THE UNIVERSITY
of EDINBURGH

Leveraging genomic risk factors for Major Depressive Disorder to provide mechanistic insights and predictive neurobiological markers

Miruna Carmen Barbu

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy (PhD)

The University of Edinburgh

2019

Abstract

Major Depressive Disorder (MDD) is a disabling, common psychiatric disorder and the leading cause of global disability. A complex combination of genetic and environmental factors gives rise to MDD, although the exact aetiology has not been identified. Genome-wide association studies (GWAS) have established that MDD has a moderate heritability of approximately 37%. MDD has in the past also been associated with abnormalities of white matter microstructure, which represents the brain's connectivity network. This network is also moderately heritable, providing rationale to investigate its relationship to MDD genetic risk.

Over recent years, there has been considerable progress in establishing genetic contributions to MDD. These advances can be harnessed, in combination with neuroimaging and epigenomics, to understand the neurobiology of the disorder. This has only recently become possible at sufficient scale with the availability of large publicly available datasets including genomic, epigenomic, and neuroimaging data.

In the current thesis, I therefore aimed to leverage genetic, epigenetic, and neuroimaging data in two large datasets, UK Biobank (N range: 6,400 – 14,800) and Generation Scotland: Scottish Family Health Study (N = 625). Specifically, I aimed to uncover links between white matter microstructure, as measured by fractional anisotropy and mean diffusivity, and (i) differential gene expression as indexed by expression quantitative trait loci (eQTLs) scores in chapter 2; here, decreased white matter integrity was found to be associated with 6 scores regulating genes previously reported to be implicated in neurological and neuropsychiatric disorders, while 2 scores regulating neurodevelopment-linked genes were associated with increased white matter integrity; (ii) MDD genetic risk stratified by the NETRIN1 Signalling Pathway, previously implicated in MDD, indexed by polygenic risk scores (PRS) in chapter 3; results indicated novel associations between the pathway-focussed PRS and decreased white matter integrity in thalamic radiations, as well as several association fibres, including superior and inferior longitudinal fasciculus; (iii) a novel whole-genome epigenetic risk score for MDD, which uncovered an association with MDD, but no significant associations with changes in white matter microstructure (chapter 4). The overall aim of the thesis was to use advanced genomic techniques to stratify

genetic function and risk and explore epigenetic risk for MDD in order to identify novel links to structural brain connectivity.

Overall, the three studies provide a strong rationale for integrating neuroimaging, genomic and epigenomic data. Specifically, findings in chapter 2 indicate the importance of *DCAKD*, *SLC35A4*, *SEC14L4*, *SRA1*, *PLEKHM1*, *UBE3C, NMT1*, and *CPNE1*, not previously found by conventional GWAS approaches. This suggests that integrating neuroimaging and genetic expression data may uncover novel associations that inform disease- or trait-specific genetic links to brain connectivity. Chapter 3 results provide a rationale for investigating the NETRIN1 Signalling Pathway and emphasise the role of thalamic connections in MDD within this biological pathway, indicating that novel associations with brain connectivity may be uncovered at a more focused level when stratifying MDD risk by biology. Finally, results from chapter 4 indicate that epigenetics play an important role in MDD risk, although further analysis including larger-scale epigenetic and neuroimaging data should be carried out to uncover the role of epigenetics in relation to brain phenotypes.

Lay summary

Major Depressive Disorder (MDD) is a common psychiatric disorder affecting approximately 4.4% of the world's population. Although a complex mixture of environmental and genetic factors plays a role in MDD, an exact cause has not been identified. MDD has been linked to changes in the wiring of the brain, which is also known to have a genetic component, making it a valid target in the investigation of MDD.

Combining neuroimaging and genetic data is useful in the investigation of MDD, as it may provide novel insights into disease mechanisms, ultimately leading to disorder categorisation and novel treatments. Despite this, large-scale studies combining both types of data have only recently become available.

The current thesis therefore presents three studies using two large datasets that comprise both genetic and neuroimaging data. In chapter 2, I looked at the genetics behind protein production, which is carried out by genes. I found that poorer connectivity was associated with genes previously known to play a role in brain-related disorders, while better connectivity was linked to those implicated in developmental processes. In chapter 3, genetic risk for MDD aggregated in a specific biological process was linked to poor connectivity between the thalamus and other parts of the brain, as well as to connections linking homologous parts of the two brain hemispheres. In chapter 4, two types of MDD risk, one coming from multiple genetic mutations, and one which may be modified by environmental factors called "epigenetic risk", were shown to additively increase risk of having MDD, but were not linked to brain connectivity in this sample, though larger studies with this specific type of "epigenetic" data are required.

The three studies show that using analysis methods that link different forms of genetic data to neuroimaging variables may elucidate the role played by a large number of genetic mutations in MDD, as well as identify specific biomarkers, improving diagnosis and treatment outcomes.

Declarations

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

The thesis includes one article submitted for publication:

Chapter 2: **Barbu MC**, Spiliopoulou A, Colombo M, McKeigue P, Clarke TK, Howard D, Adams M, Shen X, Lawrie SM, McIntosh AM, Whalley HC (2019). Expression quantitative trait loci-derived scores and white matter microstructure in UK Biobank: a novel approach to integrating genetics and neuroimaging (revision for *Translational Psychiatry*).

And one published article:

Chapter 3: **Barbu MC**, Zeng Y, Shen X, Cox S, Clarke TK, Gibson J, Adams M, Johnstone M, Haley C, Lawrie SM, Deary I, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team, McIntosh AM, Whalley HC (2019). Association of Whole-Genome and NETRIN1 Signaling Pathway–Derived Polygenic Risk Scores for Major Depressive Disorder and White Matter Microstructure in the UK Biobank. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(1), 91-100. DOI: 10.1016/j.bpsc.2018.07.006

Miruna Carmen Barbu

November, 2019

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Heather Whalley, Prof Andrew McIntosh, and Prof Stephen Lawrie. Dr Whalley has been a constant source of support, and I could always count on receiving help from her. She has been, and continues to be, a really great supervisor, always encouraging critical thinking in my research. Prof McIntosh first introduced me to the fascinating world of genetics, and I am grateful for the support and guidance I've received from him throughout the years. Prof Lawrie was my first point of contact to the field of neuroimaging in psychiatric disorders – thank you for encouraging me in my research during my PhD. I would also like to thank Dr Johnstone, who has introduced me to molecular psychiatry and provided invaluable guidance during my first year.

I would also like to thank my parents – for being available any time of day for my phone calls about my PhD, either positive or negative, and for the unwavering support they have provided since day one, and my sister, who has taught me that really hard work pays off. Finally, Philly – there are no words to describe the amount of support and love I've felt, and how much I've learned from you, from discussions about DNA methylation to the difficulty of doing psychiatric research.

Finally, I would like to thank my friends, and colleagues at Kennedy Tower - you've made this journey easier and I've felt very supported throughout the years. Shen and Toni – you are both role models for how I would like to conduct research, and I would like to thank you for always being there for me, discussing interesting findings and possible research avenues. Many thanks also to Squeaky, Makis, Cecilia, Clara, Laura(s), Melissa(s), Mat, Dave, Mark, Emma, and others I have not mentioned here.

My PhD, like any other journey, has been full of ups and downs – the people who accompanied me on this journey made it all worth it.

Table of contents

Chapter 1: Introduction

1. Major Depressive Disorder (MDD)

1.1 Definition and diagnosis

Major Depressive Disorder (MDD) is a common psychiatric disorder and the leading cause of disability worldwide. According to a report by the World Health Organization (2017), it is estimated that over 300 million individuals are affected globally, which is equivalent to 4.4% of the world's population (World Health Organization, 2017).

MDD is mainly characterized by at least one depressive episode of at least a 2-week duration, with symptoms persisting for most of the day, nearly every day. According to the Diagnostic and Statistical Manual of Mental Disorders-V (DSM-V), for a diagnosis of MDD, at least 5 of 9 symptoms must be present (Table 1). At least one of the symptoms must be either depressed mood or loss of interest or pleasure in daily activities. In addition, MDD may be further characterized using specifiers, which describe the nature of an episode (e.g. severity of episode, with mixed, melancholic, atypical, mood-congruent psychotic or mood-incongruent psychotic features, with catatonia, with peripartum onset or with a seasonal pattern). These symptoms must mark a significant change from previous functioning, such as impairment in social, educational or occupational domains, and may not be attributable to another medical condition (APA, 2013).

Due to the classification system of MDD, there are over 200 ways in which patients can meet diagnostic criteria for MDD. This means that 2 patients diagnosed with MDD can have completely different symptom profiles (Zimmerman et al., 2015). Moreover, some symptoms are alternative or opposite: a patient presenting with psychomotor agitation and insomnia meets criteria in the same way as a patient presenting with psychomotor retardation and hypersomnia (Goldberg, 2011). These factors make MDD a highly heterogeneous disorder, which may lead to difficulty in downstream analyses of the disorder.

To address this inherent heterogeneity, stratification of the disorder is needed. For instance, patients may form sub-groups comprised of different biological mechanisms. Kunugi et al. (2015) discuss three distinct biological mechanisms which may act as sub-groups of MDD. Briefly, different classes of antidepressants inhibit the reuptake of neurotransmitters in the monoamine system (serotonin, noradrenaline, dopamine), which are thought to be important biomarkers for MDD; secondly, the hypothalamic-pituitary-adrenal (HPA) axis has been shown to be disrupted in MDD, with patients showing both hyper- and hypo-cortisolism; lastly, MDD has been proposed as a chronic inflammatory disease, as shown by inflammatory markers linked to the disorder. However, research has been inconclusive and often showed opposite results when investigating these three, and other, mechanisms in MDD patients (Hodes et al., 2015; Kunugi et al., 2015), suggesting the importance for potential biological stratification of patients in future analyses.

In addition, studies suggest stratification of symptoms when assessing their association with traits of interest. Pearson et al. (2017) investigated the extent to which variation in single nucleotide polymorphism (SNPs) explained variation in 4 MDD symptom dimensions in 1,345 cases. They found that core depressives symptoms such as sad mood and anhedonia had a lower SNP heritability (14%) than symptoms such as insomnia and appetite (30% for both), although replication is needed for a more robust conclusion of this study (Pearson et al., 2017).

The approach of stratifying patients through biological systems or phenotypic similarity allows for more homogeneity within MDD when investigating specific links to biologically relevant mechanisms. This may lead to a more effective personalised medicine approach, such as tailoring treatment options to specific sub-groups (Wardenaar & de Jonge, 2013; Fried, 2017).

	Symptom
1.	Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (Note: In children and adolescents, can be irritable mood)
2.	Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation)
3.	Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month) or decrease or increase in appetite nearly every day. (Note: In children, consider failure to make expected weight gain)
4.	Insomnia or hypersomnia nearly every day
5.	Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
6.	Fatigue or loss of energy nearly every day
7.	Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
8.	Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
9.	Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

Table 1. DSM-V diagnostic criteria for MDD (APA, 2013).

1.2 Epidemiology

The average 12-month prevalence of MDD is approximately 6%, with 1 in 6 individuals affected. Although limited by recall bias and underestimation, reports show that approximately 20% of all individuals fulfill diagnosis criteria for MDD at some point in their life (Otte et al., 2016). MDD typically affects twice as many women (5.1%) as men (3.6%), at any age (WHO, 2017), with the number of MDD episodes also being more frequent in women than men (Otte et al., 2016). Moreover, the median age of onset is 25 years for both men and women, although MDD may appear at any age, and the risk period for MDD appearance ranges from mid-adolescence to mid-life (early 40s) (Otte et al., 2016).

Between 2005 and 2015, it is estimated that the number of people with an MDD diagnosis increased by 18.4%, reflecting both a growing population and an increase in the possible age groups which receive an MDD diagnosis (WHO, 2017). Briefly, nowadays, more individuals globally grow to an older age, leading to an increase in incident cases.

A recent WHO report indicated that there are some regional differences in the prevalence of MDD, ranging from 2.6% affected males in the Western Pacific region, to 5.9% affected females in the African region. Moreover, from a total of 322 million affected individuals worldwide, 9% of these are in the African region while 27% are in the South-East Asia Region (WHO, 2017). MDD also affects individuals irrespective of income. Bromet et al. (2011) investigated data from 18 countries categorized by income (N = 89,037) and found that the lifetime and 12-month prevalence was 14.6% and 5.5% in 10 high-income and 11.1% and 5.9% in 8 low-income countries, respectively, indicating that the manifestation of MDD is similar across countries, independent of income (Bromet et al., 2011).

1.3 MDD impact on everyday functioning and treatment options

The economic burden of MDD has increased through the years, with 5% attributable to suicide-related costs, 48-50% to workplace costs, and a significant 45-47% accounting for direct medical costs (Greenberg et al., 2015). MDD has a substantial impact on workplace performance, with MDD individuals missing approximately one month of work per year (McIntyre et al., 2015) Moreover, approximately 60% of individuals with MDD report impairment of functioning (Fried & Nesse, 2014). Furthermore, MDD has an effect on a range of domains which may impact individuals' capability for self-care and independent living, including homelife, social activities and relationships (Beblo et al., 2010; Rot et al., 2012; Fried & Nesse, 2014).

Given the far-reaching negative impact of MDD, numerous treatment options have been investigated in order to establish which is the most efficacious. Gartlehner et al. (2017) looked at 140 pharmacological and non-pharmacological treatment options in a review of systematic reviews and identified only 5 treatment options for which the general efficacy for MDD in an acute phase is supported by reliable evidence. Of these, cognitive-behavioural therapy seems to be the only non-pharmacological treatment with similar efficacy to second generation antidepressants, based on moderate strength evidence (Gartlehner et al., 2017). Khan et al. (2012) aimed to compare the efficacy of various treatment options in a review, including

psychotherapy, pharmacotherapy, a combination of those two, and alternative therapies. Although further research is needed, they concluded that a combination of psychotherapy and pharmacotherapy provided a slight advantage as compared to only taking antidepressants or participating in therapy.

Lastly, Cipriani et al. (2018) carried out a review and meta-analysis investigating the efficacy and acceptability of 21 antidepressant drugs in the acute treatment of adults with MDD, which included 116,477 participants across 522 trials. All 21 antidepressants were more efficacious than placebo in adults with MDD with modest effect sizes, although there was variability in their efficacy and acceptability, which indicates heterogeneity of acting drugs.

The variability in treatment for MDD mentioned above suggests that empirical research is needed in order to uncover novel targets for intervention. Genetic studies targeting specific biological pathways and genes, as well as neuroimaging studies focusing on specific brain regions associated with MDD will be needed in order to address the need for novel optimal treatment options.

1.4 Major risk factors

MDD arises as a result of a complex combination of environmental and genetic risk factors. The sections below outline some of the major risk factors, and it is important to note that these do not act in isolation. In MDD, gene-environment interactions are complex, and cumulatively act to predispose individuals to the development of the disorder throughout their lifetime (Lopizzo et al., 2015).

1.4.1 Environmental risk factors

A majority of epidemiological studies find that gender and age are highly associated with depression (Stordal et al., 2001; Brodaty et al., 2005). As indicated above, women have a two-fold increased risk of MDD, and MDD risk is known to increase with age (WHO, 2017). In addition, a variety of other sociodemographic factors increase the risk of depression. For instance, childhood is a period in life when the brain is vulnerable and sensitive due to developmental processes (Heim & Binder,

2012). Therefore, early-life stressors such as early adversity (e.g. sexual, physical, and emotional abuse and maltreatment), parental loss due to separation or death, poor paternal relationships, or maternal overprotection have all been reported to lead to an increased risk of depression (Gibb et al., 2001; Gibb, Chelminski & Zimmerman, 2007; Heim & Binder, 2012). Other environmental risk factors later in life include stressful life events, such as moving to a new house (Bhugra & Ayonrinde, 2004), a lower socio-economic status (Gavin et al., 2010), and a stressful work environment (Theorell et al., 2015).

1.4.2 Health risk factors

In addition to environmental risk factors, multiple health factors may increase the risk for depression. For instance, researchers have found that a family history of depression and co-morbidity with other psychiatric disorders (e.g. schizophrenia, bipolar disorder, anxiety disorders) may lead to an increased risk of depression. Individuals who have already experienced an episode of depression are also at increased likelihood of experiencing further episodes (Kendler et al., 2001).

Brook et al. (2002) showed that cumulative use and frequency of drug use, such as alcohol and marijuana, in childhood and early adolescence, was associated with episodes of MDD in the late 20s (Brook et al., 2002), a link uncovered in other studies as well (Nemeroff & Vale, 2005; Neupane, 2016). Moreover, food addiction has previously been linked to both MDD and depressive symptom severity (Mills et al., 2020).

Lastly, previous evidence has shown that physical conditions, such as cardiovascular disorders and type 2 diabetes, may also lead to an increased risk of depression (Beekman et al., 2000; Heim & Binder, 2012). Chronic inflammatory states and chronic pain have been consistently associated with depression prevalence (Dantzer et al., 2007; Ohayon & Schatzberg, 2003; Ohayon & Schatzberg, 2012). During system infections for instance, continual activation of the peripheral immune system may lead to the development of depressive symptoms, marking inflammation as an important risk factor for depression (Dantzer et al., 2007). Furthermore, in their study investigating the prevalence of chronic painful physical conditions and MDD,

Ohayon and Schatzberg (2010) found that 73.3% of participants who met criteria for MDD also reported chronic pain (Ohayon & Schatzberg, 2010). The studies above therefore indicate the importance of both physical and psychological health factors in MDD prevalence.

1.4.3 Genetic risk factors

Twin, adoption, and family studies

In an early meta-analysis of studies investigating genetic contributions to MDD, Sullivan et al. (2000) concluded that MDD is a heritable trait, stating that genetic effects are the most important contributor to familial aggregation. Twin studies investigating concordance rates for MDD indicate a heritability of approximately 37%, and family studies indicate that first-degree relatives of probands have a two-fold to three-fold increase in lifetime risk of developing MDD (Lohoff, 2010).

Linkage and candidate gene studies

Family, twin, and adoption studies have also provided support for the genetic contribution to MDD, and a number of linkage and candidate gene studies were conducted in the 2000s in order to identify specialised loci and genes conferring risk to MDD. However, although this type of approach was successful in the investigation of rare, Mendelian disorders with high penetrance, no major loci of large effect were reported for MDD. These studies were largely underpowered, which may have played a role in the unsuccessful results. Border et al. (2019) recently investigated 18 genes that were empirically identified by such studies to have had an association with MDD. Using new well-powered samples (N_{range} = 62,138 – 443,264), the authors showed that none of the most highly investigated polymorphisms within these 18 genes demonstrated a significant genetic contribution to the liability of MDD (Border et al., 2019).

This, and additional studies described below, has provided additional support to the hypothesis that MDD is likely to be a polygenic disorder, with thousands of loci of minor effect contributing a fraction to the liability of the disorder. Moreover, MDD has a complex genetic architecture, indicating that different sets of susceptibility

genes, interacting with environmental risk factors, confer risk of MDD (Flint & Kendler, 2014).

Rare genetic variants and MDD

Aided by recent rapid advances in genetic analysis techniques, rare genetic variants have been increasingly investigated in relation to psychiatric disorders (Cook & Scherer, 2008; Dunn et al., 2015). Specifically, previous evidence indicates that copy-number variants (CNVs), inherited or *de-novo* segments of DNA that may affect gene function through deletion or duplication, may play a role in MDD.

In a study examining CNVs in 1,693 MDD cases and 4,506 controls, Glessner et al. (2010) found 12 CNV regions that occurred more frequently in MDD cases. Among these, the most significant locus was harboring the *SLIT3* gene, which is known to be implicated in axon guidance. Rucker et al. (2013) analysed copy number variation in 2,723 individuals with recurrent depression and 5,176 controls. They found that rare deletion CNVs, specifically genic and exonic, are enriched in recurrent depression cases as compared to controls (Rucker et al., 2013).

More recently, Kendall et al. (2019) investigated 53 CNVs previously associated with neurodevelopmental disorders in 407,074 individuals (23,979 MDD cases and 383,095 controls). They found that all 53 CNVs were associated with self-reported depression, however this association was partly explained by variables such as smoking status, physical health, and alcohol consumption. Zhang et al. (2019) conducted the largest genome-wide CNV study to date in a meta-analysis of four cohorts comprised of 5,780 MDD cases and 6,626 controls, finding an enrichment of short intergenic deletions in MDD patients. This suggests that CNVs may confer risk to MDD through the deletion of regulatory mechanisms.

The studies above indicate a role played by rare genetic variants in MDD. However, the association between CNVs and MDD risk remains largely unclear and CNVs do not replicate across studies, which may be due to small sample sizes in previous studies. As such, the study of rare genetic variants in relation to MDD is still in its infancy, and further research is needed to uncover their contribution to MDD risk.

Genome-wide association studies and MDD

Genome-wide association studies (GWAS) have been an important tool in investigating the genetic architecture of MDD, as they allow researchers to identify the genetic underpinnings of MDD by investigating the association between millions of SNPs across the genome without any a priori hypothesis about the function of a gene and the phenotype of interest (McCarthy et al., 2008).

MDD is a complex genetic trait with thousands of variants each contributing a small amount to the risk for disease. Until recently, MDD GWAS did not have sufficient sample sizes to detect what is now known to be the polygenic architecture of the trait. A GWAS mega-analysis for MDD found no genome-wide significant hits in the discovery sample (9,240 cases and 9,519 controls), replication sample (6,783 cases and 50,695 controls), or any other secondary analyses (Ripke et al., 2013). In 2015, Cai et al. (2015) found two genome-wide significant loci in 5,303 MDD recurrent cases and 5,337 controls.

The success of MDD GWAS only came to be realized once sample sizes massively increased. Wray et al. (2018) found 44 risk variants associated with MDD, using 135,458 cases and 344,901 controls (Wray et al., 2018). The most recent GWAS of MDD to date, a genome-wide meta-analysis of 807,553 individuals, has identified 102 independent variants associated with depression (Howard et al., 2019). An independent replication sample of 1,306,354 individuals showed that 87 of the 102 variants continued to be significant after multiple testing correction. Genes and gene-sets uncovered in this analysis showed an association with synaptic structure and neurotransmission, highlighting prefrontal brain regions as an important area for the study of MDD (Howard et al., 2019).

Downstream genetic and epigenetic analysis approaches

The increasing power of GWAS and large number of hits have allowed for further downstream analyses, and a number of cutting-edge approaches can be used to identify the underlying biology of MDD. Given the notorious heterogeneity of MDD, there is a need for disorder stratification in order to gain a deeper understanding of the environmental and genetic impacts on the disorder. A novel way to stratify MDD is through employing genetic approaches to investigate specific links to MDD. These include, but are not limited to, polygenic risk score calculation, pathway analysis and

expression quantitative trait loci analysis. Moreover, DNA methylation analysis may be carried out to examine the impact of environmental insults on the biology of MDD.

Given that most posited biological mechanisms implicated in MDD involve neural mechanisms and brain regions, there is a need to understand the impact of MDD genetic risk factors on the brain in order to identify neurobiological markers. As such, the approaches mentioned above and discussed in more detail in the sections below, may be explored in association with neuroimaging traits. Early literature did not initially provide conclusive evidence for an association between genetic risk factors for MDD and brain phenotypes, mainly due to scarce genetic-MDD associations and limited sample sizes (Reus et al., 2017; Wigmore et al., 2017). With increasing sample sizes, associations are becoming more evident (Schmaal et al., 2016; Elliott et al., 2018; Shen et al., 2019), which further highlights heterogeneous findings. This emphasizes the importance of leveraging other genetic approaches to examine these associations. Imaging phenotypes typically studied include white matter microstructure, subcortical volumes, cortical volume, surface area, and thickness, of which white matter demonstrates moderate heritability (Elliott et al., 2018). White matter microstructure, which forms the brain's connectivity network, may be a key neurobiological marker for MDD, although findings have so far been unclear and inconsistent (Whalley et al., 2013; Shen et al., 2017; Reus et al., 2017). A description and summary of white matter microstructure and its relationship with MDD as well as genetic risk for MDD to date is presented in the next section.

2. White matter microstructure

2.1 White matter microstructure

White matter, located beneath the grey matter cortex, comprises millions of myelinated axon bundles which connect neurons in different areas of the brain, travelling along tracts (Fields, 2010). These white matter tracts are structurally classified in terms of spatial connection within the brain. More specifically, projection fibres connect higher cortical areas to subcortical regions of the brain, such as limbic system structures (e.g. amygdala, thalamus), as well as the brain stem, cerebellum and spinal cord; association fibres connect cortical areas within the same hemisphere; and commissural fibres connect homologous areas between the two hemispheres (Jellison

et al., 2004).

The myelin surrounding the axons, which gives these nerve fibres a white colour, is necessary for high-speed transmission of electrical signals. Review articles report that damage to this may result in impaired cognitive, sensory, and motor functions (Fields, 2010). Furthermore, changes in white matter microstructure have been previously associated with both normal functioning, such as learning complex tasks (Scholz et al., 2009), and psychiatric and neurological disorders, such as schizophrenia, MDD, and Alzheimer's Disease (Nasrabady et al., 2018). These findings implicate white matter microstructure in behavioural changes, indicating that perhaps psychiatric disorders arise as a result of a connection deficit within the brain, rather than being confined to a single brain region.

Diffusion tensor imaging (DTI), a specialised Magnetic Resonance Imaging (MRI) technique, is the most common method used to measure white matter microstructure. DTI allows for the measurement of both architecture and integrity of white matter tracts in both healthy and disordered brains (Assaf & Pasternak, 2008). It does this by applying a tensor which measures the three-dimensional distribution of water molecule diffusion within voxels. Temperature, presence of large molecules, myelination, and microstructural barriers such as cell membranes and axon compaction all influence the mobility of water molecules (Beaulieu et al., 2002; Jones et al., 2013). Unlike cerebrospinal fluid, in which water diffusion is isotropic (i.e. water diffusion occurs equally in any direction), water diffusion in white matter occurs along tracts, meaning it is anisotropic. As opposed to a sphere indicating an isotropic diffusion distribution, the diffusion distribution in white matter then becomes an ellipsoid, in which the main axis is the principal eigenvector, while the second and third eigenvectors are oriented perpendicularly to it (ε 1-3). The amount of diffusion along each of these eigenvectors is quantified as eigenvalues (λ 1-3) (Figure 1) (Jellison et al., 2004 ; Gerrish et al., 2014).

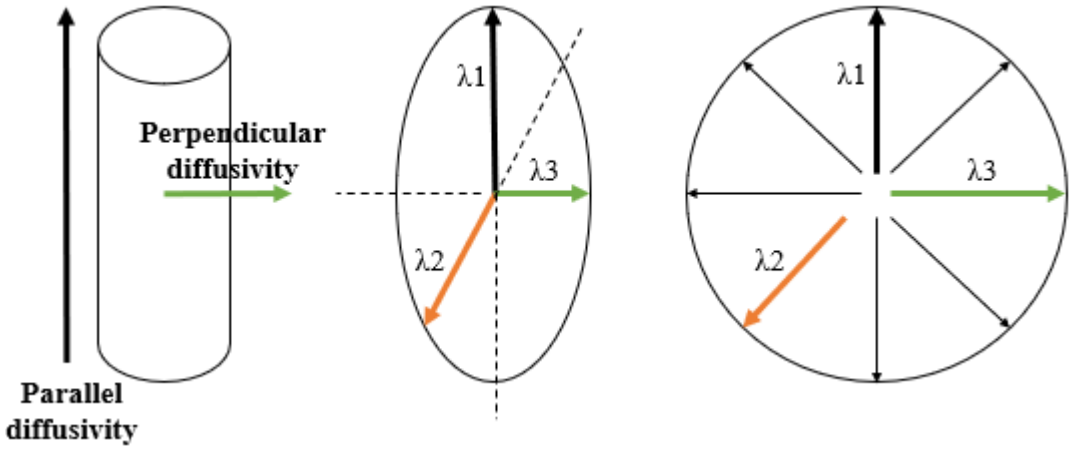


Figure 1. The diffusion tensor as a model of white matter microstructure. The figure was adapted from Jellison et al. (2004).

Two common DTI scalars are fractional anisotropy (FA) and mean diffusivity (MD), which can both be calculated using eigenvalues (Figure 2). FA measures the directionality of water diffusion from 0 (complete diffusion isotropy) to 1 (complete diffusion anisotropy). Generally, therefore, lower FA indicates decreased microstructural integrity of white matter and directionality, while higher FA represents increased white matter microstructural integrity. A major limitation of FA is crossing fibres, where different tracts with distinct orientations are present within an imaging voxel, which hamper accurate deterministic tractography of different tracts (Jbabdi et al., 2011). MD is calculated as an average of the eigenvalues and measures the magnitude of water molecule diffusion. Generally, higher MD indicates decreased white matter microstructural integrity, while lower MD indicates increased white matter microstructural integrity. Although crossing fibres affect FA more than they do MD, the scalar is sensitive to partial volume contamination in certain cases. For instance, ageing or specific disorders lead to loss of white and grey matter, which in turn may lead to cerebrospinal fluid contamination in white matter tracts which are spatially close to the ventricles (Metzler-Baddeley et al., 2011; Berlot et al., 2014).

Two additional DTI scalars providing more specific measurements of water diffusion are axial diffusivity (AD; ($\lambda 1$), which is the measurement of water molecule diffusion parallel to the tract, and radial diffusivity (RD; (($\lambda 2$ + $\lambda 3$)/2), which measures water diffusion perpendicular to the tract (Winklewski et al., 2018). The two measures may capture distinct tissue characteristics such as axonal degeneration (AD)

and demyelination (RD) and are also sensitive to issues such as crossing fibres and anisotropy decrease as a result of disorders (Alexander et al., 2008).

However, structural changes, such as demyelination, changes in neurite morphology, or increase/decrease in the dispersion of neurite orientation distribution, may contribute independently to variation within both FA and MD (Timmers et al., 2016). Newly developed measures such as neurite orientation dispersion and density imaging (NODDI) may provide additional information with regards to cellular contributors to FA and MD. NODDI provides estimates of neurite density through intra-cellular volume fraction (ICVF); extra-cellular water diffusion through isotropic volume fraction (ISOVF); and tract complexity or fanning and bending of axon bundles through orientation dispersion index (OD) (Zhang et al., 2012).

As NODDI measures may uncover additional sources of variation within FA and MD that cannot be distinguished using conventional DTI measures, there is increasing interest in using this method alongside FA and MD. Previous studies have shown they are sensitive in both healthy (Cox et al., 2016; Edwards et al., 2017) and clinical populations (Timmers et al., 2016; Rae et al., 2017), and may therefore provide more specific information with regards to changes in white matter microstructure.

Despite the limitations outlined above, FA and MD are microstructure variances that provide a more general measurement of water diffusion and directionality within white matter tracts, and have been shown to be valid and effective methods of white matter microstructure measurement (Jones et al., 2013; Shen et al., 2017). As the two DTI scalars are the most commonly reported measurements in previous studies (Jones et al., 2013), in the current thesis, findings concerning both FA and MD are presented.

Previous studies indicate that white matter microstructure is consistently heritable across tracts. Kochunov et al. (2015) investigated the heritability of FA in 481 participants, finding white matter tracts to be highly heritable, with approximately 70 – 80% of the total variance being explained by genetic factors in an additive manner. In addition, Vuoksimaa et al. (2017) examined the proportion of genetic and environmental influence on white matter microstructure, as measured by FA, MD, AD and RD, in 393 middle-aged twins, and found that genetic effects explained between 72 – 80 % of the variance in global measures of FA, although heritability differed

between individual tracts. This evidence suggests that white matter microstructure formation and maintenance is partially explained by genetic factors, enabling it as an important phenotype in downstream analyses of brain-related traits and disorders in relation to genetic information.

$$MD = \frac{\lambda 1 + \lambda 2 + \lambda 3}{3}$$

$$FA = \sqrt{\frac{(\lambda 1 - \mathrm{MD})^2 + (\lambda 2 - \mathrm{MD})^2 + (\lambda 3 - \mathrm{MD})^2}{2(\lambda 1^2 + \lambda 2^2 + \lambda 3^2)}}$$

Figure 2. Calculation of FA and MD (Alexander et al., 2007).

2.2 Probabilistic tractography and tract-based spatial statistics

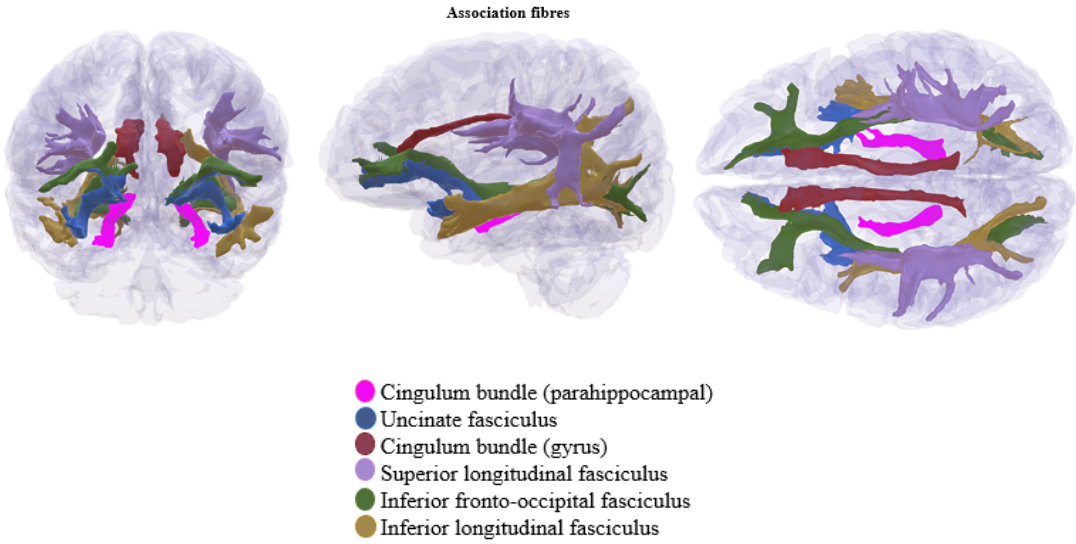
Tractography is a non-invasive method used to measure the apparent orientation and trajectory of white matter tracts *in vivo*. There exist numerous methods that allow for the characterisation of anatomical microstructure of white matter. In the present thesis, white matter tracts derived from two methods, probabilistic tractography and tract-based spatial statistics (TBSS), are presented.

Probabilistic tractography probes probability distribution of fibre orientations at each voxel. This allows for observing the probability of a given fibre moving along a specific path (Hagler et al., 2009). This method accounts for uncertainty in local fibre orientation and can reconstruct crossing fibres in a reliable way (Behrens et al., 2003; Hagler et al., 2009); however, the method is computationally demanding as it requires a large number of iterations, and prior anatomical knowledge of white matter microstructure organisation is required (Hagler et al., 2009). AutoPtx (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/AutoPtx), which is a set of scripts used to run probabilistic tractography, outputs 27 white matter tracts, 3 unilateral and 12 bilateral (Figure 3). In chapters 2 and 3, white matter tracts derived from probabilistic tractography were analysed.

Tract-based spatial statistics (TBSS; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS) aims to combine strengths of both voxel-wise and tractography-based methods (Smith et al., 2006; Yeh et al., 2009). Firstly,

subjects' FA raw images are non-linearly aligned to a standard brain space; the mean of the aligned images is then used to create a mean FA skeleton representing the centre of major white matter tracts which are common in all participants; lastly, each participant's FA data is then projected onto the mean FA skeleton, where their projected FA values are taken from the local centre of the tract in the original FA image (Smith et al., 2006). In addition to the method being less computationally intensive by reducing the number of tests it carries out (Smith et al., 2006), it also attempts to take into account issues such as tract alignment and pre-specification of tracts. Potential limitations include crossing fibres, as well as disease states, which might lead to exclusion or skewness of FA values due to, for example, reduction in grey matter volume, although images should be carefully examined during pre-processing stages in order to avoid this (Smith et al., 2006). In chapter 4, TBSS was used to derive 43 white matter tracts, 5 unilateral and 19 bilateral (Figure 4).

Although the two methods output different sets of white matter tracts, both are computationally valid and are based on connectivity and anatomical knowledge of the brain (Jones et al., 2013; Alfaro-Almagro et al., 2018).



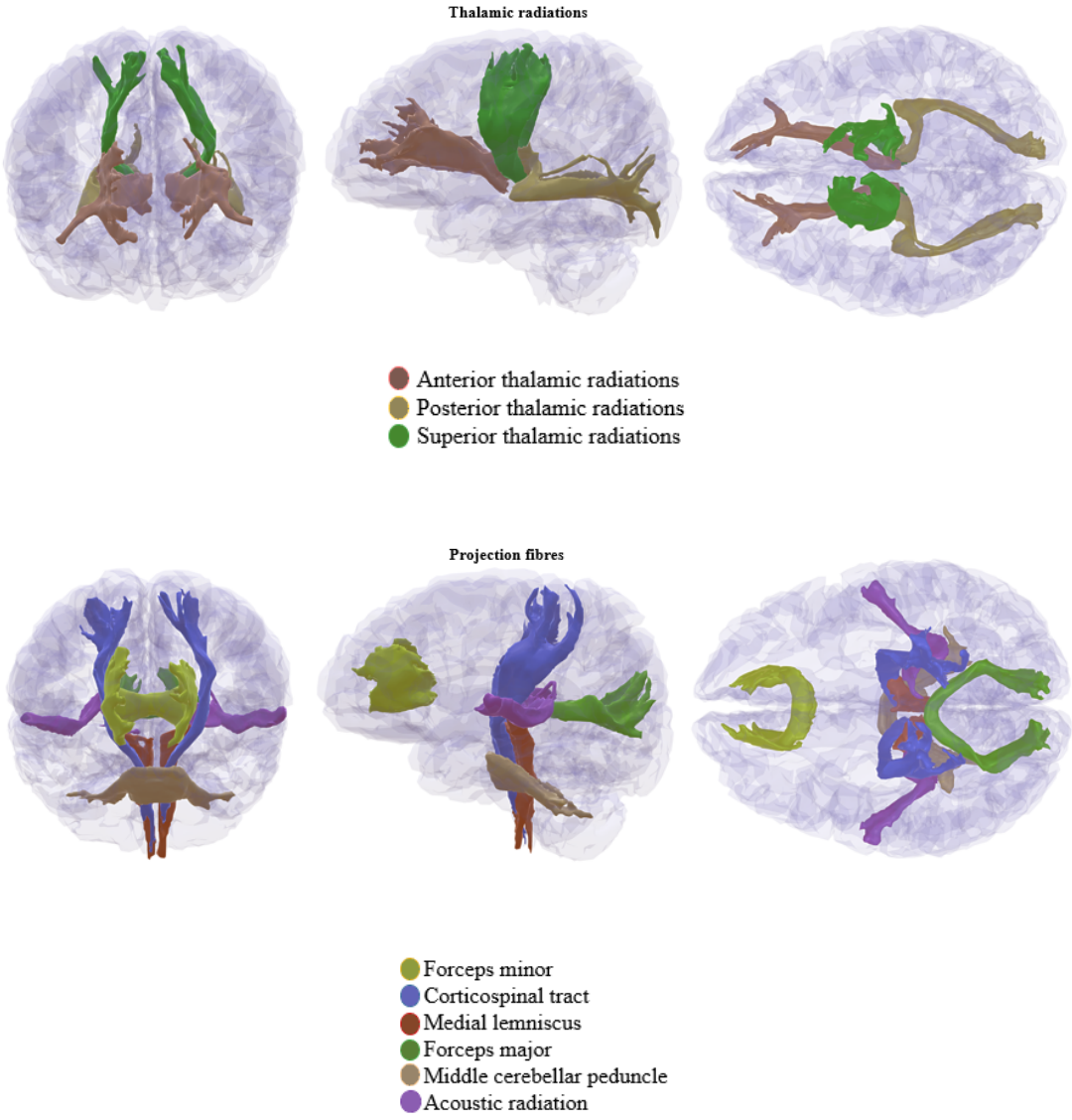


Figure 3. White matter tracts grouped in three tract categories output by AutoPtx. The images were created using Heatmapper (https://www.ccace.ed.ac.uk/research/software-resources/software).

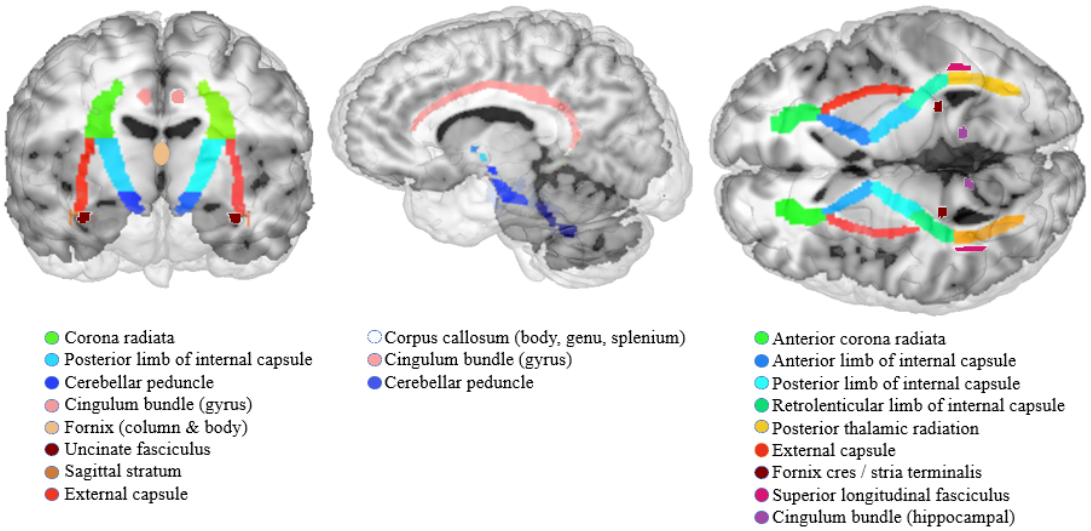


Figure 4. White matter tracts output by TBSS. The images were created using Mango (http://ric.uthscsa.edu/mango/).

2.3 White matter microstructure and MDD

White matter microstructural changes indicated by lower FA and higher MD have been associated with MDD in the past (Tham et al., 2011; Shen et al., 2017). In 2011, Tham et al. reviewed post-mortem, genetic, and neuroimaging studies of white matter microstructure abnormalities in MDD. Previous post-mortem studies mainly found white matter abnormalities in prefrontal brain regions characterised by decreases in oligodendrocyte density, a glial cell responsible for myelin production. In addition, myelin-associated genes important for processes such as axon guidance and growth, and synaptic function, were generally related to white matter abnormalities (Tham et al., 2011).

Neuroimaging studies generally reported lower FA within cortical and subcortical regions; in Tham et al.'s (2011) review, the frontal gyrus, superior longitudinal fasciculus (SLF), and the striatum were marked as specific affected tracts (Tham et al., 2011). A meta-analysis investigating DTI studies in connection to MDD found the SLF to be consistently abnormal in MDD patients as opposed to healthy individuals across studies (Murphy & Frodl, 2011). A further meta-analysis of DTI studies in MDD patients looked at research including case-control samples only. The authors found that tracts connecting the prefrontal cortex with cortical and sub-cortical areas were the most consistently identified fascicles in patients with MDD (Liao et al.,

2013).

These studies however were typically limited by sample size and heterogeneity, which hinders generalisability to wider population samples. All studies described above concluded that further analysis using much larger sample sizes would be needed in order to uncover links between genetic factors and specialised white matter tracts in MDD, as well as to identify genes which are implicated in white matter formation, maintenance, and pathology.

More recent empirical studies have attempted to address the above-mentioned limitations, and evidence exists linking lower FA and higher MD in numerous white matter tracts to MDD, both in affected individuals and those at high risk of the disorder. Whalley et al. (2013) investigated the association between white matter microstructure as measured by FA and individuals at high risk for mood disorders, quantified by a polygenic risk score (PRS) for bipolar disorder and MDD. With regards to MDD, they found a significant association between higher polygenic risk of MDD and lower FA within the parietal region of the superior and inferior longitudinal fasciculus, as well as thalamic radiations, uncinate fasciculus, and inferior fronto-occipital fasciculus (Whalley et al., 2013). In a case-control study, Shen et al. (2017) found global measures of FA, as well as thalamic radiations and association fibres, to be reduced in MDD patients as opposed to healthy individuals in a sample size of 1,087. Lower FA was also localised to individual white matter tracts, such as the left SLF, superior thalamic radiation, and forceps major. Van Velzen et al. (2019) investigated white matter anisotropy and diffusivity in 1,305 MDD cases and1,602 healthy controls across 20 samples worldwide as part of the MDD Working Group of the Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA). Within adult samples, they found significantly lower FA in MDD cases (N = 921) compared to healthy controls (N = 1,265) in 16 of the total 25 white matter tracts investigated, including parts of the corona radiata, corpus callosum, and superior and inferior fronto-occipital fasciculi. While no differences were found for AD and MD, global RD was found to be higher in MDD cases (Van Velzen et al., 2019).

Shen et al. (2019) also investigated cross-sectional and longitudinal measures of depressive symptoms and their association with white matter microstructure as measured by FA and MD in 18,959 individuals. They found that anterior thalamic

radiation was associated with all measures of depressive symptoms; several association fibre tracts, including superior and inferior longitudinal fasciculus, and projection fibre tracts, including acoustic radiation and corticospinal tract, were associated with cross-sectional measures of depressive symptoms (Shen et al., 2019).

In summary, the main findings to date indicate connections between the prefrontal cortex and sub-cortical areas, most notably the SLF and thalamic radiations. The SLF connects the frontal lobe to parietal, occipital, and temporal lobes (Schmahmann et al., 2007). As a result, it is associated with numerous higher-order cognitive functions, such as language, spatial working memory, attention, and emotion regulation (Vestergaard et al., 2011; Madhavan et al., 2014; Parkinson & Wheatly, 2014). Thalamic radiations connect the thalamus to anterior, superior, and posterior regions of the brain (Jones, 2002). The thalamus is a subcortical structure which plays an important role in sleep regulation, as well as cognitive processes such as attention, speed of information processing, and memory (Van Der Werf et al., 2001; Fama & Sullivan, 2015). White matter tracts connecting the thalamus with other cortical areas of the brain may therefore be implicated in these processes.

Deficits in these tracts may therefore reflect MDD symptomatology profiles such as insomnia or hypersomnia, inability to concentrate, mood disruptions and suicidal tendencies, providing a strong rationale for the investigation of white matter microstructure in relation to MDD (Coenen et al., 2012; Jia et al., 2014). Numerous causative paths may contribute to these symptoms, and novel opportunities allowing the combination of genetic approaches with neuroimaging traits may provide a deeper mechanistic understanding of the disorder. An overview of the genetic approaches used in this thesis is presented below.

3. Expression quantitative trait loci (eQTL)

Gene expression is the process by which genetic information is used to direct product synthesis, such as proteins for protein-coding genes, or transfer RNA, for non-protein coding genes. Within this process, some genes that produce proteins involved in important functions (i.e. breaking down glucose) are continuously expressed, while others may only be expressed as part of a specific process and at a particular time (e.g. cell differentiation) (Garcia-Sanchez & Marques-Garcia, 2016). Gene regulation, the

process that increases or represses gene expression, is vital in all living organisms, as it allows for the cell's control of structure and function, as well as cell differentiation. Moreover, it facilitates organisms' adaptability and evolution, as the cell has control over the amount of gene expression at a specific time and location (Wray, 2007).

Gene expression is one of the primary processes in converting information within the genome to observable phenotypes (Storey et al., 2007). As such, levels of expression may act as an intermediate phenotype between genetic information and observable traits, such as common diseases (McKenzie et al., 2014). Therefore, understanding the genetics of gene expression allows researchers to gain insight into the genetics of complex traits (Lee, 2018).

Expression quantitative trait loci (eQTL) are genetic variants that explain variation in gene expression, and have been characterised as *cis* (loci within 1 megabase from a gene's transcription start site) or *trans* (loci at least 5 megabases downstream or upstream of a gene's transcription start site, or on a different chromosome) (Nica & Dermitzakis, 2013) (Figure 5). GWAS of gene expression have been developed in order to identify polymorphic genetic loci influencing gene expression across the genome. Essentially, if a genotype at a specific locus is associated with an increase or decrease in the expression of a gene, this locus may act as a regulator, or eQTL (Michaelson et al., 2009), and different genotypes will lead to variation in phenotypes.

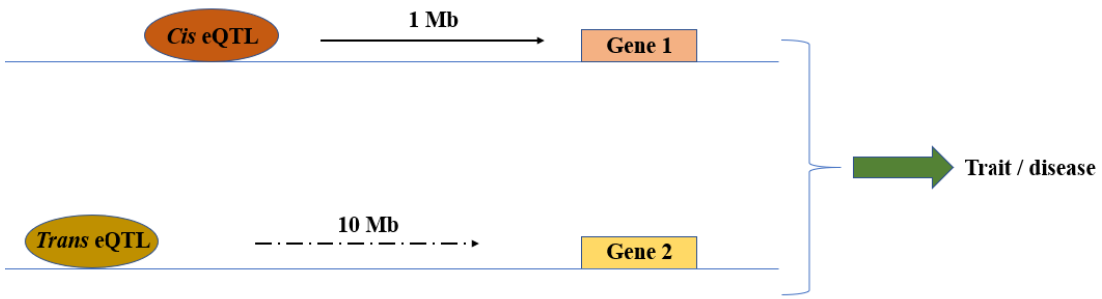


Figure 5. *Cis* and *trans* eQTL gene expression regulation. 1 Mb and 10 Mb represent the physical distance of the genome region and indicate the distance from each gene's transcription start site; the two types of eQTL regulate gene expression which in turn give rise to traits / diseases.

Findings from these methods indicate that eQTL may play a role in susceptibility to disease and may help elucidate the role of potential biological

pathways and gene sets in the manifestation of specific disorders. GWAS findings have identified significant associations between genetic variants and disease phenotypes, but our understanding of the molecular mechanisms underlying these associations is scarce. Numerous variants lie in non-protein coding regions of the genome, and therefore it could be that they influence traits through the regulation of gene expression (Fagny et al., 2017). Therefore, gaining insight into the links between eQTL and disease phenotypes may further the understanding of the causation, formation, and manifestation of these traits.

Previous studies have indicated that genetic variation may explain variance in level of gene expression in a tissue-specific manner. For example, O'Brien et al. (2018) mapped eQTL by performing deep RNA sequencing and genome-wide genotyping in 120 post-mortem foetal human brains in the second trimester, and identified eQTL conferring risk and gene expression changes mediating susceptibility to neuropsychiatric disorders, such as attention deficit hyperactivity disorder, schizophrenia, and bipolar disorder (O'Brien et al., 2018). In addition, Bhalala et al. (2018) conducted a multi-region meta-analysis to investigate whether SNPs previously associated with schizophrenia, bipolar disorder, and MDD are associated with gene expression in human brain tissue. To do this, they investigated SNPs associated with the three disorders in 11 GWAS of gene expression levels in post-mortem neurotypical brain tissue from two independent datasets, and identified 2,224 *cis* eQTL associated with expression of 40 genes (Bhalala et al., 2018).

Lastly, Zhong et al. (2019) integrated genetic associations from a recent MDD GWAS (Wray et al., 2018) and brain eQTL data to identify genes whose expression alteration may contribute to susceptibility of MDD. They found 18 genes whose perturbations may play a role in susceptibility to MDD, including *FLOT1*, whose expression was further upregulated in the brain and peripheral blood of a European sample of MDD cases, as compared to controls (Zhong et al., 2019). As shown by these studies, eQTL analysis may uncover putatively novel associations between gene expression and brain-related disorders, paving the way for further analyses implicating potentially new therapeutic targets.

However, using brain tissue in eQTL analysis poses several issues due to the nature of the tissue. Firstly, the brain is comprised of a number of cell types, hence

levels of expression throughout the brain are not likely to be uniform. Moreover, the brain is not accessible ante-mortem, and the use of tissue post-mortem introduces issues such as small sample sizes, cause of death, post-mortem interval, and gene expression differences in post-mortem as opposed to ante-mortem brains (McKenzie et al., 2014). These issues have led researchers to consider alternative approaches of investigating eQTL in relation to brain-related traits in more accessible tissues, such as whole blood (Qi et al., 2018).

Hernandez et al. (2012) sought to observe whether it is possible to use peripheral tissues such as blood to infer expression levels in the central nervous system. They examined 399 brain samples (frontal lobe and cerebral cortex) and 501 blood samples and found a small number of eQTL to be shared between the two tissues (brain and blood). They also found that some eQTL differed between the two tissues, which might be due to differences in pattern of gene expression (e.g. neuron-specific proteins being expressed). McKenzie et al. (2014) analysed eQTL overlap between 8 published brain studies and eQTL measured in blood in a large meta-analysis, finding that between 13 – 23% of eQTL overlapped between the two tissues. These studies suggest that where it is not possible to directly access the tissue of relevance, it is appropriate, with caution and an awareness of possible limitations, to use whole blood as a proxy.

Using peripheral samples to investigate gene expression levels in the brain introduces additional limitations that should be considered. Firstly, gene expression is tissue- and cell-specific, and evidence suggests that there is limited commonality among different tissue types. For instance, gene expression may be altered in both a tissue- and timing-specific manner, leading to different expression levels in peripheral tissues compared to brain (Hernandez et al., 2012). To address this, it is necessary to identify which gene expression patterns and biological processes are conserved between brain and peripheral tissues (Glatt et al., 2005).

Secondly, it is difficult to determine the overlap between brain and peripheral tissue eQTL, as data is often taken from different subjects; moreover, while brain samples are collected post-mortem, blood samples are collected in life. This in turn introduces further limitations, such as number of samples used, presence or absence of disease phenotype across samples, and different population demographics with

differential expression pattern, whose gene expression levels have been measured using different protocols (McKenzie et al., 2014).

Lastly, Sullivan et al. (2006) found a 0.5 correlation between transcripts present in whole blood and central nervous system tissues, arguing that whole blood gene expression may not be suitable for specific applications that require high tissue or transcription specificity, but may be applicable for specific sets of genes or biological pathways that are arguably expressed at a similar level across brain and peripheral tissues, or in the investigation of more general tissue-gene expression approaches (Sullivan et al., 2006).

Neuroimaging phenotypes, as measured by MRI, provide a unique opportunity to examine eQTL in association with *in vivo* brain phenotypes. Uncovering novel associations between gene expression patterns and brain structure may lead to further analysis and identification of loci that are of importance in psychiatric and neurological disorders, by linking genetic information to both specific brain regions and brain-related disorders. As both white matter microstructure and MDD are moderately heritable (Kochunov et al., 2015; Lohoff, 2010) and linked to each other (Shen et al., 2017), in Chapter 2, using genetic and neuroimaging data, the relationship between white matter phenotypes and eQTL genetic scores, previously measured in whole blood, is explored and analysed. The analysis provides evidence of changes in gene expression in relation to white matter microstructure, allowing for an insight into the relationship between previous brain- and disease-associated genes, their expression, and brain connectivity.

4. Polygenic risk scores, biological pathways, and MDD

4.1 Polygenic risk scores

GWAS allow for the identification of the genetic underpinnings of MDD reflected by the effect of multiple common genetic variants, lending support to further studies attempting to uncover the heritability of the disorder. Studies have attempted to uncover the functional impact of these variants, but a major issue in uncovering genetic links to MDD is that the variance explained by a single genetic variant is minor (Gandal et al., 2016). As such, a method of aggregating risk variants into a single variable was proposed in order to capture the additive effect of genetic variants for a

given trait, known as a PRS (Wray et al., 2008).

A PRS for a given trait is calculated from GWAS summary statistics, by summing the number of risk alleles carried by an individual in an independent dataset and weighting them by the effect size from the discovery GWAS (Euesden et al., 2015). The score can be calculated at any p-value threshold, which is chosen based on the trait it is calculated for (Euesden et al., 2015), and is described as a single continuous variable measuring genetic liability of a disorder. PRS may be used to predict an individual's risk of disease, differentiating between higher-risk and lower-risk individuals, with the average PRS being higher in cases than controls (Lewis & Vassos, 2017). PRS may also be used to investigate genetic links between two traits, by associating PRS for one trait with the phenotype for another trait. The International Schizophrenia Consortium (2009), for instance, calculated PRS for schizophrenia (3,322 cases and 3,587 controls) and were able to show that genetic risk for schizophrenia may explain some variance in bipolar disorder (1.9% and 1.4% in two independent samples).

The most direct application of PRS is to follow-up GWAS results by testing the prediction of case / control status in an independent study. However, PRS may be limited by the power of the original GWAS (Gandal et al., 2016). In MDD, GWAS results have increased in robustness and power over time, as described earlier. In the most recent MDD GWAS, Howard et al. (2019) found that PRS calculated from more than 800K individuals (246,363 cases and 561,190 controls) explained between 1.5 – 3.2% of the variance in MDD. Although this is a small proportion of the total variance explained, it is in line with previous findings concerning psychiatric disorders (Reginsson et al., 2018), and will likely further increase as sample sizes become larger and statistical methods improve.

