THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Gene Environment Interplay in Depression

Tugce Melisa Sau Chuong

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2022

Abstract

Major depressive disorder (MDD) is a common psychiatric disorder and one of the leading causes of disability worldwide. MDD is moderately heritable (~30-40%), suggesting both genetic and environmental effects are influential. Several strands of evidence point to the possible presence of gene-environment correlations and gene-environment interaction effects in MDD, although findings to date have been relatively inconsistent.

Recently, research using available genetic and environmental data on biological parents and offspring (trios), have shown that parental genetic nurturing effects are detectable using polygenic scores (PGSs) in more heritable traits such as educational attainment. Research findings also point to potential gene-by-trauma exposure interaction effects involved in MDD. It is evident that data limitations may result in power issues, confounding effects and biases, which impact reliable and accurate quantification of these effects, highlighting the need for methods that maximise statistical power and minimise bias and confounding when exploring these effects.

This thesis aims to adapt existing statistical frameworks and use of data to explore gene-environment interplay effects, which are robust to the limitations of the available data. Here, two large population-scale datasets, the UK Biobank (N~150,000) and Generation Scotland: Scottish Family Health Study (N~2680 trios), were utilised to explore genome-by-trauma exposure interaction effects in depression, as well as parental genetic nurturing effects in a range of traits including MDD.

The research aims of this thesis included (1) implementing models exploring parental genetic nurturing effects using available trio PGSs; (2) expanding these models to explore mechanisms of genetic nurturing effects with available parental phenotypic data, both addressed in chapter 2. Here, the quantification of parental genetic nurturing effects using trio PGSs was found to be reliable and robust to data limitations. However, expanding these models by including parental phenotypes

resulted in confounded effects. Simulation analyses demonstrated that the confounding was induced by power issues associated with PGSs, highlighting the need for improved measures of genetic variance. The final research aim (3) was to explore genome-by-trauma exposure interaction effects using relationship matrices capturing genetic, trauma exposure and genome-by-trauma exposure interaction similarity between participants. Genomic relationship matrices utilised all available genetic data, and thus, served as an improved representation of genetic variance. Environmental relationship matrices utilised principal components, capturing underlying dimensions of trauma exposure. A substantial proportion of MDD variance was found to be attributed to genetic, trauma exposure and genome-by-trauma interaction effects. However, little insight was inferred regarding the specificity and direction of trauma exposure involved in MDD manifestation, due to the difficulty in interpreting the underlying dimensions of trauma exposure.

The two studies provide a strong rationale for the use of improved measures of genetic and environmental components of MDD. Specifically, findings from chapter 2 highlight the need for measures that capture a substantial proportion of genetic variance to explore these complex gene-environment interplay effects. Chapter 3 results demonstrated how leveraging all available genetic data can uncover substantial effects that were previously missed. Evaluations of study designs highlight how future work can incorporate omics (e.g. methylation) data to improve measures of environmental factors, which would aid translational interpretation of results obtained from these models. Methylation data could help identify additional trait associated genetic loci as well as their mode of action, and hence, specific drug targets; which may not have been previously identified by traditional analyses such as genome-wide association studies (GWAS) not containing the relevant genetic variants.

Lay Summary

Major depressive disorder (MDD) is a common psychiatric disorder, shown to affect over 300 million individuals worldwide. Throughout the years, debates of nature vs nurture evolved to nature *and* nurture; and now, a more recent understanding that both genetic and environmental factors influence each other, and play a complex role in MDD.

Genetic factors can influence *exposure* as well as *sensitivity* to certain environments, which can go on to impact traits or diseases like MDD. This thesis sought to explore these effects using available genetic and environmental data from large-scale studies.

This thesis presents two main studies exploring the interplay between genetic and environmental factors. Chapter 2 utilised available family data, to explore whether parents can have a genetic influence on offspring traits by shaping their rearing environment (genetic nurturing). Here, genetic risk of MDD was aggregated into a single measure representing genetic factors. The main findings suggest that genetic risk variables of offspring and parents can be used to quantify the impact of genetic nurturing on offspring traits. No evidence of genetic nurturing effects in MDD was found with the available data. Data limitations result in the genetic risk measures only capturing a small fraction of the actual genetic risk. Thus, exploring these genetic risk measures alongside parental traits results in confounding, which make it difficult to explore mechanisms of genetic nurturing i.e. parental traits that shape offspring rearing environments.

In chapter 3, all available genetic and trauma exposure data are utilised to capture genetic, trauma exposure and genome-by-trauma exposure interaction similarity between unrelated individuals. Here it was found that genetics, trauma exposure and the interaction between the two played an important role in MDD. This chapter concludes that there is a need for representative and meaningful measures of trauma exposure to further explore the nature of this interaction effect.

The two studies present different methods which can leverage available genetic and environmental data to explore gene-environment interplay effects. Findings demonstrate that making use of all available genetic data can results in representative genetic measures which can aid the exploration of these effects in traits like MDD. Overall conclusions highlight the need for improved measures representing genetic and environmental measures in the exploration of these effects.

Acknowledgements

For my mother and cousin, **Serap Yildirim** and **Buse Melissa Yildirim**, my lovely little family.

I would like to thank my mother for being my greatest inspiration and support system. Thank you for constantly reminding me that I am worth more than the success I was unable to achieve. Thank you for encouraging my interest in mental health (and making sure mine was intact) which has ultimately lead me down this research path, which feels somewhat meaningful. Thank you for not forcing me to go into the fields of finance or accounting. I love and appreciate you so much.

I would like to thank **Poorya Parvizi**, for his unwavering support, motivation, kindness, genius intelligence and brilliant coding skills. Thank you for giving me R lessons for free (especially as I could not afford to pay you) and helping me EVERY time I have become stuck with analyses which has saved my soul from dying entirely. Thank you for being patient with me even when I have shocked you with my stupidity. Thank you for being my family here in Edinburgh. It has been a pleasure going through this experience with you. I love you so much cutie patootie.

I would like to thank everyone who has helped shape my path to and through this PhD starting with **Elvan Gokalp**, for being an incredible role model and one of the first people to show (through your own experience) that a PhD was a path that even existed; **Finlay Bourquin,** for being a wonderful human being who has had a huge influence on my personality and outlook on life (love you fishcake); **Aybuke Atalay** for having a positive vibe at all times and for keeping me sane by reminding me that there is more to life than this PhD. Thank you for being the main reason my final year in Edinburgh will be one I will never forget. I will miss living with you and the home we created together (we will always be family!); **Ezgi Tekdemir**, for always keeping me inspired by being a great example of someone who enjoys their job and life; **Julide** and **Ugur Kadam**, for being a source of positivity and fun at all times; **Yusuf Topaloglu**, for always listening intently to my work without ever making it obvious that you were bored and being an amazing addition to my life

with your weird general knowledge facts; **Atharva Deshmukh**, for being my ONLY friend in Edinburgh who understands my weird humour and can keep up with my near perfect accents; **Herbert Chang** (for your shockingly good K-pop breakdancing in the clubs), **Joseph Gilbody** (for all your help with the quant/pop gen assignments), **Sunny Panchal** and **Lucy Fisher** for being a part of and a constant reminder of my wonderful start to this PhD.

I would like to thank all of my friends in the Psychiatry lab; **Melissa Lewins**, for being my PhD wifey/ride or die through these 3.5 years and dancing with me randomly in the office when we'd overworked our brains; **Claire Green**, for being absolutely inspirational with all her success; **Miruna Barbu**, for being really sweet and welcoming when I first arrived; **Xueyi Shen**, for being someone I could look up to – also a great roomie for conferences; **Heather, Laura, Niamh, Anders, Ella** thank you all for helping create this lovely team. I would also like to thank everyone in the Haley lab, thank you all for all the lovely weekly catch up sessions, skittles and recipes! I would like to particularly thank **Pau Navarro** for adding extra humour to our meetings – it is always fun when you are around!

I would like to extend a massive thanks to my academic mentors, **Gursharan Kalsi**, and MSc supervisors, **Paul O'Reilly** and **Shing Wan (Sam) Choi**, who really inspired my decision to continue on to do this PhD. I would like to thank Gursharan for her enjoyable conversations and honest recount of academia. Thank you for helping me prepare for all things PhD related (and post PhD!) – I was extremely lucky to have met you. I would like to thank Paul and Sam – working with you both has been the highlight of my academic journey and has set the ultimate example for what to look out for in my future endeavours.

I would also like to thank my secondary supervision team **Andrew McIntosh, Mark Adams, Alex Kwong** and **Toni Clarke.** Andrew, for welcoming me into the lab with open arms and supervising me like one of my main supervisors. Thank you for all the positive encouragement and seeing my potential. This has positively affected my self-confidence and fuelled my motivation to work even harder. Mark, for being

super smart and being a part of our ultimate conference road-trip team. Alex, for bringing in an incredible energy to our lab. Toni, for guiding me down a slightly different research path, which has drastically shaped my PhD and research interests – I wish we had more time to work with one another.

Finally (best for last), I would like to thank the people who have been an integral part of this PhD by holding my hand through every single step of this journey: my main supervisors **Chris Haley** and **Carmen Amador**. Chris, for providing the freedom to go all out on my whims/side projects and continuing to support me through all of this. Thank you for the genius suggestions throughout my PhD, the academic conversations and support for the next stages of my life. Thank you to Carmen, for helping me with and explaining all of my (MANY) mistakes – and ultimately making me a better researcher. Thank you for being my safety net throughout this PhD. You both have been amazing examples of how to strive for the very best methods and to critically assess every step of any research project. It has been an absolute pleasure working with you both and I hope to uphold this level of research integrity throughout all my future endeavours. I feel extremely privileged to have had you both train me throughout this experience, which has genuinely been a once in a lifetime opportunity and I will be forever grateful.

From the length of this acknowledgements section, it is safe to say that it takes a village to raise a PhD student. Thank you all for accompanying me on this journey.

Declarations

I declare that this thesis is composed by myself, except where otherwise stated, and is entirely my own work. The presented work has not been submitted for any other degree or professional qualification.

……

Table of Contents

Chapter 1. Introduction		**13**
1.1.	Let's Talk About Depression	**13**
1.2.	Depression Aetiology	**16**
1.2.1.	Environmental Risk Factors	**17**
1.2.2.	Genetic Component of Depression	**19**
1.2.2.1.	Twin Design	**19**
1.2.2.2.	Candidate Genes	**21**
1.2.2.3.	Genome-Wide Association Studies	**22**
1.2.2.4.	Polygenic Scores	**24**
1.2.2.5.	Genomic Relationship Matrices	**26**
1.2.3.	Gene-Environment Interplay	**29**
1.2.3.1.	Gene-Environment Correlations	**29**
1.2.3.2.	Gene-Environment Interactions	**36**
1.3.	Thesis Statement	**40**
Chapter 2. Methodological Considerations When Using Polygenic Scores to Explore Parent-Offspring Genetic Nurturing Effects		**43**
2.1.	Abstract	**43**
2.1.1	Abstract Introduction	**43**
2.1.2	Abstract Methods	**43**
2.1.3	Abstract Results	**43**
2.1.4.	Abstract Discussion	**44**
2.2.	Introduction	**45**
2.3.	Methodology	**48**
2.3.1.	Participants	**48**
2.3.2.	Phenotypes	**49**
2.3.2.1.	Psychological Distress & Depression	**49**
2.3.2.2.	Educational Attainment	**49**
2.3.2.3.	Height	**49**
2.3.3.	Genotypes	**50**

2.3.4 Polygenic Scores **50**
2.3.5. Statistical Analyses **51**
2.3.5.1. Generation Scotland Analyses **51**
2.3.5.1.1. Regression Analyses **51**
2.3.5.1.2. Pathway Analyses **53**
2.3.5.2. Simulation Analyses **55**
2.3.5.2.1. Simulations I **55**
2.3.5.2.2. Simulations II **56**
2.4. Results **57**
2.4.1. Generation Scotland Results **57**
2.4.2. Simulations I Results **61**
2.4.3. Simulations II Results **62**
2.5. Discussion **64**
3. Genome-by-Trauma Exposure Interaction Effects in Depression **70**
3.1. Abstract **70**
3.1.1. Abstract Introduction **70**
3.1.2. Abstract Methods **70**
3.1.3. Abstract Results **70**
3.1.4. Abstract Discussion **70**
3.2. Introduction **71**
3.3. Methodology **74**
3.3.1 Participants **74**
3.3.2. Genotypes **74**
3.3.3. Phenotypes **75**
3.3.3.1. Depression & Neuroticism Phenotypes **74**
3.3.3.2. Trauma Phenotypes **75**
3.3.4. Geographical Clusters **76**
3.3.5. Covariance Matrices **78**
3.3.6. Analyses **79**

3.3.6.1. SNP Heritability & Genetic Correlations **80**

3.3.6.2. Variance Components Analyses **80**

3.4. Results **81**

3.5. Discussion **91**

4. Thesis Discussion **98**

4.1. Overview **98**

4.2. Main Findings and Contributions to the Literature **98**

4.2.1. Genetic Nurturing Effects **100**

4.2.2. Interaction Effects **101**

4.3. Strengths, Limitations and Future Work **103**

4.4. Conclusions **106**

Glossary **108**

Appendix I. Supplementary Materials for Chapter 2 'Methodological Considerations When Using Polygenic Scores to Explore Parent-Offspring Genetic Nurturing Effects' **122**

Appendix II. Supplementary Materials for Chapter 3 ' Genome-by-Trauma Interaction Effects on Depression' **155**

References **203**

Chapter 1: Introduction

1.1. Let`s Talk About Depression

Major Depressive Disorder (MDD) is a commonly experienced episodic psychiatric illness, with a global economic burden ranging from $85-210 billion across different countries (Greenberg et al, 2015; Xu et al, 2016). Prevalence rates are currently reaching ~16% (Kessler et al, 2009; Vos et al, 2016) and some reports suggest these rates are increasing (Moreno-Agostino et al, 2021; Weinberger et al, 2018).

Evidence suggests MDD affects over 300 million individuals (World Health, 2017) an increase of ~20% between the years of 2005-2015 (Vos et al, 2016). Whilst this may be a result of heightened awareness and increasing population sizes (Baxter et al, 2014; Bretschneider et al, 2018), MDD is certainly one of the leading causes of disability worldwide (Vos, 2020).

MDD typically manifests in young adulthood (Kessler & Bromet, 2013), affecting women approximately twice as frequently as men (Acciai & Hardy, 2017). Relapse or recurrence is common (Buckman et al, 2018) with evidence suggesting that the risk of future episodes increases as a function of those that have occurred previously (Bockting et al, 2015). Most importantly, MDD has been found to be associated with a large range of social and cognitive impairments. The resulting negative implications affect not only individuals living with MDD, but also their families and broader social networks (Lund et al, 2018). Implications include increased comorbidity with physical (Peres et al, 2017; Sartorius, 2017) and other psychiatric disorders (Fusar-Poli et al, 2014; Groen et al, 2020), poorer educational attainment outcomes (Wickersham et al, 2021) as well as increased risk of death due to suicide (Conejero et al, 2018).

Early intervention has been found to be particularly important in the prevention of outlined implications and for the prognosis of MDD (Ghio et al, 2014; Hung et al, 2017; Kautzky et al, 2019). However, findings point to accessibility limitations with current treatment plans (Thornicroft et al, 2017). Moreover, while there are many pharmacological and talking treatments for MDD, many people fail to respond to

the first intervention and approximately 25% fail to respond to multiple therapies (Cipriani et al, 2018; Cuijpers et al, 2020; Johnston et al, 2019; Munkholm et al, 2019).

For a clinical diagnosis of MDD, an individual will have to exhibit at least one depressive episode causing substantial dysfunction, which is characterised as a minimum two-week duration with symptoms persisting most of the day, and nearly every day of the episode. Individuals will need to exhibit five out of nine symptoms, including at least one of the cardinal symptoms, as specified by the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-V) (American Psychiatric Association, 2013) (**Table 1**). It is important to note that, whilst core features of MDD are highlighted, different symptom frequencies are observed across cultures that are not necessarily prioritised within standard instruments such as the DSM (Haroz et al, 2017). In fact, most MDD research has been conducted in Europe and North America on individuals of European ancestry, which would make extrapolation of findings to non-European ancestry problematic (Di Florio et al, 2017; Liu & Tronick, 2014).

Table 1. Major Depressive Disorder Symptoms

Symptoms
1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (Note: In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation – also known as anhedonia)
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (Note: In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day

5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

Note. Cardinal symptoms in bold.

Evidence suggests over 1000 different symptom combinations possible for a MDD diagnosis, highlighting the clinical heterogeneity of the disorder (Fried & Nesse, 2015). Patients can exhibit entirely different combinations of symptoms, sometimes within the same individual over time (**Figure 1**). Some symptoms are also bidirectional. This means that two individuals fulfilling the same symptom criterion may exhibit changes that are opposite in direction (such as changes in sleep and weight patterns) (Fried, 2017; Goldberg, 2011).

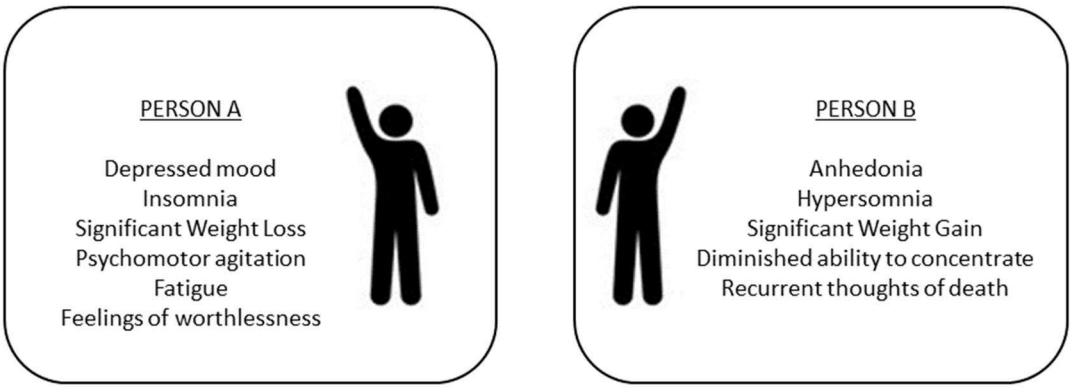


Figure 1. Major Depressive Disorder Symptom Profiles.

While, evidence shows strong correlations (0.65-0.90) between different diagnostic definitions of depression (Psychiatric Genomics Consortium and Broad depression definitions) (Jermy et al, 2021); findings have shown genetic heterogeneity between the different symptoms of depression (Thorp et al, 2020).

Genetic Heterogeneity: Different genetic mechanisms that underlie the same or similar phenotypes

This heterogeneity highlights potential statistical power issues in downstream analyses exploring MDD aetiology (Cai et al, 2020a; Levinson et al, 2014; Mullins & Lewis, 2017). Additionally, many of the symptoms are observed in other psychiatric disorders (Afzali et al, 2017; Price et al, 2019) suggesting an aetiological overlap (Anttila et al, 2018; Kendler, 1996). The ambiguities and inconsistencies on what distinguishes MDD from other mental illnesses is thought to be one of the main limitations contributing to the failure in addressing the global burden of MDD (Patel, 2017).

It is evident that further empirical research exploring the aetiology of MDD is required in order to understand how to develop efficient preventative interventions. Given the number of individuals affected, efforts to further understand the aetiology of MDD may have significant contributions to global public health.

1.2. Depression Aetiology

The debate of 'nature vs nurture' has clearly shifted to 'nature ***and*** nurture' for complex traits (Weissman, 2020). It is clear that individual differences observed in complex behavioural traits are a result of both genetic and environmental factors (Verweij et al, 2012). Furthermore, the conceptualisation of complex traits, as a continuous normal distribution of liability is now very prominent (Falconer & Mackay, 1996; Felsenstein, 2005; Wright, 1934a; b). This liability threshold model has been applied to complex discrete traits such as MDD, whereby the disorder manifests once a threshold on the liability scale is exceeded.

It is now well understood that MDD is moderately heritable with estimates ranging between 5-15% based on common genetic markers in unrelated populations and ~35-40% based on twin studies (Howard et al, 2019; Polderman et al, 2015). Findings suggest that both genetic and environmental factors play an important role (Saveanu & Nemeroff, 2012; Uher & Zwicker, 2017) and contribute to the liability of depression (Corfield et al, 2017).

Moreover, findings highlight gene-environment interplay effects; with evidence of genetic and environmental factors being correlated (Gage et al, 2016) or interacting (Assary et al, 2018) (**Figure 2**). This section will outline our understanding of the main environmental risk factors and genetic component of MDD.

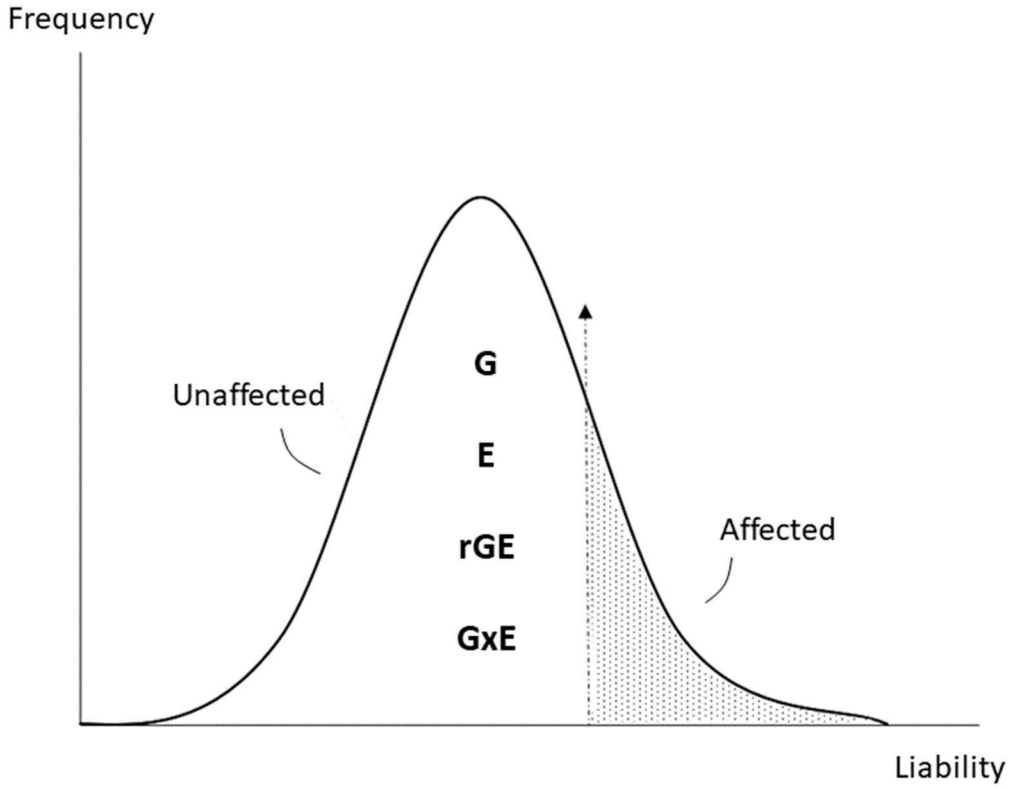


Figure 2. The Liability Threshold Model for a Discrete Complex Trait such as Depression. G, genetic variance; E, environmental variance; rGE, gene-environment correlation effects; GxE, gene-environment interaction effects.

1.2.1. Environmental Risk Factors

The most prominent known environmental risk factor identified for depression from retrospective and prospective studies is environmental adversity, in the form of trauma exposure (Dannehl et al, 2017; Kessler et al, 2010; Li et al, 2016; Vibhakar et al, 2019) and stressful life events (Hammen, 2016; Slavich & Irwin, 2014). Individuals with MDD report higher levels of adversity exposure (Negele et al, 2015), and adversity has been shown to be associated with higher severity of symptoms, longer course of illness as well as increased risk of relapse in MDD (Nanni et al, 2012; Tunnard et al, 2014). These findings, amongst others, point to a potential causal role of adversity exposure in MDD manifestation (Kendler et al, 2000; Kendler et al, 1999).

Reports suggest that females are more susceptible to adversity exposure (Hankin et al, 2015) with effects on MDD manifestation being larger than that observed in males (Gallo et al, 2018; Wamser-Nanney & Cherry, 2018); which may explain the gender gap consistently observed in depression prevalence rates (Kuehner, 2017). Some findings suggest different patterns and effects of trauma exposure between males and females (Wamser-Nanney & Cherry, 2018).

Moreover, findings have shown the effect can also differ depending on the amount (Anda et al, 2006; Hughes et al, 2017), severity (Kucharska, 2017; Lindert et al, 2014; Norman et al, 2012) and developmental period of adversity exposure (Dunn et al, 2017). It is perhaps intuitive that different forms of adversity exposure, such as sexual/physical abuse or neglect have a large effect on MDD manifestation, particularly if these are experienced in childhood or adolescence (Hoppen & Chalder, 2018; LeMoult et al, 2020). However, it is difficult to disentangle the individual effects of specific trauma exposures, as many of the trauma subtypes are overlapping and correlated with one another (de Oliveira et al, 2018).

Efforts to understand the mechanisms linking trauma exposure and MDD have consistently highlighted psychological and biological processes, such as altered emotion regulation (Christ et al, 2019), dysfunctional/negative attributional styles (Alloy et al, 2006; Mac Giollabhui et al, 2018), HPA axis dysregulation (Shea et al,

2005) as well as altered brain connectivity (Yu et al, 2019). Interestingly, findings suggest distinct profiles between MDD with and without exposure to adversity, such as higher levels of peripheral cytokines (Lu et al, 2013) and smaller hippocampal volume (Vythilingam et al, 2002) observed in individuals with both MDD and adversity exposure. Genetically informed studies have also demonstrated that MDD with trauma exposure has higher heritability than MDD without trauma exposure (Coleman et al, 2020; Peterson et al, 2018). This suggests that MDD with trauma exposure may have distinct disease mechanisms.

Heritability: Proportion of variation in a trait that is attributable to genetic variation between individuals within a population.

To complicate matters further, trauma exposure has also been found to have a genetic component (Warrier et al, 2021). Studies have found genetic correlations between trauma exposure and MDD (Coleman et al, 2020) as well as other psychiatric traits (Sartor et al, 2012). These findings support prior claims that many seemingly 'environmental' factors are also heritable and thus, the relationship between environmental and genetic factors in MDD are likely to be intertwined and complex (Akimova et al, 2020; Kendler & Baker, 2007).

1.2.2. Genetic Component of Depression

This section will outline the natural progression and evolution in the methods used to explore the genetic aetiology of depression. Each method has contributed substantially to the understanding of the genetic component of depression whilst also highlighting the need for greater sample sizes and increased statistical power in efficiently translating these genetic effects.

1.2.2.1. The Twin Design

Genetic and environmental factors are associated with individual differences in most complex behavioural traits; and phenotypic similarity is higher in individuals who are genetically similar (Verweij et al, 2012). Comparing phenotypic similarities between family members as a function of their genetic similarity, especially

monozygotic and dizygotic twins, has provided quantitative estimates of heritability for many traits. Twin studies have consistently shown that MDD is a moderately heritable trait with estimates ranging between 30-40% (Sullivan et al, 2000).

The classical twin design exploits the fact that differences observed in monozygotic twins are purely a consequence of individual (non-shared) environmental experiences, whereas both genetics and (non-shared) environmental experiences contribute to differences observed in dizygotic twins (as they only share ~50% of their genetic make-up). In contrast, both genetic and shared environmental experiences contribute to similarities observed in monozygotic and dizygotic twins (**Figure 3**) (Evans et al, 2002; Posthuma et al, 2003). Comparing twin trait correlations between mono and dizygotic twins enables the partitioning of trait variation into genetic and environmental components.

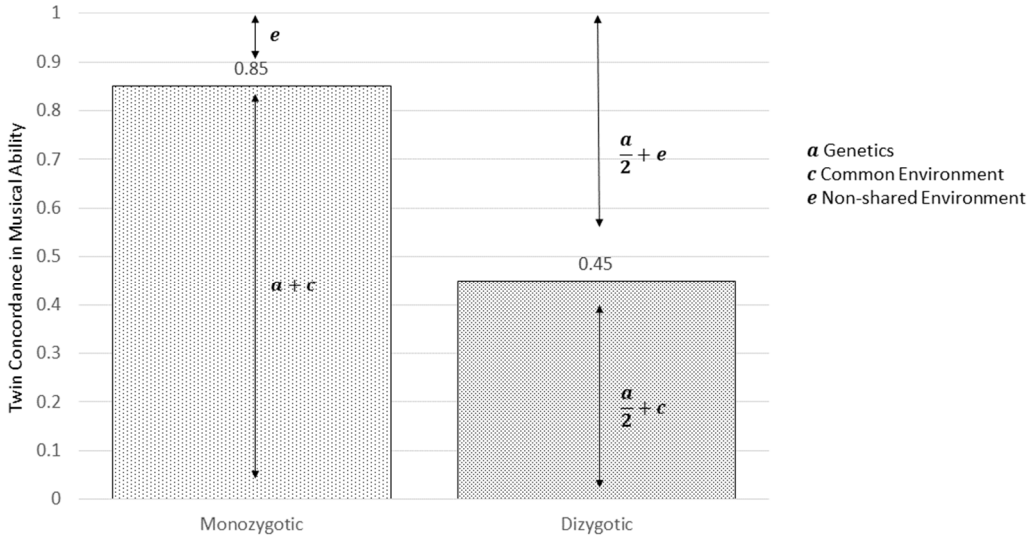


Figure 3. Twin Concordance for Musical Ability. The ACE acronyms are often used within twin study literature; a, additive genetic; c, common environment; e individual (non-shared) environment effects. By comparing monozygotic and dizygotic twin concordance, you can partition variance components. Using the concordance rates outlined, here, musical ability has $a = 0.80$, $c = 0.05$, $e = 0.15$ components, where the a component can also be thought of as a heritability estimate (proportion of difference attributable to genetic differences). Note that values provided for 'musical ability' are arbitrary.

Whilst, the classical twin design conveniently quantifies trait differences attributable to genetic and environmental components, it is limited, as it requires adaptations to model non-additive effects (Evans et al, 2002) and it does not provide insight into the molecular components of the genetic aetiology (Røysamb & Tambs, 2016). More limitations of this design will be highlighted below, in sections outlining gene-environment interplay.

1.2.2.2. Candidate Genes

To further explore the genetic component, many candidate gene studies were conducted to identify specific loci and genes increasing risk of MDD manifestation. These studies were initially thought to have high translational impact, as they were often based on underlying neurobiology or antidepressant pharmacology and could potentially highlight targetable biological mechanisms (Jia et al, 2011; Tabor et al, 2002). One widely researched candidate gene initially found to be associated with depression is the 5-HTTPLR polymorphism (Gressier et al, 2016; Xia & Yao, 2015). This polymorphism encodes the serotonin transporter protein regulating the amount and duration of serotonin in the synaptic cleft (Kobiella et al, 2011), a target of most effective antidepressants (Marazziti, 2017).

> **Polymorphism:** A gene is polymorphic when more than one allele occupies that gene's locus.
>
> **Single Nucleotide Polymorphisms (SNPs)** refer to the substitution of a single nucleotide at a specific position in the genome. Most polymorphisms have a reference and alternative allele that can have different frequencies within a population.
>
> The frequency of the less common allele is referred to as the **Minor Allele Frequency (MAF)**.

However, many limitations have been identified with the candidate gene approach, such as the 'biased' selection of mainly well researched genes (Crowe, 1993; Munafo, 2006; Zhu & Zhao, 2007) and the substantial failed replication attempts (Border et al, 2019);

indicating that findings are likely false positives, emphasised with publication bias (Bosker et al, 2011).

The most important conclusion from candidate gene studies was that complex traits are likely polygenic in nature, meaning many genetic variants are influential in complex trait manifestation (Assary et al, 2018; Culverhouse et al, 2018).

1.2.2.3. Genome-Wide Association Studies

The limitations of the candidate gene approach resulted in a transition to methods prioritising the exploration of multiple genetic variants, such as, genome-wide associations studies (GWAS) (Duncan et al, 2019). GWAS aim to identify associations between common genome-wide genetic variants, in the form of single nucleotide polymorphisms (SNPs), and traits of interest. SNPs are the most common form of measurable genetic variation seen within the human population and are relatively inexpensive to genotype. Statistical analyses that leverage SNPs are also capable of capturing a larger proportion of genetic variation (Clarke & Cooper, 2010). Associations are explored using univariate regression analyses with the formula;

$$y = \beta_0 + \beta_1 SNPx + \varepsilon$$

Where y represents the value of the trait being explored; β_0 refers to the intercept term; β_1 refers to the regression coefficient representing the strength of association (effect size) between SNP x and phenotype y; ε refers to the error term representing noise and variance not explained by SNP x. Note that models are often more complex and consider other fixed effects, such as age and sex.

As many variants are explored, resulting in multiple testing, a stringent genome-wide p-value significance threshold of 5×10^{-8} is specified (Pe'er et al, 2008). Results of GWAS are often presented in the form of Manhattan plots (**Figure 4**).

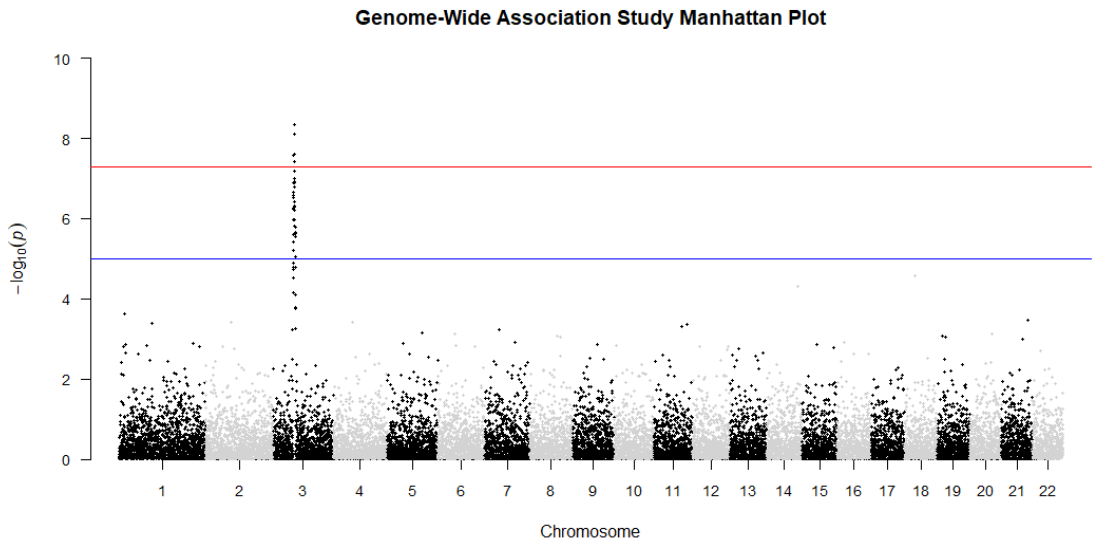


Figure 4. Genome-Wide Association Manhattan Plot. Each individual point represents a single nucleotide polymorphism (SNP) explored. The x-axis represent the chromosomes (1:22) the SNPs are positioned on. The y-axis represents the $-\log_{10}$ p-value representing significance of SNP and trait association. The blue and red lines represent genome-wide suggestive and significant thresholds respectively. SNPs exceeding the red line will be considered genome-wide significant, and thus, associated with trait of interest. Note that the plot was created on example GWAS data, 'gwasResults', provided by the 'qqman' software package in R.

GWAS on MDD initially yielded disappointing findings with no identified genome-wide significant variants despite comparable sample sizes (~10k) to other trait GWAS studies yielding many trait-associated variants (Hek et al, 2013; Mullins & Lewis, 2017; Sullivan et al, 2013).

Interestingly, two significant loci (Cai et al, 2015) were identified with fewer case and control sample sizes when exploring a more homogenous subtype of depression: recurrent MDD in Han Chinese women (although these variants were not identified in a recent East-Asian ancestry depression GWAS (Giannakopoulou et al, 2021)). Similarly, meta-analysis attempts have shown that a substantial increase in sample size can also yield genome-wide significant loci i.e. 44 risk variants identified with 480,259 individuals (Wray et al, 2018a). Evidence suggests that the rate of identified risk variants will accelerate once a threshold sample size is exceeded (Visscher et al, 2012); which is supported by findings from more recent depression GWAS meta-analyses identifying 102 risk variants with 807,533 individuals (Howard et al, 2019). Furthermore, evidence suggests that many risk

variants are yet to be identified for all complex traits, pointing to the continuing need for increasing sample sizes (Wray et al, 2018b).

Greater statistical power for genetic discovery in GWAS may, in theory, be achieved by using more detailed phenotyping of MDD and stratification into more aetiologically homogenous subtypes. However, this approach has not been repeatedly demonstrated and further increasing sample sizes remains the demonstrably effective approach to increasing variant discovery. To some extent, it is difficult to simultaneously increase sample sizes and in-depth phenotyping, as the latter is often costly and time-consuming (Lu et al, 2014). Thus, recently there has been increased use of minimal phenotyping, which has resulted in the pooling of massive sample sizes with several different definitions of depression. Whilst minimal phenotyping depression definitions have been shown to be highly correlated with other diagnostic definitions of depression (Jermy et al, 2022), evidence suggests the identified risk variants lack specificity for MDD (Cai et al, 2020b).

Minimal Phenotyping: Diagnosis of trait/disease reliant on self-reported diagnosis by a healthcare professional.

Nonetheless, advantages of GWAS, such as the exploration of variants without the need for prior knowledge paves the way for the identification of novel loci and biological mechanisms, which may be influential and potentially targetable (Tam et al, 2019). Moreover, GWAS have not only reinforced the notion that complex traits are highly polygenic, findings highlight that risk variants are likely to each have very small effects on traits of interest, particularly psychiatric traits like MDD (Howard et al, 2019; Wray et al, 2014).

1.2.2.4. Polygenic Scores

GWAS to date, have identified numerous risk variants associated with MDD (Howard et al, 2019). A single identified risk variant is unlikely to be informative when it comes to understanding the individual's disease risk, as effect sizes of identified risk variants are so small (Lewis & Vassos, 2020). However, the possibility

of incorporating all risk variants into an aggregate risk score has given rise to the computation of polygenic scores (PGSs). PGSs are calculated by summing and weighting the number of risk variants an individual has into a single value from information provided by GWAS summary statistics (Chatterjee et al, 2016; Choi et al, 2020; Torkamani et al, 2018).

PGSs are a simple and useful way of utilising genome-wide genetic data to explain trait variance. PGSs have been used in the prediction of many diseases outcomes (Lambert et al, 2019; Maher, 2015) resulting in valid evaluations of their clinical utility and validity. PGS validity is strongly dependent on the nature of the GWAS from which it is derived; many GWAS with increasing sample sizes are observed, however, these do not reflect ethnic diversity (Yanes et al, 2020). This has resulted in poor prediction of disease in non-European ancestry populations using current PGSs (Dikilitas et al, 2020), which really highlights the potential to exacerbate already prevalent health disparities (Martin et al, 2019). Other highlighted limitations include the lack of gene-gene interactions captured by PGSs. This has resulted in research re-evaluating the additive model proposed for complex traits (Moonesinghe et al, 2011; Nelson et al, 2013). However, it is important to note that findings provide no consistent evidence for gene-gene interaction effects in complex traits (Visscher & Wray, 2015) measured on the liability scale.

Additive Genetic Model: This model suggests that genetic variants exert influence on a trait in a linear or additive manner, as opposed to the variants interacting with one another.

Despite these concerns, there is a consensus that PGSs are useful as a genomic research tool. PGSs can use SNPs with association p-values greater than the genome-wide significance threshold, meaning the likely many causal variants yet to be identified as genome-wide significant can be utilised (Ni et al, 2021). Evidence has consistently shown the predictive ability when distinguishing between case-controls, increased when more SNPs are utilised in the computation of PGSs (Howard et al, 2019; Khera et al, 2018; Mavaddat et al, 2019; Musliner et al, 2019).

However, it is important to note that, to date, PGSs only capture a fraction of trait genetic variance (Dudbridge, 2013; Mostafavi et al, 2020) which contributes to the notion that the use of PGSs in clinical settings may be premature (Yanes et al, 2020). Current depression PGSs have only captured ~2-3% of MDD trait variance (Howard et al, 2019; Wray et al, 2018a) which is far less than the ~30-40% heritability estimates proposed by twin studies (Polderman et al, 2015; Sullivan et al, 2000). These findings may be a result of minimal phenotyping issues i.e. the use of broader definitions of depression (Cai et al, 2020b), introduced above, resulting in inaccurate SNP association effect sizes which go onto impact the predictive ability of PGSs. Yet, even traits with relatively low measurement error, that have highly successful GWAS findings, such as height, have yielded PGSs that capture only ~25% of the trait variance (Yengo et al, 2018), which again, is less than the ~85% heritability estimate proposed for height from twin studies (Silventoinen et al, 2003).

This discrepancy has highlighted a more concerning issue of missing heritability (Eichler et al, 2010; Manolio et al, 2009). Efforts to understand the underlying causes of missing heritability have pointed to potential gene-environment interplay effects resulting in over-estimated heritability in twin studies (Young, 2019).

Missing Heritability: This phenomenon refers to the discrepancy in variance attributable to measured single genetic variations and estimates of heritability of traits proposed by twin and other family studies.

1.2.2.5. Genomic Relationship Matrices

More recently, attempts to utilise all genotyped SNPs in the form of genomic-relationship matrices (GRMs) has yielded really exciting findings, suggesting heritability is not necessarily 'missing', but more so in 'hiding' (Yang et al, 2011). A simplistic overview of this methodology is that all genotyped SNPs are incorporated as random effects into a mixed linear model, outputting variance components, and

thus, providing a quantitative estimate of trait variance attributable to all genotyped SNPs, a.k.a. SNP heritability.

GRMs are constructed as a matrix capturing pairwise genomic similarity between any two individuals within a sample (**Figure 5**). Incorporating information on all genotyped SNPs as a single matrix allows the circumvention of statistical challenges associated with modelling many variants individually. Genomic similarity is calculated using a straightforward formula (**Equation 1**), which takes into account the number of variant copies each individual has as well as the frequency of the variant observed in the explored population. The formula is obtained from Genome-wide Complex Trait Analysis (GCTA) software, often used to create these matrices (VanRaden, 2008; Yang et al, 2011).

Random Effects: These effects refer to predictor variables where one may be interested in making inferences using the distribution of values as opposed to testing the differences of values.

In contrast, **Fixed Effects**, refer to predictor variables where one is interested in the difference of values.

Models incorporating random effects can provide insight into and control for variance attributable to the random effects of interest.

Equation 1: $A_{jk} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i}{2p_i(1-p_i)} \right)$

With A_{jk} being the genomic similarity between individuals j and k. Where N is the total number of SNPs; x_{ij} and x_{ik} are the number of copies of the reference allele for the i^{th} SNP in individual j and k respectively; p_i is the frequency of the reference allele.

Individuals	1	2	3	4…
1	0.989			
2	**0.489**	0.999		
3	0.001	0.034	0.978	
4…	0.023	**0.126**	0.031	0.964

Figure 5. Example Genomic Relationship Matrix. This is a symmetric matrix around the diagonal cells. Each cell represents pairwise genomic similarity between individuals within a sample (calculated using Equation 1). The diagonal values refer to the genomic similarity of the individuals themselves and thus, are ~1 (as they should share 100% of their genome, with themselves). Note that values >1 may represent the effect of inbreeding. The off-diagonal values represent genomic similarity between different individuals. Values that are bolded aim to show that similarity values can also provide insight into family relations e.g. ~0.5/0.125 would represent parent/sibling and cousin relations respectively. Please note, this matrix is for example purposes only, and plausible values were arbitrarily chosen.

The mixed linear model in which a GRM is fitted as a random effect, can be solved using a genomic-relatedness based restricted maximum likelihood (GREML) estimation (Yang et al, 2011). As opposed to fixed effect models, outputted results provide insight into the variance (and thus, size/importance) attributable to *each* random effect, however, further analyses would be required to explore the nature and direction of these associations.

As with all methods, the GREML approach is not free of limitations. Similar to PGSs, the GREML approach is highly reliant on the rigor of the GWAS utilised. The choice of sample and measurement error within phenotyping have been shown to bias estimates (Kumar et al, 2016).

Findings from research implementing this approach have identified SNP heritability estimates of ~20-30% for MDD (Lee et al, 2013; Peterson et al, 2018), which is a substantial improvement from the variance accounted for by PGSs (2-3%), yet still slightly lower than estimates obtained from twin studies (30-40%). An explanation for this could lie in the fact that not all genetic variants are genotyped and thus, included in GRMs, resulting in missing genetic variance. However, research simulating imputed/whole-genome sequenced data have computed GRMs and further shown that variance accounted for by the genetic variants is unlikely to reflect values proposed by twin studies; which further points to the overestimation of heritability from this design (Yang et al, 2015).

Genotype Data: Refers to measures of common SNPs across the genome.

Imputation: Refers to the statistical inference of unobserved/unmeasured SNPs using **phased haplotypes** (resolved groups of alleles in an organism that are inherited together from a single parent).

Whole Genome Sequencing (WGS): is a method that can determine the exact sequence of the whole genome.

1.2.3. Gene-Environment Interplay

Exploring gene-environment interplay can be useful in uncovering missing heritability as well as by providing translational findings by highlighting targetable risk environments and at-risk groups within a population (Jaffee & Price, 2008). Different forms of gene-environment interplay have been proposed. These can broadly be thought of as gene-environment correlations (rGE) and/or gene-environment interactions (GxE). The different forms of gene-environment interplay, evaluations of the current literature and methodological concerns are outlined below.

1.2.3.1. Gene-Environment Correlations

Evidence suggests that an individual's genotype for a trait is also correlated with important environmental influences (Gage et al, 2016). These gene-environment

correlations (rGEs) can be thought of as an individual`s genetic propensity to be *exposed* to environments which go onto impact the manifestation of a trait (Plomin, 2013).

Three separate processes have been proposed to explain how rGEs occur; passive, evocative and active rGEs (Plomin, 2014). A relatively simple trait, such as musical ability, can pose as a good analogy to explain each process (**Figure 6**). A form of passive rGEs known as ***parental genetic nurturing effects***, refer to associations between an offspring's genotype and the environment they are reared in, as both are influenced by parental genotypes. In the case of musical ability, parent genetic effects can influence offspring musical talent by forming the offspring's own genome (genetic talent) and/or shaping a musical rearing environment (e.g. presence of musical instruments which parents themselves may play), further nurturing offspring musical ability.

In contrast, evocative rGEs refer to the association between an individual's genetically influenced behaviour evoking a response from others; such as a musically talented individual gaining the attention of a music teacher, which in turn nurtures their musical ability.

Finally, active rGEs refer to the association between an individual's genetic propensity and the environments they select for themselves. Continuing with the musical ability analogy, this can be conceptualised as a musically talented individual seeking out music lessons, which in turn, nurtures their musical ability (Price & Jaffee, 2008).

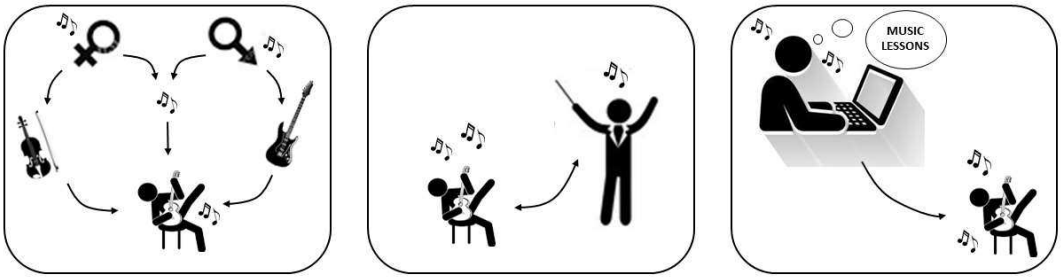


Figure 6. Gene-Environment Correlation in Musial Ability. Left, *passive* gene-environment correlation; Middle, *evocative* gene-environment correlation; Right, *active* gene-environment correlations.

There is a strong consensus that risk environments, like stressful life events or trauma exposure, have a causal effect on disease manifestation; however, if the same genetic markers are influential in the experience of risk environments as well as disease manifestation, then this association will be confounded. Having an intrinsic understanding of whether the associations between 'environmental' factors and traits are causal, non-causal or potentially more intricate with bi-directional effects is important for the development of effective interventions (Knafo & Jaffee, 2013). It is evident that these risk environments have a genetic component, but are also thought to have an environmentally mediated effect on MDD (Jaffee & Price, 2007).

Frameworks which suggest leveraging the wealth of animal data, which includes vast amounts of genotypic and phenotypic data of many generations, have suggested evidence of parental genetic nurturing effects can be quantified by exploring parental genome effects on offspring traits, over and above the effect of the offspring's own genome (Bijma, 2014; Falconer & Mackay, 1996; Griffing, 1967; Willham, 1963).

Researchers have attempted to explore these effects within human traits by leveraging available twin and children of twin family designs (Beam & Turkheimer, 2013; Wilkinson et al, 2013). Findings have implied the effect of rGEs by showing that environmental measures are heritable (Rutter et al, 2006), which have also been supported by identified associated genetic variants from GWAS findings;

particularly environmental factors shown to exacerbate risk of MDD such as stressful life events (Bolinskey et al, 2004; Power et al, 2013) and trauma exposure (Richmond-Rakerd et al, 2019; Warrier et al, 2021). Moreover, genetic correlations between these heritable environments and MDD have supported the notion of potential rGE effects involved in MDD manifestation (Clarke et al, 2018; Coleman et al, 2020).

Research exploring intergenerational data have also pointed to potential parental genetic nurturing effects (passive rGE) by showing that parental recurrent depression increases risk of offspring depression manifestation (Jaffee et al, 2021; Josefsson et al, 2019; Rutter et al, 2006) which may be mediated by maltreatment or neglectful behaviour (Yang et al, 2018).

Moreover, twin study findings have also shown that the proportion of environmental and genetic components of traits can change throughout the life-course, e.g. traits showing varying heritability estimates at different ages (increased heritability in young adulthood as opposed to childhood), potentially highlighting shifts to different forms of rGEs in later life (Hufer et al, 2020). This highlights the ever-changing properties of rGEs. In order to fully understand rGE effects across the life-span, high quality data from longitudinal and prospective study designs will be needed, which unavoidably makes the exploration of these effects much more complex.

Adoption designs are particularly powerful, as conventionally parental genetic nurturing effects (passive rGEs) are largely eliminated, due to adoptive parents and offspring being genetically unrelated. Associations between offspring traits and adoptive parent traits and/or the rearing environments they shape are likely to be a result of either environmental parenting effects or evocative rGE processes (Perlstein & Waller, 2020). Adoption study designs with available adoptive and biological parent data (546 families) have shown antisocial behaviour of biological parents to be significantly associated with negative parenting elicited by adoptive parents, pointing to the presence of evocative rGE (Klahr et al, 2017). Findings from

studies (320 families) utilising this design can also be highly translational by highlighting targetable behaviours; findings have shown that biological maternal attention-deficit/hyperactivity was associated with hostile parenting elicited by adoptive parents, with this effect being mediated by offspring impulsivity (Harold et al, 2013).

Moreover, the adoption design has also yielded fruitful results providing evidence of parental genetic nurturing effects utilising molecular genetic data. One study highlighted associations between adoptive parental PGSs and adopted offspring traits (Domingue & Fletcher, 2020), suggesting parental genetic effects on offspring traits that are not via the offspring's own genome (as adoptive parents and offspring are likely to be unrelated). Another study compared PGS and trait associations between adopted and non-adopted individuals, with results suggesting that PGSs account for greater variance in non-adopted individual's traits, likely due to the added variance attributable to parental genetic nurturing effects at play (Cheesman et al, 2020b). The adoption design has provided great insight into the role of rGEs in the manifestation of complex traits, including MDD (Lehto et al, 2020). However, smaller sample sizes become a major limitation when aiming to quantify rGEs (Philibert, 2006), especially as these effects are thought to be much smaller than the main genetic effects involved in trait manifestation (Kong et al, 2018).

These rGE effects have also been implied from studies using candidate genes, such as variations in the 5-HTTPLR being associated with lower child self-control, which in turn was associated with negative parenting (Pener-Tessler et al, 2013) amongst other similar findings (Brody et al, 2017; Feldman & Bakermans-Kranenburg, 2017). However, limitations of candidate gene studies outlined above, still apply.

In contrast, analyses exploring associations between PGSs and environmental variables, such as urban living (Maxwell et al, 2021), have highlighted that PGSs may be useful in the exploration of rGE effects. In fact, many studies have utilised PGSs in the exploration of rGEs in mediation models. Mediators explored were various

environmental factors through which genetic effects may influence a given trait, with PGSs representing genetic measures (Avinun & Hariri, 2019; Shao et al, 2021; Stephan et al, 2020; Warrier & Baron-Cohen, 2019). These findings have been interpreted as evidence of rGEs, with additional insight into the mechanism through which rGE effects occur. However, similar to the limitations associated with candidate genes, the limitations of PGSs highlight issues in the validity of these models, and thus, the interpretation of findings. The fact that PGSs capture only a fraction of the genetic variance in any given trait, and that most mediating environmental factors explored are heritable, with shared genetic variance with the traits being explored, suggests crucial associations are genetically confounded (**Figure 7**) (Akimova et al, 2020; Pingault et al, 2019).

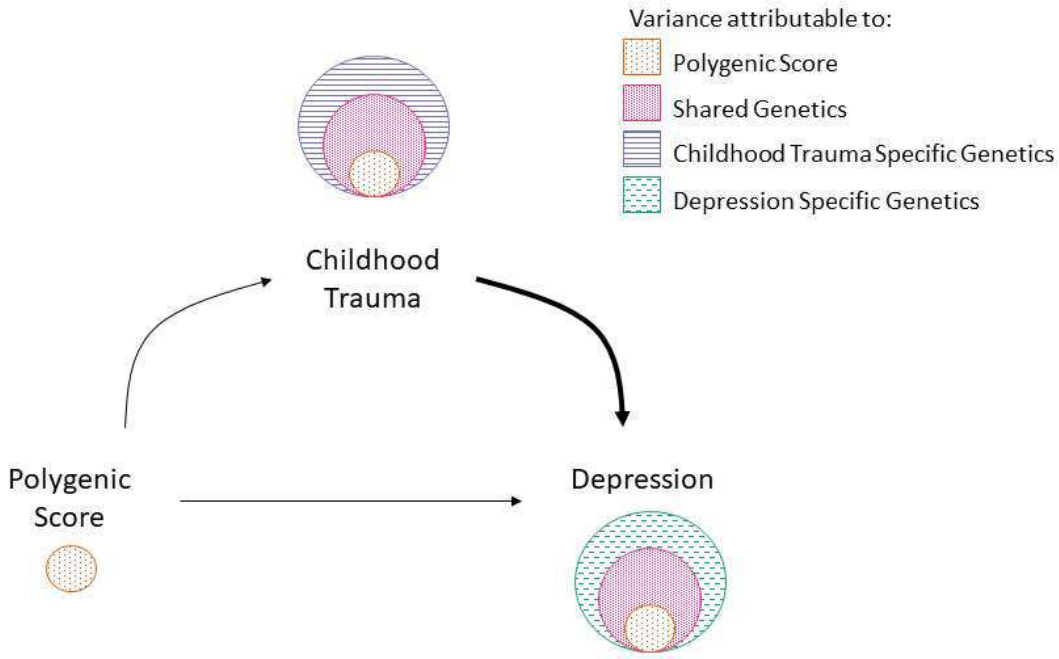


Figure 7. Methodological Considerations when exploring Gene-Environment Correlations using Polygenic Scores; Genetic Confounding. Here depression is the example outcome, and childhood trauma exposure is the mediator. As childhood trauma and depression share genetic variance that is unlikely to be captured entirely by the polygenic score, the highlighted path is confounded. When the highlighted path is significant, a potentially inaccurate interpretation of results would be that PGSs have a direct and a childhood trauma exposure mediated effect on depression.

Moreover, these mediation models may also be limited as paths can be affected by collider bias. If the exposure variables are independently caused by the predictor and outcome variables, then the exposure variable will be termed a 'collider variable'. Including a collider within the model will then bias crucial pathways i.e. the association between the predictor and outcome variable (**Figure 8**). In contrast to the genetic confounding effect described above, this bias will be amplified as PGSs become better measures of genetic variance (Akimova et al, 2020; Pingault et al, 2019).

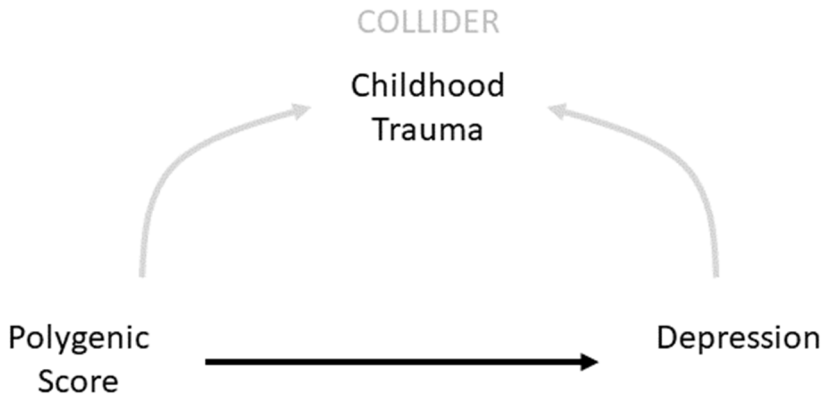


Figure 8. Methodological Considerations when Exploring Gene-Environment Correlations using Polygenic Scores; Collider Bias. Here depression is the example outcome, and childhood trauma exposure is the mediator. It may be possible that Childhood Trauma is independently caused by the polygenic score and Depression, which would make this a collider variable. When this collider is modelled, biased/distorted associations may arise between predictors and outcome variables (the Polygenic Score and Depression), which is a result of collider bias and not a genuine association.

These models, although potentially confounded and biased, are a good starting point as statistical models actually quantifying the genetic effect via the 'environment' are needed as results can provide useful insight into rGE mechanisms. New software have been developed aiming to circumvent the methodological challenges associated with PGS predictive validity, confounding and collider bias, such as Gsens (Pingault et al, 2021), which may yield findings with greater reliability.

Furthermore, Kong and colleagues (2018) propose a formula leveraging phased genetic information in the form of PGSs derived from transmitted and non-transmitted (from parent to offspring) alleles to quantify parental genetic nurturing effects (passive rGEs) on educational attainment. This method is yet to be applied to depression or other psychiatric traits, but findings can point to an estimate of effect or variance attributable to genetic nurturing effects. As mentioned above, the highly translational component of research exploring rGE effects lies in the identification of targetable 'environmental' factors or, in this particular case, parental behaviours. More information on the statistical challenges associated with PGSs in the exploration of parental genetic nurturing effects can be found in Chapter 2 of this thesis.

These parental genetic nurturing effects have been explored using mixed linear models and family data. In addition to GRMs capturing genetic similarity in the offspring generation, including GRMs capturing genetic similarity between unrelated offspring mothers or fathers, can provide insight into whether additional variance is uncovered. In this case, variance attributable to parent genetic variation, over and above variance attributable to offspring genetic variation can be thought of as parental genetic nurturing. Some parental genetic nurturing effect variance, ~14%, has been quantified for depression (Cheesman et al, 2020a), although other attempts with similar methodology have shown non-significant findings (Jami et al, 2020). These methods require vast amounts of intergenerational genetic and phenotypic data, which is far less prevalent than that observed for unrelated individuals. Moreover, it is important to note that different genetic nurturing (rGE) effects will occur within siblings, half-siblings and other family members and ideally should be modelled (Bijma, 2010), which further highlight statistical power issues in accurately detecting these effects.

1.2.3.2. Gene-Environment Interactions

Gene-environment interactions (GxE) are intrinsically different to rGEs. These effects can be conceptualised as a genetic influence on *sensitivity* to environments,

which go onto exacerbate trait/disease manifestation. Continuing with the same analogy used to outline rGE processes, an interaction effect may present as the observed musical ability difference between talented and untalented individuals (determined by genetic propensity), being exacerbated (in principal, moderated which could mean exacerbated or reduced) when both individuals are exposed to an environment that promotes musical ability e.g. access to music lessons (**Figure 9**).

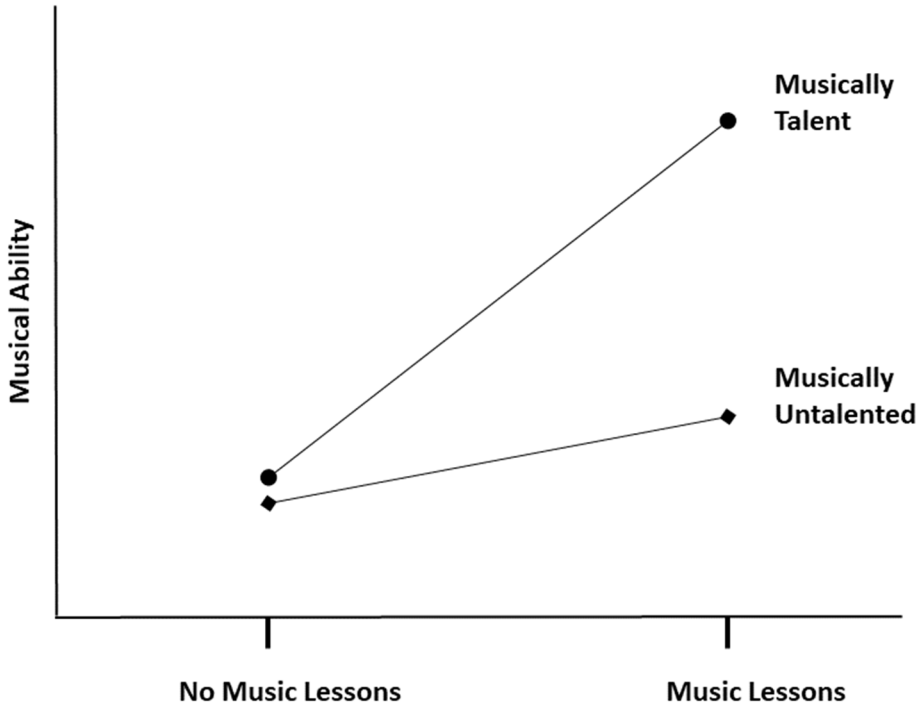


Figure 9. Gene-Environment Interactions in Musical Ability. In principle, interaction effects can come in many forms. Outcomes could be exacerbated or reduced, for example genetic appropriate musical training may make individuals more similar.

The diathesis-stress model is a conceptual framework used to explain GxE effects within disease manifestation. This model suggests that the effects of environmental risk factors, like adversity exposure, only result in psychopathology (i.e. individual's exceeding a threshold, please see liability threshold model outlined in section 1.2) when combined with inherent vulnerability to disease. Suggesting some individuals

to be more vulnerable and sensitive to adversity exposure than others (Assary et al, 2018; Colodro-Conde et al, 2018).

A substantial amount of research has highlighted the role of several environmental factors such as socio-economic status, urban environment, minority status and trauma exposure in the manifestation of MDD (Hoy-Ellis & Fredriksen-Goldsen, 2017; Jaffee, 2017; Johnson et al, 2016). Family members typically share many of the listed environments. However, twin study estimates of MDD variance attributable to shared environmental experiences are rather small (Polderman et al, 2015), which has formed the basis of the shared environment paradox. It is thought that if the effects of these environments are conditional on genetic liability, then the variance captured would be attributed to the genetic component, thus, inflating estimates of heritability (Taylor, 2007). Hence, exploring GxE interaction effects may be useful in addressing the missing heritability problem (Werme et al, 2021) as well as identifying risk environments, but more importantly, at-risk individuals to target with interventions.

Unsurprisingly, the search for GxE effects was initiated with candidate gene studies. An extremely exciting finding was the significant interaction effect between the 5-HTTPLR short variant and exposure to childhood adversity exacerbating risk of MDD manifestation (Caspi et al, 2003). Similar findings were observed with the brain derived neurotrophic factor gene variants and stressful life events (Hosang et al, 2014). However, as outlined above, failed replication attempts (Border et al, 2019) alongside the polygenic nature of these traits, with variants of small effects (Wray et al, 2014), have deemed findings from these studies obsolete.

Classic GWAS have been expanded to explore interaction effects, resulting in the application of genome-wide by environment interaction studies (GWEIS). These models further incorporate the effect of an environmental factor and the SNP by environment interaction into the regression analyses.

$$y = \beta_0 + \beta_1 SNPx + \beta_2 Env + \beta_3 SNPx \times Env + \varepsilon$$

Where y represents the value of the trait being explored; β_0 refers to the intercept term; β_1 refers to the regression coefficient representing the strength of association between (main effect of) SNP x and phenotype y; β_2 refers to the main effect of the environmental factor and phenotype y; β_3 refers to the regression coefficient representing the strength of association (effect sizes) between the SNP and environment interaction and phenotype y; ε refers to the error term representing noise.

