation, spurious GxE results can still be found when e.g. the disorder liability would be non-normally distributed, thereby reemphasizing the fragile nature of tests for interaction.[13] In his work, Eaves did not provide means to disentangle spurious results from true interaction effects, but he predicted that interaction effects would be common rather than specific when they would be attributable to a non-normally distributed disorder liability. This pattern was not found in the years following 2005, although a considerable number of GxE studies were conducted.[2] In 2014, Keller showed that many GxE studies might have overestimated or underestimated interaction effects by improperly correcting for covariates. That is, covariates were typically included only for their main effects whereas their interaction effects ought to also be included (covariates times G and covariates times E).[14] The analyses in **Chapter 4** did not comply with these recommendations, but in **Chapter 5** the analyses were corrected also for the interaction effects of the covariates and showed a similar positive interaction effect within NESDA.

Future perspective on GxE interaction research in MDD

The non-consistent findings in PRSxCT interaction in MDD do not suggest that no SNPxCT interaction effects may exist in MDD. However, the question is how to best test for SNPxCT effects in MDD. The risk for publication bias with candidate genes (or SNPs) underlines the importance of a hypothesis free approach.[2] Assuming that interaction effects, if existing, will be of the same magnitude as main SNP effects, GWAS samples with CT information of tens of thousands of cases and controls will likely be needed to reach genome-wide significance. Alternatively, a two-step approach could be applied to increase power by first selecting SNPs with e.g. a correlation with CT in the combined case-control sample and second testing these SNPs for interaction with CT in predicting MDD.[15] The increase in power from a two-step approach can be considerably,[15] and I think these analyses should be conducted within the context of MDD, although I wouldn't be surprised if no consistent interaction effects would be detected given the large genetic heterogeneity of MDD.[11]

Assessing CT information in additional numbers of MDD cases and controls will assist SNPxCT research, but also PRSxCT analyses. Notably, the PRS applied in Chapter 4 and Chapter 5 were based on discovery GWAS results from samples with unknown mixtures of individuals exposed and unexposed to CT. This complicates interpretation of results and reduces power of interaction analyses. It would be preferable to have two distinct discovery GWASs, one exclusively based on unexposed individual and one exclusively based on exposed individuals. Unfortunately, current sample sizes did not allow for this approach, but these analyses might be of particular interest. There is a third possible advantage of assessing CT information in large numbers of cases and controls: a GWAS in unexposed individuals (approximately 75% of the original sample depending on the definition of CT) may provide increased power to detect SNPs associated to MDD, because unexposed individuals require more genetic risk on average to become affected.

An important challenge will come from the choice of the environmental factor to test, a choice for which no hypothesis-free work-around is available. I have argued that CT is a plausible candidate, because of its strong and lifelong impact on MDD risk, and because exposure to CT has been hypothesized to distinguish a neurobiological distinct subgroup of MDD.[9] However, many other environmental conditions can also be tested, such as stressful life-events later in life, socioeconomic status or air pollution, inevitably introducing (hidden) multiple-testing burden risking false-positive findings. One might argue to adjust the level of significance according to the number of environmental factors tested, but this seems an unfeasible approach, in particular when tests for different environmental factors are presented in different papers. Indeed, a more plausible approach would be to emphasize the importance of independent replication, and to regard single-study results as no more than hypothesis forming.

In addition to the challenges with respect to power and choice of environmental risk factor, research on GxE interaction in MDD is further complicated and confused by the methodological challenges intrinsic to GxE testing. A choice needs to be made whether to test for interaction as departure from multiplicativity or as departure from additivity. Although the latter has been hypothesized to be more in line with meaningful biological interpretation, most interaction analyses test for interaction as departure from multiplicativity as these readily follow from logistic regression. The challenges and potential methodological pitfalls do not create optimism to test for GxE-interaction in MDD. However, when a well-replicated GxE interaction effect would be found, with well-understood biological interpretation with respect to e.g. gene-

inactivation by methylation, this would be of major importance for understanding the complex pathophysiology of MDD. Taken all together, I think two-step SNPxE analyses should be conducted in the short term,[15] but I fear that the current MDD data may present too many challenges to properly test for SNPxE interaction. Therefore, I recommend that researches who collect samples for GWAS studies consider to collect environmental information in a uniform manner to prepare for solid SNPxE testing on a large scale in the years to come. Alternatively, I would advice researcher to at least facilitate follow-up of study participants and obtain permission for record linkage.

Depression, educational attainment, and genetic correlation

In **Chapter 3**, data of approximately 25,000 individuals were applied to test whether the phenotypic link between lower educational attainment (EA) and increased MDD risk could be partly attributable to shared genetic effects (or LD between effective loci on MDD and EA). Therefore, three methods were applied: bivariate genomic relatedness matrix restricted maximum likelihood (GREML) analyses, EA polygenic risk scores predicting MDD, MDD polygenic risk score predicting EA, and SNP effect concordance analysis (SECA). None of these methods showed consistent evidence for genetic correlation between MDD and EA, indicating that genetic effects are not expected to explain much of the phenotypic link between MDD and EA. Notably, these analyses required considerable computational time, in particular the GREML analyses on the genomic relatedness matrix including 312,512,500 elements. Since then, a new method, LD score regression, has been developed illustrating the great progress in genetic research, because tests for genetic correlation can now be conducted within one minute with LD score regression based on GWAS summary statistics only.[16] Indeed, although the analyses described in **Chapter 3** resulted in a paper on itself in 2014,[6] the first LD score regression paper in 2015 presented the genetic correlations between 24 traits, i.e. 276 times the number of genetic correlation estimates. This study also found no evidence for genetic correlation between years of education and MDD risk.[16] Interestingly, the Social Science Genetics Consortium recently found a negative genetic correlation between neuroticism and educational attainment of -0.42 (SE=0.07, p=2.8e-8),[17] which may seem to contrast the finding of **Chapter 3**. This difference could be attributable to the more accurate estimation of genetic effects on educational attainment as sample size has increased from approximately 125,000[18] to approximately 300,000,[17] and because MDD and neuroticism are closely linked but not quite the same with genetic correlation estimated at 0.66 (SE=0.09,

p=2.8e-14)[19] and at 0.75 (SE=0.027) between depressive symptoms and neuroticism.[20]

LD score regression has considerably narrowed the need for bivariate GREML and PRS analyses when aiming to test for the genetic correlation between two traits. LD score regression has slightly less power than GREML analyses and is not feasible when GWAS results are based on a small sample with less than around 5,000 individuals.[16,21] For samples including less than 5,000 individuals, bivariate GREML and PRS analyses can be considered, given that individual level genotype data are available. For samples with less than around 3,000 to 4,000 individuals, bivariate GREML may be underpowered,[22] and tests for genetic correlation can then be performed with PRS-analysis, given that adequately powered discovery GWAS results are available for one of the traits. Notably, in **Chapter 3**, the phenotypic link between lower EA and MDD risk was confirmed with phenotypic data, the analyses of which cannot be replaced with LD score regression.