In addition to being influenced by the original GWAS, there are a number of additional factors that may influence PRS accuracy. Firstly, heterogeneity between training and testing samples may have an effect on the accuracy of PRS. Secondly, there may be a lack of diversity in populations used to derive PRS, as the majority of genetic studies used for GWA consist of European populations. As disease-associated risk alleles may significantly differ in frequency between populations, this may lead to misestimations of disease risk when applied to populations other than those used in

the derivation of PRS (De La Vega & Bustamante, 2018).

Moreover, GWAS are typically only used to identify common genetic variants with a small contribution to disease risk. However, as mentioned previously, other genetic variants, such as copy number variants of high penetrance, may contribute to disease risk. Thus, individuals carrying rare, but not common risk alleles, may not show high genetic risk for a disorder according to PRS (Fullerton & Nurnberg, 2019). Lastly, PRS assumes that genetic risk conferred by common alleles of small effect is additive and does not yet consider complex epistatic relationships between risk variants, which may differ between individuals based on their genetic profile (Fullerton & Nurnberg, 2019).

Even with the success of recent GWAS, at the moment, PRS are unlikely to have clinical utility as a single variable. However, their usefulness may increase when associated and combined with environmental or other genetic risk factors, such as rare risk variants or DNA methylation (Lewis & Vassos, 2017). PRS calculated for MDD have previously been associated with traits of importance in MDD, including childhood trauma (Peyrot et al., 2014), depressive symptoms and psychological distress (Musliner et al., 2015), body mass index and obesity (Clarke et al., 2015), and the personality trait neuroticism (De Moor et al., 2015). These studies show that combining PRS for MDD with other known environmental risk factors may aid in increasing the variance explained in MDD, as well as uncovering interaction effects between genetic and environmental risk factors.

To date, there is scarce evidence of the association between PRS for MDD and brain-related phenotypes. This may be in part due to unsuccessful results from past GWAS, as well as the lack of, until recently, large datasets consisting of both genetic and neuroimaging data. Large cohorts such as UK Biobank (Sudlow et al., 2015) and Generation Scotland (Smith et al., 2006; Smith et al., 2013), which contain both types of data in a large number, have allowed researchers to gain novel insights into the association between the two.

The fact that white matter microstructure is moderately heritable and the most recent GWAS shows an enrichment of risk loci in brain regions (Howard et al., 2019) points to the importance of relating this phenotype to MDD genetic risk in order to inform risk prediction models. Whalley et al. (2013) investigated the association

between polygenic risk for individuals at high risk of mood disorders (MDD and bipolar disorder) and FA in 70 high-risk cases and 62 controls. They found a negative association between PRS for MDD and several white matter tracts, including SLF, inferior longitudinal fasciculus, and inferior fronto-occipital fasciculus (IFOF). Shen et al. (2019) provided a comprehensive analysis of PRS for MDD associated with 210 behavioural and 278 neuroimaging traits in a discovery (N = 10,674) and replication sample (N = 11,214). For white matter tracts as measured by FA and MD, they found MDD PRS to be associated with lower global white matter integrity, as well as regional tracts within association fibres and thalamic radiation. Several individual white matter tracts were also associated with higher MDD PRS, including lower FA in the SLF, posterior thalamic radiation, and forceps major; and higher MD in anterior and superior thalamic radiation, SLF, IFOF, cingulate gyrus, and forceps minor (Shen et al., 2019).

Findings from the above studies indicate that white matter microstructure is globally and regionally disrupted in those at higher genetic risk of MDD (Shen et al., 2017; Whalley et al., 2013). This evidence shows that MDD is perhaps a connectivity-based disorder and may not be localised to a specific brain region or tract. As such, stratification of genetic risk factors may be needed to uncover, for instance, risk conferred by variants localised to a specific genomic region, and whether these are associated with more specific white matter tracts based on their functionality, which is discussed below.

4.2 Biological pathway specific PRS

PRS may be further stratified in terms of biological functionality of loci and genes, clustered in biological pathways. Biological pathways are defined as a series of actions and reactions among molecules inside a cell which lead to a change in the cell (e.g. turning genes on and off, producing molecules such as proteins). Numerous databases have been created that aggregate and describe biological processes and structures in which genes and proteins are involved; this has been helpful in identifying gene-sets which take part in the same biological process and pathway (Khatri et al., 2012).

When running GWAS, conducting pathway analysis may provide additional

information about the combined effect and behaviour of multiple risk variants in relation to a trait or disease, allowing researchers to gain mechanistic insight into disrupted molecular and biological mechanisms in relation to disorders (Herold et al., 2012), as well as identify possible pharmacological targets (Sullivan & Posthuma, 2015). Sullivan and Posthuma (2015) conducted a review of 42 studies investigating biological pathways in five major psychiatric disorders, finding that biological pathways converge in schizophrenia and bipolar disorder, but not in other psychiatric disorders. In MDD, larger samples were needed at the time, reflecting the scarce GWAS results driven by a low sample size.

Methods to conduct pathway analysis, as well as ever-increasing sample sizes, have since provided more success in identifying biological pathways in relation to MDD. Howard et al. (2019) conducted pathway analysis on the MDD GWAS results using MAGMA (Multi-marker Analysis of GenoMic Annotation), a tool that identified genes in biological pathways, and investigated the significance of association between each pathway and depression utilizing p-values for each gene. Using pathway information from the Gene Ontology Consortium, they found 14 biological pathways enriched for depression, of which 8 were cellular components in the nervous system, and 6 were biological processes implicated in behaviour. This information aided in uncovering several biological pathways involved in depression, including enrichment in synaptic structure and activity, and response and behaviour to external stimuli (Howard et al., 2019).

Zeng et al. (2016) integrated regional heritability analysis and pathway analysis in order to identify MDD-specific biological pathways in two independent samples (Generation Scotland$_N$ = 6,455; Psychiatric Genomics Consortium MDD$_N$ = 18,759). Regional heritability analysis is generally applied in order to identify specific genomic regions which contribute a significant amount of heritability to a trait, more so than other regions. Pathway analysis is applied to identify related proteins within biological pathways in relation to traits of interest (Zeng et al., 2016). The authors found that of 1,035 biological pathways across numerous databases investigated, one pathway, the NETRIN1 Signaling Pathway, was significant in both datasets. In addition, Zeng et al. calculated pathway-specific PRS tailored to the NETRIN1 Signaling Pathway and compared these with whole-genome (minus SNPs within the pathway) PRS in the

prediction of MDD. They found that NETRIN1-specific PRS explained 0.216% of the variance in MDD, while whole-genome PRS explained 0.198%, indicating that this region alone explained more variance than PRS derived from the rest of the genome.

Their study provided evidence that biological pathway-tailored PRS may provide novel avenues for research in MDD and showed that there is rationale to investigate the relationship between brain phenotypes and the NETRIN1 Signaling Pathway, which is known to be implicated in thalamo-cortical axon guidance (Bonnin et al., 2007). As such, in Chapter 3, I investigate the association between MDD PRS calculated for SNPs within the NETRIN1 Signalling Pathway, as well as PRS calculated for SNPs outside the pathway, and global, regional, and individual white matter tracts. The study provides evidence of novel associations implicating functionally similar risk variants for MDD in white matter microstructure.

As sample sizes increase in cohorts combining neuroimaging and genetic data, it is necessary to examine novel links between specific genetic variants, biological pathways, and neurobiological factors in MDD. Despite its inherent heterogeneity, localised effects may be uncovered in relation to MDD. Investigating both genome-wide and pathway-specific PRS in white matter microstructure is therefore important for refining the genetic and biological mechanisms underlying MDD, and their effects on neurobiological phenotypes.

5. DNA methylation in MDD

In the context of traits and disease phenotypes, equally important to the human genome is a mechanism used by cells to determine at what point in space and time a gene is expressed. This mechanism is known as an epigenetic process, and is known to induce changes during cell division, such as altered patterns of gene expression within a specific cell type, thereby influencing the genome without changing its sequence (Tost, 2009).

DNA methylation (DNAm), one of many epigenetic processes, alters gene expression through the addition of methyl groups at cytosine-phosphate-guanine (CpG) sites, chemically changing DNA, and is situated at the intersection between genetic and environmental factors (Robertson, 2005). This process is led by DNA methyltransferases (DNMTs), which are the main family that catalyse the transfer of

methyl groups to DNA (Jaenisch & Bird, 2003). Approximately 75% of CpG dinucleotides are methylated at varying levels throughout the genome of mammals, marking DNAm as a promising biomarker in the context of differentially methylated CpG sites in association with phenotypes, at least partially (Tost, 2009). Fraga et al. (2005) found that twins' epigenomes are indistinguishable during early life, but as monozygotic twins advance in life, their DNAm signatures are different. Their results indicate that epigenetic mechanisms may provide an insight into how different phenotypes may arise even with the same genome as a starting point (Fraga et al., 2005).

Patterns in DNAm are particularly susceptible to change as a result of environmental stimuli, such as lifestyle factors. Joehanes et al. (2016) conducted a meta-analysis of previous studies investigating DNAm in relation to cigarette smoking behaviour in 15,907 individuals across 16 cohorts. They found a remarkable epigenome-wide influence on smoking, consisting of 18,760 CpG sites annotated to more than 7,000 genes. Mendelson et al. (2017) conducted an epigenome-wide association study (EWAS), the epigenome equivalent of a GWAS, of body mass index (BMI) in more than 3,700 individuals and led a replication analysis in a further 4,000 individuals. They found that BMI was associated with 83 differentially methylated CpG sites. Lastly, Liu et al. (2018) conducted an EWAS of alcohol consumption in 13,317 participants across 13 cohorts and identified 144 CpG sites highly predictive in the discrimination between heavy alcohol drinkers and non-drinkers. As the results of the above study indicate, differentially methylated CpG sites associated with various environmental factors may in future act as biomarkers to advance our understanding of molecular mechanisms implicated in the phenotypes.

Alterations in DNAm also exist in the manifestations of disease phenotypes (Robertson, 2005; Bergman & Cedar, 2013). Cancer is one of the most studied diseases in its relationship with epigenetic modifications. One acting mechanism is both hypomethylation (decrease of methylation) across the entire genome within tumours, and hypermethylation (increase of methylation) in specific regions and genes which act as tumour suppressors. This increase in promoter regions of tumour suppressor genes has been associated with transcriptional silencing, thus giving rise to tumour development (Jones & Baylin, 2002; Baylin, 2005).

Epigenetic modifications and their impact on complex psychiatric disorders have also been investigated. Gene-specific hypo- and hypermethylation has been found in schizophrenia, bipolar disorder, and autism spectrum disorder (Grayson & Guidotti, 2013; Klengel et al., 2014). DNAm alterations in relation to pathological states may act as the link between genotype and phenotype. Therefore, investigating the epigenetic impact on disease susceptibility loci may in future be of clinical and therapeutic relevance.

In recent years, DNAm has also been investigated in relation to MDD, and has been posited to play a role in the susceptibility of the disorder via dysregulation of gene expression catalysed by both environmental and genetic risk factors (Dalton et al., 2014). For instance, early life stress has been posited to act as a mechanism of lifelong changes in gene expression. Franklin et al. (2010) showed, in mice, that chronic and unpredictable situations where the mother is separated during the early post-natal timeframe leads to depressive-like symptoms and modifies the offspring's behavioural responses to novel environments as well as altered DNAm at several genes in the germline (Franklin et al., 2010). Their results indicate that early-life stress modifies behaviours and alters the epigenetic profile across generations through hypo- and hyper-methylation.

Studies investigating specific genetic loci have indicated several genes that may be of interest to depression from a DNAm standpoint. These include *BDNF*, which is known to regulate neuronal plasticity and neurotransmitter signalling (Roth et al., 2009); *SLC6A4*, which transmits serotonin from synaptic spaces to pre-synaptic neurons (Kang et al., 2013); and the glucocorticoid receptor gene *NR3C1*, which is important within the stress response system (Watkeys et al., 2018). Differential DNAm at these specific sites may elucidate specific links to MDD (Li et al., 2019).

Reviews examining the relationship between DNAm alterations and depression show that EWAS findings have not generally been replicated across studies, but this might be due to a number of factors, such as small sample sizes or heterogeneity of analyses (Dalton et al., 2014; Januar et al., 2015). Recently, Jovanova et al. (2018) ran an EWAS of depressive symptoms in a middle-aged and elderly sample of 7,948 individuals across 9 cohorts and attempted replication in an independent sample of 3,308 individuals in 2 further studies. They found 3 CpG sites

to be associated with depressive symptoms. These included sites at *CDC42BPB*, which plays a role in the regulation of cytoskeleton organisation, cell migration, and regulation of neurite outgrowth; *ARHGEF3*, which plays a role in axon guidance through co-expression with other gene families; and a third site situated in an intergenic region and is associated with *SEMA4B*, which in turn interacts with *PSD-52* to promote synapse maturation (Jovanova et al., 2018). All three CpG sites seem to be implicated in axon guidance, leading to conclude that this pathway may be disrupted in MDD.

Furthermore, Aberg et al. (2018) ran an EWAS of CpG-SNPs, defined as CpG sites which are created or destroyed by SNPs, to investigate whether they contribute to risk of MDD in 1,132 individuals (320 controls; 812 cases) and found 27 CpG sites that were suggestively associated with MDD. Among the key genes at these sites are *ASIC2*, which plays a role in neurotransmission; *DCC*, which is implicated in axon guidance and neurite outgrowth in developing neurons; and *ROBO2*, which also participates in axon guidance and cell migration (Aberg et al., 2018). Their findings complement those of Jovanova et al. (2018) and further confirm that the axon guidance pathway may be a putative disrupted pathway in MDD.

The findings from the studies described above indicate that DNAm plays an important role in MDD. However, research studies have been hindered by the complexity of DNAm, small sample sizes, and heterogeneity of analysis and phenotype, as well as hundreds of thousands of individual CpG sites across the epigenome. Therefore, investigation of DNAm through EWASs often poses the same issues as a GWAS study. As such, a DNAm risk score may be created, which acts in the same manner as a PRS. Such risk scores have shown to be successful in the investigation of other traits in the past.

For instance, Shah et al. (2015) investigated whether the contribution of DNAm profiles are associated with body mass index (BMI) and height independently of genotypic information (Shah et al., 2015). The authors first conducted an EWAS for both BMI and height in two independent cohorts ($N_{Discovery}$ = 1,366; $N_{Validation}$ = 750). They also calculated DNAm profile scores, a weighted sum of methylation level at associated CpG sites, in the validation dataset based on observed associations in the discovery dataset and vice versa, and determined whether the scores were associated

with the two traits in addition to PRS (Shah et al., 2015). They found that the DNAm score, PRS, and the two combined accounted for 7%, 8%, and 14% of the variance in BMI, respectively, in one of the cohorts, and 5%, 9%, and 13%, respectively, in the second cohort. The DNAm score did not account for much variation in height, which is consistent with previous literature indicating a larger genetic influence for height (Shah et al., 2015).

McCartney et al. (2018) used penalised regression models to train DNAm predictors for ten health and lifestyle factors, including BMI, total cholesterol, HDL cholesterol, LDL with remnant cholesterol, total:HDL cholesterol ratio, waist-to-hip ratio, percentage body fat, and self-reported alcohol consumption and smoking status (N = 5,087). They then developed DNAm scores and PRS for the ten traits in an independent sample (N = 895). They found that DNAm predictors explained a high proportion of variance in smoking (60.9%), medium proportion of variance in BMI, alcohol consumption, and HDL cholesterol (12.5 – 15.6%) and a small proportion of variance for the rest of the traits (0.6 – 4.5%). The DNAm scores and PRS additively explained the most variance in each trait (McCartney et al., 2018). The study showed that DNAm predictors are able to predict various traits as well as add to variance explained when combined with a genetic predictor, indicating a strong rationale to study DNAm scores in relation to other traits and disease phenotypes.

A DNAm score for MDD has recently been developed to investigate whether DNAm explains variance in both prevalent (N = 1,780) and incident (N = 1,607) MDD in an additive manner to PRS (Barbu et al., 2019). It was found that the DNAm score explained 1.75% and 0.52% of the variance in prevalent and incident MDD, respectively. In prevalent MDD, the combined DNAm score and PRS explained 3.99% of the variance. Furthermore, when accounting for lifestyle factors, including BMI, smoking status, pack years, and alcohol consumption, the DNAm score explained 0.68% of the variance, as opposed to 1.75% on its own. This indicates that the DNAm score effect is attenuated by lifestyle factors, however the score is still independent in its prediction of MDD. The study showed that there is rationale for investigating a DNAm score in relation to MDD, and provides a basis for relating DNAm scores for MDD in relation to other traits which might in turn be associated with the disorder.

White matter microstructure has previously been associated with DNAm

alterations at specific sites across the epigenome. A review detailing imaging genetic studies in MDD indicated previous studies which found altered DNAm at specific genes associated with structural changes in the brain. The genes included *SLC6A4*, where methylation was associated with hippocampal grey matter; *OXTR* methylation level, which was associated with amygdala responsiveness; DNAm at *NR3C1* and hippocampal volume; and *BDNF* methylation level, which was associated with anterior corona radiata structure alterations (Won & Ham, 2016).

This indicates that part of the effect of DNAm at specific loci and genes on MDD may be exerted through brain phenotypes. What is more, one of the key disrupted pathways uncovered in DNAm investigations of MDD is the axon guidance pathway. Two of the genes found by Aberg et al. (2018) participate in the NETRIN1 Signalling Pathway, which guides axons from the thalamus to other parts of the brain in neurodevelopment (Tang & Kalil, 2005), and which is investigated in the current thesis in terms of aggregated MDD genetic risk in relation to white matter microstructure. These findings combined indicate that an MDD DNAm risk score could have predictive ability in relation to white matter microstructure; a significant risk score-white matter association would aid in developing and determining neurobiological markers on which DNAm acts. This would have both clinical and therapeutic relevance, and could lead to advancements in the treatment and diagnosis of MDD. As such, in chapter 4, I investigate the association between a DNAm risk score and whole-genome PRS and global and individual white matter tracts, as measured by FA and MD. The study aids in advancing research relating to DNAm associated with both MDD and white matter microstructure.

6. Neuroimaging and genetic & epigenetic research – past studies and current thesis

Following the sequencing of the first human genome, the genetics field has led to important advances in the understanding of heritable traits. Most importantly, when investigating disease, genetic loci and genes offer a mechanistic insight of the disease and allow for the identification of high-risk individuals. Moreover, genetic variants aid in uncovering molecular and cellular processes acting within diseases. Similarly, neuroimaging technological advancements have provided a unique ability to identify

structural and functional processes within the brain, and complement the genetic approach by aiding in the identification of neural systems and brain circuitry (Hariri et al., 2006). Therefore, imaging genetics provide a unique opportunity to gain an understanding of biological, chemical, and molecular mechanisms, as well as specific pathways modulating variation in traits and disorders.

Neuroimaging genetics provide an avenue to investigate the structural and functional impact of polymorphisms on brain traits, ultimately leading to an understanding of aberrant or neurotypical behavioural manifestations. Due to the fact that genes give rise to both brain function and structure, responsible for the development of cognitive and behavioural processes, genetic variation may indirectly impact behavioural traits through neural systems. In this way, neuroimaging traits may act as endophenotypes, or the path from genotype to phenotype. Furthermore, mapping genetic variants in association with specific neural phenotypes also allows for the identification of candidate genes and their neural impact *in vivo* (Scharinger et al., 2010; Bigos & Weinberger, 2010).

GWAS have led to important discoveries in relation to numerous traits and diseases along the years. Although the method's clinical utility is still in its infancy, genetic variants identified through GWAS may serve as biomarkers for imaging phenotypes. For instance, Elliott et al. (2018) carried out GWAS for 3,144 imaging-derived phenotypes (IDP), covering the entire brain, including white matter connectivity, in more than 8,428 individuals. Of the total 3,144 phenotypes, 1,578 showed significant SNP heritability, indicating that brain traits are generally heritable. Within diffusion MRI, tractography-based IDPs generally showed lower heritability than tract-skeleton-based IDPs, indicating that different modalities and pre-processing pipelines may vary in their genetic underpinnings (Elliott et al., 2018).

For polygenic traits such as MDD, the amount of phenotypic variance explained by single SNPs is small, while a large number of SNPs is thought to underlie risk for complex disorders. PRS aggregate the contribution of a large number of SNPs, and can be used to test the genetic overlap between MDD and brain traits. In this way, novel associations between PRS and disorders may be uncovered by specific brain phenotypes (Dima & Breen, 2015). What is more, genetic research advancements now

allow for the use of pathway-based approaches to investigate functionally related SNPs aggregated within a single biological pathway in relation to brain phenotypes. Inkster et al. (2010) investigated a pathway contributing to risk of MDD, the Wnt signalling pathway, in relation to grey matter volume, in 1,022 MDD patients and 1,000 healthy individuals. They found that numerous polymorphisms within the genes showed genotype-by-MDD interactions with regional grey matter volume (Inkster et al., 2010). These findings lend support to the use of candidate pathway approaches in the investigation of neuroimaging phenotypes.

Lastly, epigenetic modifications are ideal candidates in the investigation of brain-related phenotypes, as they reflect the direct influence of environmental factors (Lancaster et al., 2018). As such, using neuroimaging traits to examine the relationship between DNAm, for instance, and aberrant and neurotypical traits and behaviours has become popular in the past years. However, unique challenges accompany this research approach, not dissimilar to GWAS approaches. For instance, there are approximately 28 million CpG sites along the human haploid genome, leading to difficulty in investigating each site (Lancaster et al., 2018). The use of a DNAm score has proved to be useful in the past, in terms of aggregating sites into a single, continuous measure.

In addition to challenges mentioned above, past studies have encountered a number of difficulties in analysing and interpreting findings. One of the most common issues is sample size; when investigating neuroimaging genetics within a complex disorder, such as MDD, this issue is threefold. Firstly, sample sizes used within studies may not reflect the general population, due to phenotypic heterogeneity of MDD; patients may have different symptom manifestations, making specific associations difficult to assess and interpret. Moreover, in the investigation of both genetic and neuroimaging data in MDD, very large samples are needed in order to account for genetic heterogeneity and number of neuroimaging phenotypes. Until recently, a combination of both types of data within the same individuals had not been achieved.

A further issue in the investigation of neuroimaging genetics in MDD is accuracy of inferences and assumptions when looking at GWAS downstream analyses. In this context, pathway-based approaches are powerful as they determine how

biologically informative findings stemming from combined neuroimaging and genetic data are. However, the accuracy of these inferences is limited by previous information used to identify the pathway and its relationship to the function it carries out.

Finally, DNAm and gene expression analyses pose unique challenges, as they are both dynamic, tissue- and cell-specific, and variable (Fazzari & Greally, 2004; McKenzie et al., 2014). Therefore, both data types can only be assessed and investigated within the physical brain post-mortem, which in turn introduces its own challenges, such as differential gene expression post- as opposed to ante-mortem, and a possible heterogeneous sample limited in size. Neuroimaging phenotypes therefore provide a novel, non-invasive method of investigating genetic and epigenetic impact on the brain *in vivo*.

Therefore, in the current thesis, I address some of the issues mentioned above by aiming to uncover links between white matter microstructure and differential gene expression, as well as to identify its role in relation to (1) genetic risk stratified by biological function and (2) whole-genome epigenetic risk of MDD. The overall aim of the thesis was to stratify genetic and epigenetic risk for MDD and identify novel genetic links to structural brain connectivity.

I utilise neuroimaging genetics approaches in two large projects, UK Biobank and Generation Scotland: Scottish Family Mental Health (GS:SFHS). UK Biobank is a large, population-based health resource aiming to prevent, diagnose, and treat numerous disorders by investigating genetic and environmental risk factors in middle and old age (Sudlow et al., 2015). The prospective study comprises 502,617 individuals aged 40-69 years whose genetic and environmental (e.g. lifestyle factors, medication intake) data were collected between 2006 and 2010 (http://www.ukbiobank.ac.uk/). A total of 488,363 individuals were genotyped using two arrays, the UK BiLEVE and the UK Biobank Axiom arrays (Bycroft et al., 2018). At the time of the current thesis, approximately 20,000 individuals have neuroimaging data across a number of modalities, including structural, diffusion, and functional. This number will in time increase to 100,000 participants, making UK Biobank a unique resource for investigating neurobiological markers of disease in association with genetic and environmental factors.

GS:SFHS is a family-based population study investigating the genetics of health and disease in approximately 24,000 individuals across Scotland aged 18 – 98 years, with baseline data collected between 2006 and 2011 (Smith et al., 2006; Smith et al., 2011). Data include environmental factors (e.g. lifestyle, medication intake) as well as genetic. Genome-wide DNAm data was also profiled from blood samples, marking GS:SFHS as one of the largest cohorts with available DNAm data. A subset of individuals, as part of Stratifying Resilience and Depression Longitudinally (STRADL), were followed-up, with the project aiming to further assess mental health, especially depression. Neuroimaging data was also collected for over 1,000 individuals within the STRADL subset (Navrady et al., 2017). The data make it possible for researchers to investigate neuroimaging phenotypes in relation to a vast amount of data, including DNAm, in a large number of individuals.

In the current thesis, I first start by investigating eQTL in relation to white matter microstructure in order to explore its genetic underpinnings. I applied a PRS derived from eQTL GWAS, with each score acting as a genetic proxy for the expression of a single gene. I found that expression scores of 8 genes were significantly associated with white matter microstructure after correction for multiple comparisons across scores and DTI metrics. More specifically, genes whose expression was linked to better white matter microstructural integrity were previously associated with developmental neural processes, such as neurite outgrowth; genes whose expression was linked to worse white matter microstructural integrity were previously associated with neuropsychiatric and neurological disorders (Chapter 2).

Having lent support to white matter microstructure being genetically linked to differential expression patterns, I next investigated genetic risk for MDD aggregated within a biological pathway, the NETRIN1 Signalling Pathway, which had previously been associated with MDD, and its relationship to white matter microstructure. I calculated PRS for SNPs within and outside the pathway, and compared the two PRS lists in their association with white matter tracts. Findings indicated that the PRS aggregated within the NETRIN1 pathway was associated with large tracts connecting frontal-to-occipital areas of the brain, such as the superior and inferior longitudinal fasciculus. Most interesting was its association with thalamic radiations, both

regionally and individually, as the biological pathway itself guides axons from the thalamus to the rest of the cortex in neurodevelopment (Chapter 3).

Finally, to investigate increased epigenetic risk for MDD, I calculated an epigenome-wide DNAm risk score as well as a genome-wide PRS and associated both with white matter microstructure. While both risk scores were associated with MDD, supporting previous findings of an epigenetic signature of MDD, neither was associated with white matter tracts, globally or individually (Chapter 4). The results indicated the need for larger sample sizes in neuroimaging epigenetic studies, reflecting a similar pattern to genetic fields, which may in future prove to be more successful. Finally, the thesis ends with a summary of the main findings, strengths and limitations, and directions for future study in chapter 5.

Chapter 2: Expression quantitative trait loci-derived scores and white matter microstructure in UK Biobank: a novel approach to integrating genetics and neuroimaging

1. Chapter Introduction

Previous GWAS of white matter microstructure reported it to be moderately heritable, indicating a genetic component contributing to white matter formation (Elliot et al., 2018). However, gene expression-based data has not previously been investigated in relation to white matter tracts in large sample sizes. Novel insight into expression changes in relation to brain connectivity may be gained and downstream analyses investigating brain-related traits and disorders (e.g. cognition, psychiatric disorders) may be interrogated as a result of this exploration. Therefore, the aim of the current study was to utilise a novel approach to identify genetic underpinnings of white matter microstructure, globally at whole-brain level, and with increasing regional specificity, in order to form a basis for future MDD genetic risk-associated studies in relation to brain connectivity. This chapter investigates the association between genetic proxies of gene expression for specific genes and white matter microstructure. In total, 6,457 eQTL scores, each representing the genetic profile of a single gene's expression, were calculated for N = 14,518 individuals with FA data and N = 14,485 individuals with MD data in UK Biobank. The study has been summarised in a manuscript entitled, "Expression quantitative trait loci-derived scores and white matter microstructure in UK Biobank: a novel approach to integrating genetics and neuroimaging", and is under review at *Translational Psychiatry* (https://doi.org/10.1101/646646). As the first author, I designed the experiment, carried out all the analyses, and wrote the manuscript for publication.

2. Manuscript

2.1 Abstract

Expression quantitative trait loci (eQTL) are genetic variants associated with gene expression. Using genome-wide genotype data, it is now possible to impute gene expression using eQTL mapping efforts. This approach can be used to analyse

previously unexplored relationships between gene expression and heritable *in-vivo* measures of human brain structural connectivity.

Using large-scale eQTL mapping studies, 6,457 gene expression scores (eQTL scores) were computed using genome-wide genotype data in UK Biobank, where each score represents a genetic proxy measure of gene expression. These scores were then tested for associations with two diffusion tensor imaging measures, fractional anisotropy (N_{FA}=14,518) and mean diffusivity (N_{MD}=14,485), representing white matter microstructural integrity.

FDR-corrected significant associations were found between 8 eQTL scores and structural connectivity phenotypes, including global and regional measures ($\beta_{absolute}$ FA=0.0339-0.0453; MD=0.0308-0.0381) and individual tracts ($\beta_{absolute}$ FA=0.0320-0.0561; MD=0.0295-0.0480). The loci within these eQTL scores have been reported to regulate expression of genes involved in various brain-related processes and disorders, such as neurite outgrowth and Parkinson's disease (*DCAKD*, *SLC35A4*, *SEC14L4*, *SRA1*, *NMT1*, *CPNE1*, *PLEKHM1*, *UBE3C*).

Our findings indicate that eQTL scores are associated with measures of *in-vivo* brain connectivity and provide novel information, not previously found by conventional genome-wide association studies. Although the role of expression of these genes regarding white matter microstructural integrity is not yet clear, these findings suggest it may be possible, in future, to map potential trait- and disease-associated eQTL to *in-vivo* brain connectivity and better understand the mechanisms of psychiatric disorders and brain traits, and their associated imaging findings.

2.2 Introduction

Expression quantitative trait loci (eQTL) are genetic variants which are proximally (cis) or distally (trans) associated with variation in the expression of genes (Nica & Dermitzakis, 2013). Previous animal and human studies have found that changes in gene expression lead to phenotypic variation, including adaptive phenotypic changes and evolutionary developments. In humans, for instance, cis-regulatory mutations lead to differences in lactase (*LCT*) gene expression, resulting in

lactase persistence in adulthood (Wray, 2007). With respect to psychiatric disorders, major depressive disorder (MDD) and bipolar disorder have been associated with decreased expression of prodynorphin messenger RNA (mRNA), which is involved in regulation of mood and expressed in limbic-related areas within the brain (e.g. amygdala, hippocampus) (Hurd, 1996; Hurd, 2002; Gandal et al., 2018). These findings indicate the importance of cis-regulatory mutations and variations in trait evolution.

Variation in gene regulation leads to differences in individual phenotypes, indicating that eQTL may play a role in susceptibility to disease (De Jong et al., 2012; Luo et al., 2015). To test this hypothesis, methods which combine gene expression data with genome-wide association studies (GWAS) summary statistics have been developed. These approaches may provide further insight into the potential causal pathways and genes involved in specific disorders, or predict the regulatory roles of single nucleotide polymorphisms (SNPs) in linkage disequilibrium (LD) with previously associated variants (Gilad et al., 2008). Previous studies have found that genetic variation may explain some of the variance in levels of gene expression in human tissues, including post-mortem brain tissue (Stranger et al., 2005; Hernandez et al., 2012; Ramasamy et al., 2014; Zhu et al., 2016). In one such study, Zou et al. (2012) conducted an expression genome-wide association study (eGWAS) on post-mortem brains of individuals with Alzheimer's disease (AD) and other brain pathologies (non-AD; including progressive supranuclear palsy). They found 2,980 *cis*SNPs associated with both AD and non-AD conditions. By investigating brain eQTL in post-mortem tissue therefore, researchers have been able to discover associations between gene expression and disease states in the brain.

Using brain tissue in order to investigate gene expression levels is however problematic, due to limitations such as small sample sizes and possible expression level differences in post-mortem versus ante-mortem brains (McKenzie et al., 2014). As such, alternative approaches have therefore been investigated. One such approach is using eQTL measured from whole blood gene expression as a proxy for brain gene expression; an approach supported by important benefits such as greater sample size and easier accessibility (Qi et al., 2018). Although it is recommended that wherever possible gene expression levels should be measured in a tissue-specific manner,

considerable overlap has been demonstrated between blood and brain eQTL, indicating the validity of the approach (McKenzie et al., 2014).

Neuroimaging measures provide a novel opportunity to investigate whether eQTL are significantly associated with *in vivo* brain phenotypes, and thereby increasing our knowledge of the role of eQTL in the wider context of psychiatric disorders. White matter microstructure, as measured by diffusion tensor imaging (DTI), is consistently heritable across tracts (Kochunov et al., 2015; Vuoksimaa et al., 2017; Sprooten et al., 2014) and is compromised in several psychiatric disorders. Generally, decreased microstructural integrity of white matter is characterised by lower directionality of water molecule diffusion (reduced fractional anisotropy, FA) and less constrained water molecule diffusion (increased mean diffusivity, MD). Consistent findings across studies have indicated higher MD and lower FA in individuals suffering from MDD, for example (Whalley et al., 2013; Shen et al., 2017). Investigating the regulatory loci associated with white matter microstructure in health and disease may aid in the detection of molecular mechanisms influencing disease through aberrant structural brain connectivity.

Within the current study, eQTL scores were derived based on two well-powered whole-blood eQTL studies (Westra et al., 2013; Gusev et al., 2016). GENOSCORES, a database of filtered summary statistics of publicly-available GWAS covering multiple phenotypes, including gene expression, was used to calculate eQTL scores (https://pm2.phs.ed.ac.uk/genoscores/).

The resultant eQTL-based genetic scores can be considered proxies for the expression of particular genes, which can then be tested for association with traits of interest. Here, their association with white matter microstructure as measured by FA and MD was analysed in UK Biobank using participants from the October 2018 UK Biobank neuroimaging release (N_{FA} = 14,518; N_{MD} = 14,485). The purpose of the study was to utilise a novel approach to investigate associations between regulatory SNPs and white matter microstructure. This approach could lead to further specialised investigation into psychiatric and neurological disorders, as well as other brain-related traits, such as cognition and behaviour.

2.3 Methods and materials

2.3.1 UK Biobank (UKB)

UK Biobank is a health resource aiming to prevent, diagnose and treat numerous disorders. It is comprised of 502,617 individuals whose genetic and environmental data (e.g. lifestyle, medications) were collected between 2006 and 2010 in the United Kingdom (http://www.ukbiobank.ac.uk/). UKB received ethical approval from the Research Ethics Committee (reference: 11/NW/0382). This study has been approved by the UKB Access Committee (Project #4844). Written informed consent was obtained from all participants.

2.3.2 Study population – neuroimaging measures

In the current study, individuals were excluded if they participated in studies such as the Psychiatric Genomics Consortium (PGC) MDD GWAS or Generation Scotland (Scottish Family Health Study) to remove overlap of genetic samples.

From the total of 502,617 individuals participating in UK Biobank, a subset was invited to attend neuroimaging assessments following the initial appointment. A total of 14,506 individuals who were part of the latest UK Biobank neuroimaging release (May 2018) were used in the current chapter. The age at the imaging assessment here ranged from 44.58-80.25 (mean: 62.69 +/- 7.48), of which 47.91% were men.

The current study used two DTI scalars, FA and MD. DTI data pre-processing and quality checking included correction for eddy currents and head motion in the scanner, outlier-slices correction, as well as grand distortion correction. FA maps were used to generate tract masks, using probabilistic tractography analysis as part of the AutoPtx package in FSL (Mori et al., 2002). A total of 27 tracts were generated, of which 12 were bilateral and 3 unilateral; weighted mean FA and MD were then calculated for each tract and these were used as variables in the current chapter.

Images were acquired, pre-processed, and quality controlled by UK Biobank using FMRIB Software Library (FSL) packages through a standard protocol (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1977). All data inconsistent with

scanner settings and that did not pass initial quality control were excluded from current analyses (Alfaro-Almagro et al., 2018). Moreover, individuals whose global measures for FA and MD lay more than three standard deviations from the sample mean were excluded (Shen et al., 2017; Barbu et al., 2019). This resulted in 14,518 individuals with FA values (N_{female} = 7,561 (52%); N_{male} = 6,957 (48%); mean age: 63.14 +/- 7.4; age range: 45.92 – 80.67) and 14,485 individuals with MD values (N_{female} = 7,552 (52%); N_{male} = 6,933 (48%); mean age: 63.12 +/- 7.39; age range: 45.92 – 80.67).

Tables 1 and 2 below detail general mental health for all individuals with FA and MD values, as taken from the mental health questionnaire administered to all UK Biobank participants (http://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=137).

	Ever sought/received professional help for mental distress	Ever suffered mental distress preventing usual activities
FA		
Yes	4,180 (M=1,490)	3,603 (M=1,313)
No	6,481 (M=3,481)	6,957 (M=3,596)
Do not know	18 (M=10)	121 (M=71)
Prefer not to answer	13 (M=6)	11 (M=7)
MD		
Yes	4,175 (M=1,486)	3,598 (M=1,308)
No	6,462 (M=3,469)	6,939 (M=3,586)
Do not know	18 (M=10)	120 (M=70)
Prefer not to answer	13 (M=6)	11 (M=7)

Table 1. Mental distress reported with the on-line mental health questionnaire; M=male; column headers indicate questions asked in the questionnaire; mental health data is not available for all participants with FA and MD measures.

	Mental health problems ever diagnosed by a professional
FA and MD ($Total_n$=7)	
Female	Depression (N=1) Psychological over-eating or binge-eating (N=1) Anxiety, nerves or generalized anxiety disorder (N=1) Agoraphobia (N=1) ADD/ADHD (N=1)
Male	Anxiety, nerves or generalized anxiety disorder (N=2)

Table 2. Mental health conditions present within both FA and MD samples; 7 individuals with FA (N=14,518) and MD (N=14,485) have previously been diagnosed with mental health conditions.

2.3.3 Genotyping and eQTL score calculation

A total of 488,363 UKB blood samples (N female = 264,857; N male = 223,506; http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22001) were genotyped using the UK BiLEVE array (N = 49,949; http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149600) and the UK Biobank Axiom array (N = 438,417; http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149601). Details of genotyping and quality control are described in more detail by Hagenaars et al. (2016) and Bycroft et al. (2017).

From GENOSCORES, eQTL analysis summary statistics from two studies of whole-blood eQTL were used (Westra et al., 2013; Gusev et al., 2016). Briefly, Gusev et al. (2016) developed a novel approach aimed at identifying associations between gene expression and complex traits in cases where gene expression level is not directly measured. These authors reported eQTL based on a sample of 1,414 individuals with whole-blood expression measured using the Illumina HumanHT-12 version 4 Expression BeadChip. Westra et al. (2013) performed a large eQTL meta-analysis in 5,311 samples across 7 studies from peripheral blood, with gene expression measured using Illumina whole-genome Expression BeadChips (HT12v3, HT12v4 or H8v2 arrays). Their aim was to investigate the magnitude of the effect of cis and trans SNPs on gene expression, as well as to observe whether mapping eQTL in peripheral blood could uncover biological pathways associated with complex traits and disease. Further details of data acquisition and protocols are described in more detail in the two studies (Westra et al., 2013; Gusev et al., 2016).

Before being imported into the GENOSCORES database, summary statistics were filtered at a liberal p-value < 1E-4 (0.0001). A total of 10,884 eQTL scores (N Gusev study = 3,801; N Westra study = 7,083) were computed for individuals included in the imaging sample (N_{FA}: 14,518; N_{MD}: 14,485) from the SNPs found in GENOSCORES, using a p-value threshold of 1E-5 (0.00001). Overlapping eQTL scores between the two studies (i.e. scores for which SNPs affect expression of the same gene in both studies) were then excluded by only including the score where a SNP had the lowest p-value, i.e. most significant association. The final eQTL score list was 6,457 (N Gusev study = 3,286; N Westra study = 3,171). These scores were

used as input variables in subsequent statistical analyses (Appendix 1: Figure S4 provides a summary of the score derivation process).

Briefly, eQTL scores were computed as a sum of the genotypes for an individual (g, scored as 0, 1, 2 copies of the reference allele) weighted by the effect size estimate (βt) for the trait of interest *t.* In order to adjust for LD, vector βt was pre-multiplied by the generalized inverse of the SNP-SNP correlation matrix R estimated from the 1000 Genomes reference panel, limited to the individuals with European ancestry.

The formula to compute the eQTL score for trait *t* for an individual (*i*) is therefore:

$$score(i,t) = g_i \beta_t R^{-1}$$

2.3.4 Magnetic resonance imaging (MRI) acquisition

In the current study, imaging-derived phenotypes (IDPs) produced by UKB were used. MRI acquisition and pre-processing procedures for white matter tracts were performed by UKB using standardised protocols (https://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf). Briefly, images were acquired in Manchester (N_{FA} = 12,248; N_{MD} = 12,221) and Newcastle (N_{FA} = 2,270; N_{MD} = 2,264) on a standard Siemens Skyra 3T scanner with a 32-channel radio-frequency (RF) receive head coil and later pre-processed using the FMRIB Software Library (FSL), and parcellation of white matter tracts was conducted using AutoPtx (Alfaro-Almagro et al., 2018). Individual white matter tracts belonging to each tract category can be observed in Appendix 1, Table S13.

Owing to the fact that head position and RF coil in the scanner may affect data quality and subsequent pre-processing, three scanner brain position variables were also generated by UKB, with the aim of being used as confounding variables in subsequent analyses. These are lateral brain position – X (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25756), transverse brain position – Y (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25757) and longitudinal brain

position – Z (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25758). The three variables were included as covariates in the statistical analysis described below.

2.3.5 Statistical methods

All analyses were conducted using R (version 3.2.3) in a Linux environment. Generalised linear mixed models (function "lme" in package "nlme") were used for bilateral brain regions, which were included as dependent variables. The eQTL scores were included as independent variables separately in each model, with additional covariates: age, age^2, sex, fifteen genetic principal components to control for population stratification, three MRI head position coordinates, MRI site and genotype array, while hemisphere was included as a within-subject variable. For unilateral tracts, as well as global measures and white matter tract categories of FA and MD, also included in the models as dependent variables, a general linear model (function "lm") was used, using the same covariates as above, without hemisphere included as a separate term, and again including the eQTL scores as independent variables separately in each model.

For global measures and white matter tract categories of FA and MD, principal component analysis (PCA) was applied on the white matter tracts of interest (all 27 for global measures; 12 for association fibres; 6 for thalamic radiations; 9 for projection fibres) in order to extract a latent measure. Scores of the first unrotated component were extracted and set as dependent variables in general linear models. False discovery rate (FDR) correction using the "p.adjust" function in R ($q < 0.05$) was applied across the eQTL scores and the individual white matter tracts ($N_{tests} = 98,855$), and separately across eQTL scores and global and tract categories ($N_{tests} = 25,828$).

2.4 Results

There were several eQTL scores that showed significant associations with a number of global measures, tract categories, and white matter tracts post FDR correction (Table 3; Figure 1a & 1b and Figure 2a & 2b; Appendix 1: Tables S5 – S12). In total, 25 scores were significantly associated with FA values ($\beta_{absolute} = 0.0320$-

0.0561) and 24 scores with MD values ($\beta_{absolute}$ = 0.0295-0.0480) in several tracts (these are fully detailed in Appendix 1: Tables S1, S2, S3 and S4; Figure S1 and S2). Among these scores, 8 were associated with white matter tracts measured by both FA and MD. The primary findings reported in this thesis section focus on these 8 overlapping scores (consistent with the submitted paper), as these were considered to provide the most consistent information with regards to gene expression within white matter tracts as measured by two different DTI scalars (see tables 4 and 5).

Score name & eQTL type	N SNPs in score	Regulated gene	Study from which score is calculated	Gene function
DCAKD_eQTL_cis	8	DCAKD	Gusev et al.	Expressed in glioma; ubiquitous expression in brain; implicated in a number of psychiatric and neurological disorders (Latourelle et al., 2012; Gonzalez-Lozano et al., 2016; Schizophrenia Working Group, 2018; Butler et al., 2015)
SLC35A4_eQTL_cis	12	SLC35A4	Gusev et al.	Expressed in brain (Sosicka et al., 2017)
SEC14L4_eQTL_cis	1	SEC14L4	Westra et al.	Specific function not yet determined; may be implicated in neurodegeneration (Curwin et al., 2008)
SRA1_eQTL_cis	15	SRA1	Westra et al.	Involved in regulation of many NR (nuclear receptor) and non-NR activities (e.g. chromatin organisation); may be associated with idiopathic hypogonadotropic hypogonadism (Kotan et al., 2016; Bianco et al., 2009)
NMT1_eQTL_cis	7	NMT1	Westra et al.	Ubiquitous expression in brain; may be implicated in brain tumours (Deng et al., 2018; Lu et al., 2005; Ducker et al., 2005)
CPNE1_eQTL_cis	1	CPNE1	Westra et al.	May regulate molecular events at the interface of the cell membrane and cytoplasm; expressed during brain development and implicated in neurite outgrowth in rats (Kim et al., 2018; Park et al., 2012; Park et al., 2014)
PLEKHM1_eQTL_cis	5	PLEKHM1	Gusev et al.	Protein encoded by this gene is important for bone resorption; may play critical

				role in vesicular transport in the osteoclast (Fujiwara et al., 2016; McEwan et al., 2015)
UBE3C_eQTL_cis	4	UBE3C	Westra et al.	Expressed in brain; may be implicated in Parkinson's disease (Garriock et al., 2010; Filatova et al., 2014)

Table 3. Information regarding eQTL scores with significant associations for both FA and MD-measured tracts.

The effect of the 8 scores on FA measures of white matter microstructure

Score, White Matter Tracts	Effect size, β	SD	t value	p value	p value, FDR corrected
DCAKD eQTL score					
Global FA	-0.0367	0.0079	-4.6474	3.39E-06	0.0088
Thalamic radiations	-0.0403	0.0080	-5.0378	4.77E-07	0.0025
Superior longitudinal fasciculus (SLF)	-0.0386	0.0077	-5.0327	4.89E-07	0.0040
Anterior thalamic radiations (ATR)	-0.0429	0.0076	-5.6798	1.37E-08	0.0002
Forceps minor	-0.0471	0.0078	-6.0115	1.88E-09	0.0001
SLC35A4 eQTL score					
Global FA	-0.0403	0.0079	-5.0996	3.45E-07	0.0022
Association fibres	-0.0347	0.0079	-4.4036	1.07E-05	0.0198
Projection fibres	-0.0453	0.0079	5.7612	8.52E-09	0.0002
Corticospinal tract	-0.0326	0.0074	-4.3945	1.12E-05	0.0337
Acoustic radiation	-0.0326	0.0069	-4.7044	2.57E-06	0.0133
Inferior longitudinal fasciculus (ILF)	-0.0335	0.0076	-4.3887	1.15E-05	0.0337
Superior longitudinal fasciculus (SLF)	-0.0367	0.0077	-4.7887	1.69E-06	0.0103
Forceps minor	-0.0561	0.0078	-7.1754	7.56E-13	7.32E-08
SEC14L4 eQTL score					
Global FA	-0.0420	0.0079	-5.3199	1.05E-07	0.0011
Association fibres	-0.0358	0.0079	-4.5425	5.60E-06	0.0121
Thalamic radiations	-0.0388	0.0080	-4.8429	1.29E-06	0.0048
Projection fibres	-0.0416	0.0079	5.2850	1.27E-07	0.0011
Corticospinal tract	-0.0320	0.0074	-4.3116	1.63E-05	0.0416
Posterior thalamic radiation	-0.0352	0.0075	-4.7014	2.61E-06	0.0133
Superior longitudinal fasciculus (SLF)	-0.0392	0.0077	-5.1143	3.19E-07	0.0028
Inferior longitudinal fasciculus (ILF)	-0.0419	0.0076	-5.4773	4.39E-08	0.0006
Forceps minor	-0.0456	0.0078	-5.8270	5.76E-09	0.0001
SRA1 eQTL score					
Projection fibres	-0.0339	0.0079	4.3032	1.69E-05	0.0273
Forceps minor	-0.0462	0.0078	-5.8981	3.76E-09	0.0001

NMT1 eQTL score					
Anterior thalamic radiations (ATR)	0.0324	0.0076	4.2863	1.83E-05	0.0429
Forceps minor	0.0352	0.0078	4.4956	6.99E-06	0.0271
CPNE1 eQTL score					
Forceps minor	0.0338	0.0078	4.3185	1.58E-05	0.0416
Forceps major	0.0436	0.0081	5.3818	7.49E-08	0.0009
PLEKHM1 eQTL score					
Forceps minor	-0.0347	0.0078	-4.4321	9.40E-06	0.0337
UBE3C eQTL score					
Forceps minor	-0.0382	0.0078	-4.8721	1.12E-06	0.0077

Table 4. Significant associations between eQTL scores and global measures, category, and individual white matter tracts (FA); the first column indicates standardised effect size (β); FDR = false discovery rate; for each score, tracts are arranged from global to individual tracts.

The effect of the 8 scores on MD measures of white matter microstructure

Score, White Matter Tracts	Effect size, β	SD	t value	p value	p value, FDR corrected
DCAKD eQTL score					
Global MD	0.0404	0.0075	5.3762	7.72E-08	0.0015
Thalamic radiations	0.0327	0.0072	4.5625	5.09E-06	0.0132
Association fibres	0.0381	0.0077	4.9643	6.97E-07	0.0045
Acoustic radiation	0.0295	0.0069	4.2470	2.18E-05	0.0472
Uncinate fasciculus	0.0314	0.0068	4.6086	4.09E-06	0.0162
Cingulate gyrus	0.0352	0.0073	4.7887	1.69E-06	0.0085
Inferior longitudinal fasciculus (ILF)	0.0377	0.0073	5.1766	2.29E-07	0.0024
Anterior thalamic radiations (ATR)	0.0403	0.0070	5.7964	6.92E-09	0.0003
Inferior fronto-occipital fasciculus (IFOF)	0.0410	0.0075	5.4805	4.31E-08	0.0008
Superior longitudinal fasciculus (SLF)	0.0415	0.0076	5.4902	4.08E-08	0.0008
Forceps minor	0.0480	0.0076	6.3085	2.89E-10	2.76E-05
SLC35A4 eQTL score					
Global MD	0.0308	0.0075	4.0893	4.35E-05	0.0423
Inferior longitudinal fasciculus (ILF)	0.0362	0.0073	4.9676	6.86E-07	0.0044
Forceps minor	0.0432	0.0076	5.6773	1.39E-08	0.0004
SEC14L4 eQTL score					
Global MD	0.0326	0.0075	4.3299	1.50E-05	0.0277
Cingulate gyrus	0.0328	0.0073	4.4648	8.07E-06	0.0248
Acoustic radiation	0.0339	0.0069	4.8778	1.08E-06	0.0060
Forceps minor	0.0348	0.0076	4.5604	5.15E-06	0.0188
SRA1 eQTL score					
Forceps minor	0.0353	0.0076	4.6349	3.60E-06	0.0155

NMT1 eQTL score					
Global MD	-0.0328	0.0075	-4.3626	1.29E-05	0.0257
Inferior longitudinal fasciculus (ILF)	-0.0311	0.0073	-4.2695	1.97E-05	0.0447
Inferior fronto-occipital fasciculus (IFOF)	-0.0335	0.0075	-4.4845	7.36E-06	0.0234
Anterior thalamic radiations (ATR)	-0.0339	0.0070	-4.8703	1.13E-06	0.0060
Superior longitudinal fasciculus (SLF)	-0.0343	0.0076	-4.5355	5.79E-06	0.0204
Forceps minor	-0.0392	0.0076	-5.1537	2.59E-07	0.0025
CPNE1 eQTL score					
Global MD	-0.0366	0.0075	-4.8650	1.16E-06	0.0050
Association fibres	-0.0368	0.0077	-4.7868	1.71E-06	0.0063
Inferior longitudinal fasciculus (ILF)	-0.0309	0.0073	-4.2303	2.35E-05	0.0497
Superior longitudinal fasciculus (SLF)	-0.0356	0.0076	-4.7055	2.56E-06	0.0116
PLEKHM1 eQTL score					
Global MD	0.0330	0.0075	4.3859	1.16E-05	0.0250
Thalamic radiations	0.0296	0.0072	4.1282	3.68E-05	0.0423
Association fibres	0.0318	0.0077	4.1395	3.50E-05	0.0423
Forceps minor	0.0334	0.0076	4.3876	1.15E-05	0.0297
Superior longitudinal fasciculus (SLF)	0.0342	0.0076	4.5223	6.17E-06	0.0210
Anterior thalamic radiations (ATR)	0.0356	0.0070	5.1101	3.26E-07	0.0028
UBE3C eQTL score					
Forceps minor	0.0331	0.0076	4.3465	1.39E-05	0.0349
Inferior fronto-occipital fasciculus (IFOF)	0.0332	0.0075	4.4413	9.01E-06	0.0268

Table 5. Significant associations between eQTL scores and individual white matter tracts (MD); the first column indicates standardised effect size (β); FDR = false discovery rate; for each score, tracts are arranged from global to individual tracts.

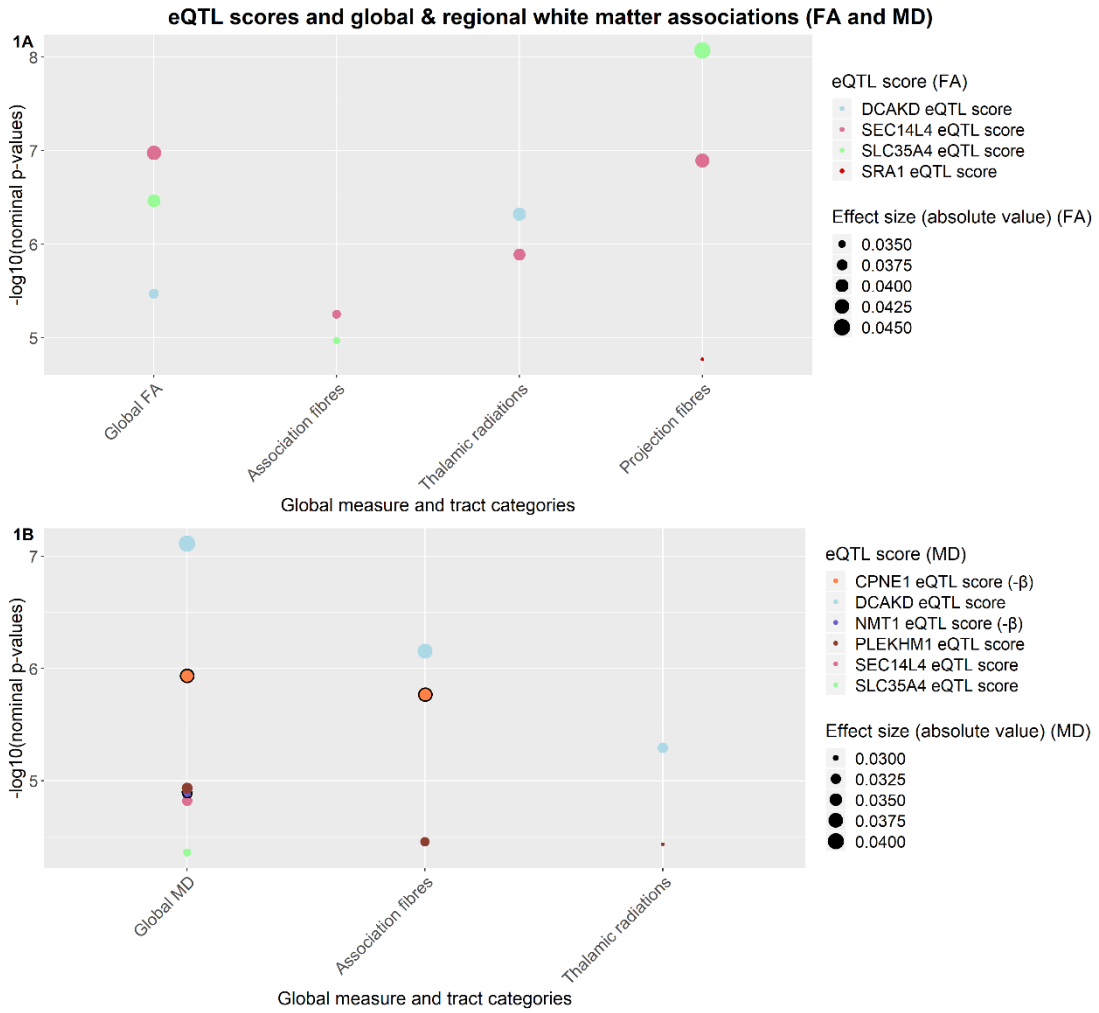


Figure 1 (A and B). Indicates nominal p-values between each of the 8 scores (shown in legend entitled "eQTL score") and global and tract category measures (noted on the x-axis; FA = fractional anisotropy (figure 1A, top); MD = mean diffusivity (figure 1B, bottom), note for 1B there were no significant relationships with projection fibres). All values in the figure met FDR correction. Two of the scores with the circular black border around the points (CPNE1 and NMT1) had an effect size in the opposite direction to all other scores (also indicated by -β for MD in figure legend). The colours of the plot points indicate the score to which they belong. Magnitude of standardised effect is shown in the legend entitled "Effect size (absolute values)".