GWEIS have allowed for the exploration of GxE effects without pre-existing hypotheses on the genetic variants utilised. However, due to the requirement of *both* genetic and high quality environmental measures, sample sizes in GWEIS are far smaller than GWAS (Uher & Zwicker, 2017); which may explain the lack of genome-wide significant interaction findings, especially in comparison to GWAS (Arnau-Soler et al, 2019b; Coleman et al, 2020; Dunn et al, 2016; Werme et al, 2021). To date, only a handful of GWEIS have been conducted for MDD, and so the robustness of current findings with replication attempts are yet to be determined (Werme et al, 2021).

The use of PGSs when exploring GxE effects have also failed to provide further insight into the role of GxE in depression and other psychiatric traits. Whilst exciting initial findings suggested significant PGS and stressful life event/trauma exposure interaction effects on MDD (Mullins et al, 2016; Peyrot et al, 2014), opposing effects of interactions on disease outcome highlighted the urgent need for replication studies. Similar to candidate gene studies, replication attempts highlighted the probability of significant results being false positive findings (Kendall et al, 2021; Peyrot et al, 2018). Issues with PGS predictive ability and validity, as seen with models exploring rGE effects, also yield methodological challenges when exploring GxE effects (Akimova et al, 2020).

PGSs have been utilised in mixed linear models alongside GRMs to address the limitations associated with PGSs capturing only a fraction of phenotypic variance; with findings highlighting only nominally significant PGS-by-Environment effects,

explaining very little additional variance (<2%) (Arnau-Soler et al, 2019a; Gillett et al, 2021). It seems evident that the results of research exploring GxE using PGSs are limited. GxE are very small effects (Assary et al, 2018) and PGSs are greatly underpowered, making it difficult to explore these effects.

An additional methodological complication that is often overlooked lies in the dichotomous/categorical measurement of traits like MDD, where individuals are often classified as cases or controls. The diathesis-stress framework and liability threshold model used together can highlight this complication. Whilst on the observed scale it may seem that individuals with greater genetic risk show exacerbated MDD symptoms when exposed to adversity, suggesting a GxE effect, the nature of this effect may in fact be *additive* on the liability scale (**Figure 10**).

The diathesis stress theory highlights that the same amount of *additive* environmental risk is likely to push individuals with greater genetic liability over the threshold as opposed to individuals with lower genetic liability. As individuals with greater genetic liability are more likely to be classified as cases, exploring these effects using case and control status may result in essentially additive effects (on the liability scale) to be interpreted as GxE effects (on the observed scale) which further complicates the already challenging interpretation of findings.

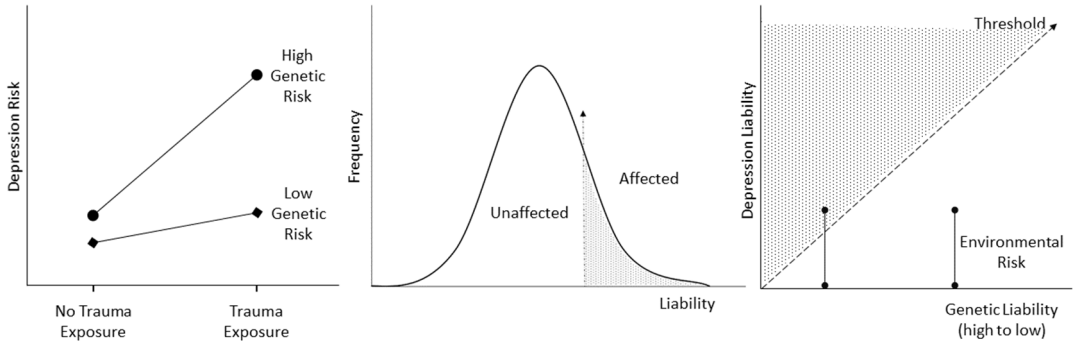


Figure 10. Methodological Considerations when Exploring Gene-Environment Interaction Effects. Left, shows how exacerbated risk of depression when exposed to adverse environments can be interpreted as an interaction effect on the observed scale. Middle, shows the continuous normal liability scale of traits, which are dichotomous on the observed scale. Right, shows the additive effect of the same amount of environmental risk resulting in individuals with greater genetic liability exceeding the threshold of disease.

1.3. Thesis Statement

There are vast amounts of research findings within the literature highlighting the importance of both the genetic and environmental component of depression. As outlined throughout this chapter, much has been learnt over the years, about the aetiology of depression. The presence of interplay between these components seem intuitive, and the literature with research utilising twin study and molecular genetic method designs support these intuitions. However, the extent to which these findings can be considered reliable evidence or proof of gene-environment correlation or interactions effects is questionable, as outlined above.

It is obvious that both forms of gene-environment interplay are highly complex, and in reality, this complexity is even further amplified; as in many cases, it is likely that rGE and GxE effects co-occur. Disentangling the two can be complicated, and disregarding one or the other, can result in further biased findings (Lau & Eley, 2008).

Moreover, this chapter has outlined a lot of methodological concerns associated with the genetic components regarding gene-environment interplay effects; however, it is important to note, that other environmental variables may also have confounding and biasing effects when modelling gene-environment interplay. Evidently, the genetic and environmental components consist of effects that are small, correlated, interacting, but most importantly, part of a much wider network of effects. Whilst these are extremely important issues in gene-environment interplay exploration, this thesis will mainly focus on highlighting and addressing the main methodological concerns raised throughout this chapter.

This thesis entails two main chapters, chapters 2 and 3. Chapter 2 is focused on gene-environment correlation effects, specifically parental genetic nurturing effects. Research aims consisted of (1) utilising PGSs in models exploring genetic nurturing effects from parent to offspring; using up to 2860 biological parent-offspring (trio) data available from Generation Scotland; (2) expanding models to

explore parental traits, which may impact offspring rearing environments, and thus, offspring traits. A variety of traits, with differing heritability and PGS power were explored. These findings were compared to results using simulated trio data and PGSs, which highlighted specific paths within these models that are affected by confounding. This chapter highlights that only the use of essentially perfect measures of genetic variance will negate the methodological concerns of the outlined models, which is highly unfeasible.

Chapter 3 is focused on the other form of gene-environment interplay; gene-environment interactions. The research aim (3) was to implement a variance components analysis to explore GxE effects. Three different relationship matrices; genetic, environmental and gene-by-environment interaction, were computed. The GxE relationship matrices were derived from the former two using matrix algebra, which also ensured the effect explored was in fact, *multiplicative* as opposed to *additive* (please see section 1.2.3.2 Gene-Environment Interactions for more information on this methodological concern). Findings indicate a large proportion of variance in depression can be attributed to genome-by-trauma exposure interaction effects, which much of the literature have hoped to, yet failed to show with methods utilising PGSs. This method is relatively powerful as it utilised all available genotyped data and a large sample size (~150,000) from UK Biobank, which has subsequently yielded high impact findings. Chapter 3 also includes a discussion on how this method can be further adapted to explore the mechanisms, such as biological processes, which may underlie the interaction effects.

The final chapter, chapter 4, entails a summary of the main findings and what these have contributed to our knowledge and understanding of methodology exploring gene-environment interplay effects. Moreover, chapter 4 includes an important evaluation of strengths, limitations and directions for future work with the available methods.

Chapter 2: Methodological Considerations When Using Polygenic Scores to Explore Parent-Offspring Genetic Nurturing Effects

2.1. Abstract

2.1.1. Background

Research has begun to explore the effects of parental environments or behaviours (parent nurturing) on offspring outcomes using polygenic scores (PGSs). Findings suggest that the presence of parental genetic nurturing effects can be detected with PGSs. However, concerns have been mooted about the potential for bias due to confounding when expanding these models to incorporate nurturing via parental phenotypes.

2.1.2. Methods

The GS cohort is a family health based genetic epidemiology study with genetic and environmental data from 20,000+ individuals aged 18-98 across Scotland. 2680 biological parent-offspring trios with genotyped data are available. Here, depression, educational attainment and height PGSs were generated for trios using the largest genome-wide association study (GWAS) meta-analyses available for each trait.

Analyses consist of regression and pathway models incorporating PGSs for both parents and offspring to explore direct (offspring PGS) and genetic nurturing (parental PGS) effects on psychological distress, depression, educational attainment and height. Genetic nurturing via parental phenotypes (the same trait as was being explored in the offspring) was also incorporated into the models. To explore sources of bias within these models, simulation analyses of 10,000 trios using combinations of PGS predictive accuracy and accounted variance were conducted.

2.1.3. Results

Evidence for genetic nurturing effects was explored with comparisons of regression models incorporating offspring PGSs with those incorporating both offspring and

parental PGSs. The latter model accounted for significantly greater variance in educational attainment with positive estimates of parental PGSs on offspring educational attainment. This was not the case for the traits psychological distress, depression and height; providing evidence for genetic nurturing effects for educational attainment but not for the other traits. However, results from models incorporating parental phenotypic information together with parental and offspring PGSs appeared to provide biased estimates, with significant negative estimates of parental PGSs on offspring for height.

The analysis of simulated data demonstrated that models incorporating only parent and offspring PGSs resulted in genetic nurturing effects that were unbiased. However, the incorporation of parental phenotypes as explanatory variables in these models shows that, despite the absence of simulated genetic nurturing effects, the use of PGSs that fail to capture the entirety of trait genetic variance results in positive estimates of parent phenotype mediated genetic nurturing effect and negative estimates of residual genetic nurturing effects. These biases were exacerbated as PGS accuracy and accounted variance decreases and are similar to the results observed from the analysis of real data.

2.1.4. Discussion

Results suggest that models combining parental and offspring PGSs can be used to detect genetic nurturing effects in complex traits without the need to discriminate between transmitted and non-transmitted parental alleles. Power to detect genetic nurturing effects is likely to be greater for traits that are more heritable with relatively greater PGS accuracy and accounted variance, as is the case for educational attainment.

Exploring genetic nurturing via parental phenotypes can point to targets for interventions. However, results demonstrate that in the absence of genuine genetic nurturing effects; limitations related to PGS predictive ability, coupled with parental phenotypes capturing larger proportions of trait genetic variance, result in false

positive genetic nurturing effects when regression or pathway models additionally incorporate parental phenotypes.

Findings suggest a relatively simple approach utilising parental and offspring PGSs can be used to identify evidence for parental genetic nurturing. Potential future directions include utilising varying PGSs between parents and offspring within these models to explore genetic nurturing effects of different parental phenotypes.

2.2. Introduction

Many traits and disorders are subject to genetic and environmental effects. Evidence of gene environment correlations (rGE) i.e. where an individual's genotype for a trait is also correlated with environmental influences, highlight that these effects may not be independent of one another (Gage et al, 2016).

Different labels have been used to explain these effects across different fields. Here, the term *parental genetic nurturing* is used when referring to rGE effects arising between biological parents and offspring (trios). Parents can have a *direct* genetic effect on offspring phenotypes by transmitting half of their genome to form the offspring's own genome. Parents can also have a *genetic nurturing* effect, as parent genetics contribute to parental phenotypes, which in turn shape the offspring's own rearing environment (Kong et al, 2018) **(*Figure 1*)**. It is important to note that a range of parental phenotypes can have a genetic nurturing effect on a given offspring phenotype. However, within this study, models conducted were based on the assumption that the same trait is involved in nurturing and the offspring trait.

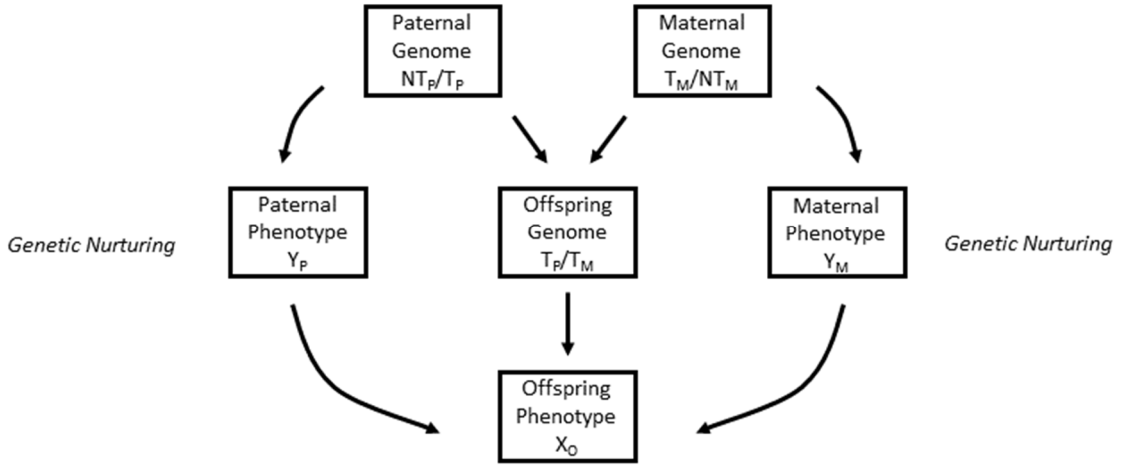


Figure 1. Direct genetic and genetic nurturing effects. Figure and legend adapted from Kong and colleagues (2018). T_P and T_M denote, respectively, the alleles transmitted from the father and mother. NT_P and NT_M denote the paternal and maternal alleles that are not transmitted. The paths show that the transmitted alleles can influence the phenotype of the offspring, X_O, through a direct path. The paths also show that both transmitted and non-transmitted alleles can influence the parent phenotypes, Y_P and Y_M, and through them have a genetic nurturing effect on the offspring phenotype, X_O. Whilst X is an individual trait of interest, Y may include a range of phenotypes.

The use of human family data within genome-wide association studies (GWAS) aiming to explore trait associated genetic variants, have demonstrated that heritability estimates for many complex human traits, such as depression, height, and educational attainment, are upwardly biased (Yang et al, 2015). This bias is attributed, but not limited to, unaccounted genetic nurturing effects (Howe et al, 2021; Yang et al, 2015; Young, 2019). Animal designs have provided theoretical frameworks on how to quantify genetic nurturing effects by modelling parental genome effects on offspring traits, over and above the (direct genetic) effects of transmitted alleles (Falconer & Mackay, 1996; Griffing, 1967; Mackay et al, 1999; Willham, 1963). These mixed model methods exploring variance components using available trio data have been implemented using available human data, showing evidence of variance attributable to genetic nurturing effects in educational attainment as well as depression and anxiety disorders (Cheesman et al, 2020a; Young et al, 2018), although findings from studies with similar designs have shown non-significant findings (Jami et al, 2020).

The use and accuracy of polygenic scores (PGSs), derived from a weighted additive

sum of trait associated alleles, has increased substantially over the last decade. Methods employing PGSs have suggested genetic nurturing effects in educational attainment by showing that scores were twice as predictive within non-adopted individuals in comparison to adopted individuals (Cheesman et al, 2020b) as well as adoptive parental PGSs showing significant associations with adopted offspring education outcomes (Domingue & Fletcher, 2020). Studies exploring trio PGSs, whereby direct genetic effects are controlled for by using offspring PGSs, provide evidence for potential genetic nurturing effects in educational attainment (Morris et al, 2020). Moreover, the use of phased family data have allowed for the identification and calculation of PGSs for alleles transmitted and non-transmitted from parent to offspring (Kong et al, 2008). Findings using the non-transmitted PGSs demonstrate that genetic nurturing effects are present, albeit much smaller than direct genetic effects (Bates et al, 2018; Kong et al, 2018).

The translational advantage of available datasets with both genetic and environmental information from unrelated individuals and families, lies in the ability to quantify and identify behaviours or environmental factors through which genetic or genetic nurturing effects are observed. This information can provide actionable findings that can influence preventative treatment or policy. This has resulted in a steady accumulation of studies utilising PGSs within mediation analyses to explore particular environmental factors through which genetic effects may influence a given trait/disorder (Avinun, 2019; Avinun & Hariri, 2019; Shao et al, 2021; Stephan et al, 2020; Warrier & Baron-Cohen, 2019). However, limitations associated with PGS predictive validity (i.e. accuracy of allele effect sizes and low numbers of associated genetic variants) are still present and have resulted in scores capturing only a fraction of the phenotypic variance for complex human traits (Dudbridge, 2013; Mostafavi et al, 2020). This, coupled with the lack of independence between genetic and environmental factors poses a challenge for statistical modelling of environment mediated PGS-trait associations (Akimova et al, 2020; Pingault et al, 2019). This methodological challenge is even more apparent when exploring parent phenotypes and genetic nurturing effects, as parent and

offspring phenotypes are inevitably influenced by direct parental genetic effects. Consequently, the seemingly environmental variable (i.e. parent phenotype) is both heritable and genetically correlated with the offspring phenotype.

In this study, four different complex phenotypes; psychological distress, depression, educational attainment and height were explored. Phenotypes substantially differ in terms of heritability estimates and PGS predictive validity. Data from up to 2680 biological parent-offspring trios within the Generation Scotland sample were utilised within analyses. Relatively straightforward regression and pathway models were used to explore parental direct (offspring PGSs) and genetic nurturing effects (parental PGSs) on offspring phenotypes using trio PGSs. This study sought to demonstrate the aforementioned methodological challenges introduced with the expansion of these models with parent phenotype mediated genetic nurturing paths. Simulations with absent genetic nurturing effects were conducted to demonstrate the effects of genetic confounding on estimates of genetic nurturing paths utilising PGSs and parental phenotypes.

2.3. Methodology

2.3.1. Participants

Participants were identified from Generation Scotland: The Scottish Family Health Study (GS). GS is a family- and population-based study that recruited individuals from 2006 until 2011. Currently GS has genetic, environmental and phenotypic data on over 20,000 individuals aged 18-99 (Smith et al, 2013). To explore genetic nurturing effects from both parents to offspring, the study sample was limited to individuals with both biological parents genotyped (trios) resulting in 2680 parent-offspring trios. This sample includes 1596 singleton offspring, with the remaining offspring having at least one sibling within the sample. These sibling effects were controlled for by either including them as random effects into mixed linear models, where possible, or by including only 1 sibling in the analysis sample. The analysis samples were limited to trios with available phenotypic data on psychological distress, depression, educational attainment and height.

2.3.2. Phenotypes

2.3.2.1. Psychological Distress & Depression

Psychological distress and depression are moderately heritable traits, with twin study estimates ranging from 20-44% (Rijsdijk et al, 2003) and 30-45% (Fernandez-Pujals et al, 2015; Sullivan et al, 2000) respectively. The traits were assessed using the General Health Questionnaire (GHQ-28), a screening tool used to identify non-psychotic and minor psychiatric disorders (Sterling, 2011). The GHQ-28 is well-validated and provides assessments of somatic symptoms, anxiety and insomnia, social dysfunction and severe depression (Goldberg & Hillier, 1979; Koeter, 1992). Each subsection is assessed using 7 behavioural items with a 4-point Likert scale (0-4) exploring the frequencies of experience: “not at all”, “no more than usual”, “rather more than usual”, “much more than usual”. This results in a minimum score of 0 and maximum score of 28 for depression (using only the severe depression questions) and 84 for psychological distress (using all questions).

2.3.2.2. Educational Attainment

Educational attainment was available as a 1-10 point ordinal scale with higher values representing greater years of education. Educational attainment was included in these analyses as significant genetic nurturing effect associations have been found within the literature (Bates et al, 2018; Cheesman et al, 2020b; Kong et al, 2018; Morris et al, 2020). Furthermore, educational attainment has relatively high heritability estimates of ~66% (de Zeeuw et al, 2016; Schwabe et al, 2017) which is likely to yield greater statistical power in comparison to psychological distress and depression.

2.3.2.3. Height

Height was measured in centimetres (cm). Similar to educational attainment, height was explored as this is a trait with high heritability, with estimates showing approximately 85% of variation to be attributable to genetic differences (Silventoinen et al, 2003). Height has also yielded well powered GWAS (Yengo et al,

2018) resulting in PGSs with relatively high predictive validity i.e. accounting for a greater proportion of phenotypic variance in comparison to depression and educational attainment PGSs

Sample sizes for trios with available phenotypic data are 2,315; 2,280; 2,361 and 2,648 for psychological distress, depression, educational attainment and height, respectively. Information on sample demographics for each phenotype as well as information on age and sex distribution in the whole GS and trio samples are available in (Appendix I Table 1).

2.3.3. Genotypes

Genetic data, single nucleotide polymorphisms (SNPs), are obtained from blood samples collected using standard operating procedures and genotyped using the IlluminaHumanOmniExpressExome-8v1.0 BeadChip and Infinium chemistry capturing 700,000 genome-wide SNPs and 250,000 exome SNPs (Gunderson, 2009). Quality control (QC) of genotyped SNPs consisted of the exclusion of SNPs with missingness > 2% and a Hardy Weinberg Equilibrium (HWE) p-value of $\leq 1 \times 10^{-6}$. SNPs with minor allele frequency (MAF) <0.01 and individuals with >3% missing genotypes were excluded from analyses. After QC procedures 561,125 SNPs were included in analyses.

2.3.4 Polygenic Scores (PGSs)

Depression, educational attainment and height PGSs for GS participants were created using allelic weight information from separate GWAS meta-analyses with relatively large sample sizes **(Table 1)**. Depression PGSs were used in models exploring both psychological distress and depression. To avoid over-fitting, GS participants were removed from these GWAS meta-analyses. The publically accessible version of educational attainment GWAS was utilised (23&Me data within the educational attainment GWAS meta-analyses were removed). Multiple PGSs were created using SNPs with varying p-value trait association thresholds (ranging from $p < 5x10^{-8}$ to 1). As GS is primarily a family-based study, the 1000G linkage disequilibrium (LD) reference was used for the clumping of SNPs. PRSice2

(Choi et al, 2020) with the default clumping (LD r2 < 0.1 across a 250 kb window) algorithm was used to compute the PGSs.

As multiple phenotypes are explored, identifying which set of PGSs is the most suitable for downstream analyses becomes intricate. For example, utilising the PGS that accounts for the greatest variance in the available GS dataset can lead to PGSs being overfit to the data, and thus, biased results. Additionally, the SNP association p-value threshold resulting in the best performing PGS can be different (and thus, the number of SNPs utilised in the calculation of PGSs may also be different) for each phenotype. To ensure uniform methodology, a principal components analysis is conducted on the different sets of PGSs created for each phenotype. The first principal component (PC) is utilised as the trait PGS. The principal component was found to be highly correlated with all different sets of PGSs, and each PGS loaded onto the PC equally. Information on the loadings and correlations as well as PGS and phenotype associations can be found in Appendix I Table 2. This method has been shown to produce PGSs that have increased predictive ability, and more importantly, are not overfit (Coombes et al, 2020). The genome-wide association studies utilised to train the generated PGSs are presented **(Table 1)**.

Table 1. Generation Scotland Polygenic Risk Score Demographics

Polygenic Score (PGS) Trait	GWAS	GWAS N
Depression	(Howard et al, 2019)	807,533
Educational Attainment	(Lee et al, 2018)	750,403
Height	(Yengo et al, 2018)	693,529

Abbreviations. GWAS, genome-wide association study; N, sample size; SNP, single nucleotide polymorphism.

2.3.5 Statistical Analyses

2.3.5.1 Generation Scotland Analyses

All statistical analyses were conducted in R (Team, 2020). Initial analyses explored phenotype and PGS associations using regression analyses (Appendix I Table 2.2).

Pearson's correlation tests were conducted to explore trio PGS and phenotype correlations (Appendix I Table 3). For subsequent regression and pathway models, PGSs were standardised to have a mean of 0, and standard deviation of 1; all variables were then pre-corrected for covariates; the first 20 genetic PCs of offspring and parents (to control for population stratification and cryptic relatedness), offspring and parental age and offspring sex.

2.3.5.1.1. Regression Analyses

Mixed effect multiple regression analyses were conducted (function 'lmer', package 'lmerTest') within R. Four separate models were explored and compared for the different phenotypes of interest (**Table 2**):

Table 2. Regression Model Equations

Model	Equation
I	$oPheno \sim oPGS + Sib$
II	$oPheno \sim oPGS + mPGS + pPGS + Sib$
III	$oPheno \sim oPGS + mPheno + pPheno + Sib$
IV	$oPheno \sim oPGS + mPGS + pPGS + mPheno + pPheno + Sib$

Abbreviations. Pheno, Phenotype; PGS, polygenic score; m/p/o, maternal, paternal, offspring respectively; Sib, sibling effect.

Model *I* aimed to capture direct genetic effects, i.e. the risk-associated effect of the offspring's own genome on its phenotype, by using the offspring PGSs. In contrast, model *II* aimed to capture both direct and parental genetic effects over and above the direct genetic risk i.e. genetic nurturing effects. If model *II* has a statistically better fit than model *I*, then any additional accounted variance can be attributed to parental genetic nurturing effects. Models *III* and *IV* aimed to provide extra insight using parental phenotype effects as well as provide a comparison for the subsequent pathway models outlined below.

Offspring phenotypes (*oPheno* in models outlined in **Table 2**), psychological distress, depression, height and educational attainment, were included as dependent variables. Trio depression, height and educational attainment PGSs

(*oPGS, mPGS, pPGS)* were included as predictors. Parental phenotypes (*mPheno, pPheno)* were also included as predictors within the models.

Sibling effects were modelled as varying intercepts (random effects) for each sibling group to account for other sources of correlation between sibling phenotypes and avoid pseudoreplication. Sibling groups were identified as trio offspring who had the same parents.

Likelihood ratio tests were explored (function 'anova') to compare the model fit of the four different models for each phenotype. Results include false discovery rate (FDR) adjustments for multiple testing (function 'p.adjust', q-value threshold <0.05). This was to control for the 5 comparisons conducted (Appendix I Table 5).

The regression models were also explored using singleton, male and female offspring samples separately.

2.3.5.1.2. Pathway Analyses

Pathway analyses were conducted to explore the parental genetic and phenotypic effects on offspring traits. Pathway analyses can be thought of as an extension to multiple regressions where 'chains' of influence and covariance structures can be incorporated into the analyses (Streiner, 2005). The key utility of pathway analyses is that the effect and proportion of variance captured by specific paths can be estimated. **Figure 2** presents a visualisation of the difference between multiple regression and pathway models.

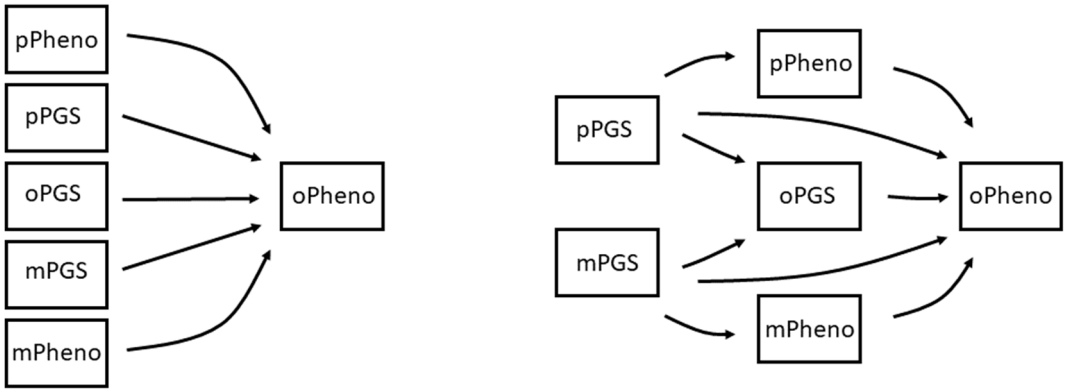


Figure 2. Visualisation of Multiple Regression and Pathway Models. PGS, polygenic score. Left, shows associations between the independent variables (offspring and parental PGSs and parental phenotypes) and the dependent variable (offspring phenotype). Right, shows associations between independent variables with the dependent variable whilst also capturing associations (mediation) between the independent variables forming pathways.

Pathway analyses were conducted using the software package, Lavaan (Rosseel, 2012; 2018). Two pathway models with varying levels of complexity, simple and extended, were explored (**Figure 3**).

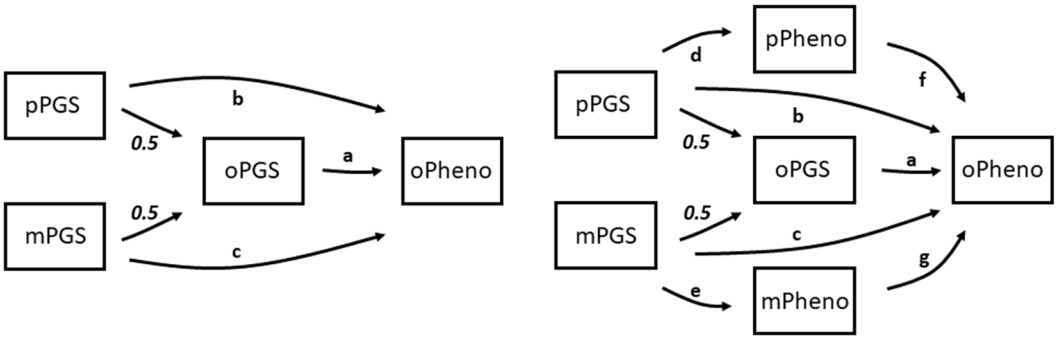


Figure 3. Simple and Extended Pathway Models. PGSs, polygenic scores; m, maternal; p, paternal; o, offspring. Left, Simple Model. Right, Extended Model. Associations between parental PGSs and offspring PGSs are fixed to 0.5 as parents and offspring share ~50% of their genome. Paths *0.5*a* represent direct genetic effects of parental PGSs on offspring phenotype. Paths *b* and *c* represent genetic nurturing effects of parental PGSs on offspring phenotype. Paths *d*f* and *e*g* represent the effect of parental PGS on offspring phenotype mediated by the parental phenotypes.

Similar to the regression analyses, offspring psychological distress, depression, height and educational attainment (*oPheno)* as well as trio depression, height and educational attainment PGSs (*oPGS, mPGS,* pPGS) were included as predictors within the simple pathway analyses. The extended model builds upon this by

further including parental phenotypes (*mPheno, pPheno*). Associations within the simple and extended pathway models are outlined in **Table 3**.

Table 3. Pathway Model Equations

Pathway	Equation
a	$oPheno \sim oPGS$
b	$oPheno \sim pPGS$
c	$oPheno \sim mPGS$
d	$pPheno \sim pPGS$
e	$mPheno \sim mPGS$
f	$oPheno \sim pPheno$
g	$oPheno \sim mPheno$
h	$oPGS \sim pPGS$
i	$oPGS \sim mPGS$

Abbreviations. Pheno, Phenotype; PGS, polygenic score; m/p/o, maternal, paternal, offspring respectively.

The coefficients of associations between parental PGSs and offspring PGSs are fixed to 0.5, as parents and offspring share ~50% of their genome. Pathway models were also conducted where all path coefficients were freely estimated, using singleton offspring (to eliminate sibling effects), only female and male offspring as well as for maternal and paternal duos separately.

2.3.5.2. Simulation Analyses

Simulations were conducted to explore how results of the regression and pathway models outlined above differed when using PGSs with varying levels of predictive ability. 10,000 trio members with respective genetic, phenotypic and PGS variables were simulated. These simulations did not include any genetic nurturing effects.

2.3.5.2.1. Simulations I

Simulated phenotypic variables had a variance of 1. Additive genetic variance was computed as the heritability multiplied by the phenotypic variance. Genetic variance was represented as two separate entities; tagged genetic variance, aiming to capture variance attributable to genotyped variants, and non-tagged genetic variance, aiming to capture variance attributable to non-genotyped variants. All remaining variance was attributed to environmental variance.

The parental genetic variables were constructed as the sum of tagged and non-tagged genetic counterparts. Tagged genetic variable values were simulated from a normal distribution with a mean of zero, and variance equal to the tagged genetic variance. Non-tagged genetic variable values were simulated from a normal distribution with a mean of zero, and variance equal to the non-tagged genetic variance. Similarly, the offspring genetic variables were constructed as the sum of tagged and non-tagged genetic counterparts. Offspring tagged and non-tagged genetic variables were simulated as the average of the summed respective tagged and non-tagged maternal and paternal genetic variables, with the addition of a respective tagged and non-tagged segregation variable aiming to capture variability that occurs from random segregation of genes observed during meiosis (Yanowitz, 2010). The tagged and non-tagged segregation variable values were simulated from a normal distribution with a mean of zero, and variance equal to half the tagged and non-tagged genetic variance, respectively.

Separate environmental variables were constructed for each member of the trio. These variables were simulated from a normal distribution with a mean of zero, and a variance equal to the environmental variance.

The phenotypic variables for each member of the trio were then constructed as the sum of the respective trio member's genetic and environmental variable values. Polygenic score variables were derived from the trio member's respective tagged genetic component.

Different scenarios, with 15 replications each, were conducted with a range of plausible trait heritabilities (0.3, 0.6, 0.9) and proportion of tagged genetic variances (0.2, 0.6, 1) in order for the findings to have broad applicability to the study of a wide variety of traits/complex trait architectures.

2.3.5.2.1. Simulations II

As mentioned above, PGSs are informed using information on variants found to be associated with traits of interest. Underpowered GWAS can also result in inaccurate beta coefficients for variant-trait associations, which will impact the power of PGSs.

Moreover, these biases associated with inaccurate variant effect sizes, are likely to be correlated between individual's with greater genotypic similarity e.g. family members.

To incorporate this level of complexity to the simulated data, parental noise variables were simulated from a normal distribution with a mean of zero, and variance equal to the noise variance. Offspring noise variables consisted of an independent component, which was simulated from a normal distribution with a mean of zero, and variance equal to half of the noise variance. The offspring noise variables consisted of an additional component correlated with the parental noise variables; to capture correlated inaccuracies/bias effects between individual's with greater genotypic similarity. This was simulated as the average of parental noise variables. Trio PGSs were then calculated as the sum of their respective PGS and noise variable.

Different levels of noise were explored (0-1, by 0.1 increments), with 0 representing complete accuracy in effect sizes of variants obtained from GWAS. Here, the value 1 does not represent complete noise. Creating a scenario where PGSs are entirely noise with no signal was not possible with the simulation strategy utilised in this study.

https://github.com/melisachuong/rGE_Depression

2.4. Results

2.4.1. Generation Scotland Results

Regression analyses exploring PGS associations with respective phenotypes show that all PGSs (created using different SNP association p-value thresholds) are positively and significantly associated with their respective phenotypes within the whole, offspring, maternal and paternal GS sample, with the exception of psychological distress and depression. Results show less evidence for these associations when utilising depression PGSs informed by only SNPs exceeding the most stringent thresholds (e.g. genome-wide significance) (Appendix I Table 3).

Correlations between the trio PGSs are in line with expectations. Offspring PGSs are correlated ~0.5 with parental PGSs. Furthermore, weak, yet, significant correlations between height and educational attainment PGSs are observed between the parents, consistent with evidence suggesting the presence of assortative mating (Conley et al, 2016; Hugh-Jones et al, 2016). Similarly, strong evidence of correlations between all trio members' phenotypes was observed (**Table 4**).

Table 4. Biological Parent-Offspring (Trio) Polygenic Score and Phenotype Correlations

Phenotypes	Pairwise Correlations	Polygenic Scores		Phenotypes	
		Pearson's R	P-Value	Pearson's R	P-Value
Psychological Distress	O-M			0.111	0.000
	O-P			0.084	0.000
	M-P			0.119	0.000
Depression	O-M	0.491	0.000	0.050	0.015
	O-P	0.488	0.000	0.056	0.006
	M-P	-0.021	0.272	0.043	0.036
Educational Attainment	O-M	0.551	0.000	0.238	0.000
	O-P	0.552	0.000	0.224	0.000
	M-P	0.124	0.000	0.414	0.000
Height	O-M	0.553	0.000	0.353	0.000
	O-P	0.527	0.000	0.368	0.000
	M-P	0.099	0.000	0.308	0.000

Abbreviations: O, Offspring; M, Maternal; P, Paternal.

The results of the four different regression models exploring direct and genetic nurturing effects are available (Appendix I Table 4). Notable results include negative and highly significant β coefficient estimates for parental height PGSs when included in models alongside parental height phenotypes (model *IV*). Interestingly, these estimates are positive and non-significant when included in models without parental phenotypes (model *III*). Log-likelihood comparisons of model *I* and *II*

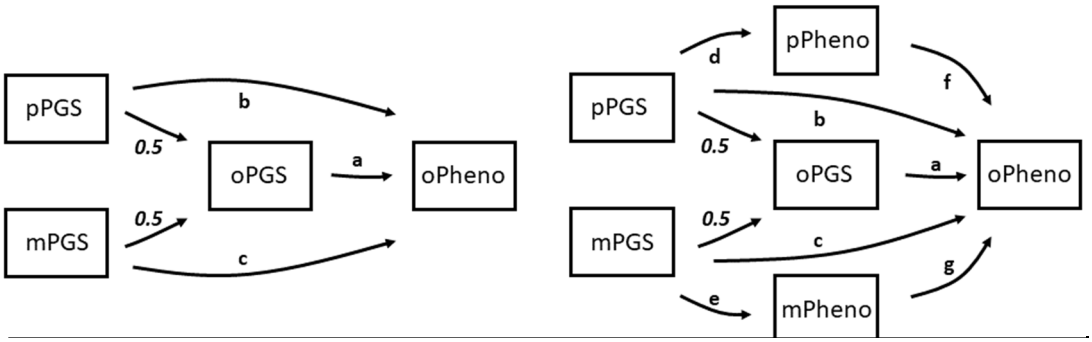
sought to identify genetic nurturing effects captured by parental PGSs. Results suggest that model *II* (including parental PGSs) accounts for greater variance and is significantly better than model *I* for educational attainment only (**Table 5**) (For all model comparison results see Appendix I Table 5).

Table 5. Regression Model I and II Log-Likelihood Comparisons

Phenotype	Model	logLik	Chisq	Df	P-Value	FDR
Psychological Distress	I	-7969.203				
	II	-7968.833	0.740	2.000	0.691	1.000
Depression	I	-4232.776				
	II	-4230.556	4.441	2.000	0.109	0.233
Educational Attainment	I	-4036.847				
	II	-4030.877	11.939	2.000	0.003	0.006
Height	I	-7985.51				
	II	-7981.57	7.87	2.00	0.02	0.04

Abbreviations. logLik; Log Likelihood; Chisq, Chi Squared Value; Df, degrees of freedom; FDR, False Discovery Rate.

Path β coefficient estimates, standard errors and p-values from the simple and extended pathway analyses are presented in **Figure 4**.



Phenotype	Label	Simple Model			Extended Model		
		Estimate	SE	P	Estimate	SE	P
Psychological Distress	a	0.678	0.243	0.005	0.693	0.241	0.004
	b	-0.103	0.210	0.625	-0.162	0.208	0.437
	c	0.149	0.218	0.494	0.008	0.217	0.969
	d				0.453	0.146	0.002

	e				0.922	0.170	0.000
	f				0.077	0.024	0.001
	g				0.113	0.022	0.000
Depression	a	0.093	0.050	0.064	0.091	0.050	0.069
	b	-0.053	0.043	0.216	-0.058	0.043	0.179
	c	0.061	0.045	0.175	0.054	0.045	0.228
	d				0.106	0.030	0.000
	e				0.128	0.033	0.000
	f				0.062	0.025	0.012
	g				0.043	0.024	0.074
Educational Attainment	a	0.199	0.041	0.000	0.192	0.040	0.000
	b	0.117	0.037	0.002	0.074	0.037	0.044
	c	0.095	0.036	0.009	0.024	0.036	0.499
	d				0.237	0.033	0.000
	e				0.285	0.028	0.000
	f				0.098	0.019	0.000
	g				0.144	0.021	0.000
Height	a	3.336	0.152	0.000	3.402	0.124	0.000
	b	0.190	0.131	0.148	-1.344	0.119	0.000
	c	0.417	0.135	0.002	-1.217	0.123	0.000
	d				3.263	0.106	0.000
	e				3.172	0.101	0.000
	f				0.386	0.016	0.000
	g				0.392	0.017	0.000

Figure 4. Simple and Extended Pathway Models Results. SE, Standard Error; P, P-value. Above, Simple and Extended Pathway Models. Paths between parent PGSs and offspring phenotypes via offspring PGSs represent direct genetic effects of parental PGSs on offspring phenotype. The β coefficients of paths between parent PGSs and offspring PGSs are fixed to 0.5 as it is expected for parents to pass on 50% of their genomic make-up to their offspring. Paths *b* and *c* represent genetic nurturing effects of parental PGSs on offspring phenotype. Paths *d*f* and *e*g* represent the effect of parental PGS on offspring phenotype mediated by the parental phenotypes. Below, table of pathway model results for depression, psychological distress, educational attainment and height.

The simple and extended models show no evidence of genetic nurturing effects (paths *b* and *c*) on psychological distress. Interestingly paths *d* (β=0.453, P<0.05)

and *e* (β=0.992, P<0.001), exploring parental PGS and parental phenotype associations, whilst expected to be similar, seem to be quite different, with maternal associations being ~2 fold greater than paternal associations. This pattern is also observed for depression.

Both the simple and extended pathway model results show a statistically significant negative paternal genetic nurturing (path *b*) effect on offspring depression; suggesting that as paternal genetic risk increases by one standard deviation, offspring depression decreases by β=0.053, P<0.05/β=0.058, P<0.05 standard deviations, respectively. The extended model further shows significant positive associations between the parental PGSs and parental depression (paths *d* and *e*, β=0.106, P<0.001 and β =0.128, P<0.001, respectively), as well as between parental depression and offspring depression (path *f* and *g*, β=0.062, P<0.05, β = 0.043, P=0.05, respectively). In fact, this pattern of significant positive genetic nurturing via parental phenotypes, is observed across all explored phenotypes.

The simple model results for educational attainment suggest clear evidence of both paternal and maternal genetic nurturing effects on offspring educational attainment (paths *b* and *c,* β=0.117, P<0.001, β=0.095, P<0.001 respectively*)* which support and build upon the findings of the regression model comparisons. Interestingly paths representing the association between the trio PGSs and their respective phenotypes in the extended model, are quite different to each other, with the parental PGS and parental phenotypic associations being almost ~2 fold greater than the offspring PGS and offspring phenotypic association (paths *a, d* and *e*; β=0.192, P<0.001; β=0.237, P<0.001; β=0.285, P<0.001, respectively). Furthermore, the inclusion of parental phenotypes within the extended model seem to result in a decrease in paternal genetic nurturing effect (from β=0.117, P<0.001 to β=0.074, P<0.001) and the maternal genetic nurturing effect becoming redundant (paths *b* and *c*) i.e. non-significant.

The simple pathway model results suggest a positive, yet, small maternal genetic nurturing effect on offspring height (β=0.417, P<0.05). Similar to regression model

IV results, the extended model results show the parental genetic nurturing effect (paths *b* and *c)* to become negative and highly significant (β=-1.344, P<0.001; β= -1.217, P<0.001), which is contradictory to the results from the simple model.

Patterns of results, and estimates are similar when all coefficients are freely estimated, as well as when samples are limited to singletons, female offspring, male offspring, maternal duos and paternal duos (Appendix I Table 6).

2.4.2. Simulation I Results

Regression and pathway analyses are conducted using simulated trio data. Each analysis is repeated 15 times. As no genetic nurturing effects are simulated, it is expected that paths *b, c, f* and *g* of the pathway models will be approximately 0 and non-significant when PGSs are adequately capturing genetic variance (**Figure 5**).

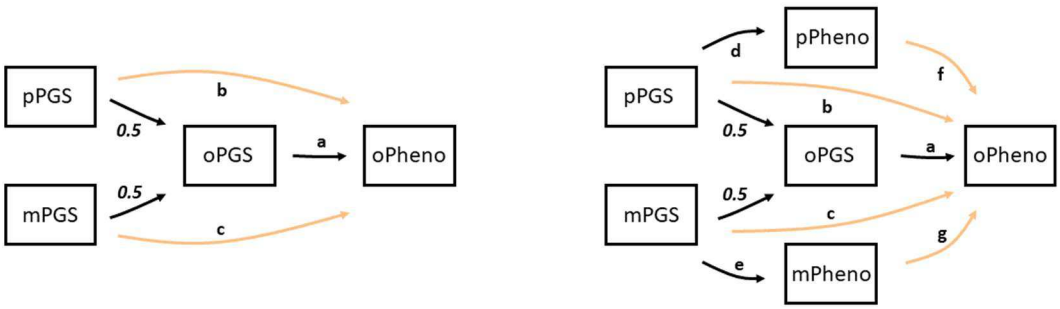


Figure 5. Simple and Extended Pathway Models. Pheno, Phenotypye; PGS, Polygenic Score; O, Offspring; P, Paternal; M, Maternal. Paths between parent PGSs and offspring phenotypes via offspring PGSs represent direct genetic effects of parental PGSs on offspring phenotype. The β coefficients of paths between parent PGSs and offspring PGSs are fixed to 0.5 as it is expected for parents to pass on 50% of their genomic make-up to their offspring. Paths *b* and *c* represent genetic nurturing effects of parental PGSs on offspring phenotype. Paths *d*f* and *e*g* represent the effect of parental PGS on offspring phenotype mediated by the parental phenotypes. Coloured pathways highlight association coefficients expected to be 0 and non-significant in the absence of genetic nurturing effects.

Figure 6 presents β coefficient boxplots of paternal genetic nurturing paths highlighted in **Figure 5**. This figure shows that when the PGSs are perfect measures of the genetic variance, i.e. tags the entirety of trait genetic variance, the estimated coefficients are as expected, ~0. In all other scenarios, extended model estimates for path *b* are always negative and estimates for *f* are always positive. These effects are exacerbated as trait heritability increases and variance captured by PGSs

decreases. Results from regression models, and maternal genetic nurturing paths show similar β coefficient estimate patterns (Appendix I Figure 1).

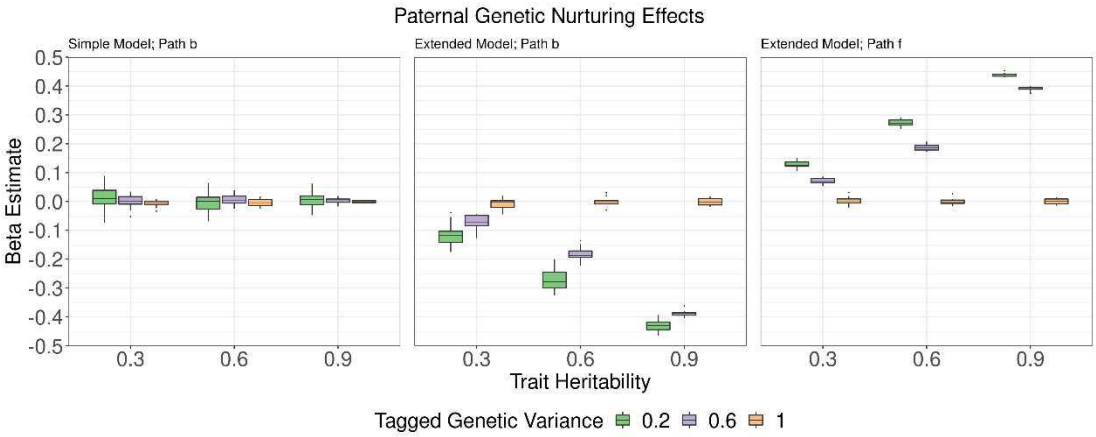


Figure 6. Paternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of simulated trait heritability; y- axis presents the beta coefficients of genetic nurturing effects. The left and middle plots presents boxplots representing β coefficient estimates of path *b*; the association between offspring phenotype and paternal PGS from the simple and extended pathway models respectively. The right plot presents boxplots representing β coefficient estimates of path *f* from the extended pathway model; the association between offspring phenotype and paternal phenotype.

2.4.3. Simulation II Results

Figure 7 presents β coefficient boxplots of paternal genetic nurturing paths highlighted in **Figure 5**. Here, an additional level of complexity is simulated; noise associated with the PGSs that arise from inaccurate effect sizes of SNPs obtained from GWAS summary statistics. This figure shows that when the PGSs are perfect measures of the genetic variance, i.e. tags the entirety of trait genetic variance, and noise is essentially 0, the estimated coefficients are as expected, ~0. In all other scenarios, extended model estimates for path *b* are always negative and estimates for *f* are always positive. These effects are exacerbated as noise associated with the PGS *decreases* and trait heritability *increases*. Results from regression models, and maternal genetic nurturing paths show similar β coefficient estimate patterns (Appendix I Figure 2).

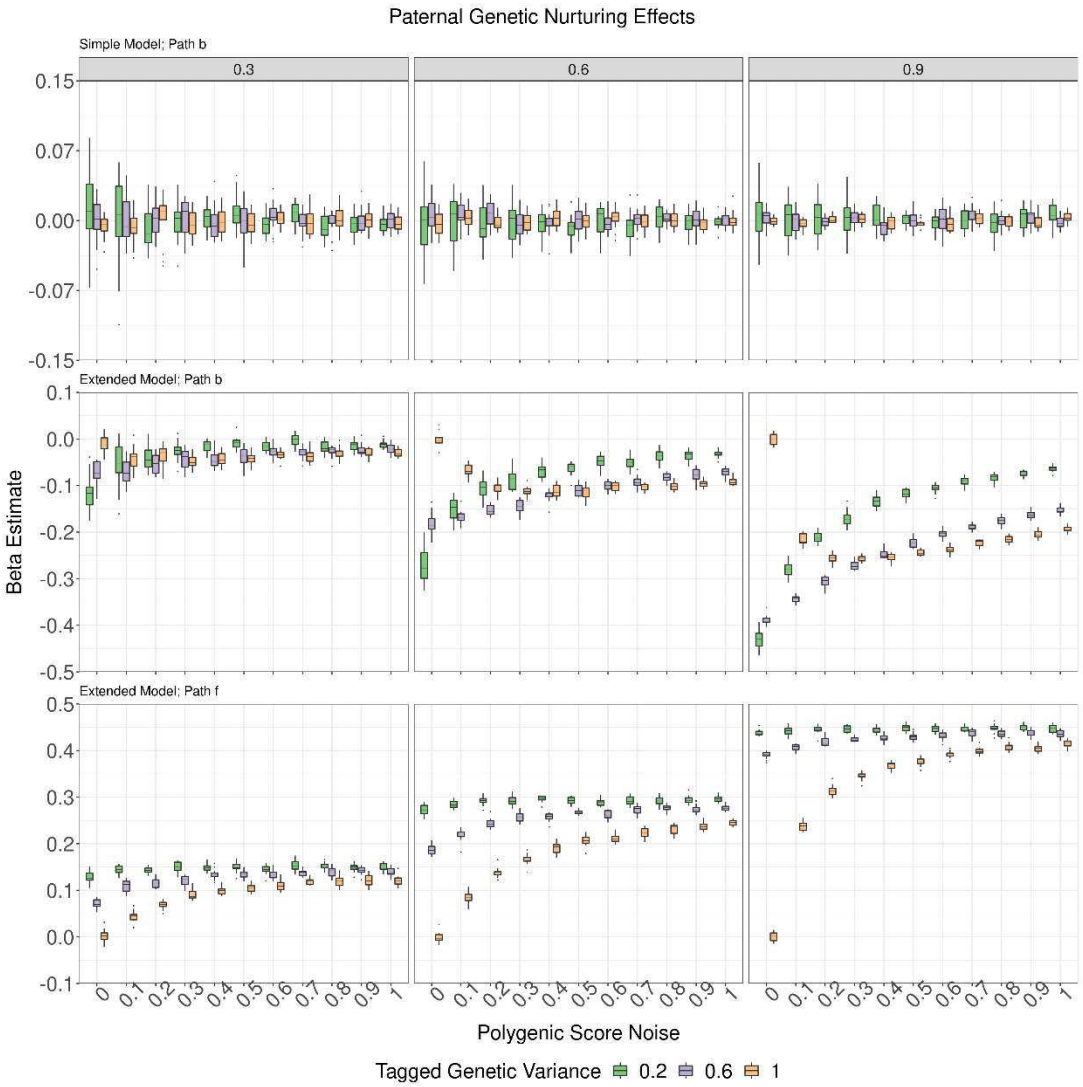


Figure 7. Paternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of noise associated with PGSs, y- axis presents the beta coefficients of genetic nurturing effects, facets present varying levels of trait heritability. The top and middle plots presents boxplots representing β coefficient estimates of path *b*; the association between offspring phenotype and paternal PGS from the simple and extended pathway models respectively. The bottom plot presents boxplots representing β coefficient estimates of path *f* from the extended pathway model; the association between offspring phenotype and paternal phenotype.

2.5. Discussion

The present study utilises polygenic scores (PGSs) informed by larger international GWAS consortia and 2680 Generation Scotland trios. Regression and pathway models using available GS data found contradicting results. Models with only trio

PGSs (regression models *III* and simple pathway models) showed evidence of genetic nurturing effects for educational attainment only. However, the addition of parental phenotypes into models (regression models *IV* and extended pathway models) showed parental phenotype mediated genetic nurturing effects for *all* phenotypes explored. Moreover, highly significant and negative β estimates were observed for both parental PGSs in models (paths *b* and *c* of the extended pathway models) exploring height, suggesting residual genetic nurturing effects.

Analyses using simulated data showed that utilising PGSs, that did not capture the entirety of the genetic variance, resulted in multiple confounded effects when models included both parental PGSs and parental phenotypes as predictors to explore mechanisms of genetic nurturing. Parental phenotypes were always positively associated with the offspring phenotypes (paths *f* and *g*), despite expectations of null associations (as genetic nurturing effects were *not* simulated). This is not counter-intuitive, as significant associations may arise between parental phenotypes and offspring phenotypes due to shared genetic variance that is not adequately accounted for by the PGSs within the models.

Interestingly, results showed that, parental PGS association estimates (representing residual genetic nurturing effects) were always negative, again despite expectations of null associations. Results further highlight that these upward and downward biases were exacerbated as PGSs became poorer measures of genetic variance and trait heritability increased. These biases were only eliminated when PGSs were simulated to capture the entirety of trait genetic variance. Whilst the direction of some confounded effects were surprising, the pattern shown is in line with what has been presented within the literature (Akimova et al, 2020).

Analyses using simulated data with the added complexity of PGS noise components, additionally showed that the biased effects were further exacerbated as noise *decreased*. These results likely reflect that, as PGSs become noisier variables, real signal and thus, associations between variables will diminish. Hence, the biased effects were more apparent when associations were detected using PGSs with

greater accuracy. In contrast, results of simpler models, excluding parental phenotypes, showed the expected null β coefficients for associations representing genetic nurturing effects. These effects were consistent across the range of simulated trait heritability, and PGS accounted variance.

Results using simulated data highlight the confounded effects in the models including parental phenotypes. Confounded effects can explain the discrepancy observed between simple and extended model results using GS data. In fact, GS height extended model results were highly comparable to the pattern of results using simulated data, which may point to the lack of/negligible genetic nurturing effects as suggested by the GS height simple model results. In contrast, results from analyses using GS psychological distress, depression and educational attainment data being dissimilar to that observed from analyses using simulated data, may suggest genuine genetic nurturing effects at play.

Generation Scotland results showed some counter-intuitive findings, such as the negative paternal genetic nurturing effects observed for depression and psychological distress. These could also represent spurious effects due to collinearity, overfitting or lack of statistical power. The depression PGSs are considerably poor predictors of psychological distress and depression, and this is more apparent in the paternal samples (Appendix I Table 2.2).The extended pathway models using GS data, also included discrepancies observed in the association effect sizes between respective trio PGS and phenotype associations. Maternal PGS-phenotype associations were much larger than that observed for fathers when exploring psychological distress and depression, suggesting potential sex interaction effects, which may require further investigation. Similarly, parental PGS-phenotype associations were larger than observed offspring PGS-phenotype associations for educational attainment. This may be attributable to the fact that genetic nurturing effects play a significant role in educational attainment, and the absence of grandparent PGSs within the models may result in parental PGS-phenotype associations encompassing additional genetic nurturing effects. These effects were accounted for within offspring PGS-phenotype associations, as parent

data is included in the models. Whilst these findings highlight future research directions and effects to consider within the models, the results using simulated data clearly show that the extended models are confounded. It is evident that without essentially perfect PGSs, accurate quantification of genetic nurturing effects from extended models is not possible due to confounded effects.

Overall, results suggest that the regression and simple pathway models exploring offspring PGS (direct genetic effects) and additional parental PGS (genetic nurturing effects) associations with offspring phenotypes, pose a straightforward and unbiased method to explore genetic nurturing effects.

Log-likelihood comparisons of regression models (*I* and *II*) and simple pathway model results suggested evidence of parental genetic nurturing effects for educational attainment, but not other explored phenotyped. The educational attainment regression model with all three trio member's PGS accounted for greater variance and showed a significantly better fit in comparison to the model utilising only the offspring PGS. This supports much of the findings from the literature suggesting genetic nurturing effects are detectable using educational attainment PGSs (Bates et al, 2018; Kong et al, 2018).

Some simple pathway model results were surprising, such as the significant *negative* paternal genetic nurturing effect (path *b*) on offspring depression. Whilst these findings were unexpected, some explanations could pave way for these results, such as, depression risk or experience resulting in greater awareness within parents, thus, resulting in behaviours that promote offspring well-being. Alternatively, the presence of depression or depression risk could be correlated with paternal absence and/or maternal compensation, which may limit negative influence of paternal genetic risk on the offspring's rearing environment. These speculations would need to be empirically explored for clearer conclusions. Another surprising result was the significant maternal genetic nurturing effect (path *c*) on offspring height. Although this effect was small and only just statistically significant, potential explanations include; maternal genetic effects on breast milk nutrition or

diet style which in turn can have an impact on offspring malnutrition levels, and consequently, offspring height. Again, these speculations would require further empirical evidence for clearer conclusions.

Future work can include the adaptation of these models to utilise data on different traits between parents and offspring e.g. parent educational attainment PGS and offspring depression PGS to explore genetic nurturing effects of parental educational attainment on offspring depression. It is important to keep in mind differences in GWAS power, and thus, discrepant predictive validity of these PGSs may result in further biases. Ideally, additional simulation analyses should be conducted to explore the impact of differences in parent and offspring PGS power on results.

As with all research designs, some limitations should be considered when interpreting the results presented. Here trio data and noise associated with inaccurate GWAS SNP effect sizes were simulated using random sampling from pre-defined variances, as opposed to simulating individual genotypes. This has resulted in a simplistic representation of the effects shared between biological parent-offspring trios. In reality, genetic nurturing effects may be at play via both parental transmitted and non-transmitted alleles, which is a level of complexity not modelled here. Thus, potential differences in how biases may arise within these specific paths were not observed. Simulating GWAS, and individual genotypes would have provided a more accurate representation of how variables are formed and the impact of inaccuracies of SNP effect sizes, in real life.

Another element not formally explored is the impact of assortative mating within these models. In fact, evidence of assortative mating has been observed for many complex behavioural traits, including phenotypes explored within this study (Hugh-Jones et al, 2016; Mathews & Reus, 2001; Stulp et al, 2017). Height, in particular, can be thought of as a useful control variable as the literature shows strong evidence of assortative mating (Stulp et al, 2017), yet limited evidence of genetic nurturing effects. Results of GS height analyses show no strong evidence of genetic

nurturing, which suggests that any bias arising from assortative mating effects is likely to be very small.

Moreover, it is important to note that shared environmental effects may also drive significant associations and confounding effects within the extended models. Identifying and modelling these effects would also require extensive research.

This study demonstrates that pathway models are a simple and useful method to explore and quantify genetic nurturing effects using offspring phenotypic and trio PGS data. However, it is important to be aware of biases which arise when including parental phenotype mediated genetic nurturing paths. The improvement of PGSs will likely fall short for the level required for clear interpretation of these models. These results suggest alternative methods should be utilised when exploring factors through which genetic nurturing effects may be at play.

Chapter 3: Genome-by-Trauma Interaction Effects on Depression

3.1. Abstract

3.1.1. Background

Self-reported trauma exposure has consistently been found to be a risk factor for Major Depressive Disorder (MDD) and several studies have reported interactions with genetic liability. To date, most studies have examined interaction effects with trauma exposure using genome-wide variants or polygenic scores, typically capturing less than 3% of phenotypic risk variance. We sought to re-examine genome-by-trauma interaction effects using genetic measures utilising all available genotyped data and thus, maximising accounted variance.

3.1.2. Methods

Measures of self-reported depression, neuroticism and trauma exposure for 148,129 participants with genome-wide SNP data are available from the UK Biobank study. Here, we used a mixed-model statistical approach utilising genetic, trauma exposure and genome-by-trauma exposure interaction similarity matrices to explore sources of variation in depression and neuroticism.

3.1.3. Findings

Our approach estimated the heritability of MDD to be approximately 0.160 [SE 0.016]. Subtypes of self-reported trauma exposure (catastrophic, adult, childhood and full trauma) accounted for a significant proportion of trait variance ranging from 0.056 [SE 0.013] to 0.176 [SE 0.025]. The proportion of MDD risk variance accounted for by significant genome-by-trauma interaction ranged from 0.074 [SE 0.006] to 0.201 [SE 0.009]. Results from sex-specific analyses found genome-by-trauma interaction variance estimates approximately 5-fold greater for MDD in males than in females.

3.1.4. Discussion

This is the first study to utilise all genome-wide SNP data when exploring genome-by-trauma interaction effects in MDD and presents evidence that interaction effects are influential in depression manifestation. This effect accounts for greater trait variance within males which points to potential differences in depression aetiology between the sexes. The methodology utilised within this study can be extrapolated to other environmental factors to identify modifiable risk environments and at-risk groups to target with interventions.

3.2. Introduction

Depression is a highly prevalent psychiatric disorder with a lifetime risk of ~16% (Kessler et al, 2009; Kessler et al, 2005) and is a leading cause of disability worldwide (Vos et al, 2012). Twin studies have shown moderate heritability estimates ranging between 30-40% (Polderman et al, 2015) suggesting both genetic and environmental factors are influential. In fact, genome-wide association studies (GWAS) have uncovered many replicated single nucleotide polymorphisms (SNPs) associated with depression, however, SNP-based heritability estimates of 5-10% (Howard et al, 2019; Wray et al, 2018a) are much lower than estimates obtained from twin studies which has highlighted the issue of missing heritability. Explanations for missing heritability include inflation of heritability estimates attributable to shared environmental effects as well as gene-environment interplay (Eichler et al, 2010; Falconer & Mackay, 1996; Manolio et al, 2009; Visscher et al, 2008).

Some environmental factors such as self-reported trauma exposure have been found to play a role in the aetiology of depression; with case-control studies suggesting individuals diagnosed with Major Depressive Disorder (MDD) report higher levels of trauma exposure (Moskvina et al, 2007; Nanni et al, 2012; Negele et al, 2015). In turn, trauma exposure in childhood has been consistently associated with adverse outcomes including increased risk of MDD in adolescence and adulthood (Hopfinger et al, 2016; Huh et al, 2017). These findings suggest a potentially causal role for trauma in depression's aetiology. Additionally, research has shown self-reported trauma exposure and MDD to be genetically correlated,

suggesting shared genetic risk factors for both (Coleman et al, 2020), or potentially more complex interplay effects of genetics and trauma exposure on depression manifestation.

One key form of interplay that has been explored is the interaction effect of genetics and trauma exposure on depression manifestation. Gene-by-environment interaction effects refer to the differential effects of an environmental exposure on traits in individuals with differing genotypes. This can be conceptualised as an individual's genetic sensitivity to certain environments, which may result in an exacerbated risk of a disorder (Mullins et al, 2016; Uher, 2014). Initial studies explored candidate genes such as the serotonin transporter gene (5-HTT) and found a significant short allele putative interaction with childhood trauma and its effect on depression (Caspi et al, 2003). However, the validity of candidate gene study findings has been brought into question by lack of replication in further studies (Culverhouse et al, 2018; Fergusson et al, 2011) as well as an improved understanding of the polygenic nature of depression and similar traits (Flint & Kendler, 2014; Levinson et al, 2014).

This has resulted in gene-by-trauma interaction research utilising genome-wide variants and polygenic scores (PGSs); genetic measures that can be calculated for each individual by identifying, weighting and summing genotyped risk variants found to be associated with depression (Demirkan et al, 2011; Dudbridge, 2013). However, the few studies that have explored variant-by-trauma interaction effects have yielded minimal findings (Arnau-Soler et al, 2018; Arnau-Soler et al, 2019b; Coleman et al, 2020; Werme et al, 2021). Moreover, studies of PGS-by-trauma as well as stressful life event (SLE) interactions within depression have yielded inconsistent findings. Whilst some studies have highlighted sex differences and found significant interaction effects associated with MDD outcomes (Colodro-Conde et al, 2018; Mullins et al, 2016; Peyrot et al, 2014), some replication attempts reported null findings (Gillett et al, 2021; Shao et al, 2021) and follow up meta-analyses have suggested that reported findings were likely to be false positives (Peyrot et al, 2018).

