



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2017

The Dual Faces of Misery

Arden Moscati
VCU

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Mental Disorders Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/4809>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

The Dual Faces of Misery

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy at Virginia Commonwealth University.

By Arden Moscati

Master of Arts, University of Oxford, 2013

Bachelor of Arts, University of Oxford, 2008

Dissertation Advisor: Kenneth S. Kendler, MD

Rachel Brown Banks Distinguished Professor of Psychiatry

Director of The Virginia Institute for Psychiatric and Behavioral Genetics

Virginia Commonwealth University

Richmond, Virginia

February 23rd, 2017

Acknowledgement

I have many thanks to give, for this project would not have been possible without the guidance, good will, and support of many individuals important to me.

I wish to express my thanks to Dr. Kenneth Kendler for agreeing to mentor the chief portion of my graduate studies. His calm support and thoughtful advice enabled me to finish every project I began, while still finding time to explore interests beyond the tasks of any given day. I look forward to continuing the studies of philosophy, psychiatric theory, and Shakespeare that I undertook with him. Key to this project, as well, was his help in navigating the limitless bounds of my imagination to find a dissertation idea that could actually be completed as a graduate student!

I owe a great debt to my committee, Drs. Lindon Eaves, Ananda Amstadter, Brien Riley and Silviu-Alin Bacanu, as well. They are not only valued scientific advisors, but also, happily, esteemed coauthors, and, I would like to think, friends. I must commend and thank them for their patience with the vicissitudes that this project has seen over the years of its inception and completion. Especial thanks are due to Dr. Lindon Eaves, as my first graduate mentor, who instilled in me a deep reverence for the search for truth that is science, as well as a keen sensitivity (and aversion) to sophistry.

My fellow psychiatric, behavioral, and statistical genetics students (and honorary PBSG) have been tantamount to a second family during this graduate program, and the magnitude of my gratitude is beyond expression. Suffice to say that intellectually, emotionally, socially, and practically, I have rarely relied on a group of friends more, and the support I've been given has been unfailing. My dear Team Awesome (Jeanne Savage and Elizabeth Do), I couldn't ask to be part of a more superb cohort, and I earnestly demand we continue playing, working and researching together eternally.

The many collaborators and informal mentors I've found at the Virginia Institute for Psychiatric and Behavioral Genetics have taught me everything of worth I know about science, and I hope to make your investment in me worthwhile as best I can.

Finally, all those involved in the collection and creation of the data and analytic tools used in this project are given thanks as well, for their work directly enabled my own.

Table of Contents

List of Figures	v
List of Tables.....	vii
List of Abbreviations	viii
Abstract.....	x
Preamble.....	1
Chapter 1: Introduction	4
I. Public health significance of anxiety and depressive disorders	4
II. Diagnostic History of MD.....	5
III. Diagnostic History of GAD	10
IV. History of Neuroticism	13
V. Epidemiology of MD	14
VI. Epidemiology of GAD.....	18
VII. Relationship of Neuroticism to anxiety and depressive disorders.....	19
VIII. Biometrical Modeling of MD and GAD	20
IX. Genotypic studies of MD and GAD	23
X. Scope of dissertation project.....	28

Chapter 2: Exploration of psychosocial risk factors in MD, GAD, and neuroticism.....	32
I. Methods	32
II. Results	36
III. Discussion	49
Chapter 3: Investigation into potential genetic variants related to MD, GAD, and neuroticism.....	52
I. Methods	52
II. Results	56
III. Discussion	94
Chapter 4: Estimation of genetic variance, correlation, and cross-prediction in MD, GAD, neuroticism.....	99
I. Methods	99
II. Results	102
III. Discussion	108
Chapter 5: Conclusions.....	111
List of References	124
Vita	143

List of Figures

Figure 1. CONVERGE Marital Termination	38
Figure 2. CONVERGE Social Support.....	39
Figure 3. CONVERGE Childhood Stressful Life Events	40
Figure 4. CONVERGE Adulthood Stressful Life Events	41
Figure 5. S4S Relationship Satisfaction.....	45
Figure 6. S4S Social Activities	46
Figure 7. S4S Social Support.....	47
Figure 8. S4S Stressful Life Events	48
Figure 9. MD BOLT-LMM Results in CONVERGE	58
Figure 10. MD q-q Plot in CONVERGE	59
Figure 11. CONVERGE MD Chromosome 10 Locus	60
Figure 12. CONVERGE MD Chromosome 19 Locus	61
Figure 13. MD+GAD BOLT-LMM Results in CONVERGE	62
Figure 14. MD+GAD q-q Plot in CONVERGE.....	63
Figure 15. Neuroticism BOLT-LMM Results in CONVERGE	64
Figure 16. Neuroticism q-q Plot in CONVERGE	65

