WIDE METHODS APPLIED TO STUDY GENE-ENVIRONMENT INTERPLAY

Coleman, Jonathan Richard Iain

Awarding institution:
King's College London

GENOME-WIDE METHODS APPLIED TO STUDY GENE-ENVIRONMENT INTERPLAY

Jonathan Richard Iain Coleman

Thesis submission for the degree of

Doctor of Philosophy

Student Number 1223375

Supervised by

Professor Thalia Eley and Dr Gerome Breen

Abstract

Information-rich genomic data have transformed the study of genetic variants but have affected investigations of gene-environment interplay less, partly due to the multiple testing involved in genome-wide interaction studies. This thesis explores alternative uses of genome-wide techniques to investigate gene-environment interplay.

Genetic associations with individual differences in response to an environment can be examined by performing genome-wide association studies in individuals with a shared exposure. Cognitive behavioural therapy is a controllable environment that can be studied prospectively. Genetic variants and RNA transcript expression were used to predict therapeutic outcome. No significant predictors were identified, suggesting that effects are likely to be small.

Genome-wide association studies remain underpowered to detect small effects, despite increasingly large cohorts. Polygenic risk scores incorporating variants below traditional thresholds of statistical significance can capture true signal. These scores can act as a proxy for the effect of the genome in genome-by-environment interaction studies, and were used in this thesis to dissect the observed increase in body mass index in individuals with depression. Results suggest that this relationship is likely to result

primarily from causes other than the additive effects of common genetic variation.

Polygenic risk scores were also used to assess the effects of social environmental and genetic influences on body mass index before and during adolescence, using a risk score primarily derived from adult participants. Positive associations between this risk score and adolescent body mass index phenotypes suggest a stable genetic influence on body mass. Social environmental influences on body mass had small effects, with weak evidence for an interaction between socioeconomic status and genetic risk influencing body mass.

Statistical limitations on genomic analyses can be reduced by using alternative methods to complement genome-wide interaction studies. These approaches provide insight into the interactive effects of the genome and the environment on behavioural phenotypes.

Contents

Table of Figures

Acknowledgements

I am indebted to my supervisors Prof Thalia Eley and Dr Gerome Breen, for their continual guidance and insight throughout the course of my PhD. I am also grateful to the many people within the Social, Genetic and Developmental Psychiatry Centre who have made the last four years a wonderful experience, particularly the members of the EDiT lab and BRC Genomics and Bioinformatics Cores for their patience, assistance and cakes.

Chapter 5 reports research conducted using the UK Biobank Resource, and I acknowledge the considerable effort involved in producing this resource, and the invaluable contribution of the participants. Chapter 6 reports research conducted using the Twins Early Development Study, and I also acknowledge the ongoing contribution of the participants in TEDS and their families, as well as the work of the TEDS team, particularly Eva Krapohl, Andrew McMillan and Prof Robert Plomin.

This thesis represents independent research part supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author and co-authors, and not necessarily those of the NHS, the NIHR or the Department of Health.

I thank the Institute of Psychiatry, Psychology and Neuroscience, and the Alexander von Humboldt Foundation (Germany) for jointly funding my PhD.

Statement of work

The machine learning analyses included in Chapter 4 were the work of Dr Héléna Gaspar. The replication of results in Generation Scotland included in Chapter 5 and Appendix III was performed by Dr Toni-Kim Clarke. All other work presented in this thesis is the work of the author, either alone or in collaboration with others as detailed below.

Work presented in Chapters 3 and 4 was undertaken as part of the Genes for Treatment (Child) and Genes for Treatment (Adult) projects investigating biological predictors of response to cognitive behavioural therapy. DNA genotyping and RNA expression measurements were performed by the BRC Genomics Core at the IoPPN, with contributions from the author and Susanna Roberts. Quality control and analyses were performed by the author. The phenotypes investigated in Chapter 3 were derived (by the author) from work by Dr Robert Keers as part of the wider project.

Analyses presented in Chapter 5 were performed by the author using a depression phenotype derived from the UK Biobank data in collaboration with Dr Mark Adams and others at the University of Edinburgh as part of a broader project on the genetics of depression.

Chapter 6 used quality-controlled and imputed genotype data from the TEDS project, prepared by Eva Krapohl.

Chapter 1: Introduction

1.1 Quantitative genetics, molecular genetics, and the relative roles of genetics and environment

1.1.1. Quantitative genetics describes genetic and environmental influences on human behaviour as a complex trait

The desire to understand the aetiology of behaviour has been a major theme within the field of genetics since its inception. A vital contribution to this endeavour was the creation of quantitative genetics, which seeks to understand the relative contributions of genetic and environmental influences to trait variance. The development of quantitative genetic theory reconciled the particulate genetic inheritance suggested by the work of Mendel with the more complex patterns of inheritance observed by biometricians such as Galton (Galton, 1869; Mendel, 1866). Quantitative genetics proposes that apparently complex traits result from the combinatorial effects of many factors, each of which obeys Mendelian laws of inheritance. These factors combine to generate a distribution of liability that underlies variation in complex traits (Fisher, 1918; Plomin, DeFries, Knopik, *et al*, 2012; Sarkar & Pfeifer, 2006; Wright, 1921). Observed within-family similarities in complex traits can be explained by the combination of multiple genetic variants with additive effects in the presence of additional factors (Fisher, 1918). Such non-genetic factors are usually referred to as

environmental influences, although there is also a contribution from non-additive genetic effects and from the random stochastic error that inaccurate measurement and observation necessarily incur. As such, the word "environment" can refer both to this "inferred component" and to specific effects stemming from the environment to which individuals are exposed ("measured environments"). This latter concept is predominantly used within this thesis – where the intended meaning is not obvious from context, it is explicitly stated.

The quantitative genetic model assumes the underlying influences on complex traits sum to a continuous distribution of liability, regardless of the presentation of the trait. In the case of discrete phenotypes (such as mental illnesses), a liability-threshold model can explain the apparent dichotomy, with the disorder manifesting above a certain amount of risk (the threshold; Falconer, 1965). This threshold could be defined by multiple factors, both genetic and environmental. If liability is modelled as a normal distribution, its variance can be partitioned among constituent factors (Figure 1; Plomin, DeFries, Knopik, *et al*, 2012). At the simplest level, two main components can be defined: the genetic component (also referred to as broad-sense heritability) and the environmental component. Again, it is important to note that environment in this context is defined negatively as all of the variance in

the phenotype not explained by genetic factors, rather than the summed effects of multiple environmental influences.

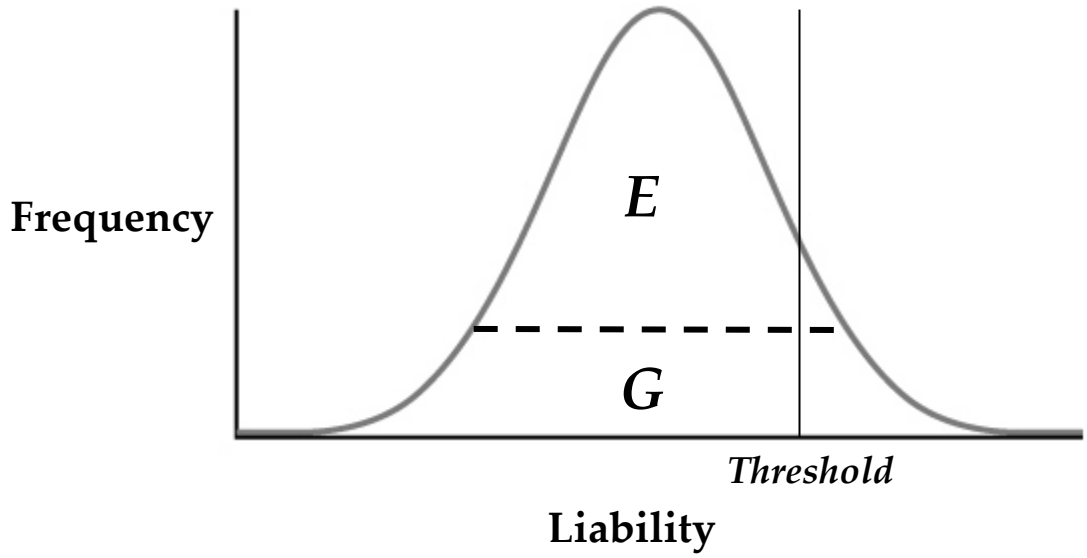


Figure 1: Liability-threshold model for a discrete trait, showing the normal distribution of risk and the threshold, to the right of which the trait presents. Imagined contributions of genetic (below) and environmental (above) influences are also shown.

Related individuals shared genetic variants, and the degree of sharing (relatedness) can be determined from genetic theory. Relatedness is important to quantitative genetic methods because it allows the genetic and environmental components of variance to be inferred from the observed phenotypic variance. An individual is expected to share half of their genome with each of their parents, and with any full siblings, due to the independent assortment of chromosomes during gamete formation (Alberts, 2002;

Mendel, 1866). Identical twins, however, develop from a single zygote, and as such share all of their genome (ignoring rare early mutational events).

This sharing is crucial to the design of one of the most prominent quantitative genetic methods, the classical twin study (reviewed in Boomsma, Busjahn & Peltonen, 2002). The correlation between the phenotypes of pairs of identical twins (which can be observed empirically) is the sum of genetic and shared environmental components (that is, factors that make the twins more alike). If any pairs have discordant phenotypes, the influencing factor must stem from the environmental component (because their genetic component is shared). Non-identical twins share only half of their genome (on average), so the genetic contribution to the correlation between pairs of non-identical twins is expected to be proportionally lower. From these observations and theories, the environmental and genetic components of variance can be inferred (Neale & Cardon, 1992; Plomin, DeFries, Knopik, *et al*, 2012).

1.1.2 Molecular genetics enables the dissection of the heritable component

Although quantitative genetic approaches can demonstrate that genetic effects influence behavioural traits, they cannot be used to determine which exact areas of the genome contribute. That is the preserve of molecular genetic methods. Initially, such methods relied either on the linkage of broad regions of the genome with outcomes of interest (linkage studies) or on

targeted examination of the effects of specific mutations (association studies). However, both of these approaches have considerable limitations. Linkage studies use variable number tandem repeats (repeated sequences of a few nucleotides), which are limited in number across the genome, resulting in low resolution and restricting linkage of traits to sections of chromosomes only, rather than to specific nucleotides (Botstein, White, Skolnick, *et al*, 1980). Association studies can examine the genome at a much finer scale by assessing the effects of single nucleotide polymorphisms (SNPs). Initially, however, they were limited by the number of SNPs identified and by the technological capacity to assay them. As a result, early association studies focussed on candidate genes. However, this approach relies on potentially erroneous assumptions about which genes are relevant to a given trait, and such studies often only assess a few variants out of the many that may influence the action of the gene (Dick, Agrawal, Keller, *et al*, 2015; Munafo, 2006). In the absence of likely effect size estimates, the cohorts examined were usually too small (in the context of effect size estimates from later genomic studies; Dick, Agrawal, Keller, *et al*, 2015). As a result, many of the findings from candidate gene studies have failed to replicate in more recent, genomic studies, and reported effect sizes have been over-estimated, a phenomenon known as winner's curse (Chabris, Hebert, Benjamin, *et al*, 2012; Dunn, Brown, Dai, *et al*, 2015; Hirschhorn, Lohmueller, Byrne, *et al*, 2002; Ioannidis, 2005; Ioannidis, 2014; Zollner & Pritchard, 2007).

The use of molecular genetics in the study of behaviour was altered considerably by the technological and statistical innovation that occurred during the Human Genome Project and following its completion (Lander, Linton, Birren, *et al*, 2001; McCarthy, Das, Kretzschmar, *et al*, 2016; Venter, Adams, Myers, *et al*, 2001). The consequent development of microarray-based genome-wide genotyping and a more sophisticated knowledge of the multiple testing load made genome-wide association studies (GWAS) possible (Klein, Zeiss, Chew, *et al*, 2005; Wellcome Trust Case Control, 2007). GWAS combine the agnosticism of linkage studies with the precision of single-variant association studies. Large sample sizes are required to identify associated loci; consequently, psychiatric genomics has been driven almost from the outset by large-scale collaborations (Sullivan, 2010).

1.1.3 The "four laws of behavior genetics" describe the relative importance of genetic and environmental components

The accumulated evidence from quantitative and molecular studies in human behavioural genetics was pithily summarised as the three "laws of behavior genetics", with a fourth law recently added. These laws are:

- Behaviour is heritable (Turkheimer, 2000; Turkheimer & Gottesman, 1991)
- Shared environments are less important than shared genetics (Turkheimer, 2000)

- Non-shared environments contribute considerably to phenotypic variance (Turkheimer, 2000).
- Variance in behavioural traits is associated with many genetic variants with very small individual effects (Chabris, Lee, Cesarini, *et al*, 2015).

The success of behavioural genomics inspired the suggestion of the fourth law; although many behavioural traits await genomic study, the evidence from those studied to date argues that polygenicity is expected. The polygenic model (incorporating variation across the allele frequency spectrum) explains findings to date better than competing models, such as a high excess of common or of rare mutations (Gratten, Wray, Keller, *et al*, 2014; Sullivan, Daly & O'Donovan, 2012; Visscher, Goddard, Derks, *et al*, 2012; Yang, Visscher & Wray, 2010). Unlike the more seriously proposed fourth law, the original laws were deliberately simplistic and designed to highlight the probabilistic nature of genetic results, and the inherent biases of the methods used (Turkheimer, 2000). Nevertheless, the laws are an accurate summary of the field. Meta-analysis of twin studies conducted between 1958 and 2012 report an average effect of shared environmental component of 17% across all mental and behavioural disorders (819 traits, 1599455 twin pairs), compared to an average heritability of 47% and therefore an average effect of the non-shared environmental component of

36% (Polderman, Benyamin, de Leeuw, *et al*, 2015). The effect of non-shared environment appears to be substantial. However, this component contains stochastic error, and as such its importance could reflect the imperfection of the model (Plomin, 1994; Plomin & Daniels, 2010; Turkheimer, 2000; Turkheimer & Waldron, 2000).

1.2 Genetic and environmental influences on the disorders of interest

Psychiatric genomics has begun to yield valuable results, despite the initial disappointment that resulted from the unexpectedly small effect sizes of individual variants associated with psychiatric phenotypes. Schizophrenia has been the pioneer disorder, with the most recent large international meta-analysis reporting associations at 108 genetic loci (Schizophrenia Working Group of the Psychiatric Genomics, 2014). Theoretical estimations from this data suggest there are likely to be thousands of associated loci (although estimation is biassed by the assumed genetic architecture). The recent completion of large-scale genotyping projects is expected to boost considerably the power to detect these loci (Aas, Blokland, Chawner, *et al*, 2016; Gratten, Wray, Keller, *et al*, 2014).

1.2.1. Studying the genetics of major depressive disorder is constrained by heterogeneity

Genomics has been slower to produce results in other psychiatric disorders, particularly those with lower heritability. The archetypal example of this is depression; the heritability of depression as estimated from twin studies is approximately 40%, around half that estimated in schizophrenia (Polderman, Benyamin, de Leeuw, *et al*, 2015). Estimates differ by the classification of depression – "depressive episode" has a reported heritability around 30%, compared to an estimate of 45% for the chronic "recurrent depressive disorder" (Polderman, Benyamin, de Leeuw, *et al*, 2015). Furthermore, the average correlation between dizygotic twin pairs for recurrent depressive disorder was less than half that between monozygotic twin pairs, indicating the potential for non-additive genetic effects (Polderman, Benyamin, de Leeuw, *et al*, 2015).

Until recently, genomics had been largely unsuccessful in depression. Despite a comparable cohort size to the successful efforts in schizophrenia (roughly nine thousand cases and nine thousand controls), the initial PGC mega-analysis of depression did not identify any variants at genome-wide levels of significance (Major Depressive Disorder Working Group of the Psychiatric, Ripke, Wray, *et al*, 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011). Although genomic studies have been

relatively unsuccessful in identifying genetic variants associated with depression, there have been hundreds of variants suggested by hypothesis-driven "candidate gene" analyses (Lopez-Leon, Janssens, Gonzalez-Zuloeta Ladd, *et al*, 2008). Variants have been implicated near genes associated with neurotransmission, both transporters (such as *SLC6A3*/*DAT* and *SLC6A4*/*5HTT*) and receptors (including the dopamine receptors *DRD2* and *DRD4*), as well as with signal transduction (like the G-protein *GNB3*), risk processes (including *MTHFR*, involved in folate metabolism), and comorbid disorders (such as *APOE* in Alzheimer's disease; Lopez-Leon, Janssens, Gonzalez-Zuloeta Ladd, *et al*, 2008). However, very few such variants are supported by evidence from GWAS (Dunn, Brown, Dai, *et al*, 2015).

The success of genomics in depression has been limited by both the lower heritability and the phenotypic heterogeneity of the disorder. Although ostensibly a single disorder, depression is frequently split into subtypes according to the recurrence of the disorder or the symptoms with which individuals present. Recurrence is conflated with severity, duration and impairment – recurrent depression is usually more severe, longer lasting and causes a greater reduction in quality of life, including a higher occurrence of suicide attempts (Kessler, Zhao, Blazer, *et al*, 1997; Merikangas, Wicki & Angst, 1994). Depression can present with features including low mood, irritation, loss of pleasure (anhedonia), weight change, disrupted

sleep and activity, fatigue, guilt, loss of concentration and suicidality (Association, 2013). A diagnosis of depression requires as a minimum unusual and impairing low mood or anhedonia, as well as at least four out of the other seven features described above (Association, 2013). As such, a diagnosis of depression can result from 105 different combinations of symptoms. Furthermore, certain of these features, particularly weight and appetite, have no inherent direction in diagnosis. Depression featuring reductions in these areas is termed "typical" because such reductions are conceptually more similar to other depressive symptoms like loss of pleasure and low mood. Conversely, a vegetative state featuring weight gain and reduced activity is characteristic of atypical depression (Davidson, Miller, Turnbull, *et al*, 1982). Latent class factor analysis can distinguish separate classes within depression that can be labelled as typical and atypical, and which vary in severity (Kendler, Eaves, Walters, *et al*, 1996; Sullivan, Kessler & Kendler, 2014). The relationship between these classes and the clinically described subtypes is unclear, however, and such analyses usually also identify intermediate classes that show a mixture of depressive features. Subtypes may be arbitrary groupings of a continuous spectrum, effectively the result of defining multiple thresholds on the underlying liability distribution (Kessler, Zhao, Blazer, *et al*, 1997; Merikangas, Wicki & Angst, 1994). As such, depression demonstrates clinical heterogeneity, and is

composed of grouped sets of features that are distinguishable within studied cohorts of individuals with depression.

A further potential source of heterogeneity in meta-analyses of genetic studies of depression is the influence of measured environmental factors. Predominant among such factors are stressful life events, a broad range of influences that can include the death of loved ones, difficulties in social relationships, injury or illness, and historical or present neglect or abuse (Holmes and Rahe, 1967; Brown & Harris, 1978). The occurrence of such events has been consistently and robustly observed at a greater rate in individuals with depression compared to population controls (Mazure, 1998; Nanni, Uher & Danese, 2012). As such events tend to occur prior to the onset of depression, they are frequently hypothesised to have a causal role. This hypothesis has been tested by contrasting independent events (that is, events out of the individual's control) and dependent events, which the individual influences to some extent (Kendler, Karkowski & Prescott, 1999). Individuals with depression had experienced more independent stressful life events than population controls, suggesting a causal effect. However, the relationship was noticeably stronger when considering dependent events, suggesting that there is also a non-causal component, such as could be produced by a gene-environment correlation in which individuals genetically predisposed to depression are more likely to generate dependent stressful life events, even

prior to depression (Kendler, Karkowski & Prescott, 1999). Such environmental factors may influence not just the onset and course of disease, but the effectiveness of treatment. For example, a meta-analysis of published studies suggests childhood maltreatment is associated with poorer response to pharmacotherapy (with or without psychotherapy) in depression (Nanni, Uher & Danese, 2012).

Beyond stressful life events, there is considerable evidence that the social environment (that is, an individual's interactions with others) many influence depression, particularly in childhood. Studies of expressed emotion (the hostility or warmth towards an individual from others, usually family members), although initially most closely associated with schizophrenia, have been extended to study depression (Hooley, Orley & Teasdale, 1986; Vaughn & Leff, 1976). Such studies generally conclude that individuals with depression were more likely to relapse after experiencing a hostile, critical social environment (Wearden, Tarrier, Barrowclough *et al*, 2000; Hooley, 2007). An association with relapse does not imply an association with the aetiology of the disorder. The evidence supporting a causal role for critical expressed emotion in depression is mixed. A prospective study of a child cohort found children in high-criticism families were more likely to develop depressive symptoms than those in low-criticism families, suggesting expressed emotion may be a risk factor for the

development, as well as the maintenance, of depression (Burkhouse, Uhrlass, Stone *et al*, 2012). Similarly, high rates of maternal criticism have been reported in cohorts of children with or at risk for depression (Silk, Ziegler, Whalen *et al*, 2009; Tompson, Pierre, Dingman Boger *et al*, 2010). However, other longitudinal studies have highlighted the potential for reciprocal effects. Associations have been identified (albeit inconsistently) between past critical parenting and later childhood depression, and between earlier depression and increased harsh parenting, with the latter effects arguably the stronger (Hale, Keijsers, Klimstra *et al*, 2011, Nelemans, Hale, Branje, *et al*, 2014; but see null findings from Frye & Garber, 2005). It should also be noted that this relationship is confounded with shared genetics between parent and offspring (Lau & Eley, 2008). There therefore appears to be a complex relationship between critical social relationships and depression.

The influence of the parent-child relationship is an important component of the general effect of the social environment and of expressed emotion. Parenting is a multi-faceted and complex phenomenon that can be conceptualised in a number of different ways. For example, one aspect of parental strategy is the level of control within the parent-child relationship, which in turn has sub-dimensions encompassing withdrawal of the parent from the child, hostility towards the child, and warmth in parent-child interactions (McLeod, Weisz & Wood, 2007). This complexity has spawned a

broad literature on the effects of parenting on juvenile depression, but this literature is difficult to reconcile to identify consistent findings, partly as a result of the wide array of parenting constructs. Nonetheless, multiple meta-analyses have suggested an effect of hostile parenting associated with increased rates of depression in young people (McLeod, Weisz & Wood, 2007; Yap, Pilkington, Ryan *et al,* 2014).

Variability in the definition of depression, and the influence of environmental factors like stressful life events, can create heterogeneity that reduces the power of meta-analyses of genetic studies. Examining the genetic correlations between independent GWA studies of schizophrenia results in very high correlations ≈ 0.9; despite clinical concerns about heterogeneity in the presentation of schizophrenia, there appears to be at least genetic consistency between studies (Cross-Disorder Group of the Psychiatric Genomics, 2013; Gratten, Wray, Keller, *et al*, 2014). However, performing the same analysis in the PGC depression cohorts yields much lower correlations ≈ 0.55, reflecting phenotypic heterogeneity (Cross-Disorder Group of the Psychiatric Genomics, 2013; Gratten, Wray, Keller, *et al*, 2014).

Recently, however, depression genomics has begun to produce results. The most recent PGC mega-analysis (with a considerable increase in the number of individuals studied) has identified a number of associated variants (although the final details were still unpublished at the submission

of this thesis; Lewis, 2015). Other significant associations have been identified by taking alternative approaches to increasing power. A GWAS in a population cohort of ~300,000 European individuals identified 15 loci at genome-wide significance in meta-analysis with existing data, demonstrating the power of large cohorts (Hyde, Nagle, Tian, *et al*, 2016). Actively reducing heterogeneity has also yielded success in studying depression, which demonstrates that increasing phenotype specificity can yield better power even if the cohort size is reduced (Traylor, Markus & Lewis, 2015). For example, selecting cases with an age-of-onset later than 27 years identified a variant on chromosome 3 which reached genome-wide significance (Power, Tansey, Buttenschon, *et al*, 2016). Restricting the depression case group to women with severe melancholic depression in a Chinese sample yielded two variants at genome-wide significance (Consortium, 2015a). Contrasting approaches to increasing power (interrogating larger cohorts and directly addressing heterogeneity) are yielding insights into the specific genetic variants underlying variance in depression.

1.2.2. The genetics of anxiety disorders may mirror those of depression

As well as high heterogeneity, depression also exhibits high comorbidity with other psychiatric disorders, particularly anxiety (Kessler, Berglund, Demler, *et al*, 2003). The two disorders are frequently studied

together as internalising disorders (in contrast with the externalising disorders: attention-deficit hyperactivity disorder, conduct disorder and oppositional-defiant disorder), and the DSM-5 has a specific diagnosis of anxious depression (Association, 2013). Furthermore, there is evidence of high genetic overlap between depression and generalised anxiety disorder, suggesting that the different presentation of these disorders may result primarily from environmental influences (that is, those not attributable to the additive effect of genetics; Kendler, 1996; Kendler, Gardner, Gatz, *et al*, 2007; Kendler, Neale, Kessler, *et al*, 1992b).

Anxiety also mirrors the heterogeneity seen in depression. Unlike depression, pathological anxiety is typically separated into distinct disorders: generalised anxiety disorder, panic disorder, agoraphobia, social anxiety disorder and specific phobias, as well as a childhood-specific disorder, selective mutism. Traditionally, obsessive-compulsive disorder and post-traumatic stress disorder were included within the anxiety disorders. However, the distinct characteristics of these disorders resulted in their separation from the anxiety disorders (and from each other) in DSM-5 (Association, 2013). Given the relative recency of this decision, many studies of anxiety disorders have included OCD and PTSD. A further disorder, separation anxiety disorder, was moved into the anxiety disorders from the childhood-onset disorders during the creation of DSM-5, an

acknowledgement of evidence that this disorder occurs (and has its onset) in adults as well as children (Association, 2013; Shear, Jin, Ruscio, *et al*, 2006; Silove, Alonso, Bromet, *et al*, 2015).

The diagnostic boundaries between the anxiety disorders are particularly blurred in childhood, where affected individuals are frequently comorbid for several anxiety disorders, and transition between disorders is common (Rapee, Schniering & Hudson, 2009). Phobias, social anxiety disorder, and separation anxiety disorder typically have their onset in childhood, while generalised anxiety disorder, panic disorder and agoraphobia show an average age of onset in adolescence or early adulthood (Kessler, Angermeyer, Anthony, *et al*, 2007; Kessler, Berglund, Demler, *et al*, 2005; Shear, Jin, Ruscio, *et al*, 2006). Not all juvenile anxiety disorders persist into adulthood but, because childhood represents a crucial developmental period, they are a risk factor for physical and psychiatric illnesses in later life, including adult anxiety disorder and major depressive disorder (Bardone, Moffitt, Caspi, *et al*, 1998; Gregory, Caspi, Moffitt, *et al*, 2007; Lewinsohn, Holm-Denoma, Small, *et al*, 2008). However, a caveat applies, namely that juvenile anxiety disorders are highly comorbid both within the anxiety disorders and with other disorders (such as major depressive disorder, conduct disorder and attention-deficit hyperactivity disorder) and as such associations between childhood anxiety and later adverse outcomes may be

confounded (Angold, Costello & Erkanli, 1999; Costello, Egger & Angold, 2005).

Heritability estimates for the anxiety disorders as a group and for the disorders individually are typically 30-40%, again mirroring depression (Hettema, Neale & Kendler, 2001; Kendler, Myers, Prescott, *et al*, 2001; Kendler, Neale, Kessler, *et al*, 1992a; Polderman, Benyamin, de Leeuw, *et al*, 2015; Scaini, Belotti & Ogliari, 2014; Shimada-Sugimoto, Otowa & Hettema, 2015; Tambs, Czajkowsky, Roysamb, *et al*, 2009; Van Houtem, Laine, Boomsma, *et al*, 2013). Attempts to identify specific genetic loci that contribute to this heritability have followed a familiar pattern, with equivocal candidate gene studies gradually giving way to better-powered genomic studies. The majority of well-powered studies have focussed on panic disorder; a recent meta-analysis identified 107 genes investigated in candidate gene studies (and GWAS), of which 23 variants in 20 genes had been studied regularly enough to justify meta-analysis (Howe, Buttenschon, Bani-Fatemi, *et al*, 2016; McGrath, Weill, Robinson, *et al*, 2012; Shimada-Sugimoto, Otowa & Hettema, 2015). Of these, variants in the *COMT* (rs4680) and *TMEM132D* (rs7309727/rs11060369) genes were significantly associated with panic disorder (although only in European participants) after controlling for multiple testing (Howe, Buttenschon, Bani-Fatemi, *et al*, 2016). For the majority of associations from the candidate literature, replication has

either not been attempted or has failed (Howe, Buttenschon, Bani-Fatemi, *et al*, 2016; Shimada-Sugimoto, Otowa & Hettema, 2015). This conclusion extends to the broader anxiety genetics literature – many studies have been published, but their results are contradictory, and no robust associations can be discerned (McGrath, Weill, Robinson, *et al*, 2012).

Furthermore, as with depression and schizophrenia, findings from the candidate gene literature have failed to replicate in the emerging GWAS literature. A caveat to this is that anxiety genomics is still in its infancy. A GWAS of phobia in 11000 individuals did not identify any associations at genome-wide significance (Walter, Glymour, Koenen, *et al*, 2013). In contrast, a series of studies of panic disorder in European samples have implicated the aforementioned rs7309727 variant in *TMEM132D* (p=1.05x10^{-8}, OR= 1.45 (95% CI: 1.20–1.72); Erhardt, Akula, Schumacher, *et al*, 2012; Erhardt, Czibere, Roeske, *et al*, 2011). Some additional support for this association comes from a family-based genome-wide linkage study of a broad anxiety phenotype in a Mexican-American cohort, which identified the 12q24.32-q24.33 region (which contains TMEM132D, amongst other candidates) at genome-wide significance (Hodgson, Almasy, Knowles, *et al*, 2016). Additional analyses of this variant in a Japanese cohort initially suggested no effect of this variant in Japanese samples (Erhardt, Akula, Schumacher, *et al*, 2012). However, recent analyses tentatively suggest that there may be a gene-by-gene interaction.

The HLA-DRB1*13:02 allele is significantly enriched in panic disorder cases compared to controls within this cohort ($p = 2.50x10^{-4}$; Shimada-Sugimoto, Otowa, Miyagawa, *et al*, 2015). Examining only individuals without this allele identified a significant association between panic disorder and rs7309727 ($p = 5.02x10^{-6}$; Shimada-Sugimoto, Otowa, Miyagawa, *et al*, 2016). Additional analyses in this Japanese sample have also implicated a variant in the *TMEM16B* gene (Otowa, Kawamura, Nishida, *et al*, 2012). However, the findings in Japanese cohorts to date stem from a single cohort, and as such require independent replication. The *COMT* variant discussed above has not reached genome-wide significance in GWAS of panic disorder to date.

Recently, GWAS of anxiety disorders as a heterogeneous group have emerged, including a large meta-analysis with a cohort size approaching those of the early PGC studies, which used both a case-control and a quantitative factor score approach to defining the phenotype (Otowa, Hek, Lee, *et al*, 2016; Otowa, Maher, Aggen, *et al*, 2014). These approaches yielded separate loci at genome-wide significance in multi-gene loci on chromosomes 3 and 2 respectively (case-control: rs1709393, $p = 1.65x10^{-8}$; factor score: rs1067327, $p = 2.86x10^{-8}$), with nominal significance for each locus in the alternative analysis (Otowa, Hek, Lee, *et al*, 2016).

Moderate heritability implies that a considerable portion of the variance in anxiety results from influences beyond the additive effect of

genetic variants. One potential source of such variance is the combined effect of specific environments. Perhaps unsurprisingly, many of the risk factors for depression also appear to be associated with anxiety. For example, a higher rate of stressful life events has been reported in individuals with anxiety disorders. The cumulative impact of multiple life events, particularly those related to threat, loss and poor health, are enriched in individuals suffering from panic disorder (Klauke, Deckert, Reif, *et al*, 2010). Associations with other specific life events, particularly childhood sexual abuse and violence, have been reported in specific phobia and social anxiety (Magee, 1999). Prospective studies have suggested that children with anxiety disorders experience a greater number of negative life events in the twelve months prior to the onset of their disorder than do unaffected children (Goodyer, Wright and Altham, 1988). In childhood, difficulties in social relationships, particularly between the child and their parents, have been associated with a broad anxiety phenotype (Goodyer, Wright & Altham, 1990; Van Der Bruggen, Stams & Bögels, 2008), as has experiencing multiple stressful life events and having poorer general health (Ford, Goodman & Meltzer, 2004; Phillips, Hammen Brennan *et al*, 2005). As such, stressful life events and the early-life social environment appear to contribute to anxiety as well as to depression.

The effect of the early environment on the development of anxiety has been of considerable interest to investigators, partly because several anxiety disorders have their onset in childhood (Kessler, Angermeyer, Anthony, *et al*, 2007; Kessler, Berglund, Demler, *et al*, 2005). Again, parenting represents a particular area of interest. The influences of warmth and control show similar effects to those observed in depression, although the evidence for a negative influence of lack of warmth is less consistent, and the role of over-control is more prominent (McLeod, Weisz & Wood, 2007; Murray, Creswell & Cooper, 2009). In addition, parents provide a model for children. In this way, parental displays of anxious behaviour or of anxious interpretations of ambiguous stimuli can lead to increased anxiety in children (although a child-to-parent effect could also contribute; Eley, McAdams, Rijsdijk *et al*¸2015; Moore, Whaley & Sigman, 2014; Muris, Steernmen, Merckelbach *et al*, 1996; Suveg, Zeman, Flannery-Schroeder *et al* 2005).

1.2.3. Behavioural genetics may be informative in understanding the aetiology of body mass index

Genomic studies in depression and anxiety are beginning to yield interesting findings, but progress has been limited, not least due to the subjective nature of the phenotypes. Genomics has had more rapid success in other complex traits and has begun to provide slightly unexpected insights. For example, body mass index (BMI) is superficially an anthropomorphic

trait but may have a behavioural component to its aetiology. Evidence from neuroendocrinological approaches suggests that BMI is closely linked to energy homoeostasis, regarding both the storage and use of energy in the body and the behavioural control of satiety and appetite (Llewellyn & Wardle, 2015; Lustig, 2001). These components are not mutually exclusive, and the importance of both has been supported by secondary genomic analyses of BMI (Locke, Kahali, Berndt, *et al*, 2015; Speliotes, Willer, Berndt, *et al*, 2010).

Gene-based analyses suggest a separable behavioural component to BMI when compared to other weight-related phenotypes. Body fat distribution (measured independently of BMI) was associated with genetic variants proximal to genes involved in adiposity (Shungin, Winkler, Croteau-Chonka, *et al*, 2015). In contrast, variants associated with BMI were enriched for genes highly expressed in the hypothalamus and pituitary gland, and the hippocampus and limbic system (Locke, Kahali, Berndt, *et al*, 2015; Shungin, Winkler, Croteau-Chonka, *et al*, 2015). These systems were previously implicated in the control of BMI by animal studies and clinical observations. For example, individuals with damaged hypothalami show increased appetite, and the appetitive hormone leptin binds to hypothalamic neurones to regulate metabolism via the melanocortic system (Anand & Brobeck, 1951; Farooqi, 2014; Schwartz, Woods, Porte, *et al*, 2000; Zhang,

Proenca, Maffei, *et al*, 1994). Disruptions to this system (in the form of rare gene-altering mutations) cause severe obesity in humans, although they do not explain the heritable component of BMI alone (Locke, Kahali, Berndt, *et al*, 2015; Montague, Farooqi, Whitehead, *et al*, 1997). Evidence from studies of both common and rare genetic variation thus supports an appetitive influence on BMI.

Additional support comes from examining the genome as a whole. The heritability detectable from GWAS analysis can be partitioned into functional components by annotating genetic variants to tissue-specific epigenetic marks. Performing this analysis in BMI identified a significant enrichment of central nervous system cell types, reinforcing the role of neural processes in BMI (Finucane, Bulik-Sullivan, Gusev, *et al*, 2015; Speliotes, Willer, Berndt, *et al*, 2010).

Finally, genetic correlations have been examined between BMI and multiple traits (Bulik-Sullivan, Finucane, Anttila, *et al*, 2015). When compared to psychiatric traits, BMI has a significant positive correlation with anorexia and schizophrenia, and a significant negative correlation with ADHD (Anttila, Bulik-Sullivan, Finucane, *et al*, 2016; Bulik-Sullivan, Finucane, Anttila, *et al*, 2015). The original paper describing the method of assessing genetic correlation reported only the anorexia correlation, while all three were reported in a second paper, which used a larger BMI cohort

(Anttila, Bulik-Sullivan, Finucane, *et al*, 2016; Bulik-Sullivan, Finucane, Anttila, *et al*, 2015). This latter study also tentatively suggests that BMI is correlated more strongly with traditional psychiatric phenotypes than with neurological disorders. However, the relationship between BMI and psychiatric traits is still emerging.

BMI has attractive properties as a phenotype for genetic study, compared to psychiatric disorders. It is a continuous phenotype and can be calculated objectively (via measurements of height and weight that have external validity) in large numbers of individuals. Evidence from twin studies suggests that BMI has a heritability around 60% (Polderman, Benyamin, de Leeuw, *et al*, 2015). However, designs, assessing correlations between distant relatives suggest estimates from twin and family studies may be inflated due to unmodelled effects such as assortative mating, and that the heritability of BMI could be closer to 30-40% (Hemani, Yang, Vinkhuyzen, *et al*, 2013; Visscher, McEvoy & Yang, 2010; Yang, Bakshi, Zhu, *et al*, 2015; Zaitlen, Kraft, Patterson, *et al*, 2013). These arguments concern the fundamental design of quantitative genetic studies, and as such extend to all traits.

Genomics has been successful in studying BMI, with the most recent meta-analysis identifying 97 significant loci (Locke, Kahali, Berndt, *et al*, 2015). These loci capture ≈ 2.7% of the variance in BMI (≈ 4-10% of the genetic

component, depending on the estimate of heritability), which compares favourably to schizophrenia (108 loci, ≈ 3.4% variance and ≈ 5% heritability, respectively; (Locke, Kahali, Berndt, *et al*, 2015; Schizophrenia Working Group of the Psychiatric Genomics, 2014). The genetic architecture of BMI also seems to mirror that of schizophrenia in part, with a common genetic component comprising many variants of small effect, and a single locus of larger effect. In the case of schizophrenia this lies in the major histocompatibility complex (MHC) gene cluster, while the strongest association in BMI GWAS is found in the *FTO* gene (Frayling, Timpson, Weedon, *et al*, 2007; Locke, Kahali, Berndt, *et al*, 2015; Schizophrenia Working Group of the Psychiatric Genomics, 2014). Curiously, there is now robust evidence that neither of these prominent associations acts via the genes in which they lie. The variant lying in the MHC appears to affect the *C4A* and *C4B* complement components rather than the more well-known *HLA* genes, while that near *FTO* alters the expression of the *IRX3* transcription factor (Sekar, Bialas, de Rivera, *et al*, 2016; Smemo, Tena, Kim, *et al*, 2014).

Unlike depression and anxiety, in which stressful life events are the most robustly-associated environmental influences, BMI is influenced by a broad range of environments (likely due to the metabolic aspects of the phenotype). Central among these is an energy-positive (as opposed to energy-balanced) environment, characterised by lower energy use from a

sedentary lifestyle with less physical activity and a diet rich in fat and sugar (Bray, 2004). Accurate measurement of the impact of these influences is confounded by the complexity of both concept and measurement in diet and activity, and reciprocal effects between the energy-positive environment and higher BMI (Rennie & Wareham, 1998; Ruel, 2003; Wareham, van Sluijs & Ekelund, 2005). A systematic review of the evidence for an effect of inactivity suggests overweight individuals typically engage in less physical activity, but that this pattern is limited by the confounding and reverse causality described above (Wareham, van Sluijs & Ekelund, 2005). Robust population-level evidence exists to support higher BMI in individuals following a meat-rich (Spencer, Appleby, Davey, *et al*, 2003) or sugar-rich diet (Te Morenga, Mallard & Mann, 2013), although this is similarly limited by uncontrollable confounding.

Beyond the energy-positive environment, there is a growing body of research suggesting that poorer sleep quality is associated with greater BMI independent of diet, with a number of hypotheses proposed to explain this association, including increased energy demands, alterations to the appetitive system, desynchrony of circadian rhythms, and comorbidity with psychological distress (Cespedes, Hu, Redline, *et al*¸2016; Chan, 2017; Chaput, 2014). The association between poor sleep quality and increased BMI is further supported by evidence of a genetic correlation of 0.15 between

BMI and insomnia symptoms from a recent sleep quality GWAS (Lane, Liang, Vlasac *et al,* 2017). While not as predominant as in depression and anxiety, an increased incidence of stressful life events has also been reported with increased BMI, independent of sleep quality and activity levels (Sampasa-Kanyinga & Chaput, 2017).

1.3 Gene-environment interplay

1.3.1. Genetic and environmental influences are not independent

The accumulated evidence argues that both genetic and environmental factors have important influences on human behavioural traits. The inherent assumption is that their actions are independent, which may not be the case; a proportion of variance in behavioural traits could be attributable to the combined effects of genetics and environment. This relationship has been discussed for over a century, with much of the debate concerning the relative importance of different types of interplay, the appropriate way to model interactions, and the relevance of statistical interactions to functional effects (Garrod, 1902, as described in Hunter, 2005; Kendler & Gardner, 2010; Rutter, 2010).

The interdependent effects of genetic and environmental factors are usually framed as gene-environment interactions and gene-environment correlations. Gene-environment interactions describe a case where the effect

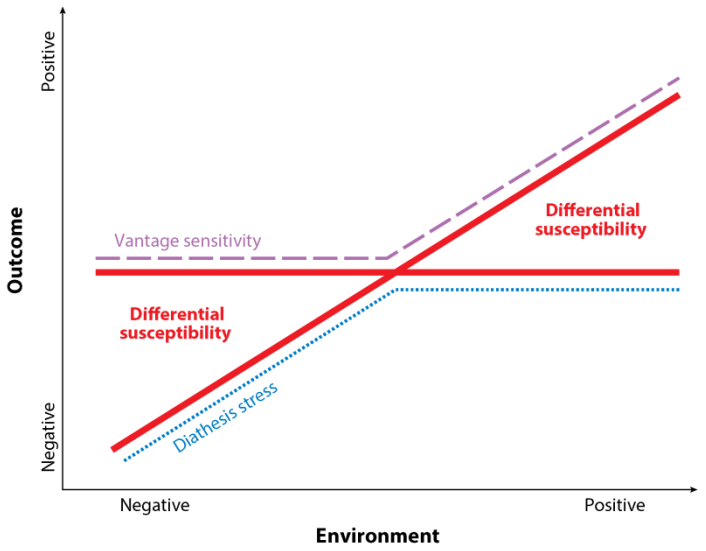
of one set of factors differs depending on the other. For example, exposure to sunlight has a greater effect on skin cancer risk in those with lighter skin (Adami, Hunter & Trichopoulos, 2008; Hunter, 2005). Gene-environment correlations describe instances when specific pairings of genetic and environmental influences co-occur at a rate greater than that expected by chance (Plomin, DeFries & Loehlin, 1977). One such correlation affects musical achievement; the frequency of practising an instrument is a heritable phenotype that in turn affects achievement (Hambrick & Tucker-Drob, 2015). Note that the examples given above refer to the influence of specific environments, rather than the broader meaning generally used in quantitative genetics. However, gene-environment interplay can also be studied using a quantitative genetic approach and modelling the environmental influence as all influences not attributable to the additive effects of genetics.

Gene-environment interplay can bias the results of genetic association studies (including GWAS) when not modelled explicitly (Cooley, Clark & Folsom, 2014; Marigorta & Gibson, 2014). To give an extreme example, a genetic locus in the *CHRNA5-A3-B4* gene cluster is robustly associated with an increased number of cigarettes per day in smokers (Ware, van den Bree & Munafo, 2011). The effect of this genetic variant is conditional entirely on whether the individual smokes (Hirschhorn, Lohmueller, Byrne, *et al*, 2002).

Such an association is impossible to discover in non-smokers and is a perfect gene-environment correlation (which is confounding). Non-independence between the effects of genetic variants and the environment can result in incorrect interpretations of evidence. For example, the *CHRNA5-A3-B4* locus is associated with lung cancer. The locus might be pleiotropic, having effects on both phenotypes. If this were true, there could also be a gene-environment interaction, with smokers having an increased risk of lung cancer if they carry this variant (Dudbridge & Fletcher, 2014). However, the weight of evidence favours confounding as the explanation, suggesting that the association results from the inclusion (intentional or otherwise) of smokers in lung cancer GWAS (Dudbridge & Fletcher, 2014; Gage, Davey Smith, Ware, *et al*, 2016; Hallden, Sjogren, Hedblad, *et al*, 2016; VanderWeele, Asomaning, Tchetgen Tchetgen, *et al*, 2012).

Classically, the effects of gene-environment interactions have been modelled in a diathesis-stress framework, in which bearing certain genetic variants (the diathesis) places an individual at risk of negative outcomes when exposed to negative environments or stress (Bleuler, 1963; Meehl, 1962; Rosenthal, 1963). This idea has been criticised as only accounting for the negative role of genes (Belsky & Pluess, 2009). Instead, a differential susceptibility hypothesis is posited, extending the diathesis-stress model to acknowledge the potential for a genetic predisposition to positive responses

in positive environments, a concept termed vantage sensitivity (Pluess & Belsky, 2013). Rather than vulnerability genes, leading to negative outcomes in negative environments, the differential susceptibility hypothesis suggests the existence of sensitivity genes, which increase responsivity to the environment in general (Figure 2; Bakermans-Kranenburg & van IJzendoorn, 2015). As such, individuals with a high load of sensitivity genes would be vulnerable to negative outcomes in negative environments, but would also benefit in positive environments. This hypothesis has found some support in human and animal studies (Bakermans-Kranenburg & van IJzendoorn, 2015; Kastner, Richter, Lesch, *et al*, 2015).



Bakermans-Kranenburg MJ, van IJzendoorn MH. 2015.
Annu. Rev. Psychol. 66:381–409

Figure 2: Alternative theories of gene effect in gene-environment interactions.
Reproduced from Bakermans-Kranenburg and van IJzendoorn (2015)

Gene-environment interactions have been a topic of interest within human behavioural genetics for decades, and a moderately-sized literature on the effects of individual variants in a variety of environments has been produced (Dick, Agrawal, Keller, *et al*, 2015; Duncan & Keller, 2011). However, the majority of these have made use of the candidate gene approach and show the same lack of power and inflated effect size estimates that limits genetic associations studies using this method (Munafo, 2015). The limitations on power are even greater in gene-environment studies, as the incorporation of an environmental exposure (with an inherent assumption of relevance) increases the level of multiple testing (Dick, Agrawal, Keller, *et al*, 2015; Duncan & Keller, 2011; Munafo, 2015).

Arguably the most robust candidate gene-environment interaction in psychiatric genetics is also the most controversial. The serotonin transporter promoter polymorphism (5HTTLPR) was initially implicated in the differential effect of life stress on depression in a longitudinal population cohort. Carriers of one or two copies of the short allele (which is associated with reduced transporter activity) were more likely to exhibit a variety of depression-related phenotypes following stressful life events than were carriers of two long alleles (Caspi, Sugden, Moffitt, *et al*, 2003; Heils, Teufel, Petri, *et al*, 1996). Since this first study, over 81 studies have attempted to replicate the finding, with mixed success (Sharpley, Palanisamy, Glyde, *et al*,

2014). The robustness of this interaction has become controversial, with some meta-analyses suggesting no effect and others confirming the original finding, with disagreement over the correct criteria for including studies in each meta-analysis (Duncan & Keller, 2011; Karg, Burmeister, Shedden, *et al*, 2011; Munafo, Durrant, Lewis, *et al*, 2009; Risch, Herrell, Lehner, *et al*, 2009; Sharpley, Palanisamy, Glyde, *et al*, 2014; Uher & McGuffin, 2010). The most recent meta-analysis reported a significant increase in depressive outcomes in S allele carriers exposed to stress (p=9x10^{-7}), robust to the nature and measurement of the stressor and the design of the study (Sharpley, Palanisamy, Glyde, *et al*, 2014). A number of criticisms of this evidence have previously been advanced. These include the potential risk of publication bias and undisclosed multiple testing, the observation that larger studies were more likely to be negative, and heterogeneity of study design and analysis (Munafo, Durrant, Lewis, *et al*, 2009; Munafò, Zammit & Flint, 2014; Risch, Herrell, Lehner, *et al*, 2009). The most recent meta-analysis addresses these concerns in part (Sharpley, Palanisamy, Glyde, *et al*, 2014). It seems unlikely that the observed result could be explained by publication bias: the data showed a fail-safe ratio of 45 non-published negative studies for every study included (Sharpley, Palanisamy, Glyde, *et al*, 2014). Negative studies in the meta-analysis had a significantly larger sample size than positive studies, and studies using self-report data to assess depression and stress (rather than an interview or objective measures) were more likely to be negative; these

factors are inter-related, as has been previously noted (Sharpley, Palanisamy, Glyde, *et al*, 2014; Uher & McGuffin, 2010). Nonetheless, there are several potential confounders (such as inter-study heterogeneity) that remain difficult to control (Sharpley, Palanisamy, Glyde, *et al*, 2014). Furthermore, reliance on a single meta-analysis is inadvisable, because differences in approach can translate to differences in outcome, as is apparent from the example of the 5HTTLPR (Taylor & Munafo, 2016). The validity of 5HTTLPR meta-analyses was recently tested using p-curve analysis (which investigates the distribution of p-values in studies reporting positive findings; Karg, Burmeister, Shedden, *et al*, 2011; Simonsohn, Nelson & Simmons, 2014; Taylor & Munafo, 2016). This alternative method provided weak evidence for an effect of the 5HTTLPR, which was very dependent on the lowest reported p-values. Given these results, the remaining doubt, and current understanding of the nature of genetic influences in psychiatry, the effect of the 5HTTLPR is still not fully resolved. The conclusion made by one commentary on the subject is probably the most reasonable: after addressing the confounding from various methodological artefacts, there appears to be a real, but probably small, effect of the 5HTTLPR (McGuffin, Alsabban & Uher, 2011).

1.3.2. Limitations on genome-wide analysis of gene-environment interplay

Using a genomic approach could allow gene-environment interactions to be studied with the hypothesis-neutral philosophy that has yielded success in exploring main effects on behaviour. However, this presents some problems in practice. Unlike the case for genetic traits, it is difficult to see how an agnostic approach could be taken in selecting environmental exposures, as there is no obvious finite amount of possible factors. Including multiple environmental influences in genome-wide interaction analyses increases the number of tests considerably. Assuming that a million independent signals result from genetic variation, every environmental factor tested adds a million new tests (Dudbridge & Gusnanto, 2008; The International HapMap, 2005). A two-step approach, whereby variants with main effects are identified and then tested for gene-environment interactions could reduce this burden (Ege & Strachan, 2013). However, the relative paucity of robustly identified main effects in behavioural phenotypes has limited the application of this process to date, and applying this method makes an inherent assumption that variants that interact with environmental influences will also show a main effect, which is contentious (Domingue & Boardman, 2016).

A genome-wide study of gene-environment interaction would not necessarily require all possible environments to be examined. It might be of

interest to assess the effects of genome-wide genetic variation on a specific environmental exposure. The choice of exposure would be hypothesis-driven, violating the agnostic approach of genome-wide studies in general. Therefore, proposing candidate environments requires an empirically robust association between the environmental and the trait under study (whether directly or via a gene-environment interaction; Dick, Agrawal, Keller, *et al*, 2015). In general, candidate environments fulfil this requirement better than candidate genetic variants. For example, there is robust evidence that childhood maltreatment is associated with later life depression (Nanni, Uher & Danese, 2012).

1.4 Alternative methods for leveraging genomic data to study gene-environment interplay

While genome-wide interaction studies (GWIS or GEWIS) are possible to perform (and have the potential to yield valuable results), they demonstrate some limitations. As such, it might be more informative to study these effects using alternative study designs, and leveraging the data-rich output from GWAS (Boardman, Domingue, Blalock, *et al*, 2014; Thomas, 2010). Although some of the caveats will remain, others can be reduced or removed with the analytical design. Within this thesis, two approaches are explored, making use of genomic and phenotypic data at varying depths. Specifically, a GWAS is conducted within a controllable environment

(cognitive behavioural therapy for anxiety disorders), to attempt to identify genetic variants predisposing to sensitivity to the environment (Chapter 3; Coleman, Lester, Keers, *et al*, 2016). This paradigm is then extended to using gene expression data, which represents a natural integration of genetic and environmental influences (Chapter 4; Coleman, Lester, Roberts, *et al*, 2016). Secondly, polygenic risk scores are used as proxies for the genetic component of variance in depression and BMI to dissect the correlation between these traits into genetic and environmental components (Chapter 5). Finally, more detailed measures of the social environment are used to explore genome-environment interactions predicting BMI in adolescence directly, using polygenic risk scores to reduce the multiple testing burden (and so increase power; Chapter 6).

1.4.1. GWAS in a controllable environment: the genetics of response to cognitive behavioural therapy

One of the major issues concerning gene-environment interactions is the difficulty in controlling the environment. It is not uncommon for critics of gene-environment studies (and of behavioural genetics in general) to compare investigations in humans to the more interpretable "common-garden" methods of studying animal behaviour (Gottlieb, 2003; Turkheimer, 2000). However, controllable environments exist (Eley, 2014). Studying the effects of genetics within a cohort who have all received a given exposure

can be informative about the genetic basis of differential response. This approach has been used with success in GWAS of trauma-exposed individuals to identify variants associated with the development of post-traumatic stress disorder (Ashley-Koch, Garrett, Gibson, *et al*, 2015; Guffanti, Galea, Yan, *et al*, 2013; Logue, Baldwin, Guffanti, *et al*, 2013; Nievergelt, Maihofer, Mustapic, *et al*, 2015; Solovieff, Roberts, Ratanatharathorn, *et al*, 2014; Stein, Chen, Ursano, *et al*, 2016; Sumner, Pietrzak, Aiello, *et al*, 2014; Wolf, Rasmusson, Mitchell, *et al*, 2014; Xie, Kranzler, Yang, *et al*, 2013). However, trauma is not a controllable exposure, which introduces heterogeneity into the approach.

Cognitive behavioural therapy (CBT) represents a positive environmental exposure that is more predictable and controllable than the majority of non-experimental environmental exposures studied. CBT is a psychological treatment in which the recipient (in a controlled, supportive atmosphere) identifies negative cognitive processes and behaviours underlying distressing outcomes and learns techniques to mitigate against these negative schema and relieve distress (Beck, 2005; Hofmann, Asmundson & Beck, 2013). Furthermore, it is possible to design prospective studies of response because CBT is provided in predictable sessions. The last decade has seen an increase in the use of cognitive behavioural therapy as a treatment for internalising disorders. For example, the UK's Improving

Access to Psychological Treatment program aims to offer CBT to at least 15% of those suffering from anxiety or depression in the community (Baldwin, Anderson, Nutt, *et al*, 2014; Bandelow, Lichte, Rudolf, *et al*, 2014; Bandelow, Lichte, Rudolf, *et al*, 2015; Clark, 2011; NCCMH, 2011). The increasing use of CBT has resulted in part from evidence that the technique is effective and demonstrates cost-efficacy. Meta-analyses of randomised control trials of CBT for anxiety in children have shown a significantly greater response immediately post-treatment compared to untreated waitlist controls (remission of disorder: odds ratio (OR) = 7.85, reduction in anxiety symptoms: standardised mean difference (SMD) = -0.98; James, James, Cowdrey, *et al*, 2013). Analyses of the adult literature show similar results comparing CBT to placebo (remission of disorder: OR = 4.06, reduction in anxiety symptoms: SMD = -0.73; sign of SMD reversed for comparison with child data; Hofmann & Smits, 2008). Estimates of efficacy are usually greater when considering completing participants only versus including all participants in intention-to-treat analyses. For example, the intention-to-treat results from the adult literature for disorder remission (OR = 1.84) and reduction in anxiety symptoms (SMD = -0.33) are smaller, although still significant (Hofmann & Smits, 2008).

Treating individuals with anxiety using CBT is effective compared to no treatment at all. Nevertheless, remission rates are rarely if ever 100%. CBT

treatment for childhood anxiety shows an average remission rate of 56% post-treatment, with very weak evidence suggesting a small increase with time after treatment (James, James, Cowdrey, *et al*, 2013). Similar remission rates have been reported in a meta-analysis of CBT for adults, 49.5% post-treatment rising to 53.6% at follow-up (Loerinc, Meuret, Twohig, *et al*, 2015). Remission rates are greater than those reported for placebo, waitlist or treatment as usual conditions, which range from 14-28% (Hofmann, Asnaani, Vonk, *et al*, 2012).

Analysable variance in response and remission following CBT exists between individuals, and a number of predictors of response have been posited (Haby, Donnelly, Corry, *et al*, 2006; Hudson, Keers, Roberts, *et al*, 2015; Mausbach, Moore, Roesch, *et al*, 2010; Olatunji, Cisler & Tolin, 2010). Meta-regression across 33 studies of CBT for depression, generalised anxiety disorder, and panic disorder assessed the effect of disorder, disorder severity, treatment type, treatment intensity, and therapist characteristics, as well as factors related to the conduct of the trial (Haby, Donnelly, Corry, *et al*, 2006). Of these potential covariates, only higher disorder severity was a significant predictor of (poorer) response. However, this study did not investigate some potential predictors suggested elsewhere. Engagement with therapy, in the form of homework compliance, was a significant predictor of outcome across disorders and within the anxiety disorders in a separate

meta-analysis (Mausbach, Moore, Roesch, *et al*, 2010). In another study, psychiatric comorbidity did not show an effect on treatment outcome at post-treatment or follow-up for anxiety disorders in general, but a negative correlation between comorbidity and treatment outcome was reported in mixed anxiety disorder specifically (Olatunji, Cisler & Tolin, 2010). These findings contrast with results from the cohort studied in Chapter 3, in which comorbid mood and externalising disorders predicted poorer outcome at post-treatment and at follow-up (Hudson, Keers, Roberts, *et al*, 2015). In addition, individuals with social anxiety disorder had a significantly poorer response than those with generalised anxiety disorder, as did participants whose illness was more initially more severe (Hudson, Keers, Roberts, *et al*, 2015). However, these latter results came from a mega-analysis of children receiving cognitive behavioural therapy for a range of anxiety disorders, incorporating studies from eleven sites across the globe. The strength of evidence for different predictors of CBT response differs between adults and children, and so differences between this study and others are unsurprising (Rapee, Schniering & Hudson, 2009). The potential for confounding by uncontrolled factors (such as the effect of the therapist) also cannot be excluded (Hudson, Keers, Roberts, *et al*, 2015). This last point extends to the literature in general – the many influencing factors and inherent heterogeneity in studying CBT limits the ability of meta-analysis and mega-analysis to detect predictors (Taylor, Abramowitz & McKay, 2012).

Nevertheless, the evidence suggests that a number of factors, particularly initial disorder severity and engagement with treatment, may predict differential response to CBT in anxiety disorders.

Given the uncertainty of phenotype-level predictors of treatment response, and the frequent time-specificity of such influences, examining genetic variation has value in understanding CBT response, as well as in broader behavioural genetic theory. Unlike environmental factors, the DNA code is fixed from birth, and as such does not alter with experience or time, suggesting genetic variants might predict response more generally and with greater stability than environmental alternatives.

The genetic study of response to CBT for internalising disorders has focussed on individual variants (the candidate gene approach), a field termed "therapygenetics" in one of the first such papers (Eley, Hudson, Creswell, *et al*, 2012). To date, variation near 19 genes has been investigated in studies of CBT or related psychotherapies for internalising disorders (Lester & Eley, 2013; Lueken, Zierhut, Hahn, *et al*, 2016). The most promising association, and the one most publications have explored, is the 5HTTLPR. Even for this variant, findings have been equivocal, with studies identifying improved response, poorer response, or no effect of the S allele or SS genotype (Andersson, Ruck, Lavebratt, *et al*, 2013; Bockting, Mocking, Lok, *et al*, 2013; Bryant, Felmingham, Falconer, *et al*, 2010; Cicchetti, Toth & Handley,

2015; Eley, Hudson, Creswell, *et al*, 2012; Felmingham, Dobson-Stone, Schofield, *et al*, 2013; Furmark, Carlbring, Hammer, *et al*, 2010; Hedman, Andersson, Ljotsson, *et al*, 2012; Kohen, Cain, Buzaitis, *et al*, 2011; Lester, Roberts, Keers, *et al*, 2016; Lonsdorf, Ruck, Bergstrom, *et al*, 2010; Sakolsky, 2011; Wang, Harrer, Tuerk, *et al*, 2009). In part, this may be attributable to a failure to control for the nearby rs25531 variant (which also reduces the effect of the long allele, making it functionally similar to the short allele; Nakamura, Ueno, Sano, *et al*, 2000). This is compounded by inconsistent use of a genotypic model (SS vs LS vs LL), a recessive model (SS vs others) or an allelic model (S vs L). The study of the 5HTTLPR, and the therapygenetics literature in general, is limited by its reliance on the candidate gene method and the associated tendency to explore a few variants in a limited set of genes, using a small cohort (Chabris, Hebert, Benjamin, *et al*, 2012; Ioannidis, 2005; Munafo, 2006). As such, the veracity of reported associations is unclear. The relative paucity of studies and the high heterogeneity in the design and nature thereof makes meta-analysis of findings difficult (which would, in turn, allow assessment of potential publication bias; Munafo, Clark & Flint, 2004).

The therapygenetics literature reflects the limitations of the candidate gene method. Given the robust associations identified by GWAS in the aetiology of behavioural traits, moving the therapygenetics field into

genomics has two objectives. Firstly, it provides a test of the exposure-only GWAS paradigm, which is valuable for assessing the effects of genetic variants on differential response to the environment. It should also have the practical benefit of improving the reliability of therapygenetic discoveries. Accordingly, Chapter 3 presents the first GWAS of CBT response study, using a cohort of 980 children recruited from sites across the world (Coleman, Lester, Keers, *et al*, 2016; Hudson, Keers, Roberts, *et al*, 2015).

1.4.2. Gene expression could be informative in exposed-only genome-wide studies

Genetic variants are attractive as possible predictors of CBT response, and genome-wide genotype data have applications beyond individual-variant association studies, such as polygenic risk scoring and the calculation of genetic correlations with other traits. Nonetheless, genetic effects are often difficult to interpret (Ward & Kellis, 2012). One of the most interesting findings from genomic studies is that the majority of significant variants lie in non-coding areas of the genome, suggesting some causal variants may exert their effects through altering the regulation of transcription or translation, the processes by which the DNA code determines the production of RNA and proteins (Hindorff, Sethupathy, Junkins, *et al*, 2009; Maurano, Humbert, Rynes, *et al*, 2012). However, it is not straightforward to identify the causal variant tagged by an association, nor to infer its functional effect.

As such, discovering a significant variant does not necessarily mean understanding why it is significant (Westra, Peters, Esko, *et al*, 2013).

Given the apparently regulatory role of many implicated genetic variants, it could be valuable to investigate differential RNA levels resulting from alterations in DNA transcription. Analysing the expression of genes shifts focus from the effects of a single variant (usually with only a minor influence on a highly polygenic trait) to those of many variants. Transcripts with significant differences in expression should be expected to explain more phenotypic variance than individual variants, and represent a potential biological marker of the combined effect of genetic and environmental influences (Emilsson, Thorleifsson, Zhang, *et al*, 2008; Rockman & Kruglyak, 2006). Reported raw correlations between RNA expression levels and protein abundance are modest (ranging between 0.3-0.6 in a study of more than 6000 proteins across 12 tissues; Wilhelm, Schlegl, Hahne *et al*, 2014). This may represent an effect of the rate of action of the translational machinery, differing between tissues and between specific gene products (Wilhelm, Schlegl, Hahne *et al*, 2014). Although the abundance of RNA transcripts and proteins differs between tissues, the gene-specific ratio of RNA to protein is consistent (Wilhelm, Schlegl, Hahne *et al*, 2014). Protein abundances predicted from RNA expression weighted by this ratio correlate strongly

with observed protein levels ($r > 0.8$; Edfors, Danielsson, Hallström, *et al*, 2016; Wilhelm, Schlegl, Hahne *et al*, 2014).

However, the analysis of RNA transcripts is limited by spatial and temporal restrictions to their effects. Changes in transcription are subject to a "hit-and-run" effect, whereby relevant changes are only detectable at the time they are occurring and in specific tissues (Stanworth, Roberts, Sharpe, *et al*, 1995). This latter issue (termed the tissue of interest problem) is a major limitation in investigating transcriptional and epigenetic biomarkers for behavioural traits (Davies, Volta, Pidsley, *et al*, 2012; Heijmans & Mill, 2012; Roth, Hevezi, Lee, *et al*, 2006; Sullivan, Fan & Perou, 2006). Brain tissue would be the ideal substrate in investigations of behaviour, as it is the most likely site for relevant biological variation. As extracting brain tissue from living subjects has obvious technical and ethical impediments, proxy tissues must be used. On the other hand, while tissue specificity may limit the insight that behavioural studies of gene expression can have into brain biology, this may not affect the study of response to cognitive behavioural therapy to a great degree. Robust associations between peripheral gene expression and response to therapy would have predictive value, even if they offered limited biological insight.

Variation in gene expression can be studied within the context of CBT response to provide a stronger indicator of the combined effects of genetic

and environmental variation. Only one study (from the cohort studied in Chapter 4) has examined gene expression in the context of CBT response for anxiety disorders. A genome-wide examination of change in expression across exposure treatment for panic disorder and specific phobia showed no association of treatment response with individual gene expression nor with co-expressed clusters (Roberts, Wong, Keers, *et al*, Under Review). Other studies have focussed on treatment response in PTSD and depression. An association between increased expression of *FKBP5* and response to CBT in PTSD has been reported in two independent candidate studies, albeit in small samples (Levy-Gigi, Szabo, Kelemen, *et al*, 2013; Yehuda, Daskalakis, Desarnaud, *et al*, 2013). Combined scores from a biomarker panel (previously associated with depression) were assessed in individuals with depression before and after CBT (Keri, Szabo & Kelemen, 2014a; Le-Niculescu, Kurian, Yehyawi, *et al*, 2009). After CBT, scores in cases were significantly greater, although still significantly lower than those in controls. Change in score was associated with a change in depressive severity on the Hamilton Depression questionnaire (Keri, Szabo & Kelemen, 2014a). The same group also assessed the effects of pro-inflammatory markers (Tol-like receptor genes and *NF-κβ*) on CBT for depression (Keri, Szabo & Kelemen, 2014b). A decrease in *TLR4* and *NF-κβ* expression (but not *TLR2* expression) was observed in treated individuals compared to controls, along with a dose-response effect, with greater reduction in pro-inflammatory markers associated with better

response to CBT. Gene expression analyses in CBT response are in their infancy and show early promise, but initial findings have come from candidate studies, and replication in larger, hypothesis-neutral studies is required.

Studying RNA transcripts and DNA variation individually has benefits and limitations for predicting differential response to CBT. Variation in DNA is effectively unchanging between tissues and does not alter with time, providing a source of predictors that can be validly assessed at any point before or after treatment. In comparison, the expression of RNA transcripts is spatially and temporally restricted and as such must be assessed prior to treatment, with careful consideration of which tissue to assess. However, statistical associations between a phenotypic outcome and DNA variation are often difficult to translate into specific causal variants and functional effects. In comparison, RNA transcript levels can be used to predict protein levels, which has clearer biological relevance for understanding the aetiology, and potentially modulation, of treatment response.

Combining these analyses can increase power and yield a greater understanding of the role of both sources of variation (Ritchie, Holzinger, Li, *et al*, 2015). If significant variants from a GWAS are also associated with gene expression, this suggests a potential function (and a route for further

investigation). Conversely, understanding which genetic variants alter the expression of a significant transcript may offer insights into the differential control of transcription.

RNA transcripts provide an indicator of the integration of genetic and environmental influences. As such, gene expression analyses within the context of CBT response could provide insight into the combined effects of genes and a controllable positive environment. Integrating genetic variation into this model in the context of expression quantitative trait locus analyses could provide further genomic context. In Chapter 4, a cohort of adults undergoing CBT for anxiety disorders, who provided DNA and RNA samples from whole blood, were investigated to assess the combined role of genetic variants and RNA transcripts (Chapter 4; Coleman, Lester, Roberts, *et al*, 2016).

1.4.3. Polygenic risk scores as a proxy for the genetic contribution to variance

Genomic studies provide a data-rich output, which can be used to perform additional analyses beyond the initial association study. A popular example of this higher-order approach is polygenic risk scoring (PRS), in which genetic variants within the study of interest (the *target*) are weighted by their association with a second phenotype, usually obtained from a separate association study (the *base*; Dudbridge, 2013; Euesden, Lewis & O'Reilly, 2015; International Schizophrenia, Purcell, Wray, *et al*, 2009; Wray,

Lee, Mehta, *et al*, 2014). The PRS strategy has two primary applications: it can be used within-trait to validate a phenotype, or across traits to compare the relationship between two phenotypes. It is this second use exemplified within this thesis.

1.4.4. Dissecting the genetic component of phenotypic correlation using polygenic risk scores

As discussed above, a major limiting factor in depression genomics to date has been the phenotypic heterogeneity of the disorder. Appetite dysregulation and altered weight are features of major depressive disorder, and the direction of this dysregulation has been proposed to generate a typical-atypical spectrum (Davidson, Miller, Turnbull, *et al*, 1982; Sullivan, Kessler & Kendler, 2014). The precise relationship between depression and BMI is unclear. Studies suggest a bidirectional, longitudinal pattern whereby high BMI predicts later, persistent depression and depression predicts later obesity, although the strength of the relationship depends on the precise measurement of depression and obesity (Gibson-Smith, Bot, Paans, *et al*, 2016; Luppino, de Wit, Bouvy, *et al*, 2010). Reported effect sizes are relatively small (largest OR = 1.58 for depression predicting obesity), which is concordant with inconsistencies seen between individual studies (Faith, Matz & Jorge, 2002; Luppino, de Wit, Bouvy, *et al*, 2010). Meta-analysis of these individual studies suggests there is a significant interaction, such that

individuals with depression typically have higher BMI, although there is also evidence for a U-shaped relationship (de Wit, Luppino, van Straten, *et al*, 2010; de Wit, van Straten, van Herten, *et al*, 2009; Luppino, de Wit, Bouvy, *et al*, 2010).

The phenotypic relationship between BMI and depression is complex, and the causal direction is unclear. Genetic variation is a potential source of shared aetiology. Twins studies of the relationship have suggested a small genetic correlation exists, estimated around 12% (Afari, Noonan, Goldberg, *et al*, 2010; Jokela, Berg, Silventoinen, *et al*, 2016; Schur, Godfrey, Dansie, *et al*, 2013). A series of studies have examined the potential genetic correlation between BMI and depression, largely focussing on genetic variants associated with BMI (Hung, Breen, Czamara, *et al*, 2015; Hung, Rivera, Craddock, *et al*, 2014; Jokela, Elovainio, Keltikangas-Jarvinen, *et al*, 2012; Rivera, Cohen-Woods, Kapur, *et al*, 2012; Samaan, Anand, Zhang, *et al*, 2013; Samaan, Lee, Gerstein, *et al*, 2015). Although approaches have explored a variety of variants both individually and in concert, they have provided only tentative evidence that genetic variants associated with BMI are associated with depression, primarily focussing on variants around the *FTO* and *TAL1* genes (Samaan, Anand, Zhang, *et al*, 2013; Samaan, Lee, Gerstein, *et al*, 2015). Variants in the *FTO* gene and a polygenic risk score of 32 BMI-associated variants have both shown increased association with BMI in depression

cases over controls, suggesting the depression status may interact with genetic risk to increase BMI (Hung, Breen, Czamara, *et al*, 2015; Rivera, Cohen-Woods, Kapur, *et al*, 2012).

Using polygenic risk scores as a proxy for the genetic component of each trait allows the assessment of the role of genetics and of environment (that is, influences beyond the additive effect of genetics) independently and together. If there were a shared aetiology between the two disorders, one would expect significant cross-disorder prediction using polygenic risk analyses, as well as within-trait genetic correlations. The absence of such a genetic relationship would argue instead for a more prominent role of other factors. These could be shared environmental effects, for instance an environmental insult that increased BMI as well as causing depression, or this could reflect the causal effect of one disorder on the other (such as higher BMI causing depression). Alternatively, there could still be a genetic relationship, but one that involved unmodelled factors like rare variants not captured by the common variants comprising the risk score. Identifying an interaction (for example between the polygenic risk for BMI and depression status, affecting BMI) might suggest a shared genetic sensitivity to the environment underlying the phenotypic correlation.

The role of genetics in this context was recently assessed in two independent studies. In the first, the association between polygenic risk

scores and depression was assessed in a cohort of depressed individuals stratified into typical and atypical conditions (Milaneschi, Lamers, Peyrot, *et al*, 2016). While case status in individuals with typical depression was only associated with polygenic risk scores from psychiatric disorders, polygenic risk scores from metabolic traits (including BMI) were associated with atypical depression (Milaneschi, Lamers, Peyrot, *et al*, 2016). The second publication, from the Generation Scotland study, explored the relationship between depression and BMI (in the context of neuroticism and general health) using polygenic risk scores from the 2010 GIANT BMI and the PGC major depressive disorder analyses (Clarke, Hall, Fernandez-Pujals, *et al*, 2015). Polygenic risk scores predicted within trait but did not predict across trait. A significant interaction was identified between BMI polygenic risk and depression status predicting BMI, and stratified analyses demonstrated an increased effect of polygenic risk for BMI predicting BMI within depression cases.

To add to this growing literature, I sought to perform similar analyses in the UK Biobank cohort, which provides a large sample independent of the previously studied cohorts. More powerful polygenic risk scores have also become available following the publication of these previous studies (Locke, Kahali, Berndt, *et al*, 2015). Accordingly, Chapter 6 presents an analysis of the

association between BMI and depression, including their respective polygenic risk scores in the model as well as their interaction terms.