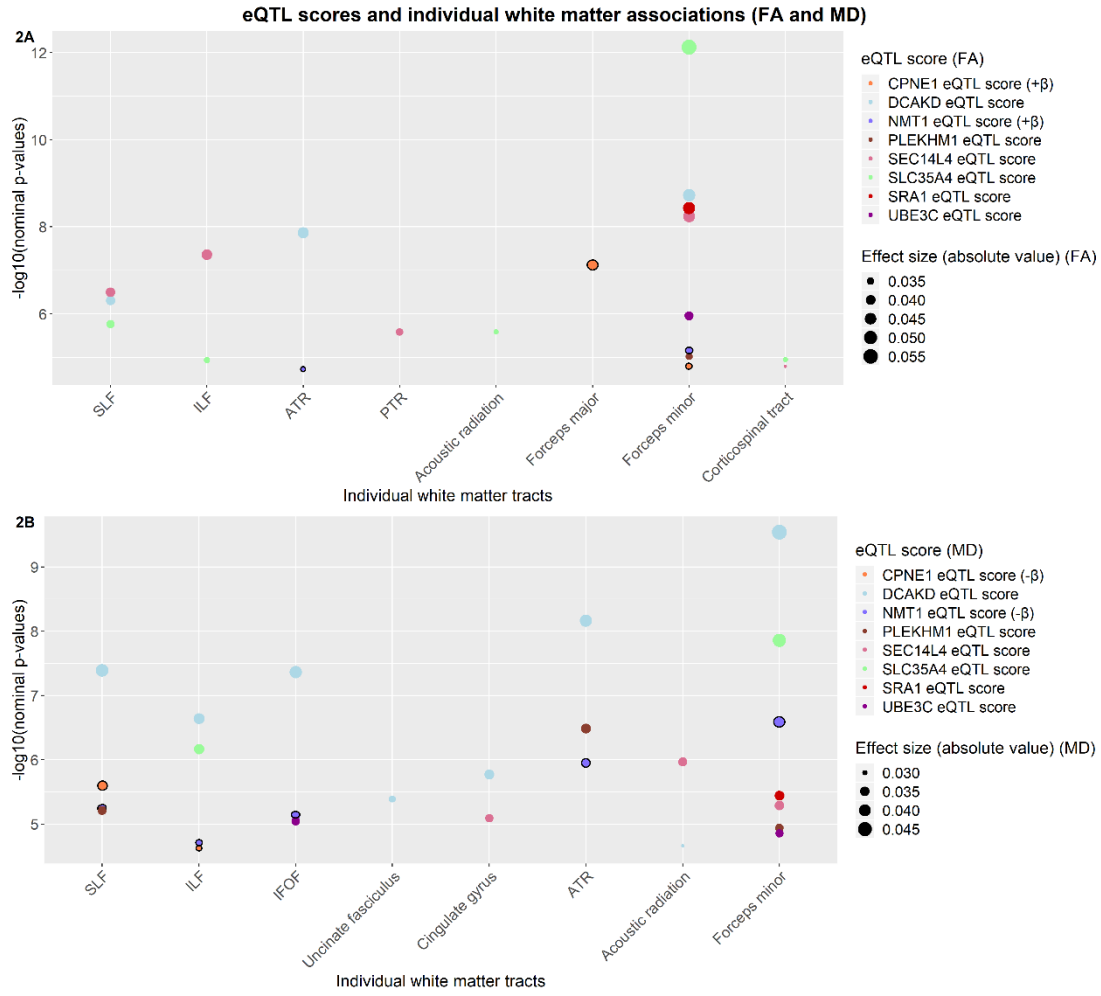


Figure 2 (A and B). Indicates nominal p-values between each of the 8 scores (shown in legend entitled "eQTL score") and individual white matter tracts (noted on the x-axis; FA = fractional anisotropy (figure 2A, top); MD = mean diffusivity (figure 2B, bottom)). SLF = superior longitudinal fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; ATR = anterior thalamic radiations; PTR = posterior thalamic radiations). All values in the figure met FDR correction. Two of the scores with the circular black border around the points (CPNE1 and NMT1) had an effect size in the opposite direction to all other scores (+β and -β for FA and MD, respectively in figure legend). The colours of the plot points indicate the score to which they belong. Magnitude of standardised effect is shown in the legend entitled "Effect size (absolute values)".

Allen Brain Atlas gene expression pattern

The Allen Brain Atlas is a multi-modal atlas of gene expression across brain regions, integrating structure, function, and gene expression data to aid in the investigation of the human brain in health and disease (Shen et al., 2012). For the current chapter, the atlas was used to investigate the 8 significantly-associated eQTL scores in terms of gene expression patterns across brain regions in 6 donors (Table 6; Figures 3 and 4).

Allen Brain Atlas Donor	Demographic characteristics
H0351.2001	24 years, Male, African American
H0351.2002	39 years, Male, African American
H0351.1009	57 years, Male, Caucasian
H0351.1012	31 years, Male, Caucasian
H0351.1015	49 years, Female, Hispanic
H0351.1016	55 years, Male, Caucasian

Table 6. Descriptive statistics of the 6 donors included in the Allen Brain Atlas (Shen et al., 2012).

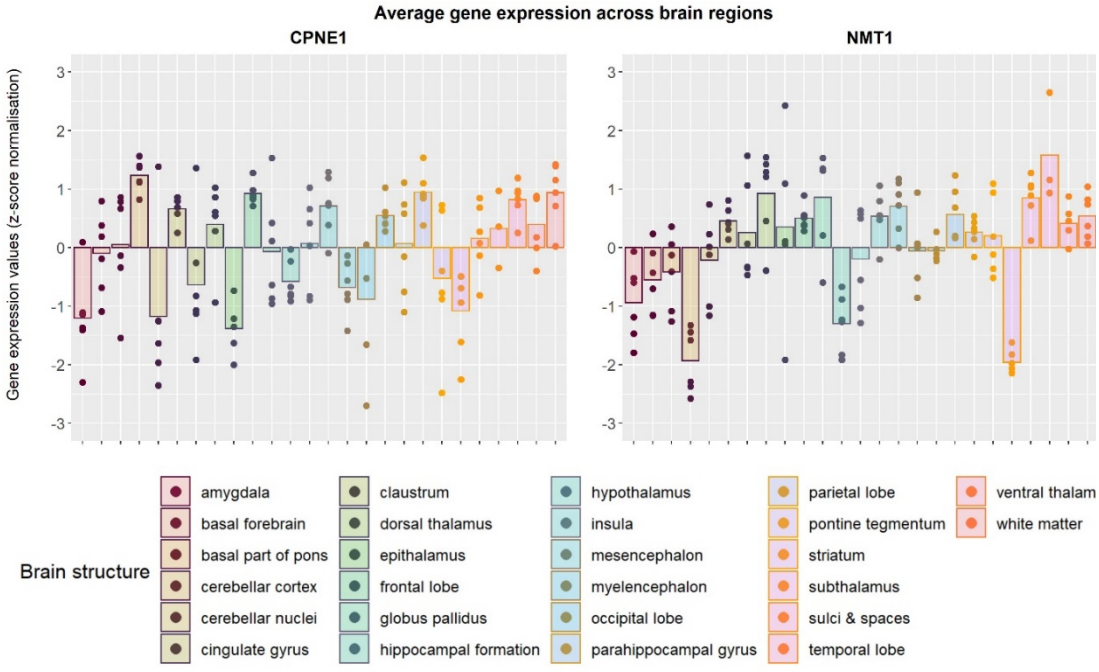


Figure 3. Average gene expression patterns across brain regions in N = 6 donors for 2 neurodevelopment-linked genes. Points on the plot indicate participants' own gene expression level, while the bars indicate the mean gene expression value across all donors. The y-axis indicates gene expression values (normalized z-scores). Brain structures are indicated in the legend.

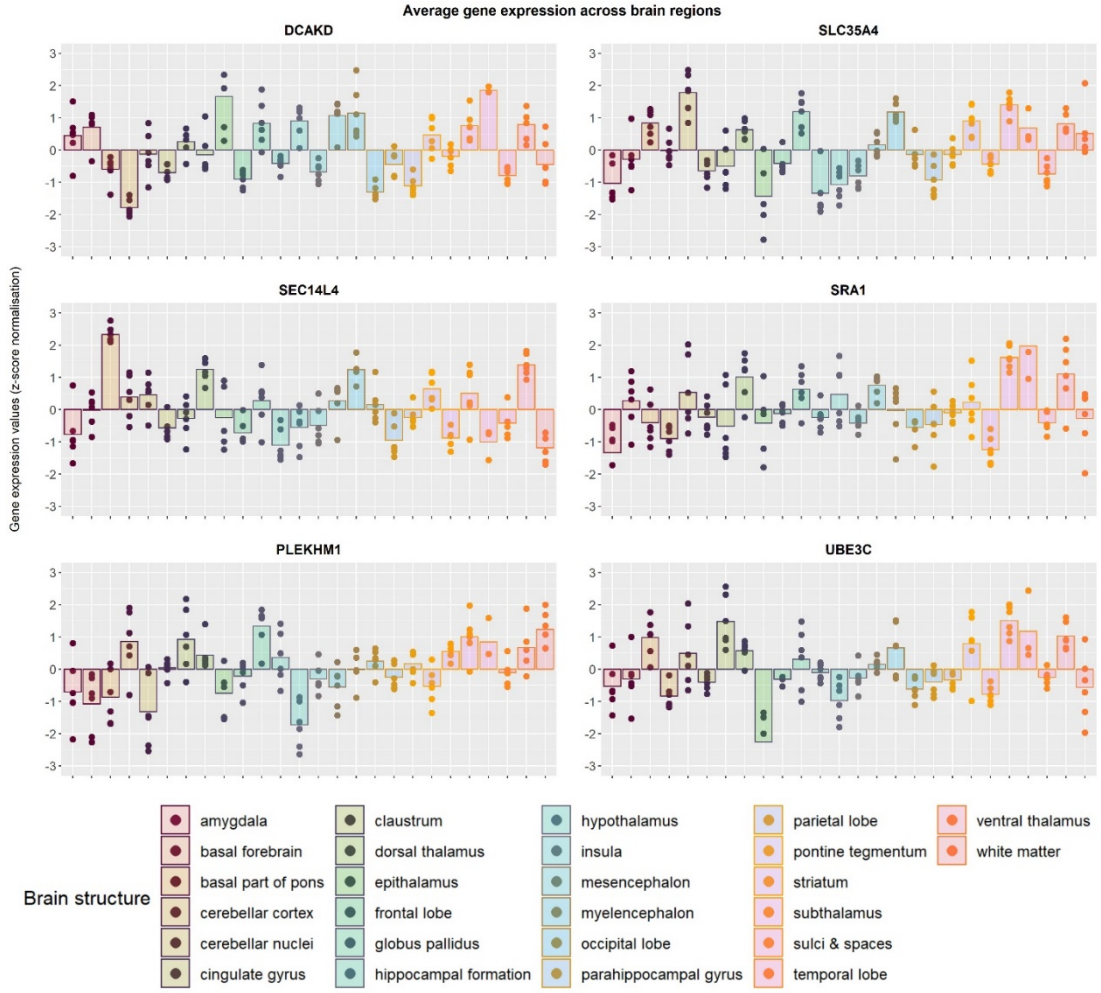


Figure 4. Average gene expression patterns across brain regions in N = 6 donors for 6 disease-linked genes. Points on the plot indicate participants' own gene expression level, while the bars indicate the mean gene expression value across all donors. The y-axis indicates gene expression values (normalized z-scores). Brain structures are indicated in the legend.

Genome-wide associations between score SNPs and white matter tracts

Using a previously published GWAS of imaging traits (Elliott et al., 2018), the association between the SNPs comprising each of the 8 scores ($N_{total} = 53$; SNP list can be found in Appendix 1: Table S14) with those found previously for the white matter tracts of interest (i.e. the tracts which showed post-FDR significant associations) were investigated. This SNP look-up was performed in order to observe whether our analysis of eQTL scores, comprising SNPs which together regulate the expression of a single gene, yielded any novel associations with white matter tracts which were not previously found in conventional GWAS.

The Brain Imaging Genetics (BIG) database (http://big.stats.ox.ac.uk/) was used to extract the effect size and p-value of each SNP of interest as associated with the white matter tracts of interest, as provided in Elliott et al. (2018). As GWAS for global and tract category measures were not performed in the original study, these GWAS were performed as part of the current project (i.e. GWAS for global measures, association fibres, thalamic radiations and projection fibres). Our GWAS parameters and quality check procedures are described in more detail in Appendix 1. P-values and effect size of each SNP for each individual white matter tract of interest (left and right hemispheres separately from Elliott et al., 2018), as well as for global and tract categories (run locally), are also contained in Appendix 1: Figure S3. Briefly, only one SNP across two eQTL scores (SLC35A4; SRA1) was previously found to reach genome-wide significance with forceps minor (FA), projection fibres (FA) and global FA (GWAS run locally): rs2237077.

2.5 Discussion

The current study utilised a novel approach to investigate whether eQTL scores, corresponding to the expression of specific genes in whole blood, were significantly and specifically associated with white matter tracts in N > 14,000 individuals. Significant associations were found in white matter microstructure as measured by both FA and MD for a number of scores ($FA_{N\ scores} = 25$; $MD_{N\ scores} = 24$). Of these, 8 scores were found to be significantly associated with various white matter tracts as measured by both FA and MD. In particular, the largest effect was seen for

the association between forceps minor (FA) and the eQTL score for *SLC35A4,* and across several tracts measured by MD for the eQTL score for *DCAKD*. Although these eQTL were derived from whole blood, there is evidence of expression in the brain for some of the genes, outlined in further detail below. These findings also provided novel information not previously found by conventional genome-wide association studies.

All 8 scores were associated with white matter microstructural integrity of the forceps minor as measured by FA (7 of which were also associated with MD values). The forceps minor forms the anterior part of the corpus callosum, connecting homologous regions of the prefrontal cortex between hemispheres. It is postulated to be involved in numerous cognitive and behavioural skills, such as decision making, social behaviour, and language (Miller et al., 2001). This connection therefore implicates forceps minor in a wide range of cognitive skills, and damage to the tract has been associated with neuropsychiatric and neurological disorders, such as multiple sclerosis and depression (Gobbi et al., 2014; Mamiya et al., 2018).

2.5.1 Global and individual tract findings – largest associations

The two genes with the largest associations were *DCAKD,* globally and across numerous tracts as measured by higher MD, and *SCL35A4* across tracts measured by lower FA, with a peak in projection fibres, localised to forceps minor. *DCAKD* is a protein coding gene which is ubiquitously expressed in brain, among other tissues (Latourelle et al., 2012). Previous evidence using mouse models indicates expression of this gene has a putative role in neurodevelopment (Gonzalez-Lozano et al., 2016), and is associated with a number of psychiatric and neurological disorders, including schizophrenia, autism spectrum disorder, and Parkinson's disease (Latourelle et al., 2012; Schizophrenia Working Group, 2018; Butler et al., 2015). Evidence for involvement in autism spectrum disorder comes from Butler et al. (2015), who compiled a list of clinically relevant genes for the disorder, with *DCAKD* among the participating susceptibility genes. Expression of *DCAKD* was also found to be implicated in Parkinson's disease (Latourelle et al., 2012), a disorder previously associated, along with other characteristic neurobiological features, with lower white matter integrity in tracts within the temporal, parietal and occipital lobes (Auning et al., 2014).

SLC35A4 belongs to the *SLC35* family, members of which act as transporters of nucleotide sugars, and is known to be expressed in brain (Sosicka et al., 2017). There is limited knowledge about its specific function, although a recent review investigating the subcellular localization and topology of *SLC35A4* demonstrated that it localizes mainly to the Golgi apparatus (Sosicka et al., 2017).

2.5.2 Disease-linked genes - lower FA & higher MD (decreased white matter integrity)

Four genes identified through eQTL methods (*SRA1, UBE3C, SEC14L4, PLEKHM1*) were associated with lower FA within several individual tracts pertaining to projection and association fibres, as well as with higher global MD. *SRA1* encodes both non-coding and protein-coding RNAs, is implicated in the regulation of numerous nuclear receptor activities, such as metabolism and chromatin organization, and is known to be expressed in the brain. Kotan et al. (2016) posited that *SRA1* plays a role in the initiation of puberty in humans by finding that inactivating *SRA1* variants were associated with idiopathic hypogonadotropic hypogonadism (IHH) in three independent families. IHH is a rare genetic disorder caused by the inability of the hypothalamus to secrete gonadotropin-releasing hormones (GnRH) or by the inability of GnRH to act on pituitary gonadotropes (Bianco et al., 2009). These previous results might link the association of *SRA1* with projection fibres, which connect the cerebral cortex to the spinal cord and brainstem, as well as to other centres of the brain (e.g. thalamus).

UBE3C contains ubiquitin-protein ligase (E3), an enzyme which accepts ubiquitin from E2 before transferring it to the target lysine; ubiquitin targets proteins for degradation via the proteasome. *UBE3C* is expressed in numerous tissues, including the brain, and has been previously associated with some neuropsychiatric-related phenotypes. For instance, Garriock et al. (2010) performed a GWAS to determine the association between genetic variation and Citalopram response. Although not genome-wide significant, their top finding was a SNP in proximity to *UBE3C* and was found to be associated with antidepressant response and MDD remission (rs6966038, p = 4.65e-07 and p = 3.63E-07, respectively) (Garriock et al., 2010). Moreover, Filatova et al. (2014) studied the expression of genes within the ubiquitin-proteasome protein degradation system, which is implicated in Parkinson's

disease, in mice with MPTP-induced pre-symptomatic and early symptomatic stages of Parkinson's disease. They found decreased expression in the striatum and the substantia nigra of mice, which may lead to a decrease in performance of the system. This may in turn lead to accumulation of abnormal and toxic proteins which guide neuronal cell death (Filatova et al., 2014).

The specific function of *SEC14L4* has not yet been determined, although the protein encoded by it is similar to a protein encoded by the *SEC14* gene in saccharomyces cerevisiae, which is essential to the biogenesis of Golgi-derived transport vesicles. Curwin and McMaster (2008) found that mutations in several *SEC14* domain-containing proteins in humans may be implicated in neurodegeneration, although it is not clear what the role of *SEC14L4* is within this context. Lastly, *PLEKHM1* is important in bone resorption, may be involved in vesicular transport in the osteoclast, and is weakly expressed in the brain. Although mutations in this gene have been associated with numerous phenotypes (Fujiwara et al., 2016; McEwan et al., 2015), none were neuropsychiatric-related.

2.5.3 Development-linked genes - higher FA & lower MD (increased white matter integrity)

Two of the eight genes (*CPNE1, NMT1*) were associated with higher FA and lower MD, indicating increased white matter integrity, associated with increased expression level as quantified by the corresponding eQTL.

CPNE1, which is thought to regulate molecular events at the cell membrane and cytoplasm, has previously been found to mediate several neuronal differentiation processes by interacting with intracellular signalling molecules. *CPNE1* has also been found to be highly expressed during brain development, indicating that it might be implicated in earlier developmental stages of neuronal function (Kim et al., 2018). Furthermore, C2 domains of *CPNE1*, calcium-dependent phospholipid-binding motives, have been shown to be implicated in neurite outgrowth of hippocampal progenitor HiB5 cells, which are hippocampal cell lines derived from the hippocampal analgen of E16 rat (Park et al., 2012; Park et al., 2014). *CPNE1* expression was associated here with two tracts within projection fibres (FA) and with regional association fibres (MD), which link the cortex to lower brain areas. In mouse and

human models, these findings may be of use when investigating neurite outgrowth from the hippocampus, which is part of the limbic system, an area located beneath the cortex.

NMT1 (N-myristoyltransferase) catalyzes the transfer of myristate (a rare 14-carbon saturated fatty acid) from CoA to proteins, and is expressed in numerous tissues, including ubiquitously in the brain. It has been found that *NMT1* is required for early mouse development, mainly due to its role in early embryogenesis (Deng et al., 2018). Expression of this gene has also been implicated in human brain tumours (Lu et al., 2005) and tumour cell proliferation (Ducker et al., 2005). In our study, *NMT1* was associated with tracts within thalamic radiations and projection fibres (FA) and global MD.

2.5.4 General Discussion

The current study employed a novel strategy of investigating a direct association between eQTL scores and white matter tracts to uncover a relationship between specific regulatory variants and brain connectivity. Together, our findings indicate that increases in expression of these genes may be implicated in several processes which may directly or indirectly alter white matter microstructure, each with localised, pronounced effects in specific tracts. Further, while some of the significant associations had connections with other brain-related traits, such as neurite outgrowth or psychiatric and neurological disorders, others did not. Interestingly, decreased white matter microstructure integrity, as marked by lower FA and higher MD, was associated with eQTL scores which regulate expression of genes implicated in neuropsychiatric and neurological disorders. Conversely, increased white matter integrity, as marked by higher FA and lower MD, was associated with *CPNE1* and *NMT1*, which are important in developmental processes such as neurite outgrowth. In addition, encouragingly, regions of the corpus callosum (i.e. the forceps minor), the largest and arguably most reliably measured white matter tract in the brain, was demonstrated to be associated with all 8 scores for FA, and 7 for MD. These findings together suggest that utilising this approach to associate eQTL scores with white matter microstructure may add to previous research which found associations between genes and these brain-related traits and disorders. These genes or eQTL for them might indirectly implicate brain connectivity through other processes in which they participate.

The current study has several strengths and some potential limitations. First, to our knowledge, this study is the first one to compute eQTL scores for specific gene transcripts and attempt to associate them with white matter tract integrity *in vivo*. Moreover, our analysis consisted of a population-based sample of N > 14,000 individuals recruited to the UKB, large enough to make our findings robust and generalizable to other samples within the same age range, background and ethnicity. Lastly, our findings revealed novel associations which were not previously found in GWAS (Elliott et al., 2018; GWAS of g measures run locally), indicating a potential to use such scores for further discovery analyses.

However, a potential limitation in this study is calculation of scores for data taken from whole blood, although there is previous evidence indicating that whole blood can be used as a proxy for brain eQTL, important for study of *in vivo* brain traits (McKenzie et al., 2014).

In summary, our results suggest that expression of the genes discussed above alter white matter microstructure and could facilitate the manifestation of numerous brain-related traits. Uncovering specific markers leading to the formation, maintenance and pathology of white matter could enable downstream analyses to elucidate links between genetics and neuroimaging in neurological and psychiatric disorders, as well as other brain-related traits.

3. Chapter conclusion

This study provided novel associations between gene expression-based eQTL scores and white matter microstructure, not previously identified by conventional genome-wide association studies. The finding that gene expression of previously disease linked-genes is associated with decreased white matter integrity, and previously development-linked genes are associated with increased white matter integrity, across two DTI scalars, indicates that the brain phenotype may in future be utilised to link genotype to disease phenotype. This chapter laid the foundation for the next two chapters, in which attempts were made to elucidate the link between stratified genetic risk of MDD to specific white matter tracts.

Chapter 3: Association of whole-genome and NETRIN1 signaling pathway-derived polygenic risk scores for Major Depressive Disorder and white matter microstructure in UK Biobank

1. Chapter introduction

As indicated in chapter 2, white matter microstructure phenotypes are linked to differential gene expression patterns, either in disease- or health-related traits. PRS have previously shown their utility in predicting psychiatric disorders in analyses including white matter microstructure (Whalley et al., 2013; Shen et al., 2017). However, while providing information relating to variance explained by additive genetic variants, a whole-genome PRS is limited in its ability to provide specific mechanistic insight into disease phenotypes (Dudbridge, 2013).

In this chapter, I attempt to look beyond whole-genome PRS and exploratory candidate gene pathways. I explore the relationship between PRS derived for a biological pathway, previously identified by large-scale data-driven genetic analyses and which participates both in neurodevelopment and manifestation of MDD (Zeng et al., 2016), and white matter microstructure. Chapter 3 therefore aims to investigate, using PRS, the association between white matter microstructure and genetic risk of MDD localised to one pathway, and uses a whole-genome PRS (excluding variants within the identified pathway) as a control risk score. PRS were calculated for 6,401 individuals with FA data and 6,390 individuals with MD data in UK Biobank. The study has been summarised in a manuscript entitled, "Association of whole-genome and NETRIN1 signaling pathway-derived polygenic risk scores for Major Depressive Disorder and white matter microstructure in UK Biobank" and has been published in *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. I am the first author of this manuscript, I designed the experiment, ran data analysis, and wrote the manuscript.

2. Manuscript

2.1 Abstract

Background: Major Depressive Disorder (MDD) is a clinically heterogeneous psychiatric disorder with a polygenic architecture. Genome-wide association studies have identified a number of risk-associated variants across the genome, and growing evidence of NETRIN1 pathway involvement. Stratifying disease risk by genetic variation within the NETRIN1 pathway may provide important routes for identification of disease mechanisms by focusing on a specific process excluding heterogeneous risk-associated variation in other pathways. Here, associations between MDD polygenic risk scores derived from the NETRIN1 signalling pathway (NETRIN1-PRS) and the whole genome excluding NETRIN1 pathway genes (genomic-PRS) with white matter microstructure were tested.

Methods: Two diffusion tensor imaging measures were used, fractional anisotropy (FA) and mean diffusivity (MD), in the most up-to-date UK Biobank neuroimaging data release (FA: N = 6,401; MD: N = 6,390).

Results: Findings included significantly lower FA in the superior longitudinal fasciculus (β = -0.035, pcorrected = 0.029) and significantly higher MD in a global measure of thalamic radiations (β = 0.029, pcorrected = 0.021), as well as higher MD in the superior (β = 0.034, pcorrected = 0.039) and inferior (β = 0.029, pcorrected = 0.043) longitudinal fasciculus and in the anterior (β = 0.025, pcorrected = 0.046) and superior (β = 0.027, pcorrected = 0.043) thalamic radiation associated with NETRIN1-PRS. Genomic-PRS was also associated with lower FA and higher MD in several tracts.

Conclusions: Our findings indicate that variation in the NETRIN1 signaling pathway may confer risk for MDD through effects on a number of white matter tracts.

2.2 Introduction

Major Depressive Disorder (MDD) is a common and frequently disabling psychiatric disorder and a leading cause of disability worldwide (Otte et al., 2016). MDD is known to result from a complex combination of environmental and genetic factors (Bromet et al., 2011; Zeng et al., 2016), with a moderate heritability of approximately 37% (Sullivan et al., 2000; Belmaker & Agam, 2008; Ripke et al., 2013).

Genome-wide association studies (GWAS) suggest that at least part of MDD's heritability is due to the cumulative effect of alleles of small effect size (Hek et al., 2013; Lubke et al., 2012) and have identified a number of risk-associated genetic variants across the genome (Ripke et al., 2013; Hek et al., 2013; Converge Consortium, 2015; Hyde et al., 2016; Mullins & Lewis, 2017). Significant findings for GWAS analyses can also be annotated to specific biological pathways, revealing underlying cellular and molecular mechanisms.

Following several GWAS, the Psychiatric Genomics Consortium (PGC) have identified an aggregation of variants in several specific biological pathways (Network T, 2015; Jia et al., 2012). In MDD, Zeng et al. (2017) combined pathway and regional heritability analysis in two independent samples and reported that the NETRIN1 signalling pathway was involved in the genetic aetiology of MDD. Moreover, polygenic risk scores (PRS) calculated for this pathway alone more accurately predicted MDD in one of the cohorts compared to PRS calculated for the whole genome. Genetic variation within the NETRIN1 signalling pathway may therefore capture more aetiologically circumscribed liability for MDD that is less susceptible to heterogeneous influences from other biological pathways.

Animal studies have previously indicated that NETRIN1, by binding to and activating NETRIN1 receptors such as 'Deleted in Colorectal Cancer' (DCC), plays an important role in commissural and cortical axon guidance (Serafini et al., 1996). More recently, DCC was identified as playing a crucial role in thalamic axonal growth, confirming that interaction of NETRIN1 with DCC leads to successful axon growth during central nervous system development (Castillo-Paterna et al., 2015). GWAS of

other traits related to MDD have also shown an aggregation of variants in the NETRIN1 pathway (Manitt et al., 2013; Ward et al., 2017).

Previous studies have attempted to investigate psychiatric disorders by examining relevant quantitative traits such as brain structure or function (Reus et al., 2017). Differences in white matter (WM) integrity as measured by diffusion tensor imaging (DTI) have been found between MDD patients and healthy participants in numerous studies, although findings have been widely inconsistent (Shen et al., 2017; Klimes-Dougan et al., 2010; Korgaonkar et al., 2011). For example, Shen et al. (2017) found significantly lower global white matter integrity in association fibres and thalamic radiations, as measured by fractional anisotropy (FA), in MDD patients compared to healthy individuals. More specifically, they also found lower FA in the left superior longitudinal fasciculus, superior thalamic radiations and forceps major tracts in MDD patients. Lower WM integrity as measured by FA has also been found in adolescents with MDD as compared to age-matched healthy individuals (Klimes-Dougan et al., 2010; Korgaonkar et al., 2011).

It has previously been shown that the NETRIN1 signaling pathway is associated with MDD and white matter microstructure (Zeng et al., 2017). Therefore, the current study sought to investigate the association between MDD risk-associated variants in the NETRIN1 signaling pathway and white matter integrity. Polygenic risk scores for pathway SNPs (NETRIN1-PRS) and SNPs excluded from the pathway (genomic-PRS) were created and their association with WM integrity as measured by FA and mean diffusivity (MD) was tested using the most up-to-date genetic and imaging data available (N after exclusion steps: FA = 6,401; MD = 6,390) from UK Biobank (UKB). It was hypothesized that NETRIN1-PRS would be significantly associated with WM integrity, after adjustment for genomic-PRS, indicating a potential role of the pathway in MDD pathophysiology.

2.3 Methods and Materials

2.3.1 UK Biobank

The UKB study consists of 502,617 community-dwelling individuals who were recruited between 2006 and 2010 in the United Kingdom (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=200). UKB received ethical approval from the Research Ethics Committee (reference: 11/NW/0382). This study has been approved by the UKB Access Committee (Project #4844). Written informed consent was obtained from all participants.

2.3.2 Study population

In the most recent UKB imaging data release, 8,839 individuals (N female = 4,639; N male = 4,200; mean age: 62.54 +/- 7.42 years; age range: 45.17 – 79.33) completed DTI assessment, and a quality check by UKB. In addition to this, for the current study, individuals were excluded if they participated in studies from the PGC MDD GWAS (Wray & Sullivan, 2017) or Generation Scotland (Scottish Family Health Study), or if they happened to be related, as the PGC MDD GWAS dataset was used in order to calculate PRS. Moreover, individuals whose FA and MD values were greater than three standard deviations above/below the mean were not included in the study (Appendix 2: Tables S4 and S5). This resulted in 6,401 individuals with FA values (N female = 3,334; N male = 3,067; mean age: 62.60 +/- 7.37; age range: 45.92 – 78.42; N control: 3,736; N case: 2,512) and 6,390 individuals with MD values (N female = 3,327; N male = 3,063; mean age: 62.58 +/- 7.36; age range: 45.92 – 78.42; N control: 3,729; N case: 2,508), excluding 19 and 30 individuals with FA and MD values from a total of 6,420, respectively. Details of data exclusion as well as participant information for the full dataset (N = 6,420) are shown in Appendix 2: Tables S1 and S2.

2.3.3 The NETRIN1 signalling pathway and SNP annotation

The NETRIN1 pathway is implicated in axon guidance, by binding to and activating receptors such as *DCC* during neurodevelopment, where axon navigation is

guided by extracellular axon guidance cues (Braisted et al., 2000). Figure 1 below indicates the NETRIN1-dependent axon guidance pathway process.

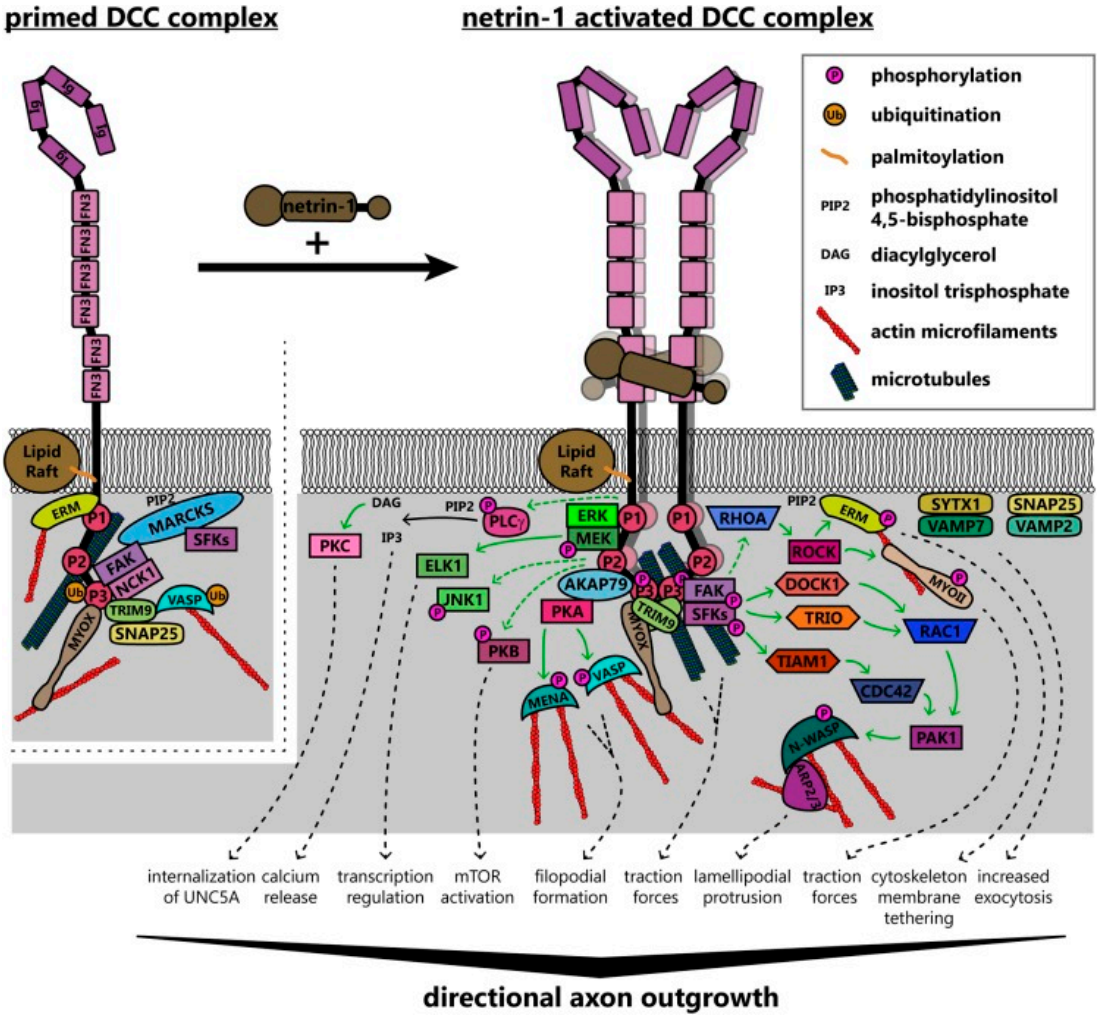


Figure 1. Model of signalling pathways and interactions downstream of *DCC* in the NETRIN1-dependent axon guidance pathway, as shown in Boyer and Gupton (2018). As shown in the figure, *DCC* interacts with enzymes and adaptor proteins in the absence of NETRIN1, which can initiate responses to ligand binding. Valency is increased by NETRIN1 through multimerization of *DCC* homodimers. Intracellular domains of the receptors are thus brought into close apposition, which forms a scaffolding for recruitment and activation of proteins. In the figure, solid green arrows indicate direct activation steps, and dashed green arrows represent known connections. The pathways modify the intracellular environment together to promote directional axon growth in response to *Netrin1* (Boyer & Gupton, 2018).

Genic SNPs found in the NETRIN1 signaling pathway as taken from Zeng et al.'s (2017) study (N genes = 43; gene list is presented in Appendix 2: Table S3) and genic SNPs excluded from the pathway were annotated using the program ANNOVAR. ANNOVAR is a biostatistical tool used to annotate genetic variants to functional genomic regions (Yang & Wang, 2015). In the current study, a gene-based annotation was performed for SNPs used in the largest available GWAS of MDD (N=461,134, of which 130,664 were MDD cases), carried out by the Psychiatric Genomics Consortium (Wray & Sullivan, 2017), which includes summary statistics from the personal genetics company 23andMe, Inc. (Hyde et al., 2016). Gene boundaries were defined as an extended region of 20 kb from transcription start sites and transcription end sites. After SNPs were annotated to genes, they were further mapped to the NETRIN1 signalling pathway. All protein-coding genes within this file were annotated in reference to hg 19. Intergenic SNPs were not included in the annotated files. The resulting output file included: function of each SNP, gene name, chromosome number, start position, end position, reference and alternative alleles, odds ratio, standard error and p-value for each variant.

Following functional annotation, a file containing the 43 gene names included in the NETRIN1 signaling pathway was used as an input in order to extract gene-based SNPs located in the pathway. For the genomic-PRS, all gene-based SNPs excluding those implicated in the NETRIN1 signaling pathway were extracted. The two files were then used as input for creation of PRS.

2.3.4 Genotyping and PRS profiling

A total of 488,363 UKB blood samples (N female = 264,857; N male = 223,506; http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22001), were genotyped using two different arrays: UK BiLEVE array (N = 49,949) (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149600) and UK Biobank Axiom array (N = 438,417) (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149601). Details of genotyping and quality control are described in more detail by Hagenaars et al. (2016) and Bycroft et al. (2017).

Using the largest available GWAS of MDD, PRS for each individual were computed using PRSice (Euesden et al., 2014), at five p-value thresholds (0.01, 0.05, 0.1, 0.5, 1) by adding the number of risk alleles and weighting them by the strength of association with MDD. PRS were created both from SNPs annotated to the NETRIN1 signalling pathway and from SNPs from the rest of the genome, thus resulting in separate PRS lists. PRS were created both with and without clump-based pruning of SNPs in linkage disequilibrium (r2 = 0.25, 250km window). The primary analysis reported in this manuscript concerns unpruned SNPs, owing to the potential of causal variants within the NETRIN1 pathway to be in LD with other variants, and uses SNPs which met a significance level of p = 0.5, in line with previous studies (Purcell et al., 2009; Whalley et al., 2016). Secondary analyses with other PRS p-value thresholds, as well as with LD pruned SNPs, are presented in Appendix 2, Tables S6 – S21.

2.3.5 MRI acquisition

In the present study, imaging-derived phenotypes (IDPs) produced by UKB were used. MRI acquisition and pre-processing procedures for FA and MD values of white matter tracts were performed by UKB using standardised protocols (https://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf). Briefly, images were collected on a single Siemens Skyra 3.0 T scanner with a standard Siemens 32-channel head coil and were pre-processed using FSL packages; parcellation of white matter tracts was conducted using AutoPtx (Alfaro-Almagro et al., 2017).

Summary data were composed of tract-averaged FA and MD values for 15 major white matter tracts, of which 12 are bilateral and three are unilateral. The white matter tracts were also categorised into three separate subsets, as follows: association fibres: inferior fronto-occipital fasciculus, uncinate fasciculus, cingulum bundle (gyrus and parahippocampal), superior and inferior longitudinal fasciculus; thalamic radiation fibres: anterior, superior and posterior thalamic radiations; projection fibres: forceps major and minor, corticospinal tract, acoustic radiation, medial lemniscus and middle cerebellar peduncle. Global measures of FA and MD are referred to as general factors of FA and MD (gFA and gMD, respectively).

Exclusion criteria comprised removal of scans with severe normalisation problems by UKB. Moreover, individuals whose FA and MD values were higher than three standard deviations from the sample mean were also excluded. Descriptive statistics for the full dataset with outliers included and excluded are also presented in Appendix 1: Tables S1 and S2. Lastly, due to the fact that the position of the head and radio-frequency coil in the scanner may affect data quality as well as IDPs, three scanner brain position variables which may be used as confounding variables in subsequent analyses were generated by UKB: lateral brain position – X (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25756), transverse brain position –Y (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25757) and longitudinal brain position – Z (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25758). The three variables were included as covariates in the statistical analysis described below.

2.3.6 Statistical methods

All analyses were conducted using R (version 3.2.3) in a Linux environment. In order to test the association between the NETRIN1 signaling pathway- and genomic pathway-derived unpruned PRS lists, repeated measures linear mixed-effects models (function "lme" in package "nlme") were used for 12 bilateral brain regions, correcting for hemisphere, with age, age^2, sex, fifteen genetic principal components, three MRI head position coordinates and genotype array set as covariates. For unilateral tracts, global measures of FA and MD, and tract categories, a general linear model (function "lm") was used, using the same covariates as above, and without hemisphere included as a separate term in the model. All models included both the genomic-PRS and the NETRIN1-PRS as predictor variables.

First, the association between unpruned PRS (both NETRIN1-PRS and genomic-PRS) and global white matter integrity was tested. Principal component analysis (PCA) was then applied on the 27 white matter tracts (12 tracts in both the right and left hemisphere and three unilateral tracts) in order to extract a latent measure. Scores of the first unrotated component of FA and MD (variance explained = 37.52% for FA and 38.83% for MD) were extracted and set as the dependent variable in a

general linear model in order to test association with both NETRIN1-PRS and genomic-PRS.

The three categories of white matter tracts were examined by applying PCA on the regions involved in each, as a substantial proportion of white matter microstructural properties shows substantial commonality across these pathways (Cox et al., 2016). Scores of the first unrotated component of FA and MD were similarly extracted and set as dependent variables in general linear modelling, as above. Variance explained for each white matter tract subset was as follows: association fibres: 45.36% (FA), 50.76% (MD); thalamic radiations: 60.85% (FA), 73.40% (MD); projection fibres: 35.54% (FA), 29.28% (MD).

Lastly, the association between PRS (both NETRIN1-PRS and genomic-PRS) and each individual white matter tract (N = 15) was tested, using a repeated-effect linear model for the 12 bilateral tracts and a random-effect general linear model for the three unilateral tracts.

False discovery rate correction was applied separately for the 15 individual tracts and for global and tract category values.

2.3.7 Permutation analysis

In order to establish that the effect of the NETRIN1 pathway-derived PRS on WM integrity as measured by FA and MD was not due to chance, a circular genomic permutation method developed by Cabrera et al. (2012) was applied to the pathway SNP genotypes. The permutation approach uses GWAS SNP association results to identify the significance of pathway associations while accounting for the linkage disequilibrium structure of SNPs. As such, for a given GWAS, all SNPs are placed in what is called a "circular genome" based on their location. The complete set of p-values derived from the GWAS SNP associations are then permuted in a rotational fashion with respect to the genomic locations of the SNPs. This allows SNPs to retain the same position within the genome and with respect to each other, but gain new random association p-values. Once simulated p-values are assigned, a Fisher's combination test is used to calculate joint gene p-values. The method was developed

to increase understanding of gene-sets and pathways implicated in traits without generating pathway associations that are false-positive (Cabrera et al., 2012).

In this study, this was done by placing all SNPs in the whole genome (excluding those in the NETRIN1 pathway) in a circular genome, according to their location. One thousand SNP lists with the same set size as the NETRIN1 pathway were permuted using the method described above and 1000 PRS lists were created, which were then fitted in linear mixed-effects and general linear models, depending on the white matter tract tested, and their association with five white matter tracts and one tract category, found to be significantly associated with NETRIN1, was tested.

2.4 Results

Results presented below are significant specifically to each pathway. White matter tracts showing a significant association with both the NETRIN1-PRS and the genomic-PRS pathways are described in Appendix 2. Results for all individual white matter tracts, tract categories and global measures can be found in tables 1-4 and figures 2-5.

2.4.1 The effect of unpruned NETRIN1-PRS & genomic-PRS on measures of white matter integrity – FA (N = 6,401)

	White matter tracts	NETRIN1-PRS					genomic-PRS				
		Effect size (β)	SD	p value	p corrected (FDR)	R2	Effect size (β)	SD	p value	p corrected (FDR)	R2
AF	CGC	-0.025	0.011	0.020	0.152	0.062	-0.019	0.011	0.069	0.115	0.038
	PHC	-0.008	0.011	0.435	0.544	0.007	-0.020	0.011	0.061	0.115	0.040
	IFOF	-0.023	0.011	0.046	0.172	0.053	-0.028	0.012	0.016	0.060	0.076
	ILF	-0.023	0.011	0.043	0.172	0.054	-0.024	0.012	0.040	0.115	0.056
	SLF	-0.036	0.012	0.002	**0.030**	0.128	-0.023	0.012	0.047	0.115	0.053
	UF	-0.019	0.011	0.081	0.202	0.102	-0.032	0.011	0.003	**0.043**	0.102
TR	ATR	-0.022	0.011	0.057	0.172	0.048	-0.015	0.011	0.190	0.238	0.023
	PTR	-0.014	0.011	0.205	0.308	0.020	-0.022	0.011	0.054	0.115	0.047
	STR	-0.006	0.012	0.622	0.718	0.003	-0.015	0.012	0.213	0.244	0.022
PF	AR	0.003	0.011	0.759	0.814	0.001	-0.013	0.011	0.228	0.244	0.016
	CST	0.002	0.011	0.863	0.863	0.000	-0.018	0.011	0.103	0.154	0.034
	ML	-0.009	0.010	0.400	0.544	0.008	-0.003	0.010	0.803	0.803	0.001
	Fmaj	-0.016	0.012	0.193	0.308	0.024	-0.032	0.012	0.009	**0.043**	0.100
	Fmin	-0.018	0.012	0.135	0.262	0.032	-0.032	0.012	0.009	**0.043**	0.099
	MCP	-0.018	0.012	0.140	0.262	0.032	-0.019	0.012	0.125	0.170	0.035

Table 1. The effect of NETRIN1-PRS & genomic-PRS at PRS threshold 0.5 on individual white matter tracts (FA values). The first column for each PRS indicates standardised effect size (β). Statistically significant p-values after false discovery rate correction for each pathway individually are shown in bold. R^2 = estimate of variance explained by each pathway in %.

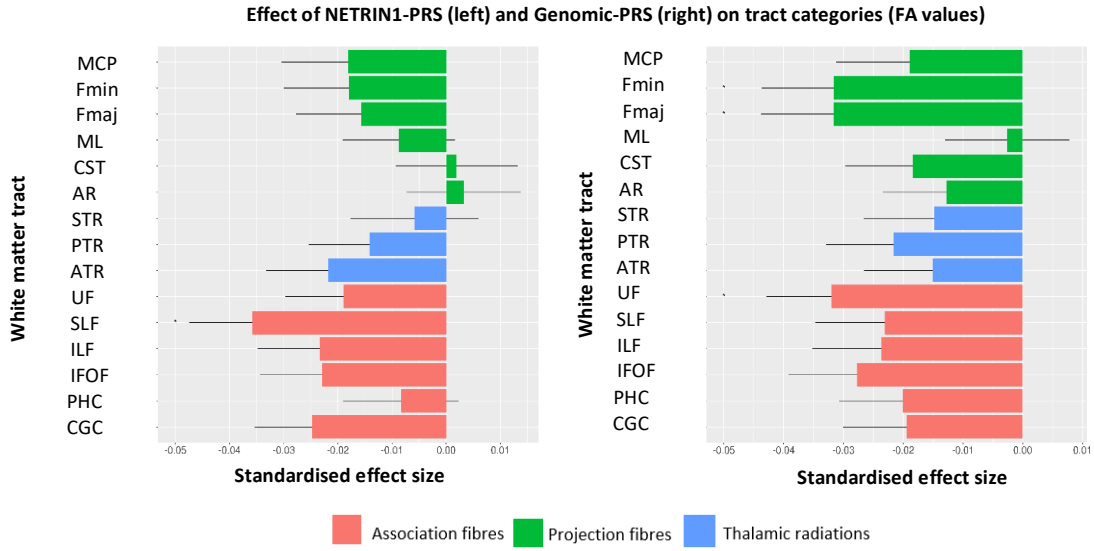


Figure 2. The effect of NETRIN1-PRS & genomic-PRS on FA values of white matter tracts. The x-axis indicates the standardised effect size of each pathway's PRS; the y-axis indicates the white matter tracts. The legend indicates the tract category belonging to each white matter tract. The error bar represents standard deviation of mean.

	NETRIN1-PRS					genomic-PRS				
	Effect size(β)	Standard deviation	p value	p corrected (FDR)	R2	Effect size(β)	Standard deviation	p value	p corrected (FDR)	R2
gFA	-0.026	0.012	0.028	0.056	0.068	-0.033	0.012	0.006	**0.011**	0.109
AF	-0.033	0.012	0.006	0.023	0.107	-0.034	0.012	0.005	0.011	0.113
TR	-0.018	0.012	0.138	0.185	0.032	-0.022	0.012	0.064	0.064	0.050
PF	-0.011	0.012	0.366	0.366	0.012	-0.029	0.012	0.016	**0.021**	0.083

Table 2. The effect of NETRIN1-PRS & genomic-PRS at PRS threshold 0.5 on global FA and 3 white matter tract categories. The first column for each PRS indicates standardised effect size (β). Statistically significant p-values after false discovery rate correction for each pathway individually are shown in bold. R^2 = estimate of variance explained by each pathway in %.

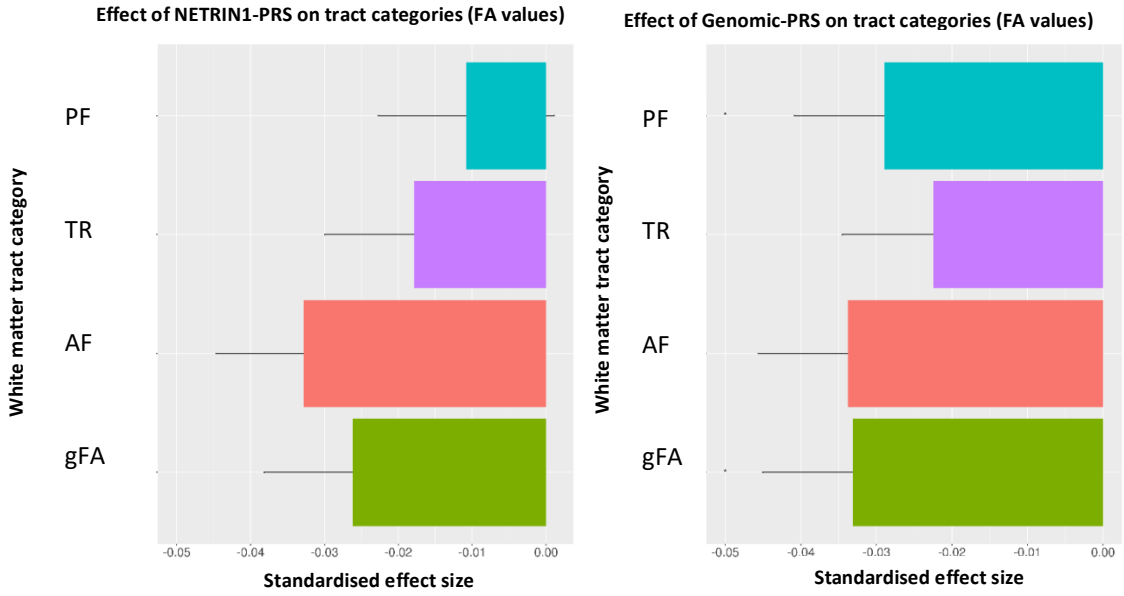


Figure 3. The effect of NETRIN1-PRS & genomic-PRS on FA values of tract categories and global FA. The x-axis indicates the standardised effect size of each pathway's PRS; the y-axis indicates the tract categories. The error bar represents standard deviation of mean.

Global measures

Lower global FA (gFA) was significantly associated with higher genomic-PRS (β = -0.033, pcorrected = 0.011) only.

Tract categories

The association between NETRIN1-PRS and Genomic-PRS and three subsets of white matter tracts (association fibres, thalamic radiations and projection fibres) was then tested. Significantly lower FA values in projection fibres were found for genomic-PRS (β = -0.028, pcorrected = 0.020) only.

Individual white matter tracts

Lastly, the effect of NETRIN1-PRS and genomic-PRS on WM integrity in 15 individual white matter tracts was investigated. NETRIN1-PRS, but not genomic-PRS, was associated with significantly lower FA in the superior longitudinal fasciculus (β = -0.035, pcorrected = 0.029).

In the genomic-PRS, there was significantly lower FA in the forceps major (β = -0.031, pcorrected = 0.043), forceps minor (β = -0.031, pcorrected = 0.043) and uncinate fasciculus (β = - 0.031, pcorrected = 0.043). None of these tracts showed significant associations with NETRIN1-PRS.

2.4.2 The effect of unpruned NETRIN1-PRS & genomic-PRS on measures of white matter integrity – MD (N = 6,390)

	White matter tracts	NETRIN1-PRS					genomic-PRS				
		Effect size (β)	SD	p value	p corrected (FDR)	R2	Effect size (β)	SD	p value	p corrected (FDR)	R2
AF	CGC	0.020	0.011	0.061	0.130	0.040	0.035	0.011	0.001	**0.014**	0.124
	PHC	-0.002	0.011	0.861	0.861	0.000	0.033	0.011	0.002	**0.014**	0.107
	IFOF	0.027	0.011	0.014	0.047	0.075	0.031	0.011	0.005	0.019	0.098
	ILF	0.029	0.011	0.009	**0.043**	0.086	0.025	0.011	0.027	0.067	0.061
	SLF	0.034	0.011	0.003	**0.039**	0.116	0.024	0.011	0.033	0.071	0.058
	UF	0.018	0.010	0.090	0.168	0.085	0.029	0.010	0.005	**0.019**	0.084
TR	ATR	0.025	0.011	0.016	**0.047**	0.065	0.021	0.011	0.043	0.080	0.046
	PTR	0.025	0.011	0.020	0.050	0.062	0.002	0.011	0.876	0.876	0.000
	STR	0.027	0.010	0.006	**0.043**	0.074	0.018	0.010	0.077	0.096	0.031
PF	AR	0.004	0.010	0.708	0.772	0.002	0.019	0.011	0.064	0.087	0.038
	CST	0.016	0.011	0.162	0.221	0.025	0.022	0.011	0.055	0.082	0.047
	ML	0.004	0.011	0.721	0.772	0.001	0.004	0.011	0.692	0.741	0.002
	Fmaj	0.018	0.012	0.135	0.203	-0.026	0.028	0.012	0.018	0.055	0.019
	Fmin	0.019	0.012	0.101	0.168	-0.063	0.023	0.012	0.050	0.082	-0.051
	MCP	0.013	0.012	0.290	0.363	0.016	0.010	0.012	0.394	0.455	0.010

Table 3. The effect of NETRIN1-PRS & genomic-PRS at PRS threshold 0.5 on individual white matter tracts (MD values). The first column for each PRS indicates standardised effect size (β). Statistically significant p-values after false discovery rate correction for each pathway individually are shown in bold. R^2 = estimate of variance explained by each pathway in %.

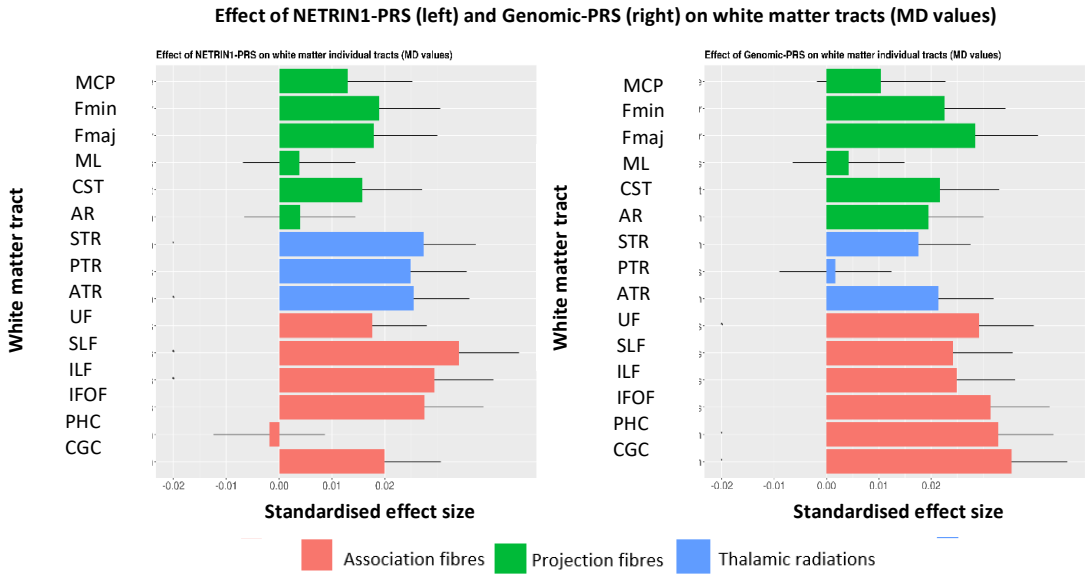


Figure 4. The effect of NETRIN1-PRS & genomic-PRS on MD values of white matter tracts. The x-axis indicates the standardised effect size of each pathway's PRS; the y-axis indicates the white matter tracts. The legend indicates the tract category belonging to each white matter tract. The error bar represents standard deviation of mean.