Figure 17. MD BOLT-LMM Results in MGS.....	67
Figure 18. MD q-q Plot in MGS.....	68
Figure 19. MD+GAD BOLT-LMM Results in MGS.....	69
Figure 20. MD+GAD q-q Plot in MGS.....	70
Figure 21. GAD BOLT-LMM Results in MGS.....	71
Figure 22. GAD q-q Plot in MGS.....	72
Figure 23. MGS GAD Chromosome 8.....	73
Figure 24. MGS GAD Chromosome 13 Locus.....	74
Figure 25. Neuroticism BOLT-LMM Results in MGS.....	75
Figure 26. Neuroticism q-q Plot in MGS.....	76
Figure 27. Depressive Symptoms PLINK Results in S4S (EAS).....	78
Figure 28. Depressive Symptoms q-q Plot in S4S (EAS).....	79
Figure 29. Depressive Symptoms BOLT-LMM Results in S4S (EUR).....	80
Figure 30. Depressive Symptoms q-q Plot in S4S (EUR).....	81
Figure 31. Neuroticism BOLT-LMM Results in S4S (EAS).....	82
Figure 32. Neuroticism q-q Plot in S4S (EAS).....	83
Figure 33. Neuroticism BOLT-LMM Results in S4S (EUR).....	84
Figure 34. Neuroticism q-q Plot in S4S (EUR).....	85
Figure 35. S4S (EUR) Neuroticism Chromosome 4 Locus.....	86
Figure 36. Stratified histogram of neuroticism values with respect to MD status.....	106

List of Tables

Table 1. DSM-IV / DSM-IV-TR Criteria for Major Depressive Episode	7
Table 2. DSM-IV / DSM-IV-TR Criteria for Major Depressive Disorder, Recurrent.....	9
Table 3. DSM-IV / DSM-IV-TR Criteria for Generalized Anxiety Disorder	12
Table 4. Sample Items from Neuroticism Scales in CONVERGE and S4S	17
Table 5. Previous Genetic findings from GWAS in MD, GAD, and Neuroticism	25
Table 6. Sample characteristics of datasets	31
Table 7 Significant Gene-based MAGMA Results for MD in CONVERGE	88
Table 8. Significant Gene-based MAGMA Results for MD+GAD in CONVERGE	89
Table 9. Significant Gene-based MAGMA Results for Neuroticism in CONVERGE	91
Table 10. Significant Gene-based MAGMA Results for GAD in MGS.	92
Table 11. Significant Gene-based MAGMA Results for Neuroticism in S4S (EAS).....	93
Table 12. Univariate heritability estimates from GCTA and LDSC	105
Table 13. Associations between LDpred polygenic scores and outcomes at each prior level – proportion of causal alleles	107

List of Abbreviations

BFI	Big Five Inventory (see Reference 101)
BUHMBOX	Breaking Up Heterogeneous Mixture Based On Cross-locus correlations
CI	Confidence Interval
CIDI	Composite International Diagnostic Interview (see Ref. 99)
CIDI-SF	Composite International Diagnostic Interview, Short Form (see Ref. 103)
CONVERGE		China, Oxford, and VCU Experimental Research on Genetic Epidemiology
CpG	DNA where a cytosine (C) nucleotide is followed by guanine (G)
DNA	Deoxyribonucleic Acid
DSM	Diagnostic and Statistical Manual of Mental Disorders (see Refs. 26, 27)
EAS	East Asian ancestry group
EUR	European ancestry group
GAD	Generalized Anxiety Disorder
GCTA	Genomic Complex Trait Analysis (see Ref. 89)
GWAS	Genome-Wide Association Study
LD	Linkage Disequilibrium
LDSC	Linkage Disequilibrium Score Regression (see Ref. 90)

LEC	Life Events Checklist (see Ref. 106)
MAGMA	Multi-marker Analysis of Genomic Annotation (see Ref. 121)
MD	Major Depression / Major Depressive Disorder
MD+GAD.....	Comorbid Major Depression and Generalized Anxiety
MGS	Molecular Genetics of Schizophrenia sample (see Ref. 104)
P.....	P-value
PBSG	Psychiatric, Behavioral, and Statistical Genetics
q-q plot	Quantile-quantile plot
RNA	Ribonucleic Acid
S4S	Spit for Science, the VCU Student Survey
SCL-90.....	Symptom Checklist 90 (see Ref. 105)
SE	Standard Error
SES	Socioeconomic Status
USA.....	The United States of America
VATSPSUD	Virginia Adult Twin Study of Psychiatric and Substance Use Disorders
VCU	Virginia Commonwealth University

Abstract

THE DUAL FACES OF MISERY

By
Arden Moscati, M. A. (Oxon)

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2017

Dissertation Advisor: Kenneth S. Kendler, MD

Rachel Brown Banks Distinguished Professor of Psychiatry

Director of The Virginia Institute for Psychiatric and Behavioral Genetics

Major Depression (MD) and Generalized Anxiety Disorder (GAD) are psychiatric disorders that arise from dysfunction of the core human capacities for emotion. Sapience is inextricably bound up with the potential for feelings of regret, worry and concern. When these emotions lead to clinically significant impairment or distress, they may result in one or both of the disorders of MD and GAD. The occurrence of MD and GAD in the same person, known as comorbidity, is remarkably high; substantially higher than would be expected by chance.

MD and GAD have been studied since the mid-20th century, resulting in a substantial body of literature. The personality trait of neuroticism is also known to correlate highly with these disorders. This project was designed to compare the etiological structure of MD and GAD using a range of psychosocial and genetic methods in three datasets, while also assessing the correlated trait of neuroticism. Results are used to inform theoretical formulation of an approximate model of comorbidity for the two disorders.

Psychosocial findings suggest that MD and GAD have similar relationships with most risk factors, and that neuroticism displays results consistent with it composing a portion of the liability to MD and GAD.

Efforts to detect specific genetic loci involved in the etiology of MD and GAD are modestly successful. Two genome-wide significant variants were found for MD (one

already identified in the literature); two for GAD, and one for neuroticism. There were also a number of significant genomic regions for each outcome.