An explanation for the inconsistent findings may lie in the predictive accuracy and validity of PGSs (Eichler et al, 2010). PGSs build upon the information provided by GWAS, which still have limited statistical power for detecting trait associated genetic variants and their effect sizes (Dudbridge, 2013; Mostafavi et al, 2020; Nicholls et al, 2020). Power limitations of GWAS, and thus PGSs, are greater for traits that also have a substantial environmental component, such as depression, as opposed to traits with higher genetic aetiology (Mullins & Lewis, 2017). This is reflected in the fact that current PGSs capture less than 3% (Howard et al, 2019; Wray et al, 2018a) of the phenotypic variance proposed for depression from both twin studies and SNP-based heritability estimates.

More importantly, although research exploring the genetic aetiology of trauma exposure is still in its infancy, it is becoming increasing clear that self-reported trauma is a heritable environment (Warrier et al, 2021), with both SNP-based and twin study heritability estimates ranging between ~20-50% (Jang et al, 2001; Lyons et al, 1993; Stein et al, 2002). Furthermore, evidence suggests that depression and trauma exposure are genetically correlated (Coleman et al, 2020) which highlights a lack of independence between genetic and seemingly 'environmental' variables; posing a serious challenge in the statistical modelling and interpretation of these PGS-by-trauma interaction effects (Akimova et al, 2020).

One way to circumvent the issues associated with PGSs outlined above, is to make use of genetic measures that are capturing greater variance. Genomic relationship matrices (GRMs) have been able to do this by utilising all genotyped SNPs (Yang et al, 2011). Including matrices representing genetic, environmental and interaction effect similarity within a sample, in mixed linear models, can provide estimates of the genetic, environmental and genome-by-environment interaction components of trait variance (Amador et al, 2021).

Here, we estimated the contribution of trauma exposure and its interaction with genetic variation to depression as well as neuroticism. We chose to explore neuroticism as this trait has been found to have a greater genetic component

(Docherty et al, 2016; Vukasovic & Bratko, 2015) and has a strong link with MDD suggesting the exploration of neuroticism to be useful in understanding the genetic aetiology of both of these traits (Kendler et al, 2006; Luciano et al, 2018). As the existing literature has highlighted the role of trauma exposure in sex differences observed in MDD, we also explored these effects in males and females separately (Kendler & Gardner, 2014).

Here we show that utilising the entirety of genotyped genetic variants can improve statistical power in the exploration of genome-by-trauma interaction effects. More importantly, our findings suggest that genome-by-trauma interaction effects may play a much larger role in depression manifestation than previously thought.

3.3. Methodology

3.3.1. Participants

Participants were identified from the UK Biobank (UKB), a national study exploring genetic and environmental determinants of health using individuals recruited from 22 different centres across the United Kingdom (Allen et al, 2014; Sudlow et al, 2015). A follow up Mental Health Questionnaire (MHQ) was administered assessing common mental health disorders including trauma experience (Davis & Hotopf, 2019). In this study we had access to individual level data from 148,129 participants who completed the MHQ with available genetic, trauma experience, depressive symptoms and/or neuroticism information. All individuals included were from European ancestries.

The UKB study received ethical approval from the NHS National Research Ethics Service (reference: 11/NW/0382) and all participants provided written informed consent. This study has been approved by the UKB Access Committee (Project #4844).

3.3.2. Genotypes

Two separate array chips with 95% SNP overlap, were used to genotype UKB participants; the Applied Biosystems™ UK BiLEVE Axiom™ Array by Affymetrix1

captures 807 411 SNPs and the Applied Biosystems™ UK Biobank Axiom™ Array captures 825,927 SNPs (Bycroft et al, 2018). Quality control of genotyped SNPs consisted of the exclusion of SNPs with missingness > 2% and a Hardy-Weinberg Equilibrium test $p < 10^{-6}$. SNPs with minor allele frequency (MAF) < 0.05 and individuals with > 2% missing genotypes were excluded from analyses. A total of 414,584 common SNPs across the 22 autosomes were included in analyses.

3.3.3. Phenotypes

All phenotypes were defined using retrospective self-reported responses to questions assessing help-seeking behaviour, depressive symptoms and trauma experience. Field codes and more information on classifications are available in (Appendix II Table 1).

3.3.3.1. Depression & Neuroticism Phenotypes

Broad depression was defined using self-reported help seeking behaviour for mental health difficulties. Case and control status was determined from the response to Touchscreen Questionnaire questions administered during initial recruitment; 'Have you ever seen a general practitioner (GP) for nerves, anxiety, tension or depression?' or 'Have you ever seen a psychiatrist for nerves, anxiety, tension or depression? Individuals responding 'Yes' to either and 'No' to both questions were classified as cases and controls, respectively.

CIDI depression was defined using questions from the Composite International Diagnostic Interview Short Form (CIDI) (Kessler & Ustun, 2004). These were administered within the online follow-up Mental Health Questionnaire (MHQ), directly assessing the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)(American Psychiatric Association, 2013) symptomatic and functional criteria for MDD, and are associated with greater MDD specificity (Cai et al, 2020b).

Neuroticism was measured using the 12-item Neuroticism scale of the revised short form Eysenck Personality Questionnaire (EPQ-R) (Eysenck et al, 1985). Response

options were ‘Yes’, ‘No’, ‘Do not know’, ‘Prefer not to answer’. Participant summed scores of ‘Yes’ responses (ranging from 0-12) were used in downstream analyses.

3.3.3.2. Trauma Phenotypes

Participants were administered a 16-item questionnaire relating to traumatic experiences as a part of the MHQ. Five items explored childhood trauma using the Childhood Trauma Screener (Glaesmer et al, 2013; Grabe et al, 2012); five items explored adult trauma using an equivalent screener developed by the UKB Mental Health steering group (Davis et al, 2020); and the remaining six questions explored catastrophic trauma using questions related to events that often trigger post-traumatic stress disorder (Coleman et al, 2020). More information on questions and responses can be found in the (Appendix II Table 2). **Table 1** shows samples sizes of individuals included in statistical analyses with complete trauma exposure, depression and neuroticism information.

Table 1. Sample sizes of UKB participants with complete trauma, depression and neuroticism information.

	CIDI Depression		Broad Depression		Neuroticism
Trauma	**Case**	**Control**	**Case**	**Control**	
Full	32,015	81,809	45,093	82,370	11,6995
Childhood	33,532	85,405	47,583	85,894	12,1670
Adult	33,262	83,722	46,886	84,234	11,9914
Catastrophic	33,657	85,704	47,850	86,127	12,2108

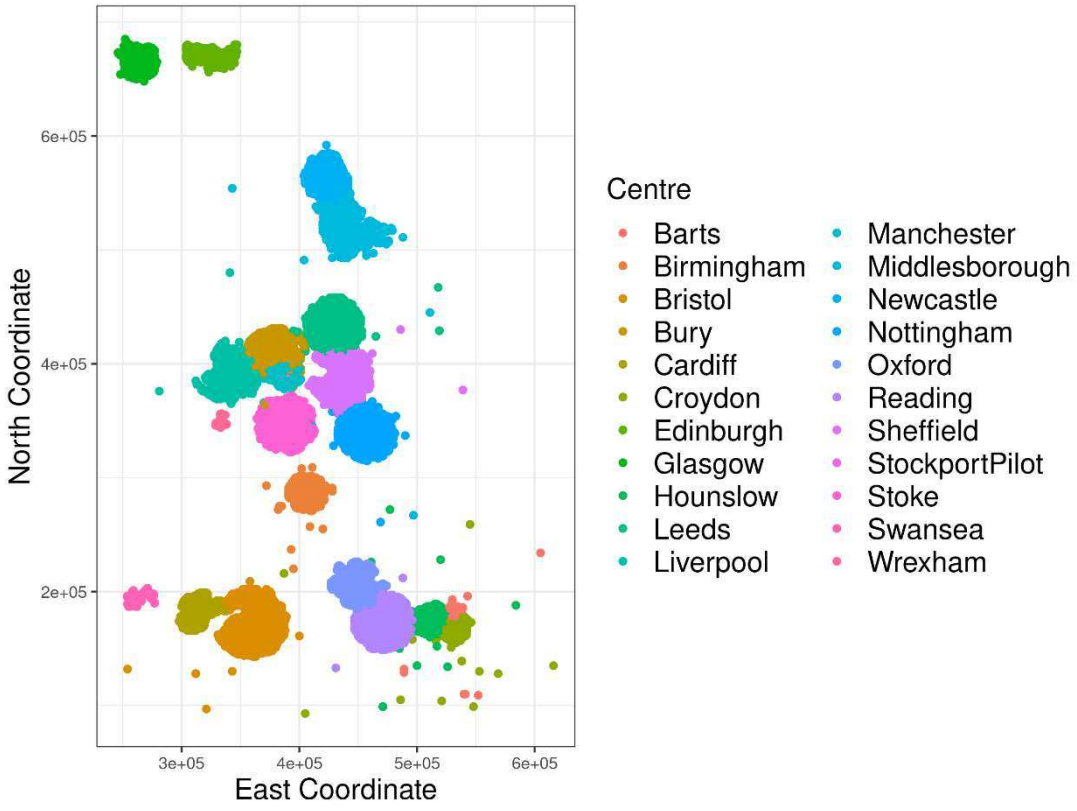
Abbreviations. CIDI, Composite International Diagnostic Inventory.

Individual regression analyses exploring associations between each trauma item and depression/neuroticism outcomes with age and sex as covariates (Appendix 2 Table 2), suggested most trauma items to be significantly associated with three phenotypes. Hence, all items were utilised in the formation of trauma phenotypes. Appropriately weighting and combining the trauma items to form a trauma variable is challenging. In order to capture separate and independent dimensions of trauma, principal components (PCs) of complete responses to all trauma items (full trauma) as well as the three sub-categories of trauma items (childhood, adult, catastrophic

trauma) were obtained. Trauma items were re-scaled to have a mean of one and unit variance before PCs were extracted using the 'prcomp' function in R 4.0.2 (Team, 2020). For more information on depression/neuroticism associations with trauma PCs and PC loadings see Appendix II Table 3-4.

3.3.4. Geographical Clusters

The number of participants with available depression/neuroticism and trauma information is large (~150k), and thus, analysing the pairwise covariance matrices jointly for the whole cohort was computationally intractable. To work around these computational issues, participants were split into five different clusters based on the geographical location of their recruitment centres (north, midnorth, midsouth, southwest and southeast regions in the UK) (**Figure 2**) and all subsequent analyses were replicated across the five clusters; for more information on cluster demographic characteristics see (Appendix II Table 5-6).



Clusters	Centre	Sample Size

North	Glasgow	4503
	Edinburgh	5592
	Newcastle	9771
	Middlesborough	6214
Mid North	Leeds	12332
	Bury	6915
	Manchester	3982
	Liverpool	8969
Mid South	Wrexham	151
	Stoke	4304
	StockportPilot	294
	Sheffield	9290
	Nottingham	10193
	Birmingham	6945
South West	Swansea	660
	Cardiff	4731
	Bristol	14642
South East	Oxford	4897
	Reading	10904
	Hounslow	9542
	Croydon	8860
	Barts	4438

***Figure 2*.** UK Biobank Geographical Clusters. Above, figure shows the North and East co-ordinates of participant residence and colours represent the recruitment centres for each participant. Below, table shows geographical clusters, respective centres and sample sizes.

3.3.5. Covariance Matrices

Covariance matrices were used in downstream analyses to explore genetic, environmental and genome-by-environment components of variation.

Genomic Relationship Matrices (G): represents genetic similarity between individuals. I computed one G for all individuals in each of the geographical clusters using genetic variants that passed quality control (see section 3.3.2 Genotypes) using GCTA v1.91.4beta (Yang et al, 2011). Supplementary Gs were computed using *only* unrelated individuals (individuals with genomic similarity values <0.05).

Trauma Environmental Relationship Matrices (E): represents similarity between individuals based on their trauma eigenvectors (PCs).

These matrices were computed using OSCA v0.45 (Zhang et al, 2019). OSCA software was developed for the analysis of omics/methylation data and provides 3 separate algorithms which can be used when computing the matrices. For more information on the algorithms see Appendix II Section A. Zhang and colleagues (2019) highlighted, with simulation analyses, that the algorithms used had negligible impact on the outputted estimates. As the trauma eigenvectors (PCs) utilised here are not intrinsically omics/methylation data, separate analyses were conducted using the 3 separate algorithms to explore the impact this would have on results. Similar to what was suggested by Zhang et al, mixed linear model findings showed negligible differences in outputted estimates for proportion of CIDI depression variance attributable to Es computed using the different algorithms (Appendix II Table 7). Matrices used for further downstream analyses were computed using the default algorithm (algorithm 1).

Separate Es were computed using full trauma, childhood, adult and catastrophic trauma eigenvectors. To control for genetic covariance between trauma exposure and the phenotypes of interest (depression/neuroticism), an E was computed using full trauma eigenvectors pre-corrected for the full sample G.

Additional Es were computed using the first eigenvector (PC1), as this accounted for the greatest variance in our phenotypes of interest, and separately, the second and third eigenvector of full trauma to explore how results differed. Subsequent PCs of full trauma were not separately explored as the variance explained by these PCs was minimal (Appendix II Table 4).

Gene-Environment Interaction Matrices (GxE): represent shared genome-by-trauma interaction effects. These were computed by multiplying Gs and Es using a cell-by-cell multiplication, also known as the hadamard product (Amador et al, 2021; Heckerman et al, 2016; Nivard et al, 2016). To obtain the GxE interaction matrices, information on similarity between individuals at both the genetic and environmental (trauma) level is required.

3.3.6. Analyses

3.6.1. SNP Heritability & Genetic Correlations

Using the first eigenvector (PC1) for the subtypes of trauma, the SNP heritability of each trauma variable was explored. Heritability estimates of the trauma variables were obtained by fitting the trauma variables as the dependent variable and Gs as random effects within a mixed linear model framework (resulting in estimates of trauma variance attributable to G), see model 1 below. The genetic correlations between trauma and depression variables were explored. Bivariate genetic correlations were explored using the moment-based method, Haseman-Elston regression analyses (which were computationally less exhaustive), where standard errors were calculated using a leave-one-individual-out jackknife technique. Age, sex, genotyping array and the first 15 principal components of the GRM were included as covariates.

3.6.2. Variance Components Analyses

Variance components of depression/neuroticism were explored within mixed linear model frameworks. Here, we incorporate the computed matrices, and thus, trauma as a random effect. Trauma exposure has been predominantly explored as a fixed effect within the literature (Coleman et al, 2020; Gillett et al, 2021; Peyrot et al, 2014; Peyrot et al, 2018). Here, trauma exposure is incorporated into models as a random effect. Four models were explored with varying levels of complexity:

1. $y = X\beta + g + \varepsilon$
2. $y = X\beta + e + \varepsilon$
3. $y = X\beta + g + e + \varepsilon$
4. $y = X\beta + g + e + g \times e + \varepsilon$

Where y is a $n \times 1$ vector of observed depression/neuroticism phenotypes; β is a vector of fixed effects (which include age, sex, genotyping array and the first 15 principal components of the full sample G) and X is its design matrix; g is a $n \times 1$ vector of SNP effects (representing additive genetic effects) with $g \sim (0, G\sigma_G^2)$; e is a $n \times 1$ vector representing common environmental effects of childhood, adult, catastrophic or all trauma with $e \sim (0, E\sigma_E^2)$; $g \times e$ is a $n \times 1$ vector representing

interactions between genetic and trauma effects with $g \times e \sim (0,\ G \circ E\sigma^2_{G \circ E})$; and ε is a $n \times 1$ vector of residual effects.

Estimates of variance attributable to the G, E, and GxE components are obtained from analyses using CIDI and Broad depression as dependent variables and are converted to the liability scale within GCTA (Yang et al, 2011). As UKB is a population-based study, we utilised the prevalence rates observed within the whole sample. Prevalence rates used were 0.28 and 0.35 for joint sex analyses; 0.35 and 0.43 for female analyses; 0.19 and 0.27 for male analyses in each case for CIDI and broad depression, respectively.

Analyses were repeated using only unrelated individuals. All analyses were replicated across the five geographical cluster samples using GCTA v1.91.4beta and results (estimates of variance components) were meta-analysed using R package `metafor`.

https://github.com/melisachuong/GxE_Depression

3.4. Results

Results presented are meta-analysed estimates from analyses conducted within each geographical cluster, specified above. Cluster specific results are relatively homogenous and can be found in Appendix II tables.

Initial analyses explored the SNP heritability of full trauma and sub-categories of trauma. Trauma variables meta-analysed SNP heritability estimates were; full trauma 0.17 [SE 0.008], childhood trauma 0.15 [SE 0.008], adult trauma 0.063 [SE 0.008] and catastrophic trauma 0.11 [SE 0.008]. Results from each cluster can be found in Appendix II Table 8. Similar results were obtained when analyses were conducted using only unrelated individuals (Appendix II Table 8). All genetic correlations between the trauma variables and depression/neuroticism phenotypes are shown in Table 2. Genetic correlations between trauma variables and broad depression and neuroticism phenotypes were modest; in contrast, stronger genetic correlations between trauma variables and the CIDI depression phenotype were

observed (**Table 2**). It is important reiterate here, that the CIDI depression phenotype questionnaire did not overlap with trauma exposure questionnaires. Results from each cluster can be found in the Appendix II Table 9.

Table 2. Trauma and Depression/Neuroticism Genetic Correlations

	CIDI Depression		Broad Depression		Neuroticism	
Trauma	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
Full	0.632	0.085	0.390	0.082	0.333	0.064
Childhood	0.605	0.091	0.337	0.090	0.332	0.071
Adult	0.647	0.134	0.358	0.132	0.274	0.105
Catastrophic	0.536	0.104	0.310	0.098	0.204	0.081

Abbreviations: SE, Standard Error; CIDI, Composite International Diagnostic Interview.

Figure 3 shows the meta-analysed estimates for the proportion of phenotypic variance explained by the different sources included in the mixed linear models. All estimates for proportion of variance explained by all components were statistically significant. Log-likelihood ratio tests (LRTs) suggested that the inclusion of trauma (E) and genome-by-trauma (GxE) interaction components improve model fit. Figure 3 shows differences in results when Es were computed using eigenvectors of full trauma, childhood, adult and catastrophic trauma separately.

Full results from these analyses, including estimates, standard errors (SEs), LRTs can be found in (Appendix II Tables 10-13). Results using full trauma Es with unrelated participants show similar results (Appendix II Table 14).

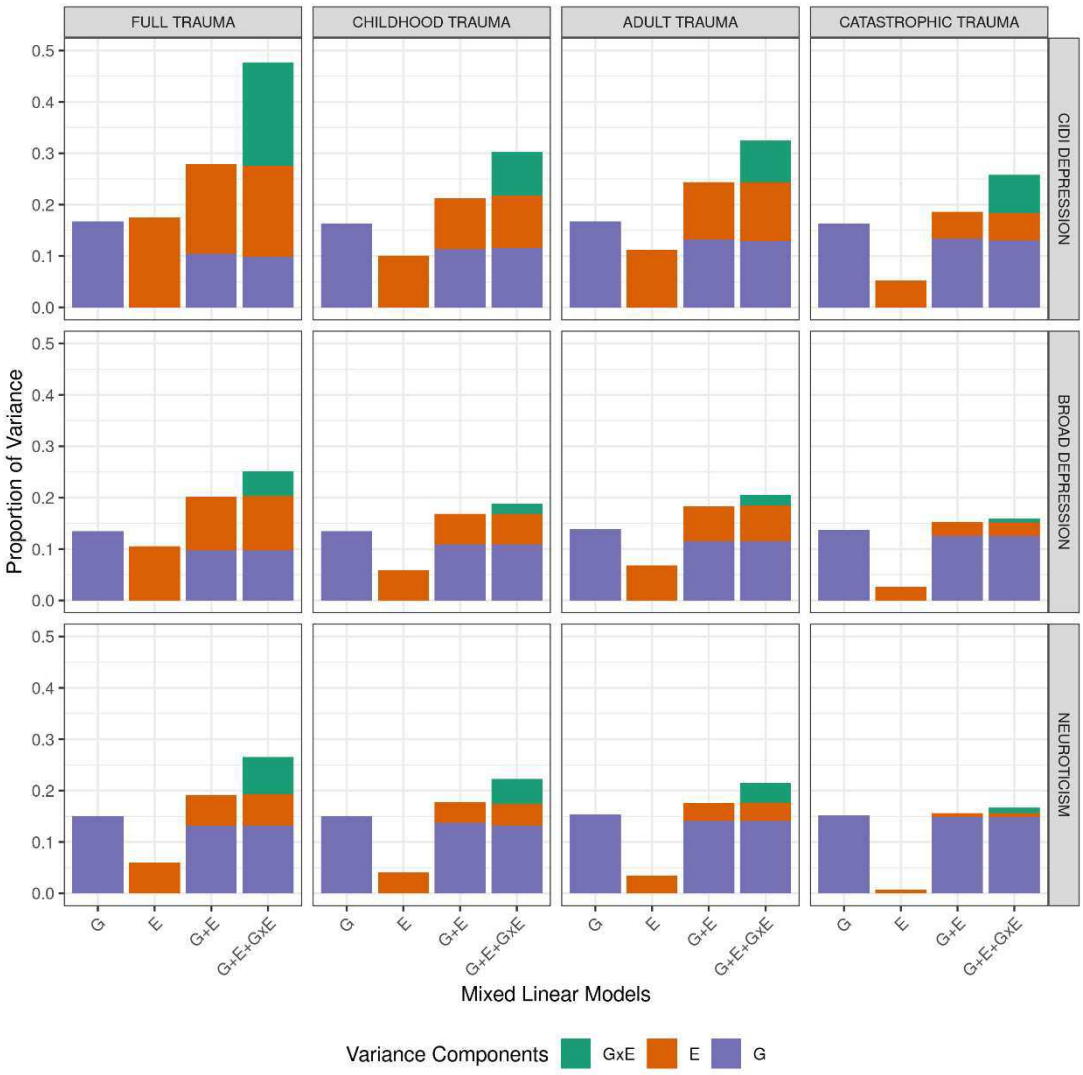


Figure 3. Proportion of Phenotypic Variance Explained by Genetic, Environmental and Interaction Sources in UK Biobank. Here, facets show results of mixed linear models including environmental relationship matrices computed using all principal components (PCs) of full trauma, childhood, adult and catastrophic trauma. The proportion of variance explained is specified on the y-axis. The sources of variation: G, Genomic; E, Environmental; GxE, Gene-by-Environment Interactions are specified with different colours and their corresponding models are specified on the x-axis.

Heritability estimates (i.e. proportion of variance accounted for by the G) of CIDI depression were stable across samples with complete trauma, childhood, adult and catastrophic trauma data separately, ~0.16 [SE 0.016]. The meta-analysed estimates for proportion of CIDI depression variance attributable to the Es were 0.18 [SE 0.025] = full trauma, 0.101 [SE 0.027] = childhood trauma, 0.113 [SE 0.03] = adult trauma, 0.05 [SE 0.013] = catastrophic trauma, respectively. Genome-by-trauma

interactions contributed significantly to the variance if CIDI depression. Estimates for proportion of CIDI depression variance attributable to the interaction effect were highest when exploring full trauma 0.201 [SE 0.009] and were 0.084 [SE 0.006], 0.081 [SE 0.005], 0.074 [SE 0.006] when respectively exploring childhood, adult and catastrophic trauma separately.

Significant, although, smaller meta-analysed heritability estimates were observed for broad depression ~0.135 [SE 0.014]. The meta-analysed estimates for proportion of broad depression variance attributable to the Es were 0.105 [SE 0.016], 0.059 [SE 0.016], 0.069 [SE 0.019], 0.027 [SE 0.007] when Es captured full trauma, childhood, adult and catastrophic trauma, respectively. Estimates for the proportion of broad depression variance attributable to the interaction effects are also smaller; 0.048 [SE 0.006]; 0.020 [SE 0.004], 0.021 [SE 0.004], 0.008 [SE 0.005] when respectively exploring full trauma, childhood, adult and catastrophic trauma separately.

Estimates for proportion of neuroticism variance attributable to the different sources included in the mixed linear models were also smaller than estimates obtained for CIDI depression. The meta-analysed heritability estimates observed were ~0.15 [SE 0.009]. The meta-analysed estimates for proportion of neuroticism variance attributable to the Es were 0.060 [SE 0.009], 0.041 [SE 0.011], 0.035 [SE 0.010], 0.007 [SE 0.002] when Es captured full trauma, childhood, adult and catastrophic trauma, respectively. Estimates for the proportion of neuroticism variance attributable to the interaction effects were 0.072 [SE 0.004]; 0.049 [SE 0.003], 0.038 [SE 0.003], 0.010 [SE 0.003] when respectively exploring full trauma, childhood, adult and catastrophic trauma separately.

Figure 4 shows the difference in estimates for proportion of phenotypic variance explained by the different sources included in the mixed linear models. The Es were computed using all eigenvectors (PCs) of full trauma, eigenvectors of full trauma that were pre-corrected for the full sample GRM (to control for genetic covariance between trauma exposure and the three phenotypes) and using only the first

eigenvector (PC1) of full trauma. Full results from these analyses, including estimates, SEs, LRTs can be found in (Appendix II Table 10, 15-16).

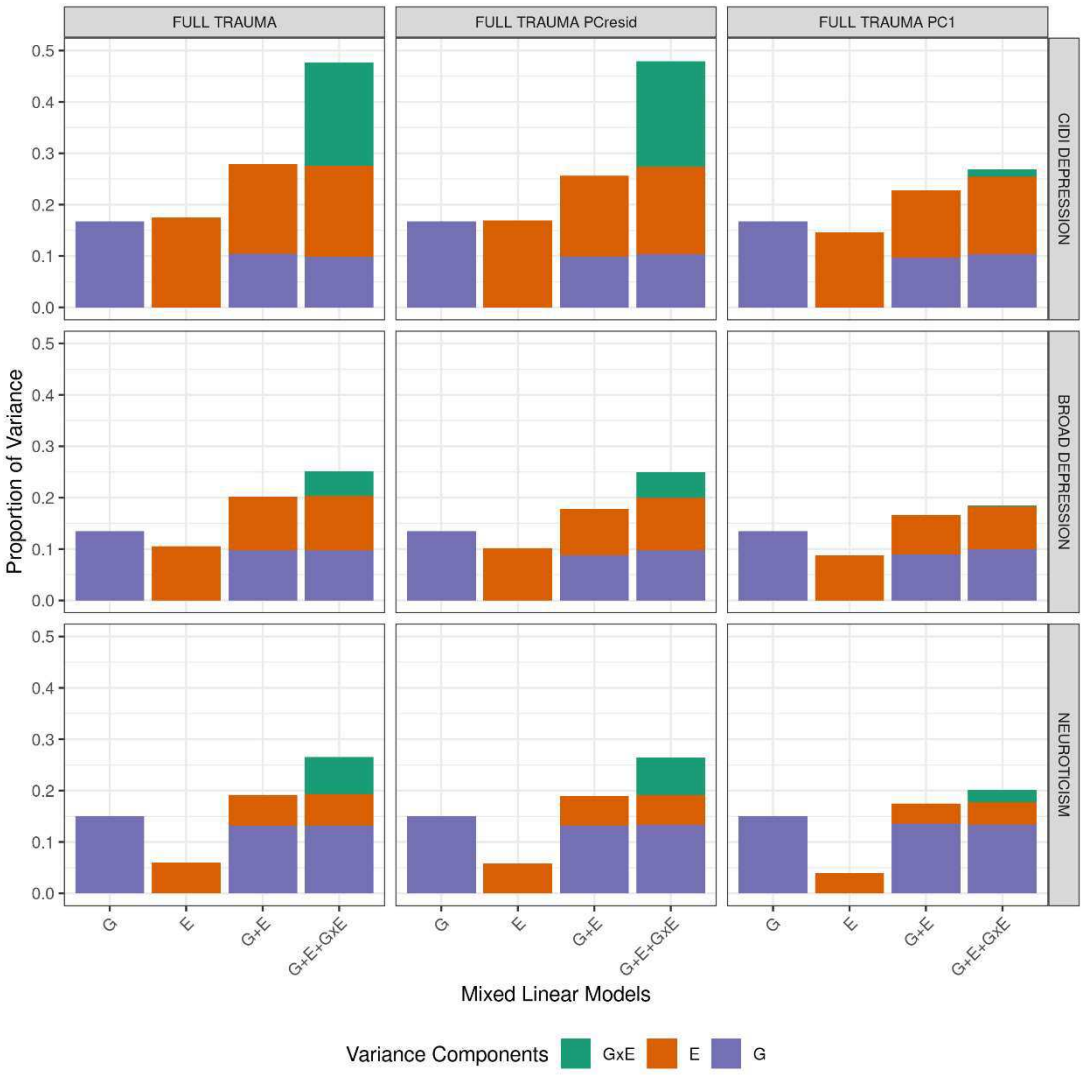


Figure 4. Proportion of Phenotypic Variance Explained by Genetic, Environmental and Interaction Sources in UK Biobank. CIDI, Composite International Diagnostic Inventory; PCresid, Full Trauma principal components pre-corrected for the full sample genomic relationship matrix (G). Here, the first column shows results of mixed linear models including an environmental relationship matrix (E) computed using all eigenvectors/principal components (PCs) of Full Trauma. The second column shows results of mixed linear models including an E computed using all PCs of Full Trauma that were pre-corrected for the full G. The final column shows results of mixed linear models including an E computed using only the first eigenvector (PC 1) of full trauma. The rows show results for each phenotype of interest; CIDI depression, broad depression and neuroticism. The proportion of variance explained is specified on the y-axis. The sources of variation: G, Genomic; E, Environmental; GxE, Gene-by-Environment Interactions are specified with different colours and their corresponding models are specified on the x-axis.

Meta-analysed estimates for proportion of phenotypic variance attributable to the E computed using all eigenvectors (PCs) of full trauma was very similar to the estimate of variance attributable to the E computed using eigenvectors of full trauma pre-corrected for the full sample G.

In contrast, estimates for the proportion of phenotypic variance attributable to the Es computed using the first eigenvector (PC1) of full trauma were smaller and statistically non-significant. Estimates were 0.146 [SE 0.085], 0.088 [SE 0.053], 0.039 [SE 0.024] for CIDI depression, broad depression and neuroticism respectively. Estimates of the interaction effects were substantially attenuated; 0.014 [SE 0.003], 0.002 [SE 0.002], 0.025 [SE 0.002] for CIDI depression, broad depression and neuroticism respectively, with the broad depression result being statistically non-significant.

Figure 5 shows the meta-analysed estimates for the proportion of phenotypic variance explained by the different sources included in the mixed linear models. The Es computed using all eigenvectors (PCs), the first (PC1), second (PC2) and third (PC3) eigenvector of full trauma separately. These analyses were conducted to explore which eigenvector of full trauma is involved in genome-by-trauma exposure interaction effects observed in depression/neuroticism phenotypes. Full results of these analyses, including estimates, SEs, LRTs can be found in (Appendix II Table 10, 16-18).

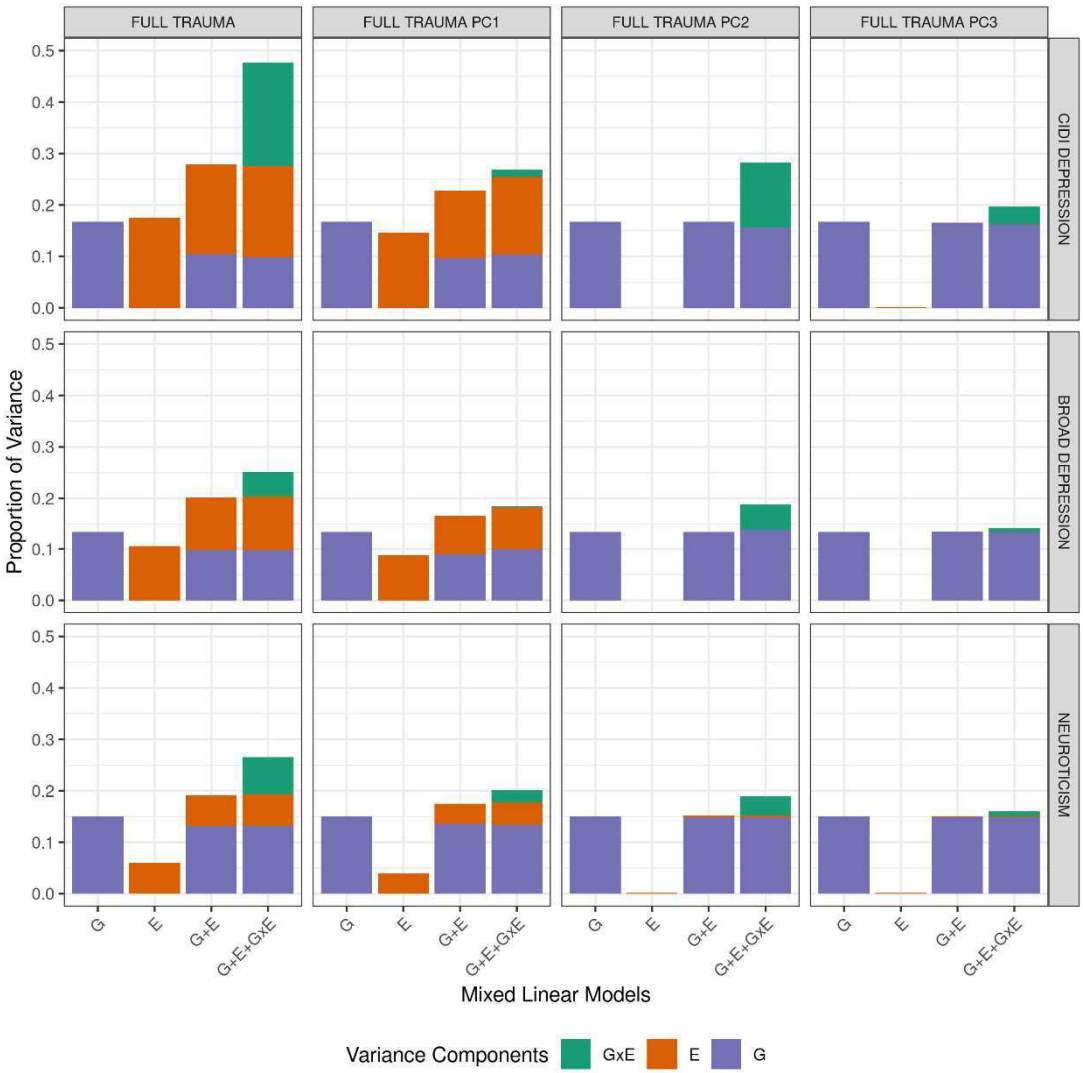


Figure 5. Proportion of Phenotypic Variance Explained by Genetic, Environmental and Interaction Sources in UK Biobank. Here, facets show results of mixed linear models including environmental relationship matrices computed using all principal components (PCs), PC 1, PC 2 and PC 3 of full trauma. The proportion of variance explained is specified on the y-axis. The sources of variation: G, Genomic; E, Environmental; GxE, Gene-by-Environment Interactions are specified with different colours and their corresponding models are specified on the x-axis.

Meta-analysed estimates for the proportion of phenotypic variance attributable to Es computed using the second and third (separately) eigenvector of full trauma were minuscule and statistically non-significant. In contrast, estimates for the proportion of variance attributable to the interaction effects were statistically significant. Estimates for the proportion of variance attributable to the interaction effect, where the E was computed using the second eigenvector of full trauma

(PC2), were 0.127 [SE 0.005], 0.048 [SE 0.004], 0.038 [SE 0.002] for CIDI depression, broad depression and neuroticism, respectively. The estimates for the proportion of variance attributable to the interaction effect, where E was computed using the third eigenvector of full trauma (PC3) were smaller, 0.034 [SE 0.004], 0.007 [SE 0.003], 0.010 [SE 0.002] for CIDI depression, broad depression and neuroticism respectively.

Table 3 shows the trauma exposure item loadings on to the first, second and third eigenvector (PC1-PC3) of full trauma. Loadings for all remaining eigenvectors including eigenvectors of childhood, adult and catastrophic trauma separately are available in (Appendix II Table 4).

Table 3. Full Trauma Principal Component 1:3 Trauma Item Loadings

Trauma	UKB Field	PC1	PC2	PC3
Childhood	f.20489.0.0	-0.34068	0.339934	-0.0855
	f.20488.0.0	-0.32443	0.256105	0.128531
	f.20487.0.0	-0.36301	0.239188	0.047123
	f.20490.0.0	-0.27858	-0.04304	0.136042
	f.20491.0.0	-0.22568	0.375387	-0.26115
Adult	f.20522.0.0	-0.16343	0.248746	-0.28518
	f.20523.0.0	-0.34053	-0.40938	-0.09764
	f.20521.0.0	-0.35621	-0.36228	-0.10418
	f.20524.0.0	-0.31619	-0.40128	-0.10728
	f.20525.0.0	-0.16113	0.175999	-0.31323
Catastrophic	f.20531.0.0	-0.29624	-0.1881	0.168518
	f.20529.0.0	-0.12538	0.016109	0.368206
	f.20526.0.0	-0.09418	0.068867	0.391261
	f.20530.0.0	-0.08016	0.092043	0.458256
	f.20528.0.0	-0.04181	0.038541	0.129188
	f.20527.0.0	-0.04853	0.125786	0.363544

Abbreviations. UKB, UK Biobank; PC, Principal Component.

Evidently, PC1 shows a pattern of high magnitude and negative loadings of the trauma exposure items. In contrast, PC2 shows a high magnitude and positive loading of childhood trauma items, high magnitude and negative loading of adult trauma items as well as a low magnitude loading of catastrophic trauma. Finally,

PC3 shows a pattern of high magnitude and positive loading of catastrophic trauma items, low magnitude and predominantly positive loading of childhood trauma as well as a low magnitude and negative loading of adult trauma.

Finally, **Figure 6** shows the meta-analysed estimates for the proportion of phenotypic variance explained by the different sources included in the mixed linear models. The E was computed using all eigenvectors of full trauma. Analyses were conducted using the joint-sex, female only and male only samples. Full results of these analyses including estimates, SEs, LRTs can be found in (Appendix II Table 10, 19-20). Similar estimates were observed when exploring analyses using only unrelated individuals, also available in (Appendix II Table 15, 21-22). Note here, that Figure 6 is scaled (y-axis from 0 to 1) differently to Figures 3-5 (y-axis 0-0.5).

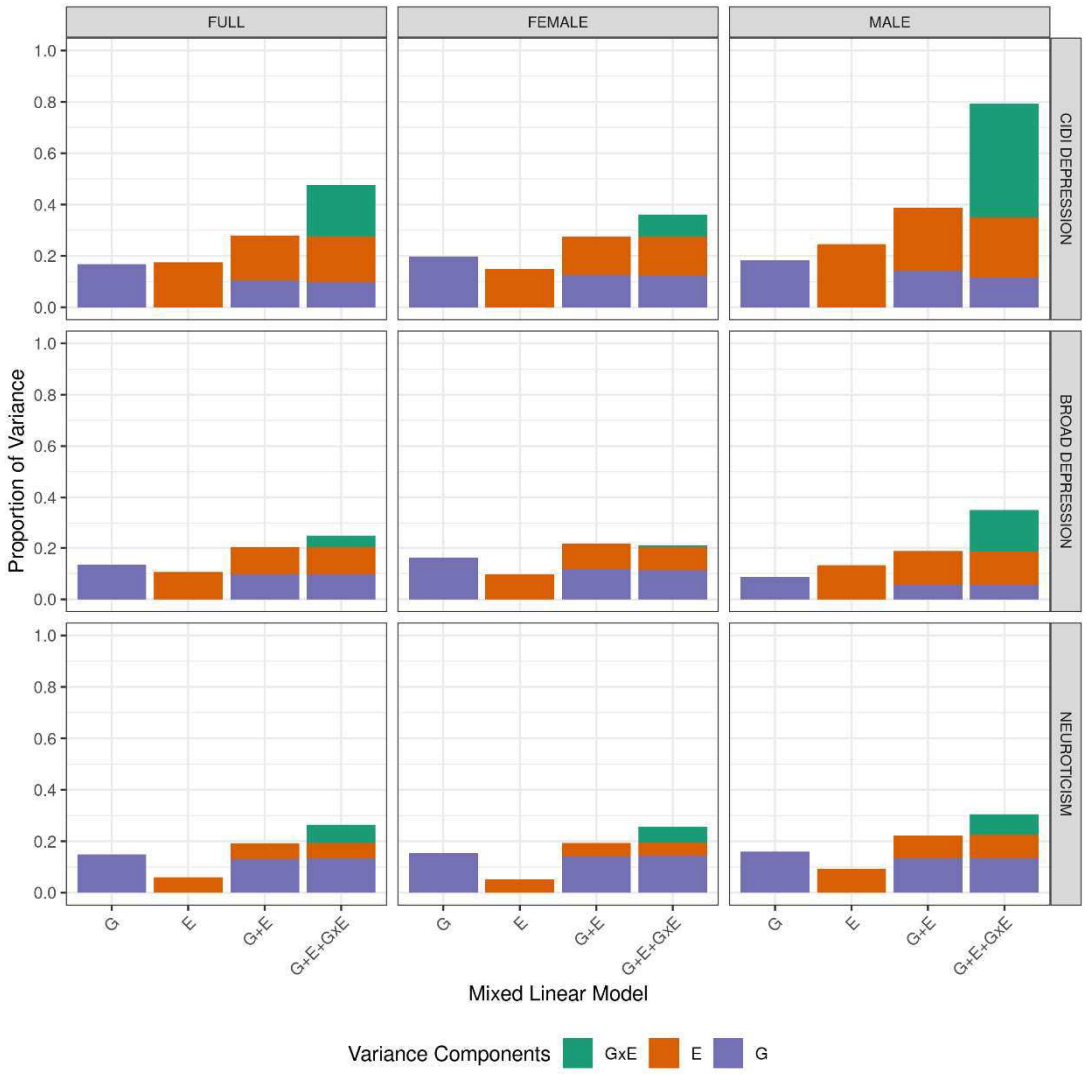


Figure 6. Proportion of Phenotypic Variance Explained by Genetic, Environmental and Interaction Sources in UK Biobank. Here, mixed linear models including environmental relationship matrices computed using all principal components (PCs) full trauma. The facets represent results from the full (joint-sex), female only and male only samples. The proportion of variance explained is specified on the y-axis. The sources of variation: G, Genomic; E, Environmental; GxE, Gene-by-Environment Interactions are specified with different colours and their corresponding models are specified on the x-axis.

Heritability estimates in females were higher than estimates observed in the full sample; 0.198 [SE 0.029], 0.162 [SE 0.024], 0.154 [SE 0.017] for CIDI depression, broad depression and neuroticism, respectively. In contrast, the estimates for proportion of variance captured by trauma exposure, were lower; 0.15 [SE 0.022], 0.098 [SE 0.015], 0.051 [SE 0.008] for CIDI depression, broad depression and neuroticism, respectively. Similarly, estimates of the genome-by-trauma exposure

interaction variance were lower, with statistically non-significant results observed in broad depression. Estimates were 0.086 [SE 0.009], 0.007 [SE 0.007], 0.061 [SE 0.005] for CIDI depression, broad depression and neuroticism, respectively.

Heritability estimates in males were slightly different to estimates observed in the full sample; 0.184 [SE 0.043], 0.086 [SE 0.034], 0.16 [SE 0.020], for CIDI depression, broad depression and neuroticism, respectively. In contrast to results observed in females, the estimates for proportion of variance attributable to trauma exposure were greater; 0.247 [SE 0.036], 0.132 [SE 0.020], 0.092 [SE 0.014] for CIDI depression, broad depression and neuroticism, respectively. Estimates of the genome-by-trauma exposure interaction variance were substantially greater than estimates observed in the full sample and female only sample. Estimates were 0.441 [SE 0.018], 0.165 [SE 0.013], 0.080 [SE 0.007] for CIDI depression, broad depression and neuroticism respectively.

A consistent pattern observed in results were large LRT values when including Es into the mixed linear models. This suggests a substantial improvement in model fit when including the Es. Whilst the estimates and standard errors observed for the proportion of variance explained by the Es suggest statistical significance, they do not suggest very *strong* significance. This is discrepant with the substantially large LRT values obtained by inclusion of the ERMs in the models. This discrepancy is explored in Appendix II Section B.

3.5. Discussion

This study sought to provide an update to the literature on genome-by-trauma exposure interaction effects within depression and depression related traits. In contrast to the studies in the literature, this project explores variance components as opposed to fixed effect estimates of genome-by-trauma interactions. Fixed effect analyses, as seen in the genome-by-trauma interaction literature on depression, mostly explore the regression of depression on trauma and genome-by-trauma interaction variables. In contrast, analyses of variance components rely on similarity within individuals. We compute genetic (using all genotyped variants), trauma

exposure and genome-by-trauma exposure interaction effect similarity to explore trait variance attributable to these effects. We explore general (full trauma) and specific (childhood, adult, catastrophic) measures of self-reported trauma exposure.

Initial mixed linear model (MLM) results highlight the heritable component of the trauma measures, supporting findings from the literature (Warrier et al, 2021). Results suggest moderate genetic correlations between the trauma and depression/neuroticism variables. Genetic correlations are 2-fold greater with CIDI depression as opposed to broad depression and neuroticism. These findings suggest the presence of shared genetic risk factors between trauma exposure and specifically depression. This may highlight a limitation as it may be possible that individuals more extreme on the depression scale are more likely remember or perceive an event as traumatic, inducing this genetic correlation.

Mixed linear model results provide estimates of depression and neuroticism SNP heritability (proportion of variance captured by the common additive genetic variants) and estimates for the proportion of variance captured by trauma exposure in line with previous literature (Jermy et al, 2021; Realo et al, 2017; Schneider et al, 2021). Findings also suggest that the sub-categories; childhood, adult and catastrophic trauma explain a substantial proportion of CIDI depression variance. Additional analyses accounting for all other sub-categories of trauma exposure would be necessary to explore individually explained variance.

These results are the first to show estimates for proportion of depression variance captured by genome-by-trauma interaction effects that are highly significant and of relatively large magnitude (e.g. variance of interaction effects in CIDI depression; 0.074 [SE 0.006]- 0.201 [SE 0.009]). Variance captured by interaction effects are shown to be approximately the same magnitude as the variance captured by self-reported trauma exposure itself (0.05 [SE 0.013]- 0.18 [SE 0.025]). These are consistent (although estimates are smaller) across broad depression and neuroticism phenotypes (see **Figure 3**). These findings suggest the effects of trauma exposure and genome-by-trauma exposure interactions perhaps play a

larger and more specific role in CIDI depression as opposed to broader psychological distress or neuroticism.

The smaller estimates, particularly for variance attributable to genome-by-trauma interaction effects, observed from models including Es computed using only the first eigenvector (PC1) of all trauma items, suggest that genome-by-trauma interaction effects are distributed across the different dimensions of self-reported trauma exposure (see **Figure 5**). The inclusion of lower order self-reported trauma exposure eigenvectors, show that trauma exposure PC2 is particularly influential in genome-by-trauma interaction effects on depression. The PC loadings, suggest that the first PC of trauma exposure is capturing the presence/absence of trauma exposure. The magnitude and direction of loadings suggest PC2 to be capturing the differentiation between exposures to childhood versus adult trauma. Finally, loadings suggest PC3 to be capturing exposure to catastrophic trauma. This may explain observed individual differences in depression manifestation after trauma exposure. Whilst the loadings are informative, it is difficult to formally explore and interpret findings using dimensions as opposed to raw linearly meaningful measures of trauma exposure (e.g. larger values meaning greater trauma exposure).

Exploring variance components of depression/neuroticism within males and females separately indicates that the proportion of variance captured by trauma exposure and genome-by-trauma interaction effects differ substantially between the sexes. Variance component estimates suggest a slightly larger role of genetics within females, and trauma exposure within males. Estimates of variance attributable to genome-by-trauma interaction effects are approximately 5-fold greater for CIDI depression in males. This pattern, although estimates are lower, was also observed for broad depression and neuroticism phenotypes. These results, alongside the evident prevalence differences, highlight the need to explore these effects within the sexes separately. Findings suggest that trauma exposure and sensitivity to trauma exposure accounts for greater variance in depression/neuroticism outcomes within males. These results should be

interpreted with caution, as technical possibilities e.g. having a greater number of female depression cases vs male cases may influence results. Further analyses would be needed to explore direction of effects, which may also provide insight into observed depression prevalence differences between the sexes.

The research design of this study consists of intrinsic replication, as analyses are repeated across five geographical cluster samples. These within-sample replications, while not independent samples, yielded relatively consistent estimates and standard errors, increasing confidence in the results produced.

Findings provide evidence of differential effects of trauma exposure dependent on differences in individual genetic liability to depression. Results can highlight both modifiable environments for preventative measures as well as at-risk groups that should be prioritised for these interventions. It is evident that the use of all genome-wide SNP data yields substantial findings when exploring GxE effects as opposed to much of the literature making use of individual genome-wide variants and polygenic scores (Arnau-Soler et al, 2018; Coleman et al, 2020; Peyrot et al, 2014; Peyrot et al, 2018). This method can be further enhanced by exploring individual trauma (e.g. neglect, physical abuse etc.) measures to gain a better understanding of the effect of specific trauma experiences.

As with all research, this study has several limitations that need to be considered when interpreting results. Here, we utilise mixed linear models to estimate variance components where the majority of our outcome variables are binary traits (depression). This highlights potential biases in results due to ascertainment biases, which may be circumvented with utilising methods adapted to deal with binary traits (Golan et al, 2014; Weissbrod et al, 2018). Moreover, principal components of trauma exposure items are utilised in analyses. Whilst this circumvented the challenges of appropriately weighting specific trauma exposure items, it has increased difficulty in interpretation of results. Incorporating measures of trauma exposure with greater objectivity, e.g. omics measures (DNA methylation) may be able to further clarify the relationship between genetic liability and trauma

exposure that is less susceptible to reporting bias. For instance, the availability of methylation data has and continues to increase substantially, and can be used as good, and in some cases far better proxy measures of environments; as seen with smoking (Amador et al, 2021; Andersen et al, 2017). Trauma exposure being heritable, as shown with results in this study and findings from the literature (Warrier et al, 2021), highlights the suitability of using available methylation data for trauma exposure. Moreover, evidence suggests there may be a distinct methylation profile observed in individuals reporting trauma exposure (Matosin et al, 2017; Mehta et al, 2017). Findings, particularly genome-by-environment interaction effects utilising this form of data, can then be dissected to explore biological pathways with non-additive effects on outcomes that can be targeted directly.

Additionally, whilst the genomic relationship matrices computed utilise all available genotyped SNPs, and subsequently account for much greater variance than PGSs; discrepancies between twin study heritability and SNP heritability estimates of depression are still apparent. It is important to note, however, that twin study estimates may be biased upward, and thus, real heritability estimates are likely to fall between SNP heritability and twin study heritability estimates (Yang et al, 2015). As mentioned above, our results show that our environmental variables; full trauma and the sub-categories of trauma have moderate heritability estimates and these are genetically correlated with our outcome phenotypes. This highlights that our environmental measures (environmental relationship matrices) capture both genetic and environmental variances. As our genetic variable is not capturing the *entirety* of the genetic variance within depression/neuroticism outcomes, the variance captured by the trauma (and sub-trauma) measures may capture residual genetic variance. Simulation findings from the literature suggests that making use of imputed or whole-genome sequencing genetic data for GRMs can uncover a further substantial proportion of genetic variance (Yang et al, 2015) which would be useful in addressing this limitation (although it is important to note that it is difficult to exclude this possibility completely), and aiding clear interpretation of results.

Furthermore, it is evident that gene-by-environment and gene-environment correlations co-occur (Lau & Eley, 2008) and are difficult to tease apart (Akimova et al, 2020). Here, to control, as much as possible, for genetic covariance (and thus, potentially gene-by-environment correlations) between our environmental and outcome variables, we also explored measures of trauma pre-corrected for the available genetic measure (GRMs). The differences in estimates were negligible. However, it is important to note, similar to the aforementioned limitation, that this effect can be more accurately controlled for with an improved genetic measure (e.g. genomic relationship matrix utilising imputed or whole-genome sequenced data).

Our trauma exposure, depression and neuroticism variables were measured using retrospective self-report. Furthermore, measures of trauma exposure and CIDI depression were obtained later than measures of broad depression and neuroticism with the follow up UKB mental health questionnaire. This indicates potential measurement error within our variables (Coleman et al, 2020). However, self-report is likely the most effective method to explore these phenotypes in studies the size of UKB. As mentioned above, making use of omics data as proxy environments or additional measures of environmental variables may be useful in avoiding limitations associated with measurement error.

Whilst UKB provides relatively homogenous data, evidence suggests a healthy volunteer bias with participation being associated with greater mental health (Adams et al, 2020), overall health and higher socioeconomic status (Fry et al, 2016) This may impact the generalisability of results to the whole population and other ethnicities as well as clinically-defined cases of depression.

In conclusion, our findings provide empirical evidence of depression/neuroticism variance attributable to genome-by-trauma interaction effects. We further highlight that the magnitude of these effects are much larger for males in comparison to females. These findings can be further explored to identify both risk groups and

modifiable environments/biological pathways which yield greater risk of depression manifestation, which would be useful in personalised/preventative interventions.

Chapter 4: Discussion

4.1. Overview

Various methods have been used within the literature highlighting the potential impact of gene-environment interplay effects in depression. Studies have recently shown that biological parent-offspring genetic nurturing effects are detectable using current polygenic scores (PGSs) in complex traits such as educational attainment (Bates et al, 2018; Kong et al, 2018; Morris et al, 2020). These designs have yet to be adapted and implemented using available depression data.

In contrast, efforts to explore gene-by-trauma exposure interaction effects in depression have yielded inconsistent findings using various methods (Assary et al, 2018; Culverhouse et al, 2018; Peyrot et al, 2018). I believe these studies are underpowered, so findings are currently unconvincing. Therefore, the overarching aim of this thesis was to implement methods that accurately and effectively explore these gene-environment interplay effects in depression.

Two main analysis chapters reported analyses exploring gene-environment correlation and interaction effects, separately. Research aims for gene-environment correlation effects were (1) to accurately quantify biological parent-offspring genetic nurturing effects using available PGSs; (2) to highlight confounding and biases that arise when mechanisms of genetic nurturing are explored using available parental phenotypes. Research aims for gene-environment interaction effects were (3) to maximise statistical power and use all available genome-wide SNP data to explore genome-by-trauma interaction effects in depression.

Detailed discussions specific to each research aim are available within the analysis chapters. This chapter will serve as a broader discussion, evaluating the main findings, their contributions to the available literature, their strengths and limitations of the methods, suggestions for future work and provide an overall conclusion of this thesis.

4.2. Main Findings and Contributions to the Literature

4.2.1. Genetic Nurturing Effects

Findings from the literature highlighted the need to explore transmitted and non-transmitted PGSs to quantify biological parent-offspring (trio) genetic nurturing effects in complex traits such as educational attainment (Bates et al, 2018; Kong et al, 2018). In the current thesis, parental genetic nurturing effects were explored using available biological parent-offspring phenotypes and full PGSs.

One of the main findings (in Chapter 2) suggested that the use of regression and pathway models including full trio PGSs, provided a straightforward quantification of parental genetic nurturing effects on offspring phenotypes. This is particularly useful, as computing full PGSs and thus implementing these models is relatively straightforward. In contrast, estimating separate PGSs based on alleles transmitted and non-transmitted from parents to offspring is more complex, as this relies on high quality phasing as well as reliable and relevant genetic reference panels a.k.a haplotypes.

The regression and pathway models are a good starting point to filter relevant phenotypes with sufficient statistical power to detect and quantify the impact of genetic nurturing effects using available PGSs and sample sizes. Results can signify phenotypes with complex underlying mechanisms, which may require further exploration. Chapter 2 findings suggested clear evidence of genetic nurturing for educational attainment but not for depression or the other explored traits. However, non-significant findings for depression could be due to insufficiently precise PGSs.

Understanding whether parental genetic nurturing effects are involved in disease aetiology is useful for planning interventions and mitigating strategies. Identifying the mechanisms of genetic nurturing effects may yield highly translational findings i.e. identifying targetable parental behaviours or parent behaviour influenced environments, promoting early intervention. The findings of chapter 2 concluded that data limitations meant exploring parent phenotype mediated genetic nurturing effects resulted in confounded effects within the models explored. Current PGSs

capture only a fraction of trait genetic variance (Dudbridge, 2013; Mostafavi et al, 2020), particularly for traits like major depressive disorder (MDD) (Howard et al, 2019; Wray et al, 2018a), meaning genetic correlations between parental and offspring phenotypes are not adequately accounted for and thus, estimates are susceptible to confounding. Associations will always arise between parental phenotypes and offspring phenotypes, which can lead to potentially erroneous interpretations of significant parental genetic nurturing effects via the parental phenotype from the models outlined in chapter 2. Moreover, the confounding is also shown to impact estimates of residual parental genetic nurturing effects, suggesting most components of the pathway and regression models are biased, and thus, accurate and reliable interpretation of results is difficult.

The pathway and regression models are relatively straightforward to implement, and less computationally intensive in comparison to models incorporating large similarity matrices. However, it is evident that the level of accuracy and predictive validity required of PGSs for reliable results from the models exploring parental genetic nurturing mechanisms with parental phenotypes, is difficult to achieve. This would require all trait associated variants to be genotyped and sufficient statistical power for GWAS to *accurately* identify all relevant associations. This highlights substantial advantages accompanying the use of genomic relationship matrices (GRMs), as these utilise all available genotyped genome-wide SNP variants, and account for a substantially larger proportion of variance than PGSs for most traits (Yang et al, 2011).

4.2.2. Interaction Effects

Interaction effects between genetic risk and trauma exposure have been repeatedly explored within the literature (Assary et al, 2018). Inconsistent findings have been highlighted, likely fuelled by false positive candidate gene associations and imprecise PGSs (Border et al, 2019; Culverhouse et al, 2018; Werme et al, 2021). To address the previously highlighted limitations associated with PGSs, genome-by-

trauma interaction effects were explored using genomic and trauma exposure similarity matrices in chapter 3.

Findings from the literature highlight that trauma exposure is complex, multi-faceted and heritable trait (Warrier et al, 2021). Various studies have shown trauma exposure to be genetically correlated with depression (Coleman et al, 2020), which was supported by results conducted within this thesis. These findings indicate potential gene-environment correlation effects, which would need to be controlled to gain an accurate understanding of genome-by-trauma interaction effects.

Adequately combining data on trauma exposure to form a single representative variable is not straightforward (Coleman et al, 2020; Warrier et al, 2021). Understanding and quantifying the importance of different trauma exposures is complex research in itself (de Oliveira et al, 2018). Hence, in this thesis, principal components (PCs) were computed to capture the underlying dimensions of trauma exposure. This method maximised statistical power and ensured all available trauma exposure data was utilised. However, the use of PCs, as opposed to raw or meaningfully combined trauma exposure data, resulted in findings that were challenging to interpret.

Findings showed that genome-by-trauma interaction effects accounted for a substantial proportion of variance in depression, approximately the same magnitude as the trauma exposure effects on their own. Moreover, the magnitude of these effects emphasised the role and complexity of trauma exposure on depression manifestation. However, insight in to which trauma exposures were influential in interaction effects was limited due to the nature of PCs capturing latent underlying dimensions. Nonetheless, these findings are particularly exciting, as studies within the depression gene-by-trauma interaction literature have failed to show consistent evidence of interaction effects similar to the scale shown with results in chapter 3 (Assary et al, 2018). Replicating this study design within other datasets would be useful to evaluate the robustness and validity of results.

The results yielded from this design highlights the importance of strong genetic risk measures and point to the need for well represented/measured environmental variables.

4.3. Strengths, Limitations and Future Work

Obvious strengths of this thesis include the use of two population-scale cohort datasets; UK Biobank (UKB) and Generation Scotland (GS), resulting in analysis samples of up to 145,000 participants with available genetic and phenotypic data. Limitations included the retrospective self-report of many variables explored and potentially increased risk of recall bias (Coleman et al, 2020). However, as mentioned before in this thesis, self-report is likely to be the most effective method to explore these phenotypes in studies the size of UKB and GS. Optimistically, large multi-generational family cohorts with longitudinal designs are being formed, aiming to capture both retrospective and prospective measures of traits/environments e.g. Avon Longitudinal Study of Parents and Children (ALSPAC) (Fraser et al, 2013) and the Norwegian Mother and Child Cohort Study (MoBA) (Magnus et al, 2016). The latter cohort in particular, not only yields the large sample sizes needed for analyses conducted within this thesis, MoBa also provide high quality genetic and environmental data from *family* members which is particularly useful for future exploration of parental genetic nurturing effects.

Models using trio PGSs provide robust quantifications of parental genetic nurturing effects for traits with large sample sizes and nurturing effects of a reasonable size (Kong et al, 2018). It is evident that the limitations of PGSs prevent the extension of models with parental phenotypes for further exploration of genetic nurturing mechanisms. However, future work can entail the adaptation of these models to explore genetic nurturing effects using PGSs for different traits in parents than the trait of interest in the offspring. Results from these analyses would provide insight into mechanisms of genetic nurturing without modelling specific parental phenotypes, and thus avoid the confounding effects demonstrated in chapter 2 of this thesis. These models will likely require cautious interpretation as variations in power of PGSs (accuracy and amount of variance captured) between different traits

are unavoidable (Cesarini & Visscher, 2017). Ideally, additional supplementary analyses using simulated data, to explore whether differences in PGS power induce biases in effect sizes, should be conducted.

The lack of predictive ability of current PGSs, particularly in psychological/psychiatric traits, has been termed missing heritability and is outlined in more detail in the introduction of this thesis. Evidence suggests potential upward biases in heritability estimates from family studies and that utilising all available genotyped data in mixed linear models can uncover much more variance (Yang et al, 2015). Overall, these results and findings from the literature, suggest that despite being computationally exhaustive, computing and implementing genomic relationship matrices in models is a fruitful method for future studies exploring mechanisms of genetic nurturing. The use of GRMs capturing offspring genomic (direct genetic effects) and parental genomic similarity (additional genetic nurturing effects) has shown a substantial proportion of depression and anxiety variance attributable to genetic nurturing effects (Cheesman et al, 2020a). However, similar study designs have yielded non-significant findings (Jami et al, 2020). This discrepancy may be attributable to sample size, and thus, statistical power limitations; as current family cohorts with available genetic and phenotypic human data are relatively small. Nonetheless, making use of GRMs may yield greater statistical power when exploring genetic nurturing effects than the particularly low powered PGSs in traits such as depression. Additional inclusions of parent phenotypes or other environmental covariates within these study designs can provide insight into whether these variables play a significant role in the parental genetic nurturing mechanisms.

Mixed effect models including these similarity matrices have some limitations that need to be understood. These matrices capture greater phenotypic variance than PGSs, but are still less than pedigree based heritability estimates, meaning they do not capture the entirety of genetic variance in complex traits (Yang et al, 2015; Yang et al, 2011). Thus, these results still require cautious interpretation. Additionally, although computing and including genomic relationship matrices in the offspring

generation (capturing direct genetic effects) and parental generation (capturing genetic nurturing effects) is a solid and powerful design, they are still incomplete. Improvement can be achieved by including GRMs within siblings, grandparents, other relatives and even spouses to capture and control for additional genetic nurturing effects via other family members. As mentioned above, the current data availability for these models are substantially limited for human complex traits (Kong et al, 2018).

Nonetheless, chapter 3 results of analyses exploring genome-by-trauma interaction effects demonstrate that the use of GRMs as opposed to PGSs also yield promising findings when exploring effects other parental genetic nurturing. In contrast, the difficulty in interpretation of results using trauma exposure PCs, highlighted limitations associated with the measurement and representation of environmental variables. For many complex, seemingly environmental traits, such as, trauma exposure, there are many layers e.g. heritability, type of trauma and severity (Warrier et al, 2021) that need to be appropriately captured.

As mentioned in chapter 3, making use of omics data, such as methylation data, as proxy environments or additional measures of environmental variables may be useful in avoiding limitations associated with measurement error (Amador et al, 2021; Andersen et al, 2017). The complexity of trauma exposure, an evidently heritable environmental variable, is particularly suitable to be explored using methylation data, although it is important to note that this research area is still in its infancy. Identified methylation sites can give insight into genetic variants, and thus, biological mechanisms that have not been previously identified by more traditional genome-wide association study (GWAS) methods; due to the genetic variants not being genotyped or included in these analyses. The genome-by-environment interaction effects explored using methylation data can give insight into biological mechanisms with non-additive effects on diseases which may be particularly useful when developing treatment interventions.