	NETRIN1-PRS					genomic-PRS				
	Effect size(β)	SD	p value	p corrected (FDR)	R2	Effect size(β)	SD	p value	p corrected (FDR)	R2
gMD	0.028	0.011	0.016	0.031	0.076	0.034	0.011	0.003	0.007	0.111
AF	0.022	0.012	0.058	0.077	0.048	0.042	0.012	0.000	**0.001**	0.172
TR	0.030	0.011	0.005	**0.021**	0.089	0.013	0.011	0.218	0.218	0.017
PF	0.021	0.012	0.077	0.077	0.045	0.029	0.012	0.017	**0.023**	0.081

Table 4. The effect of NETRIN1-PRS & genomic-PRS at PRS threshold 0.5 on global MD and 3 white matter tract subsets. The first column for each PRS indicates standardised effect size (β). Statistically significant p-values after false discovery rate correction for each pathway individually are shown in bold. R^2 = estimate of variance explained by each pathway in %.

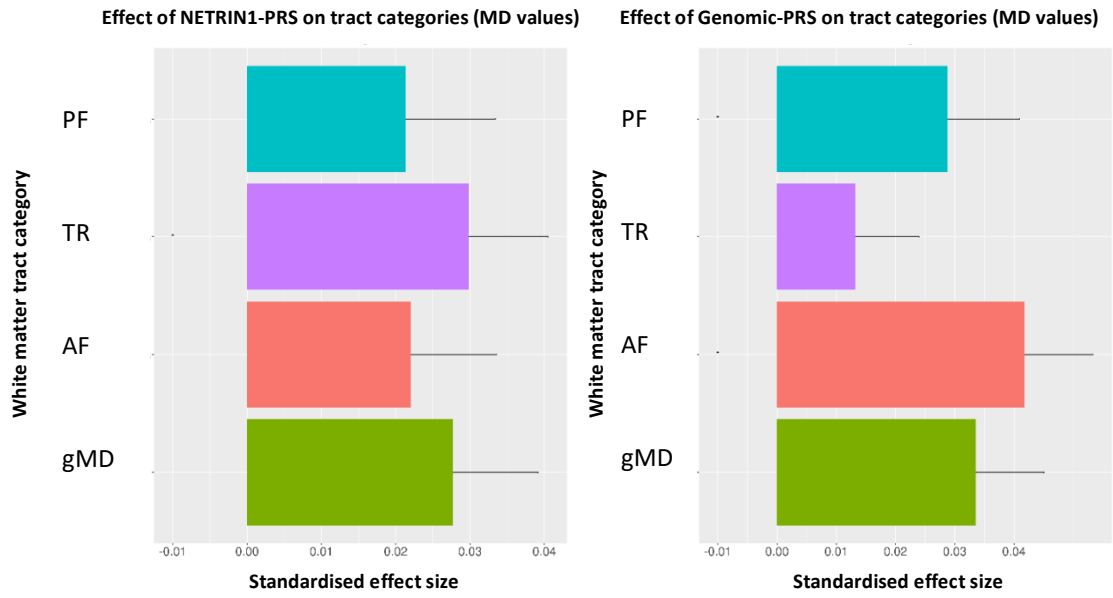


Figure 5. The effect of NETRIN1-PRS & genomic-PRS on MD values of tract categories and global MD. The x-axis indicates the standardised effect size of each pathway's PRS; the y-axis indicates the tract categories. The error bar represents standard deviation of mean.

Tract categories

MD values for association fibres (β = 0.041, pcorrected = 0.001) and projection fibres (β = 0.028, pcorrected = 0.023) were found to be significantly higher for genomic-PRS, but not NETRIN1-PRS. MD values for thalamic radiations were found to be significantly higher in the NETRIN1-PRS (β = 0.029, pcorrected = 0.021), whereas there was no significant association with genomic-PRS.

Individual white matter tracts

Within the 15 individual white matter tracts, numerous areas were significantly associated with both the NETRIN1-PRS and genomic-PRS. With regards to NETRIN1-PRS, MD values were significantly higher in the inferior longitudinal fasciculus (β = 0.029, pcorrected = 0.043), superior longitudinal fasciculus (β = 0.034, pcorrected = 0.039), and in the anterior (β = 0.025, pcorrected = 0.046) and superior (β = 0.027, pcorrected = 0.043) thalamic radiations. All of these significant associations were specific for NETRIN1-PRS.

In the genomic-PRS, there were significantly higher MD values in the cingulate gyrus (β = 0.035, pcorrected = 0.013) and parahippocampal (β = 0.032, pcorrected = 0.014) part of cingulum and in the uncinate fasciculus (β = 0.029, pcorrected = 0.018).

2.4.3 Permutation analysis

NETRIN1-PRS, but not genomic-PRS, were found to be individually significantly associated with white matter microstructure in the following white matter tracts: superior longitudinal fasciculus as measured by lower FA; superior and inferior longitudinal fasciculus and anterior and superior thalamic radiations, as well as thalamic radiations tract category, as measured by higher MD. Therefore, an additional circular genomic permutation analysis was performed and it was found that the variance explained by NETRIN1-PRS in these tracts was significantly higher than expected by chance (table 5).

White matter tract	Effect size of regression NETRIN1 pathway	Regression NETRIN1 pathway t-score	NETRIN1 calculated permutation p value
Superior longitudinal fasciculus (FA)	-0.035	-3.093	0.004
Superior longitudinal fasciculus (MD)	0.034	3.008	0.004
Inferior longitudinal fasciculus (MD)	0.029	2.624	0.014
Anterior thalamic radiations (MD)	0.025	2.419	0.023
Superior thalamic radiations (MD)	0.027	2.757	0.007
Thalamic radiations (MD)	0.029	2.785	0.008

Table 5. Permutation results for NETRIN1-PRS at PRS threshold 0.5 on 5 significant white matter tracts and one significant tract category. The first column indicates standardised effect size (β) and the second column indicates t-scores.

2.5 Discussion

The present study aimed to investigate whether PRS calculated from the NETRIN1 signalling pathway are significantly and specifically associated with WM integrity while simultaneously modelling genomic-PRS in more than 6,000 individuals. Significant differences were found in white matter integrity in both NETRIN1-PRS and genomic-PRS, for both FA and MD values. Regarding FA values, for NETRIN1-PRS, but not for genomic-PRS, a significant association was observed in the superior longitudinal fasciculus. NETRIN1-PRS alone were significantly associated with higher generalised thalamic radiations as measured by MD, as well as higher MD in the superior and inferior longitudinal fasciculus, and the anterior and superior thalamic radiations. Genomic-PRS were also significantly associated with FA and MD values in several tracts.

One of the main findings in our paper was both a reduction of FA and an increase of MD in the SLF in relation to NETRIN1-PRS. The SLF, a tract in association fibres, connects the frontal, temporal, parietal and occipital lobes, and has been shown to be highly involved in MDD (Wu et al., 2011; Cole et al., 2012). FA reductions in the SLF have also been found in previous studies combining genetic and neuroimaging techniques (Whalley et al., 2013), further indicating that the tract might be an important biomarker of MDD. In addition to this finding, there was also an increase in MD values in the ILF, a tract connecting the temporal and occipital lobes. Key areas in these two lobes include the amygdala and hippocampus, which are known to be implicated in emotion processing, a process which is disrupted in MDD (Ritchey et al., 2011). Previous studies have found disrupted white matter integrity in this tract in association with MDD using FA, indicating that it may play an important role in the pathophysiology of MDD (Whalley et al., 2013).

An MD increase in the thalamic radiations tract category was also found. Thalamic radiations connect the thalamus with numerous cortical areas (Cabrera et al., 2012; Braisted et al., 2000), and are connected to various cognitive processes, such as attention and wakefulness (Bonnin et al., 2007). Thalamocortical axons play an important role during development, as their projection from the dorsal thalamus (DT) transmit sensory information to the neocortex (Braisted et al., 2000). Thalamic radiations have previously been linked to MDD in numerous studies. For instance, a

decrease in FA was found in the TR subset in a large UKB sample comparing 335 MDD patients with 754 healthy individuals (Shen et al., 2017). This tract subset was also found to be significantly associated with higher PRS, indicating that there is a link between the sets of tracts and a potential genetic predisposition to MDD (Whalley et al., 2013).

NETRIN1, and its receptor DCC, one of the genes in the NETRIN1-pathway, have been previously implicated in thalamic axonal growth. NETRIN1 promotes growth of thalamocortical axons by binding to and activating DCC, which is expressed in the DT. Moreover, NETRIN1 has been shown to enhance axonal growth in explants of the DT, as well as providing guidance from the DT to the cortex (Braisted et al., 2000). It has also been found that serotonin, which is highly implicated in MDD, modulates the effect of NETRIN1 on embryonic thalamocortical axons (Braisted et al., 2000; Bonnin et al., 2007; Clasca et al., 2016). The active involvement of NETRIN1 in thalamocortical axonal growth, therefore, may explain our findings, and further confirms that there is a potential link between a biological pathway and specific neurobiological markers in MDD.

Several other tracts also showed a significant association of FA (individually in forceps major and minor and uncinate fasciculus, and in global measures of FA and projection fibres) and MD (individually in cingulate part of the cingulum, parahippocampal part of cingulum and uncinate fasciculus, and in global measures of association and projection fibres) with genomic-PRS, most of which have also been previously associated with MDD (Shen et al., 2017; Whalley et al., 2013). This evidence further confirms that there is an association between genetic predisposition to MDD and disruptions in white matter integrity, also for variants that lie outside the NETRIN1-DCC pathway. As such, these findings suggest that both PRS lists affect integrity across the white matter tracts, each with localized, pronounced effect in specific tracts.

The current study has several strengths and a few potential limitations. First of all, it is the largest combined genetic and neuroimaging study investigating the effect of PRS derived from a specific biological pathway on white matter integrity, to our knowledge. Moreover, our analysis consisted of a population-based sample of

ambulant individuals recruited to UKB. Our findings might therefore be robust and generalizable to other samples within a certain age range, although studies such as UKB are not immune to biases associated with study participation, such as collider bias (Day et al., 2016).

In addition to the large sample, the fact that NETRIN1-PRS are derived from only 43 genes, comprising approximately 0.215% of the genes in the whole genome (N = ~ 20,000) suggests that MDD risk associated variation exerts a disproportionate influence on white matter microstructure. Our findings are also further supported by permutation analysis. The association between the NETRIN1 pathway and white matter integrity is therefore likely to reflect the importance of a specific pathway in the pathophysiology of MDD.

The NETRIN1 signaling pathway has previously been found to be implicated in MDD (Zeng et al., 2017). The current study found specific neurobiological structural connectivity markers associated with this biological pathway. To our knowledge, the current study is the first one to note an association between PRS derived specifically from the NETRIN1 signaling pathway and several white matter tracts in a large genetic and neuroimaging dataset. This indicates that these brain structures may be involved in the manifestation of genetic risk of MDD and ultimately the aetiology of the disorder.

3. Chapter conclusion

In this study, a PRS calculated from SNPs within a single biological pathway was significantly associated with global, regional, and individual white matter microstructure across two DTI scalars, FA and MD. Most interesting is the association between PRS and regional and individual thalamic radiations, which lends support to the hypothesis that focusing on biological pathways with specific functions may elucidate the mechanistic genetic underpinnings of MDD. Finding such connections may in future aid in conducting more focused analyses to detect gene-sets which are defined by biologically functional mechanisms in an effort to identify treatment targets and neurobiological markers.

Chapter 4: Genetic and epigenetic prediction of Major Depressive Disorder and associations with white matter microstructure in Generation Scotland

1. Chapter introduction

In chapter 3, an MDD PRS calculated using multiple variants with a similar biological function, as well as cumulative genome-wide variants (whole-genome PRS), were associated with white matter microstructure. In complex, multifactorial disorders, equally important to the investigation of genetic risk are environmental factors and their impact on biology. One way to measure the effect of multiple environmental risk factors objectively is by looking at the epigenome (McCartney et al., 2018). Investigating the relationship between MDD and DNAm, which is environmentally modifiable, is especially important as MDD is in part the result of environmental risk factors.

Previously, both MDD and a higher genetic risk for MDD have been associated with disruptions in white matter microstructure (Shen et al., 2017; Shen et al., 2019), indicating that these disruptions are present throughout the manifestation of the disorder. Recently, a higher MDD DNAm risk score (hereafter named MRS) was associated with MDD, as well as with an archive of lifestyle factors known to impact MDD, such as smoking, BMI, and alcohol consumption (Barbu et al., 2019). Given that white matter microstructural disruptions are generally present in MDD, the aim of chapter 4 is to explore whether a higher MRS for MDD also plays a role in the above-mentioned structural deficits, as well as to identify whether this role is additive to MDD PRS. This was investigated in 621 individuals with FA data and 623 individuals with MD data in Generation Scotland (GS). Due to the fact that the current sample is smaller than in Barbu et al. (2019) (N = 625 versus N = 1,780), the ability of the MRS to predict MDD diagnosis here was also tested. The study is presented as a paper entitled, “Genetic and epigenetic prediction of Major Depressive Disorder and associations with white matter microstructure in Generation Scotland”, and is now ready for submission. As the first author, I designed the experiment, ran analyses, and wrote the manuscript.

2. Manuscript

2.1 Abstract

Background: Major Depressive Disorder (MDD) is among the most prevalent psychiatric disorders, resulting from a combination of genetic and environmental factors. Higher MDD genetic risk has been associated with white matter microstructural disruptions in association fibres and thalamic radiations. DNA methylation (DNAm), an environmentally modifiable epigenetic process, has recently been associated with MDD and lifestyle factors that play a role in MDD. It is therefore important to identify whether DNAm also plays a role in white matter microstructural disruptions in MDD, and whether this role is additive to MDD genetic risk. The current study aims to explore this relationship by using a DNAm risk score (MRS) in conjunction with a polygenic risk score (PRS).

Methods: First, penalised regression was used to train an MRS for MDD based on epigenome-wide methylation at CpG sites in N = 4,211 individuals. Next, in an independent test sample, MDD MRS and PRS were used to investigate associations with white matter microstructure as measured by FA (N = 621) and MD (N = 623) and MDD diagnosis (N = 625) and to explore whether the two risk scores acted additively.

Results: MRS ($\beta = 0.143$, $p = 0.0002$) and PRS ($\beta = 0.084$, $p = 0.039$) separately explained 2.11% and 0.69% of the variance in MDD, respectively and together explained 2.13% of the variance in the disorder (MRS: $\beta = 0.144$, $p = 0.0002$; PRS: $\beta = 0.084$, $p = 0.033$). The AUC for the MRS and PRS were 0.63 and 0.56, respectively. Neither score however was significantly associated with white matter microstructure, globally or regionally (MRS: $FA_{\beta\ range}$: 0.002 – -0.039; $MD_{\beta\ range}$: -0.002 – -0.075; PRS: $FA_{\beta\ range}$: -0.0006 – -0.078; $MD_{\beta\ range}$: -0.0006 – 0.041). The greatest effect sizes were for MRS and MD in the anterior corona radiata ($\beta = -0.075$) and PRS and FA in the superior fronto-occipital fasciculus ($\beta = -0.078$).

Conclusion: Both MRS and PRS for MDD were significantly associated with MDD, together explaining 2.13% of the variance in the disorder. However, neither score was significantly associated with white matter microstructure at corrected significant levels, indicating that larger sample sizes will be needed in future to

elucidate the link between MDD, its epigenetic signature, and white matter microstructure.

2.2 Introduction

Major Depressive Disorder (MDD) is the leading cause of global disability worldwide, currently affecting around 300 million individuals (WHO, 2017). Although the exact cause is unknown, it is thought to result from a complex combination of genetic and environmental risk factors (Otte et al., 2016).

Twin-based heritability studies indicate estimates of around 37% (Sullivan et al., 2000), and a proportion of this heritability is explained by the cumulative effect of common alleles of small effect size, as shown by genome-wide association studies (GWAS; Ripke et al., 2013; Wray et al., 2017). A recent meta-analysis investigated three large GWAS of depression (N = 807,553) and found 102 independent variants associated with depression, enriching our understanding of risk-associated variants across the genome (Howard et al., 2019).

A useful method of investigating this additive effect of common risk-associated variants is through the creation of polygenic risk scores (PRS). These are computed by adding risk alleles for a certain trait at an optimised p-value threshold, and weighing them by the strength of their association with the trait of interest. This method is especially useful in aiding downstream analyses, by associating a single score, which depicts an individual's overall risk at a given p-value threshold, with other factors known to relate to a specific trait. Although useful, the amount of variance PRS explain, particularly for MDD, is small. For instance, polygenic risk for MDD explains 1.5 – 3.2% of the phenotypic variance in MDD (Howard et al., 2019).

As shown in chapter 3, PRS for MDD may be used to explore associations with neuroimaging traits that may be implicated in MDD, such as white matter microstructure. Chapter 3 showed that MDD PRS comprising SNPs that form a specific biological pathway are associated with thalamic radiations when compared with a PRS comprising SNPs in the rest of the genome. Other studies have shown whole-genome MDD PRS associations with regional and global reductions in

microstructural integrity, with some indications of regional specificity, including in thalamic radiations and association fibres (Shen et al., 2017; Shen et al., 2019). These findings suggest that disruptions in white matter microstructure are present in MDD.

There is evidence that along with risk ascribed to specific genetic variants, effects of gene expression and regulation are also important in the manifestation of disorders and associated traits (Peedicayil et al., 2007). DNA methylation (DNAm) is one such epigenetic mechanism affecting gene expression whereby chemical changes to DNA occur, through the addition of methyl groups at cytosine-phosphate-guanine (CpG) nucleotide base pairings (Robertson, 2005). DNAm is essential for normal development, is tissue- and cell-specific, is involved in gene expression and regulation without altering DNA sequence, and can be influenced by both genetic and environmental factors (Jaenisch & Bird, 2003).

There is indeed strong evidence that such differential DNAm changes occur in complex disorders and traits (Cordova-Palomera et al., 2018). In MDD, a recent meta-analysis of multi-ethnic epigenome-wide association studies (EWAS) in multiple cohorts (N = 11,256) found three CpGs significantly associated with depressive symptoms. These sites included *CDC42BPB*, which plays a role in the regulation of cytoskeleton organisation, cell migration, and regulation of neurite outgrowth; *ARHGEF3*, which plays a role in axon guidance through co-expression with other gene families; and a third site situated in an intergenic region and associated with *SEMA4B*, which in turn interacts with *PSD-52* to promote synapse maturation (Jovanova et al., 2018). Interestingly, all three CpG sites seem to be implicated in axon guidance, suggesting a role played by DNAm in brain connectivity in the presence of depressive symptoms. These results, together with findings in chapter 3 where disruptions in white matter microstructural thalamic radiations were linked to a polygenic risk score comprising SNPs that form an axonal guidance pathway, further indicate that this pathway may be disrupted in MDD.

Recently, studies have derived DNAm predictors to predict MDD in independent testing samples. For instance, Clark et al. (2019) found a significant association between blood DNAm from 581 MDD patients at baseline with MDD status 6 years later. Using machine learning methods, they trained a DNAm risk score (MRS) of MDD to predict MDD status 6 years later. They found that MRS could

discriminate between MDD cases and controls with an area under the curve (AUC) of 0.74 (Clark et al., 2019). Barbu et al. (2019) trained an MRS on 1,223 MDD cases and 1,824 healthy individuals which was then tested in 1,780 independent individuals (363 cases and 1,417 controls). The MRS was significantly associated with MDD status, explaining 1.75% of the variance in the disorder, and was independent of PRS, which explained 2.40% of the variance in MDD. Together, the two risk scores explained 3.99% of the variance in MDD. Moreover, the MRS was significantly associated with a number of lifestyle factors implicated in MDD, such as smoking status, pack years, and alcohol consumption (Barbu et al., 2019).

Examining quantitative traits relevant to psychiatric disorders, such as brain structure and function, may elucidate mechanisms through which genetic risk and DNAm may act. Previous evidence indicates an association between MDD PRS and disrupted white matter integrity (Shen et al., 2019) as well as associations between MDD MRS and MDD-associated lifestyle factors (Barbu et al., 2019). Given these associations connecting MDD MRS to MDD-related traits, the current chapter firstly aims to explore whether a higher MRS for MDD is also associated with disrupted white matter microstructure, as well as to investigate whether the MDD MRS acts additively in relation to the PRS. The ability of the MRS to predict MDD status was also tested.

To achieve this, MDD PRS were trained on N = 807,579 from the most recent MDD GWAS (Howard et al., 2018) and MDD MRS were trained on N = 4,211 in Generation Scotland (GS) (McCartney et al., 2018), both the largest samples of genetic and DNAm currently available for MDD. The two scores were tested for associations with MDD status (N = 625), as well as white matter microstructure as measured by fractional anisotropy (FA; N = 621) and mean diffusivity (MD; N = 623), two white matter measures derived from diffusion tensor imaging (DTI) as discussed elsewhere in this thesis, in independent samples from the Stratifying Resilience and Depression Longitudinally (STRADL) cohort, a subsample of GS (Navrady et al., 2017; Habota et al., 2019). The purpose of the current study was to determine the proportion of individual and additive variance explained by an MDD PRS and MDD MRS in MDD and in white matter microstructure as measured by FA and MD.

2.3 Methods and Materials

2.3.1 Study populations

Training sample - GS

GS is a large, family-based epidemiological study and a health resource aiming to investigate the genetics of health and disease in approximately N = 24,000 individuals across Scotland, aged 19-98 years. Data was collected between 2006 and 2011, with 98.1% of the study population having available genetic data. At the time of the current study, N = 5,087 individuals had DNAm measures derived (McCartney et al., 2018). Environmental data (e.g. lifestyle, demographics) was also present in a high proportion of the study participants (Smith et al., 2006; Smith et al., 2013). GS received ethical approval from NHS Tayside Research Ethics Committee (REC reference number 05/S1401/89). Written consent was obtained from all participants.

Testing sample - STRADL

STRADL is a project aimed at studying the aetiology and stratification of depression, and was achieved through re-contacting individuals who previously participated in GS and further obtaining data on mental health, specifically depression. A total of N = 9,618 individuals responded at the re-contact recruitment stage and were assessed on numerous mental health and lifestyle measures; 1,095 were contacted for scanning, and 625 provided usable DTI data at the time of the current study. Details of recruitment and study information have been reported previously (Navrady et al., 2017). STRADL is supported by the Wellcome Trust through a Strategic Award (reference 104036/Z/14/Z). Written consent at each stage of the study was obtained from all participants.

Figure 1 below provides a flowchart detailing the participants used in the study.

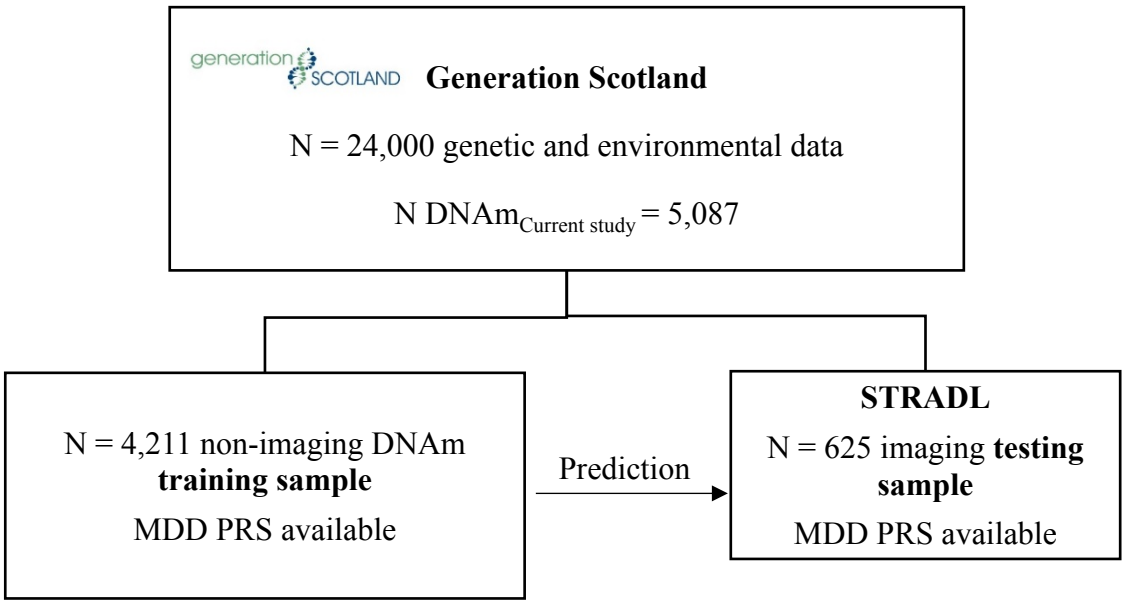


Figure 1. Flowchart detailing the number of participants with MDD PRS, MDD MRS and imaging data available. MDD PRS training sample has been taken from the most recent MDD GWAS (N = 807,579) (Howard et al., 2018) and is therefore not shown in the flowchart.

2.3.2 MDD diagnosis

The axis-I Structured Clinical Interview of the Diagnostic and Statistical Manual, version IV (SCID) was administered to participants who answered "yes" to either of two screening questions for MDD diagnosis at baseline. MDD status was measured prospectively by remote paper questionnaire between 4 and 10 years after baseline assessment (2015-2016) using the Composite International Diagnostic Interview - Short Form (CIDI-SF).

Healthy individuals used in the control group were defined as those who answered "no" to the two screening questions and did not fulfil criteria for a diagnosis of current or previous MDD following the SCID interview and CIDI-SF remote follow-up assessment. Individuals fulfilling criteria for schizophrenia or bipolar disorder, or who self-reported these diagnoses, were excluded from both case and control groups.

2.3.3 Genotyping and PRS profiling

A total of N = 20,195 individuals in GS were genotyped using the Illumina OmniExpress BeadChip. Individuals with a call rate < 98%, SNPs with a genotype rate < 98%, minor allele frequency < 1%, and p-value < 10^{-6} Hardy-Weinberg equilibrium were removed from the initial dataset. Following this, imputation was performed using the Sanger Imputation Service with the Haplotype Reference Consortium panel v1, resulting in 19,997 individuals with genome-wide data (Nagy et al., 2017; Howard et al., 2019).

Briefly, using the largest available depression GWAS (Howard et al., 2019), MDD PRS for N = 18,977 individuals were computed using Plink v1.90b4 (Chang et al., 2015) using SNPs that met a significance level of $p \leq 0.05$, in line with previous studies which have shown that this threshold explained the most variance in MDD status. Clumping was applied using a linkage disequilibrium $r^2 < 0.1$ and a 500-kb window.

2.3.4 Methylation preparation and DNAm prediction

At the time of the current study, a total of 5,087 individuals in GS had genome-wide DNAm data profiled from blood samples using the Illumina Human-MethylationEPIC BeadChip. These individuals were part of a single batch. ShinyMethyl (Fortin et al., 2014) was used to exclude samples where predicted sex mismatched recorded sex, as well as to plot the log median intensity of methylated and unmethylated signals per array; where outlying values were subsequently excluded. WaterRmelon (Pidsley et al., 2013) was then used to remove samples in which > 1% of cytosine-guanine dinucleotides had a detection p-value > 0.05; probes with a beadcount of < 3 in more than 5% samples; and probes in which > 0.5% of samples had a detection p-value > 0.05 (McCartney et al., 2018). These steps left N = 5,087 participants for analysis.

Training dataset

The final number of individuals with DNAm data used in the training dataset, following outlier exclusion as indicated above and exclusion of individuals with DTI

data, was N = 4,211. CpG sites measured in these individuals were input as independent variables in a LASSO penalised regression model using the "glmnet" function in R. Depression status was regressed on age, sex, and ten genetic principal components, and the extracted residuals from this model were input as the dependent variable in the LASSO regression model. Tenfold cross-validation was applied, and the mixing parameter was set to 1 for our LASSO penalty (Friedman et al., 2010).

Testing dataset

Using the set of CpG sites selected from the penalised regression, MRS were calculated in the testing dataset (a subset of STRADL participants who had complete PRS, DTI and DNAm data, N = 625) by summing the weights estimated in the training set. This resulted in a single continuous variable for each participant, with a higher score corresponding to a higher MRS of MDD.

2.3.5 Magnetic Resonance Imaging (MRI) acquisition and pre-processing

In the current study, DTI imaging-derived phenotypes (IDPs) pre-processed and produced locally were used. MRI acquisition was performed in two sites in Scotland, Aberdeen and Dundee.

Aberdeen

Data was acquired using a Philips Achieva 3T TX-series scanner (Philips Healthcare, Best, Netherlands) at the University of Aberdeen, with a 32-channel phased-array head coil with a back-facing mirror (software version 5.1.7; gradients with maximum amplitude 80 mT/m and maximum slew rate 100 T/m/s) (Romaniuk et al., 2019).

Dundee

Data was acquired using a Siemens 3T Prisma-FIT (Siemens Healthineers, Erlangen, Germany) at the University of Dundee, with 20 channel head and neck coil and a back-facing mirror (software version VE11, gradient with max amplitude 80 mT/m and maximum slew rate 200 T/m/s) (Habota et al., 2019).

Pre-processing – quality check and tract-based spatial statistics (TBSS)

Standard tools available from FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki) were used to quality check and exclude abnormal scans from downstream analyses. All quality checking steps were performed separately for the two scanning centres, as the acquired number of volumes differed between them ($N_{volume\ Aberdeen}$: 73; $N_{volume\ Dundee}$: 72). These included (1) correcting for eddy current-induced distortions and subject movement in the scanner; (2) skull stripping using BET at a threshold of 0.2; (3) using DTIFIT in order to compute diffusion tensor characteristics (i.e. principal eigenvectors or V1, V2, V3; eigenvalues or L1, L2, L3; fractional anisotropy (FA), mean diffusivity (MD), and others); and (4) visually checking the quality of FA images at this stage in order to exclude distorted images.

TBSS was carried out according to the ENIGMA DTI protocol (http://enigma.ini.usc.edu/protocols/dti-protocols/) for both scanning centres. Briefly, images were first slightly eroded in order to remove brain-edge artefacts as well as other outlying measures. All images were then nonlinearly registered to the ENIGMA template and all subjects were taken into 1x1x1mm standard space. A mean of all registered FA images was then calculated, in order to create a white matter skeleton. At this step, images were visually inspected in order to exclude badly registered images. Finally, a recommended threshold of FA > -0.049 was used in order to project the aligned FA data for each participant onto the skeleton created earlier. This final step created an individual FA skeleton image per subject. ROI extraction analyses using protocols provided by ENIGMA (http://enigma.ini.usc.edu/protocols/dti-protocols/) were then performed, in order to extract IDPs, including FA and MD.. This resulted in 5 unilateral tracts and 19 bilateral tracts, as well as an average measure, for all 4 DTI scalars noted above (for a list of all white matter tracts, see table 1 below). The tracts are based on the Johns-Hopkins University (JHU) DTI-based white matter atlas (Mori et al., 2005). The final number of participants before merging with PRS and MRS, following quality check and exclusion criteria, was N = 968 (details of the exclusion process can be found in Appendix 3: Figure S1).

White matter tract	Abbreviation
Average FA/MD*	aMD
Global FA/MD*	gMD
Cingulum (hippocampus)	CGH
Cingulum (cingulate gyrus)	CGC
Fornix*	FX
Fornix (cres) / Stria terminalis	FX / ST
Inferior fronto-occipital fasciculus	IFO
Superior fronto-occipital fasciculus	SFO
External capsule	EC
Superior longitudinal fasciculus	SLF
Sagittal striatum	SS
Uncinate fasciculus	UNC
Body of corpus callosum*	BCC
Genu of corpus callosum*	GCC
Splenium of corpus callosum*	SCC
Corpus callosum*	CC
Corona radiata	CR
Internal capsule	IC
Anterior corona radiata	ACR
Posterior corona radiata	PCR
Superior corona radiata	SCR
Corticospinal tract	CST
Anterior limb of internal capsule	ALIC
Posterior limb of internal capsule	PLIC
Posterior thalamic radiation	PTR
Retrolenticular limb of internal capsule	RLIC

Table 1. White matter tracts used as dependent variables in statistical analyses outlined below. * = unilateral tracts.

2.3.6 Statistical methods

All analyses were conducted using R (version 3.2.3) in a Linux environment. As GS is a family-based study, with at least one family member participating in the study (McCartney et al., 2018), ASReml-R was used in order to account for relatedness within the sample, by including pedigree information as a random effect in each model.

Association of MRS and PRS with MDD

MDD was regressed on PRS; MRS; and PRS and MRS in three separate ASReml-R models. Covariates for these models included age, sex, ten genetic principal components to control for population stratification, and smoking status and smoking pack years, as it has been shown that cigarette smoking is a strong modifier of DNAm (Lee & Pausova, 2013; Joehanes et al., 2016; McCartney et al, 2018). In

addition, using the "ROCR" R package, the predictive ability of PRS and MRS in MDD was plotted using a Receiver Operating Characteristic (ROC) curve, representing the sensitivity and specificity of the scores in relation to MDD.

Association of MRS and PRS with IDPs (FA and MD)

Firstly, for global measures of FA and MD, principal component analysis (PCA) was applied on the white matter tracts of interest (N_{tracts} = 38; for a list of the tracts included in the PCA, see Appendix 3: Table S1) in order to extract a latent measure. Scores of the first unrotated component were extracted and set as dependent variables in ASReml-R. MRS and PRS were included as independent variables, with additional covariates: sex, age, age^2, ten genetic principal components, smoking status, smoking pack years, and MRI site.

Each white matter tract (N = 24; 5 unilateral and 19 bilateral) was then included as dependent variables in separate ASReml-R models. MRS and PRS were included as independent variables, with all covariates listed above, and for bilateral tracts only, hemisphere.

2.3.7 Descriptive statistics

Table 2 provides descriptive statistics relating to the training sample. Tables 3 and 4 below provide descriptive statistics relating to the testing sample.

Descriptive statistics – training sample (N = 4,211)

Variables	**Descriptive statistics**
Depression status	
Cases (%)	1,036 (25%)
Controls (%)	3,175 (75%)
Sex	
Female (%)	2,619 (62%)
Male (%)	1,592 (38%)
Age	
Mean +/- SD, range	47.83 +/- 14.42, 18 - 95
Smoking status	
Current smoker Pack years (mean +/- SD, range)	835 19.98 +/- 18, 0.03 – 120
Former smokers who quit under a year ago Pack years (mean +/- SD, range)	117 17.31 +/- 18.36, 0.004 – 88.80
Former smokers who quit over a year ago Pack years (mean +/- SD, range)	1,082 14.31 +/- 16.60, 0.01 – 116
Never smoked tobacco Pack years (mean +/- SD, range)	2,177 -
Unsure of smoking status Pack years (mean +/- SD, range)	- -
Total	4,211

Table 2. Descriptive statistics of individuals included in training dataset; SD = standard deviation; number of pack-years = (packs smoked per day) x (years as a smoker).

Descriptive statistics – testing sample (MDD; N = 625)

Variables	**Descriptive statistics**
Depression status	
Cases (%)	122 (20%)
Controls (%)	503 (80%)
Sex	
Female (%)	378 (60%)
Male (%)	247 (40%)
Age	
Mean +/- SD, range	52.81 +/- 9.12, 20 - 72
Smoking status	
Current smoker Pack years (mean +/- SD, range)	78 25.37 +/- 18.52, 0.36 – 79.55
Former smokers who quit under a year ago Pack years (mean +/- SD, range)	12 23.54 +/- 12.68, 3.33 – 46.20
Former smokers who quit over a year ago Pack years (mean +/- SD, range)	195 16.28 +/- 18.78, 0.004 – 107.60
Never smoked tobacco Pack years (mean +/- SD, range)	333 -
Unsure of smoking status Pack years (mean +/- SD, range)	7 -
Total	625

Table 3. Descriptive statistics of individuals included in testing dataset (MDD); SD = standard deviation; number of pack-years = (packs smoked per day) x (years as a smoker).

Descriptive statistics – testing sample (FA; N = 621 and MD; N = 623)

Variables	**Fractional anisotropy (FA)**	**Mean diffusivity (MD)**
Depression status		
Cases (%)	122 (20%)	121 (19%)
Controls (%)	499 (80%)	502 (81%)
Sex		
Female (%)	376 (60%)	377 (60%)
Male (%)	245 (40%)	246 (40%)
Age		
Mean +/- SD, range	52.77 +/- 9.13, 20 - 72	52.78 +/- 9.10, 20 - 72
Smoking status		
Current smoker Pack years (mean +/- SD, range)	78 25.37 +/- 18.52, 0.36 – 79.55	78 25.37 +/- 18.52, 0.26 – 79.55
Former smokers who quit under a year ago Pack years (mean +/- SD, range)	12 23.54 +/- 12.68, 3.33 – 46.20	12 23.54 +/- 12.68, 3.33 – 46.20
Former smokers who quit over a year ago Pack years (mean +/- SD, range)	195 16.28 +/- 18.78, 0.004 – 107.60	194 16.31 +/- 18.82, 0.004 – 107.60
Never smoked tobacco Pack years (mean +/- SD, range)	330 -	332 -
Unsure of smoking status Pack years (mean +/- SD, range)	6 -	7 -
Total	621	623

Table 4. Descriptive statistics of individuals included in testing dataset (FA and MD); individuals whose global measures for FA and MD lay more than three standard deviations (SD) from the sample mean were excluded; number of pack-years = (packs smoked per day) x (years as a smoker).

2.4 Results

In the LASSO penalised regression model, 256 CpG sites with the lambda value corresponding to the minimum mean cross-validated error were extracted and applied to CpG sites in the independent testing sample (Friedman et al., 2010) (see Appendix 3: Table S2 for a list of CpG sites and their regression weights).

2.4.1 Association of MRS and PRS with MDD

ASReml-R models showed that both MRS ($\beta = 0.1433$, $p = 0.0002$, $R^2 = 2.11\%$) and PRS ($\beta = 0.0839$, $p = 0.0387$, $R^2 = 0.69\%$) explained a small proportion of variance in MDD. The model including both MRS ($\beta = 0.144$, $p = 0.0002$) and PRS ($\beta = 0.084$, $p = 0.033$) explained the most variance ($R^2 = 2.13\%$), though this was not significantly greater than MRS alone. Information relating to this can be viewed in Figure 2. The AUC of the MRS was 0.63, while the AUC of the PRS was 0.56 (Figure 3).

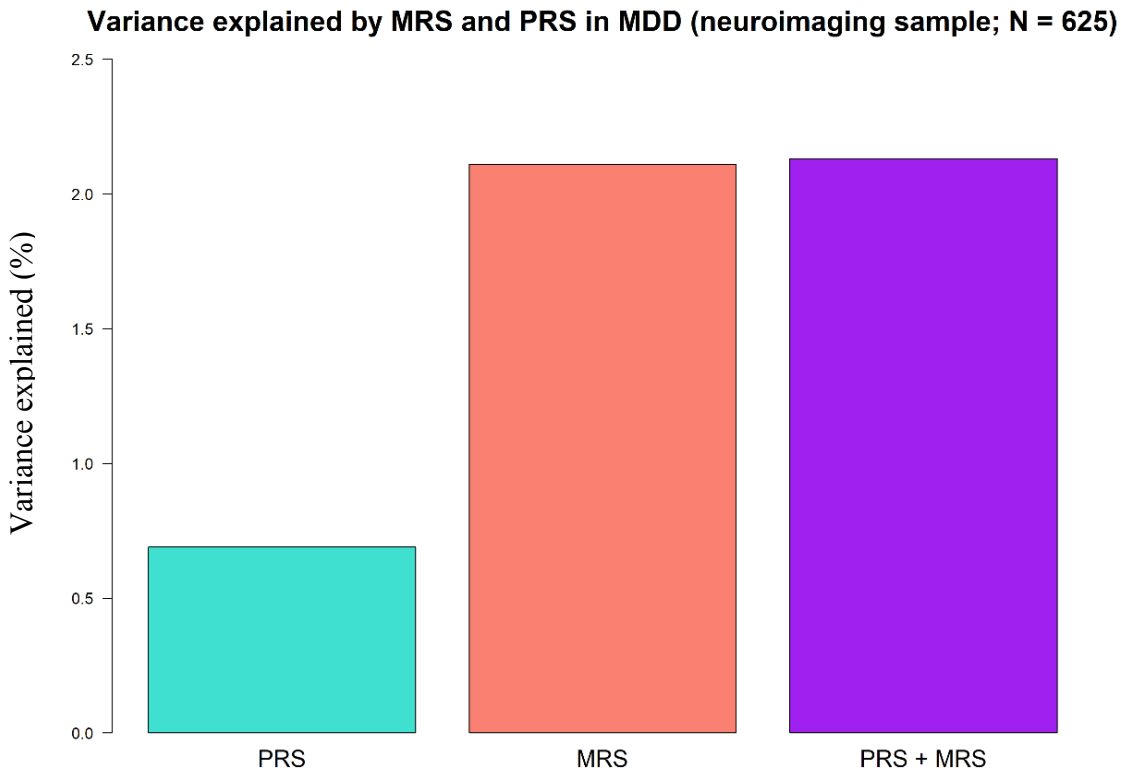


Figure 2. PRS and MRS prediction of MDD in neuroimaging sample (N = 625); variance explained (R^2) is shown as follows: PRS (blue), MRS (salmon), and additive PRS and MRS (violet) in the bar graphs above.

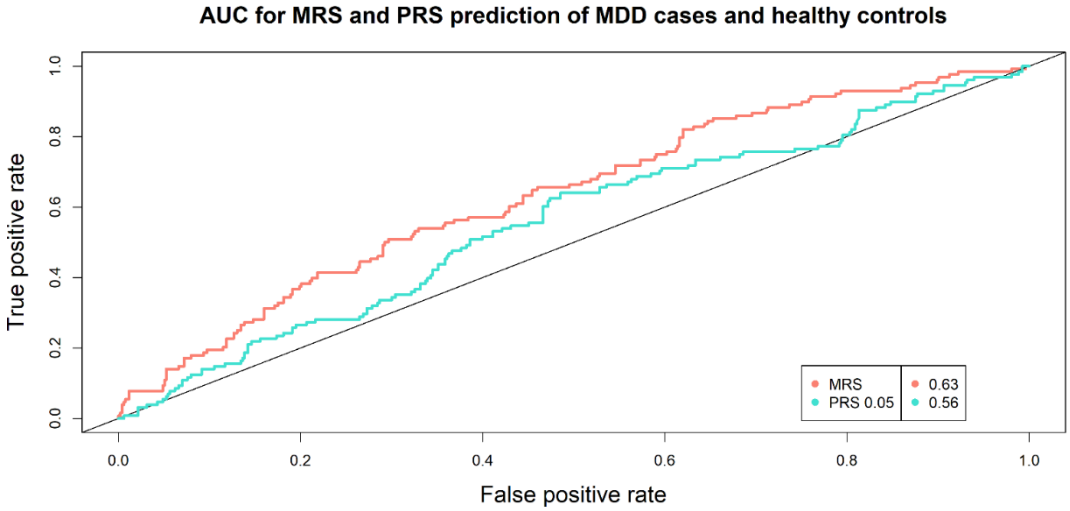


Figure 3. Receiver Operating Characteristic (ROC) curve indicating the sensitivity and specificity of MRS and PRS for MDD. The legend shows the AUC estimates for MRS and PRS.

2.4.2 Association of MRS and PRS with FA and MD

Fractional anisotropy

One white matter tract was found to be nominally significantly associated with PRS (FA; superior fronto-occipital fasciculus: β = -0.077, p = 0.022). Table 5 contains standardised effect size (β), standard error, nominal p-value, and R^2 for the association of both MRS and PRS with global measures and individual white matter tracts.

White matter tract	**MRS**				**PRS**			
	Effect size, β	**SD**	**P-value (nominal)**	**R^2 (%)**	**Effect size, β**	**SD**	**P-value (nominal)**	**R^2 (%)**
ACR	0.015	0.038	0.693	0.023	-0.014	0.034	0.673	0.021
ALIC	-0.004	0.039	0.901	0.003	-0.04	0.035	0.255	0.167
CGC	0.01	0.038	0.776	0.012	-0.058	0.034	0.093	0.341
CGH	0.023	0.033	0.488	0.054	-0.011	0.029	0.7	0.013
CR	0.004	0.038	0.914	0.002	-0.025	0.034	0.469	0.065
CST	0.007	0.04	0.856	0.005	0.018	0.035	0.612	0.033
EC	-0.027	0.038	0.469	0.08	-0.037	0.034	0.281	0.14
FX / ST	-0.015	0.032	0.64	0.024	-0.052	0.029	0.077	0.272
IC	-0.011	0.035	0.736	0.015	-0.032	0.031	0.302	0.108
IFO	-0.032	0.037	0.383	0.108	0.005	0.033	0.874	0.003
PCR	0.024	0.039	0.534	0.061	-0.012	0.034	0.713	0.017
PLIC	-0.014	0.034	0.676	0.021	-0.026	0.03	0.388	0.07
PTR	0.024	0.037	0.514	0.061	0.005	0.033	0.869	0.003
RLIC	-0.007	0.034	0.818	0.006	-0.013	0.03	0.667	0.018
SCR	-0.022	0.04	0.58	0.052	-0.031	0.036	0.387	0.101
SFO	0.003	0.037	0.925	0.001	-0.077	0.033	**0.022**	0.598
SLF	-0.021	0.039	0.577	0.05	0.004	0.035	0.901	0.002
SS	0.019	0.037	0.597	0.041	-0.042	0.033	0.207	0.186
UF	0.012	0.039	0.759	0.015	-0.0004	0.035	0.987	0
CC	-0.026	0.039	0.503	0.071	-0.029	0.035	0.406	0.086
BCC	-0.038	0.042	0.362	0.152	-0.043	0.037	0.253	0.189

GCC	-0.019	0.039	0.612	0.041	-0.016	0.035	0.64	0.027
SCC	-0.002	0.033	0.949	0	-0.006	0.03	0.821	0.005
FX	-0.009	0.037	0.802	0.009	-0.056	0.033	0.094	0.324
Global FA	0.001	0.036	0.96	0	-0.038	0.032	0.242	0.153
Average FA	-0.014	0.038	0.712	0.02	-0.033	0.034	0.328	0.113

Table 5. The association between MRS and PRS with white matter tracts (FA). Nominally significant p-values are shown in bold. The first column for MRS and PRS indicates standardised effect size (β). R^2 = estimate of variance explained by each pathway in %.

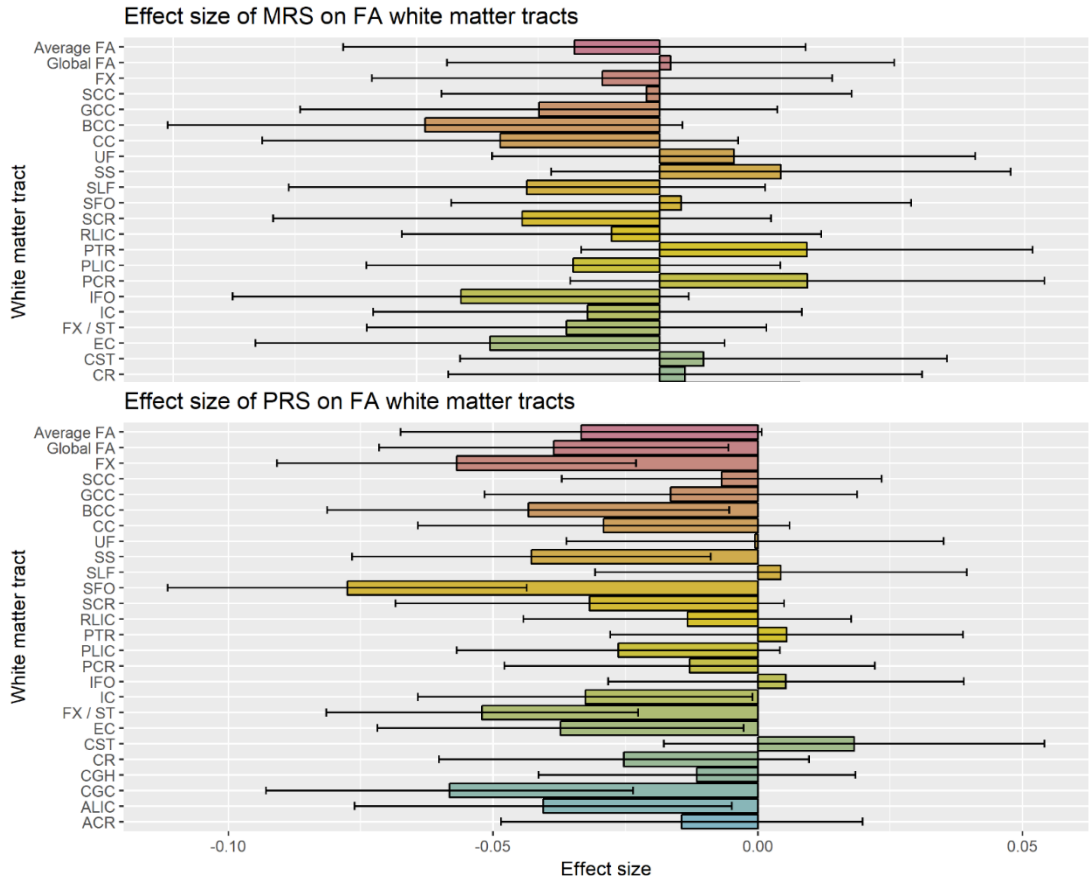


Figure 4. The effects of MRS (above) and PRS (below) on fractional anisotropy (FA) values of white matter tracts. The x-axis indicates the standardized effect size of each score association, and the y-axis indicates the white matter tracts. The error bar represents the SD of the mean.

Mean diffusivity

Table 6 contains standardised effect size (β), standard error, nominal p-value, and R^2 for the association of both MRS and PRS with global measures and individual white matter tracts.

White matter tract	**MRS**				**PRS**			
	Effect size, β	**SD**	**P-value (nominal)**	**R^2 (%)**	**Effect size, β**	**SD**	**P-value (nominal)**	**R^2 (%)**
ACR	-0.074	0.041	0.068	0.578	0.002	0.036	0.945	0.001
ALIC	-0.061	0.038	0.113	0.39	0.04	0.034	0.24	0.166
CGC	-0.054	0.034	0.115	0.31	0.026	0.03	0.391	0.071
CGH	-0.013	0.028	0.641	0.018	0.024	0.025	0.345	0.059
CR	-0.064	0.039	0.107	0.428	0.018	0.035	0.61	0.033
CST	0.024	0.04	0.544	0.059	0.002	0.035	0.933	0.001
EC	-0.035	0.037	0.346	0.129	-0.005	0.033	0.874	0.003
FX / ST	-0.01	0.03	0.738	0.011	0.008	0.027	0.741	0.008
IC	-0.044	0.029	0.138	0.208	0.014	0.026	0.594	0.021
IFO	-0.008	0.035	0.816	0.007	0.04	0.031	0.194	0.167
PCR	-0.055	0.036	0.13	0.31	0.027	0.032	0.388	0.078
PLIC	-0.022	0.028	0.441	0.05	-0.004	0.025	0.874	0.002
PTR	-0.049	0.035	0.168	0.248	0	0.031	0.985	0
RLIC	-0.034	0.028	0.23	0.122	0.006	0.025	0.796	0.004
SCR	-0.037	0.038	0.329	0.147	0.026	0.034	0.437	0.072
SFO	-0.002	0.039	0.951	0.001	-0.008	0.035	0.802	0.008
SLF	-0.011	0.036	0.763	0.013	0.023	0.032	0.475	0.055
SS	-0.05	0.035	0.154	0.268	0.005	0.031	0.856	0.003
UF	-0.052	0.039	0.188	0.276	0.015	0.035	0.661	0.024
CC	-0.016	0.037	0.663	0.027	0.029	0.033	0.377	0.088
BCC	0.007	0.04	0.853	0.006	0.038	0.035	0.281	0.149
GCC	-0.048	0.038	0.205	0.241	0.019	0.034	0.574	0.037
SCC	-0.016	0.035	0.642	0.028	0.017	0.031	0.572	0.032

FX	-0.011	0.04	0.78	0.013	0.013	0.036	0.703	0.019
Global MD	-0.032	0.039	0.411	0.106	0.019	0.035	0.578	0.038
Average MD	-0.017	0.037	0.646	0.03	0.022	0.032	0.502	0.05

Table 6. The association between MRS and PRS with white matter tracts (MD). The first column for MRS and PRS indicates standardised effect size (β). R^2 = estimate of variance explained by each pathway in %.

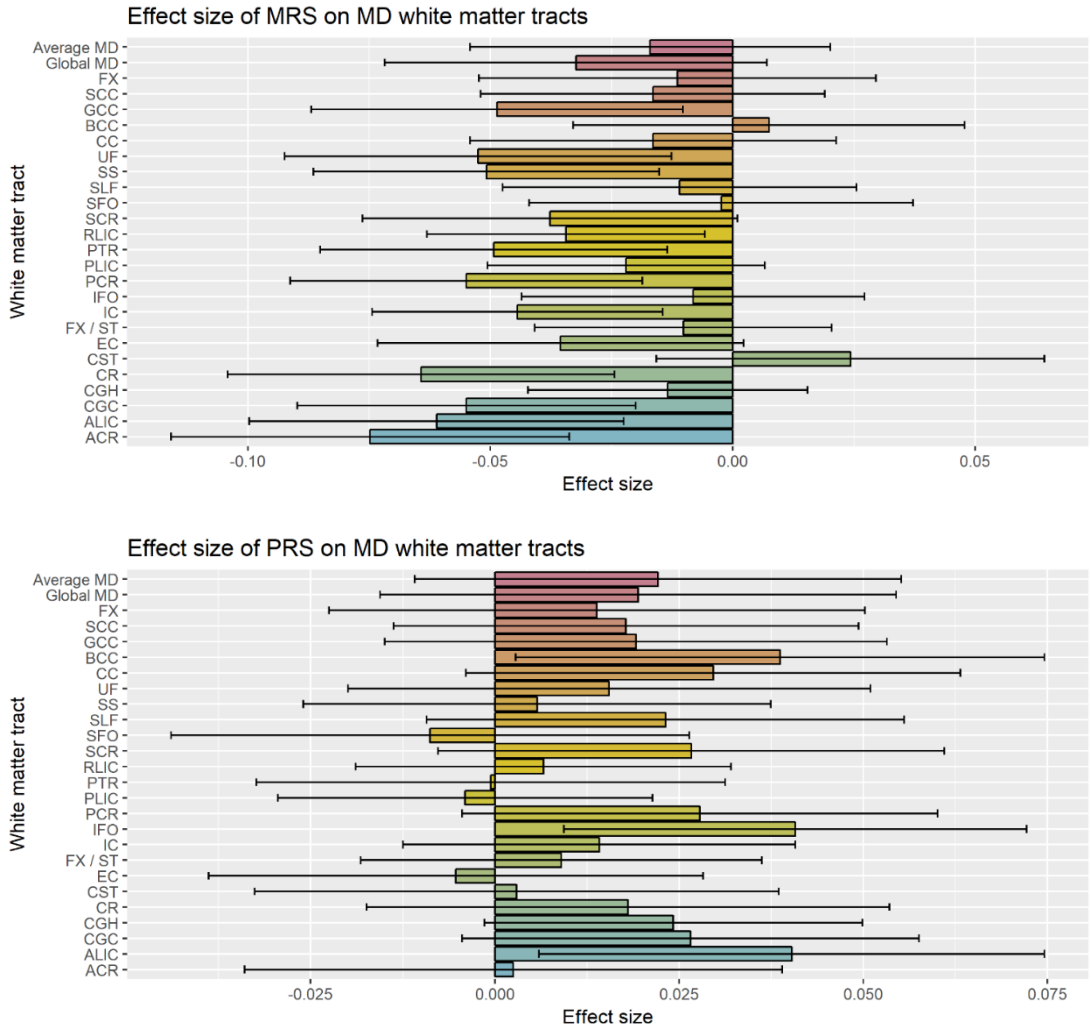


Figure 5. The effects of MRS (above) and PRS (below) on mean diffusivity (MD) values of white matter tracts. The x-axis indicates the standardized effect size of each score association, and the y-axis indicates the white matter tracts. The error bar represents the SD of the mean.

2.5 Discussion

The aim of the current study was to investigate whether MDD poly-epigenetic risk scores were significantly associated with MDD and changes in white matter microstructure as measured by FA and MD in > 600 individuals and to observe whether these associations are independent from MDD polygenic risk scores.