The use of aggregate genetic methods to estimate heritability based on genotypes was less successful. Estimation was only successful in one sample of the three, and produced modest estimates of heritability (0.2-0.25) for MD and comorbid MD+GAD. Genetic correlation was estimated to be very high between neuroticism and MD.

Models of comorbidity are evaluated in light of these results, and a model comprising multiple liability distributions, one shared entirely by MD and GAD, and two additional correlated ones for the two disorders, with reciprocal phenotypic causation, is deemed most consistent with observed evidence.

Preamble

The hallmark of humanity is what might be termed 'sapience', the ability to reason, feel, and be aware of one's own awareness. It has fueled the mastery of humans over Earth, and enabled such achievements as walking on the moon, curing disease and shared experiences through all manner of storytelling media. As may be expected from a faculty with such potential, however, complexities arise. The precise constellation of values on all the myriad trait distributions that comprise the phenomenon of consciousness possessed by each human individual can be conceived as random draws from the probability space of all human minds. Thanks to genetic and experiential diversity, the values are, overall, slowly honed towards increased success in the world with each generation. Some of the traits shaped by this process in humans may include a strong ability to understand others as intentional agents, known as theory of mind (1) and a prosocial, cooperative motivation (2). This progress takes millennia, though, and trends that may have contributed to humanity's early evolutionary success have consequences that cause distress in the state of contemporary society. One example is the metabolic efficiency that enabled early humans to store energy enough to survive even with scant nutritional resources; in present society, the easy access to food (and differing compositions of macronutrients) causes problems such as obesity in

many people (3). An even more basic trend is the characteristic of human nature that leads one to care; caring about other people, evinced in the prosocial motivation, perhaps (2), as well as caring about the content of our experiences, and future goals and hopes. This drive to be emotionally involved with others and the content of our thoughts and perceptions played a crucial role in humanity's ability to band together and protect one another from any number of physical threats, as well as seeking better and more enjoyable ways to interact with the world through an extensive accumulated culture. The unfortunate drawback of this capacity for caring, however, is that when people, events or circumstances deviate from our wishes, we may suffer emotional pain. "The pain that the world is not as you want it to be" (4).

It is important to note that emotion is not unique to humankind, evident for instance in the fact that the processing of different types of emotional vocalizations seems to occur in a similar fashion and in comparable brain areas in humans and dogs (5), whose most recent common ancestor lived between 90 and 100 million years ago (6). But emotions are more varied and elaborate in the sapient human race, a development necessary to achieve the culture that allows the extreme improvement in the quality from the earliest human lives to those in the present day. A useful definition may be: "Culture is an information-based system that allows people to live together and satisfy their needs." (7). Culture, with respect to this broad definition, is not exclusively human either, given examples such as the idea of washing potatoes before eating them being passed down in generations of monkeys on the Japanese island of Koshima (7). But monkeys and other species capable of generating culture (such as apes, whales, dolphins, and elephants) may very well also qualify for sapience (8). Emotions, while not

unique to humans, are key to healthy functioning in humankind. Individuals who lack the capacity to experience emotions, due to a brain injury for instance, show imprudent judgment and often act very impulsively, causing many problems for themselves (9). Of course, even with a full range of emotions, troubles can arise.

Separation from a loved one or being trapped in an awful situation may cause anguish to such extent that nothing seems worthwhile any longer. In certain people, such feelings may arise even without a clear contributing incident – we call this experience depression, and the most common form is major depression, a debilitating and frequently recurrent psychiatric disorder. Another distressing experience that can arise from human emotional faculties is the worry that can arise when one has been exposed to unpleasant experiences in life, and then expects more bad news to lurk around every corner. That state of constant concern in an individual of failing to meet their responsibilities, of not performing their job well or not protecting their loved ones. In some, the content of the worries changes often, but its presence is immutable. If the worry is persistent and problematic, we call it anxiety, and when the conceptual content of the anxiety changes in a free-floating manner, it may fall under the diagnosis of generalized anxiety disorder, a chronically aversive condition. As one may expect, occurrences of depressive and anxious feelings frequently dovetail, with worry causing more unhappiness, and episodes of depression sparking worries about any number of failing relationships, occupational roles or other topics. While anxiety and depressive disorders have been studied for ages, there is still little certainty about the exact etiological relationship between them. This manuscript details some efforts to unravel the intricate webs of misery woven between major depression and generalized anxiety.

Chapter 1: Introduction

I. Public health significance of anxiety and depressive disorders:

While public health programs often are designed to reduce mortality, or deaths due to a particular cause, also important is the amount of non-lethal disease burden. This can include functional impairment, disability and economic consequences, and due in part to their contributions to these outcomes, the impact of depressive and anxiety disorders is great indeed.

Major depression (MD) is the most common psychiatric disorder, and is nearly always associated with functional impairment, with almost 60% reporting severe or very severe role impairment (10). In terms of disability, depression and anxiety disorders are among the highest contributors to years lived with disability around the world, with MD the 2nd greatest cause (behind low back pain), and anxiety disorders 7th, above such problems as migraine, diabetes, and osteoarthritis (11). Generalized anxiety disorder (GAD), one of the most chronic anxiety disorders, is also a leading cause of workplace disability in the United States of America (USA), which of course affects economic productivity negatively, as well (12). In terms of overall economic effects, the impact of MD is substantial, costing \$83.1 billion in the USA during the year 2000 (13), while

anxiety disorders had a cost of \$42.3 billion in 1990 (14). GAD, in particular, has a pronounced negative effect on global well-being and life satisfaction (15).