Moreover, identifying and utilising methylation sites with lower heritability can be useful in exploring the *non-genetic* component of heritable *environments*, like trauma exposure. This would also minimise the impact of genetic correlations between traits explored. It is important to note that environmental factors can also impact methylation, and as shown throughout this thesis, many environments have a genetic component; so this effect cannot be entirely eliminated.

The suggestions for future work outlined above have the potential to yield high impact and highly translational findings. For example, identifying methylation sites which interact with genetic risk can be further dissected to explore potentially novel influential genes and biological pathways (Wang et al, 2019) . This level of insight in depression (or any disease) manifestation can then evolve into research on drug-target discovery and biological stratification of depression subtypes.

4.5. Conclusions

Depression is a complex disorder with evidence indicating the importance of gene-environment interplay effects. Underpowered, inconsistent and confounded findings within the literature, point to the need for models that can accurately explore these effects and remain robust to data limitations that will take time to resolve. It is obvious that advanced methods, such as the computation of PGSs are useful and powerful in the differentiation of disease outcomes (Howard et al, 2019). This thesis contributes two studies that demonstrate how PGSs can fall short when aiming to explore more complex gene-environment interplay effects. Making use of all available genetic data in the form of similarity matrices result in improved representative measures of genetic variance, which subsequently yield promising findings. Evidence and evaluations of limitations emphasised within this thesis can be used to guide future study designs. Moreover, increased efforts to collect large-scale family data and omics measurements indicate an upcoming data rich era, where these suggestions can be realised. Implementing models that can accurately quantifying gene-environment interplay effects and provide insight into these

mechanisms involved in depression, may be impactful on future healthcare and personalised medicine.

Glossary

Active Gene-Environment Correlation (rGE):

Active rGE occurs when an individual selects environments based on genetically influenced traits, resulting in a correlation between the individual's genotype and the chosen environment.

Additive model:

The additive model, also known as the additive genetic model, suggests that genetic variation can exert influence on a trait in a linear or additive manner, as opposed to the genetic variants interacting with one another.

Aetiology:

Aetiology refers to the underlying cause, or set of causes of a disease or trait.

Allele:

An allele is one of often two (or more) versions of a gene. An individual inherits two alleles for each gene, one from each parent. Alleles can come in different sizes; at the lowest level an allele can refer to a single nucleotide polymorphism (SNP) (Smigielski et al, 2000) and at the higher level, an allele can be up to several thousand base-pairs (Elston et al, 2012).

Base-pair:

Refers to the fundamental unit of double-stranded nucleic acids forming the DNA double helix. A base-pair refers to the nucleotide bases (adenine, thymine, cytosine and guanine) bound together by hydrogen bonds.

Beta Coefficient:

The beta coefficient or estimate represents the degree of change in an outcome variable (dependent variable) for every 1-unit change in the predictor variable (independent variable).

Biological trios:

Biological parent-offspring relationships consisting of the mother, father and a single offspring.

Bivariate:

Bivariate analyses involve two outcome/dependent variables.

Collider bias:

A collider variable is a variable that is independently influenced by the predictor and dependent variable within a model. Including the collider variable (also referred to as conditioning on a collider variable) can result in a distorted association between the predictor and dependent variable. This is different to a confounder variable, which induces biases when *not* modelled.

Complex traits:

Complex traits are also known as quantitative traits. These traits are often polygenic, meaning several genes are involved in the manifestation of these traits.

Confounder/Confounding:

A confounder is a variable that independently influences a predictor and dependent variable within a model. Not including the confounder within the model can result in a distorted association between the predictor and dependent variable. This is different to a collider variable, which induces biases when *modelled*.

Correlation Effect:

Correlation refers to associations. This is a measure of the extent to which two variables are related. A positive correlation refers to when one variable increases as the other variable increases, or alternatively, decreases as the other variable decreases. In contrast, a negative correlation would refer to when one variable increases as the other variable decreases (or vice versa).

Dependent Variable:

The dependent variable can also be referred to as the outcome variable. This is the variable that is being measured or tested. These variables depend on other factors (the independent variables/predictors).

Deoxyribonucleic Acid (DNA):

A molecule composed of two strands consisting of nucleic acid bases. These are bound together via hydrogen bonds and form a double helix which carry the genetic sequences for the development, functioning and growth of all organisms.

Diathesis Stress:

The diathesis stress model is the explanation that a disorder or trait is the result of an interaction between the genetic predisposition and stress (often caused by life events).

Direct Genetic Effect:

This effect refers to when the phenotype of an individual is directly influences by their own genes.

Dizygotic:

Refers to twins derived from two separate eggs (also known as non-identical).

Duos:

Biological parent-offspring relationships consisting of a single parent (mother OR father) and a single offspring.

Dysfunctional Attributional Styles:

Attributional styles, also known as cognitive styles, refer to the attribution/explanation of events. A dysfunctional attributional style refers to when an individual attributes negative events to internal characteristics.

Effect Size:

Effect sizes provide insight in to how meaningful the relationship between variables or difference between groups is. Whilst effects may be statistically significant, larger effect sizes provide insight into the practical significance of these relationships.

Eigenvectors:

Geometrically speaking, principal components represent the directions of the data that explain the maximal amount of variance. These can be thought of as underlying dimensions of the data. Each vector, or principal component can also be termed an eigenvector.

Epidemiology:

This is a field which explores incidence, distribution and possible control of diseases/factors relating to health.

Evocative Gene-Environment Correlation (rGE):

Evocative rGE occurs when an individual's genetically influenced trait evokes a reaction from others shaping the individual's environment, resulting in a correlation between the individual's genotype and their environment.

False Positive:

This is when the result of a test is incorrectly indicating a significant condition/effect.

Fixed Effects:

Fixed effects are fixed or non-random quantities. In most biostatistics applications a fixed effects model refers to a regression model, where the group means are fixed.

Gene Environment Correlations (rGEs):

This phenomenon refers to when an individual's genotype and their environments are correlated. This is due to the genetic effects of the individual shaping the environment in which they are exposed to.

Gene-Gene Interactions:

Gene-gene interactions are also known as epistasis. This is when the effect of one gene on a disease is modified by another gene or several other genes.

Gene:

A distinct sequence of nucleotide bases within the DNA which code for proteins.

Genetic Heterogeneity:

Genetic heterogeneity refers to when two or more genes produce the same or similar phenotypes.

Genetically Correlated:

This refers to the proportion of variance that two traits share due to genetic aetiology.

Genome:

The complete set of genes or genetic material present in a cell or organism.

Genome-wide:

Spanning across the whole genome

Genome-Wide Association Study (GWAS):

Is an observational study of genome-wide genetic variants in different individuals to see if any variant is associated with a trait.

Genomic Relationship Matrix (GRM):

Also known as a genomic similarity matrix, refers to the pairwise genetic similarity of individuals within a sample. Genomic similarity is measured using genotyped data, and accounts for number of variant copies and the frequency of the variant within a population.

Genotype:

Refers to the genetic material of an individual (or organism).

Genotyped data:

Refers to the measure of specific genetic variants spanning across the genome.

Genomic Restricted Maximum Likelihood Estimation (GREML):

The estimation of genomic parameters (covariance, heritability and correlation) for a linear mixed model using restricted (or residual/reduced) maximum likelihood estimation (REML).

In statistics, maximum likelihood estimation is a method of estimating parameters of an assumed probability distribution, given an observed dataset. The likelihood function is maximised so that the observed data is most probable.

In contrast, the restricted maximum likelihood estimation used to estimate variance components, uses a transformed dataset (contrasts calculated from the data). The likelihood function is then calculated from the probability distribution of these contrasts. For the estimation of variance components the transformation is such that the fixed effects do not bias the estimates of variance and covariance components.

Genome-Wide by Environment Interaction Studies (GWEIS):

Is an observational study of genome-wide genetic variants in different individuals and the interaction of these variants with a specified environmental variable to see if any variant and/or the interaction effect of the variant and environmental variable is associated with a trait.

Heidy Weinberg Equilibrium:

This is a principle stating that genetic variation within a population will remain constant from one generation to the next. When mating is random in a large population, the law predicts that both genotype and allele frequencies will remain constant as they are in equilibrium. Mutations, natural selection, non-random

mating, genetic drift and gene-flow can disrupt the hardy Weinberg equilibrium as these alter gene frequencies within a population.

Heritability:

This is a population estimate, referring to the proportion of phenotypic variation that can be attributable to genetic differences in individuals.

Heterogeneity:

Refers to diversity in the character/content of a trait.

Hippocampal Volume:

The hippocampus is a brain region associated with memory alongside other brain functions. The volume is a measure of this brain region.

Homogenous:

Refers to the quality or state of a trait being the same or all of the same kind.

Incidence/Prevalence:

In epidemiology, the incidence or prevalence refers to the rate of new cases of a disease/disorder occurring within a specific population.

Independent Variable:

The independent variable can also be referred to as the predictor variable. This is the variable that can be manipulated or changes and is assumed to have a direct effect on the dependent variable.

Interaction Effect:

A variable that has an interaction effect will have a different effect on the dependent variable, depending on the level of some third variable (the moderater). This is also known as a moderation effect.

Jackknife Technique:

This approach is a re-sampling method used to estimate the variance and bias of a large population. It involves aggregating the results of repeated analyses datasets size (N-1) each omitting one individual from the sample to estimate a parameter in a dataset of N observations.

Liability Scale:

The liability scale often refers to the underlying normal distribution of traits. It is assumed that categorical or binary traits have an underlying normal distribution, whereby a threshold/s on this liability scale determines the observed category of the trait. In these cases, the prevalence of the trait/disease is taken in to consideration when transforming these data to the liability scale.

Locus/Loci:

The particular position/location of a genetic variant.

Longitudinal Studies:

A type of correlational research in which researchers observe and collect data on a number of variables repeatedly over time, without trying to influence those variables.

Log-likelihood Ratio Test (LRT):

In statistics, the LRT is an assessment of goodness of fit/model fit/model utility of two competing statistical models.

Major Depressive Disorder (MDD):

Is a common psychiatric disorder classified based on the presence of one of either two cardinal symptoms: persistent low mood and anhedonia (the loss of pleasure in activities that often induce pleasure) alongside a range of other symptoms. These symptoms should be present for the most part of the day and week, for at least two weeks. These symptoms should be causing substantial dysfunction. These criteria form the basis of a clinical diagnosis of MDD.

Mediation:

In statistics a mediation model seeks to explain the mechanism or process that underlies an observed relationship between an independent variable and a dependent variable via the inclusion of a third variable (a mediator).

Meta-Analysis:

The examination of data from several independent studies/analyses of the same subject to determine overall trends.

Methylation:

This refers to a simple biochemical process, where a methyl group is transferred to the C5 position of the cytosine nucleotide base. This process regulates gene expression (aka the gene being copied and transcribed to form proteins).

Minimal Phenotyping:

The diagnosis of a trait/disease reliant on self-reported diagnosis by a healthcare professional.

Minor Allele Frequency (MAF):

The frequency of the less common genetic variant within a population.

Missing Heritability:

This phenomenon refers to the discrepancy in heritability estimates obtained from twin/family studies and molecular genetic studies.

Mixed Linear Model:

Also known as mixed-effects models. These are statistical models that contain both fixed and random effects.

Model fit/utility:

Also known as the goodness of fit of a model. Model fit refers to how well the model fits the observed data.

Moderation:

A variable that has an interaction/moderation effect will have a different effect on the dependent variable, depending on the level of some third variable (the moderater).

Molecular Genetics:

The study of the molecular structure of the DNA and its cellular activities.

Moment-Based Method:

Is a method of estimation of population parameters. It starts by expressing the population moments (i.e. the expected values) as the sample moments and ultimately solving for the parameters.

Monozygotic:

Refers to twins derived from a single egg split in two (also known as identical).

Non-additive Genetic Effects:

Refers to the effects of genes that are not linear of additive on a given trait/disease. Non-additive genetic effects can entail gene-gene interactions and gene-environment interactions.

Observed Scale:

The observed scale refers to how the data presents itself, as opposed to the liability scale which refers to the underlying quantitative normal distribution of a given trait which may be binary/categorical on the observed scale.

Omics Data:

This encompasses data from the disciplines of genomic, proteomics, metabolomics, metagenomics, transcriptomics. The overall aim is to collectively characterise and quantify the impact and function of biological molecules within an organism.

Ordinal Scale:

Where variables are named/labelled with a specific order. E.g. higher values represent a higher level of the trait/variable being measured.

Parental Genetic Nurturing:

Refers to the indirect effect of parental genetics on offspring phenotypes by shaping/nurturing the offspring's rearing environment, as opposed to a direct genetic effect (passing on genes to form offspring's own genome, which goes on to directly influence offspring traits).

Passive Gene-Environment Correlation (rGE):

Also known as parental genetic nurturing effects. Passive rGE occurs when the genetically influenced trait of a parent shapes the offspring's environment, resulting in a correlation between the offspring's genotype and the environment they find themselves in (due to offspring's genome being a derivation of their parent's genome).

Peripheral Cytokines:

Cytokines are small proteins often involved in the activity of the immune system. Peripheral cytokines refer to cytokines affected by the peripheral immune system.

Phased Data:

The process of phasing entails the resolution of haplotypes (groups of alleles in an organism that are inherited together from a single parent). Phased data can provide insight in to maternally or paternally inherited alleles.

Phenotype:

The observable characteristic of an individual resulting from the interaction of its genotype and environment.

Polygenic:

A trait is polygenic when 2 or more genes are involved in its genetic aetiology.

Polygenic Risk Scores:

Also known as polygenic scores. This refers to an aggregate genetic measure capturing the total number of genetic variants an individual has assessing their heritable risk of developing a particular disease (or outcome).

Preventative Treatment:

Refers to clinical services, programmes or interventions with the overall aim of disease prevention.

Principal Components:

Geometrically speaking, principal components represent the directions of the data that explain the maximal amount of variance. These can be thought of as underlying dimensions of the data.

Proxy Measure:

A proxy is an indirect measure of a variable, which is itself strongly correlated to the outcome.

Psychopathology:

The scientific study of mental disorders.

Publication Bias:

Refers to the tendency of studies with significant results being published, as opposed to non-significant results. This is also known as reporting bias.

P-Values:

Is a measure of the probability that an observed effect could have occurred just by random chance. The lower the p-value the greater the statistical significance of the observed effect.

Random Effects:

Random effects are predictor variables from which inferences can be made using the distribution of the values rather than testing the differences of values.

Random Segregation:

Refers to the law that states that two alleles of a single trait will separate randomly to form a gamete.

Reporting Bias:

Refers to the tendency of studies with significant results being published, as opposed to non-significant results.

Risk Variant:

Refers to a genetic variant which increases risk of a disease or trait.

Shared Environment Paradox:

Family members typically share many environments, however, twin/family study estimates of shared environments are often small forming the basis for the shared environment paradox.

Similarity Matrices:

Refers to matrices capturing pair-wise similarity of individuals within a sample. These measures can capture genomic, environmental and even genome-by-environment interaction similarity within a sample.

Simulations:

Simulated data refer to random data points from a computer model as opposed to measured real data points.

Singleton Offspring:

A single, or one pair of sibling offsprings

SNP Heritability:

The proportion of phenotypic variance attributable to the differences in single nucleotide polymorphisms (SNPs).

Single Nucleotide Polymorphism (SNP):

The most common form of variation within the human genome. This refers to variation in the nucleotide base at a specific location within the DNA.

Standard Deviation:

A statistical parameter measuring how values differ from the mean of the total values.

Standard Errors:

A measure of statistical accuracy of an estimate. This is equal to the standard deviation of the theoretical distribution of a large population of such estimates.

Synaptic Cleft:

Refers to the space that separates two neurons. It forms a junction where nerve impulses can be passed from one neuron to the other.

Transmitted/non-transmitted:

Genetic variants, i.e. nucleotide alleles are transmitted from both parents to form the offspring's own genome. Approximately 50% of a parent's genome is transmitted to form the offspring genome. The parental allele that forms part of the offspring's genome is referred to as the transmitted allele. The variants of the parental genome that does *not* form part of the offspring's genome is referred to as non-transmitted alleles.

Variance:

This is a measure of dispersion that takes into account the spread of all data points within a dataset.

Appendix I. Supplementary Materials for Chapter 2 'Methodological Considerations When Using Polygenic Scores to Explore Parent-Offspring Genetic Nurturing Effects'

Table 1. Generation Scotland Sample Demographics

TRIO SAMPLE					
Phenotype	**Sample**	**Min**	**Mean**	**Max**	**Total N**
Psychological Distress	**Offspring**	2.000	15.514	52.000	2315.00
	Maternal	3.000	15.635	51.000	
	Paternal	1.000	13.869	47.000	
Depression	**Offspring**	0.000	0.659	10.000	2280.00
	Maternal	0.000	0.576	10.000	
	Paternal	0.000	0.659	10.000	
Educational Attainment	**Offspring**	0.000	5.287	10.000	2361.00
	Maternal	1.000	4.313	10.000	
	Paternal	0.000	4.324	10.000	
Height	**Offspring**	145.000	171.217	200.000	2648.00
	Maternal	144.200	161.226	180.500	
	Paternal	152.800	174.678	195.600	
Age	**Offspring**	18.000	31.450	61.000	2680.00
	Maternal	38.000	58.561	85.000	
	Paternal	38.000	60.467	88.000	
Phenotype	**Sample**	**Female**	**Male**		**Total N**
Sex	**Offspring**	1517	1163		2680
WHOLE SAMPLE					
Phenotype	**Sample**	**Min**	**Mean**	**Max**	**Total N**
Psychological Distress	**Whole**	0.000	16.010	84.000	19695.00
Depression	**Whole**	0.000	0.911	21.000	19682.00
Educational Attainment	**Whole**	0.000	4.689	10.000	22781.00
Height	**Whole**	114.500	168.100	214.000	21450.00
Age	**Whole**	18.000	47.400	99.000	20027.00
Phenotype	**Sample**	**Female**	**Male**		**Total N**
Sex	**Whole**	11805	8227		20032

Abbreviations: N, sample size; Min, minimum value; Max, maximum value.

Table 2.1. Generation Scotland Polygenic Score (PGS) Principal Component Information

Trait	PGS	Loading Values	Correlation	R2	SD
Height	**X5e.08**	0.307	0.938	0.931	3.051
	X5e.07	0.312	0.952	0.931	3.051
	X5e.06	0.316	0.965	0.931	3.051
	X5e.05	0.319	0.974	0.931	3.051
	X0.0005	0.322	0.981	0.931	3.051
	X0.005	0.322	0.982	0.931	3.051
	X0.05	0.319	0.974	0.931	3.051
	X0.5	0.315	0.961	0.931	3.051
	X0.75	0.315	0.960	0.931	3.051
	X1	0.315	0.960	0.931	3.051
Educational Attainment	**X5e.08**	0.264	0.709	0.720	2.684
	X5e.07	0.287	0.771	0.720	2.684
	X5e.06	0.308	0.826	0.720	2.684
	X5e.05	0.325	0.872	0.720	2.684
	X0.0005	0.335	0.899	0.720	2.684
	X0.005	0.338	0.907	0.720	2.684
	X0.05	0.333	0.894	0.720	2.684
	X0.5	0.322	0.864	0.720	2.684
	X0.75	0.321	0.863	0.720	2.684
	X1	0.321	0.862	0.720	2.684
Depression	**X5e.08**	0.267	0.671	0.630	2.511
	X5e.07	0.283	0.709	0.630	2.511
	X5e.06	0.305	0.765	0.630	2.511
	X5e.05	0.323	0.811	0.630	2.511
	X0.0005	0.334	0.839	0.630	2.511
	X0.005	0.340	0.853	0.630	2.511
	X0.05	0.336	0.842	0.630	2.511
	X0.5	0.323	0.811	0.630	2.511
	X0.75	0.322	0.809	0.630	2.511
	X1	0.322	0.809	0.630	2.511

Abbreviations. Loading Values refer to how each polygenic score (PGS) loads onto principal component 1 (PC1). Correlation refers to the correlation between each PGS and PGC PC1. R2 refers to how much variance the PGS PC1 accounts for. SD refers to the standard deviation of PGS PC1

Table 2.2. Generation Scotland Biological Parent-Offspring Polygenic Score (PGS (Principal 1) and Phenotypic Correlations

		PGSpc				
Polygenic Score	**Combination**	**Pearson's R**	**Lower CI**	**Upper CI**	**DF**	**P-Value**
HEIGHT	**Offspring-Maternal**	0.553	0.526	0.579	2678	0.000
	Offspring-Paternal	0.527	0.499	0.554	2678	0.000
	Maternal-Paternal	0.099	0.061	0.136	2678	0.000
EDUCATIONAL ATTAINMENT	**Offspring-Maternal**	0.551	0.524	0.577	2678	0.000
	Offspring-Paternal	0.552	0.525	0.578	2678	0.000
	Maternal-Paternal	0.124	0.087	0.161	2678	0.000
PSYCHOLOGICAL DISTRESS	**Offspring-Maternal**	Depression PGSs used for Psychological Distress				
	Offspring-Paternal					
	Maternal-Paternal					
DEPRESSION	**Offspring-Maternal**	0.491	0.462	0.519	2678	0.000
	Offspring-Paternal	0.488	0.459	0.516	2678	0.000
	Maternal-Paternal	-0.021	-0.059	0.017	2678	0.272
		Phenotypes				
Polygenic Score	**Combination**	**Pearson's R**	**Lower CI**	**Upper CI**	**DF**	**P-Value**
HEIGHT	**Offspring-Maternal**	0.353	0.319	0.386	2650	0.000
	Offspring-Paternal	0.368	0.335	0.401	2662	0.000
	Maternal-Paternal	0.308	0.273	0.342	2658	0.000
EDUCATIONAL ATTAINMENT	**Offspring-Maternal**	0.238	0.200	0.275	2468	0.000
	Offspring-Paternal	0.224	0.186	0.262	2430	0.000
	Maternal-Paternal	0.414	0.380	0.446	2468	0.000
PSYCHOLOGICAL DISTRESS	**Offspring-Maternal**	0.111	0.071	0.151	2354	0.000
	Offspring-Paternal	0.084	0.044	0.124	2361	0.000
	Maternal-Paternal	0.119	0.079	0.158	2366	0.000
DEPRESSION	**Offspring-Maternal**	0.050	0.010	0.091	2323	0.015
	Offspring-Paternal	0.056	0.016	0.097	2333	0.006
	Maternal-Paternal	0.043	0.003	0.083	2350	0.036

Abbreviations. Off, Offspring; Mat, Maternal; Pat, Paternal; *r,* Pearson's R Coefficient; CI, Confidence Interval; T, T-Value; DF, Degrees of Freedom; P, P-Value.

Table 3.1. Psychological Distress and Depression Polygenic Score (PGS) Associations

Sample	Threshold	Estimate	StdErr	t	p	n	AdjR2
Whole	X5e.08	295.475	70.112	4.214	0.000	18067.000	0.018
	X5e.07	412.376	89.783	4.593	0.000		0.018
	X5e.06	903.082	141.195	6.396	0.000		0.019
	X5e.05	1541.157	230.235	6.694	0.000		0.019
	X0.0005	3581.814	403.649	8.874	0.000		0.021
	X0.005	8825.491	801.687	11.009	0.000		0.023
	X0.05	23595.531	1830.290	12.892	0.000		0.026
	X0.5	64652.410	5087.370	12.708	0.000		0.025
	X0.75	76585.298	6117.374	12.519	0.000		0.025
	X1	85766.887	6846.635	12.527	0.000		0.025
	PGSpc1	0.778	0.063	12.292	0.000		0.025
Offspring	X5e.08	401.041	188.451	2.128	0.033	2433.000	0.005
	X5e.07	549.916	239.800	2.293	0.022		0.005
	X5e.06	738.999	375.580	1.968	0.049		0.005
	X5e.05	1544.662	626.828	2.464	0.014		0.006
	X0.0005	3689.572	1085.712	3.398	0.001		0.008
	X0.005	7655.400	2146.657	3.566	0.000		0.008
	X0.05	17154.875	5014.528	3.421	0.001		0.008
	X0.5	48284.430	14021.667	3.444	0.001		0.008
	X0.75	57913.940	16807.199	3.446	0.001		0.008
	X1	64846.188	18829.348	3.444	0.001		0.008
	PGSpc1	0.666	0.172	3.870	0.000		0.009
Maternal	X5e.08	216.720	184.852	1.172	0.241	2420.000	0.022
	X5e.07	417.673	230.515	1.812	0.070		0.022
	X5e.06	1114.083	358.172	3.110	0.002		0.025
	X5e.05	2365.224	594.371	3.979	0.000		0.027
	X0.0005	5044.169	1047.752	4.814	0.000		0.030
	X0.005	12611.830	2102.799	5.998	0.000		0.035
	X0.05	30236.538	4816.353	6.278	0.000		0.037
	X0.5	86848.917	13389.729	6.486	0.000		0.038
	X0.75	105986.962	16102.776	6.582	0.000		0.038
	X1	117921.202	18032.301	6.539	0.000		0.038
	PGSpc1	1.056	0.167	6.335	0.000		0.037
Paternal	X5e.08	73.279	163.597	0.448	0.654	2443.000	0.023
	X5e.07	383.796	209.191	1.835	0.067		0.024
	X5e.06	836.785	327.422	2.556	0.011		0.026
	X5e.05	703.520	535.185	1.315	0.189		0.024
	X0.0005	1767.583	949.604	1.861	0.063		0.024
	X0.005	3817.780	1876.669	2.034	0.042		0.025
	X0.05	14814.831	4367.236	3.392	0.001		0.028
	X0.5	44222.209	12145.404	3.641	0.000		0.028
	X0.75	49923.237	14620.446	3.415	0.001		0.028
	X1	55975.996	16349.593	3.424	0.001		0.028
	PGSpc1	0.462	0.148	3.126	0.002		0.027

Abbreviations. Threshold; Single Nucleotide Polymorphism (SNP)-Trait Association P-Value; SE, standard error; T, T-Value; P, P-Value; adjR2, adjusted R^2; n, sample size.

Table 3.2. Depression and Depression Polygenic Score (PGS) Associations

Sample	Threshold	Estimate	StdErr	t	p	n	AdjR2
Whole	X5e.08	54.639	14.829	3.685	0.000	17921.000	0.002
	X5e.07	75.434	18.985	3.973	0.000		0.002
	X5e.06	180.330	29.864	6.038	0.000		0.003
	X5e.05	251.461	48.763	5.157	0.000		0.003
	X0.0005	630.921	85.491	7.380	0.000		0.004
	X0.005	1377.238	169.942	8.104	0.000		0.005
	X0.05	3540.885	387.974	9.127	0.000		0.006
	X0.5	9790.479	1078.515	9.078	0.000		0.006
	X0.75	11648.157	1296.918	8.981	0.000		0.006
	X1	13024.519	1451.532	8.973	0.000		0.006
	PGSpc1	0.125	0.013	9.304	0.000		0.006
Offspring	X5e.08	61.472	39.149	1.570	0.116	2412.000	0.006
	X5e.07	112.484	49.864	2.256	0.024		0.008
	X5e.06	188.939	78.064	2.420	0.016		0.008
	X5e.05	262.231	130.599	2.008	0.045		0.007
	X0.0005	627.276	226.210	2.773	0.006		0.009
	X0.005	849.599	446.862	1.901	0.057		0.007
	X0.05	1678.818	1043.295	1.609	0.108		0.007
	X0.5	5417.077	2917.922	1.856	0.064		0.007
	X0.75	6551.811	3497.320	1.873	0.061		0.007
	X1	7276.736	3917.641	1.857	0.063		0.007
	PGSpc1	0.092	0.036	2.573	0.010		0.008
Maternal	X5e.08	76.896	34.571	2.224	0.026	2408.000	0.013
	X5e.07	93.438	43.156	2.165	0.030		0.013
	X5e.06	134.438	67.188	2.001	0.046		0.013
	X5e.05	193.755	111.815	1.733	0.083		0.013
	X0.0005	555.354	196.722	2.823	0.005		0.015
	X0.005	1415.084	396.080	3.573	0.000		0.017
	X0.05	3606.804	906.801	3.978	0.000		0.018
	X0.5	9544.614	2524.828	3.780	0.000		0.017
	X0.75	11959.894	3036.326	3.939	0.000		0.018
	X1	13230.442	3399.940	3.891	0.000		0.018
	PGSpc1	0.125	0.031	3.984	0.000		0.018
Paternal	X5e.08	84.577	33.437	2.529	0.011	2435.000	0.013
	X5e.07	123.075	42.790	2.876	0.004		0.014
	X5e.06	278.728	66.988	4.161	0.000		0.018
	X5e.05	241.231	109.542	2.202	0.028		0.013
	X0.0005	537.697	194.253	2.768	0.006		0.014
	X0.005	859.798	384.116	2.238	0.025		0.013
	X0.05	2578.381	895.574	2.879	0.004		0.014
	X0.5	8213.412	2488.929	3.300	0.001		0.015
	X0.75	9495.341	2995.690	3.170	0.002		0.015
	X1	10730.908	3350.053	3.203	0.001		0.015
	PGSpc1	0.112	0.030	3.699	0.000		0.016

Abbreviations. Threshold; Single Nucleotide Polymorphism (SNP)-Trait Association P-Value; SE, standard error; T, T-Value; P, P-Value; adjR2, adjusted R^2; n, sample size.

Table 3.3. Educational Attainment and Educational Attainment Polygenic Score (PGS) Associations

Sample	Threshold	Estimate	StdErr	t	p	n	AdjR2
Whole	X5e.08	848.765	47.773	17.766	0.000	18886.000	0.081
	X5e.07	1346.026	64.788	20.776	0.000		0.087
	X5e.06	2089.115	90.247	23.149	0.000		0.092
	X5e.05	3341.409	130.056	25.692	0.000		0.097
	X0.0005	5717.884	201.048	28.440	0.000		0.104
	X0.005	10638.398	340.342	31.258	0.000		0.112
	X0.05	21894.917	677.135	32.335	0.000		0.115
	X0.5	56983.613	1686.241	33.793	0.000		0.119
	X0.75	67624.071	2006.790	33.698	0.000		0.119
	X1	75292.775	2235.338	33.683	0.000		0.119
	PGSpc1	0.368	0.011	33.610	0.000		0.119
Offspring	X5e.08	688.882	120.295	5.727	0.000	2556.000	0.019
	X5e.07	1321.166	162.790	8.116	0.000		0.031
	X5e.06	1907.251	224.627	8.491	0.000		0.034
	X5e.05	3098.036	328.375	9.434	0.000		0.040
	X0.0005	5186.328	513.987	10.090	0.000		0.045
	X0.005	9515.897	858.060	11.090	0.000		0.052
	X0.05	19507.264	1694.260	11.514	0.000		0.056
	X0.5	49370.929	4207.475	11.734	0.000		0.058
	X0.75	58631.074	5009.309	11.704	0.000		0.057
	X1	65144.308	5585.959	11.662	0.000		0.057
	PGSpc1	0.334	0.028	12.026	0.000		0.060
Maternal	X5e.08	818.889	125.241	6.539	0.000	2584.000	0.073
	X5e.07	1269.028	168.424	7.535	0.000		0.078
	X5e.06	2086.106	233.794	8.923	0.000		0.086
	X5e.05	3606.506	335.784	10.741	0.000		0.099
	X0.0005	6569.774	521.027	12.609	0.000		0.113
	X0.005	12465.303	889.100	14.020	0.000		0.125
	X0.05	25287.085	1761.891	14.352	0.000		0.128
	X0.5	65972.797	4354.760	15.150	0.000		0.135
	X0.75	77883.841	5193.098	14.998	0.000		0.134
	X1	86745.597	5785.138	14.995	0.000		0.134
	PGSpc1	0.416	0.029	14.525	0.000		0.130
Paternal	X5e.08	836.795	144.511	5.791	0.000	2550.000	0.038
	X5e.07	1340.789	196.503	6.823	0.000		0.043
	X5e.06	2318.167	270.151	8.581	0.000		0.053
	X5e.05	3616.808	395.587	9.143	0.000		0.057
	X0.0005	5906.999	616.553	9.581	0.000		0.059
	X0.005	11649.834	1036.313	11.242	0.000		0.072
	X0.05	23742.365	2031.630	11.686	0.000		0.075
	X0.5	66281.762	5111.366	12.968	0.000		0.086
	X0.75	77993.285	6095.831	12.795	0.000		0.085
	X1	86982.347	6791.416	12.808	0.000		0.085
	PGSpc1	0.411	0.034	12.255	0.000		0.080

Abbreviations. Threshold; Single Nucleotide Polymorphism (SNP)-Trait Association P-Value; SE, standard error; T, T-Value; P, P-Value; adjR2, adjusted R^2; n, sample size.

Table 3.4. Height and Height Polygenic Score (PGS) Associations

Sample	Threshold	Estimate	StdErr	t	p	n	AdjR2
Whole	X5e.08	32709.796	399.844	81.806	0.000	19914	0.670
	X5e.07	38193.155	460.148	83.002	0.000		0.673
	X5e.06	46300.845	543.843	85.136	0.000		0.677
	X5e.05	57920.103	666.251	86.934	0.000		0.681
	X0.0005	76543.665	859.618	89.044	0.000		0.685
	X0.005	110630.511	1215.567	91.011	0.000		0.689
	X0.05	182531.880	1994.708	91.508	0.000		0.690
	X0.5	358559.912	3927.376	91.298	0.000		0.690
	X0.75	406378.214	4455.279	91.213	0.000		0.689
	X1	439181.696	4815.577	91.200	0.000		0.689
	PGSpc1	3.579	0.039	92.937	0.000		0.693
Offspring	X5e.08	34092.142	1109.467	30.728	0.000	2667	0.655
	X5e.07	39423.595	1273.825	30.949	0.000		0.656
	X5e.06	47657.014	1491.387	31.955	0.000		0.662
	X5e.05	59201.538	1827.497	32.395	0.000		0.665
	X0.0005	77505.527	2369.663	32.707	0.000		0.667
	X0.005	112469.941	3368.871	33.385	0.000		0.671
	X0.05	183814.225	5527.434	33.255	0.000		0.670
	X0.5	361674.026	10931.198	33.086	0.000		0.669
	X0.75	410437.725	12394.995	33.113	0.000		0.669
	X1	443448.378	13401.777	33.089	0.000		0.669
	PGSpc1	3.641	0.106	34.197	0.000		0.675
Maternal	X5e.08	30138.733	1070.467	28.155	0.000	2664	0.286
	X5e.07	35396.205	1229.135	28.798	0.000		0.294
	X5e.06	42102.580	1449.794	29.040	0.000		0.297
	X5e.05	53145.001	1791.184	29.670	0.000		0.304
	X0.0005	70679.973	2313.858	30.546	0.000		0.314
	X0.005	103767.778	3262.874	31.803	0.000		0.329
	X0.05	174231.027	5273.611	33.038	0.000		0.343
	X0.5	341896.621	10467.295	32.663	0.000		0.339
	X0.75	386384.058	11872.170	32.545	0.000		0.338
	X1	417047.944	12840.374	32.479	0.000		0.337
	PGSpc1	3.360	0.103	32.564	0.000		0.338
Paternal	X5e.08	29952.380	1097.568	27.290	0.000	2676	0.263
	X5e.07	35379.644	1261.240	28.051	0.000		0.272
	X5e.06	44467.951	1489.598	29.852	0.000		0.294
	X5e.05	55203.333	1842.622	29.959	0.000		0.295
	X0.0005	73409.993	2411.859	30.437	0.000		0.301
	X0.005	107255.574	3436.426	31.211	0.000		0.310
	X0.05	178748.028	5610.530	31.859	0.000		0.318
	X0.5	344985.888	11084.200	31.124	0.000		0.309
	X0.75	392069.956	12568.523	31.195	0.000		0.310
	X1	423736.156	13585.146	31.191	0.000		0.310
	PGSpc1	3.417	0.108	31.746	0.000		0.316

Abbreviations. Threshold; Single Nucleotide Polymorphism (SNP)-Trait Association P-Value; SE, standard error; T, T-Value; P, P-Value; adjR2, adjusted R^2; n, sample size.

Table 4 presents results of regression models *I-IV* for each phenotype. Analyses were repeated for singleton (siblings excluded), female only and male only samples. An additional model 0 (a covariate model) was included to explore the impact of the main covariates; age and sex. Analyses conducted on samples including siblings were mixed effect regression models. Two separate R^2 values, marginal and conditional, were obtained from these analyses, with the former indicating variance attributable to fixed effects and the latter indicating variance attributable to all predictors. Table 4 is presented in separate sections for each phenotype explore; psychological distress, depression, educational attainment and height.

Table 4.1.1. Generation Scotland Psychological Distress Regression Models; Full Trio and Singleton Samples

		All Trio Sample							Singleton Sample				
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	t value	Pr(>|t|)	AdjR2
Model 1	Intercept	0.022	0.165	1276.032	0.134	0.893	0.007	0.108	-0.114	0.203	-0.561	0.575	0.005
	Offspring PGS	0.697	0.179	2083.431	3.896	0.000	0.007	0.108	0.623	0.225	2.773	0.006	0.005
Model 2	Intercept	0.023	0.165	1270.873	0.143	0.887	0.007	0.108	-0.115	0.203	-0.563	0.574	0.003
	Offspring PGS	0.675	0.242	2282.843	2.788	0.005	0.007	0.108	0.609	0.312	1.953	0.051	0.003
	Maternal PGS	0.138	0.225	1695.962	0.614	0.539	0.007	0.108	-0.015	0.281	-0.052	0.959	0.003
	Paternal PGS	-0.081	0.214	1701.877	-0.380	0.704	0.007	0.108	0.040	0.269	0.148	0.883	0.003
Model 3	Intercept	0.036	0.163	1276.476	0.223	0.824	0.024	0.114	-0.113	0.201	-0.563	0.573	0.027
	Offspring PGS	0.618	0.178	2075.371	3.479	0.001	0.024	0.114	0.517	0.223	2.319	0.021	0.027
	Maternal Phenotype	0.112	0.023	1366.218	4.959	0.000	0.024	0.114	0.121	0.028	4.396	0.000	0.027
	Paternal Phenotype	0.080	0.025	1286.474	3.180	0.002	0.024	0.114	0.106	0.031	3.404	0.001	0.027
Model 4	Intercept	0.037	0.163	1271.793	0.229	0.819	0.025	0.114	-0.115	0.201	-0.572	0.568	0.026
	Offspring PGS	0.689	0.240	2285.598	2.868	0.004	0.025	0.114	0.606	0.308	1.965	0.050	0.026
	Maternal PGS	-0.004	0.224	1689.349	-0.016	0.987	0.025	0.114	-0.167	0.279	-0.599	0.549	0.026
	Paternal PGS	-0.145	0.212	1695.438	-0.685	0.494	0.025	0.114	-0.028	0.266	-0.107	0.915	0.026
	Maternal Phenotype	0.112	0.023	1359.510	4.921	0.000	0.025	0.114	0.123	0.028	4.430	0.000	0.026
	Paternal Phenotype	0.081	0.025	1278.312	3.210	0.001	0.025	0.114	0.107	0.031	3.408	0.001	0.026

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est, beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2; adjR2, adjusted R^2 value.

Table 4.1.2. Generation Scotland Psychological Distress Regression Models; Full Trio Female and Male samples

		Male Only Sample							Female Only Sample						
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c
Model 1	**Intercept**	0.044	0.238	1002.000	0.185	0.853	0.013	0.013	0.031	0.221	879.567	0.139	0.889	0.004	0.205
	Offspring PGS	0.965	0.264	1002.000	3.663	0.000	0.013	0.013	0.512	0.240	1248.402	2.132	0.033	0.004	0.205
Model 2	**Intercept**	0.039	0.239	1000.000	0.165	0.869	0.015	0.015	0.024	0.221	880.318	0.110	0.913	0.005	0.211
	Offspring PGS	1.223	0.360	1000.000	3.396	0.001	0.015	0.015	0.207	0.326	1274.672	0.634	0.526	0.005	0.211
	Maternal PGS	-0.107	0.333	1000.000	-0.321	0.748	0.015	0.015	0.409	0.300	1086.131	1.362	0.173	0.005	0.211
	Paternal PGS	-0.417	0.304	1000.000	-1.374	0.170	0.015	0.015	0.257	0.294	1099.433	0.876	0.381	0.005	0.211
Model 3	**Intercept**	0.068	0.238	1000.000	0.286	0.775	0.021	0.021	0.033	0.217	870.660	0.154	0.877	0.032	0.206
	Offspring PGS	0.947	0.263	1000.000	3.603	0.000	0.021	0.021	0.368	0.238	1241.881	1.544	0.123	0.032	0.206
	Maternal Phenotype	0.059	0.034	1000.000	1.706	0.088	0.021	0.021	0.151	0.029	928.612	5.140	0.000	0.032	0.206
	Paternal Phenotype	0.078	0.038	1000.000	2.036	0.042	0.021	0.021	0.080	0.033	897.125	2.457	0.014	0.032	0.206
Model 4	**Intercept**	0.062	0.238	998.000	0.260	0.795	0.023	0.023	0.029	0.217	870.184	0.135	0.893	0.032	0.210
	Offspring PGS	1.260	0.359	998.000	3.509	0.000	0.023	0.023	0.194	0.323	1279.049	0.601	0.548	0.032	0.210
	Maternal PGS	-0.180	0.333	998.000	-0.542	0.588	0.023	0.023	0.206	0.298	1073.627	0.692	0.489	0.032	0.210
	Paternal PGS	-0.464	0.303	998.000	-1.529	0.127	0.023	0.023	0.174	0.290	1089.553	0.601	0.548	0.032	0.210
	Maternal Phenotype	0.059	0.035	998.000	1.712	0.087	0.023	0.023	0.149	0.030	925.258	5.028	0.000	0.032	0.210
	Paternal Phenotype	0.081	0.038	998.000	2.123	0.034	0.023	0.023	0.079	0.033	895.544	2.417	0.016	0.032	0.210

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est; beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2.

Table 4.2.1. Generation Scotland Depression Regression Models; Full Trio and Singleton Samples

		All Trio Sample							Singleton Sample				
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	t value	Pr(>|t|)	AdjR2
Model 1	**Intercept**	-0.010	0.034	1233.601	-0.294	0.769	0.003	0.080	-0.028	0.041	-0.689	0.491	0.004
	Offspring PGS	0.095	0.037	2024.728	2.588	0.010	0.003	0.080	0.111	0.045	2.459	0.014	0.004
Model 2	**Intercept**	-0.009	0.034	1224.456	-0.282	0.778	0.005	0.079	-0.028	0.041	-0.675	0.500	0.003
	Offspring PGS	0.091	0.050	2260.601	1.814	0.070	0.005	0.079	0.080	0.063	1.285	0.199	0.003
	Maternal PGS	0.062	0.046	1637.253	1.341	0.180	0.005	0.079	0.057	0.057	1.001	0.317	0.003
	Paternal PGS	-0.049	0.044	1643.043	-1.112	0.266	0.005	0.079	0.009	0.054	0.172	0.863	0.003
Model 3	**Intercept**	-0.009	0.034	1232.832	-0.282	0.778	0.008	0.083	-0.029	0.041	-0.718	0.473	0.007
	Offspring PGS	0.088	0.037	2023.843	2.384	0.017	0.008	0.083	0.103	0.045	2.283	0.023	0.007
	Maternal Phenotype	0.045	0.025	1376.111	1.840	0.066	0.008	0.083	0.042	0.029	1.448	0.148	0.007
	Paternal Phenotype	0.062	0.025	1312.997	2.425	0.015	0.008	0.083	0.063	0.030	2.063	0.039	0.007
Model 4	**Intercept**	-0.009	0.033	1223.881	-0.269	0.788	0.009	0.082	-0.029	0.041	-0.704	0.482	0.006
	Offspring PGS	0.088	0.050	2259.412	1.776	0.076	0.009	0.082	0.078	0.062	1.250	0.212	0.006
	Maternal PGS	0.055	0.046	1631.314	1.192	0.233	0.009	0.082	0.050	0.057	0.879	0.379	0.006
	Paternal PGS	-0.054	0.044	1639.208	-1.226	0.220	0.009	0.082	0.005	0.054	0.088	0.930	0.006
	Maternal Phenotype	0.042	0.025	1366.125	1.692	0.091	0.009	0.082	0.040	0.029	1.373	0.170	0.006
	Paternal Phenotype	0.064	0.025	1303.313	2.502	0.012	0.009	0.082	0.063	0.031	2.064	0.039	0.006

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est, beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2; adjR2, adjusted R^2 value.

Table 4.2.2. Generation Scotland Depression Regression Models; Full Trio Female and Male samples

		Male Only Sample							Female Only Sample						
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c
Model 1	**Intercept**	0.001	0.052	985.000	0.027	0.979	0.005	0.005	-0.016	0.043	815.521	-0.369	0.712	0.002	0.147
	Offspring PGS	0.132	0.057	985.000	2.300	0.022	0.005	0.005	0.068	0.047	1202.245	1.444	0.149	0.002	0.147
Model 2	**Intercept**	0.002	0.052	983.000	0.029	0.977	0.006	0.006	-0.016	0.043	812.117	-0.373	0.709	0.006	0.149
	Offspring PGS	0.144	0.078	983.000	1.846	0.065	0.006	0.006	0.045	0.065	1269.872	0.698	0.486	0.006	0.149
	Maternal PGS	0.026	0.072	983.000	0.361	0.718	0.006	0.006	0.099	0.059	1028.477	1.675	0.094	0.006	0.149
	Paternal PGS	-0.048	0.066	983.000	-0.729	0.466	0.006	0.006	-0.048	0.058	1041.973	-0.842	0.400	0.006	0.149
Model 3	**Intercept**	0.002	0.052	983.000	0.031	0.975	0.006	0.006	-0.016	0.043	816.605	-0.372	0.710	0.014	0.151
	Offspring PGS	0.131	0.057	983.000	2.286	0.022	0.006	0.006	0.048	0.047	1206.798	1.010	0.313	0.014	0.151
	Maternal Phenotype	0.036	0.039	983.000	0.930	0.353	0.006	0.006	0.058	0.031	977.735	1.884	0.060	0.014	0.151
	Paternal Phenotype	-0.002	0.041	983.000	-0.040	0.968	0.006	0.006	0.106	0.032	756.997	3.298	0.001	0.014	0.151
Model 4	**Intercept**	0.002	0.052	981.000	0.032	0.975	0.007	0.007	-0.016	0.043	811.534	-0.369	0.712	0.017	0.152
	Offspring PGS	0.147	0.078	981.000	1.875	0.061	0.007	0.007	0.038	0.064	1269.541	0.585	0.559	0.017	0.152
	Maternal PGS	0.021	0.073	981.000	0.291	0.771	0.007	0.007	0.081	0.059	1023.870	1.387	0.166	0.017	0.152
	Paternal PGS	-0.050	0.066	981.000	-0.756	0.450	0.007	0.007	-0.058	0.057	1042.331	-1.019	0.309	0.017	0.152
	Maternal Phenotype	0.036	0.039	981.000	0.905	0.366	0.007	0.007	0.052	0.031	972.394	1.684	0.093	0.017	0.152
	Paternal Phenotype	0.001	0.041	981.000	0.014	0.989	0.007	0.007	0.107	0.032	751.968	3.317	0.001	0.017	0.152

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est; beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2.

Table 4.3.1. Generation Scotland Educational Attainment Regression Models; Full Trio and Singleton Samples

		All Trio Sample							Singleton Sample				
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	t value	Pr(>|t|)	AdjR2
Model 1	Intercept	0.001	0.030	1296.212	0.048	0.962	0.038	0.257	-0.024	0.036	-0.676	0.499	0.032
	Offspring PGS	0.282	0.030	2249.461	9.425	0.000	0.038	0.257	0.262	0.037	7.018	0.000	0.032
Model 2	Intercept	0.001	0.030	1291.757	0.021	0.984	0.046	0.260	-0.024	0.036	-0.683	0.495	0.036
	Offspring PGS	0.191	0.040	2164.136	4.775	0.000	0.046	0.260	0.155	0.054	2.873	0.004	0.036
	Maternal PGS	0.089	0.038	1788.640	2.338	0.020	0.046	0.260	0.087	0.047	1.865	0.062	0.036
	Paternal PGS	0.113	0.039	1756.950	2.895	0.004	0.046	0.260	0.118	0.047	2.483	0.013	0.036
Model 3	Intercept	0.002	0.029	1308.985	0.051	0.959	0.087	0.270	-0.025	0.035	-0.712	0.476	0.087
	Offspring PGS	0.222	0.030	2257.320	7.455	0.000	0.087	0.270	0.184	0.037	4.959	0.000	0.087
	Maternal Phenotype	0.150	0.023	1346.310	6.535	0.000	0.087	0.270	0.150	0.027	5.589	0.000	0.087
	Paternal Phenotype	0.105	0.020	1319.930	5.205	0.000	0.087	0.270	0.122	0.024	5.075	0.000	0.087
Model 4	Intercept	0.001	0.029	1304.792	0.044	0.965	0.089	0.271	-0.024	0.035	-0.705	0.481	0.087
	Offspring PGS	0.187	0.039	2194.718	4.736	0.000	0.089	0.271	0.143	0.052	2.736	0.006	0.087
	Maternal PGS	0.015	0.038	1767.708	0.404	0.686	0.089	0.271	0.010	0.046	0.212	0.832	0.087
	Paternal PGS	0.068	0.038	1757.924	1.772	0.077	0.089	0.271	0.073	0.046	1.571	0.117	0.087
	Maternal Phenotype	0.148	0.023	1329.814	6.388	0.000	0.089	0.271	0.150	0.027	5.507	0.000	0.087
	Paternal Phenotype	0.101	0.020	1306.047	4.967	0.000	0.089	0.271	0.118	0.024	4.911	0.000	0.087

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est, beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2; adjR2, adjusted R^2 value.

Table 4.3.2. Generation Scotland Educational Attainment Regression Models; Full Trio Female and Male samples

		Male Only Sample							Female Only Sample						
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c
Model 1	Intercept	0.013	0.045	765.420	0.294	0.769	0.038	0.284	0.002	0.039	908.954	0.063	0.950	0.048	0.317
	Offspring PGS	0.282	0.045	1017.755	6.287	0.000	0.038	0.284	0.319	0.039	1300.398	8.087	0.000	0.048	0.317
Model 2	Intercept	0.009	0.045	763.842	0.203	0.839	0.045	0.284	0.002	0.038	905.635	0.056	0.955	0.057	0.321
	Offspring PGS	0.192	0.061	944.089	3.120	0.002	0.045	0.284	0.217	0.054	1232.777	4.056	0.000	0.057	0.321
	Maternal PGS	0.050	0.057	939.433	0.869	0.385	0.045	0.284	0.125	0.050	1115.193	2.528	0.012	0.057	0.321
	Paternal PGS	0.151	0.059	915.777	2.537	0.011	0.045	0.284	0.087	0.050	1147.758	1.748	0.081	0.057	0.321
Model 3	Intercept	0.009	0.044	768.592	0.212	0.833	0.106	0.297	0.003	0.038	913.313	0.083	0.934	0.080	0.326
	Offspring PGS	0.204	0.044	1015.431	4.606	0.000	0.106	0.297	0.267	0.040	1303.689	6.723	0.000	0.080	0.326
	Maternal Phenotype	0.192	0.034	746.753	5.585	0.000	0.106	0.297	0.118	0.029	938.561	4.036	0.000	0.080	0.326
	Paternal Phenotype	0.120	0.030	731.700	3.949	0.000	0.106	0.297	0.087	0.026	882.240	3.341	0.001	0.080	0.326
Model 4	Intercept	0.008	0.043	764.495	0.174	0.862	0.110	0.296	0.003	0.038	909.361	0.079	0.937	0.083	0.327
	Offspring PGS	0.188	0.060	963.986	3.145	0.002	0.110	0.296	0.210	0.053	1241.473	3.960	0.000	0.083	0.327
	Maternal PGS	-0.053	0.057	932.827	-0.934	0.350	0.110	0.296	0.073	0.050	1099.958	1.476	0.140	0.083	0.327
	Paternal PGS	0.096	0.058	909.012	1.659	0.097	0.110	0.296	0.051	0.049	1141.991	1.033	0.302	0.083	0.327
	Maternal Phenotype	0.199	0.035	741.833	5.718	0.000	0.110	0.296	0.111	0.030	925.398	3.745	0.000	0.083	0.327
	Paternal Phenotype	0.116	0.031	731.534	3.773	0.000	0.110	0.296	0.083	0.026	872.165	3.181	0.002	0.083	0.327

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est; beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2.

Table 4.4.1. Generation Scotland Height Regression Models; Full Trio and Singleton Samples

		All Trio Sample							Singleton Sample				
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	t value	Pr(>|t|)	AdjR2
Model 1	**Intercept**	-0.033	0.118	1522.830	-0.278	0.781	0.306	0.618	-0.054	0.129	-0.421	0.674	0.325
	Offspring PGS	3.662	0.107	2643.517	34.077	0.000	0.306	0.618	3.738	0.135	27.625	0.000	0.325
Model 2	**Intercept**	-0.030	0.117	1521.405	-0.252	0.801	0.316	0.622	-0.054	0.129	-0.417	0.676	0.327
	Offspring PGS	3.442	0.137	2194.963	25.106	0.000	0.316	0.622	3.427	0.196	17.471	0.000	0.327
	Maternal PGS	0.396	0.147	2067.676	2.695	0.007	0.316	0.622	0.398	0.175	2.278	0.023	0.327
	Paternal PGS	0.166	0.141	1976.215	1.179	0.238	0.316	0.622	0.183	0.164	1.117	0.264	0.327
Model 3	**Intercept**	0.005	0.092	1433.575	0.055	0.956	0.514	0.610	-0.022	0.108	-0.201	0.841	0.521
	Offspring PGS	2.491	0.100	2561.336	24.883	0.000	0.514	0.610	2.391	0.126	19.020	0.000	0.521
	Maternal Phenotype	0.331	0.017	1539.822	19.036	0.000	0.514	0.610	0.335	0.020	16.435	0.000	0.521
	Paternal Phenotype	0.310	0.016	1520.743	19.293	0.000	0.514	0.610	0.304	0.019	16.021	0.000	0.521
Model 4	**Intercept**	0.008	0.088	1449.616	0.086	0.931	0.540	0.619	-0.010	0.105	-0.092	0.926	0.551
	Offspring PGS	3.430	0.122	2552.040	28.105	0.000	0.540	0.619	3.476	0.160	21.707	0.000	0.551
	Maternal PGS	-1.226	0.131	1902.353	-9.351	0.000	0.540	0.619	-1.263	0.161	-7.865	0.000	0.551
	Paternal PGS	-1.358	0.124	1818.077	-10.907	0.000	0.540	0.619	-1.321	0.150	-8.806	0.000	0.551
	Maternal Phenotype	0.393	0.019	1480.067	20.880	0.000	0.540	0.619	0.397	0.022	17.816	0.000	0.551
	Paternal Phenotype	0.387	0.017	1479.731	22.249	0.000	0.540	0.619	0.372	0.021	18.084	0.000	0.551

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est, beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2; adjR2, adjusted R^2 value.

Table 4.4.2. Generation Scotland Height Regression Models; Full Trio Female and Male samples

		Male Only Sample							Female Only Sample						
		Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c	Estimate	Std. Error	df	t value	Pr(>|t|)	R2m	R2c
Model 1	**Intercept**	-0.016	0.170	874.252	-0.093	0.926	0.304	0.608	-0.081	0.143	1083.115	-0.567	0.571	0.316	0.648
	Offspring PGS	3.769	0.169	1139.943	22.254	0.000	0.304	0.608	3.650	0.138	1495.984	26.365	0.000	0.316	0.648
Model 2	**Intercept**	-0.016	0.170	874.538	-0.092	0.927	0.309	0.610	-0.078	0.143	1079.952	-0.544	0.586	0.324	0.650
	Offspring PGS	3.507	0.233	1045.632	15.029	0.000	0.309	0.610	3.418	0.179	1274.741	19.060	0.000	0.324	0.650
	Maternal PGS	0.326	0.227	1072.541	1.440	0.150	0.309	0.610	0.435	0.180	1328.411	2.411	0.016	0.324	0.650
	Paternal PGS	0.231	0.211	1024.977	1.097	0.273	0.309	0.610	0.113	0.177	1295.464	0.637	0.524	0.324	0.650
Model 3	**Intercept**	0.005	0.138	844.669	0.035	0.972	0.507	0.589	-0.008	0.116	1045.962	-0.067	0.947	0.516	0.641
	Offspring PGS	2.345	0.157	1108.542	14.943	0.000	0.507	0.589	2.506	0.128	1474.235	19.548	0.000	0.516	0.641
	Maternal Phenotype	0.323	0.026	900.976	12.447	0.000	0.507	0.589	0.337	0.022	1100.819	15.287	0.000	0.516	0.641
	Paternal Phenotype	0.352	0.024	889.649	14.515	0.000	0.507	0.589	0.284	0.020	1088.314	13.995	0.000	0.516	0.641
Model 4	**Intercept**	0.019	0.133	818.640	0.146	0.884	0.534	0.596	-0.012	0.112	1059.947	-0.107	0.915	0.544	0.658
	Offspring PGS	3.384	0.198	1139.129	17.107	0.000	0.534	0.596	3.451	0.155	1437.585	22.224	0.000	0.544	0.658
	Maternal PGS	-1.250	0.204	994.606	-6.116	0.000	0.534	0.596	-1.196	0.165	1250.792	-7.242	0.000	0.544	0.658
	Paternal PGS	-1.320	0.186	884.536	-7.109	0.000	0.534	0.596	-1.402	0.162	1236.299	-8.647	0.000	0.544	0.658
	Maternal Phenotype	0.383	0.028	854.563	13.630	0.000	0.534	0.596	0.397	0.024	1078.352	16.440	0.000	0.544	0.658
	Paternal Phenotype	0.416	0.026	822.976	16.082	0.000	0.534	0.596	0.366	0.022	1096.502	16.394	0.000	0.544	0.658

Abbreviations. o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic scores; pheno, phenotype; Est; beta coefficient estimate; SE, standard error; DF, degrees of freedom; T, T-Value, P, P-Value; R2m, marginal R^2; R2c, conditional R^2.

Table 5 presents log-likelihood comparisons of regression models *I*-*IV* for each phenotype (function ’anova()’ in R). Results included log-likelihood values and chi-squared values for mixed effect models. False discovery rate corrected P-Values were provided for each analyses. Table 5 is presented in separate sections for each sample explored; full trios, singletons, female only and male only samples.

Table 5. Log Likelihood Comparison of Regression Models I-IV in Full Biological Parent-Offspring (Trio) Sample

		Whole Trio Sample							
	Comparison	**AIC**	**BIC**	**logLik**	**deviance**	**Chisq**	**DF**	**Pr(>Chisq)**	**FDR**
HEIGHT	**Model 1 vs Model 2**	15979.02	16002.54	-7985.51	15971.02	NA	NA	NA	NA
		15975.15	16010.44	-7981.57	15963.15	7.87	2.00	0.02	0.04
	Model 1 vs Model 3	15979.02	16002.54	-7985.51	15971.02	NA	NA	NA	NA
		15279.849	15315.138	-7633.924	15267.849	703.166	2.000	0.000	0.000
	Model 1 vs Model 4	15979.015	16002.541	-7985.508	15971.015	NA	NA	NA	NA
		15120.398	15167.451	-7552.199	15104.398	866.617	4.000	0.000	0.000
	Model 2 vs Model 4	15975.148	16010.438	-7981.574	15963.148	NA	NA	NA	NA
		15120.398	15167.451	-7552.199	15104.398	858.750	2.000	0.000	0.000
	Model 3 vs Model 4	15279.849	15315.138	-7633.924	15267.849	NA	NA	NA	NA
		15120.398	15167.451	-7552.199	15104.398	163.451	2.000	0.000	0.000
EDUCATIONAL ATTAINMENT	**Model 1 vs Model 2**	8081.693	8104.760	-4036.847	8073.693	NA	NA	NA	NA
		8073.754	8108.355	-4030.877	8061.754	11.939	2.000	0.003	0.006
	Model 1 vs Model 3	8081.693	8104.760	-4036.847	8073.693	NA	NA	NA	NA
		7983.457	8018.058	-3985.728	7971.457	102.237	2.000	0.000	0.000
	Model 1 vs Model 4	8081.693	8104.760	-4036.847	8073.693	NA	NA	NA	NA
		7984.300	8030.435	-3984.150	7968.300	105.393	4.000	0.000	0.000
	Model 2 vs Model 4	8073.754	8108.355	-4030.877	8061.754	NA	NA	NA	NA
		7984.300	8030.435	-3984.150	7968.300	93.454	2.000	0.000	0.000
	Model 3 vs Model 4	7983.457	8018.058	-3985.728	7971.457	NA	NA	NA	NA
		7984.300	8030.435	-3984.150	7968.300	3.156	2.000	0.206	0.413
PSYCHOLOGICAL DISTRESS	**Model 1 vs Model 2**	15946.407	15969.395	-7969.203	15938.407	NA	NA	NA	NA
		15949.667	15984.150	-7968.833	15937.667	0.740	2.000	0.691	1.000
	Model 1 vs Model 3	15946.407	15969.395	-7969.203	15938.407	NA	NA	NA	NA
		15912.005	15946.488	-7950.003	15900.005	38.401	2.000	0.000	0.000
	Model 1 vs Model 4	15946.407	15969.395	-7969.203	15938.407	NA	NA	NA	NA
		15915.488	15961.465	-7949.744	15899.488	38.919	4.000	0.000	0.000
	Model 2 vs Model 4	15949.667	15984.150	-7968.833	15937.667	NA	NA	NA	NA
		15915.488	15961.465	-7949.744	15899.488	38.178	2.000	0.000	0.000
	Model 3 vs Model 4	15912.005	15946.488	-7950.003	15900.005	NA	NA	NA	NA
		15915.488	15961.465	-7949.744	15899.488	0.517	2.000	0.772	1.000
DEPRESSION	**Model 1 vs Model 2**	8473.552	8496.480	-4232.776	8465.552	NA	NA	NA	NA
		8473.111	8507.503	-4230.556	8461.111	4.441	2.000	0.109	0.233
	Model 1 vs Model 3	8473.552	8496.480	-4232.776	8465.552	NA	NA	NA	NA
		8467.831	8502.223	-4227.916	8455.831	9.721	2.000	0.008	0.028
	Model 1 vs Model 4	8473.552	8496.480	-4232.776	8465.552	NA	NA	NA	NA
		8467.533	8513.388	-4225.766	8451.533	14.020	4.000	0.007	0.028
	Model 2 vs Model 4	8473.111	8507.503	-4230.556	8461.111	NA	NA	NA	NA
		8467.533	8513.388	-4225.766	8451.533	9.578	2.000	0.008	0.028
	Model 3 vs Model 4	8467.831	8502.223	-4227.916	8455.831	NA	NA	NA	NA
		8467.533	8513.388	-4225.766	8451.533	4.299	2.000	0.117	0.233

Abbreviations. logLik, Log Likelihood value; chisq, χ^2 value; DF, degrees of freedom; P, P-Value; FDR, false discovery rate corrected P-Value.

Table 6 presents results of the simple and extended pathway models for each phenotype. The pathway models were repeated where all path coefficients were freely estimated. This was specifically to check that paths between parent and offspring polygenic scores (PGS) were ~0.50 as expected. Similar to regression models, pathway models were repeated using singleton, female only and male only samples. Additionally, duo models (maternal-offspring and paternal-offspring models) were conducted to explore whether the exclusion of one parent had an impact on results. Table 6 is presented in separate sections for each repeated analysis outlined above.