1.4.5. Polygenic risk-environment interactions

Using polygenic risk scores as proxies for the genetic contribution to variance in a trait provides insight into the environmental contributions, but only inasmuch as the variance not accounted for by the genetic component is assumed to be environmental. If a well-defined environmental influence is available in a genomic dataset, however, it is possible to explore direct polygenic risk-by-environment interactions. Polygenic interaction studies are becoming increasing popular, and have been applied to test the interaction of genetic risk and stressful life events on the development of depression (Mullins, Power, Fisher, *et al*, 2016; Musliner, Seifuddin, Judy, *et al*, 2015; Peyrot, Milaneschi, Abdellaoui, *et al*, 2014; Vrshek-Schallhorn, Stroud, Mineka, *et al*, 2015). An interaction between PRS and childhood trauma has been reported, but failed to replicate in an independent cohort, potentially reflecting differences in study design (Mullins, Power, Fisher, *et al*, 2016; Peyrot, Milaneschi, Abdellaoui, *et al*, 2014). In behavioural genetics more generally, this approach has provided evidence that lifetime trauma exposure interacts with genetic risk for externalising behaviours such that highly traumatised individuals at high genetic risk show poorer working memory than others with low genetic risk (Sadeh, Wolf, Logue, *et al*, 2015). A

pair of studies on externalising behaviour and problematic alcohol use in adolescents suggests that peer deviance and low parental monitoring exacerbate genetic risk for both phenotypes (Salvatore, Aliev, Bucholz, *et al*, 2015; Salvatore, Aliev, Edwards, *et al*, 2014). Further investigations report that polygenic risk for alcohol dependence is associated with reduced verbal ability in individuals from socially deprived areas; that early adoption of heavy smoking in adolescence mediates the effect of genetic risk on smoking problems in adulthood; and that traumatic events and low neighbourhood social cohesion are associated with more cigarette smoking in individuals with high genetic risk (Belsky, Moffitt, Baker, *et al*, 2013; Clarke, Smith, Gelernter, *et al*, 2016; Meyers, Cerda, Galea, *et al*, 2013).

Although the use of this approach is growing and providing interesting insights, the limitations and assumptions of measuring and analysing a candidate environment in gene-environment interaction previously referenced still apply (Dick, Agrawal, Keller, *et al*, 2015). Obtaining a high-quality environmental measure in a cohort of sufficient size to perform genome-wide analysis requires a considerable investment. BMI is a useful phenotype in which to explore polygenic risk-environment interactions, as the relative ease of defining the phenotype allows greater investment in defining environmental measures. This ease of definition has also contributed to large GWAS analyses, enabling the generation of

powerful polygenic risk scores (Felix, Bradfield, Monnereau, *et al*, 2016; Locke, Kahali, Berndt, *et al*, 2015). Furthermore, many potential candidate environments might influence BMI. Previous studies have examined the effect on BMI of genetic risk interacting with education (no interaction); intake of sugary drinks (greater effect of genetic risk with more frequent intake); diet (greater effect of genetic risk in diets higher in saturated fat and fried food); and physical activity (smaller effect of genetic risk in more active, and less sedentary, individuals, although negative findings have also been reported; Ahmad, Rukh, Varga, *et al*, 2013; Casas-Agustench, Arnett, Smith, *et al*, 2014; Johnson, Ong, Elks, *et al*, 2014; Li, Zhao, Luan, *et al*, 2010; Liu, Walter, Marden, *et al*, 2015; Qi, Chu, Kang, *et al*, 2014; Qi, Chu, Kang, *et al*, 2012; Qi, Li, Chomistek, *et al*, 2012; Reddon, Gerstein, Engert, *et al*, 2016).

Exploring such interactions may be easier in children and adolescents, where environmental influences tend to be more restricted and controlled (due to the reduced autonomy of childhood). One example of such an environment is parenting style, which can influence child BMI via multiple pathways. Direct influences include parental control of child energy intake (through the regulation of eating behaviours) and energy use (such as by providing opportunities for exercise; Davison & Birch, 2001; Rhee, 2008). Parental style can also affect child BMI through more indirect means, including influencing child energy-related and self-regulatory behaviours

(Rhee, 2008; Rhee, Lumeng, Appugliese, *et al*, 2006). There is evidence that authoritative parenting (characterised by warmth and clear discipline) is associated with positive child energy behaviours (increased physical activity, healthier diet and lower BMI; Shloim, Edelson, Martin, *et al*, 2015). A recent study in the Twins Early Development Study (TEDS) cohort demonstrated an association between juvenile emotional difficulties and the interaction between colder, more punitive parenting style and a genetic risk score proposed to capture sensitivity to the environment (Keers, Coleman, Lester *et al*, 2016). As such, parental style can be considered an environment that has genetically-sensitive effects on behavioural phenotypes. The cohort and measure used in this study are also used in Chapter 6. Building on this sensitivity effect, the evidence suggesting parental style affects psychiatric phenotypes such as depression and anxiety in childhood, and the proposed role for parenting style in influencing childhood BMI, Chapter 6 investigates whether the effects of parental style on childhood BMI are influenced by the genetic influences.

However, parenting style is only one aspect of the more general influence of the social environment on juvenile BMI, and so there may be benefits to using a broader measure, such as parental SES, to capture this general effect. Previous studies, using a lifetime measure of socioeconomic status have suggested a greater effect of genetic risk and so higher BMI in

individuals with lower socioeconomic status (Liu & Guo, 2015). This is also reflected in epidemiological studies, including those in adolescents and young adults. Studies tend to report lower SES in late childhood associated with higher BMI, and the persistence and potentially the strengthening of this effect into adulthood (Braddon, Rodgers, Wadsworth, *et al*, 1986; Hardy, Wadsworth & Kuh, 2000; Monteiro, Moura, Conde, *et al,* 2004; Parsons, Power, Logan, *et al*, 1999; Shrewsbury & Wardle, 2008; Sobal & Stunkard, 1989; Sundquist & Johansson, 1998; Wang, Kim, Gonzalez, *et al*, 2007). However, the magnitude and direction of effects varies across epidemiological studies, most likely due to differences in related factors such as ethnicity, age and gender, and due to variability in the definition of SES. There are also apparent cohort effects, with null results more common in cohorts ascertained less recently, and the opposite direction of effect (that is, higher BMI associated with higher SES) more commonly seen in developing compared with developed countries (Blane, Hart, Smith, *et al*, 1996; Laitinen, Power & Jarvelin, 2001; Lauderdale & Rathouz, 2000; Monteiro, Moura, Conde, *et al,* 2004; Parsons, Power & Manor, 2001; Power & Moynihan, 1987; Shrewsbury & Wardle, 2008; Sobal & Stunkard, 1989). Nonetheless, in contemporary juvenile cohorts from developed countries, lower SES appears to be usually associated with higher BMI, and Chapter 6 seeks (as a secondary analysis) to investigate whether this relationship differs with the genetic predisposition to BMI.

Given that the social environment and genetics may both influence BMI through behavioural processes, it may be informative to investigate this genome-environment interaction. Therefore, adding to the growing polygenic risk-environment interactions literature in BMI, Chapter 6 examines the effect of the interaction of specific (parental style) and general effects of the social environment (parental SES) with genetic risk to predict BMI in late childhood and adolescence.

1.5 Summary and Aims

There is convincing evidence that both genetic variation and environmental factors affect behavioural traits, including psychiatric disorders and their treatment. It seems very unlikely that such influences act in isolation, but progress in understanding how such factors interact has been slow, in part due to concerns over the low power provided by genome-wide interactional studies. The chapters of this thesis explore alternative ways to use the data-rich output of genomic studies to investigate such effects, with the aim of adding to genetic theory and beginning to inform the pragmatic goal of improved decision-making in mental health.

Chapter 3 describes a genome-wide association study of response to treatment with cognitive behavioural therapy. This aims to identify any common variants present in the genome that can be captured using a low-coverage microarray, and which have a large effect (considered in an

additive model) on response to a shared environmental exposure. There are also three secondary aims of the study: to replicate positive associations from the candidate gene therapygenetic literature (some of which were reported with large effect sizes); to quantify the common additive genetic influence on CBT response via GCTA-GREML; and to establish whether polygenic risk scores generated from schizophrenia (the most powerful psychiatric polygenic risk scores available), depression (the psychiatric disorder with genomic data most closely associated with CBT response in patients with anxiety), and anti-depressant response (the phenotype with genomic data most similar to CBT response) were associated with response to CBT.

Chapter 4 extends the paradigm used in Chapter 3 to include individual differences in RNA expression, which result from the integration of genetic and environmental influences. Specifically, it examines the association of genetic variants and RNA expression levels (individually and together) with response to CBT. The principal aims of the study are to identify whether the level of expression of any RNA transcript is associated with CBT response; whether data- and literature-defined groups of transcripts are associated with CBT response; and to identify genetic variants that interact with CBT response to predict RNA expression. The ability of polygenic risk scores derived from Chapter 3 to predict response to CBT in

this cohort is tested as a secondary aim. Similarly, the association of PRS derived from one treatment group to predict response in the other is tested.

In the first two results chapters, genetic variants are assayed genome-wide, but primarily modelled as individual influences on response to an environmental exposure. The latter two results chapters examine polygenic risk scores as proxies for the overall effect of genetic variation genome-wide (although this is limited to common effects modelled additively). Chapter 5 uses polygenic risk scores to control for the genetic component of the phenotypic relationship between BMI and depression. If controlling for this component attenuates the association between BMI and depression, this implies that the relationship is principally genetic, and does not involve a substantial contribution from non-genetic influences ("environmental" influences, although including non-additive genetic effects and effects from rare variants). Specifically, this chapter aimed to investigate whether modelling the genetic influences on BMI and depression as optimised polygenic risk scores from the largest publically-available GWAS of these traits attenuated the association between BMI and depression observed in the absence of modelled genetic effects. Furthermore, this analysis also aimed to identify interactions between increased genetic risk for BMI and depression status associated with variance in BMI, and similar interactions associated with depression status. Secondary aims test whether the common-

variant heritability of BMI differs between depression cases and controls, and whether a genetic correlation exists between BMI and depression.

Finally, Chapter 6 seeks to assess directly the interaction between the genome-wide effect of common variants (modelled as a polygenic risk score) and a specific environment, parental warmth and punitive discipline, on pre-adolescent BMI, and on the change in BMI across adolescence. The main aim of the study is to identify whether the child's genetic predisposition to higher BMI is differentially associated with the BMI phenotypes of interest dependent on parental style. A secondary aim was to establish whether these effects differed according to the sex of the child. In the course of these analyses, it became clear that investigating socio-economic status as a broader measure of the social environment is valuable. Accordingly, secondary analyses are performed using parental socio-economic status at the birth of the child in the place of parenting, with the same aims.

This thesis seeks to exemplify that genome-wide genotype data can be used to examine the relationship between genetic and environmental influences on traits of psychiatric interest, in ways other than performing a genome-wide gene-environment interactional study. Specifically it examines genetic predictors of differential response to an environmental exposure (Chapter 3), the association of RNA expression on response to an environmental exposure (with and without genetic variation; Chapter 4), the

importance of the genetic component (compared to the environmental component) in the relationship between depression and BMI (Chapter 5) and assesses the interaction between genetic effects combined across the genome and a specific environment on BMI phenotypes in childhood (Chapter 6).

Chapter 2: Methodology

2.1 Genome-wide genetics

2.1.1. Linkage disequilibrium underlies genomic studies

The phenomenon of linkage disequilibrium is essential to the design and understanding of genome-wide studies. During the process of cell division, chromosome pairs align at the cell equator and are then pulled to opposite ends of the dividing cell (Remak, 1862; Weismann, Poulton, Schönland, *et al*, 1891). This separation can cause part of each chromosome to break off the main body and reanneal to the sister chromosome, a process called recombination (Holliday, 1964). Over multiple generations, repeated recombination events occur, resulting in a chromosome made up of stretches of DNA that originated on different chromosomes, known as haplotype blocks, linkage regions or regions of high linkage disequilibrium (Daly, Rioux, Schaffner, *et al*, 2001). Breakages are not equally likely along the length of the chromosome, resulting in recombination hotspots (Myers & Stahl, 1994; Sun, Treco, Schultes, *et al*, 1989). Consequently, the size of these linkage regions varies. Variants in linkage disequilibrium violate Mendel's law of independent assortment; sections of the genome in the same region are more likely to be inherited together than expected given their frequencies in the population (Bateson, Saunders & Punnett, 1906; Geiringer, 1944; Morgan, 1917). The identities of variants in linkage disequilibrium are highly

correlated, such that the identity of one variant can be used to predict another. This enables the genetic mapping of genomes and the dissection of heritability.

2.1.2. Whole genome sequencing enabled the development of microarrays

The Human Genome Project, and the genome-wide sequence data that resulted from it, was fundamental to the development of microarray-based genome-wide genetic studies for two broad reasons (Gabriel, Schaffner, Nguyen, *et al*, 2002; Lander, Linton, Birren, *et al*, 2001; Myers, Bottolo, Freeman, *et al*, 2005; Venter, Adams, Myers, *et al*, 2001). Firstly, it identified thousands of SNPs and their surrounding sequence, enabling the design of SNP-detecting probes that could be chemically bonded to a microarray to assay genetic variation genome-wide (Gunderson, Steemers, Lee, *et al*, 2005; Sapolsky, Hsie, Berno, *et al*, 1999; Steemers, Chang, Lee, *et al*, 2006). Furthermore, whole-genome sequence data provided an improved reference for mapping linkage disequilibrium across the genome. Consequently, the identity of millions of SNPs not present on microarrays can now be imputed (estimated with a quantifiable degree of confidence) from the known identities of assayed SNPs with which they are in linkage disequilibrium. A better understanding of the architecture of variation in the genome enabled the estimation of an appropriate threshold to control for the multiple testing inherent in genome-wide studies (Dudbridge & Gusnanto,

2008; Myers, Bottolo, Freeman, *et al*, 2005). The conventional genome-wide significance threshold of $p=5x10^{-8}$ was derived from this theoretical work, and has been established through its use by the large collaborations that have come to dominate the field (Collins & Sullivan, 2013; Fadista, Manning, Florez, *et al*, 2016; Psychiatric, Cichon, Craddock, *et al*, 2009; Sullivan, 2010).

2.1.3. Microarray-based genome-wide association studies

Whole-genome sequencing directly genotypes many more variants than sparse, proxy-reliant microarray-based genotyping. However, it is currently prohibitively expensive for the assessment of the tens of thousands of individuals required for the study of common complex behavioural traits (Corvin, Craddock & Sullivan, 2010; Spencer, Su, Donnelly, *et al*, 2009). Microarray-based genotyping, followed by imputation to a sequenced reference panel, can accurately estimate the identities of millions of genetic variants across the genome comparatively cheaply (Marchini & Howie, 2010; Spencer, Su, Donnelly, *et al*, 2009). For this reason, microarray-based GWAS has driven the rapid development of the field of psychiatric and behavioural genetics in the last decade (Smoller, 2014). It is central to the work presented in this thesis, both directly in the generation of new genomic data in Chapters 3 and 4, and indirectly in the use of existing data in Chapters 5 and 6.

Assaying many variants also provides large amounts of data that can be used to control for confounds such as between sample relatedness, quality of genotyping, and sample contamination. These data also allow studies to progress beyond individual association tests and integrate external information from other genomic datasets to examine associations between the phenotype and higher-order genetic variation. In this way, the genome-wide approach can assess the role of genetic variation ranging from individual variants to studying the effect of the entire genome (Visscher, Brown, McCarthy, *et al*, 2012; Wray, Goddard & Visscher, 2007).

Generating high-quality data from the output of microarray-based genotyping requires a careful process of both automated and manual inspection of variants (recalling), as well as several quality control steps (Anderson, Pettersson, Clarke, *et al*, 2010; Teo, 2008; Weale, 2010). These processes are described in the Appendix to this chapter (Appendix I), published as Coleman, Euesden, Patel, *et al* (2016). Specifically, this paper describes and justifies a process of excluding variants and individuals according to thresholds related to minor allele frequency (or count); missing variant calls; deviations from the expected pattern of Hardy-Weinberg equilibrium; sample relatedness; incorrect assignment of gender; heterozygosity of variants; and population stratification. It then describes the

process of imputing thousands (or millions) of variants, and describes secondary quality control and basic association analyses.

2.1.4. Measuring inter-individual relatedness using genomic data

The studies within this thesis use more advanced techniques than those described in Appendix I, which is a guide to the essential steps to perform a GWAS, not an exhaustive review of the field (which has also advanced considerably since the submission of that protocol). Appendix I describes the strategy for controlling for population stratification using principal components from genotyped variants (having removed variants in linkage disequilibrium; Menozzi, Piazza & Cavalli-Sforza, 1978; Price, Patterson, Plenge, *et al*, 2006). Although principal component analysis is an effective means by which to control for population stratification, a finer-scale method has since been developed based on genomic-relatedness matrices (GRMs; Kang, Sul, Service, *et al*, 2010; Yang, Benyamin, McEvoy, *et al*, 2010; Yang, Lee, Goddard, *et al*, 2011; Yang, Zaitlen, Goddard, *et al*, 2014). In this approach, individuals are scored at each SNP according to how different the number of reference alleles they carry is from the average in the cohort, weighted by the heterozygosity of the variant (Yang, Lee, Goddard, *et al*, 2011). These scores are summed across all variants to give an overall score per individual, which can then be compared to assess how similar any two individuals are within the cohort. In this way, GRMs provide an alternative

and more sensitive method of identifying sample relatedness and population stratification than does principal component analysis (Kang, Sul, Service, *et al*, 2010; Yang, Benyamin, McEvoy, *et al*, 2010; Yang, Lee, Goddard, *et al*, 2011). However, principal component analysis captures additional variance, resulting from confounding effects independent of sample relatedness such as genotyping batch effects. As such, some studies now recommend including fixed effects from principal components as well as the random effect of the GRM in mixed association models, or conducting more sophisticated integration of the two approaches (Yang, Zaitlen, Goddard, *et al*, 2014; Zhang & Pan, 2015).

The relatedness coefficient from the GRM can be entered into mixed linear models as a random effect to perform association analyses (MLMA; Yang, Zaitlen, Goddard, *et al*, 2014). In MLMA, the association of each variant with the phenotype of interest is examined in the context of the gross genetic similarity between individuals (Kang, Sul, Service, *et al*, 2010; Yang, Zaitlen, Goddard, *et al*, 2014). As a further refinement to this approach, regions of the genome close to the variant of interest can be excluded from the calculation of the GRM, increasing power (Listgarten, Lippert, Kadie, *et al*, 2012; Yang, Zaitlen, Goddard, *et al*, 2014). The specific example of this general approach used in this thesis is the leave-one-chromosome-out approach, in which GRMs are calculated from the whole genome excluding

the chromosome bearing the variant of interest (Yang, Lee, Goddard, *et al*, 2011).

Despite the recent success of genomic studies, variants identified at genome-wide significance generally explain only a small proportion of heritability (Manolio, Collins, Cox, *et al*, 2009). However, this "missing heritability" problem can be explored by using GRMs. By providing a measure of the small degrees of relatedness between population samples (compared to the higher relatedness of family members), GRMs allow estimation of the genetic component of variance via a process referred to as genomic relatedness matrix-based restricted estimation of maximum likelihood or GREML (Yang, Lee, Goddard, *et al*, 2011). The effects of all SNPs are entered into a mixed linear model as random effects weighted by the GRM, and the components of variance estimated by an iterative algorithm designed to maximise the likelihood of the estimate (Yang, Lee, Goddard, *et al*, 2011). The ability to estimate variance in this manner frees heritability estimation from requiring related samples (Visscher, Yang & Goddard, 2010). However, the resulting estimates are limited to the proportion of variance able to be captured by variants on microarrays (Visscher, Yang & Goddard, 2010). Additionally, the original method relies on several assumptions concerning the allele frequency distribution of causal variants and the effect of linkage disequilibrium. Violations of these

assumptions will bias the results of GREML studies. Improvements have been made to address these biases, by using multiple GRMs stratified by allele frequency and average linkage disequilibrium (Yang, Bakshi, Zhu, *et al*, 2015).

2.1.5. Secondary analyses using summary statistics (LD score regression)

Although GRM-based methods have been extremely influential and valuable to the field, they require individual-level genotype data to produce. While this is feasible for local datasets (such as those generated in Chapters 3 and 4), this becomes more burdensome for combining multiple studies, where data sharing may be limited by logistical or ethical concerns, or by the computational burden of analysing large amounts of genotype data (Finucane, Bulik-Sullivan, Gusev, *et al*, 2015; Loh, Bhatia, Gusev, *et al*, 2015). These difficulties motivated the development of methods that use summary statistics from association studies.

One particularly prominent method that has emerged in the last few years is LD score regression (Bulik-Sullivan, Finucane, Anttila, *et al*, 2015; Bulik-Sullivan, Loh, Finucane, *et al*, 2015; Finucane, Bulik-Sullivan, Gusev, *et al*, 2015; Loh, Bhatia, Gusev, *et al*, 2015; Loh, Tucker, Bulik-Sullivan, *et al*, 2015; Yang, Bakshi, Zhu, *et al*, 2015). Variants in linkage disequilibrium with causal variants show increased effect sizes in association studies (Pritchard & Przeworski, 2001; Sham, Cherny, Purcell, *et al*, 2000; Yang, Weedon, Purcell,

et al, 2011). In regions where many variants are in high linkage disequilibrium, the chances of one of those variants being causal (and so increasing the effect size of all other variants) is greater than in areas of sparse linkage disequilibrium. There is, therefore, a positive correlation between the average linkage disequilibrium of a variant (its LD score) and its effect size. The p-values obtained from a GWAS are expected to follow a uniform null distribution, and deviation from this expectation is known as genomic inflation. Inflation can result from the combined effect of true signals and from confounding processes such as population stratification or batch effects. These confounding effects are uncorrelated with linkage disequilibrium, and so, by regressing the effect sizes of variants on their LD score, it is possible to differentiate the polygenic effect of true signal (as captured by the LD score) from inflation due to confounds. This separation is the principle purpose of LD score regression, but it has a variety of other applications, including estimating heritability from genome-wide data (or from specific regions of the genome) and assessing genetic correlations between traits (Bulik-Sullivan, Finucane, Anttila, *et al*, 2015; Bulik-Sullivan, Loh, Finucane, *et al*, 2015; Loh, Bhatia, Gusev, *et al*, 2015).

LD score regression is a valuable approach and represents one method in an expanding arsenal of summary statistic-based methods. However, although such methods require less complex data to use than

genotype-reliant methods, they show reduced power (Bulik-Sullivan, Loh, Finucane, *et al*, 2015). As such, the most appropriate approach is determined both by data availability and by the intended analysis. In cases where genotypic data is available, using both methods (for example, to estimate heritability) can be valuable as a means of technical validation.

2.1.6. Polygenic risk scoring

The work in this thesis (particularly the latter two chapters) makes regular use of one particular secondary analysis from GWAS data. Polygenic risk scoring (PRS) combines both the genotype-level and summary statistic-based approaches discussed above (although versions of the technique using summary statistics only have been developed; Johnson, 2013; Palla & Dudbridge, 2015).

Even with the large sample sizes available to international consortia, genomic studies of behavioural phenotypes remain underpowered to detect all associated variants (International Schizophrenia, Purcell, Wray, *et al*, 2009; Palla & Dudbridge, 2015; Wray, Lee, Mehta, *et al*, 2014). Each variant within a study theoretically falls into one of four groups: true positives (variants capturing real effects that are called as significant), false positives (variants of no effect called as significant by chance), true negatives (variants of no effect that are not significantly associated with the phenotype), and false negatives (variants of real effect that miss significance within the study). The

conventional threshold of genome-wide significance ($p=5x10^{-8}$) aims to reduce the number of false positives, at the expense of increasing false negatives. As such, there may be valuable information beneath the significance threshold (false negatives). Incrementally increasing this threshold may gradually capture more signal, but at the expense of increasing statistical noise (true negatives becoming false positives; Figure 3).

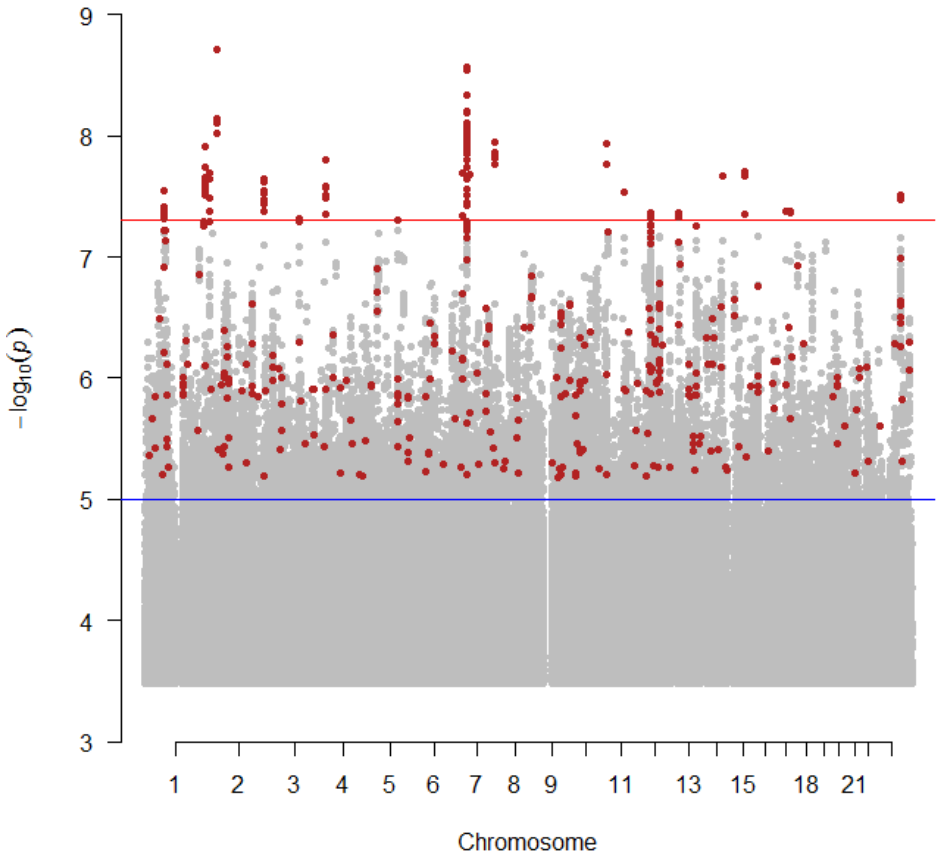


Figure 3: Imagined distribution of true positives (red) in GWAS results, showing the potential signal accessible beneath the conventional significance threshold (red line).

To create a polygenic risk score, variants from an existing GWAS (the *base*) are grouped by the p-value, and each effect allele is weighted by its effect size (Wray, Lee, Mehta, *et al*, 2014). Within the dataset of interest (the *target*), the weighted alleles are summed to give a total genetic risk score per individual. This score is then regressed on the target phenotype to assess the relationship between genetic risk for the base phenotype (determined by the p-values and effect sizes used in generating the score) and the target phenotype. Typically, this process is performed multiple times, with increasingly relaxed thresholds for including variants in the score (Wray, Lee, Mehta, *et al*, 2014). Within this thesis, polygenic risk scores were generated on genotyped data (or hard-called imputed data), using mean imputation for missing genotypes, and calculating the score as the mean average of each per-allele score (Chang, Chow, Tellier *et al*, 2015; Purcell, Neale, Todd-Brown *et al*, 2007).

As discussed in the introduction, PRS has two uses, within-trait and cross-trait. As an example, if the target were a case-control study of schizophrenia, results from the most recent PGC schizophrenia mega-analysis could be used as the base. The genetic risk would be expected to predict schizophrenia case status in the target dataset; were this not the case, it might then raise questions about the genetics (or diagnosis) of schizophrenia within the target cohort. The same logic applies when

investigating across traits, but in this instance, the results may be more informative about pleiotropy and shared genetic aetiology between the base and target phenotypes.

Examining the pattern of association between different risk scores and the phenotype can be informative about the target sample and about the base. Assuming the association study used as the base were well-powered, different traits would show different patterns. For a truly Mendelian trait, where a single variant captures all of the heritability (pi0 → 1 in Figure 4), risk scores constructed at the lowest significance thresholds would explain all of the genetic variance, and this would gradually decline as the threshold was increased and the signal was lost in the accumulating noise. At the opposite extreme, a highly polygenic trait underlain by many variants of small effect (pi0 → 0 in Figure 4) would show a different risk profile, one in which the proportion of variance explained gradually increased with more variants until an unknown tipping point was reached. At this point, the noise added by raising the threshold would negate the increase in signal (Figure 4; Dudbridge, 2013). As the power of the base study increases, the likelihood that any false negative becomes a true positive increases, and as such, the PRS profile of the polygenic trait is expected to become more like that of the Mendelian trait (Dudbridge, 2013; International Schizophrenia, Purcell, Wray, *et al*, 2009; Wray, Lee, Mehta, *et al*, 2014).

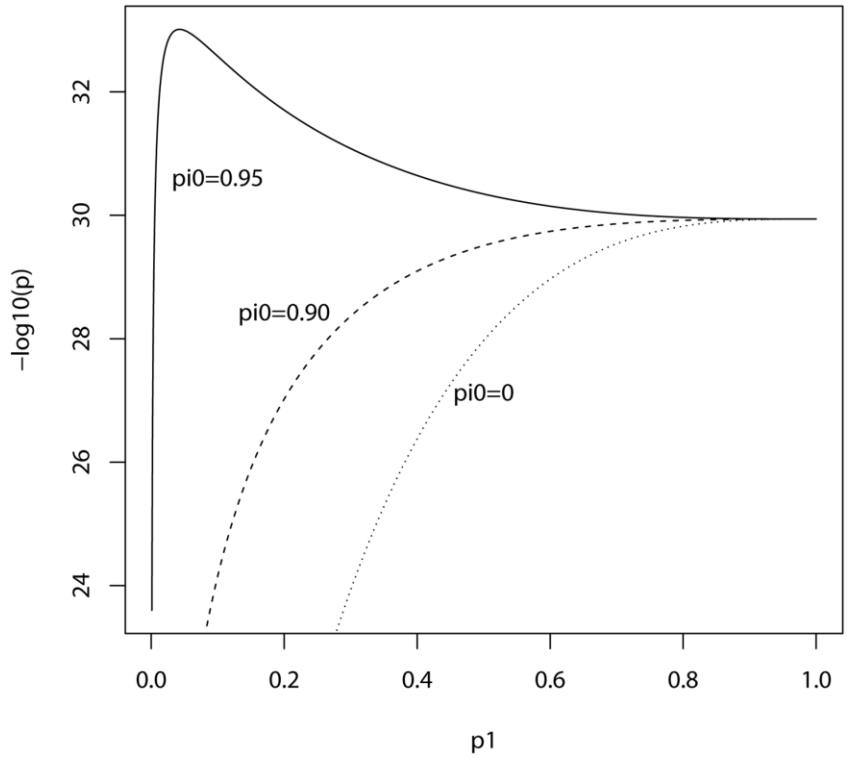


Figure 4: Expected p-values (y-axis) of multiple risk scores including variants with p-values between 0-p1 (x-axis), assuming different proportions of variants with no effect on the phenotype (pi0) and moderate (but not full) power to detect true positives. Reproduced from (Dudbridge, 2013).

The pattern of multiple risk scores can, therefore, be informative. In other instances, it is useful to define the optimal risk score (that is, the one that describes the most variance in the target). If a limited set of risk scores are assessed, it is likely that the optimal score will be missed, or that a misleading pattern is described. For this reason, high-resolution scoring approaches were developed, which extend the basic PRS method to examine thousands of scores, and apply a suitable correction for multiple testing (Euesden, Lewis & O'Reilly, 2015). In Chapters 3 and 4, risk scores are used

as proxies of the genetic component of variance, and as such it is valuable to identify the optimal risk score using high-resolution approaches.

2.2 Interaction modelling with linear and logistic regression

The second half of this thesis relies on modelling interactions between polygenic risk scores (as proxies for the effect of the genome) and measures of the environment (both inferred components and measured environments). Although the linear and logistic regression methods that underlie this approach are robust, commonplace statistical approaches, there are some considerations that must be made in performing such studies.

A statistical interaction is best defined (in this case) as a deviation from an additive function of genetic and environmental effects, on a given scale (Plomin, DeFries & Loehlin, 1977; Thomas, 2004). This does not translate directly into a specific mathematical operation. For example, many studies model interactions as multiplicative effects, where the effects are modelled on the logarithmic scale, and so the combined risk of both influences is expected to be the product of the risk from each influence alone. This is an analytical choice, rather than an inherent part of studying gene-environment interplay (Kendler & Gardner, 2010; Yang & Khoury, 1997). Particularly in the field of public health, modelling gene-environment interactions as additive (where significant interactions deviate from the sum of the risks from each influence) can be informative (Li & Chambless, 2007).

The appropriate way to model a given interaction is not always obvious. One solution would be to model all interactions as multiplicative effects (Thomas, 2010). This would be akin to standard practices in main-effects GWAS, where the multiplicative model of SNP effects has come to dominate. It would also have the benefit of reducing the multiple testing incurred by testing multiple models of interaction (Lewis & Knight, 2012; Zammit, Lewis, Dalman, *et al*, 2010). However, an inappropriate model may lead to false positive findings, or prevent the discovery of real effects (Rothman, Greenland & Walker, 1980). Furthermore, rejecting the null hypothesis in a study of gene-environment interaction only means that the interaction deviates from the modelled expectation – this may be entirely uninformative about the mechanism of any interaction that exists (Kendler & Gardner, 2010; Zammit, Lewis, Dalman, *et al*, 2010).

Modelling gene-environment interactions is scale-dependent, and so relies on the construction of the genetic and environmental variables. For example, environmental variables can range in the length, type, and precise measure of the exposure (Dick, Agrawal, Keller, *et al*, 2015). Especially in studies of human behaviour, the environmental variable may not have an obvious scale; while weight has a clear interval scale (kg), the severity of depression is a complex phenomenon usually described with an arbitrary ordinal scale (Dick, Agrawal, Keller, *et al*, 2015; Falconer & Mackay, 1996;

Lynch & Walsh, 1998). The nature of the scale is extremely important – for example, a multiplicative interaction between gene and environment becomes additive if the environmental variable is log-transformed (which may be an arbitrary decision), potentially changing the interpretation entirely (Dick, Agrawal, Keller, *et al*, 2015). Such variability in the environmental exposure produces a source of heterogeneity that will at best reduce the power of large-scale meta-analysis, and may result in true interactive effects between specific environments and genetic variants being lost (Dick, Agrawal, Keller, *et al*, 2015). Finally, most (if not all) of the statistical concerns surrounding candidate gene-environment interactions also apply to genome-wide studies (Dick, Agrawal, Keller, *et al*, 2015).

Gene-environment interaction studies have traditionally included covariates within their models but frequently have included these as main effects only (Keller, 2014; Yzerbyt, Muller & Judd, 2004). To control for the effects of these covariates on the interaction term, it is necessary to include gene x covariate and environment x covariate effects as well, with the resulting increase in terms in the model and degradation of evidence for the interaction term (Keller, 2014; Yzerbyt, Muller & Judd, 2004). Although this correction was proposed to address off-target gene-environment correlations in single-variant gene-by-environment interaction studies, the logic extends to genome-by-environment interaction studies as well.

Gene-by-environment interaction models assume that the genetic and environmental components are independent (Dudbridge & Fletcher, 2014). If this assumption is false, the results of the interaction may be biassed such that a spurious interaction could be generated or a true interaction obscured (Dudbridge & Fletcher, 2014; Jaffee & Price, 2007; Purcell, 2002). Accordingly, it is necessary to test for gene-environment correlations when exploring gene-environment interactions, and to interpret the results appropriately (Jaffee & Price, 2007).

2.3 Gene expression analyses

Much as is the case for genomic studies, whole-genome expression studies have benefited from increased knowledge from the Human Genome Project and from technological advancement, particularly the development of DNA microarray technology (Kuhn, Baker, Chudin, *et al*, 2004; Schena, Shalon, Davis, *et al*, 1995). Genome-wide investigations require certain considerations regardless of the exact biology studied. Just as GWAS requires large sample sizes and consistency in the way that DNA is obtained, so does genome-wide gene expression analysis. Although the sample size requirement may be smaller (because there are fewer RNA transcripts compared to genetic variants), the requirement for consistency is arguably greater. There are two reasons for this. The tissue-specificity issue previously discussed constrains analyses to a single tissue or else requires adequate

control for cross-tissue analysis (Heijmans & Mill, 2012). Single-stranded RNA is also a less stable molecule than double-stranded DNA, and as such requires more immediate analysis following extraction, with a resultant increase in the potential for technical artefacts (batch effects; Leek, Scharpf, Bravo, *et al*, 2010; Scherer, 2009; Tsui, Ng & Lo, 2002).

Genome-wide expression analysis involves assessing associations between the phenotype of interest and synthesised DNA complementary to RNA transcripts. Typically, gene expression studies have focussed on a differential expression approach, assuming a dichotomous phenotype or different cell types for comparison (Ritchie, Phipson, Wu, *et al*, 2015). However, the analysis of continuous phenotypes is straightforward, and at the simplest level can be performed with correlation or regression analyses.

In addition to analysing individual transcripts, network-based examinations of the combined effects of multiple genes can be performed (Butte & Kohane, 2000; Eisen, Spellman, Brown, *et al*, 1998; Zhang & Horvath, 2005). The gene sets studied can be drawn from the literature or through observing how the empirical data clusters (Butte & Kohane, 2000; Eisen, Spellman, Brown, *et al*, 1998; Subramanian, Tamayo, Mootha, *et al*, 2005; Zhang & Horvath, 2005). Multiple sources of information can be used to construct and define sets of related genes, including genomic features (such as DNase hypersensitivity sites), co-expression of RNA transcripts, or

interactions of the resulting proteins. These approaches have the benefit of highlighting patterns of enrichment that are obscured by focussing on single transcripts, and which may be biologically relevant. Collapsing the results from many individual data points into fewer sets also reduces the number of tests it is necessary to perform on the data.

The analysis presented in Chapter 4 uses two methods of network analysis to examine data-driven and literature-driven groupings respectively. Weighted Gene Network Co-Expression Analysis (WGCNA) is a suite of analysis methods built around a data-driven clustering algorithm that groups genes with similar expression into modules (although this method can cluster any data associated with a continuous metric; Langfelder & Horvath, 2008; Zhang & Horvath, 2005). This approach has become popular as a means of assessing inter-gene correlations because it relies heavily on networks, which are a familiar concept in biology and the mathematics of which can accurately describe observed biological patterns (Barabasi & Albert, 1999; Bergmann, Ihmels & Barkai, 2004; Jeong, Tombor, Albert, *et al*, 2000; Langfelder & Horvath, 2008; Tanaka, 2005; Zhang & Horvath, 2005).

WGCNA is built around a weighted network in which expressed transcripts are the nodes. Transcripts are connected to each other by edges, with each edge weighted by a continuous value (the adjacency of the two

nodes). In the simplest instance, this adjacency is determined by the correlation between expression levels of the two transcripts, raised to a power that optimises scale-free topology while still retaining a high number of connections between nodes (Langfelder & Horvath, 2008; Zhang & Horvath, 2005). Once a network is constructed, co-expression modules can be defined by linking strongly-connected genes together in a hierarchical manner (Langfelder, Zhang & Horvath, 2008). Associations can then be determined between the modules and the phenotype, with clear reductions in multiple testing compared to transcript-level analyses.

Modules provide a means of dimensional reduction in analysing expression data, and can also be biologically meaningful. Gene annotations from the literature (described in repositories such as Gene Ontology) can be used to identify enrichment for biological pathways in modules, as is implemented in WGCNA (Ashburner, Ball, Blake, *et al*, 2000; Langfelder & Horvath, 2008). Such analysis is not limited to modules, and many programs exist to conduct literature-driven clustering or pathway analysis at the level of individual transcripts. The method used in Chapter 4 is GOrilla (Eden, Navon, Steinfeld, *et al*, 2009). Unlike many methods for literature-based annotation, GOrilla can identify GO term enrichment in ranked lists of genes without requiring separate target and background sets (Eden, Navon, Steinfeld, *et al*, 2009). This is achieved by an algorithm that takes all listed

genes as the background set, defines all possible target sets with a stepwise inclusion of genes from the top of the list, and then identifies the target set that gives the strongest enrichment (Eden, Lipson, Yogev, *et al*, 2007). The p-value for this enrichment is then corrected for the multiple thresholding the algorithm involves (Eden, Lipson, Yogev, *et al*, 2007; Eden, Navon, Steinfeld, *et al*, 2009). Accordingly, GOrilla identifies enrichment of GO terms at the top of a ranked list, which is valuable for identifying potential biological signals of interest in the absence of a clearly defined target set. However, because GOrilla uses rankings rather than any weighting (for example, by the effect size of the association to the phenotype), the results of the enrichment analysis require careful interpretation, taking into consideration how the ranking was performed. In addition, the ranking of genes could be performed agnostic to the direction of effect (such as ranking by ascending p-value) or with regard to the direction of effect (such as by ranking on Pearson's r; Hong, Zhang, Li *et al*, 2013). Different rankings may yield different results. The published analyses in Chapter 4 used p-value based ranking; additional analyses ranking on Pearson's r are included in Appendix III.

Further insight into the importance of individual or group differences in expression in the context of a given phenotype can be gained by integrating genomic data into the analysis (which in turn provides new

information about the genomic data; Ritchie, Holzinger, Li, *et al*, 2015; Zhu, Zhang, Hu, *et al*, 2016). This can be achieved by context-dependent expression trait locus (eQTL) analysis. eQTL studies are effectively a special case of multivariate GWAS, using the expression of all RNA transcripts as the phenotypes (Jansen & Nap, 2001; Schadt, Monks, Drake, *et al*, 2003). Investigations can be restricted to associations between transcripts and variants local to the relevant coding region (*cis*-eQTLs), or can be truly genome-wide (incorporating *trans*-eQTLs). Performing such studies requires a considerable number of tests, especially in the case of genome-wide eQTL studies; however, efficient software for performing such interactions has been developed, such as the MatrixEQTL R package used in this thesis (Shabalin, 2012).

Genotype-environment interactions are integral to biology, and eQTL associations tend to be of large effect and detectable in cohorts comprising only tens of individuals (Monks, Leonardson, Zhu, *et al*, 2004; Morley, Molony, Weber, *et al*, 2004). However, it is the relationship between eQTLs and phenotypes that is of most interest. Such relationships can be used to annotate significant findings from association analyses; identifying a significant SNP from GWAS as an eQTL might provide functional insight.

It would be useful to perform a full interaction analysis, assessing whether genetic variation and differential gene expression together influence

a separate phenotype. However, this increases the number of tests performed. Every factor in a genome-wide interaction analysis incurs a million additional tests. Assuming twenty thousand transcripts are tested, billions of tests must be performed. For traits such as BMI, where measurement is straightforward and there are minimal ascertainment differences in participants, this level of multiple testing might be acceptable. In traits where sample sizes are limited and ascertainment is more complex (including most behavioural phenotypes), focussing on interactions between local SNPs and transcripts (or examining only *cis*-eQTLs and their associated transcript) can reduce the number of tests required (Consortium, 2015b). However, this comes at the expense of missing potential long-distance interactions. An alternative approach, limiting analyses only to SNPs already robustly associated with the disorder, has been successful in physical disorders but is limited by the current small numbers of associated SNPs in behavioural traits (Westra, Peters, Esko, *et al*, 2013).

Alternatively, genotype-by-phenotype interactions could be used to predict gene expression. These effects identify eQTLs that are context-dependent, and the effects tend to be considerably smaller and harder to detect than those of eQTLs in general. Selecting a subset of interactions to examine increases the power to detect significant interactions (although at the expense of missing true interactions that are not selected). Such selection

could be determined by location (favouring local *cis*- interactions over distant *trans*- interactions), or by the effects of individual components, such as including only those eQTLs where either the genotype or the transcript is associated with a main effect or including only known eQTLs (Hernandez, Nalls, Moore*, et al*, 2012).

Chapter 3

BJPsych | The British Journal of Psychiatry
1–8. doi: 10.1192/bjp.bp.115.168229

Genome-wide association study of response to cognitive–behavioural therapy in children with anxiety disorders

Jonathan R. I. Coleman, Kathryn J. Lester, Robert Keers, Susanna Roberts, Charles Curtis, Kristian Arendt, Susan Bögels, Peter Cooper, Cathy Creswell, Tim Dalgleish, Catharina A. Hartman, Einar R. Heiervang, Katrin Hötzel, Jennifer L. Hudson, Tina In-Albon, Kristen Lavallee, Heidi J. Lyneham, Carla E. Marin, Richard Meiser-Stedman, Talia Morris, Maaike H. Nauta, Ronald M. Rapee, Silvia Schneider, Sophie C. Schneider, Wendy K. Silverman, Mikael Thastum, Kerstin Thirlwall, Polly Waite, Gro Janne Wergeland, Gerome Breen* and Thalia C. Eley*

Background
Anxiety disorders are common, and cognitive–behavioural therapy (CBT) is a first-line treatment. Candidate gene studies have suggested a genetic basis to treatment response, but findings have been inconsistent.

Aims
To perform the first genome-wide association study (GWAS) of psychological treatment response in children with anxiety disorders ($n = 980$).

Method
Presence and severity of anxiety was assessed using semi-structured interview at baseline, on completion of treatment (post-treatment), and 3 to 12 months after treatment completion (follow-up). DNA was genotyped using the Illumina Human Core Exome-12v1.0 array. Linear mixed models were used to test associations between genetic variants and response (change in symptom severity) immediately post-treatment and at 6-month follow-up.

Results
No variants passed a genome-wide significance threshold ($P = 5 \times 10^{-8}$) in either analysis. Four variants met criteria for suggestive significance ($P < 5 \times 10^{-6}$) in association with response post-treatment, and three variants in the 6-month follow-up analysis.

Conclusions
This is the first genome-wide therapygenetic study. It suggests no common variants of very high effect underlie response to CBT. Future investigations should maximise power to detect single-variant and polygenic effects by using larger, more homogeneous cohorts.

Declaration of interest
R.M.R., J.L.H. and H.J.L. are co-authors of the Cool Kids program but receive no direct payments. C. Creswell is co-author of books in the 'Overcoming' series and receives royalties. W.K.S. is author of the Anxiety Disorders Interview Schedule for Children and receives royalties. G.B. is a consultant in pre-clinical genetics for Eli Lilly.

Anxiety disorders are the most common psychiatric disorders, with a lifetime prevalence of ~30%.[1] They are a major cause of global disability, and impose considerable economic burdens on society.[2,3] They commonly have their onset in childhood or adolescence and have been linked to the occurrence of later disorders, including depression and conduct disorder.[1,4] Adults with anxiety disorders show rates of childhood anxiety diagnoses significantly above baseline.[5] Given this potential gateway effect, and the distress caused by these disorders, there is a need to optimise and understand treatment effectiveness in childhood.

Cognitive–behavioural therapy (CBT) is a first-line treatment for anxiety disorders in the UK, with 59% remission reported immediately post-treatment.[6,7] Despite this high reported efficacy, variability exists in patient response that may be influenced in part by genetic variants. Multiple studies examining the genetics of differential response to psychological therapies (therapygenetics[8]) have been undertaken, and variants in seven genes (*5HTT/SLC6A4*, *TPH2*, *MAOA*, *COMT*, *NGF*, *BDNF* and *GRIK4*) have been implicated at least once in studies of CBT for anxiety disorders.[9] However, findings have proven difficult to replicate,[10] and the direction of effects found inconsistent. These problems may result from the low power of small cohort sizes, resulting in a high rate of false positives, and a narrow focus on a few genes that may have limited relevance to the phenotype.

Genome-wide association studies (GWAS) provide a hypothesis-neutral alternative, agnostic to prior assumptions of relevance and with the potential to discover novel findings at a single variant level. By analysing thousands of variants across the genome, GWAS yield more information than the candidate gene approach, allowing for the acknowledgement and control of confounds such as ancestry and the quality of genotyping. Genome-wide information can also be used to investigate associations between phenotypic change and different levels of the genetic architecture, including the effect of all variants in a given gene, and the effect of all genotyped variants across the genome. However, the explicit requirement for multiple testing correction in GWAS imposes a need for large sample sizes.

Although GWAS have not been used to study response to CBT, they have shown early promise in studying anxiety disorders. Genetic influences on the development of anxiety disorders may indicate processes underlying treatment response, and provide interesting genetic candidates.[11] A detailed review of the genetics of anxiety disorders is available elsewhere.[12] In brief, one variant, rs7309727 (*TMEM132D*), was associated with panic disorder in a cohort of European ancestry ($P = 1.1 \times 10^{-8}$, odds ratio (OR) = 1.45 (95% CI 1.20–1.72).[13] A variant in the *TMEM16B*

*These authors contributed equally to the work

gene was reported at genome-wide significance in a Japanese cohort with panic disorder, but was not significant in replication analyses.[14] Two GWAS of post-traumatic stress disorder (PTSD) have identified variants at genome-wide significance in the *TLL1* gene (rs6812849, $P = 3.13 \times 10^{-9}$, OR not reported)[15] and *PRTFDC1* (rs6482463, $P = 2.04 \times 10^{-9}$, OR = 1.47 (95% CI 1.35–1.59)).[16] However, these results require replication in larger studies; for example, variants in the *RORA* gene previously implicated in a GWAS of PTSD failed to attain significance in a larger replication effort.[17] No significant findings from the anxiety literature to date had previously been considered in candidate gene studies.[12]

To our knowledge, this is the first GWAS to examine response to psychological therapy in any disorder, and the first to examine treatment response of any kind in anxiety disorders. Participants were drawn from the Genes for Treatment (GxT) study, an international, multisite investigation of clinical, demographic and genetic predictors of response to CBT for anxiety in childhood and adolescence.[10,18] Two analyses of association between single nucleotide polymorphisms (SNPs) and response to CBT were conducted, investigating change in symptom severity between baseline and immediately post-treatment (post-treatment), and between baseline and 6 months after treatment cessation (follow-up).

Method

Study design and sample

A detailed description of the participants and the treatment programmes from which they were drawn is provided elsewhere (online supplemental material).[18] In brief, participants provided DNA for the GxT study between 2005 and 2013, at 11 sites across the USA, Australia and Western Europe. Children and adolescents (5–17 years old, 94% aged 5–13) were included if they met DSM-IV criteria[19] for a primary anxiety disorder diagnosis, with further psychiatric diagnoses made as appropriate. Exclusion criteria were significant physical or intellectual impairment, and the presence of psychotic symptoms. All participants completed a full course of individual-based CBT (with or without parental involvement), group-based CBT or guided self-help either as part of a trial or as treatment as usual within a clinical research department. All treatments were manualised and treatment protocols across all sites were comparable for core elements of CBT including teaching of coping skills, cognitive restructuring, and exposure.

Assessments were made using the Anxiety Disorders Interview Schedule for DSM-IV, Parent and Child Versions (ADIS-IV-C/P),[20] except at Bochum (Germany) and Basel (Switzerland) where the German equivalent, Kinder-DIPS,[21] was used. All participants were assessed prior to and immediately after treatment, with further assessments made at 3-, 6- or 12-month follow-up where possible. Output from the ADIS (or equivalent) was converted into Clinical Severity Ratings (CSR) on a scale of 0–8. A diagnosis was made when the child met the diagnostic criteria and received a CSR of 4 or more, usually based on a composite of parent and child report. Diagnoses were made from the ADIS for multiple anxiety disorders, and primary status allocated to the most severe, defined as the highest CSR, with ties resolved by clinical judgement (online Table DS1(b) and (c)).

To minimise differential assessment across sites, raters at Reading (UK), Oxford (UK) and Aarhus (Denmark) all received training in evaluation from the Sydney (Australia) site, and clinicians at Aarhus received additional training in the ADIS from W.K.S., principal investigator of the Florida (USA) site. As such, standardised assessments were made for at least 85% of the analysed sample (for further details see the online supplement).

Definition of the treatment response phenotype

As in previous analyses of the GxT sample, outcome was assessed across two periods: baseline to post-treatment and baseline to follow-up. Although dichotomised treatment outcomes are often used in clinical decision making in treatment response, a continuous measure of change in severity provides substantially more power for analyses.[22]

Response post-treatment was therefore defined as percentage change in CSR score between baseline and immediately following treatment. Percentage change, rather than absolute change, was used as it has been shown to better reflect clinical ratings of improvement by its successful use in pharmacogenetics GWAS.[23] For follow-up analyses, a range of time points were available; assessments taken at the 6-month time point were used, as these were the most complete ($n = 483$). Missing data at this time point was imputed using the best linear unbiased estimates from linear mixture models fitted to the GxT data as part of analyses predicting response from clinical variables alone.[18] The mixture models included the linear and quadratic effects of time as well as gender, age, primary diagnosis, treatment type and the random effects of individual and trial (for a full explanation, see Hudson *et al*[18]). This allowed us to compute response at follow-up as the percentage improvement in CSR score from baseline to 6 months after the end of treatment. Analyses were performed on residual scores generated from a linear regression of the percentage change measure adjusted for baseline severity, age, gender, treatment type, diagnosis and trial.

Both sets of residual scores were created as output variables from our previous study, which found a number of significant non-genetic influences on treatment outcome (online supplement).[18]

DNA extraction and genotyping

DNA was collected and extracted using standard protocols, from buccal swabs[24] and saliva kits (OG-500 / PrepitL2P, DNAgenotek, Kanata, Canada). Sample preparation (including concentration and quantification) prior to genotyping is described in the online supplement. Genotyping was performed on Illumina HumanCoreExome-12v1.0 microarrays (Illumina, San Diego, California, USA), using a standard protocol.[25] Samples were genotyped in two batches, and randomized by site on each microarray.

Quality control

SNPs were mapped to build version 37/hg19 of the human genome. Initial genotype calls were made with GenCall software (GenomeStudio, Illumina, San Diego, California, USA), reprocessed to remove poorly performing samples, re-clustered, and manually recalled where appropriate. Further recalling, targeted at improving the identification of rare variants (such as the exonic content of the microarray) was performed using ZCall.[26] Following recalling, the data were transferred to a multinode computing cluster, and quality control was performed following previously published protocols (online supplement).

Quality controlled data were imputed to the December 2013 release of the 1000 Genomes Project reference (for autosomes; March 2012 release for the X chromosome[27]), using the posterior-sampling method in IMPUTE2 with concurrent phasing.[28] SNPs imputed with an info metric >0.8 and a minor allele frequency (MAF) $>1\%$ were considered best-guess genotypes, and converted back to PLINK binary format using GTOOL (Freeman and Marchini, available at www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html). SNPs with a genotype probability of <0.9 were set as missing, and those present in $<98\%$ of the sample were excluded from the analysis.

Statistical analysis

Two analyses were performed, examining adjusted percentage change in CSR score from baseline to post-treatment, and from baseline to 6-month follow-up, as described above. Principal component analysis (PCA) of the genotype data was performed to attempt to control for population stratification. However, this yielded components that were not sensitive to differences in outcome. This was likely due to the quantitative nature of the phenotype, the fact that multiple covariates were controlled for in constructing the phenotype, and because participants were drawn from a variety of sites across the globe (online supplement). Accordingly, PCA was deemed unsuitable for controlling for population stratification, prompting the adoption of mixed linear modelling for the association analyses (MLMA). MLMA uses genome-wide genotype data to derive a genomic relationship matrix (GRM), which is used to control for genetic similarity between participants as a random effect.[29]

MLMA association analysis was performed in GCTA, using the *mlma-loco* option for autosomes and the *mlma* option for the X chromosome (online supplement).[30] For each SNP in the study, percentage change in CSR was regressed on the number of copies of the reference allele of the SNP (0, 1 or 2), weighted by its additive effect. A random effect of genetic similarity (from the GRM) was included as a covariate, as were fixed effects of sample concentration at genotyping, sample type (buccal swab or saliva), and ultrafiltration status (whether the sample was filtered in preparation for genotyping; online supplement). Using the assumptions of this approach, power for the GWAS was estimated using the Genetic Power Calculator.[31] The sample of 980 participants has 80% power to detect a variant explaining ~4% of variance and 1% power to detect variants explaining 1%.

Results from the association analysis were clumped according to *P*-value using PLINK.[32,33] Each clump is represented by a sentinel SNP (that with the lowest *P*-value in the clump), and contains all SNPs in linkage disequilibrium with the sentinel ($R^2 > 0.25$, within 250kb of the sentinel). One imputed sentinel SNP in the 6-month follow-up analysis was on the borderline of genome-wide significance (rs72850669, $P = 7.54 \times 10^{-8}$), and was re-genotyped *post hoc* (LGC Genomics, Teddington, UK). This showed the genotype calling of rs72850669 was unreliable (data not shown), and it was removed from the analyses.

To assess the ability of the GWAS to replicate previous findings, the association of SNPs implicated in CBT response in previous candidate gene studies was examined.[9] Exploratory secondary analyses were performed to assess the combined effects of SNPs on response (details can be found in the online supplement). The proportion of variance in CSR change across time accounted for by all the SNPs in the study was assessed with univariate genomic-relatedness-matrix restricted maximum likelihood (GREML), performed in GCTA using the GRM derived for the GWAS. Polygenic risk score profiling was used to investigate the ability of external data-sets to predict CBT response, using risk profiles from publicly available GWAS of major depressive disorder[34] and schizophrenia,[35] as well as from a meta-analysis of response to antidepressants.[36] To test the ability of the GxT data to predict response to CBT, five analyses were performed. Participants with generalised anxiety disorder, separation anxiety, social phobia and specific phobia, and those from the Reading (UK) site, were separately removed from the dataset and risk profiles derived from the remaining participants. Each profile was then used to predict outcome in the relevant set of removed participants.

Ethics

All trials and collection of samples were approved by site-specific human ethics and biosafety committees. Parents provided informed consent, children provided assent. The storage and analysis of DNA was approved by the King's College London Psychiatry, Nursing and Midwifery Research Ethics Sub-Committee.

Results

Sample and SNP exclusions are shown in online Fig. DS1. Phenotype and high-quality genotype data were available for 939 participants in the analysis post-treatment, with an additional 41 participants available for analysis at 6-month follow-up ($n = 980$). Baseline demographic information for these 980 participants is described in online Table DS1(a). The position of the samples on principal component axes derived from the HapMap reference populations suggests 92% of the sample are of White Western European ancestry.[37] A total of 260824 common SNPs passed quality control, which rose to 3017604 SNPs when imputed genotypes were added.

No SNPs were found at formal genome-wide significance for either analysis (all $P > 5 \times 10^{-8}$). In the post-treatment analysis, four independent clumps passed threshold for suggestive significance ($P < 5 \times 10^{-6}$; Table 1 and Fig. 1). Quantile–quantile plots show no departure from the chi-squared distribution of *P*-values expected under the null hypothesis, suggesting there is no underlying inflation of association statistics by uncontrolled confounds (lambda median = 0.972, Fig. 2). Three independent clumps were suggestive of significance in the 6-month follow-up analysis (Table 2 and Fig. 3), with no evidence of inflation (lambda = 1.02, Fig. 4). All clumps with $P < 1 \times 10^{-4}$ are displayed in online Table DS2.

A secondary analysis with increased power was performed restricted to nine SNPs previously associated with response to CBT in candidate gene studies (five other SNPs have been previously implicated in CBT response, but did not pass quality control). Assuming a significance threshold of 0.005455 (0.05/9), none of the nine previously associated SNPs was significant (Table 3 and online supplement). The sample had 80% power to detect an SNP accounting for 1.4% of variance at this significance threshold, suggesting any effect of these SNPs in this data-set is smaller than this.

Table 1 Independent clumps associated with cognitive–behavioural therapy response at post-treatment with $P < 5 \times 10^{-6}$

Sentinel SNP	CHR	Clump BP	Sentinel SNP *P*	Sentinel SNP MAF	Sentinel SNP information	Genes +/–100kb
rs10881475	1	108113663–108203647	2.45×10^{-6}	0.187	0.993	NTNG1, VAV3
rs11834041	12	128232721–128239057	3.50×10^{-6}	0.135	Genotyped	–
rs12464559	2	152498699–152679462	4.09×10^{-6}	0.0410	0.941	NEB, ARL5A, CACNB4
rs881301	8	38322346–38332318	4.46×10^{-6}	0.403	Genotyped	WHSC1L1, LETM2, FGFR1, C8orf86

SNP, single nucleotide polymorphism; CHR, chromosome; BP, base pair; MAF, minor allele frequency.

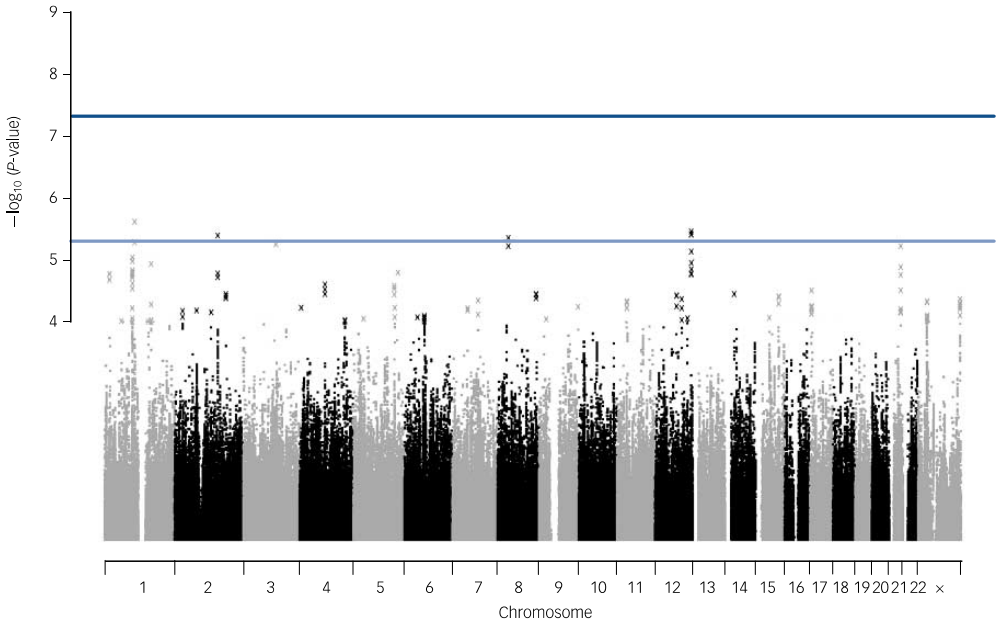


Fig. 1 Manhattan plot of genetic associations with cognitive–behavioural therapy response baseline to post-treatment.

X-axis shows the top million most associated single nucleotide polymorphisms, arranged by position on the chromosome. Lines show conventional thresholds for genome-wide significance ($P = 5 \times 10^{-8}$) and suggestive significance ($P = 5 \times 10^{-6}$).

Table 2 Independent clumps associated with cognitive–behavioural therapy response at 6-month follow-up with $P < 5 \times 10^{-6}$

Sentinel SNP	CHR	Clump BP	Sentinel SNP *P*	Sentinel SNP MAF	Sentinel SNP information	Genes +/–100kb
rs72711240	4	135657189–135695807	4.49×10^{-7}	0.0269	0.903	–
rs9875578	3	13707416–13810670	1.43×10^{-6}	0.424	0.994	FBLN2, WNT7A
rs6813264	4	146509970–146631854	4.68×10^{-6}	0.410	Genotyped	SMAD1, MMAA, C4orf51, ZNF827

SNP, single nucleotide polymorphism; CHR, chromosome; BP, base pair; MAF, minor allele frequency.

Exploratory secondary analyses (GREML, gene-wide analyses and polygenic risk score profiling) were performed. No significant estimate of SNP heritability could be obtained from GREML, and the effect of adding principal components was minimal. In the post-treatment analysis, all estimates were non-significant. In the 6-month follow-up data the highest estimate was 0.0797 (95% CI –0.194 to 0.35) without principal components. The power of univariate GREML in the sample was estimated for a range of true heritabilities.[38] Power ranged from 9 to 46% assuming true heritability between 0.2 and 0.6. To achieve 80% power within this range of heritabilities will require 1450–4450 samples (for heritabilities between 0.6 and 0.2).

Polygenic risk score profiling failed to generate predictions that were consistently significant, either for external GWAS or in the internal predictions of response.

Discussion

Main findings

We report the first genome-wide association study of psychological therapy. Although no region reached genome-wide significance, the single SNP and polygenic results are consistent with the wider literature of treatment genetics in psychiatry, given the sample size studied. Genome-wide significant variants detected in GWAS of psychiatric phenotypes have shown small effect sizes (with the exception of late-onset dementia), requiring thousands of participants to discover. The pattern of results in psychiatric genomics to date suggests that a critical number of participants (varying by disorder) are required before robust findings begin to be made. In studies of schizophrenia, this critical number was ~9000 cases.[39] Our results, although preliminary, suggest response to CBT could be a complex phenotype at the early point of this trajectory, although the critical sample size is not yet clear.

The purpose of this study was to identify genetic variants capable of predicting change in symptom severity during treatment. No common, high-effect SNPs were identified, suggesting that it is very unlikely a single variant could be used as a predictor. This also places an upper bound on expected effect sizes in studies of CBT response. This is relevant considering that neither GWAS replicated previous findings from the literature. This does not appear to be due to insufficient statistical power. For example, the *COMT* val158met polymorphism (rs6265) was reported to account for 8% of variance in CBT response in adults with panic disorder, well above the 4% of variance explained for which this GWAS was powered.[40] Failure to replicate previous findings from the candidate gene literature has proved common in psychiatric genetics, whereas GWAS is proving more

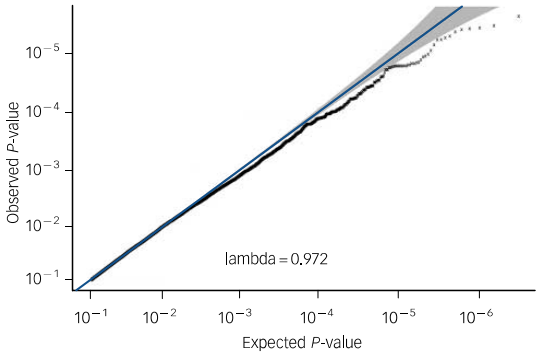


Fig. 2 Quantile–quantile plot of *P*-values (pruned for linkage disequilibrium) from genetic associations with cognitive–behavioural therapy response post-treatment.

X-axis shows spread of *P*-values expected under the null chi-squared distribution. Y-axis shows observed data. Grey region shows rough 95% confidence intervals around each point on the line x = y. Lambda median is a measure of inflation of the observed distribution of associations compared with expected null distribution. Lambda ⩽1 implies no inflation.

reliable.[35,41] The failure to replicate any published variants suggests previous assumptions of gene relevance may be erroneous, resulting from underpowered candidate gene studies that overestimated the likely effect sizes of studied variants, and that reported variants are likely to be false positives, or to have effect sizes inflated due to winner's curse.[42] Proximity to a gene does not imply an effect on gene expression, so the failure to replicate the effects of candidate SNPs does not exclude a role for candidate genes, as the SNPs assessed may not capture true functional variation.

Not all candidate variants are SNPs, and one limitation of GWAS is the difficulty of assessing structural variants not captured by the probes on microarrays. For example, we cannot comment on the previously reported role of the *MAOA*-u variable number tandem repeat in CBT response.[43] Nor could we assess the effect of the *5HTTLPR* variant of *SLC6A4,* previously associated with remission from anxiety disorders at follow-up; however, we directly genotyped this variant in this cohort, and were unable to replicate our earlier finding.[8,10]

Although small when compared with high-profile studies such as the PGC studies in schizophrenia and depression,[34,35] our sample is similar in size to studies in the depression pharmacogenetic literature.[23,44] The first of these used a multistage design ($n = 1532$) and identified several associations at nominal significance, but none remained significant after correction for multiple testing.[44] The second ($n = 706$) found one genome-wide significant locus (for response to nortryptiline treatment) and six loci at suggestive significance across four subanalyses.[23] More recent meta-analyses were unable to find genome-wide significant variants.[36] However, a significant GREML estimate of SNP-chip heritability of 42% (95% CI 6%–78%) was identified, suggesting useful information about treatment response can be obtained at the whole-genome level.[45] Future studies in psychological therapy-genetics should aim to build a cohort of sufficient size to estimate SNP-chip heritability and bivariate genetic correlations, enabling further comparison with pharmacogenetic studies. Such a cohort could act as a target data-set for polygenic risk scoring, exploring the predictive value of variants associated with potentially relevant phenotypes assessed in other GWAS.

Limitations

There are parallels between the antidepressant GWAS literature and this study, including the necessity of combining many studies to obtain sufficient participants for analysis. Herein, we examined a naturalistic clinical cohort, drawn from CBT trials or from treatment as usual. As all participants received CBT, there was no placebo group for comparison. Therefore, the results may

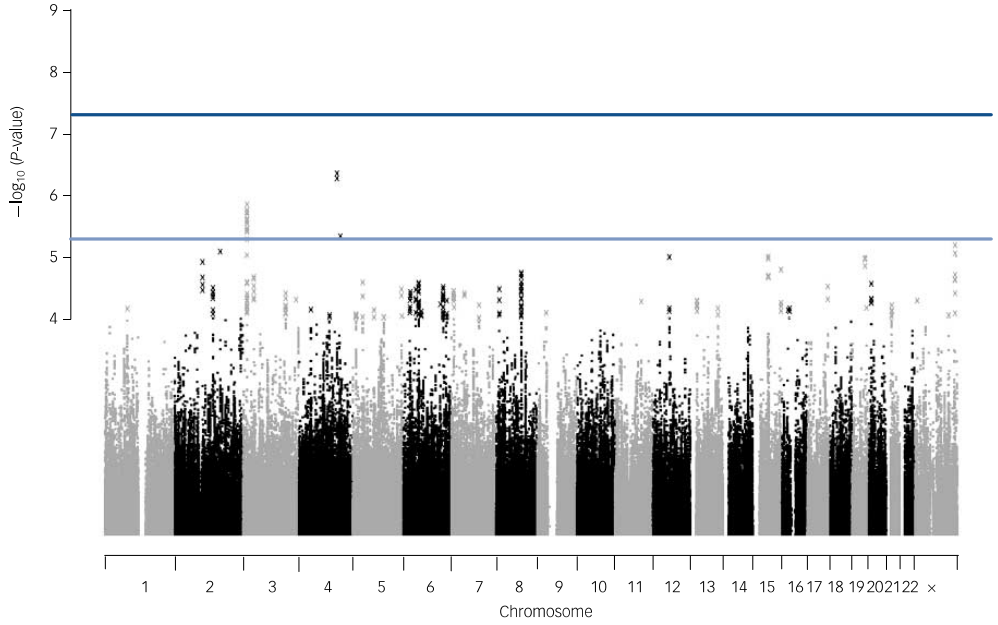


Fig. 3 Manhattan plot of genetic associations with cognitive–behavioural therapy response baseline to 6 months after treatment.

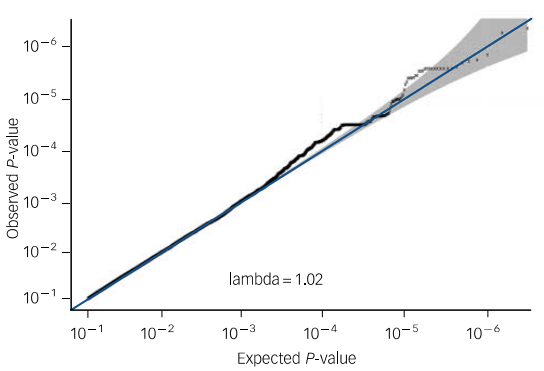


Fig. 4 Quantile–quantile plot of *P*-values from genetic associations with cognitive–behavioural therapy response baseline to 6-month follow-up, including lambda median.

reflect natural regression to the mean, rather than an effect of treatment. Theoretically, a parallel GWAS of change in severity could be performed on wait-list controls to identify associations with regression to the mean. Results from the GWAS of CBT response could be weighted by the likelihood that any given association resulted from regression to the mean. However, this would require deliberate non-treatment of thousands of wait-list controls over a period of at least 7 months for the purpose of comparison only. As CBT is effective in this age group, with significant improvement seen in treated groups relative to wait-list controls, non-treatment would raise serious ethical concerns.[7] The aim of therapygenetics is to discover predictors of differential response to treatment. These predictors need not capture a treatment effect *per se*; they may describe processes separate to treatment that nonetheless lead to better (or worse) response. Nevertheless, in the absence of a control group, this study specifically examines the association between genetic variation and change in CSR across the period of CBT treatment and follow-up, not the biological mechanism of response to CBT.

The naturalistic nature of the cohort creates heterogeneity, including differences in the details of the treatment given, the target disorder of the treatment, and several participant characteristics. The effectiveness of CBT is influenced by a variety of environmental factors. Some of these can be considered within the design, such as treatment type, diagnosis and severity. Others are less easily accounted for, including therapeutic alliance and other social influences, which may only be partly controlled for by the inclusion of trial as a covariate.[18,46] This reduces the statistical power of analyses, but should not be viewed as an argument against therapygenetics. The ability to offer personalised advice to patients about treatment could avoid considerable amounts of unnecessary distress and expense. Obtaining a set of genes able to assist in clinical prediction will require a cohort that is powerful enough to detect true variants while remaining clinically representative. Thus, a degree of heterogeneity is unavoidable in studying response to CBT, and similar difficulties in pharmacogenetic GWAS suggest this limitation applies to treatment response genomics more generally.

Combining data from trials at multiple sites necessitated compromises in study design. Participants were included if they completed treatment, but drop-out from treatment is common and likely to be related to poorer response. As such, future studies should aim to include severity data for non-completing participants. This would require appropriate modelling of the treatment period, and the proportion of the treatment process completed, before participation ceased. Similarly, combining measurements from different sites and from participants with varying diagnoses required the use of a general, widely applicable outcome measure. The ADIS fit these requirements well, but relies on clinical judgement derived from parent and child report. It may be less sensitive to the effects of CBT than a self-report measure, and be more vulnerable to site-specific biases. However, a suitable diagnosis-general self-report scale was unavailable, and standardising outcomes to combine multiple diagnosis-specific scales is likely to lead to a generalised and difficult-to-interpret result.