The effects of MD and GAD are not limited to reduced quality of life or productivity, as both disorders are associated with a range of medical conditions, some of which may be life-threatening. Depression increases risk for myocardial infarction (16), stroke (17), hypertension (18), and depressed individuals are approximately eleven times more likely to attempt suicide than those without depression (19). GAD is associated with a host of ailments as well, including pain conditions, such as arthritis, migraine, and back pain (20), a four-fold increase in risk of peptic ulcer disease (21), coronary heart disease (22), metabolic disorders such as diabetes (23), and chronic obstructive pulmonary disease (24).

In addition to the public health consequences associated with MD and GAD individually, there is a high degree of comorbidity (occurrence of both diseases in the same person) between them. The sequelae of this comorbid state are often worse than either disorder individually. Individuals with comorbid MD and GAD have higher impairment (25), increased disability, utilization of health care and suicidality (26). While MD and GAD both present significant public health concern, the two disorders have quite different histories in psychiatric research and practice.

II. Diagnostic History of MD:

Depression has been characterized in many ways over the years, but there are similarities in its current characterization to behavioral maladies in the earliest recorded human history – for example, accounts of symptoms of insomnia, loss of appetite, and

impaired concentration in ancient Babylon (27). Perhaps one of the closest conceptual ancestors to our modern conception of depression is the ancient Greek idea of melancholia, though as described by Hippocrates it is substantially more pervasive than contemporary depression (28). Similar syndromes can be found in most periods and regions of human society, but much of the explanation was spiritual and religious in origin, claiming possession and witchcraft as causes. Eventually, theories about what we now called depression returned to roughly correspond to a conception of mental illness, the shift being best captured in what was the first concerted effort to detail the risk factors for and treatment of depression: *The Anatomy of Melancholy* (29). While the tome is at least as much literary as scientific, its breadth reflected the diverse and multifactorial nature of melancholy as it was conceived then.

The transition from the diagnosis of ‘melancholia’ to the term ‘depression’ was in large part due to Emil Kraepelin, in his theoretical construct of manic-depressive insanity (including specifically depressive states) (30). Perhaps the fullest expression of the Kraepelinian approach to diagnoses via syndrome or longitudinal patterns of symptoms was realized in the eventual codification of the operationalized criteria for the diagnosis of psychiatric disorders into such reference texts as the *Diagnostic and Statistical Manual of Mental Disorders* (31, 32), which shares many common origins and similarities in content to the psychiatric portion of the *International Classification of Diseases* (33). DSM-IV criteria for major depression (or DSM-IV-TR criteria, which are the same) were administered in the datasets used for the present analysis to establish the MD diagnoses – see Tables 1 and 2.

Table 1. DSM-IV / DSM-IV-TR Criteria for Major Depressive Episode

<p>A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) Depressed mood or (2) loss of interest or pleasure. Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.</p>	
	(1) Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) 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 made by others)
	(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 gains.
	(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
<p>B. The symptoms do not meet criteria for a Mixed Episode (see p. 335*).</p>	
<p>C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.</p>	
<p>D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism).</p>	
<p>E. The symptoms are not better accounted for by Bereavement, i.e., after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation.</p>	

* [Description of a Mixed Episode]

A. The criteria are met both for a Manic Episode and for a Major Depressive Episode (except for duration) nearly every day during at least a 1-week period.

B. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.

C. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).

Note: Mixed-like episodes that are clearly caused by somatic antidepressant treatment (e.g., medication, electroconvulsive therapy, light therapy) should not count toward a diagnosis of Bipolar I Disorder.

Table 2. DSM-IV / DSM-IV-TR Criteria for Major Depressive Disorder, Recurrent

A. Presence of two or more Major Depressive Episodes. Note: To be considered separate episodes, there must be an interval of at least 2 consecutive months in which criteria are not met for a Major Depressive Episode.
B. The Major Depressive Episodes are not better accounted for by Schizoaffective Disorder and are not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified.
C. There has never been a Manic Episode, a Mixed Episode, or a Hypomanic Episode. Note: This exclusion does not apply if all of the manic-like, mixed-like, or hypomanic-like episodes are substance or treatment induced or are due to the direct physiological effects of a general medical condition.

III. Diagnostic History of GAD:

The origins of anxiety as a cognitive state are perhaps even older than depression, as it is associated with physiological arousal and stress responses in general, such as the fight-or-flight response (34), which is present in most vertebrates and other animals. Anxiety, as a counterpoint to depression, is commonly understood to have some adaptive value at moderate levels, including reducing risk of accidents (35), and even fostering trust in others (36). It is also, as mentioned earlier, intimately related to the human capacity for reasoning, with some even claiming that anxiety is the default state of the human brain (when not suppressed by regulatory signals) (37), and that anxiety evolved in a close relationship with intelligence (38). Nevertheless, too much anxiety is clearly pathological, and this has been known to history for nearly as long as depression's troubles: anxiety symptoms can be seen in the texts of Hippocrates (especially phobia – of flute music at night!) (39), and the Anatomy of Melancholy also contains descriptions of anxiety (29). But the historical conceptions of anxiety did not approach the modern understanding of the condition until more recently.