Table 6.1.1. Generation Scotland Psychological Distress Pathway Models

				All Paths Freely Estimated						Singleton Sample					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.149	0.219	0.495	-0.280	0.578	32638.743	-0.015	0.282	0.959	-0.567	0.538	19972.774
	OPHENO	PPGS	b	-0.103	0.208	0.621	-0.510	0.305	32638.743	0.040	0.269	0.883	-0.487	0.566	19972.774
Simple	OPGS	MPGS	i	0.505	0.015	0.000	0.474	0.535	32638.743						
	OPGS	PGS	h	0.484	0.015	0.000	0.456	0.513	32638.743						
	OPHENO	OPGS	a	0.678	0.243	0.005	0.202	1.155	32638.743	0.609	0.311	0.050	-0.001	1.220	19972.774
	OPHENO	MPGS	c	0.008	0.218	0.969	-0.419	0.435	63503.257	-0.167	0.279	0.550	-0.715	0.380	38852.000
	OPHENO	PPGS	b	-0.162	0.206	0.432	-0.566	0.242	63503.257	-0.028	0.266	0.915	-0.550	0.493	38852.000
	OPGS	MPGS	i	0.505	0.015	0.000	0.474	0.535	63503.257						
	OPGS	PGS	h	0.484	0.015	0.000	0.456	0.513	63503.257						
Extended	OPHENO	OPGS	a	0.693	0.241	0.004	0.220	1.165	63503.257	0.606	0.308	0.049	0.003	1.209	38852.000
	PPHENO	PPGS	d	0.453	0.146	0.002	0.167	0.738	63503.257	0.535	0.184	0.004	0.174	0.896	38852.000
	MPHENO	MPGS	e	0.922	0.170	0.000	0.589	1.255	63503.257	0.903	0.224	0.000	0.465	1.341	38852.000
	OPHENO	PPHENO	f	0.077	0.024	0.001	0.030	0.125	63503.257	0.107	0.031	0.001	0.046	0.168	38852.000
	OPHENO	MPHENO	g	0.113	0.022	0.000	0.070	0.156	63503.257	0.123	0.028	0.000	0.068	0.177	38852.000
Model	**IV**	**DV**	**Label**	**Female Only Sample**						**Male Only Sample**					
	OPHENO	MPGS	c	0.360	0.290	0.215	-0.209	0.930	18406.595	-0.107	0.330	0.746	-0.754	0.540	14232.508
	OPHENO	PPGS	b	0.170	0.286	0.553	-0.391	0.731	18406.595	-0.417	0.308	0.176	-1.022	0.187	14232.508
Simple	OPGS	MPGS	i												
	OPGS	PGS	h												
	OPHENO	OPGS	a	0.225	0.330	0.496	-0.422	0.871	18406.595	1.223	0.359	0.001	0.519	1.927	14232.508
	OPHENO	MPGS	c	0.163	0.288	0.572	-0.402	0.727	36001.178	-0.180	0.330	0.584	-0.826	0.466	27501.819
	OPHENO	PPGS	b	0.094	0.283	0.739	-0.460	0.648	36001.178	-0.464	0.307	0.131	-1.066	0.139	27501.819
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	0.207	0.325	0.524	-0.430	0.845	36001.178	1.260	0.358	0.000	0.559	1.961	27501.819
	PPHENO	PPGS	d	0.534	0.204	0.009	0.135	0.933	36001.178	0.360	0.207	0.082	-0.046	0.766	27501.819
	MPHENO	MPGS	e	1.085	0.229	0.000	0.636	1.534	36001.178	0.703	0.253	0.005	0.207	1.199	27501.819
	OPHENO	PPHENO	f	0.078	0.031	0.013	0.017	0.139	36001.178	0.081	0.038	0.033	0.006	0.156	27501.819
	OPHENO	MPHENO	g	0.150	0.028	0.000	0.095	0.206	36001.178	0.059	0.034	0.086	-0.008	0.126	27501.819

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here, paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Table 6.1.2. Generation Scotland Psychological Distress Pathway Models

				Maternal Duo						Paternal Duo					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.185	0.207	0.371	-0.220	0.591	21414.337	-0.150	0.199	0.450	-0.539	0.239	21393.577
	OPHENO	PPGS	b												
Simple	OPGS	MPGS	i												
	OPGS	PGS	h							0.771	0.201	0.000	0.377	1.166	21393.577
	OPHENO	OPGS	a	0.610	0.200	0.002	0.217	1.003	21414.337						
	OPHENO	MPGS	c	0.070	0.207	0.735	-0.335	0.475	37076.790						
	OPHENO	PPGS	b							-0.184	0.198	0.354	-0.572	0.205	36618.704
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	0.610	0.199	0.002	0.220	1.000	37076.790	0.752	0.201	0.000	0.359	1.145	36618.704
	PPHENO	PPGS	d							0.467	0.147	0.001	0.179	0.754	36618.704
	MPHENO	MPGS	e	0.962	0.171	0.000	0.626	1.297	37076.790						
	OPHENO	PPHENO	f							0.092	0.024	0.000	0.044	0.139	36618.704
	OPHENO	MPHENO	g	0.120	0.022	0.000	0.077	0.163	37076.790						

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here, paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Table 6.2.1. Generation Scotland Depression Pathway Models

				All Paths Freely Estimated						Singleton Sample					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.061	0.045	0.178	-0.028	0.149	24909.272	0.057	0.057	0.316	-0.054	0.168	15215.697
	OPHENO	PPGS	b	-0.053	0.043	0.211	-0.137	0.030	24909.272	0.009	0.054	0.863	-0.096	0.115	15215.697
Simple	OPGS	MPGS	i	0.509	0.016	0.000	0.478	0.539	24909.272						
	OPGS	PGS	h	0.484	0.015	0.000	0.455	0.513	24909.272						
	OPHENO	OPGS	a	0.093	0.050	0.064	-0.005	0.191	24909.272	0.080	0.062	0.198	-0.042	0.203	15215.697
	OPHENO	MPGS	c	0.054	0.045	0.231	-0.034	0.143	40478.210	0.050	0.057	0.378	-0.061	0.161	24898.561
	OPHENO	PPGS	b	-0.058	0.043	0.175	-0.142	0.026	40478.210	0.005	0.054	0.930	-0.101	0.110	24898.561
	OPGS	MPGS	i	0.509	0.016	0.000	0.478	0.539	40478.210						
	OPGS	PGS	h	0.484	0.015	0.000	0.455	0.513	40478.210						
Extended	OPHENO	OPGS	a	0.091	0.050	0.070	-0.007	0.188	40478.210	0.078	0.062	0.210	-0.044	0.200	24898.561
	PPHENO	PPGS	d	0.106	0.030	0.000	0.047	0.164	40478.210	0.106	0.039	0.006	0.031	0.182	24898.561
	MPHENO	MPGS	e	0.128	0.033	0.000	0.064	0.192	40478.210	0.130	0.043	0.003	0.045	0.216	24898.561
	OPHENO	PPHENO	f	0.062	0.025	0.012	0.014	0.110	40478.210	0.063	0.030	0.038	0.003	0.123	24898.561
	OPHENO	MPHENO	g	0.043	0.024	0.074	-0.004	0.090	40478.210	0.040	0.029	0.169	-0.017	0.097	24898.561
Model	**IV**	**DV**	**Label**	**Female Only Sample**						**Male Only Sample**					
	OPHENO	MPGS	c	0.089	0.057	0.119	-0.023	0.201	13941.200	0.026	0.072	0.715	-0.115	0.167	10962.730
	OPHENO	PPGS	b	-0.058	0.056	0.302	-0.169	0.052	13941.200	-0.048	0.067	0.473	-0.180	0.083	10962.730
Simple	OPGS	MPGS	i												
	OPGS	PGS	h												
	OPHENO	OPGS	a	0.051	0.065	0.435	-0.077	0.178	13941.200	0.144	0.078	0.064	-0.008	0.297	10962.730
	OPHENO	MPGS	c	0.073	0.057	0.201	-0.039	0.185	22861.648	0.021	0.072	0.769	-0.120	0.162	17610.596
	OPHENO	PPGS	b	-0.068	0.056	0.225	-0.179	0.042	22861.648	-0.050	0.067	0.456	-0.181	0.081	17610.596
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	0.042	0.065	0.513	-0.084	0.169	22861.648	0.147	0.078	0.060	-0.006	0.300	17610.596
	PPHENO	PPGS	d	0.160	0.042	0.000	0.078	0.241	22861.648	0.042	0.043	0.328	-0.042	0.125	17610.596
	MPHENO	MPGS	e	0.141	0.044	0.001	0.056	0.226	22861.648	0.113	0.049	0.022	0.017	0.209	17610.596
	OPHENO	PPHENO	f	0.107	0.031	0.001	0.046	0.167	22861.648	0.001	0.041	0.989	-0.079	0.080	17610.596
	OPHENO	MPHENO	g	0.049	0.030	0.103	-0.010	0.107	22861.648	0.036	0.039	0.364	-0.041	0.112	17610.596

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here, paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Table 6.2.2. Generation Scotland Depression Pathway Models

				Maternal Duo						Paternal Duo					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.080	0.043	0.061	-0.004	0.163	13851.696						
	OPHENO	PPGS	b	0.057	0.041	0.164	-0.023	0.138	13851.696	-0.073	0.041	0.075	-0.153	0.007	13847.147
Simple	OPGS	MPGS	i												
	OPGS	PGS	h												
	OPHENO	OPGS	a	0.057	0.041	0.164	-0.023	0.138	13851.696	0.131	0.041	0.002	0.050	0.212	13847.147
	OPHENO	MPGS	c	0.074	0.043	0.085	-0.010	0.157	21707.918						
	OPHENO	PPGS	b							-0.078	0.041	0.056	-0.158	0.002	21562.602
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	0.058	0.041	0.159	-0.023	0.139	21707.918	0.128	0.041	0.002	0.047	0.209	21562.602
	PPHENO	PPGS	d							0.105	0.030	0.000	0.046	0.163	21562.602
	MPHENO	MPGS	e	0.130	0.033	0.000	0.066	0.194	21707.918						
	OPHENO	PPHENO	f							0.064	0.025	0.009	0.016	0.113	21562.602
	OPHENO	MPHENO	g	0.046	0.024	0.055	-0.001	0.093	21707.918						

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here, paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Table 6.3.1. Generation Scotland Educational Attainment Pathway Models

				All Paths Freely Estimated						Singleton Sample					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.095	0.036	0.009	0.024	0.166	25491.715	0.087	0.047	0.061	-0.004	0.178	15584.390
	OPHENO	PPGS	b	0.117	0.037	0.002	0.045	0.189	25491.715	0.118	0.048	0.014	0.024	0.211	15584.390
Simple	OPGS	MPGS	i	0.493	0.015	0.000	0.464	0.523	25491.715						
	OPGS	PGS	h	0.486	0.015	0.000	0.456	0.517	25491.715						
	OPHENO	OPGS	a	0.199	0.041	0.000	0.118	0.280	25491.715	0.155	0.054	0.004	0.049	0.260	15584.390
	OPHENO	MPGS	c	0.024	0.036	0.497	-0.046	0.095	41966.263	0.010	0.046	0.830	-0.080	0.100	25679.412
	OPHENO	PPGS	b	0.074	0.036	0.043	0.002	0.145	41966.263	0.073	0.047	0.118	-0.019	0.165	25679.412
	OPGS	MPGS	i	0.493	0.015	0.000	0.464	0.523	41966.263						
	OPGS	PGS	h	0.486	0.015	0.000	0.456	0.517	41966.263						
Extended	OPHENO	OPGS	a	0.192	0.040	0.000	0.113	0.271	41966.263	0.143	0.052	0.006	0.041	0.246	25679.412
	PPHENO	PPGS	d	0.237	0.033	0.000	0.172	0.302	41966.263	0.233	0.042	0.000	0.150	0.316	25679.412
	MPHENO	MPGS	e	0.285	0.028	0.000	0.230	0.340	41966.263	0.271	0.036	0.000	0.201	0.341	25679.412
	OPHENO	PPHENO	f	0.098	0.019	0.000	0.061	0.134	41966.263	0.118	0.024	0.000	0.072	0.165	25679.412
	OPHENO	MPHENO	g	0.144	0.021	0.000	0.102	0.186	41966.263	0.150	0.027	0.000	0.097	0.203	25679.412
Model	**IV**	**DV**	**Label**	**Female Only Sample**						**Male Only Sample**					
	OPHENO	MPGS	c	0.130	0.048	0.007	0.036	0.225	14231.544	0.052	0.056	0.349	-0.057	0.161	11261.232
	OPHENO	PPGS	b	0.088	0.048	0.069	-0.007	0.183	14231.544	0.153	0.058	0.008	0.039	0.267	11261.232
Simple	OPGS	MPGS	i												
	OPGS	PGS	h												
	OPHENO	OPGS	a	0.206	0.055	0.000	0.099	0.314	14231.544	0.193	0.063	0.002	0.070	0.316	11261.232
	OPHENO	MPGS	c	0.082	0.048	0.089	-0.012	0.177	23564.413	-0.052	0.055	0.344	-0.159	0.055	18408.233
	OPHENO	PPGS	b	0.054	0.048	0.264	-0.040	0.148	23564.413	0.099	0.057	0.079	-0.011	0.210	18408.233
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	0.198	0.054	0.000	0.092	0.304	23564.413	0.189	0.060	0.002	0.071	0.308	18408.233
	PPHENO	PPGS	d	0.213	0.044	0.000	0.127	0.300	23564.413	0.270	0.051	0.000	0.171	0.369	18408.233
	MPHENO	MPGS	e	0.277	0.038	0.000	0.201	0.352	23564.413	0.293	0.041	0.000	0.212	0.374	18408.233
	OPHENO	PPHENO	f	0.082	0.024	0.001	0.034	0.129	23564.413	0.118	0.029	0.000	0.061	0.174	18408.233
	OPHENO	MPHENO	g	0.107	0.028	0.000	0.053	0.162	23564.413	0.193	0.033	0.000	0.128	0.258	18408.233

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here all models have paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Table 6.3.2. Generation Scotland Educational Attainment Pathway Models

				Maternal Duo						Paternal Duo					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.070	0.035	0.042	0.002	0.138	13791.861						
	OPHENO	PPGS	b							0.096	0.035	0.006	0.027	0.165	13849.532
Simple	OPGS	MPGS	i												
	OPGS	PGS	h												
	OPHENO	OPGS	a	0.270	0.035	0.000	0.202	0.338	13791.861	0.260	0.034	0.000	0.193	0.327	13849.532
	OPHENO	MPGS	c	0.017	0.035	0.624	-0.051	0.085	21796.310						
	OPHENO	PPGS	b							0.068	0.035	0.052	-0.001	0.137	22519.100
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	0.248	0.034	0.000	0.182	0.315	21796.310	0.235	0.034	0.000	0.169	0.301	22519.100
	PPHENO	PPGS	d							0.293	0.035	0.000	0.225	0.362	22519.100
	MPHENO	MPGS	e	0.357	0.030	0.000	0.299	0.415	21796.310						
	OPHENO	PPHENO	f							0.141	0.018	0.000	0.106	0.176	22519.100
	OPHENO	MPHENO	g	0.183	0.020	0.000	0.143	0.223	21796.310						

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here all models have paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Table 6.4.1. Generation Scotland Height Pathway Models

				All Paths Freely Estimated						Singleton Sample					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.417	0.136	0.002	0.150	0.685	35712.723	0.398	0.170	0.019	0.064	0.731	21372.173
	OPHENO	PPGS	b	0.190	0.129	0.140	-0.063	0.442	35712.723	0.183	0.168	0.276	-0.146	0.513	21372.173
Simple	OPGS	MPGS	i	0.516	0.014	0.000	0.488	0.544	35712.723						
	OPGS	PGS	h	0.469	0.014	0.000	0.442	0.496	35712.723						
	OPHENO	OPGS	a	3.336	0.152	0.000	3.038	3.633	35712.723	3.427	0.195	0.000	3.044	3.810	21372.173
	OPHENO	MPGS	c	-1.217	0.124	0.000	-1.460	-0.973	66633.530	-1.263	0.156	0.000	-1.569	-0.957	39909.318
	OPHENO	PPGS	b	-1.344	0.117	0.000	-1.574	-1.114	66633.530	-1.321	0.152	0.000	-1.619	-1.022	39909.318
	OPGS	MPGS	i	0.516	0.014	0.000	0.488	0.544	66633.530						
	OPGS	PGS	h	0.469	0.014	0.000	0.442	0.496	66633.530						
Extended	OPHENO	OPGS	a	3.402	0.124	0.000	3.159	3.645	66633.530	3.476	0.159	0.000	3.163	3.788	39909.318
	PPHENO	PPGS	d	3.263	0.106	0.000	3.055	3.470	66633.530	3.277	0.138	0.000	3.008	3.547	39909.318
	MPHENO	MPGS	e	3.172	0.101	0.000	2.973	3.370	66633.530	3.249	0.129	0.000	2.996	3.501	39909.318
	OPHENO	PPHENO	f	0.386	0.016	0.000	0.355	0.418	66633.530	0.372	0.020	0.000	0.332	0.411	39909.318
	OPHENO	MPHENO	g	0.392	0.017	0.000	0.358	0.426	66633.530	0.397	0.022	0.000	0.353	0.440	39909.318
Model	**IV**	**DV**	**Label**	**Female Only Sample**						**Male Only Sample**					
	OPHENO	MPGS	c	0.502	0.172	0.003	0.165	0.838	20123.947	0.298	0.216	0.168	-0.126	0.722	15594.184
	OPHENO	PPGS	b	0.180	0.171	0.293	-0.155	0.515	20123.947	0.196	0.205	0.340	-0.206	0.597	15594.184
Simple	OPGS	MPGS	i												
	OPGS	PGS	h												
	OPHENO	OPGS	a	3.252	0.194	0.000	2.873	3.632	20123.947	3.450	0.242	0.000	2.976	3.924	15594.184
	OPHENO	MPGS	c	-1.178	0.157	0.000	-1.487	-0.870	37553.428	-1.264	0.196	0.000	-1.649	-0.879	29093.020
	OPHENO	PPGS	b	-1.357	0.156	0.000	-1.663	-1.051	37553.428	-1.316	0.185	0.000	-1.678	-0.954	29093.020
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	3.409	0.158	0.000	3.099	3.718	37553.428	3.372	0.197	0.000	2.985	3.759	29093.020
	PPHENO	PPGS	d	3.359	0.142	0.000	3.082	3.637	37553.428	3.147	0.159	0.000	2.834	3.459	29093.020
	MPHENO	MPGS	e	3.181	0.131	0.000	2.923	3.438	37553.428	3.159	0.158	0.000	2.849	3.469	29093.020
	OPHENO	PPHENO	f	0.363	0.021	0.000	0.323	0.404	37553.428	0.416	0.025	0.000	0.367	0.464	29093.020
	OPHENO	MPHENO	g	0.400	0.023	0.000	0.356	0.444	37553.428	0.382	0.027	0.000	0.329	0.435	29093.020

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here, paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Table 6.4.2. Generation Scotland Height Pathway Models

				Maternal Duo						Paternal Duo					
Model	**IV**	**DV**	**Label**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**	**Estimate**	**Std Error**	**P-Value**	**Lower CI**	**Upper CI**	**AIC**
	OPHENO	MPGS	c	0.365	0.128	0.004	0.114	0.616	22537.423						
	OPHENO	PPGS	b							0.087	0.124	0.483	-0.156	0.330	22627.794
Simple	OPGS	MPGS	i												
	OPGS	PGS	h												
	OPHENO	OPGS	a	3.460	0.126	0.000	3.212	3.707	22537.423	3.603	0.124	0.000	3.359	3.847	22627.794
	OPHENO	MPGS	c	-1.120	0.131	0.000	-1.377	-0.863	37799.210						
	OPHENO	PPGS	b	3.307	0.114	0.000	3.084	3.529	37799.210	-1.356	0.124	0.000	-1.600	-1.112	38272.723
	OPGS	MPGS	i												
	OPGS	PGS	h												
Extended	OPHENO	OPGS	a	3.307	0.114	0.000	3.084	3.529	37799.210	3.396	0.111	0.000	3.180	3.613	38272.723
	PPHENO	PPGS	d							3.385	0.108	0.000	3.174	3.597	38272.723
	MPHENO	MPGS	e	3.322	0.103	0.000	3.119	3.524	37799.210						
	OPHENO	PPHENO	f							0.457	0.017	0.000	0.424	0.491	38272.723
	OPHENO	MPHENO	g	0.473	0.019	0.000	0.436	0.510	37799.210						

Abbreviations. DV, dependent variable; IV, independent variable; o,m,p, offspring, maternal, paternal, respectively; PGS, polygenic score; pheno, phenotype; Est, estimate; SE, standard error; P, P-Value; LCI, lower confidence interval; UCI, upper confidence interval. Here, paths h and i fixed to 0.5, these represent associations between parent PGSs and offspring PGSs.

Figure 1 presents β coefficient boxplots of parental genetic nurturing paths highlighted in **chapter 2 Figure 5**. Regression and pathway models were conducted using data simulated with the **chapter 2 simulations I** strategy. Figure 1 is presented in separate sections for regression and pathway models as well as maternal and paternal genetic nurturing effects. This figures shows that when the PGSs were perfect measures of the genetic variance, i.e. tags the entirety of trait genetic variance, the estimated coefficients were as expected, ~0. In all other scenarios, extended model estimates for path *b* were always negative and estimates for *f* were always positive. These effects were exacerbated as trait heritability increases and variance captured by PGSs decreases.

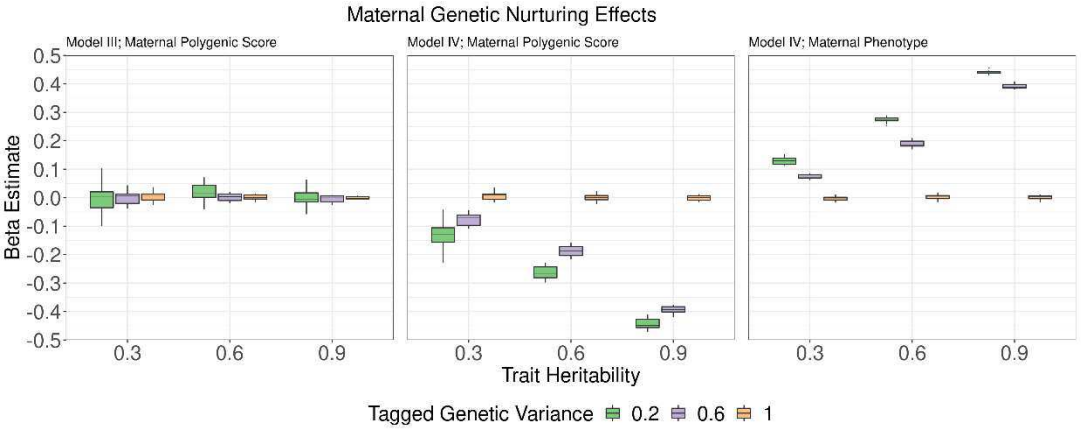


Figure 1.1. Regression Model Maternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of simulated trait heritability; y- axis presents the beta coefficients of genetic nurturing effects. The left and middle plots presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal PGS from regression model *III* and *IV*, respectively. The right plot presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal phenotype regression model *IV*.

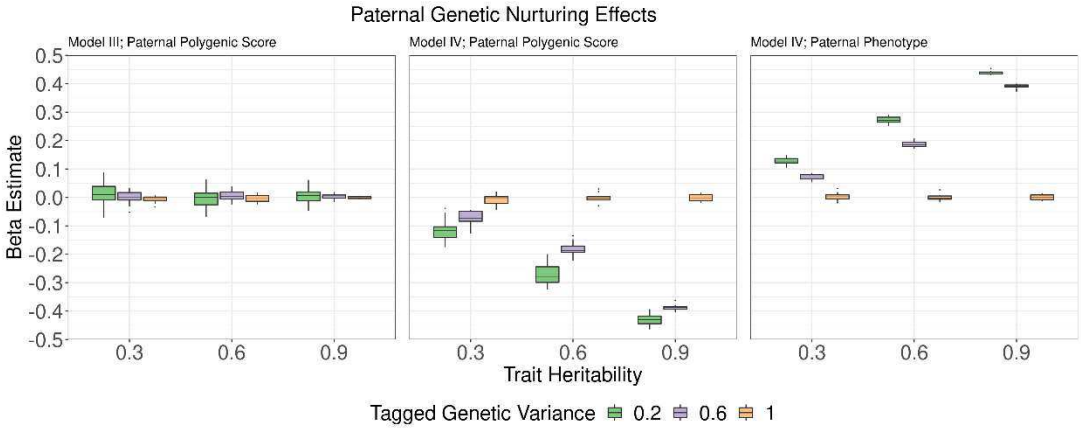


Figure 1.2. Regression Model Paternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of simulated trait heritability; y- axis presents the beta coefficients of genetic nurturing effects. The left and middle plots presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal PGS from regression model *III* and *IV*, respectively. The right plot presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal phenotype regression model *IV*.

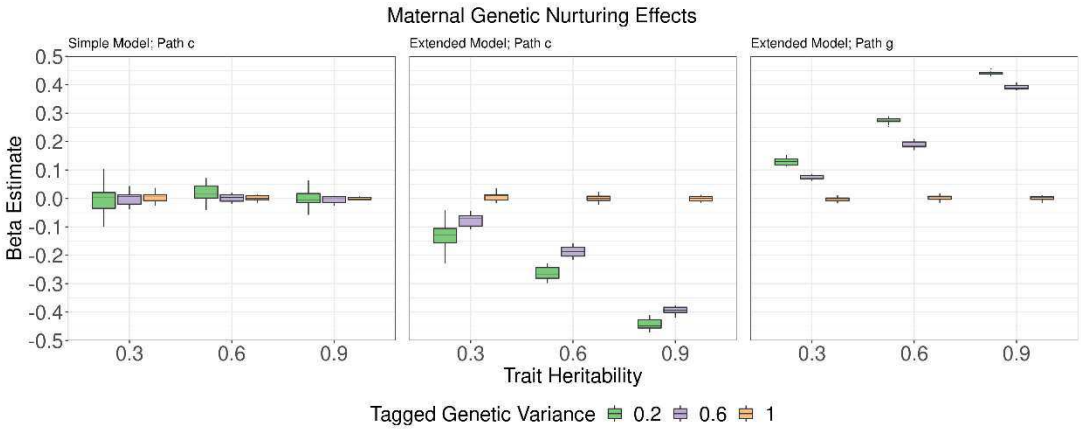


Figure 1.3. Pathway Model Maternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of simulated trait heritability; y- axis presents the beta coefficients of genetic nurturing effects. The left and middle plots presents boxplots representing β coefficient estimates of path *b*; the association between offspring phenotype and paternal PGS from the simple and extended pathway models respectively. The right plot presents boxplots representing β coefficient estimates of path *f* from the extended pathway model; the association between offspring phenotype and paternal phenotype.

Figure 2 presents β coefficient boxplots of parental genetic nurturing paths highlighted in **chapter 2 Figure 5**. Regression and pathway models were conducted using data simulated with the **chapter 2 simulations II** strategy. This strategy included an additional level of complexity; noise associated with the PGSs that arise from inaccurate effect sizes of SNPs obtained from GWAS summary statistics. This figure shows that when the PGSs were perfect measures of the genetic variance, i.e. tags the entirety of trait genetic variance, and noise was essentially 0, the estimated coefficients were as expected, ~0. In all other scenarios, extended model estimates for path *b* were always negative and estimates for *f* were always positive. These effects were exacerbated as noise associated with the PGS decreases and trait heritability increases.

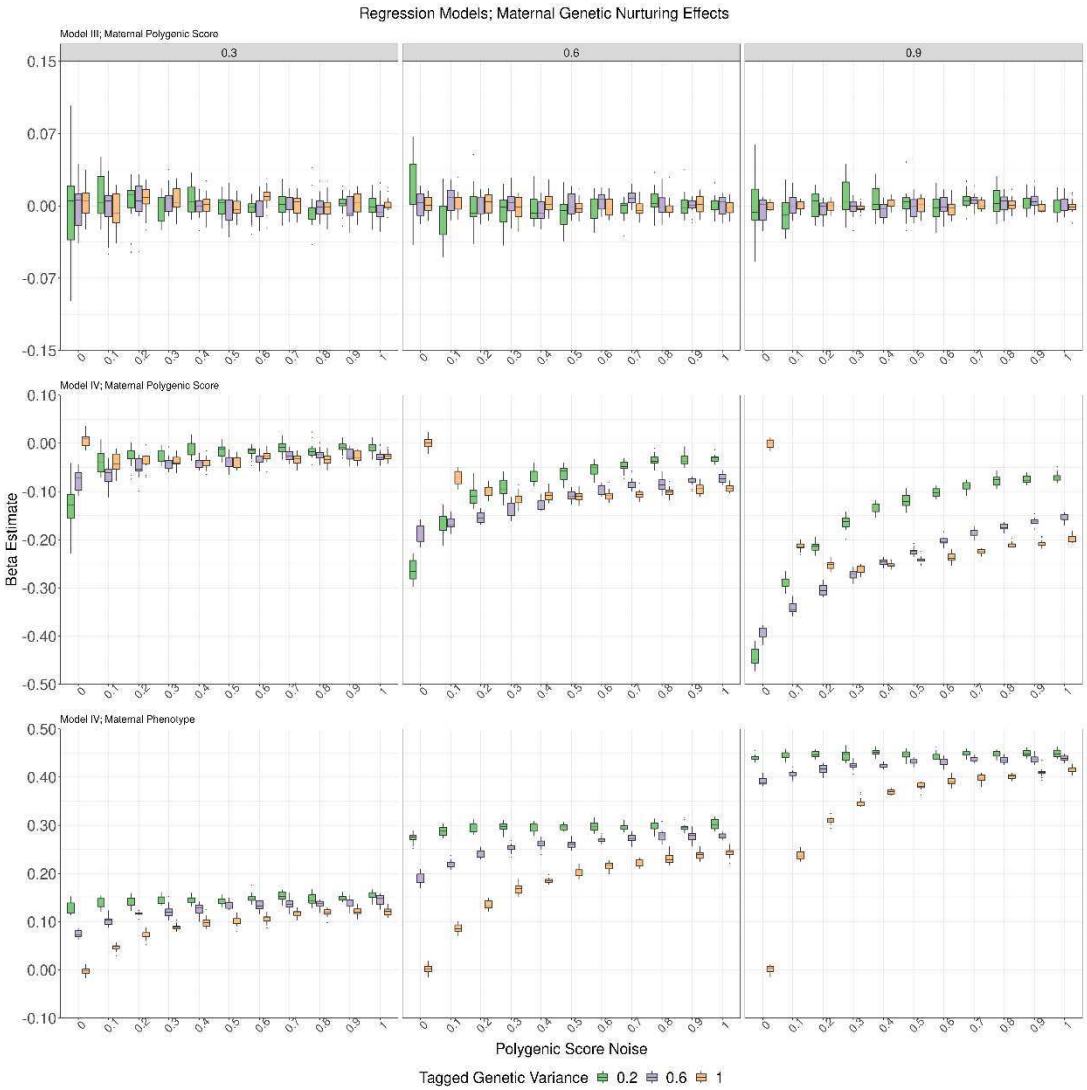


Figure 2.1. Regression Model Maternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of noise associated with PGSs, y- axis presents the beta coefficients of genetic nurturing effects, facets present varying levels of trait heritability. The top and middle plots presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal PGS from regression model *III* and *IV*, respectively. The bottom plot presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal phenotype regression model *IV.*

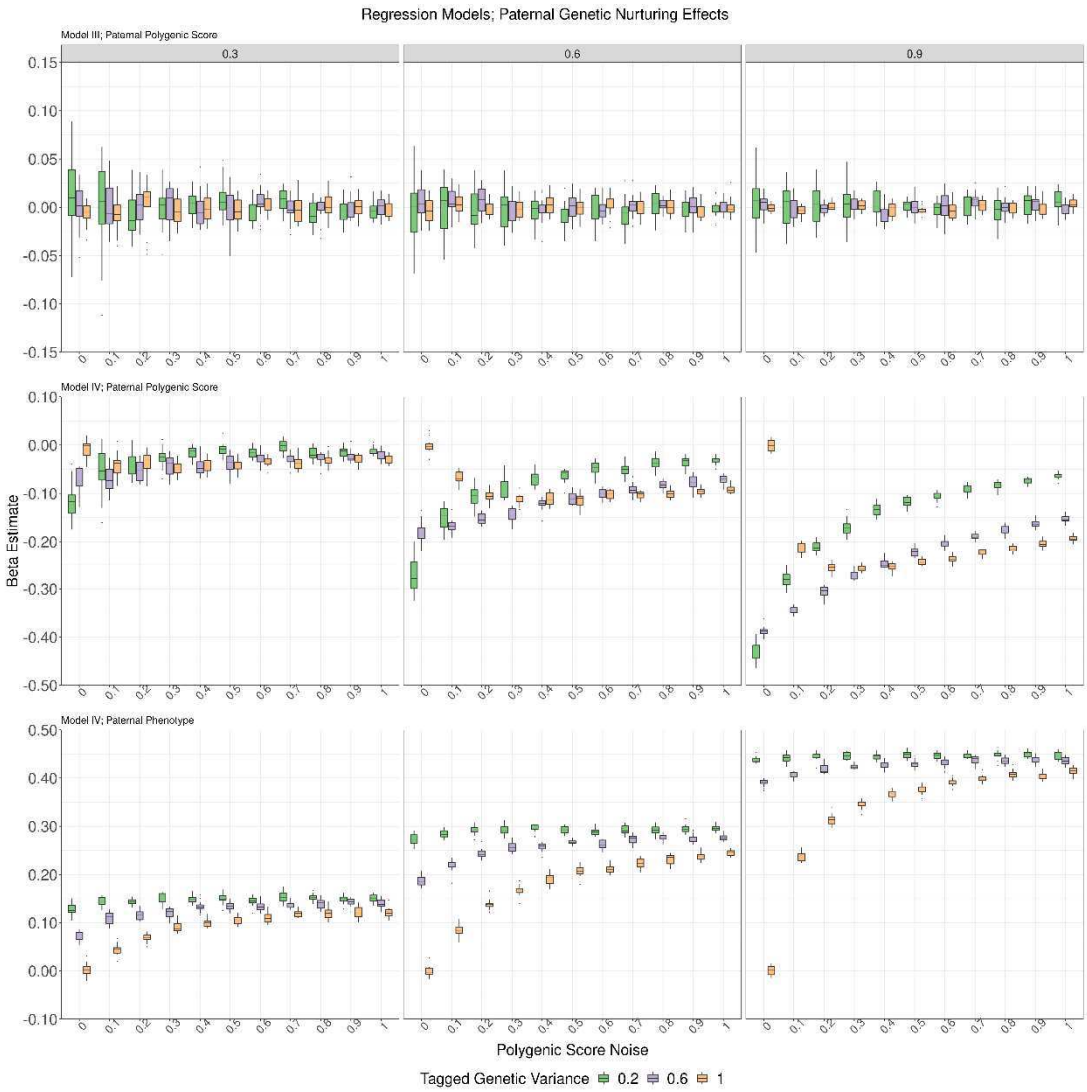


Figure 2.1. Regression Model Paternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of noise associated with PGSs, y- axis presents the beta coefficients of genetic nurturing effects, facets present varying levels of trait heritability. The top and middle plots presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal PGS from regression model *III* and *IV*, respectively. The bottom plot presents boxplots representing β coefficient estimates of the association between offspring phenotype and maternal phenotype regression model *IV.*

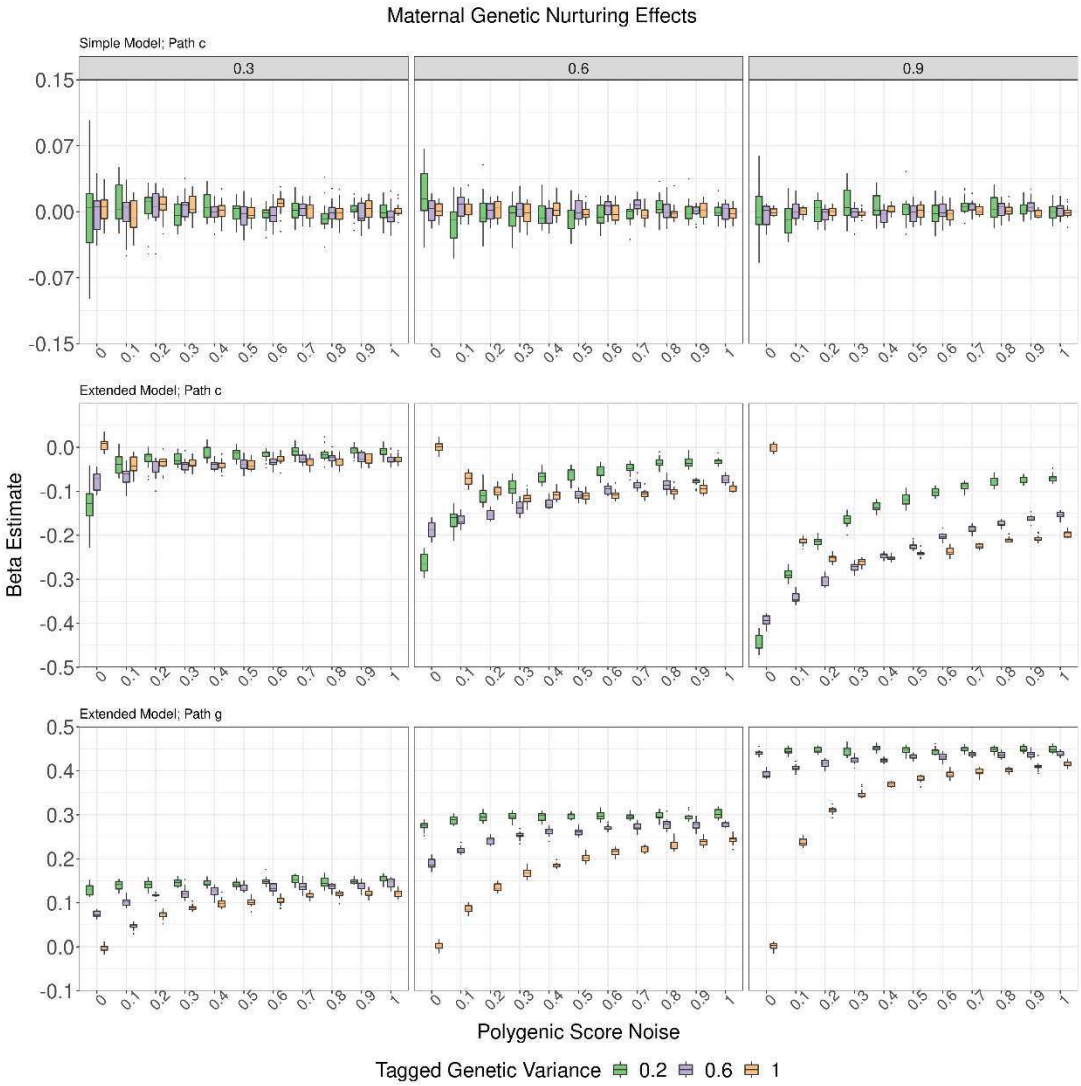


Figure 2.3. Pathway Model Maternal Genetic Nurturing Effects. Boxplots represent β coefficients from 15 replication analyses. Colours present simulated polygenic scores (PGS) capturing varying levels of trait genetic variance. X-axis presents varying levels of noise associated with PGSs, y- axis presents the beta coefficients of genetic nurturing effects, facets present varying levels of trait heritability. The top and middle plots presents boxplots representing β coefficient estimates of path b; the association between offspring phenotype and paternal PGS from the simple and extended pathway models respectively. The bottom plot presents boxplots representing β coefficient estimates of path f from the extended pathway model; the association between offspring phenotype and paternal phenotype.

Appendix II. Supplementary Materials for Chapter 3 'Genome-by-Trauma Interaction Effects on Depression'

Table 1.1. Broad Depression UK Biobank Questions and Field Codes

UKB FIELD	QUESTION	RESPONSE OPTIONS	CRITERIA
f.2090.0.0	Seen doctor (GP) for nerves, anxiety, tension or depression	Do not know Prefer not to answer No Yes	Participants who responded 'Yes' to either question are classified as cases. Participants who repsonded 'No' to both questions are classified as controls. Participants who responded 'Prefer not to answer' or 'Do not know' were omitted from further analyses.
f.2100.0.0	Seen a psychiatrist for nerves, anxiety, tension or depression		

Table 1.2. Neuroticism UK Biobank Questions and Field Codes

UKB FIELD	QUESTION	RESPONSE OPTIONS	CRITERIA
f.1920.0.0	Does your mood often go up and down?		Participants are given a sumscore of neuroticism symptom between 0-12
f.1930.0.0	Do you ever feel 'just miserable' for no reason?		
f.1940.0.0	Are you an irritable person?		
f.1950.0.0	Are your feelings easily hurt?		
f.1960.0.0	Do you often feel 'fed-up'?	Do not know	
f.1970.0.0	Would you call yourself a nervous person?	Prefer not to answer	
f.1980.0.0	Are you a worrier?	No	
f.1990.0.0	Would you call yourself tense or 'highly strung'?	Yes	
f.2000.0.0	Do you worry too long after an embarrassing experience?		
f.2010.0.0	Do you suffer from 'nerves'?		
f.2020.0.0	Do you often feel lonely?		
f.2030.0.0	Are you often troubled by feelings of guilt?		

Table 1.3. Composite International Diagnostic Inventory (CIDI) Depression UK Biobank Questions and Field Codes

UKB FIELD	QUESTION	RESPONSE OPTIONS	CRITERIA
f.20446.0.0 (cardinal)	Ever had prolong feelings of sadness or depression		Participants who reported at least 5 symptoms, including one cardinal symptom AND reported 'Most of the day' or more affected during worst episode of depression [f20436], Depressed 'almost every day' or more during worst episode of depression [f20439] and 'More than a little impact' or more on normal roles during worst period of depression [f20440] were classified as cases. All other participants were classified as controls.
f.20441.0.0 (cardinal)	Ever had prolonged loss of interest in normal activities		
f.20449.0.0	Feelings of tiredness during worst episode of depression	Prefer not to answer	
f.20532.0.0	Sleep change during worst episode of depression	No	
f.20435.0.0	Difficulty concentrating during worst episode of depression	Yes	
f.20450.0.0	Feelings of worthlessness during worst episode of		
f.20437.0.0	Thoughts of death during worst episode of depression		
f.20536.0.0	Weight change during worst episode of depression	Do not know Prefer not to answer Same/Dieting Gained Lost Gained and lost	
f.20436.0.0	Most of the day or more affected during worst episode of depression	Do not know Prefer not to answer Less than half of the day About half of the day Most of the day All day long	
f.20439.0.0	Depressed (almost) every day during worst episode of depression	Do not know Prefer not to answer Less often Almost every day Every day	
f.20440.0.0	More than a little impact on normal roles during worst period of depression	Prefer not to answer Not at all A little Somewhat A lot	

Table 1.4. Exclusion Criteria UK Biobank Questions and Field Codes

UKB FIELD	QUESTION	RESPONSE OPTIONS	CRITERIA
f.20544.0.0	Report any mental health problems diagnosed by a professional	Prefer not to answer ADD/ADHD Agoraphobia Anorexia nervosa Anxierty, nerves or GAD Autism, Asperger's or ASD Binge eating Bulimia nervosa Depression Mania, hypomania, bipolar OCD Panic attacks Any other phobia Personality disorder Any other type of psychosis Schizophrenia Social anxiety or social phobia	Cases self-reporting diagnoses of schizophrenia, other psychoses or bipolar disorder were excluded from analyses. Cases reporting any mental health disorder, hospitalisation due to mood disorders, taking antidepressant medication or met previouslt defined criteria for a mood disorder are exlcuded from analyses.
f.20002.0.0	Report depression in previous interview with psychiatric nurse	Depression	
f.20126.0.0	Meet previous criteria for depression or bipolar disorder	No bipolar or depression Bipolar I disorder Bipolar II disorder Probable recurrent major depression Single probable major depression	
f.41202.0.0 f.41204.0.0	Have a hospital inpatient ICD10 code for mood disorder	F30-F39	
f.20003.0.0	Report use of anti-depressant medication* at baseline	1140879616, 1140921600, 1140879540, 1140867878, 1140916282, 1140909806, 1140867888, 1141152732, 1141180212, 1140879634, 1140867876, 140882236, 1141190158, 1141200564, 1140867726, 1140879620, 1140867818, 1140879630, 1140879628, 1141151946, 1140867948, 1140867624, 1140867756, 1140867884, 1141151978, 1141152736, 1141201834, 1140867690, 1140867640, 1140867920, 1140867850, 1140879544, 1141200570, 1140867934, 1140867758, 1140867914, 1140867820, 1141151982, 1140882244, 1140879556, 1140867852, 1140867860, 1140917460, 1140867938, 1140867856, 1140867922, 1140910820, 1140882312, 1140867944, 1140867784, 1140867812, 1140867668, 1140867940	

Table 2 presents associations between UK Biobank trauma exposure questions and the phenotypes of interest; broad depression, CIDI depression and neuroticism. Trauma exposure question responses were 'Rarely True', 'Sometimes True', 'Often True', 'Always True'. Trauma exposure was explored as a categorical variable, and covariates; age and sex, were included in analyses. Table 2 was presented in separate sections for each trauma exposure sub-category (childhood, adult and catastrophic trauma exposure) and each phenotype of interest. Tables include β coefficient estimates of associations, standard errors, T and P-Values as well as phenotypic variance accounted for by each model.

Table 2.1.1. Childhood Trauma Exposure and CIDI Depression Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Felt loved f.20489.0.0	Intercept	0.691	0.009	75.093	0.000	0.084		
	Sex(M)	-0.151	0.002	-60.453	0.000			
	Age	-0.007	0.000	-43.205	0.000			
	Rarely True	0.032	0.003	10.809	0.000		31163	64492
	Sometimes True	0.174	0.004	48.808	0.000		18948	
	Often True	0.289	0.006	46.798	0.000		5292	
	Always True	0.354	0.011	32.234	0.000		1583	
Hit hard f.20488.0.0	Intercept	0.696	0.009	74.646	0.000	0.066		
	Sex(M)	-0.156	0.003	-62.029	0.000			
	Age	-0.007	0.000	-39.996	0.000			
	Rarely True	0.077	0.004	18.975	0.000		12954	99403
	Sometimes True	0.192	0.005	37.210	0.000		7692	
	Often True	0.306	0.014	22.128	0.000		1002	
	Always True	0.381	0.018	21.569	0.000		613	
Felt hated by family member f.20487.0.0	Intercept	0.661	0.009	71.538	0.000	0.084		
	Sex(M)	-0.140	0.002	-55.860	0.000			
	Age	-0.006	0.000	-38.304	0.000			
	Rarely True	0.121	0.005	22.864	0.000		7113	103745
	Sometimes True	0.247	0.005	47.964	0.000		7546	
	Often True	0.379	0.010	36.418	0.000		1748	
	Always True	0.420	0.011	36.751	0.000		1447	
Sexually molested f.20490.0.0	Intercept	0.723	0.009	77.618	0.000	0.058		
	Sex(M)	-0.143	0.003	-56.056	0.000			
	Age	-0.007	0.000	-42.388	0.000			
	Rarely True	0.119	0.006	19.467	0.000		5363	110669
	Sometimes True	0.185	0.007	24.952	0.000		3585	
	Often True	0.321	0.019	17.224	0.000		552	
	Always True	0.371	0.021	18.032	0.000		453	
Someone to take to the doctor f.20491.0.0	Intercept	0.756	0.009	81.211	0.000	0.054		
	Sex(M)	-0.150	0.003	-59.037	0.000			
	Age	-0.007	0.000	-45.083	0.000			
	Rarely True	0.051	0.004	12.240	0.000		12565	102320
	Sometimes True	0.193	0.008	24.001	0.000		3046	
	Often True	0.247	0.015	16.535	0.000		871	
	Always True	0.036	0.009	3.891	0.000		2362	

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.1.2. Childhood Trauma Exposure and Broad Depression Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Felt loved f.20489.0.0	Intercept	0.482	0.009	50.787	0.000	0.0557		
	Sex(M)	-0.159	0.003	-62.030	0.000			
	Age	-0.002	0.000	-10.883	0.000		34817	
	Rarely True	0.030	0.003	9.817	0.000		21700	71645
	Sometimes True	0.163	0.004	45.123	0.000		6033	
	Often True	0.248	0.006	39.683	0.000		1808	
	Always True	0.307	0.011	27.694	0.000			
Hit hard f.20488.0.0	Intercept	0.494	0.010	51.409	0.000	0.0394		
	Sex(M)	-0.163	0.003	-63.344	0.000			
	Age	-0.001	0.000	-8.652	0.000			
	Rarely True	0.058	0.004	13.988	0.000		14461	
	Sometimes True	0.150	0.005	28.486	0.000		8614	115923
	Often True	0.243	0.014	16.907	0.000		1077	
	Always True	0.304	0.018	16.515	0.000		655	
Felt hated by family member f.20487.0.0	Intercept	0.462	0.010	48.382	0.000	0.0521		
	Sex(M)	-0.149	0.003	-58.208	0.000			
	Age	-0.001	0.000	-6.958	0.000			
	Rarely True	0.105	0.005	19.515	0.000		8060	
	Sometimes True	0.208	0.005	39.833	0.000		8617	115923
	Often True	0.308	0.011	28.759	0.000		1942	
	Always True	0.337	0.012	28.669	0.000		1602	
Sexually molested f.20490.0.0	Intercept	0.512	0.010	53.375	0.000	0.0362		
	Sex(M)	-0.152	0.003	-58.681	0.000			
	Age	-0.002	0.000	-10.320	0.000			
	Rarely True	0.092	0.006	14.941	0.000		6104	
	Sometimes True	0.159	0.007	21.236	0.000		4105	123692
	Often True	0.252	0.019	13.482	0.000		637	
	Always True	0.294	0.021	13.884	0.000		497	
Someone to take to the doctor f.20491.0.0	Intercept	0.541	0.010	56.586	0.000	0.0341		
	Sex(M)	-0.158	0.003	-61.040	0.000			
	Age	-0.002	0.000	-13.003	0.000			
	Rarely True	0.043	0.004	10.250	0.000		14154	
	Sometimes True	0.163	0.008	20.074	0.000		3469	114362
	Often True	0.208	0.015	13.828	0.000		991	
	Always True	0.050	0.009	5.449	0.000		2665	

Abbreviations. Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.1.3. Childhood Trauma Exposure and Neuroticism Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Felt loved f.20489.0.0	Intercept	6.224	0.065	95.865	0.000	0.065		
	Sex(M)	-0.806	0.018	-45.839	0.000			
	Age	-0.043	0.001	-37.757	0.000			
	Rarely True	0.411	0.021	19.482	0.000		31162	65609
	Sometimes True	1.265	0.025	50.862	0.000		19649	
	Often True	1.882	0.043	44.040	0.000		5561	
	Always True	2.326	0.075	31.033	0.000		1709	
Hit hard f.20488.0.0	Intercept	6.423	0.066	97.244	0.000	0.038		
	Sex(M)	-0.825	0.018	-46.302	0.000			
	Age	-0.041	0.001	-35.360	0.000			
	Rarely True	0.302	0.029	10.425	0.000		13076	
	Sometimes True	0.922	0.036	25.622	0.000		8042	101038
	Often True	1.542	0.097	15.895	0.000		1034	
	Always True	1.588	0.121	13.123	0.000		662	
Felt hated by family member f.20487.0.0	Intercept	6.153	0.066	93.790	0.000	0.053		
	Sex(M)	-0.738	0.018	-41.679	0.000			
	Age	-0.038	0.001	-33.305	0.000			
	Rarely True	0.796	0.037	21.237	0.000		7218	
	Sometimes True	1.399	0.036	38.806	0.000		7882	105257
	Often True	2.021	0.072	28.205	0.000		1881	
	Always True	2.101	0.079	26.694	0.000		1556	
Sexually molested f.20490.0.0	Intercept	6.551	0.066	99.386	0.000	0.034		
	Sex(M)	-0.766	0.018	-42.663	0.000			
	Age	-0.043	0.001	-36.928	0.000			
	Rarely True	0.383	0.043	8.990	0.000		5596	
	Sometimes True	0.858	0.051	16.821	0.000		3856	112289
	Often True	1.193	0.126	9.448	0.000		609	
	Always True	1.642	0.142	11.594	0.000		484	
Someone to take to the doctor f.20491.0.0	Intercept	6.712	0.066	102.226	0.000	0.035		
	Sex(M)	-0.791	0.018	-44.267	0.000			
	Age	-0.046	0.001	-39.571	0.000			
	Rarely True	0.451	0.029	15.318	0.000		12589	
	Sometimes True	0.994	0.056	17.681	0.000		3148	104329
	Often True	1.511	0.102	14.853	0.000		941	
	Always True	0.568	0.065	8.800	0.000		2376	

Abbreviations. Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.2.1. Adult Trauma Exposure and CIDI Depression Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Been in a confiding relationship as an adult f.20522	Intercept	0.715	0.009	75.706	0.000	0.059		
	Sex(M)	-0.154	0.003	-60.213	0.000			
	Age	-0.007	0.000	-42.679	0.000			
	Rarely True	0.037	0.003	10.489	0.000		20921	61655
	Sometimes True	0.110	0.003	31.359	0.000		20954	
	Often True	0.172	0.006	27.685	0.000		5411	
	Always True	0.025	0.005	5.303	0.000		9971	
Physical violence by partner or ex-partner as an adult f.20523.0.0	Intercept	0.694	0.009	74.613	0.000	0.067		
	Sex(M)	-0.135	0.003	-53.316	0.000			
	Age	-0.007	0.000	-40.616	0.000			
	Rarely True	0.121	0.005	23.199	0.000		7426	106872
	Sometimes True	0.211	0.006	34.379	0.000		5316	
	Often True	0.290	0.015	19.850	0.000		898	
	Always True	0.375	0.013	27.943	0.000		1069	
Belittlement by partner or ex-partner as an adult f.20521.0.0	Intercept	0.624	0.009	67.840	0.000	0.100		
	Sex(M)	-0.113	0.003	-44.950	0.000			
	Age	-0.006	0.000	-37.857	0.000			
	Rarely True	0.124	0.004	29.156	0.000		11332	93821
	Sometimes True	0.242	0.004	56.669	0.000		11455	
	Often True	0.377	0.009	44.246	0.000		2611	
	Always True	0.452	0.009	50.437	0.000		2368	
Sexual interference by partner or ex-partner without consent as an adult f.20524.0.0	Intercept	0.718	0.009	77.249	0.000	0.061		
	Sex(M)	-0.133	0.003	-51.894	0.000			
	Age	-0.007	0.000	-42.200	0.000			
	Rarely True	0.172	0.007	23.281	0.000		3657	115014
	Sometimes True	0.275	0.010	28.960	0.000		2179	
	Often True	0.338	0.024	14.245	0.000		339	
	Always True	0.449	0.022	20.014	0.000		381	
Able to pay rent/mortgage as an adult f.20525.0.0	Intercept	0.728	0.009	77.961	0.000	0.060		
	Sex(M)	-0.146	0.003	-57.404	0.000			
	Age	-0.007	0.000	-42.672	0.000			
	Rarely True	0.106	0.005	22.113	0.000		8966	103689
	Sometimes True	0.274	0.008	32.599	0.000		2791	
	Often True	0.273	0.016	17.098	0.000		757	
	Always True	0.011	0.007	1.637	0.102		4047	

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.2.2. Adult Trauma Exposure and Broad Depression Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Been in a confiding relationship as an adult f.20522	Intercept	0.497	0.010	51.306	0.000	0.039		
	Sex(M)	-0.161	0.003	-61.764	0.000			
	Age	-0.002	0.000	-10.886	0.000			
	Rarely True	0.042	0.004	11.866	0.000		23516	
	Sometimes True	0.105	0.004	29.784	0.000		23832	68306
	Often True	0.165	0.006	26.400	0.000		6173	
	Always True	0.034	0.005	6.994	0.000		11188	
Physical violence by partner or ex-partner as an adult f.20523.0.0	Intercept	0.483	0.010	50.427	0.000	0.044		
	Sex(M)	-0.144	0.003	-55.719	0.000			
	Age	-0.001	0.000	-8.561	0.000			
	Rarely True	0.116	0.005	22.003	0.000		8489	
	Sometimes True	0.190	0.006	30.850	0.000		6151	119238
	Often True	0.278	0.015	18.971	0.000		1038	
	Always True	0.315	0.014	23.250	0.000		1219	
Belittlement by partner or ex-partner as an adult f.20521.0.0	Intercept	0.428	0.010	44.934	0.000	0.064		
	Sex(M)	-0.127	0.003	-48.942	0.000			
	Age	-0.001	0.000	-6.229	0.000			
	Rarely True	0.109	0.004	25.140	0.000		12900	
	Sometimes True	0.207	0.004	47.617	0.000		13171	104338
	Often True	0.319	0.009	36.916	0.000		3005	
	Always True	0.370	0.009	40.606	0.000		2705	
Sexual interference by partner or ex-partner without consent as an adult f.20524.0.0	Intercept	0.508	0.010	53.160	0.000	0.038		
	Sex(M)	-0.144	0.003	-55.107	0.000			
	Age	-0.002	0.000	-10.171	0.000			
	Rarely True	0.150	0.007	20.047	0.000		4171	
	Sometimes True	0.215	0.010	22.449	0.000		2483	128661
	Often True	0.313	0.024	13.003	0.000		384	
	Always True	0.353	0.023	15.126	0.000		409	
Able to pay rent/mortgage as an adult f.20525.0.0	Intercept	0.519	0.010	54.064	0.000	0.037		
	Sex(M)	-0.155	0.003	-59.628	0.000			
	Age	-0.002	0.000	-11.155	0.000			
	Rarely True	0.080	0.005	16.240	0.000		10042	
	Sometimes True	0.226	0.008	26.788	0.000		3195	115814
	Often True	0.239	0.016	14.807	0.000		854	
	Always True	0.038	0.007	5.296	0.000		4581	

Abbreviations. Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.2.3. Adult Trauma Exposure and Neuroticism Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Been in a confiding relationship as an adult f.20522	Intercept	6.349	0.066	95.820	0.000	0.048		
	Sex(M)	-0.808	0.018	-45.172	0.000			
	Age	-0.043	0.001	-37.495	0.000			
	Rarely True	0.267	0.024	10.930	0.000		21360	
	Sometimes True	0.877	0.024	35.873	0.000		21288	63266
	Often True	1.542	0.043	35.486	0.000		5487	
	Always True	0.662	0.033	19.849	0.000		9936	
Physical violence by partner or ex-partner as an adult f.20523.0.0	Intercept	6.484	0.066	98.057	0.000	0.034		
	Sex(M)	-0.749	0.018	-41.605	0.000			
	Age	-0.042	0.001	-36.159	0.000			
	Rarely True	0.249	0.037	6.784	0.000		7700	
	Sometimes True	0.757	0.042	17.870	0.000		5741	108165
	Often True	0.931	0.100	9.274	0.000		970	
	Always True	1.393	0.091	15.376	0.000		1198	
Belittlement by partner or ex-partner as an adult f.20521.0.0	Intercept	6.045	0.066	91.735	0.000	0.054		
	Sex(M)	-0.628	0.018	-34.987	0.000			
	Age	-0.038	0.001	-33.456	0.000			
	Rarely True	0.721	0.030	23.680	0.000		11535	
	Sometimes True	1.240	0.030	41.386	0.000		12150	94691
	Often True	1.543	0.059	26.116	0.000		2822	
	Always True	1.806	0.062	29.261	0.000		2590	
Sexual interference by partner or ex-partner without consent as an adult f.20524.0.0	Intercept	6.550	0.066	99.450	0.000	0.033		
	Sex(M)	-0.740	0.018	-40.757	0.000			
	Age	-0.043	0.001	-36.875	0.000			
	Rarely True	0.555	0.052	10.588	0.000		3708	
	Sometimes True	1.023	0.066	15.540	0.000		2307	116944
	Often True	1.368	0.162	8.464	0.000		372	
	Always True	1.493	0.153	9.764	0.000		416	
Able to pay rent/mortgage as an adult f.20525.0.0	Intercept	6.573	0.066	99.695	0.000	0.036		
	Sex(M)	-0.778	0.018	-43.366	0.000			
	Age	-0.043	0.001	-37.496	0.000			
	Rarely True	0.472	0.034	13.830	0.000		9011	
	Sometimes True	1.164	0.058	20.059	0.000		2950	105706
	Often True	1.522	0.109	13.937	0.000		815	
	Always True	0.390	0.050	7.866	0.000		4087	

Abbreviations. Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.3.1. Catastrophic Trauma Exposure and CIDI Depression Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Victim of sexual assault f.20531.0.0	Intercept	0.692	0.009	74.197	0.000	0.066		
	Sex(M)	-0.130	0.003	-50.529	0.000			
	Age	-0.007	0.000	-41.146	0.000			
	Yes, nw12 mths	0.183	0.004	50.125	0.000		17222	103230
	Yes, w 12 mths	0.193	0.040	4.848	0.000		119	
Victim of physically violent crime f.20529.0.0	Intercept	0.719	0.009	76.182	0.000	0.051		
	Sex(M)	-0.161	0.003	-62.895	0.000			
	Age	-0.007	0.000	-40.652	0.000			
	Yes, nw12 mths	0.075	0.003	22.860	0.000		22490	98642
	Yes, w 12 mths	0.102	0.019	5.429	0.000		546	
Been in serious accident believed to be life-threatening f.20526.0.0	Intercept	0.742	0.009	79.672	0.000	0.051		
	Sex(M)	-0.159	0.003	-62.494	0.000			
	Age	-0.007	0.000	-43.033	0.000			
	Yes, nw12 mths	0.100	0.004	22.964	0.000		11447	109967
	Yes, w 12 mths	0.098	0.024	4.142	0.000		342	
Witnessed sudden violent death f.20530.0.0	Intercept	0.743	0.009	79.726	0.000	0.051		
	Sex(M)	-0.162	0.003	-63.035	0.000			
	Age	-0.007	0.000	-43.197	0.000			
	Yes, nw12 mths	0.084	0.004	22.416	0.000		16076	104955
	Yes, w 12 mths	0.064	0.017	3.691	0.000		655	
Diagnosed with life-threatening illness f.20528.0.0	Intercept	0.770	0.009	82.523	0.000	0.050		
	Sex(M)	-0.155	0.003	-60.841	0.000			
	Age	-0.008	0.000	-46.119	0.000			
	Yes, nw12 mths	0.075	0.004	20.386	0.000		16877	102040
	Yes, w 12 mths	0.066	0.009	7.504	0.000		2541	
Been involved in combat or exposed to war-zone f.20527.0.0	Intercept	0.757	0.009	81.216	0.000	0.047		
	Sex(M)	-0.155	0.003	-60.441	0.000			
	Age	-0.007	0.000	-43.838	0.000			
	Yes, nw12 mths	0.036	0.007	5.272	0.000		4298	117377
	Yes, w 12 mths	0.029	0.047	0.613	0.540		89	

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.3.2. Catastrophic Trauma Exposure and Broad Depression Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Victim of sexual assault f.20531.0.0	Intercept	0.486	0.010	50.520	0.000	0.041		
	Sex(M)	-0.142	0.003	-54.083	0.000			
	Age	-0.002	0.000	-9.193	0.000			
	Yes, nw12 mths	0.149	0.004	40.390	0.000		19575	115219
	Yes, w 12 mths	0.202	0.041	4.936	0.000		132	
Victim of physically violent crime f.20529.0.0	Intercept	0.509	0.010	52.497	0.000	0.032		
	Sex(M)	-0.167	0.003	-64.180	0.000			
	Age	-0.002	0.000	-9.154	0.000			
	Yes, nw12 mths	0.061	0.003	18.161	0.000		19575	110680
	Yes, w 12 mths	0.097	0.019	5.116	0.000		623	
Been in serious accident believed to be life-threatening f.20526.0.0	Intercept	0.529	0.010	55.287	0.000	0.031		
	Sex(M)	-0.166	0.003	-63.759	0.000			
	Age	-0.002	0.000	-10.954	0.000			
	Yes, nw12 mths	0.074	0.004	16.711	0.000		12638	123323
	Yes, w 12 mths	0.095	0.024	3.941	0.000		385	
Witnessed sudden violent death f.20530.0.0	Intercept	0.533	0.010	55.632	0.000	0.030		
	Sex(M)	-0.165	0.003	-63.096	0.000			
	Age	-0.002	0.000	-11.215	0.000			
	Yes, nw12 mths	0.041	0.004	10.621	0.000		17559	118005
	Yes, w 12 mths	0.037	0.018	2.098	0.036		724	
Diagnosed with life-threatening illness f.20528.0.0	Intercept	0.550	0.010	57.394	0.000	0.031		
	Sex(M)	-0.162	0.003	-62.640	0.000			
	Age	-0.002	0.000	-13.466	0.000			
	Yes, nw12 mths	0.055	0.004	14.791	0.000		19051	114029
	Yes, w 12 mths	0.068	0.009	7.628	0.000		2898	
Been involved in combat or exposed to war-zone f.20527.0.0	Intercept	0.539	0.010	56.338	0.000	0.029		
	Sex(M)	-0.161	0.003	-61.726	0.000			
	Age	-0.002	0.000	-11.408	0.000			
	Yes, nw12 mths	-0.005	0.007	-0.744	0.457		4642	131632
	Yes, w 12 mths	-0.015	0.049	-0.299	0.765		93	

Abbreviations. Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 2.3.3. Catastrophic Trauma Exposure and Neuroticism Regression Model Associations

Trauma	Predictor	Est	SE	T	P	AdjR2	Ncase	Ncontrol
Victim of sexual assault f.20531.0.0	Intercept	6.455	0.066	97.406	0.000	0.034		
	Sex(M)	-0.724	0.018	-39.870	0.000			
	Age	-0.042	0.001	-36.288	0.000			
	Yes, nw12 mths	0.643	0.025	25.228	0.000		18064	104622
	Yes, w 12 mths	0.779	0.286	2.723	0.006		118	
Victim of physically violent crime f.20529.0.0	Intercept	6.519	0.067	97.857	0.000	0.031		
	Sex(M)	-0.848	0.018	-47.031	0.000			
	Age	-0.041	0.001	-35.609	0.000			
	Yes, nw12 mths	0.327	0.023	14.081	0.000		22874	100423
	Yes, w 12 mths	0.492	0.131	3.764	0.000		572	
Been in serious accident believed to be life-threatening f.20526.0.0	Intercept	6.649	0.066	101.066	0.000	0.030		
	Sex(M)	-0.828	0.018	-46.035	0.000			
	Age	-0.043	0.001	-37.381	0.000			
	Yes, nw12 mths	0.234	0.030	7.665	0.000		11692	111898
	Yes, w 12 mths	0.364	0.167	2.177	0.030		349	
Witnessed sudden violent death f.20530.0.0	Intercept	6.688	0.066	101.595	0.000	0.029		
	Sex(M)	-0.809	0.018	-44.778	0.000			
	Age	-0.044	0.001	-37.690	0.000			
	Yes, nw12 mths	-0.021	0.027	-0.790	0.430		16292	106902
	Yes, w 12 mths	-0.054	0.120	-0.449	0.653		682	
Diagnosed with life-threatening illness f.20528.0.0	Intercept	6.731	0.066	102.202	0.000	0.030		
	Sex(M)	-0.818	0.018	-45.719	0.000			
	Age	-0.045	0.001	-38.710	0.000			
	Yes, nw12 mths	0.246	0.026	9.543	0.000		17347	103619
	Yes, w 12 mths	0.369	0.061	6.057	0.000		2699	
Been involved in combat or exposed to war-zone f.20527.0.0	Intercept	6.684	0.066	101.769	0.000	0.030		
	Sex(M)	-0.804	0.018	-44.646	0.000			
	Age	-0.043	0.001	-37.617	0.000			
	Yes, nw12 mths	-0.171	0.049	-3.505	0.000		4280	119584
	Yes, w 12 mths	-0.496	0.342	-1.450	0.147		83	

Abbreviations. Sex(M), Male; w/nw, within/not within; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2, adjusted R^2 value; n, sample size.