Future directions

This study represents the first GWAS of psychological therapy. Although no genome-wide significant findings emerged, the spread of significance in the associations captured is similar to other early general psychiatric and pharmacogenetic GWAS. The best approach in the immediate future is to increase the sample size available through combining existing cohorts in mega- and meta-analyses. Such a cohort would allow replication of the findings presented in this paper to be attempted, which currently is not possible due to the lack of an independent cohort of suitable

Table 3 Genome-wide association study *P*-values of single nucleotide polymorphisms (SNPs) previously associated with cognitive–behavioural therapy response.[12,a]

Gene	SNP	*P* (post-treatment)	*P* (follow-up)
SLC6A4	rs25531	Imputed with info <0.8	Imputed with info <0.8
HTR2A	rs6311	0.4717	0.9692
	rs6313	0.5451	0.8109
	rs6314	Imputed with info <0.8	Imputed with info <0.8
	rs7997012	Completeness after imputation <0.98	Completeness after imputation <0.98
TPH2	rs4570625	Completeness after imputation <0.98	Completeness after imputation <0.98
COMT	rs4680	0.7699	0.5956
NGF	rs6330	0.5093	0.4559
BDNF	rs6265 (val158met)	0.3408	0.9078
	rs7934165	0.5231	0.9880
	rs1519480	0.8211	0.5013
	rs11030104	0.3158	0.9675
GRIN2B	rs1019385	Imputed with info <0.8	Imputed with info <0.8
GRIK4	rs1954787	0.1315	0.1914

a. No *P*-value is significant after multiple testing correction.

size. However, individual variants are likely to have small effect sizes, so future studies should utilise higher-order approaches such as polygenic risk scoring and GREML to leverage the predictive effects of the whole genome. This would also provide an estimate of heritability, which is difficult to obtain through traditional family-based approaches. If the heritability of CBT response were around 30% (similar to that of anxiety disorders), a high-powered polygenic risk score could capture 10–15% of variance, which could be clinically useful when combined with known environmental risk factors.[47] However, creating such a score will require a sample size of at least 10 000, which would involve considerable effort to obtain.

Alternative approaches may also yield interesting findings. Response to CBT is a behavioural change following exposure to a positive environment, so epigenetic studies investigating how these exposures influence gene expression via DNA methylation will be informative.[48] Similarly, it will be useful to examine changes in gene transcript expression across treatment and in the longer term. Used in parallel to these approaches, studying specific genetic variants remains a potential method of predicting response to CBT (and understanding its biological basis) and genome-wide investigations represent the most promising avenue in which to focus the gathering momentum of therapygenetics.[49]

Jonathan R. I. Coleman, MSc, King's College London, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, UK; **Kathryn J. Lester**, DPhil, King's College London, IoPPN, MRC SGDP Centre, UK, and School of Psychology, University of Sussex, UK; **Robert Keers**, PhD, **Susanna Roberts**, MSc, King's College London, IoPPN, MRC SGDP Centre, UK; **Charles Curtis**, MSc, King's College London, IoPPN, MRC SGDP Centre, UK, and National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, UK; **Kristian Arendt**, PhD, Department of Psychology and Behavioural Sciences, Aarhus University, Denmark; **Susan Bögels**, PhD, Research Institute Child Development and Education, University of Amsterdam, The Netherlands; **Peter Cooper**, DPhil, School of Psychology and Clinical Language Sciences, University of Reading, UK, and Department of Psychology, Stellenbosch University, South Africa; **Cathy Creswell**, DClinPsy, PhD, School of Psychology and Clinical Language Sciences, University of Reading, UK; **Tim Dalgleish**, PhD, MRC Cognition & Brain Sciences Unit, Cambridge, UK; **Catharina A. Hartman**, PhD, Department of Psychiatry, University of Groningen, University Medical Center Groningen, The Netherlands; **Einar R. Heiervang**, PhD, Institute of Clinical Medicine, University of Oslo, Norway; **Katrin Hötzel**, PhD, Department of Psychology, Ruhr-Universität Bochum, Germany; **Jennifer L. Hudson**, PhD, Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia; **Tina In-Albon**, PhD, Clinical Child and Adolescent Psychology, Universität Koblenz-Landau, Germany; **Kristen Lavallee**, PhD, Department of Psychology, University of Basel, Switzerland; **Heidi J. Lyneham**, PhD, Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia; **Carla E. Marin**, PhD, Yale University, Child Study Center, New Haven, Connecticut, USA; **Richard Meiser-Stedman**, PhD, MRC Cognition & Brain Sciences Unit, Cambridge, UK; **Talia Morris**, BPsych, Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia; **Maaike H. Nauta**, PhD, Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, The Netherlands; **Ronald M. Rapee**, PhD, Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia; **Silvia Schneider**, PhD, Department of Psychology, Ruhr-Universität Bochum, Germany; **Sophie C. Schneider**, BPsych, Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia; **Wendy K. Silverman**, PhD, Yale University, Child Study Center, New Haven, Connecticut, USA; **Mikael Thastum**, PhD, Department of Psychology and Behavioural Sciences, Aarhus University, Denmark; **Kerstin Thirlwall**, DClinPsy, **Polly Waite**, DClinPsy, School of Psychology and Clinical Language Sciences, University of Reading, UK; **Gro Janne Wergeland**, MD, Department of Child and Adolescent Psychiatry, Haukeland University Hospital, Bergen, and Anxiety Disorders Research Network, Haukeland University Hospital, Norway; **Gerome Breen**, PhD, King's College London, IoPPN, MRC SGDP Centre, UK, and National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, UK; **Thalia C. Eley**, PhD, King's College London, IoPPN, MRC SGDP Centre, UK

Correspondence: Professor Thalia Eley, Social, Genetic and Developmental Psychiatry Centre (MRC), Institute of Psychiatry, Psychology and Neuroscience – PO80, DeCrespigny Park, Denmark Hill, London SE5 8AF, UK. Email: thalia.eley@kcl.ac.uk)

First received 14 Apr 2015, final revision 30 Sep 2015, accepted 14 Dec 2015

Funding

Combined study supported by UK Medical Research Council grant G0901874/1 to T.C.E. Individual trials support by Australian Research Council grant DP0878609 to J.L.H., R.M.R. and T.C.E.; Australian NHMRC grants to R.M.R., J.L.H. and H.J.L. (1027556), H.J.L, J.L.H and R.M.R (488505), and J.L.H and R.M.R. (382008); TrygFonden grant (7-10-1391) to M.T.; Edith og Godtfred Kirk Christiansens Fond grant (21-5675) to M.T.; Swiss National Science Foundation grant (105314-116517) to S.S., Western Norway Regional Health Authority grant to E.R.H. (911366); UK Medical Research Council Clinical Fellowship (G0802821) to R.M-S.; NIMH R01 (MH079943) to W.K.S.; UK NIHR grants to C. Creswell and P.C. (PB-PG-0110-21190) and P.C. and C. Creswell (PB-PG-0107-12042); UK Medical Research Council Grants to P.C. and C. Creswell (09-800-17), K.T., P.C. and C. Creswell (G0802326), P.W., C. Creswell and P.C. (G1002011), and C. Creswell (G0601874). Grant 09/800/17 was managed by National Institute for Health Research (NIHR) on behalf of the MRC-NIHR partnership. This study presents independent research part-funded by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Acknowledgements

The authors would like to thank: Irma Knuistingh Neven, Marianne Bjerregaard Madsen, Lisbeth Jørgensen, Carmen Adornetto, Kristine Fonnes Griffin, Krister Fjermestad, Elisa Kulewski, Jonas Großekathöfer, Nora Dirks, Evelyn Kmelnitski, Sabrina Heuser, Adrian Boyle, Clare Dixon, Harma Moorlag, Nienke Boersma, Sanne Hogendoorn, Sue Cruddace, Marie Weber, Zoe Hughes, Kiri Clarke, Francoise Hentges, Liz White, Nynke Wagenaar, Luci Motoca, Yesenia Rodriguez, Klaudia Perreira, Erin Hedemann, Cristina Del Busto, Gillian O'Neill, Lucy Lindley, Rachel Banham, Nora Delvendahl, Hjalti Jonsson, Judith Blatter-Meunier, Chantal Herren, Odd E. Havik, Karen Krause, Anna McKinnon, Patrick Smith and Yasmin Rey; and all participants in the study.

References

1 Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; **62**: 593–602.

2 Fineberg NA, Haddad PM, Carpenter L, Gannon B, Sharpe R, Young AH, et al. The size, burden and cost of disorders of the brain in the UK. *J Psychopharm* 2013; **27**: 761–70.

3 Baxter AJ, Vos T, Scott KM, Ferrari AJ, Whiteford HA. The global burden of anxiety disorders in 2010. *Psychol Med* 2014; **44**: 2363–74.

4 Bittner A, Egger HL, Erkanli A, Costello EJ, Foley DL, Angold A. What do childhood anxiety disorders predict? *J Child Psychol Psychiatry* 2007; **48**: 1174–83.

5 Gregory A, Caspi A, Moffitt T, Koenen K, Eley T, Poulton R. Juvenile mental health histories of adults with anxiety disorders. *Am J Psychiatry* 2007; **164**: 301–8.

6 National Collaborating Centre for Mental H. *Common Mental Health Disorders: Identification and Pathways to Care*. British Psychological Society & Royal College of Psychiatrists, 2011.

7 James AC, James G, Cowdrey FA, Soler A, Choke A. Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database Syst Rev* 2013; **6**: CD004690.

8 Eley TC, Hudson JL, Creswell C, Tropeano M, Lester KJ, Cooper P, et al. Therapygenetics: the 5HTTLPR and response to psychological therapy. *Mol Psychiatry* 2012; **17**: 236–7.

9 Lester KJ, Eley TC. Therapygenetics: using genetic markers to predict response to psychological treatment for mood and anxiety disorders. *Biol Mood Anxiety Disord* 2013; **3**: 1–16.

10 Lester KJ, Roberts S, Keers R, Coleman JRI, Breen G, Wong CCY, et al. Non-replication of the association between 5HTTLPR and response to psychological therapy for child anxiety disorders. *Br J Psychiatry* 2016, **208**: 182–8.

11 Uher R. The implications of gene-environment interactions in depression: will cause inform cure? *Mol Psychiatry* 2008; **13**: 1070–8.

12 Shimada-Sugimoto M, Otowa T, Hettema JM. Genetics of anxiety disorders: genetic epidemiological and molecular studies in humans. *Psychiatry Clin Neurosci* 2015; **69**: 388–401.

13 Erhardt A, Akula N, Schumacher J, Czamara D, Karbalai N, Muller-Myhsok B, et al. Replication and meta-analysis of TMEM132D gene variants in panic disorder. *Transl Psychiatry* 2012; **2**: e156.

14 Otowa T, Kawamura Y, Nishida N, Sugaya N, Koike A, Yoshida E, et al. Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. *Transl Psychiatry* 2012; **2**: e186.

15 Xie P, Kranzler HR, Yang C, Zhao H, Farrer LA, Gelernter J. Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. *Biol Psychiatry* 2013; **74**: 656–63.

16 Nievergelt CM, Maihofer AX, Mustapic M, Yurgil KA, Schork NJ, Miller MW, et al. Genomic predictors of combat stress vulnerability and resilience in U.S. Marines: a genome-wide association study across multiple ancestries implicates PRTFDC1 as a potential PTSD gene. *Psychoneuroendocrinology* 2015; **51**: 459–71.

17 Guffanti G, Ashley-Koch AE, Roberts AL, Garrett ME, Solovieff N, Ratanatharathorn A, et al. No association between RORA polymorphisms and PTSD in two independent samples. *Mol Psychiatry* 2014; **19**: 1056–7.

18 Hudson JL, Keers R, Roberts S, Coleman JRI, Breen G, Arendt K, et al. Clinical predictors of response to cognitive-behavioral therapy in pediatric anxiety disorders: the Genes for Treatment (GxT) Study. *J Am Acad Child Adolesc Psychiatry* 2015; 54: 454–63.

19 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (4th edn) (DSM-IV)*. APA, 1994.

20 Silverman WK, Albano AM. *Anxiety Disorders Interview Schedule (ADIS-IV) Child and Parent Interview Schedules: 5 of Each*. Graywind Publications, 1996.

21 Schneider S, Unnewehr S, Margraf J. *Diagnostisches interview psychischer Störungen im Kindes-und Jugendalter (Kinder-DIPS) [Diagnostic Interview for Mental Disorders in Childhood and Adolescence]*. Springer, 2009.

22 Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry* 2002; **47**: 262–6.

23 Uher R, Perroud N, Ng MY, Hauser J, Henigsberg N, Maier W, et al. Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. *Am J Psychiatry* 2010; **167**: 555–64.

24 Freeman B, Smith N, Curtis C, Huckett L, Mill J, Craig I. DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behav Genet* 2003; **33**: 67–72.

25 Gunderson KL, Steemers FJ, Ren H, Ng P, Zhou L, Tsan C, et al. Whole-genome genotyping. In *Methods Enzymol* (eds K Alan, O Brian): 359–76. Academic Press, 2006.

26 Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 2012; **28**: 2543–5.

27 The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.

28 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.

29 Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 2014; **46**: 100–6.

30 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.

31 Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; **19**: 149–50.

32 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.

33 Chang C, Chow C, Tellier L, Vattikuti S, Purcell S, Lee J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7.

34 Ripke S, Lewis CM, Lin D-Y, Wray N. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 2012; **18**: 497–511.

35 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–7.

36 Tansey KE, Guipponi M, Perroud N, Bondolfi G, Domenici E, Evans D, et al. Genetic predictors of response to serotonergic and noradrenergic antidepressants in major depressive disorder: a genome-wide analysis of individual-level data and a meta-analysis. *PLoS Med* 2012; **9**: e1001326.

37 Consortium IH. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52–8.

38 Visscher PM, Hemani G, Vinkhuyzen AAE, Chen G-B, Lee SH, Wray NR, et al. Statistical power to detect genetic (Co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* 2014; **10**: e1004269.

39 Hyman S. Mental health: depression needs large human-genetics studies. *Nature* 2014; **515**: 189–91.

40 Lonsdorf TB, Ruck C, Bergstrom J, Andersson G, Ohman A, Lindefors N, et al. The COMTval158met polymorphism is associated with symptom relief during exposure-based cognitive-behavioral treatment in panic disorder. *BMC Psychiatry* 2010; **10**: 9.

41 Bosker F, Hartman C, Nolte I, Prins B, Terpstra P, Posthuma D, et al. Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry* 2010; **16**: 516–32.

42 Ioannidis JPA. Genetic associations: false or true? *Trends Mol Med* 2003; **9**: 135–8.

43 Reif A, Richter J, Straube B, Höfler M, Lueken U, Gloster A, et al. MAOA and mechanisms of panic disorder revisited: from bench to molecular psychotherapy. *Mol Psychiatry* 2014; **19**: 122–8.

44 Ising M, Lucae S, Binder EB, Bettecken T, Uhr M, Ripke S, et al. A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch Gen Psychiatry* 2009; **66**: 966–75.

45 Tansey KE, Guipponi M, Hu X, Domenici E, Lewis G, Malafosse A, et al. Contribution of common genetic variants to antidepressant response. *Biol Psychiatry* 2013; **73**: 679–82.

46 Hudson JL, Kendall PC, Chu BC, Gosch E, Martin E, Taylor A, et al. Child involvement, alliance, and therapist flexibility: process variables in cognitive-behavioural therapy for anxiety disorders in childhood. *Behav Res Ther* 2014; **52**: 1–8.

47 Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013; **9**: e1003348.

48 Roberts S, Lester KJ, Hudson JL, Rapee RM, Creswell C, Cooper PJ, et al. Serotonin tranporter methylation and response to cognitive behaviour therapy in children with anxiety disorders. *Transl Psychiatry* 2014; **4**: e444.

49 Eley TC. The future of therapygenetics: where will studies predicting psychological treatment response from genomic markers lead? *Depress Anxiety* 2014; **31**: 617–20.

Addendum: GRMs and principal components

The analyses in Chapter 3 used a random effect derived from a genomic relatedness matrix (GRM) to control for population stratification in the multi-ethnic sample. The rationale behind this was that the multiple ethnicities in the sample resulted in principal components (PCs) that primarily separated samples of African or East Asian ancestry from those of White Western European (WWE) ancestry (Figures DS2a and DS2b). Accordingly, the WWE samples (which represent 92% of the cohort) were poorly separated by PCs. One approach to addressing this would have been to remove samples that lay outside the WWE cluster on the PC plot (red box in Figure DS2b), and re-run PC analysis to obtain WWE-specific components for use in the final analysis. Controlling for gross genetic similarity between samples using the GRM offered a means to address better both fine-scale and broad-scale inter-sample relatedness (compared to PC-based control; Wang, Hu & Peng, 2013) whilst retaining sample size (and so statistical power).

The use of GRMs to control for population stratification has become much more common since the analysis in Chapter 3 was performed, and the validity of using GRMs in the presence of population stratification has come under greater scrutiny. Although GRM-based control is generally robust in the presence of population stratification, it fails in the presence of extreme differences in allele frequency between populations (Price, Zaitlen, Reich *et*

al, 2010). As such, this method may not appropriately control for population stratification in this instance. The robustness of GRM-based correction in the case of unbalanced population stratification (that is, the situation in Chapter 3) has not been explicitly tested. As such, it may be the case that the results of Chapter 3 are limited by inadequate control for population stratification, and that outlier removal would have been a better strategy for analysis in this instance. Accordingly, I performed sensitivity analyses, limiting the cohort to WWE samples only. This yields results that are largely concordant with those reported - 5/7 loci have $p<10^{-5}$ in pre to post-treatment analyses in the full cohort and limited to the WWE samples only, and 6/7 loci have $p<10^{-5}$ in the analyses to follow-up. All loci reported in Tables 1 and 2 have $p < 5x10^{-5}$ in the WWE analysis. The conclusions of the analysis are not substantially biased due to the method used, although the precision of the associations differs when considering the WWE samples only. In the analysis limited to WWE samples, as in the analyses in the full sample, there are no variants of large effect, and there are a number of loci at a suggestive level of significance, although any true signal cannot be disentangled from associations due to chance. Loci with $p<1x10^{-4}$ in the sensitivity analyses are provided in Tables DS3a and DS3b, in the Addendum to Appendix II.

Chapter 4:

THE WORLD JOURNAL OF BIOLOGICAL PSYCHIATRY, 2016
http://dx.doi.org/10.1080/15622975.2016.1208841

ORIGINAL INVESTIGATION

Separate and combined effects of genetic variants and pre-treatment whole blood gene expression on response to exposure-based cognitive behavioural therapy for anxiety disorders

Jonathan R.I. Coleman[a], Kathryn J. Lester[a,b], Susanna Roberts[a], Robert Keers[a,c], Sang hyuck Lee[a,d], Simone De Jong[a,d], Héléna Gaspar[a,d], Tobias Teismann[e], André Wannemüller[e,f], Silvia Schneider[e], Peter Jöhren[f], Jürgen Margraf[e], Gerome Breen[a,d]* and Thalia C. Eley[a,d]*

[a]King's College London, Institute of Psychiatry, Psychology and Neuroscience, MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, London, UK; [b]School of Psychology, University of Sussex, UK; [c]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK; [d]National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, London, UK; [e]Mental Health Research and Treatment Center, Ruhr-Universität Bochum, Bochum, Germany; [f]Dental Clinic Bochum, Bochum, Germany

ABSTRACT

Objectives: Exposure-based cognitive behavioural therapy (eCBT) is an effective treatment for anxiety disorders. Response varies between individuals. Gene expression integrates genetic and environmental influences. We analysed the effect of gene expression and genetic markers separately and together on treatment response.

Methods: Adult participants ($n \leq 181$) diagnosed with panic disorder or a specific phobia underwent eCBT as part of standard care. Percentage decrease in the Clinical Global Impression severity rating was assessed across treatment, and between baseline and a 6-month follow-up. Associations with treatment response were assessed using expression data from 3,233 probes, and expression profiles clustered in a data- and literature-driven manner. A total of 3,343,497 genetic variants were used to predict treatment response alone and combined in polygenic risk scores. Genotype and expression data were combined in expression quantitative trait loci (eQTL) analyses.

Results: Expression levels were not associated with either treatment phenotype in any analysis. A total of 1,492 eQTLs were identified with $q < 0.05$, but interactions between genetic variants and treatment response did not affect expression levels significantly. Genetic variants did not significantly predict treatment response alone or in polygenic risk scores.

Conclusions: We assessed gene expression alone and alongside genetic variants. No associations with treatment outcome were identified. Future studies require larger sample sizes to discover associations.

ARTICLE HISTORY
Received 19 February 2016
Revised 27 May 2016
Accepted 28 June 2016

KEYWORDS
Genetics; psychotherapy; anxiety disorders; gene expression; exposure therapy

Introduction

Anxiety disorders are the most common group of mental illnesses, with lifetime prevalence estimates ranging between 10 and 30% (Kessler et al. 2007; Michael et al. 2007). They are an economic burden on society and the sixth largest cause of disability globally (Fineberg et al. 2013; Baxter et al. 2014). Suffering from an anxiety disorder is distressing, with affected individuals reporting adverse effects on quality of life comparable to sufferers of major depressive disorder, and in excess of the population norm (Mendlowicz and Stein 2000).

Treatment of anxiety disorders uses a variety of pharmacological and psychological modalities (National Collaborating Centre for Mental Health 2011). Exposure-based cognitive behavioural therapy (eCBT) is a common treatment, and shows large effect sizes across the anxiety disorders, comparable to or better than those obtained by anxiolytic medication (Norton and Price 2007; Stewart and Chambless 2009; Barlow et al. 2013; Cuijpers et al. 2013; Margraf and Zlomuzica 2015). During eCBT, participants confront the object of their anxiety (whether literally, referred to as in vivo, or through imagination or virtual reality, referred to as in sensu), within a carefully managed and supportive environment. They identify the cognitive and behavioural processes underlying their anxious response, and develop strategies to mitigate against these negative schema and to cope with their anxiety (Otto et al.

2004). Rates of response (in terms of a reduction in symptom severity) and of remission (no longer meeting diagnostic criteria) vary between specific disorders and studies, but are invariably less than 100% (Ballenger 1999; Olatunji et al. 2010; Hofmann et al. 2012; Loerinc et al. 2015).

Numerous influences have been proposed to lead to poorer treatment outcome, including high initial severity, Axis I and Axis II comorbidity, illness duration, low expectancy of treatment success, poor treatment compliance and therapeutic alliance, and general interpersonal difficulties (Newman et al. 2013). However, studies disagree on the importance and validity of such predictors (Taylor et al. 2012; Olatunji et al. 2013; Schneider et al. 2015). The success of any specific treatment for a given participant is difficult to predict. This is relevant given the high costs (both economic and emotional) of pursuing unsuccessful treatment (Otto et al. 2000). It is of clear interest to develop reliable predictors of treatment response.

Genetic variants represent a potential source of predictors. The study of such variants (termed therapygenetics) has largely been confined to candidate gene studies (Eley et al. 2012; Lester and Eley 2013). However, these findings have proven difficult to replicate, and the direction of effect found has been inconsistent between studies (Lester et al. 2016). Recently, we published a genome-wide association study (GWAS) of therapy response in a cohort of children with anxiety disorders (Coleman et al. 2016). Although underpowered to identify the small-effect variants typical of behavioural phenotypes, sufficient power was available to test some effect sizes reported in the therapygenetics literature. No variants were found at conventional genome-wide significance, and candidate variants were not replicated. Therefore, the effects of individual genetic variants on response to CBT are likely to be small, and the predictive effects of such variants are likely to be negligible when used alone.

Studying the differential expression of gene transcripts may be more useful for predicting treatment response. Multiple factors affect gene expression, potentially including genetic variants and environmental influences. Gene expression represents a biologically relevant means of combining genetic and environmental variation to predict response to CBT for anxiety disorders. Two studies have found an association between increased expression of FKBP5 and response to CBT for post-traumatic stress disorder (PTSD; Levy-Gigi et al. 2013; Yehuda et al. 2013). A recent analysis of change across treatment including a subset of the cohort presented within this paper showed no association between treatment response and individual gene expression, nor when expression was clustered according to similarities in expression in the data (Roberts et al. under review).

This investigation combines genetic and gene-expression approaches to predict response to eCBT. It assesses the interaction of differential gene expression at baseline (both of individual transcripts, and using data- and literature-driven clustering methods) and genetic variation to assess the outcome of eCBT for panic disorder (PD) and specific phobias (SPs).

Method

Participants and therapeutic procedure

Two hundred and forty-four participants diagnosed with PD or a SP completed one of four eCBT treatment programmes at the Mental Health Research and Treatment Center, Ruhr-Universität Bochum, Germany as part of standard care. In all programmes, diagnoses were made according to DSM-IV criteria using the Diagnostisches Interview bei psychischen Störungen (DIPS) and Mini-DIPS, structured interviews with well-established reliability, validity and patient acceptance (Margraf 1994; In-Albon et al. 2008; Suppiger et al. 2008; Suppiger et al. 2009; Bruchmuller et al. 2011; Schneider and Margraf 2011). All treatment programmes featured core elements of exposure therapy, including psychoeducation, applied relaxation and exposure (in vivo or in sensu). Specifics of each treatment programme are described below. All treatments were regularly supervised by experienced senior clinicians using audio-visual recordings in order to ensure treatment protocol integrity.

Individuals diagnosed with a SP of receiving dental treatment, not secondary to a separate diagnosis (such as PTSD or injection phobia), were treated in a dental anxiety-specific (DA) programme (Wannemuller and Johren 2015). Treatment was given in five weekly sessions comprising an initial diagnostic and psychoeducation session, a session developing relaxation techniques, and three in sensu exposure sessions related to dental treatment. Participants were not excluded on the basis of concurrent treatment with anxiolytic medication.

Participants with a SP not primarily associated with dental fear were treated in a longer-term programme covering up to 30 sessions, split into five initial sessions of diagnosis and psychoeducation, and 25 sessions of in vivo exposure (relevant to their SP) with elements of cognitive restructuring. Participants were excluded from the study if they were using anxiolytic medication.

Participants with a primary diagnosis of PD with agoraphobia, or agoraphobia alone, were randomised

either to eCBT (PD-CBT; akin to the SP group) or to an exposure-alone condition without any element of cognitive restructuring (PD-exposure [EXP]; Clinical Trials: NCT01680327). Participants in both conditions were excluded if they were using anxiolytic medication. Bodily sensation was used as the specific exposure stimulus for participants suffering from PD. However, as there were no patients with PD without agoraphobia in the trial, interoceptive exposure was always combined with in vivo exposure.

Prior to receiving exposure, immediately following completion of the treatment programme, and at a follow-up assessment approximately 6 months after treatment completion, all participants completed a range of questionnaire measures. In addition, peripheral blood was drawn for DNA and RNA extraction.

Phenotype definition

Treatment response was defined as percentage improvement in the clinician-rated severity scale of the Clinical Global Impression-Severity (CGI-S) rating, and was examined pre-treatment to post-treatment, and pre-treatment to follow-up. The CGI-S ranges from 1 to 7, with a score of 1 representing no symptoms of concern and a score of 7 representing extremely severe illness requiring hospitalisation (Guy 1976). The scale was chosen as it was used in all treatment groups, and was expected to capture severity in a disorder-independent fashion.

The CGI-S was rescaled to range from 0 to 6 to allow outcome to be defined as percentage decrease in severity across time (as this has previously been used successfully in pharmacogenetic GWAS) with 100% indicating full remission (Uher et al. 2010). For both phenotypes, correlations were calculated between percentage improvement and a variety of covariates: age, gender, severity at baseline, presence of comorbid mental disorders, number of treatment sessions attended, treatment period (days between pre-treatment and post-treatment assessment), follow-up period (days between post-treatment assessment and follow-up assessment), use of psychoactive medication at pre-treatment, use of any other medication pre-treatment, body mass index (BMI) and whether the participant smoked. Although the use of concurrent anxiolytic medication was an exclusion criterion for the SP, PD-CBT and PD-EXP groups, some participants were using other medications which may have a psychotropic effect, so this covariate was not restricted to the DA group (Table 1).

Of these covariates, severity at baseline, presence of comorbid mental disorders, use of psychoactive medication and follow-up period were correlated with at least one phenotype in the whole cohort (Table 2). In secondary examinations within each treatment group, treatment period was associated with at least one phenotype in both the PD-CBT ($P = 0.014$) and SP groups ($P = 0.012$). BMI was weakly associated ($P = 0.0424$) with response at post-treatment in the PD-CBT group; however, as this effect was not seen in any other group nor in the whole cohort, BMI was not included as a covariate.

Table 2. Correlations between clinical covariates and treatment response phenotypes for the whole cohort ($N = 187$).

	Post-treatment		Follow-up	
Variable	*r*	*P*	*r*	*P*
Age (years)	−0.0536	0.469	0.0212	0.816
Gender	0.045	0.543	0.0497	0.585
Baseline CGI severity	0.112	0.128	0.303	**$7x10^{-4}$**
Treatment sessions	0.129	0.0818	0.106	0.255
Treatment duration (days)	0.117	0.111	0.0189	0.835
Follow-up duration (days)	−0.178	**0.0409**	−0.0411	0.6614
Psychoactive medication (use)	−0.205	**0.00521**	−0.203	**0.0243**
Other medication (use)	−0.0483	0.514	−0.117	0.200
Mental comorbidity (yes/no)	−0.210	**0.00407**	−0.188	**0.0379**
Body mass index	0.0572	0.440	0.0008	0.993
Smoker (yes/no)	−0.0317	0.687	0.0146	0.880

Correlations with nominal significance ($P < 0.05$) are highlighted in bold.

Table 1. Demographic and treatment information on participants with genotype and/or expression data.

Variable	WC	DA	SP	PD-CBT	PD-EXP	Test	Stat	*P*
N	187	95	38	25	29	–	–	–
Age in years (Mean [SD])	39.2 [11.4]	40.5 [10.4]	37.8 [13.2]	38.4 [11.9]	37.4 [11.9]	ANOVA	0.831	0.478
Gender (N male [%])	67 [35.8]	35 [36.8]	9 [23.7]	13 [52.0]	10 [34.5]	Chi square	5.35	0.148
Baseline CGI-S (Mean [SD])	4.70 [1.13]	4.83 [1.27]	4.16 [0.973]	4.80 [0.707]	4.86 [0.915]	ANOVA	3.80	**0.0112**[a]
Treatment duration in days (Mean [SD])	200 [184]	47.6 [32.8]	340 [152]	351 [118]	383 [137]	ANOVA	151	**3.15×10^{-49}**[b]
Follow-up duration in days (Mean [SD])	215 [62.7]	249 [72]	191 [41.4]	190 [35.4]	191 [44.7]	ANOVA	11.4	**4.32×10^{-7}**[c]
Psychoactive medication at baseline (*N* taking [%])	20 [10.7]	18 [18.9]	1 [2.63]	1 [4.00]	0 [0.00]	Fisher's exact test		**0.00247**[d]
Mental disorder comorbidities (*N* [%])	72 [38.5]	46 [48.4]	8 [21.1]	9 [36.0]	9 [31.0]	Chi square	9.58	**0.0225**[e]

Post hoc *t*-tests (variances assumed unequal; Bonferroni corrected significance threshold = 0.00834; significant results in bold).
[a]SP lower: vs. DA: $t = -3.29$, $P = 0.00143$; vs. PD-CBT: $t = -3.03$, $P = 0.00361$; vs. PD-EXP: $t = -3.04$, $P = 0.00350$.
[b]DA shorter: vs. SP: $t = -11.7$, $P = 2.79 \times 10^{-14}$; vs. PD-CBT: $t = -12.8$, $P = 1.93 \times 10^{-12}$; vs. PD-EXP: $t = -13.0$, $P = 1.19 \times 10^{-13}$.
[c]DA longer: vs. SP: $t = 4.96$, $P = 3.19 \times 10^{-6}$; vs. PD-CBT: $t = 4.86$, $P = 6.93 \times 10^{-6}$; vs. PD-EXP: $t = 4.48$, $P = 2.74 \times 10^{-5}$.
[d]Higher rate in DA: vs. SP: $t = 3.38$, $P = 9.47 \times 10^{-4}$; vs. PD-CBT: $t = 2.63$, $P = 0.0103$; vs. PD-EXP: $t = 4.69$, $P = 9.33 \times 10^{-6}$.
[e]Higher rate in DA: vs. SP: $t = 3.24$, $P = 0.00174$; vs. PD-CBT: $t = 1.12$, $P = 0.269$; vs. PD-EXP: $t = 1.71$, $P = 0.0930$.
WC, whole cohort; DA, dental anxiety; SP, specific phobia; PD-CBT, panic disorder CBT; PD-EXP, panic disorder exposure.

The phenotypes for analysis were defined as the residuals from two linear mixed regressions investigating change in severity between pre-treatment and post-treatment, and pre-treatment and follow-up. Percentage decrease in severity was regressed on fixed effects of baseline severity, presence of comorbid mental disorders, use of psychoactive medication, treatment period (and follow-up period in the analysis pre-treatment to follow-up), and a higher-order random effect of treatment group (to account for differences between treatment groups).

Genotyping

DNA was extracted from peripheral blood drawn pre-treatment using FlexiGene DNA Kits, following the protocol provided by the manufacturer (QIAGEN, Manchester, UK). DNA concentration was quantified using spectrophotometry (NanoDrop 1000, NanoDrop, Wilmington, DE, USA), and samples diluted to 40 μl at a concentration of 75 ng/μl for genotyping. Genotyping was performed using the Illumina PsychChip microarray (Illumina, San Diego, CA, USA), a modified version of the Illumina HumanCoreExome microarray with additional content of interest in psychiatric genomics. All laboratory procedures were performed at the Institute of Psychiatry, Psychology and Neuroscience, King's College London.

Genotype quality control

Quality control was performed following a previously published protocol (Coleman et al. 2015). In brief, genotype data were called using Illumina GenomeStudio software, with manual recalling where appropriate. Rare variants were recalled using ZCall (Goldstein et al. 2012). Variants were removed from the analysis if they were rare (minor allele frequency <0.05), present in $<99\%$ of individuals, or deviated substantially from Hardy–Weinberg equilibrium (Hardy–Weinberg test $P < 1 \times 10^{-5}$). Individuals were excluded if they had genotype calls for $<99\%$ of variants, where reported gender differed from that indicated by the genotypes, or if genome-wide estimates of heterozygosity >3 standard deviations from the sample mean. Additional exclusions were made if the individual showed cryptic relatedness to other individuals in the study (identity by descent [IBD] > 0.1875) or had an average proportion of variants shared IBD with the cohort as a whole >6 standard deviations above the cohort mean.

Following quality control, variants were imputed to the Phase 3 release from the 1000 Genomes Project, using IMPUTE2 with concurrent phasing (1000GenomesConsortium 2012; Howie et al. 2012). X chromosome variants were imputed using the March 2012 Phase 1 release (1000GenomesConsortium 2012; Howie et al. 2012). Imputed variants were imported into PLINK2 for analysis, and filtered to remove uncertain variants (posterior-probability <0.8) and poorly imputed variants (info <0.8) (Chang et al. 2015). Following hard-calling, variants present in $<98\%$ of the cohort were dropped from analysis.

Gene expression

Whole blood samples were drawn at pre-treatment using PAXgene blood RNA tubes. Blood RNA was isolated and purified using the PAXgene Blood miRNA Kit according to the manufacturer's protocol using the QIAcube (QIAGEN). RNA quality was measured using spectrophotometry (NanoDrop 1000, NanoDrop) and integrity using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Genome-wide expression levels were measured from 750 ng total RNA using the Illumina HumanHT-12v4 Expression BeadChip (Illumina).

Gene expression quality control

Raw expression data were processed following internal pipelines (available at https://github.com/snewhouse/BRC_MH_Bioinformatics). Samples with detection rates dissimilar from the rest of the cohort were identified and removed in GenomeStudio (Illumina). Raw data were imported into R for quality control primarily using the lumi package (Du et al. 2008; Team 2012). Expression data were background corrected using module-based background correction for BeadArrays (Ding et al. 2008). Probes with an expression level >2 standard deviations above the background mean were defined as detected. *XIST* gene expression (specific to females) and Y chromosome gene expression (specific to males) was compared to reported gender and gender inferred from genotyping, and discordant samples removed. Expression data were log2 transformed and normalised using robust splines normalisation from the lumi package (Du et al. 2008; Schmid et al. 2010). Sample co-expression relationships were assessed, and samples with connectivity <2 standard deviations from the cohort mean were excluded (Oldham et al. 2012). Associations between covariates and the first principal component of the expression data were assessed using stepwise linear regression bootstrapped 100 times, with randomised order of covariates in the regression. Covariates included batch variables (expression microarray, sample position on microarray, date of RNA extraction, date of expression measurement, machine used in RNA isolation, RNA integrity (RIN)

value, RNA yield, amplified concentration of RNA, whether the sample required additional treatment to remove DNA and whether the blood sample was the first or second drawn) and demographic covariates (BMI and smoking). The effect of associated covariates was regressed out of the expression data, using the sva ComBat package in the case of categorical variables (extraction date) and linear regression (with RcppArmadillo) in the case of continuous variables (RIN value, RNA yield, amplified concentration of RNA) (Johnson et al. 2007; Eddelbuettel and Sanderson 2014). Probes detected in $<80\%$ of the sample were removed. As expression data were generated from whole blood without assessment of cellular composition, deconvolution methods implemented in CellMix were used to assess the origin of RNA transcripts before and after differentially expressed probes were selected (Gaujoux and Seoighe 2013). Correlations between the estimated final proportions of leukocytes (neutrophils, lymphocytes and monocytes) and the two CBT response phenotypes were calculated. Additional exclusion of probes was performed to allow combined analysis with genotyped variants. Specifically, probes were excluded if they were not annotated in the ENSEMBL hg19 build, if they contained any genetic variant genotyped in the cohort, or if they did not map to a unique site on the genome. Probes were identified using nucleotide universal identifiers (nuIDs), which are unique to the DNA sequence of the probe (Du et al. 2007).

Statistical analysis

Following quality control, the association of genome-wide genotyping data with both response phenotypes was assessed in GWAS. The participants in the study were of Central or Eastern European ancestry. Genomic estimation of ancestry was established using principal components analysis performed in EIGENSOFT (Price et al. 2006). No principal component was correlated with either of the phenotypes at a level greater than chance. To account for finer-scale population stratification, analyses were run using a linear mixed model incorporating a random effect of gross genetic similarity between individuals (the *mlma-loco* option in GCTA; Yang et al. 2011). Results were clumped in PLINK2, pruning all variants in linkage disequilibrium ($r^2 > 0.25$, $\pm$ 250 kb) of a variant with a lower *P* value.

Genotype information was used as a target dataset in polygenic risk scoring. Specifically, the results of a previous GWAS of CBT response in children were used to predict both phenotypes in the whole cohort using PRSice, which performs high-resolution polygenic risk scoring to identify the most predictive risk score (Euesden et al. 2015; Coleman et al. 2016). Further GWAS were performed on the cohort minus individuals treated for DA, and the results from these subset GWAS were used to predict response to treatment in the DA subgroup.

Probe-level expression data were imported into R, and analysed using weighted gene correlation network analysis (WGCNA; Langfelder and Horvath 2008). Data-driven clustering of co-expressed probes was performed using an automatically-constructed signed network from the *blockwiseModules* function in WGCNA (details on this procedure are provided in the Supplemental Material available online; Langfelder and Horvath 2008). Correlations between individual probes and both response phenotypes, and between WGCNA module eigengenes and response phenotypes, were calculated. Local false discovery rates were calculated to account for multiple testing using the qvalue package in R (Dabney et al. 2004).

Probe-level correlations were mapped to HUGO gene names, ranked according to significance and used in gene ontology (GO) enrichment analysis in GOrilla (Eden et al. 2009). Where multiple probes mapped to the same gene, the highest-ranked was retained. Details of the enrichment analysis performed by GOrilla are provided in the Supplemental Material. Significance was set as the Bonferroni correction for the 8746 GO terms tested ($P = 5.72 \times 10^{-6}$), with results reported below $P = 5 \times 10^{-4}$. Results were pruned for redundancy in REViGO, with results with $>50\%$ dispensability dropped (Supek et al. 2011).

Probe-level expression data were combined using a machine-learning approach in WEKA, to assess the viability of prediction from expression probe data alone (Hall et al. 2009). Classical machine learning algorithms were used to predict outcome using the full dataset (3,233 expression probes, 166 participants for baseline to post-treatment analysis, 110 participants for baseline to follow-up analysis). Five approaches were used: mean prediction with ZeroR; inverse distance weighting with a nearest neighbours algorithm (kNN), with and without subset evaluation; linear kernel-based regression with regression SVM (SMOReg) and a 500-tree Random Forest algorithm. Multiple algorithms were chosen as they optimise different aspects of the learning process. All analyses were performed using 10-fold cross-validation (splitting the cohort into 80% training and 20% test subsamples), repeated 5 times.

Probe-level expression data and genotype data were imported into R for eQTL analyses using the MatrixEQTL package (Shabalin 2012). All transcripts

captured by the assessed probes were mapped to the hg19 build of the human genome. Analyses were performed using a two-stage design. C*is*-eQTLs were calculated independent of the phenotype, using the *modelLINEAR* option and genotypes from a window ± 100 kb of the transcript. Linkage-independent results were obtained by clumping using PLINK2 (250 kb window, $r^2 < 0.25$), and by performing conditional eQTL analyses (Chang et al. 2015). Both techniques identified the same sentinel SNPs (data not shown). Clumped results were retained for the second, phenotype-dependent, stage. SNP-by-treatment response interactions predicting expression change were assessed for both phenotypes, using the *modelLINEAR_CROSS* option in MatrixEQTL, to investigate whether the effect of eQTLs in the data differed in relation to treatment response.

Power analyses for the expression analyses were performed using the pwr package in R.

Ethics

Ethics approval for this study was received from the Ethics Committee at the Faculty of Psychology, Ruhr-Universität Bochum, from the London-Bentham NRES Committee and from the King's College London Psychiatry, Nursing and Midwifery Research Ethics Sub-Committee. All participants provided informed consent. The study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Results

Phenotype data were available on 187 participants (185 at post-treatment; 122 at follow-up). Following quality control, genotype data were available on 3,343,497 variants (267,037 genotyped) for 181 participants for the post-treatment analysis (122 were available for the analysis at follow-up). Data from 3,233 expression probes were available on 166 (110) participants. Both data types were available on 162 (110) participants.

Demographics and clinical covariates

Demographic data on the cohort are displayed in Table 1. Individuals in the SP group had lower baseline severity than all other groups. Groups also differed by mental disorder comorbidity, with individuals in the DA group exhibiting more comorbidity than other groups, and significantly more than the SP group (details of mental comorbidities are provided in Supplemental Table 1, available online). As expected, there was a higher rate of psychoactive medication use in the DA group compared to all others. Treatment duration also differed significantly across the groups, with shorter treatment in the DA group than in all others. Follow-up duration was significantly longer in the DA group. All covariates showing inter-group differences were included as covariates when defining the treatment response phenotypes (as was a random effect of treatment group; Table 2).

Changes in CGI from pre-treatment to post-treatment and to follow-up are described in Table 3. All treatments were generally effective, with most participants improving on the CGI-S between pre-treatment and post-treatment, and between pre-treatment and follow-up. However, there was considerable variance in the percentage change shown between individuals. Demographic differences between response groups following treatment are described for the whole cohort in Supplemental Table 2 (available online). Significantly lower baseline severity and higher comorbidity was observed in those deteriorating compared to those improving. No other significant differences were observed.

GWAS and polygenic risk score analysis

Results from both GWAS are shown in Supplemental Table 3 and Supplemental Figures 1 and 2 (available online). No variants passed the threshold for genome-wide significance ($P = 5 \times 10^{-8}$), but three independent

Table 3. Treatment response as percentage change in CGI-S, and grouped by improvement (percentage change positive), no change, and deterioration (percentage change negative), in the whole cohort and each treatment.

	Whole cohort	DA	SP	PD-CBT	PD-EXP	Test	Stat	*P*
Response post-treatment								
N	185	95	37	24	29	–	–	–
% change in CGI (Mean [SD])	67.4 [34.6]	62.0 [32.1]	77 [39.5]	67.1 [37.1]	72.9 [31.8]	ANOVA	2.01	0.115
Improved (*N* [%])	165 [89.2]	84 [88.4]	33 [89.2]	21 [87.5]	27 [93.1]	Fisher's exact test		0.613
No change (*N* [%])	16 [8.65]	10 [10.5]	2 [5.41]	2 [8.33]	2 [6.90]			
Deteriorated (*N* [%])	4 [2.16]	1 [1.05]	2 [5.41]	1 [4.17]	0 [0.00]			
Response at follow-up								
N	122	54	32	17	19	–	–	
% change in CGI (Mean [SD])	59.7 [46.2]	52.3 [54.5]	71.4 [37.9]	71.5 [30.7]	49.8 [41.2]	ANOVA	1.90	0.134
Improved (*N* [%])	101 [82.8]	42 [77.8]	28 [87.5]	16 [94.1]	15 [78.9]	Fisher's exact test		0.641
No change (*N* [%])	11 [9.02]	5 [9.26]	3 [9.38]	1 [5.88]	2 [10.5]			
Deteriorated (*N* [%])	10 [8.20]	7 [13.0]	1 [3.13]	0 [0.00]	2 [10.5]			

Groups did not differ on treatment response by either measure.

loci in the analysis to post-treatment and four loci in the analysis to follow-up reached a suggestive level of significance ($P < 5 \times 10^{-6}$). Quantile-quantile plots indicated no substantial genomic inflation in either analysis.

Polygenic risk score analysis from an independent GWAS of response to CBT in children failed to predict response in the whole cohort with $P < 0.001$ (Supplemental Table 4a, available online, threshold adjusted for multiple testing; Euesden et al. 2015). Prediction between the DA treatment group and all other treatment groups explained more variance in outcome than the analysis using the independent GWAS, but predictors were not significant (Supplemental Table 4b, available online). Further discussion of the GWAS and PRSice analyses are included in the Supplemental Material available online.

Individual expression probes

No probes were significantly associated with either phenotype after correcting for multiple testing (all $q > 0.05$; Table 4). The probes with the lowest q values in this analysis showed no overlap with those reported in a parallel analysis of this cohort, examining change in expression over the course of treatment (Roberts et al. under review).

Power analyses indicated the analyses have 80% power to detect associations capturing at least 14.8% (post-treatment), and 19.6% (follow-up) of variance respectively, where $\alpha = 1.55 \times 10^{-5}$ (Bonferroni correction for 3,233 tests).

Data-driven network-based analyses

Clustering by co-expression patterns yielded eight network modules ranging from 750 to 63 probes and a further "grey" module of 459 probes that did not fall into any cluster. Although different clusters showed associations with a variety of sample characteristics, no cluster was associated with either treatment response phenotype (all $P > 0.05$; Figure 1).

Literature-driven GO analysis

HUGO gene names were assigned to 2,652 probes associated with at least one GO term (process, function or component). No significant pathways were found after correction for multiple testing (all $P > 5.72 \times 10^{-6}$). Following removal of redundant GO terms, five processes and one function were associated with $P < 5 \times 10^{-4}$ in the analysis from baseline to post-treatment. From baseline to follow-up, eight processes and two functions were associated with $P < 5 \times 10^{-4}$ (Supplemental Table 5, available online).

Classical machine learning analyses

Classical machine learning methods did not outperform the null model in either analysis. The most effective model was random forest classification (root-mean-square error [RMSE]: 31.3, post-treatment; 42.7, follow-up) but this did not outperform ZeroR, which predicts the mean (RMSE: 30.6, post-treatment; 42.3, follow-up).

Expression quantitative trait loci

Expression quantitative trait loci (eQTL) analysis identified 42,868 *cis*-eQTLs with $q < 0.05$, independent of phenotype. Following the removal of variants in linkage disequilibrium with more strongly associated eQTLs, 1,492 variants were present with $q < 0.05$ (Table 5, Supplemental Table 5, available online). Phenotype-dependent analyses of the interaction between these variants and treatment response predicting expression levels yielded no associations with $q < 0.05$ (Table 6). One interaction was identified with $q < 0.2$ (rs10498246 × treatment response baseline to follow-up, predicting *SP110* (probe nuID:

Table 4. Largest correlations between individual expression probes and the treatment response phenotypes.

Associations between expression probes and treatment outcome					
Probe nuID	Gene	WGCNA module	Pearson's *r*	*P*	*q*
Baseline – post-treatment					
TkiT0uUa.K4LZ5M7h4	*FDFT1*	blue	0.282	2.34×10^{-4}	0.756
0Z7unqF.KAuA5K4ggU	*FDFT1*	grey	0.241	0.00175	1
Eqx.SxEEVcI.VLrWJI	*IL18RAP*	grey	0.237	0.00211	1
Te4VV0giY1VcQvr17E	*RNASE6*	grey	−0.216	0.00515	1
QuynqD354KD6IAXvnk	*YIPF4*	grey	−0.214	0.00550	1
Baseline – follow-up					
TXm4UjVovoAQ4ApVQo	*MYC*	grey	−0.346	2.17×10^{-4}	0.702
Krrborr9LqDhB.rPoo	*HNRNPA1P33*	brown	−0.294	0.00180	1
Ew_iK7UunWqIb0nFeE	*AIF1*	grey	0.265	0.00518	1
6dFQSN.UitTroIYwV4	*MAL*	grey	−0.240	0.0115	1
T0upGOh1A5dC87MXtU	*PPP6C*	turquoise	0.235	0.0136	1

WGCNA modules refer to the data-driven clusters to which each probe belongs.

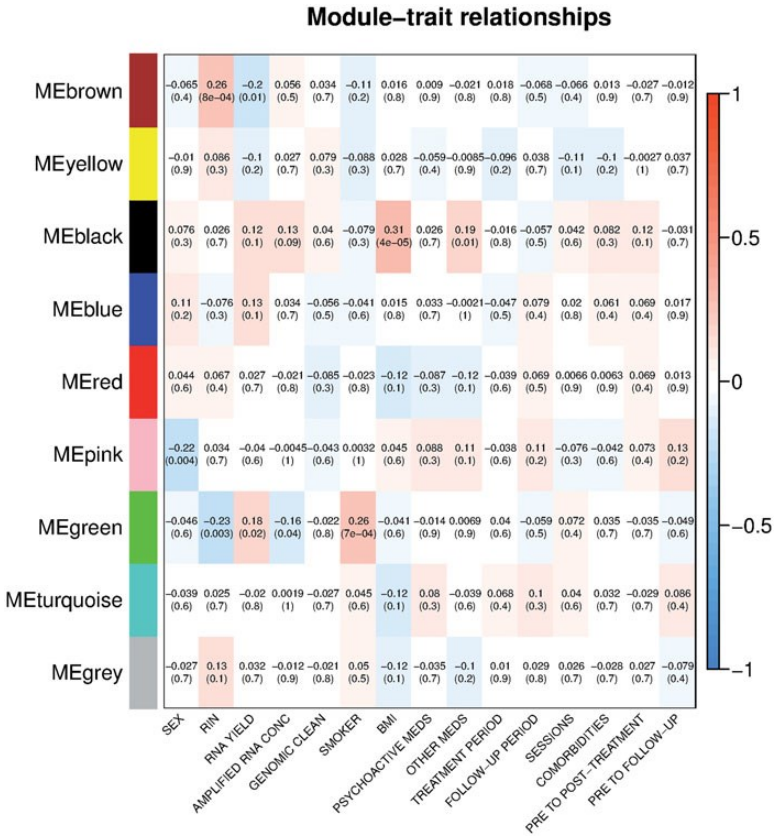


Figure 1. Correlations between expression profiles of module eigengenes from WGCNA and treatment phenotypes (and covariates). Positive correlations are shown in red, negative correlations in blue, with colour intensity indicating strength of correlation. No module expression profile is associated with a treatment response phenotype (all $P > 0.05$).

Table 5. Raw and clumped results from the expression QTL analysis, by false discovery rate.

Linkage-dependent and -independent blood eQTLs		
FDR q threshold	No. of variants	No. of linkage-independent variants
0.01	26,566	788
0.05	42,868	1492
0.1	54,795	2159
0.5	61,799	2503

fcV3S0U75lf1e3op0U) expression, $B = -0.0041$, $P = 2.23 \times 10^{-5}$, $q = 0.103$).

Discussion

We performed genome-wide analysis of genetic variation and pre-treatment gene expression to assess independent and combined effects on response to CBT for anxiety disorders in a cohort of adult participants. This is the first analysis to integrate this data in studying psychological treatment response, and (together with a companion paper; Roberts et al. under review) is an analysis of the largest psychological treatment cohort in which gene expression analyses have been performed. Despite this, no variants or expression profiles were associated (at a genome-wide level of significance) with treatment response across the treatment period or at a 6-month follow-up.

The cohort is larger than previous studies of the effect of gene expression on response to CBT in anxiety disorder (Levy-Gigi et al. 2013; Yehuda et al. 2013). However, it is clear that this study is underpowered to detect all but the largest effects on response, and that robust prediction requires larger cohort sizes. Integrating data from two different approaches (that is, genotyping and gene expression) increases power, but requires two sets of quality control, resulting in fewer samples with full data available (Ritchie et al. 2015). Obtaining a large sample size for a study such as this is non-trivial. Prospective recruitment results in a high rate of attrition as participants withdraw from treatment or are lost to follow-up. Furthermore, this attrition is likely to be related to poor treatment response.

Table 6. Top six results from eQTL-outcome interactions predicting expression level.

eQTL x treatment response predicting gene expression					
SNP	Probe nuID	Gene	β	*P*	*q*
Baseline – post-treatment					
rs11260538	94gYDdn0tHeWCmeGk0	*SDF4*	0.00394	7.86×10^{-4}	1
rs3129996	llGfH57t5ug93Xe1XU	*KIAA1949*	−0.00431	0.00126	1
rs16965033	onsnvop.hKDoejReHU	*HERPUD1*	−0.00528	0.00155	1
rs3743888	9Sft35eUe7g2mGiR5E	*AXIN1*	0.00270	0.00163	1
rs11850781	NoXN6F3SR7AMv_v_6Q	*NIN*	−0.00577	0.00164	1
Baseline – follow-up					
rs10498246	fcV3S0U75lf1e3op0U	*SP110*	−0.00414	2.23×10^{-5}	0.103
rs6701295	cXl3ddwDJC3qA16ri4	*SMG5*	0.00249	2.26×10^{-4}	0.523
rs1737046	Tt5huq2hqZcdZqzRSc	*HCG4*	−0.00318	3.99×10^{-4}	0.616
rs4602357	6olooHit00T3lmfo5U	*CEP63*	0.00251	5.94×10^{-4}	0.688
rs12343854	fpnvXlHteCO4OrrGP0	*SEMA4D*	0.00231	9.24×10^{-4}	0.784

No interactions are significant at $q < 0.05$.

The aim of this investigation was to study genetic and transcriptomic correlates of response to exposure-based therapy, which may act across diagnostic boundaries. We sought to increase power by recruiting from treatment studies for two disorders with differing treatment procedures. Studying anxiety disorders as a heterogeneous group has been effective in genomics (Otowa et al. 2016). However, combining groups increases heterogeneity, partially negating the increased power from the enlarged sample size. The disorders studied are conceptually distinct, and treatment is tailored to the needs of the participant, differing between and within diagnostic groups. Combining across disorder groups incurs disorder-specific differences, such as the lower baseline severity of the SP group and higher comorbidity in the DA group herein. These differences reflect the varying nature of the disorders and recruitment to treatment – for example, the high comorbidity of the DA group is likely to result from secondary consequences of avoiding dental treatment, such as a phobia of vomiting or social anxiety about visiting the dentist. Although we have sought to control for this heterogeneity statistically, it limits the conclusions of this investigation. Furthermore, many social and environmental influences on treatment response have been proposed, and the covariates controlled for within this analysis cannot correct for all possible confounds. Nevertheless, investigating biological correlates of therapy requires a pragmatic approach. Cohorts of individuals receiving psychological therapies, particularly those outside of clinical trials, are prone to heterogeneity and attrition. For any biological predictor to contribute valuably to therapeutic decision-making, it must be robust to these limitations.

Although no genome-wide gene expression studies have investigated response to CBT, single-gene studies have suggested a role for differential *FKBP5* expression in response to CBT for PTSD (Levy-Gigi et al. 2013; Yehuda et al. 2013). One probe in this study, ZdI45Se3VG7s869FKo, captures expression of FKBP5, but was not associated with either outcome (baseline to post-treatment: $P = 0.0533$, $q = 0.999$; baseline to follow-up: $P = 0.607$, $q = 0.997$). However, the low power of the analysis (and differences between this cohort and those examined previously), limit strong conclusion.

Gene expression differs between different tissues and organs; expression observed in peripheral blood may not reflect that in the brain. Previous studies suggest moderate correlation between gene expression in different tissues, varying by individual genes (Sullivan et al. 2006). The emergence of reference panels such as the GTEx Portal has made in silico assessment of blood–brain expression correlations at the individual gene level viable (Consortium 2015). As such, peripheral blood gene expression can provide relevant insights into gene expression in the brain, and this will improve as further brain expression samples are added to the reference. From a pragmatic standpoint, gene expression markers of treatment response will only be useful if they can be obtained from peripheral tissues – while the effect of gene expression in brain tissues is of biological interest, it cannot be of practical utility in this case.

Assessing the severity of anxiety disorders can be performed using different rating scales, with varying characteristics. No consensus regarding the best means of measuring response to CBT exists (Loerinc et al. 2015). In this study, the CGI-S was used as a measure of clinical concern across treatment groups, allowing a single measure to be used to assess general functioning. However, this measure bears a number of limitations. It is a subjective measure of clinical judgement that may fail to capture the participant's anxiety as appropriately as a self-report measure. Treatment response is likely to involve multiple components, including reduction in fear and increase in functioning, that a single measure may not capture. One potential solution is to combine

a number of scales using different assessors and assessing different aspects of treatment response. However, this increases the complexity of the analysis and the potential for spurious results. In addition, it would be difficult to interpret in a useful manner.

Anxiety disorders are widespread and disabling, and CBT is a first-line treatment for these conditions. CBT involves a considerable investment from the recipient, and a significant minority of those receiving it do not respond adequately. Stable pre-treatment predictors of outcome are required. To date, genetic variants and gene expression levels have not provided these predictors, individually or in combination. However, this does not argue against the continued study of the biology underlying CBT response. The pattern of findings to date is consistent with the highly polygenic model that has been proposed to influence behavioural traits (Chabris et al. 2015). Although individual genetic variants seem extremely unlikely to be valuable predictors, prediction might be achieved through the combined effect of many genetic variants, at multiple levels of analysis.

Response to CBT is likely to be influenced by genes and by the environment, and continued research to define reliable environmental and clinical predictors of response is vital – genetics can only be clinically useful in the context of known environmental and clinical risk factors (Hudson et al. 2015). Studies of genetic variation, gene expression and epigenetics should either adopt a hypothesis-neutral approach (exploring variation genome-wide), or be informed by robust associations in related traits (rather than assumed biological relevance). The effects of individual transcript differences are likely to be small. For these insights to be discovered, cohorts of thousands of individuals must be treated in as homogenous a manner as possible from recruitment to the analysis of the resulting data. This is not straightforward (especially given the heterogeneity inherent to CBT) but the example of the many international consortia driving advances in complex trait genetics demonstrates such investigations can yield valuable insights.

Acknowledgements

This paper represents independent research part supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust and King's College London, as well as the Alexander von Humboldt Foundation (Germany). J.R.I.C.'s PhD is funded by the Institute of Psychiatry, Psychology and Neuroscience, and the Alexander von Humboldt Foundation. S.D.J. was funded by NARSAD Young Investigator Grant (YI 60373). We would like to thank Hamel Patel, Amos Folarin and Stephen Newhouse from the NIHR BRC Bioinformatics Core for their help and the use of their pipelines (http://core.brc.iop.kcl.ac.uk/). We acknowledge the use of the multi-node computing cluster maintained by the NIHR BRC Bioinformatics Core. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Summary statistics from all analyses are available on request from the authors.

Disclosure statement

G. Breen is a consultant in pre-clinical genetics for Eli Lilly. All other authors declare no financial interests.

ORCID

Jonathan R.I. Coleman http://orcid.org/0000-0002-6759-0944
Kathryn J. Lester http://orcid.org/0000-0002-0853-2664
Susanna Roberts http://orcid.org/0000-0002-2616-2644
Robert Keers http://orcid.org/0000-0001-7954-7583

Simone De Jong http://orcid.org/0000-0002-8002-7195
Héléna Gaspar http://orcid.org/0000-0002-5540-2707
Silvia Schneider http://orcid.org/0000-0003-0698-8411
Jürgen Margraf http://orcid.org/0000-0001-5207-7016
Gerome Breen http://orcid.org/0000-0003-2053-1792
Thalia C. Eley http://orcid.org/0000-0001-6458-0700

References

1000GenomesConsortium 2012. An integrated map of genetic variation from 1,092 human genomes. Nature. 491:56–65.

Ballenger JC. 1999. Current treatments of the anxiety disorders in adults. Biol Psychiatry. 46:1579–1594.

Barlow DH, Bullis JR, Comer JS, Ametaj AA. 2013. Evidence-based psychological treatments: an update and a way forward. Annu Rev Clin Psychol. 9:1–27.

Baxter AJ, Vos T, Scott KM, Ferrari AJ, Whiteford HA. 2014. The global burden of anxiety disorders in 2010. Psychological Medicine. 44:2363–2374.

Bruchmuller K, Margraf J, Suppiger A, Schneider S. 2011. Popular or unpopular? Therapists' use of structured interviews and their estimation of patient acceptance. Behav Ther. 42:634–643.

Chabris CF, Lee JJ, Cesarini D, Benjamin DJ, Laibson DI. 2015. The Fourth Law of Behavior Genetics. Curr Dir Psychol Sci. 24:304–312.

Chang C, Chow C, Tellier L, Vattikuti S, Purcell S, Lee J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 4:7.

Coleman JRI, Euesden J, Patel H, Folarin AA, Newhouse S, Breen G. 2015. Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. Brief Funct Genomics. Advance online publication. doi:10.1093/bfgp/elv037.

Coleman JRI, Lester KJ, Keers R, Roberts S, Curtis C, Arendt K, Bögels S, Cooper P, Creswell C, Dalgleish T, et al. 2016.

Genome-wide association study of response to cognitive behavioral therapy in children with anxiety disorders. Br J Psychiatry. Advance online publication. doi:10.1192/bjp.bp.115.168229

Consortium TG. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 348:648–660.

Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. 2013. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry. 12:137–148.

Dabney A, Storey JD, Warnes G. 2004. Q-value estimation for false discovery rate control. Medicine. 344:539–548.

Ding L-H, Xie Y, Park S, Xiao G, Story MD. 2008. Enhanced identification and biological validation of differential gene expression via Illumina whole-genome expression arrays through the use of the model-based background correction methodology. Nucleic Acids Res. 36:e58.

Du P, Kibbe WA, Lin SM. 2007. nuID: a universal naming scheme of oligonucleotides for illumina, affymetrix, and other microarrays. Biol Direct. 2:16–6152.

Du P, Kibbe WA, Lin SM. 2008. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 24:1547–1548.

Eddelbuettel D, Sanderson C. 2014. RcppArmadillo: Accelerating R with high-performance C++ linear algebra. Comput Stat Data Anal. 71:1054–1063.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 10:48.

Eley TC, Hudson JL, Creswell C, Tropeano M, Lester KJ, Cooper P, Farmer A, Lewis CM, Lyneham HJ, Rapee RM, et al. 2012. Therapygenetics: the 5HTTLPR and response to psychological therapy. Mol Psychiatry. 17:236–237.

Euesden J, Lewis CM, O'Reilly PF. 2015. PRSice: Polygenic Risk Score software. Bioinformatics. 31:1466–1468.

Fineberg NA, Haddad PM, Carpenter L, Gannon B, Sharpe R, Young AH, Joyce E, Rowe J, Wellsted D, Nutt D, et al. 2013. The size, burden and cost of disorders of the brain in the UK. J Psychopharmacol. 27:761–770.

Gaujoux R, Seoighe C. 2013. CellMix: a comprehensive toolbox for gene expression deconvolution. Bioinformatics. 29:2211–2212.

Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, O'Dushlaine C, Moran JL, Chambert K, Stevens C, et al. 2012. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics. 28:2543–2545.

Guy W. 1976. Clinical global impression scale. The ECDEU Assessment Manual for Psychopharmacology-Revised. Rockville (MD): US Department of Health Education and Welfare. (DHEW publication no. ADM 76 338:218-222).

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. 2009. The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter. 11:10–18.

Hofmann S, Asnaani A, Vonk IJ, Sawyer A, Fang A. 2012. The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. Cognit Ther Res. 36:427–440.

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 44:955–959.

Hudson JL, Keers R, Roberts S, Coleman JRI, Breen G, Arendt K, Bogels S, Cooper P, Creswell C, Hartman C, et al. 2015. Clinical Predictors of Response to Cognitive-Behavioral Therapy in Pediatric Anxiety Disorders: The Genes for Treatment (GxT) Study. J Am Acad Child Adolesc Psychiatry. 54:454–463.

In-Albon T, Suppiger A, Schlup B, Wendler S, Margraf J, Schneider S. 2008. Validität des Diagnostischen Interviews bei psychischen Störungen (DIPS für DSM-IV-TR). Z Klin Psychol Psychother (Gott). 37:33–42.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 8:118–127.

Kessler RC, Angermeyer M, Anthony JC, De Graaf RON, Demyttenaere K, Gasquet I, De Girolamo G, Gluzman S, Gureje OYE, Haro JM, et al. 2007. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry. 6:168–176.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 9:559.

Lester KJ, Eley TC. 2013. Therapygenetics: Using genetic markers to predict response to psychological treatment for mood and anxiety disorders. Biol Mood Anxiety Disord. 3:4–16.

Lester KJ, Roberts S, Keers R, Coleman JR, Breen G, Wong CC, Xu X, Arendt K, Blatter-Meunier J, Bogels S, et al. 2016. Non-replication of the association between 5HTTLPR and response to psychological therapy for child anxiety disorders. Br J Psychiatry. 208:182–188.

Levy-Gigi E, Szabó C, Kelemen O, Kéri S. 2013. Association Among Clinical Response, Hippocampal Volume, and FKBP5 Gene Expression in Individuals with Posttraumatic Stress Disorder Receiving Cognitive Behavioral Therapy. Biol Psychiatry. 74:793–800.

Loerinc AG, Meuret AE, Twohig MP, Rosenfield D, Bluett EJ, Craske MG. 2015. Response rates for CBT for anxiety disorders: Need for standardized criteria. Clin Psychol Rev. 42:72–82.

Margraf J. 1994. Mini-DIPS. Diagnostisches Kurz-Interview bei psychischen Stoerungen (Testmappe mit Handbuch und Interviewleitfaden).

Margraf J, Zlomuzica A. 2015. Changing the future, not the past: a translational paradigm shift in treating anxiety. EMBO Rep. 16:259–260.

Mendlowicz MV, Stein MB. 2000. Quality of Life in Individuals With Anxiety Disorders. Am J Psychiatry. 157:669–682.

Michael T, Zetsche U, Margraf J. 2007. Epidemiology of anxiety disorders. Psychiatry. 6:136–142.

National Collaborating Centre for Mental Health. 2011. Common mental health disorders: identification and pathways to care. Leicester: British Psychological Society; London: Royal College of Psychiatrists.

Newman MG, Llera SJ, Erickson TM, Przeworski A, Castonguay LG. 2013. Worry and Generalized Anxiety Disorder: A Review and Theoretical Synthesis of Evidence on Nature, Etiology, Mechanisms, and Treatment. Annu Rev Clin Psychol. 9:275–297.

Norton PJ, Price EC. 2007. A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorders. J Nerv Ment Dis. 195:521–531.

Olatunji BO, Cisler JM, Deacon BJ. 2010. Efficacy of Cognitive Behavioral Therapy for Anxiety Disorders: A Review of Meta-Analytic Findings. Psychiatr Clin North Am. 33:557–577.

Olatunji BO, Davis ML, Powers MB, Smits JAJ. 2013. Cognitive-behavioral therapy for obsessive-compulsive disorder: A meta-analysis of treatment outcome and moderators. J Psychiatr Res. 47:33–41.

Oldham MC, Langfelder P, Horvath S. 2012. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. BMC Syst Biol. 6:63.

Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, Bigdeli T, Aggen SH, Adkins D, Wolen A, et al. 2016. Meta-analysis of genome-wide association studies of anxiety disorders. Mol Psychiatry. Advance online publication. doi:10.1038/mp.2015.197

Otto MW, Pollack MH, Maki KM. 2000. Empirically supported treatments for panic disorder: Costs, benefits, and stepped care. J Consult Clin Psychol. 68:556.

Otto MW, Smits JA, Reese HE. 2004. Cognitive-behavioral therapy for the treatment of anxiety disorders. J Clin Psychiatry. 65:34–41.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38:904–909.

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet. 16:85–97.

Roberts S, Wong CCY, Keers R, De Jong S, Breen G, Lee SH, Coleman JRI, Teismann T, Margraf J, Schneider S, et al. Under Review. Genome-wide expression and response to exposure-based psychological therapy.

Schmid R, Baum P, Ittrich C, Fundel-Clemens K, Huber W, Brors B, Eils R, Weith A, Mennerich D, Quast K. 2010. Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3. BMC Genomics. 11:349.

Schneider RL, Arch JJ, Wolitzky-Taylor KB. 2015. The state of personalized treatment for anxiety disorders: A systematic review of treatment moderators. Clin Psychol Rev. 38:39–54.

Schneider S, Margraf J. 2011. DIPS: Diagnostisches Interview bei psychischen Störungen. Berlin Heidelberg: Springer-Verlag.

Shabalin AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 28:1353–1358.

Stewart RE, Chambless DL. 2009. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: a meta-analysis of effectiveness studies. J Consult Clin Psychol. 77:595–606.

Sullivan PF, Fan C, Perou CM. 2006. Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet. 141:261–268.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE. 6:e21800.

Suppiger A, In-Albon T, Hendriksen S, Hermann E, Margraf J, Schneider S. 2009. Acceptance of Structured Diagnostic Interviews for Mental Disorders in Clinical Practice and Research Settings. Behav Ther. 40:272–279.

Suppiger A, In-Albon T, Herren C, Bader K SS., Margraf J, 2008. Reliabilität des Diagnostischen Interviews bei Psychischen Störungen (DIPS für DSM-IV-TR) unter klinischen Routinebedingungen. Verhaltenstherapie. 18:237–244.

Taylor S, Abramowitz JS, McKay D. 2012. Non-adherence and non-response in the treatment of anxiety disorders. J Anxiety Disord. 26:583–589.

Team RC. 2012. R: A language and environment for statistical computing.

Uher R, Perroud N, Ng MY, Hauser J, Henigsberg N, Maier W, Mors O, Placentino A, Rietschel M, Souery D. 2010. Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. Am J Psychiatry. 167:555–564.

Wannemuller S, Johren M. 2015. Ein Fünf-Sitzungsprogramm zur Behandlung von Zahnbehandlungsangst. Verhaltenstherapie & Verhaltensmedizin. 36:177–196.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 88:76–82.

Yehuda R, Daskalakis NP, Desarnaud F, Makotkine I, Lehrner A, Koch E, Flory JD, Buxbaum JD, Meaney MJ, Bierer LM. 2013. Epigenetic Biomarkers as Predictors and Correlates of Symptom Improvement Following Psychotherapy in Combat Veterans with PTSD. Front Psychiatry. 4:1–14.

Chapter 5: The relationship between depression and body mass index in the UK Biobank, and the contribution of polygenic risk

5.1 Introduction

Depression and obesity are severe and growing public health concerns. Depressive disorders account for approximately 3% of the total global burden of disease and are the largest single contributor to years lived with disability (Whiteford, Degenhardt, Rehm, *et al*, 2013). While the global burden of disease has decreased in relative terms in the last twenty years, the burden of depressive disorders has increased (Murray, Vos, Lozano, *et al*, 2012). The effect of high body mass index (BMI) shows a similar pattern, accounting (indirectly) for around 4% of the global burden of disease, and increasing substantially from 1990 to 2010 (Lim, Vos, Flaxman, *et al*, 2013; Ng, Fleming, Robinson, *et al*, 2014).

Variation in BMI and depression are not independent. Although individual studies often show contradictory results, large-scale meta-analyses have repeatedly shown a small positive correlation between depression and BMI (de Wit, Luppino, van Straten, *et al*, 2010; Luppino, de Wit, Bouvy, *et al*, 2010; Scott, Bruffaerts, Simon, *et al*, 2007; Simon, Von Korff, Saunders, *et al*, 2006). Furthermore, diagnostic criteria for depression include BMI-related processes such as weight change, dysregulated appetite, and unusual sleep and activity patterns (Association, 2013). As these criteria have

no inherent directionality, both reductions (termed typical depression) and increases (atypical depression) in these aspects can be present in depression (Sullivan, Prescott & Kendler, 2002). As such, variation in BMI is not merely associated with depression, but can also be an inherent part of the diagnosis. However, the causal nature and direction of the relationship remains unclear, and many potential mediators or moderators have been suggested (Faith, Butryn, Wadden, *et al*, 2011; Faith, Matz & Jorge, 2002; Gibson-Smith, Bot, Paans, *et al*, 2016; Konttinen, Kiviruusu, Huurre, *et al*, 2014).

Insights into the relationship between depression and BMI could be gained by analysing the genetic contribution to variance in each trait, and assessing to what extent these contributions are shared. Evidence from the twin literature suggests that BMI and depression may share a small genetic component, resulting in a genetic correlation around 12% (Afari, Noonan, Goldberg, *et al*, 2010; Jokela, Berg, Silventoinen, *et al*, 2016). Studies seeking to identify genetic variants have suggested a role for common (and potentially for rare) variants, but such findings await robust replication (Jokela, Elovainio, Keltikangas-Jarvinen, *et al*, 2012; Milaneschi, Lamers, Bot, *et al*, 2015; Samaan, Lee, Gerstein, *et al*, 2015).