Neurasthenia, first described by George Miller Beard in 1869, was quite a heterogeneous condition that encompassed many contemporary anxiety symptoms (40), among its many other diverse symptoms. Anxiety also formed a basis for Freud's theories, though he refined his definition of anxiety over time (41). Kraepelin wrote on anxiety as well, calling it the most frequent of all abnormal distressing effects (42), though he did not write extensively on anxiety as a separate diagnoses from other conditions. After the first two iterations of the DSM focusing on psychodynamic theory, with anxiety disorders being designated primarily by the terms 'psychoneurotic

disorders' and 'neuroses' (43), the creation of GAD as an individual disorder in DSM-III came about due to differences in psychopharmacological treatment (44). While there have been changes to the diagnoses since its inception, primarily with regards to minimum duration and hierarchy with other disorders, the core of the condition: chronic, context-agnostic anxiety and worry that is difficult to control, persists. See Table 3 for the DSM-IV / DSM-IV-TR criteria used in this study.

Table 3. DSM-IV / DSM-IV-TR Criteria for Generalized Anxiety Disorder

A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).	
B. The person finds it difficult to control the worry.	
C. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms present for more days than not for the past 6 months). Note: Only one item is required in children.	
	<ul style="list-style-type: none"> (1) Restlessness or feeling keyed up or on edge (2) Being easily fatigued (3) Difficulty concentrating or mind going blank (4) Irritability (5) Muscle tension (6) Sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep)
D. The focus of the anxiety and worry is not confined to features of an Axis I disorder, e.g., the anxiety or worry is not about having a Panic Attack (as in Panic Disorder), being embarrassed in public (as in Social Phobia), being contaminated (as in Obsessive-Compulsive Disorder), being away from home or close relatives (as in Separation Anxiety Disorder), gaining weight (as in Anorexia Nervosa), having multiple physical complaints (as in Somatization Disorder), or having a serious illness (as in Hypochondriasis), and the anxiety and worry do not occur exclusively during Posttraumatic Stress Disorder.	
E. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.	
F. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hyperthyroidism) and does not occur exclusively during a Mood Disorder, a Psychotic Disorder, or a Pervasive Developmental Disorder.	

IV. History of Neuroticism:

Neuroticism is presently understood as a foundational factor in an individual's personality, associated with the frequent occurrence of distressing emotions such as anxiety, guilt, anger, envy, frustration and sadness (45). Its origins, (and those of other theories of defining personality features) can again be traced to Hippocrates, who developed the (even more ancient) idea of four primary temperaments (or humors) into a medical theory, though not all of the writings attributed to him include the theory (46). The four temperaments included sanguine, choleric, phlegmatic and melancholic, corresponding to four substances believed to be present in the human body (29). The traits of the melancholic temperament: despondent and serious, bear the most resemblance to modern neuroticism.

The most well-known user of the term neuroticism is likely Hans Eysenck, who introduced a two-factor model of personality in his work *The Dimensions of Personality* (47) – this was later expanded to the more well-known three-factor model (48). He also made reference to the ancient four-temperament theory, by noting that combinations of neuroticism and extraversion (his other initial personality factor) produced similar results – e.g. high neuroticism and low extraversion corresponded to the melancholic temperament. These personality traits, like temperaments, were considered to be inherent features of the individual, and essentially stable over time. While other personality theories later gained comparable acclaim, such as the five-factor model e.g. (49), Eysenck's neuroticism appears largely unchanged (as does extraversion) in most of them, reflecting the relevance of these dimensions in describing personalities. While

neuroticism, as a personality trait, may seem an odd inclusion in this investigation of MD and GAD, the connection is actually quite close; high levels of neuroticism are associated with a range of psychopathological outcomes (50), including all mood disorders, anxiety disorders, and substance use disorders. Indeed, Eysenck's conception of neuroticism may very well have been an effort to characterize a latent continuum of liability to affective disorders. The relationship of MD and GAD to neuroticism has fittingly been theorized to be due to common etiology, for instance, there is evidence that latent "genetic factors underlying individual differences in neuroticism exhibit significant overlap with the genetic risk for major depression, generalized anxiety disorder, panic disorder, and the phobias" (51). Eysenck anticipated this, too, having conducted research that showed that when a number of different assessments of various aspects of neuroticism were given (which successfully discriminated 'normal and neurotic children'), the general factor explaining overall variance was found to have a heritability of 0.81, suggesting neuroticism was mostly genetic (52, 53). This is a higher figure than more recent biometrical estimates, however, summarized in section VIII below. See Table 4 for some sample items used to measure neuroticism in this study.

V. Epidemiology of MD:

As mentioned in Section I, MD is one of the most prevalent single mental disorders, with lifetime prevalence in most countries falling between 8-12% (54). One of the most consistent findings in its assessment is a much higher prevalence in women than in men, throughout the world (55). The cause of this sex difference may be partly

physiological, but there is also evidence that structural gender discrimination such as the wage gap between men and women in the United States of America (USA) contributes to increased rates of MD in women (56). Lifetime prevalence is highest in the USA at 17% (54), but is difficult to assess reliably in some other countries, such as China. MD in China shows lower estimates from 3.6% (57) to 6.5% (58), depending on the region surveyed, though there is some evidence of underreporting in MD in China due to a cultural tendency to deny depression or focus on physiological symptoms (59). Therefore, it is likely that the true prevalence is somewhat higher, around 8% in Chinese women.