Table 3 presents associations between UK Biobank trauma exposure eigenvectors, or also known as principal components (PCs) and the phenotypes of interest; broad depression, CIDI depression and neuroticism. Trauma PCs were obtained for full, childhood, adult and catastrophic trauma separately. Full trauma PCs pre-corrected for the full sample GRMs were also explored. Table 3 was presented in separate sections for each trauma exposure sub-category; full, childhood, adult and catastrophic trauma exposure. Tables include β coefficient estimates of associations, standard errors, T and P-Values as well as phenotypic variance accounted for by each model.

Table 3.1. Childhood Trauma Exposure Principal Components and Depression/Neuroticism Regression Model Associations

	CIDI DEPRESSION (N = 128,927)					BROAD DEPRESSION (N = 486,165)					NEUROTICISM (N = 401,663)				
Predictor	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***
(Intercept)	0.698	0.009	75.527	0.000	NA	0.497	0.010	51.947	0.000	NA	6.482	0.065	99.144	0.000	NA
Age	-0.006	0.000	-37.980	0.0	0.019	-0.001	0.000	-7.541	0.000	0.002	-0.040	0.001	-35.218	0.000	0.011
Sex(M)	-0.141	0.003	-56.103	0.000	0.047	-0.150	0.003	-57.814	0.000	0.028	-0.751	0.018	-42.218	0.000	0.032
PC1	0.070	0.001	79.004	0.000	0.094	0.059	0.001	64.670	0.000	0.059	0.398	0.006	64.468	0.000	0.062
PC2	0.019	0.001	13.862	0.000	0.095	0.011	0.001	8.181	0.000	0.059	-0.031	0.010	-3.285	0.001	0.062
PC3	0.014	0.001	9.959	0.000	0.096	0.009	0.001	6.384	0.000	0.060	0.066	0.010	6.869	0.000	0.062
PC4	-0.035	0.002	-21.690	0.000	0.100	-0.037	0.002	-22.318	0.000	0.063	-0.351	0.011	-31.272	0.000	0.070
PC5	0.011	0.002	6.188	0.000	0.100	0.007	0.002	4.066	0.000	0.063	0.004	0.012	0.315	0.753	0.070

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; PC, principal component; Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; AdjR2; N, sample size. Coefficients are obtained from the full multiple regression model (including all predictors). *Represent adjusted R2 values when the specific predictor is ADDED to the model, on top of prior predictors.

Table 3.2. Adult Trauma Exposure Principal Components and Depression/Neuroticism Regression Model Associations

	CIDI DEPRESSION (N = 128,927)					BROAD DEPRESSION (N = 486,165)					NEUROTICISM (N = 401,663)				
Predictor	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***
(Intercept)	0.683	0.009	73.522	0.000	NA	0.476	0.010	49.578	0.000	NA	6.411	0.066	97.180	0.000	NA
Age	-0.006	0.000	-37.358	0.000	0.019	-0.001	0.000	-6.454	0.000	0.002	-0.040	0.001	-34.593	0.000	0.011
Sex(M)	-0.112	0.003	-43.518	0.000	0.047	-0.124	0.003	-46.860	0.000	0.028	-0.629	0.018	-34.520	0.000	0.032
PC1	0.071	0.001	77.007	0.000	0.092	0.062	0.001	65.473	0.000	0.060	0.309	0.006	48.234	0.000	0.047
PC2	0.006	0.001	5.224	0.000	0.093	0.010	0.001	8.357	0.000	0.060	0.211	0.009	24.395	0.000	0.052
PC3	-0.011	0.001	-7.885	0.000	0.093	-0.010	0.001	-6.892	0.000	0.061	-0.179	0.010	-17.947	0.000	0.055
PC4	-0.039	0.002	-23.199	0.000	0.097	-0.035	0.002	-20.764	0.000	0.064	-0.196	0.012	-16.997	0.000	0.057
PC5	0.061	0.002	31.521	0.000	0.105	0.044	0.002	22.353	0.000	0.067	0.373	0.013	27.948	0.000	0.063

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; PC, principal component; Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; N, sample size. Coefficients are obtained from the full multiple regression model (including all predictors). *Represent adjusted R2 values when the specific predictor is ADDED to the model, on top of prior predictors.

Table 3.3. Catastrophic Trauma Exposure Principal Components and Depression/Neuroticism Regression Model Associations

	CIDI DEPRESSION (N = 128,927)					BROAD DEPRESSION (N = 486,165)					NEUROTICISM (N = 401,663)				
Predictor	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2***	**Est**	**SE**	**T**	**P**	**Cumulative Adjusted R2**
(Intercept)	0.724	0.009	76.915	0.000	NA	0.509	0.010	52.401	0.000	NA	6.540	0.067	97.767	0.000	NA
Age	-0.007	0.000	-39.975	0.000	0.019	-0.001	0.000	-8.639	0.000	0.002	-0.042	0.001	-35.370	0.000	0.011
Sex(M)	-0.148	0.003	-56.024	0.000	0.047	-0.152	0.003	-56.271	0.000	0.028	-0.753	0.019	-40.154	0.000	0.032
PC1	0.046	0.001	42.547	0.000	0.062	0.031	0.001	28.004	0.000	0.035	0.101	0.008	13.284	0.000	0.031
PC2	0.043	0.001	33.730	0.000	0.070	0.041	0.001	31.211	0.000	0.042	0.214	0.009	23.734	0.000	0.035
PC3	-0.022	0.001	-17.182	0.000	0.072	-0.020	0.001	-15.156	0.000	0.044	-0.106	0.009	-11.697	0.000	0.036
PC4	0.016	0.001	11.761	0.000	0.073	0.010	0.001	7.416	0.000	0.044	0.031	0.009	3.312	0.001	0.036
PC5	-0.014	0.001	-10.466	0.000	0.074	-0.011	0.001	-7.926	0.000	0.045	-0.010	0.010	-1.041	0.298	0.036
PC6	-0.010	0.001	-7.143	0.000	0.075	-0.004	0.001	-3.043	0.002	0.045	0.020	0.010	2.036	0.042	0.036

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; PC, principal component; Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; N, sample size. Coefficients are obtained from the full multiple regression model (including all predictors). *Represent adjusted R2 values when the specific predictor is ADDED to the model, on top of prior predictors.

Table 3.4. Full Trauma Exposure Principal Components and Depression/Neuroticism Regression Model Associations

	CIDI DEPRESSION (N = 128,927)					BROAD DEPRESSION (N = 486,165)					NEUROTICISM (N = 401,663)				
Predictor	Est	SE	T	P	Cumulative Adjusted R2*	Est	SE	T	P	Cumulative Adjusted R2*	Est	SE	T	P	Cumulative Adjusted R2*
(Intercept)	0.649	0.009	69.005	0.000	NA	0.450	0.010	45.870	0.000	NA	6.282	0.067	93.516	0.000	NA
Age	-0.005	0.000	-33.160	0.000	0.019	-0.001	0.000	-3.932	0.000	0.002	-0.038	0.001	-32.367	0.000	0.011
Sex(M)	-0.116	0.003	-43.328	0.000	0.047	-0.122	0.003	-44.101	0.000	0.028	-0.615	0.019	-32.340	0.000	0.032
PC1	-0.075	0.001	-99.268	0.000	0.122	-0.064	0.001	-81.872	0.000	0.077	-0.366	0.005	-69.159	0.000	0.067
PC2	-0.006	0.001	-5.406	0.000	0.122	-0.005	0.001	-4.918	0.000	0.078	0.114	0.007	15.426	0.000	0.069
PC3	0.007	0.001	6.455	0.000	0.122	-0.003	0.001	-3.032	0.002	0.078	-0.126	0.008	-16.291	0.000	0.071
PC4	-0.004	0.001	-3.516	0.000	0.123	-0.003	0.001	-2.407	0.016	0.078	-0.006	0.008	-0.764	0.445	0.071
PC5	-0.015	0.001	-12.609	0.000	0.124	-0.013	0.001	-10.234	0.000	0.078	-0.107	0.009	-12.415	0.000	0.072
PC6	0.013	0.001	10.531	0.000	0.125	0.011	0.001	8.502	0.000	0.079	0.043	0.009	4.833	0.000	0.072
PC7	0.008	0.001	6.203	0.000	0.125	0.012	0.001	9.377	0.000	0.079	0.099	0.009	10.873	0.000	0.073
PC8	-0.002	0.001	-1.215	0.224	0.125	0.000	0.001	-0.016	0.987	0.079	0.072	0.010	7.553	0.000	0.074
PC9	-0.010	0.001	-7.643	0.000	0.125	-0.012	0.001	-8.400	0.000	0.080	-0.194	0.010	-19.849	0.000	0.077
PC10	-0.013	0.001	-9.781	0.000	0.126	-0.008	0.001	-5.861	0.000	0.080	-0.073	0.010	-7.384	0.000	0.077
PC11	-0.014	0.002	-9.478	0.000	0.127	-0.012	0.002	-7.912	0.000	0.081	-0.086	0.011	-8.015	0.000	0.078
PC12	0.046	0.002	28.633	0.000	0.133	0.044	0.002	26.448	0.000	0.086	0.341	0.011	29.976	0.000	0.085
PC13	0.011	0.002	6.475	0.000	0.133	0.009	0.002	5.535	0.000	0.086	0.009	0.011	0.787	0.431	0.085
PC14	0.013	0.002	7.325	0.000	0.133	0.008	0.002	4.203	0.000	0.086	0.039	0.012	3.133	0.002	0.085
PC15	0.018	0.002	10.150	0.000	0.134	0.015	0.002	8.228	0.000	0.086	0.043	0.013	3.384	0.001	0.085
PC16	0.042	0.002	21.569	0.000	0.138	0.026	0.002	13.468	0.000	0.088	0.242	0.013	18.073	0.000	0.088

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; PC, principal component; Sex(M), Male; Est, estimate; SE, standard error; T, T-Value; P, P-Value; N, sample size. Coefficients are obtained from the full multiple regression model (including all predictors). *Represent adjusted R2 values when the specific predictor is ADDED to the model, on top of prior predictors.

Table 4 presents trauma exposure principal component (PC) loadings. Here, values represent how each trauma exposure question loads on to the separate PCs. Loadings are presented separately for each trauma exposure sub-category; full, childhood, adult and catastrophic trauma exposure. Table 4 is presented in separate sections for each trauma exposure sub-category.

Table 4.1. Childhood Trauma Exposure Principal Component Loadings

Trauma	UKB Field	PC1	PC2	PC3	PC4	PC5
Childhood	f.20489.0.0	0.511	-0.228	0.080	-0.683	-0.463
	f.20488.0.0	0.491	0.141	0.383	0.648	-0.416
	f.20487.0.0	0.528	0.070	0.328	-0.113	0.772
	f.20490.0.0	0.310	0.734	-0.600	-0.062	-0.033
	f.20491.0.0	0.351	-0.619	-0.616	0.313	0.124

Abbreviations. UKB, UK Biobank; PC, principal component.

Table 4.2. Adult Trauma Exposure Principal Component Loadings

Trauma	UKB Field	PC1	PC2	PC3	PC4	PC5
Adult	f.20522.0.0	0.168	0.701	-0.691	0.029	-0.047
	f.20523.0.0	0.578	-0.155	0.022	-0.298	-0.743
	f.20521.0.0	0.567	-0.142	-0.071	-0.471	0.657
	f.20524.0.0	0.525	-0.157	-0.004	0.830	0.107
	f.20525.0.0	0.203	0.663	0.719	-0.004	0.042

Abbreviations. UKB, UK Biobank; PC, principal component.

Table 4.3. Catastrophic Trauma Exposure Principal Component Loadings

Trauma	UKB Field	PC1	PC2	PC3	PC4	PC5	PC6
Catastrophic	f.20531.0.0	0.263	0.723	-0.122	0.545	-0.310	-0.015
	f.20529.0.0	0.458	0.373	0.137	-0.262	0.738	0.139
	f.20526.0.0	0.479	0.039	0.033	-0.600	-0.584	0.260
	f.20530.0.0	0.527	-0.276	0.154	0.105	-0.032	-0.781
	f.20528.0.0	0.179	-0.135	-0.965	-0.054	0.124	-0.034
	f.20527.0.0	0.426	-0.493	0.106	0.511	0.043	0.549

Abbreviations. UKB, UK Biobank; PC, principal component.

Table 4.4. Full Trauma Exposure Principal Component Loadings

Trauma	UKB Field	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Childhood	f.20489.0.0	-0.341	0.340	-0.086	0.057	-0.224	0.008	-0.014	0.016	-0.148	-0.057	0.138	0.570	-0.232	-0.514	-0.120	-0.073
	f.20488.0.0	-0.324	0.256	0.129	0.120	-0.364	0.011	-0.071	-0.007	0.140	0.068	-0.293	-0.597	0.168	-0.369	0.088	0.149
	f.20487.0.0	-0.363	0.239	0.047	0.136	-0.380	0.035	-0.044	-0.029	0.038	0.031	-0.150	0.124	-0.160	0.742	0.030	-0.157
	f.20490.0.0	-0.279	-0.043	0.136	0.522	0.380	-0.008	-0.030	-0.034	-0.026	-0.018	-0.041	0.021	0.289	0.054	-0.626	0.056
	f.20491.0.0	-0.226	0.375	-0.261	-0.073	0.221	0.008	-0.010	-0.102	0.207	0.069	0.724	-0.216	0.171	0.113	0.123	0.023
Adult	f.20522.0.0	-0.163	0.249	-0.285	-0.221	0.256	-0.106	0.123	0.238	-0.654	-0.314	-0.222	-0.213	0.046	0.092	0.019	-0.023
	f.20523.0.0	-0.341	-0.409	-0.098	-0.254	-0.088	0.004	-0.037	-0.003	0.036	0.031	0.011	-0.016	0.351	-0.100	-0.018	-0.705
	f.20521.0.0	-0.356	-0.362	-0.104	-0.220	-0.141	0.019	-0.040	-0.005	-0.064	-0.013	0.016	0.270	0.364	0.086	0.138	0.651
	f.20524.0.0	-0.316	-0.401	-0.107	-0.128	0.016	0.009	-0.062	-0.033	-0.009	-0.011	0.162	-0.329	-0.692	-0.016	-0.278	0.122
	f.20525.0.0	-0.161	0.176	-0.313	-0.272	0.422	-0.020	0.062	-0.091	0.500	0.191	-0.506	0.154	-0.110	-0.026	-0.015	0.033
Catastrophic	f.20531.0.0	-0.296	-0.188	0.169	0.437	0.373	-0.020	0.019	0.018	-0.051	-0.032	-0.030	0.052	-0.173	-0.073	0.685	-0.074
	f.20529.0.0	-0.125	0.016	0.368	-0.150	0.011	-0.247	0.538	0.631	0.176	0.163	0.118	0.022	-0.021	0.020	-0.065	0.027
	f.20526.0.0	-0.094	0.069	0.391	-0.238	0.073	-0.060	0.438	-0.683	-0.261	0.200	-0.003	-0.012	-0.008	-0.004	-0.014	0.000
	f.20530.0.0	-0.080	0.092	0.458	-0.286	0.119	-0.087	-0.248	-0.073	0.271	-0.727	0.020	0.038	-0.008	0.014	-0.014	-0.002
	f.20528.0.0	-0.042	0.039	0.129	-0.095	0.086	0.956	0.167	0.139	-0.011	-0.048	-0.004	-0.014	-0.003	-0.006	-0.010	-0.003
	f.20527.0.0	-0.049	0.126	0.364	-0.271	0.204	0.006	-0.628	0.177	-0.236	0.501	0.009	0.032	-0.022	0.019	-0.002	0.001

Abbreviations. UKB, UK Biobank; PC, principal component.

Table 5 presents demographic information on individuals within each geographical cluster; North, Mid-North, Mid-South, South-West, South-East. Demographic information for each cluster was presented when clusters were formed using all individuals responding to ALL trauma exposure questions. Additional demographic information tables were provided for only unrelated individuals responding to all trauma exposure questions as well as all individuals who have responded to all childhood, adult and catastrophic trauma exposure questions separately. Table 5 was presented in separate sections for each sample.

Table 5.1. Cluster Demographics Using All Participants With Complete Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER					MIDNORTH CLUSTER					MIDSOUTH CLUSTER					SOUTHWEST CLUSTER					SOUTHEAST CLUSTER				
Total Nind	24287					29981					28961					18696					36323				
Sex	Female	Male				Female	Male				Female	Male				Female	Male				Female	Male			
	13433	10854				16595	13386				15835	13126				10533	8163				20599	15724			
Age	Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max		
	40	56.2	70.3			40.2	56.51	72.5			40.1	56.89	72.9			40.2	56.05	70.4			38.9	56.66	70.6		
	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total
CIDI Depression	3726	1814	7053	7418	20011	4913	2277	8534	9054	24778	4876	2282	7644	8795	23597	3012	1371	5565	5623	15571	5402	2342	11050	11073	29867
Broad Depression	5225	2654	7053	7418	22350	6706	3546	8513	8897	27662	6563	3365	7892	8876	26696	3928	1989	5729	5634	17280	7530	3587	11253	11105	33475
Neuroticism	Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M	
	4.28	11375	3.41	9337		4.24	13748	3.38	11438		4.33	13158	3.36	11302		4.20	8832	3.31	7019		4.06	17141	3.32	13645	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum; F, M, Female and Male, respectively.

Table 5.2. Cluster Demographics Using Unrelated Participants With Complete Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER					MIDNORTH CLUSTER					MIDSOUTH CLUSTER					SOUTHWEST CLUSTER					SOUTHEAST CLUSTER				
Total Nind	20994					26566					25691					16841					33128				
Sex	Female	Male				Female	Male				Female	Male				Female	Male				Female	Male			
	11626	9368				14729	11837				14048	11643				9473	7368				18735	14393			
Age	Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max		
	40	56.14	70.3			40.3	56.47	72.5			40.1	56.86	72.9			40.2	56.03	70.4			40.2	56.60	70.6		
	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total
CIDI Depression	3226	1575	6125	6461	17387	4395	2020	7552	8007	21974	4346	2065	6774	7775	20960	2710	1244	4988	5066	14008	4974	2131	10030	10165	27300
Broad Depression	4494	2287	6114	6395	19290	5960	3148	7538	7855	24501	5828	3002	6992	7858	23680	3551	1789	5122	5090	15552	6852	3253	10230	10195	30530
Neuroticism	Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M	
	4.25	9826	3.41	8074		4.24	12213	3.38	10118		4.32	11689	3.36	10019		4.20	7940	3.31	6351		4.06	15598	3.31	12481	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum; F, M, Female and Male, respectively.

Table 5.3. Cluster Demographics Using All Participants With Complete Childhood Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER					MIDNORTH CLUSTER					MIDSOUTH CLUSTER					SOUTHWEST CLUSTER					SOUTHEAST CLUSTER				
Total Nind	25466					31458					30463					19603					37861				
Sex	Female	Male				Female	Male				Female	Male				Female	Male				Female	Male			
	14197	11269				17561	13897				16824	13639				11126	8477				21662	16199			
Age	Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max		
	40.00	56.25	70.30			40.20	56.56	72.50			40.10	56.95	72.90			40.20	56.10	70.40			38.90	56.71	70.60		
	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total
CIDI Depression	3941	1864	7424	7754	20983	5178	2357	9007	9380	25922	5129	2375	8107	9117	24728	3159	1419	5859	5828	16265	5686	2424	11558	11371	31039
Broad Depression	5580	2759	7397	7691	23427	7147	3688	8963	9218	29016	6980	3515	8366	9204	28065	4168	2066	6026	5842	18102	7960	3720	11781	11406	34867
Neuroticism	Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M	
	4.30	11912	3.42	9653		4.26	14440	3.40	11819		4.35	13833	3.39	11675		4.21	9255	3.32	7234		4.08	17875	3.34	13974	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum; F, M, Female and Male, respectively.

Table 5.4. Cluster Demographics Using All Participants With Complete Adult Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER					MIDNORTH CLUSTER					MIDSOUTH CLUSTER					SOUTHWEST CLUSTER					SOUTHEAST CLUSTER				
Total Nind	25043					30914					29899					19246					37273				
Sex	Female	Male				Female	Male				Female	Male				Female	Male				Female	Male			
	13939	11104				17257	13657				16504	13395				10924	8322				21254	16019			
Age	Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max		
	40.00	56.20	70.30			40.20	56.51	72.50			40.10	56.92	72.90			40.20	56.09	70.40			38.90	56.68	70.60		
	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total
CIDI Depression	3887	1866	7289	7612	20654	5145	2344	8808	9196	25493	5105	2345	7898	8934	24282	3141	1405	5740	5714	16000	5613	2411	11311	11220	30555
Broad Depression	5480	2740	7262	7549	23031	7031	3632	8777	9047	28487	6891	3457	8153	9024	27525	4091	2040	5920	5727	17778	7826	3698	11527	11248	34299
Neuroticism	Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M	
	4.30	11743	3.43	9522		4.26	14225	3.41	11642		4.35	13624	3.38	11500		4.21	9117	3.33	7124		4.08	17584	3.35	13833	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum; F, M, Female and Male, respectively.

Table 5.5. Cluster Demographics Using All Participants With Catastrophic Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER					MIDNORTH CLUSTER					MIDSOUTH CLUSTER					SOUTHWEST CLUSTER					SOUTHEAST CLUSTER				
Total Nind	25590					31563					30543					19694					38009				
Sex	Female	Male				Female	Male				Female	Male				Female	Male				Female	Male			
	14262	11328				17627	13936				16850	13693				11195	8499				21733	16276			
Age	Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max			Min	Mean	Max		
	40.00	56.27	70.30			40.20	56.56	72.50			40.10	56.97	72.90			40.20	56.13	70.40			38.90	56.72	70.60		
	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total	Case F	Case M	Con F	Con M	Total
CIDI Depression	3951	1881	7454	7792	21078	5208	2361	9012	9411	25992	5166	2376	8115	9160	24817	3176	1419	5898	5842	16335	5692	2427	11593	11427	31139
Broad Depression	5613	2776	7417	7729	23535	7206	3698	8967	9245	29116	6990	3525	8377	9232	28124	4201	2074	6062	5856	18193	8023	3744	11794	11448	35009
Neuroticism	Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M		Mean F	Total F	Mean M	Total M	
	4.29	11958	3.42	9704		4.26	14494	3.40	11860		4.36	13854	3.38	11729		4.22	9298	3.31	7247		4.08	17926	3.35	14038	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum; F, M, Female and Male, respectively.

Table 6 presents demographic information on individuals within each geographical cluster; North, Mid-North, Mid-South, South-West, South-East. Demographic information for each cluster was presented when clusters were formed using all individuals responding to ALL trauma exposure questions. Table 6 was presented in separate sections for each sex.

Table 6.1. Cluster Demographics Using All Female Participants With Complete Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER			MIDNORTH CLUSTER			MIDSOUTH CLUSTER			SOUTHWEST CLUSTER			SOUTHEAST CLUSTER		
Total Nind	13433			16595			15835			10533			20599		
Age	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
	40.20	55.63	70.30	40.20	56.01	70.30	40.10	56.27	70.40	40.20	55.51	70.40	40.20	56.26	70.60
	Case	Control	Total	Case	Control	Total	Case	Control	Total	Case	Control	Total	Case	Control	Total
CIDI Depression	3726	7067	10793	4913	8534	13447	4876	7644	12520	3012	5565	8577	5402	11050	16452
Broad Depression	5225	7053	12278	6706	8513	15219	6563	7892	14455	3928	5729	9657	7530	11253	18783
Neuroticism	Mean	Total		Mean	Total		Mean	Total		Mean	Total		Mean	Total	
	4.28	11375		4.24	13748		4.33	13158		4.20	8832		4.06	17141	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum.

Table 6.2. Cluster Demographics Using Unrelated Female Participants With Complete Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER			MIDNORTH CLUSTER			MIDSOUTH CLUSTER			SOUTHWEST CLUSTER			SOUTHEAST CLUSTER		
Total Nind	11626			14729			14048			9473			18735		
Age	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
	40.20	55.56	70.30	40.30	55.98	70.30	40.10	56.25	70.40	40.20	55.48	70.40	40.20	56.20	70.60
	Case	Control	Total	Case	Control	Total	Case	Control	Total	Case	Control	Total	Case	Control	Total
CIDI Depression	3226	6125	9351	4395	7552	11947	4346	6774	11120	2710	4988	7698	4974	10030	15004
Broad Depression	4494	6114	10608	5960	7538	13498	5828	6992	12820	3551	5122	8673	6852	10230	17082
Neuroticism	Mean	Total		Mean	Total		Mean	Total		Mean	Total		Mean	Total	
	4.25	9826		4.24	12213		4.32	11689		4.20	7940		4.06	15598	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum.

Table 6.3. Cluster Demographics Using All Male Participants With Complete Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER			MIDNORTH CLUSTER			MIDSOUTH CLUSTER			SOUTHWEST CLUSTER			SOUTHEAST CLUSTER		
Total Nind	10854			13386			13126			8163			15724		
Age	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
	40.00	56.87	70.30	40.30	57.12	72.50	40.10	57.64	72.90	40.30	56.74	70.40	38.90	57.18	70.40
	Case	Control	Total	Case	Control	Total	Case	Control	Total	Case	Control	Total	Case	Control	Total
CIDI Depression	1814	7479	9293	2277	9054	11331	2282	8795	11077	1371	5623	6994	2342	11073	13415
Broad Depression	2654	7418	10072	3546	8897	12443	3365	8876	12241	1989	5634	7623	3587	11105	14692
Neuroticism	Mean	Total		Mean	Total		Mean	Total		Mean	Total		Mean	Total	
	3.41	9337		3.38	11438		3.36	11302		3.31	7019		3.32	13645	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum.

Table 6.4. Cluster Demographics Using Unrelated Male Participants With Complete Trauma Exposure Responses

DEMOGRPAHIC	NORTH CLUSTER			MIDNORTH CLUSTER			MIDSOUTH CLUSTER			SOUTHWEST CLUSTER			SOUTHEAST CLUSTER		
Total Nind	9368			11837			11643			7368			14393		
Age	**Min**	**Mean**	**Max**	**Min**	**Mean**	**Max**	**Min**	**Mean**	**Max**	**Min**	**Mean**	**Max**	**Min**	**Mean**	**Max**
	40.00	56.85	70.30	40.30	57.08	72.50	40.10	57.59	72.90	40.30	56.73	70.40	40.20	57.13	70.40
	Case	**Control**	**Total**	**Case**	**Control**	**Total**	**Case**	**Control**	**Total**	**Case**	**Control**	**Total**	**Case**	**Control**	**Total**
CIDI Depression	1575	6461	8036	2020	8007	10027	2065	7775	9840	1244	5066	6310	2131	10165	12296
Broad Depression	2287	6395	8682	3148	7855	11003	3002	7858	10860	1789	5090	6879	3253	10195	13448
Neuroticism	**Mean**	**Total**		**Mean**	**Total**		**Mean**	**Total**		**Mean**	**Total**		**Mean**	**Total**	
	3.41	8074		3.38	10118		3.36	10019		3.31	6351		3.31	12481	

Abbreviations. Nind, number of individuals; Min, minimum; Max, maximum.

Section A

OmicS-data-based Complex-trait Analysis OSCA v0.45 is a software used to analyse available omics (e.g. DNA methylation) data (Zhang et al, 2019). Three separate equations are provided to compute Omics Relationship Matrices (ORMs). In this study, we refer to ORMs as environmental relationship matrices (Es), as this is more appropriate for the data being used in this study (trauma principal components).

Algorithm 1 computes similarity by using standardised DNAm measures of all probes. Here it is important to note that the DNAm measures will be replaced with the principal component eigenvectors (PCs) of our trauma exposure measures.

Equation 1. $A_{jk} = \Sigma_i(x_{ij} - \mu_i)(x_{ik} - \mu_i) / \sigma_i^2$

With A_{jk} being the environmental similarity between individuals j and k. Where x_{ij} and x_{ik} is the **standardised** PC_i in individual j and k, respectively. μ_i and σ_i^2 are the mean and variance of PC_i over all the individuals respectively.

Algorithm 2 computes similarity by using **unstandardised** eigenvectorss and factors in the number of eigenvectors utilised. This algorithm implicitly assumes eigenvectors with smaller variance tend to have larger effects on the phenotype, and that there is no relationship between the depression/neuroticism phenotypic variance captured by the eigenvector and the variance of the eigenvector. Here, we cannot confirm these assumptions, so equation 2, was deemed inappropriate for the data we have.

Equation 2. $A_{jk} = \frac{1}{m}\Sigma_i(x_{ij} - \mu_i)(x_{ik} - \mu_i) / \sigma_i^2$

With A_{jk} being the environmental similarity between individuals j and k. Where x_{ij} and x_{ik} is the **unstandardised** PC_i in individual j and k, respectively. μ_i and σ_i^2 are the mean and variance of PC_i over all the individuals respectively and m is the number of PCs utilised.

Algorithm 3 computes similarity by iteratively standardising eigenvectors and across individuals.

Equation 3. $A_{jk} = \frac{1}{m}\Sigma_i(x_{ij} - \mu_i)(x_{ik} - \mu_i) \;/\; \Sigma_i\sigma_i^2$

With A_{jk} being the environmental similarity between individuals j and k. Where x_{ij} and x_{ik} is the **standardised** PC_i in individual j and k, respectively. μ_i and σ_i^2 are the mean and variance of PC_i over all the individuals respectively.

For our study, all algorithms were separately used to compute Es from full trauma eigenvectors (PCs), algorithm 1 and 3 were deemed the most appropriate for our data. Subtle differences were observed in outputted relationship matrices (**Figure 1**). Matrices computed using algorithm 1 present diagonal values with a mean of 1, whereas matrices computed using algorithm 3 present diagonal values of 1.

Individuals	1	2	3	4...
1	0.989			
2	0.002	0.758		
3	0.031	0.004	0.632	
4...	0.013	0.003	0.031	0.944

Individuals	1	2	3	4...
1	1.00			
2	0.002	1.00		
3	0.301	0.104	1.00	
4...	0.003	0.023	0.231	1.00

Figure SA1. Trauma Exposure Relationship Matrices. Matrix diagonals represent similarity between an individual with themselves. Offdiagonals represent pairwise similarity between individuals within the sample. Here, similarity is computed using available trauma exposure principal components using OSCA software. Left, shows a representative similarity matrix where similarity is computed using Algorithm 1. Right, shows a representative similarity matrix where similarity is computed using Algorithm 3. Please note values specified are not obtained from real matrices.

Offdiagonal values of the computed Es and genomic relationship matrices (Gs) were plotted to visualise the relationship between values (**Figure 2**).

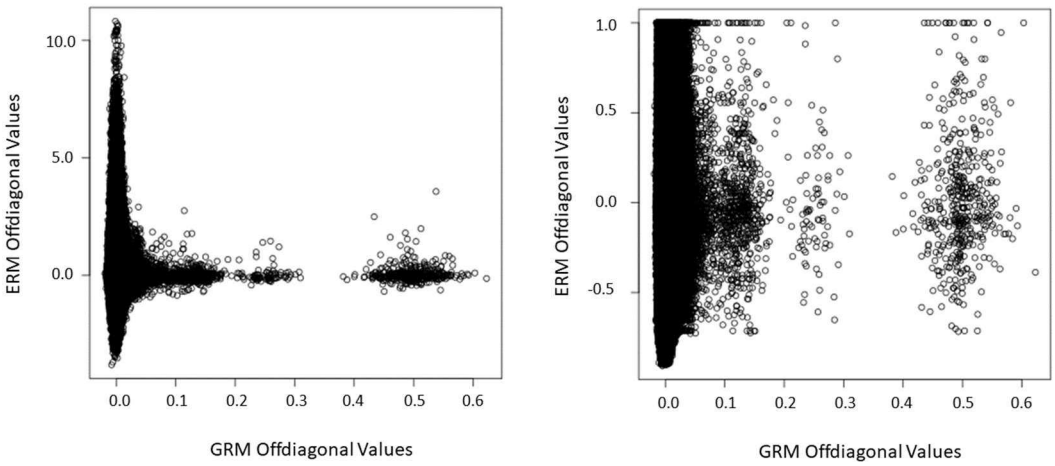


***Figure SA2*.** Offdiagonal Values of UKB North Geographical Cluster Genomic and Environmental Relationship Matrices. Left, Environmental Relationship Matrix (ERM) computed using Algorithm 1 of OSCA software. Right, ERM computed using Algorithm 3 of OSCA software. The x-axis represent offdiagonal genomic similarity values, cluster points surrounding points 0.2-0.3 and 0.5 show points of relatives. The y-axis represents offdiagonal trauma exposure similarity values.

Zhang and colleagues (2019) demonstrated, with simulation analyses, that mixed linear model anlaysis results obtained from matrices computed with algorithm 1 and 3 showed minimal differences. Similarly, we show that GREML outputs were negligibly different between the algorithms of interest (**Table 7**).

Here, algorithm 1 is opted when computing the ERMs used for downstream analyses. The mixed linear model results are similar and the plots from **figure 2** suggest values in line with expectations.

Table 7 presents results of mixed linear models utilising genomic relationship matrices, trauma exposure relationship matrices and genome-by-trauma exposure relationship matrices as random effects. Results include estimates of CIDI depression variance attributable to each random effect, standard errors, log-ratio test values and log-ratio test p-values signifying model utility. Table 7 presents results of mixed linear models whereby trauma exposure relationship matrices were computed using the 3 different algorithms available within OmicS-data-based Complex-trait Analysis OSCA software.

Table 7. Mixed Linear Model Results of Proportion of CIDI Depression Variance Attributable to Environmental Relationship Matrices Computed Using Different OSCA Algorithms

OSCA	Trait	Model	Source	NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
Algorithm				Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
		G	G	0.195	0.037	28.956	0.000	0.134	0.030	20.793	0.000	0.114	0.030	15.209	0.000	0.218	0.048	21.359	0.000	0.121	0.027	20.707	0.000
		E	E	0.145	0.047	1801.636	0.000	0.159	0.050	2528.312	0.000	0.147	0.047	2299.555	0.000	0.140	0.046	2299.555	0.000	0.159	0.051	2969.008	0.000
		G + E	G	0.146	0.034			0.084	0.027			0.070	0.027			0.141	0.043			0.063	0.024		
1	CIDI		E	0.144	0.047	1792.306	0.000	0.159	0.050	2517.636	0.000	0.147	0.047	2291.178	0.000	0.140	0.045	1373.706	0.000	0.159	0.051	2955.386	0.000
		G x E	G	0.148	0.033			0.082	0.026			0.065	0.027			0.135	0.043			0.044	0.044		
			E	0.150	0.048			0.162	0.051			0.150	0.048			0.140	0.045			0.163	0.052		
			GxE	0.205	0.018	159.722	0.000	0.141	0.015	106.471	0.000	0.126	0.014	90.449	0.000	0.166	0.019	80.735	0.000	0.259	0.023	155.156	0.000
		G	G	0.195	0.037	28.956	0.000	0.134	0.030	20.793	0.000	0.114	0.030	15.209	0.000	0.218	0.048	21.359	0.000	0.014	0.003	20.707	0.000
		E	E	0.169	0.056	1362.035	0.000	0.175	0.057	1740.767	0.000	0.153	0.050	1599.097	0.000	0.136	0.046	883.480	0.000	0.144	0.048	1948.929	0.000
		G + E	G	0.163	0.034			0.085	0.026			0.074	0.027			0.169	0.044			0.069	0.024		
3	CIDI		E	0.168	0.055	1358.782	0.000	0.174	0.057	2517.636	0.000	0.152	0.050	1591.741	0.000	0.135	0.046	877.686	0.000	0.143	0.048	1936.461	0.000
		G x E	G	0.164	0.034			0.084	0.026			0.074	0.027			0.169	0.044			0.060	0.046		
			E	0.168	0.055			0.174	0.057			0.152	0.050			0.135	0.046			0.141	0.048		
			GxE	0.065	0.092	0.445	0.252	0.119	0.075	2.387	0.061	0.025	0.073	0.111	0.370	0.021	0.120	0.028	0.434	0.077	0.132	0.335	0.281
		G	G	0.195	0.037	28.956	0.000	0.134	0.030	20.793	0.000	0.114	0.030	15.209	0.000	0.218	0.048	21.359	0.000	0.014	0.003	20.707	0.000
		E	E	0.173	0.002	1811.054	0.000	0.176	0.002	2539.357	0.000	0.180	0.002	2310.492	0.000	0.176	0.002	1396.145	0.000	0.166	0.001	2980.101	0.000
		G + E	G	0.151	0.035			0.088	0.028			0.073	0.028			0.148	0.045			0.066	0.025		
2	CIDI		E	0.094	0.032	1801.649	0.000	0.095	0.032	2528.653	0.000	0.088	0.029	2302.109	0.000	0.078	0.027	1385.673	0.000	0.093	0.031	2966.475	0.000
		G x E	G	0.154	0.035			0.087	0.028			0.068	0.028			0.142	0.045			0.048	0.046		
			E	0.096	0.032			0.096	0.032			0.089	0.030			0.078	0.027			0.097	0.033		
			GxE	0.164	0.017	110.361	0.000	0.093	0.013	49.731	0.000	0.092	0.013	51.669	0.000	0.125	0.019	45.357	0.000	0.212	0.022	103.159	0.000

Abbreviations. CIDI, Composite International Diagnostic Inventory Depression; OSCA, OmicS-based-data Complex-trait Analysis; G, Genomic Relationship Matrix; E, Environmental (trauma exposure) Relationship Matrix; GxE, Genome-by-Trauma Exposure Interaction Relationship Matrix; SE, standard error; LRT, log-ratio test; P, P-Value.

Table 8 presents heritability estimates obtained for trauma exposure obtained from Haseman-Elston regression models. The table presents estimates of variance attributable to the genomic relationship matrix, standard errors and P-Values from analyses conducted on the whole and unrelated samples separately. Heritability estimates were obtained separately for full trauma, childhood, adult and catastrophic trauma exposure.

Table 8. Trauma Exposure Heritability Estimates

		Full Sample				Unrelated Sample			
Trauma	**Cluster**	**Variance**	**SE**	**P**	**N**	**Variance**	**SE**	**P**	**N**
Full Trauma	North	0.173	0.021		24287	0.174	0.024		20994
	Midnorth	0.166	0.017		29981	0.149	0.019		26566
	Midsouth	0.163	0.017		28961	0.160	0.020		25691
	Southwest	0.159	0.026		18696	0.149	0.030		16841
	Southeast	0.177	0.014		36323	0.168	0.016		33128
	meta-A	0.169	0.008	0.000		0.161	0.009	0.000	
Childhood	North	0.156	0.021		24287	0.138	0.024		20994
	Midnorth	0.125	0.017		29981	0.108	0.019		26566
	Midsouth	0.151	0.017		28961	0.140	0.019		25691
	Southwest	0.151	0.026		18696	0.140	0.029		16841
	Southeast	0.149	0.014		36323	0.134	0.015		33128
	meta-A	0.145	0.008	0.000		0.131	0.009	0.000	
Adult	North	0.063	0.020		24287	0.073	0.024		20994
	Midnorth	0.068	0.017		29981	0.061	0.019		26566
	Midsouth	0.064	0.017		28961	0.077	0.019		25691
	Southwest	0.036	0.026		18696	0.054	0.030		16841
	Southeast	0.066	0.014		36323	0.068	0.015		33128
	meta-A	0.063	0.008	0.000		0.068	0.009	0.000	
Catastrophic	North	0.100	0.020		24287	0.099	0.024		20994
	Midnorth	0.093	0.017		29981	0.083	0.019		26566
	Midsouth	0.119	0.017		28961	0.120	0.019		25691
	Southwest	0.135	0.026		18696	0.129	0.029		16841
	Southeast	0.098	0.014		36323	0.093	0.015		33128
	meta-A	0.105	0.008	0.000		0.101	0.009	0.000	

Abbreviations. SE, standard error, P, P-Value; N, sample size; meta-A, Meta-analysed estimates of variance and SEs.

Table 9 presents estimates of genetic correlations between trauma exposure and depression/neuroticism variables. Tables present estimates, standard errors and P-Values.

Table 9. Trauma Exposure and Depression/Neuroticism Genetic Correlations

		CIDI			BROAD			NEUROTICISM		
TRAUMA	**CLUSTER**	**Est**	**SE**	**P**	**Est**	**SE**	**P**	**Est**	**SE**	**P**
TRAUMA	North	0.506	0.171		0.475	0.216		0.289	0.219	
	Midnorth	0.721	0.167		0.340	0.156		0.277	0.116	
	Midsouth	0.636	0.198		0.413	0.187		0.229	0.141	
	Southwest	0.614	0.276		0.730	0.317		0.408	0.276	
	Southeast	0.672	0.186		0.307	0.147		0.447	0.110	
	Meta-A	0.632	0.085	0.000	0.390	0.082	0.000	0.333	0.064	0.000
CHILDHOOD	North	0.430	0.181		0.374	0.226		0.229	0.230	
	Midnorth	0.765	0.187		0.404	0.177		0.369	0.131	
	Midsouth	0.668	0.229		0.465	0.216		0.282	0.155	
	Southwest	0.570	0.290		0.621	0.331		0.473	0.277	
	Southeast	0.604	0.184		0.141	0.154		0.332	0.120	
	Meta-A	0.605	0.091	0.000	0.337	0.090	0.000	0.332	0.071	0.000
ADULT	North	0.638	0.267		0.567	0.330		0.351	0.331	
	Midnorth	0.668	0.243		0.234	0.230		0.091	0.175	
	Midsouth	0.599	0.280		0.451	0.269		0.272	0.214	
	Southwest	0.956	0.748		1.336	0.950		0.416	0.658	
	Southeast	0.630	0.312		0.221	0.266		0.476	0.201	
	Meta-A	0.647	0.134	0.000	0.358	0.132	0.007	0.274	0.105	0.009
CATASTROPHIC	North	0.527	0.230		0.553	0.295		0.217	0.295	
	Midnorth	0.657	0.213		0.224	0.202		0.225	0.161	
	Midsouth	0.537	0.240		0.092	0.215		-0.080	0.166	
	Southwest	0.307	0.306		0.013	0.332		0.225	0.285	
	Southeast	0.535	0.208		0.489	0.163		0.372	0.137	
	Meta-A	0.536	0.104	0.000	0.310	0.098	0.002	0.204	0.081	0.012

Abbreviations. CIDI, Composite International Diagnostic Inventory definition of depression; Est, estimate; SE, standard error; P, P-Value; Meta-A, meta analysed values of estimates and SEs.

Tables 10-22 presents results of mixed linear models using genomic relationship matrices (Gs), environmental relationship matrices (Es) using trauma exposure eigenvectors (PCs), and genome-by-trauma exposure interaction relationship matrices (GxE) as random effects for each geographical cluster. Results include estimates of trait variance attributable to the G, E and GxE, standard errors and P-Values. Results also include meta-analysed estimates and standard errors. Each table presents results of models utilising either different samples (unrelated, female only, males only) or different Es (using full trauma exposure principal components PCs, childhood trauma exposure PCs, adult trauma exposure PCs etc.). Details of the samples and Es utilised are available in table headings.

Table 10. Mixed Linear Model Results Including Es of Full Trauma Exposure Principal Components.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.167	0.017	0.000	0.229	0.044	28.956	0.000	0.158	0.035	20.793	0.000	0.134	0.035	15.209	0.000	0.256	0.056	21.359	0.000	0.142	0.032	20.707	0.000
	E	E	0.176	0.025	0.000	0.170	0.055	1801.636	0.000	0.187	0.059	2528.312	0.000	0.173	0.055	2299.555	0.000	0.165	0.053	1384.156	0.000	0.187	0.060	2969.008	0.000
	G + E	G	0.104	0.015	0.000	0.171	0.040			0.098	0.031			0.082	0.032			0.166	0.051			0.074	0.028		
		E	0.175	0.025	0.000	0.169	0.055	1792.306	0.000	0.186	0.059	2517.636	0.000	0.172	0.055	2291.178	0.000	0.164	0.053	1373.706	0.000	0.187	0.060	2955.386	0.000
	G x E	G	0.098	0.015	0.000	0.174	0.039			0.096	0.031			0.076	0.031			0.158	0.050			0.062	0.027		
		E	0.178	0.026	0.000	0.176	0.057			0.190	0.060			0.176	0.056			0.165	0.053			0.188	0.060		
		GxE	0.201	0.009	0.000	0.241	0.021	159.722	0.000	0.165	0.017	106.471	0.000	0.148	0.017	90.449	0.000	0.195	0.023	80.735	0.000	0.294	0.020	295.840	0.000
BROAD	G	G	0.135	0.014	0.000	0.154	0.036	19.061	0.000	0.160	0.029	32.247	0.000	0.110	0.030	14.595	0.000	0.159	0.048	11.076	0.000	0.118	0.026	21.365	0.000
	E	E	0.105	0.016	0.000	0.104	0.035	1301.673	0.000	0.106	0.035	1657.122	0.000	0.100	0.034	1537.483	0.000	0.107	0.036	1021.711	0.000	0.110	0.037	2094.855	0.000
	G + E	G	0.098	0.013	0.000	0.122	0.034			0.126	0.027			0.076	0.028			0.119	0.045			0.073	0.024		
		E	0.105	0.016	0.000	0.104	0.035	1296.487	0.000	0.106	0.035	1648.216	0.000	0.100	0.033	1531.056	0.000	0.107	0.036	1017.853	0.000	0.110	0.037	2083.518	0.000
	G x E	G	0.097	0.013	0.000	0.121	0.034			0.127	0.027			0.076	0.028			0.116	0.044			0.074	0.024		
		E	0.106	0.016	0.000	0.105	0.035			0.107	0.036			0.100	0.034			0.107	0.036			0.111	0.037		
		GxE	0.048	0.006	0.000	0.051	0.015	12.320	0.000	0.027	0.013	4.269	0.019	0.021	0.013	2.648	0.052	0.072	0.018	15.423	0.000	0.087	0.014	41.422	0.000
NEURO	G	G	0.150	0.009	0.000	0.102	0.024	19.868	0.000	0.174	0.020	81.415	0.000	0.147	0.021	54.107	0.000	0.148	0.031	24.168	0.000	0.160	0.017	94.024	0.000
	E	E	0.060	0.009	0.000	0.061	0.020	1221.472	0.000	0.066	0.022	1539.876	0.000	0.061	0.021	1420.245	0.000	0.054	0.018	789.554	0.000	0.061	0.020	1906.571	0.000
	G + E	G	0.132	0.009	0.000	0.092	0.022			0.144	0.019			0.131	0.019			0.129	0.029			0.144	0.016		
		E	0.060	0.009	0.000	0.061	0.020	1220.160	0.000	0.065	0.022	1522.604	0.000	0.061	0.020	1416.278	0.000	0.054	0.018	786.084	0.000	0.061	0.020	1899.179	0.000
	G x E	G	0.132	0.009	0.000	0.091	0.022			0.147	0.019			0.134	0.019			0.131	0.029			0.143	0.016		
		E	0.061	0.009	0.000	0.062	0.021			0.066	0.022			0.062	0.021			0.054	0.019			0.062	0.021		
		GxE	0.072	0.004	0.000	0.063	0.010	51.022	0.000	0.075	0.009	80.387	0.000	0.073	0.009	76.109	0.000	0.080	0.012	54.876	0.000	0.071	0.008	84.435	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 11. Mixed Linear Model Results Including Es of Childhood Trauma Exposure Principal Components.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.163	0.016	0.000	0.219	0.042	28.960	0.000	0.140	0.033	17.998	0.000	0.144	0.034	18.766	0.000	0.243	0.054	21.238	0.000	0.144	0.030	23.045	0.000
	E	E	0.101	0.027	0.000	0.092	0.055	1015.673	0.000	0.104	0.062	1482.339	0.000	0.097	0.058	1372.832	0.000	0.102	0.061	918.242	0.000	0.112	0.067	1866.562	0.000
	G + E	G	0.113	0.015	0.000	0.174	0.040			0.095	0.031			0.099	0.032			0.171	0.051			0.090	0.029		
		E	0.100	0.027	0.000	0.091	0.055	1006.830	0.000	0.104	0.062	1473.776	0.000	0.097	0.058	1363.960	0.000	0.101	0.061	908.968	0.000	0.112	0.067	1853.858	0.000
	G x E	G	0.116	0.017	0.000	0.177	0.040			0.096	0.031			0.095	0.032			0.164	0.050			0.073	0.053		
		E	0.102	0.027	0.000	0.093	0.056			0.106	0.063			0.098	0.059			0.104	0.062			0.114	0.068		
		GxE	0.084	0.006	0.000	0.096	0.013	66.250	0.000	0.079	0.011	61.123	0.000	0.056	0.011	31.394	0.000	0.100	0.015	55.787	0.000	0.126	0.017	68.423	0.000
BROAD	G	G	0.135	0.014	0.000	0.149	0.034	19.829	0.000	0.142	0.028	27.752	0.000	0.115	0.028	17.267	0.000	0.167	0.046	13.392	0.000	0.128	0.025	27.266	0.000
	E	E	0.059	0.016	0.000	0.056	0.034	749.802	0.000	0.058	0.035	949.004	0.000	0.059	0.036	988.282	0.000	0.061	0.038	636.928	0.000	0.064	0.039	1308.214	0.000
	G + E	G	0.109	0.013	0.000	0.122	0.033			0.116	0.027			0.093	0.027			0.134	0.044			0.103	0.024		
		E	0.059	0.016	0.000	0.055	0.034	744.447	0.000	0.057	0.035	941.854	0.000	0.059	0.036	983.444	0.000	0.061	0.037	632.895	0.000	0.064	0.039	1300.637	0.000
	G x E	G	0.109	0.015	0.000	0.122	0.033			0.116	0.027			0.093	0.027			0.134	0.044			0.086	0.045		
		E	0.059	0.016	0.000	0.056	0.034			0.058	0.035			0.059	0.036			0.062	0.038			0.063	0.039		
		GxE	0.020	0.004	0.000	0.022	0.009	6.399	0.006	0.015	0.008	3.413	0.032	0.006	0.008	0.585	0.222	0.038	0.011	11.776	0.000	0.041	0.012	13.243	0.000
NEURO	G	G	0.150	0.009	0.000	0.099	0.023	20.378	0.000	0.174	0.019	89.340	0.000	0.153	0.020	62.963	0.000	0.141	0.030	23.680	0.000	0.159	0.016	98.967	0.000
	E	E	0.041	0.011	0.000	0.043	0.026	932.418	0.000	0.044	0.027	1119.272	0.000	0.042	0.025	1049.332	0.000	0.036	0.022	583.258	0.000	0.040	0.025	1326.980	0.000
	G + E	G	0.138	0.009	0.000	0.095	0.022			0.153	0.019			0.143	0.019			0.129	0.029			0.149	0.016		
		E	0.040	0.011	0.000	0.043	0.026	932.508	0.000	0.044	0.027	1105.278	0.000	0.042	0.025	1047.271	0.000	0.036	0.022	580.974	0.000	0.040	0.025	1322.971	0.000
	G x E	G	0.132	0.010	0.000	0.092	0.022			0.156	0.019			0.144	0.019			0.127	0.028			0.128	0.029		
		E	0.041	0.011	0.000	0.044	0.027			0.045	0.027			0.042	0.026			0.037	0.022			0.041	0.025		
		GxE	0.049	0.003	0.000	0.048	0.007	66.900	0.000	0.050	0.006	82.226	0.000	0.048	0.006	78.308	0.000	0.052	0.008	54.896	0.000	0.051	0.008	48.876	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 12. Mixed Linear Model Results Including Es of Adult Trauma Exposure Principal Components.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.167	0.016	0.000	0.237	0.043	33.288	0.000	0.152	0.034	20.905	0.000	0.139	0.034	17.463	0.000	0.248	0.054	21.519	0.000	0.139	0.031	21.132	0.000
	E	E	0.113	0.030	0.000	0.111	0.066	1217.822	0.000	0.123	0.072	1667.859	0.000	0.112	0.066	1476.666	0.000	0.103	0.062	845.731	0.000	0.118	0.070	1872.734	0.000
	G + E	G	0.132	0.016	0.000	0.206	0.041			0.122	0.032			0.101	0.032			0.217	0.052			0.101	0.029		
		E	0.112	0.030	0.000	0.110	0.066	1212.851	0.000	0.122	0.072	1662.461	0.000	0.111	0.066	1469.484	0.000	0.103	0.062	842.647	0.000	0.118	0.070	1864.260	0.000
	G x E	G	0.129	0.015	0.000	0.203	0.040			0.122	0.032			0.100	0.032			0.208	0.051			0.097	0.028		
		E	0.114	0.030	0.000	0.113	0.067			0.125	0.073			0.115	0.068			0.104	0.067			0.120	0.071		
		GxE	0.081	0.005	0.000	0.088	0.013	59.266	0.000	0.056	0.011	31.401	0.000	0.066	0.011	44.934	0.000	0.075	0.014	29.602	0.000	0.142	0.014	149.724	0.000
BROAD	G	G	0.139	0.014	0.000	0.167	0.035	23.626	0.000	0.153	0.028	31.392	0.000	0.118	0.029	17.801	0.000	0.157	0.046	11.612	0.000	0.125	0.025	25.795	0.000
	E	E	0.069	0.019	0.000	0.068	0.042	911.063	0.000	0.071	0.043	1140.854	0.000	0.065	0.040	1025.708	0.000	0.069	0.042	669.480	0.000	0.070	0.043	1351.899	0.000
	G + E	G	0.115	0.013	0.000	0.144	0.034			0.141	0.027			0.095	0.028			0.138	0.045			0.090	0.024		
		E	0.068	0.019	0.000	0.068	0.041	906.831	0.000	0.071	0.043	1138.513	0.000	0.065	0.040	1020.536	0.000	0.069	0.042	667.662	0.000	0.070	0.042	1341.111	0.000
	G x E	G	0.115	0.013	0.000	0.142	0.034			0.141	0.027			0.096	0.028			0.139	0.045			0.090	0.024		
		E	0.069	0.019	0.000	0.069	0.042			0.071	0.043			0.066	0.040			0.069	0.042			0.071	0.043		
		GxE	0.021	0.004	0.000	0.028	0.009	10.439	0.001	0.002	0.007	0.056	0.407	0.015	0.008	3.822	0.025	0.026	0.011	5.972	0.007	0.042	0.009	24.719	0.000
NEURO	G	G	0.154	0.009	0.000	0.098	0.023	19.166	0.000	0.175	0.020	87.683	0.000	0.162	0.020	67.484	0.000	0.147	0.030	25.455	0.000	0.165	0.017	104.280	0.000
	E	E	0.035	0.010	0.000	0.033	0.020	676.810	0.000	0.037	0.023	898.821	0.000	0.038	0.023	886.390	0.000	0.031	0.019	466.142	0.000	0.037	0.023	1194.391	0.000
	G + E	G	0.141	0.009	0.000	0.091	0.022			0.159	0.019			0.148	0.020			0.139	0.029			0.152	0.016		
		E	0.035	0.010	0.000	0.033	0.020	675.491	0.000	0.037	0.023	888.207	0.000	0.037	0.023	880.641	0.000	0.031	0.019	464.832	0.000	0.037	0.023	1185.494	0.000
	G x E	G	0.142	0.009	0.000	0.092	0.022			0.160	0.019			0.147	0.020			0.142	0.029			0.151	0.016		
		E	0.035	0.010	0.000	0.033	0.020			0.038	0.023			0.038	0.023			0.032	0.020			0.038	0.023		
		GxE	0.038	0.003	0.000	0.035	0.006	41.298	0.000	0.043	0.006	64.331	0.000	0.035	0.006	41.759	0.000	0.039	0.008	31.966	0.000	0.039	0.005	61.135	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 13. Mixed Linear Model Results Including Es of Full Catastrophic Trauma Exposure Principal Components.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.163	0.016	0.000	0.223	0.042	30.295	0.000	0.135	0.033	16.870	0.000	0.143	0.034	18.819	0.000	0.239	0.053	20.615	0.000	0.148	0.030	24.511	0.000
	E	E	0.053	0.013	0.000	0.052	0.030	547.825	0.000	0.056	0.032	825.414	0.000	0.054	0.030	764.793	0.000	0.052	0.029	444.673	0.000	0.051	0.029	811.485	0.000
	G + E	G	0.134	0.016	0.000	0.202	0.041			0.098	0.032			0.117	0.033			0.213	0.052			0.117	0.029		
		E	0.053	0.013	0.000	0.052	0.029	544.327	0.000	0.056	0.031	818.099	0.000	0.053	0.030	759.110	0.000	0.051	0.029	441.317	0.000	0.051	0.029	803.426	0.000
	G x E	G	0.130	0.016	0.000	0.203	0.040			0.094	0.032			0.116	0.033			0.209	0.052			0.110	0.029		
		E	0.053	0.013	0.000	0.053	0.030			0.056	0.032			0.054	0.030			0.051	0.029			0.051	0.029		
		GxE	0.074	0.006	0.000	0.069	0.014	24.637	0.000	0.075	0.012	39.458	0.000	0.058	0.012	24.359	0.000	0.082	0.016	27.677	0.000	0.092	0.013	53.042	0.000
BROAD	G	G	0.137	0.013	0.000	0.148	0.034	19.739	0.000	0.140	0.028	27.559	0.000	0.123	0.028	19.773	0.000	0.162	0.046	12.620	0.000	0.133	0.025	30.092	0.000
	E	E	0.027	0.007	0.000	0.030	0.017	369.925	0.000	0.029	0.016	496.765	0.000	0.023	0.013	370.424	0.000	0.026	0.015	252.145	0.000	0.029	0.017	557.863	0.000
	G + E	G	0.126	0.013	0.000	0.144	0.034			0.126	0.027			0.114	0.028			0.155	0.045			0.117	0.024		
		E	0.026	0.007	0.000	0.030	0.017	369.656	0.000	0.028	0.016	492.455	0.000	0.023	0.013	368.628	0.000	0.026	0.015	251.554	0.000	0.029	0.016	552.287	0.000
	G x E	G	0.126	0.013	0.000	0.142	0.034			0.126	0.027			0.114	0.028			0.153	0.045			0.117	0.024		
		E	0.026	0.007	0.000	0.027	0.015			0.028	0.016			0.023	0.013			0.026	0.015			0.029	0.016		
		GxE	0.008	0.005	0.090	0.000	0.010	0.000	0.500	0.012	0.009	1.639	0.100	0.000	0.009	0.000	0.494	0.020	0.013	2.522	0.056	0.011	0.010	1.203	0.136
NEURO	G	G	0.152	0.009	0.000	0.104	0.023	22.301	0.000	0.174	0.019	89.921	0.000	0.155	0.020	65.198	0.000	0.140	0.030	23.527	0.000	0.161	0.016	102.537	0.000
	E	E	0.007	0.002	0.000	0.007	0.004	134.114	0.000	0.010	0.006	243.103	0.000	0.006	0.004	145.426	0.000	0.007	0.004	94.148	0.000	0.007	0.004	200.821	0.000
	G + E	G	0.149	0.009	0.000	0.106	0.022			0.168	0.019			0.158	0.020			0.135	0.029			0.156	0.016		
		E	0.007	0.002	0.000	0.008	0.004	135.479	0.000	0.010	0.006	238.450	0.000	0.006	0.004	149.255	0.000	0.007	0.004	92.922	0.000	0.007	0.004	195.251	0.000
	G x E	G	0.149	0.009	0.000	0.105	0.022			0.168	0.019			0.160	0.020			0.136	0.029			0.156	0.016		
		E	0.007	0.002	0.000	0.008	0.005			0.010	0.006			0.006	0.004			0.007	0.004			0.007	0.004		
		GxE	0.010	0.003	0.000	0.006	0.007	0.747	0.194	0.016	0.007	5.852	0.008	0.013	0.006	3.997	0.023	0.015	0.008	3.502	0.031	0.005	0.006	0.854	0.178

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 14. Mixed Linear Model Results Including Es of Full Trauma Exposure Principal Components using Unrelated Individuals.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.154	0.019	0.000	0.207	0.052	16.837	0.000	0.155	0.040	15.213	0.000	0.146	0.040	14.014	0.000	0.212	0.064	11.228	0.000	0.118	0.034	12.292	0.000
	E	E	0.148	0.021	0.000	0.142	0.046	1539.533	0.000	0.161	0.051	2282.157	0.000	0.148	0.047	2070.893	0.000	0.137	0.045	1218.759	0.000	0.160	0.051	2751.651	0.000
	G + E	G	0.094	0.017	0.000	0.159	0.047			0.098	0.036			0.087	0.036			0.137	0.058			0.056	0.031		
		E	0.174	0.025	0.000	0.167	0.054	1534.810	0.000	0.188	0.060	2274.722	0.000	0.173	0.055	2062.838	0.000	0.161	0.052	1213.166	0.000	0.188	0.060	2742.841	0.000
	G x E	G	0.089	0.017	0.000	0.164	0.046			0.098	0.035			0.085	0.036			0.128	0.057			0.043	0.029		
		E	0.177	0.025	0.000	0.172	0.056			0.192	0.061			0.177	0.056			0.161	0.053			0.189	0.061		
		GxE	0.199	0.009	0.000	0.243	0.023	139.913	0.000	0.161	0.018	90.888	0.000	0.148	0.018	81.912	0.000	0.182	0.024	64.257	0.000	0.299	0.021	285.257	0.000
BROAD	G	G	0.119	0.015	0.000	0.159	0.043	14.241	0.000	0.146	0.033	20.110	0.000	0.122	0.034	13.515	0.000	0.123	0.044	8.017	0.002	0.093	0.023	16.805	0.000
	E	E	0.086	0.013	0.000	0.083	0.028	1113.781	0.000	0.086	0.029	1478.493	0.000	0.083	0.028	1396.299	0.000	0.087	0.029	921.338	0.000	0.091	0.030	1960.973	0.000
	G + E	G	0.097	0.015	0.000	0.137	0.040			0.114	0.031			0.085	0.032			0.134	0.051			0.068	0.026		
		E	0.106	0.016	0.000	0.103	0.035	1111.684	0.000	0.106	0.036	1472.698	0.000	0.102	0.034	1390.426	0.000	0.107	0.036	920.553	0.000	0.112	0.037	1951.169	0.000
	G x E	G	0.097	0.015	0.000	0.136	0.040			0.115	0.031			0.085	0.032			0.132	0.050			0.068	0.026		
		E	0.107	0.016	0.000	0.104	0.035			0.107	0.036			0.102	0.034			0.108	0.036			0.113	0.038		
		GxE	0.049	0.007	0.000	0.053	0.016	11.372	0.000	0.025	0.013	3.156	0.038	0.025	0.014	3.204	0.037	0.070	0.019	12.984	0.000	0.091	0.014	41.230	0.000
NEURO	G	G	0.138	0.011	0.000	0.081	0.028	8.975	0.001	0.155	0.023	49.616	0.000	0.142	0.023	38.894	0.000	0.132	0.035	14.886	0.000	0.152	0.019	70.069	0.000
	E	E	0.059	0.009	0.000	0.059	0.020	1016.525	0.000	0.066	0.022	1364.508	0.000	0.063	0.021	1294.966	0.000	0.051	0.017	667.636	0.000	0.062	0.021	1776.746	0.000
	G + E	G	0.122	0.010	0.000	0.073	0.026			0.127	0.021			0.126	0.022			0.125	0.033			0.138	0.018		
		E	0.059	0.009	0.000	0.059	0.020	1015.899	0.000	0.065	0.022	1352.817	0.000	0.063	0.021	1291.776	0.000	0.051	0.017	667.619	0.000	0.062	0.021	1771.728	0.000
	G x E	G	0.123	0.010	0.000	0.072	0.026			0.129	0.021			0.129	0.022			0.131	0.033			0.136	0.018		
		E	0.060	0.009	0.000	0.060	0.020			0.067	0.022			0.064	0.021			0.052	0.018			0.063	0.021		
		GxE	0.074	0.004	0.000	0.065	0.010	48.482	0.000	0.076	0.010	72.565	0.000	0.073	0.010	68.513	0.000	0.076	0.012	45.650	0.000	0.077	0.009	88.384	0.000