Caution is required when testing for genetic correlation with bivariate GREML for disease-traits, because of concerns raised by Golan et al about univariate GREML in estimating SNP-heritability in cases-control data.[23] In case-control data, oversampling of cases introduces correlation between genetic and environmental effects, which violates the assumptions underlying GREML analysis and results in underestimation of the SNP-heritability. Golan et al advised applying cross-product Haseman-Elston (HE) regression, which is anticipated to yield unbiased estimates of the SNP-heritability. Unfortunately, Golan et al did not address bivariate GREML analyses in the context of case-control data, and I am not aware of any clear advice in this aspect. However, I would advice caution, as it seems plausible that case-control data may also result in biased estimates from bivariate GREML analyses. The proposed method of Golan et al to correct for covariates does not naturally extend to bivariate analysis. Here, I would advice a pragmatic approach, which is to present results from both bivariate GREML analyses and bivariate HE-regression with the residuals of regression of the two traits of interest on the relevant covariates.

SNP heritability of psychiatric traits

The proportion of variance in disease risk attributable to genotyped SNPs (SNP heritability) can be assessed with a variety of methods on different types of case-control data. The methods most often applied are genomic relatedness matrix restricted maximum likelihood (GREML),[24] Haseman-Elston (HE) regression,[25] LD score regression,[21] and, less frequently, the method from So et al.[26] SNP

heritability estimates are typically lower than family-study based heritability estimates,[11] a phenomenon referred to as the missing heritability,[27] which has been hypothesized to be attributable to, for example, non-genotyped genetic risk variants.[28] Indeed recently, Yang et al. found negligible missing heritability for height and BMI when imputing rare genetic variants and applying GREML stratified for minor allele frequency and linkage disequilibrium.[29] For case-control data, a different method has been proposed that constructs haplotypes from common SNPs to tag rare variants followed by HE-regression stratified by minor allele frequency. With this method a haplotype heritably of 0.64 for schizophrenia was found compared to SNP heritability of 0.32.[30] In particular, SNP heritability estimates from case-control data also depend on the method applied as Golan et al found that GREML likely underestimates the SNP heritability, because of correlation between genetic and environmental effects introduced by oversampling of cases typical in case-control studies.[23] Golan et al found that HE regression is robust against oversampling of cases, and it can be assumed that the same holds for LD score regression, which provides similar SNP heritability estimates as HE regression.[31]

In **Chapter 6**, the consequences of study design on SNP heritability estimates were explored, with a specific focus on the designs applied in the Psychiatric Genomics Consortium.[32] As shown, the SNP heritability is likely underestimated when analyzing parent-offspring trio data, that is regularly applied in childhood onset disorders such as autism (ASD) and attention deficit/ hyperactivity disorders (ADHD). In addition, the SNP heritability will be underestimated when analyzing data of unscreened or poorly screened controls for common disorders, such as major depressive disorder. An updated equation is provided for estimating SNP heritability from data with unscreened controls (Equation 3 in **Chapter 6**). This equation corrects the expected underestimation of $(1 - Ku)^2$ for a disorder with a population prevalence K, that is studied in a sample with a proportion of unscreened controls u. Ideally, the SNP heritability underestimation from trio data would have also been captured in an equation, but this was considerably more complex theoretically, and found to depend on (the often unknown) proportion of multiplex families in the study, and the amount of assortative mating for the disorder under consideration. Nevertheless, a considerable underestimation of the SNP heritability is expected from trio data when oversampling multiplex families, or under assortative mating that has been confirmed for both ADHD and ASD (Chapter 6 Figure 1).[33] Thus, I advise against the use of trio data for SNP heritability estimation, and advice to appropriately

scale the SNP heritability when applying data with unscreened controls for common disorders.

Assortative mating

Recently, Nordsletten et al published a comprehensive study indicating a clear pattern of nonrandom mating within and across eleven major psychiatric disorders based on over 700,000 psychiatric patients from the Swedish population.[33] Although the partner-resemblances found were pronounced (e.g. a partner-correlation of 0.47 for autism), Nordsletten et al did not discuss the expected consequences for disorder prevalence and heritability in the offspring generation, despite available theory in the quantitative genetics literature. In **Chapter 7,** upper boundaries were explored of the consequences of assortative mating for psychiatric traits by applying quantitative genetic models. First of all, the correlation estimates of Nordsletten et al were found to likely overestimate the correlations in the full population, because they analyzed a study sample with oversampling of cases. For example, the partner-correlation in the full population for autism was approximated at around 0.28 rather than the 0.47 presented by Nordsletten et al. Based on several inevitable assumptions discussed in detail in **Chapter 7**, the heritability is likely to increase as a consequence of assortative mating with an upper boundary of around 5% for one generation of assortative mating and 13% for multiple generations. In addition, the population prevalence is expected to increase, with a more pronounced impact for strong assortative mating (ρ), for disorders with a low population prevalence (K), and high heritability (h_l^2). For example, a relative increase in population prevalence of around 50% would be expected for autism ($\rho = 0.28, K = 0.0015, h_l^2 = 0.8$) compared to only 0.6% for MDD ($\rho = 0.12, K = 0.15, h_l^2 = 0.35$) from one generation of assortative mating. I note that these numbers provide upper boundaries; when partner-resemblance would, for example, be partly attributable to social homogamy (partner-resemblance driven by shared environmental factors) rather than phenotypic assortment (partner-resemblance driven by the psychiatric trait), the consequences would be considerably less pronounced. Notably, assortative mating is not expected to affect GWAS results, because of the small proportion of variance explained by individual loci, as discussed in **Chapter 6**.