DNAm predictors for MDD were identified in a training dataset of 4,211 individuals and were significantly associated with MDD in an independent testing dataset. The study showed that the MRS explained 2.11% of the phenotypic variance in MDD, as compared with MDD PRS, which only explained 0.69%. Together, the two risk scores explained 2.13% of the variance in MDD.

PRS derived from GWAS have offered insight into how cumulative risk from a large number of common genetic variants of small effect relate to MDD (Howard et al., 2019). On the other hand, studies of differential DNAm in MDD have only recently become possible at sufficient scale with the availability of large datasets including epigenomic and diagnostic data. For instance, Jovanova et al. (2018) found 3 CpG sites associated with depressive symptoms at epigenome-wide significance in N = 11,256 individuals.

Moreover, MRS based on DNAm in large datasets have also recently shown interesting results implicating environmental and lifestyle factors in the relationship between DNAm and MDD. Clark et al. (2019) showed that an MDD MRS, in combination with 27 lifestyle characteristics, including smoking status, alcohol consumption, body mass index (BMI), and physical activity, could discriminate between MDD cases and controls with an area under the curve (AUC) of 0.742. Moreover, Barbu et al. (2019) showed associations between MDD MRS and MDD status as well as numerous lifestyle factors, including smoking status, pack years, alcohol consumption, and BMI, which are known to play a role in MDD (Paperwalla et al., 2004; De Wit et al., 2010; Briere et al., 2014; Opel et al., 2015). They concluded that MRS may reflect lifestyle factors, indicating that some of the variation in MDD may be explained by environmental factors through DNAm. Results here show that both MRS and PRS are significantly associated with MDD, although their contribution is not independent from each other. However, Barbu et al. (2019) show that MRS and

PRS do have an additive nature in MDD prediction, perhaps due to an increased sample size and statistical power.

The findings here and in the studies above aid in elucidating a role played by DNAm in MDD. However, unlike fixed genetic factors, DNAm changes throughout life, which may either be a cause or a consequence of altered environmental and lifestyle factors. Due to the temporal variation of DNAm, reverse causality may arise in cross-sectional studies where DNAm samples are collected at the same time as a diagnosis is made (Walton et al., 2019). In other words, it may be that individuals with MDD have differentially methylated CpG sites as a result of disorder manifestation, leading to changes in their environment and lifestyle, or that their altered lifestyle may lead to differential DNAm. It may be possible, in future, to investigate direction of causality using methods such as mendelian randomisation, which measures variation in genes with a known function to examine causality (Lawlor et al., 2008). However, this approach would need to be repeated longitudinally, due to the dynamic nature of DNAm across life. Future studies could also measure DNAm before a diagnosis is made, as this may reduce confounding by reverse causation, although this approach does not completely reverse the risk of confounding (Juvinao-Quintero et al., 2019).

In the current study, MRS and PRS were not associated with white matter tracts as measured by FA or MD. To the author's knowledge, this is the first study attempting to investigate the association between a DNAm-based risk score for MDD and white matter tracts. A previous study has shown that elevated levels of DNA methylation in *SLC6A4*, a gene previously associated with both depression and white matter microstructure, was associated with decreased FA in the body of corpus callosum in MDD patients, although this analysis was based on five focussed CpG sites within the gene (Won et al., 2016). Choi et al. (2015) also found that differential DNA methylation at four CpG sites in the *BDNF* promoter region, previously associated with both white matter microstructure and depression, was associated with decreased integrity in the right anterior corona radiata of MDD patients. Importantly, these previous studies used *a priori* hypotheses linking specific genes and methylation within them to both traits, unlike the current study, which used a less biased data-driven approach.

PRS have previously shown associations with a wide range of neuropsychiatric traits and white matter microstructure measures, as indexed by FA and MD (Shen et al., 2019). Shen et al. (2019) conducted a phenome-wide association study in 21,888 individuals to explore how PRS at different p-value thresholds associate with behavioural and neuroimaging traits. They showed that MDD PRS $p \leq 0.01$ showed the largest effect sizes in neuroimaging phenotypes (Shen et al., 2019). This indicates that the PRS p-value threshold used here ($p \leq 0.05$) may not be optimal in detecting meaningful associations, although the threshold was selected based on its ability to explain the most variance in MDD status (Howard et al., 2019). Moreover, in addition to significant associations between individual white matter tracts and MDD PRS, Shen et al. (2019) also found evidence of global and regional associations, for which effect sizes were larger. This may indicate that the effect of the two risk scores may be global rather than tract-specific, although this is not reflected in the global and average FA and MD associations with either score in the current study.

Moreover, Barbu et al. (2019) investigated whether MRS and PRS for MDD capture different exposures to behavioural and environmental phenotypes. They found that the MDD MRS was more significantly associated with sociodemographic and lifestyle measures, while the MDD PRS was more significantly associated with disease and mental health variables (Barbu et al., 2019). In the current study, both risk scores were associated with MDD, but none with white matter microstructure. Given previous evidence relating MDD PRS to decreased white matter integrity and more robust associations with mental health variables, it may be that the genetic risk score is more well-suited to identify disruptions in white matter in relation to MDD, while epigenetic risk may exert its effect on MDD through environmental modifications, rather than through changes in white matter microstructure.

Although associations are non-significant, this is also reflected in the direction of effect from each risk score; the PRS seem to relate to decreased FA and increased MD in most tracts, an indication of white matter microstructural disruptions. This reflects previous findings associating higher MDD PRS with decreased white matter microstructural integrity (Shen et al., 2019; Barbu et al., 2019). On the other hand, the MRS seem to indicate increased FA and decreased MD in a large number of tracts, suggesting increased white matter integrity. Moreover, the PRS explained a greater

proportion of the variance in multiple white matter tracts in FA (FA $R^2 = 0 – 0.6\%$; MD $R^2 = 0 – 0.17\%$), while the MRS explained a greater proportion of the variance in MD-measured tracts (FA $R^2 = 0 – 0.15\%$; MD $R^2 = 0.001 – 0.58\%$) (Tables 5 and 6).

Previous studies investigating PRS and white matter microstructure associations have used sample sizes larger than 5,000 (Shen et al., 2019; Barbu et al., 2019). In addition, lack of significant associations may be due to the current sample, which is a relatively healthy community-based sample that may not reflect severe depression or depressive symptoms. These results together indicate that a larger sample size might be needed to detect an association between increased risk of depression, both polygenic and poly-epigenetic, and white matter microstructure.

A strength of the current study is the analysis between a novel MDD MRS and white matter microstructure as measured by FA and MD. Moreover, findings revealed an association between DNAm risk and MDD, indicating a potential to use such a score for further analyses, as well as for other traits which are implicated in MDD.

In summary, results show that MDD MRS and PRS are associated with MDD. Results suggest that a larger sample may be needed to uncover robust associations between white matter microstructure and both MDD risk scores. Moreover, based on previous findings, DNAm may contribute to MDD via environmental and lifestyle factors, rather than through disruptions in white matter microstructure. Further testing and validation in clinically ascertained samples is needed, however the findings here may justify future efforts to collect DNAm in larger samples and investigate associations between DNAm risk and emotional, cognitive and other brain imaging traits related to depression.

3. Chapter conclusion

The study found an association between both MRS and PRS with MDD status, although no associations were found between the two risk scores and white matter microstructure post-FDR correction. A small-sized sample comprised of community-based, generally healthy individuals, may reflect the non-significant findings here, as well as the fact that global and regional brain connectivity associations are generally

more robust than individual tracts, as reflected by previous studies. Moreover, as DNAm is environmentally modifiable, it may be that changes in lifestyle and environment, rather than disruptions in white matter microstructure, may connect DNAm to MDD prevalence. In conclusion, larger studies comprising genetic, epigenetic, and neuroimaging data will be needed in future to examine the role of an MRS in white matter microstructure and investigate whether this score is independent to MDD PRS.

Chapter 5: Discussion

1. Introduction

The current thesis aimed to investigate the genetic relationship between white matter microstructure and gene expression, as well as to identify its association with stratified genetic and epigenetic risk for MDD. The thesis included two large-scale cohorts, UKB and GS (and sub-sample STRADL), which combine neuroimaging and genetic data, with samples ranging from 620 to 14,500.

In the past, white matter microstructure has been reported to be moderately heritable and associated with MDD, both globally and regionally. As white matter represents the brain's connectivity network, having a far-reaching structural and temporal effect, there is rationale to study its relationship to psychiatric and neurological disorders. The aims of this thesis were therefore to (1) investigate the genetic basis of gene expression changes in relation to white matter microstructure, in order to form a basis for in-depth downstream analyses of disease- and trait-linked genes; (2) stratify genetic risk for MDD by a validated biological pathway and investigate its association with white matter microstructure; and (3) objectively investigate the effect of potential environmental insults by analysing epigenetic risk of MDD in relation to the disorder and white matter microstructure. To do this, increasingly specific genetic analysis approaches were used, all of which included computing scores that aggregate the cumulative effect of multiple genetic variants and CpG sites for (i) gene expression; (ii) genetic risk for MDD; and (iii) epigenetic risk for MDD.

The three research chapters each include a discussion section which is specifically tailored to the analysis and findings presented there. Therefore, the aim of this chapter is to provide a broader discussion of the findings and how they interconnect in the investigation of MDD. The chapter then concludes with strengths and limitations, suggestions for future research, and conclusions.

2. Summary of main findings

2.1 Genetic underpinnings of gene expression in white matter microstructure – specific and global findings

To investigate the relationship between white matter microstructure and genetic risk of complex disorders, the genetic underpinnings of white matter were first explored. Although heritability of white matter microstructure has been previously established, the role of the genetic variants involved is unknown (Sprooten et al., 2014). In the current thesis, to gain understanding of the functional effects of regulatory variants, the genetic basis of gene expression was investigated in relation to white matter tracts, globally and with increasing regional specificity.

One of the main findings was the association between higher white matter microstructural integrity and genetic variants regulating neural development-linked genes, and lower white matter microstructural integrity and genetic variants regulating disease-linked genes. The genes found here are different in functionality, and findings from this chapter allow for in-depth insight into expression-based effects of regulatory loci on white matter microstructure. As a result, future studies may investigate differential genotypes at regulatory loci and differential gene expression between patients and healthy control participants in downstream analyses combining neuroimaging and genetic data.

Furthermore, while all other white matter tracts were found to be associated with genetic variants regulating gene expression of either disease- or development-linked genes, the forceps minor was found to be associated with both. Interestingly, the forceps minor forms the anterior part of the corpus callosum, and connects homologous prefrontal cortex regions between hemispheres, thus enabling communication between the two (Wakana et al., 2007). It is reported to be involved in numerous cognitive and behavioural skills, as well as neuropsychiatric and neurological disorders (Mamiya et al., 2018). This finding is therefore unsurprising, as the corpus callosum is less prone to errors during the imaging process and is arguably the largest white matter tract in the brain which connects a large number of brain

regions to each other (Hofer & Frahm, 2006). Undoubtedly, a large number of genes may be expressed in the formation, maintenance, and pathology of this tract.

In addition to these specific findings, differences were also found globally for some of the genes investigated, suggesting that the expression of some genes has a more widespread effect on white matter microstructure than others. In addition to analysis of disease states and traits, these findings may be leveraged in downstream analyses to investigate loci implicated in the formation, development, and plasticity of white matter microstructure globally.

As discussed in chapters 1 and 2, neuroimaging phenotypes provide a novel and sound opportunity to investigate the genetics of gene expression in relation to *in vivo* brain phenotypes. This method accounts for the increasing number of limitations in analysing gene expression in the brain directly, such as cause of death and post-mortem expression level differences (McKenzie et al., 2014). For a comprehensive understanding, the findings discussed here involve changes in both FA and MD across tracts in relation to regulatory loci. However, these loci may implicate FA and MD measures of white matter microstructure across different tracts, regionally or specifically, as the two scalars capture different characteristics of white matter microstructural integrity (Jones et al., 2013).

Lastly, the findings in chapter 2 uncovered novel associations which were not previously reported by GWAS (Elliott et al., 2018), suggesting that genetic loci important in white matter maintenance and pathology are regulatory. This is additionally important as future studies may leverage these regulatory loci to investigate their direct effects on both traits and disease states through the alteration of white matter microstructure.

2.2 Thalamic radiations are key neurobiological markers in stratified genetic risk for MDD

Polygenic risk of MDD has continuously been associated with white matter microstructure in the past (Whalley et al., 2017; Shen et al., 2017). However, findings have generally been inconclusive, with numerous tracts being associated with

increased risk of MDD. This has made it difficult to uncover genetic risk factors and their effect on brain connectivity in the context of MDD. Here, findings concerning thalamic radiations as both white matter tracts and stratified biological pathway process are discussed.

As discussed previously, MDD is a highly heterogeneous disorder, both clinically and biologically. Methods to stratify MDD have been considered in order to gain an understanding in the aetiology and manifestation of the disorder. Here, genetic risk for MDD was stratified based on genetic variants aggregated within a biological pathway. The third chapter found higher general MD in thalamic radiations and superior and inferior longitudinal fasciculi, as well as lower FA in superior longitudinal fasciculus associated with PRS computed using variants aggregated within the NETRIN1 Signalling Pathway. While several white matter tracts including tracts pertaining to association and projection regional fibres were associated with PRS computed from variants outside the pathway, interestingly, they were not associated with thalamic radiations.

This result is fitting as the NETRIN1 Signalling Pathway is responsible for neuronal migration and guiding axons branching from the thalamus to the rest of the cortex during neuronal development (Braisted et al., 2000). The thalamus is a subcortical structure located above the brain stem with widespread connections to both cortex & subcortical areas (Sherman, 2016). The thalamus is often referred to as the hub of the brain, as it is linked to cortical areas globally as well as to various subcortical structures, such as the hippocampus and amygdala, and uses these global connections to relay information between cortical and subcortical structures (Sherman, 2016).

Therefore, the hub is implicated in negative emotional processing, cognitive functions such as memory, executive functions, attention, and information processing, and is known to regulate states of sleep and wakefulness (Herrero, Barcia & Navarro, 2002; Saalmann & Kastner, 2011; Yousaf et al., 2018). These are all factors contributing to the MDD symptom profile (e.g. inability to concentrate, insomnia and hypersomnia, enhanced negative emotional states), so it is unsurprising that a heterogeneous disorder such as MDD is associated with a structure that is so widespread in its functionality.

Moreover, PRS confined to the NETRIN1 Signalling Pathway also showed associations with disrupted microstructural integrity in large association fibres, such as the superior and inferior longitudinal fasciculi, which connect different cortical areas across the four lobes to each other (Schmahmann et al., 2007). This finding fits well with the thalamocortical connections, indicating that stratified risk of MDD in this particular biological pathway is related to lower white matter microstructural integrity in tracts connecting cortical and subcortical regions throughout the brain.

The findings in chapter 3 indeed show a strong connection between genetic risk for MDD aggregated in a biological process and brain connectivity, both implicating the thalamus. The results indicate that stratifying MDD by biology may uncover novel insights into specific connectivity deficits related to the disorder. In downstream analyses, stratification of both symptom profiles and genetic risk may lead to specific genetic variants linked to particular symptoms. In addition, future studies may attempt to investigate functional connectivity in relation to stratified MDD genetic risk.

2.3 Whole-epigenome DNAm identified as a novel risk factor for MDD

Genetic studies have only recently garnered success in uncovering part of the genetic basis of psychiatric disorders (Howard et al., 2019). As MDD is a multifactorial, complex disorder, with both genetics and the environment playing a pivotal role in its development, it is safe to assume that research investigating the disorder would benefit from an integrated approach including both genetic and epigenetic risk factors. Moreover, MDD is reported to have a heritability of 37% based on family studies, however GWAS indicate that common genetic variants explain only part of this total heritability (Howard et al., 2019). Therefore, a proportion of variance in MDD may be explained by changes in gene expression induced by epigenetic factors. Support to this is lent in chapter 4, where both whole-genome and whole-epigenome risk explained a small proportion of variance in MDD (additive R^2 = 2.13%), indicating that epigenetic mechanisms may be important in the formation and manifestation of MDD.

Research has so far focused on DNAm alterations of specific genes posited to be associated with MDD, such as *BDNF, SLC6A4,* and *NR3C1,* although it is now widely believed that candidate genes may not be an optimal way to investigate MDD, due to its polygenicity and complex nature (Border et al., 2019). The findings in chapter 4 are one of the first to indicate that a whole-epigenome approach may be more indicative in uncovering novel risk factors for MDD. As it is one of the first studies to investigate epigenetic risk for MDD aggregated in a single variable, the research presented here provides a basis for future epigenetic-based analyses for MDD. For instance, future studies may investigate associations between environmental and lifestyle factors implicated in MDD, such as childhood trauma, smoking status, alcohol consumption, and body mass index (BMI) in relation to DNAm risk for MDD. Furthermore, DNAm signatures of antidepressants, one of the most widely-used treatments for MDD acting on biological pathways, may be investigated in future to observe whether differential DNAm exists between those who take and do not take antidepressants.

Moreover, it is perhaps unsurprising that epigenetic alterations, situated at the intersection between genetic and environmental factors, play a role in MDD. A number of lifestyle and environmental insults, such as childhood adversity, work-related stress, smoking, and alcohol, are associated with MDD, many of which may silence or activate specific genes through hyper- or hypo-methylation at promoter sites to give rise to the disorder. As this study is relatively novel due to the rarity of studies containing large DNAm data, it is presently unclear in what way epigenetic modifications influence MDD. Epigenetic alterations may well be one of the mechanisms integrating both environmental and genetic risk factors in MDD, and combined analyses that include a wide variety of environmental, genetic, and epigenetic risk factors, should be carried out.

2.4 No association revealed between MRS for MDD and white matter microstructure

The study additionally set out to investigate links between whole-epigenome and whole-genome risk for MDD and white matter microstructure. As with MDD

diagnosis, the association between DNAm alterations at specific sites across the genome posited to be related to MDD and disrupted white matter microstructure has been previously established, providing a rationale to investigate links between the two (Won & Ham, 2016). However, in the current study, there was no association between a genetic or an epigenetic risk score for MDD and white matter microstructure, despite previous evidence associating whole-genome PRS for MDD with the brain connectivity network (Whalley et al., 2013; Shen et al., 2017).

The null findings here may be due to a number of factors. Firstly, the sample size comprising non-clinically ascertained individuals used in the study (N_{FA} = 621; N_{MD} = 623) is small compared to usual neuroimaging genetics studies. Previous studies showing an association between genetic risk for MDD and white matter microstructure contained sample sizes of over 1,000 individuals (Shen et al., 2017; Shen et al., 2019). Reus et al. (2017) computed PRS for MDD, schizophrenia, and bipolar disorder, and associated them with subcortical brain volumes (N = 978) and white matter microstructure (N = 816). The authors found no link between subcortical volumes or white matter microstructure and PRS for either disorder, although their findings may be due to formerly underpowered GWAS which led to scarce common genetic variants for use in the calculation of PRS. In addition, the study used the first release of UK Biobank imaging data (Reus et al., 2017); later releases adding participants to this original number and the more successful findings indicated that the lack of findings may have been due to small sample size.

Secondly, previous studies investigating a whole-epigenome MRS in association with various traits used sample sizes of approximately 900 individuals, indicating the need for a larger sample size here (Shah et al., 2015; McCartney et al., 2018). Moreover, it would be reasonable that genetic risk for MDD would be associated with white matter microstructure, which is also moderately heritable (Elliott et al., 2018), and epigenetic risk for MDD would be associated with lifestyle and environmental factors, which partly influence the epigenome. Barbu et al. (2019) found that an MDD MRS was more significantly associated with sociodemographic and lifestyle measures, while an MDD PRS was more significantly associated with disease and mental health variables (Barbu et al., 2019). The association between an MRS for MDD and lifestyle factors, including BMI, smoking and alcohol

consumption, and self-reported antidepressant use, was also shown in Barbu et al. (2019). It therefore remains to be seen, as sample sizes increase and analysis methods advance, whether whole-epigenome MDD risk is associated with brain connectivity.

3. Strengths and limitations of the current thesis and suggestions for future research

Two major strengths for the studies conducted in this thesis are (1) the large sample size within UK Biobank, which can accommodate both biological and clinical heterogeneity of MDD and (2) a combination of neuroimaging and genetic data in these large samples. Firstly, UK Biobank is an invaluable resource combining a vast amount of data; this includes neuroimaging data collected at only two sites, thus accounting for limitations and artefacts resulting from scanning individuals across multiple sites; and genetic data, which has now been released for approximately 500,000 individuals, and permits investigation into a large number of phenotypes in relation to genotype.

Moreover, although sample size within GS is small for genetic-neuroimaging associations, it is important to note that the cohort is a rich resource containing invaluable data, including a combination of genetic, neuroimaging, and DNAm in a carefully chosen sample. In addition to this, the individuals for which neuroimaging data is available were specifically chosen to study resilience and depression (Navrady et al., 2017), which adds to the value of investigating the above-mentioned data in relation to the disorder.

As discussed in the introductory chapter, neurobiological markers implicated in psychiatric disorders may provide a mechanistic insight into the formation and manifestation of disease states. Integrating both neuroimaging and genetic data in the investigation of psychiatric disorders may therefore pave the way to further specialised studies and uncover therapeutic targets to be used for prevention and treatment.

One of the limitations present in both datasets used in this thesis is the cohorts' age range, which generally reflect older populations (Mean age: UKB: 56.52 +/- 8.09 years; STRADL respondents of GS: 50.48 +/- 13.41 years), as well as the fact that

participants in both cohorts are generally healthier and wealthier than the rest of the population. This may induce some bias in the interpretation of the results implicating MDD, as the average age of onset is 25, a much younger age than those of participants in the studies, although MDD may appear at any age (WHO, 2017). Moreover, in a study presented in the introduction, Bromel et al. (2011) showed that MDD 12-month prevalence was similar between high- and low-income countries. While these factors may not have a great impact on MDD, it is still advised to carefully consider them when interpreting findings for further analyses.

Further, although the large sample sizes used here are lauded, data from a higher number of participants still must be collected in order to be able to conduct more complex and in-depth genetic and neuroimaging analyses. Stratification by biology or genetic factors, such as biological pathways, haplotype blocks, or genetic correlations, and even more general genetic analyses such as GWAS, may need hundreds of thousands of individuals, especially in the investigation of MDD, where different combinations of genes and SNPs act together to give rise to the disorder. Replication of findings between the already-existing large studies may strengthen the conclusions made so far and encourage further studies to carefully select participants for future investigation. These findings may also be used to generate hypotheses to test in smaller, but still substantial, genetic neuroimaging studies, incorporating a discovery and replication approach.

With regards to the neuroimaging data, two tractography-based methods, probabilistic tractography and TBSS, were used in the present thesis. As discussed in the introduction, the two methods are both highly validated and sound measures of capturing white matter microstructure. However, the two methods may well have different proportions of heritability (Elliott et al., 2018), and both utilise different methods to construct and annotate white matter tracts. Replication across both methods in the studies presented was considered beyond the scope of this thesis, although in future, studies should take into account the differences between the two and attempt to investigate both.

Moreover, findings were consistently different between the two scalars investigated, FA and MD. In some instances, significant findings were associated with

one scalar, but not the other. This is not a limitation in itself, however it should be mentioned that the two scalars may capture different characteristics of white matter microstructure and should be carefully considered when drawing conclusions from studies using such measures.

Finally, overall, very small effect sizes were reported in the three studies (Largest effect size: Chapter 2: -0.0561 (FA) and 0.0480 (MD); Chapter 3: -0.036 (FA) and 0.042 (MD); Chapter 4: 0.1440 (MDD MRS). While this may not necessarily be a limitation, it does provide a rationale for better-defined phenotypes in larger groups in future studies. Especially in MDD, stratification may provide an advantage in that more specific patient groups may show greater associations with particular phenotypes.

In addition to the suggestions for future studies made above, direct implications of the current thesis to be considered by further research are threefold. Firstly, more detailed investigation should be carried out in analysing genetic underpinnings of MDD. Future studies may wish to look at localised genetic effects aggregated in different functional and biological pathways and perhaps integrate gene expression-based analyses of participating SNPs. For instance, the 8 eQTL scores uncovered in chapter 2 may further be analysed and tested in knockout animal models to investigate their possible role in MDD and white matter microstructure. Furthermore, in-depth investigation should be carried out into the NETRIN1 Signalling Pathway, and axon guidance pathways in general, as they seem to emerge in MDD analyses (Zeng et al., 2016; Aberg et al., 2018).

Moreover, novel developments in diffusion MRI measures, such as NODDI (neurite orientation dispersion and density imaging), which measures intra-neurite, extra-neurite, and cerebral spinal fluid volume fractions separately, may be employed by studies in the future to investigate more localised disruptions in white matter microstructure in relation to genetic and epigenetic risk for MDD (McCunn et al., 2019). Lastly, white matter microstructure is the brain's connectivity network, providing a complex mode of communication between brain regions. As such, it is important to add to previous literature and investigate specific functional connectivity

networks as well as cortical or subcortical areas connected by specific white matter tracts in relation to MDD genetic risk.

4. Conclusions

MDD is a highly heterogeneous disorder with an unclear aetiology. Genetic and neuroimaging links to MDD have so far been vague, indicating the need for further stratification, by biology or symptom profile, as well as development of more advanced analysis techniques incorporating both types of data. The present thesis contributes three studies that aid in the understanding of MDD at the intersection between genetics and neuroimaging. Results provide evidence of white matter microstructure associations with expression of disease- and neurodevelopment-linked genes and propose thalamic radiations as a key neurobiological factor in genetic risk aggregated to a small portion of the genome. The findings presented here also suggest that whole-epigenome risk is associated with the presence of MDD. Evidence presented here may be used to guide future studies and implement large cohorts, with an emphasis placed on neuroimaging, genetics, and epigenetics in the context of MDD.

References

Aan het Rot, M., Hogenelst, K., & Schoevers, R. A. (2012). Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*, *32*(6), 510–523. https://doi.org/10.1016/j.cpr.2012.05.007

Aberg, K. A., Shabalin, A. A., Chan, R. F., Zhao, M., Kumar, G., van Grootheest, G., … van den Oord, E. J. C. G. (2018). Convergence of evidence from a methylome-wide CpG-SNP association study and GWAS of major depressive disorder. *Translational Psychiatry*, *8*(1). https://doi.org/10.1038/s41398-018-0205-8

Alexander, A. L., Lee, J. E., Lazar, M., & Field, A. S. (2008). Diffusion Tensor Imaging of the Brain. *Neurotherapeutics*, *4*(3), 316–329.

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., … Smith, S. M. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, *166*(April 2017), 400–424. https://doi.org/10.1016/j.neuroimage.2017.10.034

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub.

Assaf, Y., & Pasternak, O. (2008). Diffusion tensor imaging (DTI)-based white matter mapping in brain research: A review. *Journal of Molecular Neuroscience*, *34*(1), 51–61. https://doi.org/10.1007/s12031-007-0029-0

Auning, E., Kjærvik, V. K., Selnes, P., Aarsland, D., Haram, A., Bjørnerud, A. et al. (2014). White matter integrity and cognition in Parkinson's disease: a cross-sectional study. *BMJ open*, *4*(1), e003976.

Barbu, M. C., Walker, R. M., Howard, D. M., Evans, K. L., Whalley, H. C., Porteous, D. J., ... & McIntosh, A. M. (2019). Epigenetic prediction of major depressive disorder. *medRxiv*, 19001123.

Barbu, M. C., Zeng, Y., Shen, X., Cox, S. R., Clarke, T. K., Gibson, J. et al. (2019). Association of Whole-Genome and NETRIN1 Signaling Pathway–Derived Polygenic

Risk Scores for Major Depressive Disorder and White Matter Microstructure in the UK Biobank. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(1), 91-100.

Baylin, S. B. (2005). DNA methylation and gene silencing in cancer. *Nature Reviews Clinical Oncology*, *2*(S1), S4.

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system - A technical review. *NMR in Biomedicine*, *15*(7–8), 435–455. https://doi.org/10.1002/nbm.782

Beblo, T., Mensebach, C., Wingenfeld, K., Schlosser, N., Rullkoetter, N., Schaffrath, C., & Driessen, M. (2010). The impact of neutral and emotionally negative distraction on memory performance and its relation to memory complaints in major depression. *Psychiatry Research*, *178*(1), 106–111. https://doi.org/10.1016/j.psychres.2009.04.012

Beekman, A. T. F., Ph, D., Beurs, E. De, Ph, D., Dyck, R. Van, Ph, D., … Ph, D. (2000). Anxiety and Depression in Later Life: Co-Occurrence and Communality of Risk Factors. *American Journal of Psychiatry*, (January), 89–95.

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Nunes, R. G., Clare, S., Matthews, P. M., … Smith, S. M. (2003). Characterization and Propagation of Uncertainty in Diffusion-Weighted MR Imaging. *Magnetic Resonance in Medicine*, *1088*, 1077–1088. https://doi.org/10.1002/mrm.10609

Belmaker RH, Agam G (2008): Major Depressive Disorder. *The New England Journal of Medicine, 358*: 55-68.

Bergman, Y., & Cedar, H. (2013). *DNA methylation dynamics in health and disease*. *20*(3). https://doi.org/10.1038/nsmb.2518

Berlot, R., Metzler-baddeley, C., Jones, D. K., & Sullivan, M. J. O. (2014). CSF contamination contributes to apparent microstructural alterations in mild cognitive impairment. *NeuroImage*, *92*, 27–35. https://doi.org/10.1016/j.neuroimage.2014.01.031

Bhalala, O. G., Nath, A. P., Inouye, M., & Sibley, C. R. (2018). Identification of

expression quantitative trait loci associated with schizophrenia and affective disorders in normal brain tissue. *PLoS Genetics*, *14*(8), 1–25. https://doi.org/10.1371/journal.pgen.1007607

Bhugra, D., & Ayonrinde, O. (2004). Depression in migrants and ethnic minorities. *Advances in Psychiatric Treatment*, *10*, 13–17.

Bianco, S. D., & Kaiser, U. B. (2009). The genetic and molecular basis of idiopathic hypogonadotropic hypogonadism. *Nature Reviews Endocrinology*, *5*(9), 569.

Bigos, K. L., & Weinberger, D. R. (2010). Imaging genetics-days of future past. *NeuroImage*, *53*(3), 804–809. https://doi.org/10.1016/j.neuroimage.2010.01.035

Bonnin, A., Torii, M., Wang, L., Rakic, P., & Levitt, P. (2007). Serotonin modulates the response of embryonic thalamocortical axons to netrin-1. *Nature Neuroscience*, *10*(5), 588–597. https://doi.org/10.1038/nn1896

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *American Journal of Psychiatry*, (July), appi.ajp.2018.1. https://doi.org/10.1176/appi.ajp.2018.18070881.

Boyer, N. P., & Gupton, S. L. (2018). Revisiting Netrin-1: one who guides (axons). *Frontiers in cellular neuroscience*, *12*, 221.

Braisted JE, Catalano SM, Stimac R, Kennedy TE, Tessier-Lavigne M, Shatz CJ, et al. (2000): Netrin-1 promotes thalamic axon growth and is required for proper development of the thalamocortical projection. *Journal of Neuroscience*, *20*(15): 5792-5801.

Brière FN, Rohde P, Seeley JR, Klein D, Lewinsohn PM. Comorbidity between major depression and alcohol use disorder from adolescence to adulthood. Comprehensive psychiatry. 2014 Apr 1;55(3):526-33.

Brodaty, H., Thompson, C., Mitchell, P., Parker, G., Austin, M., Malhi, G. (2005). Age and Gender in the Phenomenology of Depression. *American Journal of Geriatric Psychiatry*, *13*(7), 589–596. https://doi.org/10.1097/00019442-200507000-00007.

Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., Girolamo, G. De, … Kessler, R. C. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine*, *9*(1), 90. https://doi.org/10.1186/1741-7015-9-90.

Brook, D. W., Brook, J. S., Zhang, C., Cohen, P., & Whiteman, M. (2002). Drug use and the risk of major depressive disorder, alcohol dependence, and substance use disorders. *Archives of general psychiatry*, *59*(11), 1039-1044.

Butler, M. G., Rafi, S. K., & Manzardo, A. M. (2015). High-resolution chromosome ideogram representation of currently recognized genes for autism spectrum disorders. *International journal of molecular sciences*, *16*(3), 6464-6495.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. (2017): Genome-wide genetic data on~ 500,000 UK Biobank participants. *bioRxiv*, 166298.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., … Young, A. (2018). The UK Biobank resource with deep phenotyping and genomic data. https://doi.org/10.1038/s41586-018-0579-z

Cabrera CP, Navarro P, Huffman J, Wright AF, Hayward C, Campbell H et al. 2012). Uncovering Networks from Genome-Wide Association Studies via Circular Genomic Permutation. *G3: Genes, Genomics, Genetics,2*(9): 1067-1075.

Castillo-Paterna M, Moreno-Juan V, Filipchuk A, Rodríguez-Malmierca L, Susín R, López-Bendito G. (2015): DCC functions as an accelerator of thalamocortical axonal growth downstream of spontaneous thalamic activity. *EMBO reports*, e201439882.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), 7.

Chen, G., Hu, X., Li, L., Huang, X., Lui, S., Kuang, W., ... & Gong, Q. (2016). Disorganization of white matter architecture in major depressive disorder: a meta-analysis of diffusion tensor imaging with tract-based spatial statistics. *Scientific reports*, *6*, 21825.

Choi, S., Han, K. M., Won, E., Yoon, B. J., Lee, M. S., & Ham, B. J. (2015). Association of brain-derived neurotrophic factor DNA methylation and reduced white

matter integrity in the anterior corona radiata in major depression. *Journal of affective disorders*, *172*, 74-80.

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., … Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*, *391*(10128), 1357–1366. https://doi.org/10.1016/S0140-6736(17)32802-7

Clark, S. L., Hattab, M. W., Chan, R. F., Shabalin, A. A., Han, L. K., Zhao, M., ... & van Grootheest, G. (2019). A methylation study of long-term depression risk. *Molecular psychiatry*, 1-10.

Clarke, T. K., Hall, L. S., Fernandez-Pujals, A. M., MacIntyre, D. J., Thomson, P., Hayward, C., … McIntosh, A. M. (2015). Major depressive disorder and current psychological distress moderate the effect of polygenic risk for obesity on body mass index. *Translational Psychiatry*, *5*(6), e592-6. https://doi.org/10.1038/tp.2015.83

Clascá F, Porrero C, Galazo MJ, Rubio-Garrido P, Evangelio M. (2016): Anatomy and development of multispecific thalamocortical axons: implications for cortical dynamics and evolution. In *Axons and Brain Architecture:* 69-92.

Coenen, V., Panksepp, J., Hurwitz, T., Urbach, H., Madler, B. (2012). Anterior Thalamic Radiation (ATR): Imaging of Two Major Subcortical Pathways and the Dynamic Balance of Opposite Affects in. *J Neuropsychiatry Clin Neurosci*, 223–236.

Cole, J., Chaddock, C. A., Farmer, A. E., Aitchison, K. J., Simmons, A., McGuffin, P., & Fu, C. H. (2012). White matter abnormalities and illness severity in major depressive disorder. *The British Journal of Psychiatry*, *201*(1), 33-39.

Consortium, I. S. (2009). Common polygenic variation contributes to risk of schizophremia that overlaps with bipolar disease. *Nature*, *460*(7256), 748–752. https://doi.org/10.1038/nature08185.Common

Consortium, M. D. D. W. G. of the P. G. (2013). A mega-Analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, *18*(4), 497–511. https://doi.org/10.1038/mp.2012.21

CONVERGE Consortium. (2015). Sparse whole genome sequencing identifies two

loci for major depressive disorder CONVERGE Consortium 1 Europe PMC Funders Group. *Nature*, *523*(7562), 588–591. https://doi.org/10.1038/nature14659.

Cook Jr, E. H., & Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature*, *455*(7215), 919-923.

Córdova-Palomera, A., Palma-Gudiel, H., Fores-Martos, J., Tabares-Seisdedos, R., & Fañanás, L. (2018). Epigenetic outlier profiles in depression: A genome-wide DNA methylation analysis of monozygotic twins. *PloS one*, *13*(11), e0207754.

Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G et al. (2016): Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nature communications*, *7*: 13629.

Curwin, A., & McMaster, C. (2008). Structure and function of the enigmatic Sec14 domain-containing proteins and theetiology of human disease. *Future Lipidology*, *3*(4), 399-410.

Dalton, V. S., Kolshus, E., & Mcloughlin, D. M. (2014). Epigenetics and depression : return of the repressed. *Journal of Affective Disorders*, *155*, 1–12. https://doi.org/10.1016/j.jad.2013.10.028.

Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W., & Kelley, K. W. (2008). From inflammation to sickness and depression: when the immune system subjugates the brain. *Nature reviews neuroscience*, *9*(1), 46-56.

Day FR, Loh PR, Scott RA, Ong KK, Perry JR. (2016): A robust example of collider bias in a genetic association study. *The American Journal of Human Genetics*, *98*(2): 392-393.

De La Vega, F. M., & Bustamante, C. D. (2018). Polygenic risk scores: a biased prediction?. *Genome medicine*, *10*(1), 1-3.

De Jong, S., Van Eijk, K. R., Zeegers, D. W., Strengman, E., Janson, E., Veldink, J. H. et al. (2012). Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes. *European Journal of Human Genetics*, *20*(9), 1004.

De Moor, M. H. M., Van Den Berg, S. M., Verweij, K. J. H., Krueger, R. F., Luciano, M., Arias Vasquez, A., … Boomsma, D. I. (2015). Meta-analysis of genome-wide

association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry*, *72*(7), 642–650. https://doi.org/10.1001/jamapsychiatry.2015.0554

De Wit L, Luppino F, van Straten A, Penninx B, Zitman F, Cuijpers P. Depression and obesity: a meta-analysis of community-based studies. Psychiatry research. 2010 Jul 30;178(2):230-5.

Deng, L., Gao, X., Liu, B., He, X., Xu, J., Qiang, J. et al. (2018). NMT1 inhibition modulates breast cancer progression through stress-triggered JNK pathway. *Cell death & disease*, *9*(12), 1143.

Dima, D., & Breen, G. (2015). Polygenic risk scores in imaging genetics: Usefulness and applications. *Journal of Psychopharmacology*, *29*(8), 867–871. https://doi.org/10.1177/0269881115584470

Ducker, C. E., Upson, J. J., French, K. J., & Smith, C. D. (2005). Two N-myristoyltransferase isozymes play unique roles in protein myristoylation, proliferation, and apoptosis. *Molecular cancer research*, *3*(8), 463-476.

Dunn, E. C., Brown, R. C., Dai, Y., Rosand, J., Nugent, N. R., Amstadter, A. B., & Smoller, J. W. (2015). Genetic determinants of depression: recent findings and future directions. *Harvard review of psychiatry*, *23*(1), 1.

Edwards, L. J., Pine, K. J., Ellerbrock, I., Weiskopf, N., & Mohammadi, S. (2017). NODDI-DTI: estimating neurite orientation and dispersion parameters from a diffusion tensor in healthy white matter. *Frontiers in neuroscience*, *11*, 720.

Elliott, L. T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K. L., Douaud, G., … Smith, S. M. (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*, *562*(7726), 210–216. https://doi.org/10.1038/s41586-018-0571-7

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics*, *31*(9), 1466–1468. https://doi.org/10.1093/bioinformatics/btu848

Fagny, M., Paulson, J. N., Kuijjer, M. L., Sonawane, A. R., Chen, C.-Y., Lopes-

Ramos, C. M., … Platig, J. (2017). Exploring regulation in tissues with eQTL networks. *Proceedings of the National Academy of Sciences*, *114*(37), E7841–E7850. https://doi.org/10.1073/pnas.1707375114

Fama, R., Sullivan, E. V, Sciences, B., & Park, M. (2015). Thalamic structures and associated cognitive functions: Relations with age and aging. *Neuroscience Biobehav Rev*, (650), 29–37. https://doi.org/10.1016/j.neubiorev.2015.03.008.Thalamic

Fazzari, M. J., & Greally, J. M. (2015). *EPIGENOMICS : BEYOND CPG ISLANDS*. (July 2004). https://doi.org/10.1038/nrg1349

Fields, R. D. (2015). White matter in learning, cognition and psychiatric disorders. *Trends in Neurosciences*, *31*(7), 361–370.

Filatova, E. V., Shadrina, M. I., Alieva, A. K., Kolacheva, A. A., Slominsky, P. A., & Ugrumov, M. V. (2014, May). Expression analysis of genes of ubiquitin-proteasome protein degradation system in MPTP-induced mice models of early stages of Parkinson's disease. In *Doklady Biochemistry and Biophysics*(Vol. 456, No. 1, pp. 116-118). Pleiades Publishing.

Flint, J., & Kendler, K. S. (2014). The Genetics of Major Depression. *Neuron*, *81*(3), 484–503. https://doi.org/10.1016/j.neuron.2014.01.027

Fortin JP, Fertig E, Hansen K. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. F1000Research. 2014;3.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., … Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *PNAS*, *102*(30), 10–15.

Franklin, T. B., Russig, H., Weiss, I. C., Gräff, J., Linder, N., Michalon, A., … Mansuy, I. M. (2010). Epigenetic Transmission of the Impact of Early Stress Across Generations. *Biological Psychiatry*, *68*(5), 408–415. https://doi.org/10.1016/j.biopsych.2010.05.036

Fried, E. I. (2017). Moving forward : how depression heterogeneity hinders progress in treatment and research. *Expert Review of Neurotherapeutics*, *17*(5), 423–425. https://doi.org/10.1080/14737175.2017.1307737

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0090311

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, *33*(1), 1.

Fujiwara, T., Ye, S., Castro-Gomes, T., Winchell, C. G., Andrews, N. W., Voth, D. E. et al. (2016). PLEKHM1/DEF8/RAB7 complex regulates lysosome positioning and bone homeostasis. *JCI insight*, *1*(17).

Fullerton, J. M., & Nurnberger, J. I. (2019). Polygenic risk scores in psychiatry: Will they be useful for clinicians?. *F1000Research*, *8*.

Gandal, M. J., Leppa, V., Won, H., Parikshak, N. N., & Geschwind, D. H. (2016). The road to precision psychiatry: Translating genetics into disease mechanisms. *Nature Neuroscience*, *19*(11), 1397–1407. https://doi.org/10.1038/nn.4409

Gandal, M. J., Haney, J. R., Parikshak, N. N., Leppa, V., Ramaswami, G., Hartl, C., ... & Liu, C. (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, *359*(6376), 693-697.

García-Sánchez, A., & Marqués-García, F. (2016). Review of Methods to Study Gene Expression Regulation Applied to Asthma. In *Molecular Genetics of Asthma* (pp. 71-89). Humana Press, New York, NY.

Garriock, H. A., Kraft, J. B., Shyn, S. I., Peters, E. J., Yokoyama, J. S., Jenkins, G. D. et al. (2010). A genomewide association study of citalopram response in major depressive disorder. *Biological psychiatry*, *67*(2), 133-138.

Gartlehner, G., Wagner, G., Matyas, N., Titscher, V., Greimel, J., Lux, L., … Lohr, K. N. (2017). Pharmacological and non-pharmacological treatments for major depressive disorder: Review of systematic reviews. *BMJ Open*, *7*(6), 1–13. https://doi.org/10.1136/bmjopen-2016-014912

Gavin, A. R., Walton, E., Chae, D. H., Alegria, M., Jackson, J. S., & Takeuchi, D. (2010). The associations between socio-economic status and major depressive disorder among Blacks , Latinos , Asians and non-Hispanic Whites : findings from the

Collaborative Psychiatric Epidemiology Studies. *Psychological Medicine*, (2010), 51–61. https://doi.org/10.1017/S0033291709006023

Gerrish, A. C., Thomas, A. G., & Dineen, R. A. (2014). Brain white matter tracts: Functional anatomy and clinical relevance. *Seminars in Ultrasound, CT and MRI*, *35*(5), 432–444. https://doi.org/10.1053/j.sult.2014.06.003

Gibb, B., Chelminski, I., Zimmerman, M. (2007). Childhood emotional, physical, and sexual abuse, and diagnoses of depressive and anxiety disorders in adult psychiatric outpatients. *Depression and Anxiety*, *263*(October 2006), 256–263. https://doi.org/10.1002/da

Gibb, B. E., Alloy, L. B., Abramson, L. Y., Rose, D. T., Whitehouse, W. G., Donovan, P., … Tierney, S. (2001). *History of Childhood Maltreatment , Negative Cognitive Styles , and Episodes of Depression in Adulthood*. *25*(4), 425–446.

Gilad, Y., Rifkin, S. A., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in genetics*, *24*(8), 408-415.

Glatt, S. J., Everall, I. P., Kremen, W. S., Corbeil, J., Šášik, R., Khanlou, N., ... & Tsuang, M. T. (2005). Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. *Proceedings of the National Academy of Sciences*, *102*(43), 15533-15538.

Glessner, J. T., Wang, K., Sleiman, P. M., Zhang, H., Kim, C. E., Flory, J. H., ... & Mentch, F. (2010). Duplication of the SLIT3 locus on 5q35. 1 predisposes to major depressive disorder. *PloS one*, *5*(12).

Gobbi, C., Rocca, M. A., Pagani, E., Riccitelli, G. C., Pravatà, E., Radaelli, M. et al. (2014). Forceps minor damage and co-occurrence of depression and fatigue in multiple sclerosis. *Multiple Sclerosis Journal*, *20*(12), 1633-1640.

Goldberg, D. (2011). *The heterogeneity of " major depression ."* (October), 14–16.

Gonzalez-Lozano, M. A., Klemmer, P., Gebuis, T., Hassan, C., Van Nierop, P., Van Kesteren, R. E. et al. (2016). Dynamics of the mouse brain cortical synaptic proteome during postnatal brain development. *Scientific reports*, *6*, 35456.

Grayson, D. R., & Guidotti, A. (2012). The Dynamics of DNA Methylation in Schizophrenia and Related Psychiatric Disorders. *Neuropsychopharmacology*, *38*(1), 138–166. https://doi.org/10.1038/npp.2012.125

Greenberg, P. E., Fournier, A. A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *Journal of Clinical Psychiatry*, *76*(2), 155–162. https://doi.org/10.4088/JCP.14m09298

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics*, *48*(3), 245.

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C., Ritchie, S. J. et al. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N= 112 151) and 24 GWAS consortia. *Molecular psychiatry*, *21*(11), 1624.

Hagler Jr., D. J., Ahmadi, M. E., Kuperman, J., Holland, D., Carrie, R., Halgren, E., & Dale, A. M. (2009). Automated white-matter tractography using a probabilistic diffusion tensor atlas : Application to temporal lobe epilepsy. *Human Brain Mapping*, *30*(5), 1535–1547. https://doi.org/10.1002/hbm.20619.Automated

Hariri, A. R., Drabant, E. M., & Weinberger, D. R. (2006). Imaging Genetics: Perspectives from Studies of Genetically Driven Variation in Serotonin Function and Corticolimbic Affective Processing. *Biological Psychiatry*, *59*(10), 888–897. https://doi.org/10.1016/j.biopsych.2005.11.005

Heim, C., & Binder, E. B. (2012). Current research trends in early life stress and depression : Review of human studies on sensitive periods , gene – environment interactions , and epigenetics. *Experimental Neurology*, *233*(1), 102–111. https://doi.org/10.1016/j.expneurol.2011.10.032

Hek K, Demirkan A, Lahti J, Terracciano A, Teumer A, Cornelis MC, et al. (2013): A genome-wide association study of depressive symptoms. *Biological psychiatry*, *73*(7): 667-678.

Hernandez, D. G., Nalls, M. A., Moore, M., Chong, S., Dillman, A., Trabzuni, D., …

Cookson, M. R. (2012). Neurobiology of Disease Integration of GWAS SNPs and tissue speci fi c expression pro fi ling reveal discrete eQTLs for human traits in blood and brain. *Neurobiology of Disease*, *47*(1), 20–28. https://doi.org/10.1016/j.nbd.2012.03.020

Herold, C., Mattheisen, M., Lacour, A., Vaitsiakhovich, T., Angisch, M., Drichel, D., & Becker, T. (2012). Integrated genome-wide pathway association analysis with INTERSNP. *Human Heredity*, *73*(2), 63–72. https://doi.org/10.1159/000336196

Herrero, M. T., Barcia, C., & Navarro, J. (2002). Functional anatomy of thalamus and basal ganglia. *Child's Nervous System*, *18*(8), 386-404.

Hodes, G. E., Kana, V., Menard, C., Merad, M., & Russo, S. J. (2015). *Neuroimmune mechanisms of depression*. *18*(10). https://doi.org/10.1038/nn.4113

Hofer, S., & Frahm, J. (2006). Topography of the human corpus callosum revisited—comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *Neuroimage*, *32*(3), 989-994.

Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, *22*(3), 343–352. https://doi.org/10.1038/s41593-018-0326-7

Hurd Y. L. (1996) *Differential messenger RNA expression of prodynorphin and proenkephalin in the human brain. Neuroscience,* 72, 767–783

Hurd, Y. L. (2002). *Subjects with major depression or bipolar disorder show reduction of prodynorphin mRNA expression in discrete nuclei of the amygdaloid complex. Mol. Psychiatry* 7, 75–81.

Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. (2016): Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics*, *48*(9): 1031-1036.

Inkster, B., Nichols, T. E., Saemann, P. G., Auer, D. P., Holsboer, F., Muglia, P., & Matthews, P. M. (2010). Pathway-based approaches to imaging genetics association studies: Wnt signaling, GSK3beta substrates and major depression. *NeuroImage*,

53(3), 908–917. https://doi.org/10.1016/j.neuroimage.2010.02.065

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression : how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33*(march), 245–254. https://doi.org/10.1038/ng1089

Januar, V., Saffery, R., & Ryan, J. (2015). Epigenetics and depressive disorders : a review of current progress and future directions. *International Journal of Epidemiology*, (February), 1364–1387. https://doi.org/10.1093/ije/dyu273

Jbabdi, S., & Johansen-berg, H. (2011). Tractography : Where Do We Go from Here ? *Brain Connectivity*, *1*(3). https://doi.org/10.1089/brain.2011.0033

Jellison, B. J., Field, A. S., Medow, J., Lazar, M., Salamat, M. S., & Alexander, A. L. (2004). Diffusion Tensor Imaging of Cerebral White Matter: A Pictorial Review of Physics, Fiber Tract Anatomy, and Tumor Imaging Patterns. *Am J Neuroradiol*, (March), 356–369. https://doi.org/10.1038/nrn2776

Jia P, Wang L, Fanous AH, Chen X, Kendler KS, Zhao Z, et al. (2012): A bias-reducing pathway enrichment analysis of genome-wide association data confirmed association of the MHC region with schizophrenia. *Journal of medical genetics*, *49*(2): 96-103.

Jia, Z., Wang, Y., Huang, X., & Kuang, W. (2014). Impaired frontothalamic circuitry in suicidal patients with depression revealed by diffusion tensor imaging at 3 . 0 T. *39*(3), 170–177. https://doi.org/10.1503/jpn.130023

Jo, T. (2010). DNA Methylation : An Introduction to the Biology and the Disease-Associated Changes of a Promising Biomarker. *Molecular Biotechnology*, 71–81. https://doi.org/10.1007/s12033-009-9216-2

Joehanes, R., Just, A. C., Marioni, R. E., Pilling, L. C., Reynolds, L. M., Mandaviya, P. R., … Waldenberger, M. (2016). Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*, 436–447. https://doi.org/10.1161/CIRCGENETICS.116.001506

Jones, D. K., Knösche, T. R., & Turner, R. (2013). White matter integrity , fiber count , and other fallacies : The do ’ s and don ’ ts of diffusion MRI. *NeuroImage*, *73*, 239–254. https://doi.org/10.1016/j.neuroimage.2012.06.081

Jones, E. G. (2002). *Thalamic circuitry and thalamocortical synchrony*. (November), 1659–1673. https://doi.org/10.1098/rstb.2002.1168

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics*, *3*(June). https://doi.org/10.1038/nrg816

Jovanova, O. S., Nedeljkovic, I., Spieler, D., Walker, R. M., Liu, C., Luciano, M., … Amin, N. (2018). DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: Meta-analysis of multiethnic epigenome-wide studies. *JAMA Psychiatry*, *75*(9), 949–959. https://doi.org/10.1001/jamapsychiatry.2018.1725.

Juvinao-Quintero, D. L., Hivert, M. F., Sharp, G. C., Relton, C. L., & Elliott, H. R. (2019). DNA Methylation and Type 2 Diabetes: the Use of Mendelian Randomization to Assess Causality. *Current Genetic Medicine Reports*, *7*(4), 191-207.

Kang, H., Kim, J., Stewart, R., Kim, S., Bae, K., Kim, S., … Yoon, J. (2013). Progress in Neuro-Psychopharmacology & Biological Psychiatry Association of SLC6A4 methylation with early adversity , characteristics and outcomes in depression. *Progress in Neuropsychopharmacology & Biological Psychiatry*, *44*, 23–28. https://doi.org/10.1016/j.pnpbp.2013.01.006.

Kendall, K. M., Rees, E., Bracher-Smith, M., Legge, S., Riglin, L., Zammit, S., ... & Walters, J. T. R. (2019). Association of rare copy number variants with risk of depression. *JAMA psychiatry*, *76*(8), 818-825.

Kendler, K. S., Gardner, C. O., Neale, M. C., & Prescott, C. A. (2001). Genetic risk factors for major depression in men and women: Similar or different heritabilities and same or partly distinct genes? *Psychological Medicine*, *31*(4), 605–616. https://doi.org/10.1017/S0033291701003907

Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I., & Brown, W. A. (2012). A systematic review of comparative efficacy of treatments and controls for depression. *PLoS ONE*, *7*(7), 1–11. https://doi.org/10.1371/journal.pone.0041778

Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Computational Biology*, *8*(2). https://doi.org/10.1371/journal.pcbi.1002375

Klengel, T., Pape, J., Binder, E. B., & Mehta, D. (2014). Neuropharmacology The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology*, *80*, 115–132. https://doi.org/10.1016/j.neuropharm.2014.01.013

Klimes-Dougan B, Muetzel R, Mueller BA, Camchong J, Houri A, Lim KO, et al. (2010): Altered white matter microstructure in adolescents with major depression: a preliminary study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(2): 173-183.

Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T. E., … Van Essen, D. C. (2015). Heritability of fractional anisotropy in human white matter: A comparison of Human Connectome Project and ENIGMA-DTI data. *NeuroImage*, *111*, 300–311. https://doi.org/10.1016/j.neuroimage.2015.02.050

Korgaonkar MS, Grieve SM, Koslow SH, Gabrieli JD, Gordon E, Williams LM, et al. (2011): Loss of white matter integrity in major depressive disorder: Evidence using tract-based spatial statistical analysis of diffusion tensor imaging. *Human brain mapping*, *32*(12): 2161-2171.

Kotan, L. D., Cooper, C., Darcan, Ş., Carr, I. M., Özen, S., Yan, Y. et al. (2016). Idiopathic hypogonadotropic hypogonadism caused by inactivating mutations in SRA1. *Journal of clinical research in pediatric endocrinology*, *8*(2), 125.

Kunugi, H., Hori, H., & Ogawa, S. (2015). *Biochemical markers subtyping major depressive disorder*. (March), 597–608. https://doi.org/10.1111/pcn.12299

Lancaster, K., Morris, J. P., & Connelly, J. J. (2018). Neuroimaging Epigenetics: Challenges and Recommendations for Best Practices. *Neuroscience*, *370*, 88–100. https://doi.org/10.1016/j.neuroscience.2017.08.004

Latourelle, J. C., Dumitriu, A., Hadzi, T. C., Beach, T. G., & Myers, R. H. (2012). Evaluation of Parkinson disease risk variants as expression-QTLs. *PloS one*, *7*(10), e46199.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in medicine*, *27*(8), 1133-1163.

Lee, C. (2018). Genome-wide expression quantitative trait loci analysis using mixed

models. *Frontiers in Genetics*, *9*(AUG), 1–9. https://doi.org/10.3389/fgene.2018.00341

Lee, K. W., & Pausova, Z. (2013). Cigarette smoking and DNA methylation. *Frontiers in genetics*, *4*, 132.

Lewis, C. M., & Vassos, E. (2017). Prospects for using risk scores in polygenic medicine. *Genome Medicine*, *9*(1), 9–11. https://doi.org/10.1186/s13073-017-0489-y

Li, M., D'Arcy, C., Li, X., Zhang, T., Joober, R., & Meng, X. (2019). What do DNA methylation studies tell us about depression? A systematic review. *Translational psychiatry*, *9*(1), 68.