Abbreviations. ERM, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 15. Mixed Linear Model Results Including Es of Full Trauma Exposure Principal Components Pre-Corrected for Genomic Relationship Matrix.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.167	0.017	0.000	0.229	0.044	28.956	0.000	0.158	0.035	20.793	0.000	0.134	0.035	15.209	0.000	0.256	0.056	21.359	0.000	0.142	0.032	20.707	0.000
	E	E	0.170	0.024	0.000	0.165	0.053	1776.693	0.000	0.180	0.057	2490.465	0.000	0.167	0.054	2269.990	0.000	0.160	0.052	1369.547	0.000	0.180	0.058	2914.130	0.000
	G + E	G	0.099	0.014	0.000	0.179	0.040			0.087	0.027			0.074	0.027			0.170	0.051			0.078	0.028		
		E	0.157	0.023	0.000	0.163	0.053	1769.053	0.000	0.152	0.049	2480.528	0.000	0.141	0.045	2262.515	0.000	0.159	0.052	1359.598	0.000	0.179	0.058	2901.274	0.000
	G x E	G	0.103	0.015	0.000	0.182	0.039			0.101	0.031			0.082	0.031			0.162	0.050			0.066	0.027		
		E	0.171	0.025	0.000	0.169	0.055			0.182	0.058			0.169	0.054			0.159	0.052			0.179	0.058		
		GxE	0.205	0.009	0.000	0.243	0.021	158.523	0.000	0.168	0.018	107.173	0.000	0.152	0.017	92.227	0.000	0.201	0.023	83.884	0.000	0.300	0.021	300.562	0.000
BROAD	G	G	0.135	0.014	0.000	0.154	0.036	19.061	0.000	0.160	0.029	32.247	0.000	0.110	0.030	14.595	0.000	0.159	0.048	11.076	0.000	0.118	0.026	21.365	0.000
	E	E	0.102	0.015	0.000	0.101	0.034	1286.139	0.000	0.102	0.034	1635.383	0.000	0.097	0.032	1517.333	0.000	0.104	0.035	1013.179	0.000	0.105	0.035	2054.566	0.000
	G + E	G	0.088	0.012	0.000	0.124	0.034			0.104	0.022			0.061	0.022			0.118	0.045			0.072	0.024		
		E	0.091	0.014	0.000	0.100	0.034	1281.304	0.000	0.082	0.027	1626.904	0.000	0.078	0.026	1510.685	0.000	0.103	0.035	1009.162	0.000	0.105	0.035	2042.577	0.000
	G x E	G	0.098	0.013	0.000	0.123	0.034			0.128	0.027			0.076	0.028			0.116	0.045			0.072	0.024		
		E	0.102	0.015	0.000	0.101	0.034			0.102	0.034			0.096	0.032			0.104	0.035			0.106	0.035		
		GxE	0.050	0.006	0.000	0.052	0.015	12.608	0.000	0.028	0.013	4.625	0.016	0.022	0.013	2.913	0.044	0.075	0.018	16.759	0.000	0.090	0.014	42.857	0.000
NEURO	G	G	0.150	0.009	0.000	0.102	0.024	19.868	0.000	0.174	0.020	81.415	0.000	0.147	0.021	54.107	0.000	0.148	0.031	24.168	0.000	0.160	0.017	94.024	0.000
	E	E	0.058	0.009	0.000	0.059	0.020	1213.783	0.000	0.063	0.021	1518.197	0.000	0.059	0.020	1415.037	0.000	0.052	0.018	784.201	0.000	0.059	0.020	1884.675	0.000
	G + E	G	0.132	0.009	0.000	0.093	0.022			0.147	0.019			0.131	0.020			0.130	0.029			0.145	0.016		
		E	0.058	0.009	0.000	0.059	0.020	1212.593	0.000	0.062	0.021	1502.818	0.000	0.059	0.020	1410.217	0.000	0.052	0.018	780.807	0.000	0.058	0.020	1877.945	0.000
	G x E	G	0.133	0.009	0.000	0.092	0.022			0.150	0.019			0.133	0.019			0.132	0.029			0.144	0.016		
		E	0.059	0.009	0.000	0.060	0.020			0.064	0.021			0.060	0.020			0.053	0.018			0.059	0.020		
		GxE	0.073	0.004	0.000	0.063	0.010	51.355	0.000	0.076	0.009	80.835	0.000	0.074	0.009	75.554	0.000	0.081	0.012	55.533	0.000	0.073	0.008	86.113	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 16. Mixed Linear Model Results Including Es of Full Trauma Exposure Principal Component 1.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.167	0.017	0.000	0.229	0.044	28.956	0.000	0.158	0.035	20.793	0.000	0.134	0.035	15.209	0.000	0.256	0.056	21.359	0.000	0.142	0.032	20.707	0.000
	E	E	0.146	0.085	0.086	0.140	0.183	1512.750	0.000	0.148	0.192	2107.698	0.000	0.138	0.180	1884.967	0.000	0.143	0.187	1223.926	0.000	0.166	0.214	2536.516	0.000
	G + E	G	0.096	0.015	0.000	0.140	0.038			0.098	0.034			0.081	0.034			0.147	0.047			0.068	0.026		
		E	0.131	0.077	0.086	0.118	0.155	1501.103	0.000	0.148	0.192	2096.645	0.000	0.138	0.179	1876.219	0.000	0.121	0.158	1214.063	0.000	0.141	0.182	2523.859	0.000
	G x E	G	0.103	0.016	0.000	0.164	0.044			0.098	0.034			0.079	0.034			0.173	0.055			0.079	0.030		
		E	0.151	0.088	0.085	0.152	0.197			0.150	0.194			0.144	0.186			0.143	0.186			0.174	0.224		
		GxE	0.014	0.003	0.000	0.040	0.010	31.069	0.000	0.004	0.005	0.937	0.166	0.016	0.006	8.815	0.001	0.003	0.008	0.115	0.367	0.043	0.010	40.008	0.000
BROAD	G	G	0.135	0.014	0.000	0.154	0.036	19.061	0.000	0.160	0.029	32.247	0.000	0.110	0.030	14.595	0.000	0.159	0.048	11.076	0.000	0.118	0.026	21.365	0.000
	E	E	0.088	0.053	0.096	0.086	0.116	1112.191	0.000	0.087	0.116	1418.563	0.000	0.081	0.109	1286.448	0.000	0.092	0.123	898.291	0.000	0.097	0.129	1794.423	0.000
	G + E	G	0.089	0.012	0.000	0.096	0.028			0.130	0.029			0.081	0.028			0.105	0.037			0.066	0.020		
		F	0.077	0.046	0.096	0.070	0.093	1105.984	0.000	0.086	0.116	1410.377	0.000	0.081	0.109	1280.978	0.000	0.074	0.099	895.655	0.000	0.078	0.104	1785.255	0.000
	G x E	G	0.100	0.014	0.000	0.118	0.035			0.125	0.029			0.078	0.028			0.126	0.046			0.081	0.025		
		E	0.083	0.049	0.089	0.089	0.119			0.078	0.099			0.076	0.099			0.085	0.110			0.096	0.129		
		GxE	0.002	0.002	0.471	0.010	0.006	3.457	0.031	0.000	0.005	0.000	0.500	0.000	0.005	0.000	0.500	0.000	0.007	0.000	0.500	0.000	0.005	0.001	0.488
NEURO	G	G	0.150	0.009	0.000	0.102	0.024	19.868	0.000	0.174	0.020	81.415	0.000	0.147	0.021	54.107	0.000	0.148	0.031	24.168	0.000	0.160	0.017	94.024	0.000
	E	E	0.039	0.024	0.100	0.039	0.053	813.621	0.000	0.041	0.056	1002.493	0.000	0.040	0.055	969.691	0.000	0.038	0.051	575.482	0.000	0.040	0.054	1199.364	0.000
	G + E	G	0.136	0.010	0.000	0.096	0.023			0.156	0.021			0.142	0.021			0.131	0.030			0.142	0.018		
		E	0.039	0.024	0.100	0.039	0.053	812.995	0.000	0.041	0.055	992.386	0.000	0.040	0.055	971.070	0.000	0.037	0.051	572.233	0.000	0.039	0.054	1186.482	0.000
	G x E	G	0.134	0.010	0.000	0.092	0.023			0.155	0.022			0.142	0.021			0.132	0.030			0.139	0.018		
		E	0.043	0.026	0.099	0.044	0.060			0.046	0.062			0.043	0.059			0.040	0.055			0.044	0.059		
		GxE	0.025	0.002	0.000	0.027	0.005	51.053	0.000	0.027	0.005	60.822	0.000	0.020	0.004	31.918	0.000	0.026	0.006	26.177	0.000	0.026	0.004	58.352	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 17. Mixed Linear Model Results Including Es of Full Trauma Exposure Principal Component 2.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.167	0.017	0.000	0.229	0.044	28.956	0.000	0.158	0.035	20.793	0.000	0.134	0.035	15.209	0.000	0.256	0.056	21.359	0.000	0.142	0.032	20.707	0.000
	E	E	0.000	0.000	0.746	0.000	0.000	0.000	0.500	0.001	0.001	8.207	0.002	0.000	0.000	0.755	0.193	0.000	0.001	0.000	0.500	0.000	0.000	0.036	0.425
	G + E	G	0.168	0.017	0.000	0.229	0.044			0.158	0.035			0.135	0.035			0.258	0.056			0.142	0.032		
		E	0.000	0.000	0.775	0.000	0.000	0.000	0.500	0.001	0.001	8.329	0.002	0.000	0.000	0.844	0.179	0.000	0.001	0.000	0.500	0.000	0.000	0.018	0.446
	G x E	G	0.156	0.017	0.000	0.226	0.043			0.155	0.034			0.124	0.035			0.229	0.055			0.123	0.030		
		E	0.000	0.000	0.837	0.000	0.000			0.000	0.000			0.000	0.001			0.000	0.001			0.002	0.002		
		GxE	0.127	0.005	0.000	0.135	0.013	155.371	0.000	0.117	0.011	160.355	0.000	0.093	0.010	127.360	0.000	0.125	0.015	107.192	0.000	0.184	0.012	342.197	0.000
BROAD	G	G	0.135	0.014	0.000	0.154	0.036	19.061	0.000	0.160	0.029	32.747	0.000	DID NOT CONVERGE				0.159	0.048	11.076	0.000	0.118	0.026	21.365	0.000
	E	E	0.000	0.000	0.848	0.000	0.000	0.113	0.369	0.000	0.001	3.336	0.034					0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.500
	G + E	G	0.135	0.014	0.000	0.154	0.036			0.161	0.029							0.158	0.048			0.118	0.026		
		E	0.000	0.000	0.848	0.000	0.000	0.125	0.362	0.000	0.001	3.524	0.030					0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.500
	G x E	G	0.139	0.016	0.000	0.151	0.036			0.160	0.029							0.154	0.048			0.113	0.026		
		E	0.000	0.000	0.845	0.000	0.000			0.000	0.000							0.000	0.000			0.000	0.000		
		GxE	0.048	0.004	0.000	0.048	0.009	39.830	0.000	0.036	0.007	30.602	0.000					0.049	0.010	30.986	0.000	0.064	0.008	92.435	0.000
NEURO	G	G	0.150	0.009	0.000	0.102	0.024	19.868	0.000	0.174	0.020	81.415	0.000	0.147	0.021	54.107	0.000	0.148	0.031	24.168	0.000	0.160	0.017	94.024	0.000
	E	E	0.002	0.001	0.133	0.003	0.004	57.725	0.000	0.003	0.004	59.104	0.000	0.002	0.003	41.350	0.000	0.002	0.003	22.315	0.000	0.003	0.005	88.253	0.000
	G + E	G	0.149	0.009	0.000	0.102	0.024			0.172	0.020			0.146	0.021			0.146	0.031			0.161	0.017		
		E	0.002	0.001	0.134	0.003	0.004	57.814	0.000	0.003	0.004	57.380	0.000	0.002	0.003	40.391	0.000	0.002	0.002	21.571	0.000	0.003	0.005	90.553	0.000
	G x E	G	0.148	0.009	0.000	0.097	0.023			0.171	0.020			0.147	0.020			0.149	0.031			0.159	0.017		
		E	0.003	0.002	0.124	0.004	0.006			0.004	0.005			0.003	0.004			0.003	0.004			0.005	0.006		
		GxE	0.038	0.002	0.000	0.039	0.006	66.627	0.000	0.038	0.005	69.272	0.000	0.041	0.005	82.580	0.000	0.035	0.006	46.662	0.000	0.038	0.005	88.187	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 18. Mixed Linear Model Results Including Es of Full Trauma Exposure Principal Component 3.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.167	0.017	0.000	0.229	0.044	28.956	0.000	0.158	0.035	20.793	0.000	0.134	0.035	15.209	0.000	0.256	0.056	21.359	0.000	0.142	0.032	20.707	0.000
	E	E	0.001	0.001	0.209	0.001	0.001	3.964	0.023	0.002	0.003	22.508	0.000	0.001	0.002	16.506	0.000	0.003	0.004	19.385	0.000	0.002	0.002	21.260	0.000
	G + E	G	0.165	0.017	0.000	0.227	0.044			0.153	0.035			0.132	0.035			0.252	0.056			0.139	0.032		
		E	0.001	0.001	0.217	0.001	0.001	3.539	0.030	0.002	0.003	21.327	0.000	0.001	0.002	15.976	0.000	0.003	0.004	18.668	0.000	0.002	0.002	20.673	0.000
	G x E	G	0.161	0.017	0.000	0.226	0.044			0.149	0.035			0.133	0.035			0.246	0.056			0.134	0.032		
		E	0.001	0.001	0.209	0.000	0.001			0.002	0.002			0.001	0.002			0.002	0.003			0.001	0.002		
		GxE	0.034	0.004	0.000	0.033	0.010	11.534	0.000	0.035	0.009	19.884	0.000	0.021	0.008	7.319	0.003	0.042	0.012	16.178	0.000	0.048	0.009	32.184	0.000
BROAD	G	G	0.135	0.014	0.000	0.154	0.036	19.061	0.000	0.160	0.029	32.247	0.000	0.110	0.030	14.595	0.000	0.159	0.048	11.076	0.000	0.118	0.026	21.365	0.000
	E	E	0.000	0.000	0.700	0.000	0.000	1.184	0.138	0.000	0.000	0.000	0.500	0.000	0.000	0.064	0.400	0.000	0.000	0.000	0.500	0.000	0.000	0.142	0.353
	G + E	G	0.135	0.014	0.000	0.155	0.036			0.161	0.029			0.110	0.030			0.159	0.048			0.117	0.026		
		E	0.000	0.000	0.708	0.000	0.000	1.290	0.128	0.000	0.000	0.000	0.500	0.000	0.000	0.058	0.405	0.000	0.000	0.000	0.500	0.000	0.000	0.110	0.370
	G x E	G	0.135	0.014	0.000	0.155	0.036			0.161	0.029			0.110	0.030			0.157	0.048			0.117	0.026		
		E	0.000	0.000	0.728	0.000	0.000			0.000	0.000			0.000	0.000			0.000	0.000			0.000	0.000		
		GxE	0.007	0.003	0.017	0.002	0.007	0.103	0.374	0.011	0.007	2.646	0.052	0.002	0.006	0.059	0.404	0.014	0.009	2.735	0.049	0.011	0.007	2.911	0.044
NEURO	G	G	0.150	0.009	0.000	0.102	0.024	19.868	0.000	0.174	0.020	81.415	0.000	0.147	0.021	54.107	0.000	0.148	0.031	24.168	0.000	0.160	0.017	94.024	0.000
	E	E	0.001	0.001	0.161	0.003	0.004	49.305	0.000	0.002	0.002	34.623	0.000	0.002	0.004	54.317	0.000	0.001	0.001	9.983	0.001	0.001	0.002	32.638	0.000
	G + E	G	0.149	0.009	0.000	0.100	0.024			0.173	0.020			0.146	0.021			0.148	0.031			0.160	0.017		
		E	0.001	0.001	0.160	0.003	0.004	48.427	0.000	0.002	0.002	33.659	0.000	0.002	0.003	53.258	0.000	0.001	0.001	9.890	0.001	0.001	0.002	32.831	0.000
	G x E	G	0.149	0.009	0.000	0.100	0.024			0.173	0.020			0.147	0.021			0.149	0.031			0.160	0.017		
		E	0.001	0.001	0.154	0.003	0.004			0.002	0.002			0.003	0.004			0.001	0.001			0.001	0.002		
		GxE	0.010	0.002	0.000	0.001	0.005	0.020	0.443	0.010	0.004	5.127	0.012	0.011	0.004	7.660	0.003	0.016	0.006	7.299	0.003	0.013	0.004	9.792	0.001

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 19. Mixed Linear Model Results Including Es of Female Only Full Trauma Exposure Principal Components.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.198	0.029	0.000	0.322	0.074	20.003	0.000	0.231	0.059	15.459	0.000	0.182	0.062	8.956	0.001	0.323	0.094	12.044	0.000	0.088	0.051	2.962	0.043
	E	E	0.150	0.022	0.000	0.153	0.050	1164.678	0.000	0.155	0.050	1526.949	0.000	0.146	0.048	1371.836	0.000	0.140	0.046	840.290	0.000	0.156	0.051	1820.246	0.000
	G + E	G	0.125	0.026	0.000	0.265	0.068			0.161	0.054			0.114	0.056			0.223	0.086			0.015	0.046		
		E	0.149	0.022	0.000	0.153	0.050	1160.699	0.000	0.154	0.050	1520.678	0.000	0.146	0.047	1367.036	0.000	0.139	0.046	834.981	0.000	0.156	0.051	1817.387	0.000
	G x E	G	0.123	0.026	0.000	0.270	0.068			0.164	0.054			0.112	0.056			0.220	0.086			0.006	0.046		
		E	0.152	0.022	0.000	0.158	0.052			0.156	0.050			0.148	0.048			0.141	0.047			0.158	0.051		
		GxE	0.086	0.009	0.000	0.142	0.023	40.514	0.000	0.044	0.017	5.911	0.008	0.043	0.017	6.259	0.006	0.098	0.025	14.488	0.000	0.154	0.020	62.105	0.000
BROAD	G	G	0.162	0.024	0.000	0.127	0.062	4.326	0.019	0.249	0.051	25.049	0.000	0.068	0.052	1.733	0.094	0.201	0.081	6.085	0.007	0.169	0.043	16.118	0.000
	E	E	0.098	0.015	0.000	0.100	0.034	893.573	0.000	0.098	0.033	1079.739	0.000	0.094	0.032	991.726	0.000	0.099	0.034	673.226	0.000	0.099	0.033	1370.310	0.000
	G + E	G	0.119	0.023	0.000	0.082	0.058			0.180	0.048			0.041	0.049			0.143	0.076			0.140	0.040		
		E	0.098	0.015	0.000	0.100	0.034	891.270	0.000	0.097	0.033	1069.543	0.000	0.094	0.032	990.727	0.000	0.099	0.034	670.672	0.000	0.099	0.033	1367.076	0.000
	G x E	G	0.113	0.023	0.000	0.081	0.058			0.167	0.048			0.029	0.049			0.143	0.076			0.139	0.040		
		E	0.090	0.013	0.000	0.100	0.034			0.083	0.026			0.079	0.025			0.099	0.034			0.099	0.033		
		GxE	0.007	0.007	0.347	0.008	0.017	0.221	0.319	0.000	0.015	0.000	0.500	0.000	0.016	0.000	0.500	0.005	0.020	0.056	0.407	0.020	0.015	1.549	0.107
NEURO	G	G	0.154	0.017	0.000	0.164	0.043	15.340	0.000	0.147	0.036	17.099	0.000	0.161	0.037	20.109	0.000	0.159	0.054	8.932	0.001	0.150	0.030	26.136	0.000
	E	E	0.051	0.008	0.000	0.057	0.019	681.303	0.000	0.055	0.019	792.602	0.000	0.049	0.017	689.270	0.000	0.046	0.016	422.747	0.000	0.051	0.017	1003.343	0.000
	G + E	G	0.142	0.016	0.000	0.143	0.040			0.115	0.034			0.161	0.035			0.142	0.052			0.150	0.028		
		E	0.051	0.008	0.000	0.056	0.019	678.919	0.000	0.055	0.019	787.164	0.000	0.049	0.017	691.742	0.000	0.046	0.016	421.722	0.000	0.051	0.018	1006.301	0.000
	G x E	G	0.144	0.016	0.000	0.140	0.040			0.118	0.034			0.164	0.035			0.152	0.051			0.148	0.028		
		E	0.052	0.008	0.000	0.057	0.020			0.056	0.019			0.050	0.017			0.047	0.016			0.053	0.018		
		GxE	0.061	0.005	0.000	0.053	0.012	24.488	0.000	0.063	0.011	37.812	0.000	0.060	0.011	32.006	0.000	0.066	0.014	23.697	0.000	0.065	0.010	45.881	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 20. Mixed Linear Model Results Including Es of Male Only Full Trauma Exposure Principal Components.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.184	0.043	0.000	0.251	0.107	5.616	0.009	0.297	0.088	11.781	0.000	0.107	0.086	1.587	0.104	0.223	0.140	2.617	0.053	0.099	0.083	1.382	0.120
	E	E	0.247	0.036	0.000	0.205	0.068	502.832	0.000	0.258	0.081	872.405	0.000	0.248	0.078	829.870	0.000	0.252	0.082	462.914	0.000	0.301	0.094	1021.302	0.000
	G + E	G	0.143	0.038	0.000	0.199	0.097			0.240	0.077			0.079	0.075			0.118	0.123			0.094	0.072		
		E	0.246	0.035	0.000	0.205	0.068	501.577	0.000	0.258	0.081	870.559	0.000	0.248	0.078	829.409	0.000	0.251	0.082	461.228	0.000	0.302	0.094	1021.584	0.000
	G x E	G	0.116	0.035	0.001	0.187	0.092			0.200	0.073			0.058	0.072			0.121	0.118			0.056	0.066		
		E	0.235	0.034	0.000	0.195	0.066			0.255	0.081			0.238	0.076			0.237	0.079			0.281	0.089		
		GxE	0.441	0.018	0.000	0.462	0.040	163.444	0.000	0.436	0.038	189.634	0.000	0.371	0.035	147.682	0.000	0.389	0.045	95.051	0.000	0.575	0.043	296.618	0.000
BROAD	G	G	0.086	0.034	0.012	0.000	0.086	0.000	0.500	0.134	0.068	4.086	0.022	0.096	0.070	1.952	0.081	0.134	0.117	1.318	0.126	0.067	0.066	1.012	0.157
	E	E	0.132	0.020	0.000	0.128	0.045	331.562	0.000	0.130	0.044	491.207	0.000	0.126	0.043	481.217	0.000	0.132	0.046	292.452	0.000	0.149	0.050	622.781	0.000
	G + E	G	0.056	0.031	0.074	0.000	0.080			0.098	0.062			0.081	0.065			0.075	0.108			0.022	0.060		
		E	0.131	0.020	0.000	0.124	0.043	331.531	0.000	0.130	0.044	489.710	0.000	0.126	0.043	480.871	0.000	0.132	0.046	291.614	0.000	0.149	0.050	621.903	0.000
	G x E	G	0.057	0.031	0.067	0.000	0.079			0.097	0.061			0.078	0.064			0.057	0.106			0.033	0.059		
		E	0.130	0.020	0.000	0.120	0.042			0.132	0.045			0.124	0.043			0.132	0.046			0.145	0.049		
		GxE	0.165	0.013	0.000	0.132	0.029	24.632	0.000	0.162	0.026	43.114	0.000	0.145	0.026	32.739	0.000	0.189	0.035	32.356	0.000	0.207	0.027	64.694	0.000
NEURO	G	G	0.160	0.020	0.000	0.000	0.051	0.000	0.500	0.188	0.043	20.845	0.000	0.196	0.044	20.402	0.000	0.232	0.068	12.041	0.000	0.179	0.037	23.435	0.000
	E	E	0.092	0.014	0.000	0.082	0.028	514.331	0.000	0.105	0.034	749.093	0.000	0.102	0.033	752.985	0.000	0.087	0.029	364.535	0.000	0.093	0.030	915.529	0.000
	G + E	G	0.135	0.019	0.000	0.000	0.047			0.166	0.039			0.168	0.040			0.196	0.063			0.140	0.034		
		E	0.087	0.013	0.000	0.070	0.022	513.589	0.000	0.105	0.034	748.589	0.000	0.102	0.033	751.243	0.000	0.086	0.029	362.562	0.000	0.092	0.030	909.666	0.000
	G x E	G	0.135	0.018	0.000	0.000	0.046			0.169	0.038			0.165	0.039			0.193	0.062			0.142	0.034		
		E	0.090	0.013	0.000	0.076	0.025			0.107	0.035			0.102	0.033			0.084	0.029			0.093	0.031		
		GxE	0.080	0.007	0.000	0.068	0.016	25.530	0.000	0.082	0.015	34.015	0.000	0.084	0.015	38.918	0.000	0.093	0.019	28.750	0.000	0.077	0.014	33.111	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 21. Mixed Linear Model Results Including Es of Female Only Full Trauma Exposure Principal Components using Unrelated Individuals.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.175	0.033	0.000	0.264	0.088	9.476	0.001	0.220	0.068	10.649	0.001	0.223	0.071	10.338	0.001	0.181	0.107	2.920	0.044	0.076	0.057	1.831	0.088
	E	E	0.151	0.022	0.000	0.154	0.050	1013.563	0.000	0.159	0.051	1393.347	0.000	0.150	0.049	1240.789	0.000	0.141	0.047	757.497	0.000	0.156	0.051	1685.144	0.000
	G + E	G	0.103	0.029	0.000	0.211	0.081			0.155	0.061			0.134	0.064			0.102	0.098			0.005	0.051		
		E	0.151	0.022	0.000	0.154	0.050	1011.203	0.000	0.158	0.051	1389.332	0.000	0.149	0.048	1234.951	0.000	0.140	0.046	755.646	0.000	0.156	0.051	1683.321	0.000
	G x E	G	0.104	0.029	0.000	0.230	0.080			0.159	0.061			0.132	0.064			0.105	0.098			0.000	0.050		
		E	0.154	0.022	0.000	0.159	0.052			0.160	0.052			0.151	0.049			0.142	0.047			0.159	0.052		
		GxE	0.084	0.009	0.000	0.138	0.024	33.391	0.000	0.040	0.018	4.614	0.016	0.042	0.018	5.540	0.009	0.093	0.026	12.002	0.000	0.155	0.021	59.515	0.000
BROAD	G	G	0.165	0.028	0.000	0.138	0.074	3.519	0.030	0.209	0.059	13.106	0.000	0.092	0.061	2.354	0.062	0.195	0.093	4.321	0.019	0.184	0.048	15.538	0.000
	E	E	0.099	0.015	0.000	0.100	0.034	767.453	0.000	0.100	0.034	971.030	0.000	0.095	0.032	885.712	0.000	0.102	0.035	621.371	0.000	0.100	0.034	1271.425	0.000
	G + E	G	0.123	0.026	0.000	0.105	0.070			0.135	0.055			0.059	0.057			0.165	0.088			0.150	0.045		
		E	0.099	0.015	0.000	0.100	0.034	766.253	0.000	0.099	0.033	964.132	0.000	0.095	0.032	884.473	0.000	0.101	0.035	620.524	0.000	0.100	0.034	1268.129	0.000
	G x E	G	0.117	0.026	0.000	0.105	0.070			0.120	0.055			0.047	0.057			0.165	0.088			0.150	0.045		
		E	0.091	0.013	0.000	0.100	0.034			0.084	0.026			0.082	0.026			0.102	0.035			0.100	0.034		
		GxE	0.007	0.008	0.335	0.009	0.018	0.220	0.320	0.000	0.016	0.000	0.500	0.000	0.016	0.000	0.500	0.006	0.021	0.072	0.394	0.021	0.016	1.580	0.104
NEURO	G	G	0.142	0.019	0.000	0.152	0.051	9.364	0.001	0.119	0.041	8.698	0.002	0.164	0.043	15.792	0.000	0.137	0.063	4.891	0.013	0.140	0.033	18.760	0.000
	E	E	0.051	0.008	0.000	0.055	0.019	569.485	0.000	0.056	0.019	707.839	0.000	0.050	0.017	628.199	0.000	0.044	0.016	361.696	0.000	0.053	0.018	952.272	0.000
	G + E	G	0.130	0.018	0.000	0.124	0.048			0.090	0.039			0.162	0.040			0.129	0.060			0.141	0.031		
		E	0.051	0.008	0.000	0.055	0.019	566.936	0.000	0.056	0.019	704.666	0.000	0.050	0.017	629.758	0.000	0.044	0.016	361.632	0.000	0.053	0.018	954.773	0.000
	G x E	G	0.132	0.018	0.000	0.125	0.048			0.093	0.038			0.164	0.040			0.145	0.060			0.138	0.031		
		E	0.052	0.008	0.000	0.056	0.019			0.057	0.019			0.051	0.018			0.045	0.016			0.055	0.019		
		GxE	0.063	0.006	0.000	0.057	0.013	24.349	0.000	0.066	0.012	35.395	0.000	0.056	0.012	26.064	0.000	0.060	0.015	17.931	0.000	0.073	0.011	52.091	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Table 22. Mixed Linear Model Results Including Es of Male Only Trauma Exposure Principal Components using Unrelated Individuals.

Trait	Model	Source	META ANALYSIS			NORTH				MIDNORTH				MIDSOUTH				SOUTHWEST				SOUTHEAST			
			Variance	SE	P-Value	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P	Variance	SE	LRT	LRT-P
CIDI	G	G	0.197	0.048	0.000	0.224	0.124	3.377	0.033	0.369	0.100	13.950	0.000	0.129	0.097	1.781	0.091	0.230	0.159	2.115	0.073	0.089	0.092	0.921	0.169
	E	E	0.238	0.035	0.000	0.186	0.063	406.679	0.000	0.261	0.082	762.730	0.000	0.241	0.076	738.421	0.000	0.247	0.081	381.466	0.000	0.306	0.095	940.638	0.000
	G + E	G	0.157	0.043	0.000	0.193	0.113			0.298	0.089			0.103	0.086			0.127	0.140			0.081	0.080		
		E	0.237	0.035	0.000	0.186	0.063	406.324	0.000	0.260	0.082	760.648	0.000	0.241	0.076	738.121	0.000	0.247	0.081	380.185	0.000	0.306	0.095	940.730	0.000
	G x E	G	0.129	0.040	0.001	0.162	0.107			0.263	0.084			0.096	0.082			0.147	0.135			0.034	0.073		
		E	0.227	0.034	0.000	0.180	0.062			0.255	0.081			0.228	0.074			0.231	0.078			0.286	0.091		
		GxE	0.436	0.018	0.000	0.481	0.043	148.300	0.000	0.423	0.039	162.732	0.000	0.365	0.036	131.491	0.000	0.353	0.046	71.293	0.000	0.587	0.044	286.036	0.000
BROAD	G	G	0.100	0.039	0.010	0.005	0.102	0.002	0.481	0.146	0.078	3.661	0.028	0.125	0.081	2.503	0.057	0.219	0.133	2.763	0.048	0.054	0.072	0.553	0.229
	E	E	0.132	0.020	0.000	0.122	0.043	273.926	0.000	0.131	0.045	428.852	0.000	0.131	0.045	452.356	0.000	0.125	0.044	246.017	0.000	0.157	0.053	596.314	0.000
	G + E	G	0.077	0.036	0.030	0.019	0.095			0.119	0.072			0.118	0.074			0.177	0.124			0.010	0.066		
		E	0.132	0.020	0.000	0.122	0.043	273.966	0.000	0.131	0.044	428.086	0.000	0.131	0.045	452.484	0.000	0.125	0.044	245.337	0.000	0.157	0.053	595.783	0.000
	G x E	G	0.075	0.035	0.034	0.008	0.094			0.118	0.071			0.117	0.074			0.158	0.122			0.015	0.065		
		E	0.131	0.020	0.000	0.120	0.043			0.132	0.045			0.130	0.044			0.125	0.044			0.153	0.052		
		GxE	0.167	0.013	0.000	0.141	0.031	22.859	0.000	0.154	0.027	34.660	0.000	0.148	0.028	30.851	0.000	0.183	0.037	26.761	0.000	0.216	0.029	64.430	0.000
NEURO	G	G	0.149	0.023	0.000	DID NOT CONVERGE				0.192	0.049	16.366	0.000	0.180	0.050	13.491	0.000	0.169	0.077	5.099	0.012	0.162	0.041	16.004	0.000
	E	E	0.091	0.014	0.000					0.103	0.034	647.931	0.000	0.104	0.034	678.551	0.000	0.085	0.029	296.973	0.000	0.093	0.031	825.527	0.000
	G + E	G	0.150	0.023	0.000					0.174	0.044			0.162	0.045			0.156	0.071			0.123	0.037		
		E	0.095	0.016	0.000					0.103	0.034	648.313	0.000	0.103	0.034	678.791	0.000	0.085	0.029	297.011	0.000	0.093	0.030	820.723	0.000
	G x E	G	0.151	0.023	0.000					0.175	0.044			0.161	0.044			0.156	0.070			0.126	0.037		
		E	0.095	0.016	0.000					0.105	0.034			0.104	0.034			0.083	0.029			0.094	0.031		
		GxE	0.084	0.008	0.000					0.081	0.016	29.516	0.000	0.090	0.016	39.224	0.000	0.091	0.020	25.450	0.000	0.078	0.014	31.479	0.000

Abbreviations. E, environmental relationship matrix using full trauma principal components; SE, standard error; LRT, log-ratio test value; LRT-P, log-ratio test P-Value; G, genetic (Genomic Relationship Matrix); E, environmental (full trauma exposure ERM); GxE, genome-by-trauma exposure interaction.

Section B

A consistent pattern observed in results were large LRT values when including environmental relationship matrices (Es) into the mixed linear models (see **Table SB1**). This suggests a substantial improvement in model fit when including the ERMs. Whilst the estimates and standard errors observed for the proportion of variance explained by the Es suggest statistical significance, they do not suggest very *strong* significance, which is discrepant with what the LRT values suggest.

Table SB1. North Cluster Mixed Linear Model Results Including E of Full Trauma Exposure Principal Components.

			NORTH CLUSTER			
Trait	**Model**	**Source**	**Variance**	**SE**	**LRT**	**LRT-P**
CIDI	**G**	**G**	0.229	0.044	28.956	0.000
	E	**E**	0.170	0.055	1801.636	0.000
	G + E	**G**	0.171	0.040		
		E	0.169	0.055	1792.306	0.000
	G x E	**G**	0.174	0.039		
		E	0.176	0.057		
		GxE	0.241	0.021	159.722	0.000

Abbreviations. CIDI, Composite International Diagnostic Interview Depression; G, Genomic Relationship Matrix; E, Environmental Relationship Matrix; GxE, Genome-by-Trauma Exposure Relationship Matrix; LRT, Log-Likelihood Ratio Test; P, P-value.

To explore this discrepancy the log-likelihood distribution was examined. A log-likelihood function aims to identify and fit the most appropriate distribution to the data available (e.g. a normal distribution). A value from the observed data will be tested as the mean of the explored distribution. A log-likelihood value, signifying the likelihood of observing the data is obtained. A range of values are tested, and the resulting log-likelihood values are plotted to form the log-likelihood distribution. The highest value i.e. the maximum likelihood estimate, signifies the optimal position of the distribution explored to the data at hand.

Here, the discrepancy suggests a potentially unusual log-likelihood distribution which may be due to the nature of the Es used. Large LRT values would suggest the distribution to have a steep increase, whereas, the smaller standard errors would suggest a plateau surrounding the maximum estimate of the log-likelihood. To plot

the log-likelihood distribution, a range of E values and their corresponding log-likelihood value was plotted (**Figure 3**).

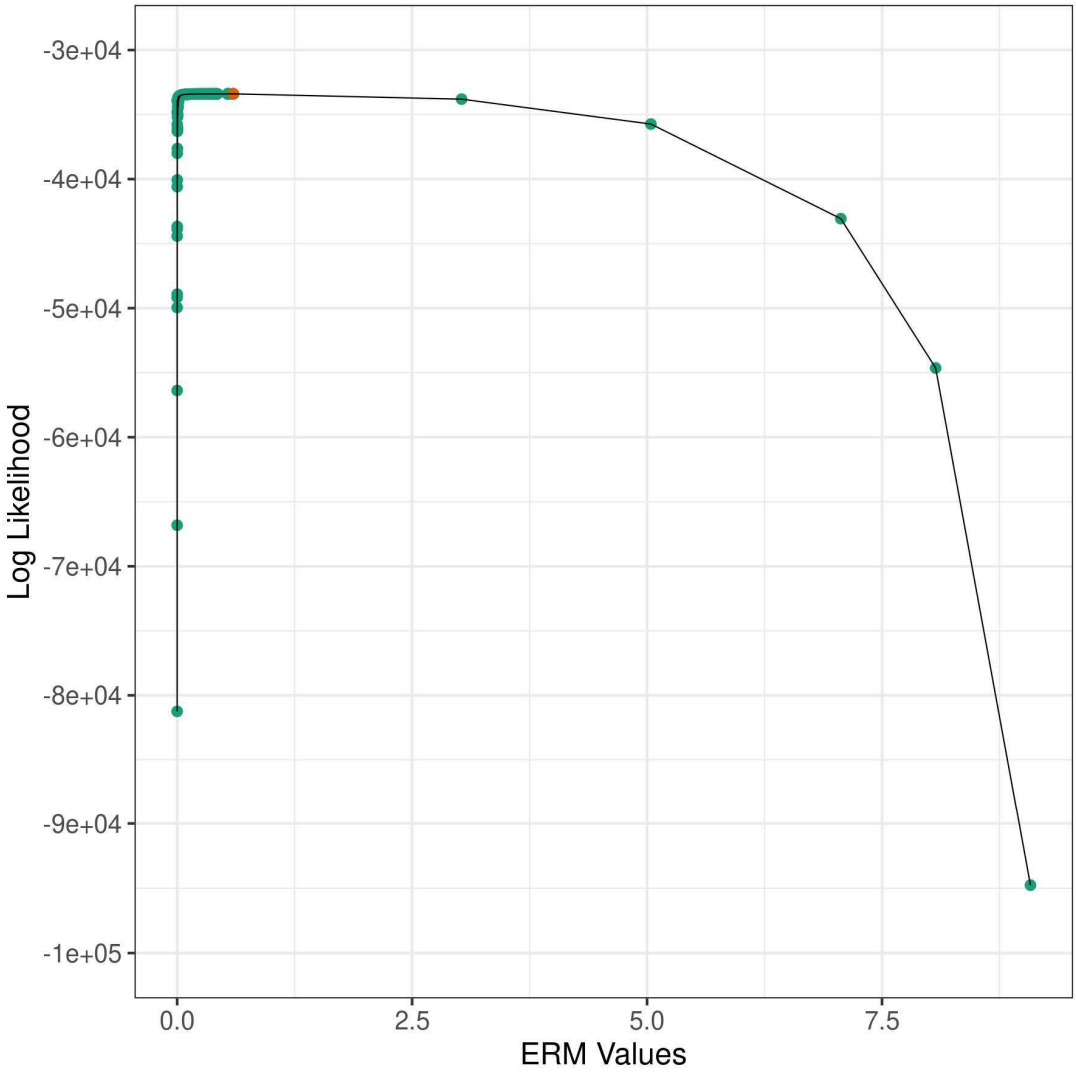


Figure SB1. Log-Likelihood Distribution of Mixed Linear Model Including the Trauma Exposure Environmental Relationship Matrix (ERM) for the UK Biobank North Geographical Cluster and Neuroticism Phenotype. The x-axis represents values of the ERM. The y-axis represents log-likelihood values. The orange point is the maximum likelihood estimate.

The distribution observed explains the discrepancy observed between the LRT values and the estimates/standard errors of the variance attributable to the E.

References:

Acciai, F. & Hardy, M. (2017) Depression in later life: A closer look at the gender gap. *Social Science Research*, 68, 163-175.

Adams, M. J., Hill, W. D., Howard, D. M., Dashti, H. S., Davis, K. A. S., Campbell, A., Clarke, T. K., Deary, I. J., Hayward, C., Porteous, D., Hotopf, M. & McIntosh, A. M. (2020) Factors associated with sharing e-mail information and mental health survey participation in large population cohorts. *Int J Epidemiol*, 49(2), 410-421.

Afzali, M. H., Sunderland, M., Teesson, M., Carragher, N., Mills, K. & Slade, T. (2017) A network approach to the comorbidity between posttraumatic stress disorder and major depressive disorder: The role of overlapping symptoms. *Journal of Affective Disorders*, 208, 490-496.

Akimova, E. T., Breen, R., Brazel, D. M. & Mills, M. C. (2020) Heritable environments: Bias due to conditioning on a collider in models with polygenic scores.

Allen, N. E., Sudlow, C., Peakman, T. & Collins, R. (2014) UK biobank data: Come and get it.

Alloy, L. B., Abramson, L. Y., Smith, J. M., Gibb, B. E. & Neeren, A. M. (2006) Role of parenting and maltreatment histories in unipolar and bipolar mood disorders: Mediation by cognitive vulnerability to depression. *Clinical Child and Family Psychology Review*, 9(1), 23-64.

Amador, C., Zeng, Y., Barber, M., Walker, R. M., Campbell, A., McIntosh, A. M., Evans, K. L., Porteous, D. J., Hayward, C., Wilson, J. F., Navarro, P. & Haley, C. S. (2021) Genome-wide methylation data improves dissection of the effect of smoking on body mass index. *PLoS Genet*, 17(9), e1009750.

American Psychiatric Association, D. S. a. A. P. A. (2013) Diagnostic and statistical manual of mental disorders: DSM-5.

Anda, R. F., Felitti, V. J., Bremner, J. D., Walker, J. D., Whitfield, C., Perry, B. D., Dube, S. R. & Giles, W. H. (2006) The enduring effects of abuse and related adverse experiences in childhood. A convergence of evidence from neurobiology and epidemiology. *Eur Arch Psychiatry Clin Neurosci*, 256(3), 174-86.

Andersen, A. M., Philibert, R. A., Gibbons, F. X., Simons, R. L. & Long, J. (2017) Accuracy and utility of an epigenetic biomarker for smoking in populations with varying rates of false self-report. *Am J Med Genet B Neuropsychiatr Genet*, 174(6), 641-650.

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D. M., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., Berr, C., Letenneur, L., Hannequin, D., Amouyel, P., Boland, A., Deleuze, J. F., Duron, E., Vardarajan, B. N., Reitz, C., Goate, A. M., Huentelman, M. J., Kamboh, M. I., Larson, E. B., Rogaeva, E., St George-Hyslop, P., Hakonarson, H., Kukull, W. A., Farrer, L. A., Barnes, L. L., Beach, T. G., Demirci, F. Y., Head, E., Hulette, C. M., Jicha, G. A., Kauwe, J. S. K., Kaye, J. A., Leverenz, J. B., Levey, A. I., Lieberman, A. P., Pankratz, V. S., Poon, W. W., Quinn, J. F., Saykin, A. J., Schneider, L. S.,

Smith, A. G., Sonnen, J. A., Stern, R. A., Van Deerlin, V. M., Van Eldik, L. J., Harold, D., Russo, G., Rubinsztein, D. C., Bayer, A., Tsolaki, M., Proitsi, P., Fox, N. C., Hampel, H., Owen, M. J., Mead, S., Passmore, P., Morgan, K., Nothen, M. M., Rossor, M., Lupton, M. K., Hoffmann, P., Kornhuber, J., Lawlor, B., McQuillin, A., Al-Chalabi, A., Bis, J. C., Ruiz, A., Boada, M., Seshadri, S., Beiser, A., Rice, K., van der Lee, S. J., De Jager, P. L., Geschwind, D. H., Riemenschneider, M., Riedel-Heller, S., Rotter, J. I., Ransmayr, G., Hyman, B. T., Cruchaga, C., Alegret, M., Winsvold, B., Palta, P., Farh, K. H., Cuenca-Leon, E., Furlotte, N., Kurth, T., et al (2018) Analysis of shared heritability in common disorders of the brain. *Science*, 360(6395), 1313-+.

Arnau-Soler, A., Adams, M. J., Clarke, T. K., MacIntyre, D. J., Milburn, K., Navrady, L., Hayward, C., McIntosh, A. & Thomson, P. A. (2019a) A validation of the diathesis-stress model for depression in Generation Scotland. *Translational Psychiatry*.

Arnau-Soler, A., Adams, M. J., Hayward, C., Thomson, P. A., Porteous, D., Campbell, A., Smith, B. H., Black, C., Padmanabhan, S., McIntosh, A., Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Agerbo, E., Air, T. M., Andlauer, T. F. M., Bacanu, S. A., Bækvad-Hansen, M., Beekman, A. T. F., Bigdeli, T. B., Binder, E. B., Blackwood, D. H. R., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J. H., Clarke, T. K., Coleman, J. R. I., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G. E., Davies, G., Deary, I. J., Degenhardt, F., Derks, E. M., Direk, N., Dolan, C. V., Dunn, E. C., Eley, T. C., Escott-Price, V., Kiadeh, F. F. H., Finucane, H. K., Forstner, A. J., Frank, J., Gaspar, H. A., Gill, M., Goes, F. S., Gordon, S. D., Grove, J., Hall, L. S., Hansen, C. S., Hansen, T. F., Herms, S., Hickie, I. B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J. J., Hougaard, D. M., Ising, M., Jansen, R., Jorgenson, E., Knowles, J. A., Kohane, I. S., Kraft, J., Kretzschmar, W. W., Krogh, J., Kutalik, Z., Li, Y., Lind, P. A., MacIntyre, D. J., MacKinnon, D. F., Maier, R. M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S. E., Mehta, D., Middeldorp, C. M., Mihailov, E., Milaneschi, Y., Milani, L., Mondimore, F. M., Montgomery, G. W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M. G., Nyholt, D. R., O'Reilly, P. F., Oskarsson, H., et al (2018) Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. *PLoS ONE*.

Arnau-Soler, A., Macdonald-Dunlop, E., Adams, M. J., Clarke, T. K., MacIntyre, D. J., Milburn, K., Navrady, L., Generation, S., Major Depressive Disorder Working Group of the Psychiatric Genomics, C., Hayward, C., McIntosh, A. M. & Thomson, P. A. (2019b) Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Transl Psychiatry*, 9(1), 14.

Assary, E., Vincent, J. P., Keers, R. & Pluess, M. (2018) Gene-environment interaction and psychiatric disorders: Review and future directions. *Semin Cell Dev Biol*, 77, 133-143.

Avinun, R. (2019) Educational Attainment Polygenic Score is Associated with Depressive Symptoms via Socioeconomic Status: A Gene-Environment-Trait Correlation.

Avinun, R. & Hariri, A. R. (2019) A polygenic score for body mass index is associated with depressive symptoms via early life stress: Evidence for gene-environment correlation. *Journal of Psychiatric Research*.

Bates, T. C., Maher, B. S., Medland, S. E., McAloney, K., Wright, M. J., Hansell, N. K., Kendler, K. S., Martin, N. G. & Gillespie, N. A. (2018) The Nature of Nurture: Using a Virtual-Parent Design to Test Parenting Effects on Children's Educational Attainment in Genotyped Families. *Twin Res Hum Genet*, 21(2), 73-83.

Baxter, A. J., Scott, K. M., Ferrari, A. J., Norman, R. E., Vos, T. & Whiteford, H. A. (2014) Challenging the myth of an "epidemic" of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depress Anxiety*, 31(6), 506-16.

Beam, C. R. & Turkheimer, E. (2013) Phenotype-environment correlations in longitudinal twin models. *Development and Psychopathology*, 25(1), 7-16.

Bijma, P. (2010) Estimating indirect genetic effects: precision of estimates and optimum designs. *Genetics*, 186(3), 1013-28.

Bijma, P. (2014) The quantitative genetics of indirect genetic effects: a selective review of modelling issues. *Heredity (Edinb)*, 112(1), 61-9.

Bockting, C. L., Hollon, S. D., Jarrett, R. B., Kuyken, W. & Dobson, K. (2015) A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clinical Psychology Review*, 41, 16-26.

Bolinskey, P. K., Neale, M. C., Jacobson, K. C., Prescott, C. A. & Kendler, K. S. (2004) Sources of individual differences in stressful life event exposure in male and female twins. *Twin Research*, 7(1), 33-38.

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F. & Keller, M. C. (2019) No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *American Journal of Psychiatry*, 176(5), 376-387.

Bosker, F. J., Hartman, C. A., Nolte, I. M., Prins, B. P., Terpstra, P., Posthuma, D., van Veen, T., Willemsen, G., DeRijk, R. H., de Geus, E. J., Hoogendijk, W. J., Sullivan, P. F., Penninx, B. W., Boomsma, D. I., Snieder, H. & Nolen, W. A. (2011) Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Molecular Psychiatry*, 16(5), 516-532.

Bretschneider, J., Janitza, S., Jacobi, F., Thom, J., Hapke, U., Kurth, T. & Maske, U. E. (2018) Time trends in depression prevalence and health-related correlates: results from population-based surveys in Germany 1997-1999 vs. 2009-2012. *BMC Psychiatry*, 18(1), 394.

Brody, G. H., Yu, T. Y., Barton, A. W., Miller, G. E. & Chen, E. (2017) Youth temperament, harsh parenting, and variation in the oxytocin receptor gene forecast allostatic load during emerging adulthood. *Development and Psychopathology*, 29(3), 791-803.

Buckman, J. E. J., Underwood, A., Clarke, K., Saunders, R., Hollon, S. D., Fearon, P. & Pilling, S. (2018) Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clinical Psychology Review*, 64, 13-38.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P. & Marchini, J. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*.

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y. H., Liang, J. Q., Song, L., Hu, J. C., Li, Q. B., Jin, W., Hu, Z. F., Wang, G. B., Wang, L. M., Qian, P. Y., Liu, Y., Jiang, T., Lu, Y., Zhang, X. Q., Yin, Y., Li, Y. R., Xu, X., Gao, J. F., Reimers, M., Webb, T., Riley, B., Bacanu, S., Peterson, R. E., Chen, Y. P., Zhong, H., Liu, Z. R., Wang, G., Sun, J., Sang, H., Jiang, G. Q., Zhou, X. Y., Li, Y., Li, Y., Zhang, W., Wang, X. Y., Fang, X., Pan, R. D., Miao, G. D., Zhang, Q. W., Hu, J., Yu, F. Y., Du, B., Sang, W. H., Li, K. Q., Chen, G. B., Cai, M., Yang, L. J., Yang, D. L., Ha, B. W., Hong, X. H., Deng, H., Li, G. Y., Li, K., Song, Y., Gao, S. G., Zhang, J. B., Gan, Z. Y., Meng, H. Q., Pan, J. Y., Gao, C. G., Zhang, K. R., Sun, N., Li, Y. H., Niu, Q. H., Zhang, Y. T., Liu, T. Q., Hu, C. M., Zhang, Z., Lv, L. X., Dong, J. C., Wang, X. P., Tao, M., Wang, X. M., Xia, J., Rong, H., He, Q., Liu, T. B., Huang, G. P., Mei, Q. Y., Shen, Z. M., Liu, Y., Shen, J. H., Tian, T., Liu, X. J., Wu, W. Y., Gu, D. H., Fu, G. Y., Shi, J. G., Chen, Y. C., Gan, X. C., Liu, L. F., Wang, L. N., Yang, F. Z., Cong, E. Z., Marchini, J., Yang, H. M., Wang, J., et al (2015) Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, 523(7562), 588-+.

Cai, N., Choi, K. W. & Fried, E. I. (2020a) Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. *Human Molecular Genetics*, 29(R1), R10-R18.

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., Clarke, T. K., Forstner, A. J., Grabe, H. J., Hamilton, S. P., Levinson, D. F., Lewis, C. M., Lewis, G., Martin, N. G., Milaneschi, Y., Mors, O., Müller-Myhsok, B., Penninx, B. W. J. H., Perlis, R. H., Pistis, G., Potash, J. B., Preisig, M., Shi, J., Smoller, J. W., Streit, F., Tiemeier, H., Uher, R., Van der Auwera, S., Viktorin, A., Weissman, M. M., Kendler, K. S. & Flint, J. (2020b) Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H. L., McClay, J., Mill, J., Martin, J., Braithwaite, A. & Poulton, R. (2003) Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*.

Cesarini, D. & Visscher, P. M. (2017) Genetics and educational attainment. *npj Science of Learning*.

Chatterjee, N., Shi, J. & Garcia-Closas, M. (2016) Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*, 17(7), 392-406.

Cheesman, R., Eilertsen, E. M., Ahmadzadeh, Y. I., Gjerde, L. C., Hannigan, L. J., Havdahl, A., Young, A. I., Eley, T. C., Njølstad, P. R. & Magnus, P. (2020a) How important are parents in the development of child anxiety and depression? A genomic analysis of parent-offspring trios in the Norwegian Mother Father and Child Cohort Study (MoBa). *BMC medicine*, 18(1), 1-11.

Cheesman, R., Hunjan, A., Coleman, J. R. I., Ahmadzadeh, Y., Plomin, R., McAdams, T. A., Eley, T. C. & Breen, G. (2020b) Comparison of Adopted and Nonadopted Individuals Reveals Gene–Environment Interplay for Education in the UK Biobank. *Psychological Science*.

Choi, S. W., Mak, T. S. H. & O'Reilly, P. F. (2020) Tutorial: a guide to performing polygenic risk score analyses. Nature Research.

Christ, C., de Waal, M. M., Dekker, J. J. M., van Kuijk, I., van Schaik, D. J. F., Kikkert, M. J., Goudriaan, A. E., Beekman, A. T. F. & Messman-Moore, T. L. (2019) Linking childhood emotional abuse and depressive symptoms: The role of emotion dysregulation and interpersonal problems. *Plos One*, 14(2).

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, H. G., Turner, E. H., Higgins, J. P. T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J. P. A. & Geddes, J. R. (2018) Comparative Efficacy and Acceptability of 21 Antidepressant Drugs for the Acute Treatment of Adults With Major Depressive Disorder: A Systematic Review and Network Meta-Analysis. *Focus (Am Psychiatr Publ)*, 16(4), 420-429.

Clarke, A. J. & Cooper, D. N. (2010) GWAS: heritability missing in action? *Eur J Hum Genet*, 18(8), 859-61.

Clarke, T. K., Zeng, Y., Navrady, L., Xia, C., Haley, C. S., Campbell, A., Navarro, P., Amador, C., Adams, M. J., Howard, D. M. & Soler, A. (2018) Genetic and environmental determinants of stressful life events and their overlap with depression and neuroticism. *Wellcome open research*, 3.

Coleman, J. R. I., Peyrot, W. J., Purves, K. L., Davis, K. A. S., Rayner, C., Choi, S. W., Hübel, C., Gaspar, H. A., Kan, C., Van der Auwera, S., Adams, M. J., Lyall, D. M., Choi, K. W., Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. F. M., Bacanu, S. A., Bækvad-Hansen, M., Beekman, A. T. F., Bigdeli, T. B., Binder, E. B., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J. H., Clarke, T. K., Coleman, J. R. I., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G. E., Davies, G., Deary, I. J., Degenhardt, F., Derks, E. M., Direk, N., Dolan, C. V., Dunn, E. C., Eley, T. C., Escott-Price, V., Kiadeh, F. F. H., Finucane, H. K., Foo, J. C., Forstner, A. J., Frank, J., Gaspar, H. A., Gill, M., Goes, F. S., Gordon, S. D., Grove, J., Hall, L. S., Hansen, C. S., Hansen, T. F., Herms, S., Hickie, I. B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J. J., Hougaard, D. M., Howard, D. M., Ising, M., Jansen, R., Jones, I., Jones, L. A., Jorgenson, E., Knowles, J. A., Kohane, I. S., Kraft, J., Kretzschmar, W. W., Kutalik, Z., Li, Y., Lind, P. A., MacIntyre, D. J., MacKinnon, D. F., Maier, R. M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S. E., Mehta, D., Middeldorp, C. M., Mihailov, E., Milaneschi, Y., Milani, L., Mondimore, F. M., Montgomery, G. W., Mostafavi, S., Mullins, N., et al (2020) Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *Molecular Psychiatry*.

Colodro-Conde, L., Couvy-Duchesne, B., Zhu, G., Coventry, W. L., Byrne, E. M., Gordon, S., Wright, M. J., Montgomery, G. W., Madden, P. A. F., Ripke, S., Eaves, L. J., Heath, A. C., Wray, N. R., Medland, S. E. & Martin, N. G. (2018) A direct test of the diathesis–stress model for depression. *Molecular Psychiatry*.

Conejero, I., Olie, E., Calati, R., Ducasse, D. & Courtet, P. (2018) Psychological Pain, Depression, and Suicide: Recent Evidences and Future Directions. *Current Psychiatry Reports*, 20(5).

Conley, D., Laidley, T. M., Boardman, J. D. & Domingue, B. W. (2016) Changing Polygenic Penetrance on Phenotypes in the 20th Century Among Adults in the US Population. *Scientific Reports*, 6.

Coombes, B. J., Ploner, A., Bergen, S. E. & Biernacka, J. M. (2020) A principal component approach to improve association testing with polygenic risk scores. *Genet Epidemiol*, 44(7), 676-686.

Corfield, E. C., Yang, Y., Martin, N. G. & Nyholt, D. R. (2017) A continuum of genetic liability for minor and major depression. *Transl Psychiatry*, 7(5), e1131.

Crowe, R. R. (1993) Candidate genes in psychiatry: an epidemiological perspective. *Am J Med Genet*, 48(2), 74-7.

Cuijpers, P., Karyotaki, E., de Wit, L. & Ebert, D. D. (2020) The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychother Res*, 30(3), 279-293.

Culverhouse, R. C., Saccone, N. L., Horton, A. C., Ma, Y., Anstey, K. J., Banaschewski, T., Burmeister, M., Cohen-Woods, S., Etain, B., Fisher, H. L., Goldman, N., Guillaume, S., Horwood, J., Juhasz, G., Lester, K. J., Mandelli, L., Middeldorp, C. M., Olié, E., Villafuerte, S., Air, T. M., Araya, R., Bowes, L., Burns, R., Byrne, E. M., Coffey, C., Coventry, W. L., Gawronski, K. A. B., Glei, D., Hatzimanolis, A., Hottenga, J. J., Jaussent, I., Jawahar, C., Jennen-Steinmetz, C., Kramer, J. R., Lajnef, M., Little, K., Zu Schwabedissen, H. M., Nauck, M., Nederhof, E., Petschner, P., Peyrot, W. J., Schwahn, C., Sinnamon, G., Stacey, D., Tian, Y., Toben, C., Van Der Auwera, S., Wainwright, N., Wang, J. C., Willemsen, G., Anderson, I. M., Arolt, V., Aslund, C., Bagdy, G., Baune, B. T., Bellivier, F., Boomsma, D. I., Courtet, P., Dannlowski, U., De Geus, E. J. C., Deakin, J. F. W., Easteal, S., Eley, T., Fergusson, D. M., Goate, A. M., Gonda, X., Grabe, H. J., Holzman, C., Johnson, E. O., Kennedy, M., Laucht, M., Martin, N. G., Munafò, M. R., Nilsson, K. W., Oldehinkel, A. J., Olsson, C. A., Ormel, J., Otte, C., Patton, G. C., Penninx, B. W. J. H., Ritchie, K., Sarchiapone, M., Scheid, J. M., Serretti, A., Smit, J. H., Stefanis, N. C., Surtees, P. G., Völzke, H., Weinstein, M., Whooley, M., Nurnberger, J. I., Breslau, N. & Bierut, L. J. (2018) Collaborative meta-Analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Molecular Psychiatry*.

Dannehl, K., Rief, W. & Euteneuer, F. (2017) Childhood adversity and cognitive functioning in patients with major depression. *Child Abuse Negl*, 70, 247-254.

Davis, K. & Hotopf, M. (2019) Mental health phenotyping in UK Biobank.

Davis, K. A. S., Coleman, J. R. I., Adams, M., Allen, N., Breen, G., Cullen, B., Dickens, C., Fox, E., Graham, N., Holliday, J., Howard, L. M., John, A., Lee, W., McCabe, R., McIntosh, A., Pearsall, R., Smith, D. J., Sudlow, C., Ward, J., Zammit, S. & Hotopf, M. (2020) Mental health in UK Biobank - development, implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis. *BJPsych Open*, 6(2), e18.

de Oliveira, I. R., Matos-Ragazzo, A. C., Zhang, Y. N., Vasconcelos, N. M., Velasquez, M. L., Reis, D., Ribeiro, M. G., da Rocha, M. M., Rosario, M. C., Stallard, P. & Cecil, C. A. M. (2018) Disentangling the mental health impact of childhood abuse and neglect: A replication and

extension study in a Brazilian sample of high-risk youth. *Child Abuse & Neglect*, 80, 312-323.

de Zeeuw, E. L., van Beijsterveldt, C. E. M., Glasner, T. J., de Geus, E. J. C. & Boomsma, D. I. (2016) Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. *Learning and Individual Differences*.

Demirkan, A., Penninx, B. W. J. H., Hek, K., Wray, N. R., Amin, N., Aulchenko, Y. S., Van Dyck, R., De Geus, E. J. C., Hofman, A., Uitterlinden, A. G., Hottenga, J. J., Nolen, W. A., Oostra, B. A., Sullivan, P. F., Willemsen, G., Zitman, F. G., Tiemeier, H., Janssens, A. C. J. W., Boomsma, D. I., Van Duijn, C. M. & Middeldorp, C. M. (2011) Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Molecular Psychiatry*.

Di Florio, A., Putnam, K., Altemus, M., Apter, G., Bergink, V., Bilszta, J., Brock, R., Buist, A., Deligiannidis, K. M., Devouche, E., Epperson, C. N., Guille, C., Kim, D., Lichtenstein, P., Magnusson, P. K. E., Martinez, P., Munk-Olsen, T., Newport, J., Payne, J., Penninx, B. W., O'Hara, M., Robertson-Blackmore, E., Roza, S. J., Sharkey, K. M., Stuart, S., Tiemeier, H., Viktorin, A., Schmidt, P. J., Sullivan, P. F., Stowe, Z. N., Wisner, K. L., Jones, I., Rubinow, D. R., Meltzer-Brody, S. & Action, P. D. (2017) The impact of education, country, race and ethnicity on the self-report of postpartum depression using the Edinburgh Postnatal Depression Scale. *Psychological Medicine*, 47(5), 787-799.

Dikilitas, O., Schaid, D. J., Kosel, M. L., Carroll, R. J., Chute, C. G., Denny, J. A., Fedotov, A., Feng, Q. P., Hakonarson, H., Jarvik, G. P., Lee, M. T. M., Pacheco, J. A., Rowley, R., Sleiman, P. M., Stein, C. M., Sturm, A. C., Wei, W. Q., Wiesner, G. L., Williams, M. S., Zhang, Y. F., Manolio, T. A. & Kullo, I. J. (2020) Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three Major Racial and Ethnic Groups. *American Journal of Human Genetics*, 106(5), 707-716.