Molecular genetic studies of BMI and depression have differed in their rate of success to date. Studies of BMI have successfully detected high-effect rare variants (such as perturbations in the leptin system) and identified

a considerable amount of the contribution of common genetic variants (Zhang, Proenca, Maffei, *et al*, 1994). Much of this latter success has been driven by meta-analyses from the GIANT consortium, the most recent of which identified 97 associated genetic loci and a genome-wide heritability estimate from genome-wide genotype data of ~20%(Locke, Kahali, Berndt, *et al*, 2015). In contrast, the progress of genomics in studying depression has been slower despite considerable research effort (Major Depressive Disorder Working Group of the Psychiatric, Ripke, Wray, *et al*, 2013). However, genome-wide significant loci are emerging. A detailed study of severe depression in Han Chinese participants yielded 2 variants at genome-wide significance (Cai, Bigdeli, Kretzschmar, *et al*, 2015). An investigation of self-reported data from users of the 23&Me consumer genotyping service (limited to those with European-ancestry) identified 15 loci in meta-analysis with previous studies from the Psychiatric Genomics Consortium (PGC; Hyde, Nagle, Tian, *et al*, 2016). The next iteration of the PGC depression genomic mega-analysis is expected to yield further associations (Lewis, 2015).

Genomic data have uses beyond identifying genome-wide significant loci. Genomic studies are typically underpowered because individual genetic variants have small effects; as a result, loci with real effects will be enriched at *p*-value thresholds below conventional significance. Although these

cannot be specifically identified, the predictive power of such loci can be leveraged through building polygenic risk scores, which use the results of established genome-wide association studies (such as those mentioned above) to weight variants in a target cohort, and thus enable the calculation of genetic risk for a disorder in a novel cohort (Dudbridge, 2013; Euesden, Lewis & O'Reilly, 2015; International Schizophrenia, Purcell, Wray, *et al*, 2009). Such risk scores can also be used to examine the relationship between disorders. A similar approach has been carried out previously in the Generation Scotland cohort, in which variance in BMI was predicted using genetic risk scores for depression and for BMI (Clarke, Hall, Fernandez-Pujals, *et al*, 2015). The study found no effect of genetic risk of depression on BMI (nor of genetic risk for BMI on depression) but did find a significant interaction between (unweighted) BMI polygenic profile scores and depressive disorder predicting BMI, with a greater effect of BMI genetic risk in the case group compared to controls.

Recently, there has been increased interest in BMI as a behavioural phenotype. Most antipsychotic and some anti-depressant drugs alter BMI, potentially implying shared biological pathways, and a combined role of metabolic and psychiatric influences has been proposed in a variety of psychiatric disorders (Fava, 2000; Khansari & Sperlagh, 2012; Yao & Reddy, 2005). One of the most intriguing pieces of evidence has emerged from the

field of complex genetics. The common additive component of the genetic contribution to variance in a trait (sometimes referred to as "chip" heritability) can be partitioned into cell-type-specific components according to the annotation of genetic variants to cell-type-specific histone modifications (Finucane, Bulik-Sullivan, Gusev, *et al*, 2015). When this procedure was performed on the results from a large meta-analysis of BMI genomic studies, significant enrichment was found for central nervous system cell types (in comparison to enrichment of adrenal and pancreatic cell types with variance in fasting glucose levels; Finucane, Bulik-Sullivan, Gusev, *et al*, 2015; Speliotes, Willer, Berndt, *et al*, 2010). This reinforces the potential for a shared behavioural genetic component between BMI and psychological or psychiatric phenotypes.

Within this study, we investigated the relationship between BMI and depression using polygenic risk scores in a large population cohort, the UK Biobank (Sudlow, Gallacher, Allen, *et al*, 2015). The results of this study add to recent analyses to provide additional insight into this complex relationship and address the relevance of genetic factors within that relationship (Clarke, Hall, Fernandez-Pujals, *et al*, 2015; Milaneschi, Lamers, Peyrot, *et al*, 2016).

5.2 Materials and methods

5.2.1 Cohort Description

The UK Biobank is a prospectively sampled population cohort of approximately 500,000 adult individuals (aged between 40-69) from the UK, with data collected on an extensive range of health-related phenotypes (Sudlow, Gallacher, Allen, *et al*, 2015). Data were made available from the baseline assessment of the full cohort, as well as a targeted re-assessment of a subset of individuals. Health data were gathered via an extensive touchscreen questionnaire, with specific details on prescription medications and health conditions obtained during interviews with a nurse. Additional information was available from electronic health records detailing inpatient hospital episodes data. Full details on the collection of the UK Biobank cohort can be found on the project website (http://www.ukbiobank.ac.uk/).

5.2.2 Phenotype definitions

Data on BMI and covariates were available from UK Biobank. Individuals reporting regular use of antipsychotic or mood stabilising medications, reporting cancer of any kind or reporting an eating disorder (all of which may alter BMI) were excluded from analysis. Raw BMI scores were transformed with a natural logarithm to increase normality.

Depression was defined as any reported primary diagnosis of depression from inpatient hospital episodes data (ICD 10 subchapters F32 and F33) or meeting criteria from a previous publication on the UK Biobank cohort (Smith, Nicholl, Cullen, *et al*, 2013). These latter criteria required participants to report (as part of their general assessment) a previous visit to a GP or psychiatrist for stress, anxiety or depression, and at least one period of depression or anhedonia lasting at least two weeks. Depression cases were excluded if they self-reported (or had a hospitalisation primarily for) bipolar disorder, psychosis, multiple personality disorder, autism or intellectual disability. Depression controls were defined as individuals who did not meet the criteria for case status, did not report depression, anxiety or the excluded conditions outlined above, and did not report taking medication with an antidepressant or anxiolytic indication. Individuals who did not provide sufficient data to establish depression case status were excluded from analysis.

5.2.3. Genotyping and imputation

Genome-wide genotyping of the cohort was performed by Affymetrix using two customised microarrays, the UK BiLEVE and UKB Axiom arrays, which have very similar content and assay over 800,000 variants. Details of the genotyping and quality control processes are available at http://www.ukbiobank.ac.uk/wp-

content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf. Genotype data were imputed to a combined reference panel of the UK10K and 1000 Genomes Phase 3. Imputation analyses were performed as part of the UK Biobank project using IMPUTE3 software; full details of this procedure are available at http://biobank.ctsu.ox.ac.uk/showcase/docs/impute_ukb_v1.pdf.

5.2.4. Genotype quality control

Preliminary quality control was performed by Affymetrix during and after genotyping, and centrally by the UK Biobank team before and following imputation, and is described in the previously noted references. Further quality control specific to this study was performed using PLINK2 and QCTOOL (Chang, Chow, Tellier, *et al*, 2015; http://www.well.ox.ac.uk/~gav/qctool). The cohort was limited to depression cases and controls who self-identified as British and were of White Western European ancestry (as inferred from principal components analysis by UK Biobank). No participants were substantially related in this subset of the cohort (all pairwise KING relatedness coefficient < 0.044, equivalent to a greater separation than third-degree relatives). Participants were retained in the analysis if they had > 98% of genotyped variants available and had genome-wide heterozygosity (as inferred from genotyped variants) within three standard deviations of the sample mean. Imputed variants were

retained for the analysis if they were common (minor allele frequency ≥ 0.01), imputed with high certainty (info ≥ 0.9) and did not deviate substantially from Hardy-Weinberg equilibrium (Hardy-Weinberg equilibrium test $p \geq 1x10^{-6}$).

5.2.5. Generation of polygenic risk scores

Polygenic risk scores (PRS) were derived for the analyses using the default settings of PRSice, which performs high-resolution scoring to identify the most predictive PRS (Euesden, Lewis & O'Reilly, 2015). The depression PRS was derived from the PGC1 mega-analysis of major depression, and the BMI PRS from the GIANT 2015 trans-ethnic analysis (Locke, Kahali, Berndt, *et al*, 2015; Major Depressive Disorder Working Group of the Psychiatric, Ripke, Wray, *et al*, 2013). Variants were retained for the analysis if present in the relevant base dataset and the UK Biobank dataset, which were merged and then clumped to address linkage disequilibrium. Scores were calculated at *p*-value thresholds extending from 0.0001 to 0.5, with intervals of 0.00005. Analyses were performed within- and across-trait. The most predictive PRS was identified by comparing the R^2 of the model containing the PRS and genotyping batch, assessment centre, and the first eight principal components (to address potential confounding by technical artefacts and population stratification) with that not including the PRS. As fewer covariates were included, PRS were calculated on a slightly larger cohort

than was assessed in the main analysis. Variance explained for depression is reported by PRSice as Nagelkerke's pseudo-R^2 (on the observed scale). This was back-converted to the Cox-Snell pseudo-R^2, and then converted to the liability scale using the software package GEAR, assuming a population prevalence of 12% (Chen, 2014; Fernandex-Pujals, Adams, Thomson, *et al*, 2015; Lee, Goddard, Wray, *et al*, 2012). Variance explained for log-BMI is on the liability scale. The multiple testing incurred by the PRSice method suggests an adjusted alpha threshold of 0.001 (derived by permutation) should be used for a single test (Euesden, Lewis & O'Reilly, 2015). To account for the four tests performed, the correlation matrix between the most-predictive PRS from each analysis was spectrally decomposed and the Nyholt-Šidák calculation of the effective number of independent tests was performed (Nyholt, 2004).

5.2.6 Analyses

Statistical analyses were performed in R. The most predictive within-trait PRS were used as proxies for the genetic components of each trait. Depression case-control status was regressed on the depression PRS, on the BMI PRS, on log-BMI as a trait, and on the interaction between the three components, using a logistic model (to assess main effects and multiplicative interactions) and a linear model (to assess additive interactions; Mullins, Power, Fisher, *et al*, 2016; Rothman, Greenland & Lash, 2008). A linear model

was used to assess the effects of BMI PRS, depression PRS, depression case status and the interaction between the dependent variables on log-BMI. All analyses featured as covariates fixed effects of sex, age in years (at baseline assessment), Townsend Deprivation Index, the first eight principal components (as derived by UK Biobank from the genotype data), and unordered factors accounting for region of birth, assessment centre and genotyping batch. The Townsend Deprivation Index is a measure of neighbourhood deprivation, where a higher score on the index represents more unemployment, less vehicle or home ownership, and more home overcrowding (Townsend, Phillimore & Beattie, 1988). Region of birth was converted from Cartesian coordinates (as received from UK Biobank) into eight factors by k-means clustering using the pamk function from the fpc R package(Kaufman & Rousseeuw, 2009; https://cran.r-project.org/web/packages/fpc/fpc.pdf). Where the model included an interaction between the trait and either or both PRS, all covariate-by-trait and covariate-by-PRS interactions were included in the model (Keller, 2014). All continuous variables were normalised for analysis. Following the initial analysis of log-BMI, the cohort was stratified into depression cases and controls to assess differential effects of PRS on log-BMI within these groups. Significance for each test was set at 0.0125 (Bonferroni correction for four tests, assessing association with depression, with BMI and stratifying the tests with BMI into cases and controls).

In addition to the linear models as described above, genome-wide association analyses (GWAS) of depression and of log-BMI (in the whole cohort and stratified by depression status) were performed. GWAS was performed using the "frequentist" option in SNPTEST and probabilistic dosage estimates from the imputed genotypes (Marchini & Howie, 2010). The variables of interest (log-BMI and depression) were separately regressed on the covariates from the linear models, and the resulting residuals used as the phenotype for GWAS. The results of GWAS were used to calculate genetic correlation and to estimate differences in heritability. Specifically, genetic correlations between the residuals for depression and for log-BMI, and between the residuals for log-BMI in depression cases and in controls, were calculated using LDScore, which also provides an estimate of heritability (Bulik-Sullivan, Loh, Finucane, *et al*, 2015). Analyses were performed without constraining the intercept of the LD score regression, and constraining the intercepts for heritability to 1 and genetic covariance to 0.0625 (the correlation between the residuals for depression and for log-BMI). Additional estimates of heritability were calculated from imputed genotypes (hard-called using the thresholds used in GWAS) using GCTA-GREML (Lee, Yang, Goddard, *et al*, 2012; Yang, Lee, Goddard, *et al*, 2011). Estimates of log-BMI heritability were compared between depression cases and controls.

Sensitivity analyses were performed to assess the importance of antidepressant or anxiolytic use in the case group. Medications were classified as weight-increasing, weight-decreasing, weight-modulating (both increasing and decreasing) or weight-neutral (Supplementary Material). Additional analyses were performed by removing all cases reporting medication use, and by adding the different categories of medication as covariates in the model.

Post-hoc power analyses were conducted in R using the *pwr* package. Cohen's f^2 at 80% and 90% power, and power to detect observed effects of each variable of interest, was calculated for each analysis (Cohen, 1988; Selya, Rose, Dierker, *et al*, 2012). Cohen's f^2 is a measure of effect size suitable for assessing the contribution of a single variable in a multiple regression (Cohen, 1992).

The BMI PRS analyses presented in this paper mirror previous work performed in the Generation Scotland cohort (Clarke, Hall, Fernandez-Pujals, *et al*, 2015). However, this previous analysis used an earlier meta-analysis of BMI GWAS (Speliotes, Willer, Berndt, *et al*, 2010). As this GWAS did not list betas for each variant, the PRS produced is unweighted, differing from that used in this analysis. As such, there may be differences between the results of Clarke et al and this analysis that are attributable to the different PRS. To test this, the analyses in this study were repeated with the

older, unweighted PRS (all betas = 1), and the analyses in Generation Scotland were repeated using the new BMI PRS (Locke, Kahali, Berndt, *et al*, 2015). Analyses in Generation Scotland were performed in 18850 individuals, of whom 2605 were defined as depression cases, with 16245 controls. Age, sex and 4 multi-dimensional scaling components were included as covariates. As in the UKBB cohort, the optimal PRS was identified using the default settings in PRSice (Euesden, Lewis & O'Reilly, 2015).

5.3 Results

5.3.1. Cohort characteristics

The interim release of the genotyping data contained genotypes for 152,734 individuals, from which 7,009 depression cases and 14,030 controls were available with full phenotypic data following quality control (Supplementary Figure 1). Imputed genotype data were available on 8,747,914 variants following quality control.

A higher proportion of depression cases than controls were female and cases tended to have higher BMI than controls. Cases were also younger and tended to live in more deprived areas (Table 1). All covariates were associated with at least one variable of interest: log(BMI), depression status or either PRS (Table 2).

5.3.2. Polygenic risk score analyses

The Nyholt-Šidák method indicated that the within-trait and cross-trait PRS analyses were largely independent (3.75 effective tests), resulting in an adjusted alpha threshold of $p = 2.67 \times 10^{-4}$ (Euesden, Lewis & O'Reilly, 2015; Nyholt, 2004). PRS were significantly associated within-trait, but not cross-trait (Table 3). The optimal within-trait PRS (bold in Table 3) were taken forward for further analyses, including calculating correlations with covariates (Table 2).

5.3.3. Effects on depression

Depression and log-BMI had a small but significant positive correlation at the phenotype level (Table 2). In the linear and logistic analyses of depression, no interactions (neither the three-way interaction between depression status, depression PRS and BMI PRS, nor the two-way interactions alone or in combination) were significantly associated with outcome when modelled as multiplicative, nor when modelled as additive (Table 4). BMI PRS was not significantly associated with depression status, although log-BMI and depression PRS were associated alone and together (Table 4).

5.3.4. Effects on log-BMI

In the analysis of log-BMI, no interactions were significantly associated with variance in log-BMI. Depression status and BMI PRS were significantly associated included together and separately (Table 5). The polygenic risk for depression was not significantly associated with log-BMI. Analyses stratified by depression case status show a nominally significant effect of depression PRS in controls, but this does not survive correction for multiple testing (Table 5).

5.3.5. Stratified heritability and genetic correlations

Heritability estimates for log-BMI did not differ substantially between depression cases (16.8%, 95% CIs: 3.56-30.1%) and controls (18.8%, 95% CIs: 11.9-25.7%), and the genetic correlation between the two groups = 1 when intercepts were not constrained. Constraining the intercepts did not alter the results substantially (cases = 19.9%, 95% CIs: 10.8-29.0%; controls = 18.5%, 95% CIs: 13.4-23.6).

When the intercepts were not constrained, the genetic correlation between depression and log-BMI was not statistically significant (r_g = 1.26%, 95 CIs: -21.5% – 24.1%). Variance explained by common genetic variants in the cohort was estimated at 11.3% for depression (95 CIs: 6.03-16.5%; liability scale, assuming population prevalence of 12%), and at 18.4% for BMI (95 CIs:

13.4-23.4%; estimates from LDScore, similar estimates from GCTA not shown). Defining the intercepts had minimal effect on the estimation of heritability (depression = 13.1%, 95% CIs: 9.6-16.6%; log-BMI = 20.5%, 95% CIs: 16.8-24.2%), but altered the estimate of genetic correlation, although this did not pass statistical significance (r_g = 10.7%, 95 CIs: -5.05% – 26.5%).

5.3.6. Sensitivity analyses concerning antidepressant use

Antidepressant or anxiolytic medication use was reported by 1368 (19.5%) individuals with depression (some of whom were taking more than one drug). Analyses were performed excluding individuals using medication (Supplementary Tables 5, 6), and including the medication categories as covariates (Supplementary Tables 7). Controlling for medication use did not alter the conclusions from any analysis.

5.3.7. Power

The main analyses were powered to detect very small effects at 90% power (f^2 = 0.000521), as were the analyses of BMI stratified by depression case status (f^2= 0.00154 and 0.000764 in cases and controls respectively). Cohen suggested f^2=0.02 should be considered "small" (Cohen, 1988).

5.3.8. Replication of Clarke et al (2015)

BMI analyses were repeated using the Speliotes GWAS as the base, and assuming all betas = 1 to create an unweighted risk score (Speliotes,

Willer, Berndt, *et al*, 2010). Overall, results did not differ from the main analysis: BMI PRS was associated with log-BMI and not with depression status, regardless of other additions to the null model, and no interactions were significant (Supplementary Tables 8 and 9). However, the interaction between BMI PRS and depression status predicting log-BMI was nominally significant with the same direction of effect as was observed in Clarke et al (2015).

Additional analyses were performed in the Generation Scotland cohort using a weighted PRS derived from the most recent GIANT BMI GWAS meta-analysis (Supplementary Table 10; Locke, Kahali, Berndt, *et al*, 2015). The BMI PRS was strongly associated with BMI in the cohort, but not with depression, and the depression PRS was associated with depression but not BMI. The interaction between BMI PRS and depression status remained significant, but had a diminished effect compared to that reported in Clarke et al (2015).

5.4. Discussion

A small but significant positive correlation was observed between BMI and depression status, such that individuals with depression have an increased BMI on average. Polygenic risk scores capture within-trait variance in the cohort, apparently independently of the phenotype-level relationship between BMI and depression. Genetic risk for increased BMI does not appear

to affect depression status, nor is there a significant interaction between PRS-BMI and PRS-MDD to predict depression status or BMI. This argues that the mechanism of association between BMI and depression is not accounted substantially by the shared additive effect of common genetic variants.

The results suggest that the association between depression and BMI does not result from a genetic relationship that would arise from a genetic correlation greater than 25% (that is, greater than the 95% confidence interval of the observed genetic correlation). This is concordant with the estimate of genetic correlation around 12% reported in the twin literature. More modest genetic correlations will be detectable when the full UK Biobank data are released. A further caveat to this is that the genetic component studied herein results from an additive model of the effects of common variants captured by (or imputed from) genome-wide microarrays. As such, a relationship resulting from rare variation or non-additive scale (dominance) or interaction (epistatic) effects would not be captured in this study. Furthermore, although the PRS used in this study were generated from the largest relevant studies published to date, they still do not capture the full contribution of additive genetic effects. The absence of a sizable genetic component to the association between depression and BMI argues for increased study of non-genetic factors, including the effects of physical

illnesses that are correlated with both traits, as well as the environmental (including social) risks driving higher BMI in the context of depression.

The effect of genomic risk on the relationship between BMI and depression has previously been examined in the Generation Scotland cohort, a similar (but largely independent) study to the UK Biobank (Clarke, Hall, Fernandez-Pujals, *et al*, 2015). The Generation Scotland study found no cross-trait associations of PRS, in accordance with the results of this study, but differed in that a significant interaction between BMI polygenic profile scores and MDD predicting BMI was identified, with higher BMI PRS in the case group compared with the controls. Although the two cohorts are similar, the BMI PRS used in the Generation Scotland study was from a previous GIANT BMI meta-analysis compared to that used in this study, and the PRS produced was a summed risk score, not weighted by beta values (which were not available in the data release). Repeating the relevant analyses in the UK Biobank and in Generation Scotland (such that both PRS were used in both studies) suggests that the use of the unweighted PRS explains some, but not all of this disparity. No distinction was made for differing symptoms of depression in either this study or in Clarke et al (2015), and inter-study differences are likely to account for the rest of the inconsistency in results. In particular, depression associated with increased appetite and weight gain ("atypical depression") might be more common in the Generation Scotland

cohort than in UK Biobank. If so, this could explain some of the observed difference between the cohorts. This is supported by a further investigation in an independent cohort, which dissected depression into typical and atypical subtypes (and more specifically low vs high appetite subtypes) and showed no effect of metabolic PRS (including BMI PRS) on depression status overall, but found an effect of BMI PRS when the atypical group was considered alone (Milaneschi, Lamers, Peyrot, *et al*, 2016).

An important difference between the results of this analysis and those of Clarke et al was the use of weighted versus unweighted polygenic risk scores (Clarke, Hall, Fernandez-Pujals, *et al*, 2015). The most appropriate score depends on the power of the analysis. The accuracy of weighted risk scores can be limited if the base dataset is underpowered to detect the true effect sizes of variants associated with the phenotype, as is usually the case in psychiatric genetics. Underpowered GWAS are vulnerable to winner's curse (the overestimation of significant effect sizes when multiple tests are performed), which biases simple weighting by effect size (Vilhjálmsson, Yang, Finucane *et al*, 2015; Shi, Park, Duan, *et al*, 2016). More sophisticated weighting of variants could be performed, such as shrinking the reported effect size to a prior distribution reflecting the genetic architecture of the trait, but this may be inaccurate if the true genetic architecture differs from that modelled (Vilhjálmsson, Yang, Finucane *et al*, 2015). Assuming simple

weighting by effect size, unweighted scores are more robust to confounds like winner's curse (Dudbridge, 2013). However, as the power of the base dataset increases, the sampling error of the weighted score reduces to zero, unlike that of the unweighted score, and weighted scores become superior (Dudbridge, 2013). Additional complexity is added by the use of risk scores to perform cross-trait analyses. The weighting of the risk score is trained to the base phenotype, and reflects the observed distribution of effects. In comparison, an unweighted score would capture only which variants are associated, without imposing any effect size distribution on them. Accordingly, the precise hypothesis tested by cross-trait PRS differs if the score is weighted ("those at greater genetic risk of the base phenotype are more likely to exhibit the target phenotype") or unweighted ("the variants associated with the base phenotype are also associated with the target phenotype"). As such, the reported differences between results obtained with the Speliotes unweighted BMI risk score (Speliotes, Willer, Berndt, *et al*, 2010) and the Locke weighted risk score (Locke, Kahali, Berndt, *et al*, 2015) reflect differences in the method used as well as the power of the base studies.

This study represents an initial study on the pilot data of the UK Biobank, which will become an increasingly valuable resource as new data are added, especially when genetic data are available on the full cohort.

However, the breadth of the cohort results in some limitations, particularly in the diagnosis of depression. Although the definition of depression used herein has a high genetic correlation with the results of the PGC MDD meta-analysis (0.738, 95 CIs: 0.404-1.07), suggesting external validity on a genetic level, the lack of a formal psychiatric diagnostic assessment makes it difficult to assess the clinical validity of depression as defined (Major Depressive Disorder Working Group of the Psychiatric, Ripke, Wray, *et al*, 2013). In addition, dissection into symptom groups was not possible, as mentioned above. Finally, although this represents a large cohort for study, the final sample size is still underpowered to detect a very small contribution to variance from the genetic factors (particularly in the case of depression), and so we cannot exclude such effects. However, future analyses in the full UK Biobank cohort could be informative about such effects,

A small but significant positive correlation exists between BMI and depression, and this does not result from a substantial effect of shared common genetic variants. The effect of genetic risk for depression or for BMI does not appear to be influenced substantially by variance in the opposite trait, although this does not replicate previous findings; this disparity may result from unmeasured cohort-specific effects. Understanding the relationship between BMI and depression, and the effects of non-genetic

factors on these traits, may provide insight into two areas of growing concern for public health.

5.5 References

Afari, N., Noonan, C., Goldberg, J.*, et al* (2010) Depression and obesity: do shared genes explain the relationship? *Depress Anxiety*, **27**, 799-806.

Association, A. P. (2013) *Diagnostic and statistical manual of mental disorders, (DSM-5®)*: American Psychiatric Pub.

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K.*, et al* (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*, **47**, 291-295.

Cai, N., Bigdeli, T. B., Kretzschmar, W.*, et al* (2015) Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, **523**, 588-591.

Chang, C., Chow, C., Tellier, L.*, et al* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.

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

Clarke, T. K., Hall, L. S., Fernandez-Pujals, A. M.*, et al* (2015) Major depressive disorder and current psychological distress moderate the effect of polygenic risk for obesity on body mass index. *Transl Psychiatry*, **5**, e592.

Cohen, J. (1988) Statistical power analysis for the behavioral sciences.

---- (1992) A power primer. *Psychological bulletin*, **112**, 155.

de Wit, L., Luppino, F., van Straten, A.*, et al* (2010) Depression and obesity: a meta-analysis of community-based studies. *Psychiatry Res*, **178**, 230-235.

Dudbridge, F. (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet*, **9**, e1003348.

Euesden, J., Lewis, C. M. & O'Reilly, P. F. (2015) PRSice: Polygenic Risk Score software. *Bioinformatics*, **31**, 1466-1468.

Faith, M. S., Butryn, M., Wadden, T. A.*, et al* (2011) Evidence for prospective associations among depression and obesity in population-based studies. *Obesity Reviews*, **12**, e438-e453.

Faith, M. S., Matz, P. E. & Jorge, M. A. (2002) Obesity-depression associations in the population. *J Psychosom Res*, **53**, 935-942.

Fava, M. (2000) Weight gain and antidepressants. *J Clin Psychiatry*, **61**, 1,478-441.

Fernandez-Pujals, A. M., Adams, M. J., Thomson, P., *et al* (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,* **16,** e0142197

Finucane, H. K., Bulik-Sullivan, B., Gusev, A.*, et al* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*, **47**, 1228-1235.

Gibson-Smith, D., Bot, M., Paans, N. P.*, et al* (2016) The role of obesity measures in the development and persistence of major depressive disorder. *J Affect Disord*, **198**, 222-229.

Hyde, C. L., Nagle, M. W., Tian, C.*, et al* (2016) Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet*, **advance online publication**.

International Schizophrenia, C., Purcell, S. M., Wray, N. R.*, et al* (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748-752.

Jokela, M., Berg, V., Silventoinen, K.*, et al* (2016) Body Mass Index and Depressive Symptoms: Testing for Adverse and Protective Associations in Two Twin Cohort Studies. *Twin Res Hum Genet*, **19**, 306-311.

Jokela, M., Elovainio, M., Keltikangas-Jarvinen, L.*, et al* (2012) Body mass index and depressive symptoms: instrumental-variables regression with genetic risk score. *Genes Brain Behav*, **11**, 942-948.

Kaufman, L. & Rousseeuw, P. J. (2009) *Finding groups in data: an introduction to cluster analysis*: John Wiley & Sons.

Keller, M. C. (2014) Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biol Psychiatry*, **75**, 18-24.

Khansari, P. S. & Sperlagh, B. (2012) Inflammation in neurological and psychiatric diseases. *Inflammopharmacology*, **20**, 103-107.

Konttinen, H., Kiviruusu, O., Huurre, T.*, et al* (2014) Longitudinal associations between depressive symptoms and body mass index in a 20-year follow-up. *Int J Obes*, **38**, 668-674.

Lee, S. H., Goddard, M. E., Wray, N. R., *et al* (2012) A better coefficient of determination for genetic profile analysis. *Genet Epidemiol,***36,** 214-224

Lee, S. H., Yang, J., Goddard, M. E.*, et al* (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, **28**, 2540-2542.

Lewis, C. M. (2015) Mega-analysis of genome-wide association studies in major depressive disorder: MDD Working Group of the Psychiatric Genomics Consortium. In *World Congress of Psychiatric Genetics*. Toronto, Canada.

Lim, S. S., Vos, T., Flaxman, A. D.*, et al* (2013) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, **380**, 2224-2260.

Locke, A. E., Kahali, B., Berndt, S. I.*, et al* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197-206.

Luppino, F. S., de Wit, L. M., Bouvy, P. F.*, et al* (2010) Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Arch Gen Psychiatry*, **67**, 220-229.

Major Depressive Disorder Working Group of the Psychiatric, G. C., Ripke, S., Wray, N. R.*, et al* (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*, **18**, 497-511.

Marchini, J. & Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nat Rev Genet*, **11**, 499-511.

Milaneschi, Y., Lamers, F., Bot, M.*, et al* (2015) Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression. *Biol Psychiatry*.

Milaneschi, Y., Lamers, F., Peyrot, W. J.*, et al* (2016) Polygenic dissection of major depression clinical heterogeneity. *Mol Psychiatry*, **21**, 516-522.

Mullins, N., Power, R. A., Fisher, H. L.*, et al* (2016) Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol Med*, **46**, 759-770.

Murray, C. J., Vos, T., Lozano, R.*, et al* (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**, 2197-2223.

Ng, M., Fleming, T., Robinson, M.*, et al* (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, **384**, 766-781.

Nyholt, D. R. (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*, **74**, 765-769.

Rothman, K. J., Greenland, S. & Lash, T. L. (2008) *Modern epidemiology*: Lippincott Williams & Wilkins.

Samaan, Z., Lee, Y. K., Gerstein, H. C.*, et al* (2015) Obesity genes and risk of major depressive disorder in a multiethnic population: a cross-sectional study. *J Clin Psychiatry*, **76**, e1611-1618.

Scott, K. M., Bruffaerts, R., Simon, G. E.*, et al* (2007) Obesity and mental disorders in the general population: results from the world mental health surveys. *Int J Obes*, **32**, 192-200.

Selya, A. S., Rose, J. S., Dierker, L. C.*, et al* (2012) A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, **3**, 111.

Shi, J., Park, J-H., Duan, J., *et al* (2016) Winner's Curse Correction and Variable Thresholding Improve Performance of Polygenic Risk Modeling Based on Genome-Wide Association Study Summary-Level Data. *PLoS Genet,* **12,** e1006493

Simon, G. E., Von Korff, M., Saunders, K.*, et al* (2006) ASsociation between obesity and psychiatric disorders in the us adult population. *Archives of General Psychiatry*, **63**, 824-830.

Smith, D. J., Nicholl, B. I., Cullen, B.*, et al* (2013) Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS ONE*, **8**, e75362.

Speliotes, E. K., Willer, C. J., Berndt, S. I.*, et al* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*, **42**, 937-948.

Sudlow, C., Gallacher, J., Allen, N.*, et al* (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**, e1001779.

Sullivan, P. F., Prescott, C. A. & Kendler, K. S. (2002) The subtypes of major depression in a twin registry. *Journal of Affective Disorders*, **68**, 273-284.

Townsend, P., Phillimore, P. & Beattie, A. (1988) *Health and deprivation: inequality and the North*: Routledge.

Whiteford, H. A., Degenhardt, L., Rehm, J.*, et al* (2013) Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*, **382**, 1575-1586.

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

Yao, J. K. & Reddy, R. D. (2005) Metabolic investigation in psychiatric disorders. *Molecular neurobiology*, **31**, 193-203.

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., *et al* (2015) Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.,* **97,** 576-592

Zhang, Y., Proenca, R., Maffei, M.*, et al* (1994) Positional cloning of the mouse obese gene and its human homologue. *Nature*, **372**, 425-432.

5.6. Tables

5.6.1. Table 1

Variable	Full Cohort	Females	Males	t-test *p*	Depression Cases	Depression Controls	t-test *p*
N	21039	10232	10807	-	7009	14030	-
Female sex (N [%])	10232 [48.6]	-	-	-	**4250 [60.6]**	**5982 [42.6]**	**$< 10^{-50}$**
Age (mean, SD)	56.9 [7.95]	**56.4 [7.85]**	**57.4 [8.01]**	**2.26×10^{-22}**	**55.9 [7.79]**	**57.4 [7.98]**	**7.92×10^{-39}**
Townsend Index (mean, SD)	-1.55 [2.78]	-1.54 [2.75]	-1.56 [2.80]	0.522	**-1.19 [2.94]**	**-1.73 [2.67]**	**1.31×10^{-37}**
BMI (mean, SD)	27.5 [4.68]	**27.0 [5.10]**	**27.9 [4.21]**	**3.65×10^{-40}**	**27.9 [5.11]**	**27.3 [4.44]**	**8.99×10^{-17}**

Table 1: Demographic variables in the full cohort, and split by gender and depression case status. Significant differences between genders and between depression cases and controls are marked in bold ($p < 0.05$, t-tests not assuming equality of variance).

5.6.2. Table 2

Variable	log(BMI)		Depression		BMI PRS		Depression PRS	
	r	*P*	*r*	*p*	*r*	*p*	*r*	*p*
log(BMI)	-	-	**0.0539**	**4.88×10^{-15}**	**0.240**	**$< 10^{-50}$**	0.0130	0.0598
Depression	**0.0539**	**4.89×10^{-15}**	-	-	0.0103	0.137	**0.0505**	**2.22×10^{-13}**
BMI PRS	**0.240**	**$< 10^{-50}$**	0.0103	0.137	-	-	0.00172	0.803
Depression PRS	0.0130	0.0598	**0.0505**	**2.22×10^{-13}**	0.00172	0.803	-	-
Male gender	**0.113**	**$< 10^{-50}$**	**-0.170**	**$< 10^{-50}$**	0.0106	0.125	-0.00262	0.704
Age (years)	**0.0477**	**4.59×10^{-12}**	**-0.0891**	**2.56×10^{-38}**	0.000138	0.984	**-0.0201**	**0.00361**
Townsend Index	**0.0871**	**$< 10^{-50}$**	**0.0912**	**$< 10^{-50}$**	**0.0317**	**4.39×10^{-6}**	**0.0165**	**0.0167**
Centre *	**-0.0509**	**1.42×10^{-13}**	**0.0365**	**1.18×10^{-7}**	**-0.0182**	**0.00846**	**0.0468**	**1.14×10^{-11}**
Birth Cluster *	**-0.0478**	**4.04×10^{-12}**	**0.0207**	**0.00273**	**-0.0245**	**0.000379**	**-0.0601**	**2.51×10^{-18}**
Batch *	**0.0277**	**6.04×10^{-5}**	-0.0133	0.0542	**0.0276**	**6.24×10^{-5}**	**-0.0167**	**0.0155**
PC1	-0.0130	0.0595	0.000341	0.961	**0.0263**	**0.000136**	**-0.0204**	**0.00301**
PC2	0.00109	0.874	0.00682	0.323	0.00835	0.226	**0.0271**	**8.24×10^{-5}**
PC3	-0.0000125	0.999	-0.00259	0.707	-0.000322	0.963	-0.00558	0.419
PC4	**0.0261**	**1.56×10^{-4}**	0.00398	0.564	0.00571	0.408	**0.115**	**$< 10^{-50}$**
PC5	**-0.0264**	**1.30×10^{-4}**	-0.00281	0.684	-0.0111	0.109	**-0.174**	**$< 10^{-50}$**
PC6	-0.00604	0.381	-0.0129	0.0614	-0.00260	0.706	0.00881	0.201
PC7	-0.00141	0.838	0.00604	0.381	-0.0112	0.106	**-0.0179**	**0.00952**
PC8	0.00133	0.847	-0.00599	0.385	**0.0310**	**7.09×10^{-6}**	**0.0903**	**$< 10^{-50}$**

Table 2: Pairwise univariate correlations between the variables under study. Correlations with $p < 0.05$ are marked in bold. For factors (marked with *), the most significant correlation is reported.

5.6.3. Table 3

Polygenic risk scoring results within and between traits				
External GWAS	**UK Biobank Phenotype**	**Best threshold**	***p*-value at best threshold**	**Variance explained (R^2)**
PGC MDD	**Depression**	**0.1811**	**8.56×10^{-15}**	**0.00310 (0.00343)**
PGC MDD	log-BMI	0.0001	0.0349	0.000202
GIANT	**log-BMI**	**0.04795**	**$< 10^{-50}$**	**0.0559**
GIANT	Depression	0.0055	8.67×10^{-4}	0.000570 (0.000631)

Table 3: Polygenic risk scoring results within and between traits. Within-trait analyses were significant (bold, $p < 2.67 \times 10^{-4}$), cross-trait analyses were non-significant. Variance explained is reported on the liability scale where the phenotype was log-BMI. Where the phenotype is depression, pseudo-R^2 transformed to the liability scale is reported, with the untransformed Nagelkerke's pseudo-R^2 in parentheses.

5.6.4. Table 4

Coefficient	B	SE	*p*
Null model	See Supplementary Table 1		
… + depression PRS	**0.112**	**0.0154**	**4.37×10^{-13}**
… + BMI PRS	0.0161	0.0151	0.287
… + log-BMI	**0.144**	**0.0151**	**1.09×10^{-21}**
… + depression PRS **… + log-BMI**	**0.111** **0.144**	**0.0154** **0.0151**	**5.66×10^{-13}** **1.41×10^{-21}**
… + depression PRS x log-BMI (Multiplicative)	-0.0287	0.0157	0.0680
… + depression PRS x log-BMI (Additive)	-0.00548	0.00332	0.0984
… + BMI PRS **… + log-BMI**	-0.0192 **0.149**	0.0156 **0.0155**	0.217 **9.17×10^{-22}**
… + BMI PRS x log-BMI (Multiplicative)	0.00268	0.0150	0.858
… + BMI PRS x log-BMI (Additive)	-8.55×10^{-4}	0.00317	0.787
… + depression PRS … + BMI PRS	**0.112** 0.0165	**0.0154** 0.0151	**4.24×10^{-13}** 0.274
… + depression PRS x BMI PRS (Multiplicative)	-0.0150	0.0158	0.343
… + depression PRS x BMI PRS (Additive)	-0.00301	0.00330	0.362

(Table 4 continued)

… + depression PRS … + BMI PRS **… + log-BMI**	**0.111** -0.0187 **0.148**	**0.0154** 0.0156 **0.0155**	**5.90×10^{-15}** 0.230 **1.27×10^{-21}**
… + depression PRS x BMI PRS (Multiplicative) … + depression PRS x log-BMI (Multiplicative) … + BMI PRS x log-BMI (Multiplicative)	-0.00830 -0.0239 0.00274	0.0164 0.0163 0.0150	0.612 0.142 0.855
… + depression PRS x BMI PRS (Additive) …+ depression PRS x log-BMI (Additive) … + BMI PRS x log-BMI (Additive)	-0.00165 -0.00450 9.08×10^{-4}	0.00340 0.00342 0.00317	0.627 0.189 0.774
… + depression PRS x BMI PRS x log-BMI (Multiplicative)	0.00795	0.0148	0.590
… + depression PRS x BMI PRS x log-BMI (Additive)	0.00142	0.00311	0.648

Table 4: Effects of adding variables and interactions to the null model (effects shown in Supplementary Table 1) predicting variance in depression status. Significant ($p < 0.0125$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

5.6.5. Table 5

Coefficient	**B**	**SE**	***p***
Null model	See Supplementary Table 2		
… + BMI PRS	**0.233**	**0.00661**	**$< 10^{-50}$**
… + depression PRS	0.00809	0.00691	0.242
… + depression	**0.144**	**0.0147**	**1.30×10^{-22}**
… + BMI PRS **… + depression**	**0.232** **0.140**	**0.00660** **0.0143**	**$< 10^{-50}$** **1.13×10^{-22}**
… + BMI PRS x depression	0.0192	0.0143	0.182
… + depression PRS **… + depression**	0.00474 **0.144**	0.00691 **0.0147**	0.493 **4.93×10^{-22}**
… + depression PRS x depression	-0.0278	0.0150	0.0642
… + BMI PRS … + depression PRS	**0.232** 0.00874	**0.00664** 0.00672	**$< 10^{-50}$** 0.193
… + BMI PRS x depression PRS	-4.24×10^{-4}	0.00686	0.951
… + BMI PRS … + depression PRS **… + depression**	**0.232** 0.00548 **0.140**	**0.00660** 0.00671 **0.0143**	**$< 10^{-50}$** 0.414 **1.90×10^{-22}**
… + depression PRS x BMI PRS … + BMI PRS x depression …+ depression PRS x depression	-1.92×10^{-5} 0.0190 -0.0244	0.00686 0.0144 0.0146	0.998 0.186 0.0941
… + depression PRS x BMI PRS x depression	0.00869	0.0140	0.536

(Table 5 continued)

Cases			
Null model	See Supplementary Table 3		
… + BMI PRS	**0.238**	**0.0116**	**< 10^{-50}**
… + depression PRS	-0.0162	0.0120	0.178
… + BMI PRS … + depression PRS	**0.238** -0.0129	**0.0116** 0.0117	**< 10^{-50}** 0.270
… + BMI PRS x depression PRS	0.00217	0.0118	0.854
Controls			
Null model	See Supplementary Table 4		
… + BMI PRS	**0.230**	**0.00807**	**< 10^{-50}**
… + depression PRS	0.0171	0.00844	0.0432
… + BMI PRS … + depression PRS	**0.230** 0.0168	**0.00807** 0.00821	**< 10^{-50}** 0.0432
… + BMI PRS x depression PRS	-0.00230	0.00853	0.788

Table 5: Effects of adding variables and interactions to the null model (effects shown in Supplementary Table 2) predicting variance in log-BMI, and stratified analyses of log-BMI within depression cases and controls. Significant ($p < 0.0125$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms.

Chapter 6: Interactions between social environment and polygenic risk scores for body mass index predicting variance in adolescent body mass index

6.1. Introduction

The prevalence of overweight and obesity is increasing in children and adolescents in developed countries, such that over 20% of individuals under the age of nineteen have a body mass index (BMI) > 25 (Ng, Fleming, Robinson, *et al*, 2014). High BMI in this period is associated with psychosocial discrimination, and with socioeconomic hardship and increased cardiometabolic morbidity in later life (Ebbeling, Pawlak & Ludwig, 2002; Gortmaker, Must, Perrin, *et al*, 1993; Hill & Silver, 1995; Reilly & Kelly, 2011; Wabitsch, 2000). Understanding the aetiology of juvenile BMI and of factors influencing change in BMI across adolescence could be informative in developing interventions, and alleviating current and future personal and economic costs (Lobstein, Jackson-Leach, Moodie, *et al*, 2015; Lustig, 2001).

There is robust evidence that variation in BMI is influenced by genetic factors, both from studies of rare variants (such as perturbations in the leptin signalling pathway) and from large genome-wide association studies in adults and in children (Chua Jr, Chung, Wu-Peng, *et al*, 1996; Felix, Bradfield, Monnereau, *et al*, 2016; Llewellyn, Trzaskowski, Plomin, *et al*, 2013; Locke,

Kahali, Berndt, *et al*, 2015; Yang & Barouch, 2007). Evidence from both neuroendocrinological and statistical genetic approaches suggest brain expressed genes may underlie variation in BMI, potentially through controlling energy homoeostasis directly within the body as well as via behavioural processes such as eating and exercise (Finucane, Bulik-Sullivan, Gusev, *et al*, 2015; Lustig, 2001).

The rapid increase in obesity in the last three decades argues for a role of environmental factors, potentially acting to mediate genetic predispositions (Ebbeling, Pawlak & Ludwig, 2002; Lustig, 2001). Parenting is one factor that can influence childhood BMI directly through diet and via learnt food-related behaviours in children, including dietary self-control and regulation of active and sedentary behaviours (Davison & Birch, 2001). However, excessive parental control over food intake behaviours can have a rebound effect when that control is relaxed, such that children over-indulge in previously restricted foodstuffs (Birch & Fisher, 1998; Fisher & Birch, 1999). Much of the research on parenting style and BMI has focussed on the related concepts of parental control and involvement, with some evidence suggesting a controlled disciplinary style and positive parent-child interactions are associated with greater control over BMI levels in childhood (Hughes, Power, Orlet Fisher, *et al*, 2005; Shloim, Edelson, Martin *et al*, 2015; Sleddens, Gerards, Thijs, *et al*, 2011; Vollmer & Mobley, 2013). The presence

of this effect is relatively consistent across studies (although not all studies report a significant effect), but the approaches taken to assessing such effects, and the reported measures of effect, are varied and inconsistent (Shloim, Edelson, Martin *et al*, 2015; Volmer & Mobley, 2013). There are a broad range of measures used in the assessment of parental style (Vollmer & Mobley, 2013), and a focus on specific relationships and populations (specifically maternal influences in White, Western, affluent populations; Gicevic, Aftosmes-Tobio, Manganello *et al*, 2016) that potentially limit the generalisability of findings in the field as a whole.

Parenting style represents one part of the wider influence of socioeconomic environment on child development (Davison & Birch, 2001). Broader measures, such as parental socioeconomic status (SES), may capture this more general influence. In the particular case of BMI, low SES has been associated with higher BMI, particularly in adolescents and young adults (Braddon, Rodgers, Wadsworth, *et al*, 1986; Hardy, Wadsworth & Kuh, 2000; Sundquist & Johansson, 1998; Wang, Kim, Gonzalez, *et al*, 2007). However, reported results vary according to gender, ethnicity and nationality, and there is a potential cohort effect, with null results more common in cohorts ascertained less recently (Blane, Hart, Smith, *et al*, 1996; Laitinen, Power & Jarvelin, 2001; Lauderdale & Rathouz, 2000; Parsons, Power & Manor, 2001; Power & Moynihan, 1987).

There is an observable difference in BMI pre-adolescence between females and males, due in part to the earlier onset of puberty in females, and there is an ongoing debate whether the genetic aetiology of pre-adolescent BMI is sex-specific (Nan, Guo, Warner, *et al*, 2012; Schousboe, Willemsen, Kyvik, *et al*, 2003). A combined analysis of twin studies examining BMI in pre-adolescence did not identify any difference in heritability between sexes, but lacked necessary data (such as opposite-sex dizygotic twin pairs) to make strong conclusions (Nan, Guo, Warner, *et al*, 2012). In contrast, a larger, multi-national study of young adult twin pairs found higher heritability for BMI in females, with results largely consistent across national studies (Schousboe, Willemsen, Kyvik, *et al*, 2003).

Evidence that genetic and environmental influences contribute to the aetiology of BMI before adolescence have prompted a considerable number of studies exploring gene-by-environment interactions (Ahmad, Varga & Franks, 2013). Of these, the interaction between variation in the *FTO* gene and physical activity is the most robust, although the functional mechanism of this interaction remains an area of active research (Andreasen, Stender-Petersen, Mogensen, *et al*, 2008; Franks, Pearson & Florez, 2013). Beyond this interaction, most studies have explored the effects of single variants in the context of many different environments (Ahmad, Varga & Franks, 2013). However, this approach has been limited due to small sample sizes (and

hence low power), inadequate sampling of variation at the genetic locus of interest, and a potentially incorrect hypothesis-driven approach (Dick, Agrawal, Keller, *et al*, 2015; Duncan & Keller, 2011). Recent studies have begun to address this criticism by using gene scores that include associated variants from genome-wide meta-analyses of BMI (Hung, Rivera, Craddock, *et al*, 2014; Qi, Chu, Kang, *et al*, 2012; Qi, Li, Chomistek, *et al*, 2012). This technique can be extended by using weighted polygenic risk scores, which use genome-wide genotypes to construct scores, weighting each variant (commonly by its effect size in genome-wide association study meta-analyses; Purcell, Wray, Stone, *et al*, 2009).

I investigated the independent and interactive effects of parental warmth and discipline and genetic influences on BMI pre-adolescence, and on the rate of change in BMI across adolescence, in a cohort of unrelated adolescents representative of the population of the United Kingdom (the Twins Early Development Study: TEDS; Haworth, Davis & Plomin, 2013; Krapohl, Rimfeld, Shakeshaft, *et al*, 2014; Llewellyn, Trzaskowski, Plomin, *et al*, 2014). The contribution of genetic factors to phenotypic variance was estimated using the most associated polygenic risk score from the largest genome-wide association study meta-analysis in BMI published to date (Locke, Kahali, Berndt, *et al*, 2015). During these analyses, it became apparent that SES contributes to the aetiology of BMI in a manner that overlaps with

the effect of parental warmth and punitive discipline. As such, secondary analyses were performed assessing the effect of SES in the place of parenting.

6.2. Materials and methods

6.2.1. Analysis sample

Data on BMI at 11 years old, child perceptions of parental warmth and punitive discipline, covariates of interest (including parental socioeconomic status) and genome-wide genotype data was available for 3414 unrelated participants from TEDS. The sample was restricted to individuals self-identifying as White Western European (Appendix V; Trzaskowski, Eley, Davis, *et al*, 2013).

6.2.2. Genotype data

Genome-wide genotyping data was obtained in two waves of genotyping, and imputed using minimac3 to the Haplotype Reference Consortium reference data (Appendix V; Fuchsberger, Abecasis & Hinds, 2014; Howie, Fuchsberger, Stephens, *et al*, 2012; McCarthy, Das, Kretzschmar, *et al*, 2016; Trzaskowski, Eley, Davis, *et al*, 2013). Details on quality control and imputation are included in Appendix V. Following QC, genotyped or imputed data from 5,147,884 variants was available on 6710 participants.

6.2.3. Polygenic risk scoring

Polygenic risk scores (PRS) were generated in the TEDS cohort using the results from the largest published meta-analysis of BMI genome-wide association studies (Locke, Kahali, Berndt, *et al*, 2015). The risk score capturing the most variance in BMI at 11 years old was obtained using the default settings in PRSice, which identifies the most predictive score by high-resolution polygenic risk scoring (Euesden, Lewis & O'Reilly, 2015). Eight principal components were included in PRSice analyses to control for population stratification. A binary variable was also included to capture differences between genotyping waves.

Alternative meta-analyses were considered for generating polygenic risk scores. These were the European subset of the same meta-analysis, and the meta-analysis of a smaller cohort of children (Felix, Bradfield, Monnereau, *et al*, 2016; Locke, Kahali, Berndt, *et al*, 2015). However, the cross-ethnic meta-analysis was selected as it was largest and expected to provide the most power. Sensitivity analyses were performed using the alternative sources, with no major differences observed (Appendix V).

Much of the literature on the effect of gene-environment interactions on BMI has examined variation in the FTO gene, particularly the variants rs1558902 and rs9939609 (which have shown strong associations with BMI in different GWAS; Ahmad, Varga & Franks, 2013; Andreasen, Stender-

Petersen, Mogensen, *et al*, 2008; Franks, Pearson & Florez, 2013; Frayling, Timpson, Weedon, *et al*, 2007; Locke, Kahali, Berndt, *et al*, 2015). To allow comparison with the literature, analyses were repeated with each variant in place of the PRS (Appendix V).

6.2.4. Phenotype definition

BMI was calculated from self-reported height and weight, and transformed using a natural logarithm to increase the normality of the distribution. Parenting was defined as the combined results from the child-report sections of the shortened Parental Feelings Questionnaire (PFQ) and the Parental Strategies Questionnaire (PSQ), which measure parental warmth and quality of parental discipline respectively (Deater-Deckard, 2000; Deater-Deckard, Dodge, Bates, *et al*, 1998). The PFQ consists of seven statements designed to assess the warmth of the parent-child relationship (for example, "I feel close to my Mum/Dad", answered *very true* / *quite true* / *not true*). Similarly, the PSQ contains four three-point scales assessing parental actions when the child misbehaved, such as "When I misbehave I am told off or shouted at", answered *not true* / *quite true* / *very true*). Both scales were scored such that higher scores reflected less parental warmth and more punitive discipline respectively. Total scores were standardised and summed to give an overall parenting style variable (Keers, Coleman, Lester, *et al*, 2016).

Covariates were included to control for the effects of age (in days) at assessment, sex, pubertal development and SES at birth, in addition to the covariates used in PRSice analyses. Pubertal development was assessed using the Petersen Pubertal Development Scale (PDS), which has five items assessing the progress of markers of puberty (Petersen, Crockett, Richards, *et al*, 1988). This includes three general questions (for example "*would you say your growth-spurt has not yet begun / barely begun / definitely begun / completed"*) and two sex-specific questions (assessing breast development and menstruation in females, and hair growth and voice deepening in males). The overall score is a mean average of these five items.

A composite measure of SES was derived at the birth of the participants based on measures of maternal and paternal qualifications and occupations, and maternal age at first childbirth, which were standardised and summed (Petrill, Pike, Tom, *et al*, 2004). Specifically, maternal and paternal qualifications at birth were scored from 1 (no qualifications) to 8 (postgraduate qualifications), and occupations were scored from 1 (unskilled) to 9 (managerial). Maternal age at first childbirth was encoded in years. Higher composite scores reflect higher SES.

6.2.5. Statistical analysis

Linear models were constructed in R to test the individual and interactive effects of parenting and genetic risk on BMI at 11. Continuous

variables and covariates (that is, all except sex and genotyping wave) were standardised to produce standardised betas. Pairwise correlations between variables and covariates were calculated to assess the impact of multicollinearity. When interactions between parenting and genetic risk were included in the linear model, all covariate-by-parenting and covariate-by-genetic risk interactions were also included (Keller, 2014).

A subset of the cohort (N = 1943) had BMI data at a later assessment (14 years old, 16 years old, or both). 154 individuals with BMI data at 16 had no age information recorded, so their age at 16 was imputed from age at 11 in a twenty-fold multiple imputation use the *mi* package in R (Graham, Olchowski & Gilreath, 2007; Rubin, 2004; Su, Gelman, Hill*, et al*, 2011). BMI was regressed on time from initial assessment in random effects models (random intercepts and random slopes, one model for each random imputation) using the *lme4* package in R (Bates, Mächler, Bolker*, et al*, 2014). The random coefficient associated with time for each individual was averaged across the twenty models. The average coefficient was then used as the phenotype in further linear models to determine the effects of genetic risk and parenting at 11 years old on change in BMI across adolescence, controlling for covariates as in the previous analysis.

Stratified secondary analyses were performed to assess sex-specific effects on BMI at 11 years and on change in BMI across adolescence. Post-hoc

power calculations were performed using the *pwr* package in R to assess the strength of evidence provided by this study. Specifically, the minimum f^2 values that the analyses had 80% and 90% power to detect were calculated, and the power of the analysis to detect observed f^2 values for social environmental variables and genetic risk were calculated (Cohen, 1988; Selya, Rose, Dierker, *et al*, 2012).

During analysis with parental warmth and discipline, it became clear that SES competed with parenting to explain variance in BMI at 11. When SES was included as a covariate in the model, the proportion of variance explained by parenting was diminished compared to when SES was not included. Analyses were thus repeated with SES as the environmental variable of interest (and parenting as a covariate). In total, twelve analyses were performed, with three basic models (full model, female-only and male-only) for two phenotypes (BMI at 11 and change in BMI across adolescence) with two environments of interest (parenting and SES).

6.2.6. Ethics

Parents provided informed consent for each part of the study before data collection. King's College London's Ethics Committee provided ethical approval.

6.3. Results

6.3.1. Demographics

Demographic data on the full cohort and subsets are available in Table 1. The subset of the cohort in which change in BMI across adolescence was assessed was significantly older than the main cohort. Although both SES and pubertal development were higher in the subset, the difference was not significant after multiple testing (Welch two sample t-test, Bonferroni correction for 17 tests, $p = 0.00294$; Table 1a). Females were significantly more developed than males, reported less harsh and punitive parenting, and had higher BMI, although the difference in BMI was not significant in the subset with multiple BMI assessments (Table 1b).

Correlations between variables included in the analyses are displayed in Supplementary Table 1 in Appendix V. Genotyping wave was strongly correlated with the first principal component ($r = 0.71$), and BMI at 11 was strongly correlated with change in BMI across adulthood ($r = -0.51$). Repeating the analyses without including genotyping wave as a covariate did not alter the conclusions of the study. BMI at 11 is also strongly correlated with the random intercepts used in the construction of the change phenotype ($r \approx 0.9$). As such, BMI at 11 is an integral part of the change phenotype, and the inclusion of this covariate is required for the proper

interpretation of these analyses. No other strong correlations were observed (all $r < 0.5$).

6.3.2. Polygenic risk scoring

Polygenic risk analyses identified a score comprised of 2321 independent variants with $p \leq 0.0032$ in the GIANT 2015 all ancestries GWAS, which predicted a significant proportion of variance in BMI at 11 years of age ($p = 4.55 \times 10^{-32}$, $R^2 = 0.0425$; Locke, Kahali, Berndt, *et al*, 2015). This is consistent with previous estimations of polygenic risk in the TEDS cohort at age 16 (Krapohl, Euesden, Zabaneh, *et al*, 2015). In cross-trait analyses, scores from the GIANT GWAS were not associated with parenting (best threshold = 0.0845, $N_{SNPS} = 18336$, $p=0.0663$, $R^2 = 9.89 \times 10^{-4}$). Similar analyses with SES identified a significant association with the BMI polygenic score when it was optimised for SES (threshold = 0.0795, $N_{SNPS} = 17580$, $p=9.18 \times 10^{-5}$, $R^2 = 0.00445$), but not when it was optimised for BMI (threshold = 0.0032, $N_{SNPS} = 2321$, $p=0.154$, $R^2 = 5.93 \times 10^{-4}$). The score optimised for BMI is reported in all main analyses.

6.3.3. BMI at 11

Higher genetic risk was associated with higher BMI at 11 years old (Table 2). A nominally significant effect of colder and more punitive parenting associated with higher BMI was observed, but was not significant

after Bonferroni correction for twelve tests ($p = 0.05/12 \approx 0.00417$). No interaction between risk and parenting was identified in the main analysis, or after stratifying by sex (Supplementary Table 2a). In secondary analyses with SES as the environment of interest, lower SES was associated with higher BMI. The effect of SES was largely independent of the effect of genetic risk; the inclusion of both variables in the model did not substantially alter the effect sizes observed when each variable was included alone. The interaction between SES and genetic risk was nominally significant, but did not survive correction for multiple testing. No sex-specific effects were observed (Supplementary Table 2b).

6.3.4. Change in BMI during adolescence

Higher genetic risk was associated with a greater increase in BMI (Table 3). Genetic risk was significantly associated with change in BMI in females but not in males (Supplementary Table 3a). However, the interaction between PRS and sex was not significant in the main analysis (p = 0.240). No interaction between genetic risk and parental warmth and discipline was observed. In secondary analyses with SES as the environment of interest, there was no significant main effect of SES. The interaction between genetic risk and SES was nominally significant when both sexes were analysed together and in females only (Supplementary Table 3b). However, neither

the interaction in the full analysis nor that in the female-only subset was significant after correction for multiple testing.

6.3.5. Power

Post-hoc power calculations suggested that the full sample was powered to detect small effects (80% power to detect Cohen's $f^2 = 0.00229$ at age 11, $f^2 = 0.00410$ for change in BMI), as were sex-stratified analyses (BMI at age 11: $f^2 = 0.00452$ and $f^2 = 0.00475$; change in BMI across adolescence: $f^2 = 0.00806$ and $f^2 = 0.00832$ for females and males respectively).

6.4. Discussion

6.4.1. Summary of findings

This study examined the relationship between genetic and social environmental effects (individually and in combination) and two BMI phenotypes: BMI prior to adolescence and the rate of change in BMI between 11 and 16. Genetic effects associated with higher BMI in the largest cohort published to date (the 2015 GIANT consortium meta-analysis) were associated with higher BMI before adolescence, and with a greater increase in BMI across adolescence (Locke, Kahali, Berndt, *et al*, 2015). In contrast, child perceptions of parental warmth and discipline were not significantly associated with pre-adolescent BMI or with change in BMI across adolescence in this study. However, lower parental SES, as a more general

measure of childhood social environment, was associated with higher BMI pre-adolescence, but not with change in BMI.

6.4.2. Limitations

The measures used in this study are unlikely to capture the full component of variance they each represent. The PRS is limited to the effects of common variants on BMI in an additive model, and only to those regions of the genome that are captured adequately by both the GIANT BMI GWAS and the TEDS study genotyping (Appendix V; Locke, Kahali, Berndt, *et al*, 2015). In addition, only a small proportion of the genetic component of variance in BMI was captured by the PRS in this study (7-14%, assuming a heritability of BMI of 30-60%; Polderman, Benyamin, de Leeuw, *et al*, 2015; Yang, Bakshi, Zhu, *et al*, 2015). Finally, these analyses used the optimal PRS (that is, the one explaining the most variance in BMI as a main effect). Multiple PRS, generated using a variety of p-value thresholds, could be used in PRS-by-environment interaction studies. Using the optimal PRS is an analytical choice akin to only examining variables with main effects in any interaction analysis.

An alternative BMI PRS (specifically, one optimised to predict SES) was significantly associated with SES. This demonstrates both that SES can be predicted from genetic data, and that there is an overlap of the genetic influences on BMI and SES (Krapohl & Plomin, 2015). The analyses in this

study used the PRS optimised for BMI (which was not significantly associated with SES) as this best captures the overall influence of the genome on BMI. The modelled interaction term then examines how this genomic effect alters in the presence of the social environment. It would be possible to use the PRS optimised for SES instead. This would focus on the genetic overlap between the two traits; however, the interaction term would then examine how this shared genetic component altered in the presence of the environment with which it is associated. It is unclear what the implications of a significant interaction would be in this case.

Height and weight were ascertained in this cohort via self-report from the participants, as part of a larger questionnaire booklet (Haworth, Davis & Plomin, 2013). Studies comparing self-reported to objectively-measured BMI report a general trend for height to be overestimated and weight to be underestimated, which consequently results in underestimates of BMI (Connor Gorber, Tremblay, Moher, *et al*, 2007). Discrepancies tend to be greater in females, and increase with weight and age (Connor Gorber, Tremblay, Moher, *et al*, 2007). Although this discrepancy has largely been observed in adults, there is also similar evidence reported in children (Goodman, Hinden & Khandelwal, 2000; Strauss, 1999). Therefore, although reported discrepancies tend to be small (the correlation with ascertained measures is approximately 0.8), the precision of the BMI calculation within

this study is likely to be impaired by the self-reported collection. Due to the breadth of phenotype collection in TEDS, it was not practical for all phenotypes to be collected via objective measurement, nor was a subsample measured to allow for correction of the bias.

A central issue of gene-environment interaction studies is the definition of the environment (Dick, Agrawal, Keller, *et al*, 2015). Although the measures of parenting and SES used in this analysis have previously been used successfully to capture their respective constructs, they both differ from possible alternatives (Keers, Coleman, Lester, *et al*, 2016; Petrill, Pike, Tom, *et al*, 2004). Previous research on the effect of parenting style on BMI has examined parental control and involvement. The concepts of punitive parental discipline and parental warmth used in this analysis are similar. However, parental control reflects aspects of both constructive and punitive discipline, whereas the discipline measure used in this analysis focusses on punitive discipline alone. As such, the parenting style measure used in this analysis differs from that used elsewhere in the BMI literature. Furthermore, parenting behaviour is highly complex and multi-faceted, and the measure of parental style used herein can only approximate the overall effect of parenting. In part, the secondary analyses performed using SES as the environment of interest reflects the need to examine the broader effects of the social environment. However, this measure is also only one means of

capturing a complex construct, and different measures of the social environment could yield different results.

Results from the study of change across adolescence need careful interpretation. The random intercepts used in the construction of the change phenotype are highly correlated with BMI at age 11 ($r \approx 0.9$). As such, the inclusion of BMI at 11 as a covariate in the analysis of BMI change largely accounts for influences on pre-adolescent BMI. The non-significant association of social environment with this phenotype may thus reflect the continuation of effects from pre-adolescence, rather than an absence of effect during adolescence.

6.4.3. Interpretation

Genetic risk, modelled as a PRS derived from a cohort mostly comprising adult participants, captures a significant amount of variance both in BMI pre-adolescence and in change in BMI across adolescence. This suggests that the genetic effects on BMI are (at least partly) stable across the lifespan. This is consistent with findings from quantitative genetic studies, which suggest a sizable component of genetic influence on BMI remains from childhood into adulthood, and with high genetic correlations ($r_g = 0.73$) reported in a meta-analysis of GWAS studies in children (Felix, Bradfield, Monnereau, *et al*, 2016; Llewellyn, Trzaskowski, Plomin, *et al*, 2014; Silventoinen & Kaprio, 2009). The TEDS cohort, as a longitudinal study,

could be of considerable value in testing this hypothesised stable component, as genetic associations with BMI would eventually be able to be tested in a within-subject, repeated measures design across the entire lifespan. Such analyses would also be able to consider a broad range of potential environments as covariates and confounds. In the shorter term, growing cohort sizes for adult and particularly for child GWAS of BMI will allow increased precision in the estimate of the genetic correlation. The development of statistical analysis techniques such as partitioned heritability could be extended to identify the precise regions of the genome common between children and adults (and, perhaps more interestingly, those that are distinct); such techniques could also be informative about the biology of these shared and distinct components (Finucane, Bulik-Sullivan, Gusev, *et al*, 2015; Finucane, Reshef, Antitila, *et al* 2017).

Stratified analyses did not suggest a sex-specific effect in this study. Although genetic risk was significantly associated with change in BMI across adolescence in females only, the absence of a significant genetic risk-by-sex interaction in the main analysis suggests this could result from measurement error alone. However, the demographic differences between females and males observed in the cohort argue that stratifying analyses by sex is appropriate in studying influences (genetic and otherwise) on BMI at this age.

The effect of parental warmth and discipline in this study was of nominal significance and did not pass correction for multiple testing. However, when SES is not included in the model, the effect is larger (and would have passed correction for multiple testing had secondary analyses with SES not been performed). The analyses presented have reasonable power. Cohen suggested $f^2 = 0.02$ as a small effect, and all post hoc power calculations show all analyses within this paper had lower f^2 than this (Cohen, 1992). As such, while we cannot exclude an effect of parental warmth and punitive discipline on BMI, these results suggest any such effect is likely to be very small.

In contrast to the effect of parenting in this analysis, parental SES was associated with BMI at 11, suggesting an effect of the social environment from sources other than parenting style alone. Furthermore, the interaction between genetic influences and SES reached a nominal level of significance in the analysis of BMI change across adolescence, and it may be of interest to explore this interaction in a larger cohort. However, conclusions from the analysis using SES must be tempered by the fact that these are secondary analyses, related to the initial hypothesis (that BMI and parenting act together to influence BMI) but not explicitly specified.

The components of variance captured by parenting and SES appear to overlap (as including one in the model diminishes the effect of the other).

The social environment is a complex construct that is likely to reflect and to be influenced by many factors in the wider environment. As such, further investigation to identify the precise component of the social environment that influences juvenile BMI would be of value.

The increasing rate of obesity is a developing public health crisis, and BMI, although imperfect, is a useful proxy for overall metabolic health (Janssen, Katzmarzyk & Ross, 2004). An improved understanding of the factors affecting BMI in late childhood and adolescence could provide useful information in addressing this crisis. It is likely that the majority of obesity does not stem from single factor causes (such as mutations in the leptin system), but rather from the upper extreme of the normal population distribution of BMI (Grarup, Sandholt, Hansen*, et al*, 2014). TEDS is a population cohort, and is not enriched for juvenile obesity. However, studying this cohort can yield insight about the aetiology of BMI within the normal distribution, which may, in turn, be informative about the extremes of that distribution.

The generalisability of genetic findings from a population cohort to the genetics of obesity relies on the assumption that the genetic factors that predispose individuals to extreme BMI influence variance across the BMI spectrum. Yet it may be that distinct genetic influences predispose to extremely low and extremely high BMI (Berndt, Gustafsson, Mägi *et al*,

2013). Other anthropometric traits show evidence for distinct contributions of (often rare or non-additive) genetic effects at the extremes. For example, extremely short individuals show less depletion of a polygenic risk score for height than would be expected from their position in the overall spectrum, indicating a stronger influence of rare or non-additive effects (genetic or otherwise) at this extreme (Chan, Holmen, Dauber *et al*, 2011). However, a reanalysis of the GIANT BMI GWAS, comparing genetic influences in the tails of the distribution with those in the distribution as a whole reported no systematic differences in the additive effects of common variants (although this does not preclude rare variant effects or effects acting in a non-additive manner; Berndt, Gustafsson, Mägi *et al,* 2013). This mirrors similar findings in young adults (Paternoster, Evans, Nohr, *et al*, 2011). Furthermore, the most recent genomic study of anorexia nervosa (which is characterised in part by extremely low BMI) identified negative genetic correlations with extremely high BMI (r_g = -0.29) and with BMI in the normal range (r_g = -0.25; Duncan, Yilmaz, Walters *et al*, 2016). Together, these data provide tentative evidence supporting a role for a shared, common, additive genetic effect across the range of BMI (although genetic correlations are not necessarily transitive; Weiner, Wigdor, Ripke *et al*, 2016).

Common, additive genetic influences on BMI may differ between children, in whom BMI is affected both by growth and by weight changes,

and adults, whose height is stable. As such, using PRS derived from studies of BMI in adult participants may fail to capture important aspects of the common genetic influence on childhood BMI (the reported genetic correlation of 0.73 notwithstanding; Felix, Bradfield, Monnereau, *et al*, 2016). One possible solution to this is to use a combined polygenic risk score from multiple sources (Krapohl, Patel, Newhouse et al, Under Review). This could be performed in a hypothesis-driven manner, such as combining BMI and height scores in adults to better capture variation in children, or machine-learning approaches could be taken to allow the most predictive score to be generated from a large initial set of polygenic risk scores (with appropriate model fitting). In both cases, it would be necessary to use statistical techniques robust to correlated variables (such as penalised regression), as it is possible that the PRS may be correlated.

This study has shown a stable effect of genetic variants (from a meta-analysis predominantly of adult genome-wide association studies of BMI) capturing variance in BMI in children entering adolescence, and also capturing variance in the trajectory of BMI growth across adolescence. SES is associated with BMI pre-adolescence, but parenting style has at most a small effect. The availability of powerful genome-wide meta-analyses and the decreasing cost of obtaining genome-wide genotype data have increased the potential for performing genome-by-environment interaction studies to

identify influential factors underlying important phenotypes in public health.

6.5. References

Ahmad, S., Varga, T. V. & Franks, P. W. (2013) Gene × Environment Interactions in Obesity: The State of the Evidence. *Hum Hered*, **75**, 106-115.

Andreasen, C. H., Stender-Petersen, K. L., Mogensen, M. S.*, et al* (2008) Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. *Diabetes*, **57**, 95-101.

Bates, D., Mächler, M., Bolker, B.*, et al* (2014) Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Berndt, S.I., Gustafsson, S., Mägi, R., *et al* (2013) Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture *Nat Genet,* **45***,* 501-512.

Birch, L. L. & Fisher, J. O. (1998) Development of eating behaviors among children and adolescents. *Pediatrics*, **101**, 539-549.

Blane, D., Hart, C. L., Smith, G. D.*, et al* (1996) Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. *BMJ*, **313**, 1434-1438.

Braddon, F. E., Rodgers, B., Wadsworth, M. E.*, et al* (1986) Onset of obesity in a 36 year birth cohort study. *Br Med J (Clin Res Ed)*, **293**, 299-303.

Chan, Y., Holmen, O. L., Dauber, A., *et al* (2011) Common Variants Show Predicted Polygenic Effects on Height in the Tails of the Distribution, Except in Extremely Short Individuals**.** *PloS Genet***, 7***,* e1002439.

Chua Jr, S. C., Chung, W. K., Wu-Peng, S.*, et al* (1996) Phenotypes of mouse diabetes and rat fatty due to mutations in the OB (leptin) receptor. *Science*, **271**, 994.

Cohen, J. (1988) Statistical power analysis for the behavioral sciences.

---- (1992) A power primer. *Psychological bulletin*, **112**, 155.

Connor Gorber, S., Tremblay, M., Moher, D., *et al* (2007) A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev,* **8,** 307-326

Davison, K. K. & Birch, L. L. (2001) Childhood overweight: a contextual model and recommendations for future research. *Obes Rev*, **2**, 159-171.

Deater-Deckard, K. (2000) Parenting and child behavioral adjustment in early childhood: a quantitative genetic approach to studying family processes. *Child Dev*, **71**, 468-484.

Deater-Deckard, K., Dodge, K. A., Bates, J. E.*, et al* (1998) Multiple risk factors in the development of externalizing behavior problems: group and individual differences. *Dev Psychopathol*, **10**, 469-493.

Dick, D. M., Agrawal, A., Keller, M. C.*, et al* (2015) Candidate gene-environment interaction research: reflections and recommendations. *Perspect Psychol Sci*, **10**, 37-59.

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

Ebbeling, C. B., Pawlak, D. B. & Ludwig, D. S. (2002) Childhood obesity: public-health crisis, common sense cure. *Lancet*, **360**, 473-482.

Euesden, J., Lewis, C. M. & O'Reilly, P. F. (2015) PRSice: Polygenic Risk Score software. *Bioinformatics*, **31**, 1466-1468.

Felix, J. F., Bradfield, J. P., Monnereau, C.*, et al* (2016) Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet*, **25**, 389-403.

Finucane, H. K., Bulik-Sullivan, B., Gusev, A.*, et al* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*, **47**, 1228-1235.

Fisher, J. O. & Birch, L. L. (1999) Restricting access to palatable foods affects children's behavioral response, food selection, and intake. *Am J Clin Nutr*, **69**, 1264-1272.

Franks, P. W., Pearson, E. & Florez, J. C. (2013) Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. *Diabetes Care*, **36**, 1413-1421.

Fraser, J., Skouteris, H., McCabe, M., *et al* (2011) Paternal influences on children's weight gain : a systematic review. *Fathering*, **9,** 252-267.

Frayling, T. M., Timpson, N. J., Weedon, M. N.*, et al* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889-894.

Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. (2014) minimac2: faster genotype imputation. *Bioinformatics*.

Gicevic, S., Aftosmes-Tobio, A., Manganello, J.A., *et al* (2016) Parenting and childhood obesity research: a quantitative content analysis of published research 2009–2015. *Pediatric Obesity*, **17**, 724-734.