Liao, Y., Huang, X., Wu, Q., Yang, C., Kuang, W., Du, M., … Gong, Q. (2013). Is depression a disconnection syndrome? Meta- analysis of diffusion tensor imaging studies in patients with MDD. *Journal of Psychiatry and Neuroscience*, *38*(1), 49–56. https://doi.org/10.1503/jpn.110180

Liu, C., Marioni, R. E., Hedman, Å. K., Pfeiffer, L., Tsai, P., Reynolds, L. M., … Boer, C. G. (2018). A DNA methylation biomarker of alcohol consumption. *Molecular Psychiatry*, (February 2016), 422–433. https://doi.org/10.1038/mp.2016.192

Lohoff, F. W. (2010). Overview of the genetics of major depressive disorder. *Current Psychiatry Reports*, *12*(6), 539–546. https://doi.org/10.1007/s11920-010-0150-6

Lopizzo, N., Chiavetto, L., Cattane, N., Plazzotta, G., Tarazi, F., Pariante, C., Riva, M., & Cattaneo, A. (2015). Gene – environment interaction in major depression : focus on experience-dependent biological systems. *Frontiers in Psychiatry*, *6*(May), 1–12. https://doi.org/10.3389/fpsyt.2015.00068

Lu, Y., Selvakumar, P., Ali, K., Shrivastav, A., Bajaj, G., Resch, L. et al. (2005). Expression of N-myristoyltransferase in human brain tumors. *Neurochemical research*, *30*(1), 9-13.

Lubke GH, Hottenga JJ, Walters R, Laurin C, de Geus EJ, Willemsen G, et al. (2012): Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biological psychiatry*, *72*(8): 707-709.

Luo, X. J., Mattheisen, M., Li, M., Huang, L., Rietschel, M., Børglum, A. D. et al. (2015). Systematic integration of brain eQTL and GWAS identifies ZNF323 as a novel schizophrenia risk gene and suggests recent positive selection based on compensatory advantage on pulmonary function. *Schizophrenia bulletin*, *41*(6), 1294-1308.

Madhavan, K, McQueeny, T., Howe, S., Shear, P., Szaflarski, J. (2014). Superior Longitudinal Fasciculus and Language Functioning in Healthy Aging. *Brain Res*, 11–22. https://doi.org/10.1016/j.brainres.2014.03.012.

Mamiya, P. C., Richards, T. L., & Kuhl, P. K. (2018). Right Forceps Minor and Anterior Thalamic Radiation Predict Executive Function Skills in Young Bilingual Adults. *Frontiers in psychology*, *9*, 118.

Manitt C, Eng C, Pokinko M, Ryan RT, Torres-Berrio A, Lopez JP, et al. (2013): DCC orchestrates the development of the prefrontal cortex during adolescence and is altered in psychiatric patients. *Translational psychiatry*, *3*(12): e338.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews. Genetics*, *9*(5), 356–369. https://doi.org/10.1038/nrg2344

McCartney, D. L., Hillary, R. F., Stevenson, A. J., Ritchie, S. J., Walker, R. M., Zhang, Q., … Marioni, R. E. (2018). Epigenetic prediction of complex traits and death. *Genome Biology*, *19*(1), 136. https://doi.org/10.1186/s13059-018-1514-1

McCunn, P., Gilbert, K. M., Zeman, P., Li, A. X., Strong, M. J., Khan, A. R., & Bartha, R. (2019). Reproducibility of Neurite Orientation Dispersion and Density Imaging (NODDI) in rats at 9.4 Tesla. *PloS one*, *14*(4), e0215974.

McEwan, D. G., Popovic, D., Gubas, A., Terawaki, S., Suzuki, H., Stadel, D. et al. (2015). PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. *Molecular cell*, *57*(1), 39-54.

McIntyre, R. S., Soczynska, J. Z., Woldeyohannes, H. O., Alsuwaidan, M. T., Cha, D. S., Carvalho, A. F., … Kennedy, S. H. (2015). The impact of cognitive impairment on perceived workforce performance: Results from the International Mood Disorders Collaborative Project. *Comprehensive Psychiatry*, *56*, 279–282.

https://doi.org/10.1016/j.comppsych.2014.08.051

McKenzie, M., Henders, A. K., Caracella, A., Wray, N. R., & Powell, J. E. (2014). Overlap of expression Quantitative Trait Loci (eQTL) in human brain and blood. *BMC Medical Genomics*, *7*(1). https://doi.org/10.1186/1755-8794-7-31

Mendelson, M. M., Marioni, R. E., Joehanes, R., Liu, C., Hedman, Å. K., Aslibekyan, S., … Ingelsson, E. (2017). Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease : A Mendelian Randomization Approach. *PLOS Medicine*, 1–30. https://doi.org/10.1371/journal.pmed.1002215

Metzler-baddeley, C., Sullivan, M. J. O., Bells, S., Pasternak, O., & Jones, D. K. (2012). How and how not to correct for CSF-contamination in diffusion MRI. *NeuroImage*, *59*(2), 1394–1403. https://doi.org/10.1016/j.neuroimage.2011.08.043

Michaelson, J. J., Loguercio, S., & Beyer, A. (2009). Detection and interpretation of expression quantitative trait loci (eQTL). *Methods*, *48*(3), 265–276. https://doi.org/10.1016/j.ymeth.2009.03.004

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1), 167-202.

Mills, J. G., Thomas, S. J., Larkin, T. A., & Deng, C. (2020). Overeating and food addiction in Major Depressive Disorder: Links to peripheral dopamine. *Appetite*, 104586.

Mori, S., Kaufmann, W. E., Davatzikos, C., Stieltjes, B., Amodei, L., Fredericksen, K., ... & Moser, H. W. (2002). Imaging cortical association tracts in the human brain using diffusion-tensor-based axonal tracking. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *47*(2), 215-223.

Mori, S., Wakana, S., Van Zijl, P. C., & Nagae-Poetscher, L. M. (2005). *MRI atlas of human white matter*. Elsevier.

Mullins N, Lewis CM. (2017): Genetics of Depression: Progress at Last. *Current Psychiatry Reports*, *19*(8): 43.

Murphy, M. L., & Frodl, T. (2011). Meta-analysis of diffusion tensor imaging studies shows altered fractional anisotropy occurring in distinct brain areas in association with depression. *Biology of Mood & Anxiety Disorders*, *1*(1), 3. https://doi.org/10.1186/2045-5380-1-3

Musliner, K. L., Seifuddin, F., Judy, J. A., Pirooznia, M., Goes, F. S., & Zandi, P. P. (2015). Polygenic risk, stressful life events and depressive symptoms in older adults: A polygenic score analysis. *Psychological Medicine*, *45*(8), 1709–1720. https://doi.org/10.1017/S0033291714002839

Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, Evenden L, Gibson J, Amador C, Howard DM, Navarro P. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. Genome medicine. 2017 Dec;9(1):23.

Nasrabady, S. E., Rizvi, B., Goldman, J. E., & Brickman, A. M. (2018). White matter changes in Alzheimer's disease: a focus on myelin and oligodendrocytes. *Acta Neuropathologica Communications*, *6*(1), 22. https://doi.org/10.1186/s40478-018-0515-3.

Navrady, L. B., Wolters, M. K., Macintyre, D. J., Clarke, T., Campbell, A. I., Murray, A. D., … Haley, C. (2018). Cohort Profile : Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up of Generation Scotland : Scottish Family Health Study (GS : SFHS). *International Journal of Epidemiology*, (July 2017), 13–14. https://doi.org/10.1093/ije/dyx115.

Nemeroff, C. B., & Vale, W. W. (2005). The neurobiology of depression: inroads to treatment and new drug discovery. *The Journal of clinical psychiatry*, *66*, 5-13.

Network T, Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. (2015): Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature neuroscience*, *18*(2): 199.

Neupane, S. P. (2016). Neuroimmune interface in the comorbidity between alcohol use disorder and major depression. *Frontiers in immunology*, *7*, 655.

Nica, A. C., & Dermitzakis, E. T. (2013). Expression quantitative trait loci: Present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*,

368(1620). https://doi.org/10.1098/rstb.2012.0362

O'Brien, H., et al. (2018). Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biology*, 1–13. https://doi.org/10.1186/s13059-018-1567-1.

Ohayon, M. M., & Schatzberg, A. F. (2003). Using chronic pain to predict depressive morbidity in the general population. *Archives of general psychiatry*, *60*(1), 39-47.

Ohayon, M. M., & Schatzberg, A. F. (2010). Chronic pain and major depressive disorder in the general population. *Journal of psychiatric research*, *44*(7), 454-461.

Opel N, Redlich R, Grotegerd D, Dohm K, Heindel W, Kugel H, Arolt V, Dannlowski U. Obesity and major depression: body-mass index (BMI) is associated with a severe course of disease and specific neurostructural alterations. Psychoneuroendocrinology. 2015 Jan 1;51:219-26.

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., … Schatzberg, A. F. (2016). Major depressive disorder. *Nature Publishing Group*, *2*(Mdd), 1–21. https://doi.org/10.1038/nrdp.2016.65

Paperwalla KN, Levin TT, Weiner J, Saravay SM. Smoking and depression. The Medical Clinics of North America. 2004 Nov;88(6):1483-94.

Park, N., Yoo, J. C., Lee, Y. S., Choi, H. Y., Hong, S. G., Hwang, E. M. et al. (2014). Copine1 C2 domains have a critical calcium-independent role in the neuronal differentiation of hippocampal progenitor HiB5 cells. *Biochemical and biophysical research communications*, *454*(1), 228-233.

Park, N., Yoo, J. C., Ryu, J., Hong, S. G., Hwang, E. M., & Park, J. Y. (2012). Copine1 enhances neuronal differentiation of the hippocampal progenitor HiB5 cells. *Molecules and cells*, *34*(6), 549-554.

Parkinson, C., & Wheatley, T. (2014). Relating Anatomical and Social Connectivity : White Matter Microstructure Predicts Emotional Empathy. *Cerebral Cortex*, (March), 614–625. https://doi.org/10.1093/cercor/bhs347

Pearson, R., Palmer, R., Brick, L., McGeary, J., Knopik, V., Beevers, C. (2017). Additive Genetic Contribution to Symptom Dimensions in Major Depressive

Disorder. *J Abnorm Psychol.*, *125*(4), 495–501. https://doi.org/10.1037/abn0000161.Additive

Peedicayil, J. (2007). The role of epigenetics in mental disorders. *Indian Journal of Medical Research*, *126*(2), 105.

Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I., & Penninx, B. W. J. H. (2014). Effect of polygenic risk scores on depression in childhood trauma. *British Journal of Psychiatry*, *205*(2), 113–119. https://doi.org/10.1192/bjp.bp.113.143081

Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC genomics. 2013 Dec;14(1):293.

Purcell SM, Wray NR, Stone JL, Visscher PM, O'donovan MC, Sullivan PF, et al. (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256): 748-752.

Qi, T. et al. (2018). Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nature Communications*, (2018). https://doi.org/10.1038/s41467-018-04558-1.

Rae, C. L., Davies, G., Garfinkel, S. N., Gabel, M. C., Dowell, N. G., Cercignani, M., ... & Critchley, H. D. (2017). Deficits in neurite density underlie white matter structure abnormalities in first-episode psychosis. *Biological psychiatry*, *82*(10), 716-725.

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R. et al. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature neuroscience*, *17*(10), 1418.

Reginsson, G. W., Ingason, A., Euesden, J., Bjornsdottir, G., Olafsson, S., Sigurdsson, E., … Stefansson, K. (2018). Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addiction Biology*, 485–492. https://doi.org/10.1111/adb.12496

Reus, L. M., Shen, X., Gibson, J., Wigmore, E., Ligthart, L., Adams, M. J., & Davies, G. (2017). Association of polygenic risk for major psychiatric illness with subcortical

volumes and white matter integrity in UK Biobank. *Nature Publishing Group*, (October 2016), 1–8. https://doi.org/10.1038/srep42140

Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, et al. (2013): A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry*, *18*(4): 497-511.

Ritchey, M., Dolcos, F., Eddington, K., Strauman, T. & Cabeza, R. (2011). Neural correlates of emotional processing in depression: Changes with cognitive behavioural therapy and predictors of treatment response. *Journal of Psychiatric Research, 45*(5), 577-587.

Robertson, K. D. (2005). Dna methylation and human disease. *Nature Reviews Genetics*, *6*(August), 597–610. https://doi.org/10.1038/nrg1655

Romaniuk, L., Sandu, A. L., Waiter, G. D., McNeil, C. J., Xueyi, S., Harris, M. A., ... & Delgado, M. R. (2019). The neurobiology of personal control during reward learning and its relationship to mood. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(2), 190-199.

Roth, T. L., Lubin, F. D., Funk, A. J., & Sweatt, J. D. (2009). Lasting Epigenetic Influence of Early-Life Adversity on the BDNF Gene. *Biological Psychiatry*, *65*(9), 760–769. https://doi.org/10.1016/j.biopsych.2008.11.028.

Rucker, J. J., Breen, G., Pinto, D., Pedroso, I., Lewis, C. M., Cohen-Woods, S., ... & Craddock, N. (2013). Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Molecular psychiatry*, *18*(2), 183-189.

Saalmann, Y. B., & Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron*, *71*(2), 209-223.

Scharinger, C., Rabl, U., Sitte, H. H., & Pezawas, L. (2010). Imaging genetics of mood disorders. *NeuroImage*, *53*(3), 810–821. https://doi.org/10.1016/j.neuroimage.2010.02.019

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2018). Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*, *173*(7), 1705-1715.

Schmaal, L., Veltman, D. J., Erp, T. G. M. Van, Sämann, P. G., Frodl, T., Jahanshad, N., … Niessen, W. J. (2016). Subcortical brain alterations in major depressive disorder : fi ndings from the ENIGMA Major Depressive Disorder working group. *Molecular Psychiatry*, (April 2015), 806–812. https://doi.org/10.1038/mp.2015.69

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., Arceuil, H. E. D., Crespigny, A. J. De, & Wedeen, V. J. (2007). *Association fibre pathways of the brain : parallel observations from diffusion spectrum imaging and autoradiography*. 630–653. https://doi.org/10.1093/brain/awl359

Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-berg, H. (2009). Training induces changes in white matter architecture. *Nature Neuroscience*, *12*(11), 1370–1371. https://doi.org/10.1038/nn.2412.

Serafini T, Colamarino SA, Leonardo ED, Wang H, Beddington R, Skarnes WC, et al. (1996): Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell*, *87*(6): 1001-1014.

Shah, S., Bonder, M. J., Marioni, R. E., Zhu, Z., McRae, A. F., Zhernakova, A., … Visscher, P. M. (2015). Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *American Journal of Human Genetics*, *97*(1), 75–85. https://doi.org/10.1016/j.ajhg.2015.05.014

Shen, X., Adams, M. J., Ritakari, T. E., Cox, S. R., Mcintosh, A. M., & Whalley, H. C. (2019). White Matter Microstructure and Its Relation to Longitudinal Measures of Depressive Symptoms in Mid- and Late Life. *Biological Psychiatry*, 1–10. https://doi.org/10.1016/j.biopsych.2019.06.011

Shen, X., Howard, D. M., Adams, M. J., Deary, I. J., Whalley, H. C., McIntosh, A. M., & 23andMe Research Team. (2019). A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *bioRxiv*, 617969.

Shen, E. H., Overly, C. C., & Jones, A. R. (2012). The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends in neurosciences*, *35*(12), 711-714.

Shen, X., Reus, L. M., Cox, S. R., Adams, M. J., Liewald, D. C., Bastin, M. E., …

McIntosh, A. M. (2017). Subcortical volume and white matter integrity abnormalities in major depressive disorder: Findings from UK Biobank imaging data. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-05507-6

Sherman, S. M. (2016). Thalamus plays a central role in ongoing cortical functioning. *Nature neuroscience*, *19*(4), 533.

Smith, B. H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S. M., … Morris, A. D. (2013). Cohort Profile : Generation Scotland : Scottish Family Health Study (GS : SFHS). The study , its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology*, (July 2012), 689–700. https://doi.org/10.1093/ije/dys084

Smith, B. H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I. J., … Morris, A. D. (2006). Generation Scotland : the Scottish Family Health Study ; a new resource for researching genes and heritability. *BMC Medical Genomics*, *9*, 1–9. https://doi.org/10.1186/1471-2350-7-74

Smith, S. M., Jenkinson, M., Johansen-berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., … Behrens, T. E. J. (2006). Tract-based spatial statistics : Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, *31*, 1487–1505. https://doi.org/10.1016/j.neuroimage.2006.02.024

Sosicka, P., Maszczak-Seneczko, D., Bazan, B., Shauchuk, Y., Kaczmarek, B., & Olczak, M. (2017). An insight into the orphan nucleotide sugar transporter SLC35A4. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1864*(5), 825-838.

Sprooten, E., Knowles, E. E., McKay, D. R., Göring, H. H., Curran, J. E., Kent Jr, J. W. et al. (2014). Common genetic variants and gene expression associated with white matter microstructure in the human brain. *Neuroimage*, *97*, 252-261.

Stordal, E., & Kru, B. (2001). *Depression in relation to age and gender in the general population : the Nord-Trøndelag Health Study (HUNT)*. (22), 210–216.

Storey, J. D., Madeoy, J., Strout, J. L., Wurfel, M., Ronald, J., & Akey, J. M. (2007). Gene-Expression Variation Within and Among Human Populations. *The American*

Journal of Human Genetics, *80*(3), 502–509. https://doi.org/10.1086/512017

Stranger, B. E., Forrest, M. S., Clark, A. G., Minichiello, M. J., Deutsch, S., Lyle, R. et al. (2005). Genome-wide associations of gene expression variation in humans. *PLoS genetics*, *1*(6), e78.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Sprosen, T. (2015). UK Biobank : An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, 1–10. https://doi.org/10.1371/journal.pmed.1001779.

Sullivan, P. F., Fan, C., & Perou, C. M. (2006). Evaluating the comparability of gene expression in blood and brain. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *141*(3), 261-268.

Sullivan, P. F., Neale, M. C., Ph, D., & Kendler, K. S. (2000). Genetic Epidemiology of Major Depression : Review and Meta-Analysis. *Am J Psychiatry*, (October), 1552–1562.

Sullivan, P. F., & Posthuma, D. (2015). Biological pathways and networks implicated in psychiatric disorders. *Current Opinion in Behavioral Sciences*, *2*(Mdd), 58–68. https://doi.org/10.1016/j.cobeha.2014.09.003

Tang, F., & Kalil, K. (2005). *Netrin-1 Induces Axon Branching in Developing Cortical Neurons by Frequency-Dependent Calcium Signaling Pathways*. *25*(28), 6702–6715. https://doi.org/10.1523/JNEUROSCI.0871-05.2005

Tham, M. W., Woon, P. S., Sum, M. Y., Lee, T. S., & Sim, K. (2011). White matter abnormalities in major depression: Evidence from post-mortem, neuroimaging and genetic studies. *Journal of Affective Disorders*, *132*(1–2), 26–36. https://doi.org/10.1016/j.jad.2010.09.013.

Theorell, T., Hammarström, A., Aronsson, G., Bendz, L. T., Grape, T., Hogstedt, C., … Hall, C. (2015). A systematic review including meta-analysis of work environment and depressive symptoms. *BMC Public Health*, 1–14. https://doi.org/10.1186/s12889-015-1954-4.

Timmers, I., Roebroeck, A., Bastiani, M., Jansma, B., Rubio-Gozalbo, E., & Zhang,

H. (2016). Assessing microstructural substrates of white matter abnormalities: a comparative study using DTI and NODDI. *PloS one*, *11*(12).

Velzen, L. S. Van, Kelly, S., Isaev, D., Aleman, A., Aftanas, L. I., & Bauer, J. (2019). White matter disturbances in major depressive disorder : a coordinated analysis across 20 international cohorts in the ENIGMA MDD working group. *Molecular Psychiatry*, (Mdd). https://doi.org/10.1038/s41380-019-0477-2

Vestergaard, M., Madsen, K. S., & Baaré, W. F. C. (2011). *White Matter Microstructure in Superior Longitudinal Fasciculus Associated with Spatial Working Memory Performance in Children*. https://doi.org/10.1162/jocn.2010.21592

Vuoksimaa, E., Panizzon, M. S., Hagler, D. J., Hatton, S. N., Fennema-Notestine, C., Rinker, D., … Kremen, W. S. (2017). Heritability of white matter microstructure in late middle age: A twin study of tract-based fractional anisotropy and absolute diffusivity indices. *Human Brain Mapping*, *38*(4), 2026–2036. https://doi.org/10.1002/hbm.23502.

Walton, E., Relton, C. L., & Caramaschi, D. (2019). Using Openly Accessible Resources to Strengthen Causal Inference in Epigenetic Epidemiology of Neurodevelopment and Mental Health. *Genes*, *10*(3), 193.

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., ... & Blitz, A. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage*, *36*(3), 630-644.

Wakana, S., Jiang, H., Nagae-Poetscher, L. M., Van Zijl, P. C., & Mori, S. (2004). Fiber tract–based atlas of human white matter anatomy. *Radiology*, *230*(1), 77-87.

Ward J, Strawbridge R, Graham N, Bailey M, Freguson A, Lyall D, et al. (2017): Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. bioRxiv, 117796.

Wardenaar, K. J., & Jonge, P. De. (2013). Diagnostic heterogeneity in psychiatry : towards an empirical solution. *Current Controversies in Psychiatry*, 2–4.

Werf, Y. D. Van Der, Tisserand, D. J., Jelle, P., Hofman, P. A. M., Vuurman, E.,

Uylings, H. B. M., & Jolles, J. (2001). Thalamic volume predicts performance on tests of cognitive speed and decreases in healthy aging A magnetic resonance imaging-based volumetric analysis. *Cognitive Brain Research*, *11*, 377–385.

Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J. et al. (2013). Systematic identification of trans eQTL as putative drivers of known disease associations. *Nature genetics*, *45*(10), 1238.

Whalley HC, Adams MJ, Hall LS, Clarke TK, Fernandez-Pujals AM, Gibson J, et al. (2016): Dissection of major depressive disorder using polygenic risk scores for schizophrenia in two independent cohorts. *Translational psychiatry*, *6*(11): e938.

Whalley, H. C., Sprooten, E., Hackett, S., Hall, L., Blackwood, D. H., Glahn, D. C., … Mcintosh, A. M. (2013). Polygenic Risk and White Matter Integrity in Individuals at High Risk of Mood Disorder. *Biological Psychiatry*, *74*(4), 280–286. https://doi.org/10.1016/j.biopsych.2013.01.027

Whalley, H. C., Sussmann, J. E., Romaniuk, L., Stewart, T., Papmeyer, M., Sprooten, E. et al. (2013). Prediction of depression in individuals at high familial risk of mood disorders using functional magnetic resonance imaging. *PloS one*, *8*(3), e57357.

Wigmore, E. M., Clarke, T., Howard, D. M., Adams, M. J., Hall, L. S., Zeng, Y., … Thomson, P. A. (2017). Do regional brain volumes and major depressive disorder share genetic architecture ? A study of Generation Scotland (n = 19 762), UK Biobank (n = 24 048) and the English Longitudinal Study of Ageing (n = 5766). *Translational Psychiatry*, (December 2016). https://doi.org/10.1038/tp.2017.148

Winklewski, Pawel J, Sabisz, A., Naumczyk, P., Jodzio, K., Szurowska, E., Szarmach, A. (2018). Understanding the Physiopathology Behind Axial and Radial Diffusivity Changes — what Do we Know ? *Frontiers in Neurology*, *9*(February). https://doi.org/10.3389/fneur.2018.00092

Won, E., & Ham, B. (2016). Progress in Neuro-Psychopharmacology & Biological Psychiatry Imaging genetics studies on monoaminergic genes in major depressive disorder. *Progress in Neuropsychopharmacology & Biological Psychiatry*, *64*, 311–319. https://doi.org/10.1016/j.pnpbp.2015.03.014

Won, E., & Ham, B. J. (2016). Imaging genetics studies on monoaminergic genes in major depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *64*, 311-319.

World Health Organization. (2017). Depression and Other Common Mental Disorders. *Who*, 24. https://doi.org/CC BY-NC-SA 3.0 IGO

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews. Genetics*, *8*(3), 206–216. https://doi.org/10.1038/nrg2063

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2008). Prediction of individual genetic risk of complex disease. *Current Opinion in Genetics & Development*, *25*, 257–263. https://doi.org/10.1016/j.gde.2008.07.006

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Sullivan, P. F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Wu, F., Tang, Y., Xu, K., Kong, L., Sun, W., Wang, F., ... & Liu, Y. (2011). Whiter matter abnormalities in medication-naive subjects with a single short-duration episode of major depressive disorder. *Psychiatry Research: Neuroimaging*, *191*(1), 80-83.

Xiao, J., He, Y., McWhinnie, C. M., & Yao, S. (2015). Altered white matter integrity in individuals with cognitive vulnerability to depression: a tract-based spatial statistics study. *Scientific reports*, *5*, 9738.

Yang H, Wang K (2015): Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols*, *10*(10): 1556-1566.

Yeh, P., Simpson, K., Durazzo, T. C., Gazdzinski, S., & Meyerhoff, D. J. (2009). Tract-based spatial statistics (TBSS) of diffusion tensor imaging data in alcohol dependence : Abnormalities of the motivational neurocircuitry. *Psychiatry Research: Neuroimaging*, *173*(1), 22–30. https://doi.org/10.1016/j.pscychresns.2008.07.012

Yousaf, T., Pagano, G., Niccolini, F., & Politis, M. (2018). Increased dopaminergic function in the thalamus is associated with excessive daytime sleepiness. *Sleep medicine*, *43*, 25-30.

Zeng, Y., Navarro, P., Fernandez-Pujals, A. M., Hall, L. S., Clarke, T. K., Thomson, P. A., … McIntosh, A. M. (2017). A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway Is Associated With Major Depressive Disorder. *Biological Psychiatry*, *81*(4), 336–346. https://doi.org/10.1016/j.biopsych.2016.04.017

Zeng Y, Navarro P, Xia C, Amador C, Fernandez-Pujals AM, Thomson PA, et al. (2016): Shared genetics and couple-associated environment are major contributors to the risk of both clinical and self-declared depression. *EBioMedicine*, *14*: 161-167.

Zhang, X., Abdellaoui, A., Rucker, J., de Jong, S., Potash, J. B., Weissman, M. M., ... & Sobell, J. (2019). Genome-wide burden of rare short deletions is enriched in major depressive disorder in four cohorts. *Biological psychiatry*, *85*(12), 1065-1073.

Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., & Alexander, D. C. (2012). NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, *61*(4), 1000-1016.

Zhong, J., Li, S., Zeng, W., Li, X., Gu, C., Liu, J., & Luo, X. (2019). Integration of GWAS and brain eQTL identi fi es FLOT1 as a risk gene for major depressive disorder. *Neuropsychopharmacology*, (February). https://doi.org/10.1038/s41386-019-0345-4.

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E. et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature genetics*, *48*(5), 481.

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). ScienceDirect How many different ways do patients meet the diagnostic criteria for major depressive disorder ? *Comprehensive Psychiatry*, *56*, 29–34. https://doi.org/10.1016/j.comppsych.2014.09.007

Zou, F., Chai, H. S., Younkin, C. S., Allen, M., Crook, J., Pankratz, V. S. et al. (2012). Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS genetics*, *8*(6), e1002707.

Appendix 1: Supplementary materials for Chapter 2: Expression quantitative trait loci-derived scores and white matter microstructure in UK Biobank: a novel approach to integrating genetics and neuroimaging

Results for scores associated with FA (N = 17; table S1) and MD (N = 16; table S2) white matter tracts separately

Score, white matter tract	Effect size	SD	t value	p value	p value, FDR corrected
ATG10_eQTL_cis, global FA	0.0341	0.0079	4.3106	1.64E-05	0.0273
SF3A1_eQTL_cis, global FA	-0.0327	0.0079	-4.1305	3.64E-05	0.0495
SMARCAL1_eQTL_cis, global FA	0.0374	0.0079	4.7354	2.21E-06	0.0071
SF3A1_eQTL_cis, association fibres	-0.0334	0.0079	-4.2386	2.26E-05	0.0344
SMARCAL1_eQTL_cis, association fibres	0.0326	0.0079	4.1375	3.53E-05	0.0495
ATG10_eQTL_cis, thalamic radiations	0.0373	0.0080	4.6587	3.21E-06	0.0088
PPP4R3A_eQTL_cis, thalamic radiations	0.0357	0.0080	4.4572	8.36E-06	0.0166
SMARCAL1_eQTL_cis, thalamic radiations	0.0394	0.0080	4.9292	8.35E-07	0.0036
CD14_eQTL_cis, projection fibres	-0.0360	0.0079	-4.5691	4.94E-06	0.0116
COG7_eQTL_cis, anterior thalamic radiation	-0.0333	0.0076	-4.4005	1.09E-05	0.0337
SMARCAL1_eQTL_cis, anterior thalamic radiation	0.0394	0.0076	5.2164	1.85E-07	0.0018
LINC01605_eQTL_trans, cingulate gyrus	-0.0337	0.0071	-4.7560	1.99E-06	0.0114
ANXA1_eQTL_cis, corticospinal tract	-0.0320	0.0074	-4.3218	1.56E-05	0.0416
ZSCAN26_eQTL_cis, forceps major	-0.0397	0.0081	-4.9048	9.45E-07	0.0070
ATG10_eQTL_cis, forceps minor	0.0360	0.0078	4.5986	4.29E-06	0.0189
CD14_eQTL_cis, forceps minor	0.0456	0.0078	5.8210	6E-09	0.0001
SHTN1 / KIAA1598_eQTL_cis, forceps minor	0.0376	0.0078	4.8050	1.56E-06	0.0101
ZNF282_eQTL_cis, forceps minor	-0.0346	0.0078	-4.4224	9.83E-06	0.0337
ENO4_eQTL_cis, forceps minor	0.0354	0.0078	4.5197	6.24E-06	0.0252
COG7_eQTL_cis, forceps minor	-0.0338	0.0078	-4.3127	1.62E-05	0.0416
SMARCAL1_eQTL_cis, forceps minor	0.0361	0.0078	4.6056	4.15E-06	0.0189
ASRGL1_eQTL_cis, inferior fronto-occipital fasciculus	0.0329	0.0077	4.2950	1.76E-05	0.0426
ATG10_eQTL_cis, inferior fronto-occipital fasciculus	0.0355	0.0077	4.6291	3.7E-06	0.0179

TMEM184B_eQTL_cis, inferior fronto-occipital fasciculus	0.0337	0.0077	4.3935	1.12E-05	0.0337
SMARCAL1_eQTL_cis, inferior longitudinal fasciculus	0.0349	0.0076	4.5704	4.91E-06	0.0207
ATG10_eQTL_cis, posterior thalamic radiation	0.0325	0.0075	4.3416	1.42E-05	0.0406
ZBTB7B_eQTL_cis, superior longitudinal fasciculus	-0.0329	0.0077	-4.2946	1.76E-05	0.0426
GPT_eQTL_cis, superior longitudinal fasciculus	0.0339	0.0077	4.4153	1.02E-05	0.0337
SMARCAL1_eQTL_cis, superior longitudinal fasciculus	0.0401	0.0077	5.2356	1.67E-07	0.0018
GPT_eQTL_cis, superior thalamic radiation	0.0337	0.0079	4.2827	1.86E-05	0.0429
AP2S1_eQTL_cis, superior thalamic radiation	0.0348	0.0079	4.4164	1.01E-05	0.0337

Table S1. eQTL scores associated only with white matter tracts as measured by FA. The first column indicates standardised effect size (β).

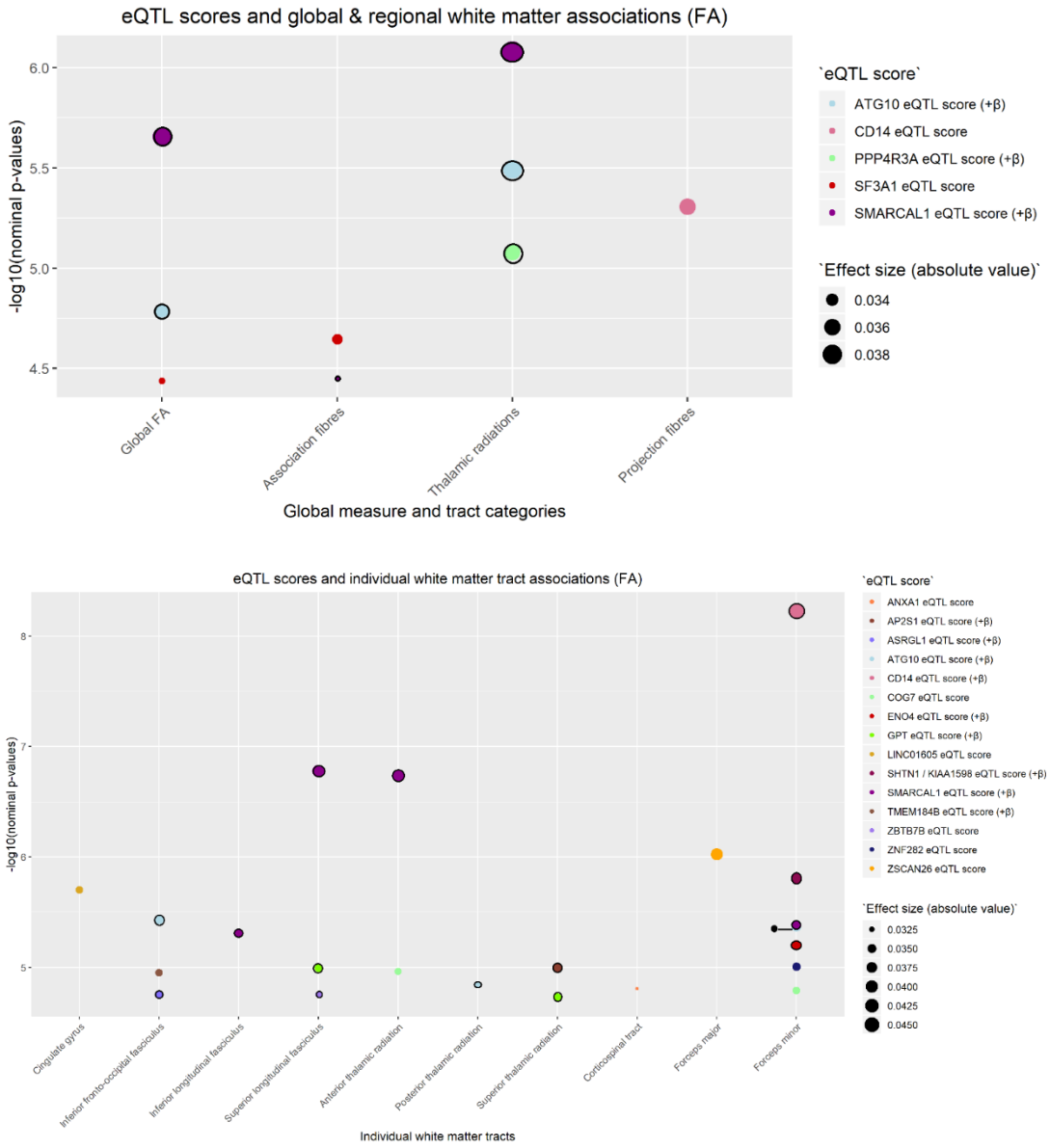


Figure S1. eQTL scores associated only with white matter tracts as measured by FA (fractional anisotropy). Indicates nominal p-values between each of the scores (shown in legend entitled "eQTL score") and global and tract category measures (noted on the x-axis). All values in the figure met FDR correction. Some of the scores with an additional line around the points had an effect size in the opposite direction to all other scores (also indicated by +β for FA in figure legend). The colours of the plot points indicate the score to which they belong. Magnitude of effect is shown in the legend entitled "Effect size (absolute values)".

Score, white matter tract	Effect size	SD	t value	p value	p value, FDR corrected
APOA1BP / NAXE_eQTL_cis, global MD	0.0311	0.0075	4.1331	3.6E-05	0.0423
BTN3A2_eQTL_cis, global MD	0.0308	0.0075	4.0853	4.42E-05	0.0423
UMPS_eQTL_cis, global MD	-0.0319	0.0075	-4.2381	2.27E-05	0.0366
CSF3R_eQTL_cis, global MD	0.0345	0.0075	4.5704	4.91E-06	0.0132
TMEM154_eQTL_cis, global MD	-0.0400	0.0076	-5.2970	1.19E-07	0.0015
APOA1BP / NAXE_eQTL_cis, association fibres	0.0326	0.0077	4.2419	2.23E-05	0.0366
BTN3A2_eQTL_cis, association fibres	0.0314	0.0077	4.0888	4.36E-05	0.0423
SAMM50_eQTL_cis, association fibres	-0.0311	0.0077	-4.0501	5.15E-05	0.0475
UMPS_eQTL_cis, association fibres	-0.0355	0.0077	-4.6219	3.84E-06	0.0124
CSF3R_eQTL_cis, association fibres	0.0377	0.0077	4.9069	9.36E-07	0.0048
TMEM154_eQTL_cis, association fibres	-0.0402	0.0077	-5.2158	1.86E-07	0.0016
HLA-C_eQTL_cis, association fibres	-0.0342	0.0077	-4.4402	9.05E-06	0.0213
MED15_eQTL_cis, thalamic radiations	0.0297	0.0072	4.1354	3.56E-05	0.0423
KANSL1_eQTL_cis, thalamic radiations	-0.0302	0.0072	-4.2041	2.64E-05	0.0401
IL18RAP_eQTL_cis, projection fibres	0.0324	0.0078	4.1621	3.17E-05	0.0423
C6orf106_eQTL_cis, projection fibres	0.0318	0.0078	4.0863	4.41E-05	0.0423
RABEPK_eQTL_cis, acoustic radiation	0.0307	0.0069	4.4197	9.95E-06	0.0287
CFDP1_eQTL_cis, anterior thalamic radiation	0.0298	0.0070	4.2749	1.92E-05	0.0447
PTPN13_eQTL_cis, anterior thalamic radiation	-0.0347	0.0070	-4.9756	6.58E-07	0.0044
KANSL1_eQTL_cis, anterior thalamic radiation	-0.0369	0.0070	-5.3005	1.17E-07	0.0016
TMEM154_eQTL_cis, anterior thalamic radiation	-0.0303	0.0070	-4.3269	1.52E-05	0.0372
UMPS_eQTL_cis, cingulate gyrus	-0.0323	0.0073	-4.3999	1.09E-05	0.0297
PLEC_eQTL_cis, forceps minor	0.0335	0.0076	4.3887	1.15E-05	0.0297
TMEM154_eQTL_cis, forceps minor	-0.0384	0.0076	-5.0145	5.38E-07	0.0043
TMEM154_eQTL_cis, inferior fronto-occipital fasciculus	-0.0330	0.0075	-4.3971	1.1E-05	0.0297
TMEM154_eQTL_cis, inferior longitudinal fasciculus	-0.0337	0.0073	-4.5974	4.32E-06	0.0164

SAMM50_eQTL_cis, parahippocampal part of cingulum	-0.0303	0.0071	-4.2755	1.92E-05	0.0447
BTN3A2_eQTL_cis, superior longitudinal fasciculus	0.0350	0.0076	4.6271	3.74E-06	0.0155
UMPS_eQTL_cis, superior longitudinal fasciculus	-0.0413	0.0076	-5.4562	4.94E-08	0.0008
TMEM154_eQTL_cis, superior longitudinal fasciculus	-0.0377	0.0076	-4.9633	7.01E-07	0.0044
PTPN13_eQTL_trans, superior thalamic radiation	-0.0336	0.0068	-4.9424	7.8E-07	0.0046
TMEM154_eQTL_cis, superior thalamic radiation	-0.0323	0.0068	-4.7333	2.23E-06	0.0106
TMEM154_eQTL_cis, uncinate fasciculus	-0.0308	0.0068	-4.5058	6.66E-06	0.0219

Table S2. eQTL scores associated only with white matter tracts as measured by MD. The first column indicates standardised effect size (β).

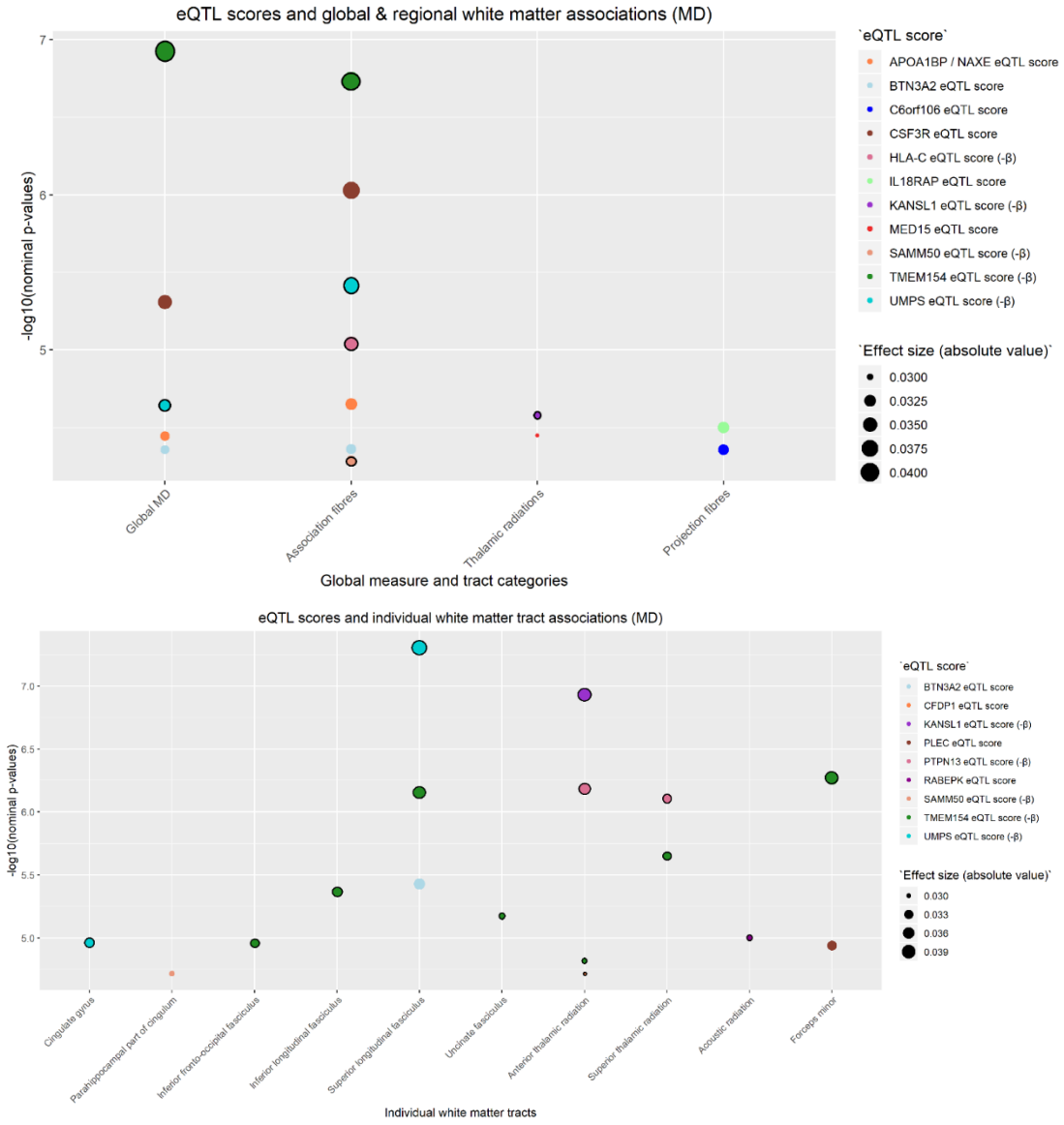


Figure S2. eQTL scores associated only with white matter tracts as measured by MD (mean diffusivity). Indicates nominal p-values between each of the scores (shown in legend entitled "eQTL score") and global and tract category measures (noted on the x-axis). All values in the figure met FDR correction. Some of the scores with an additional line around the points had an effect size in the opposite direction to all other scores (also indicated by -β for MD in figure legend). The colours of the plot points indicate the score to which they belong. Magnitude of effect is shown in the legend entitled "Effect size (absolute value)".

Brief gene look-up for genes whose expression was associated with FA (N = 17; table 1) and MD (N = 16; table 2) separately

Score name & eQTL type	N SNPs in score	Regulated gene	Study from which score is calculated	Gene function
ATG10_eQTL_cis	7	ATG10	Gusev et al.	E2-like enzyme involved in 2 ubiquitin-like modifications essential for autophagosome formation; expressed in brain (1)
SF3A1_eQTL_cis	23	SF3A1	Westra et al.	Expressed in brain; gene encodes a subunit of the splicing factor 3a protein complex (2)
SMARCAL1_eQTL_cis	1	SMARCAL1	Westra et al.	Protein encoded by this gene is a member of SWI/SNP family of proteins; members have helicase and ATPase activities and are thought to regulate transcription of certain genes by altering chromatin structure around those genes; expressed in brain; associated with Schimke immunoosseous dysplasia (3)
PPP4R3A_eQTL_cis	1	PPP4R3A	Westra et al,	Expressed in brain; may be involved in Alzheimer's disease risk (4)
CD14_eQTL_cis	18	CD14	Gusev et al.	Protein encoded by this gene is a surface antigen that is preferentially expressed on monocytes/macrophages; it cooperates with other proteins to mediate the innate immune response to bacterial lipopolysaccharide; expressed in brain (5)
COG7_eQTL_cis	1	COG7	Westra et al.	Protein encoded by this gene resides in the golgi and is part of 8 subunits of the conserved oligomeric Golgi (COG) complex; expressed in brain; mutations in gene associated with microcephaly, adducted thumbs, growth retardation, VSD and episodes of hyperthermia (6; 7)
LINC01605_eQTL_trans	1	LINC01605	Westra et al.	RNA gene; affiliated with non-coding RNA class; expression of gene associated with bladder cancer (8)
ANXA4_eQTL_cis	5	ANXA4	Gusev et al.	Little expression in brain; gene belongs to annexin family of calcium dependent phospholipid binding proteins (9)
ZSCAN26_eQTL_cis	32	ZSCAN26	Westra et al.	Expressed in brain; protein coding gene
SHTN1 / KIAA1598_eQTL_cis	5	SHTN1 / KIAA1598	Gusev et al.	Expressed in brain; involved in generation of internal asymmetric signals required for neuronal polarization and neurite outgrowth; mediated netrin-1-induced F-actin substrate coupling or clutch engagement within axon growth cone through activation of several genes & pathways (10)

ZNF282_eQTL_cis	7	ZNF282	Gusev et al.	Expressed in brain; diseases associated with gene: T-cell leukemia (11)
ENO4_eQTL_cis	7	ENO4	Westra et al.	Expressed in brain
ASRGL1_eQTL_cis	5	ASRGL1	Gusev et al.	Expressed in brain; may be involved in production of L-aspartate, which can act as an excitatory neurotransmitter in some brain regions; may be implicated in endometrioid endometrial carcinoma (12)
TMEM184B_eQTL_cis	12	TMEM184B	Gusev et al.	Expressed in brain; may be implicated in axon degeneration (13)
ZBTB7B_eQTL_cis	8	ZBTB7B	Gusev et al.	Expressed in brain; gene encodes a zinc finger-containing transcription factor that acts as a key regulator of lineage commitment of immature T-cell precursors (14)
GPT_eQTL_cis	5	GPT	Westra et al.	Little expression in brain
AP2S1_eQTL_cis	5	AP2S1	Westra et al.	One of 2 major clathrin-associated adaptor complexes, AP-2 is a heterotetramer which is associated with the plasma membrane; complex is composed of 2 large chains, 1 medium chain and 1 small chain, and the gene encodes the small chain; expressed in brain (15)

Table S3. Information regarding eQTL scores with significant associations FA-measured tracts. 1 score (LINC01605_eQTL_trans) is trans, while all others are cis.

Score name & eQTL type	**N SNPs in score**	**Regulated gene**	**Study from which score is calculated**	**Gene function**
APOA1BP / NAXE_eQTL_cis	10	APOA1BP / NAXE	Gusev et al.	Expressed in brain; diseases associated with gene: encephalopathy; brain edema (16)
BTN3A2_eQTL_cis	42	BTN3A2	Gusev et al.	May be involved in adaptive immune system response; may be involved in risk for gastric cancer (17)
UMPS_eQTL_cis	5	UMPS	Westra et al.	Encoded protein is a bifunctional enzyme that catalyzes the final 2 steps of the de novo pyrimidine biosynthetic pathway (18)
CSF3R_eQTL_cis	5	CSF3R	Westra et al.	Mutations in this gene are a cause of Kostmann syndrome / congenital neutropenia; not expressed in brain (19)
TMEM154_eQTL_cis	20	TMEM154	Westra et al.	Very little expression in brain

SAMM50_eQTL_cis	15	SAMM50	Gusev et al.	Gene encodes a component of the Sorting and Assembly Machinery of the mitochondrial outer membrane (20)
HLA-C_eQTL_cis	38	HLA-C	Westra et al.	Expressed in nearly all cells
MED15_eQTL_cis	6	MED15	Westra et al.	Expressed in brain
KANSL1_eQTL_cis	3	KANSL1	Westra et al.	Gene encodes a nuclear protein that is a subunit of 2 protein complexes involved with histone acetylation (21)
IL18RAP_eQTL_cis	12	IL18RAP	Gusev et al.	Mutations in this gene have been associated with Crohn's disease; expressed in brain (22)
C6orf106_eQTL_cis	5	C6orf106	Westra et al.	Expressed in cortex
RABEPK_eQTL_cis	9	RABEPK	Gusev et al.	Expressed in brain
CFDP1_eQTL_cis	4	CFDP1	Gusev et al.	Expressed in brain; may be implicated in coronary artery disease risk (23)
PTPN13_eQTL_trans	1	PTPN13	Westra et al.	Protein encoded by this gene is a member of the PTP family, which are signalling molecules that regulate cellular processes (e.g. cell growth, differentiation, mitotic cell cycle, oncogenic transformation); disease associated with this gene: tropical spastic paraparesis, a disease of the nervous system affecting people living near the equator; expressed in the brain (24)
EVL_eQTL_cis	1	EVL	Westra et al.	Expressed in brain; actin-associated proteins involved in processes such as axon guidance and lamellipodial and filopodial dynamics in migrating cells; enhances actin nucleation and polymerization (25)
PLEC_eQTL_cis	3	PLEC	Westra et al.	Prominent member of a protein family of proteins which interlink different elements of the cytoskeleton; expressed in a wide range of cell types and tissues (including brain) (26)

Table S4. Information regarding eQTL scores with significant associations MD-measured tracts. 1 score (PTPN13_eQTL_trans) is trans, while all others are cis.

Results for 8 scores associated with both FA and MD (N = 8).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Global FA	-0.0367	0.0079	-4.6474	3.39161E-06	0.0088
Thalamic radiations	-0.0403	0.0080	-5.0378	4.76577E-07	0.0025
Anterior thalamic radiations	-0.0429	0.0076	-5.6798	1.37465E-08	0.0002
Forceps minor	-0.0471	0.0078	-6.0115	1.88218E-09	0.0001
Superior longitudinal fasciculus	-0.0386	0.0077	-5.0327	4.89475E-07	0.0040
MD					
Global MD	0.0404	0.0075	5.3762	7.72382E-08	0.0015
Association fibres	0.0381	0.0077	4.9643	6.97256E-07	0.0045
Thalamic radiations	0.0327	0.0072	4.5625	5.09715E-06	0.0132
Acoustic radiation	0.0295	0.0069	4.2470	2.17989E-05	0.0472
Anterior thalamic radiations	0.0403	0.0070	5.7964	6.91525E-09	0.0003
Cingulate gyrus	0.0352	0.0073	4.7887	1.69554E-06	0.0085
Forceps minor	0.0480	0.0076	6.3085	2.89925E-10	2.76005E-05
Inferior fronto-occipital fasciculus	0.0410	0.0075	5.4805	4.31258E-08	0.0008
Inferior longitudinal fasciculus	0.0377	0.0073	5.1766	2.28961E-07	0.0024
Superior longitudinal fasciculus	0.0415	0.0076	5.4902	4.08256E-08	0.0008
Uncinate fasciculus	0.0314	0.0068	4.6086	4.08933E-06	0.0162

Table S5. Significant associations between DCAKD_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Global FA	-0.0403	0.0079	-5.0996	3.44595E-07	0.0022
Association fibres	-0.0347	0.0079	-4.4036	1.07241E-05	0.0198
Projection fibres	0.0453	0.0079	5.7612	8.51978E-09	0.0002
Acoustic radiation	-0.0326	0.0069	-4.7044	2.56987E-06	0.0133
Corticospinal tract	-0.0326	0.0074	-4.3945	1.11801E-05	0.0337
Forceps minor	-0.0561	0.0078	-7.1754	7.5595E-13	7.3217E-08
Inferior longitudinal fasciculus	-0.0335	0.0076	-4.3887	1.14829E-05	0.0337
Superior longitudinal fasciculus	-0.0367	0.0077	-4.7887	1.6956E-06	0.0103
MD					
Global MD	0.0308	0.0075	4.0893	4.3502E-05	0.0423
Forceps minor	0.0432	0.0076	5.6773	1.3949E-08	0.0004
Inferior longitudinal fasciculus	0.0362	0.0073	4.9676	6.8552E-07	0.0044

Table S6. Significant associations between SLC35A4_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Global FA	-0.0420	0.0079	-5.3199	1.0538E-07	0.0011
Association fibres	-0.0358	0.0079	-4.5425	5.6047E-06	0.0121
Thalamic radiations	-0.0388	0.0080	-4.8429	1.2928E-06	0.0048
Projection fibres	0.0416	0.0079	5.2850	1.275E-07	0.0011
Corticospinal tract	-0.0320	0.0074	-4.3116	1.6311E-05	0.0416
Forceps minor	-0.0456	0.0078	-5.8270	5.763E-09	0.0001
Inferior longitudinal fasciculus	-0.0419	0.0076	-5.4773	4.3905E-08	0.0006
Posterior thalamic radiation	-0.0352	0.0075	-4.7014	2.6076E-06	0.0133
Superior longitudinal fasciculus	-0.0392	0.0077	-5.1143	3.1895E-07	0.0028
MD					
Global MD	0.0326	0.0075	4.3299	1.5015E-05	0.0277
Acoustic radiation	0.0339	0.0069	4.8778	1.0844E-06	0.0060
Cingulate gyrus	0.0328	0.0073	4.4648	8.074E-06	0.0248
Forceps minor	0.0348	0.0076	4.5604	5.1479E-06	0.0188

Table S7. Significant associations between SEC14L4_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Projection fibres	0.0339	0.0079	4.3032	1.6943E-05	0.0273
Forceps minor	-0.0462	0.0078	-5.8981	3.7587E-09	0.0001
MD					
Forceps minor	0.0353	0.0076	4.6349	3.6022E-06	0.0155

Table S8. Significant associations between SRA1_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Anterior thalamic radiations	0.0324	0.0076	4.2863	1.8287E-05	0.0429
Forceps minor	0.0352	0.0078	4.4956	6.992E-06	0.0271
MD					
Global MD	-0.0328	0.0075	-4.3626	1.2941E-05	0.0257
Anterior thalamic radiations	-0.0339	0.0070	-4.8703	1.1263E-06	0.0060
Forceps minor	-0.0392	0.0076	-5.1537	2.5879E-07	0.0025
Inferior fronto-occipital fasciculus	-0.0335	0.0075	-4.4845	7.3652E-06	0.0234
Inferior longitudinal fasciculus	-0.0311	0.0073	-4.2695	1.9718E-05	0.0447
Superior longitudinal fasciculus	-0.0343	0.0076	-4.5355	5.7939E-06	0.0204

Table S9. Significant associations between NMT1_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Forceps major	0.0436	0.0081	5.3818	7.4908E-08	0.0009
Forceps minor	0.0338	0.0078	4.3185	1.5817E-05	0.0416
MD					
Global MD	-0.0366	0.0075	-4.8650	1.1564E-06	0.0050
Association fibres	-0.0368	0.0077	-4.7868	1.7111E-06	0.0063
Inferior longitudinal fasciculus	-0.0309	0.0073	-4.2303	2.3485E-05	0.0497
Superior longitudinal fasciculus	-0.0356	0.0076	-4.7055	2.5555E-06	0.0116

Table S10. Significant associations between CPNE1_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Forceps minor	-0.0347	0.0078	-4.4321	9.4015E-06	0.0337
MD					
Global MD	0.0330	0.0075	4.3859	1.1631E-05	0.0250
Association fibres	0.0318	0.0077	4.1395	3.5002E-05	0.0423
Thalamic radiations	0.0296	0.0072	4.1282	3.6762E-05	0.0423
Anterior thalamic radiations	0.0356	0.0070	5.1101	3.2604E-07	0.0028
Forceps minor	0.0334	0.0076	4.3876	1.154E-05	0.0297
Superior longitudinal fasciculus	0.0342	0.0076	4.5223	6.1651E-06	0.0210

Table S11. Significant associations between PLEKHM1_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

White Matter Tracts	Effect size	SD	t value	p value	p value, FDR corrected
FA					
Forceps minor	-0.0382	0.0078	-4.8721	1.1158E-06	0.0077
MD					
Forceps minor	0.0331	0.0076	4.3465	1.3925E-05	0.0349
Inferior fronto-occipital fasciculus	0.0332	0.0075	4.4413	9.01E-06	0.0268

Table S12. Significant associations between UBE3C_eQTL_cis and FA and MD-measured white matter tracts. The first column indicates standardised effect size (β).