Docherty, A. R., Moscati, A., Peterson, R., Edwards, A. C., Adkins, D. E., Bacanu, S. A., Bigdeli, T. B., Webb, B. T., Flint, J. & Kendler, K. S. (2016) SNP-based heritability estimates of the personality dimensions and polygenic prediction of both neuroticism and major depression: findings from CONVERGE. *Translational Psychiatry*, 6.

Domingue, B. W. & Fletcher, J. (2020) Separating Measured Genetic and Environmental Effects: Evidence Linking Parental Genotype and Adopted Child Outcomes. *Behavior Genetics*.

Dudbridge, F. (2013) Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*.

Duncan, L. E., Ostacher, M. & Ballon, J. (2019) How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology*, 44(9), 1518-1523.

Dunn, E. C., Nishimi, K., Powers, A. & Bradley, B. (2017) Is developmental timing of trauma exposure associated with depressive and post-traumatic stress disorder symptoms in adulthood? *Journal of Psychiatric Research*, 84, 119-127

Dunn, E. C., Wiste, A., Radmanesh, F., Almli, L. M., Gogarten, S. M., Sofer, T., Faul, J. D., Kardia, S. L., Smith, J. A., Weir, D. R., Zhao, W., Soare, T. W., Mirza, S. S., Hek, K., Tiemeier, H., Goveas, J. S., Sarto, G. E., Snively, B. M., Cornelis, M., Koenen, K. C., Kraft, P., Purcell, S., Ressler, K. J., Rosand, J., Wassertheil-Smoller, S. & Smoller, J. W. (2016) Genome-Wide Association Study (Gwas) and Genome-Wide by Environment Interaction Study (Gweis) of Depressive Symptoms in African American and Hispanic/Latina Women. *Depress Anxiety*, 33(4), 265-80.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H. & Nadeau, J. H. (2010) Missing heritability and strategies for finding the underlying causes of complex disease.

Elston, R. C., Satagopan, J. M. & Sun, S. (2012) Genetic terminology. *Methods Mol Biol*, 850, 1-9.

Evans, D. M., Gillespie, N. A. & Martin, N. G. (2002) Biometrical genetics. *Biol Psychol*, 61(1-2), 33-51.

Eysenck, S. B., Eysenckm H. J. & Barrett, P. (1985) A revised version of the psychoticism scale. *Personality and individual differences*, 6(1), pp.21-29.

Falconer, D. S. & Mackay, T. F. C. (1996) *Introduction to Quantitative Genetics (Fourth Edition)*.

Feldman, R. & Bakermans-Kranenburg, M. J. (2017) Oxytocin: a parenting hormone. *Current Opinion in Psychology*, 15, 13-18.

Felsenstein, J. (2005) Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360(1459), 1427-1434.

Fergusson, D. M., Horwood, L. J., Miller, A. L. & Kennedy, M. A. (2011) Life stress, 5-HTTLPR and mental disorder: Findings from a 30-year longitudinal study. *British Journal of Psychiatry*.

Fernandez-Pujals, A. M., Adams, M. J., Thomson, P., McKechanie, A. G., Blackwood, D. H. R., Smith, B. H., Dominiczak, A. F., Morris, A. D., Matthews, K., Campbell, A., Linksted, P., Haley, C. S., Deary, I. J., Porteous, D. J., MacIntyre, D. J. & McIntosh, A. M. (2015) Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in generation Scotland: Scottish family health study (GS: SFHS). *PLoS ONE*, 10(11), e0142197-e0142197.

Flint, J. & Kendler, K. S. (2014) The Genetics of Major Depression. *Neuron*, 81(5), 1214.

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M. & Lawlor, D. A. (2013) Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*, 42(1), 97-110.

Fried, E. (2017) Moving forward: how depression heterogeneity hinders progress in treatment and research. *Expert Review of Neurotherapeutics*, 17(5), 423-425.

Fried, E. I. & Nesse, R. M. (2015) Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, 172, 96-102.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N. & Allen, N. E. (2016) The representativeness of the UK Biobank cohort on a range of sociodemographic, physical, lifestyle and health-related characteristics. *Journal of Epidemiology & Community Health*, 70(1).

Fusar-Poli, P., Nelson, B., Valmaggia, L., Yung, A. R. & McGuire, P. K. (2014) Comorbid Depressive and Anxiety Disorders in 509 Individuals With an At-Risk Mental State: Impact on Psychopathology and Transition to Psychosis. *Schizophrenia Bulletin*, 40(1), 120-131.

Gage, S. H., Smith, G. D., Ware, J. J., Flint, J. & Munafò, M. R. (2016) Correction: G = E: What GWAS Can Tell Us about the Environment. *PLOS Genetics*, 12(5), e1006065-e1006065.

Gallo, E. A. G., Munhoz, T. N., Loret de Mola, C. & Murray, J. (2018) Gender differences in the effects of childhood maltreatment on adult depression and anxiety: A systematic review and meta-analysis. *Child Abuse Negl*, 79, 107-114.

Ghio, L., Gotelli, S., Marcenaro, M., Amore, M. & Natta, W. (2014) Duration of untreated illness and outcomes in unipolar depression: A systematic review and meta-analysis. *Journal of Affective Disorders*, 152, 45-51.

Giannakopoulou, O., Lin, K., Meng, X., Su, M. H., Kuo, P. H., Peterson, R. E., Awasthi, S., Moscati, A., Coleman, J. R. I., Bass, N., Millwood, I. Y., Chen, Y., Chen, Z., Chen, H. C., Lu, M. L., Huang, M. C., Chen, C. H., Stahl, E. A., Loos, R. J. F., Mullins, N., Ursano, R. J., Kessler, R. C., Stein, M. B., Sen, S., Scott, L. J., Burmeister, M., Fang, Y., Tyrrell, J., Jiang, Y., Tian, C., McIntosh, A. M., Ripke, S., Dunn, E. C., Kendler, K. S., Walters, R. G., Lewis, C. M., Kuchenbaecker, K., andMe Research Team, C. K. B. C. G. & Major Depressive Disorder Working Group of the Psychiatric Genomics, C. (2021) The Genetic Architecture of Depression in Individuals of East Asian Ancestry: A Genome-Wide Association Study. *JAMA Psychiatry*, 78(11), 1258-1269.

Gillett, A. C., Jermy, B., Lee, S. H., Pain, O., Howard, D. M., Hagenaars, S., Hanscombe, K., Coleman, J. R. & Lewis, C. M. (2021) Exploring polygenic-environment and residual-environment interactions for depressive symptoms within the UK Biobank. *medRxiv*.

Glaesmer, H., Brähler, E., Schulz, A., Freyberger, H., Grabe, H. J. & Häuser, W. (2013) The childhood trauma screener (CTS) - Development and validation of cut-off-scores for classificatory diagnostics. *Psychiatrische Praxis*.

Golan, D., Lander, E. S. & Rosset, S. (2014) Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A*, 111(49), E5272-81.

Goldberg, D. (2011) The heterogeneity of "major depression". *World Psychiatry*, 10(3), 226-228.

Goldberg, D. P. & Hillier, V. F. (1979) A scaled version of the General Health Questionnaire. *Psychological Medicine*.

Grabe, H. J., Schulz, A., Schmidt, C. O., Appel, K., Driessen, M., Wingenfeld, K., Barnow, S., Spitzer, C., John, U., Berger, K., Wersching, H. & Freyberger, H. J. (2012) [A brief instrument for the assessment of childhood abuse and neglect: the childhood trauma screener (CTS)]. *Psychiatr Prax*, 39(3), 109-15.

Greenberg, P. E., Fournier, A. A., Sisitsky, T., Pike, C. T. & Kessler, R. C. (2015) The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry*, 76(2), 155-62.

Gressier, F., Calati, R. & Serretti, A. (2016) 5-HTTLPR and gender differences in affective disorders: A systematic review. *J Affect Disord*, 190, 193-207.

Griffing, B. (1967) Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Australian journal of biological sciences*.

Groen, R. N., Ryan, O., Wigman, J. T. W., Riese, H., Penninx, B. W. J. H., Giltay, E. J., Wichers, M. & Hartman, C. A. (2020) Comorbidity between depression and anxiety: assessing the role of bridge mental states in dynamic psychological networks. *Bmc Medicine*, 18(1).

Gunderson, K. L. (2009) Whole-Genome Genotyping on Bead Arrays BT - Methods Mol Biol.

Hammen, C. (2016) Depression and stressful environments: identifying gaps in conceptualization and measurement. *Anxiety Stress Coping*, 29(4), 335-51.

Hankin, B. L., Young, J. F., Abela, J. R., Smolen, A., Jenness, J. L., Gulley, L. D., Technow, J. R., Gottlieb, A. B., Cohen, J. R. & Oppenheimer, C. W. (2015) Depression from childhood into late adolescence: Influence of gender, development, genetic susceptibility, and peer stress. *Journal of Abnormal Psychology*, 124(4), 803.

Harold, G. T., Leve, L. D., Barrett, D., Elam, K., Neiderhiser, J. M., Natsuaki, M. N., Shaw, D. S., Reiss, D. & Thapar, A. (2013) Biological and rearing mother influences on child ADHD symptoms: revisiting the developmental interface between nature and nurture. *Journal of Child Psychology and Psychiatry*, 54(10), 1038-1046.

Haroz, E. E., Ritchey, M., Bass, J. K., Kohrt, B. A., Augustinavicius, J., Michalopoulos, L., Burkey, M. D. & Bolton, P. (2017) How is depression experienced around the world? A systematic review of qualitative literature. *Social Science & Medicine*, 183, 151-162.

Heckerman, D., Gurdasani, D., Kadie, C., Pomilla, C., Carstensen, T., Martin, H., Ekoru, K., Nsubuga, R. N., Ssenyomo, G., Kamali, A., Kaleebu, P., Widmer, C. & Sandhu, M. S. (2016) Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7377-7382.

Hek, K., Demirkan, A., Lahti, J., Terracciano, A., Teumer, A., Cornelis, M. C., Amin, N., Bakshis, E., Baumert, J., Ding, J. Z., Liu, Y. M., Marciante, K., Meirelles, O., Nalls, M. A., Sun, Y. V., Vogelzangs, N., Yu, L., Bandinelli, S., Benjamin, E. J., Bennett, D. A., Boomsma, D., Cannas, A., Coker, L. H., de Geus, E., De Jager, P. L., Diez-Roux, A. V., Purcell, S., Hu, F. B., Rimm, E. B., Hunter, D. J., Jensen, M. K., Curhan, G., Rice, K., Penman, A. D., Rotter, J. I., Sotoodehnia, N., Emeny, R., Eriksson, J. G., Evans, D. A., Ferrucci, L., Fornage, M.,

Gudnason, V., Hofman, A., Illig, T., Kardia, S., Kelly-Hayes, M., Koenen, K., Kraft, P., Kuningas, M., Massaro, J. M., Melzer, D., Mulas, A., Mulder, C. L., Murray, A., Oostra, B. A., Palotie, A., Penninx, B., Petersmann, A., Pilling, L. C., Psaty, B., Rawal, R., Reiman, E. M., Schulz, A., Shulman, J. M., Singleton, A. B., Smith, A. V., Sutin, A. R., Uitterlinden, A. G., Volzke, H., Widen, E., Yaffe, K., Zonderman, A. B., Cucca, F., Harris, T., Ladwig, K. H., Llewellyn, D. J., Raikkonen, K., Tanaka, T., van Duijn, C. M., Grabe, H. J., Launer, L. J., Lunetta, K. L., Mosley, T. H., Newman, A. B., Tiemeier, H. & Murabito, J. (2013) A Genome-Wide Association Study of Depressive Symptoms. *Biological Psychiatry*, 73(7), 667-678.

Hopfinger, L., Berking, M., Bockting, C. L. H. & Ebert, D. D. (2016) Emotion regulation mediates the effect of childhood trauma on depression. *Journal of Affective Disorders*.

Hoppen, T. H. & Chalder, T. (2018) Childhood adversity as a transdiagnostic risk factor for affective disorders in adulthood: A systematic review focusing on biopsychosocial moderating and mediating variables. *Clinical Psychology Review*, 65, 81-151.

Hosang, G. M., Shiles, C., Tansey, K. E., McGuffin, P. & Uher, R. (2014) Interaction between stress and the BDNF Val66Met polymorphism in depression: a systematic review and meta-analysis. *Bmc Medicine*, 12.

Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., Coleman, J. R. I., Hagenaars, S. P., Ward, J., Wigmore, E. M., Alloza, C., Shen, X., Barbu, M. C., Xu, E. Y., Whalley, H. C., Marioni, R. E., Porteous, D. J., Davies, G., Deary, I. J., Hemani, G., Berger, K., Teismann, H., Rawal, R., Arolt, V., Baune, B. T., Dannlowski, U., Domschke, K., Tian, C., Hinds, D. A., Trzaskowski, M., Byrne, E. M., Ripke, S., Smith, D. J., Sullivan, P. F., Wray, N. R., Breen, G., Lewis, C. M. & McIntosh, A. M. (2019) Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*.

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Lind, P. A., Palviainen, T., Zee, M. D. v. d., Cheesman, R., Mangino, M., Wang, Y., Li, S., Klaric, L., Ratliff, S. M., Bielak, L. F., Nygaard, M., Reynolds, C. A., Balbona, J. V., Bauer, C. R., Boomsma, D. I., Baras, A., Campbell, A., Campbell, H., Chen, Z., Christofidou, P., Dahm, C. C., Dokuru, D. R., Evans, L. M., Geus, E. J. C. d., Giddaluru, S., Gordon, S. D., Harden, K. P., Havdahl, A., Hill, W. D., Kerr, S. M., Kim, Y., Kweon, H., Latvala, A., Li, L., Lin, K., Martikainen, P., Magnusson, P. K. E., Mills, M. C., Lawlor, D. A., Overton, J. D., Pedersen, N. L., Porteous, D. J., Reid, J., Silventoinen, K., Southey, M. C., Mallard, T. T., Tucker-Drob, E. M., Wright, M. J., Consortium, S. S. G. A., Consortium, W. F., Hewitt, J. K., Keller, M. C., Stallings, M. C., Christensen, K., Kardia, S. L. R., Peyser, P. A., Smith, J. A., Wilson, J. F., Hopper, J. L., Hägg, S., Spector, T. D., Pingault, J.-B., Plomin, R., Bartels, M., Martin, N. G., Justice, A. E., Millwood, I. Y., Hveem, K., Naess, Ø., Willer, C. J., Åsvold, B. O., Koellinger, P. D., Kaprio, J., Medland, S. E., Walters, R. G., Benjamin, D. J., Turley, P., Evans, D. M., Smith, G. D., Hayward, C., Brumpton, B., Hemani, G. & Davies, N. M. (2021) Within-sibship GWAS improve estimates of direct genetic effects. *bioRxiv*, 2021.03.05.433935-2021.03.05.433935.

Hoy-Ellis, C. P. & Fredriksen-Goldsen, K. I. (2017) Depression Among Transgender Older Adults: General and Minority Stress. *American Journal of Community Psychology*, 59(3-4), 295-305.

Hufer, A., Kornadt, A. E., Kandler, C. & Riemann, R. (2020) Genetic and environmental variation in political orientation in adolescence and early adulthood: A Nuclear Twin Family analysis. *J Pers Soc Psychol*, 118(4), 762-776.

Hugh-Jones, D., Verweij, K. J. H., St. Pourcain, B. & Abdellaoui, A. (2016) Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores. *Intelligence*.

Hughes, K., Bellis, M. A., Hardcastle, K. A., Sethi, D., Butchart, A., Mikton, C., Jones, L. & Dunne, M. P. (2017) The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. *The Lancet Public Health*.

Huh, H. J., Kim, K. H., Lee, H. K. & Chae, J. H. (2017) The relationship between childhood trauma and the severity of adulthood depression and anxiety symptoms in a clinical sample: The mediating role of cognitive emotion regulation strategies. *Journal of Affective Disorders*.

Hung, C. I., Liu, C. Y. & Yang, C. H. (2017) Untreated duration predicted the severity of depression at the two-year follow-up point. *Plos One*, 12(9).

Jaffee, S. R. (2017) Child Maltreatment and Risk for Psychopathology in Childhood and Adulthood. *Annu Rev Clin Psychol*, 13, 525-551.

Jaffee, S. R. & Price, T. S. (2007) Gene-environment correlations: a review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*, 12(5), 432-442.

Jaffee, S. R. & Price, T. S. (2008) Genotype–environment correlations: implications for determining the relationship between environmental exposures and psychiatric illness. *Psychiatry*, 7(12), 496-499.

Jaffee, S. R., Sligo, J. L., McAnally, H. M., Bolton, A. E., Baxter, J. M. & Hancox, R. J. (2021) Early-onset and recurrent depression in parents increases risk of intergenerational transmission to adolescent offspring. *Journal of Child Psychology and Psychiatry*, 62(8), 979-988.

Jami, E. S., Eilertsen, E. M., Hammerschlag, A. R., Qiao, Z., Evans, D. M., Ystrom, E., Bartels, M. & Middeldorp, C. M. (2020) Maternal and paternal effects on offspring internalizing problems: Results from genetic and family-based analyses. *Am J Med Genet B Neuropsychiatr Genet*, 183(5), 258-267.

Jang, K. L., Vernon, P. A., Livesley, W. J., Stein, M. B. & Wolf, H. (2001) Intra- and extra-familial influences on alcohol and drug misuse: a twin study of gene-environment correlation. *Addiction*, 96(9), 1307-18.

Jermy, B. S., Glanville, K. P., Coleman, J. R. I., Lewis, C. M. & Vassos, E. (2021) Exploring the genetic heterogeneity in major depression across diagnostic criteria. *Mol Psychiatry*.

Jermy, B. S., Hagenaars, S. P., Glanville, K. P., Coleman, J. R. I., Howard, D. M., Breen, G., Vassos, E. & Lewis, C. M. (2022) Using major depression polygenic risk scores to explore the depressive symptom continuum. *Psychological Medicine*, 52(1), 149-158.

Jia, P., Kao, C. F., Kuo, P. H. & Zhao, Z. (2011) A comprehensive network and pathway analysis of candidate genes in major depressive disorder. *BMC Syst Biol*, 5 Suppl 3, S12.

Johnson, S. B., Riis, J. L. & Noble, K. G. (2016) State of the Art Review: Poverty and the Developing Brain. *Pediatrics*, 137(4).

Johnston, K. M., Powell, L. C., Anderson, I. M., Szabo, S. & Cline, S. (2019) The burden of treatment-resistant depression: A systematic review of the economic and quality of life literature. *Journal of Affective Disorders*, 242, 195-210.

Josefsson, A., Vikstrom, J., Bladh, M. & Sydsjo, G. (2019) Major depressive disorder in women and risk for future generations: population-based three-generation study. *Bjpsych Open*, 5(1).

Kautzky, A., Dold, M., Bartova, L., Spies, M., Kranz, G. S., Souery, D., Montgomery, S., Mendlewicz, J., Zohar, J., Fabbri, C., Serretti, A., Lanzenberger, R., Dikeos, D., Rujescu, D. & Kasper, S. (2019) Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study. *Acta Psychiatrica Scandinavica*, 139(1), 78-88.

Kendall, K. M., Van Assche, E., Andlauer, T. F. M., Choi, K. W., Luykx, J. J., Schulte, E. C. & Lu, Y. (2021) The genetic basis of major depression. *Psychol Med*, 51(13), 2217-2230.

Kendler, K. S. (1996) Major depression and generalised anxiety disorder. Same genes, (partly)different environments--revisited. *Br J Psychiatry Suppl*(30), 68-75.

Kendler, K. S. & Baker, J. H. (2007) Genetic influences on measures of the environment: A systematic review.

Kendler, K. S., Bulik, C. M., Silberg, J., Hettema, J. M., Myers, J. & Prescott, C. A. (2000) Childhood sexual abuse and adult psychiatric and substance use disorders in women - An epidemiological and Cotwin control analysis. *Archives of General Psychiatry*, 57(10), 953-959.

Kendler, K. S. & Gardner, C. O. (2014) Sex differences in the pathways to major depression: a study of opposite-sex twin pairs. *Am J Psychiatry*, 171(4), 426-35.

Kendler, K. S., Gatz, M., Gardner, C. O. & Pedersen, N. L. (2006) Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch Gen Psychiatry*, 63(10), 1113-20.

Kendler, K. S., Karkowski, L. M. & Prescott, C. A. (1999) Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry*, 156(6), 837-841.

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., Üstün, T. B. & Wang, P. S. (2009) The global burden of mental disorders: An update from the WHO World Mental Health (WMH) surveys. *Epidemiologia e Psichiatria Sociale*.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R. & Walters, E. E. (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*, 62(6), 593-602.

Kessler, R. C. & Bromet, E. J. (2013) The epidemiology of depression across cultures. *Annu Rev Public Health*, 34, 119-38.

Kessler, R. C., McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., Aguilar-Gaxiola, S., Alhamzawi, A. O., Alonso, J., Angermeyer, M., Benjet, C., Bromet, E., Chatterji, S., de Girolamo, G., Demyttenaere, K., Fayyad, J., Florescu, S., Gal, G., Gureje, O., Haro, J. M., Hu, C. Y., Karam, E. G., Kawakami, N., Lee, S., Lepine, J. P., Ormel, J., Posada-Villa, J., Sagar, R., Tsang, A., Ustun, T. B., Vassilev, S., Viana, M. C. & Williams, D. R. (2010) Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *Br J Psychiatry*, 197(5), 378-85.

Kessler, R. C. & Ustun, T. B. (2004) The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*, 13(2), 93-121.

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T. & Kathiresan, S. (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*, 50(9), 1219-1224.

Klahr, A. M., Burt, S. A., Leve, L. D., Shaw, D. S., Ganiban, J. M., Reiss, D. & Neiderhiser, J. M. (2017) Birth and Adoptive Parent Antisocial Behavior and Parenting: A Study of Evocative Gene-Environment Correlation. *Child Dev*, 88(2), 505-513.

Knafo, A. & Jaffee, S. R. (2013) Gene-environment correlation in developmental psychopathology.

Kobiella, A., Reimold, M., Ulshöfer, D. E., Ikonomidou, V. N., Vollmert, C., Vollstädt-Klein, S., Rietschel, M., Reischl, G. & Smolka, M. N. (2011) How the serotonin transporter 5-HTTLPR polymorphism influences amygdala function: the roles of in vivo serotonin transporter expression and amygdala structure. *Translational Psychiatry*, 1(8), e37.

Koeter, M. W. J. (1992) Validity of the GHQ and SCL anxiety and depression scales: A comparative study. *Journal of Affective Disorders*.

Kong, A., Masson, G., Frigge, M. L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P. I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F., Thorsteinsdottir, U., Gudbjartsson, D. F., Stefansson, H. & Stefansson, K. (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet*, 40(9), 1068-75.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U. & Stefansson, K. (2018) The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424-428.

Kucharska, J. (2017) Sexual and Non-Sexual Trauma, Depression and Self-Esteem in a Sample of Polish Women. A Cross-Sectional Study. *Clinical Psychology & Psychotherapy*, 24(1), 186-194.

Kuehner, C. (2017) Why is depression more common among women than among men? *The Lancet Pyschiatry*, 4(2), 146-158.

Kumar, S. K., Feldman, M. W., Rehkopf, D. H. & Tuljapurkar, S. (2016) Limitations of GCTA as a solution to the missing heritability problem. *Proceedings of the National Academy of Sciences*, 113(1), E61-E70.

Lambert, S. A., Abraham, G. & Inouye, M. (2019) Towards clinical utility of polygenic risk scores. *Hum Mol Genet*, 28(R2), R133-R142.

Lau, J. Y. F. & Eley, T. C. (2008) Disentangling gene-environment correlations and interactions on adolescent depressive symptoms. *Journal of Child Psychology and Psychiatry and Allied Disciplines*.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., Yengo, L., Agee, M., Alipanahi, B., Auton, A., Bell, R. K., Bryc, K., Elson, S. L., Fontanillas, P., Hinds, D. A., McCreight, J. C., Huber, K. E., Litterman, N. K., McIntyre, M. H., Mountain, J. L., Noblin, E. S., Northover, C. A. M., Pitts, S. J., Sathirapongsasuti, J. F., Sazonova, O. V., Shelton, J. F., Shringarpure, S., Tian, C., Vacic, V., Wilson, C. H., Beauchamp, J. P., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G. B., Emilsson, V., Meddens, S. F. W., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., Vlaming, R. d., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., Bjornsdottir, G., Brandsma, J. H., Concas, M. P., Derringer, J., Furlotte, N. A., Galesloot, T. E., Girotto, G., Gupta, R., Hall, L. M., Harris, S. E., Hofer, E., Horikoshi, M., Huffman, J. E., Kaasik, K., Kalafati, I. P., Karlsson, R., Kong, A., Lahti, J., van der Lee, S. J., Leeuw, C. d., Lind, P. A., Lindgren, K. O., Liu, T., Mangino, M., Marten, J., Mihailov, E., Miller, M. B., van der Most, P. J., Oldmeadow, C., Payton, A., Pervjakova, N., Peyrot, W. J., Qian, Y., Raitakari, O., Rueedi, R., Salvi, E., Schmidt, B., Schraut, K. E., Shi, J., Smith, A. V., Poot, R. A., et al (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*.

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., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O. A., Anjorin, A., Anney, R., Anttila, V., Arking, D. E., Asherson, P., Azevedo, M. H., Backlund, L., Badner, J. A., Bailey, A. J., Banaschewski, T., Barchas, J. D., Barnes, M. R., Barrett, T. B., Bass, N., Battaglia, A., Bauer, M., Bayes, M., Bellivier, F., Bergen, S. E., Berrettini, W., Betancur, C., Bettecken, T., Biederman, J., Binder, E. B., Black, D. W., Blackwood, D. H. R., Bloss, C. S., Boehnke, M., Boomsma, D. I., Breen, G., Breuer, R., Bruggeman, R., Cormican, P., Buccola, N. G., Buitelaar, J. K., Bunney, W. E., Buxbaum, J. D., Byerley, W. F., Byrne, E. M., Caesar, S., Cahn, W., Cantor, R. M., Casas, M., Chakravarti, A., Chambert, K., Choudhury, K., Cichon, S., Cloninger, C. R., Collier, D. A., Cook, E. H., Coon, H., Cormand, B., Corvin, A., Coryell, W. H., Craig, D. W., Craig, I. W., Crosbie, J., Cuccaro, M. L., Curtis, D., Czamara, D., Datta, S., Dawson, G., Day, R., De Geus, E. J., Degenhardt, F., Djurovic, S., Donohoe, G. J., Doyle, A. E., Duan, J. B., Dudbridge, F., Duketis, E., Ebstein, R. P., Edenberg, H. J., Elia, J., Ennis, S., Etain, B., Fanous, A., Farmer, A. E., Ferrier, I. N., Flickinger, M., Fombonne, E., Foroud, T., Frank, J., Franke, B., Fraser, C., et al (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, 45(9), 984-+.

Lehto, K., Hagg, S., Lu, D., Karlsson, R., Pedersen, N. L. & Mosing, M. A. (2020) Childhood Adoption and Mental Health in Adulthood: The Role of Gene-Environment Correlations and Interactions in the UK Biobank. *Biol Psychiatry*, 87(8), 708-716.

LeMoult, J., Humphreys, K. L., Tracy, A., Hoffmeister, J. A., Ip, E. & Gotlib, I. H. (2020) Meta-analysis: Exposure to Early Life Stress and Risk for Depression in Childhood and Adolescence. *J Am Acad Child Adolesc Psychiatry*, 59(7), 842-855.

Levinson, D. F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N. R. & Sullivan, P. F. (2014) Genetic Studies of Major Depressive Disorder: Why Are There No Genome-wide Association Study Findings and What Can We Do About It? *Biological Psychiatry*, 76(7), 510-512.

Lewis, C. M. & Vassos, E. (2020) Polygenic risk scores: from research tools to clinical instruments. *Genome Med*, 12(1), 44.

Li, M., D'Arcy, C. & Meng, X. (2016) Maltreatment in childhood substantially increases the risk of adult depression and anxiety in prospective cohort studies: systematic review, meta-analysis, and proportional attributable fractions. *Psychol Med*, 46(4), 717-30.

Lindert, J., von Ehrenstein, O. S., Grashow, R., Gal, G., Braehler, E. & Weisskopf, M. G. (2014) Sexual and physical abuse in childhood is associated with depression and anxiety over the life course: systematic review and meta-analysis. *International Journal of Public Health*, 59(2), 359-372.

Liu, C. H. & Tronick, E. (2014) Prevalence and predictors of maternal postpartum depressed mood and anhedonia by race and ethnicity. *Epidemiol Psychiatr Sci*, 23(2), 201-9.

Lu, J. T., Campeau, P. M. & Lee, B. H. (2014) Genotype-phenotype correlation--promiscuity in the era of next-generation sequencing. *N Engl J Med*, 371(7), 593-6.

Lu, S., Peng, H., Wang, L., Vasish, S., Zhang, Y., Gao, W., Wu, W., Liao, M., Wang, M., Tang, H. & Li, W. (2013) Elevated specific peripheral cytokines found in major depressive disorder patients with childhood trauma exposure: a cytokine antibody array analysis. *Comprehensive Psychiatry*, 54(7), 953-961.

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T. K., Shirali, M., Harris, S. E., Marioni, R. E., Liewald, D. C., Fawns-Ritchie, C., Adams, M. J., Howard, D. M., Lewis, C. M., Gale, C. R., McIntosh, A. M. & Deary, I. J. (2018) Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*.

Lund, C., Brooke-Sumner, C., Baingana, F., Baron, E. C., Breuer, E., Chandra, P., Haushofer, J., Herrman, H., Jordans, M., Kieling, C., Medina-Mora, M. E., Morgan, E., Omigbodun, O., Tol, W., Patel, V. & Saxena, S. (2018) Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. *Lancet Psychiatry*, 5(4), 357-369.

Lyons, M. J., Goldberg, J., Eisen, S. A., True, W., Tsuang, M. T., Meyer, J. M. & Henderson, W. G. (1993) Do genes influence exposure to trauma? A twin study of combat. *Am J Med Genet*, 48(1), 22-7.

Mac Giollabhui, N., Hamilton, J. L., Nielsen, J., Connolly, S. L., Stange, J. P., Varga, S., Burdette, E., Olino, T. M., Abramson, L. Y. & Alloy, L. B. (2018) Negative Cognitive Style Interacts With Negative Life Events to Predict First Onset of a Major Depressive Episode in Adolescence via Hopelessness. *Journal of Abnormal Psychology*, 127(1), 1-11.

Mackay, T. F. C., Lynch, M. & Walsh, B. (1999) Quantitative GeneticsGenetics and Analysis of Quantitative Traits. *Evolution*.

Magnus, P., Birke, C., Vejrup, K., Haugan, A., Alsaker, E., Daltveit, A. K., Handal, M., Haugen, M., Hoiseth, G., Knudsen, G. P., Paltiel, L., Schreuder, P., Tambs, K., Vold, L. & Stoltenberg, C. (2016) Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol*, 45(2), 382-8.

Maher, B. S. (2015) Polygenic scores in epidemiology: risk prediction, etiology, and clinical utility. *Current epidemiology reports*, 2(4), 239-44.

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., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., MacKay, T. F. C., McCarroll, S. A. & Visscher, P. M. (2009) Finding the missing heritability of complex diseases.

Marazziti, D. (2017) Understanding the role of serotonin in psychiatric diseases. *F1000Res*, 6, 180.

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M. & Daly, M. J. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*.

Mathews, C. A. & Reus, V. I. (2001) Assortative mating in the affective disorders: A systematic review and meta-analysis. *Comprehensive Psychiatry*, 42(4), 257-262.

Matosin, N., Cruceanu, C. & Binder, E. B. (2017) Preclinical and Clinical Evidence of DNA Methylation Changes in Response to Trauma and Chronic Stress. *Chronic Stress (Thousand Oaks)*, 1.

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J. P., Chen, T. H., Wang, Q., Bolla, M. K., Yang, X., Adank, M. A., Ahearn, T., Aittomaki, K., Allen, J., Andrulis, I. L., Anton-Culver, H., Antonenkova, N. N., Arndt, V., Aronson, K. J., Auer, P. L., Auvinen, P., Barrdahl, M., Beane Freeman, L. E., Beckmann, M. W., Behrens, S., Benitez, J., Bermisheva, M., Bernstein, L., Blomqvist, C., Bogdanova, N. V., Bojesen, S. E., Bonanni, B., Borresen-Dale, A. L., Brauch, H., Bremer, M., Brenner, H., Brentnall, A., Brock, I. W., Brooks-Wilson, A., Brucker, S. Y., Bruning, T., Burwinkel, B., Campa, D., Carter, B. D., Castelao, J. E., Chanock, S. J., Chlebowski, R., Christiansen, H., Clarke, C. L., Collee, J. M., Cordina-Duverger, E., Cornelissen, S., Couch, F. J., Cox, A., Cross, S. S., Czene, K., Daly, M. B., Devilee, P., Dork, T., Dos-Santos-Silva, I., Dumont, M., Durcan, L., Dwek, M., Eccles, D. M., Ekici, A. B., Eliassen, A. H., Ellberg, C., Engel, C., Eriksson, M., Evans, D. G., Fasching, P. A., Figueroa, J., Fletcher, O., Flyger, H., Forsti, A., Fritschi, L., Gabrielson, M., Gago-Dominguez, M., Gapstur, S. M., Garcia-Saenz, J. A., Gaudet, M. M., Georgoulias, V., Giles, G. G., Gilyazova, I. R., Glendon, G., Goldberg, M. S., Goldgar, D. E., Gonzalez-Neira, A., Grenaker Alnaes, G. I., Grip, M., Gronwald, J., Grundy, A., Guenel, P., Haeberle, L., Hahnen, E., Haiman, C. A., Hakansson, N., Hamann, U., Hankinson, S. E., et al (2019) Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*, 104(1), 21-34.

Maxwell, J. M., Coleman, J. R. I., Breen, G. & Vassos, E. (2021) Association Between Genetic Risk for Psychiatric Disorders and the Probability of Living in Urban Settings. *Jama Psychiatry*.

Mehta, D., Bruenig, D., Carrillo-Roa, T., Lawford, B., Harvey, W., Morris, C. P., Smith, A. K., Binder, E. B., Young, R. M. & Voisey, J. (2017) Genomewide DNA methylation analysis in combat veterans reveals a novel locus for PTSD. *Acta Psychiatr Scand*, 136(5), 493-505.

Moonesinghe, R., Khoury, M. J., Liu, T. & Janssens, A. C. J. W. (2011) Discriminative accuracy of genomic profiling comparing multiplicative and additive risk models. *European Journal of Human Genetics*, 19(2), 180-185.

Moreno-Agostino, D., Wu, Y. T., Daskalopoulou, C., Hasan, M. T., Huisman, M. & Prina, M. (2021) Global trends in the prevalence and incidence of depression:a systematic review and meta-analysis. *J Affect Disord*, 281, 235-243.

Morris, T. T., Davies, N. M., Hemani, G. & Smith, G. D. (2020) Population phenomena inflate genetic associations of complex social traits. *Science Advances*, 6(16), eaay0328-eaay0328.

Moskvina, V., Farmer, A., Swainson, V., O'Leary, J., Gunasinghe, C., Owen, M., Craddock, N., McGuffin, P. & Korszun, A. (2007) Interrelationship of childhood trauma, neuroticism, and depressive phenotype. *Depression and Anxiety*.

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K. & Przeworski, M. (2020) Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, 9.

Mullins, N. & Lewis, C. M. (2017) Genetics of Depression: Progress at Last.

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., Uher, R., Cohen-Woods, S., Rivera, M., Jones, L., Jones, I., Craddock, N., Owen, M. J., Korszun, A., Craig, I. W., Farmer, A. E., McGuffin, P., Breen, G. & Lewis, C. M. (2016) Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological Medicine*.

Munafo, M. R. (2006) Candidate gene studies in the 21st century: meta-analysis, mediation, moderation. *Genes Brain and Behavior*, 5, 3-8.

Munkholm, K., Paludan-Muller, A. S. & Boesen, K. (2019) Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. *Bmj Open*, 9(6).

Musliner, K. L., Mortensen, P. B., McGrath, J. J., Suppli, N. P., Hougaard, D. M., Bybjerg-Grauholm, J., Baekvad-Hansen, M., Andreassen, O., Pedersen, C. B., Pedersen, M. G., Mors, O., Nordentoft, M., Borglum, A. D., Werge, T., Agerbo, E. & Bipolar Disorder Working Group of the Psychiatric Genomics, C. (2019) Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia With Risk for Depression in the Danish Population. *JAMA Psychiatry*, 76(5), 516-525.

Nanni, V., Uher, R. & Danese, A. (2012) Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. *American Journal of Psychiatry*.

Negele, A., Kaufhold, J., Kallenbach, L. & Leuzinger-Bohleber, M. (2015) Childhood Trauma and Its Relation to Chronic Depression in Adulthood. *Depression Research and Treatment*.

Nelson, R. M., Pettersson, M. E. & Carlborg, O. (2013) A century after Fisher: time for a new paradigm in quantitative genetics. *Trends in Genetics*, 29(12), 669-676.

Ni, G., Zeng, J., Revez, J. A., Wang, Y., Zheng, Z., Ge, T., Restuadi, R., Kiewa, J., Nyholt, D. R., Coleman, J. R. I., Smoller, J. W., Schizophrenia Working Group of the Psychiatric Genomics, C., Major Depressive Disorder Working Group of the Psychiatric Genomics, C., Yang, J., Visscher, P. M. & Wray, N. R. (2021) A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. *Biol Psychiatry*, 90(9), 611-620.

Nicholls, H. L., John, C. R., Watson, D. S., Munroe, P. B., Barnes, M. R. & Cabrera, C. P. (2020) Reaching the End-Game for GWAS: Machine Learning Approaches for the Prioritization of Complex Disease Loci. *Front Genet*, 11, 350.

Nivard, M. G., Middeldorp, C. M., Lubke, G., Hottenga, J. J., Abdellaoui, A., Boomsma, D. I. & Dolan, C. V. (2016) Detection of gene-environment interaction in pedigree data using genome-wide genotypes. *European Journal of Human Genetics*, 24(12), 1803-1809.

Norman, R. E., Byambaa, M., De, R., Butchart, A., Scott, J. & Vos, T. (2012) The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. *PLoS Med*, 9(11), e1001349.

Patel, V. (2017) Talking sensibly about depression. *Plos Medicine*, 14(4).

Pe'er, I., Yelensk, R., Altshuler, D. & Daly, M. J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology*, 32(4), 381-385.

Pener-Tessler, R., Avinun, R., Uzefovsky, F., Edelman, S., Ebstein, R. P. & Knafo, A. (2013) Boys' serotonin transporter genotype affects maternal behavior through self-control: A case of evocative gene-environment correlation. *Development and Psychopathology*, 25(1), 151-162.

Peres, M. F. P., Mercante, J. P. P., Tobo, P. R., Kamei, H. & Bigal, M. E. (2017) Anxiety and depression symptoms and migraine: a symptom-based approach research. *Journal of Headache and Pain*, 18.

Perlstein, S. & Waller, R. (2020) Integrating the study of personality and psychopathology in the context of gene-environment correlations across development. *Journal of Personality*.

Peterson, R. E., Cai, N., Dahl, A. W., Bigdeli, T. B., Edwards, A. C., Webb, B. T., Bacanu, S. A., Zaitlen, N., Flint, J. & Kendler, K. S. (2018) Molecular Genetic Analysis Subdivided by Adversity Exposure Suggests Etiologic Heterogeneity in Major Depression. *Am J Psychiatry*, 175(6), 545-554.

Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I. & Penninx, B. W. J. H. (2014) Effect of polygenic risk scores on depression in childhood trauma. *British Journal of Psychiatry*.

Peyrot, W. J., Van der Auwera, S., Milaneschi, Y., Dolan, C. V., Madden, P. A. F., Sullivan, P. F., Strohmaier, J., Ripke, S., Rietschel, M., Nivard, M. G., Mullins, N., Montgomery, G. W., Henders, A. K., Heat, A. C., Fisher, H. L., Dunn, E. C., Byrne, E. M., Air, T. A., Wray, N. R., Mattheisen, M., Trzaskowski, M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. F. M., Bacanu, S. A., Bækvad-Hansen, M., Beekman, A. T. F., Bigdeli, T. B., Binder, E. B., Blackwood, D. H. R., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J. H., Clarke, T. K., Coleman, J. R. I., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G. E., Davies, G., Deary, I. J., Degenhardt, F., Derks, E. M., Direk, N., Eley, T. C., Escott-Price, V., Farnush, Kiadeh, F. H., Finucane, H. K., Forstner, A. J., Frank, J., Gaspar, H. A., Gill, M., Goes, F. S., Gordon, S. D., Grove, J., Hall, L. S., Hansen, C. S., Hansen, T. F., Herms, S., Hicki, I. B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J. J., Hougaard, D. M., Ising, M., Jansen, R., Jorgenson, E., Kohane, I. S., Kraft, J., Kretzschmar, W. W., Krogh, J., Kutalik, Z., Li, Y., Lind, P. A., MacIntyre, D. J., MacKinnon, D. F., Maier, R. M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S. E., Mehta, D., Middeldorp, C. M., Mihailov, E., Milani, L., Mondimore, F. M., Mostafavi, S., Nauck, M., Ng, B., Nyholt, D. R., O'Reilly, P. F., et al (2018) Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. *Biological Psychiatry*.

Philibert, R. (2006) Merging genetic and environmental effects in the Iowa Adoption Studies: focus on depression. *Ann Clin Psychiatry*, 18(4), 219-22.

Pingault, J. B., Rijsdijk, F., Schoeler, T., Choi, S. W., Selzam, S., Krapohl, E., O'Reilly, P. F. & Dudbridge, F. (2021) Genetic sensitivity analysis: Adjusting for genetic confounding in epidemiological associations. *PLoS Genet*, 17(6), e1009590.

Pingault, J. B., Rijsdijk, F., Schoeler, T., Choi, S. W., Selzam, S., Krapohl, E., O'Reilly, P. F. & Dudbridge, F. (2019) Estimating the sensitivity of associations between risk factors and outcomes to shared genetic effects.

Plomin, R. (2013) Commentary: missing heritability, polygenic scores, and gene-environment correlation. *J Child Psychol Psychiatry*, 54(10), 1147-9.

Plomin, R. (2014) Genotype-Environment Correlation in the Era of DNA. *Behavior Genetics*, 44(6), 629-638.

Polderman, T. J., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M. & Posthuma, D. (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*, 47(7), 702-9.

Posthuma, D., Beem, A. L., de Geus, E. J., van Baal, G. C., von Hjelmborg, J. B., Iachine, I. & Boomsma, D. I. (2003) Theory and practice in quantitative genetics. *Twin Res*, 6(5), 361-76.

Power, R. A., Wingenbach, T., Cohen-Woods, S., Uher, R., Ng, M. Y., Butler, A. W., Ising, M., Craddock, N., Owen, M. J., Korszun, A., Jones, L., Jones, I., Gill, M., Rice, J. P., Maier, W., Zobel, A., Mors, O., Placentino, A., Rietschel, M., Lucae, S., Holsboer, F., Binder, E. B., Keers, R., Tozzi, F., Muglia, P., Breen, G., Craig, I. W., Muller-Myhsok, B., Kennedy, J. L., Strauss, J., Vincent, J. B., Lewis, C. M., Farmer, A. E. & McGuffin, P. (2013) Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychological Medicine*, 43(9), 1965-1971.

Price, M., Legrand, A. C., Brier, Z. M. F. & Hebert-Dufresne, L. (2019) The symptoms at the center: Examining the comorbidity of posttraumatic stress disorder, generalized anxiety disorder, and depression with network analysis. *Journal of Psychiatric Research*, 109, 52-58.

Price, T. S. & Jaffee, S. R. (2008) Effects of the Family Environment: Gene-Environment Interaction and Passive Gene-Environment Correlation. *Developmental Psychology*.

Realo, A., van der Most, P. J., Allik, J., Esko, T., Jeronimus, B. F., Koots-Ausmees, L., Mottus, R., Tropf, F. C., Snieder, H. & Ormel, J. (2017) SNP-Based Heritability Estimates of Common and Specific Variance in Self- and Informant-Reported Neuroticism Scales. *J Pers*, 85(6), 906-919.

Richmond-Rakerd, L. S., Trull, T. J., Gizer, I. R., McLaughlin, K., Scheiderer, E. M., Nelson, E. C., Agrawal, A., Lynskey, M. T., Madden, P. A. F., Heath, A. C., Statham, D. J. & Martin, N. G. (2019) Common genetic contributions to high-risk trauma exposure and self-injurious thoughts and behaviors. *Psychological Medicine*, 49(3), 421-430.

Rijsdijk, F. V., Snieder, H., Ormel, J., Sham, P., Goldberg, D. P. & Spector, T. D. (2003) Genetic and environmental influences on psychological distress in the population: General Health Questionnaire analyses in UK twins. *Psychol Med*, 33(5), 793-801.

Rosseel, Y. (2012) Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*.

Rosseel, Y. (2018) The lavaan tutorial. *Department of Data Analysis Ghent University (Belgium)*.

Røysamb, E. & Tambs, K. (2016) The beauty, logic and limitations of twin studies.

Rutter, M., Moffitt, T. E. & Caspi, A. (2006) Gene-environment interplay and psychopathology: Multiple varieties but real effects.

Sartor, C. E., Grant, J. D., Lynskey, M. T., McCutcheon, V. V., Waldron, M., Statham, D. J., Bucholz, K. K., Madden, P. A., Heath, A. C., Martin, N. G. & Nelson, E. C. (2012) Common heritable contributions to low-risk trauma, high-risk trauma, posttraumatic stress disorder, and major depression. *Arch Gen Psychiatry*, 69(3), 293-9.

Sartorius, N. (2017) Comorbidity of depression and diabetes: Questions recently answered and raised. *European Psychiatry*, 41, S62-S62.

Saveanu, R. V. & Nemeroff, C. B. (2012) Etiology of depression: genetic and environmental factors. *Psychiatr Clin North Am*, 35(1), 51-71.

Schneider, G., Kohnke, C., Teismann, H. & Berger, K. (2021) Childhood trauma and personality explain more variance in depression scores than sociodemographic and lifestyle factors-Results from the BiDirect Study. *Journal of Psychosomatic Research*, 147.

Schwabe, I., Janss, L. & van den Berg, S. M. (2017) Can we validate the results of twin studies? A census-based study on the heritability of educational achievement. *Frontiers in Genetics*.

Shao, N., Gong, Y., Wang, X., Wei, J., Shi, J., Ding, H., Zhang, M., Kang, C., Wang, S., Chen, L., Yu, Y. & Han, J. (2021) Effects of polygenic risk score, childhood trauma and resilience on depressive symptoms in Chinese adolescents in a three-year cohort study. *Journal of Affective Disorders*, 282, 627-636.

Shea, A., Walsh, C., Macmillan, H. & Steiner, M. (2005) Child maltreatment and HPA axis dysregulation: relationship to major depressive disorder and post traumatic stress disorder in females. *Psychoneuroendocrinology*, 30(2), 162-78.

Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D. I., Cornes, B. K., Davis, C., Dunkel, L., De Lange, M., Harris, J. R., Hjelmborg, J. V. B., Luciano, M., Martin, N. G., Mortensen, J., Nisticò, L., Pedersen, N. L., Skytthe, A., Spector, T. D., Stazi, M. A., Willemsen, G. & Kaprio, J. (2003) Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries.

Slavich, G. M. & Irwin, M. R. (2014) From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol Bull*, 140(3), 774-815.

Smigielski, E. M., Sirotkin, K., Ward, M. & Sherry, S. T. (2000) dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res*, 28(1), 352-5.

Smith, B. H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S. M., Deary, I. J., MacIntyre, D. J., Campbell, H., McGilchrist, M., Hocking, L. J., Wisely, L., Ford, I., Lindsay, R. S., Morton, R., Palmer, C. N. A., Dominiczak, A. F., Porteous, D. J. & Morris, A. D. (2013) Cohort profile: Generation scotland: Scottish family health study (GS: SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology*.

Stein, M. B., Jang, K. L., Taylor, S., Vernon, P. A. & Livesley, W. J. (2002) Genetic and environmental influences on trauma exposure and posttraumatic stress disorder symptoms: a twin study. *Am J Psychiatry*, 159(10), 1675-81.

Stephan, Y., Sutin, A. R., Luchetti, M., Caille, P. & Terracciano, A. (2020) An examination of potential mediators of the relationship between polygenic scores of BMI and waist circumference and phenotypic adiposity.

Sterling, M. (2011) General Health Questionnaire-28 (GHQ-28) Article in. *Journal of Physiotherapy*.

Streiner, D. L. (2005) Finding our way: An introduction to path analysis.

Stulp, G., Simons, M. J. P., Grasman, S. & Pollet, T. V. (2017) Assortative mating for human height: A meta-analysis. *American Journal of Human Biology*.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T. & Collins, R. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, 12(3), e1001779.

Sullivan, P. F., Daly, M. J., Ripke, S., Lewis, C. M., Lin, D. Y., Wray, N. R., Neale, B., Levinson, D. F., Breen, G., Byrne, E. M., Wray, N. R., Levinson, D. F., Rietschel, M., Hoogendijk, W., Ripke, S., Sullivan, P. F., Hamilton, S. P., Levinson, D. F., Lewis, C. M., Ripke, S., Weissman, M. M., Wray, N. R., Breuer, R., Cichon, S., Degenhardt, F., Frank, J., Gross, M., Herms, S., Hoefels, S., Maier, W., Mattheisen, M., Noethen, M. M., Rietschel, M., Schulze, T. G., Steffens, M., Treutlein, J., Boomsma, D. I., De Geus, E. J., Hoogendijk, W., Hottenga, J. J., Jung-Ying, T., Lin, D. Y., Middeldorp, C. M., Nolen, W. A., Penninx, B. P., Smit, J. H., Sullivan, P. F., van Grootheest, G., Willemsen, G., Zitman, F. G., Coryell, W. H., Knowles, J. A., Lawson, W. B., Levinson, D. F., Potash, J. B., Scheftner, W. A., Shi, J. X., Weissman, M. M., Holsboer, F., Muglia, P., Tozzi, F., Blackwood, D. H. R., Boomsma, D. I., De Geus, E. J., Hottenga, J. J., MacIntyre, D. J., McIntosh, A., McLean, A., Middeldorp, C. M., Nolen, W. A., Penninx, B. P., Ripke, S., Smit, J. H., Sullivan, P. F., van Grootheest, G., Willemsen, G., Zitman, F. G., van den Oord, E. J. C. G., Holsboer, F., Lucae, S., Binder, E., Muller-Myhsok, B., Ripke, S., Czamara, D., Kohli, M. A., Ising, M., Uhr, M., Bettecken, T., Barnes, M. R., Breen, G., Craig, I. W., Farmer, A. E., Lewis, C. M., McGuffin, P., Muglia, P., Byrne, E., Gordon, S. D., Heath, A. C., Henders, A. K., Hickie, I. B., et al (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, 18(4), 497-511.

Sullivan, P. F., Neale, M. C. & Kendler, K. S. (2000) Genetic epidemiology of major depression: Review and meta-analysis.

Tabor, H. K., Risch, N. J. & Myers, R. M. (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*, 3(5), 391-A396.

Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G. & Meyre, D. (2019) Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20(8), 467-484.

Taylor, P. J. (2007) The unreliability of high human heritability estimates and small shared effects of growing up in the same family. *Biological Theory*, 2(4), 387-397.

Team, R. C. (2020) R: A language and environment for statistical computing.

Thornicroft, G., Chatterji, S., Evans-Lacko, S., Gruber, M., Sampson, N., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Andrade, L., Borges, G., Bruffaerts, R., Bunting, B., de Almeida, J. M., Florescu, S., de Girolamo, G., Gureje, O., Haro, J. M., He, Y., Hinkov, H., Karam, E., Kawakami, N., Lee, S., Navarro-Mateu, F., Piazza, M., Posada-Villa, J., de Galvis, Y. T. & Kessler, R. C. (2017) Undertreatment of people with major depressive disorder in 21 countries. *Br J Psychiatry*, 210(2), 119-124.

Thorp, J. G., Marees, A. T., Ong, J. S., An, J. Y., MacGregor, S. & Derks, E. M. (2020) Genetic heterogeneity in self-reported depressive symptoms identified through genetic analyses of the PHQ-9. *Psychological Medicine*, 50(14), 2385-2396.

Torkamani, A., Wineinger, N. E. & Topol, E. J. (2018) The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*, 19(9), 581-590.

Tunnard, C., Rane, L. J., Wooderson, S. C., Markopoulou, K., Poon, L., Fekadu, A., Juruena, M. & Cleare, A. J. (2014) The impact of childhood adversity on suicidality and clinical course in treatment-resistant depression. *J Affect Disord*, 152-154, 122-30.

Uher, R. (2014) Gene-environment interactions in common mental disorders: An update and strategy for a genome-wide search.

Uher, R. & Zwicker, A. (2017) Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry*, 16(2), 121-129.

VanRaden, P. M. (2008) Efficient methods to compute genomic predictions. *J Dairy Sci*, 91(11), 4414-23.

Verweij, K. J., Mosing, M. A., Zietsch, B. P. & Medland, S. E. (2012) Estimating heritability from twin studies. *Methods Mol Biol*, 850, 151-70.

Vibhakar, V., Allen, L. R., Gee, B. & Meiser-Stedman, R. (2019) A systematic review and meta-analysis on the prevalence of depression in children and adolescents after exposure to trauma. *Journal of Affective Disorders*, 255, 77-89.

Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. (2012) Five years of GWAS discovery. *Am J Hum Genet*, 90(1), 7-24.

Visscher, P. M., Hill, W. G. & Wray, N. R. (2008) Heritability in the genomics era - concepts and misconceptions. *Nature Reviews Genetics*, 9(4), 255-266.

Visscher, P. M. & Wray, N. R. (2015) Concepts and Misconceptions about the Polygenic Additive Model Applied to Disease. *Human Heredity*, 80(4), 165-170.

Vos, T. (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019 (vol 396, pg 1204, 2020). *Lancet*, 396(10262), 1562-1562.

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., Carter, A., Casey, D. C., Charlson, F. J., Chen, A. Z., Coggeshall, M., Cornaby, L., Dandona, L., Dicker, D. J., Dilegge, T., Erskine, H. E., Ferrari, A. J., Fitzmaurice, C., Fleming, T., Forouzanfar, M. H., Fullman, N., Gething, P. W., Goldberg, E. M., Graetz, N., Haagsma, J. A., Johnson, C. O., Kassebaum, N. J., Kawashima, T., Kemmer, L., Khalil, I. A., Kinfu, Y., Kyu, H. H., Leung, J. N., Liang, X. F., Lim, S. S., Lopez, A. D., Lozano, R., Marczak, L., Mensah, G. A., Mokdad, A. H., Naghavi, M., Nguyen, G., Nsoesie, E., Olsen, H., Pigott, D. M., Pinho, C., Rankin, Z., Reinig, N., Salomon, J. A., Sandar, L., Smith, A., Stanaway, J., Steiner, C., Teeple, S., Thomas, B. A., Troeger, C., Wagner, J. A., Wang, H. D., Wanga, V., Whiteford, H. A., Zoeckler, L., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., Abraham, B., Abubakar, I., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Ackerman, I. N., Adebiyi, A. O., Ademi, Z., Adou, A. K., Afanvi, K. A., Agardh, E. E., Agarwal, A., Kiadaliri, A. A., Ahmadieh, H., Ajala, O. N., Akinyemi, R. O., Akseer, N., Al-Aly, Z., Alam, K., Alam, N. K. M., Aldhahri, S. F., Alegretti, M. A., Alemu, Z. A., Alexander, L. T., Alhabib, S., Ali, R., Alkerwi, A., Alla, F., Allebeck, P., Al-Raddadi, R., Alsharif, U., Altirkawi, K. A., Alvis-Guzman, N., Amare, A. T., Amberbir, A., et al (2016) Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*, 388(10053), 1545-1602.

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., Aboyans, V., Abraham, J., Ackerman, I., Aggarwal, R., Ahn, S. Y., Ali, M. K., Almazroa, M. A., Alvarado, M., Anderson, H. R., Anderson, L. M., Andrews, K. G.,

Atkinson, C., Baddour, L. M., Bahalim, A. N., Barker-Collo, S., Barrero, L. H., Bartels, D. H., Basáñez, M. G., Baxter, A., Bell, M. L., Benjamin, E. J., Bennett, D., Bernabé, E., Bhalla, K., Bhandari, B., Bikbov, B., Abdulhak, A. B., Birbeck, G., Black, J. A., Blencowe, H., Blore, J. D., Blyth, F., Bolliger, I., Bonaventure, A., Boufous, S., Bourne, R., Boussinesq, M., Braithwaite, T., Brayne, C., Bridgett, L., Brooker, S., Brooks, P., Brugha, T. S., Bryan-Hancock, C., Bucello, C., Buchbinder, R., Buckle, G., Budke, C. M., Burch, M., Burney, P., Burstein, R., Calabria, B., Campbell, B., Canter, C. E., Carabin, H., Carapetis, J., Carmona, L., Cella, C., Charlson, F., Chen, H., Cheng, A. T. A., Chou, D., Chugh, S. S., Coffeng, L. E., Colan, S. D., Colquhoun, S., Colson, K. E., Condon, J., Connor, M. D., Cooper, L. T., Corriere, M., Cortinovis, M., De Vaccaro, K. C., Couser, W., Cowie, B. C., Criqui, M. H., Cross, M., Dabhadkar, K. C., Dahiya, M., Dahodwala, N., Damsere-Derry, J., Danaei, G., Davis, A., De Leo, D., Degenhardt, L., Dellavalle, R., Delossantos, A., Denenberg, J., Derrett, S., Des Jarlais, D. C., Dharmaratne, S. D., et al (2012) Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*.

Vukasovic, T. & Bratko, D. (2015) Heritability of personality: A meta-analysis of behavior genetic studies. *Psychol Bull*, 141(4), 769-85.

Vythilingam, M., Heim, C., Newport, J., Miller, A. H., Anderson, E., Bronen, R., Brummer, M., Staib, L., Vermetten, E., Charney, D. S., Nemeroff, C. B. & Bremner, J. D. (2002) Childhood trauma associated with smaller hippocampal volume in women with major depression. *Am J Psychiatry*, 159(12), 2072-80.

Wamser-Nanney, R. & Cherry, K. E. (2018) Children's trauma-related symptoms following complex trauma exposure: Evidence of gender differences. *Child Abuse Negl*, 77, 188-197.

Wang, Y., Franks, J. M., Whitfield, M. L. & Cheng, C. (2019) BioMethyl: an R package for biological interpretation of DNA methylation data. *Bioinformatics*, 35(19), 3635-3641.

Warrier, V. & Baron-Cohen, S. (2019) Childhood trauma, life-time self-harm, and suicidal behaviour and ideation are associated with polygenic scores for autism. *Molecular Psychiatry*.

Warrier, V., Kwong, A. S. F., Luo, M., Dalvie, S., Croft, J., Sallis, H. M., Baldwin, J., Munafo, M. R., Nievergelt, C. M., Grant, A. J., Burgess, S., Moore, T. M., Barzilay, R., McIntosh, A., van, I. M. H. & Cecil, C. A. M. (2021) Gene-environment correlations and causal effects of childhood maltreatment on physical and mental health: a genetically informed approach. *Lancet Psychiatry*, 8(5), 373-386.

Weinberger, A. H., Gbedemah, M., Martinez, A. M., Nash, D., Galea, S. & Goodwin, R. D. (2018) Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychol Med*, 48(8), 1308-1315.

Weissbrod, O., Flint, J. & Rosset, S. (2018) Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics. *Am J Hum Genet*, 103(1), 89-99.

Weissman, M. M. (2020) Is depression nature or nurture? *American Journal of Psychiatry*, 177(5), 376-377.

Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C. A. (2021) Correction to: Genome-wide gene-environment interactions in neuroticism: an exploratory study across 25 environments. *Transl Psychiatry*, 11(1), 207.

Wickersham, A., Sugg, H. V. R., Epstein, S., Stewart, R., Ford, T. & Downs, J. (2021) Systematic Review and Meta-analysis: The Association Between Child and Adolescent Depression and Later Educational Attainment. *Journal of the American Academy of Child and Adolescent Psychiatry*, 60(1), 105-118.

Wilkinson, P. O., Trzaskowski, M., Haworth, C. M. A. & Eley, T. C. (2013) The role of gene-environment correlations and interactions in middle childhood depressive symptoms. *Development and Psychopathology*, 25(1), 93-104.

Willham, R. L. (1963) The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics*

pp.18-27.

World Health, O. (2017) Depression and other common mental disorders: global health estimates. *World Health Organization*.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F. & Middeldorp, C. M. (2014) Research Review: Polygenic methods and their application to psychiatric traits.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S. A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J. H., Clarke, T. K., Coleman, J. I. R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G. E., Crowley, C. A., Dashti, H. S., Davies, G., Deary, I. J., Degenhardt, F., Derks, E. M., Direk, N., Dolan, C. V., Dunn, E. C., Eley, T. C., Eriksson, N., Escott-Price, V., Kiadeh, F. H. F., Finucane, H. K., Forstner, A. J., Frank, J., Gaspar, H. A., Gill, M., Giusti-Rodríguez, P., Goes, F. S., Gordon, S. D., Grove, J., Hall, L. S., Hannon, E., Hansen, C. S., Hansen, T. F., Herms, S., Hickie, I. B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J. J., Hougaard, D. M., Hu, M., Hyde, C. L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J. A., Kohane, I. S., Kraft, J., Kretzschmar, W. W., Krogh, J., Kutalik, Z., Lane, J. M., Li, Y., Li, Y., Lind, P. A., Liu, X., Lu, L., MacIntyre, D. J., MacKinnon, D. F., Maier, R. M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S. E., Mehta, D., Middeldorp, C. M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F. M., Montgomery, G. W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., et al (2018a) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*.

Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. (2018b) Common disease is more complex than implied by the core gene omnigenic model. *Cell*, 173(7), 1573-1580.

Wright, S. (1934a) An Analysis of Variability in Number of Digits in an Inbred Strain of Guinea Pigs. *Genetics*, 19(6), 506-36.

Wright, S. (1934b) The results of crosses between inbred strains of guinea pigs, differing in number of digits. *Genetics*, 19(6), 0537-0551.

Xia, L. & Yao, S. (2015) The Involvement of Genes in Adolescent Depression: A Systematic Review. *Front Behav Neurosci*, 9, 329.

Xu, J., Wang, J., Wimo, A. & Qiu, C. (2016) The economic burden of mental disorders in China, 2005-2013: implications for health policy. *BMC Psychiatry*, 16, 137.

Yanes, T., McInerney-Leo, A. M., Law, M. H. & Cummings, S. (2020) The emerging field of polygenic risk scores and perspective for use in clinical care. *Human Molecular Genetics*, 29(R2), R165-R176.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A., Lee, S. H., Robinson, M. R., Perry, J. R., Nolte, I. M., van Vliet-Ostaptchouk, J. V., Snieder, H., LifeLines Cohort, S., Esko, T., Milani, L., Magi, R., Metspalu, A., Hamsten, A., Magnusson, P. K., Pedersen, N. L., Ingelsson, E., Soranzo, N., Keller, M. C., Wray, N. R., Goddard, M. E. & Visscher, P. M. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*, 47(10), 1114-20.

Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. (2011) GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1), 76-82.

Yang, M. Y., Font, S. A., Ketchum, M. & Kim, Y. K. (2018) Intergenerational transmission of child abuse and neglect: Effects of maltreatment type and depressive symptoms. *Children and Youth Services Review*, 91, 364-371.

Yanowitz, J. (2010) Meiosis: Making a break for it.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J. & Visscher, P. M. (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Human Molecular Genetics*.

Young, A. I. (2019) Solving the missing heritability problem. *PLoS Genet*, 15(6), e1008222.

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., Masson, G., Thorsteinsdottir, U., Stefansson, K. & Kong, A. (2018) Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics*.

Yu, M. C., Linn, K. A., Shinohara, R. T., Oathes, D. J., Cook, P. A., Duprat, R., Moore, T. M., Oquendo, M. A., Phillips, M. L., McInnis, M., Fava, M., Trivedi, M. H., McGrath, P., Parsey, R., Weissman, M. M. & Sheline, Y. I. (2019) Childhood trauma history is linked to abnormal brain connectivity in major depression (vol 116, pg 8582, 2019). *Proceedings of the National Academy of Sciences of the United States of America*, 116(26), 13146-13146.

Zhang, F., Chen, W., Zhu, Z., Zhang, Q., Nabais, M. F., Qi, T., Deary, I. J., Wray, N. R., Visscher, P. M., McRae, A. F. & Yang, J. (2019) OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol*, 20(1), 107.

Zhu, M. J. & Zhao, S. H. (2007) Candidate gene identification approach: Progress and challenges. *International Journal of Biological Sciences*, 3(7), 420-427.