How can these modeled consequences be interpreted in the context of empirical data? First, it should be noted that the increase in prevalence is most pronounced for rare disorders, and a prevalence increase from parental to offspring generation of e.g. 1.0% to 1.5% would be hard to detect in empirical

data given the standard errors around these estimates. Second, the presented consequence provide upper boundaries; when the partner-resemblance found by Nordsletten et al would e.g. be partly attributable to social homogamy, the consequences would be considerable less. Third, other factors also affect the population prevalence, such as reduced fecundity in psychiatric patients, which might reduce the consequences of assortative mating to some extent. Notably however, assortative mating and reduced fecundity are not expected to balance each other in the long term: the consequences of assortative mating asymptotically reach equilibrium after several generations, whether the consequences of the reduced fecundity do not. As a general limitation, it should further be noted that modeling was based on several inevitable assumptions discussed in detail in **Chapter 7**. To conclude, the modeled consequences of assortative mating are difficult to test in empirical data, but suggest that the consequences of assortative mating are at most modest for the heritability but may be considerable for the population prevalence of rare psychiatric disorders. A challenge for future research will be to test what proportion of partner resemblance can be attributed to phenotypic assortment, secondary assortment, social homogamy and marital interaction. With the emergence of large-scale population based samples including genotyped spouse-data, opportunities may present to address this question.

Next to within disorder partner-resemblance, Nordsletten et al also found evidence for across-disorder partner-resemblance.[33] Phillips et al explained that such across-disorder partner-resemblance may be attributable to across-disorder assortment, or to within-disorder assortment in addition to within-person correlation between both disorders.[34] Nordsletten et al did not distinguish between these two scenarios, but Van Grootheest et al and found that across-trait partner-resemblance with respect to obsessive-compulsive, anxious and depressive symptoms was attributable to both across-trait assortment and within-person correlation.[35] Notably, Wesseldijk et al found that across-trait partner-resemblance was more pronounced in parents of individuals affected with a psychiatric disorder.[36] To model the genetic consequences of across-disorder partner-resemblance is complex, but it could be hypothesized that this phenomenon relates to the genetic correlation between different psychiatric disorders in one way or the other. First, the genetic correlation between psychiatric disorders can be hypothesized to reflect a general underlying liability for all (or several) psychiatric disorders,[11] and individuals might get affected with e.g. either schizophrenia (SCZ) or bipolar disorder (BIP) based on environmental effects or disorder-specific genetic effects. When partners would assort based on

this general psychiatric liability, they might present as affected with different disorders assessed as across-disorder assortative mating. In other words, this first hypothesis could be thought of as "across-disorder partner-resemblance as consequence of assortment based on general psychiatric liability (detected as genetic correlation)." Nevertheless, another hypothesis might be in line with the opposite. Suppose that individuals would, indeed, mate based on distinct liabilities for actual different disorders, such as e.g. SCZ and BIP (now hypothesized entirely distinct disorders). Under this scenario, the risk alleles of those with high liability for SCZ would get together with the risk alleles for those with high liability for BIP, which would result in a correlation in the offspring generation between the effective loci on SCZ and the effective loci on BIP. This correlation between effective loci would be detected as genetic correlation between SCZ and BIP: in other words, the "genetic correlation would be attributable to across-disorder assortative mating." I am not aware of methods suitable to distinguish between these two and other hypotheses, but this might be a challenge for future research.

Recent successes in psychiatric genetics

Since my PhD trajectory commenced in 2011, research on psychiatric genetics has made great progress. In 2015, the CONVERGE consortium identified and replicated two loci associated to MDD in a Chinese female sample comprising 5,337 screened controls and 5,303 cases with recurrent episodes of MDD.[37] The success in this relatively small sample has been hypothesized to be attributable to the homogeneous cases, inclusion of female only and use of the same genotyping platform for the whole sample. This approach contrasts the more heterogeneous PGC sample including 9240 cases and 9519 controls that found no loci.[4] Another recent success was obtained with a very different strategy: in a total of 180,281 individuals from European ancestry two loci were linked to depressive symptoms, which were heterogeneously assessed across contributing cohorts.[20] The relevance of these latter findings for clinical diagnosis of MDD is not certain yet, but these findings may suggest that including enough individuals balances phenotypic heterogeneity. More interestingly, at the print of this thesis, the unpublished PGC-MDD meta-analysis comprising tens of thousands of cases from European ancestry also detected several of loci associated to MDD.[38]

The recent successes in genetic research on SCZ might hint to the successes to be expected in MDD, as research on SCZ seems to be a couple of years ahead of MDD (possibly owing to SCZ larger heritability,[39] more homogeneous nature,[11] and smaller population prevalence).[40] In 2011, seven loci

had been detected for SCZ[41] (which can roughly be compared to the current stage of genetic research in MDD), increasing to 108 loci identified in 2014 in a sample including 36,989 cases.[42] Moreover, in 2016 Sekar et al found that the top finding for SCZ is likely to reflect different levels of complement component 4 (C4) in individuals with SCZ, while in mice C4 was found to mediate synaptic pruning in postnatal development.[43] The finding of Sekar et al has been considered the first 'inroad into the molecular etiology of SCZ' that might potentially lead to new therapies in the future.[44]

From a more skeptical point of view, one might wonder how much clinical relevance can be expected from genetic loci individually explaining typically less than 1% of variance in disease risk. Professor Lander, first author of the Human Genome Project,[45] replied to this in a masterclass that I attended in November 2015, and noted that statins are important drugs in preventing cardiovascular disease that also link to a SNP explaining very little variance; rs12916 in the HMG-CoA reductase gene explains only 1.6% of variance in cholesterol levels.[46] So, although SNPs that have been (and will be) identified for MDD explain only a very small proportion of variance in disease risk, they may still point to relevant biologically pathways that could possibly initiate novel pharmacotherapeutic development in the coming decades.

Past challenges and the road to the future

At the time of commencement of my PhD, research on genetic effects on MDD and other psychiatric disorders had been characterized by single study findings that were not replicated. Findings from linkage studies and candidate gene studies were inconsistent,[47,48] and research on GxE with candidate genes has led to publication bias without any robust findings.[2] Naturally, until recently large scale GWAS studies were not achievable, and in the candidate gene era it was like seeking a needle in a haystack without the appropriate instrument. This metaphor is further strengthened by the current knowledge about the small effect sizes of individual loci, that typically explain less than 1% of variance in disease risk. The interaction between polygenic MDD risk and childhood trauma found in **Chapter 4** can best be viewed as a single cohort phenomenon given the non-replication in **Chapter 5.** Personally, I do not think research on GxE analyses will be the best way forward for genetic research in MDD in the near future. Nevertheless, I would advise for GWAS cohorts to collect uniformly environmental information as much as possible to prepare for solid GxE testing in the further future.