Goodman, E., Hinden, E. R., Khandelwal, S. (2000) Accuracy of Teen and Parental Reports of Obesity and Body Mass Index. *Pediatrics,* **106,** 52-58

Gortmaker, S. L., Must, A., Perrin, J. M.*, et al* (1993) Social and economic consequences of overweight in adolescence and young adulthood. *N Engl J Med*, **329**, 1008-1012.

Graham, J. W., Olchowski, A. E. & Gilreath, T. D. (2007) How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, **8**, 206-213.

Grarup, N., Sandholt, C. H., Hansen, T.*, et al* (2014) Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia*, **57**, 1528-1541.

Hardy, R., Wadsworth, M. & Kuh, D. (2000) The influence of childhood weight and socioeconomic status on change in adult body mass index in a British national birth cohort. *Int J Obes Relat Metab Disord*, **24**, 725-734.

Haworth, C., Davis, O. S. & Plomin, R. (2013) Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics*, **16**, 117-125.

Hill, A. J. & Silver, E. K. (1995) Fat, friendless and unhealthy: 9-year old children's perception of body shape stereotypes. *Int J Obes Relat Metab Disord*, **19**, 423-430.

Howie, B., Fuchsberger, C., Stephens, M.*, et al* (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*, **44**, 955-959.

Hughes, S. O., Power, T. G., Orlet Fisher, J.*, et al* (2005) Revisiting a neglected construct: parenting styles in a child-feeding context. *Appetite*, **44**, 83-92.

Hung, C. F., Rivera, M., Craddock, N.*, et al* (2014) Relationship between obesity and the risk of clinically significant depression: Mendelian randomisation study. *Br J Psychiatry*, **205**, 24-28.

Janssen, I., Katzmarzyk, P. T. & Ross, R. (2004) Waist circumference and not body mass index explains obesity-related health risk. *The American Journal of Clinical Nutrition*, **79**, 379-384.

Keers, R., Coleman, J. R., Lester, K. J.*, et al* (2016) A Genome-Wide Test of the Differential Susceptibility Hypothesis Reveals a Genetic Predictor of Differential Response to Psychological Treatments for Child Anxiety Disorders. *Psychother Psychosom*, **85**, 146-158.

Keller, M. C. (2014) Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biol Psychiatry*, **75**, 18-24.

Krapohl, E., Euesden, J., Zabaneh, D.*, et al* (2015) Phenome-wide analysis of genome-wide polygenic scores. *Mol Psychiatry*.

Krapohl, E., Patel, H., Newhouse, S., *et al* (Under Review) Multi-polygenic score (MPS) models predict 11% variation in educational achievement and 5% in general cognitive ability.

Krapohl, E. & Plomin, R. (2015) Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol Psychiatry*.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G.*, et al* (2014) The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences*.

Laitinen, J., Power, C. & Jarvelin, M. R. (2001) Family social class, maternal body mass index, childhood body mass index, and age at menarche as predictors of adult obesity. *American Journal of Clinical Nutrition*, **74**, 287-294.

Lauderdale, D. S. & Rathouz, P. J. (2000) Body mass index in a US national sample of Asian Americans: effects of nativity, years since immigration and socioeconomic status. *International Journal of Obesity*, **24**, 1188-1194.

Llewellyn, C. H., Trzaskowski, M., Plomin, R.*, et al* (2013) Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis. *Int J Obes*, **37**, 1506-1509.

---- (2014) From modeling to measurement: developmental trends in genetic influence on adiposity in childhood. *Obesity (Silver Spring)*, **22**, 1756-1761.

Lobstein, T., Jackson-Leach, R., Moodie, M. L.*, et al* (2015) Child and adolescent obesity: part of a bigger picture. *Lancet*, **385**, 2510-2520.

Locke, A. E., Kahali, B., Berndt, S. I.*, et al* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197-206.

Lustig, R. H. (2001) The neuroendocrinology of childhood obesity. *Pediatr Clin North Am*, **48**, 909-930.

McCarthy, S., Das, S., Kretzschmar, W.*, et al* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*, **advance online publication**.

Nan, C., Guo, B., Warner, C.*, et al* (2012) Heritability of body mass index in pre-adolescence, young adulthood and late adulthood. *Eur J Epidemiol*, **27**, 247-253.

Ng, M., Fleming, T., Robinson, M.*, et al* (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, **384**, 766-781.

Parsons, T. J., Power, C. & Manor, O. (2001) Fetal and early life growth and body mass index from birth to early adulthood in 1958 British cohort: longitudinal study. *BMJ*, **323**, 1331-1335.

Paternoster, L., Evans, D. M., Aagaard Nohr, E., *et al* (2011) Genome-Wide Population-Based Association Study of Extremely Overweight Young Adults – The GOYA Study. *PloS One,* **6**, e24303

Petersen, A. C., Crockett, L., Richards, M.*, et al* (1988) A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc*, **17**, 117-133.

Petrill, S. A., Pike, A., Tom, P.*, et al* (2004) Chaos in the home and socioeconomic status are associated with cognitive development in early childhood: Environmental mediators identified in a genetic design. *Intelligence*, **32**, 445-460.

Polderman, T. J., Benyamin, B., de Leeuw, C. A.*, et al* (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*, **47**, 702-709.

Power, C. & Moynihan, C. (1987) Social class and changes in weight-for-height between childhood and early adulthood. *International Journal of Obesity*, **12**, 445-453.

Purcell, S. M., Wray, N. R., Stone, J. L.*, et al* (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748-752.

Qi, Q., Chu, A. Y., Kang, J. H.*, et al* (2012) Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med*, **367**, 1387-1396.

Qi, Q., Li, Y., Chomistek, A. K.*, et al* (2012) Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men. *Circulation*, **126**, 1821-1827.

Reilly, J. J. & Kelly, J. (2011) Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *Int J Obes (Lond)*, **35**, 891-898.

Rubin, D. B. (2004) *Multiple imputation for nonresponse in surveys*: John Wiley & Sons.

Schousboe, K., Willemsen, G., Kyvik, K. O.*, et al* (2003) Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin Res*, **6**, 409-421.

Selya, A. S., Rose, J. S., Dierker, L. C.*, et al* (2012) A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, **3**, 111.

Silventoinen, K. & Kaprio, J. (2009) Genetics of tracking of body mass index from birth to late middle age: evidence from twin and family studies. *Obesity facts*, **2**, 196-202.

Sleddens, E. F., Gerards, S. M., Thijs, C.*, et al* (2011) General parenting, childhood overweight and obesity-inducing behaviors: A review. *International journal of pediatric obesity*, **6**, e12-e27.

Shloim, N., Edelson, L. R., Martin, N. *et al* (2015) Parenting Styles, Feeding Styles, Feeding Practices, and Weight Status in 4–12 Year-Old Children: A Systematic Review of the Literature**.** *Front. Psychol.***, 6,** 1849

Strauss, R. S. (1999) Comparison of measured and self-reported weight and height in a cross-sectional sample of young adolescents. *Int J Obes Relat Metab Disord,* **23,** 904-908

Su, Y. S., Gelman, A., Hill, J.*, et al* (2011) Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. *Journal of Statistical Software*, **45**, 1-31.

Sundquist, J. & Johansson, S. E. (1998) The influence of socioeconomic status, ethnicity and lifestyle on body mass index in a longitudinal study. *Int J Epidemiol*, **27**, 57-63.

Trzaskowski, M., Eley, T. C., Davis, O. S.*, et al* (2013) First genome-wide association study on anxiety-related behaviours in childhood. *PLoS ONE*, **8**, e58676.

Vollmer, R.L. & Mobley, A. R. (2013) Parenting styles, feeding styles, and their influence on child obesogenic behaviors and body weight. A review. *Appetite,* **71,** 232-241.

Wabitsch, M. (2000) Overweight and obesity in European children: definition and diagnostic procedures, risk factors and consequences for later health outcome. *Eur J Pediatr*, **159 Suppl 1**, S8-13.

Wake, M., Nicholson, J.M., Hardy, P., *et al* (2007) Preschooler obesity and parenting styles of mothers and fathers: Australian national population study. *Pediatrics,* **120**, e1520-1527.

Wang, M. C., Kim, S., Gonzalez, A. A.*, et al* (2007) Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *J Epidemiol Community Health*, **61**, 491-498.

Yang, J., Bakshi, A., Zhu, Z.*, et al* (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*, **47**, 1114-1120.

Yang, R. & Barouch, L. A. (2007) Leptin signaling and obesity: cardiovascular consequences. *Circ Res*, **101**, 545-559.

6.6. Tables

6.6.1. Table 1a

Variable	Full Cohort	BMI Change Subset	t-test *p*
N	3414	1943	-
Female sex (%)	51.3	53.7	0.0881
Age (mean, SD)	**11.3 (0.695)**	**11.5 (0.616)**	**1.52×10^{-27}**
SES at birth (mean, SD)	0.235 (0.961)	0.289 (0.958)	0.0474
Pubertal development (mean, SD)	1.69 (0.554)	1.74 (0.567)	0.00310
BMI (mean, SD)	17.8 (3.07)	17.9 (3.101)	0.188
Parenting style (mean, SD)	0 (1.71)	-0.0733 (1.70)	0.131
Change in BMI (mean, SD)	-	0.580 (0.0995)	-

Table 1a: Demographic data of the full cohort and subset in which change in BMI was studied. Age was significantly higher in the subset after multiple testing correction; SES and pubertal development were greater in the subset, but did not pass multiple testing correction (Welch two sample t-test, $p = 0.05/17 = 0.00294$).

Table 1b

Variable	Full Cohort		t-test *p*
	Females	**Males**	
N	1750	1664	-
Age (mean, SD)	11.3 (0.693)	11.2 (0.697)	0.550
SES at birth (mean, SD)	0.232 (0.957)	0.238 (0.965)	0.862
Pubertal development (mean, SD)	**1.85 (0.589)**	**1.52 (0.457)**	**1.79×10^{-72}**
BMI (mean, SD)	**17.9 (3.22)**	**17.6 (2.90)**	**0.00179**
Parenting style (mean, SD)	**-0.152 (1.69)**	**0.160 (1.73)**	**9.87×10^{-8}**
Variable	**BMI Change Subset**		**t-test *p***
	Females	**Males**	
N	1043	900	-
Age (mean, SD)	11.4 (0.633)	11.5 (0.584)	0.126
SES at birth (mean, SD)	0.294 (0.947)	0.283 (0.970)	0.805
Pubertal development (mean, SD)	**1.90 (0.600)**	**1.56 (0.465)**	**1.06×10^{-42}**
BMI (mean, SD)	18.0 (3.33)	17.8 (2.83)	0.206
Parenting style (mean, SD)	**-0.234 (1.67)**	**0.113 (1.72)**	**7.53×10^{-6}**
Change in BMI (mean, SD)	0.581 (0.103)	0.580 (0.0951)	0.764

Table 1b: Demographic data of the full cohort and subset in which change in BMI was studied, stratified by sex. Females were significantly more developed and reported less harsh and punitive parenting, and exhibited high BMI (the last of which only in the full cohort; Welch two sample t-tests, $p = 0.05/17 = 0.00294$).

6.6.2. Table 2

Coefficient	B	SE	*p*	Adjusted R^2
BMI at 11 years old, with parenting				
Null model	Supplementary Table 2a			0.0667
(Null model) + Parental style	0.0378	0.0167	0.0239	0.0678
(Null model) **+ BMI PRS**	**0.210**	**0.0162**	**1.59×10^{-37}**	0.110
(Null model) + Parental style **+ BMI PRS**	0.0360 **0.210**	0.0163 **0.0162**	0.0273 **1.83×10^{-37}**	0.111
(Null model) + Parental style x BMI PRS	0.00642	0.0172	0.709	0.113
BMI at 11 years old, with SES				
Null model	Supplementary Table 2b			0.0628
Null model + **SES**	**-0.0729**	**0.0167**	**1.33×10^{-5}**	0.0678
Null model + **BMI PRS**	**0.211**	**0.0162**	**7.95×10^{-38}**	0.107
Null model **+ SES** **+ BMI PRS**	**-0.0682** **0.210**	**0.0163** **0.0162**	**3.11×10^{-5}** **1.83×10^{-37}**	0.111
Null model + SES x BMI PRS	-0.0336	0.0165	0.0413	0.112

Table 2: Effects of adding variables and interactions to the null model predicting variance in BMI at 11 years old, with parental style (top) and SES (bottom) as the environments of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

6.6.3. Table 3

Coefficient	B	SE	*p*	Adjusted R^2
BMI change, with parenting				
Null model	Supplementary Table 3a			0.277
(Null model) + Parental style	-0.00848	0.0196	0.666	0.277
(Null model) **+ BMI PRS**	**0.0902**	**0.0197**	**4.96×10^{-6}**	0.284
(Null model) + Parental style **+ BMI PRS**	-0.00969 **0.0903**	0.0195 **0.0197**	0.620 **4.84×10^{-6}**	0.284
(Null model) + Parental style x BMI PRS	0.000551	0.0207	0.979	0.285
BMI change, with SES				
Null model	Supplementary Table 3b			0.277
Null model + SES	0.0106	0.0196	0.591	0.277
Null model **+ BMI PRS**	**0.0904**	**0.0197**	**4.70×10^{-6}**	0.284
Null model + SES **+ BMI PRS**	0.00965 **0.0903**	0.0195 **0.0197**	0.622 **4.84×10^{-6}**	0.284
Null model + SES x BMI PRS	-0.0494	0.0205	0.0159	0.289

Table 3: Effects of adding variables and interactions to the null model (uppermost line) predicting variance in the linear trajectory of change in BMI between 11 and 16 years old, with parental style (top) and SES (bottom) as the environment of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

Chapter 7: Discussion

7.1 Summary of the work

The work presented in this thesis seeks to demonstrate the value of information-rich genomic data in exploring questions about the interplay of genetic and environmental influences on behaviour. Within that overarching aim, it has explored two specific themes: employing exposed-only GWAS to identify biological predictors of response to CBT, and assessing genome-environment relationships using polygenic risk scores.

The genetic study of response to CBT (therapygenetics) has relied almost entirely on the candidate gene method (Eley, Hudson, Creswell, *et al*, 2012; Lester & Eley, 2013; Lueken, Zierhut, Hahn, *et al*, 2016). The GWAS described in Chapter 3 was the first in the field, and the genome-wide studies that follow in Chapter 4 extend these investigations to gene expression (along with a sister paper that investigated change in expression across the course of treatment; Roberts, Wong, Keers, *et al*, Under Review). Chapter 3 partially achieved its initial aim (to identify any large-effect common variants able to be captured using a low-coverage microarray). The analysis presented had sufficient power to examine large effects on response to CBT (4% of phenotypic variance), and the null results suggest such effects are unlikely to exist. However, the analysis was underpowered to detect all but the largest effects from common variants seen elsewhere in the

behavioural genetic literature. For example, the 97 loci significantly associated with BMI only explain 2.7% of the variance combined, and the 108 loci significant loci in schizophrenia explain 3.4% of the variance (Locke, Kahali, Berndt, *et al*, 2015; Schizophrenia Working Group of the Psychiatric Genomics, 2014). Although the APOE locus accounts for approximately 5% of the phenotypic variance in late-onset Alzheimer's disease, this is an unusually large effect for a complex behavioural trait (Escott-Price, Shoai, Pither *et al*, 2017). If common variants of similar effect influence response to CBT, the analysis in Chapter 3 did not detect them, most likely due to a lack of power. The results of Chapter 3 are concordant with a polygenic model of behavioural phenotypes, comprising many variants of very small individual effect (Chabris, Lee, Cesarini, *et al*, 2015; Visscher, Goddard, Derks, *et al*, 2012). However, they are also concordant with the absence of a genetic component to CBT. The analysis did not achieve its secondary aim to quantify such a component, and polygenic risk scores derived from relevant psychiatric phenotypes were not associated with treatment outcome. Significant associations with individual SNPs had been previously reported in the candidate gene literature, some of which were tested in Chapter 3 as a secondary aim. The lack of effect shown by these variants calls into question the generalizability of these earlier findings (Lester & Eley, 2013).

Chapter 4 sought to increase power to detect significant predictors by combining two sources of information, genetic variants and measures of gene expression (Ritchie, Holzinger, Li, *et al*, 2015). Again, the analysis had sufficient power only to examine large effects on response to CBT, and the null results suggest such effects are unlikely to exist. The absence of a significant predictor in these results argues that no single genetic locus accounts for large amounts of variance in response to CBT, whether this quantified as the expression level of the related RNA transcript, or as a context-dependant eQTL. In addition, increasing power through dimensional reduction (specifically, grouping transcript expression levels by literature-driven and data-driven means) did not yield associations, and polygenic risk scoring between the results of Chapters 3 and 4 did not yield significant predictors. As such, Chapter 4 did not achieve its aims to identify correlates of treatment response, beyond casting doubt on the existence of loci of large effect.

Polygenic risk scores can be used as a proxy for the overall effect of common genetic variation on phenotypic variation (in an additive model; McGrath, Mortensen, Visscher, *et al*, 2013). This was the principal aim of Chapter 5, which sought to use risk scores to assess the genetic contribution to the complex relationship between BMI and depression, using a large population cohort. A significant correlation was observed, such that

individuals with depression had higher BMI on average, but polygenic risk scores were associated with BMI, and with depression, only in within-trait analyses. The correlation between the traits was not attenuated by concurrently modelling the genetic influences on either trait, neither through diminishing the main effects nor through the presence of a significant gene-by-trait interaction. As such, the results of this analysis did not support the hypothesis that the relationship is affected by a shared common additive genetic component. The strength of this conclusion is limited by the small amounts of variance captured by the polygenic risk scores used, although similar estimates of the heritability of BMI in depression cases and controls provides further counter-evidence. The estimated genetic correlation between depression and BMI was also not significantly different from zero, yet this may be limited by the power of the study. Thus, the combined evidence from the study suggests that any shared common additive genetic component between BMI and depression is small. This suggests the increased BMI observed in depression cases results from another cause, of which many candidates exist, including shared physical comorbidities or a (potentially bidirectional) causal relationship, although the influence of gene-environment correlation and rare or non-additive genetic effects cannot be excluded (Luppino, de Wit, Bouvy, *et al*, 2010).

The approach taken in Chapter 5 examines the relationship between broad measures of the overall influence of genetic and non-genetic factors. Polygenic scores may also be valuable in assessing the interaction of effects from candidate environments with those from the genome as a whole, and this concept was tested in Chapter 6, assessing the independent and combined impact of social environments and genomic risk on BMI in late childhood and across adolescence. Known genomic influences on BMI were associated with both pre-adolescent BMI and change in BMI across adolescence. These influences were identified from a study composed predominantly of adults; as such, this may suggest a stable genetic component to variance in BMI, although this was not tested explicitly within Chapter 6. The initial aim of the investigation was focussed on parental style, which showed only a small effect on BMI pre-adolescence (such that the effect was not significant after correction for multiple testing), and did not affect change in BMI across adolescence, nor did this effect differ significantly dependent on the polygenic risk score for BMI. However, the inclusion of measures of parental socioeconomic status in the model diminished the effect of parenting style. A more punitive parenting style was significantly correlated with lower socioeconomic status in the cohort, and as such the components of variance captured by these measures are not independent of one another, and may reflect a more general effect of the social environment. Repeating the analyses with socioeconomic status (as a

secondary analysis) identified an effect on BMI pre-adolescence. The effect of interactions between genomic risk and socioeconomic status on pre-adolescent BMI and change in BMI across adolescence were of nominal significance, although not sufficiently strong to pass correction for multiple testing. Together, these argue for a general effect of social environment on BMI in adolescence (which may involve parental style in part, but must have contributions from other sources), which is smaller than the combined effect of the genome, and which differ only minimally in the context of the polygenic risk score for BMI. Stratifying the analyses by gender, as a secondary aim, did not yield qualitatively different results.

7.2 Limitations and points of debate

The effects of individual genetic variants on behaviour are very small, and (at least at the moment) the summed effects of those variants only explain part of the estimated heritability (Dudbridge, 2013; Wray, Yang, Hayes, *et al*, 2013). Similarly, individual environmental influences, while larger in effect than those of individual SNPs, are often smaller than the overall effect of the genome (although this is confounded by differences in the measurement of genetic and environmental effects; Dick, Agrawal, Keller, *et al*, 2015). The power required to estimate small effects is considerable, and they are vulnerable to confounding (Dudbridge, 2013). The

studies within this thesis attempt to address these concerns, but are each impaired by them to some extent.

Genetic research into CBT suffers considerably from a lack of participants (Lester & Eley, 2013; Lueken, Zierhut, Hahn, *et al*, 2016). Although the studies described in Chapters 3 and 4 are drawn from the largest cohorts of their type to date, they are not an exception to this. To estimate heritability, and to identify individual genetic variants and gene transcripts associated with response to psychological therapies, will require the recruitment of thousands of participants (Visscher, Hemani, Vinkhuyzen, *et al*, 2014). This is non-trivial in the study of psychological therapies. Using a prospective design in examining response to CBT is beneficial for studying genetics in the context of an environmental exposure, but it results in a considerable degree of attrition. Participants drop out of psychological therapy for a wide variety of reasons. Poor therapy response is a major cause of dropout, and this is of clear relevance to studying differential response (Loerinc, Meuret, Twohig, *et al*, 2015). Studies can be designed to allow the appropriate inclusion of participants who drop out (through intention-to-treat analyses and last-observation-carried-forward), but this comes at an increased financial burden, which in turn limits the cohort size attainable.

Alongside attrition, the power of therapygenetic cohorts is limited by heterogeneity. Psychological therapy is focussed on the needs of the

participant, meaning that the same therapy given to different participants will result in a different therapeutic experience, even when ostensibly the same program is delivered by the same therapist (Rice & Greenberg, 1984; Shirk & Saiz, 1992). Obtaining the necessary cohort size to conduct genetic studies is likely to require combining multiple smaller cohorts, with participants treated by different therapists, and experiencing a variety of different therapeutic regimens. Again, the studies described in Chapters 3 and 4 are vulnerable to this limitation. However, this must be balanced against the generalisability of findings. The provision of psychological therapy is riven with heterogeneity; any predictor must be robust to that heterogeneity. As such, the brute force approach of overwhelming heterogeneity with sample size, while crude, may be the most appropriate approach in this instance.

Chapters 3 and 4 study the phenotype of "response to psychological therapy". However, this is far from an objective phenotype. No gold-standard approach to the definition of "therapy response" as a phenotype exists, and even the precise terms are not well-defined (Creswell & Waite, 2016; Loerinc, Meuret, Twohig, *et al*, 2015). "Response" and "remission" are occasionally used interchangeably, but usually refer to distinct features. Response is often defined as a change in a given measure from baseline, but the exact size of the change required, and whether this is measured in

absolute units or as a percentage, varies (Loerinc, Meuret, Twohig, *et al*, 2015; Prien, Carpenter & Kupfer, 1991). In contrast, "remission" is usually a dichotomous variable meaning that the participant has passed some low threshold of severity. This distinction means that, depending on the definitions used in the study, it is often possible for an individual to respond but not remit or *vice versa* (Loerinc, Meuret, Twohig, *et al*, 2015). Remission often implies the individual no longer meet diagnostic criteria, although even this simple definition shows considerable heterogeneity in application, with the diagnosis in question ranging from the primary source of therapeutic concern to all possible inclusion criteria for the study (Creswell & Waite, 2016). Given the high comorbidity seen in individuals with anxiety disorders (particularly children), these different meanings of remission introduce heterogeneity when comparing between studies, and this is compounded yet further when comparing studies of response and remission (Creswell & Waite, 2016; Waite & Creswell, 2014). Furthermore, although CBT studies are based on a core framework, programs vary in length and may focus on different aspects of treatment (Barrett, Turner & Lowry-Webster, 2000; Bögels, 2008; Kendall & Hedtke, 2006; Schneider, Blatter-Meunier, Herren, *et al*, 2013). Again, the resulting heterogeneity hampers combined analysis across programs (Creswell & Waite, 2016; Haby, Donnelly, Corry, *et al*, 2006; Hudson, Keers, Roberts, *et al*, 2015; Loerinc, Meuret, Twohig, *et al*, 2015).

As a direct result of the way in which they are defined, response and remission are measure-dependent (Prien, Carpenter & Kupfer, 1991). Even within this thesis, however, the two empirical chapters studying psychological therapy response employ different measures of the phenotype, and this reflects the use of a variety of measures in studies more generally (Loerinc, Meuret, Twohig, *et al*, 2015). This variation has definable effects; those studies combining multiple measures of the same type report significantly lower response rates on average, as do intention-to-treat analyses compared to those limited to treatment completers only (Loerinc, Meuret, Twohig, *et al*, 2015). Combining measures using different reporters or measurement techniques (for example, a diagnostic scale and a behavioural marker) increases the validity of the outcome measure, as does the independent assessment of response (Loerinc, Meuret, Twohig, *et al*, 2015).

As such, convincing arguments have been made that the most powerful and robust method to study response is to combine multiple measures (preferably from different modalities) to create a single response variable (Loerinc, Meuret, Twohig, *et al*, 2015). It will be noted that this is not the approach taken within this thesis. Power is an important concern in studying psychological therapy response, and Chapters 3 and 4 used a continuous measure of response (rather than creating a dichotomy) for this

reason. However, response was encoded using a single measure in both cases. The combined measure of response suggested in the literature may be more valid as a measure of improvement following therapy. However, unless the set of measures that are combined to create this composite are used as standard across studies, the gain in power from this combination is arguably negated by the loss in interpretability and comparability between studies. If the outcome of therapygenetic studies is to be useful, it must be understandable outside of the context of a single study (Eley, 2014; Paulus, 2015). This said, it is likely that the phenotype definition in Chapters 3 and 4 is anti-conservative, resulting in better treatment outcomes than would be obtained using a combined measure of response or intention-to-treat analyses.

Individuals recruited to receive psychological therapy are usually severely unwell (hence their need for therapy). This generates a potential confound, namely regression to the mean (Barnett, van der Pols & Dobson, 2005). Illness severity fluctuates over time, and individuals seeking therapy are more likely to be at the severe end of that fluctuation. As such, when they are assessed several weeks later (following the end of their therapy), they may be less severely ill simply because they have descended from the peak of their fluctuation, rather than due to the therapy. In studying the efficacy of therapy, this can be (and has been) countered using a matched waitlist

control group (Hofmann & Smits, 2008; James, James, Cowdrey, *et al*, 2013). The same approach could be taken to account for regression to the mean when investigating predictors of response to therapy. However, allocating individuals to a provably inferior treatment condition (such as a waitlist) would be ethically unsound, as scientific gains would come at the expense of prolonged distress to the participants (Emanuel & Miller, 2001). The studies contained in Chapters 3 and 4 are thus exposed-only studies, where all participants received treatment. The results must be interpreted carefully. The studies examined change in illness severity from the onset of a shared exposure to its completion (and beyond into follow-up), but this change is not necessarily a response to that exposure. However, a robust predictor of treatment response would have value regardless of its relevance to the underlying mechanisms. This reflects a general point concerning statistical modelling, namely that models are rarely "right", but can be useful – a robust predictor of regression to the mean would be largely irrelevant to the mechanisms of therapy response, but would be of considerable value in clinical decision making (Box, 1976; Paulus, 2015).

A further consequence of the absence of a control group in Chapters 3 and 4 is that all participants are exposed to the specific environment (CBT). Genetic variants (or differences in RNA transcript expression) might be associated with change in severity independently of the environmental

exposure (including variants associated with regression to the mean as discussed above). Accordingly, such case-only studies cannot truly investigate gene-environment interactions, as the environmental exposure does not vary between individuals (For much the same reason, such studies also do not examine gene-environment correlation.) A gene-environment interaction can be inferred by assuming that the change in severity results only from exposure to CBT – if this is true, then a genetic association with change in severity is an association with exposure, and as such represents an instance of gene-environment interplay. However, to demonstrate this robustly would require performing a parallel GWAS in an untreated control group and comparing the results, which has ethical impediments as discussed above. One possibility to address this would be to take a naturalistic approach. CBT is commonly over-subscribed, and so waitlists exist independently of the necessity for a control condition, and could be used for parallel genetic studies. However, there would be potential limitations of this approach, not least that there may be a negative "nocebo" effect of being placed on waitlist (Furukawa, Noma, Caldwell, *et al*, 2014).

In general, genomics can overwhelm the subtlety of behavioural phenotypes by brute force. With enough power and robust statistical methods, genetic effects can be identified in spite of heterogeneity produced by confounders not included in the model (Manchia, Cullis, Turecki, *et al*,

2013). Response to cognitive behavioural therapy is an example of this: response results from factors relating not just to the recipient, but also to the therapist (and to the interaction between them; Ackerman & Hilsenroth, 2003; Ilardi & Craighead, 1994; Martin, Garske & Davis, 2000). However, with enough power, it should be possible to identify genetic effects (which are recipient-based) regardless of the context of the therapist.

Chapter 4 studied RNA transcripts as a measure of the combined effect of genetic and environmental influences on response to therapy. It extended therapygenetic investigations a step closer to studying the multifactorial and highly interactive network of influences that is likely to underlie therapeutic response. However, there are a number of factors that make RNA transcripts less attractive than DNA variation as a source of biological predictors. Chief among these is the dynamic nature of RNA expression. Unlike genetic variation (which is stable across the lifespan, and effectively identical between tissues), RNA expression levels can show considerable temporal variability within an individual (Raj & van Oudenaarden, 2008). In addition to stochastic variability, variation in gene expression can result from confounds such as temperature or general health. Statistical noise thus has a greater effect on differences in gene expression than on differences in genetic variation between individuals. Gene expression is also spatially variable, and not all tissues will be of interest in

studying behavioural phenotypes. There remains controversy concerning which of blood, saliva and buccal cells represent the best proxy for variation in brain tissues (Davies, Volta, Pidsley, *et al*, 2012; Lowe, Gemma, Beyan, *et al*, 2013; Smith, Kilaru, Klengel, *et al*, 2015; Sullivan, Fan & Perou, 2006). Brain tissue would be preferable, but is inaccessible as a living tissue. Blood is inferior to brain tissue, but is feasible to use in living participants and shows more consistency in collection compared to buccal cells (Hansen, Simonsen, Nielsen, *et al*, 2007). Saliva provides a less invasive method of collection than blood and is also more consistent than buccal cells. When combined, the temporal and spatial variability of gene expression partly negates the increased power individual gene expression measures might be expected to show over individual genetic variants. The logistic and technical difficulties of collecting RNA samples are likely to limit the possible cohort size attainable, rendering expression studies inferior to genetic studies in this context.

The latter chapters of the thesis make use of polygenic risk scoring to explore genome-environment interactions. Specifically, the optimal risk score is used, as determined by high-resolution scoring across 10000 thresholds (Euesden, Lewis & O'Reilly, 2015). The use of the optimal score as a proxy for the genome is an analytical choice, and a range of risk scores could be used instead (at the cost of increasing the number of tests performed). Using

the optimal score is conceptually similar to the general two-stage strategy of only testing for interactions between variables whose main effects pass a given threshold for significance (Bourgon, Gentleman & Huber, 2010; Dai, Kooperberg, Leblanc, *et al*, 2012; Ege & Strachan, 2013; Kooperberg & Leblanc, 2008). Such filtering strategies increase power at the expense of neglecting potential interactions with variables that do not pass the filter (Domingue & Boardman, 2016). Testing for interactions using a genome-wide risk score differs from testing for interactions using individual variants because all possible risk scores are correlated (as is addressed in the use of the empirically-derived family-wise error correction from Euesden, Lewis & O'Reilly, 2015). Therefore, using the optimal risk score explicitly tests the hypothesis that the overall genetic component of phenotypic variance interacts with an environmental effect, while controlling appropriately for multiple testing.

Optimisation maximises the phenotypic variance explained by the polygenic risk score as a main effect. Nonetheless, none of the scores used in this thesis captures a majority of the heritability of the relevant trait. Even scores generated from the GIANT BMI GWAS (over 300,000 individuals) into the UK Biobank (21039 individuals in Chapter 5) only captured around 5.6% of the variance (9-18% of the estimated 30-60% of variance accounted for by genetic influences; Polderman, Benyamin, de Leeuw, *et al*, 2015; Yang,

Bakshi, Zhu, *et al*, 2015). As such, polygenic risk scores can only be considered weak proxies for the effect of the genome at present. This is, to an extent, a temporary limitation that will be reduced by the use of larger base GWAS in polygenic scoring, as is demonstrated by the increase in variance explained by polygenic risk scores in more recent schizophrenia GWAS (Dudbridge, 2013; Schizophrenia Working Group of the Psychiatric Genomics, 2014). However, the construction of polygenic risk scores, particularly the accumulation of errors through summing the effects of many variants, may limit the amount of variance they can ever explain (Dudbridge, 2013; Wray, Yang, Hayes, *et al*, 2013). An alternative would be to use whole-genome regression to construct genetic best linear unbiased predictors (gBLUPS), which estimate the overall genomic influence on a trait as a random effect from a GRM (de los Campos, Vazquez, Fernando, *et al*, 2013). However, this method requires individual-level genotype data for prediction (rather than summary statistics), which limits its generalisability and applicability between datasets (Moser, Lee, Hayes, *et al*, 2015; Speed & Balding, 2014).

The study of genetic effects in this thesis focusses on common variants present on (or imputable from) microarrays. Furthermore, the effect of these variants is modelled additively, both when considered individually in GWAS, and when combined in genome-environment interaction analyses. It

is assumed that the effects of the genome can be modelled as a single risk score, constructed from common variants. However, this assumption does not take into account the role of rare variation, nor of non-additive effects.

There is a growing appreciation that variants across the allelic frequency spectrum contribute to the genetic component of variance (Gratten, Wray, Keller *et al*, 2014; Hoischen, Krumm & Eichler, 2014). Much of the genome-wide assessment of rare variants has come from whole exome sequencing studies, with the most in-depth studies in psychiatric genomics focussing on schizophrenia and autism (Hoischen, Krumm & Eichler, 2014). Current cohort sizes are not yet sufficient to make broad statements about the role of rare variation in behavioural disorders, but findings from these initial studies suggests they might contribute considerably to the genetic component of variance. In reference to depression specifically, it has required very large cohort sizes for common variant discovery to begin, and so it is likely that even larger cohorts will be needed to understand the genome-wide contribution of rare variation in this disorder. An analysis of the genetic component of variance in BMI suggested that rare (MAF < 0.01) variants able to be imputed from genome-wide microarrays capture around 5% of variance in BMI (Yang, Bakshi, Zhu *et al,* 2015). Given that the same study estimated the heritability of BMI to be approximately 27%, and that around 32% of variation at rare variants couldn't be captured by imputation,

this rare component in BMI might capture as much as a third of the total heritability (Yang, Bakshi, Zhu *et al*, 2015). As such, it is reasonable to expect a sizable proportion of the genomic influence on the traits studied in this thesis will stem from rare variation inadequately assayed by the methods used herein.

Similarly, non-additive genetic effects may influence the traits examined in this thesis. For example, a meta-analysis of twin studies reported the average correlation between dizygotic twin pairs for recurrent depressive disorder was less than half that of monozygotic twin pairs, suggesting that the correlation between identical twins may be increased by non-additive genetic effects (Polderman, Benyamin, de Leeuw, *et al*, 2015). This finding is potentially confounded by the age of assessment in the constituent twin studies of this meta-analysis. Specifically, studies in children and adolescents were more likely to report dizygotic twin pair correlations indicative of shared environmental effects, whereas studies in adults were more likely to support non-additive genetic effects. All of data in the recurrent depressive analysis assessed by Polderman et al (2015) comes from adults, compared to 54% in the depressive episode analysis (the results of which are most consistent with a model of additive genetics alone), and 14% in the "other anxiety disorders" analysis (which reported a pattern of twin correlations supporting shared environmental effects; Polderman,

Benyamin, de Leeuw, *et al*, 2015). One explanation for this may be that there is a stable non-additive genetic effect, which is masked in studies of children by the presence of shared environmental effects. As each member of a twin pair becomes exposed to different environments across adolescence, the shared effect of environment on a given trait may decline, allowing the non-additive environmental effect to be observed. This hypothesis is difficult to test with the twin method, which lacks sufficient sources of variance to fit the full ACDE model (Neale & Cardon, 1992). However, studies of genome-wide gene-gene interactions are becoming increasing tractable thorough computational and analytical developments (Wei, Hemani & Haley, 2014). As such, exploring the role of non-additive genetic effects on depression from a genotype-level perspective may represent a valuable, if challenging, avenue for future research.

The interaction models used in chapters 5 and 6 assume that the interacting variables are independent. Gene-environment correlation violates this assumption, and could bias the results and interpretation of the interaction term (Dudbridge & Fletcher, 2014; Jaffee & Price, 2007; Purcell, 2002). The effects of gene-environment interactions are likely to be small, and as such even correlations that do not meet significance may have an effect on the interaction term (Jaffee & Price, 2007). Given the relative novelty of polygenic-risk-by-environment interaction studies, it remains unclear to

what degree correlations will affect interaction terms when the genetic variable captures variation genome-wide. Furthermore, genetic effects from genome-wide association studies are heavily dependent on the estimation of the effect of causal variants via marker variants in linkage disequilibrium. This introduces a misclassification problem, whereby the true effect of a locus is estimated with a degree of error due to imperfect linkage disequilibrium between the causal variant and its markers (Dudbridge & Fletcher, 2014). This misclassification can lead to spurious marker-by-environment interactions when the causal variant is associated with the environment of interest, even when no causal variant-by-environment interaction exists (Dudbridge & Fletcher, 2014). Polygenic scores are constructed from multiple markers, and so the misclassification problem is also likely to have an effect on polygenic-risk-by-environment interactions. As such, and despite the fact that the correlations between the interacting variables in both chapters were non-significant, it is possible that the effects of the interaction models in this thesis are confounded by gene-environment correlations. It would be valuable to define theoretically the underlying mathematics of this bias in polygenic-risk-by-environment interactions.

A motivating factor behind the studies presented in Chapters 5 and 6 was the developing public health crisis caused by increasing rates of obesity. However, some of the negative consequences of obesity may result from

generally poor physical health, and the suitability of BMI as a measure of this phenotype is debatable. BMI is biassed by muscle mass (which is heavier than fat mass) and oversimplifies the complexities of human body shape (Nevill, Stewart, Olds, *et al*, 2006). Alternative phenotypes, such as waist circumference or fat mass may better reflect the health risk represented by obesity (Janssen, Katzmarzyk & Ross, 2004). Nonetheless, high BMI is associated with the negative health outcomes from obesity, and also with comorbidities such as depression (Luppino, de Wit, Bouvy, *et al*, 2010; Taylor, Ebrahim, Ben-Shlomo, *et al*, 2010). Given the relative ease of defining BMI, and the need for pragmatic measures to obtain the required cohort sizes for genomic studies, BMI remains a reasonable proxy for studying the genetics of obesity in the context of comorbidity and environmental influences (Taylor, Ebrahim, Ben-Shlomo, *et al*, 2010).

7.3 The work in the context of the field, and future directions suggested

Studying response to therapy represents a prospective, controllable approach to the exposed-only genomic method, which has been successful in the study of post-traumatic stress disorder (where the exposure is unpredictable and uncontrollable; Ashley-Koch, Garrett, Gibson, *et al*, 2015; Guffanti, Galea, Yan, *et al*, 2013; Logue, Baldwin, Guffanti, *et al*, 2013; Nievergelt, Maihofer, Mustapic, *et al*, 2015; Solovieff, Roberts, Ratanatharathorn, *et al*, 2014; Stein, Chen, Ursano, *et al*, 2016; Sumner,

Pietrzak, Aiello, *et al*, 2014; Wolf, Rasmusson, Mitchell, *et al*, 2014; Xie, Kranzler, Yang, *et al*, 2013). The use of robust statistical genetic methods in therapygenetics should improve the quality of evidence produced by such studies. This is not of abstract importance; anxiety is a considerable burden on the individual and society both financially and regarding quality of life, and can be a chronic issue from childhood into adulthood (Baxter, Vos, Scott, *et al*, 2014; Bittner, Egger, Erkanli, *et al*, 2007; Fineberg, Haddad, Carpenter, *et al*, 2013; Gregory, Caspi, Moffitt, *et al*, 2007; Remes, Brayne, van der Linde, *et al*, 2016). Improving the quality of treatment for anxiety, especially in childhood, would have broad and important societal and public health effects.

"Therapygenomic" studies can have practical utility in providing potential prognostic information for individuals undergoing psychological therapy, as well as value in demonstrating theoretical concepts. Perhaps the key example of this lies in determining the extent of the genetic component to response to CBT, including demonstrating that such a component exists. In behavioural genetics more broadly, this has been achieved using family-based methods, particularly twin studies. However, in the case of psychological therapy response, such methods are limited by the availability of appropriate participants. Obtaining large cohorts from clinical therapy settings is difficult; limiting recruitment to parent-child triads or twin pairs

(to allow heritability to be estimated) would be sufficient to make any such study impractical. There would also be discordance between related individuals at the disorder level, as well as the treatment response level, and appropriately modelling this would be non-trivial (Plomin & Haworth, 2010; Tansey, Guipponi, Hu, *et al*, 2013). An alternative approach would be to give psychological therapy to twin pairs or family members without a target disorder, which could alleviate both of these limitations. However, it may prove difficult to demonstrate that these genetic effects generalise to therapy-seeking individuals. Alternatively, an epidemiological approach might be undertaken, potentially with a genetic component. For example, electronic health records listing individuals receiving psychological therapy could be combined with either genetic data from biobanking efforts or with additional registers of family relationships to allow the assessment of heritability or familiality (which would be conflated with shared environment). However, there are likely to be few, if any, suitable datasets for such a study, controlling for confounding effects such as initial severity could be challenging, and the possible outcome measures would potentially be limited to overarching measures of quality of life such as employment status. Genomic data provides a potentially workable solution to identifying a heritable component to CBT response, via approaches such as GREML or LD score regression (Bulik-Sullivan, Loh, Finucane, *et al*, 2015; Yang, Lee, Goddard, *et al*, 2011). Although this will require cohort sizes in the

thousands to achieve sufficient power, it seems more achievable than the alternatives.

Both investigations in Chapters 3 and 4 demonstrate that it is feasible to study therapygenetics using genomic methods, improving the standard of evidence for identifying predictors of treatment response. This is crucial to identifying associated variants robustly, as has been demonstrated by the general failure to replicate the findings of candidate gene studies in genomic studies, and the consistent direction of effect for associated variants across the component studies in large GWAS meta-analyses (Bosker, Hartman, Nolte*, et al*, 2011; Farrell, Werge, Sklar*, et al*, 2015; Schizophrenia Working Group of the Psychiatric Genomics, 2014). Adopting the genomic approach is not straightforward, but many earlier limitations (particularly cost) have been removed by technological developments across the last decade. The primary barriers to future genomic studies are cohort size and heterogeneity. The former could be addressed by the formation of an international collaboration of clinical studies, such as have yielded success in other areas of behavioural genomics. However, the heterogeneity involved in such an effort would be high. The development of large psychological therapy programs and the increasing popularity of internet-delivered therapies provide an interesting alternative. Such programs generally have fewer exclusion criteria than randomised control trials (RCTs; Clark, 2011;

Cromarty, Drummond, Francis, *et al*, 2016; Hepgul, King, Amarasinghe, *et al*, 2016). Although individual studies of this type are likely to experience more heterogeneity than do RCTs, greater accessibility provides larger numbers, and as such there would be less need to combine multiple trials, resulting in potentially less heterogeneity overall. However, power is in part a trade-off between sample size and heterogeneity, and if sample sizes were sufficiently large, combining different types of programs could be worthwhile despite the increase in heterogeneity.

The most appropriate biological substrate for studying treatment response remains up for debate. As argued above, gene transcripts are likely to be less useful than genetic variants; however, other alternatives exist. Epigenetic marks, particularly DNA methylation but also histone modifications, show some of the variability that limits the use of gene expression measures but are more stable over time. As such they may represent a useful intermediate between gene expression and genetic variants that is worthy of further exploration. However, the appropriate tissue in which to study epigenetic marks is not obvious, and would need to be taken into consideration when designing any future study. This is especially important given the relatively novelty of the field of epigenomics, and the likely need for thousands of samples to obtain adequate power to detect the effects of individual epigenetic marks.

Study design is the primary means to increase statistical power to identify real effects. However, this can also be achieved (arguably to a lesser extent) through analytical methods, as was attempted in Chapter 4. Dimensional reduction techniques aim to combine the variable-rich output of genomic studies in meaningful ways, and in doing so reduce the multiple testing load of such investigations. Data-driven or literature-driven clustering (such as are implemented in WGCNA or GOrilla respectively) can provide insight into the combined effects of high-dimensional data. Typically these methods are used on gene expression or DNA methylation data, but they could be extended to incorporate genetic variation, or even to incorporate both biological and environmental variables. This has clear implications in the context of gene-environment interplay – clusters featuring both environmental and biological factors may provide valuable targets for further study. Alternatively, dimensional reduction could be extended by identifying the main component of variance within clusters. For example, the first principal component of variance in a gene expression module from WGCNA could be used as an endophenotype in GWAS analyses. This might enable the identification of genetic influences on the action of a biological pathway in general, rather than focussing on the expression of a single gene. However, adopting any dimensional reduction methods requires an appreciation of the underlying statistical model. For example, while WGCNA has become a popular method for data-driven

clustering, it assumes that biological networks of interaction are best modelled with scale-free topology, which is disputed (Khanin & Wit, 2006; Stumpf, Wiuf & May, 2005).

Attempts to increase power through dimensional reduction also underlie the use of polygenic risk scores in this thesis. Rather than examining thousands of independent variants, risk scoring allows a single proxy for the effect of the genome to be defined. Chapters 5 and 6 differ in the way they treat the environment. In Chapter 5, environmental effects are not explicitly modelled, but instead are reflected in the robustness of the relationship between BMI and depression, even when the effect of the polygenic risk score is taken into account. This has parallels with the manner in which environment is defined in quantitative genetic analyses. In contrast, Chapter 6 explicitly defines environmental effects. Although only two environmental effects were defined in this study, multiple environments could be included in this approach, limited only by the increased burden of multiple testing. As such, the method used in Chapter 6 is equivalent to a GWIS with the genetic component collapsed into a single variable.

The subject explored in Chapter 5 has considerable external relevance – both depression and being overweight contribute considerably and increasingly to the burden of ill health (Lim, Vos, Flaxman, *et al*, 2013; Murray, Vos, Lozano, *et al*, 2012; Ng, Fleming, Robinson, *et al*, 2014;

Whiteford, Degenhardt, Rehm, *et al*, 2013). Furthermore, understanding the relationship between these traits could inform treatment - there is evidence to suggest a differential effect of monoamine oxidase inhibitory antidepressants on depression depending on weight status (Liebowitz, Quitkin, Stewart, *et al*, 1988; Quitkin, Stewart, McGrath, *et al*, 1993). Equally, different approaches might be taken to manage weight in individuals with depression if there is a direct causal relationship between the traits than if they are both common outcomes from stem from an alternative cause (for example, both might result from a low-quality diet; Lopresti, Hood & Drummond, 2013; Penninx, 2016). However, Chapter 5 is far from the first study of this question. Aside from the decades of study of the phenotypic and clinical relationship between depression and BMI, there is a growing literature regarding the genetic component of this relationship. Chapter 5 employs a similar method to a recent paper from the Generation Scotland study and includes attempts to cross-replicate findings from Generation Scotland and UK Biobank (Clarke, Hall, Fernandez-Pujals, *et al*, 2015). Chief among these was the presence of a significant interaction such that genetic influences on BMI were greater in individuals with depression than in controls, which was observed in Generation Scotland but not in UK Biobank. While this was in part due to an analytical artefact (the use of different polygenic risk scores), the fact that this disparity remained when the same approach was used in both studies suggests that the composition of the

depression case group differed between UK Biobank and Generation Scotland. The results from Chapter 5 in the context of the field thus suggest that studying the relationship between BMI and depression may first require dissection of the depression group into more homogeneous subsets. This is supported by other recent work, which suggested that genetic risk for metabolic phenotypes was associated with depression in individuals with atypical depression, but not in those with typical depression (Milaneschi, Lamers, Bot, *et al*, 2015; Milaneschi, Lamers, Peyrot, *et al*, 2016). Depression is far from being the only heterogeneous behavioural phenotype. The necessity to identify homogeneous subgroups is a major theme in behavioural genetics, and this extends to the study of gene-environment interplay.

Chapter 6 demonstrates an important point concerning BMI development, which may be informative about behavioural development in general. Within the study, there are two related phenotypes, BMI at age 11 and the slope of change between 11 and 16 (modelled linearly). The latter phenotype is dependent on the first, and so BMI at 11 was included as a covariate in the analysis of change across adolescence. As a result, the effects of variables of interest (in this instance, genetic risk and social environmental factors) before 11 years old are captured by the inclusion of BMI at 11 as a covariate, and so some insight into the stability of these effects can be gleaned from the results. In the instance of BMI, this suggested relatively

stable influences of parenting and SES, and a genetic influence with both stable and novel components (because a PRS derived from an adult sample shows an effect both at 11 and across adolescence, despite the inclusion of BMI at 11 as a variable in the latter analysis). This provides an interesting corroboration of traditional longitudinal modelling approaches, which also argue for both stable and time-specific effects of genetics on BMI (Haworth, Carnell, Meaburn, *et al*, 2008; Hjelmborg, Fagnani, Silventoinen, *et al*, 2008).

7.4 Gene-environment interplay in the behavioural genomic era

Decades of behavioural genetic research strongly argues that both genetic and non-genetic influences have important roles in behaviour. Although specific, robust examples of gene-environment interaction are few, it seems unlikely that these different components act entirely independently. Genome-wide association studies have strengthened the study of genetic factors as main effects, providing multiple examples of common variants that influence behavioural traits. The field of gene-environment interplay is now beginning to catch up, and to adopt genome-wide methods.

Within this thesis, I have explored alternatives to genome-wide interaction studies (GWIS), which are the most direct way of using genomics to study gene-environment interplay. The potential to examine multiple environmental influences using GWIS is limited by power. However, it has become feasible to perform GWIS focussed on specific environmental

influences, and such studies may prove valuable in identifying interactions (Dunn, Wiste, Radmanesh, *et al*, 2016; Winkler, Justice, Graff, *et al*, 2015). Although focussing on a single environment violates the hypothesis-neutral ethos of genomic studies, there is a broader evidence basis for candidate environments than for candidate genetic variants, including the role of childhood maltreatment in depression and cannabis use in schizophrenia (Green, Young & Kavanagh, 2005; Nanni, Uher & Danese, 2012).

Genomic data is information-rich, and alternative approaches that use genomics complement the GWIS approach. Fundamentally, the aim of such research is to identify genetic variants or an overall genomic effect that is contingent on environmental context. Exposed-only GWAS achieve this aim directly, by using differential response to a shared environment as the phenotype. Cognitive behavioural therapy provides a theoretically attractive phenotype for pursuing such studies. The studies presented in Chapters 3 and 4 demonstrate that genomic studies can be used to investigate genetic associations with response to cognitive behavioural therapy as an environmental exposure. The results provide an exclusionary upper threshold for proposed effect sizes; future studies must be designed with sufficient power to test the likely small effect sizes that can be inferred from behavioural genomic studies more generally. Much the same conclusions can be applied to the study of RNA transcripts in Chapter 4.

Using polygenic risk scores to dissect relationships between phenotypes allows environmental effects to be inferred by the removal of genetic effects, and as such gives insight into the broad effect of (additive) genetic and non-genetic factors on that relationship. Chapter 5 demonstrated that this approach can be informative of the likely components of the relationship between depression and BMI. Such an approach will become more effective as polygenic risk scores capture a greater proportion of variance, and could be extended to incorporate alternative methods to use genotype data as a proxy for the overall effect of the genome (such as the summarised GBLUPs proposed by Speed & Balding (2014)).

Finally, studying genetic risk-by-environment interactions allows a direct test of the effect of environmental context, while reducing the multiple testing inherent in the GWIS approach. Polygenic risk-by-environment interaction studies of this kind are becoming popular, and Chapter 6 adds to this literature. However, it also demonstrates the difficulties of combining the cohort sizes needed for genomic study with the precise environmental definitions needed to robustly assess gene-environment interplay.

The concept of gene-environment interplay is intuitively attractive, and fits within the broader theory of systems biology - that is, that influences on biological phenotypes are not independent, but exist within a broad network of many interacting components. However, the evidence to date in support

of gene-environment interplay in behavioural phenotypes, including from the studies in this thesis, is underwhelming. There are few, if any, robust gene-by-environment interactions in the behavioural literature. The 5HTTLPR-by-stressful life event interaction associated with depression may be an exception, but the variability in the environments studied, the different genotypic models used in such analyses, and the amount of analytical variability in meta-analysis of this association is sufficient to prevent a strong conclusion on the veracity of that association (Sharpley, Palanisamy, Glyde *et al*, 2014; Taylor & Munafo, 2016). In part, this lack of evidence reflects the slow movement of the field as a whole away from the candidate gene method, the limitations of which (focus on a small number of genes, tendency to use small sample sizes, vulnerability to winner's curse) have already been discussed in this thesis (Dick, Agrawal, Keller *et al* 2015). This is compounded further by the general difficulties that large-scale environmental research presents - adopting a gold-standard measure of a specific environment and obtaining this on sufficient participants to capture small genetic influences has proved a considerable challenge.

However, some of the issues that have impaired gene-environment research in the past can be addressed by adopting a genome-wide approach. The polygenic risk score-by-environment approach, although it did not yield conclusive results when used in this thesis, removes the necessity to select a

genetic variant of interest. As the base datasets used to develop risk scores become larger, and risk scores begin to capture more variance, risk score-by-environment interactions may become more meaningful. A caveat to this is that the current approach effectively assumes genetic homogeneity of the trait under study (at least in within-trait analyses), which may not be the case - the risk score as a whole may not interact with a given environment even if a subset of variants truly do. However, the growing general interest in the partitioning of variance components is likely to spread to PRS, and this may provide opportunities for extending the PRS-by-environment interaction paradigm to account for genetic heterogeneity in traits under study.

A further issue that has limited candidate gene-environment interaction studies has been a limited appreciation of multiple testing and the low prior probability that a given interaction is truly associated with the outcome. Although it was not the focus of this thesis, the GWIS approach is now becoming tractable, and yielding findings (Dunn, Wiste, Radmanesh *et al*, 2016). Too few GWIS studies have been done to date to know whether such findings will prove robust, but the general replicability of findings from main effects GWAS gives reason to be hopeful that single-variant genome-wide approaches, in concert with variant-grouping methods like PRS, may identify gene-environment interactions (Vinkhuyzen & Wray, 2015).

By incorporating environmental measurements into genetic studies (and vice versa), it might be possible to gain important insights into the fundamental, and highly complex, nature of behaviour. It is unlikely that a single investigation will ever serve to answer a question of biological interest because no method is a perfect model of the underlying system, but applying multiple approaches can provide a richer evidence base than using a single method alone. Through such cumulative steps, we will begin to understand behaviour.

Chapter 8: References for Chapters 1, 2 and 7

Aas, M., Blokland, G. A., Chawner, S. J.*, et al* (2016) Summaries of plenary, symposia, and oral sessions at the XXII World Congress of Psychiatric Genetics, Copenhagen, Denmark, 12-16 October 2014. *Psychiatr Genet*, **26**, 1-47.

Ackerman, S. J. & Hilsenroth, M. J. (2003) A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clin Psychol Rev*, **23**, 1-33.

Adami, H.-O., Hunter, D. J. & Trichopoulos, D. (2008) *Textbook of cancer epidemiology*: Oxford University Press.

Afari, N., Noonan, C., Goldberg, J.*, et al* (2010) Depression and obesity: do shared genes explain the relationship? *Depress Anxiety*, **27**, 799-806.

Ahmad, S., Rukh, G., Varga, T. V.*, et al* (2013) Gene x physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. *PLoS Genet*, **9**, e1003607.

Alberts, B. (2002) *Molecular biology of the cell* (4th edn). New York: Garland Science.

Anand, B. K. & Brobeck, J. R. (1951) Hypothalamic control of food intake in rats and cats. *Yale J Biol Med*, **24**, 123-140.

Anderson, C. A., Pettersson, F. H., Clarke, G. M.*, et al* (2010) Data quality control in genetic case-control association studies. *Nat Protoc*, **5**, 1564-1573.

Andersson, E., Ruck, C., Lavebratt, C.*, et al* (2013) Genetic polymorphisms in monoamine systems and outcome of cognitive behavior therapy for social anxiety disorder. *PLoS ONE*, **8**, e79015.

Angold, A., Costello, E. J. & Erkanli, A. (1999) Comorbidity. *J Child Psychol Psychiatry*, **40**, 57-87.

Anttila, V., Bulik-Sullivan, B., Finucane, H. K.*, et al* (2016) Analysis of shared heritability in common disorders of the brain. *bioRxiv*, 048991.

Ashburner, M., Ball, C. A., Blake, J. A.*, et al* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**, 25-29.

Ashley-Koch, A. E., Garrett, M. E., Gibson, J.*, et al* (2015) Genome-wide association study of posttraumatic stress disorder in a cohort of Iraq-Afghanistan era veterans. *J Affect Disord*, **184**, 225-234.

Association, A. P. (2013) *Diagnostic and statistical manual of mental disorders, (DSM-5®)*: American Psychiatric Pub.

Bakermans-Kranenburg, M. J. & van IJzendoorn, M. H. (2015) The hidden efficacy of interventions: genexenvironment experiments from a differential susceptibility perspective. *Annual Review of Psychology, Vol 66*, **66**, 381-409.

Baldwin, D. S., Anderson, I. M., Nutt, D. J.*, et al* (2014) Evidence-based pharmacological treatment of anxiety disorders, post-traumatic stress disorder and obsessive-compulsive disorder: a revision of the 2005 guidelines from the British Association for Psychopharmacology. *J Psychopharmacol*, **28**, 403-439.

Bandelow, B., Lichte, T., Rudolf, S.*, et al* (2014) The diagnosis of and treatment recommendations for anxiety disorders. *Dtsch Arztebl Int*, **111**, 473-480.

---- (2015) The German guidelines for the treatment of anxiety disorders. *European Archives of Psychiatry and Clinical Neuroscience*, **265**, 363-373.

Barabasi, A. L. & Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509-512.

Bardone, A. M., Moffitt, T. E., Caspi, A.*, et al* (1998) Adult physical health outcomes of adolescent girls with conduct disorder, depression, and anxiety. *J Am Acad Child Adolesc Psychiatry*, **37**, 594-601.

Barnett, A. G., van der Pols, J. C. & Dobson, A. J. (2005) Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, **34**, 215-220.

Barrett, P., Turner, C. & Lowry-Webster, H. (2000) *Friends for Children Group Leader's Manual*: Australian Academic Press.

Bateson, W., Saunders, E. R. & Punnett, R. C. (1906) Further experiments on inheritance in sweet peas and stocks: Preliminary account. *Proceedings of the Royal Society of London Series B-Containing Papers of a Biological Character*, **77**, 236-238.

Baxter, A. J., Vos, T., Scott, K. M.*, et al* (2014) The global burden of anxiety disorders in 2010. *Psychol Med*, **44**, 2363-2374.

Beck, A. T. (2005) The current state of cognitive therapy: a 40-year retrospective. *Arch Gen Psychiatry*, **62**, 953-959.

Belsky, D. W., Moffitt, T. E., Baker, T. B.*, et al* (2013) Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. *JAMA Psychiatry*, **70**, 534-542.

Belsky, J. & Pluess, M. (2009) Beyond diathesis stress: differential susceptibility to environmental influences. *Psychol Bull*, **135**, 885-908.

Bergmann, S., Ihmels, J. & Barkai, N. (2004) Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol*, **2**, E9.

Bittner, A., Egger, H. L., Erkanli, A.*, et al* (2007) What do childhood anxiety disorders predict? *J Child Psychol Psychiatry*, **48**, 1174-1183.

Blane, D., Hart, C. L., Smith, G. D.*, et al* (1996) Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. *BMJ*, **313**, 1434-1438.

Bleuler, M. (1963) Conception of schizophrenia within the last fifty years and today [abridged]. *Proceedings of the Royal Society of Medicine*, **56**, 945.

Boardman, J. D., Domingue, B. W., Blalock, C. L.*, et al* (2014) Is the gene-environment interaction paradigm relevant to genome-wide studies? The case of education and body mass index. *Demography*, **51**, 119-139.

Bockting, C. L., Mocking, R. J., Lok, A.*, et al* (2013) Therapygenetics: the 5HTTLPR as a biomarker for response to psychological therapy? *Mol Psychiatry*, **18**, 744-745.

Bögels, S. (2008) Behandeling van angststoornissen bij kinderen en adolescenten met het cognitief-gedragstherapeutisch protocol Denken+ Doen= Durven [Treatment of anxiety disorders in children and adolescents with cognitive behavioral protocol Thinking + Doing = Daring]: Bohn Stafleu van Loghum. [Dutch]

Boomsma, D., Busjahn, A. & Peltonen, L. (2002) Classical twin studies and beyond. *Nat Rev Genet*, **3**, 872-882.

Bosker, F. J., Hartman, C. A., Nolte, I. M.*, et al* (2011) Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry*, **16**, 516-532.

Botstein, D., White, R. L., Skolnick, M.*, et al* (1980) Construction of a Genetic-Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am J Hum Genet*, **32**, 314-331.

Bourgon, R., Gentleman, R. & Huber, W. (2010) Independent filtering increases detection power for high-throughput experiments. *Proc Natl Acad Sci U S A*, **107**, 9546-9551.

Box, G. E. P. (1976) Science and Statistics. *J Am Stat Assoc*, **71**, 791-799.

Braddon, F.E., Rodgers, B., Wadsworth M. E., *et al* (1986) Onset of obesity in a 36 year birth cohort study. *Br Med J,* **293**, 299-303

Bray, G. A. (2004) Obesity is a chronic, relapsing neurochemical disease**.** *International Journal of Obesity,* **28,** 34-38.

Brown, G.W. & Harris, T. (1978) Social origins of depression: A study of psychiatric disorder in women. London, UK: Routledge.

Bryant, R. A., Felmingham, K. L., Falconer, E. M.*, et al* (2010) Preliminary evidence of the short allele of the serotonin transporter gene predicting poor response to cognitive behavior therapy in posttraumatic stress disorder. *Biol Psychiatry*, **67**, 1217-1219.

Bulik-Sullivan, B., Finucane, H. K., Anttila, V.*, et al* (2015) An atlas of genetic correlations across human diseases and traits. *Nat Genet*, **47**, 1236-1241.

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K.*, et al* (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*, **47**, 291-295.

Burkhouse, K. L., Uhrlass, D. J., Stone, L. B. *et al* (2012) Expressed Emotion-Criticism and Risk of Depression Onset in Children**.** *J Clin. Child Adolesc Psychol,* **41,** 771-777

Butte, A. J. & Kohane, I. S. (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In *Pac Symp Biocomput*, pp. 418-429: Citeseer.

Casas-Agustench, P., Arnett, D. K., Smith, C. E.*, et al* (2014) Saturated fat intake modulates the association between an obesity genetic risk score and body mass index in two US populations. *J Acad Nutr Diet*, **114**, 1954-1966.

Caspi, A., Sugden, K., Moffitt, T. E.*, et al* (2003) Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**, 386-389.

Cespedes, E. M., Hu, F., B., Redline, S., *et al*., (2016) Chronic insufficient sleep and diet quality: Contributors to childhood obesity**.** *Obesity (Silver Spring),* **24**, 184-190**.**

Chabris, C. F., Hebert, B. M., Benjamin, D. J.*, et al* (2012) Most reported genetic associations with general intelligence are probably false positives. *Psychol Sci*, **23**, 1314-1323.

Chabris, C. F., Lee, J. J., Cesarini, D.*, et al* (2015) The Fourth Law of Behavior Genetics. *Curr Dir Psychol Sci*, **24**, 304-312.

Chaput, J-P. (2014) Sleep patterns, diet quality and energy balance. *Physiology & Behavior,* **134,** 86-91.

Chan, W. S. (2017) Delay discounting and response disinhibition moderate associations between actigraphically measured sleep parameters and body mass index. *Journal of Sleep Research,* **26,** 21-29.

Chang, C., Chow, C., Tellier, L.*, et al* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.

Cicchetti, D., Toth, S. L. & Handley, E. D. (2015) Genetic moderation of interpersonal psychotherapy efficacy for low-income mothers with major depressive disorder: implications for differential susceptibility. *Dev Psychopathol*, **27**, 19-35.

Clark, D. M. (2011) Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *Int Rev Psychiatry*, **23**, 318-327.

Clarke, T. K., Hall, L. S., Fernandez-Pujals, A. M.*, et al* (2015) Major depressive disorder and current psychological distress moderate the effect of polygenic risk for obesity on body mass index. *Transl Psychiatry*, **5**, e592.

Clarke, T. K., Smith, A. H., Gelernter, J.*, et al* (2016) Polygenic risk for alcohol dependence associates with alcohol consumption, cognitive function and social deprivation in a population-based cohort. *Addict Biol*, **21**, 469-480.

Coleman, J. R. I., Euesden, J., Patel, H.*, et al* (2016) Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Brief Funct Genomics*, **15**, 298-304.

Coleman, J. R. I., Lester, K. J., Keers, R.*, et al* (2016) Genome-wide association study of response to cognitive-behavioural therapy in children with anxiety disorders. *Br J Psychiatry*, **209,** 236-243.

Coleman, J. R. I., Lester, K. J., Roberts, S.*, et al* (2016) Separate and combined effects of genetic variants and pre-treatment whole blood gene expression on response to exposure-based cognitive behavioural therapy for anxiety disorders. *World J Biol Psychiatry*, 1-34.

Collins, A. L. & Sullivan, P. F. (2013) Genome-wide association studies in psychiatry: what have we learned? *Br J Psychiatry*, **202**, 1-4.

Consortium, C. (2015a) Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, **523**, 588-591.

Consortium, G. T. (2015b) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648-660.

Cooley, P., Clark, R. & Folsom, R. (2014) Assessing Gene-Environment Interactions in Genome-Wide Association Studies: Statistical Approaches. In *RTI Research Report Series*: RTI International.

Corvin, A., Craddock, N. & Sullivan, P. F. (2010) Genome-wide association studies: a primer. *Psychol Med*, **40**, 1063-1077.

Costello, E. J., Egger, H. L. & Angold, A. (2005) The developmental epidemiology of anxiety disorders: phenomenology, prevalence, and comorbidity. *Child Adolesc Psychiatr Clin N Am*, **14**, 631-648, vii.

Creswell, C. & Waite, P. (2016) Assessing treatment outcomes in RCTs for childhood anxiety disorders. In *8th World Congress of Behavioural and Cognitive Therapies*. Melbourne.

Cromarty, P., Drummond, A., Francis, T.*, et al* (2016) NewAccess for depression and anxiety: adapting the UK Improving Access to Psychological Therapies Program across Australia. *Australas Psychiatry*.

Cross-Disorder Group of the Psychiatric Genomics, C. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*, **45**, 984-994.

Dai, J. Y., Kooperberg, C., Leblanc, M.*, et al* (2012) Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*, **99**, 929-944.

Daly, M. J., Rioux, J. D., Schaffner, S. F.*, et al* (2001) High-resolution haplotype structure in the human genome. *Nat Genet*, **29**, 229-232.

Davidson, J. R., Miller, R. D., Turnbull, C. D.*, et al* (1982) Atypical depression. *Arch Gen Psychiatry*, **39**, 527-534.

Davies, M. N., Volta, M., Pidsley, R.*, et al* (2012) Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol*, **13**, R43.

Davison, K. K. & Birch, L. L. (2001) Childhood overweight: a contextual model and recommendations for future research. *Obes Rev*, **2**, 159-171.

de los Campos, G., Vazquez, A. I., Fernando, R.*, et al* (2013) Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet*, **9**, e1003608.

de Wit, L., Luppino, F., van Straten, A.*, et al* (2010) Depression and obesity: a meta-analysis of community-based studies. *Psychiatry Res*, **178**, 230-235.

de Wit, L. M., van Straten, A., van Herten, M.*, et al* (2009) Depression and body mass index, a u-shaped association. *BMC Public Health*, **9**, 14.

Dick, D. M., Agrawal, A., Keller, M. C.*, et al* (2015) Candidate gene-environment interaction research: reflections and recommendations. *Perspect Psychol Sci*, **10**, 37-59.

Domingue, B. W. & Boardman, J. D. (2016) Using Genome Wide Estimates of Heritability to Examine the Relevance of Gene-Environment Interplay. *bioRxiv*, 037861.

Dudbridge, F. (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet*, **9**, e1003348.
Dudbridge, F. & Fletcher, O. (2014) Gene-environment dependence creates spurious gene-environment interaction. *Am J Hum Genet*, **95**, 301-307.
Dudbridge, F. & Gusnanto, A. (2008) Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol*, **32**, 227-234.
Duncan, L. E. & Keller, M. C. (2011) A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry*, **168**, 1041-1049.
Dunn, E. C., Brown, R. C., Dai, Y.*, et al* (2015) Genetic determinants of depression: recent findings and future directions. *Harv Rev Psychiatry*, **23**, 1-18.
Dunn, E. C., Wiste, A., Radmanesh, F.*, et al* (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**, 265-280.
Eden, E., Lipson, D., Yogev, S.*, et al* (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol*, **3**, e39.
Eden, E., Navon, R., Steinfeld, I.*, et al* (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48.
Edfors, F., Danielsson, F., Hallström, B.M.*, et al* (2016) Gene-specific correlation of RNA and protein levels in human cells and tissue**s.** *Mol. Syst. Biol.,* **12,** 883.
Ege, M. J. & Strachan, D. P. (2013) Comparisons of power of statistical methods for gene-environment interaction analyses. *Eur J Epidemiol*, **28**, 785-797.
Eisen, M. B., Spellman, P. T., Brown, P. O.*, et al* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, **95**, 14863-14868.
Eley, T. C. (2014) The future of therapygenetics: where will studies predicting psychological treatment response from genomic markers lead? *Depress Anxiety*, **31**, 617-620.
Eley, T. C., Hudson, J. L., Creswell, C.*, et al* (2012) Therapygenetics: the 5HTTLPR and response to psychological therapy. *Mol Psychiatry*, **17**, 236-237.
Eley, T. C, McAdams, T. A., Rijsdijk, F. V. *et al* (2015) The Intergenerational Transmission of Anxiety: A Children-of-Twins Study. *Am J Psych*, **172**, 630-637.
Emanuel, E. J. & Miller, F. G. (2001) The ethics of placebo-controlled trials--a middle ground. *N Engl J Med*, **345**, 915-919.
Emilsson, V., Thorleifsson, G., Zhang, B.*, et al* (2008) Genetics of gene expression and its effect on disease. *Nature*, **452**, 423-428.
Erhardt, A., Akula, N., Schumacher, J.*, et al* (2012) Replication and meta-analysis of TMEM132D gene variants in panic disorder. *Transl Psychiatry*, **2**, e156.
Erhardt, A., Czibere, L., Roeske, D.*, et al* (2011) TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. *Mol Psychiatry*, **16**, 647-663.
Escott-Price, V., Shoai, M., Pither, R., *et al* (2017) Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiology of Aging,* **49**, 214-e7.
Euesden, J., Lewis, C. M. & O'Reilly, P. F. (2015) PRSice: Polygenic Risk Score software. *Bioinformatics*, **31**, 1466-1468.
Fadista, J., Manning, A. K., Florez, J. C.*, et al* (2016) The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet*, **24**, 1202-1205.
Faith, M. S., Matz, P. E. & Jorge, M. A. (2002) Obesity-depression associations in the population. *J Psychosom Res*, **53**, 935-942.

Falconer, D. S. (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics*, **29**, 51-76.

Falconer, D. S. & Mackay, T. F. (1996) Introduction to quantitative genetics. *Harlow: Longman Group Ltd*.

Farooqi, I. S. (2014) Defining the neural basis of appetite and obesity: from genes to behaviour. *Clin Med (Lond)*, **14**, 286-289.

Farrell, M. S., Werge, T., Sklar, P.*, et al* (2015) Evaluating historical candidate genes for schizophrenia. *Mol Psychiatry*, **20**, 555-562.

Felix, J. F., Bradfield, J. P., Monnereau, C.*, et al* (2016) Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet*, **25**, 389-403.

Felmingham, K. L., Dobson-Stone, C., Schofield, P. R.*, et al* (2013) The brain-derived neurotrophic factor Val66Met polymorphism predicts response to exposure therapy in posttraumatic stress disorder. *Biol Psychiatry*, **73**, 1059-1063.

Fineberg, N. A., Haddad, P. M., Carpenter, L.*, et al* (2013) The size, burden and cost of disorders of the brain in the UK. *J Psychopharmacol*, **27**, 761-770.

Finucane, H. K., Bulik-Sullivan, B., Gusev, A.*, et al* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*, **47**, 1228-1235.

Finucane, H K., Reshef, Y., Anttila, V. *et al* (2017) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types**.** *bioRxiv*, 103069.

Fisher, R. A. (1918) XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, **52**, 399-433.

Ford, T., Goodman, R. & Meltzer, H. (2004) The relative importance of child, family, school and neighbourhood correlates of childhood psychiatric disorder. *Social Psychiatry and Psychiatric Epidemiology*, **39,** 487-496.

Frayling, T. M., Timpson, N. J., Weedon, M. N.*, et al* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889-894.

Frye, A. A. & Garber, J. (2005) The Relations Among Maternal Depression, MaternalCriticism, and Adolescents' Externalizingand Internalizing Symptoms. *Journal of Abnormal Child Psychology,* **33,** 1-11.

Furmark, T., Carlbring, P., Hammer, S.*, et al* (2010) Effects of serotonin transporter and tryptophan hydroxylase-2 gene variation on the response to cognitive-bahavior therapy in patients with social anxiety disorder.

Furukawa T. A., Noma H., Caldwell D. M., *et al* (2014) Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand,* **130,** 181-192.

Gabriel, S. B., Schaffner, S. F., Nguyen, H.*, et al* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225-2229.

Gage, S. H., Davey Smith, G., Ware, J. J.*, et al* (2016) G = E: What GWAS Can Tell Us about the Environment. *PLoS Genet*, **12**, e1005765.