It is also clear that there are a number of risk factors with a powerful impact on the risk of developing MD. Perhaps the strongest are social relationships and stressful life events. The connection between romantic relationships and depression seems intuitive. That is, being in a fulfilling relationship is a goal for many people, and indeed marriage (the most well-studied type of romantic relationship) is long known to be protective against depression (60). Even among college students, being in a committed dating relationship seems to have some protective effect against psychopathology (61), though the impact in that particular study was not significant for depressive symptoms in males.

Stressful life events, a catch-all term for common generally aversive experiences, such as death of a loved one, physical or sexual abuse, romantic separation or financial difficulties, are highly associated with the onset of depression (62). There are differential effects on how depressogenic (depression-causing) an individual event may be however, based on context, and what type of impact it may have; events that involve

loss or humiliation tend to be the most depressogenic (63). Certain severe events also qualify as traumatic events, and, unfortunately, these are quite common. The majority of the world's population, as well as a majority of most countries worldwide have experienced at least one traumatic event, though the precise prevalence varies from country to country, such as 52.5% of individuals in China, to 82.7% of individuals in the USA (64).

There are, of course, many other risk factors for MD, including low socioeconomic status (SES) or poverty (65), use of some psychoactive substances, such as alcohol (66) and methamphetamines (67), and more, but the effects of social support and life stress are among the strongest predictors of MD.

Table 4. Sample Items from Neuroticism Scales in CONVERGE* and S4S**

Sample:	Items:
CONVERGE*	Would you call yourself a nervous person?
	Do you ever feel "just miserable" for no reason?
	Does your mood often go up and down?
S4S*	I see myself as someone who is relaxed, handles stress well. [Reversed]
	I see myself as someone who worries a lot.
	I see myself as someone who is emotionally stable, not easily upset. [Reversed]

*China, Oxford, and VCU Experimental Research on Genetic Epidemiology – Items presented are English translations from the Mandarin Chinese original items. Yes/no responses.

** Spit for Science, the VCU Student Survey – Items rated as 1-5 for agreement with the statement (from 'disagree strongly' to 'agree strongly')

VI. Epidemiology of GAD:

GAD is less common than MD, but its often chronic nature renders it a lifetime burden to many who suffer from it. In the USA, lifetime prevalence is 4.2-12.7% depending on required minimum duration, though notably a number of clinical features (such as onset, persistence, severity, and comorbidity patterns) are similar between shorter duration cutoffs and the DSM-IV duration requirement of 6 months (68). Like MD, anxiety disorders such as GAD are about twice as prevalent in females as males, and are less common among individuals in Asian cultures than in Western cultures (69). The sex difference in GAD appears to be partially mediated by wage disparity between men and women in the USA, if to a somewhat less pronounced degree than in MD (56). The prevalence of GAD in China is estimated at slightly over 1% (70), though again, there is evidence of a certain reluctance in Chinese individuals to disclose psychological symptoms during face-to-face interviews (71).

Risk factors for GAD are quite similar to those of MD; romantic relationships such as marriages are protective against GAD, though this effect is proportional to the quality of the marital relationship, and relationships with other family members are still important as well (72). Stressful life events, such as natural disasters, childhood abuse, and family problems, seem to have a strong impact on the onset and maintenance of GAD (73). The events that have elements of loss or danger are particularly likely to predispose to GAD (63), which is intuitive when you consider the defining feature of worry – being exposed to danger or threat is clearly anxiogenic. As mentioned in the last section, traumatic events such as these are far from rare, with a majority of people worldwide reporting at least one traumatic experience (64). Substance use can also

contribute to GAD, including common stimulants such as nicotine (74), and caffeine (75).

VII. Relationship of Neuroticism to anxiety and depressive disorders:

Neuroticism is perhaps the most frequently assessed personality trait, and with good reason. Levels of this trait are associated with psychopathological outcomes quite strongly (50), and could therefore be used as a trait-level index of susceptibility to internalizing disorders. It is not, however, to be properly thought of a 'risk factor' for internalizing psychopathology in the usual sense, as there is strong evidence that common etiological origins are shared between neuroticism and such disorders as MD and GAD. Firstly, the conceptual similarity between neuroticism and the disorders of MD and GAD is evident even in their composite measurement instruments (compare diagnostic criteria from Tables 1-3 with sample neuroticism items in Table 4). Latent additive genetic factors contributing to neuroticism appear to account for a sizeable portion of the genetic variance of most anxiety and depressive disorders (51). Latent unique environmental factors have also have been shown to overlap between neuroticism and MD, as well as all anxiety disorders save for specific phobias (51). This indicates that, rather than being a separate temperamental characteristic whose presence aggravates the psychological equilibrium of a person, and tends to plunge them into depression or anxiety, it instead more likely constitutes a portion of whatever baseline vulnerability there may be to internalizing psychopathology. In other words, neuroticism is better classified as 'diathesis' than 'stress' (76).

However, investigation of neuroticism as a separate construct from mood and anxiety disorders may be useful, as levels of the trait are often fairly stable from early life, and therefore may precede the onset of many disorders, including not just psychiatric ones but also such medical conditions as cardiovascular disease (77). One theory (which dates at least to Eysenck himself), suggests that neuroticism is the dispositional manifestation of the level of activity in the limbic system, the reactivity of the sympathetic nervous system, and corresponds to sensitivity to environmental stimuli (78) Indeed, neuroticism has been found to prospectively predict the onset of mood and anxiety disorders (more so than other correlated psychiatric outcomes, such as substance use disorders), and especially comorbid mood and anxiety disorders (79).