Rather, for the next couple of years, I think progress can be expected from collaboration in the Psychiatric Genomics Consortium, and from the large-scale population samples, such as UK Biobank[49] and iPSYCH. These population samples include many MDD cases based on MDD's high prevalence of around 15%, but are less suitable for rarer disorders such as schizophrenia with a prevalence of around 1%. The overall motto to me seems to regard single-study findings as hypothesis forming, and to aim for the largest possible sample size with subsequent independent replication. For the future, when sufficient causal loci have been identified, I think attempts can be made to integrate gene-findings with gene-expression and epigenetic information. This approach may help to improve classification of MDD and other psychiatric disorders from a more biologically informed point of view than the current DSM classification. I further anticipate that the rapidly increasing number of genetic correlation estimates will help to elucidate the relation between MDD, other disorders, and non-disease traits.

I think these are exciting times for genetic research in MDD. The GWAS sample size has increased drastically from 5763 cases by 2011[47] to tens of thousands of cases in 2016, the exciting results of which are anticipated to be published shortly.[38] In parallel, statistical methods were developed to effectively analyze this vast amount of data. An important method is LD score regression, which allows analyses of summary statistics to assess the variance explained by genotyped SNPs,[21] and the genetic correlation between any two traits.[16] It seems to me that the skepticism from a couple of years ago about the lack of GWAS findings,[50] which had followed the initial excitement following completion of the Human Genome Project in 2001,[45] has now been replaced again with more realistic optimism. I, for one, am looking forward to the years to come.

REFERENCES

1. Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science *301*, 386–389.

2. Duncan, L.E., and Keller, M.C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am. J. Psychiatry *168*, 1041–1049.

3. Peyrot, W.J., Milaneschi, Y., Abdellaoui, A., Sullivan, P.F., Hottenga, J.J., Boomsma, D.I., and Penninx, B.W.J.H. (2014). Effect of polygenic risk scores on depression in childhood trauma. Br. J. Psychiatry *205*, 113–119.

4. Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G.,

Byrne, E.M., Blackwood, D.H.R., Boomsma, D.I., Cichon, S., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.

5. Demirkan, A., Penninx, B.W.J.H., Hek, K., Wray, N.R., Amin, N., Aulchenko, Y.S., Dyck van, R., Geus de, E.J.C., Hofman, A., Uitterlinden, a G., et al. (2011). Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol. Psychiatry *16*, 773–783.

6. Peyrot, W.J., Lee, S.H., Milaneschi, Y., Abdellaoui, A., Byrne, E.M., Esko, T., de Geus, E.J.C., Hemani, G., Hottenga, J.J., Kloiber, S., et al. (2015). The association between lower educational attainment and depression owing to shared genetic effects? Results in ~25 000 subjects. Mol. Psychiatry.

7. Clarke, H., Flint, J., Attwood, a S., and Munafò, M.R. (2010). Association of the 5- HTTLPR genotype and unipolar depression: a meta-analysis. Psychol. Med. *40*, 1767–1778.

8. Kendler, K.S., and Gardner, C.O. (2010). Interpretation of interactions: guide for the perplexed. Br. J. Psychiatry *197*, 170–171.

9. Teicher, M.H., and Samson, J. a (2013). Childhood maltreatment and psychopathology: A case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. Am. J. Psychiatry *170*, 1114–1133.

10. Mullins, N., Power, R.A., Fisher, H.L., Hanscombe, K.B., Euesden, J., Iniesta, R., Levinson, D.F., Weissman, M.M., Potash, J.B., Shi, J., et al. (2015). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. Psychol. Med. 1–12.

11. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S. V, Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.

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

13. Eaves, L.J. (2006). Genotype x Environment interaction in psychopathology: fact or artifact? Twin Res. Hum. Genet. *9*, 1–8.

14. Keller, M.C. (2014). Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biol. Psychiatry *75*, 18–24.

15. Gauderman, W.J., Zhang, P., Morrison, J.L., and Lewinger, J.P. (2013). Finding novel genes by testing G × E interactions in a genome-wide association study. Genet. Epidemiol. *37*, 603–613.

16. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R.,

Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet.
17. Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S.F.W., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature.
18. Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science (80-.). *340*, 1467–1471.
19. Gale, C.R., Hagenaars, S.P., Davies, G., Hill, W.D., Liewald, D.C.M., Cullen, B., Penninx, B.W., Boomsma, D.I., Pell, J., McIntosh, A.M., et al. (2016). Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank. Transl. Psychiatry *6*,.
20. Okbay, A., Baselmans, B.M.L., De Neve, J.-E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linnér, R.K., Rietveld, C.A., Derringer, J., et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet.
21. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
22. Visscher, P.M., Hemani, G., Vinkhuyzen, A.E., Chen, G.-B., Lee, S.H., Wray, N.R., Goddard, M.E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet. *10*, e1004269.
23. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: Inferring the contribution of common variants. Proc. Natl. Acad. Sci. U. S. A. *111*, E5272–E5281.
24. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am. J. Hum. Genet. *88*, 294–305.
25. Haseman, J.K., and Elston, R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. *2*, 3–19.
26. So, H.-C., Li, M., and Sham, P.C. (2011). Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genet. Epidemiol. *35*, 447–456.
27. Maher, B. (2008). The case of the missing heritability. *456*,.

28. Manolio, T. a, Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L. a, Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.
29. Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J. V, et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. *47*, 1114–1120.
30. Bhatia, G., Gusev, A., Loh, P.-R., Vilhjálmsson, B.J., Ripke, S., Purcell, S., Stahl, E., Daly, M., de Candia, T.R., Kendler, K.S., et al. (2015). Haplotypes of common SNPs can explain missing heritability of complex diseases. bioRxiv 022418.
31. Bulik-Sullivan, B. (2015). Relationship between LD Score and Haseman-Elston Regression. bioRxiv.
32. Sullivan, P.F. (2010). The psychiatric GWAS consortium: big science comes to psychiatry. Neuron *68*, 182–186.
33. Nordsletten, A.E., Larsson, H., Crowley, J.J., Almqvist, C., Lichtenstein, P., and Mataix-Cols, D. (2016). Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. JAMA Psychiatry.
34. Phillips, K., Fulker, D.W., Carey, G., and Nagoshi, C.T. (1988). Direct marital assortment for cognitive and personality variables. Behav. Genet. *18*, 347–356.
35. van Grootheest, D.S., van den Berg, S.M., Cath, D.C., Willemsen, G., and Boomsma, D.I. (2008). Marital resemblance for obsessive-compulsive, anxious and depressive symptoms in a population-based sample. Psychol. Med. *38*, 1731–1740.
36. Wesseldijk, L.W., Dieleman, G.C., Lindauer, R.J.L., Bartels, M., Willemsen, G., Hudziak, J.J., Boomsma, D.I., and Middeldorp, C.M. (2016). Spousal resemblance in psychopathology: A comparison of parents of children with and without psychopathology. Eur. Psychiatry *34*, 49–55.
37. Cai, N., Bigdeli, T.B., Kretzschmar, W., Li, Y., Liang, J., Song, L., Hu, J., Li, Q., Jin, W., Hu, Z., et al. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature *523*, 588–591.
38. Psychiatric Genomics Consortium. Major Depressive Disorder: wave 2 GWAS results. In Preperation.
39. Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat. Rev. Genet. *13*, 537–551.
40. Yang, J., Wray, N.R., and Visscher, P.M. (2010). Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet. Epidemiol. *34*, 254–257.

41. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, Lin DY, Duan J, Ophoff RA, Andreassen OA, Scolnick E, Cichon S, St Clair D, Corvin A, Gurling H, Werge T, Rujescu D, Blackwood DH, Pato CN, Malhotra AK, Purcell S, Dudbridge F, Neale BM, R, G.P. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976.

42. Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

43. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al. (2016). Schizophrenia risk from complex variation of complement component 4. Nature *advance on*,.

44. Dhindsa, R.S., and Goldstein, D.B. (2016). Schizophrenia: From genetics to physiology at last. Nature *530*, 162–163.

45. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

46. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.

47. Wray, N.R., Pergadia, M.L., Blackwood, D.H.R., Penninx, B.W.J.H., Gordon, S.D., Nyholt, D.R., Ripke, S., MacIntyre, D.J., McGhee, K. a, Maclean, a W., et al. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. Mol. Psychiatry *17*, 36–48.

48. Boomsma, D.I., Willemsen, G., Sullivan, P.F., Heutink, P., Meijer, P., Sondervan, D., Kluft, C., Smit, G., Nolen, W. a, Zitman, F.G., et al. (2008). Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur. J. Hum. Genet. *16*, 335–342.

49. Allen, N.E., Sudlow, C., Peakman, T., and Collins, R. (2014). UK biobank data: come and get it. Sci. Transl. Med. *6*, 224ed4.

50. Sullivan, P. (2012). Don't give up on GWAS. Mol. Psychiatry *17*, 2–3.

NEDERLANDSE SAMENVATTING

In het eerste deel van dit proefschrift wordt de complexe relatie tussen genetische effecten en omgevingsfactoren bij het ontstaan van depressie onderzocht met behulp van empirische data van proefpersonen. Het is bekend dat zowel genen als bepaalde omgevingsinvloeden het risico op depressie verhogen. Het is echter onduidelijk of beide een onafhankelijk effect hebben, of er sprake is van interactie (een synergistisch effect van genen en omgeving), of dat de gemeten omgevingsinvloeden mogelijk een genetische basis delen met depressie (stress uit de omgeving kan samenhangen met persoonlijkheidskenmerken die genetisch bepaald kunnen zijn). In het tweede deel van dit proefschrift wordt een meer theoretische benadering gevolgd. De meeste genetische studies vergelijken mensen met een psychiatrische stoornis met mensen waarbij de stoornis expliciet is uitgesloten (zogeheten *controles*). Sommige studies volgen echter een andere strategie met bijvoorbeeld vader-moeder-kind trio data, of met data waarbij niet bij iedereen de aan- of afwezigheid van de psychiatrische stoornis is nagegaan. Deze genetische studies zullen andere resultaten opleveren, zoals in dit proefschrift wordt onderzocht. Het is bekend dat partners die samen een kind krijgen op elkaar lijken wat psychische kwetsbaarheid betreft (in het Engels aangeduide als *assortative mating*). In dit proefschrift worden de consequenties hiervan onderzocht op de erfelijkheid en prevalentie van psychiatrische stoornissen in de generatie van de kinderen met genetische modellen.

Het serotonine transporter gen is het meest onderzochte kandidaat-gen bij depressie (*kandidaat* wil zeggen gebaseerd op een a priori hypothese). Het lijkt voor de hand te liggen dat het onderscheid tussen de lange en korte variant van het lengte polymorfisme van dit gen (5-HTTLPR) van belang zou kunnen zijn voor depressie, omdat een belangrijke groep van antidepressieve medicatie aangrijpt op de serotonine transporter. Een befaamde studie van Caspi e.a. vond dat 5-HTTLPR meer effect had op het risico op depressie in personen die jeugdtrauma hadden doorgemaakt, of, met andere woorden, dat 5-HTTLPR en jeugdtrauma een synergistisch effect op depressie hadden waarbij hun gecombineerde effect groter was dan de som van de individuele effecten (gen-omgevingsinteractie). Het 5-HTTLPR polymorfisme werd onderzocht in **Hoofdstuk 2** in 1593 mensen met depressie en 1411 gezonde controles van de Nederlandse Studie naar Depressie en Angst (NESDA) en het Nederlands Tweelingen Register (NTR). Vier uitkomstmaten werden gedefinieerd: depressie, depressie met suïcidaliteit, chronische depressie, en het beloop van depressie (chronisch tegenover niet-chronisch). Ten aanzien van deze vier uitkomstmaten werden geen aanwijzingen

gevonden voor een interactie-effect van 5-HTTLPR met jeugdtrauma, maar ook niet met levenslange stressvolle levensgebeurtenissen, recente stressvolle levensgebeurtenissen, seksueel misbruik, of opleidingsniveau. Het gen-omgevingsinteractie effect dat Caspi beschreef werd dus niet gevonden in dit Nederlandse cohort.