GWAS quality check and parameters

In order to determine whether any score SNPs were previously associated with global and tract category measures of interest (i.e. tract categories & global measures found to be significantly associated with the 8 eQTL scores), 8 GWAS were run locally (the 3 tract categories: association fibres, thalamic radiations, and projection fibres for both FA and MD, and global measures for FA and MD). BGENIE (1) was used to conduct the association analysis and excluded related participants (up to the third degree using the KING toolset (2)), as well as those who also participated in Generation Scotland and PGC MDD GWAS. Only variants with a minor allele frequency (MAF) > 0.001 (0.1%), SNP information score (quality of imputation) > 0.1, and Hardy-Weinberg equilibrium (HWE) p-value >= 1e-6 were examined. Sex, age, the first 8 principal components, genotyping array, and three head position coordinates were fitted as covariates in the analysis.

The output summary statistics files contain information with regards to the chromosome, SNP ID, p-value and effect size of association with each phenotype. The SNPs significantly associated with the tracts of interest were noted and the effect size and p-value for each was extracted.

Individual white matter tracts and tract category to which they belong

Fractional anisotropy and mean diffusivity
Association fibres
Inferior fronto-occipital fasciculus
Uncinate fasciculus
Parahippocampal cingulum
Cingulate gyrus
Superior longitudinal fasciculus
Inferior longitudinal fasciculus
Thalamic radiations
Anterior thalamic radiation
Posterior thalamic radiation
Superior thalamic radiation
Projection fibres
Acoustic radiation
Medial lemniscus
Forceps major*
Forceps minor*
Middle cerebellar peduncle*
Corticospinal tract
Global FA & global MD
Table S13. White matter tracts, global white matter and tract categories for FA and MD. * indicates unilateral regions.

P-values and effect size of each SNP for each individual white matter tract results (Elliott et al., 2018) and global and regional results (run locally); figure produced locally

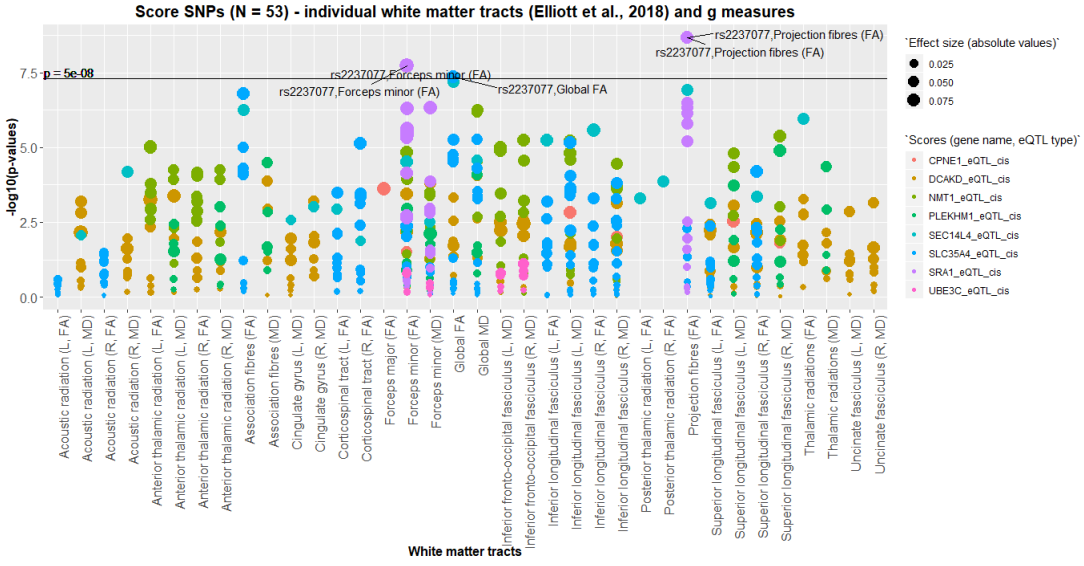


Figure S3. GWA between SNPs, individual white matter tracts of interest (Elliott et al., 2018) and global and tract category measures (run locally). Each point on the plot corresponds to one SNP. White matter tracts of interest are noted on the x-axis (L = left; R = right; FA = fractional anisotropy; MD = mean diffusivity). The 8 colours of the plot points indicate the score to which they belong (shown in "Scores (gene name, eQTL type)" legend). Magnitude of effect is shown in the legend entitled "Effect size (absolute values)". The horizontal line indicates genome-wide significance (5e-8).

P-values and effect size for each SNP in association with gene expression as taken from GENOSCORES; these values were obtained in the two discovery datasets used in the current study (Gusev et al., 2016; Westra et al., 2013).

Chromosome	SNP	Gene	Effect size	P-value
17	rs4793119	DCAKD	0.188	3.98E-08
17	rs17682536	DCAKD	-0.114	1.20E-13
17	rs962888	DCAKD	0.284	4.14E-19
17	rs9898793	DCAKD	0.343	3.53E-240
17	rs2040558	DCAKD	-0.171	1.73E-52
17	rs2239921	DCAKD	0.367	1.03E-28
17	rs3744760	DCAKD	0.365	2.47E-184
17	rs4986172	DCAKD	0.093	1.07E-24
5	rs269783	SLC35A4	-0.114	6.55E-11
5	rs13175916	SLC35A4	-0.025	1.33E-09
5	rs2237077	SLC35A4	0.322	0
5	rs1862176	SLC35A4	0.223	0
5	rs6860077	SLC35A4	0.210	0
5	rs17286676	SLC35A4	-0.041	9.04E-97
5	rs250430	SLC35A4	0.087	4.09E-86
5	rs250429	SLC35A4	0.208	0
5	rs12517200	SLC35A4	0.061	7.19E-298
5	rs1583005	SLC35A4	0.138	5.452E-06
5	rs2286394	SLC35A4	-0.055	4.40E-39
5	rs3733709	SLC35A4	-0.110	5.77E-24
22	rs2267161	SEC14L4	-0.093	6.96E-06
5	rs2237077	SRA1	-0.025	4.55E-19
5	rs1862176	SRA1	-0.040	1.72E-39
5	rs6860077	SRA1	-0.039	3.91E-41
5	rs1835959	SRA1	-0.087	1.14E-20
5	rs250430	SRA1	-0.087	1.12E-20
5	rs250429	SRA1	-0.039	2.05E-42

5	rs2569163	SRA1	0.048	3.13E-24
5	rs778582	SRA1	-0.016	2.08E-34
5	rs12517200	SRA1	-0.013	6.77E-34
5	rs1583005	SRA1	-0.005	1.10E-11
5	rs2530241	SRA1	0.001	3.92E-14
5	rs801186	SRA1	0.010	3.28E-18
5	rs801171	SRA1	0.001	5.85E-15
5	rs2531360	SRA1	0.000	6.17E-14
5	rs2240696	SRA1	-0.007	6.31E-12
17	rs9898793	NMT1	0.015	3.69E-08
17	rs4793172	NMT1	0.029	1.41E-10
17	rs2239916	NMT1	0.035	1.17E-14
17	rs1053739	NMT1	0.032	2.77E-13
17	rs3744760	NMT1	0.032	2.85E-10
17	rs12946454	NMT1	0.028	8.33E-09
17	rs4986172	NMT1	0.030	5.50E-09
6	rs4324798	CPNE1	0.181	1.79E-06
17	rs9898793	PLEKHM1	0.394	6.38E-78
17	rs2239921	PLEKHM1	0.304	6.23E-13
17	rs3744760	PLEKHM1	0.247	2.54E-79
17	rs4986172	PLEKHM1	-0.138	3.73E-06
17	rs1552458	PLEKHM1	0.317	2.91E-47
7	rs17646960	UBE3C	0.342	1.49E-20
7	rs1182398	UBE3C	-0.285	1.07E-33
7	rs1182393	UBE3C	-0.298	9.37E-43
7	rs2527866	UBE3C	-0.297	2.95E-06

Table S14. Associations between SNPs found in the 8 eQTL scores and gene expression ($N_{total} = 53$); effect size and p-values are taken from the two GWAS studies (Gusev et al., 2016; Westra et al., 2013) available in GENOSCORES.

eQTL score computation process

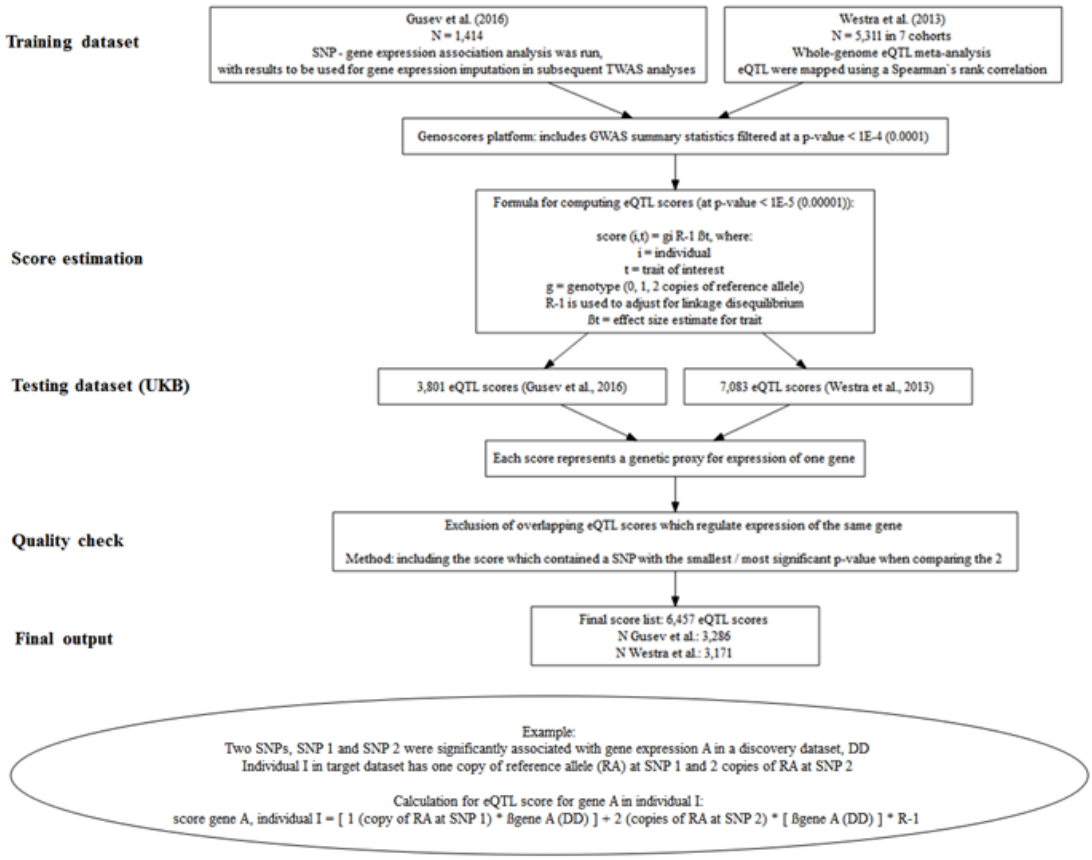


Figure S4. eQTL score computation process.

Supplementary material references

1. Phillips, A. R., Suttangkakul, A., & Vierstra, R. D. (2008). The ATG12-conjugating enzyme ATG10 is essential for autophagic vesicle formation in Arabidopsis thaliana. *Genetics*, *178*(3), 1339-1353.
2. Sharma, S., Wongpalee, S. P., Vashisht, A., Wohlschlegel, J. A., & Black, D. L. (2014). Stem–loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes & development*, *28*(22), 2518-2531.
3. Bansbach, C. E., Bétous, R., Lovejoy, C. A., Glick, G. G., & Cortez, D. (2009). The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes & development*, *23*(20), 2405-2414.
4. Christopher, L., Napolioni, V., Khan, R. R., Han, S. S., Greicius, M. D., & Alzheimer's Disease Neuroimaging Initiative. (2017). A variant in PPP4R3A protects against alzheimer-related metabolic decline. *Annals of neurology*, *82*(6), 900-911.
5. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J., & Mathison, J. C. (1990). CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science*, *249*(4975), 1431-1433.
6. Steet, R., & Kornfeld, S. (2006). COG-7-deficient human fibroblasts exhibit altered recycling of Golgi proteins. *Molecular biology of the cell*, *17*(5), 2312-2321.
7. Morava, E., Zeevaert, R., Korsch, E., Huijben, K., Wopereis, S., Matthijs, G. et al. (2007). A common mutation in the COG7 gene with a consistent phenotype including microcephaly, adducted thumbs, growth retardation, VSD and episodes of hyperthermia. *European Journal of Human Genetics*, *15*(6), 638.
8. Qin, Z., Wang, Y., Tang, J., Zhang, L., Li, R., Xue, J. et al. (2018). High LINC01605 expression predicts poor prognosis and promotes tumor progression via up-regulation of MMP9 in bladder cancer. *Bioscience reports*, *38*(5), BSR20180562.

9. Massé, K. L., Collins, R., Bhamra, S., Seville, R. A., & Jones, E. (2007). Anxa4 genes are expressed in distinct organ systems in xenopus laevis and tropicalis but are functionally conserved. *Organogenesis*, *3*(2), 83-92.
10. Ergin, V., Erdogan, M., & Menevse, A. (2015). Regulation of shootin1 gene expression involves ngf-induced alternative splicing during neuronal differentiation of PC12 cells. *Scientific reports*, *5*, 17931.
11. Yeo, S. Y., Ha, S. Y., Yu, E. J., Lee, K. W., Kim, J. H., & Kim, S. H. (2014). ZNF282 (Zinc finger protein 282), a novel E2F1 co-activator, promotes esophageal squamous cell carcinoma. *Oncotarget*, *5*(23), 12260.
12. Edqvist, P. H. D., Huvila, J., Forsström, B., Talve, L., Carpén, O., Salvesen, H. B. et al. (2015). Loss of ASRGL1 expression is an independent biomarker for disease-specific survival in endometrioid endometrial carcinoma. *Gynecologic oncology*, *137*(3), 529-537.
13. Bhattacharya, M. R., Geisler, S., Pittman, S. K., Doan, R. A., Weihl, C. C., Milbrandt, J. et al. (2016). TMEM184b promotes axon degeneration and neuromuscular junction maintenance. *Journal of Neuroscience*, *36*(17), 4681-4689.
14. Wang, L., Wildt, K. F., Castro, E., Xiong, Y., Feigenbaum, L., Tessarollo, L. et al. (2008). The zinc finger transcription factor Zbtb7b represses CD8-lineage gene expression in peripheral CD4+ T cells. *Immunity*, *29*(6), 876-887.
15. Nesbit, M. A., Hannan, F. M., Howles, S. A., Reed, A. A., Cranston, T., Thakker, C. E. et al. (2013). Mutations in AP2S1 cause familial hypocalciuric hypercalcemia type 3. *Nature genetics*, *45*(1), 93.
16. Spiegel, R., Shaag, A., Shalev, S., & Elpeleg, O. (2016). Homozygous mutation in the APOA1BP is associated with a lethal infantile leukoencephalopathy. *Neurogenetics*, *17*(3), 187-190.
17. Zhu, M., Yan, C., Ren, C., Huang, X., Zhu, X., Gu, H. et al. (2017). Exome array analysis identifies variants in SPOCD1 and BTN3A2 that affect risk for gastric cancer. *Gastroenterology*, *152*(8), 2011-2021.
18. Evans, D. R., & Guy, H. I. (2004). Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway. *Journal of Biological Chemistry*, *279*(32), 33035-33038.

19. Maxson, J. E., Gotlib, J., Pollyea, D. A., Fleischman, A. G., Agarwal, A., Eide, C. A. et al. (2013). Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML. *New England Journal of Medicine*, *368*(19), 1781-1790.
20. Rhee, H. W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A. et al. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, *339*(6125), 1328-1331.
21. Zollino, M., Orteschi, D., Murdolo, M., Lattante, S., Battaglia, D., Stefanini, C. et al. (2012). Mutations in KANSL1 cause the 17q21. 31 microdeletion syndrome phenotype. *Nature genetics*, *44*(6), 636.
22. Zhernakova, A., Festen, E. M., Franke, L., Trynka, G., van Diemen, C. C., Monsuur, A. J. et al. (2008). Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *The American Journal of Human Genetics*, *82*(5), 1202-1210.
23. Gertow, K., Sennblad, B., Strawbridge, R. J., Öhrvik, J., Zabaneh, D., Shah, S. et al. (2012). Identification of the BCAR1-CFDP1-TMEM170A locus as a determinant of carotid intima-media thickness and coronary artery disease risk. *Circulation: Cardiovascular Genetics*, *5*(6), 656-665.
24. Zhu, J. H., Chen, R., Yi, W., Cantin, G. T., Fearns, C., Yang, Y. et al. (2008). Protein tyrosine phosphatase PTPN13 negatively regulates Her2/ErbB2 malignant signaling. *Oncogene*, *27*(18), 2525.
25. Wills, Z., Bateman, J., Korey, C. A., Comer, A., & Van Vactor, D. (1999). The tyrosine kinase Abl and its substrate enabled collaborate with the receptor phosphatase Dlar to control motor axon guidance. *Neuron*, *22*(2), 301-312.
26. Niwa, T., Saito, H., Imajoh-ohmi, S., Kaminishi, M., Seto, Y., Miki, Y. et al. (2009). BRCA2 interacts with the cytoskeletal linker protein plectin to form a complex controlling centrosome localization. *Cancer science*, *100*(11), 2115-2125.
27. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K. et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203.

28. Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics, 26(22), 2867-2873.

Appendix 2: Supplementary materials for Chapter 3: Association of whole-genome and NETRIN1 signaling pathway-derived polygenic risk scores for Major Depressive Disorder and white matter microstructure in UK Biobank

Supplementary notes

- Demographic data concerning complete dataset of individuals with DTI values
- Descriptive statistics of imaging phenotype
- NETRIN1 signalling pathway gene list
- Demographic data and FA descriptive statistics of individuals excluded from the study (N = 19)
- Demographic data and MD descriptive statistics of individuals excluded from the study (N = 30)
- Statistical analysis of FA and MD values containing:
    1. Unpruned NETRIN1- and genomic-PRS with outliers excluded (6,401 for FA and 6,390 for MD) at all 5 thresholds (0.01, 0.05, 0.1, 0.5, 1) and full sample (6,420) at threshold 0.5
    2. Pruned NETRIN1- and Genomic-PRS with outliers excluded (6,401 for FA and 6,390 for MD) at all 5 thresholds (0.01, 0.05, 0.1, 0.5, 1) and full sample (6,420) at threshold 0.5
- White matter tracts significantly associated with both NETRIN1-PRS and genomic-PRS.
    1. Fractional anisotropy
    2. Mean diffusivity

Demographic data concerning complete dataset of individuals with DTI values

Complete dataset (N = 6,420): N female = 3,345; N male = 3,075; mean age: 62.62 +/- 7.37 years; age range: 45.92 – 78.42

Descriptive statistics of imaging phenotype

Fractional Anisotropy				
	Full dataset (N = 6,420)		Outliers excluded dataset (N = 6,401)	
White matter tract	Mean	SD	Mean	SD
Cingulate gyrus part of cingulum (left)	0.535	0.035	0.535	0.033
Cingulate gyrus part of cingulum (right)	0.497	0.034	0.498	0.033
Parahippocampal part of cingulum (left)	0.314	0.029	0.314	0.028
Parahippocampal part of cingulum (right)	0.313	0.030	0.313	0.030
Inferior fronto-occipital fasciculus (left)	0.475	0.024	0.476	0.022
Inferior fronto-occipital fasciculus (right)	0.465	0.021	0.465	0.020
Inferior longitudinal fasciculus (left)	0.460	0.021	0.460	0.019
Inferior longitudinal fasciculus (right)	0.451	0.020	0.451	0.018
Superior longitudinal fasciculus (left)	0.440	0.022	0.440	0.020
Superior longitudinal fasciculus (right)	0.423	0.021	0.424	0.019
Uncinate fasciculus (left)	0.388	0.024	0.388	0.235
Uncinate fasciculus (right)	0.390	0.021	0.390	0.020
Anterior thalamic radiation (left)	0.399	0.019	0.399	0.017
Anterior thalamic radiation (right)	0.392	0.019	0.392	0.017
Posterior thalamic radiation (left)	0.458	0.022	0.458	0.020
Posterior thalamic radiation (right)	0.455	0.022	0.456	0.020

Superior thalamic radiation (left)	0.422	0.019	0.423	0.018
Superior thalamic radiation (right)	0.422	0.020	0.422	0.018
Acoustic radiation (left)	0.419	0.023	0.420	0.021
Acoustic radiation (right)	0.411	0.022	0.412	0.020
Corticospinal tract (left)	0.545	0.024	0.545	0.022
Corticospinal tract (right)	0.539	0.025	0.539	0.022
Medial lemniscus (left)	0.419	0.024	0.419	0.023
Medial lemniscus (right)	0.422	0.025	0.422	0.024
Forceps major	0.580	0.029	0.580	0.027
Forceps minor	0.465	0.022	0.465	0.020
Middle cerebellar peduncle	0.481	0.031	0.481	0.029

Table S1. Descriptive statistics of FA values (mean and standard deviation). The full dataset contains 6,420 individuals, while the outliers-excluded dataset contains 6,401 individuals.

Mean Diffusivity				
	Full dataset (N = 6,420)		Outliers excluded dataset (N = 6,390)	
White matter tract	Mean	SD	Mean	SD
Cingulate gyrus part of cingulum (left)	0.0007	0.00003	0.0007	0.00002
Cingulate gyrus part of cingulum (right)	0.0007	0.00003	0.0007	0.00002
Parahippocampal part of cingulum (left)	0.0008	0.00006	0.0008	0.00005
Parahippocampal part of cingulum (right)	0.0008	0.00006	0.0008	0.00005
Inferior fronto-occipital fasciculus (left)	0.0008	0.00003	0.0008	0.00003
Inferior fronto-occipital fasciculus (right)	0.0008	0.00003	0.0008	0.00003
Inferior longitudinal fasciculus (left)	0.0008	0.00003	0.0008	0.00003
Inferior longitudinal fasciculus (right)	0.0008	0.00003	0.0008	0.00003
Superior longitudinal fasciculus (left)	0.0007	0.00003	0.0007	0.00003

Superior longitudinal fasciculus (right)	0.0007	0.00003	0.0007	0.00003
Uncinate fasciculus (left)	0.0008	0.00004	0.0008	0.00003
Uncinate fasciculus (right)	0.0008	0.00003	0.0008	0.00003
Anterior thalamic radiation (left)	0.0007	0.00003	0.0007	0.00003
Anterior thalamic radiation (right)	0.0007	0.00003	0.0007	0.00003
Posterior thalamic radiation (left)	0.0008	0.00004	0.0008	0.00004
Posterior thalamic radiation (right)	0.0008	0.00004	0.0008	0.00004
Superior thalamic radiation (left)	0.0007	0.00003	0.0007	0.00002
Superior thalamic radiation (right)	0.0007	0.00003	0.0007	0.00002
Acoustic radiation (left)	0.0007	0.00004	0.0007	0.00003
Acoustic radiation (right)	0.0007	0.00004	0.0007	0.00003
Corticospinal tract (left)	0.0007	0.00002	0.0007	0.00002
Corticospinal tract (right)	0.0007	0.00002	0.0007	0.00002
Medial lemniscus (left)	0.0009	0.00004	0.0009	0.00003
Medial lemniscus (right)	0.0009	0.00004	0.0009	0.00003
Forceps major	0.0009	0.00005	0.0009	0.00005
Forceps minor	0.0008	0.00003	0.0008	0.00003
Middle cerebellar peduncle	0.0007	0.00006	0.0007	0.00006

Table S2. Descriptive statistics of MD values (mean and standard deviation). The full dataset contains 6,420 individuals, while the outliers-excluded dataset contains 6,390 individuals.

NETRIN1 signalling pathway gene list

Gene name	**Description**
UNC5D	unc-5 homolog D (C. elegans)
HFE2	hemochromatosis type 2 (juvenile)
DCC	deleted in colorectal carcinoma
DOCK1	dedicator of cytokinesis 1
UNC5B	unc-5 homolog B (C. elegans)
ABLIM3	actin binding LIM protein family, member 3
FYN	FYN oncogene related to SRC, FGR, YES
RGMB	RGM domain family, member B

ABLIM1	actin binding LIM protein 1
MYO10	myosin X
NCK1	NCK adaptor protein 1
NEO1	neogenin 1
PITPNA	phosphatidylinositol transfer protein, alpha
PLCG1	phospholipase C, gamma 1
PRKCQ	protein kinase C, theta
RGMA	RGM domain family, member A
TRPC7	transient receptor potential cation channel
PTK2	PTK2 protein tyrosine kinase 2
RAC1	ras-related C3 botulinum toxin substrate 1 percursor
NTN4	netrin 4
ROBO1	roundabout, axon guidance receptor, homolog 1
SIAH1	seven in absentia homolog 1 (Drosophila)
SIAH2	seven in absentia homolog 2 (Drosophila)
SLIT1	slit homolog 1 (Drosophila)
SLIT3	slit homolog 3 (Drosophila)
SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene
TRIO	triple functional domain (PTPRF interacting)
TRPC3	transient receptor potential cation channel
TRPC4	transient receptor potential cation channel
TRPC5	transient receptor potential cation channel
TRPC6	transient receptor potential cation channel
LOC730030	---
LOC730221	---
LOC730335	---
LOC730221	---
LOC730030	---
EZR	ezrin
UNC5C	unc-5 homolog C (C. elegans)
WASL	Wiskott-Aldrich syndrome-like
UNC5A	unc-5 homolog A (C. elegans)
SLIT2	slit homolog 2 (Drosophila)
NTN1	netrin 1
CDC42	cell division cycle 42 (GTP binding protein)

Table S3. Gene list and brief gene description included in the NETRIN1 signalling pathway, composed of 43 genes.

Demographic data and FA descriptive statistics of individuals excluded from the study (N = 19)

N female = 11; N male = 8; mean age: 69.26 +/- 4.53 years; age range: 58.92 – 77.42

Fractional Anisotropy		
	Outlier dataset (N = 19)	
White matter tract	Mean	SD
Cingulate gyrus part of cingulum (left)	0.407	0.149
Cingulate gyrus part of cingulum (right)	0.388	0.144
Parahippocampal part of cingulum (left)	0.246	0.091
Parahippocampal part of cingulum (right)	0.254	0.095
Inferior fronto-occipital fasciculus (left)	0.354	0.127
Inferior fronto-occipital fasciculus (right)	0.354	0.127
Inferior longitudinal fasciculus (left)	0.348	0.124

Inferior longitudinal fasciculus (right)	0.338	0.122
Superior longitudinal fasciculus (left)	0.325	0.117
Superior longitudinal fasciculus (right)	0.309	0.112
Uncinate fasciculus (left)	0.296	0.106
Uncinate fasciculus (right)	0.301	0.107
Anterior thalamic radiation (left)	0.306	0.110
Anterior thalamic radiation (right)	0.306	0.109
Posterior thalamic radiation (left)	0.358	0.127
Posterior thalamic radiation (right)	0.350	0.126
Superior thalamic radiation (left)	0.335	0.119
Superior thalamic radiation (right)	0.336	0.120
Acoustic radiation (left)	0.324	0.116
Acoustic radiation (right)	0.320	0.116
Corticospinal tract (left)	0.436	0.156
Corticospinal tract (right)	0.431	0.155
Medial lemniscus (left)	0.353	0.127
Medial lemniscus (right)	0.353	0.130
Forceps major	0.460	0.166
Forceps minor	0.346	0.125
Middle cerebellar peduncle	0.381	0.171

Table S4. Descriptive statistics of FA values (mean and standard deviation) for individuals excluded from the study (N = 19).

Demographic data and MD descriptive statistics of individuals excluded from the study (N = 30)

N female = 18; N male = 12; mean age: 70.29 +/- 4.66 years; age range: 58.92 – 77.42

Mean Diffusivity		
	Outlier dataset (N = 30)	
White matter tract	Mean	SD
Cingulate gyrus part of cingulum (left)	0.0007	0.0002
Cingulate gyrus part of cingulum (right)	0.0007	0.0002
Parahippocampal part of cingulum (left)	0.0009	0.0002
Parahippocampal part of cingulum (right)	0.0009	0.0002
Inferior fronto-occipital fasciculus (left)	0.0008	0.0002
Inferior fronto-occipital fasciculus (right)	0.0008	0.0002
Inferior longitudinal fasciculus (left)	0.0008	0.0002
Inferior longitudinal fasciculus (right)	0.0008	0.0002
Superior longitudinal fasciculus (left)	0.0008	0.0002
Superior longitudinal fasciculus (right)	0.0008	0.0002
Uncinate fasciculus (left)	0.0008	0.0002
Uncinate fasciculus (right)	0.0008	0.0002
Anterior thalamic radiation (left)	0.0008	0.0002
Anterior thalamic radiation (right)	0.0008	0.0002
Posterior thalamic radiation (left)	0.0009	0.0002
Posterior thalamic radiation (right)	0.0009	0.0002
Superior thalamic radiation (left)	0.0007	0.0002
Superior thalamic radiation (right)	0.0007	0.0002
Acoustic radiation (left)	0.0008	0.0002
Acoustic radiation (right)	0.0008	0.0002

Corticospinal tract (left)	0.0007	0.0002
Corticospinal tract (right)	0.0007	0.0002
Medial lemniscus (left)	0.0008	0.0002
Medial lemniscus (right)	0.0008	0.0002
Forceps major	0.0009	0.0002
Forceps minor	0.0008	0.0002
Middle cerebellar peduncle	0.0007	0.0002

Table S5. Descriptive statistics of MD values (mean and standard deviation) for individuals excluded from the study (N = 30).

Statistical analysis of FA and MD values containing:

Unpruned NETRIN1- and genomic-PRS with outliers included (6,420) and outliers excluded (6,401 for FA and 6,390 for MD) at all 5 thresholds (0.01, 0.05, 0.1, 0.5, 1)

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 acoustic radiation	-0.002	0.011	-0.166	0.868
NETRIN1 anterior_thalamic_radiation	-0.021	0.011	-1.833	0.067
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.013	0.011	-1.255	0.209
NETRIN1 parahippocampal_part_of_cingulum	-0.006	0.011	-0.526	0.599
NETRIN1 corticospinal_tract	-0.019	0.011	-1.701	0.089
NETRIN1 inferior_fronto_occipital_fasciculus	-0.019	0.012	-1.647	0.100
NETRIN1 inferior_longitudinal_fasciculus	-0.021	0.012	-1.813	0.070
NETRIN1 medial_lemniscus	-0.008	0.010	-0.735	0.462
NETRIN1 posterior_thalamic_radiation	-0.011	0.011	-0.981	0.326
NETRIN1 superior_longitudinal_fasciculus	-0.026	0.012	-2.254	0.024
NETRIN1 superior_thalamic_radiation	-0.015	0.012	-1.251	0.211
NETRIN1 uncinate_fasciculus	-0.018	0.011	-1.680	0.093
NETRIN1 bl.FA.wm.forceps_major	-0.017	0.012	-1.409	0.159
NETRIN1 bl.FA.wm.forceps_minor	-0.011	0.012	-0.934	0.351
NETRIN1 bl.FA.wm.middle_cerebellar_peduncle	-0.029	0.012	-2.333	0.020
Genomic acoustic_radiation	-0.011	0.011	-1.033	0.301
Genomic anterior_thalamic_radiation	-0.015	0.012	-1.315	0.188
Genomic cingulate_gyrus_part_of_cingulum	-0.016	0.011	-1.528	0.127
Genomic parahippocampal_part_of_cingulum	-0.019	0.011	-1.779	0.075
Genomic corticospinal_tract	-0.008	0.011	-0.666	0.505
Genomic inferior_fronto_occipital_fasciculus	-0.023	0.012	-2.008	0.045
Genomic inferior_longitudinal_fasciculus	-0.023	0.012	-1.959	0.050
Genomic medial_lemniscus	0.003	0.010	0.306	0.760
Genomic posterior_thalamic_radiation	-0.021	0.011	-1.873	0.061
Genomic superior_longitudinal_fasciculus	-0.026	0.012	-2.244	0.025
Genomic superior_thalamic_radiation	-0.010	0.012	-0.867	0.386
Genomic uncinate_fasciculus	-0.028	0.011	-2.545	0.011
Genomic bl.FA.wm.forceps_major	-0.037	0.012	-3.042	0.002
Genomic bl.FA.wm.forceps_minor	-0.031	0.012	-2.600	0.009
Genomic bl.FA.wm.middle_cerebellar_peduncle	-0.009	0.012	-0.730	0.465
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.009	0.011	0.819	0.413
NETRIN1 anterior_thalamic_radiation	-0.015	0.011	-1.282	0.200
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.011	0.011	-1.065	0.287

NETRIN1 parahippocampal_part_of_cingulum	-0.013	0.011	-1.197	0.232
NETRIN1 corticospinal_tract	-0.003	0.011	-0.276	0.782
NETRIN1 inferior_fronto_occipital_fasciculus	-0.005	0.011	-0.473	0.636
NETRIN1 inferior_longitudinal_fasciculus	-0.010	0.011	-0.841	0.400
NETRIN1 medial_lemniscus	-0.005	0.010	-0.456	0.649
NETRIN1 posterior_thalamic_radiation	-0.002	0.011	-0.205	0.838
NETRIN1 superior_longitudinal_fasciculus	-0.015	0.012	-1.265	0.206
NETRIN1 superior_thalamic_radiation	-0.001	0.012	-0.049	0.961
NETRIN1 uncinate_fasciculus	-0.009	0.011	-0.876	0.381
NETRIN1 bl.FA.wm.forceps_major	-0.008	0.012	-0.644	0.520
NETRIN1 bl.FA.wm.forceps_minor	-0.005	0.012	-0.397	0.691
NETRIN1 bl.FA.wm.middle_cerebellar_peduncle	-0.018	0.012	-1.461	0.144
Genomic acoustic_radiation	-0.012	0.011	-1.151	0.250
Genomic anterior_thalamic_radiation	-0.017	0.011	-1.459	0.145
Genomic cingulate_gyrus_part_of_cingulum	-0.019	0.011	-1.739	0.082
Genomic parahippocampal_part_of_cingulum	-0.019	0.011	-1.771	0.077
Genomic corticospinal_tract	-0.012	0.011	-1.037	0.300
Genomic inferior_fronto_occipital_fasciculus	-0.026	0.012	-2.292	0.022
Genomic inferior_longitudinal_fasciculus	-0.026	0.012	-2.252	0.024
Genomic medial_lemniscus	0.000	0.010	0.037	0.970
Genomic posterior_thalamic_radiation	-0.026	0.011	-2.357	0.018
Genomic superior_longitudinal_fasciculus	-0.029	0.012	-2.500	0.012
Genomic superior_thalamic_radiation	-0.015	0.012	-1.296	0.195
Genomic uncinate_fasciculus	-0.030	0.011	-2.725	0.006
Genomic bl.FA.wm.forceps_major	-0.037	0.012	-3.083	0.002
Genomic bl.FA.wm.forceps_minor	-0.034	0.012	-2.834	0.005
Genomic bl.FA.wm.middle_cerebellar_peduncle	-0.012	0.012	-0.983	0.326
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.005	0.011	0.443	0.658
NETRIN1 anterior_thalamic_radiation	-0.018	0.011	-1.580	0.114
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.016	0.011	-1.528	0.127
NETRIN1 parahippocampal_part_of_cingulum	-0.006	0.011	-0.580	0.562
NETRIN1 corticospinal_tract	-0.004	0.011	-0.314	0.753
NETRIN1 inferior_fronto_occipital_fasciculus	-0.013	0.011	-1.090	0.276
NETRIN1 inferior_longitudinal_fasciculus	-0.013	0.012	-1.091	0.275
NETRIN1 medial_lemniscus	-0.016	0.010	-1.569	0.117
NETRIN1 posterior_thalamic_radiation	-0.005	0.011	-0.481	0.631
NETRIN1 superior_longitudinal_fasciculus	-0.024	0.012	-2.065	0.039
NETRIN1 superior_thalamic_radiation	-0.010	0.012	-0.827	0.408
NETRIN1 uncinate_fasciculus	-0.008	0.011	-0.756	0.450
NETRIN1 bl.FA.wm.forceps_major	-0.014	0.012	-1.145	0.252
NETRIN1 bl.FA.wm.forceps_minor	-0.011	0.012	-0.934	0.350
NETRIN1 bl.FA.wm.middle_cerebellar_peduncle	-0.013	0.012	-1.049	0.294
Genomic acoustic_radiation	-0.012	0.011	-1.147	0.251
Genomic anterior_thalamic_radiation	-0.014	0.011	-1.186	0.236
Genomic cingulate_gyrus_part_of_cingulum	-0.018	0.011	-1.699	0.089
Genomic parahippocampal_part_of_cingulum	-0.017	0.011	-1.552	0.121
Genomic corticospinal_tract	-0.016	0.011	-1.379	0.168
Genomic inferior_fronto_occipital_fasciculus	-0.025	0.012	-2.177	0.030
Genomic inferior_longitudinal_fasciculus	-0.024	0.012	-2.121	0.034
Genomic medial_lemniscus	0.001	0.010	0.139	0.890
Genomic posterior_thalamic_radiation	-0.022	0.011	-2.002	0.045
Genomic superior_longitudinal_fasciculus	-0.026	0.012	-2.267	0.023
Genomic superior_thalamic_radiation	-0.014	0.012	-1.188	0.235

Genomic uncinate_fasciculus	-0.032	0.011	-2.915	0.004
Genomic bl.FA.wm.forceps_major	-0.033	0.012	-2.753	0.006
Genomic bl.FA.wm.forceps_minor	-0.031	0.012	-2.572	0.010
Genomic bl.FA.wm.middle_cerebellar_peduncle	-0.018	0.012	-1.465	0.143
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.003	0.011	0.294	0.769
NETRIN1 anterior_thalamic_radiation	-0.023	0.011	-2.051	0.040
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.029	0.011	-2.720	0.007
NETRIN1 parahippocampal_part_of_cingulum	-0.007	0.011	-0.692	0.489
NETRIN1 corticospinal_tract	0.001	0.011	0.120	0.905
NETRIN1 inferior_fronto_occipital_fasciculus	-0.024	0.011	-2.070	0.039
NETRIN1 inferior_longitudinal_fasciculus	-0.023	0.011	-1.978	0.048
NETRIN1 medial_lemniscus	-0.008	0.010	-0.757	0.449
NETRIN1 posterior_thalamic_radiation	-0.015	0.011	-1.360	0.174
NETRIN1 superior_longitudinal_fasciculus	-0.035	0.012	-3.017	0.003
NETRIN1 superior_thalamic_radiation	-0.006	0.012	-0.517	0.605
NETRIN1 uncinate_fasciculus	-0.019	0.011	-1.799	0.072
NETRIN1 forceps_major	-0.016	0.012	-1.333	0.183
NETRIN1 forceps_minor	-0.018	0.012	-1.537	0.124
NETRIN1 middle_cerebellar_peduncle	-0.016	0.012	-1.294	0.196
Genomic acoustic_radiation	-0.013	0.011	-1.230	0.219
Genomic anterior_thalamic_radiation	-0.016	0.011	-1.386	0.166
Genomic cingulate_gyrus_part_of_cingulum	-0.021	0.011	-1.943	0.052
Genomic parahippocampal_part_of_cingulum	-0.022	0.011	-2.022	0.043
Genomic corticospinal_tract	-0.018	0.011	-1.604	0.109
Genomic inferior_fronto_occipital_fasciculus	-0.028	0.012	-2.444	0.015
Genomic inferior_longitudinal_fasciculus	-0.025	0.012	-2.135	0.033
Genomic medial_lemniscus	-0.004	0.010	-0.401	0.689
Genomic posterior_thalamic_radiation	-0.022	0.011	-1.923	0.054
Genomic superior_longitudinal_fasciculus	-0.022	0.012	-1.927	0.054
Genomic superior_thalamic_radiation	-0.014	0.012	-1.202	0.229
Genomic uncinate_fasciculus	-0.032	0.011	-2.957	0.003
Genomic forceps_major	-0.031	0.012	-2.589	0.010
Genomic forceps_minor	-0.031	0.012	-2.573	0.010
Genomic middle_cerebellar_peduncle	-0.020	0.012	-1.585	0.113

Table S6. The effect of unpruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on individual white matter tracts (FA) (N = 6,401).

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.025	0.012	-2.065	0.039
NETRIN1 Association fibres	-0.024	0.012	-2.024	0.043
NETRIN1 Thalamic radiations	-0.020	0.012	-1.615	0.106
NETRIN1 Projection fibres	-0.024	0.012	-1.963	0.050
Genomic gFA	-0.029	0.012	-2.431	0.015
Genomic Association fibres	-0.031	0.012	-2.574	0.010
Genomic Thalamic radiations	-0.020	0.012	-1.685	0.092
Genomic Projection fibres	-0.021	0.012	-1.716	0.086
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.012	0.012	-1.030	0.303
NETRIN1 Association fibres	-0.016	0.012	-1.333	0.183
NETRIN1 Thalamic radiations	-0.007	0.012	-0.590	0.555
NETRIN1 Projection fibres	-0.007	0.012	-0.552	0.581
Genomic gFA	-0.033	0.012	-2.776	0.006

Genomic Association fibres	-0.034	0.012	-2.845	0.004
Genomic Thalamic radiations	-0.026	0.012	-2.128	0.033
Genomic Projection fibres	-0.025	0.012	-2.073	0.038
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.018	0.012	-1.494	0.135
NETRIN1 Association fibres	-0.020	0.012	-1.684	0.092
NETRIN1 Thalamic radiations	-0.014	0.012	-1.125	0.261
NETRIN1 Projection fibres	-0.012	0.012	-1.032	0.302
Genomic gFA	-0.032	0.012	-2.656	0.008
Genomic Association fibres	-0.032	0.012	-2.728	0.006
Genomic Thalamic radiations	-0.022	0.012	-1.820	0.069
Genomic Projection fibres	-0.026	0.012	-2.201	0.028
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.027	0.012	-2.288	0.022
NETRIN1 Association fibres	-0.034	0.012	-2.903	0.004
NETRIN1 Thalamic radiations	-0.019	0.012	-1.590	0.112
NETRIN1 Projection fibres	-0.011	0.012	-0.881	0.379
Genomic gFA	-0.034	0.012	-2.824	0.005
Genomic Association fibres	-0.035	0.012	-2.927	0.003
Genomic Thalamic radiations	-0.023	0.012	-1.863	0.062
Genomic Projection fibres	-0.029	0.012	-2.443	0.015

Table S7. The effect of unpruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on tract categories (FA) (N = 6,401).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.002	0.011	0.222	0.824
NETRIN1 anterior_thalamic_radiation	-0.021	0.012	-1.800	0.072
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.024	0.011	-2.199	0.028
NETRIN1 parahippocampal_part_of_cingulum	-0.008	0.011	-0.731	0.465
NETRIN1 corticospinal_tract	0.001	0.011	0.125	0.900
NETRIN1 inferior_fronto_occipital_fasciculus	-0.022	0.012	-1.899	0.058
NETRIN1 inferior_longitudinal_fasciculus	-0.021	0.012	-1.853	0.064
NETRIN1 medial_lemniscus	-0.009	0.010	-0.826	0.409
NETRIN1 posterior_thalamic_radiation	-0.013	0.011	-1.162	0.245
NETRIN1 superior_longitudinal_fasciculus	-0.034	0.012	-2.897	0.004
NETRIN1 superior_thalamic_radiation	-0.006	0.012	-0.466	0.641
NETRIN1 uncinate_fasciculus	-0.019	0.011	-1.698	0.090
NETRIN1 forceps_major	-0.014	0.012	-1.197	0.231
NETRIN1 forceps_minor	-0.018	0.012	-1.489	0.136
NETRIN1 middle_cerebellar_peduncle	-0.016	0.012	-1.270	0.204
Genomic acoustic_radiation	-0.016	0.011	-1.464	0.143
Genomic anterior_thalamic_radiation	-0.018	0.012	-1.530	0.126
Genomic cingulate_gyrus_part_of_cingulum	-0.020	0.011	-1.859	0.063
Genomic parahippocampal_part_of_cingulum	-0.022	0.011	-2.042	0.041
Genomic corticospinal_tract	-0.022	0.012	-1.878	0.060
Genomic inferior_fronto_occipital_fasciculus	-0.030	0.012	-2.579	0.010
Genomic inferior_longitudinal_fasciculus	-0.026	0.012	-2.258	0.024
Genomic medial_lemniscus	-0.006	0.011	-0.580	0.562
Genomic posterior_thalamic_radiation	-0.025	0.011	-2.224	0.026
Genomic superior_longitudinal_fasciculus	-0.025	0.012	-2.095	0.036
Genomic superior_thalamic_radiation	-0.018	0.012	-1.487	0.137
Genomic uncinate_fasciculus	-0.034	0.011	-3.111	0.002

Genomic forceps_major	-0.034	0.012	-2.781	0.005
Genomic forceps_minor	-0.033	0.012	-2.717	0.007
Genomic middle_cerebellar_peduncle	-0.023	0.012	-1.828	0.068

Table S8. The effect of unpruned NETRIN1- and Genomic-PRS at threshold 0.5 on individual white matter tracts (FA) (N = 6,420).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.002	0.001	-2.197	0.028
NETRIN1 Association fibres	-0.002	0.001	-2.762	0.006
NETRIN1 Thalamic radiations	-0.001	0.000	-1.482	0.138
NETRIN1 Projection fibres	0.000	0.001	-0.904	0.366
Genomic gFA	-0.002	0.001	-2.769	0.006
Genomic Association fibres	-0.002	0.001	-2.836	0.005
Genomic Thalamic radiations	-0.001	0.000	-1.855	0.064
Genomic Projection fibres	-0.001	0.001	-2.415	0.016

Table S9. The effect of unpruned NETRIN1- and Genomic-PRS at threshold 0.5 on tract categories (FA) (N = 6,420).

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.008	0.011	0.772	0.440
NETRIN1 anterior_thalamic_radiation	0.018	0.011	1.694	0.090
NETRIN1 cingulate_gyrus_part_of_cingulum	0.013	0.011	1.257	0.209
NETRIN1 parahippocampal_part_of_cingulum	-0.007	0.011	-0.621	0.535
NETRIN1 corticospinal_tract	0.003	0.011	0.270	0.787
NETRIN1 inferior_fronto_occipital_fasciculus	0.021	0.011	1.905	0.057
NETRIN1 inferior_longitudinal_fasciculus	0.019	0.011	1.727	0.084
NETRIN1 medial_lemniscus	0.007	0.011	0.659	0.510
NETRIN1 posterior_thalamic_radiation	0.016	0.011	1.466	0.143
NETRIN1 superior_longitudinal_fasciculus	0.023	0.011	2.046	0.041
NETRIN1 superior_thalamic_radiation	0.016	0.010	1.589	0.112
NETRIN1 uncinate_fasciculus	0.011	0.010	1.033	0.302
NETRIN1 bl.MD.wm.forceps_major	0.013	0.012	1.083	0.279
NETRIN1 bl.MD.wm.forceps_minor	0.022	0.012	1.946	0.052
NETRIN1 bl.MD.wm.middle_cerebellar_peduncle	0.003	0.012	0.239	0.811
Genomic acoustic_radiation	0.015	0.011	1.453	0.146
Genomic anterior_thalamic_radiation	0.020	0.011	1.878	0.060
Genomic cingulate_gyrus_part_of_cingulum	0.038	0.011	3.529	0.000
Genomic parahippocampal_part_of_cingulum	0.030	0.011	2.846	0.004
Genomic corticospinal_tract	0.030	0.011	2.654	0.008
Genomic inferior_fronto_occipital_fasciculus	0.032	0.011	2.879	0.004
Genomic inferior_longitudinal_fasciculus	0.029	0.011	2.618	0.009
Genomic medial_lemniscus	0.012	0.011	1.145	0.252
Genomic posterior_thalamic_radiation	0.016	0.011	1.493	0.135
Genomic superior_longitudinal_fasciculus	0.028	0.011	2.490	0.013
Genomic superior_thalamic_radiation	0.023	0.010	2.320	0.020
Genomic uncinate_fasciculus	0.033	0.010	3.148	0.002
Genomic bl.MD.wm.forceps_major	0.033	0.012	2.733	0.006
Genomic bl.MD.wm.forceps_minor	0.020	0.012	1.692	0.091
Genomic bl.MD.wm.middle_cerebellar_peduncle	0.004	0.012	0.362	0.718
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.006	0.010	-0.561	0.575

NETRIN1 anterior_thalamic_radiation	0.015	0.011	1.426	0.154
NETRIN1 cingulate_gyrus_part_of_cingulum	0.014	0.011	1.354	0.176
NETRIN1 parahippocampal_part_of_cingulum	0.004	0.011	0.347	0.729
NETRIN1 corticospinal_tract	-0.003	0.011	-0.226	0.821
NETRIN1 inferior_fronto_occipital_fasciculus	0.015	0.011	1.303	0.193
NETRIN1 inferior_longitudinal_fasciculus	0.017	0.011	1.538	0.124
NETRIN1 medial_lemniscus	0.002	0.011	0.160	0.873
NETRIN1 posterior_thalamic_radiation	0.016	0.011	1.509	0.131
NETRIN1 superior_longitudinal_fasciculus	0.023	0.011	1.998	0.046
NETRIN1 superior_thalamic_radiation	0.014	0.010	1.420	0.156
NETRIN1 uncinate_fasciculus	0.008	0.010	0.752	0.452
NETRIN1 bl.MD.wm.forceps_major	0.014	0.012	1.172	0.241
NETRIN1 bl.MD.wm.forceps_minor	0.015	0.012	1.292	0.196
NETRIN1 bl.MD.wm.middle_cerebellar_peduncle	-0.002	0.012	-0.138	0.890
Genomic acoustic_radiation	0.021	0.011	1.959	0.050
Genomic anterior_thalamic_radiation	0.025	0.011	2.359	0.018
Genomic cingulate_gyrus_part_of_cingulum	0.040	0.011	3.734	0.000
Genomic parahippocampal_part_of_cingulum	0.033	0.011	3.108	0.002
Genomic corticospinal_tract	0.034	0.011	2.999	0.003
Genomic inferior_fronto_occipital_fasciculus	0.037	0.011	3.327	0.001
Genomic inferior_longitudinal_fasciculus	0.032	0.011	2.890	0.004
Genomic medial_lemniscus	0.012	0.011	1.091	0.275
Genomic posterior_thalamic_radiation	0.016	0.011	1.527	0.127
Genomic superior_longitudinal_fasciculus	0.032	0.011	2.819	0.005
Genomic superior_thalamic_radiation	0.028	0.010	2.812	0.005
Genomic uncinate_fasciculus	0.032	0.010	3.116	0.002
Genomic bl.MD.wm.forceps_major	0.032	0.012	2.663	0.008
Genomic bl.MD.wm.forceps_minor	0.024	0.012	2.103	0.036
Genomic bl.MD.wm.middle_cerebellar_peduncle	0.006	0.012	0.515	0.607
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.005	0.010	-0.458	0.647
NETRIN1 anterior_thalamic_radiation	0.020	0.011	1.868	0.062
NETRIN1 cingulate_gyrus_part_of_cingulum	0.014	0.011	1.334	0.182
NETRIN1 parahippocampal_part_of_cingulum	-0.007	0.011	-0.710	0.478
NETRIN1 corticospinal_tract	0.002	0.011	0.204	0.838
NETRIN1 inferior_fronto_occipital_fasciculus	0.020	0.011	1.800	0.072
NETRIN1 inferior_longitudinal_fasciculus	0.020	0.011	1.832	0.067
NETRIN1 medial_lemniscus	0.011	0.011	1.018	0.309
NETRIN1 posterior_thalamic_radiation	0.018	0.011	1.638	0.102
NETRIN1 superior_longitudinal_fasciculus	0.030	0.011	2.611	0.009
NETRIN1 superior_thalamic_radiation	0.021	0.010	2.073	0.038
NETRIN1 uncinate_fasciculus	0.009	0.010	0.879	0.379
NETRIN1 forceps_major	0.017	0.012	1.407	0.159
NETRIN1 forceps_minor	0.018	0.012	1.597	0.110
NETRIN1 middle_cerebellar_peduncle	0.004	0.012	0.298	0.766
Genomic acoustic_radiation	0.022	0.011	2.107	0.035
Genomic anterior_thalamic_radiation	0.023	0.011	2.143	0.032
Genomic cingulate_gyrus_part_of_cingulum	0.038	0.011	3.601	0.000
Genomic parahippocampal_part_of_cingulum	0.033	0.011	3.098	0.002
Genomic corticospinal_tract	0.032	0.011	2.802	0.005
Genomic inferior_fronto_occipital_fasciculus	0.034	0.011	3.081	0.002
Genomic inferior_longitudinal_fasciculus	0.030	0.011	2.689	0.007
Genomic medial_lemniscus	0.005	0.011	0.489	0.625
Genomic posterior_thalamic_radiation	0.009	0.011	0.884	0.377

Genomic superior_longitudinal_fasciculus	0.030	0.011	2.617	0.009
Genomic superior_thalamic_radiation	0.024	0.010	2.442	0.015
Genomic uncinate_fasciculus	0.034	0.010	3.320	0.001
Genomic forceps_major	0.028	0.012	2.358	0.018
Genomic forceps_minor	0.021	0.012	1.783	0.075
Genomic middle_cerebellar_peduncle	0.008	0.012	0.666	0.505
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.004	0.010	0.347	0.729
NETRIN1 anterior_thalamic_radiation	0.028	0.011	2.669	0.008
NETRIN1 cingulate_gyrus_part_of_cingulum	0.022	0.011	2.023	0.043
NETRIN1 parahippocampal_part_of_cingulum	0.000	0.011	-0.023	0.981
NETRIN1 corticospinal_tract	0.017	0.011	1.525	0.127
NETRIN1 inferior_fronto_occipital_fasciculus	0.028	0.011	2.551	0.011
NETRIN1 inferior_longitudinal_fasciculus	0.029	0.011	2.553	0.011
NETRIN1 medial_lemniscus	0.005	0.011	0.428	0.669
NETRIN1 posterior_thalamic_radiation	0.026	0.011	2.453	0.014
NETRIN1 superior_longitudinal_fasciculus	0.033	0.011	2.953	0.003
NETRIN1 superior_thalamic_radiation	0.027	0.010	2.763	0.006
NETRIN1 uncinate_fasciculus	0.020	0.010	1.900	0.058
NETRIN1 forceps_major	0.018	0.012	1.519	0.129
NETRIN1 forceps_minor	0.021	0.012	1.791	0.073
NETRIN1 middle_cerebellar_peduncle	0.011	0.012	0.890	0.373
Genomic acoustic_radiation	0.019	0.011	1.841	0.066
Genomic anterior_thalamic_radiation	0.021	0.011	2.021	0.043
Genomic cingulate_gyrus_part_of_cingulum	0.036	0.011	3.332	0.001
Genomic parahippocampal_part_of_cingulum	0.034	0.011	3.223	0.001
Genomic corticospinal_tract	0.023	0.011	1.997	0.046
Genomic inferior_fronto_occipital_fasciculus	0.032	0.011	2.828	0.005
Genomic inferior_longitudinal_fasciculus	0.025	0.011	2.262	0.024
Genomic medial_lemniscus	0.005	0.011	0.470	0.639
Genomic posterior_thalamic_radiation	0.002	0.011	0.142	0.887
Genomic superior_longitudinal_fasciculus	0.024	0.011	2.156	0.031
Genomic superior_thalamic_radiation	0.018	0.010	1.804	0.071
Genomic uncinate_fasciculus	0.030	0.010	2.844	0.004
Genomic forceps_major	0.029	0.012	2.447	0.014
Genomic forceps_minor	0.021	0.012	1.858	0.063
Genomic middle_cerebellar_peduncle	0.012	0.012	0.965	0.335

Table S10. The effect of unpruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on individual white matter tracts (MD) (N = 6,390).