VIII. Biometrical Modeling of MD, GAD and Neuroticism:

The high prevalence, and debilitating effects of MD and GAD have led many researchers to seek better understanding of the contributions of different types of factors to their variation in the populace. Neuroticism, as a personality trait easily measured via self-report, is also well-studied. Due to the evidence of common associated factors between the three, there have also been efforts to characterize the degree of overlap. Perhaps the greatest contributions to the quantification of latent factors in these constructs are from the twin methodology, comparing trait correlations among monozygotic and dizygotic twin pairs to assess the contributions of latent additive genetic factors to the trait.

It is worth noting that an implicit assumption in many conceptions of MD and GAD is that of a liability model (80), where there is an continuous distribution of

propensity to MD and another distribution of propensity to GAD. Under this model, disease occurs when a person's liability score reaches a certain threshold in that distribution, but there is nevertheless some (recondite) significance in smaller increases in liability, often in terms of an increase in risk.

The heritability of MD as assessed by twin studies is approximately 0.42 in women of European ancestry, and 0.29 in European men, with a genetic correlation of 0.63 between the sexes (indicating a portion of genetic influences acting in a sex-limited fashion), and when constrained to equality, the estimated heritability in both sexes together was 0.38 (81). These estimates differ under certain circumstances, for instance in densely affected pedigrees with high risk of MD, heritability has been estimated at 0.67 (82). The heritability of MD in other populations, such as the Chinese population, is understudied in twin work, though there is some evidence that heritability is comparable between European and East Asian populations (83).

Heritability of GAD is often estimated in about the same range, at about 0.32 (84), though this appears not to differ between the sexes as the MD estimate does (85). There is evidence, however, of a threshold difference between men and women, however (86), suggesting that, though the contributions of genetic factors may be similar in both sexes, the level of liability required to develop GAD is lower in women than in men, which is consistent with the higher prevalence of GAD in women (69). Again, GAD is understudied in populations of non-European descent, and therefore twin estimates of GAD heritability in other populations such as that of China are not readily available.

Neuroticism, though a personality trait rather than a psychiatric disorder, is frequently estimated via twin studies as having a heritability of about the same magnitude as MD and GAD, or slightly higher, ranging from about 0.40-0.60 (87). It does not seem as though there are appreciable sex differences in the heritability of neuroticism, however, unlike MD (88). Neuroticism, like GAD, has not been the subject of sufficient biometrical research in non-European populations, so estimates of heritability in Chinese populations are unavailable.

Given the high comorbidity between MD and GAD, as well as the strong correlation between both disorders and neuroticism, it may perhaps be expected that there is overlap in the latent factors involved in their etiology. Most biometrical studies of MD and GAD find a correlation in their genetic factors indistinguishable from unity (86, 89-91), though at least one has found an estimate of high (0.74 genetic correlation) but not total overlap in males (91). Notably, the latent environmental correlation is more modest, and seems to be higher in males at 0.65 (91), than females at 0.40 (91, 92).

Neuroticism also seems to display some common genetic and environmental influences with MD and GAD. Latent genetic correlation with neuroticism is estimated at approximately 0.2 with MD, and 0.24 with GAD (51). Latent unique environmental correlation with neuroticism is somewhat smaller, at about 0.15 with MD, and 0.16 with GAD (51). This may seem surprisingly low, compared with the strong phenotypic relationship between neuroticism and both MD and GAD, but notably, it has been estimated that there is no latent genetic variance in neuroticism that is unshared with mood and anxiety disorders – i.e. all of neuroticism's genetic influences also predispose to MD, GAD or other anxiety disorders (51).

IX. Genotypic studies of MD, GAD and Neuroticism:

More recently, investigations into these disorders and neuroticism have begun using measured genetic information as technology and statistical methods develop in concert. To date, only a relatively small amount of understanding has been gained through these sorts of investigations, but the results suggest that progress will be incremental in these (and most) psychiatric outcomes of interest. Future discovery of any as yet unknown single genetic variants (or single genes) explaining a majority of trait variance in any of MD, GAD or neuroticism is, based on existing evidence, unlikely. Nevertheless, important insights about biological mechanisms of these disorders are more often the goal of these studies, and in this respect there has been some success.

Perhaps the most direct interrogation of the human genome for etiologically-relevant variation is the genome-wide association study (GWAS). In brief, this compares individuals' unique genotypes at a great many polymorphic sites (loci where multiple alleles, or alternative nucleotide sequences for the same portion of a chromosome) in the genome to their phenotypic values at whatever the trait of interest may be. If there is a statistical relationship that is sufficiently strong, and it is borne out in subsequent investigations (called replication), the relevant variant may be considered to be a marker of a genomic region likely to be involved in the outcome. GWAS have been performed on MD, GAD, and Neuroticism many times, but for various reasons, only in the last couple years have there been truly promising results.

Notably, the first genome-wide significant (after multiple testing correction for the vast number of polymorphic variants included in the average GWAS) loci for MD were

found in one of the samples used in this paper; the all-female Han Chinese CONVERGE sample (93). GAD has been examined in GWAS less frequently, but one recent study found a variant that reached significance for a GAD symptom score in the primary analysis, but did not replicate (94); this study also was conducted on one of the relatively few samples composed of Latino individuals. Neuroticism has had somewhat better success, with a large meta-analysis finding 11 associated variants, including two that were attributable to previously characterized inversions (a type of genomic rearrangement in which the normal sequence order is reversed; these are especially difficult to detect using standard mapping methods), due in large part to greatly increased sample size over most similar studies (95).