De bekende relatie tussen een verhoogd risico op depressie en lager opleidingsniveau werd onderzocht in **Hoofstuk 3** in data van ongeveer 25.000 deelnemers aan de internationale samenwerking in het *Psychiatric Genomics Consortium*, een samenwerking tussen verschillende instituten uit 6 landen (PGC wave 1). Een verhoogd risico op depressie werd bevestigd in deelnemers met een lager opleidingsniveau, zoals uit eerder onderzoek al bekend was. Vervolgens werd onderzocht of deze relatie verklaard kon worden door overeenkomsten tussen de genetische effecten op depressie en de genetische effecten op lager opleidingsniveau. Hiervoor werden drie methoden toegepast op informatie over 884.105 *'single nucleotide polymorphisms'*, zogenaamde SNPs: kleine stukjes DNA die de genetische verschillen tussen mensen representeren. De eerste methode maakte gebruik van polygenetische risico scores die het aantal risico SNPs dat iemand bij zich draagt samenvatten, waarbij in een onafhankelijke data set is getest wat de risico SNPs zijn. Depressie werd niet voorspeld door de polygenetische risico scores voor opleidingsniveau (het aantal 'risico' SNPs voor lager opleidingsniveau), en opleidingsniveau werd niet voorspeld door polygenetische risico scores voor depressie (het aantal risico SNPs voor depressie). Met de tweede methode, bivariate *genomic-relationship-matrix restricted maximum likelihood* (GREML), werd een inconsistente zwakke negatieve genetische correlatie gevonden. Met de derde methode, *SNP effect concordance analysis* (SECA), werden geen overeenkomsten gevonden tussen SNP effecten op depressie en SNP effecten op opleidingsniveau. Al met al laten deze resultaten zien dat het onwaarschijnlijk is dat de relatie tussen depressie en lager opleidingsniveau verklaard wordt door gedeelde genetische effecten, alhoewel een kleine gedeelde genetische basis niet kon worden uitgesloten. Dit betekent dat de relatie tussen lager opleidingsniveau en depressie waarschijnlijk verklaard wordt doordat een lager opleidingsniveau het risico op depressie verhoogt, doordat depressie een effect heeft op het behaalde opleidingsniveau, of doordat een derde factor een invloed heeft op beide.

Tot 2014 had het onderzoek naar gen-omgevingsinteractie in depressie zich met name gericht op kandidaat-genen, zoals 5-HTTLPR. Maar recent genetisch onderzoek heeft ons geleerd dat het risico van depressie niet simpelweg te verklaren is door slechts een paar genen, maar waarschijnlijk door

kleine effecten van duizenden SNPs. Het ligt daarom voor de hand om gen-omgevingsinteracties te testen met het gezamenlijke effect van alle gemeten SNPs tegelijkertijd. In **Hoofdstuk 4** werden daarom polygenetische risico scores geconstrueerd voor 1645 mensen met een depressie en 340 gezonde controles uit NESDA gebaseerd op de SNP effecten van de *Psychiatric Genomics Consortium* (PGC wave 1). Deze polygenetische risico scores hadden een groter effect op depressie in die individuen die jeugdtrauma hadden meegemaakt. Dit suggereerde dat gen-omgevingsinteractie optreedt bij veel SNPs. De resultaten wezen op een multiplicatief interactie-effect (gecombineerde effect groter dan *het product* van de afzonderlijke effecten) en ook op een additief interactie-effect (gecombineerde effect groter dan *de som* van de afzonderlijke effecten).

Het interactie-effect tussen polygenetische risico scores en jeugdtrauma werd vervolgens onderzocht in **Hoofdstuk 5** met data van 3024 mensen met een depressie en 2741 gezonde controles uit zeven cohorten van het *Psychiatric Genomics Consortium* (PGC wave 2). Jeugdtrauma had een vergelijkbaar en groot effect op depressie in alle zeven cohorten, en dit effect was vergelijkbaar in mannen en vrouwen. De gen-omgevingsinteractie effecten waren echter verschillend in de zeven cohorten: in een cohort (NESDA) werd een positief interactie-effect gevonden (zoals in Hoofdstuk 4), in een ander cohort een negatief interactie-effect, en in de overige vijf cohorten werden geen interactie-effecten gevonden. In zijn geheel genomen is er daarom geen wetenschappelijk bewijs voor interactie tussen polygenetische risico scores en jeugdtrauma. De resultaten van Hoofdstuk 5 illustreren de grote verscheidenheid tussen de verschillende depressie-cohorten, en suggereren dat de bevinding van Hoofdstuk 4 het best geïnterpreteerd kan worden als een uniek fenomeen in NESDA.

In het tweede deel van dit proefschrift werd een meer theoretische benadering gevolgd om methodologische aspecten van genetisch onderzoek te belichten. Genetisch onderzoek naar psychiatrische stoornissen wordt over het algemeen gedaan door het vergelijken van mensen met een psychiatrische stoornis (bijvoorbeeld depressie) met controles die de stoornis in principe niet hebben. De eigenschappen van de controles verschillen van studie tot studie. In **Hoofdstuk 6** werden twee soorten controles onderzocht met betrekking tot (1) hun vermogen (*power*) om een geassocieerde SNP te detecteren en (2) hun schattingen van de SNP-erfelijkheid (dat deel van variatie in het ziekte risico dat toe te schrijven is aan genoomwijd gemeten SNPs). Ten eerste, vader-moeder-kind trio data worden regelmatig gebruikt in de cohorten die bijdragen aan het onderzoek naar aandachtstekort-hyperactiviteitstoornis (ADHD) en autisme binnen het *Psychiatric Genomics Consortium*. Met trio data worden de SNPs van

een kind met ADHD of autisme vergeleken met de SNPs van zijn of haar ouders. Trio data zijn noodzakelijk om nieuwe mutaties te vinden (SNPs bij het kind die niet bij de ouders voorkomen), maar kunnen leiden tot verminderde power en lagere schattingen van de SNP-erfelijkheid in vergelijking met datasets waarin mensen met ADHD (of autisme) vergeleken worden met gezonde controles die geen familie zijn. Deze verminderde power en lagere SNP-erfelijkheid schattingen zijn toe te schrijven aan (i) de potentiële over-representatie van families met meer dan één aangedaan familielid, en (ii) *assortative mating* (gelijkenis in psychiatrische kwetsbaarheid tussen partners die samen een kind krijgen). De tweede onderzochte studieopzet, waarbij een verminderde power en lagere schatting van de SNP-erfelijkheid verwacht wordt, is die waarbij gebruik gemaakt wordt van controles waarbij het voorkomen van de psychiatrische aandoening niet consequent is uitgesloten (niet-gescreende controles). Dit is in het bijzonder relevant voor studies naar psychiatrische aandoeningen die vaak voorkomen, omdat die meer mensen met de aandoening onder de controles zullen hebben. Ter illustratie, voor depressie (wat voorkomt in ongeveer 15% van de populatie) zal de SNP-erfelijkheid onderschat worden met 28% als bij geen van de controles depressie is uitgesloten. Dit leidde tot een formule om de schatting van de SNP-erfelijkheid aan te passen voor het gebruik van niet-gescreende controles (Formule 3 in **Hoofdstuk 6**). Als het doel is de polygenetische effecten op een psychiatrische stoornis te onderzoeken, dan is het advies om geen gebruik te maken van trio data en om de schatting van de SNP-erfelijkheid aan te passen als niet-gescreende controles worden gebruikt.