Galton, F. (1869) *Hereditary genius*: Macmillan and Company.

Garrod, A. E. (1902) The incidence of alkaptonuria a study in chemical individuality. *Lancet*, **2**, 1616-1620.

Geiringer, H. (1944) On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics*, **15**, 25-57.

Gibson-Smith, D., Bot, M., Paans, N. P.*, et al* (2016) The role of obesity measures in the development and persistence of major depressive disorder. *J Affect Disord*, **198**, 222-229.

Goodyer, I.M., Wright, C. & Altham, P.M.E. (1988) Maternal adversity and recent stressful life events in anxious and depressed children. *J Child Psychol Psychiatry*, **29**, 651-667.

Goodyer, I.M., Wright, C. & Altham, P.M.E. (1990) The friendships and recent life events of anxious and depressed school-age children. *Br J Psychiatry*, **156,** 689-698

Gottlieb, G. (2003) On making behavioral genetics truly developmental. *Human Development*, **46**, 337-355.

Gratten, J., Wray, N. R., Keller, M. C.*, et al* (2014) Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat Neurosci*, **17**, 782-790.

Green, B., Young, R. & Kavanagh, D. (2005) Cannabis use and misuse prevalence among people with psychosis. *Br J Psychiatry*, **187**, 306-313.

Gregory, A. M., Caspi, A., Moffitt, T. E.*, et al* (2007) Juvenile mental health histories of adults with anxiety disorders. *Am J Psychiatry*, **164**, 301-308.

Guffanti, G., Galea, S., Yan, L.*, et al* (2013) Genome-wide association study implicates a novel RNA gene, the lincRNA AC068718.1, as a risk factor for post-traumatic stress disorder in women. *Psychoneuroendocrinology*, **38**, 3029-3038.

Gunderson, K. L., Steemers, F. J., Lee, G.*, et al* (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet*, **37**, 549-554.

Haby, M. M., Donnelly, M., Corry, J.*, et al* (2006) Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: a meta-regression of factors that may predict outcome. *Australian and New Zealand Journal of Psychiatry*, **40**, 9-19.

Hale III, W. W., Keijsers, L., Klimstra, T. A. *et al* (2011) How does longitudinally measured maternal expressed emotion affect internalizing and externalizing symptoms of adolescents from the general community? *J Child Psychol Psychiatry,* **52,** 1174-1183.

Hallden, S., Sjogren, M., Hedblad, B.*, et al* (2016) Gene variance in the nicotinic receptor cluster (CHRNA5-CHRNA3-CHRNB4) predicts death from cardiopulmonary disease and cancer in smokers. *J Intern Med*, **279**, 388-398.

Hambrick, D. Z. & Tucker-Drob, E. M. (2015) The genetics of music accomplishment: evidence for gene-environment correlation and interaction. *Psychon Bull Rev*, **22**, 112-120.

Hansen, T. V., Simonsen, M. K., Nielsen, F. C.*, et al* (2007) Collection of blood, saliva, and buccal cell samples in a pilot study on the Danish nurse cohort: comparison of the response rate and quality of genomic DNA. *Cancer Epidemiol Biomarkers Prev*, **16**, 2072-2076.

Hardy, R., Wadsworth, M. & Kuh, D. (2000) The influence of childhood weight and socioeconomic status on change in adult body mass index in a British national birth cohort. *Int J Obes Relat Metab Disord*, **24**, 725-734.

Haworth, C. M., Carnell, S., Meaburn, E. L.*, et al* (2008) Increasing heritability of BMI and stronger associations with the FTO gene over childhood. *Obesity (Silver Spring)*, **16**, 2663-2668.

Hedman, E., Andersson, E., Ljotsson, B.*, et al* (2012) Clinical and genetic outcome determinants of Internet- and group-based cognitive behavior therapy for social anxiety disorder. *Acta Psychiatr Scand*, **126**, 126-136.

Heijmans, B. T. & Mill, J. (2012) Commentary: The seven plagues of epigenetic epidemiology. *Int J Epidemiol*, **41**, 74-78.

Heils, A., Teufel, A., Petri, S.*, et al* (1996) Allelic variation of human serotonin transporter gene expression. *J Neurochem*, **66**, 2621-2624.

Hemani, G., Yang, J., Vinkhuyzen, A.*, et al* (2013) Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *Am J Hum Genet*, **93**, 865-875.

Hepgul, N., King, S., Amarasinghe, M.*, et al* (2016) Clinical characteristics of patients assessed within an Improving Access to Psychological Therapies (IAPT) service: results from a naturalistic cohort study (Predicting Outcome Following Psychological Therapy; PROMPT). *BMC Psychiatry*, **16**, 52.

Hernandez, D. G., Nalls, M. A., Moore, M.*, et al* (2012) Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol Dis*, **47**, 20-28.

Hettema, J. M., Neale, M. C. & Kendler, K. S. (2001) A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am J Psychiatry*, **158**, 1568-1578.

Hindorff, L. A., Sethupathy, P., Junkins, H. A.*, et al* (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, **106**, 9362-9367.

Hirschhorn, J. N., Lohmueller, K., Byrne, E.*, et al* (2002) A comprehensive review of genetic association studies. *Genet Med*, **4**, 45-61.

Hjelmborg, J., Fagnani, C., Silventoinen, K.*, et al* (2008) Genetic influences on growth traits of BMI: a longitudinal study of adult twins. *Obesity (Silver Spring)*, **16**, 847-852.

Hodgson, K., Almasy, L., Knowles, E. E.*, et al* (2016) Genome-wide significant loci for addiction and anxiety. *Eur Psychiatry*, **36**, 47-54.

Hofmann, S. G., Asmundson, G. J. & Beck, A. T. (2013) The science of cognitive therapy. *Behav Ther*, **44**, 199-212.

Hofmann, S. G., Asnaani, A., Vonk, I. J.*, et al* (2012) The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognit Ther Res*, **36**, 427-440.

Hofmann, S. G. & Smits, J. A. (2008) Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *J Clin Psychiatry*, **69**, 621-632.

Holliday, R. (1964) A mechanism for gene conversion in fungi. *Genetical Research*, **5**, 282.

Holmes T. H. & Rahe, R. H. (1967) The Social Readjustment Rating Scale. *J Psychosom Res,* **2**, 213-218

Hong, G., Zhang, W., Li, H.,*et al* (2013) Separate enrichment analysis of pathways for up- and downregulated genes. *J. R. Soc. Interface*, **11,** 20130950.

Hooley, J. M., Orley, J. & Teasdale, J. D. (1986) Levels of expressed emotion and relapse in depressed patients. *Br J Psychiatry,* **148**, 642-647.

Howe, A. S., Buttenschon, H. N., Bani-Fatemi, A.*, et al* (2016) Candidate genes in panic disorder: meta-analyses of 23 common variants in major anxiogenic pathways. *Mol Psychiatry*, **21**, 665-679.

Hudson, J. L., Keers, R., Roberts, S.*, et al* (2015) Clinical Predictors of Response to Cognitive-Behavioral Therapy in Pediatric Anxiety Disorders: The Genes for Treatment (GxT) Study. *J Am Acad Child Adolesc Psychiatry*, **54**, 454-463.

Hung, C. F., Breen, G., Czamara, D.*, et al* (2015) A genetic risk score combining 32 SNPs is associated with body mass index and improves obesity prediction in people with major depressive disorder. *BMC Med*, **13**, 86.

Hung, C. F., Rivera, M., Craddock, N.*, et al* (2014) Relationship between obesity and the risk of clinically significant depression: Mendelian randomisation study. *Br J Psychiatry*, **205**, 24-28.

Hunter, D. J. (2005) Gene-environment interactions in human diseases. *Nat Rev Genet*, **6**, 287-298.

Hyde, C. L., Nagle, M. W., Tian, C.*, et al* (2016) Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet*, **advance online publication**.

Ilardi, S. S. & Craighead, W. E. (1994) The Role of Nonspecific Factors in Cognitive-Behavior Therapy for Depression. *Clinical Psychology-Science and Practice*, **1**, 138-156.

International Schizophrenia, C., Purcell, S. M., Wray, N. R.*, et al* (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748-752.

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

---- (2014) How to make more published research true. *PLoS Med*, **11**, e1001747.

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

James, A. C., James, G., Cowdrey, F. A.*, et al* (2013) Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database Syst Rev*, **6**, CD004690.

Jansen, R. C. & Nap, J. P. (2001) Genetical genomics: the added value from segregation. *Trends Genet*, **17**, 388-391.

Janssen, I., Katzmarzyk, P. T. & Ross, R. (2004) Waist circumference and not body mass index explains obesity-related health risk. *The American Journal of Clinical Nutrition*, **79**, 379-384.

Jeong, H., Tombor, B., Albert, R.*, et al* (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651-654.

Johnson, T. (2013) gtx: Genetics ToolboX. *R package version 0.0*, **8**.

Johnson, W., Ong, K. K., Elks, C. E.*, et al* (2014) Modification of genetic influences on adiposity between 36 and 63 years of age by physical activity and smoking in the 1946 British Birth Cohort Study. *Nutr Diabetes*, **4**, e136.

Jokela, M., Berg, V., Silventoinen, K.*, et al* (2016) Body Mass Index and Depressive Symptoms: Testing for Adverse and Protective Associations in Two Twin Cohort Studies. *Twin Res Hum Genet*, **19**, 306-311.

Jokela, M., Elovainio, M., Keltikangas-Jarvinen, L.*, et al* (2012) Body mass index and depressive symptoms: instrumental-variables regression with genetic risk score. *Genes Brain Behav*, **11**, 942-948.

Kang, H. M., Sul, J. H., Service, S. K.*, et al* (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*, **42**, 348-354.

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

Kastner, N., Richter, S. H., Lesch, K. P.*, et al* (2015) Benefits of a "vulnerability gene"? A study in serotonin transporter knockout mice. *Behav Brain Res*, **283**, 116-120.

Keers, R., Coleman, J. R. I., Lester, K. J., *et al* (2016) A Genome-Wide Test of the Differential Susceptibility Hypothesis Reveals a Genetic Predictor of Differential Response to Psychological Treatments for Child Anxiety Disorders. *Psychother Psychosom,* **85**, 146-158

Keller, M. C. (2014) Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biol Psychiatry*, **75**, 18-24.

Kendall, P. C. & Hedtke, K. A. (2006) *The coping cat workbook*: Workbook Pub.

Kendler, K. S. (1996) Major depression and generalised anxiety disorder - Same genes, (partly) different environments - Revisited. *British Journal of Psychiatry*, **168**, 68-75.

Kendler, K. S., Eaves, L. J., Walters, E. E.*, et al* (1996) The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Arch Gen Psychiatry*, **53**, 391-399.

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

Kendler, K. S., Gardner, C. O., Gatz, M.*, et al* (2007) The sources of co-morbidity between major depression and generalized anxiety disorder in a Swedish national twin sample. *Psychol Med*, **37**, 453-462.

Kendler, K. S., Karkowski, L. M. & Prescott, C. A. (1999) Causal Relationship Between Stressful Life Events and the Onset of Major Depression**.** *Am J Psychiatry*, **156,** 837-841

Kendler, K. S., Myers, J., Prescott, C. A.*, et al* (2001) The genetic epidemiology of irrational fears and phobias in men. *Arch Gen Psychiatry*, **58**, 257-265.

Kendler, K. S., Neale, M. C., Kessler, R. C.*, et al* (1992a) The genetic epidemiology of phobias in women. The interrelationship of agoraphobia, social phobia, situational phobia, and simple phobia. *Arch Gen Psychiatry*, **49**, 273-281.

---- (1992b) Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Arch Gen Psychiatry*, **49**, 716-722.

Keri, S., Szabo, C. & Kelemen, O. (2014a) Blood biomarkers of depression track clinical changes during cognitive-behavioral therapy. *J Affect Disord*, **164**, 118-122.

---- (2014b) Expression of Toll-Like Receptors in peripheral blood mononuclear cells and response to cognitive-behavioral therapy in major depressive disorder. *Brain Behav Immun*, **40**, 235-243.

Kessler, R. C., Angermeyer, M., Anthony, J. C.*, et al* (2007) Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, **6**, 168-176.

Kessler, R. C., Berglund, P., Demler, O.*, et al* (2003) The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*, **289**, 3095-3105.

Kessler, R. C., Berglund, P., Demler, O.*, et al* (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*, **62**, 593-602.

Kessler, R. C., Zhao, S., Blazer, D. G.*, et al* (1997) Prevalence, correlates, and course of minor depression and major depression in the National Comorbidity Survey. *J Affect Disord*, **45**, 19-30.

Khanin, R. & Wit, E. (2006) How scale-free are biological networks. *J Comput Biol*, **13**, 810-818.

Klein, R. J., Zeiss, C., Chew, E. Y.*, et al* (2005) Complement factor H polymorphism in age-related macular degeneration. *Science*, **308**, 385-389.

Kohen, R., Cain, K. C., Buzaitis, A.*, et al* (2011) Response to Psychosocial Treatment in Poststroke Depression Is Associated With Serotonin Transporter Polymorphisms. *Stroke*, **42**, 2068-2070.

Kooperberg, C. & Leblanc, M. (2008) Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol*, **32**, 255-263.

Kuhn, K., Baker, S. C., Chudin, E.*, et al* (2004) A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res*, **14**, 2347-2356.

Laitinen, J., Power, C. & Jarvelin, M. R. (2001) Family social class, maternal body mass index, childhood body mass index, and age at menarche as predictors of adult obesity. *American Journal of Clinical Nutrition*, **74**, 287-294.

Lander, E. S., Linton, L. M., Birren, B.*, et al* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

Lane, J. M., Liang, J., Vlasac, I. *et al* (2017) Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.,* **49,** 274-281.

Langfelder, P. & Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Langfelder, P., Zhang, B. & Horvath, S. (2008) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, **24**, 719-720.

Lau, J. Y. F. & Eley, T.C. (2008) Disentangling gene-environment correlations and interactions on adolescent depressive symptoms. *J Child Psychol Psychiatry,* **49,** 142-150.

Lauderdale, D. S. & Rathouz, P. J. (2000) Body mass index in a US national sample of Asian Americans: effects of nativity, years since immigration and socioeconomic status. *International Journal of Obesity*, **24**, 1188-1194.

Le-Niculescu, H., Kurian, S., Yehyawi, N.*, et al* (2009) Identifying blood biomarkers for mood disorders using convergent functional genomics. *Molecular Psychiatry*, **14**, 156-174.

Leek, J. T., Scharpf, R. B., Bravo, H. C.*, et al* (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*, **11**, 733-739.

Lester, K. J. & Eley, T. C. (2013) Therapygenetics: Using genetic markers to predict response to psychological treatment for mood and anxiety disorders. *Biol Mood Anxiety Disord*, **3**, 4.

Lester, K. J., Roberts, S., Keers, R.*, et al* (2016) Non-replication of the association between 5HTTLPR and response to psychological therapy for child anxiety disorders. *Br J Psychiatry*, **208**, 182-188.

Levy-Gigi, E., Szabo, C., Kelemen, O.*, et al* (2013) Association among clinical response, hippocampal volume, and FKBP5 gene expression in individuals with posttraumatic stress disorder receiving cognitive behavioral therapy. *Biol Psychiatry*, **74**, 793-800.

Lewinsohn, P. M., Holm-Denoma, J. M., Small, J. W.*, et al* (2008) Separation anxiety disorder in childhood as a risk factor for future mental illness. *J Am Acad Child Adolesc Psychiatry*, **47**, 548-555.

Lewis, C. M. (2015) Mega-analysis of genome-wide association studies in major depressive disorder: MDD Working Group of the Psychiatric Genomics Consortium. In *World Congress of Psychiatric Genetics*. Toronto, Canada.

Lewis, C. M. & Knight, J. (2012) Introduction to genetic association studies. *Cold Spring Harb Protoc*, **2012**, 297-306.

Li, R. & Chambless, L. (2007) Test for additive interaction in proportional hazards models. *Ann Epidemiol*, **17**, 227-236.

Li, S., Zhao, J. H., Luan, J.*, et al* (2010) Physical activity attenuates the genetic predisposition to obesity in 20,000 men and women from EPIC-Norfolk prospective population study. *PLoS Med*, **7**, e1000332.

Liebowitz, M. R., Quitkin, F. M., Stewart, J. W.*, et al* (1988) Antidepressant specificity in atypical depression. *Arch Gen Psychiatry*, **45**, 129-137.

Lim, S. S., Vos, T., Flaxman, A. D.*, et al* (2013) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, **380**, 2224-2260.

Listgarten, J., Lippert, C., Kadie, C. M.*, et al* (2012) Improved linear mixed models for genome-wide association studies. *Nat Methods*, **9**, 525-526.

Liu, H. & Guo, G. (2015) Lifetime Socioeconomic Status, Historical Context, and Genetic Inheritance in Shaping Body Mass in Middle and Late Adulthood. *Am Sociol Rev*, **80**, 705-737.

Liu, S. Y., Walter, S., Marden, J.*, et al* (2015) Genetic vulnerability to diabetes and obesity: does education offset the risk? *Soc Sci Med*, **127**, 150-158.
Llewellyn, C. & Wardle, J. (2015) Behavioral susceptibility to obesity: Gene-environment interplay in the development of weight. *Physiol Behav*, **152**, 494-501.
Locke, A. E., Kahali, B., Berndt, S. I.*, et al* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197-206.
Loerinc, A. G., Meuret, A. E., Twohig, M. P.*, et al* (2015) Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clin Psychol Rev*, **42**, 72-82.
Logue, M. W., Baldwin, C., Guffanti, G.*, et al* (2013) A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. *Mol Psychiatry*, **18**, 937-942.
Loh, P. R., Bhatia, G., Gusev, A.*, et al* (2015) Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet*, **47**, 1385-1392.
Loh, P. R., Tucker, G., Bulik-Sullivan, B. K.*, et al* (2015) Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*, **47**, 284-290.
Lonsdorf, T. B., Ruck, C., Bergstrom, J.*, et al* (2010) The COMTval158met polymorphism is associated with symptom relief during exposure-based cognitive-behavioral treatment in panic disorder. *BMC Psychiatry*, **10**, 99.
Lopez-Leon, S., Janssens, A. C., Gonzalez-Zuloeta Ladd, A. M.*, et al* (2008) Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry*, **13**, 772-785.
Lopresti, A. L., Hood, S. D. & Drummond, P. D. (2013) A review of lifestyle factors that contribute to important pathways associated with major depression: Diet, sleep and exercise. *Journal of Affective Disorders*, **148**, 12-27.
Lowe, R., Gemma, C., Beyan, H.*, et al* (2013) Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics*, **8**, 445-454.
Lueken, U., Zierhut, K. C., Hahn, T.*, et al* (2016) Neurobiological markers predicting treatment response in anxiety disorders: A systematic review and implications for clinical application. *Neurosci Biobehav Rev*, **66**, 143-162.
Luppino, F. S., de Wit, L. M., Bouvy, P. F.*, et al* (2010) Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Arch Gen Psychiatry*, **67**, 220-229.
Lustig, R. H. (2001) The neuroendocrinology of childhood obesity. *Pediatr Clin North Am*, **48**, 909-930.
Lynch, M. & Walsh, B. (1998) *Genetics and analysis of quantitative traits*: Sinauer Sunderland, MA.
Major Depressive Disorder Working Group of the Psychiatric, G. C., Ripke, S., Wray, N. R.*, et al* (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*, **18**, 497-511.
Manchia, M., Cullis, J., Turecki, G.*, et al* (2013) The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS ONE*, **8**, e76295.
Manolio, T. A., Collins, F. S., Cox, N. J.*, et al* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747-753.
Magee, W. J. (1999) Effects of negative life experiences on phobia onset. *Soc Psychiatry Psychiatr Epidemiol*, **34**, 343-351
Marchini, J. & Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nat Rev Genet*, **11**, 499-511.
Marigorta, U. M. & Gibson, G. (2014) A simulation study of gene-by-environment interactions in GWAS implies ample hidden effects. *Front Genet*, **5**, 225.

Martin, D. J., Garske, J. P. & Davis, M. K. (2000) Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *J Consult Clin Psychol*, **68**, 438-450.

Maurano, M. T., Humbert, R., Rynes, E.*, et al* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190-1195.

Mausbach, B. T., Moore, R., Roesch, S.*, et al* (2010) The Relationship Between Homework Compliance and Therapy Outcomes: An Updated Meta-Analysis. *Cognitive therapy and research*, **34**, 429-438.

McCarthy, S., Das, S., Kretzschmar, W.*, et al* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*, **advance online publication**.

McGrath, J. J., Mortensen, P. B., Visscher, P. M.*, et al* (2013) Where GWAS and epidemiology meet: opportunities for the simultaneous study of genetic and environmental risk factors in schizophrenia. *Schizophr Bull*, **39**, 955-959.

McGrath, L. M., Weill, S., Robinson, E. B.*, et al* (2012) Bringing a developmental perspective to anxiety genetics. *Dev Psychopathol*, **24**, 1179-1193.

McGuffin, P., Alsabban, S. & Uher, R. (2011) The truth about genetic variation in the serotonin transporter gene and response to stress and medication. *Br J Psychiatry*, **198**, 424-427.

McLeod, B. D., Weisz, J. R. & Wood, J. J. (2007) Examining the association between parenting and childhood depression: A meta-analysis. *Clinical Psychology Review,* **27**, 986-1003.

Meehl, P. E. (1962) Schizotaxia, Schizotypy, Schizophrenia. *American Psychologist*, **17**, 827-838.

Mendel, G. (1866) Versuche über Pflanzenhybriden [Experiments in Plant Hybridization]. *Verhandlungen des naturforschenden Vereines in Brunn 4: 3*, **44**. [English translation available at http://www.mendelweb.org/Mendel.html]

Menozzi, P., Piazza, A. & Cavalli-Sforza, L. (1978) Synthetic maps of human gene frequencies in Europeans. *Science*, **201**, 786-792.

Merikangas, K. R., Wicki, W. & Angst, J. (1994) Heterogeneity of depression. Classification of depressive subtypes by longitudinal course. *Br J Psychiatry*, **164**, 342-348.

Meyers, J. L., Cerda, M., Galea, S.*, et al* (2013) Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit Neighborhood Health Study. *Transl Psychiatry*, **3**, e290.

Milaneschi, Y., Lamers, F., Bot, M.*, et al* (2015) Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression. *Biol Psychiatry*.

Milaneschi, Y., Lamers, F., Peyrot, W. J.*, et al* (2016) Polygenic dissection of major depression clinical heterogeneity. *Mol Psychiatry*, **21**, 516-522.

Monks, S. A., Leonardson, A., Zhu, H.*, et al* (2004) Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet*, **75**, 1094-1105.

Montague, C. T., Farooqi, I. S., Whitehead, J. P.*, et al* (1997) Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, **387**, 903-908.

Monteiro, C. A., Moura, E. C., Conde W. L., *et al* (2004) Socioeconomic status and obesity in adult populations of developing countries: a review. *Bull World Health Organ*, **82**, 940-946

Moore P.S., Whaley S.E. & Sigman M. (2004) Interactionsbetween mothers and children: impacts of maternal andchild anxiety. *Journal of Abnormal Psychology* **113,** 471–476

Morgan, T. H. (1917) The theory of the gene. *American Naturalist*, **51**, 513-544.

Morley, M., Molony, C. M., Weber, T. M.*, et al* (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743-747.

Moser, G., Lee, S. H., Hayes, B. J.*, et al* (2015) Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet*, **11**, e1004969.

Mullins, N., Power, R. A., Fisher, H. L.*, et al* (2016) Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol Med*, **46**, 759-770.

Munafo, M. R. (2006) Candidate gene studies in the 21st century: meta-analysis, mediation, moderation. *Genes Brain Behav*, **5 Suppl 1**, 3-8.

---- (2015) Understanding the candidate gene x environment interaction debate: epistemological or evidential divide? *Int J Epidemiol*, **44**, 1130-1132.

Munafo, M. R., Clark, T. G. & Flint, J. (2004) Assessing publication bias in genetic association studies: evidence from a recent meta-analysis. *Psychiatry Res*, **129**, 39-44.

Munafo, M. R., Durrant, C., Lewis, G.*, et al* (2009) Gene X environment interactions at the serotonin transporter locus. *Biol Psychiatry*, **65**, 211-219.

Munafò, M. R., Zammit, S. & Flint, J. (2014) Practitioner Review: A critical perspective on gene–environment interaction models–what impact should they have on clinical perceptions and practice? *J Child Psychol Psychiatry*, **55**, 1092-1101.

Muris, P., Steernmen, P., Merckelbach, H., *et al* (1996) The role of parental fearfulness and modeling in children's fear**.** *Behavior Research and Therapy,* **34,** 265–268.

Murray, C. J., Vos, T., Lozano, R.*, et al* (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**, 2197-2223.

Murrary, L., Creswell, C. & Cooper, P. J. (2009) The development of anxiety disorders inchildhood: an integrative review. *Psychol Med,* **39**, 1413-1423.

Musliner, K. L., Seifuddin, F., Judy, J. A.*, et al* (2015) Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis. *Psychol Med*, **45**, 1709-1720.

Myers, R. S. & Stahl, F. W. (1994) Chi and the RecBC D enzyme of Escherichia coli. *Annu Rev Genet*, **28**, 49-70.

Myers, S., Bottolo, L., Freeman, C.*, et al* (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science*, **310**, 321-324.

Nakamura, M., Ueno, S., Sano, A.*, et al* (2000) The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. *Molecular Psychiatry*, **5**, 32-38.

Nanni, V., Uher, R. & Danese, A. (2012) Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am J Psychiatry*, **169**, 141-151.

NCCMH (2011) *Common mental health disorders : identification and pathways to care*. Leicester: British Psychological Society ; London : Royal College of Psychiatrists.

Neale, M. & Cardon, L. (1992) *Methodology for genetic studies of twins and families*: Springer Science & Business Media.

Nelemans, S. A., Hale III, W. W., Branje, S. J. T., *et al* (2014) Maternal Criticism and Adolescent Depressive and Generalized Anxiety Disorder Symptoms: A 6-Year Longitudinal Community Study. *Journal of Abnormal Child Psychology,* **42,** 755-766.

Nevill, A. M., Stewart, A. D., Olds, T.*, et al* (2006) Relationship between adiposity and body size reveals limitations of BMI. *Am J Phys Anthropol*, **129**, 151-156.

Ng, M., Fleming, T., Robinson, M.*, et al* (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, **384**, 766-781.

Nievergelt, C. M., Maihofer, A. X., Mustapic, M.*, et al* (2015) Genomic predictors of combat stress vulnerability and resilience in U.S. Marines: A genome-wide association study across multiple ancestries implicates PRTFDC1 as a potential PTSD gene. *Psychoneuroendocrinology*, **51**, 459-471.

Olatunji, B. O., Cisler, J. M. & Tolin, D. F. (2010) A meta-analysis of the influence of comorbidity on treatment outcome in the anxiety disorders. *Clin Psychol Rev*, **30**, 642-654.

Otowa, T., Hek, K., Lee, M.*, et al* (2016) Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry*.

Otowa, T., Kawamura, Y., Nishida, N.*, et al* (2012) Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. *Transl Psychiatry*, **2**, e186.

Otowa, T., Maher, B. S., Aggen, S. H.*, et al* (2014) Genome-wide and gene-based association studies of anxiety disorders in European and African American samples. *PLoS ONE*, **9**, e112559.

Palla, L. & Dudbridge, F. (2015) A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. *Am J Hum Genet*, **97**, 250-259.

Parsons, T. J., Power, C. & Manor, O. (2001) Fetal and early life growth and body mass index from birth to early adulthood in 1958 British cohort: longitudinal study. *BMJ*, **323**, 1331-1335.

Parsons, T. J., Power, C. & Logan, S. (1999) Childhood predictors of adult obesity: a systematic review. *International Journal of Obesity*, **23,** S1-S107

Paulus, M. P. (2015) Pragmatism Instead of Mechanism: A Call for Impactful Biological Psychiatry. *JAMA Psychiatry*, **72**, 631-632.

Penninx, B. W. (2016) Depression and cardiovascular disease: Epidemiological evidence on their linking mechanisms. *Neurosci Biobehav Rev*.

Peyrot, W. J., Milaneschi, Y., Abdellaoui, A.*, et al* (2014) Effect of polygenic risk scores on depression in childhood trauma. *Br J Psychiatry*, **205**, 113-119.

Phillips N.K., Hammen C.L., Brennan P.A., *et al* (2005). Early adversity and the prospective prediction ofdepressive and anxiety disorders in adolescents. *Journal of Abnormal Child Psychology* **33,** 13–24

Plomin, R. (1994) The Emanuel Miller Memorial Lecture 1993. Genetic research and identification of environmental influences. *J Child Psychol Psychiatry*, **35**, 817-834.

Plomin, R. & Daniels, D. (2010) Why are children in the same family so different from one another? *Behavioral and Brain Sciences*, **10**, 1.

Plomin, R., DeFries, J. C., Knopik, V. S.*, et al* (2012) *Behavioral genetics* (6th edn). New York, NY: Worth Publishers.

Plomin, R., DeFries, J. C. & Loehlin, J. C. (1977) Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull*, **84**, 309-322.

Plomin, R. & Haworth, C. M. (2010) Genetics and Intervention Research. *Perspect Psychol Sci*, **5**, 557-563.

Pluess, M. & Belsky, J. (2013) Vantage sensitivity: individual differences in response to positive experiences. *Psychol Bull*, **139**, 901-916.

Polderman, T. J., Benyamin, B., de Leeuw, C. A.*, et al* (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*, **47**, 702-709.

Power, C. & Moynihan, C. (1987) Social class and changes in weight-for-height between childhood and early adulthood. *International Journal of Obesity*, **12**, 445-453.

Power, R. A., Tansey, K. E., Buttenschon, H. N.*, et al* (2016) Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. *Biol Psychiatry*.

Price, A. L., Patterson, N. J., Plenge, R. M.*, et al* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, **38**, 904-909.

Prien, R. F., Carpenter, L. L. & Kupfer, D. J. (1991) The definition and operational criteria for treatment outcome of major depressive disorder. A review of the current research literature. *Arch Gen Psychiatry*, **48**, 796-800.

Pritchard, J. K. & Przeworski, M. (2001) Linkage disequilibrium in humans: models and data. *Am J Hum Genet*, **69**, 1-14.

Psychiatric, G. C. C. C., Cichon, S., Craddock, N.*, et al* (2009) Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*, **166**, 540-556.

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

Purcell, S., Neale, B., Todd-Brown, K.*, et al* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559-575.

Qi, Q., Chu, A. Y., Kang, J. H.*, et al* (2014) *Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies*.

Qi, Q., Chu, A. Y., Kang, J. H.*, et al* (2012) Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med*, **367**, 1387-1396.

Qi, Q., Li, Y., Chomistek, A. K.*, et al* (2012) Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men. *Circulation*, **126**, 1821-1827.

Quitkin, F. M., Stewart, J. W., McGrath, P. J.*, et al* (1993) Columbia atypical depression: a subgroup of depressives with better response to MAOI than to tricyclic antidepressants or placebo. *The British Journal of Psychiatry*.

Raj, A. & van Oudenaarden, A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, **135**, 216-226.

Rapee, R. M., Schniering, C. A. & Hudson, J. L. (2009) Anxiety disorders during childhood and adolescence: origins and treatment. *Annu Rev Clin Psychol*, **5**, 311-341.

Reddon, H., Gerstein, H. C., Engert, J. C.*, et al* (2016) Physical activity and genetic predisposition to obesity in a multiethnic longitudinal study. *Sci Rep*, **6**, 18672.

Remak, R. (1862) Translations: On the Embryological Basis of the Cell-Theory. *Journal of Cell Science*, **s2-2**, 277-284.

Remes, O., Brayne, C., van der Linde, R.*, et al* (2016) A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav*, **6**, e00497.

Rennie, K. L. & Wareham, N. J. (1998) The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls. *Public Health Nutr.,* **1,** 265-271.

Rhee, K. (2008) Childhood overweight and the relationship between parent behaviors, parenting style ' and family functioning. *Annals of the American Academy of Political and Social Science*, **615**, 12-37.

Rhee, K. E., Lumeng, J. C., Appugliese, D. P.*, et al* (2006) Parenting styles and overweight status in first grade. *Pediatrics*, **117**, 2047-2054.

Rice, L. & Greenberg, L. (1984) The new research paradigm. *Patterns of change*, 7-25.

Risch, N., Herrell, R., Lehner, T.*, et al* (2009) Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA*, **301**, 2462-2471.

Ritchie, M. D., Holzinger, E. R., Li, R.*, et al* (2015) Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet*, **16**, 85-97.
Ritchie, M. E., Phipson, B., Wu, D.*, et al* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43**, e47.
Rivera, M., Cohen-Woods, S., Kapur, K.*, et al* (2012) Depressive disorder moderates the effect of the FTO gene on body mass index. *Mol Psychiatry*, **17**, 604-611.
Roberts, S., Wong, C. C. Y., Keers, R.*, et al* (Under Review) Genome-wide expression and response to exposure-based psychological therapy.
Rockman, M. V. & Kruglyak, L. (2006) Genetics of global gene expression. *Nat Rev Genet*, **7**, 862-872.
Rosenthal, D. (1963) A Suggested Conceptual Framework.
Roth, R. B., Hevezi, P., Lee, J.*, et al* (2006) Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics*, **7**, 67-80.
Rothman, K. J., Greenland, S. & Walker, A. M. (1980) Concepts of interaction. *Am J Epidemiol*, **112**, 467-470.
Ruel, M. T. (2003) Operationalizing Dietary Diversity: A Review of Measurement Issues and Research Priorities. *J.Nutr.,* **133**, 3911S-3926S
Rutter, M. (2010) Gene-environment interplay. *Depress Anxiety*, **27**, 1-4.
Sadeh, N., Wolf, E. J., Logue, M. W.*, et al* (2015) Polygenic Risk for Externalizing Psychopathology and Executive Dysfunction in Trauma-Exposed Veterans. *Clinical Psychological Science*, **4**, 545-558.
Sakolsky, D. (2011) Serotonin Transporter Variation and Treatment Response in the Child/Adolescent Anxiety Multimodal Study (CAMS). In *AACAP+ CACAP Joint Annual Meeting (October 18-23, 2011)*: AACAP.
Salvatore, J. E., Aliev, F., Bucholz, K.*, et al* (2015) Polygenic risk for externalizing disorders: Gene-by-development and gene-by-environment effects in adolescents and young adults. *Clin Psychol Sci*, **3**, 189-201.
Salvatore, J. E., Aliev, F., Edwards, A. C.*, et al* (2014) Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. *Genes (Basel)*, **5**, 330-346.
Samaan, Z., Anand, S. S., Zhang, X.*, et al* (2013) The protective effect of the obesity-associated rs9939609 A variant in fat mass- and obesity-associated gene on depression. *Mol Psychiatry*, **18**, 1281-1286.
Samaan, Z., Lee, Y. K., Gerstein, H. C.*, et al* (2015) Obesity genes and risk of major depressive disorder in a multiethnic population: a cross-sectional study. *J Clin Psychiatry*, **76**, e1611-1618.
Sampasa-Kanyinga, H. & Chaput, J-P. (2017) Associations among self-perceived work and life stress, trouble sleeping, physical activity, and body weight among Canadian adults. *Preventive Medicine,* **96**, 16-20
Sapolsky, R. J., Hsie, L., Berno, A.*, et al* (1999) High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet Anal*, **14**, 187-192.
Sarkar, S. & Pfeifer, J. (2006) The Philosophy of Science: An Encyclopedia.
Scaini, S., Belotti, R. & Ogliari, A. (2014) Genetic and environmental contributions to social anxiety across different ages: a meta-analytic approach to twin data. *J Anxiety Disord*, **28**, 650-656.
Schadt, E. E., Monks, S. A., Drake, T. A.*, et al* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297-302.
Schena, M., Shalon, D., Davis, R. W.*, et al* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-470.
Scherer, A. (2009) *Batch effects and noise in microarray experiments: sources and solutions*: John Wiley & Sons.

Schizophrenia Psychiatric Genome-Wide Association Study, C. (2011) Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*, **43**, 969-976.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421-427.

Schneider, S., Blatter-Meunier, J., Herren, C.*, et al* (2013) The efficacy of a family-based cognitive-behavioral treatment for separation anxiety disorder in children aged 8-13: a randomized comparison with a general anxiety program. *J Consult Clin Psychol*, **81**, 932-940.

Schur, E., Godfrey, K. M., Dansie, E.*, et al* (2013) Can familial factors account for the association of body mass index with poor mental health in men or women? *General Hospital Psychiatry*, **35**, 502-507.

Schwartz, M. W., Woods, S. C., Porte, D., Jr.*, et al* (2000) Central nervous system control of food intake. *Nature*, **404**, 661-671.

Sekar, A., Bialas, A. R., de Rivera, H.*, et al* (2016) Schizophrenia risk from complex variation of complement component 4. *Nature*, **530**, 177-183.

Shabalin, A. A. (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, **28**, 1353-1358.

Sham, P., Cherny, S., Purcell, S.*, et al* (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *The American Journal of Human Genetics*, **66**, 1616-1630.

Sharpley, C. F., Palanisamy, S. K., Glyde, N. S.*, et al* (2014) An update on the interaction between the serotonin transporter promoter variant (5-HTTLPR), stress and depression, plus an exploration of non-confirming findings. *Behav Brain Res*, **273**, 89-105.

Shear, K., Jin, R., Ruscio, A. M.*, et al* (2006) Prevalence and correlates of estimated DSM-IV child and adult separation anxiety disorder in the National Comorbidity Survey Replication. *Am J Psychiatry*, **163**, 1074-1083.

Shimada-Sugimoto, M., Otowa, T. & Hettema, J. M. (2015) Genetics of anxiety disorders: Genetic epidemiological and molecular studies in humans. *Psychiatry Clin Neurosci*, **69**, 388-401.

Shimada-Sugimoto, M., Otowa, T., Miyagawa, T.*, et al* (2015) Immune-related pathways including HLA-DRB1(*)13:02 are associated with panic disorder. *Brain Behav Immun*, **46**, 96-103.

Shimada-Sugimoto, M., Otowa, T., Miyagawa, T.*, et al* (2016) Polymorphisms in the TMEM132D region are associated with panic disorder in HLA-DRB1*13:02-negative individuals of a Japanese population. *Hum Genome Var*, **3**, 16001.

Shirk, S. R. & Saiz, C. C. (1992) Clinical, Empirical, and Developmental Perspectives on the Therapeutic Relationship in Child-Psychotherapy. *Dev Psychopathol*, **4**, 713-728.

Shloim, N., Edelson, L. R., Martin, N.*, et al* (2015) Parenting Styles, Feeding Styles, Feeding Practices, and Weight Status in 4-12 Year-Old Children: A Systematic Review of the Literature. *Front Psychol*, **6**, 1849.

Shrewsbury, V. & Wardle, J. (2008) Socioeconomic Status and Adiposity in Childhood: A Systematic Review of Cross-sectional Studies 1990–2005. *Obesity (Silver Spring)*, **16,** 275-284.

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C.*, et al* (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, **518**, 187-196.

Silk, J.S., Ziegler, M. L., Whalen, D. J., *et al* (2009) Expressed Emotion in Mothers of Currently Depressed, Remitted, High-Risk, and Low-Risk Youth: Links to Child Depression Status and Longitudinal Course. *J Clin. Child Adolesc Psychol,* **1,** 36-47.

Silove, D., Alonso, J., Bromet, E.*, et al* (2015) Pediatric-Onset and Adult-Onset Separation Anxiety Disorder Across Countries in the World Mental Health Survey. *Am J Psychiatry*, **172**, 647-656.

Simonsohn, U., Nelson, L. D. & Simmons, J. P. (2014) P-curve: a key to the file-drawer. *J Exp Psychol Gen*, **143**, 534-547.

Smemo, S., Tena, J. J., Kim, K. H.*, et al* (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, **507**, 371-375.

Smith, A. K., Kilaru, V., Klengel, T.*, et al* (2015) DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain. *Am J Med Genet B Neuropsychiatr Genet*, **168B**, 36-44.

Smoller, J. W. (2014) Psychiatric genetics and the future of personalized treatment. *Depress Anxiety*, **31**, 893-898.

Sobal, J. & Stunkard, A. J. (1989) Socioeconomic status and obesity: a review of the literature. *Psychol. Bull.,* **105,** 260**–**275**.**

Solovieff, N., Roberts, A. L., Ratanatharathorn, A.*, et al* (2014) Genetic association analysis of 300 genes identifies a risk haplotype in SLC18A2 for post-traumatic stress disorder in two independent samples. *Neuropsychopharmacology*, **39**, 1872-1879.

Speed, D. & Balding, D. J. (2014) MultiBLUP: improved SNP-based prediction for complex traits. *Genome Res*, **24**, 1550-1557.

Speliotes, E. K., Willer, C. J., Berndt, S. I.*, et al* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*, **42**, 937-948.

Spencer, C. C., Su, Z., Donnelly, P.*, et al* (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet*, **5**, e1000477.

Spencer, E. A., Appleby, P. N., Davey, G. K., *et al* (2003) Diet and body mass index in 38 000 EPIC-Oxford meat-eaters, fish-eaters, vegetarians and vegans. *International Journal of Obesity*, **27,** 728-734.

Stanworth, S. J., Roberts, N. A., Sharpe, J. A.*, et al* (1995) Established epigenetic modifications determine the expression of developmentally regulated globin genes in somatic cell hybrids. *Mol Cell Biol*, **15**, 3969-3978.

Steemers, F. J., Chang, W., Lee, G.*, et al* (2006) Whole-genome genotyping with the single-base extension assay. *Nat Methods*, **3**, 31-33.

Stein, M. B., Chen, C. Y., Ursano, R. J.*, et al* (2016) Genome-wide Association Studies of Posttraumatic Stress Disorder in 2 Cohorts of US Army Soldiers. *JAMA Psychiatry*, **73**, 695-704.

Stumpf, M. P., Wiuf, C. & May, R. M. (2005) Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc Natl Acad Sci U S A*, **102**, 4221-4224.

Subramanian, A., Tamayo, P., Mootha, V. K.*, et al* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, **102**, 15545-15550.

Sundquist, J. & Johansson, S. E. (1998) The influence of socioeconomic status, ethnicity and lifestyle on body mass index in a longitudinal study. *Int J Epidemiol*, **27**, 57-63.

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

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

Sullivan, P. F., Fan, C. & Perou, C. M. (2006) Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet*, **141B**, 261-268.

Sullivan, P. F., Kessler, R. C. & Kendler, K. S. (2014) Latent class analysis of lifetime depressive symptoms in the national comorbidity survey.

Sumner, J. A., Pietrzak, R. H., Aiello, A. E.*, et al* (2014) Further support for an association between the memory-related gene WWC1 and posttraumatic stress disorder: results from the Detroit Neighborhood Health Study. *Biol Psychiatry*, **76**, e25-26.

Sun, H., Treco, D., Schultes, N. P.*, et al* (1989) Double-Strand Breaks at an Initiation Site for Meiotic Gene Conversion. *Nature*, **338**, 87-90.

Suveg, C., Zeman, J., Flannery-Schroeder, E., *et al* (2005) Emotion socialization in families of children with an anxiety disorder**.** *Journal of Abnormal Child Psychology,* **33,** 145–155.

Tambs, K., Czajkowsky, N., Roysamb, E.*, et al* (2009) Structure of genetic and environmental risk factors for dimensional representations of DSM-IV anxiety disorders. *Br J Psychiatry*, **195**, 301-307.

Tanaka, R. (2005) Scale-rich metabolic networks. *Phys Rev Lett*, **94**, 168101.

Tansey, K. E., Guipponi, M., Hu, X.*, et al* (2013) Contribution of common genetic variants to antidepressant response. *Biol Psychiatry*, **73**, 679-682.

Taylor, A. E., Ebrahim, S., Ben-Shlomo, Y.*, et al* (2010) Comparison of the associations of body mass index and measures of central adiposity and fat mass with coronary heart disease, diabetes, and all-cause mortality: a study using data from 4 UK cohorts. *Am J Clin Nutr*, **91**, 547-556.

Taylor, A. E. & Munafo, M. R. (2016) Triangulating meta-analyses: the example of the serotonin transporter gene, stressful life events and major depression. *BMC Psychol*, **4**, 23.

Taylor, S., Abramowitz, J. S. & McKay, D. (2012) Non-adherence and non-response in the treatment of anxiety disorders. *J Anxiety Disord*, **26**, 583-589.

Te Morenga, L., Mallard, S. & Mann, J. (2013) Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. *BMJ,* **346,** e7492

Teo, Y. Y. (2008) Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Curr Opin Lipidol*, **19**, 133-143.

The International HapMap, C. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299-1320.

Thomas, D. (2010) Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet*, **11**, 259-272.

Thomas, D. C. (2004) *Statistical methods in genetic epidemiology*: Oxford University Press.

Tompson, M. C., Pierre, C. B., Dingman Boger, K. *et al* (2010) Maternal Depression, Maternal Expressed Emotion, and Youth Psychopathology. *Journal of Abnormal Child Psychology,* **1,** 105-117.

Traylor, M., Markus, H. & Lewis, C. M. (2015) Homogeneous case subgroups increase power in genetic association studies. *Eur J Hum Genet*, **23**, 863-869.

Tsui, N. B., Ng, E. K. & Lo, Y. M. (2002) Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin Chem*, **48**, 1647-1653.

Turkheimer, E. (2000) Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, **9**, 160-164.

Turkheimer, E. & Gottesman, I. I. (1991) Is H 2= 0 a null hypothesis anymore? *Behavioral and Brain Sciences*, **14**, 410-411.

Turkheimer, E. & Waldron, M. (2000) Nonshared environment: a theoretical, methodological, and quantitative review. *Psychol Bull*, **126**, 78-108.

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

Van Der Bruggen, C. O., Stams, G. J. J. M. & Bögels, S. M. (2008) Research Review: The relation between child and parent anxiety and parental control: a meta-analytic review. *J Child Psychol Psychiatry*, **49,** 1257-1269

Van Houtem, C. M., Laine, M. L., Boomsma, D. I.*, et al* (2013) A review and meta-analysis of the heritability of specific phobia subtypes and corresponding fears. *J Anxiety Disord*, **27**, 379-388.

VanderWeele, T. J., Asomaning, K., Tchetgen Tchetgen, E. J.*, et al* (2012) Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction. *Am J Epidemiol*, **175**, 1013-1020.

Vaughn, C.E. & Leff, J. P. (1976) The influence of family and social factors on the course of psychiatric illness: A comparison of schizophrenic and depressed neurotic patients. *Br J Psychiatry*, **129,** 125-137.

Venter, J. C., Adams, M. D., Myers, E. W.*, et al* (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.

Vinkhuyzen, A. A. E. & Wray, N. R. (2015) Novel directions for G x E analysis in psychiatry. *Epidemiol. Psychiatr. Sci,* **24**, 12-19.

Visscher, P. M., Brown, M. A., McCarthy, M. I.*, et al* (2012) Five years of GWAS discovery. *Am J Hum Genet*, **90**, 7-24.

Visscher, P. M., Goddard, M. E., Derks, E. M.*, et al* (2012) Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatry*, **17**, 474-485.

Visscher, P. M., Hemani, G., Vinkhuyzen, A. A.*, et al* (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet*, **10**, e1004269.

Visscher, P. M., McEvoy, B. & Yang, J. (2010) From Galton to GWAS: quantitative genetics of human height. *Genet Res (Camb)*, **92**, 371-379.

Visscher, P. M., Yang, J. & Goddard, M. E. (2010) A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res Hum Genet*, **13**, 517-524.

Vrshek-Schallhorn, S., Stroud, C. B., Mineka, S.*, et al* (2015) Additive genetic risk from five serotonin system polymorphisms interacts with interpersonal stress to predict depression. *J Abnorm Psychol*, **124**, 776-790.

Waite, P. & Creswell, C. (2014) Children and adolescents referred for treatment of anxiety disorders: differences in clinical characteristics. *J Affect Disord*, **167**, 326-332.

Walter, S., Glymour, M. M., Koenen, K.*, et al* (2013) Performance of polygenic scores for predicting phobic anxiety. *PLoS ONE*, **8**, e80326.

Wang, M. C., Kim, S., Gonzalez, A. A.*, et al* (2007) Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *J Epidemiol Community Health*, **61**, 491-498.

Wang, K., Hu, X., & Peng, Y., (2013) An analytical comparison of the principal component method and the mixed effects model for association studies in the presence of cryptic relatedness and population stratification. *Human heredity,* **76,** 1-9.

Wang, Z., Harrer, J., Tuerk, P.*, et al* (2009) 5-HTTLPR influence PTSD treatment outcome. In *Biol Psychiatry*, pp. 119S-119S: ELSEVIER SCIENCE INC 360 PARK AVE SOUTH, NEW YORK, NY 10010-1710 USA.

Ward, L. D. & Kellis, M. (2012) Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol*, **30**, 1095-1106.

Ware, J. J., van den Bree, M. B. & Munafo, M. R. (2011) Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: a meta-analysis. *Nicotine Tob Res*, **13**, 1167-1175.

Wareham, N. J., van Sluijs, E. M. F. & Ekelund, U. (2005) Physical activity and obesity prevention: a review of the current evidence. *Proc Natl Acad Sci U S A,* **64,** 229-247

Weale, M. E. (2010) Quality control for genome-wide association studies. *Methods Mol Biol*, **628**, 341-372.

Wearden, A. J., Tarrier, N., Barrowclough, C., *et al* (2000) A review of expressed emotion research in health care. *Clinical Psychology Review,* **20,** 633-666.

Wei, W-H., Hemani, G. & Haley, C. S. (2014) Detecting epistasis in human complex traits. *Nat Rev Genet,* **15**, 722-733.

Weismann, A., Poulton, E. B., Schönland, S.*, et al* (1891) *Essays upon heredity and kindred biological problems*: Clarendon press.

Wellcome Trust Case Control, C. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661-678.

Westra, H. J., Peters, M. J., Esko, T.*, et al* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*, **45**, 1238-U1195.

Whiteford, H. A., Degenhardt, L., Rehm, J.*, et al* (2013) Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*, **382**, 1575-1586.

Wilhelm, M., Schlegl, J., Hahne, H., *et al* (2014) Mass-spectrometry-based draft of the human proteome. *Nature,* **509,** 582-587.

Winkler, T. W., Justice, A. E., Graff, M.*, et al* (2015) The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet*, **11**, e1005378.

Wolf, E. J., Rasmusson, A. M., Mitchell, K. S.*, et al* (2014) A genome-wide association study of clinical symptoms of dissociation in a trauma-exposed sample. *Depress Anxiety*, **31**, 352-360.

Wray, N. R., Goddard, M. E. & Visscher, P. M. (2007) Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res*, **17**, 1520-1528.

Wray, N. R., Lee, S. H., Mehta, D.*, et al* (2014) Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry*, **55**, 1068-1087.

Wray, N. R., Yang, J., Hayes, B. J.*, et al* (2013) Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet*, **14**, 507-515.

Wright, S. (1921) Systems of mating. I. The biometric relations between parent and offspring. *Genetics*, **6**, 111-123.

Xie, P., Kranzler, H. R., Yang, C.*, et al* (2013) Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. *Biol Psychiatry*, **74**, 656-663.

Yang, J., Bakshi, A., Zhu, Z.*, et al* (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*, **47**, 1114-1120.

Yang, J., Benyamin, B., McEvoy, B. P.*, et al* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*, **42**, 565-569.

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

Yang, J., Visscher, P. M. & Wray, N. R. (2010) Sporadic cases are the norm for complex disease. *Eur J Hum Genet*, **18**, 1039-1043.

Yang, J., Weedon, M. N., Purcell, S.*, et al* (2011) Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet*, **19**, 807-812.

Yang, J., Zaitlen, N. A., Goddard, M. E.*, et al* (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*, **46**, 100-106.

Yang, Q. & Khoury, M. J. (1997) Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. *Epidemiol Rev*, **19**, 33-43.
Yap, M. B. H., Pilkington, P. D., Ryan, S. M. *et al* (2014) Parental factors associated with depression and anxiety in young people: A systematic review and meta-analysis. *Journal of Affective Disorders,* **156,** 8-23.
Yehuda, R., Daskalakis, N. P., Desarnaud, F.*, et al* (2013) Epigenetic Biomarkers as Predictors and Correlates of Symptom Improvement Following Psychotherapy in Combat Veterans with PTSD. *Front Psychiatry*, **4**, 118.
Yzerbyt, V. Y., Muller, D. & Judd, C. M. (2004) Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, **40**, 424-431.
Zaitlen, N., Kraft, P., Patterson, N.*, et al* (2013) Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet*, **9**, e1003520.
Zammit, S., Lewis, G., Dalman, C.*, et al* (2010) Examining interactions between risk factors for psychosis. *Br J Psychiatry*, **197**, 207-211.
Zhang, B. & Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, **4**, Article17.
Zhang, Y. & Pan, W. (2015) Principal component regression and linear mixed model in association analysis of structured samples: competitors or complements? *Genet Epidemiol*, **39**, 149-155.
Zhang, Y., Proenca, R., Maffei, M.*, et al* (1994) Positional cloning of the mouse obese gene and its human homologue. *Nature*, **372**, 425-432.
Zhu, Z., Zhang, F., Hu, H.*, et al* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*, **48**, 481-487.
Zollner, S. & Pritchard, J. K. (2007) Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*, **80**, 605-615.

Appendix I:

OXFORD

Briefings in Functional Genomics, 15(4), 2016, 298–304

doi: 10.1093/bfgp/elv037
Advance Access Publication Date: 5 October 2015
Review paper

Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray

Jonathan R. I. Coleman, Jack Euesden, Hamel Patel, Amos A. Folarin, Stephen Newhouse and Gerome Breen

Corresponding author: Jonathan Coleman, Social, Genetic and Developmental Psychiatry Centre (MRC), Institute of Psychiatry, Psychology and Neuroscience - PO80, DeCrespigny Park, Denmark Hill, London, SE5 8AF, UK. Tel.: +442078480411; E-mail: jonathan.coleman@kcl.ac.uk

Abstract

The decreasing cost of performing genome-wide association studies has made genomics widely accessible. However, there is a paucity of guidance for best practice in conducting such analyses. For the results of a study to be valid and replicable, multiple biases must be addressed in the course of data preparation and analysis. In addition, standardizing methods across small, independent studies would increase comparability and the potential for effective meta-analysis. This article provides a discussion of important aspects of quality control, imputation and analysis of genome-wide data from a low-coverage microarray, as well as a straight-forward guide to performing a genome-wide association study. A detailed protocol is provided online, with example scripts available at https://github.com/JoniColeman/gwas_scripts.

Key words: GWAS; methods; low-coverage microarray; imputation; analysis

Introduction

Genome-wide association studies (GWAS) are widely used to assess the impact of common genetic variation on a variety of phenotypes [1, 2]. Low-cost microarrays designed to assay thousands of variants and to be imputable to millions, such as the Illumina HumanCoreExome microarray (Illumina, San Diego, CA, USA), have increased the accessibility of this technology. Although the rapid development and falling cost of whole-genome sequencing is likely to reduce the use of GWAS in the long term, the costs of running a GWAS are currently an order of magnitude smaller than those for sequencing, suggesting GWAS will remain an important technique into the near future [3].

However, there is a paucity of information on best practice for using the data resulting from microarray-based genotyping. Excellent theoretical and practical protocols for the quality control of genome-wide genotype data exist [4, 5], and most commonly used software have well-constructed user manuals, but structured advice to guide analysis is missing from the literature. To date, a

Jonathan Coleman is a PhD student at the MRC Social, Genetic and Developmental Psychiatry Centre (SGDP), using genomic methods to explore differential response to psychological treatments for anxiety disorders.
Jack Euesden is a PhD student at the SGDP. His interests include developing new methods to understand the genetic architecture of, and epidemiological relationship between, psychiatric and other medical disorders.
Hamel Patel is a PhD student at the SGDP and the National Institute for Health Research Biomedical Research Centre for Mental Health (NIHR BRC MH) Bioinformatics Core, South London and Maudsley NHS Trust. He is involved in developing bioinformatics pipelines and protocols for the analysis of genotyping and sequencing data.
Amos Folarin is a senior software developer and bioinformatician at the NIHR BRC MH Bioinformatics Core, using bioinformatics for drug screening, target identification and disease analysis.
Stephen Newhouse is a senior bioinformatician at the NIHR BRC MH Bioinformatics Core, with a focus on translational bioinformatics and the genetics of complex disorders.
Gerome Breen is a senior lecturer at the SGDP, and Theme Lead for the Genomics and Biomarkers and BioResource for Mental and Neurological Health themes at the NIHR BRC MH. His research interests are the genetics of complex psychiatric and co-morbid disorders.

considerable proportion of the analysis of such data has been concentrated within large consortia (such as the Psychiatric Genomics Consortium), with experienced analysts and in-house protocols [6, 7]. However, such guidance is not easily available to groups outside these consortia. As the accessibility of genome-wide data increases, so must the accessibility of advice on its analysis. Furthermore, a standardized approach would increase comparability between studies, facilitating further investigations such as meta-analysis and augmenting the value of each individual study [8].

The choices made in conducting the analysis of genotype data affect the final result. At worst, poor quality control can lead to systematic biases in outcome and increased false-positive (and false-negative) associations [4]. However, the effects can be more nuanced; for example, association testing using a mixed linear model may use a genetic relatedness matrix (GRM) to control for gross genetic similarity between individuals [9, 10]. The precise pairwise relationships will differ subtly depending on whether the GRM is made using the genotype data before or after imputation (as well as on the programme used), and so the results of the association study will also differ slightly. Neither choice in this context is wrong, but the choice made has consequences, and as such needs to be considered and reported [11].

Recently, we performed the first genome-wide association study of response to cognitive behavioural therapy, using the HumanCoreExome microarray (Coleman et al, Under Review). In this protocol, we have used that experience to provide suggestions for the quality control, imputation and analysis of data from this microarray, assuming careful recalling of the raw intensity data has been performed. The steps are likely to be applicable to data from other arrays, with the caveat that differences in array content may require alteration of the various thresholds discussed. The analysis of genome-wide data remains a data-driven activity, and, where appropriate, we have provided advice on making informed choices from the data. Furthermore, we recommend consulting graphical representations of the data when defining thresholds.

Pre-analytical procedures: genotyping, calling and recalling

This protocol describes the basic analytical steps required to conduct a genome-wide association study; it is expected that DNA genotyping and genotype recalling have already been performed. In this context, genotyping refers to the hybridization of genomic DNA to oligonucleotide probes targeted at a polymorphic region, and the extension of these probes to encompass this region. This extension uses chemically labelled nucleotides that are specific to the different alleles of the polymorphism and that bind either red or green fluorescent agents, which can be read using a fluorescence-sensitive scanner. The end product of genotyping is the raw intensity data of these fluorescent agents at each polymorphic site [12]. To determine the identity of the alleles at these sites, the raw intensity data must be called—clusters of samples with similar intensities are identified, and the clusters are labelled according to the design of the microarray. This initial calling is performed by automated software—however, the algorithms to perform this calling sometimes fail to identify valid clusters, especially when patterns of clustering are unusual. As such, some clusters must be identified by manual recalling by a bioinformatician. Recalling is an extremely important step—badly called genotypes create biases that severely impair the quality control and analysis of data. The complexities of genotyping and recalling are beyond the scope of this protocol, but guidance is available from array manufacturers and as referenced in the online protocol [13].

Considerations in conducting a study

The value of any finding in molecular genetics is reliant on the ability to replicate it in an independent cohort, and the first step to successful replication is to minimize the likelihood that reported findings are false positives. Given that thousands of variants are assessed in a GWAS, and the potential for random error in genotyping and recalling (as discussed above), it is necessary to impose stringent thresholds on the quality of data to be taken forward to analysis [4]. Pre-analytical steps partly inform these thresholds. When a more variable method of collection has been used, it is advisable to consider more stringent quality control parameters; for example, collection using buccal swabs produces poorer quality DNA than extractions from whole blood or saliva [14].

Quality control: selecting variants by allele frequency

Following genotyping and the recalling of genotypes, most GWAS studies begin by filtering the variants by the frequency of the less-common allele (minor allele frequency or MAF). Variant MAF has many effects on later analysis, as allele frequency is associated with time since mutation, the structure of local linkage disequilibrium (LD) and the relative size of the association statistic [15, 16]. The chances of an error in genotype calling increase with decreasing MAF, as the certainty of manual and automatic clustering falls with fewer variants in each cluster [17]. At the most extreme level, if all but one variant cluster together, it is difficult to assess whether the lone variant is truly a different genotype, or whether it is a missed call. For this reason, the rarest variants should be discarded from the analysis. What constitutes 'rare' depends on the size of the studied cohort—assuming perfect Hardy–Weinberg equilibrium, the minor allele of a variant with MAF = 0.1% would be expected to be present in 19 heterozygotes and 1 homozygote in a cohort of 10 000 individuals, but only one or two heterozygotes would be expected in a cohort of 1000 individuals. In smaller cohorts, a more stringent MAF cut-off is recommended, as the minor allele count will be lower, which limits the value of conclusions from the analysis of these variants. For the smallest studies, where fewer than 1000 individuals are investigated, a cut-off of 5% should be considered—this is in line with the analysis program GenAbel, for example, which uses a minor allele count of 5 as its cut-off [18]. Typically, many studies define rare single nucleotide polymorphisms (SNPs) as having a MAF <1%, which has historical roots in the HapMap project [19]. It is worth noting that the exonic content of the HumanCoreExome chip was specifically designed to target coding variants, with much of this content having a population MAF <1% [17]. Therefore, using this microarray in smaller cohorts and imposing a MAF cut-off of 1% or higher will result in discarding most of the exonic content.

Quality control: removing variants and samples with missing data

It is necessary to remove rare variants from GWAS because the certainty of the genotype call is reduced by their low minor allele count. Even in common variants, however, genotyping and genotype recalling are subject to technical error, with the result that a proportion of variants and samples are of low quality, and should be removed from the analysis. Removal of such missing variants and samples is best conducted in an iterative manner, removing variants genotyped in <90% of the samples,

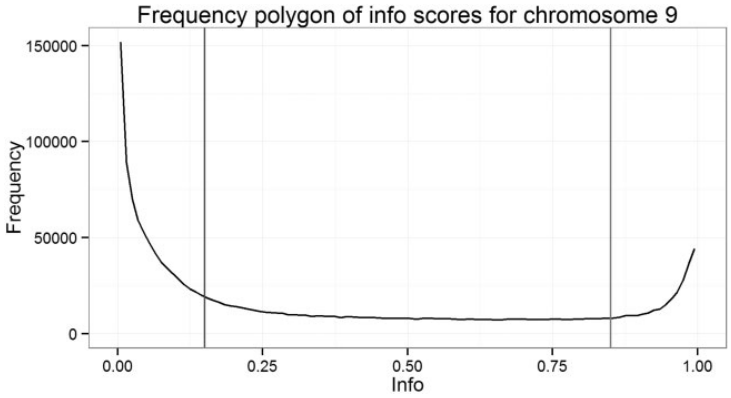


Figure 1. Frequency polygon showing the number of variants at each info value post-imputation, including poor-quality variants to be excluded (info <0.15) and higher-quality variants that should be kept (info >0.85).

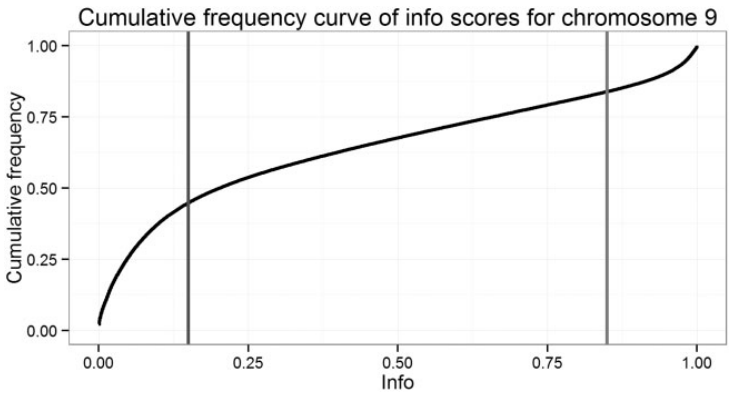


Figure 2. Cumulative frequency curve showing the same data as Figure 1.

then samples with <90% of variants and continuing with increasing stringency to a user-defined final threshold (typically in the range of 95–99% completeness, depending on the required stringency of quality control). This has benefits over removing all variants and samples beneath the final threshold, as fewer samples are lost using the iterative procedure (at the expense of a slight increase in variant exclusions).

Quality control: assessing deviation from Hardy-Weinberg equilibrium

Thresholds that identify missing variants do not necessarily exclude miscalled variants. For example, clustering algorithms can incorrectly define a group of samples as heterozygous. One method to detect this is to evaluate the deviation from Hardy–Weinberg equilibrium at each variant. Although such deviations can be caused by processes that may be of interest within the study, such as selection pressure, the expected size of such deviations is small. Setting the threshold for the *P*-value of the Hardy–Weinberg test to be low ($P < 1 \times 10^{-5}$) decreases the probability of excluding deviations that result from processes of interest. In case-control studies, it is recommended to remove SNPs deviant in controls only (this is the default behaviour in PLINK2). Deviations from Hardy–Weinberg equilibrium as a result of genotyping artefacts are not expected to differ between cases and controls, but biologically relevant deviations are more likely to occur in cases [5]. The threshold for the *P*-value cut-off can be determined empirically, by examining the spread of *P*-values from the Hardy–Weinberg test in the data, and selecting a threshold under which there are a greater number of variants than expected by chance (in our experience, with small data sets, this is typically around $P = 1 \times 10^{-5}$).

Quality control: pruning for LD and removing related samples

The initial quality control steps described above correct for the random errors introduced by genotyping and recalling. Further steps are required to address cryptic structure, the presence of similarities between individuals independent of the phenotype under study, which present a source of potential bias in the outcome of association tests. Such structure is commonly envisaged as two interconnected concepts, high relatedness between individuals (determined by the proportion of their genomes identical-by-descent—IBD) and population stratification. The presence of structure is inferred from examining genome-wide genotype data. However, the phenomenon of LD can exaggerate or obscure similarities, as a shared region of high LD results in more shared

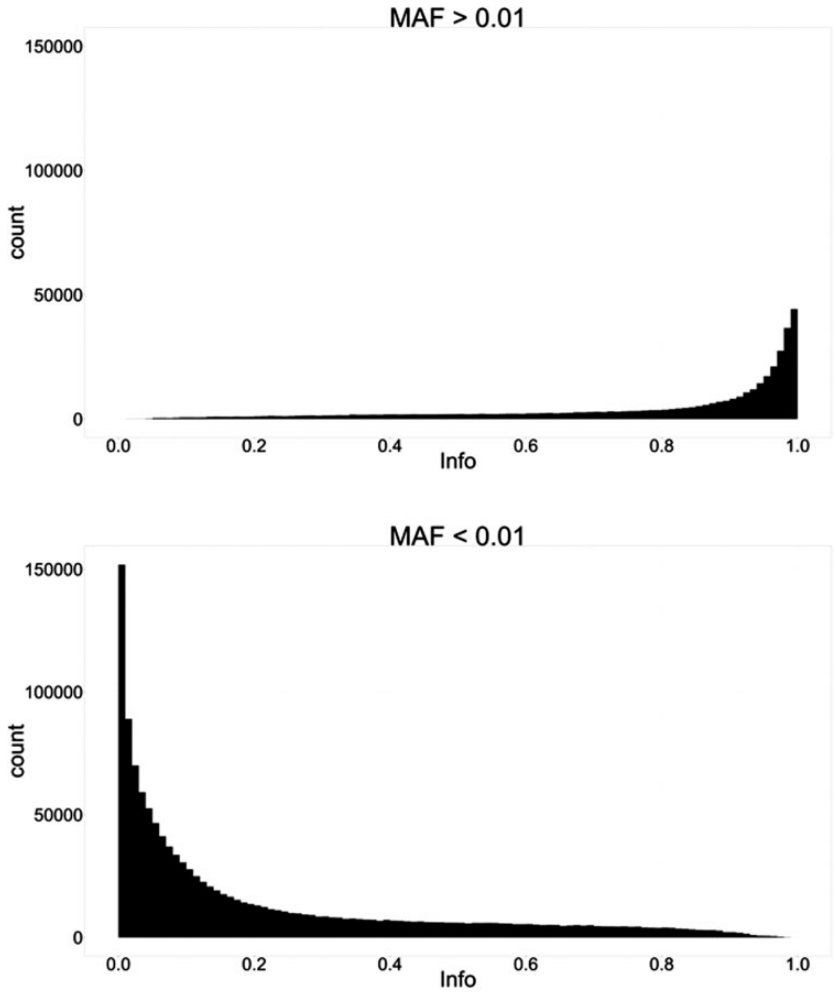


Figure 3. Histograms of the info metric of imputed variants on chromosome 9, split by MAF at 0.01.

variants than one of low LD, even if the two regions are the same size. Accordingly, it is necessary to prune the data for LD before assessing IBD and population stratification. This can be achieved using a pairwise comparison method, comparing each possible pair of variants in a given window of variants and removing one of the pair if the LD between them is above a given cut-off. This protocol uses a window of 1500 variants, shifted by 10% for each new round of comparisons, and a threshold of $R^2 > 0.2$. The window size of 1500 variants corresponds to the large, high LD chromosome 8 inversion, while the shift of 10% represents a trade-off between efficiency and thoroughness [5].

Once an LD-pruned data set is obtained, individuals can be compared pairwise to establish the proportion of variants they share identical-by-state (IBS). Closely related individuals share more of their genome than a randomly chosen pair of individuals from the population, and are likely to be more phenotypically similar. As a result, including closely related individuals can skew analysis; genetic variants shared because of close relatedness can become falsely associated with phenotypic similarity that also results from close relatedness.

With a sufficiently homogeneous cohort assayed at thousands of variants, IBS information can be used to infer variants that are shared identical-by-descent (IBD) [20]. Individuals with an IBD metric (pi-hat) > 0.1875 (halfway between a second and third degree relative [4]) should be removed, as well as individuals with unusually high average IBD with all other individuals, which may indicate sample contamination or genotype recalling error leading to too many heterozygote calls [20]. The IBD threshold suggested here is designed to remove the most closely related individuals, while avoiding removing large numbers of samples through being overly stringent. It is worth noting that some downstream analysis programs impose much more severe IBD cut-offs (GREML estimation in GCTA, which produces an estimate of heritability from all assayed variants, uses 0.025), while other analyses account for between-sample relatedness as part of the analysis [9, 21]. What quality control is appropriate depends on the nature of the cohort, the question being asked and the analysis methods intended to be used.

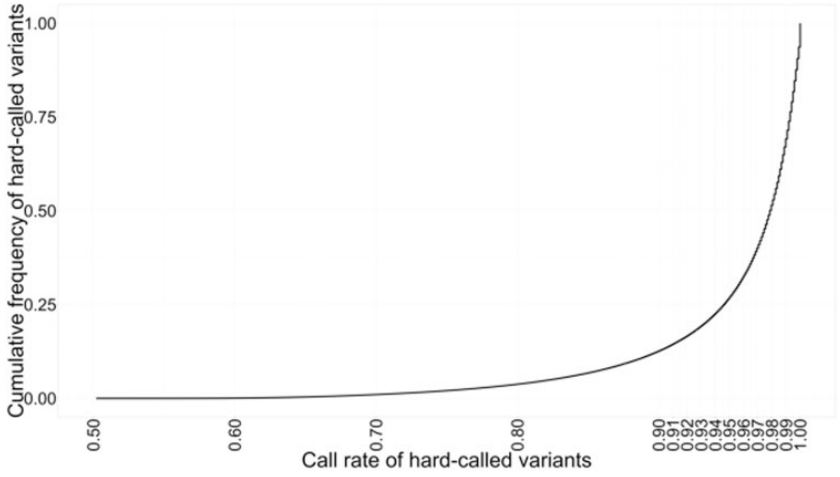


Figure 4. Cumulative frequency plot of call rate of hard-called imputed SNPs (genome wide).

Quality control: confirming sample gender and assessing the inbreeding coefficient

Samples whose reported gender differs from that suggested by their genes are likely to have been assigned the wrong identity. This leads to reduced power, as the sample's genotype becomes effectively randomized in respect to the phenotype. The average homozygosity of variants on the X chromosome (the X-chromosome *F* statistic) can be used to indicate sample gender. Much as it confounds estimates of IBD, patterns of LD will also impair chromosome-specific (and genome-wide) tests of homozygosity, and so it is necessary to perform this test following pruning for LD. The *F* statistic is a function of the deviation of the observed number of heterozygote variants from that expected under Hardy–Weinberg equilibrium. In males, $F \approx 1$, because all X chromosome variants are hemizygous, and so no heterozygotes are observed. Females are expected to have lower values of *F*, distributed normally around 0 [22]. However, this is an imprecise measure—female subjects with high *F* have been reported in the 1000 Genomes reference population (https://www.cog-genomics.org/plink2/basic_stats). As such, it is recommended that the <0.2 *F* threshold for females (as used by PLINK) is treated as guidance, and that further checks (such as counting the number of Y chromosome SNPs with data) are made, and that the phenotypic gender of discordant samples is confirmed with the collecting site where possible [20, 23].

In addition to using a chromosome-specific homozygosity check to confirm gender, a whole-genome *F* should also be calculated. This statistic is also referred to as an 'inbreeding coefficient', as inbreeding results in reduced numbers of heterozygotes. Individuals with particularly high or low inbreeding coefficients should be removed from analyses, as this is likely to be an artefact caused by genotyping error. However, caution is advised when studying cohorts in which consanguineous relationships are common, as high inbreeding coefficients are expected in these samples.