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.018	0.012	1.574	0.116
NETRIN1 Association fibres	0.013	0.012	1.086	0.277
NETRIN1 Thalamic radiations	0.019	0.011	1.781	0.075
NETRIN1 Projection fibres	0.013	0.012	1.087	0.277
Genomic gMD	0.037	0.012	3.248	0.001
Genomic Association fibres	0.043	0.012	3.707	0.000
Genomic Thalamic radiations	0.022	0.011	2.027	0.043
Genomic Projection fibres	0.026	0.012	2.180	0.029
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.016	0.011	1.380	0.168
NETRIN1 Association fibres	0.015	0.012	1.320	0.187

NETRIN1 Thalamic radiations	0.018	0.011	1.669	0.095
NETRIN1 Projection fibres	0.004	0.012	0.322	0.748
Genomic gMD	0.041	0.011	3.607	0.000
Genomic Association fibres	0.047	0.012	4.033	0.000
Genomic Thalamic radiations	0.025	0.011	2.334	0.020
Genomic Projection fibres	0.030	0.012	2.478	0.013
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.018	0.011	1.596	0.111
NETRIN1 Association fibres	0.013	0.012	1.106	0.269
NETRIN1 Thalamic radiations	0.022	0.011	2.055	0.040
NETRIN1 Projection fibres	0.011	0.012	0.920	0.358
Genomic gMD	0.038	0.011	3.342	0.001
Genomic Association fibres	0.046	0.012	3.934	0.000
Genomic Thalamic radiations	0.020	0.011	1.822	0.069
Genomic Projection fibres	0.029	0.012	2.391	0.017
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.029	0.011	2.524	0.012
NETRIN1 Association fibres	0.023	0.012	2.014	0.044
NETRIN1 Thalamic radiations	0.031	0.011	2.944	0.003
NETRIN1 Projection fibres	0.020	0.012	1.686	0.092
Genomic gMD	0.034	0.011	2.974	0.003
Genomic Association fibres	0.043	0.012	3.666	0.000
Genomic Thalamic radiations	0.013	0.011	1.229	0.219
Genomic Projection fibres	0.030	0.012	2.494	0.013

Table S11. The effect of unpruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on tract categories (MD) (N = 6,390).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.005	0.011	0.484	0.628
NETRIN1 anterior_thalamic_radiation	0.023	0.011	2.171	0.030
NETRIN1 cingulate_gyrus_part_of_cingulum	0.019	0.011	1.682	0.093
NETRIN1 parahippocampal_part_of_cingulum	0.000	0.011	-0.004	0.997
NETRIN1 corticospinal_tract	0.014	0.012	1.232	0.218
NETRIN1 inferior_fronto_occipital_fasciculus	0.025	0.011	2.242	0.025
NETRIN1 inferior_longitudinal_fasciculus	0.027	0.011	2.377	0.017
NETRIN1 medial_lemniscus	0.003	0.011	0.288	0.774
NETRIN1 posterior_thalamic_radiation	0.024	0.011	2.213	0.027
NETRIN1 superior_longitudinal_fasciculus	0.030	0.011	2.649	0.008
NETRIN1 superior_thalamic_radiation	0.024	0.010	2.345	0.019
NETRIN1 uncinate_fasciculus	0.017	0.011	1.559	0.119
NETRIN1 forceps_major	0.019	0.012	1.599	0.110
NETRIN1 forceps_minor	0.019	0.012	1.592	0.111
NETRIN1 middle_cerebellar_peduncle	0.012	0.012	0.984	0.325
Genomic acoustic_radiation	0.010	0.011	0.949	0.342
Genomic anterior_thalamic_radiation	0.011	0.011	1.009	0.313
Genomic cingulate_gyrus_part_of_cingulum	0.021	0.011	1.852	0.064
Genomic parahippocampal_part_of_cingulum	0.027	0.011	2.485	0.013
Genomic corticospinal_tract	0.009	0.012	0.800	0.424
Genomic inferior_fronto_occipital_fasciculus	0.019	0.011	1.630	0.103
Genomic inferior_longitudinal_fasciculus	0.013	0.011	1.138	0.255
Genomic medial_lemniscus	-0.004	0.011	-0.378	0.705
Genomic posterior_thalamic_radiation	-0.006	0.011	-0.505	0.613

Genomic superior_longitudinal_fasciculus	0.013	0.012	1.113	0.266
Genomic superior_thalamic_radiation	0.007	0.011	0.653	0.514
Genomic uncinate_fasciculus	0.017	0.011	1.588	0.112
Genomic forceps_major	0.020	0.012	1.624	0.104
Genomic forceps_minor	0.012	0.012	0.982	0.326
Genomic middle_cerebellar_peduncle	0.005	0.012	0.437	0.662

Table S12. The effect of unpruned NETRIN1- and Genomic-PRS at threshold 0.5 on tract categories (MD) (N = 6,420).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 gMD	3.4E-06	1.4E-06	2.4E+00	1.6E-02
NETRIN1 Association fibres	2.0E-06	1.1E-06	1.9E+00	5.8E-02
NETRIN1 Thalamic radiations	2.2E-06	7.9E-07	2.8E+00	5.4E-03
NETRIN1 Projection fibres	1.4E-06	8.0E-07	1.8E+00	7.7E-02
Genomic gMD	4.2E-06	1.4E-06	2.9E+00	3.5E-03
Genomic Association fibres	3.9E-06	1.1E-06	3.6E+00	3.3E-04
Genomic Thalamic radiations	9.8E-07	7.9E-07	1.2E+00	2.2E-01
Genomic Projection fibres	1.9E-06	8.1E-07	2.4E+00	1.7E-02

Table S13. The effect of unpruned NETRIN1- and Genomic-PRS at threshold 0.5 on tract categories (MD) (N = 6,320).

Pruned NETRIN1- and Genomic-PRS with outliers included (6,420) and outliers excluded (6,401 for FA and 6,390 for MD) at all 5 thresholds (0.01, 0.05, 0.1, 0.5, 1)

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.004	0.011	-0.349	0.727
NETRIN1 anterior_thalamic_radiation	-0.020	0.011	-1.709	0.087
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.008	0.011	-0.704	0.482
NETRIN1 parahippocampal_part_of_cingulum	-0.007	0.011	-0.641	0.521
NETRIN1 corticospinal_tract	-0.022	0.011	-1.923	0.055
NETRIN1 inferior_fronto_occipital_fasciculus	-0.024	0.012	-2.058	0.040
NETRIN1 inferior_longitudinal_fasciculus	-0.024	0.012	-2.047	0.041
NETRIN1 medial_lemniscus	-0.012	0.010	-1.131	0.258
NETRIN1 posterior_thalamic_radiation	-0.014	0.011	-1.238	0.216
NETRIN1 superior_longitudinal_fasciculus	-0.028	0.012	-2.408	0.016
NETRIN1 superior_thalamic_radiation	-0.020	0.012	-1.677	0.094
NETRIN1 uncinate_fasciculus	-0.023	0.011	-2.156	0.031
NETRIN1 bl.FA.wm.forceps_major	-0.014	0.012	-1.137	0.255
NETRIN1 bl.FA.wm.forceps_minor	-0.009	0.012	-0.784	0.433
NETRIN1 bl.FA.wm.middle_cerebellar_peduncle	-0.028	0.012	-2.257	0.024
Genomic acoustic_radiation	-0.010	0.011	-0.942	0.346
Genomic anterior_thalamic_radiation	-0.017	0.011	-1.440	0.150
Genomic cingulate_gyrus_part_of_cingulum	-0.008	0.011	-0.785	0.432
Genomic parahippocampal_part_of_cingulum	0.009	0.011	0.810	0.418
Genomic corticospinal_tract	-0.010	0.011	-0.889	0.374
Genomic inferior_fronto_occipital_fasciculus	-0.014	0.012	-1.249	0.212
Genomic inferior_longitudinal_fasciculus	-0.012	0.012	-1.023	0.306
Genomic medial_lemniscus	0.000	0.010	0.010	0.992

Genomic posterior_thalamic_radiation	-0.011	0.011	-0.965	0.335
Genomic superior_longitudinal_fasciculus	-0.013	0.012	-1.080	0.280
Genomic superior_thalamic_radiation	-0.016	0.012	-1.332	0.183
Genomic uncinate_fasciculus	-0.019	0.011	-1.793	0.073
Genomic bl.FA.wm.forceps_major	-0.013	0.012	-1.086	0.278
Genomic bl.FA.wm.forceps_minor	-0.018	0.012	-1.475	0.140
Genomic bl.FA.wm.middle_cerebellar_peduncle	0.017	0.012	1.369	0.171
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.008	0.011	0.770	0.441
NETRIN1 anterior_thalamic_radiation	-0.012	0.011	-1.047	0.295
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.007	0.011	-0.627	0.531
NETRIN1 parahippocampal_part_of_cingulum	-0.014	0.011	-1.335	0.182
NETRIN1 corticospinal_tract	-0.002	0.011	-0.146	0.884
NETRIN1 inferior_fronto_occipital_fasciculus	-0.007	0.011	-0.590	0.555
NETRIN1 inferior_longitudinal_fasciculus	-0.010	0.011	-0.865	0.387
NETRIN1 medial_lemniscus	-0.006	0.010	-0.574	0.566
NETRIN1 posterior_thalamic_radiation	-0.003	0.011	-0.304	0.761
NETRIN1 superior_longitudinal_fasciculus	-0.015	0.012	-1.290	0.197
NETRIN1 superior_thalamic_radiation	-0.003	0.012	-0.275	0.783
NETRIN1 uncinate_fasciculus	-0.011	0.011	-1.030	0.303
NETRIN1 bl.FA.wm.forceps_major	-0.004	0.012	-0.292	0.770
NETRIN1 bl.FA.wm.forceps_minor	-0.002	0.012	-0.178	0.858
NETRIN1 bl.FA.wm.middle_cerebellar_peduncle	-0.015	0.012	-1.200	0.230
Genomic acoustic_radiation	-0.005	0.011	-0.462	0.644
Genomic anterior_thalamic_radiation	-0.010	0.011	-0.901	0.367
Genomic cingulate_gyrus_part_of_cingulum	-0.004	0.011	-0.350	0.726
Genomic parahippocampal_part_of_cingulum	0.001	0.011	0.103	0.918
Genomic corticospinal_tract	-0.014	0.011	-1.272	0.203
Genomic inferior_fronto_occipital_fasciculus	-0.016	0.011	-1.351	0.177
Genomic inferior_longitudinal_fasciculus	-0.015	0.011	-1.281	0.200
Genomic medial_lemniscus	-0.006	0.010	-0.569	0.569
Genomic posterior_thalamic_radiation	-0.019	0.011	-1.716	0.086
Genomic superior_longitudinal_fasciculus	-0.012	0.012	-1.076	0.282
Genomic superior_thalamic_radiation	-0.019	0.012	-1.596	0.110
Genomic uncinate_fasciculus	-0.017	0.011	-1.557	0.119
Genomic bl.FA.wm.forceps_major	-0.013	0.012	-1.093	0.275
Genomic bl.FA.wm.forceps_minor	-0.014	0.012	-1.186	0.236
Genomic bl.FA.wm.middle_cerebellar_peduncle	-0.003	0.012	-0.271	0.786
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.005	0.011	0.452	0.652
NETRIN1 anterior_thalamic_radiation	-0.017	0.011	-1.442	0.149
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.013	0.011	-1.238	0.216
NETRIN1 parahippocampal_part_of_cingulum	-0.007	0.011	-0.681	0.496
NETRIN1 corticospinal_tract	-0.003	0.011	-0.225	0.822
NETRIN1 inferior_fronto_occipital_fasciculus	-0.016	0.012	-1.381	0.167
NETRIN1 inferior_longitudinal_fasciculus	-0.014	0.012	-1.221	0.222
NETRIN1 medial_lemniscus	-0.018	0.010	-1.730	0.084
NETRIN1 posterior_thalamic_radiation	-0.007	0.011	-0.601	0.548
NETRIN1 superior_longitudinal_fasciculus	-0.026	0.012	-2.205	0.027
NETRIN1 superior_thalamic_radiation	-0.010	0.012	-0.871	0.384
NETRIN1 uncinate_fasciculus	-0.010	0.011	-0.896	0.370
NETRIN1 forceps_major	-0.012	0.012	-1.004	0.316
NETRIN1 forceps_minor	-0.013	0.012	-1.041	0.298

NETRIN1 middle_cerebellar_peduncle	-0.011	0.012	-0.922	0.356
Genomic acoustic_radiation	0.001	0.011	0.080	0.936
Genomic anterior_thalamic_radiation	0.002	0.011	0.146	0.884
Genomic cingulate_gyrus_part_of_cingulum	-0.001	0.011	-0.075	0.940
Genomic parahippocampal_part_of_cingulum	-0.002	0.011	-0.188	0.851
Genomic corticospinal_tract	-0.015	0.011	-1.345	0.179
Genomic inferior_fronto_occipital_fasciculus	-0.008	0.011	-0.723	0.469
Genomic inferior_longitudinal_fasciculus	-0.009	0.012	-0.760	0.447
Genomic medial_lemniscus	-0.001	0.010	-0.131	0.896
Genomic posterior_thalamic_radiation	-0.009	0.011	-0.770	0.441
Genomic superior_longitudinal_fasciculus	-0.007	0.012	-0.580	0.562
Genomic superior_thalamic_radiation	-0.011	0.012	-0.951	0.342
Genomic uncinate_fasciculus	-0.017	0.011	-1.572	0.116
Genomic forceps_major	-0.008	0.012	-0.632	0.528
Genomic forceps_minor	-0.006	0.012	-0.518	0.605
Genomic middle_cerebellar_peduncle	-0.016	0.012	-1.281	0.200
PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.006	0.011	0.520	0.603
NETRIN1 anterior_thalamic_radiation	-0.021	0.011	-1.811	0.070
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.023	0.011	-2.201	0.028
NETRIN1 parahippocampal_part_of_cingulum	-0.006	0.011	-0.583	0.560
NETRIN1 corticospinal_tract	0.002	0.011	0.204	0.839
NETRIN1 inferior_fronto_occipital_fasciculus	-0.021	0.011	-1.824	0.068
NETRIN1 inferior_longitudinal_fasciculus	-0.021	0.012	-1.790	0.074
NETRIN1 medial_lemniscus	-0.011	0.010	-1.061	0.289
NETRIN1 posterior_thalamic_radiation	-0.011	0.011	-0.981	0.327
NETRIN1 superior_longitudinal_fasciculus	-0.035	0.012	-3.031	0.002
NETRIN1 superior_thalamic_radiation	-0.006	0.012	-0.521	0.603
NETRIN1 uncinate_fasciculus	-0.018	0.011	-1.702	0.089
NETRIN1 forceps_major	-0.009	0.012	-0.740	0.459
NETRIN1 forceps_minor	-0.013	0.012	-1.071	0.284
NETRIN1 middle_cerebellar_peduncle	-0.017	0.012	-1.363	0.173
Genomic acoustic_radiation	-0.005	0.011	-0.488	0.625
Genomic anterior_thalamic_radiation	-0.007	0.011	-0.607	0.544
Genomic cingulate_gyrus_part_of_cingulum	-0.008	0.011	-0.780	0.435
Genomic parahippocampal_part_of_cingulum	-0.013	0.011	-1.189	0.235
Genomic corticospinal_tract	-0.022	0.011	-1.926	0.054
Genomic inferior_fronto_occipital_fasciculus	-0.018	0.011	-1.581	0.114
Genomic inferior_longitudinal_fasciculus	-0.014	0.012	-1.242	0.214
Genomic medial_lemniscus	-0.011	0.010	-1.055	0.291
Genomic posterior_thalamic_radiation	-0.015	0.011	-1.346	0.178
Genomic superior_longitudinal_fasciculus	-0.012	0.012	-1.014	0.311
Genomic superior_thalamic_radiation	-0.016	0.012	-1.381	0.167
Genomic uncinate_fasciculus	-0.023	0.011	-2.172	0.030
Genomic forceps_major	-0.015	0.012	-1.270	0.204
Genomic forceps_minor	-0.014	0.012	-1.184	0.237
Genomic middle_cerebellar_peduncle	-0.016	0.012	-1.334	0.182
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.006	0.011	0.554	0.579
NETRIN1 anterior_thalamic_radiation	-0.022	0.011	-1.896	0.058
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.026	0.011	-2.428	0.015
NETRIN1 parahippocampal_part_of_cingulum	-0.006	0.011	-0.558	0.577
NETRIN1 corticospinal_tract	-0.001	0.011	-0.057	0.954

NETRIN1 inferior_fronto_occipital_fasciculus	-0.020	0.011	-1.765	0.078
NETRIN1 inferior_longitudinal_fasciculus	-0.019	0.011	-1.629	0.103
NETRIN1 medial_lemniscus	-0.011	0.010	-1.020	0.308
NETRIN1 posterior_thalamic_radiation	-0.011	0.011	-0.991	0.322
NETRIN1 superior_longitudinal_fasciculus	-0.034	0.012	-2.959	0.003
NETRIN1 superior_thalamic_radiation	-0.007	0.012	-0.582	0.560
NETRIN1 uncinate_fasciculus	-0.018	0.011	-1.635	0.102
NETRIN1 forceps_major	-0.008	0.012	-0.678	0.497
NETRIN1 forceps_minor	-0.013	0.012	-1.116	0.264
NETRIN1 middle_cerebellar_peduncle	-0.015	0.012	-1.195	0.232
Genomic acoustic_radiation	-0.005	0.011	-0.502	0.616
Genomic anterior_thalamic_radiation	-0.009	0.011	-0.755	0.450
Genomic cingulate_gyrus_part_of_cingulum	-0.010	0.011	-0.976	0.329
Genomic parahippocampal_part_of_cingulum	-0.015	0.011	-1.373	0.170
Genomic corticospinal_tract	-0.021	0.011	-1.826	0.068
Genomic inferior_fronto_occipital_fasciculus	-0.018	0.011	-1.588	0.112
Genomic inferior_longitudinal_fasciculus	-0.014	0.011	-1.231	0.218
Genomic medial_lemniscus	-0.011	0.010	-1.044	0.296
Genomic posterior_thalamic_radiation	-0.013	0.011	-1.151	0.250
Genomic superior_longitudinal_fasciculus	-0.010	0.012	-0.848	0.396
Genomic superior_thalamic_radiation	-0.015	0.012	-1.263	0.207
Genomic uncinate_fasciculus	-0.024	0.011	-2.174	0.030
Genomic forceps_major	-0.014	0.012	-1.167	0.243
Genomic forceps_minor	-0.012	0.012	-1.017	0.309
Genomic middle_cerebellar_peduncle	-0.017	0.012	-1.344	0.179

Table S14. The effect of pruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on individual white matter tracts (FA) (N = 6,401).

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.026	0.012	-2.186	0.029
NETRIN1 Association fibres	-0.025	0.012	-2.066	0.039
NETRIN1 Thalamic radiations	-0.022	0.012	-1.853	0.064
NETRIN1 Projection fibres	-0.025	0.012	-2.098	0.036
Genomic gFA	-0.015	0.012	-1.226	0.220
Genomic Association fibres	-0.013	0.012	-1.068	0.285
Genomic Thalamic radiations	-0.018	0.012	-1.488	0.137
Genomic Projection fibres	-0.009	0.012	-0.766	0.444
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.011	0.012	-0.943	0.346
NETRIN1 Association fibres	-0.015	0.012	-1.245	0.213
NETRIN1 Thalamic radiations	-0.008	0.012	-0.635	0.526
NETRIN1 Projection fibres	-0.004	0.012	-0.367	0.714
Genomic gFA	-0.017	0.012	-1.385	0.166
Genomic Association fibres	-0.013	0.012	-1.074	0.283
Genomic Thalamic radiations	-0.021	0.012	-1.740	0.082
Genomic Projection fibres	-0.015	0.012	-1.283	0.200
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.018	0.012	-1.518	0.129
NETRIN1 Association fibres	-0.020	0.012	-1.720	0.085
NETRIN1 Thalamic radiations	-0.014	0.012	-1.147	0.251
NETRIN1 Projection fibres	-0.012	0.012	-0.981	0.327
Genomic gFA	-0.010	0.012	-0.855	0.393

Genomic Association fibres	-0.008	0.012	-0.714	0.476
Genomic Thalamic radiations	-0.008	0.012	-0.666	0.505
Genomic Projection fibres	-0.013	0.012	-1.105	0.269
PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.023	0.012	-1.966	0.049
NETRIN1 Association fibres	-0.031	0.012	-2.567	0.010
NETRIN1 Thalamic radiations	-0.016	0.012	-1.327	0.184
NETRIN1 Projection fibres	-0.008	0.012	-0.668	0.504
Genomic gFA	-0.021	0.012	-1.794	0.073
Genomic Association fibres	-0.020	0.012	-1.656	0.098
Genomic Thalamic radiations	-0.017	0.012	-1.376	0.169
Genomic Projection fibres	-0.024	0.012	-1.983	0.047
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.024	0.012	-1.991	0.047
NETRIN1 Association fibres	-0.031	0.012	-2.585	0.010
NETRIN1 Thalamic radiations	-0.017	0.012	-1.387	0.166
NETRIN1 Projection fibres	-0.009	0.012	-0.715	0.475
Genomic gFA	-0.021	0.012	-1.793	0.073
Genomic Association fibres	-0.021	0.012	-1.741	0.082
Genomic Thalamic radiations	-0.016	0.012	-1.296	0.195
Genomic Projection fibres	-0.023	0.012	-1.899	0.058

Table S15. The effect of pruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on tract categories (FA) (N = 6,401).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.002	0.011	0.198	0.843
NETRIN1 anterior_thalamic_radiation	-0.022	0.012	-1.922	0.055
NETRIN1 cingulate_gyrus_part_of_cingulum	-0.024	0.011	-2.234	0.025
NETRIN1 parahippocampal_part_of_cingulum	-0.007	0.011	-0.644	0.520
NETRIN1 corticospinal_tract	-0.001	0.011	-0.100	0.920
NETRIN1 inferior_fronto_occipital_fasciculus	-0.023	0.012	-1.957	0.050
NETRIN1 inferior_longitudinal_fasciculus	-0.022	0.012	-1.865	0.062
NETRIN1 medial_lemniscus	-0.013	0.011	-1.240	0.215
NETRIN1 posterior_thalamic_radiation	-0.013	0.011	-1.120	0.263
NETRIN1 superior_longitudinal_fasciculus	-0.035	0.012	-3.029	0.002
NETRIN1 superior_thalamic_radiation	-0.008	0.012	-0.693	0.488
NETRIN1 uncinate_fasciculus	-0.020	0.011	-1.841	0.066
NETRIN1 forceps_major	-0.011	0.012	-0.908	0.364
NETRIN1 forceps_minor	-0.016	0.012	-1.299	0.194
NETRIN1 middle_cerebellar_peduncle	-0.017	0.012	-1.360	0.174
Genomic acoustic_radiation	-0.008	0.011	-0.773	0.439
Genomic anterior_thalamic_radiation	-0.010	0.012	-0.894	0.371
Genomic cingulate_gyrus_part_of_cingulum	-0.010	0.011	-0.925	0.355
Genomic parahippocampal_part_of_cingulum	-0.014	0.011	-1.309	0.191
Genomic corticospinal_tract	-0.025	0.012	-2.154	0.031
Genomic inferior_fronto_occipital_fasciculus	-0.021	0.012	-1.758	0.079
Genomic inferior_longitudinal_fasciculus	-0.017	0.012	-1.436	0.151
Genomic medial_lemniscus	-0.015	0.011	-1.382	0.167
Genomic posterior_thalamic_radiation	-0.018	0.011	-1.586	0.113
Genomic superior_longitudinal_fasciculus	-0.014	0.012	-1.193	0.233
Genomic superior_thalamic_radiation	-0.019	0.012	-1.613	0.107
Genomic uncinate_fasciculus	-0.025	0.011	-2.284	0.022
Genomic forceps_major	-0.019	0.012	-1.521	0.128

Genomic forceps_minor	-0.017	0.012	-1.370	0.171
Genomic middle_cerebellar_peduncle	-0.019	0.012	-1.545	0.122

Table S16. The effect of pruned NETRIN1- and Genomic-PRS at threshold 0.5 on individual white matter tracts (FA) (N = 6,420).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 gFA	-0.002	0.001	-1.966	0.049
NETRIN1 Association fibres	-0.002	0.001	-2.567	0.010
NETRIN1 Thalamic radiations	-0.001	0.000	-1.327	0.184
NETRIN1 Projection fibres	0.000	0.001	-0.668	0.504
Genomic gFA	-0.002	0.001	-1.794	0.073
Genomic Association fibres	-0.001	0.001	-1.656	0.098
Genomic Thalamic radiations	-0.001	0.000	-1.376	0.169
Genomic Projection fibres	-0.001	0.001	-1.983	0.047

Table S17. The effect of unpruned NETRIN1- and Genomic-PRS at threshold 0.5 on tract categories (FA) (N = 6,420).

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	0.006	0.011	0.544	0.586
NETRIN1 anterior_thalamic_radiation	0.008	0.011	0.773	0.439
NETRIN1 cingulate_gyrus_part_of_cingulum	0.013	0.011	1.241	0.215
NETRIN1 parahippocampal_part_of_cingulum	-0.012	0.011	-1.116	0.264
NETRIN1 corticospinal_tract	-0.002	0.011	-0.206	0.837
NETRIN1 inferior_fronto_occipital_fasciculus	0.017	0.011	1.482	0.138
NETRIN1 inferior_longitudinal_fasciculus	0.017	0.011	1.543	0.123
NETRIN1 medial_lemniscus	0.013	0.011	1.225	0.220
NETRIN1 posterior_thalamic_radiation	0.008	0.011	0.773	0.439
NETRIN1 superior_longitudinal_fasciculus	0.018	0.011	1.556	0.120
NETRIN1 superior_thalamic_radiation	0.008	0.010	0.851	0.395
NETRIN1 uncinate_fasciculus	0.014	0.010	1.387	0.165
NETRIN1 bl.MD.wm.forceps_major	0.009	0.012	0.741	0.459
NETRIN1 bl.MD.wm.forceps_minor	0.016	0.012	1.390	0.165
NETRIN1 bl.MD.wm.middle_cerebellar_peduncle	-0.004	0.012	-0.350	0.726
Genomic acoustic_radiation	-0.004	0.011	-0.353	0.724
Genomic anterior_thalamic_radiation	0.019	0.011	1.833	0.067
Genomic cingulate_gyrus_part_of_cingulum	0.024	0.011	2.263	0.024
Genomic parahippocampal_part_of_cingulum	0.008	0.011	0.715	0.475
Genomic corticospinal_tract	0.012	0.011	1.041	0.298
Genomic inferior_fronto_occipital_fasciculus	0.019	0.011	1.711	0.087
Genomic inferior_longitudinal_fasciculus	0.008	0.011	0.758	0.449
Genomic medial_lemniscus	0.001	0.011	0.117	0.907
Genomic posterior_thalamic_radiation	-0.001	0.011	-0.100	0.920
Genomic superior_longitudinal_fasciculus	0.017	0.011	1.503	0.133
Genomic superior_thalamic_radiation	0.018	0.010	1.831	0.067
Genomic uncinate_fasciculus	0.023	0.010	2.213	0.027
Genomic bl.MD.wm.forceps_major	0.014	0.012	1.138	0.255
Genomic bl.MD.wm.forceps_minor	0.018	0.012	1.602	0.109
Genomic bl.MD.wm.middle_cerebellar_peduncle	0.010	0.012	0.821	0.411
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.010	0.010	-0.947	0.344
NETRIN1 anterior_thalamic_radiation	0.006	0.011	0.571	0.568

NETRIN1 cingulate_gyrus_part_of_cingulum	0.014	0.011	1.273	0.203
NETRIN1 parahippocampal_part_of_cingulum	0.001	0.011	0.082	0.935
NETRIN1 corticospinal_tract	-0.006	0.011	-0.575	0.565
NETRIN1 inferior_fronto_occipital_fasciculus	0.010	0.011	0.883	0.377
NETRIN1 inferior_longitudinal_fasciculus	0.014	0.011	1.274	0.203
NETRIN1 medial_lemniscus	0.004	0.011	0.395	0.693
NETRIN1 posterior_thalamic_radiation	0.011	0.011	0.993	0.321
NETRIN1 superior_longitudinal_fasciculus	0.018	0.011	1.549	0.122
NETRIN1 superior_thalamic_radiation	0.007	0.010	0.741	0.458
NETRIN1 uncinate_fasciculus	0.008	0.010	0.780	0.435
NETRIN1 bl.MD.wm.forceps_major	0.011	0.012	0.937	0.349
NETRIN1 bl.MD.wm.forceps_minor	0.009	0.012	0.752	0.452
NETRIN1 bl.MD.wm.middle_cerebellar_peduncle	-0.007	0.012	-0.600	0.549
Genomic acoustic_radiation	0.004	0.010	0.428	0.669
Genomic anterior_thalamic_radiation	0.018	0.011	1.754	0.079
Genomic cingulate_gyrus_part_of_cingulum	0.021	0.011	1.993	0.046
Genomic parahippocampal_part_of_cingulum	0.013	0.011	1.222	0.222
Genomic corticospinal_tract	0.019	0.011	1.674	0.094
Genomic inferior_fronto_occipital_fasciculus	0.024	0.011	2.168	0.030
Genomic inferior_longitudinal_fasciculus	0.013	0.011	1.160	0.246
Genomic medial_lemniscus	0.008	0.011	0.724	0.469
Genomic posterior_thalamic_radiation	0.001	0.011	0.091	0.928
Genomic superior_longitudinal_fasciculus	0.017	0.011	1.492	0.136
Genomic superior_thalamic_radiation	0.017	0.010	1.735	0.083
Genomic uncinate_fasciculus	0.015	0.010	1.418	0.156
Genomic bl.MD.wm.forceps_major	0.016	0.012	1.298	0.194
Genomic bl.MD.wm.forceps_minor	0.020	0.012	1.703	0.089
Genomic bl.MD.wm.middle_cerebellar_peduncle	0.010	0.012	0.791	0.429
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.008	0.011	-0.727	0.468
NETRIN1 anterior_thalamic_radiation	0.014	0.011	1.297	0.195
NETRIN1 cingulate_gyrus_part_of_cingulum	0.014	0.011	1.352	0.176
NETRIN1 parahippocampal_part_of_cingulum	-0.009	0.011	-0.899	0.369
NETRIN1 corticospinal_tract	-0.001	0.011	-0.071	0.943
NETRIN1 inferior_fronto_occipital_fasciculus	0.018	0.011	1.641	0.101
NETRIN1 inferior_longitudinal_fasciculus	0.020	0.011	1.774	0.076
NETRIN1 medial_lemniscus	0.015	0.011	1.377	0.169
NETRIN1 posterior_thalamic_radiation	0.014	0.011	1.334	0.182
NETRIN1 superior_longitudinal_fasciculus	0.027	0.011	2.413	0.016
NETRIN1 superior_thalamic_radiation	0.016	0.010	1.652	0.099
NETRIN1 uncinate_fasciculus	0.010	0.010	0.991	0.321
NETRIN1 forceps_major	0.014	0.012	1.205	0.228
NETRIN1 forceps_minor	0.015	0.012	1.284	0.199
NETRIN1 middle_cerebellar_peduncle	0.004	0.012	0.359	0.720
Genomic acoustic_radiation	0.002	0.011	0.161	0.872
Genomic anterior_thalamic_radiation	0.011	0.011	1.023	0.306
Genomic cingulate_gyrus_part_of_cingulum	0.014	0.011	1.306	0.191
Genomic parahippocampal_part_of_cingulum	0.016	0.011	1.537	0.124
Genomic corticospinal_tract	0.013	0.011	1.167	0.243
Genomic inferior_fronto_occipital_fasciculus	0.014	0.011	1.228	0.219
Genomic inferior_longitudinal_fasciculus	0.006	0.011	0.516	0.606
Genomic medial_lemniscus	-0.003	0.011	-0.323	0.747
Genomic posterior_thalamic_radiation	-0.008	0.011	-0.791	0.429
Genomic superior_longitudinal_fasciculus	0.009	0.011	0.832	0.405

Genomic superior_thalamic_radiation	0.008	0.010	0.776	0.438
Genomic uncinate_fasciculus	0.013	0.010	1.246	0.213
Genomic forceps_major	0.009	0.012	0.751	0.453
Genomic forceps_minor	0.010	0.012	0.852	0.394
Genomic middle_cerebellar_peduncle	0.009	0.012	0.712	0.476
PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.002	0.010	-0.187	0.852
NETRIN1 anterior_thalamic_radiation	0.020	0.011	1.854	0.064
NETRIN1 cingulate_gyrus_part_of_cingulum	0.017	0.011	1.586	0.113
NETRIN1 parahippocampal_part_of_cingulum	-0.009	0.011	-0.813	0.416
NETRIN1 corticospinal_tract	0.008	0.011	0.694	0.488
NETRIN1 inferior_fronto_occipital_fasciculus	0.025	0.011	2.193	0.028
NETRIN1 inferior_longitudinal_fasciculus	0.025	0.011	2.221	0.026
NETRIN1 medial_lemniscus	0.004	0.011	0.338	0.736
NETRIN1 posterior_thalamic_radiation	0.020	0.011	1.878	0.060
NETRIN1 superior_longitudinal_fasciculus	0.030	0.011	2.603	0.009
NETRIN1 superior_thalamic_radiation	0.020	0.010	2.051	0.040
NETRIN1 uncinate_fasciculus	0.015	0.010	1.421	0.155
NETRIN1 forceps_major	0.014	0.012	1.181	0.237
NETRIN1 forceps_minor	0.014	0.012	1.202	0.229
NETRIN1 middle_cerebellar_peduncle	0.009	0.012	0.759	0.448
Genomic acoustic_radiation	-0.002	0.010	-0.177	0.860
Genomic anterior_thalamic_radiation	0.013	0.011	1.216	0.224
Genomic cingulate_gyrus_part_of_cingulum	0.014	0.011	1.359	0.174
Genomic parahippocampal_part_of_cingulum	0.018	0.011	1.678	0.093
Genomic corticospinal_tract	0.002	0.011	0.205	0.838
Genomic inferior_fronto_occipital_fasciculus	0.014	0.011	1.238	0.216
Genomic inferior_longitudinal_fasciculus	0.006	0.011	0.495	0.621
Genomic medial_lemniscus	-0.002	0.011	-0.204	0.839
Genomic posterior_thalamic_radiation	-0.010	0.011	-0.972	0.331
Genomic superior_longitudinal_fasciculus	0.008	0.011	0.667	0.505
Genomic superior_thalamic_radiation	0.006	0.010	0.582	0.561
Genomic uncinate_fasciculus	0.010	0.010	0.960	0.337
Genomic forceps_major	0.014	0.012	1.202	0.230
Genomic forceps_minor	0.016	0.012	1.425	0.154
Genomic middle_cerebellar_peduncle	0.011	0.012	0.870	0.384
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.003	0.010	-0.333	0.739
NETRIN1 anterior_thalamic_radiation	0.022	0.011	2.070	0.039
NETRIN1 cingulate_gyrus_part_of_cingulum	0.018	0.011	1.698	0.089
NETRIN1 parahippocampal_part_of_cingulum	-0.006	0.011	-0.608	0.543
NETRIN1 corticospinal_tract	0.009	0.011	0.789	0.430
NETRIN1 inferior_fronto_occipital_fasciculus	0.024	0.011	2.176	0.030
NETRIN1 inferior_longitudinal_fasciculus	0.023	0.011	2.018	0.044
NETRIN1 medial_lemniscus	0.004	0.011	0.355	0.723
NETRIN1 posterior_thalamic_radiation	0.020	0.011	1.875	0.061
NETRIN1 superior_longitudinal_fasciculus	0.029	0.011	2.576	0.010
NETRIN1 superior_thalamic_radiation	0.021	0.010	2.132	0.033
NETRIN1 uncinate_fasciculus	0.016	0.010	1.562	0.118
NETRIN1 forceps_major	0.013	0.012	1.067	0.286
NETRIN1 forceps_minor	0.016	0.012	1.403	0.161
NETRIN1 middle_cerebellar_peduncle	0.008	0.012	0.664	0.507
Genomic acoustic_radiation	-0.003	0.010	-0.318	0.750

Genomic anterior_thalamic_radiation	0.013	0.011	1.202	0.229
Genomic cingulate_gyrus_part_of_cingulum	0.014	0.011	1.292	0.196
Genomic parahippocampal_part_of_cingulum	0.019	0.011	1.780	0.075
Genomic corticospinal_tract	0.003	0.011	0.258	0.796
Genomic inferior_fronto_occipital_fasciculus	0.013	0.011	1.167	0.243
Genomic inferior_longitudinal_fasciculus	0.005	0.011	0.471	0.638
Genomic medial_lemniscus	-0.004	0.011	-0.396	0.692
Genomic posterior_thalamic_radiation	-0.013	0.011	-1.172	0.241
Genomic superior_longitudinal_fasciculus	0.006	0.011	0.549	0.583
Genomic superior_thalamic_radiation	0.005	0.010	0.464	0.643
Genomic uncinate_fasciculus	0.010	0.010	0.983	0.326
Genomic forceps_major	0.015	0.012	1.282	0.200
Genomic forceps_minor	0.012	0.012	1.044	0.296
Genomic middle_cerebellar_peduncle	0.012	0.012	1.006	0.314

Table S18. The effect of pruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on individual white matter tracts (MD) (N = 6,390).

PGRS THRESHOLD: 0.01	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.011	0.012	0.998	0.318
NETRIN1 Association fibres	0.009	0.012	0.737	0.461
NETRIN1 Thalamic radiations	0.010	0.011	0.896	0.370
NETRIN1 Projection fibres	0.005	0.012	0.429	0.668
Genomic gMD	0.018	0.011	1.546	0.122
Genomic Association fibres	0.020	0.012	1.690	0.091
Genomic Thalamic radiations	0.011	0.011	1.041	0.298
Genomic Projection fibres	0.015	0.012	1.205	0.228
PGRS THRESHOLD: 0.05	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.010	0.011	0.844	0.399
NETRIN1 Association fibres	0.012	0.012	0.995	0.320
NETRIN1 Thalamic radiations	0.010	0.011	0.916	0.360
NETRIN1 Projection fibres	-0.003	0.012	-0.268	0.789
Genomic gMD	0.021	0.011	1.798	0.072
Genomic Association fibres	0.022	0.012	1.913	0.056
Genomic Thalamic radiations	0.012	0.011	1.091	0.276
Genomic Projection fibres	0.019	0.012	1.595	0.111
PGRS THRESHOLD: 0.1	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.015	0.011	1.327	0.184
NETRIN1 Association fibres	0.011	0.012	0.970	0.332
NETRIN1 Thalamic radiations	0.017	0.011	1.583	0.114
NETRIN1 Projection fibres	0.010	0.012	0.796	0.426
Genomic gMD	0.012	0.011	1.064	0.287
Genomic Association fibres	0.018	0.012	1.539	0.124
Genomic Thalamic radiations	0.001	0.011	0.120	0.904
Genomic Projection fibres	0.012	0.012	1.010	0.312
PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.020	0.011	1.783	0.075
NETRIN1 Association fibres	0.015	0.012	1.328	0.184
NETRIN1 Thalamic radiations	0.023	0.011	2.169	0.030
NETRIN1 Projection fibres	0.014	0.012	1.171	0.242
Genomic gMD	0.012	0.011	1.045	0.296
Genomic Association fibres	0.018	0.012	1.526	0.127

Genomic Thalamic radiations	0.001	0.011	0.050	0.960
Genomic Projection fibres	0.014	0.012	1.174	0.240
PGRS THRESHOLD: 1	Value	Std. Error	t-value	p-value
NETRIN1 gMD	0.021	0.011	1.829	0.068
NETRIN1 Association fibres	0.016	0.012	1.412	0.158
NETRIN1 Thalamic radiations	0.024	0.011	2.263	0.024
NETRIN1 Projection fibres	0.013	0.012	1.060	0.289
Genomic gMD	0.011	0.011	0.979	0.328
Genomic Association fibres	0.018	0.012	1.533	0.125
Genomic Thalamic radiations	-0.001	0.011	-0.091	0.928
Genomic Projection fibres	0.015	0.012	1.206	0.228

Table S19. The effect of pruned NETRIN1- and Genomic-PRS at thresholds 0.01, 0.05, 0.1, 0.5 and 1 on tract categories (MD) (N = 6,390).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 acoustic_radiation	-0.002	0.011	-0.223	0.824
NETRIN1 anterior_thalamic_radiation	0.015	0.011	1.434	0.151
NETRIN1 cingulate_gyrus_part_of_cingulum	0.013	0.011	1.153	0.249
NETRIN1 parahippocampal_part_of_cingulum	-0.008	0.011	-0.762	0.446
NETRIN1 corticospinal_tract	0.004	0.012	0.341	0.733
NETRIN1 inferior_fronto_occipital_fasciculus	0.020	0.011	1.764	0.078
NETRIN1 inferior_longitudinal_fasciculus	0.020	0.011	1.779	0.075
NETRIN1 medial_lemniscus	0.001	0.011	0.059	0.953
NETRIN1 posterior_thalamic_radiation	0.017	0.011	1.588	0.112
NETRIN1 superior_longitudinal_fasciculus	0.023	0.011	2.036	0.042
NETRIN1 superior_thalamic_radiation	0.015	0.010	1.456	0.145
NETRIN1 uncinate_fasciculus	0.012	0.011	1.102	0.271
NETRIN1 forceps_major	0.014	0.012	1.146	0.252
NETRIN1 forceps_minor	0.011	0.012	0.949	0.342
NETRIN1 middle_cerebellar_peduncle	0.007	0.012	0.573	0.567
Genomic acoustic_radiation	-0.009	0.011	-0.817	0.414
Genomic anterior_thalamic_radiation	0.003	0.011	0.261	0.794
Genomic cingulate_gyrus_part_of_cingulum	0.003	0.011	0.234	0.815
Genomic parahippocampal_part_of_cingulum	0.010	0.011	0.980	0.327
Genomic corticospinal_tract	-0.008	0.012	-0.656	0.512
Genomic inferior_fronto_occipital_fasciculus	0.002	0.011	0.164	0.870
Genomic inferior_longitudinal_fasciculus	-0.005	0.011	-0.427	0.669
Genomic medial_lemniscus	-0.009	0.011	-0.790	0.430
Genomic posterior_thalamic_radiation	-0.018	0.011	-1.651	0.099
Genomic superior_longitudinal_fasciculus	-0.002	0.012	-0.187	0.851
Genomic superior_thalamic_radiation	-0.004	0.011	-0.335	0.738
Genomic uncinate_fasciculus	0.000	0.011	-0.035	0.972
Genomic forceps_major	0.005	0.012	0.440	0.660
Genomic forceps_minor	0.004	0.012	0.352	0.725
Genomic middle_cerebellar_peduncle	0.006	0.012	0.455	0.649

Table S20. The effect of unpruned NETRIN1- and Genomic-PRS at threshold 0.5 on individual white matter tracts (MD) (N = 6,420).

PGRS THRESHOLD: 0.5	Value	Std. Error	t-value	p-value
NETRIN1 gMD	2.53E-06	1.42E-06	1.78E+00	7.47E-02
NETRIN1 Association fibres	1.43E-06	1.08E-06	1.33E+00	1.84E-01
NETRIN1 Thalamic radiations	1.71E-06	7.87E-07	2.17E+00	3.01E-02
NETRIN1 Projection fibres	9.36E-07	8.00E-07	1.17E+00	2.42E-01
Genomic gMD	1.50E-06	1.43E-06	1.04E+00	2.96E-01
Genomic Association fibres	1.66E-06	1.09E-06	1.53E+00	1.27E-01
Genomic Thalamic radiations	3.99E-08	7.93E-07	5.02E-02	9.60E-01
Genomic Projection fibres	9.47E-07	8.06E-07	1.17E+00	2.40E-01

Table S21. The effect of unpruned NETRIN1- and Genomic-PRS at threshold 0.5 on tract categories (MD) (N = 6,420).

Results depicted in tables S6 – S21 indicate secondary analyses which complement the primary analyses. These consist firstly of the effect unpruned NETRIN1- and genomic-PRS on FA and MD values, conducted on both the full dataset (N = 6,420) and the dataset with excluded outliers (N = 6,401 and 6,390 for FA and MD, respectively). Secondly, the effect of pruned NETRIN1- and genomic-PRS on FA and MD values was also investigated, again conducted on both the full dataset and dataset with excluded outliers. The analyses consist of PRS at all five p-value thresholds (0.01, 0.05, 0.1, 0.5 and 1). A similar pattern is observed for significance in white matter tracts associated with both PRS lists across PRS thresholds within the dataset with outliers removed. Some tracts remain significant within the full sample dataset as compared to the sample with outliers removed at PRS threshold 0.5, however there is a trend towards more significant results when outliers are removed. Please refer to tables S8-S9, S12-S13, S16-S17 and S20-S21 for an account of results at PRS threshold 0.5 within the full dataset, which are directly comparable to the primary results depicted in the manuscript (PRS threshold 0.5 with outliers removed).

White matter tracts significantly associated with both NETRIN1-PRS and genomic-PRS.

Fractional anisotropy

Tract categories

Significantly lower FA values in association fibres were found for both NETRIN1-PRS (β = -0.032, $p_{corrected}$ = 0.023) and genomic-PRS (β = -0.033, $p_{corrected}$ = 0.011).

Mean diffusivity

Global measures

Significantly higher gMD was associated with both NETRIN1-PRS (β = 0.027, $p_{corrected}$ = 0.031) and genomic-PRS (β = 0.033, $p_{corrected}$ = 0.006).

Individual white matter tracts

Significantly higher MD in the inferior fronto-occipital fasciculus was found for both NETRIN1-PRS (β = 0.027, $p_{corrected}$ = 0.046) and genomic-PRS (β = 0.031, $p_{corrected}$ = 0.018).

Appendix 3: Supplementary materials for Chapter 4: Genetic and epigenetic prediction of Major Depressive Disorder and associations with white matter microstructure in Generation Scotland

Details of exclusion process in TBSS pre-processing pipeline

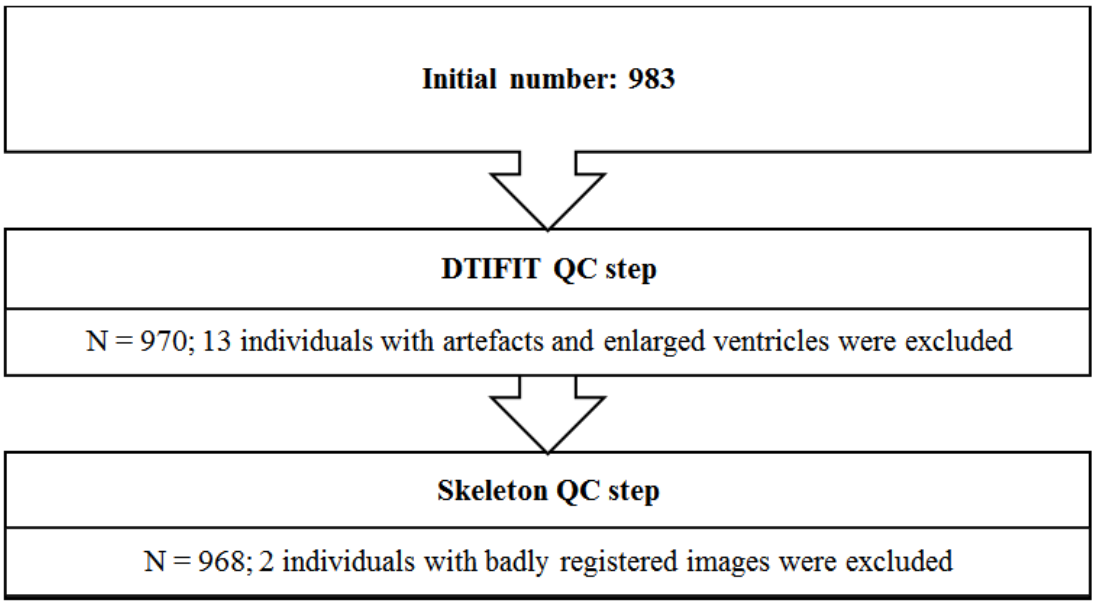


Figure S1. At time of pre-processing, 983 individuals had raw DTI data available. The figure above indicates the two visual QC steps. At the DTIFIT QC step, the main reason for exclusion was enlarged ventricles in individuals. These individuals were excluded as in later steps in the pre-processing pipeline, a mean FA image would be compiled, which could be skewed due to enlarged ventricles. At the skeleton QC step, two images were flipped, and would be unable to be passed through the pipeline.

Individual white matter tracts included in global FA and MD PCA analyses

White matter tract	Abbreviation
Cingulum (hippocampus)	CGH
Cingulum (cingulate gyrus)	CGC
Fornix*	FX
Fornix (cres) / Stria terminalis	FX / ST
Inferior fronto-occipital fasciculus	IFO
Superior fronto-occipital fasciculus	SFO
External capsule	EC
Superior longitudinal fasciculus	SLF
Sagittal striatum	SS
Uncinate fasciculus	UNC
Body of corpus callosum*	BCC
Genu of corpus callosum*	GCC
Splenium of corpus callosum*	SCC
Anterior corona radiata	ACR
Posterior corona radiata	PCR
Superior corona radiata	SCR
Corticospinal tract	CST
Anterior limb of internal capsule	ALIC
Posterior limb of internal capsule	PLIC
Posterior thalamic radiation	PTR
Retrolenticular limb of internal capsule	RLIC

Table S1. White matter tracts used as dependent variables in statistical analyses outlined below. * = unilateral tracts.

The following tracts were excluded from the global FA and MD PCA derivation: (1) corpus callosum; (2) corona radiata; and (3) internal capsule. This is because TBSS outputs subsets of these three tracts, as well as the entire tract, as indicated below. Including the subsets as well as the entire tract output would have resulted in over-inclusion of these regions; including the three sub-sets of tracts within these overall tracts aided in observing whether there is an association of the two risk scores with specific, individualised white matter tracts, rather than a more regional estimate comprising all three.

1. For corpus callosum: body, genu, and splenium of corpus callosum.
2. For corona radiata: anterior, posterior, and superior corona radiata.
3. For internal capsule: anterior, posterior, and retrolenticular limbs of internal capsule.

CpG site	Beta	CpG site	Beta
cg20116804	-0.015421634	cg18751657	0.027288385
cg15971980	0.006942908	cg04821375	0.024358727
cg01049205	0.067276206	cg02634584	0.757509208
cg12736206	0.036873976	cg19143959	-0.266215326
cg22225420	0.278735199	cg02203922	-0.053844111
cg11044575	0.113164134	cg03903647	-0.353215677
cg24254177	0.26284975	cg25610515	0.019056411
cg22407822	0.001719539	cg02822381	0.069731732
cg20984994	-0.028034876	cg18200311	-0.003902515
cg26063721	-0.064925727	cg17943757	-0.13338057
cg24173182	0.093449601	cg11463427	0.228010283
cg10539371	-0.266530585	cg27653901	-0.21325932
cg17250537	0.114100798	cg08744097	-0.075534229
cg25985659	-0.016420323	cg13483916	0.113879106
cg02576528	-0.008732521	cg25821785	0.045204034
cg21124940	0.125069705	cg26621790	-2.09E-05
cg01038738	0.09211571	cg10435816	-0.114915234
cg21022949	-0.122922677	cg09490565	0.064310355
cg18197594	-0.037357869	cg01170758	0.743479881
cg18090197	0.275197436	cg13278241	-0.244617463
cg14728380	0.000910739	cg17775332	-0.046708676
cg15248828	0.004431601	cg27404676	0.08042557
cg03859186	0.182232759	cg20528583	-7.91E-05
cg13247663	-0.131885152	cg15770238	0.036076248
cg13529291	0.172056972	cg07920739	-0.032823893
cg14996929	-0.053869123	cg07861790	-0.064235819
cg04191989	0.134906691	cg01950844	0.2704649
cg09906991	-0.003536151	cg07971952	0.025230404
cg06482498	-0.002781115	cg06157334	-0.05587217
cg18035255	-0.234119678	cg05592146	-0.02029348
cg22044566	-0.052043377	cg14210405	0.065082199
cg02459042	-0.005130736	cg26579032	-0.03597194
cg02055264	0.020186238	cg07323350	-0.322253068
cg25242471	0.095306294	cg21562656	-0.073532389
cg07548512	-0.389017708	cg10139443	-0.908790082
cg10928544	0.00597106	cg03736774	0.181528807
cg24072885	0.174686087	cg03827626	-0.495993086
cg08464831	-0.077103695	cg09614389	0.001449656
cg17054674	-0.109492909	cg25600478	-0.013482982
cg26172211	-0.75833086	cg04029366	-0.08658678
cg26038465	0.065443581	cg25394505	0.075190223
cg06833732	-0.032487218	cg13494933	0.012216425

cg16081176	0.099548626	cg26720682	0.032298882
cg24601536	0.084188501	cg10401489	-0.01351181
cg27332938	-0.075262664	cg12962542	-0.132560179
cg20674014	0.204549134	cg26422761	-0.072979904
cg23986470	0.110633334	cg12461092	0.012583705
cg08873940	0.390595732	cg00298921	-0.025104852
cg16605431	-0.067886882	cg21974358	-0.057762198
cg26416971	0.017162464	cg04349815	-0.048879526
cg10438391	0.017884374	cg19421526	0.059908283
cg01305745	0.182066587	cg05924543	0.236377006
cg18355902	-0.108723476	cg08912860	0.071151194
cg26099134	0.049722806	cg22024931	0.007790324
cg22237300	-0.078493498	cg23817627	0.035000194
cg20545941	-0.214689878	cg13999210	0.105935225
cg14443301	-0.070153977	cg21621114	-0.03694387
cg10451078	-0.162129625	cg03839794	-0.028866007
cg20984053	-0.13668675	cg12160741	0.012219106
cg20273485	-0.119750057	cg08805821	0.044670337
cg05176970	0.012997037	cg03055837	-0.053989788
cg07244098	-0.067200919	cg24456846	0.066426026
cg02613370	0.129967025	cg01297383	-0.065075788
cg03079761	-0.105643912	cg25949304	0.07527537
cg12138286	0.186379821	cg05621218	-0.018744792
cg12654519	-0.218628548	cg21292008	-0.076694204
cg24583766	0.042427484	cg15207669	0.063024294
cg25707767	0.002092093	cg23214464	-0.011578311
cg04772025	-0.007247445	cg04758026	0.007067428
cg13463245	-0.104253112	cg09865955	-0.038167735
cg19866673	-0.03368796	cg13751872	0.001763831
cg16761754	0.01285452	cg15046935	0.067951209
cg14375923	-0.039366583	cg12609526	0.13350385
cg09910998	0.028668133	cg19698976	0.005413523
cg24948792	-0.090876298	cg00828721	0.03947539
cg17537844	0.00997141	cg01494348	0.033054139
cg12140144	0.056099238	cg27168858	-0.028599887
cg10515332	0.160491599	cg26146184	0.086514453
cg02770534	-0.017156512	cg00555420	-0.022829456
cg07296835	0.035630636	cg27129029	0.022538802
cg06781788	-0.026267442	cg24367957	-0.030035942
cg18944924	-0.068730008	cg11507780	-0.004280805
cg03230711	-0.037727384	cg20821187	-0.030688714
cg24185124	0.219559925	cg09552652	-0.060135871
cg05070690	-0.160686747	cg24769830	0.001766541

cg17983217	0.070366023	cg22539189	-0.050789338
cg20711828	0.163479128	cg09935388	-0.041843214
cg07733920	-0.046450328	cg15849154	0.005918876
cg00287370	-0.713805334	cg22738642	-0.017391434
cg21601837	0.031436265	cg14157549	0.038867665
cg22210337	0.01654544	cg06360820	-0.011082584
cg09320113	0.192740454	cg08800396	-0.004683139
cg07163389	-0.174219338	cg18815120	-0.002973236
cg05828191	-0.085560334	cg02082929	0.008323224
cg22430972	0.860620796	cg09557034	-0.093897544
cg08821669	-0.014756084	cg00344422	-0.007259721
cg24252746	-0.709664177	cg00318111	0.08602903
cg00474840	-0.505416185	cg21549285	-0.060031127
cg18811093	-0.054353706	cg05141400	-0.018150327
cg23889772	0.028348977	cg24425727	-0.002068791
cg04884395	-0.216310791	cg09768983	-0.002262711
cg19047068	-0.07625118	cg19806221	0.005584262
cg16088894	-0.019705825	cg07576632	-0.017198847
cg15673187	-0.110154478	cg07801181	-0.011335721
cg14337472	-0.237370309	cg02211983	0.000787545
cg23878564	0.040093287	cg17542176	-0.087123853
cg14959820	0.094109238	cg25189904	-0.052305954
cg07076105	-0.080253123	cg16685388	0.005120511
cg08747591	-0.171259744	cg07620573	-0.053164589
cg06819963	0.060736712	cg15374515	-0.080333691
cg24754199	0.244701222	cg17109042	-0.069025755
cg02505588	0.045103077	cg16322792	0.0101152
cg16442574	-0.118103539	cg14164492	0.027446766
cg17517128	-0.02734914	cg16434510	-0.026626517
cg23273834	0.103155906	cg22507558	-0.045030822
cg02813644	-0.147179559	cg27304415	0.006514712
cg21848117	0.067060087	cg21243459	-0.030101755
cg24476033	0.028594309	cg17568934	-0.006180372
cg07761822	0.066217342	cg24487940	-0.066777461
cg14070323	0.036294258	cg01859717	-0.218157298
cg13422261	0.225832858	cg08754268	-0.042172831
cg01893681	-0.024661955	cg11864774	-0.048926955
cg21033440	0.054716338	cg10185424	-0.017677734
cg05544413	-0.057264219	cg14156792	-0.002430209
cg25982965	-0.021451257	cg15611176	0.152441437
cg14556303	-0.020044439	cg07211220	-0.105345409
cg26326298	0.341406175	cg06754079	-0.002567777

Table S2. CpG sites selected by the LASSO penalised regression model for the MDD DNAm predictor and their beta weights.