Onderzoek heeft herhaaldelijk laten zien dat partners die samen een kind krijgen op elkaar lijken wat betreft psychiatrische kwetsbaarheid; dat wil zeggen, partners van mensen met een psychiatrische stoornis hebben een grotere kans die aandoening te hebben dan partners van gezonde controles. In het Engels wordt dit fenomeen vaak aangeduid als *assortative mating*. De verwachte consequenties van deze partner-gelijkenis zijn niet berekend, terwijl hier theorie voor beschikbaar is. In **Hoofdstuk 7** werden daarom grenzen berekend voor de verwachte genetische consequenties van partner-gelijkenis met betrekking tot psychiatrische kwetsbaarheid. Het is niet bekend hoe het komt dat partners op elkaar lijken: mogelijk dat mensen het prettig vinden om samen te zijn met iemand die hen begrijpt en dat daardoor psychiatrische kwetsbaarheid gedeeld wordt, maar het zou bijvoorbeeld ook zo kunnen zijn dat partners elkaar tegen komen in de buurt waar ze beide wonen met dezelfde risicofactoren voor het ontwikkelen van een psychiatrische stoornis. De verwachte consequenties van partner-gelijkenis hangen af van de oorzaak van deze gelijkenis, en zijn het meest

uitgesproken als partners elkaar (onbewust) kiezen op basis van een gedeelde psychiatrische kwetsbaarheid. De verwachte consequenties bestaan uit een verhoogde erfelijkheid en verhoogde populatie-prevalentie in de generatie van de kinderen, dat wil zeggen dat er in de generatie van de kinderen meer mensen met de psychiatrische stoornis verwacht worden. In het bijzonder wordt voor geen van de psychiatrische aandoeningen verwacht dat de erfelijkheid met meer dan 5% zal toenemen als gevolg van één generatie met *assortative mating* (of 13% over meerdere generaties). De populatie-prevalentie zal het meest stijgen voor aandoeningen die zeldzaam zijn en een hoge erfelijkheid hebben: de prevalentie van autisme zou bijvoorbeeld tot 1.5 keer kunnen toenemen door één generatie met *assortative mating* (of 2.4 keer na meerdere generaties). Een ander fenomeen van psychiatrische aandoeningen is dat is gebleken dat mensen met psychiatrische aandoeningen over het algemeen genomen minder kinderen krijgen dan mensen zonder een psychiatrische aandoeningen: dit fenomeen wordt selectie genoemd. Selectie verlaagt de populatie-prevalentie in de generatie van de kinderen, terwijl *assortative mating* deze verhoogd. In Hoofdstuk 7 werd berekend dat de gecombineerde effecten van selectie en *assortative mating* elkaar gedurende een paar generaties min of meer in evenwicht kunnen houden, maar dat het effect van selectie op de langere termijn belangrijker is. Andere theorieën zijn daarom beter toegerust om te verklaren waarom psychiatrische stoornissen blijven bestaan ondanks selectie (deze theorieën zijn geen onderwerp van dit proefschrift). Het is belangrijk te benadrukken dat de beschreven modellen berusten op meerdere onvermijdelijke aannames. Met deze belangrijke beperking in het achterhoofd, lijken de modellen uit Hoofdstuk 7 te suggereren dat de consequenties van *assortative mating* voor de generatie van de kinderen beperkt zijn met betrekking tot de erfelijkheid, maar dat de consequenties voor de populatie-prevalentie aanzienlijk kunnen zijn, in het bijzonder voor zeldzame aandoeningen met een hoge erfelijkheid.

In **Hoofdstuk 8** worden de implicaties van dit proefschrift besproken. In Hoofdstuk 2 werd geen gen-omgevingsinteractie gevonden tussen 5-HTTLPR en jeugdtrauma in depressie. In combinatie met andere onafhankelijke onderzoeken lijkt het daarom zeer onwaarschijnlijk dat de befaamde resultaten van Caspi en collegae te extrapoleren zijn naar anderen cohorten. In Hoofdstuk 3 werd geen overeenkomst gevonden tussen de genetische effecten op depressie en lager opleidingsniveau. In meer recente onafhankelijk onderzoeken wordt gesuggereerd dat er wel een kleine genetische overeenkomst zou kunnen zijn, maar de belangrijkste conclusie blijft dat de relatie tussen depressie en lager opleidingsniveau voor het grootste deel niet bepaald wordt door gedeelde

genetische effecten. Het meest waarschijnlijk is dat de relatie verklaard wordt doordat een lager opleidingsniveau het risico op depressie verhoogd, of doordat een derde factor een invloed heeft op beide. In Hoofdstuk 4 werd gen-omgevingsinteractie gevonden tussen jeugdtrauma en alle risico SNPs gezamenlijk (samengevat in polygenetische risico scores), maar dit interactie effect werd niet gevonden in zes onafhankelijke cohorten (Hoofdstuk 5). De discrepantie tussen de bevindingen uit Hoofdstuk 4 en de bevindingen uit Hoofdstuk 5 illustreren het belang van onafhankelijke replicatie van genetische resultaten, maar illustreren eveneens de kwetsbaarheid van statistische analyses van interactie effecten. In Hoofdstuk 6 worden aanbevelingen gedaan voor genetische studies die gestart zullen worden, en worden handvatten gegeven om resultaten van genetisch onderzoek te interpreteren. In het bijzonder werd een formule afgeleid om de SNP erfelijkheid te corrigeren als de controles niet gecontroleerd zijn op de onderzochte stoornis. In Hoofdstuk 7 werden de verwachte consequenties van partner-gelijkenis (*assortative mating*) met betrekking tot psychiatrische kwetsbaarheid op de erfelijkheid en prevalentie van psychiatrische stoornissen in de generatie van de kinderen onderzocht. Een uitdaging voor toekomstig onderzoek zal zijn om de oorzaken van deze partner-gelijkenis te ontrafelen, omdat de verwachte consequenties hier in hoge mate van afhankelijk zijn. In grote lijnen kan echter verwacht worden dat de consequenties het meest uitgesproken zullen zijn voor zeldzame aandoeningen met een hoge erfelijkheid.

SUMMARY OF THESIS FINDINGS

This thesis aimed to study the complex link between genetic effects and measured environmental risk factors in major depressive disorder (MDD) in empirical data, and to explore boundaries of the consequences of two Genome-Wide Association study (GWAS) designs and assortative mating from a theoretical perspective. Both genes and environment affect MDD risk, but it remains unclear whether both act independently, whether they interact, or whether environmental risk might actually reflect shared genetic effects between MDD risk and e.g. behavioral traits that might increase environmental stress.