Quality control: controlling for population stratification

Similarities exist between the false genotype–phenotype correlations created by close between-sample relatedness and those created by population stratification, where phenotypic and genotypic similarity are correlated because of geographical location, rather than a true association. A variety of methods exist to control for population stratification, of which the most common is to perform principal component analysis on the genome-wide data, and then use the resulting components as covariates in association analysis. However, there is little guidance as to which components to choose, and this is often determined empirically in individual studies through piecemeal inclusion of principal components into the analysis until measures of genomic inflation fall below a chosen threshold (usually until the genomic inflation statistic lambda ≈ 1 [24]). We suggest an alternative, regressing principal components on outcome directly, and keeping only those that explain variance in the outcome at a rate above chance for use as covariates in the GWAS. This then leaves the question of what should be done if no component is associated with outcome. Recent computational developments have enabled an alternative means of control through the construction of genomic relatedness matrices [11]. This method compares the deviation of each individual from the population mean at each variant in the data set, and then compares individuals pairwise to establish a value for overall genetic similarity. This can then be entered into the analysis as a random variable in a mixed linear regression, and has the benefit of capturing population variance at a finer-scale level than principal component analysis [11] (for an in-depth discussion of the comparison between principal component analysis and genetic relatedness matrices, see [25]).

Imputation to the 1000 Genomes reference population

The main benefits of the HumanCoreExome as a low-cost microarray are twofold. First, the exonic content allows rare coding variation to be assayed in large numbers of samples without the high costs of sequencing these variants [26]. However, this relies on large sample sizes to allow for reliable calling of the genotypes. The value of the array in smaller cohorts is in providing an inexpensive means to assay thousands of variants that are in high LD with a considerably greater number. To make effective use of the array in this manner requires imputation of the data to

a reference population, most commonly the 1000 Genomes Reference [27]. However, the advent of large-scale sequencing studies such as UK10K (http://www.uk10k.org/) and Genomics England (http://www.genomicsengland.co.uk/), and the increasing availability of sequence data on specific populations, is likely to result in alterations to imputation practice in the near future.

The online protocol uses IMPUTE2 [28, 29] to impute to the full 1000 Genomes Reference population. This is performed without pre-phasing, as there is evidence that this is the most accurate method (albeit somewhat slower than pre-phasing; http://blog.goldenhelix.com/?p=1911). It also assumes access to a multi-node computing cluster, although jobs could be run sequentially (with considerable increases in computational time). The imputed data that result from these methods are provided in a probabilistic 'dosage' format, which is an attractive format from a statistical perspective, as it allows for the variable certainty of each imputed call to be considered within the association model. Programs exist that allow for the direct use of dosage data in association analyses, such as SNPTEST and ProbABEL (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/old/snptest.html; [30]). However, this format remains computationally burdensome at present—for example, it is not yet possible to store dosage data as a file input type in PLINK, akin to the PLINK binary format. As such, the protocol converts these probabilistic calls to binary 'hard' calls, marking less-certain calls as missing. This increases downstream flexibility at the expense of losing the more informative probabilistic calls. With increasing computational sophistication, it is likely that the use of dosage data as an input file type will become possible and commonplace; to this end, readers are advised to consult the PLINK2 website (https://www.cog-genomics.org/plink2/).

Post-imputation quality control: monomorphic, rare and missing variants

Following imputation, data are provided for a large number of variants (83 million in the latest release of the 1000 Genomes Project). As such, there is a necessity to perform post-imputation quality control. Monomorphic variants should be removed (MAF = 0), as well as variants that are extremely rare in the cohort (see the earlier discussion of MAF removals). IMPUTE2 provides an 'info' score related to the quality of the imputation for each variant. Different sources recommend different thresholds to exclude poorly imputed data. The selection of this threshold should be made taking into account the overall quality of the data (poor-quality data require greater quality control, and so a higher info threshold should be used). The best method is to plot a frequency curve (Figure 1) or cumulative distribution (Figure 2) of the info score and assign the threshold at the inflexion point. For example, the graphs below show most of the worst-performing variants have info<0.15, and there is an enrichment of high-quality variants with info>0.85. The threshold chosen should fall between these two. There is a relationship between MAF and info, and it is valuable to examine these metrics together—rarer variants usually show lower info scores, and often the appropriate cut-off is obvious from plotting info in MAF bins (Figure 3). In this example, a MAF cut-off of 0.01 appears to remove most of the SNPs with low info scores. Finally, it is necessary to exclude variants missing in multiple samples when using hard-called data, as variants imputed with a certainty below threshold are marked as missing rather than being excluded. Defining the threshold for completeness again benefits from plotting the data: in the example shown in Figure 4, a cut-off of 98% completeness appears to be an acceptable trade-off between retaining variants in the analysis and reducing the variation in sample size between analyses of each variant. Again, the threshold chosen should be informed by the necessary stringency of the quality control and the proposed downstream analysis.

Association analyses

The final step presented in this protocol is to perform the association analysis itself. The exact analysis performed depends on the research question being investigated and the covariates included. The flexibility of PLINK2 for running multiple statistical models and including covariates in a variety of different ways, coupled with a user-friendly implementation, arguably means it remains the first choice for performing analyses. However, many other programs exist, and it is worthwhile investigating whether a piece of software particularly suited to the planned analysis is available. The introduction of mixed linear model association analysis is an example of this, allowing for an approach to control for population structure that is as yet not available in PLINK2, although the implementation of GCTA code into PLINK2 is expected in the near future [9, 11, 23]. The development of association analysis software is an active area of research, with programs such as FaST-LMM and BOLT-LMM providing alternative implementations to GCTA [31, 32].

Conclusion

GWAS remains a valuable technique for understanding the role of genetic variants in explaining phenotypic variation, and is likely to persist as an affordable alternative as the field moves into the sequencing era. The analysis of thousands of variants allows novel findings to be made, and targets for replication to be established. Minimizing false-positive findings from GWAS will allow for more efficient use of research effort through reducing the likelihood of failed replication.

This protocol is intended as an introduction to the concepts and processes of analysing novel data from microarrays—quality control, imputation and analysis are areas of constant statistical and computational innovation, and advanced techniques that may be more appropriate for a given data set are regularly posited in the literature. We hope that the provision of this simple protocol will ensure the general standard of GWAS remains high, and will simplify the combination of independent studies into the collaborative meta-analyses that have become a hallmark of success in genomics.

Key Points

- Replication, including combining individual studies in meta-analyses is central to genomics.
- Well-executed recalling and quality control of genotype data reduces biases within GWAS studies and increases the probability of successful replication.
- Quality control, imputation and analysis of genotype data are data-driven activities.
- The protocol provided with this article provides a straightforward introduction to the basics of GWAS that will increase standardization of GWAS studies between different groups.
- Example scripts are provided at https://github.com/JoniColeman/gwas_scripts.

Acknowledgements

The authors would like to acknowledge the work of the developmental teams behind PLINK2, GCTA and IMPUTE2. In addition, publicly available scripts from Mike Weale and from Timothée Flutre are used in the protocol. We are grateful for the advice and support of the Statistical Genetics Unit at KCL, and the NIHR-BRC Bioinformatics group.

Funding

This study presents independent research part-funded by the National Institute for Health Research Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. JRIC's PhD is partly funded by the Institute of Psychiatry, Psychology and Neuroscience, and partly by the Alexander von Humboldt Foundation.

References

1. Hofker MH, Fu J, Wijmenga C. The genome revolution and its role in understanding complex diseases. *Biochim Biophys Acta* 2014;**1842**:1889–95.
2. Visscher Peter M, Brown Matthew A, McCarthy Mark I, *et al.* Five years of GWAS discovery. *Am J Hum Genet* 2012;**90**:7–24.
3. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet* 2013;**14**:415–26.
4. Anderson CA, Pettersson FH, Clarke GM, *et al.* Data quality control in genetic case-control association studies. *Nat Protoc* 2010;**5**:1564–73.
5. Weale ME. Quality control for genome-wide association studies. *Methods Mol Biol* 2010;**628**:341–72.
6. Corvin A, Craddock N, Sullivan PF. Genome-wide association studies: a primer. *Psychol Med* 2010;**40**:1063–77.
7. Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* 2010;**68**:182–6.
8. de Bakker PIW, Ferreira MAR, Jia X, *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008;**17**:R122–8.
9. Yang J, Lee SH, Goddard ME, *et al.* GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;**88**:76–82.
10. Kang HM, Sul JH, Service SK, *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 2010;**42**:348–54.
11. Yang J, Zaitlen NA, Goddard ME, *et al.* Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 2014;**46**:100–6.
12. Steemers FJ, Chang W, Lee G, *et al.* Whole-genome genotyping with the single-base extension assay. *Nat Methods* 2006;**3**:31–3.
13. Gunderson KL, Steemers FJ, Ren H, *et al.* Whole-genome genotyping. In: Kimmel A, Oliver, B (eds). *Methods in Enzymology*. San Diego, CA: Academic Press, 2006, 359–76.
14. Hansen TvO, Simonsen MK, Nielsen FC, *et al.* Collection of blood, saliva, and buccal cell samples in a pilot study on the danish nurse cohort: comparison of the response rate and quality of genomic DNA. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:2072–6.
15. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999;**22**:139–44.
16. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 2001;**20**:4–16.
17. Goldstein JI, Crenshaw A, Carey J, *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis., *Bioinformatics* 2012;**28**:2543–5.
18. Aulchenko YS, Ripke S, Isaacs A, *et al.* GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;**23**:1294–6.
19. The International HapMap C. A haplotype map of the human genome. *Nature* 2005;**437**:1299–320.
20. Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
21. Visscher PM, Andrew T, Nyholt DR. Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. *Eur J Hum Genet* 2008;**16**:387–90.
22. Hartl D, Clark HG. Principles of population genetics. Sunderland, MA: Sinauer Associates, 2006.
23. Chang C, Chow C, Tellier L, *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015;**4**:7.
24. Dadd T, Weale ME, Lewis CM. A critical evaluation of genomic control methods for genetic association studies. *Genet Epidemiol* 2009;**33**:290–8.
25. Zhang Y, Pan W. Principal component regression and linear mixed model in association analysis of structured samples: competitors or complements? *Genet Epidemiol* 2014;**39**:149–55.
26. Wagner MJ. Rare-variant genome-wide association studies: a new frontier in genetic analysis of complex traits. *Pharmacogenomics* 2013;**14**:413–24.
27. 1000 Genomes Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491**:56–65.
28. Howie B, Fuchsberger C, Stephens M, *et al.* Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012;**44**:955–9.
29. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;**5**:e1000529.
30. Aulchenko Y, Struchalin M, van Duijn C. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 2010;**11**:134.
31. Loh PR, Tucker G, Bulik-Sullivan BK, *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;**47**:484–90.
32. Lippert C, Listgarten J, Liu Y, *et al.* FaST linear mixed models for genome-wide association studies. *Nat Methods* 2011;**8**:833–5.

Appendix II:

Data supplement to Coleman et al. Genome-wide association study of response to cognitive–behavioural therapy in children with anxiety disorders. Br J Psychiatry doi: 10.1192/bjp.bp.115.168229

Supplemental material

Site information

Unless otherwise specified, clinical trials included all primary anxiety disorder diagnoses. All sites made secondary anxiety disorder diagnoses where appropriate.

Sydney, Australia

Participants aged 6-18 were recruited from the Centre for Emotional Health, Macquarie University, Sydney. All participants completed the Cool Kids program(1), with 10-12 family sessions involving the parents (the majority of which were conducted in groups; 8% of the sample's DNA were collected retrospectively). Variations on this treatment program include a subgroup from previous randomized trials who received group, individual or phone-based CBT sessions(2, 3); participants from a guided self-help trial with phone support for children in rural Australia(4); a group from a trial with additional parental anxiety management (5); and those recruited from an ongoing randomized trial of progressive allocation to treatment (Stepped Care).

Reading and Oxford, UK

Participants aged 5-18 were recruited jointly from Reading and Oxford from eight trials at the Berkshire Child Anxiety Clinic (University of Reading) and the Oxfordshire Primary Child and Adolescent Mental Health Service. Participants received treatment in three main themes; one focusing on children with anxious mothers; a set of trials using a parent-guided self-help CBT program; and an online CBT program for adolescents.

The Mother and Child (MaCh) project(6). Children whose mother also had a current anxiety disorder completed an 8 session manual-based CBT treatment based on the Cool Kids

program(7). The mothers of these children also received extra sessions focusing on their own anxiety and on mother-child interactions.

Overcoming. Children were treated with a parent-guided self-help CBT program, comprised of the same primary components as the Cool Kids program (7, 8). This consisted of 2-4 in-person sessions and 2-4 telephone sessions. A sub-set of this group with a primary anxiety disorder diagnosis of Social Phobia also received targeted Cognitive Bias Modification Training (CBM-I,(9)). Additionally, participants with highly anxious parents (screened using DASS or by meeting ADIS criteria) were randomized to groups in a trial including additional sessions for the parents which focused on strategies for tolerating children's negative emotions. In Oxford, treatment was based on the same basic program, and delivered by primary health workers as part of a feasibility trial(10).

BRAVE. The final treatment group completed a therapist-supported online CBT program for adolescents (BRAVE), consisting of 10 sessions, half with 5 additional parent sessions and half without parent sessions.

Aarhus, Denmark

Participants aged 7-17 years were recruited from the Department of Psychology and Behavioural Sciences, Aarhus University, and all anxiety disorder diagnoses were included. Participants received CBT using the Cool Kids manual (including the adolescent version where appropriate (7, 11)). Participants came from two groups; one aged 7-17, from a trial including treatment and waitlist conditions; and another group aged 7-12 from a trial comparing efficacy of traditional group-based treatment with Cool Kids versus a guided self-help version with clinician support (bibliotherapy). In both trials only participants that received in-person CBT were included.

Bergen, Norway

Participants aged 5-13 were recruited from the child part of the "Assessment and Treatment – Anxiety in Children and Adults" study, Haukeland University Hospital, Bergen. Patients referred to outpatient mental health clinics in Western Norway, with a primary diagnosis of separation anxiety, social phobia, or generalized anxiety, received group or individual treatment with the FRIENDS program (4th edition(12, 13)) in a randomized control trial comparing active treatment with a waitlist condition(14).

Bochum, Germany

Participants aged 5-18 were recruited from the Research and Treatment Centre for Mental Health, Ruhr-Universität Bochum. Participants received either exposure-based CBT (8-25 sessions, with sessions occurring at least every 2 weeks), the Coping Cat program (15), or a family-based version of CBT specifically designed to target separation anxiety disorder (TAFF (16, 17)). Diagnoses were provided separately for parent- and child-report. The primary diagnosis was selected as being the most severe from either reporter. If the most severe disorder reported by each was of equal severity but was a different diagnosis, the parent-reported diagnosis was selected.

Basel, Switzerland

Participants aged 5-13 (all with a primary diagnosis of Separation Anxiety Disorder) were recruited from the Faculty of Psychology, University of Basel. All participants took part in a randomized control trial comparing a family-based version of CBT specifically designed to target separation anxiety disorder (TAFF (16, 17)with Coping Cat(15)). All participants received 16 sessions over 12 weeks.

Groningen, The Netherlands

Participants aged 8 to 17 were recruited from the Department of Child and Adolescent Psychiatry, University of Groningen. All participants were treated within a randomized control trial of Coping Cat (Dutch version (18)) including 12 individual child sessions and 2 parent sessions.

Florida, USA

Participants aged 7 to 16 (including all primary anxiety disorder diagnoses except PTSD) were recruited from the Child Anxiety and Phobia Program, Florida International University, Miami. All participants received 12 to 14 hour-long sessions of individual manualized CBT. Additionally, two conditions included parental involvement focusing on different parent skills (Relationship Skills Training or Reinforcement Skills Training).

Cambridge, UK

Participants aged 8-17 were recruited from the MRC Cognition and Brain Sciences Unit, Cambridge, UK. Participants were taking part in the ASPECTS trial, which recruited individuals exposed to a recent (i.e. in the previous six months) traumatic stressor (i.e. any event that involve the threat of death, severe injury, or threat to bodily integrity, or witnessing such an event). Those that developed PTSD were randomized to a 10-week waitlist or individual PTSD-specific CBT(19), which consisted of up to 10 sessions over a 10 week period. Only participants that received treatment were included.

Amsterdam, The Netherlands

Participants aged 10-14 were recruited from the Academic Treatment Centre for Parent and Child, University of Amsterdam UvA Minds and received either 12 weeks of CBT in individual sessions or 8 weeks of CBT in group sessions, according to the Dutch protocol Discussing + Doing = Daring(20). Diagnoses were provided separately for parent- and child-report with the primary diagnosis selected from these data by the trial manager.

Assessment of treatment response

At all sites, an experienced diagnostician trained the independent assessors using observation, feedback and supervision, and clearly specified guidelines for allocating diagnoses and CSRs were used. Inter-site consistency between the two largest sites, Sydney and Reading/Oxford (hereafter referred to as Reading), was established through initial training of assessors at Reading using video-recorded assessments from Sydney. In addition, detailed guidance provided by the Sydney site was used in assessments at Reading throughout the study. The principal investigator at the Aarhus site (Mikael Thastum) was trained in Sydney, and assessors in Aarhus received additional training from the principal investigator at the Florida site (Wendy Silverman). As such, treatment response for participants at these four sites, which comprise 85% of the sample, was assessed with a consistent methodology. Within-site inter-rater reliability for the primary anxiety diagnosis ranged from 0.72-1.00, demonstrating that inter-rater agreement was high.

Clinical Severity Ratings across time (and number of participants assessed) by site are shown in Supplementary Table 1c. Overall, mean severity decreased from pre-treatment to post-treatment, and then roughly plateaued across the three follow-up assessments.

However, the results at each follow-up assessment are dependent on which sites performed the assessment; therefore, this should not be considered a general trajectory of treatment response. Similarly, although the mean CSR at each assessment varies between sites, the 95% confidence intervals of each mean overlap, suggesting mean CSRs do not vary significantly. The follow-up phenotype presented in this paper is imputed from this information, as described in the main text.

Non-genetic influences on treatment outcome

A diagnosis of specific phobia was associated with poorer response (percentage change in CSR score over time) and non-remission (CSR>4) at post-treatment, and a diagnosis of social phobia was associated with poorer outcome on both measures at post-treatment and at follow-up (both compared to a diagnosis of generalized anxiety disorder). Comorbid mood and externalizing disorders predicted poorer outcomes at both time-points, and parental psychopathology (self-reported anxious and depressive symptoms) interacted with time since treatment, showing little effect post-treatment but associated with poorer response at follow-up. For further information, see (21).

Sample preparation

DNA concentration was quantified before genotyping by fluorometry using PicoGreen (Invitrogen). Samples below 50ng/ul were concentrated using ultrafiltration and re-suspension. 3600ng of each sample (usually as 300ul at 12ng/ul, although this was adjusted as sample characteristics dictated) was dispensed using a customized Beckman FX robot, and then pipetted via a manual multichannel pipette into a 96-well filtration plate, which captured DNA fragments above 500bp (Multiwell 96-well PCR clean-up plate, Millipore). Samples were filtered under 750mBar of pressure until wells were dry. Following filtration, samples were re-suspended in 40ul of Tris-EDTA buffer with vigorous shaking, and DNA concentration re-quantified using spectroscopy (Nanodrop). Samples with concentration above 50ng/ul continued to genotyping on the Illumina Human Core Exome-12v1.0 microarray, which assays approximately 250 000 common SNPs and 250 000 exomic SNPs located across the genome.

Quality control

In addition to recalling of rare variants with ZCall, recalling was also performed in Opticall (22). The two methods were concordant for 99.78% of cases.

Quality control post-recalling was performed in PLINK (23) and PLINK2 (24), with reference to previously published protocols (25, 26). SNPs were excluded if the frequency of the minor allele was <5%, or if the frequencies of both alleles were out of Hardy-Weinberg equilibrium, with a threshold of $p<10^{-5}$. Samples and SNPs were excluded if call rate was <99%. Samples were excluded if phenotypic gender was inconsistent with X-chromosome homozygosity (F-statistic), if genome-wide heterozygosity was >3 standard deviations from the sample mean, if more than 18.75% of variants were shared by descent (pi-hat) between two samples, or if the average pi-hat of the sample differed from the mean by >6 standard deviations (Supplementary Figure 1). Reported sample gender was compared with X chromosome heterozygosity calculated from genotypes. Male samples are expected to be homozygous for X chromosome SNPs, while females are expected to be heterozygous – the standard PLINK thresholds of >0.8 and <0.2 respectively were used as guidance. Two samples were just outside these thresholds, but were retained as their phenotypic gender matched that suggested by the genotypes.

Principal component analysis (PCA) was performed in EIGENSTRAT (27, 28) on the dataset, pruned for linkage disequilibrium (25). Specifically, SNPs were compared pairwise in windows of 1500 SNPs, and one of each pair removed if $R^2 > 0.2$, and the procedure repeated after a shift of 150 SNPs (23). Initially, PCA was performed with the intention of using principal components to control for population stratification within the dataset. However, the use of quantitative phenotypes from which site differences had been regressed, combined with the fact that participants were recruited from across the globe, prevented the use of principal components for this purpose. The top 100 principal components were not associated with either phenotype beyond a level expected by chance. However, the principal components capture the different ethnicities in the sample, confirming participant self-reported ancestry. The majority (92.4%) of the sample are of White Western European descent (Supplementary Figure 2a, 2b; Supplementary Table 1). The recent development of software to perform mixed linear model association analyses in

genome-wide data provided a better alternative to control for background genetic similarity between individuals (29).

Association analyses were performed on phenotypes indicative of sample quality (sample concentration at entry into genotyping, and whether the sample was collected as a buccal swab or as saliva) as a quality control step. QQ plots were generated using R (script adapted from M. Weale, available at http://sites.google.com/site/mikeweale) and lambda-median values calculated to assess inflation. SNPs showing a lower *p*-value than expected under the null (those below thresholds $p<0.01$ and $p<0.001$, respectively) for either sample quality phenotype were excluded from the final analysis.

Statistical analysis

GWAS was performed using mixed linear model association analysis (MLMA), which derives a genomic relationship matrix (GRM) from genome-wide genotype data, and uses it to model the overall genetic contribution to phenotypic correlation between participants as a random effect. The *mlma-loco* option in GCTA was used to perform a leave-one-chromosome-out marker-excluded analysis on the autosomes, in which the GRM was produced excluding variants on the same chromosome at the SNP being tested. This prevents any effect of the variant of interest being partly captured by the GRM (which would reduce the measured effect of the variant). X-chromosome SNPs were assessed using the *mlma* option and a GRM produced from all autosomes. The X chromosome results were then merged with the autosomal data.

The ability of the GWAS to replicate previous findings was explored. Variants previously implicated in CBT response in mood disorders were examined, as well as further variants in *HTR2A* that have been linked to anxiety disorders more generally (see Table 2). Fourteen SNPs were identified, of which nine passed quality control in the GWAS, none of which was nominally associated with either phenotype (all $p>0.05$). Other variants, such as VNTRs in *SLC6A4* (STin2) and *MAOA* cannot be captured by GWAS. This is also true of the *SLC6A4* 5HTTLPR, which was explored elsewhere (30). In addition to individual assessment, the effect of the SNPs as a set in a linear regression in PLINK was examined. This regression used the same phenotypes and covariates as the main GWAS analyses, but used 10 PCs to control

for further confounds. The effect of the set was not significant (p=1). However, population stratification was not controlled for in this analysis, as it is not currently possible to include a set-based test in the MLMA-GWAS, so it is possible the results of the set-based test were population-confounded.

The GRM produced in the main analysis from all autosomes was used to perform univariate genomic-relatedness-matrix restricted maximum likelihood (GREML) estimation. GREML estimates the heritability captured by the SNPs investigated within the study; this is a fraction of the total heritability in the phenotype, as genotyping will not capture the full effect of variants in imperfect linkage disequilibrium with genotyped SNPs (31). GREML was performed with iterative inclusion of zero to twenty principal components.

Polygenic risk score profiling (implemented in PRSice (32)) was used to investigate the predictive power of the dataset. For each dataset, SNP positions were converted to hg19 where necessary and SNPs not present in the GxT GWAS discarded. The remaining SNPs were clumped by the top p-value using PLINK, such that no SNP that remained was in linkage disequilibrium (r^2>0.1, distance <250kb) with a more significant SNP (33). Risk profiles were created in PLINK, using SNPs with external GWAS p ranging from 0.0001 to 0.5, in increments of 0.00005. Risk was weighted by multiplying risk allele number by beta or log(OR), depending on the dataset. The proportion of variance (adjusted R^2) was calculated from a linear regression of score on outcome for each p-value threshold.

Leave-one-out polygenic risk score profile analyses was performed to test prediction within the dataset. In separate analyses, participants with GAD, separation anxiety disorder, social phobia and specific phobias were secondarily excluded from the data, and MLMA analysis performed on the remaining participants. Profile scores were calculated using the method described above, and the resulting profiles used to predict response in the excluded individuals. The same technique was also used to predict response in participants from Reading, using a profile derived from the participants at other sites.

Supplemental references

1. Rapee RM, Lyneham HJ, Schniering CA, Wuthrich VM, Abbott MA, Hudson JL, et al. The Cool Kids® Child and Adolescent Anxiety Program. Sydney: Centre for Emotional Health, Macquarie University; 2006.

2. Hudson JL, Rapee RM, Deveney C, Schniering CA, Lyneham HJ, Bovopoulos N. Cognitive-Behavioral Treatment Versus an Active Control for Children and Adolescents With Anxiety Disorders: A Randomized Trial. J Am Acad Child Adolesc Psychiatry. 2009;48(5):533-44.

3. Rapee RM, Abbott MJ, Lyneham HJ. Bibliotherapy for children with anxiety disorders using written materials for parents: A randomized controlled trial. J Consult Clin Psychol. 2006;74(3):436 - 44.

4. Lyneham HJ, Rapee RM. Evaluation of therapist-supported parent-implemented CBT for anxiety disorders in rural children. Behav Res Ther. 2006;44(9):1287-300.

5. Hudson JL, Newall C, Rapee RM, Lyneham HJ, Schniering CC, Wuthrich VM, et al. The Impact of Brief Parental Anxiety Management on Child Anxiety Treatment Outcomes: A Controlled Trial. Journal of Clinical Child and Adolescent Psychology. 2014;43(3):370-80.

6. Creswell C, Cruddace S, Gerry S, McIntosh A, Mollison J, Murray L, et al. Treatment of childhood anxiety disorder in the context of maternal anxiety disorder: A randomised controlled trial. Health Technol Assess.

7. Lyneham HJ, Abbott MJ, Wignall A, Rapee RM. The Cool Kids Anxiety Treatment Programme. Sydney MUARU, Macquarie University; 2003.

8. Thirlwall K, Cooper PJ, Karalus J, Voysey M, Willetts L, Creswell C. Treatment of child anxiety disorders via guided parent-delivered cognitive-behavioural therapy: randomised controlled trial. The British Journal of Psychiatry. 2013;203(6):436 - 44. Epub 2013/08/24.

9. Vassilopoulos SP, Banerjee R, Prantzalou C. Experimental modification of interpretation bias in socially anxious children: Changes in interpretation, anticipated interpersonal anxiety, and social anxiety symptoms. Behav Res Ther. 2009;47(12):1085-9.

10. Creswell C, Hentges F, Parkinson M, Sheffield P, Willetts L, Cooper P. Feasibility of guided cognitive behaviour therapy (CBT) self-help for childhood anxiety disorders in primary care. Mental Health in Family Medicine. 2010;7(1):49 - 57.

11. Rapee RM, Lyneham HJ, Schniering CA, Wuthrich VM, Abbott M, Hudson J, et al. Cool Kids "Chilled" Adolescent Anxiety Program2006.

12. Barrett PM, Farrell LJ, Ollendick TH, Dadds M. Long-term outcomes of an Australian universal prevention trial of anxiety and depression symptoms in children and youth: an evaluation of the friends program. Journal of Clinical Child and Adolescent Psychology. 2006;35(3):403-11.
13. Barrett P. Friends for Life Program–Group Leader's Workbook for Youth. Brisbane: Australian Academic Press; 2005.
14. Wergeland GJH, Fjermestad KW, Marin CE, Haugland BS-M, Bjaastad JF, Oeding K, et al. An effectiveness study of individual vs. group cognitive behavioral therapy for anxiety disorders in youth. Behaviour Research and Therapy. 2014;57(0):1-12.
15. Kendall PC. Treating anxiety disorders in children: results of a randomized clinical trial. J Consult Clin Psychol. 1994;62(1):100-10.
16. Schneider S, Blatter-Meunier J, Herren C, In-Albon T, Adornetto C, Meyer A, et al. The Efficacy of a Family-Based Cognitive-Behavioral Treatment for Separation Anxiety Disorder in Children Aged 8-13: A Randomized Comparison With a General Anxiety Program. J Consult Clin Psychol. 2013;81(5):932-40.
17. Schneider S, Lavallee K. Separation Anxiety Disorder. In: Essau CA, Ollendick T, editors. The Wiley-Blackwell Handbook of The Treatment of Childhood and Adolescent Anxiety Chichester: Wiley-Blackwell; 2013. p. 301-34.
18. Nauta M, Scholing A. Cognitieve gedragstherapie bij kinderen en jongeren met angststoornissen: een protocol van 12 sessies. Handleiding voor de therapeut. Groningen: Rijksuniversiteit Groningen (Klinische en Ontwikkelingspsychologie); 1998.
19. Smith P, Yule W, Perrin S, Tranah T, Dalgleish T, M. CD. Cognitive-Behavioral Therapy for PTSD in Children and Adolescents: A Preliminary Randomized Controlled Trial. J Am Acad Child Adolesc Psychiatry. 2007;46(8):1051-61.
20. Bögels SM. Behandeling van angststoornissen bij kinderen en adolescenten: Springer; 2008.
21. Hudson JL, Keers R, Roberts S, Coleman JRI, Breen G, Arendt K, et al. Clinical Predictors of Response to Cognitive-Behavioral Therapy in Pediatric Anxiety Disorders: The Genes for Treatment (GxT) Study. J Am Acad Child Adolesc Psychiatry. 2015.
22. Shah TS, Liu JZ, Floyd JAB, Morris JA, Wirth N, Barrett JC, et al. optiCall: A robust genotype-calling algorithm for rare, low frequency and common variants. Bioinformatics. 2012.

23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007;81(3):559-75.
24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. arXiv preprint arXiv:14104803. 2014.
25. Weale ME. Quality control for genome-wide association studies. Methods in Molecular Biology. 2010;628:341-72.
26. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature Protocols. 2010;5(9):1564-73. Epub 2010/11/19.
27. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLoS Genet. 2006;2:e190.
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904-9.
29. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. 2014;46(2):100-6.
30. Lester KJ, Roberts S, Keers R, Jonathan R.I. Coleman, Gerome Breen, Chloe Wong, et al. The 5HTTLPR and response to psychological therapy – an update. Br J Psychiatry. In Press.
31. Visscher PM, Yang J, Goddard ME. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). Twin research and human genetics : the official journal of the International Society for Twin Studies. 2010;13(6):517-24. Epub 2010/12/15.
32. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31(9):1466-8. Epub 2015/01/01.
33. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748-52.

Table DS1(a) Demographic details for the 980 participants included in the follow-up GWAS

Site	**N**	**% Female**	**Mean Age (95% CI)**	**White Western European ancestry (N, %)**
Reading	229	55.02	9.57 (6.02-13.12)	208 (91%)
Sydney	467	53.10	9.42 (5.56-13.28)	435 (93%)
Oxford	14	57.14	9.21 (6.37-12.06)	14 (100%)
Florida	25	48.00	9.24 (4.95-13.53)	13 (52%)
Aarhus	96	59.38	11.12 (5.98-16.27)	93 (97%)
Amsterdam	3	0.00	12.67 (9.61-15.72)	3 (100%)
Groningen	25	56.00	11.64 (5.62-17.66)	24 (96%)
Bochum	37	56.76	11.22 (5.72-16.72)	34 (92%)
Basel	38	52.63	8.42 (4.19-12.65)	38 (100%)
Bergen	36	61.11	11.44 (7.38-15.51)	35 (97%)
Cambridge	10	70.00	13.4 (8.79-18.01)	10 (100%)
Total	980	54.69	9.82 (5.39-14.25)	906 (92%)

Table DS1(b) Treatment and diagnosis of the 980 participants included in the follow-up GWAS

Site	Treatment			Primary Anxiety Diagnosis				
	Individual CBT	Group CBT	Guided Self-Help	SAD	Social Phobia	Specific Phobia	GAD	Other Anxiety Disorder
Reading	103	0	126	57	48	40	67	17
Sydney	24	382	61	64	92	31	247	33
Oxford	0	0	14	5	6	1	1	1
Florida	25	0	0	9	5	3	6	2
Aarhus	1	95	0	25	13	16	27	15
Amsterdam	1	2	0	1	1	1	0	0
Groningen	25	0	0	5	11	3	4	2
Bochum	37	0	0	9	11	13	3	0
Basel	38	0	0	38	0	0	0	1
Bergen	20	16	0	11	16	0	9	0
Cambridge	10	0	0	0	0	0	0	10
Total	284	495	201	224	203	108	364	81

Table DS1(c) Mean Clinical Severity Rating and 95% confidence intervals for the participants split by site and assessment

Site	Severity by assessment									
	Pre		Post		3 months		Six months		12 months	
	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
Reading	5.64 (4.07-7.21)	229	2.69 (-2.05-7.44)	227	-	-	1.90 (-2.65-6.45)	143	2.11 (-2.70-6.91)	76
Sydney	6.33 (4.57-8.09)	467	3.21 (-0.33-6.75)	432	2.85 (-1.54-7.25)	41	2.78 (-0.63-6.19)	324	2.76 (-1.29-6.81)	46
Oxford	5.64 (3.79-7.50)	14	2.36 (-2.64-7.36)	14	-	-	0.00 (0.00-0.00)	2	-	-
Florida	6.84 (4.34-9.34)	25	2.72 (-0.84-6.27)	25	-	-	-	-	5.50 (2.04-8.96)	4
Aarhus	6.45 (3.97-8.93)	96	2.71 (-2.64-8.06)	96	1.97 (-3.19-7.14)	92	-	-	1.40 (1.07-1.72)	7
Amsterdam	5.00 (3.00-7.00)	3	5.00 (-3.72-13.72)	3	-	-	-	-	-	-
Groningen	6.24 (4.48-8.00)	25	2.75 (-0.37-5.87)	25	0.43 (-2.51-3.38)	23	-	-	-	-
Bochum	6.86 (4.65-9.08)	37	2.00 (-2.40-6.40)	34	1.63 (1.33-1.93)	17	1.57 (-2.63-5.78)	14	1.52 (1.23-1.81)	21
Basel	5.92 (4.42-7.42)	38	2.18 (-0.37-4.73)	38	-	-	-	-	4.67 (2.36-6.98)	3
Bergen	6.81 (4.42-9.19)	36	4.80 (0.25-9.35)	35	-	-	-	-	3.58 (-1.50-8.65)	33
Cambridge	6.40 (4.05-8.75)	10	2.24 (-0.41-4.89)	10	-	-	-	-	-	-
Total	6.20 (4.20-8.20)	980	2.96 (-1.28-7.20)	939	1.94 (-2.72-6.61)	173	2.47 (-1.43-6.37)	483	2.54 (-1.98-7.07)	190

Addendum: Discussion of inter-site variability in response to CBT

As can be seen in Table DS1(c), although a general pattern of response to CBT can be observed (decreasing across treatment, and then roughly plateauing across follow-up, as described under "Assessment of treatment response" above), there is considerable inter-site variability. There are multiple potential reasons why this variability may exist, as there are a number of differences between sites. The type of treatment delivered, and the specific anxiety disorder of the participants varied between sites (Table DS1(b)). Although all sites used manualised forms of CBT, therapy was delivered by site-specific therapists, confounding therapist-specific influences with site. The number of participants varied considerably between sites, and as such the average severity at smaller sites is likely to be more affected by participants who responded particularly well or poorly (Table DS1(c)). Within the GWAS analysis contained in Chapter 3, variables for treatment type, disorder, and trial (which is nested within site) were included as covariates in order to mitigate against this variability in treatment response.

Table DS2 Clumps with association *p*-value < $1x10^{-4}$ in the GWAS, extending Tables 1 and 2

a) Independent clumps associated with CBT response post-treatment with *p*<$1x10^{-4}$						
Sentinel SNP	**CHR**	**Clump BP**	**Sentinel SNP *p***	**Sentinel SNP MAF**	**Sentinel SNP Info**	**Genes +/- 100kb**
rs10881475	1	108113663-108203647	$2.45x10^{-6}$	0.187	0.993	NTNG1, VAV3
rs11834041	12	128232821-128239057	$3.50x10^{-6}$	0.135	Genotyped	-
rs12464559	2	152498699-152679462	$4.09x10^{-6}$	0.0410	0.941	NEB, ARL5A, CACNB4
rs881301	8	38322346-38332318	$4.46x10^{-6}$	0.403	Genotyped	WHSC1L1, LETM2, FGFR1, C8orf86
rs16823934	3	115335684-115340900	$5.62x10^{-6}$	0.238	Genotyped	GAP43
rs460214	21	39962001-40059734	$6.01x10^{-6}$	0.269	0.988	ERG
rs11581859	1	99095611-99393710	$9.18x10^{-6}$	0.218	0.981	SNX7, LPPR5
rs3856211	1	166021956-166047333	$1.18x10^{-5}$	0.394	Genotyped	FAM78B
rs12188300	5	158829527-158848071	$1.61x10^{-5}$	0.0801	Genotyped	IL12B
rs2095842	1	18283857-18297688	$1.71x10^{-5}$	0.231	Genotyped	-
rs2619372	4	90710099-90779823	$2.53x10^{-5}$	0.279	0.994	SNCA, MMRN1
rs4705334	5	145822073-145904225	$2.64x10^{-5}$	0.166	Genotyped	TCERG1, GPR151, PPP2R2B

rs17106850	5	146905987-146920247	$6.02x10^{-5}$	0.169	0.998	DPYSL3, JAKMIP2
rs73127355	7	53180775-53653377	$6.04x10^{-5}$	0.0200	0.930	POM121L12
rs433156	2	77589901-77627119	$6.59x10^{-5}$	0.368	Genotyped	LRRTM4
rs35048888	2	28683174-28689459	$6.72x10^{-5}$	0.498	0.992	FOSL2, PLB1
rs148631369	2	128804780-128929492	$7.06x10^{-5}$	0.0110	0.927	SAP130, UGGT1, HS6ST1
rs6900853	6	71618855-71729332	$8.14x10^{-5}$	0.306	Genotyped	SMAP1, B3GAT2
rs35884480	6	46519020-46632594	$8.49x10^{-5}$	0.0587	Genotyped	RCAN2, CYP39A1 , SLC25A27, TDRD6, PLA2G7, ANKRD66
rs143836403	15	48728634-48941542	$8.66x10^{-5}$	0.0820	0.951	DUT, FBN1, CEP152
rs4766728	12	114711649-114725149	$8.88x10^{-5}$	0.152	0.988	TBX5
rs7734294	5	36689181-36768602	$9.01x10^{-5}$	0.197	Genotyped	SLC1A3
rs1336336	9	26759980-26918113	$9.17x10^{-5}$	0.474	Genotyped	CAAP1, PLAA, IFT74, LRRC19
rs6536613	4	162668979-162729203	$9.47x10^{-5}$	0.0230	0.931	FSTL5
rs12410507	1	60899849-61041875	$9.72x10^{-5}$	0.177	0.978	-
rs59085393	1	156374432-156390617	$9.88x10^{-5}$	0.0390	0.949	CCT3, RHBG, MEF2D

b) Independent clumps associated with CBT response at six-month follow-up with $p<1x10^{-4}$						
Sentinel SNP	**CHR**	**Clump BP**	**Sentinel SNP *p***	**Sentinel SNP MAF**	**Sentinel SNP Info**	**Genes +/- 100kb**
rs72711240	4	135657189-135695807	$4.49x10^{-7}$	0.0269	0.903	-
rs9875578	3	13707416 - 13810670	$1.43x10^{-6}$	0.424	0.994	FBLN2, WNT7A
rs6813264	4	146509970-146631854	$4.68x10^{-6}$	0.410	Genotyped	SMAD1, MMAA, C4orf51, ZNF827
rs12850751	X	145130635-145161195	$6.64x10^{-6}$	0.0655	0.952	-
rs13432654	2	162300286-162411997	$8.40x10^{-6}$	0.0939	Genotyped	PSMD14, TBR1, SLC4A10
rs76635837	15	53613961-53636281	$1.00x10^{-5}$	0.0376	0.956	-
rs1795708	12	58750680-58836631	$1.04x10^{-5}$	0.344	Genotyped	-
rs7257625	19	46468703-46474428	$1.05x10^{-5}$	0.189	Genotyped	FOXA3, IRF2BP1, MYPOP, NANOS2, NOVA2, CCDC61, PGLYRP1, IGFL4
rs17025778	2	98637504-98701594	$1.23x10^{-5}$	0.0821	Genotyped	TMEM131, VWA3B
rs56090036	15	99052579-99054173	$1.65x10^{-5}$	0.0457	0.931	FAM169B
rs111589871	8	89764480-90195838	$1.87x10^{-5}$	0.0459	0.955	-

rs73060838	3	37982687-38221526	$2.18x10^{-5}$	0.0487	0.970	CTDSPL, VILL, PLCD1, DLEC1, ACAA1, MYD88, OXSR1, SLC22A13
rs11949603	5	36361696-36383780	$2.67x10^{-5}$	0.307	0.994	RANBP3L
rs7766941	6	54310901-54702870	$2.70x10^{-5}$	0.339	0.991	T1NAG, FAM83B
rs6133736	20	9627908-9726640	$2.79x10^{-5}$	0.133	0.968	PAK7
rs55776604	17	73362147-73411596	$3.11x10^{-5}$	0.0532	0.965	MRPS7, MIF4GD, SLC25A19, GRB2, KIAA0195, CASKIN2
rs10484917	6	142038521-142110406	$3.14x10^{-5}$	0.122	0.978	-
rs61470941	2	136393157-136747085	$3.24x10^{-5}$	0.0958	0.984	R3HDM1, UBXN4, LCT, MCM6, DARS
rs11784693	8	11527910-11832769	$3.40x10^{-5}$	0.291	Genotyped	GATA4, NEIL2, FDFT1, CTSB, DEFB136, DEFB135, DEFB134, DEFB130
rs13163544	5	174069668-174126415	$3.44x10^{-5}$	0.426	Genotyped	MSX2
rs9472259	6	44291641-	$3.50x10^{-5}$	0.327	0.989	SLC29A1,

		44355423				HSP90AB1, SLC35B2, NFKBIE, TMEM151B, TCTE1, AARS2, SPATS1, CDC5L
rs6971364	7	8417400-8453313	$3.69x10^{-5}$	0.438	0.993	NXPH1
rs2690112	6	25288549-25328790	$3.81x10^{-5}$	0.372	0.985	LRRC16A
rs1486171	7	46172701-46211646	$3.97x10^{-5}$	0.392	0.996	-
rs6804426	3	151676820-151780935	$4.00x10^{-5}$	0.224	0.988	SUCNR1
rs13237987	7	9842272-9875208	$4.83x10^{-5}$	0.278	0.994	-
rs4686487	3	188341678	$5.03x10^{-5}$	0.199	Genotyped	LPP
rs114726046	6	24058226-24083141	$5.16x10^{-5}$	0.0130	0.819	NRSN1, DCDC2
rs11155986	6	154875787-154953972	$5.21x10^{-5}$	0.244	Genotyped	CNKSR3
rs4770433	13	23892555-23916736	$5.27x10^{-5}$	0.439	Genotyped	SGCG, SACS
rs12855797	X	10723386	$5.28x10^{-5}$	0.125	Genotyped	MID1
rs7131178	11	93322831-93473333	$5.46x10^{-5}$	0.181	Genotyped	SMCO4, CP295, TAF1D, c11orf54, MED17, VSTM5
rs202245865	6	132282553-132336972	$6.03x10^{-5}$	0.00980	0.828	ENPP1, CTGF
rs7784698	7	98253847-	$6.17x10^{-5}$	0.0608	0.993	NPTX2

		98311136				
rs56118623	21	19063114-19085866	$6.21x10^{-5}$	0.0906	0.946	CXADR, BTG3, c21orf91
rs12985380	19	51850290-51869346	$6.91x10^{-5}$	0.475	Genotyped	SIGLECL1, IGLON5, VSIG10L, ETFB, CLDND2, NKG7, LI2, c19orf84, SIGLEC10, SIGLEC8
rs4417554	16	27028555-27034201	$6.97x10^{-5}$	0.417	0.983	c16orf82
rs875104	13	97981705-98028784	$7.04x10^{-5}$	0.115	0.980	MBNL2, RAP2A
rs1279690	1	81066500-81154515	$7.13x10^{-5}$	0.300	Genotyped	-
rs115613292	4	43199190-43330931	$7.40x10^{-5}$	0.170	0.979	-
rs6453323	5	76726202-76877496	$7.42x10^{-5}$	0.364	Genotyped	PDE8B, WDR41, OTP
rs8047148	16	22255898-22377003	$7.45x10^{-5}$	0.225	Genotyped	VWA3A, EEF2K, POLR3E, CDR2
rs321505	6	64381461-64741820	$7.91x10^{-5}$	0.407	0.996	PTP4A1, PHF3, EYS
rs9393387	6	23274466-23320458	$8.11x10^{-5}$	0.497	Genotyped	-
rs17289116	9	32454368-32546117	$8.33x10^{-5}$	0.206	0.977	ACO1, DDX58, TOPORS, NDUFB6
rs6862501	5	12611030-12778499	$8.72x10^{-5}$	0.155	0.973	-

rs2343115	4	109070672-109111726	$8.99x10^{-5}$	0.462	Genotyped	LEF1
rs6608068	X	122425522-122503729	$9.08x10^{-5}$	0.184	Genotyped	GRIA3
rs75403290	5	175607631-175839232	$9.33x10^{-5}$	0.0203	0.910	FAM153B, SIMC1, KIAA1191, ARL10, NOP16, CLTB, FAF2
rs62312236	4	108955150-109017528	$9.58x10^{-5}$	0.0594	0.984	CYP2U1, HADH, LEF1
rs26571	5	111189290-111668828	$9.70x10^{-5}$	0.0428	0.958	NREP, EPB41L4A

Table DS3 Clumps with association *p*-value < $1x10^{-4}$ in the GWAS limited to White Western European individuals

a) Independent clumps associated with CBT response at post-treatment with p <10^{-4}						
Sentinel SNP	**CHR**	**Clump BP**	**Sentinel SNP**			**Genes +/- 100kb**
			P	**MAF**	**Info**	
rs11581859	1	99095611-99310566	$2.26x10^{-6}$	0.225	0.981	SNX7 LPPR5
rs10881475	1	108181596-108181947	$4.52x10^{-6}$	0.182	0.993	NTNG1 VAV3
rs460214	21	39975924-40008668	$4.56x10^{-6}$	0.261	0.988	ERG
rs7138026	12	128232821-128239057	$6.40x10^{-6}$	0.122	0.936	-
rs16823934	3	115335684	$8.38x10^{-6}$	0.24	Genotyped	GAP43
rs12188300	5	158829527	$9.31x10^{-6}$	0.0844	Genotyped	IL12B
rs688067	X	151284910-151313926	$9.71x10^{-6}$	0.148	0.975	MAGEA5 MAGEA10 GABRA3
rs142445243	3	861255-873247	$1.12x10^{-5}$	0.334	0.988	-
rs35048888	2	28684316-28689459	$1.39x10^{-5}$	0.493	0.992	FOSL2 PLB1
rs78885728	11	34970164-35015437	$2.08x10^{-5}$	0.0738	0.969	APIP PDHX
rs10777556	12	94309145-94316320	$2.16x10^{-5}$	0.0519	Genotyped	CRADD
rs34141319	9	139147174	$2.89x10^{-5}$	0.144	Genotyped	LHX3 QSOX2 GPSM1
rs12464559	2	152597660-152632574	$3.02x10^{-5}$	0.0392	0.941	NEB ARL5A CACNB4
rs881301	8	38332249-38332318	$3.15x10^{-5}$	0.407	Genotyped	WHSC1L1 LETM2

						FGFR1 C8orf86
rs433156	2	77627119	$3.25x10^{-5}$	0.366	Genotyped	LRRTM4
rs11636318	15	81625385- 81637284	$3.73x10^{-5}$	0.237	Genotyped	IL16 STARD5 TMC3
rs245607	5	162117403- 162156149	$4.03x10^{-5}$	0.338	0.999	-
rs2093933	1	109721684- 109723188	$4.19x10^{-5}$	0.221	Genotyped	TMEM167BS CARNA2 C1orf194 KIAA1324 SARS CELSR2 PSRC
rs11770698	7	90349848- 90364067	$4.32x10^{-5}$	0.384	0.987	CDK14
rs2095842	1	18297407- 18297688	$4.35x10^{-5}$	0.239	Genotyped	-
rs2506818	X	33809508- 33869539	$4.57x10^{-5}$	0.199	0.975	-
rs12785983	11	92741266- 92742731	$4.63x10^{-5}$	0.301	Genotyped	MTNR1B
rs34580908	11	93233511- 93249941	$5.00x10^{-5}$	0.163	0.987	CCDC67 SMCO4
rs2619372	4	90724869- 90740878	$5.88x10^{-5}$	0.269	0.994	SNCA MMRN1
rs6433860	2	181623822- 181626750	$5.90x10^{-5}$	0.289	0.943	-
rs9983768	21	30616480	$6.83x10^{-5}$	0.0613	Genotyped	MAP3K7CLB ACH1
rs1529692	5	145822515- 145841466	$7.10x10^{-5}$	0.164	0.954	TCERG1 GPR151
rs73127355	7	53421770- 53466859	$7.14x10^{-5}$	0.0202	0.930	-
rs4939881	18	47161733	$7.56x10^{-5}$	0.413	0.985	LIPG
rs17106850	5	146906766	$7.73x10^{-5}$	0.165	Genotyped	DPYSL3

						JAKMIP2
rs1556113	10	19116610-19169229	7.81×10^{-5}	0.386	0.993	-
rs3892710	6	32682862	8.09×10^{-5}	0.170	Genotyped	HLA-DQA1 HLA-DQB1 HLA-DQA2 HLA-DQB2 HLA-DOB
rs4591151	16	72354029	8.54×10^{-5}	0.0112	0.858	-
rs10978931	9	110346728	8.62×10^{-5}	0.431	Genotyped	KLF4
rs111988532	12	76161146-76170322	8.72×10^{-5}	0.00730	0.855	-
rs141980060	2	128737920	8.86×10^{-5}	0.0118	0.834	AMMECR1L SAP130
rs6465600	7	97139357	9.38×10^{-5}	0.325	Genotyped	-
rs17133411	10	4730637-4731224	9.61×10^{-5}	0.119	0.964	-
rs727675	14	31733642	9.65×10^{-5}	0.424	Genotyped	HECTD1 HEATR5A
rs9882669	3	127578497	9.75×10^{-5}	0.174	0.982	MGLL KBTBD12
rs11118645	1	221150673	9.97×10^{-5}	0.122	0.972	HLX

b) Independent clumps associated with CBT response at follow-up with p <10^{-4}						
Sentinel SNP	**CHR**	**Clump BP**	**Sentinel SNP**			**Genes +/- 100kb**
			P	**MAF**	**Info**	
rs72711240	4	135657189-135695807	$1.16x10^{-6}$	0.0286	0.903	-
rs6509245	19	46468703-46474428	$3.28x10^{-6}$	0.184	0.971	FOXA3 IRF2BP1 MYPOP NANOS2 NOVA2 CCDC61 PGLYRP1 IGFL4
rs9380865	6	39389992-39666704	$3.80x10^{-6}$	0.0303	0.909	
rs9875578	3	13689452-13810670	$4.43x10^{-6}$	0.419	0.994	FBLN2 WNT7A
rs7169126	15	53613961-53636281	$6.32x10^{-6}$	0.0368	0.960	-
rs13432654	2	162300286-162411997	$7.25x10^{-6}$	0.0966	Genotyped	PSMD14 TBR1 SLC4A10
rs9472259	6	44291641-44355423	$1.97x10^{-5}$	0.323	0.989	SLC29A1 HSP90AB1 SLC35B2 NFKBIE TMEM151B TCTE1 AARS2 SPATS1 CDC5L
rs6813264	4	146524560-146631854	$2.13x10^{-5}$	0.401	Genotyped	SMAD1 MMAA C4orf51 ZNF827
rs9393387	6	23274466-23320458	$2.26x10^{-5}$	0.489	Genotyped	-

rs1279690	1	81066500-81154515	$2.41x10^{-5}$	0.289	Genotyped	-
rs4770433	13	23892555-23916736	$2.61x10^{-5}$	0.444	Genotyped	SGCG SACS
rs7766941	6	54310901-54721617	$2.61x10^{-5}$	0.334	0.991	T1NAG FAM83D
rs61755441	3	37982687-38221526	$3.44x10^{-5}$	0.0497	Genotyped	CTDSPL VILL PLCD1 DLEC1 ACAA1 MYD88 OXSR1 SLC22A13
rs12836210	X	145130635-145161195	$3.51x10^{-5}$	0.0587	0.971	-
rs9956331	18	62581797	$4.13x10^{-5}$	0.493	Genotyped	-
rs6608068	X	122425522-122503729	$4.50x10^{-5}$	0.178	Genotyped	GRIA3
rs6804426	3	151676820-151780935	$4.83x10^{-5}$	0.225	0.988	SUCNR1
rs17025778	2	98637504-98701594	$4.96x10^{-5}$	0.0828	Genotyped	TMEM131 VWA3B
rs4686487	3	188341678	$5.42x10^{-5}$	0.194	Genotyped	LPP
rs6453323	5	76717417-76877496	$5.81x10^{-5}$	0.356	Genotyped	PDE8B WDR41 OTP
rs10484917	6	142038521-142110406	$6.02x10^{-5}$	0.119	0.978	-
rs111589871	8	89764480-90194404	$6.08x10^{-5}$	0.0496	0.955	-
rs3213871	2	136393157-136747085	$6.28x10^{-5}$	0.0977	Genotyped	R3HDM1 UBXN4 LCT MCM6 DARS
rs1795708	12	58750680-	$6.50x10^{-5}$	0.343	Genotyped	-

		58836631				
rs114726046	6	24021811-24083141	$6.92x10^{-5}$	0.0134	0.819	NRSN1 DCDC2
rs11155986	6	154875787-154953972	$7.22x10^{-5}$	0.247	Genotyped	CNKSR3
rs7784698	7	98253847-98311136	$7.54x10^{-5}$	0.0546	0.993	NPTX2
rs74439728	6	39670986-39698821	$7.81x10^{-5}$	0.0333	0.954	KIF6 DAAM2
rs6971364	7	8417400-8453313	$7.82x10^{-5}$	0.444	0.993	NXPH1
rs7942333	11	22389762-22525721	$7.97x10^{-5}$	0.265	0.981	ANO5 SLC17A6
rs56090036	15	99052579-99054173	$8.07x10^{-5}$	0.046	0.931	FAM169B
rs12985380	19	51850290-51869346	$8.19x10^{-5}$	0.461	Genotyped	SIGLECL1 IGLON5 VSIG10L ETFB CLDND2 NKG7 LI2 c19orf84 SIGLEC10 SIGLEC8
rs7131178	11	93322831-93473333	$8.27x10^{-5}$	0.177	Genotyped	SMCO4 CP295 TAF1D c11orf54 MED17 VSTM5
rs875104	13	97981705-98028784	$8.34x10^{-5}$	0.112	0.980	MBNL2 RAP2A
rs4527055	17	64210757-64331957	$8.64x10^{-5}$	0.0854	0.966	CEP112 APOH PRKCA

rs2343115	4	109017528-109111726	$8.66x10^{-5}$	0.479	Genotyped	HADH LEF1
rs56118623	21	19063114-19085866	$8.69x10^{-5}$	0.0886	0.946	CXADR BTG3 c21orf91
rs6817483	4	135454396-135648802	$8.76x10^{-5}$	0.190	0.978	-
rs8064192	16	55164542-55184874	$9.15x10^{-5}$	0.480	Genotyped	-
rs2334201	16	87562882-87608253	$9.22x10^{-5}$	0.404	Genotyped	ZCCHC14 JPH3
rs871644	1	18283857-18297688	$9.51x10^{-5}$	0.238	0.990	-
rs77413226	18	74515796-74653603	$9.91x10^{-5}$	0.0112	0.895	ZNF236 MBP

Fig. DS1 Exclusion of samples (top) and single nucleotide polymorphisms (bottom).

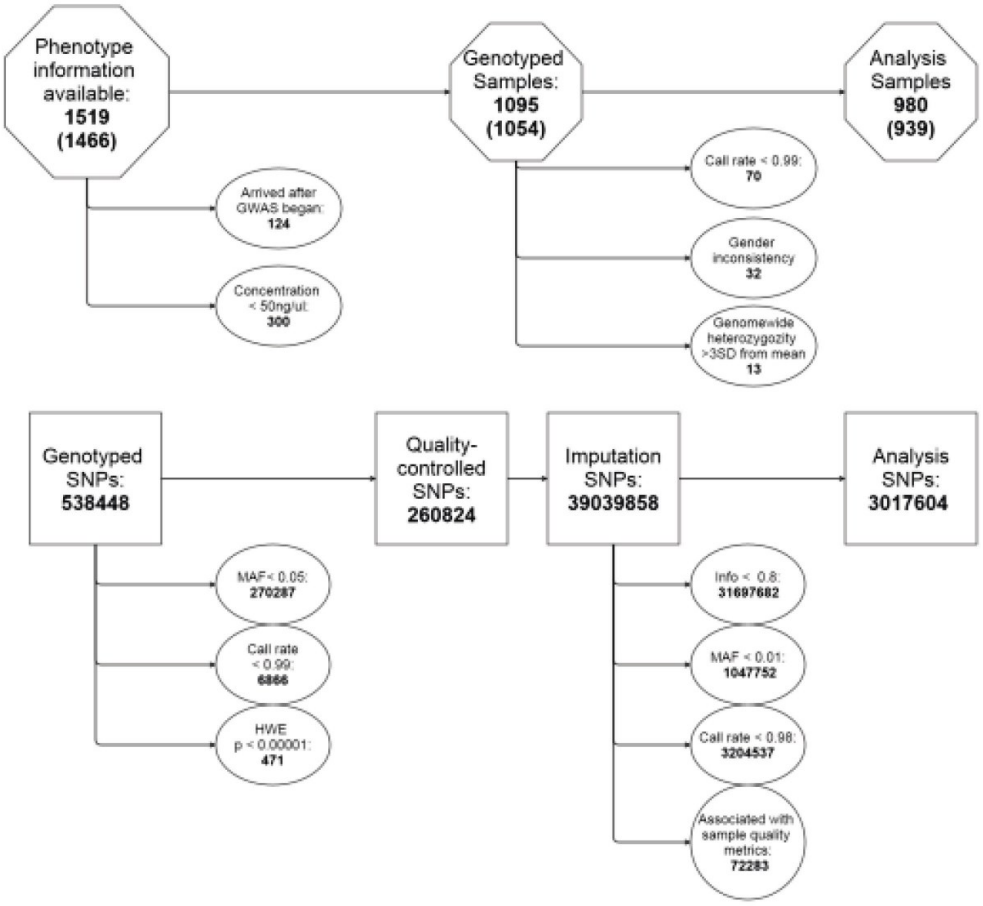


Fig. DS2(a) Samples projected on the first two principal components derived from the study samples.

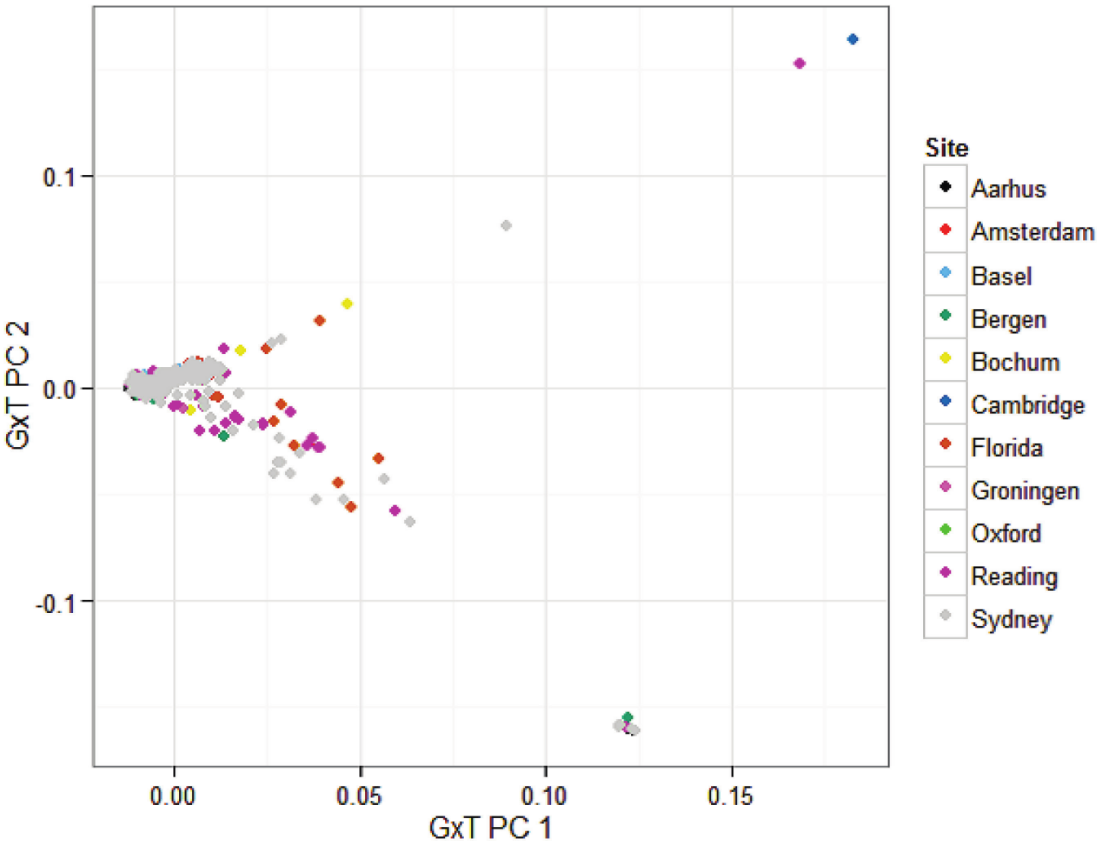
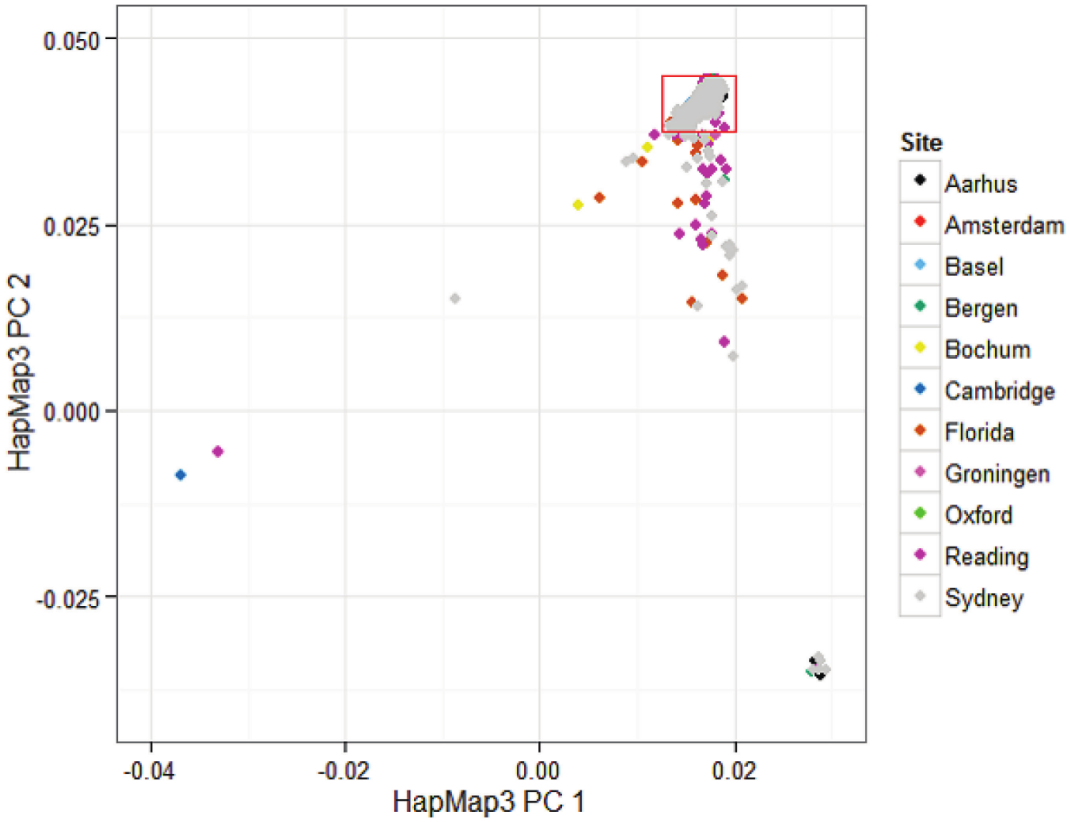


Fig. DS2(b) Samples projected on the first two principal components derived from the HapMap3 samples, showing that the majority cluster in a White Western European group (red box), with admixed samples descending down to East Asian ancestry (right), and to African ancestry (left).



Appendix III: Supplementary information from Chapter 4: Separate and combined effects of genetic variants and pre-treatment whole blood gene expression on response to exposure-based cognitive behavioural therapy for anxiety disorders.

III.I Supplementary Methods

III.I.I. Construction of the co-expression matrix in WGCNA

Data-driven clustering of co-expressed probes was performed using an automatically-constructed signed network from the *blockwiseModules* function in WGCNA. An unsigned topological overlap matrix and a signed network were specified to obtain expression modules with a shared direction of effect within the modules. The adjacency function for the network used a soft thresholding power of 13, obtained from the network topology analysis function (*pickSoftThreshold*), which identifies the lowest power at which the fit of the network to scale-free topology has $R^2>0.9$. A dendrogram of probe relationships was constructed using average linkage hierarchical clustering, and modules defined via a dynamic hybrid tree cutting approach (Langfelder and Horvath 2008). The cutting threshold for module definition was set to 0.25 and a minimum module size of 30 was specified. Genes were assigned to modules according to the significance of their correlation with the module eigengene

(that is, *pamRespectsDendro*=FALSE and *reassignThreshold*=0). All other arguments to the *blockwiseModules* function were set as default (Langfelder and Horvath 2008).

III.I.II. Enrichment analysis in GOrilla

GOrilla identifies enrichment in ranked lists of genes by calculating a minimum hypergeometric score. The ranked list is split into a target set (the first x genes) and the background set (the entire list). This is performed iteratively, adding each gene into the target set until all genes are included. The enrichment score is calculated from the optimal target set, using a hypergeometric distribution, correcting for the multiple thresholding involved in the method (Eden *et al* 2009). GOrilla reports false discovery rate q-values; however, the distribution of p-values resulting from these analyses was skewed and could not be controlled appropriately using false discovery rate. Accordingly, significance was set as the Bonferroni correction for the 8746 GO terms tested (p=5.72x10^{-6}).

Addendum to III.I.II Up-regulated and down-regulated genes

Following the publication of chapter 4, it was suggested that ranking by signed effect size (in this case, Pearson's r) may be more powerful for assessing gene-set enrichment than ranking by p-value. Although I am unconvinced of the evidence for increased power from this method (at least in the case of

continuous data), it does allow for additional exploration of gene set enrichment. Accordingly, the GOrilla analyses in Chapter 4 were re-run, ranking by Pearson's r, and calculating scores from the top of the list to the bottom (capturing up-regulated genes; Supplementary Table 5b) and bottom-to-top (capturing down-regulated genes; Supplementary Table 5c). As the Gene Ontology database had been updated since the original analyses, GOrilla analyses ranking by p-value were also re-run (Supplementary Table 5a, which differs from the published Supplementary Table 5).

Performing these analyses allows dissection of the reported gene set findings - for example, the enrichment of genes involved in apoptosis signalling in the post-treatment analyses appears to be driven more specifically by the up-regulation of cysteine-type endopeptidase activity. New gene sets are also implicated, including a trend towards the down-regulation of histone acetylation in post-treatment. However, it should be noted that (with one exception), the findings of the up-regulated and down-regulated gene set enrichment analyses are only nominally significant (as were the original analyses), and that the increased number of enriched gene sets is partly a function of the increased number of tests performed. The most striking difference between the two methods of assessing gene set enrichment is that the up-regulation of immune system process genes was significantly associated

with response at follow-up, even taking into account the number of tests performed. This gene set was originally nominally associated with response at follow-up, although this effect was diminished when the analyses ranking by p-value were re-run. Although this is interesting, conclusions concerning this association should be cautious, particularly given the extensive roles played by the immune system, and the broad, exploratory nature of the analyses in Chapter 4.

III.I.III. GWAS and polygenic risk scoring analyses

No individual genetic variants were identified with genome-wide significance (Supplementary Table 3). This was expected, and is in line with a previous GWAS of CBT response (Coleman *et al* 2016). This cohort was too small to provide the necessary power to detect the anticipated small effect sizes of individual genetic variants underlying treatment response to psychiatric disorders. Associations of potential biological interest were identified near *ADCY2* and *GDNF*. The rs17826816 variant in adenylate cyclase 2 (*ADCY2*) has previously been implicated in a large GWAS of bipolar disorder (Muhleisen *et al* 2014). Glial cell-line derived neurotrophic factor (*GDNF*) is a signalling molecule, expressed in the cerebellum and involved in the promotion of dopamine uptake (Lin *et al* 1993). Although both of these regions are plausibly

involved in psychiatric phenotypes, the small sample size and lack of genome-wide significance means these associations with treatment response could be the result of chance alone. More broadly, this is supported by the lack of concordance between these results and those from our previous GWAS of CBT response (see below).

The purpose of including genetic data in this analysis was twofold: to enrich expression analyses (as discussed in the main text), and to enable polygenic risk score analysis between this cohort and our previous study of CBT response in children (Supplementary Table 4a; Coleman *et al* 2016). However, results from our previous study did not significantly predict variance in treatment response in this analysis. There are a number of potential explanations for this. Power estimation in polygenic risk scoring relies on a large number of variables, and accordingly power estimates in this analysis should be treated with caution. Assuming an underlying true heritability of response to CBT of 30%, perfect genetic correlation between the two samples, a highly polygenic model in which 95% of variants have an effect, and an alpha threshold of 0.001, estimates using polygenescore in R suggests the follow-up results from the child study have ~5% power to predict response at follow-up in the adult study (the most predictive estimate in this analysis; Dudbridge 2013; Palla and Dudbridge 2015). Therefore, we can conclude that this analysis was underpowered in general. In

addition, the two analyses are dissimilar. The participants in this paper were adults, recruited from Germany, entirely of White Western European ancestry, and undergoing treatment for panic disorder and specific phobias. The treatment featured a considerable element of exposure, and the treatment response phenotype was derived from the Clinical Global Impressions severity scale. In comparison, the children in the previous analysis were recruited from across the globe, mostly not German (although predominantly were of White Western European ancestry), and were undergoing treatment for a range of anxiety disorders, very few of which were panic disorder (and it is likely that the specific phobias the children were treated for differed considerably from those for which the adults were treated). Treatments were not primarily exposure-based, and the response phenotype was derived from the Anxiety Disorders Interview Schedule (Silverman and Albano 1996). Consequently, there are a variety of differences between the two groups which may explain the lack of prediction. We cannot exclude the possibility that no genetic component to CBT response exists; however, there are sufficient alternative reasons for the observed lack of concordance (and the absence of a genetic component would be sufficiently surprising) that we cannot conclude that no such component exists.

To reduce the difference between the two groups compared using polygenic risk scoring, a within-cohort analysis was undertaken, using the results from the dental treatment group to predict response in all other groups (Supplementary Table 4b). The results here, while non-significant, are superficially more promising, explaining a higher proportion of variance and being less likely to have occurred by chance. However, it would be wrong to conclude anything substantial from these results – the analysis is very likely to be underpowered due to the small sample sizes, and the optimisation provided by the PRSice method results in the best prediction being reported. This combination is likely to result in false positives, despite the rigorous correction for multiple testing recommended.

III.II. Supplementary References

Coleman JRI, Lester KJ, Keers R, Roberts S, Curtis C, Arendt K, Bögels S, Cooper P, Creswell C, Dalgleish T et al. . 2016. Genome-wide association study of response to cognitive behavioral therapy in children with anxiety disorders. . British Journal of Psychiatry.

Dudbridge F. 2013. Power and Predictive Accuracy of Polygenic Risk Scores. PLoS Genet 9(3):e1003348.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10(1):48.

Euesden J, Lewis CM, O'Reilly PF. 2015. PRSice: Polygenic Risk Score software. Bioinformatics 31(9):1466-8.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9(1):559.

Lin LF, Doherty DH, Lile JD, Bektesh S, Collins F. 1993. GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. Science 260(5111):1130-2.

Muhleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J, Mattheisen M, Forstner AJ, Schumacher J, Breuer R et al. . 2014. Genome-wide association study reveals two new risk loci for bipolar disorder. Nat Commun 5.

Palla L, Dudbridge F. 2015. A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. The American Journal of Human Genetics 97(2):250-259.

Silverman WK, Albano AM. 1996. The Anxiety Disorders Interview Schedule for Children-IV (Child and Parent Versions). New York Oxford University Press.

III.III. Supplementary Tables

III.III.I Supplementary Table 1

Comorbidity	WC	DA	SP	PD-CBT	PD-EXP
N	72	46	8	9	9
Panic disorder with agoraphobia	7	6	1	0	0
Panic disorder without agoraphobia	5	4	1	0	0
Agoraphobia without panic disorder	5	5	0	0	0
Specific phobia	30	13	7	5	5
Social anxiety disorder	13	6	0	3	4
Generalised anxiety disorder	1	1	0	0	0
Major depressive disorder	13	6	1	1	5
Post-traumatic stress disorder	3	3	0	0	0
Substance abuse	7	7	0	0	0
Hypochondriasis	5	2	0	2	1
Personality disorder NOS	3	3	0	0	0
Bulimia nervosa	2	2	0	0	0
Somatisation disorder	1	1	0	0	0
Mild intellectual disability	1	1	0	0	0
Insomnia	1	0	0	1	0

Supplementary Table 1: Comorbidities in the whole cohort, and by treatment group. Some individuals had multiple mental comorbidities, so individual comorbidities do not sum to N. NOS = not otherwise specified.

III.III.II. Supplementary Table 2

Response post-treatment	**Improved**	**No Change**	**Deteriorated**	**Test**	**Stat**	***p***
N	165	16	4	-	-	-
Age in years (Mean [SD])	38.9 [11.5]	38.9 [9.65]	43.3 [13.6]	ANOVA	0.284	0.753
Gender (N male [%])	61 [37.0]	4 [25.0]	1 [25.0]	Fisher's exact test		0.646
Baseline CGI severity (Mean [SD])	4.74 [1.11]	4.50 [1.37]	3.75 [0.500]	ANOVA	1.78	0.172
Treatment duration in days (Mean [SD])	198 [181]	157 [190]	284 [229]	ANOVA	0.836	0.435
Follow-up duration in days (Mean [SD])	213 [57.5]	235 [65.4]	262 [173]	ANOVA	1.62	0.202
Psychoactive medication at baseline (N taking [%])	15 [9.09]	4 [25.0]	1 [25.0]	Fisher's exact test		0.0742
Mental disorder comorbidities (N [%])	61 [37.0]	8 [50.0]	2 [50.0]	Fisher's exact test		0.547
Response at follow-up	**Improved**	**No Change**	**Deteriorated**	**Test**	**Stat**	***P***
N	101	11	10	-	-	-
Age in years (Mean [SD])	39.1 [12.1]	39.3 [12.7]	39.7 [13.2]	ANOVA	0.00977	0.990
Gender (N male [%])	35 [34.7]	4 [36.4]	2 [20.0]	Fisher's exact test		0.748
Baseline CGI severity (Mean [SD])	4.69 [1.00]	4.36 [1.29]	3.80 [0.422]	ANOVA	3.99	**0.0211 ***
Treatment duration in days (Mean [SD])	225 [186]	273 [244]	141 [130]	ANOVA	1.34	0.265
Follow-up duration in days (Mean [SD])	212 [50.6]	204 [57.1]	232 [115]	ANOVA	0.681	0.508
Psychoactive medication at baseline (N taking [%])	7 [6.93]	1 [9.09]	3 [30.0]	Fisher's exact test		0.0446
Mental disorder comorbidities (N [%])	34 [33.7]	3 [27.3]	8 [80.0]	Fisher's exact test		**0.0144 †**

Supplementary Table 2: Demographic information on the whole cohort, split by response to treatment group.
Post-hoc t-tests (variances assumed unequal; Bonferroni-corrected threshold $p = 0.0125$; significant differences in bold):
* Deteriorated lower: **vs improved: t=-5.37, p=2.44x10^{-5}**, vs no change: t=-1.37, p=0.194
† Higher rate in deteriorated: **vs improved: t=3.28, p= 0.00708**, vs no change: t=2.72, p=0.0136)

III.III.III. Supplementary Table 3

a) Independent clumps associated with CBT response at post-treatment with $p<5x10^{-6}$								
Sentinel SNP	**A1**	**CHR**	**Clump BP**	**Sentinel SNP**				**Genes +/- 100kb**
				Z	***p***	**MAF**	**Info**	
rs373527574	T	7	52998001 - 53008188	-5.11	$3.27x10^{-7}$	0.0225	0.927	POM121L12
rs17826816	G	5	7519298 - 7583156	-4.66	$3.14x10^{-6}$	0.262	0.996	ADCY2
rs7298068	T	12	124015832 - 124200135	-4.63	$3.68x10^{-6}$	0.0395	0.919	RILPL2, SNRNP35, RILPL1, MIR3908, TMED2, DDX55, EIF2B1, GTF2H3, TCTN2, ATP6V0A2, DNAH10
b) Independent clumps associated with CBT response at six-month follow-up with $p<5x10^{-6}$								
Sentinel SNP	**A1**	**CHR**	**Clump BP**	**Sentinel SNP**				**Genes +/- 100kb**
				Z	***p***	**MAF**	**Info**	
rs145019082	T	4	19231227 - 19480588	-4.86	$1.19x10^{-6}$	0.103	0.973	-
rs11959616	T	5	37912995 - 37915720	-4.73	$2.27x10^{-6}$	0.164	Gen.	GDNF
rs55749034	G	6	169498498-169546063	4.60	$4.20x10^{-6}$	0.374	0.992	THBS2
rs9381793	A	6	49261273-49494241	-4.60	$4.27x10^{-6}$	0.361	Gen.	MUT, CENPQ, GLYATL3, C6orf141, RHAG

Supplementary Table 3: Variants with $p<5x10^{-6}$ in the GWAS from pre-treatment to a) post-treatment and b) follow-up. Variants in linkage disequilibrium ($r^2 > 0.25$) with a more associated variant are not shown. Negative Z scores indicate worse response with each effect allele (A1). Gen. = Genotyped SNP

III.III.IV. Supplementary Tables 4a and 4b

a) Polygenic risk scores child GWAS -> Whole cohort				
Base: Child CBT	**Target: Adult CBT**	**Best threshold**	***p*-value**	**Variance explained (R^2)**
Baseline – Post-treatment	Baseline – Post-treatment	0.10445	0.329	0.00536
Baseline – Follow-up	Baseline – Post-treatment	0.0021	0.221	0.00848
Baseline – Post-treatment	Baseline – Follow-up	0.0315	0.123	0.0199
Baseline – Follow-up	Baseline – Follow-up	0.0012	0.0372	0.0364

Supplementary Table 4a: Variance explained by the most predictive polygenic risk scores from a GWAS of CBT response in children predicting response in the whole cohort (Chapter 3). No threshold passes the recommended $\alpha = 0.001$ for a single test (Euesden *et al* 2015).

b) Polygenic risk scores DA <-> other treatment groups				
Base: SP + PD-CBT + PD-EXP	**Target: DA**	**Best threshold**	***p*-value**	**Variance explained (R^2)**
Baseline – Post-treatment	Baseline – Post-treatment	$9x10^{-4}$	0.0458	0.0427
Baseline – Post-treatment	Baseline – Follow-up	0.06545	0.263	0.0240
Baseline – Follow-up	Baseline – Post-treatment	0.01155	0.0408	0.0447
Baseline – Follow-up	Baseline – Follow-up	$1.5x10^{-4}$	0.258	0.0246
Base: DA	**Target: SP + PD-CBT + PD-EXP**	**Best threshold**	***p*-value**	**Variance explained (R^2)**
Baseline – Post-treatment	Baseline – Post-treatment	0.18425	0.0523	0.0436
Baseline – Post-treatment	Baseline – Follow-up	0.01475	0.0219	0.0770
Baseline – Follow-up	Baseline – Post-treatment	0.01425	0.0266	0.0565
Baseline – Follow-up	Baseline – Follow-up	0.0196	0.0764	0.0468

Supplementary Table 4b: Variance explained by the most predictive polygenic risk scores between the DA group and all others. No threshold passes the recommended $\alpha = 0.001$ for a single test (Euesden *et al* 2015).

III.III.V. Supplementary Table 5a

GO Term	Description	*p*	Enrichment	# Genes
GO Process terms nominally associated with treatment response baseline-post-treatment				
GO:2001233	regulation of apoptotic signaling pathway	$7.23x10^{-5}$	6.69	9/34
GO:0009966	regulation of signal transduction	$8.39x10^{-5}$	1.68	62/184
GO:0045338	farnesyl diphosphate metabolic process*	$3.77x10^{-4}$	2652	1/1
GO:0072044	collecting duct development	$4.22x10^{-4}$	68	2/39
GO:1903902	positive regulation of viral life cycle	$4.98x10^{-4}$	11.84	5/35
GO Function terms nominally associated with treatment response baseline-post-treatment				
GO:0004311	farnesyltranstransferase activity*	$3.77x10^{-4}$	2652	1/1
GO Process terms nominally associated with treatment response baseline-follow-up				
GO:0010715	regulation of extracellular matrix disassembly	$7.74x10^{-5}$	156	2/17
GO:0043966	histone H3 acetylation	$7.93x10^{-5}$	25.38	4/38
GO:0002376	immune system process	$8.13x10^{-5}$	1.26	215/743
GO:0050920	regulation of chemotaxis	$8.54x10^{-5}$	2.37	19/851
GO:0048678	response to axon injury	$1.17x10^{-4}$	52	3/17
GO:0006915	apoptotic process involved in patterning of blood vessels	$1.71x10^{-4}$	106.08	2/25
GO:0006955	immune response	$2.49x10^{-4}$	1.77	48/312
GO:0050921	response to wounding	$2.51x10^{-4}$	21.83	4/18
GO:0007166	cell surface receptor signaling pathway	$3.27x10^{-4}$	1.37	122/555
GO:0060033	anatomical structure regression	$4.62x10^{-4}$	15.98	3/166
GO Function terms nominally associated with treatment response baseline-follow-up				
GO:0045295	gamma-catenin binding	$1.71x10^{-4}$	106.08	2/25
GO:0019899	enzyme binding	$3.76x10^{-4}$	3.29	13/25
GO:0022857	transmembrane transporter activity	$3.90x10^{-4}$	1.90	37/523

Supplementary Table 5a: Gene ontology terms with $p<5x10^{-4}$ in either analysis. Final column (b/n) shows the optimal number of top genes from the ranking (n) to maximise the enrichment of genes from the pathway (b). * A single gene (FDFT1) gene set - all GO terms referring to this gene set are collapsed into these exemplars.

Supplementary Table 5b

GO Term	Description	*p*	Enrichment	# Genes
GO Process terms nominally associated with treatment response baseline-post-treatment considering up-regulated genes				
GO:2001267	regulation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway	$1.02x10^{-4}$	44.33	3/30
GO:1903306	negative regulation of regulated secretory pathway	$4.51x10^{-4}$	21.45	3/93
GO:0033003	regulation of mast cell activation	$4.99x10^{-4}$	10.44	5/91
GO Process terms nominally associated with treatment response baseline-post-treatment considering down-regulated genes				
GO:0035065	regulation of histone acetylation	$1.44x10^{-4}$	3.32	10/729
GO:0036498	IRE1-mediated unfolded protein response	$1.79x10^{-4}$	2.57	16/752
GO:0045338	farnesyl diphosphate metabolic process*	$3.76x10^{-4}$	2660	1/1
GO Function terms nominally associated with treatment response baseline-post-treatment considering down-regulated genes				
GO:0002376	protein disulphide isomerase activity	$2.16x10^{-4}$	133	2/10
GO:0004311	farnesyltranstransferase activity*	$3.76x10^{-4}$	2660	1/1
GO Component terms nominally associated with treatment response baseline-post-treatment considering down-regulated genes				
GO:0005737	cytoplasm	$4.16x10^{-4}$	1.17	319/919

Supplementary Table 5b: Gene ontology terms with $p<5x10^{-4}$ in the post-treatment analysis, assessing up-regulated and down-regulated genes separately. * A single gene (FDFT1) gene set - all GO terms referring to this gene set are collapsed into these exemplars.

Supplementary Table 5c

GO Term	Description	*p*	Enrichment	# Genes
GO Process terms nominally associated with treatment response baseline-follow-up considering up-regulated genes				
GO:0002376	**immune system process**	**$1.47x10^{-7}$**	**1.47**	**153/456**
GO:0002252	immune effector process	$1.39x10^{-5}$	1.57	90/456
GO:0048583	regulation of response to stimulus	$2.06\ x10^{-5}$	1.21	298/908
GO:0006887	exocytosis	$2.34x10^{-5}$	1.42	117/818
GO:1902600	hydrogen ion transmembrane transport	$2.98x10^{-5}$	2.92	19/481
GO:0048678	response to axon injury	$3.54x10^{-5}$	73.89	3/12
GO:0006818	hydrogen transport	$4.69x10^{-5}$	2.70	21/481
GO:0009966	regulation of signal transduction	$8.11x10^{-5}$	1.24	226/908
GO:0006952	defence response	$9.35x10^{-5}$	1.65	65/435
GO:0051258	protein polymerisation	$1.02x10^{-4}$	3.32	12/601
GO:0044763	single organism cellular process	$2.48x10^{-4}$	1.16	298/457
GO:0051340	regulation of ligase activity	$4.41x10^{-4}$	66.5	2/40
GO:0006935	chemotaxis	$4.88x10^{-4}$	2.03	28/622
GO:0035455	response to interferon-alpha	$4.96x10^{-4}$	4.47	8/397
GO Function terms nominally associated with treatment response baseline-follow-up considering up-regulated genes				
GO:0005507	copper ion binding	$2.90x10^{-4}$	38.55	3/23
GO:0015078	hydrogen ion transmembrane transporter activity	$3.75x10^{-4}$	2.69	17/481
GO Component terms nominally associated with treatment response baseline-post-treatment considering up-regulated genes				
GO:0098796	membrane protein complex	$8.03x10^{-5}$	1.72	60/461
GO:0098800	inner mitochondrial membrane protein complex	$3.42x10^{-4}$	2.33	22/545
GO:0044425	membrane part	$3.49x10^{-4}$	1.41	92/186

GO Process terms nominally associated with treatment response baseline-follow-up considering down-regulated genes				
GO:0090304	nucleic acid metabolic process	$3.88x10^{-5}$	1.27	222/706
GO:0002885	positive regulation of hypersensitivity	$5.03x10^{-5}$	32.44	3/82
GO:1901360	organic cyclic compound metabolic process	$1.93x10^{-4}$	1.21	270/748
GO:1904837	beta-catenin-TCF complex assembly	$2.91x10^{-4}$	152	2/5
GO:0006396	RNA processing	$3.38x10^{-4}$	1.46	86/737
GO:0008543	fibroblast growth factor receptor signalling pathway	$3.92x10^{-4}$	18.67	4/30
GO:0006413	translational initiation	$4.59x10^{-4}$	1.90	32/711
GO Function terms nominally associated with treatment response baseline-follow-up considering down-regulated genes				
GO:0003676	nucleic acid binding	$5.71x10^{-5}$	1.24	243/711
GO:0003723	RNA binding	$8.19x10^{-5}$	1.32	163/711

Supplementary Table 5c: Gene ontology terms with $p<5x10^{-4}$ in the follow-up analysis, assessing up-regulated and down-regulated genes separately.

III.III.VI. Supplementary Table 6

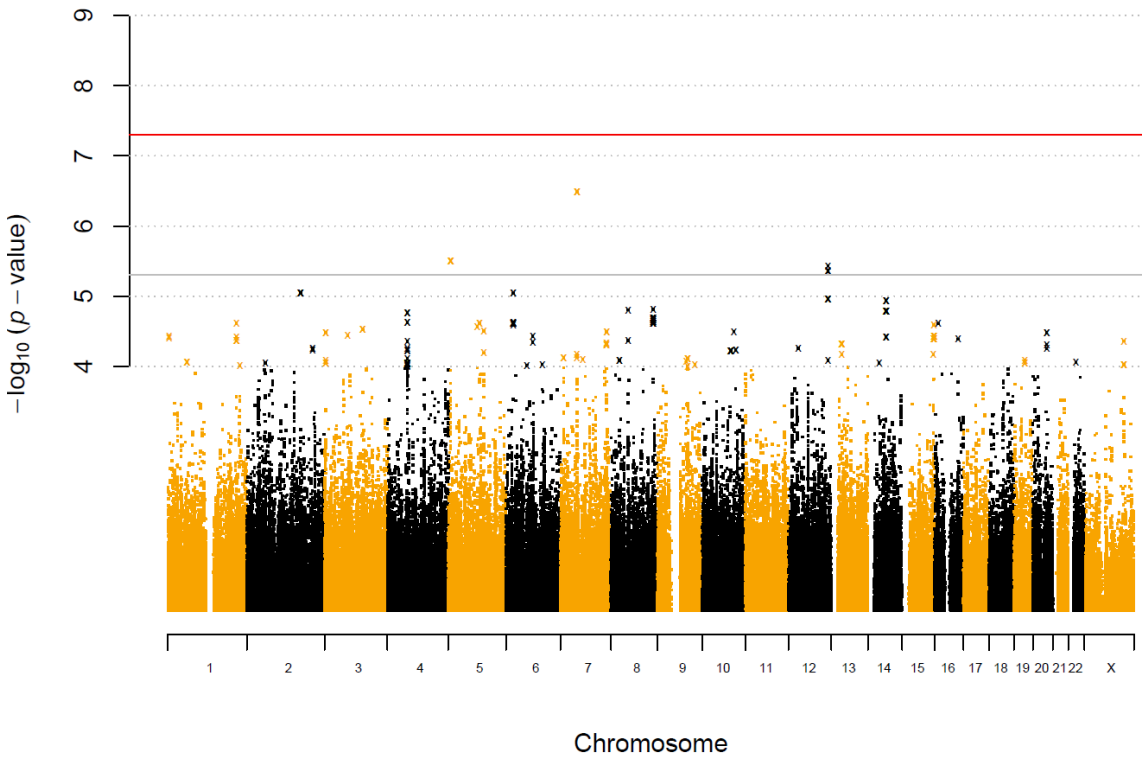
Removed for space concerns – available at

http://www.tandfonline.com/doi/suppl/10.1080/15622975.2016.1208841?scroll=top

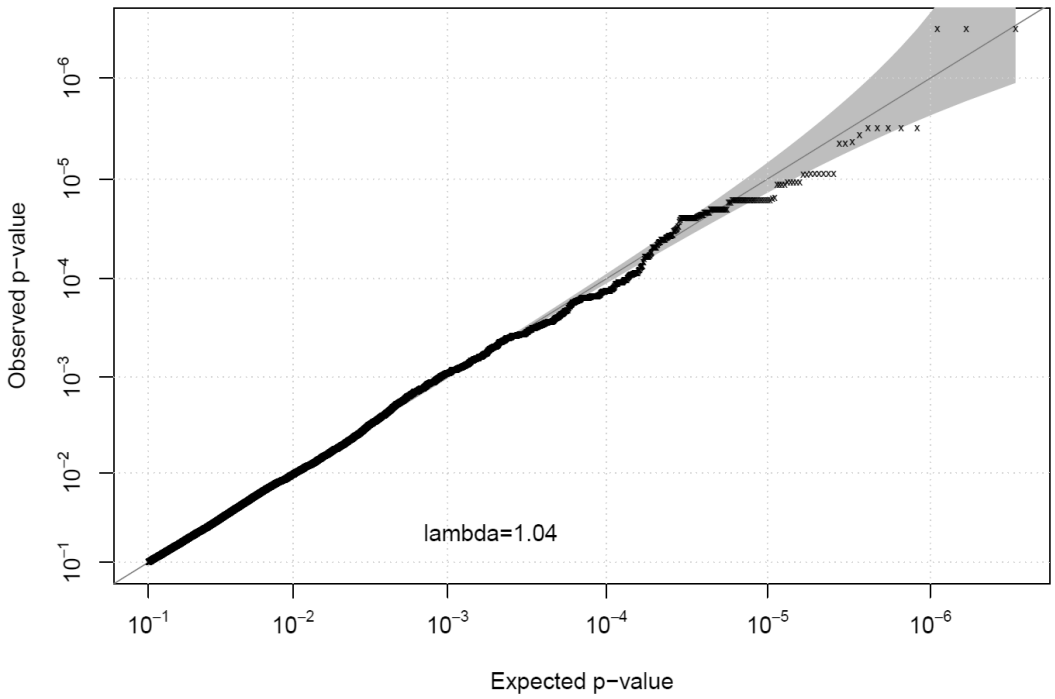
Supplementary Table 6: Linkage-independent blood eQTLs with $q<0.05$.

III.IV. Supplementary Figures

III.IV.I. Supplementary Figure 1 (Post-treatment)

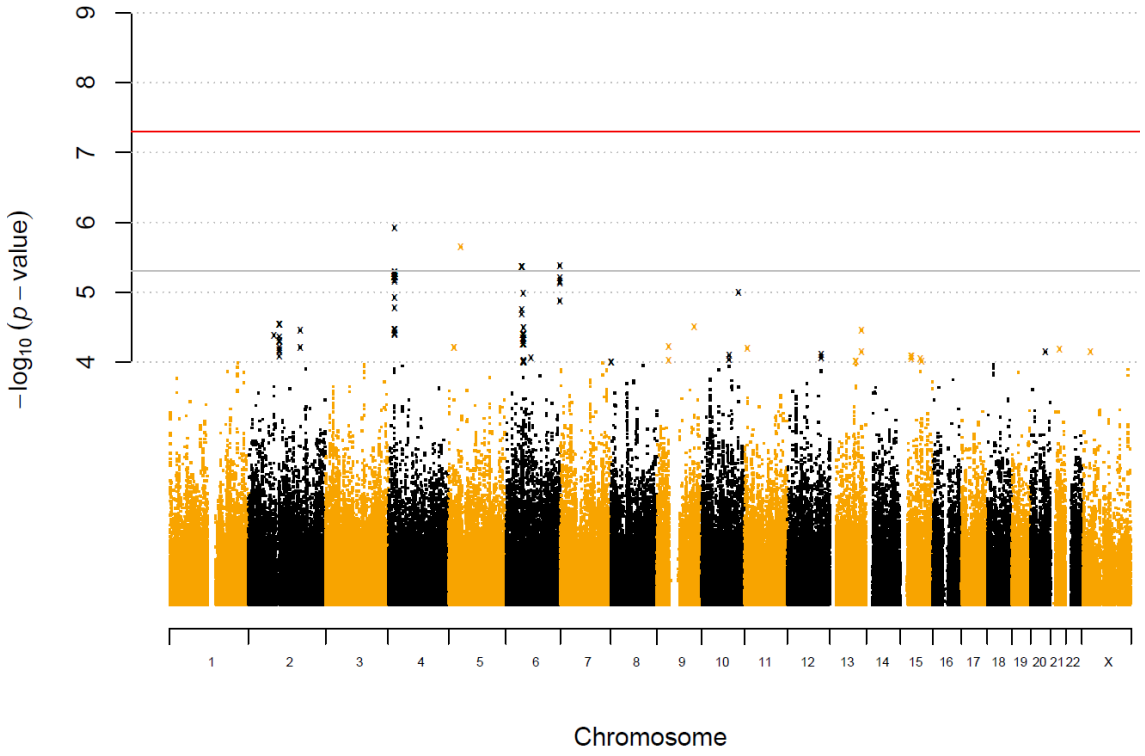


Supplementary Figure 1a: Manhattan plot of associations between treatment response baseline to post-treatment and genetic variants. X-axis shows position of genetic variants by chromosome. Y-axis is –log *p*-value, with top line showing the threshold for genome-wide significance ($p = 5x10^{-8}$), and bottom suggestive significance ($p = 5x10^{-6}$).

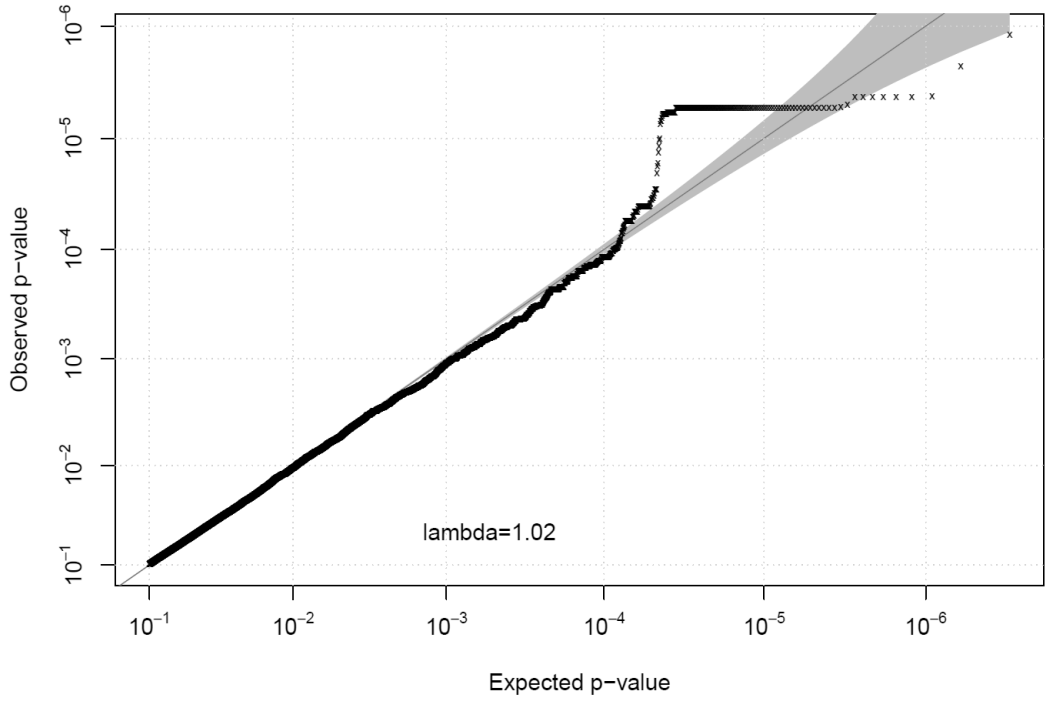


Supplementary Figure 1b: Quantile-quantile plot of associations between treatment response baseline to post-treatment and genetic variants. X-axis shows log *p*-value distribution expected under the null hypothesis. Y-axis shows observed log *p*-values. Lambda median is a measure of genomic inflation. Lambda ≈ 1, indicating minimal inflation due to confounds. Observed p-values do not deviate from the distribution expected under the null hypothesis.

III.IV.II. Supplementary Figure 2 (Follow-up)



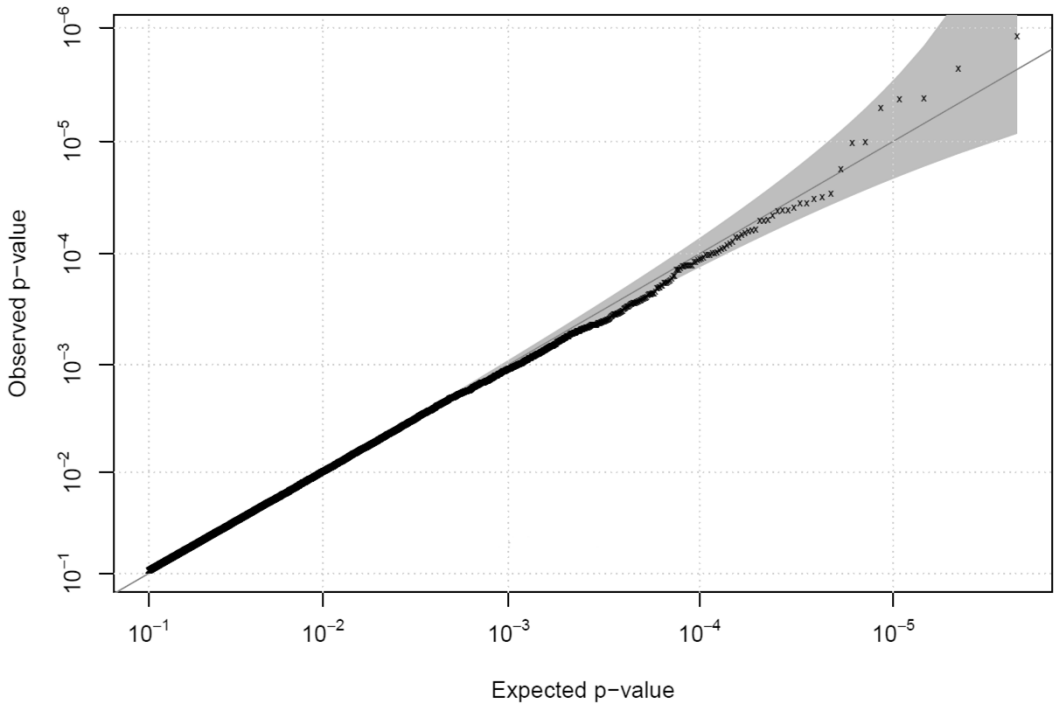
Supplementary Figure 2a: Manhattan plot of associations between treatment response baseline to follow-up and genetic variants.



Supplementary Figure 2b: Quantile-quantile plot of associations between treatment response baseline to follow-up and genetic variants.

Note that this plot includes all assayed variants (that is, variants in high linkage disequilibrium are retained), in order to assess the distribution of the observed results. 106 variants on chromosome 4 are in very high linkage disequilibrium and have an association p-value of 5.37×10^{-6}. Accordingly, they appear as a single point in the Manhattan plot (Supplementary Figure 2a), but form a kink in this QQ plot.

Compare Supplementary Figure 2c below, which removes variants in high linkage disequilibrium with more strongly associated variants.



Supplementary Figure 2c: Supplementary Figure 2b replotted after removing all variants in very high linkage disequilibrium with a more strongly associated variant ($r^2 > 0.99$, ± 250kb).

Appendix IV: Supplementary information from Chapter 5: The relationship between depression and body mass index in the UK Biobank, and the contribution of polygenic risk

IV.I Supplementary Notes

IV.I.I. Antidepressant use

Medication was classified into weight-increasing (amitriptyline, paroxetine, mirtazapine, clomipramine, nortriptyline and imipramine; N = 409), weight-decreasing (fluoxetine and venlafaxine; N = 348), weight-modulating (citalopram, sertraline, trazodone, duloxetine and fluvoxamine; N = 489) and weight-neutral (dosulepin/dothiepin, escitalopram, lofepramine, trimipramine, reboxetine, flupentixol/flupenthixol, doxepin, tranylcypromine, tryptophan, hypericum, buspirone, chlordiazepoxide, diazepam, lorazepam and nitrazepam; N=229). Drugs were assigned a weight effect if such was reported as frequent (or more common) in the SIDER drug side-effects database (with confirmation by searching the associated warning literature; Kuhn, Letunic, Jensen, *et al*, 2016). If a drug was not present in the database, assessment was made by a PubMed literature search for "drug AND (weight AND (gain OR loss))" and inspection of the resulting articles for reported weight effects.

IV.I.II. BMI PRS

The BMI PRS in this study was derived from the all ancestries analyses from the GIANT BMI consortium (Locke, Kahali, Berndt*, et al*, 2015). This consortium also published analyses in a European-only cohort (Locke, Kahali, Berndt*, et al*, 2015). The analyses in this paper were repeated using this European-only PRS, with no effect on the conclusions drawn.

IV.II. Supplementary References

Kuhn, M., Letunic, I., Jensen, L. J.*, et al* (2016) The SIDER database of drugs and side effects. *Nucleic Acids Res*, **44**, D1075-1079.

Locke, A. E., Kahali, B., Berndt, S. I.*, et al* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197-206.

IV.III. Supplementary Tables

IV.III.I. Supplementary Table 1

Coefficient	B	SE	*p*
Male gender	**-0.364**	**0.0152**	**$<10^{-50}$**
Age (years)	**-0.158**	**0.0151**	**8.55×10^{-26}**
Townsend Index	**0.189**	**0.0152**	**1.73×10^{-35}**
Centre *	1.86	1.25	0.137
Birth Cluster *	**0.386**	**0.112**	**5.53×10^{-4}**
Batch *	0.279	0.122	0.0221
PC1	0.00121	0.0156	0.938
PC2	0.0127	0.0152	0.403
PC3	0.00308	0.0157	0.845
PC4	0.00266	0.0208	0.898
PC5	-0.00806	0.0211	0.702
PC6	-0.0279	0.0157	0.0746
PC7	0.0158	0.0161	0.326
PC8	-0.0283	0.0177	0.110

Supplementary Table 1: Effects of covariates in the null model predicting variance in depression status. Significant ($p < 0.0125$) effects in bold.

IV.III.II. Supplementary Table 2

Coefficient	B	SE	*p*
Male gender	**0.106**	**0.00682**	**$<10^{-50}$**
Age (years)	**0.0477**	**0.00686**	**$3.73x10^{-12}$**
Townsend Index	**0.0954**	**0.00700**	**$3.70x10^{-42}$**
Centre *	**0.634**	**0.212**	**0.00283**
Birth Cluster *	**0.144**	**0.0515**	**0.00525**
Batch *	**0.206**	**0.0557**	**$2.13x10^{-4}$**
PC1	-0.00426	0.00706	0.546
PC2	$1.52x10^{-4}$	0.00687	0.982
PC3	0.00869	0.00712	0.222
PC4	0.00814	0.00940	0.387
PC5	-0.0197	0.00953	0.0385
PC6	-0.00540	0.00706	0.444
PC7	-0.00585	0.00726	0.326
PC8	-0.00312	0.00793	0.110

Supplementary Table 2: Effects of covariates in the null model predicting variance in log-BMI. Significant ($p < 0.0125$) effects in bold.

IV.III.III. Supplementary Table 3

Coefficient	B	SE	*p*
Male gender	**0.0711**	**0.0119**	**2.38×10^{-9}**
Age (years)	**0.0395**	**0.0120**	**9.83×10^{-4}**
Townsend Index	**0.0919**	**0.0122**	**6.46×10^{-14}**
Centre *	0.775	0.339	0.0222
Birth Cluster *	0.155	0.0948	0.103
Batch *	**0.320**	**0.0983**	**0.00113**
PC1	-0.0159	0.0123	0.196
PC2	-0.00850	0.0120	0.479
PC3	0.00656	0.0125	0.599
PC4	0.0119	0.0165	0.471
PC5	-0.0273	0.0167	0.103
PC6	-6.66×10^{-5}	0.0123	0.996
PC7	-0.0146	0.0126	0.246
PC8	0.00182	0.0139	0.896

Supplementary Table 3: Effects of covariates in the null model predicting variance in log-BMI in depression cases. Significant ($p < 0.0125$) effects in bold.

IV.III.IV. Supplementary Table 4

Coefficient	B	SE	*p*
Male gender	**0.141**	**0.00833**	**$<10^{-50}$**
Age (years)	**0.0596**	**0.00835**	**9.81×10^{-13}**
Townsend Index	**0.0868**	**0.00853**	**3.21×10^{-24}**
Centre *	0.618	0.258	0.0164
Birth Cluster *	0.123	0.0593	0.0381
Batch *	0.151	0.0684	0.0271
PC1	0.00263	0.00861	0.760
PC2	0.00464	0.00838	0.580
PC3	0.00978	0.00866	0.259
PC4	0.00449	0.0115	0.695
PC5	-0.0153	0.0116	0.189
PC6	-0.00688	0.00862	0.425
PC7	-0.00212	0.00889	0.811
PC8	-0.00411	0.00968	0.671

Supplementary Table 4: Effects of covariates in the null model predicting variance in log-BMI in depression controls. Significant ($p < 0.0125$) effects in bold.

IV.III.V. Supplementary Table 5

Coefficient	B	SE	*p*
Null model	See Supplementary Table 1		
… + depression PRS	**0.108**	**0.0165**	**5.32×10^{-11}**
… + BMI PRS	0.0137	0.0162	0.395
… + log-BMI	**0.104**	**0.0161**	**9.28×10^{-11}**
… + depression PRS **… + log-BMI**	**0.108** **0.103**	**0.0165** **0.0161**	**7.44×10^{-11}** **1.30×10^{-10}**
… + depression PRS x log-BMI (Multiplicative)	-0.0202	0.0167	0.227
… + depression PRS x log-BMI (Additive)	-0.00349	0.00331	0.292
… + BMI PRS **… + log-BMI**	-0.0116 **0.107**	0.0167 **0.0165**	0.486 **1.06×10^{-10}**
… + BMI PRS x log-BMI (Multiplicative)	9.82×10^{-4}	0.0160	0.951
… + BMI PRS x log-BMI (Additive)	5.44×10^{-4}	0.00318	0.864
… + depression PRS … + BMI PRS	**0.108** 0.0141	**0.0165** 0.0162	**5.23×10^{-11}** 0.385
… + depression PRS x BMI PRS (Multiplicative)	-0.0155	0.0170	0.362
… + depression PRS x BMI PRS (Additive)	0.00271	0.00330	0.412
… + depression PRS … + BMI PRS **… + log-BMI**	**0.108** -0.0111 **0.106**	**0.0165** 0.0167 **0.0165**	**7.61×10^{-11}** 0.506 **1.54×10^{-10}**
… + depression PRS x BMI PRS (Multiplicative) … + depression PRS x log-BMI (Multiplicative) …+ BMI PRS x log-BMI (Multiplicative)	0.0116 -0.0151 6.65×10^{-4}	0.0161 0.0173 0.0176	0.510 0.382 0.967
… + depression PRS x BMI PRS (Additive) … + depression PRS x log-BMI (Additive) …+ BMI PRS x log-BMI (Additive)	-0.00202 -0.00254 5.12×10^{-4}	0.00340 0.00341 0.00318	0.553 0.457 0.872
… + depression PRS x BMI PRS x log-BMI (Multiplicative)	0.00628	0.0157	0.689
… + depression PRS x BMI PRS x log-BMI (Additive)	0.00101	0.00310	0.746

Supplementary Table 5: Effects of adding variables and interactions to the null model predicting variance in depression status, excluding individuals on medication. Significant ($p < 0.0125$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms.

IV.III.VI. Supplementary Table 6

Coefficient	B	SE	*p*
Null model	See Supplementary Table 2		
… + BMI PRS	**0.232**	**0.00684**	**$< 10^{-50}$**
… + depression PRS	0.0115	0.00715	0.107
… + depression	**0.105**	**0.0158**	**3.41×10^{-11}**
… + BMI PRS **… + depression**	**0.232** **0.101**	**0.00683** **0.0153**	**$< 10^{-50}$** **3.95×10^{-11}**
… + BMI PRS x depression	0.00147	0.0154	0.339
… + depression PRS **… + depression**	0.00934 **0.104**	0.00715 **0.0158**	0.192 **5.39×10^{-11}**
… + depression PRS x depression	-0.0200	0.0161	0.215
… + BMI PRS … + depression PRS	**0.232** 0.0120	**0.00684** 0.00695	**$< 10^{-50}$** 0.083
… + BMI PRS x depression PRS	-3.09×10^{-4}	0.00710	0.965
… + BMI PRS … + depression PRS **… + depression**	**0.232** 0.00993 **0.100**	**0.00683** 0.00695 **0.0154**	**$< 10^{-50}$** 0.153 **6.47×10^{-11}**
… + depression PRS x BMI PRS … + BMI PRS x depression …+ depression PRS x depression	-3.31×10^{-4} 0.0146 -0.0163	0.00711 0.0154 0.0157	0.963 0.343 0.300
… + depression PRS x BMI PRS x depression	0.00907	0.0151	0.548
Cases			
Null model	See Supplementary Table 3		
… + BMI PRS	**0.238**	**0.0129**	**$< 10^{-50}$**
… + depression PRS	-0.00830	0.0135	0.538
… + BMI PRS … + depression PRS	**0.238** -0.00472	**0.0129** 0.0131	**$< 10^{-50}$** 0.719
… + BMI PRS x depression PRS	-3.41×10^{-4}	0.0133	0.979

Supplementary Table 6: Effects of adding variables and interactions to the null model predicting variance in log-BMI (in the whole sample and depression cases only), excluding individuals on medication. Significant ($p < 0.0125$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms.

IV.III.VII. Supplementary Table 7

Coefficient	B	SE	*p*
Null model	See Supplementary Table 2		
…+ weight increaser	**0.308**	**0.0497**	**5.92×10^{-10}**
… + weight decreaser	**0.299**	**0.0533**	**2.10×10^{-8}**
… + weight modulator	**0.272**	**0.0452**	**1.84×10^{-9}**
… + weight neutral	0.0585	0.0670	0.383
… + BMI PRS	**0.232**	**0.00660**	**$< 10^{-50}$**
… + depression PRS	0.00659	0.00690	0.340
… + depression	**0.105**	**0.0157**	**1.75×10^{-11}**
… + BMI PRS	**0.232**	**0.00659**	**$< 10^{-50}$**
… + depression	**0.102**	**0.0152**	**1.80×10^{-11}**
… + BMI PRS x depression	0.0153	0.0153	0.317
… + depression PRS	0.00449	0.00690	0.515
… + depression	**0.105**	**0.0157**	**2.25×10^{-11}**
… + depression PRS x depression	-0.0218	0.0160	0.171
… + BMI PRS	**0.232**	**0.00660**	**$< 10^{-50}$**
… + depression PRS	0.00729	0.00670	0.277
… + BMI PRS x depression PRS	-2.43×10^{-4}	0.00684	0.972
… + BMI PRS	**0.232**	**0.00659**	**$< 10^{-50}$**
… + depression PRS	0.00526	0.00670	0.433
… + depression	**0.102**	**0.0152**	**2.41×10^{-11}**
… + depression PRS x BMI PRS	2.12×10^{-4}	0.00685	0.975
… + BMI PRS x depression	0.0153	0.0153	0.317
…+ depression PRS x depression	-0.0194	0.0155	0.211
… + depression PRS x BMI PRS x depression	0.00954	0.0140	0.496

(Supplementary Table 7 continued)

Cases			
Null model	See Supplementary Table 3		
…+ weight increaser	**0.229**	**0.0508**	**6.89×10^{-6}**
… + weight decreaser	**0.219**	**0.0545**	**6.11×10^{-5}**
… + weight modulator	**0.192**	**0.0466**	**3.73×10^{-5}**
… + weight neutral	-9.15×10^{-4}	0.0680	0.999
… + BMI PRS	**0.238**	**0.0115**	**$< 10^{-50}$**
… + depression PRS	-0.0169	0.0120	0.159
… + BMI PRS	**0.238**	**0.0115**	**$< 10^{-50}$**
… + depression PRS	**-0.0135**	**0.0117**	**0.245**
… + BMI PRS x depression PRS	0.00273	0.0118	0.817

Supplementary Table 7: Effects of adding variables and interactions to the null model predicting variance in log-BMI (in the whole sample and stratified by depression status), adding covariates for medication status. Significant ($p < 0.0125$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms.

IV.III.VIII. Supplementary Table 8

Coefficient	B	SE	*p*
Null model	See Supplementary Table 1		
… + BMI PRS (Speliotes)	0.00642	0.0151	0.670
… + BMI PRS (Speliotes) **… + log-BMI**	-0.0221 **0.149**	0.0154 **0.0154**	0.151 **4.32×10^{-22}**
… + BMI PRS (Speliotes) x log-BMI (Multiplicative)	0.0141	0.0153	0.357
… + BMI PRS (Speliotes) x log-BMI (Additive)	0.00316	0.00324	0.330
… + depression PRS … + BMI PRS (Speliotes)	**0.0758** 0.00650	**0.0153** 0.0151	**7.01×10^{-7}** 0.667
… + depression PRS x BMI PRS (Speliotes) (Multiplicative)	-0.00596	0.0155	0.700
… + depression PRS x BMI PRS (Speliotes) (Additive)	-0.00142	0.00323	0.661
… + depression PRS … + BMI PRS (Speliotes) **… + log-BMI**	**0.0785** -0.0223 **0.150**	**0.0153** 0.0154 **0.0154**	**2.92×10^{-7}** 0.148 **1.85×10^{-22}**
… + depression PRS x BMI PRS (Speliotes) (Multiplicative) … + depression PRS x log-BMI (Multiplicative) … + BMI PRS (Speliotes) x log-BMI (Multiplicative)	-7.48×10^{-4} -0.0164 0.0153	0.0158 0.0160 0.0154	0.962 0.307 0.321
… + depression PRS x BMI PRS (Speliotes) (Additive) …+ depression PRS x log-BMI (Additive) … + BMI PRS (Speliotes) x log-BMI (Additive)	-4.09×10^{-4} -0.00298 0.00335	0.00328 0.00338 0.00325	0.901 0.378 0.302
… + depression PRS x BMI PRS (Speliotes) x log-BMI (Multiplicative)	-0.00556	0.0152	0.714
… + depression PRS x BMI PRS (Speliotes) x log-BMI (Additive)	-0.00118	0.00320	0.712

Supplementary Table 8: Effects of adding variables and interactions to the null model (effects shown in Supplementary Table 2) predicting variance in depression status, with the BMI PRS from Speliotes et al (2010). Significant ($p < 0.05$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms.

IV.III.IX. Supplementary Table 9

Coefficient	B	SE	*p*
Null model	See Supplementary Table 2		
… + BMI PRS (Speliotes)	**0.188**	**0.00668**	**$< 10^{-50}$**
… + BMI PRS (Speliotes) **… + depression**	**0.188** **0.143**	**0.00666** **0.0145**	**$< 10^{-50}$** **5.07×10^{-23}**
… + BMI PRS x depression	0.0280	0.0145	0.0542
… + BMI PRS (Speliotes) … + depression PRS	**0.188** -0.0137	**0.00668** 0.00675	**$< 10^{-50}$** 0.0427
… + BMI PRS x depression PRS	-0.00762	0.00678	0.261
… + BMI PRS (Speliotes) … + depression PRS **… + depression**	**0.188** -0.0160 **0.144**	**0.00666** 0.00674 **0.0145**	**$< 10^{-50}$** 0.0178 **2.36×10^{-23}**
… + depression PRS x BMI PRS (Speliotes) … + BMI PRS (Speliotes)x depression …+ depression PRS x depression	-0.00720 0.0288 -0.0167	0.00677 0.0146 0.0148	0.287 0.0478 0.257
… + depression PRS x BMI PRS (Speliotes) x depression	-0.00628	0.0141	0.657
Cases			
Null model	See Supplementary Table 3		
… + BMI PRS (Speliotes)	**0.200**	**0.0116**	**$< 10^{-50}$**
… + BMI PRS (Speliotes) **… + depression PRS**	**0.199** **-0.0302**	**0.0116** **0.0117**	**$< 10^{-50}$** **0.0100**
… + BMI PRS x depression PRS	-0.0101	0.0117	0.389
Controls			
Null model	See Supplementary Table 4		
… + BMI PRS (Speliotes)	**0.182**	**0.00816**	**$< 10^{-50}$**
… + BMI PRS (Speliotes) … + depression PRS	**0.182** -0.00767	**0.00816** 0.00826	**$< 10^{-50}$** 0.354
… + BMI PRS (Speliotes) x depression PRS	-0.0613	0.00841	0.466

Supplementary Table 9: Effects of adding variables and interactions to the null model predicting variance in log-BMI, with the BMI PRS from Speliotes et al (2010), and stratifying by depression case status. Significant ($p < 0.0125$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms.

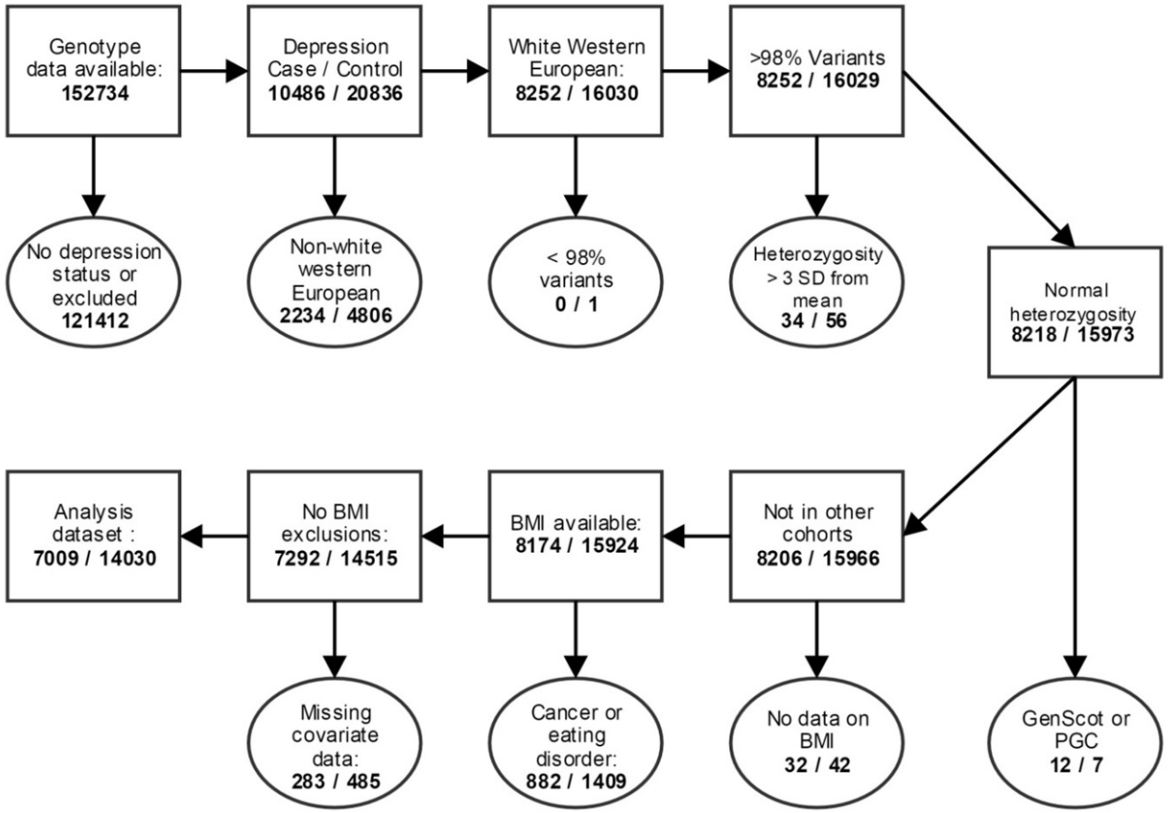
IV.III.X. Supplementary Table 10

Coefficient	B	SE	*p*
Depression			
BMI PRS	0.005	0.003	0.062
Depression PRS	**0.01**	**0.003**	**1.01×10^{-5}**
log-BMI			
BMI PRS	**0.26**	**0.007**	**$< 10^{-50}$**
Depression PRS	0.01	0.008	0.072
BMI PRS x depression	**0.046**	**0.019**	**0.016**
log-BMI in depression cases			
BMI PRS	**0.31**	**0.02**	**1.84×10^{-49}**
log-BMI in depression controls			
BMI PRS	**0.25**	**0.008**	**$< 10^{-50}$**

Supplementary Table 10: Analyses from the Generation Scotland cohort, using the BMI PRS derived from Locke et al (2015). For the BMI PRS x depression interaction, MDD and covariate-by-PRS and covariate-by-depression interactions were included (not shown).

IV.IV. Supplementary Figures

IV.IV.I. Supplementary Figure 1



Supplementary Figure 1: Sample exclusions during quality control

Appendix V: Supplementary information from Chapter 6: Interactions between social environment and polygenic risk scores for body mass index predicting variance in adolescent body mass index

V.I Supplementary Methods

V.I.I Genotyping protocol and quality control (as described in Selzam, Krapohl, von Stumm, *et al*, 2016)

Genome-wide genotype data was obtained in two waves of collection. In the first wave, DNA from 3,665 samples was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. Samples were successfully hybridised to Affymetrix Gene Chip 6.0 SNP genotyping arrays using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalised and preprocessed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

In the second wave, DNA for 4,649 individuals was extracted from saliva samples and hybridised to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility. The raw image data from the array were normalised, preprocessed, and filtered in GenomeStudio following internal protocols (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). Prior to genotype calling, 869 multi-mapping SNPs and 353 samples with call rate <.95 were removed. ZCall was used to augment the genotype calling for samples and SNPs that passed the initial QC (Goldstein, Crenshaw, Carey, *et al*, 2012).

After initial quality control and genotype calling, the same quality control was separately performed on samples from both waves using PLINK, R and vcftools (Chang, Chow, Tellier, *et al*, 2015; Danecek, Auton, Abecasis, *et al*, 2011; Purcell, Neale, Todd-Brown, *et al*, 2007; Team, 2014).

Samples were removed from subsequent analyses on the basis of call rate (<0.99), suspected non-European ancestry, heterozygosity, array signal intensity, and relatedness. SNPs were excluded if the minor allele frequency was <.05%, if more than 1% of genotype data were missing, or if the Hardy Weinberg p-value was lower than 10-5. Non-autosomal markers and insertion-deletions were removed. Association between the SNP and the

array, batch, or plate on which samples were genotyped was calculated; SNPs with an effect p-value < 10^{-3} were excluded. A total sample of 6,710 samples, with 3,617 individuals and 600,034 SNPs genotyped on Illumina and 3,093 individuals and 525,859 SNPs genotyped on Affymetrix remained after quality control.

Genotypes from the two arrays were separately imputed using the Haplotype Reference Consortium (McCarthy, Das, Kretzschmar, *et al*, 2016) and Minimac3 1.0.13 (Fuchsberger, Abecasis & Hinds, 2014; Howie, Fuchsberger, Stephens, *et al*, 2012) available on the Michigan Imputation Server as reference data. A series of quality checks was performed before data from the two waves were merged (e.g. array effects, allele frequencies by imputation quality). For the present analyses we limited our analyses to variants genotyped or imputed at info > 0.95 on both arrays, allele frequency difference between arrays smaller than 5%, and Hardy Weinberg p-value was greater than 10^{-5}. Using these criteria, 5,147,884 genotyped and well-imputed SNPs were retained for the analyses.

Principal component analysis was performed on a subset of 42,859 common (MAF>5%) autosomal SNPs found on the HapMap3 data (Consortium, 2010), after stringent pruning to remove markers in high linkage disequilibrium (r^2 > 0.1) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

V.II Supplementary Notes

V.II.I Sensitivity analyses using alternative PRS

The main analyses used a PRS derived from the all-ancestries analysis of the GIANT adult BMI GWAS meta-analysis (Locke, Kahali, Berndt, *et al*, 2015). There are good justifications to consider the European-only analyses from the adult GWAS (as the TEDS sample is predominantly of White Western European ancestry) and a recent child BMI GWAS meta-analysis as alternative base GWAS for generating PRS (Felix, Bradfield, Monnereau, *et al*, 2016; Locke, Kahali, Berndt, *et al*, 2015; Trzaskowski, Eley, Davis, *et al*, 2013). Analyses were repeated using these PRS. Conclusions from the replication with the European subset PRS differed in that the main effect of socioeconomic status on change in BMI across adolescence in males was not significant when PRS was in the model ($p = 0.00451$). Using the child BMI PRS from Felix et al (2016) did not alter the conclusions from the main analyses.

V.II.II. Interaction analyses with FTO rs1558902

To enable comparison to this previous literature, analyses were re-run using the number of A alleles of rs1558902 (and rs9939609) in place of the polygenic risk score. Conclusions from the analysis of rs1558902 and rs9939609 did not differ, as these variants are in strong linkage

disequilibrium ($r^2 > 0.9$), and so only results from rs9939609 are shown. Results were similar to those obtained with the full polygenic risk score, but showed smaller effects (Supplementary Tables 4 and 5).

V.III Supplementary References

Chang, C., Chow, C., Tellier, L.*, et al* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.

Consortium, I. H. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52-58.

Danecek, P., Auton, A., Abecasis, G.*, et al* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Felix, J. F., Bradfield, J. P., Monnereau, C.*, et al* (2016) Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet*, **25**, 389-403.

Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. (2014) minimac2: faster genotype imputation. *Bioinformatics*.

Goldstein, J. I., Crenshaw, A., Carey, J.*, et al* (2012) zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics (Oxford, England)*, **28**, 2543-2545.

Howie, B., Fuchsberger, C., Stephens, M.*, et al* (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*, **44**, 955-959.

Keller, M. C. (2014) Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biol Psychiatry*, **75**, 18-24.

Locke, A. E., Kahali, B., Berndt, S. I.*, et al* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197-206.

McCarthy, S., Das, S., Kretzschmar, W.*, et al* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*, **advance online publication**.

Purcell, S., Neale, B., Todd-Brown, K.*, et al* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559-575.

Selzam, S., Krapohl, E., von Stumm, S.*, et al* (2016) Predicting educational achievement from DNA. *Mol Psychiatry*.

Team, R. C. (2014) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

Trzaskowski, M., Eley, T. C., Davis, O. S.*, et al* (2013) First genome-wide association study on anxiety-related behaviours in childhood. *PLoS ONE*, **8**, e58676.

V.IV. Supplementary Tables

V.IV.I. Supplementary Table 1

	log(BMI)	BMI Change	BMI PRS	Parenting	SES	Sex	Age	Puberty	Wave	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
log(BMI)	-	**$3.3x10^{-135}$**	**$1.9x10^{-37}$**	**0.026**	**$4.0\ x10^{-8}$**	**0.0029**	**$2.3x10^{-18}$**	**$1.1x10^{-43}$**	0.93	0.60	0.90	0.63	0.36	0.46	0.44	**0.035**	0.88
BMI Change	**-0.52**	-	0.26	0.054	**0.0016**	0.77	**0.0089**	**$1.9x10^{-5}$**	0.22	**0.024**	0.56	0.83	**0.0051**	0.79	0.63	**0.042**	0.80
BMI PRS	**0.22**	-0.026	-	0.43	0.13	0.54	0.35	0.078	0.81	**0.037**	0.80	0.45	0.77	0.77	0.81	0.80	0.48
Parenting	**0.038**	-0.044	0.013	-	**$3.7x10^{-8}$**	**$9.7x10^{-8}$**	**0.0025**	0.63	0.97	0.91	0.61	0.93	0.71	0.14	0.45	0.37	0.86
SES	**-0.094**	**0.072**	-0.026	**-0.094**	-	0.86	**$1.8x10^{-5}$**	**$2.3x10^{-4}$**	0.14	0.18	0.19	**$1.4\ x10^{-5}$**	0.91	0.42	0.10	0.78	0.33
Sex	**-0.051**	-0.0067	0.010	**0.091**	0.0030	-	0.55	**$9.2x10^{-72}$**	0.12	0.071	0.32	0.24	0.98	0.75	0.25	0.13	0.76
Age	**0.15**	**-0.059**	-0.016	**-0.052**	**-0.073**	-0.01	-	**$3.7x10^{-77}$**	0.31	0.14	0.60	0.58	0.28	0.92	0.77	0.86	0.32
Puberty	**0.24**	**-0.097**	0.030	-0.0083	**-0.063**	**-0.30**	**0.31**	-	0.57	0.58	0.88	0.25	0.47	0.50	0.43	0.45	0.98
Wave1	-0.0016	0.028	-0.0042	$-6.0\ x10^{-4}$	0.025	-0.027	-0.017	-0.0099	-	**$< 10^{-150}$**	**$5.1x10^{-92}$**	**0.00032**	0.22	0.95	0.10	0.31	0.28
PC1	-0.0090	**0.051**	**-0.036**	0.0020	0.023	-0.031	0.025	0.0096	**0.72**	-	0.79	0.18	0.56	0.26	0.22	0.79	0.23
PC2	0.0022	-0.013	0.0044	0.0088	-0.022	-0.017	0.0091	-0.0027	**0.34**	-0.0045	-	0.63	0.70	0.40	0.29	0.28	0.12
PC3	0.0083	0.0049	-0.013	-0.0016	**0.074**	0.020	-0.0096	-0.020	**0.061**	-0.023	0.0081	-	0.19	0.42	0.10	0.70	0.68
PC4	0.016	**-0.064**	0.0050	-0.0064	-0.0019	$-4.6x10^{-4}$	-0.018	-0.012	-0.021	-0.0099	-0.0066	0.023	-	0.25	0.86	0.92	0.87
PC5	-0.013	0.0059	0.0051	0.025	-0.014	-0.0054	0.0018	-0.012	0.001	0.019	-0.015	-0.014	-0.020	-	0.94	0.68	0.60
PC6	0.013	0.011	-0.0041	-0.013	0.028	0.020	0.0051	-0.014	-0.028	-0.021	-0.018	0.028	-0.0031	0.0012	-	0.84	0.28
PC7	**-0.036**	**-0.046**	0.0044	0.015	0.0047	0.026	-0.0030	-0.013	-0.017	-0.0046	-0.018	-0.0066	-0.0018	0.0071	-0.0034	-	0.13
PC8	0.0026	0.0058	-0.012	0.0030	0.017	-0.0052	-0.017	$-4.0x10^{-4}$	0.018	0.021	-0.026	-0.0072	0.0030	-0.0089	0.019	0.026	-

Supplementary Table 1: Correlations between phenotypes and variables in the analyses (Pearson pairwise product-moment correlation; lower triangle) and associated *p*-values (upper triangle). Nominally significant correlations are marked in bold ($p < 0.05$).

V.IV.II Supplementary Table 2a

BMI at 11 years old, with parenting										
		Full cohort (N = 3414)			Females (N = 1750)			Males (N = 1664)		
Coefficient		B	SE	*p*	B	SE	*p*	B	SE	*p*
Null Model	Sex	0.0260	0.0349	0.457	-	-	-	-	-	-
	Age	**0.0795**	**0.0176**	**$6.12x10^{-6}$**	0.0346	0.0248	0.162	**0.126**	**0.0249**	**$4.90x10^{-7}$**
	SES	**-0.0766**	**0.0167**	**$4.42x10^{-6}$**	-0.0740	0.0230	**0.00131**	**-0.0796**	**0.0243**	**0.00106**
	Pubertal development	**0.211**	**0.0183**	**$5.44x10^{-30}$**	0.276	0.0247	**$4.38x10^{-28}$**	**0.106**	**0.0249**	**$2.30x10^{-5}$**
	Wave	0.0598	0.0566	0.291	0.151	0.0780	0.0532	-0.0566	0.0826	0.493
	PC1	-0.0312	0.0260	0.230	-0.0443	0.0353	0.211	-0.0107	0.0386	0.782
	PC2	-0.00982	0.0191	0.607	-0.0429	0.0262	0.102	0.0267	0.0280	0.340
	PC3	0.0149	0.0167	0.372	0.0190	0.0231	0.410	0.00822	0.0244	0.736
	PC4	0.0193	0.0166	0.245	0.00107	0.0229	0.963	0.0402	0.0241	0.0954
	PC5	-0.0101	0.0166	0.541	-0.00566	0.0229	0.805	-0.0148	0.0241	0.538
	PC6	0.0169	0.0166	0.307	0.0246	0.0229	0.281	0.00705	0.0241	0.770
	PC7	-0.0329	0.0166	0.0472	-0.0190	0.0229	0.407	-0.0461	0.0240	0.0553
	PC8	0.00575	0.0166	0.728	0.0181	0.0229	0.431	-0.0155	0.0241	0.520
Null model + parental style		0.0378	0.0167	0.0239	0.0413	0.0230	0.0730	0.0364	0.0242	0.133
Null model + **BMI PRS**		**0.210**	**0.0162**	**$1.59x10^{-37}$**	**0.192**	**0.0224**	**$3.04x10^{-17}$**	**0.231**	**0.0234**	**$2.68x10^{-22}$**
Null model + Parental style		0.0360	0.0163	0.0273	0.0407	0.0225	0.0712	0.0330	0.0235	0.160
+ **BMI PRS**		**0.210**	**0.0162**	**$1.83x10^{-37}$**	**0.191**	**0.0224**	**$3.03x10^{-17}$**	**0.230**	**0.0234**	**$3.19x10^{-22}$**
Null model + Parental style x BMI PRS		0.00642	0.0172	0.709	-0.00627	0.0240	0.794	0.0296	0.0252	0.241

Supplementary Table 2a: Effects of adding variables and interactions to the null model (uppermost line) predicting variance in BMI at 11 years old, with parenting as the environment of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

Supplementary Table 2b

BMI at 11 years old, with socioeconomic status										
		Full cohort (N = 3414)			Females (N = 1750)			Males (N = 1664)		
Coefficient		B	SE	*p*	B	SE	*p*	B	SE	*p*
Null Model	Sex	0.0187	0.0351	0.594	-	-	-	-	-	-
	Age	**0.0868**	**0.0176**	**$8.39x10^{-7}$**	0.0401	0.0248	0.107	**0.135**	**0.0250**	**$7.18x10^{-8}$**
	Parenting	0.0448	0.0167	0.00726	0.0490	0.0229	0.0326	0.0426	0.0242	0.0784
	Pubertal development	**0.212**	**0.0184**	**$2.43x10^{-30}$**	**0.279**	**0.0247**	**$1.22x10^{-28}$**	**0.106**	**0.0250**	**$2.46x10^{-5}$**
	Wave	0.0581	0.0567	0.306	0.142	0.0781	0.0687	-0.0507	0.0829	0.540
	PC1	-0.0329	0.0261	0.207	-0.0421	0.0354	0.235	-0.0170	0.0387	0.660
	PC2	-0.00838	0.0191	0.661	-0.0384	0.0262	0.143	0.0247	0.0280	0.379
	PC3	0.00950	0.0167	0.57	0.0143	0.0231	0.536	0.00178	0.0243	0.942
	PC4	0.0199	0.0166	0.229	0.00277	0.0229	0.904	0.0399	0.0242	0.0991
	PC5	-0.0102	0.0166	0.538	-0.00473	0.0229	0.837	-0.0164	0.0241	0.497
	PC6	0.0156	0.0166	0.349	0.0254	0.0229	0.268	0.00354	0.0242	0.884
	PC7	-0.0338	0.0166	0.0418	-0.0200	0.0229	0.382	-0.0466	0.0241	0.0532
	PC8	0.00455	0.0166	0.784	0.0174	0.0230	0.449	-0.0173	0.0241	0.473
Null model + **SES**		**-0.0729**	**0.0167**	**$1.33x10^{-5}$**	**-0.0694**	**0.0231**	**0.00274**	**-0.0766**	**0.0243**	**0.00167**
Null model + **BMI PRS**		**0.211**	**0.0162**	**$7.95x10^{-38}$**	**0.192**	**0.0225**	**$2.38x10^{-17}$**	**0.232**	**0.0235**	**$1.75x10^{-22}$**
Null model + **SES** + **BMI PRS**		**-0.0682** **0.210**	**0.0163** **0.0162**	**$3.11x10^{-5}$** **$1.83x10^{-37}$**	**-0.0663** **0.191**	**0.0227** **0.0224**	**0.00349** **$3.03x10^{-17}$**	**-0.0701** **0.230**	**0.0237** **0.0234**	**0.00309** **$3.19x10^{-22}$**
Null model + SES x BMI PRS		-0.0336	0.0165	0.0413	-0.0230	0.0233	0.324	-0.0382	0.0239	0.111

Supplementary Table 2b: Effects of adding variables and interactions to the null model (uppermost line) predicting variance in BMI at 11 years old, with socioeconomic status as the environment of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

V.IV.III Supplementary Table 3a

BMI change across adolescence, with parenting										
		Full cohort (N = 1943)			Females (N = 1043)			Males (N = 900)		
Coefficient		B	SE	*p*	B	SE	*p*	B	SE	*p*
Null Model	**BMI at 11**	**-0.528**	**0.0201**	**$1.56x10^{-130}$**	**-0.546**	**0.0275**	**$1.56x10^{-74}$**	**-0.505**	**0.0296**	**$1.55x10^{-56}$**
	Sex	-0.0281	0.0410	0.493	-	-	-	-	-	-
	Age	0.0178	0.0205	0.386	-0.0138	0.0286	0.628	0.0547	0.0298	0.0673
	SES	0.0113	0.0196	0.565	-0.0120	0.0264	0.650	0.0407	0.0294	0.167
	Pubertal development	0.0126	0.0218	0.562	0.0206	0.0293	0.482	0.0128	0.0298	0.669
	Wave	-0.0135	0.0665	0.839	0.0749	0.0901	0.406	-0.130	0.0992	0.190
	PC1	0.0353	0.0303	0.244	-0.00625	0.0404	0.877	0.0890	0.0461	0.0536
	PC2	-0.0157	0.0224	0.483	-0.00365	0.0303	0.904	-0.0306	0.0335	0.362
	PC3	0.0105	0.0197	0.594	-0.00693	0.0266	0.794	0.0290	0.0295	0.326
	PC4	**-0.0588**	**0.0193**	**0.00242**	-0.0593	0.0263	0.0243	-0.0536	0.0290	0.0650
	PC5	-0.00878	0.0193	0.650	-0.0234	0.0263	0.373	0.00648	0.0288	0.822
	PC6	0.00908	0.0193	0.639	0.0134	0.0262	0.611	-0.00208	0.0290	0.943
	PC7	**-0.0710**	**0.0194**	**0.000251**	-0.0666	0.0263	0.0114	-0.0708	0.0289	0.0144
	PC8	0.00997	0.0193	0.606	0.00450	0.0262	0.864	0.0207	0.0289	0.473
Null model + parental style		-0.00848	0.0196	0.666	-0.0147	0.0265	0.580	$-9.79x10^{-5}$	0.0291	0.997
Null model + BMI PRS		**0.0902**	**0.0197**	**$4.96x10^{-6}$**	**0.105**	**0.0266**	**$8.12x10^{-5}$**	0.0738	0.0295	0.0124
Null model + parental style **+ BMI PRS**		-0.00969 **0.0903**	0.0195 **0.0197**	0.620 **$4.84x10^{-6}$**	-0.0159 **0.105**	0.0263 **0.0266**	0.546 **$7.94x10^{-5}$**	-0.00134 0.0738	0.0290 0.0295	0.963 0.0125
Null model + Parental style x BMI PRS		0.000551	0.0207	0.979	0.0209	0.0285	0.463	-0.0416	0.0309	0.179

Supplementary Table 3a: Effects of adding variables and interactions to the null model (uppermost line) predicting change in BMI across adolescence, with parenting as the environment of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

Supplementary Table 3b

BMI change across adolescence, with socioeconomic status										
		Full cohort (N = 1943)			Females (N = 1043)			Males (N = 900)		
Coefficient		B	SE	*p*	B	SE	*P*	B	SE	*p*
Null Model	**BMI at 11**	**-0.528**	**0.0200**	**2.99×10^{-131}**	**-0.544**	**0.0275**	**3.36×10^{-74}**	**-0.509**	**0.0295**	**1.11×10^{-57}**
	Sex	-0.0265	0.0412	0.521	-	-	-	-	-	-
	Age	0.0162	0.0206	0.432	-0.0146	0.0286	0.611	0.0516	0.0301	0.0862
	Parental style	-0.00936	0.0196	0.632	-0.0133	0.0264	0.613	-0.00262	0.0291	0.928
	Pubertal development	0.0125	0.0218	0.569	0.0206	0.0293	0.484	0.0104	0.0299	0.727
	Wave	-0.0121	0.0664	0.855	0.0754	0.0901	0.403	-0.131	0.0993	0.189
	PC1	0.0351	0.0303	0.246	-0.00668	0.0404	0.869	0.0915	0.0461	0.0475
	PC2	-0.0161	0.0224	0.470	-0.00342	0.0303	0.910	-0.0307	0.0336	0.360
	PC3	0.0111	0.0196	0.572	-0.00733	0.0265	0.783	0.0326	0.0294	0.267
	PC4	**-0.0587**	**0.0193**	**0.00246**	-0.0591	0.0263	0.0249	-0.0520	0.0290	0.0738
	PC5	-0.00854	0.0193	0.659	-0.0228	0.0263	0.387	0.00784	0.0289	0.786
	PC6	0.00938	0.0193	0.628	0.0128	0.0262	0.624	-4.49×10^{-5}	0.0290	0.999
	PC7	**-0.0708**	**0.0194**	**0.000262**	-0.0670	0.0263	0.0109	-0.0709	0.0289	0.0144
	PC8	0.00955	0.0193	0.621	0.00486	0.0262	0.853	0.0198	0.0289	0.493
Null model + SES		0.0106	0.0196	0.591	-0.0135	0.0265	0.612	0.0407	0.0295	0.168
Null model + BMI PRS		**0.0904**	**0.0197**	**4.70×10^{-6}**	**0.105**	**0.0266**	**8.19×10^{-5}**	0.0739	0.0295	0.0124
Null model + SES **+ BMI PRS**		0.00965 **0.0903**	0.0195 **0.0197**	0.622 **4.84×10^{-6}**	-0.0152 **0.105**	0.0264 **0.0266**	0.565 **7.94×10^{-5}**	0.0405 0.0738	0.0294 0.0295	0.168 0.0125
Null model + SES x BMI PRS		-0.0494	0.0205	0.0159	-0.0724	0.0282	0.0103	-0.0152	0.0311	0.626

Supplementary Table 3b: Effects of adding variables and interactions to the null model (uppermost line) predicting change in BMI across adolescence, with socioeconomic status as the environment of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

V.IV.IV. Supplementary Table 4

Coefficient	B	SE	*p*	Adjusted R^2
BMI at 11 years old, with parenting				
Null model	Supplementary Table 2a			0.0667
(Null model) + Parental style	0.0378	0.0167	0.0239	0.0678
(Null model) + **rs9939609**	**0.150**	**0.0239**	**4.64×10^{-10}**	0.0770
(Null model) + Parental style + **rs9939609**	0.0374 **0.149**	0.0166 **0.0239**	0.0247 **5.28×10^{-10}**	0.0781
(Null model) + Parental style x rs9939609	-0.0208	0.0245	0.395	0.0800
BMI at 11 years old, with SES				
Null model	Supplementary Table 2b			0.0628
Null model + **SES**	**-0.0729**	**0.0167**	**1.33×10^{-5}**	0.0678
Null model + **rs9939609**	**0.149**	**0.0240**	**6.08×10^{-10}**	0.0734
Null model + **SES** + **rs9939609**	**-0.0714** **0.149**	**0.0167** **0.0239**	**1.89×10^{-5}** **5.28×10^{-10}**	0.0781
Null model + SES x rs9939609	-0.0237	0.0242	0.328	0.0794

Supplementary Table 4: Effects of adding variables and interactions to the null model predicting variance in BMI at 11 years old, with parenting as the environment of interest and FTO variant rs9939609 as the genotype of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).

V.IV.V. Supplementary Table 5

Coefficient	B	SE	*p*	Adjusted R^2
BMI change, with parenting				
Null model	Supplementary Table 3a			0.277
(Null model) + Parental style	-0.00848	0.0196	0.666	0.277
(Null model) + rs9939609	0.0450	0.0285	0.114	0.277
(Null model) + Parental style + rs9939609	-0.00702 0.0452	0.0196 0.0285	0.720 0.112	0.277
(Null model) + Parental style x rs9939609	0.00651	0.0289	0.821	0.274
BMI change, with SES				
Null model	Supplementary Table 3b			0.277
Null model + SES	0.0106	0.0196	0.591	0.277
Null model + rs9939609	0.0454	0.0285	0.111	0.277
Null model + SES + rs9939609	0.0121 0.0452	0.0196 0.0285	0.616 0.112	0.277
Null model + SES x rs9939609	0.0180	0.0293	0.541	0.279

Supplementary Table 5: Effects of adding variables and interactions to the null model predicting variance in BMI at 11 years old, with SES as the environment of interest and FTO variant rs9939609 as the genotype of interest. Significant ($p<0.00417$) terms are in bold. Interactions include all main effects, covariates and covariate interaction terms (Keller, 2014).