The most studied candidate gene in MDD is the serotonin transporter gene. The length polymorphism in the promoter region of this gene (5-HTTLPR) has been hypothesized to influence MDD risk, because an important group of antidepressant drugs acts on the serotonin transporter. A Science paper by Caspi et al suggested that this gene mainly exerts its influence when persons had been exposed to childhood trauma.[1] In **Chapter 2**, the 5-HTTLPR polymorphism was analyzed in 1593 cases and 1411 controls from the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR). In these individuals, four different outcome measures were defined: lifetime MDD, suicidal MDD, chronic MDD, and course of MDD (chronic versus non-chronic). No evidence was found for either direct effects of 5-HTTLPR on these outcome measures or interaction effects between 5-HTTLPR and five environmental risk factors for MDD: lifetime stressful life-events, recent stressful life-events, sexual abuse, childhood trauma, and educational attainment (as proxy for social economic status associated with increased stress).

In **Chapter 3**, the relation between MDD and educational attainment was investigated in approximately 25,000 individuals from the Psychiatric Genomics Consortium (PGC wave 1) with additional Dutch and Estonian data. An increased risk for MDD was confirmed in individuals with lower educational attainment. Subsequently, the possible contribution of shared genetic effects to this link was assessed with three different methods applying data of 884,105 autosomal common single-nucleotide polymorphisms (SNPs). Firstly, polygenic risk scores (PRS) based on GWAS results on education attainment in ~120,000 individuals (EA-PRS) did not affect MDD risk, and PRS based on MDD GWAS results in ~20,000 individuals (MDD-PRS) did not affect EA. Secondly, a non-consistent weak significant negative genetic correlation was found with bivariate genomic-relationship-matrix restricted maximum likelihood (GREML). Thirdly, no concordance was found in either significance or direction of SNP effects across MDD GWAS and EA GWAS results based on SNP effect concordance analysis

(SECA). To conclude, these findings indicate that it is unlikely that shared genetic effects explain a large proportion of the link between MDD risk and lower education attainment, but a small genetic contribution to this deleterious link could not be excluded.

Up to 2014, research on gene-by-environment (GxE) interaction in MDD had mainly focused on candidate genes, such as 5-HTTLPR. However, with the emergence of MDD cohorts with genome wide SNP data, novel methods were developed that allowed to tag genome-wide genetic MDD risk with polygenic risk scores (PRS). In **Chapter 4**, PRS were constructed in 1645 cases and 340 controls from NESDA based on discovery results from the large Psychiatric Genomics Consortium (PGC wave 1). These PRS were found to have an increased impact on MDD risk in individuals exposed to childhood trauma (CT), which suggested gene-by-environment interaction on a genome-wide scale. This interaction-effect was found both as departure from multiplicativity (combined impact of PRS and CT larger than the *product* of the individual effects) and as departure from additivity (combined impact larger than the *sum* of the individual effects), the latter of which has been hypothesized to be more plausible from a biological perspective.

In **Chapter 5**, the interaction between polygenic risk scores (PRS) and childhood trauma (CT) was further tested in seven cohorts from PGC (wave 2) with CT information available in 3,024 cases and 2,741 controls. CT had consistent impact across cohorts, with similar impact in males and females. However, the interaction effects were heterogeneous with a positive interaction effect in NESDA (as in Chapter 4), negative interaction effect in the Radiant-UK study, and no interaction in the other contributing five cohorts, resulting in no overall evidence for interaction between PRS and CT in MDD. The results from Chapter 5 illustrate the heterogeneity of MDD, and suggest that the results of Chapter 4 can best be interpreted as a single cohort phenomenon.

The focus was switched from analyzing empirical data to theoretic work in **Chapter 6.** Two GWAS study designs applied in the Psychiatric Genomics Consortium (PGC) were considered with respect to their power of SNP association analysis and SNP-heritability estimates (proportion of population-variance in disease-risk attributable to genome-wide common SNPs). First, parent-affected-offspring trio data are regularly applied in the subgroups of the PGC analyzing autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD). Trio data are essential to detect *de novo* mutations, but its use may results in reduced power in association analysis and underestimation of the SNP-heritability compared to analyses in case screened-control data. This difference is attributable to (i) potential oversampling of multiplex families (with more than

one affected offspring) and to (ii) assortative mating, which describes the correlation between mating partners in vulnerability for ADHD or ASD. Second, the use of poorly or unscreened controls for common disorders will result in decreased power in association analysis and decreased SNP-heritability estimates. In particular, for MDD with a lifetime prevalence of around 15%, the anticipated underestimation of the SNP-heritability was analytically derived at 28% when none of the controls would be screened. An updated equation was provided to properly scale the SNP-heritability when including unscreened controls (Equation 3 in **Chapter 6**). When aiming to analyze the polygenic effects in psychiatric disorders, it is advisable not to use trio data and to properly scale the SNP-heritability when applying data with unscreened controls.

Research has found significant partner-resemblances for psychiatric disorders, that is, mating partners are more often concordant in psychiatric disorder-status than expected by chance. This phenomenon is often referred to as assortative mating. However, the potential consequences of these partner-resemblances have not been quantified and have been left implicit despite available theory in the quantitative genetics literature. Therefore, in **Chapter 7**, boundaries were quantified for the anticipated consequences for disorder prevalence and heritability under various inevitable assumptions. The consequences are most pronounced when partner-resemblance is attributable to phenotypic assortment (partner-resemblance driven by the psychiatric trait), and are reflected in increased population prevalence and heritability in the offspring generation. From the first generation in which assortative mating takes place, the consequences add generation after generation to reach equilibrium asymptotically over generations. Because of this equilibrium, assortative mating is unlikely to balance the impact of reduced fecundity of psychiatric patients in the long term, as analytically derived in **Chapter 7**. Modeling suggests that the heritability of none of the psychiatric disorders considered is likely to increase with more than 5% from one generation of assortative mating (or 13% over several generations). The population prevalence will increase most for rare disorders with high heritability, such as the prevalence of ASD that might maximally increase 1.5-fold after one generation of assortative mating (or 2.4-fold over several generations). While emphasizing the limitations inherent to the inevitable model assumptions, genetic theory suggests that the consequences of assortative mating are, at most, modest for the heritability, but may be considerable for the population prevalence.