Due to the difficulty and expense in assessing clinical disorders such as MD and GAD in a large number of individuals as GWAS requires for adequate power, many researchers have opted to evaluate genetic associations for related phenotypes with close conceptual and statistical relationship with the psychiatric disorders. In the same study that found the neuroticism loci above, the outcome of depressive symptom count was also tested, resulting in 2 independent variants (95). Another study used anxiety sensitivity, a dispositional fear of arousal sensations, as a proxy for GAD, finding a significant result despite a comparatively small sample size (96). An alternative strategy involves pooling of multiple related outcomes to maximize genetic variance, resulting in a significant result for an 'any anxiety disorder' dichotomous phenotype, and another for a factor-analytic anxiety score in a meta-analytic study assessing not only GAD, but also phobias and panic disorder (97). A brief summary of genes that contain associated variants with any of these outcomes is listed for reference in Table 5.

Table 5. Previous Genetic findings from GWAS in MD, GAD, and Neuroticism

Outcome	Genes	Variants	Reference
MD (CONVERGE)	SIRT1 LHPP	rs12415800 rs35936514	(93)
MD (23AndMe)	TMEM161B VRK2 L3MBTL2 NEGR1 MEF2C RERE HACE1 / LIN28B SORCS3 OLFM4 PAX5 None None MLF1 None None None None	rs10514299 rs1518395 rs2179744 rs11209948 rs454214 rs301806 rs1475120 rs10786831 rs12552 rs6476606 rs8025231 rs12065553 rs1656369 rs4543289 rs2125716 rs2422321 rs7044150	(98)
GAD Symptoms	THBS2	rs78602344*	(94)
Neuroticism	LINC00529 KANSL1 None None SBF2 CELF4 PAFAH1B1 None LOC102724048 None LINGO1	rs2572431 rs193236081 rs10960103 rs4938021 rs139237746 rs1557341 rs12938775 rs12961969 rs35688236 rs2150462 rs12903563	(95)
Depressive Symptoms	KSR2 DCC	rs7973260 rs62100776	(95)
Anxiety Sensitivity	RBFOX1	rs13334105	(96)
Any Anxiety Disorder	LOC152225	rs1709393	(97)
Anxiety Factor Score	CAMKMT	rs1067327	(97)

* – result did not replicate.

In addition to GWAS, a genotypic analogue to twin-based heritability estimation is possible through the use of such techniques as Genomic Complex Trait Analysis (99) or GCTA, and Linkage Disequilibrium Score Regression (100) or LDSC. These methods use different strategies to estimate the overall level of variance attributable to additive genetic effects in measured genetic variants, and can also estimate genetic correlation between pairs of traits. See Chapter 4 for a more detailed explanation of each. MD heritability has been estimated at 0.41 (transformed to the liability scale) using LDSC (100), which is comparable to twin estimates. This last finding is somewhat anomalous, however, given the estimate is comparable to, or perhaps in excess of the coed twin estimate of 0.38 for MD heritability (81); the twin estimate should represent an upper bound on the range of snp-based heritability for a given trait, due to including, in addition to all additive genetic effects of any magnitude (101), any unmodeled dominance or epistatic effects (102), interaction between additive genetic effects and shared environmental effects, and correlation between additive genetic effects and unique environmental effects (103). Also see the discussion of Chapter 4.

GAD has not notably been assessed thus far with these methods in its dichotomous diagnostic form, but a GAD symptom score has demonstrated a low heritability of 0.07, using a method similar to GCTA (94). Neuroticism appears to have a low heritability as estimated by these methods as well, about 0.09 (using LDSC) in European populations (95), and 0.10 (using GCTA) in the Chinese population (104). Similarly meager estimates are found for depressive symptoms using LDSC, at 0.05

(95), the any anxiety disorder outcome using GCTA, at 0.14, and the anxiety factor score, also using GCTA, at 0.11 (97).

While the differences between empirical estimates of heritability generated by GCTA and LDSC and latent estimates produced by biometrical studies are troubling to the extent that they are both designed to estimate 'additive genetic heritability', they properly include different sets of genetic effects. As mentioned above, the twin estimates include the effects of rare variants, genetic interactions, and gene-environment interplay, among other sources of variance, whereas the genotype-based methods include only the genetic effects of variants provided as input. Therefore, the heritability estimates provided by GCTA and LDSC are often lower bounds of true narrow-sense heritability (variance due to all and only additive genetic effects). Essentially, the heritability from genotypic methods is, generally, 'narrower' than narrow-sense heritability, and biometrical heritability estimates are often 'broader' than narrow-sense heritability.

Even in the absence of sufficient sample size to estimate genetic correlation using measured genotypes, which can be substantial for both LDSC and GCTA, genetic relationships between phenotypes can be evaluated using polygenic scoring methods, such as the Purcell method (105), which prunes association results for variants using linkage disequilibrium (LD) to select a set of independent loci that are filtered by significance to predict an outcome, and LDpred (106), which uses a Gaussian mixture prior parameter (proportion of causal effects) and empirical estimation of heritability to infer the posterior mean effect sizes (weights) of variants using a Markov Chain Monte Carlo Gibbs Sampler. Early polygenic scoring work suggested MD risk scores could

predict anxiety symptoms to some degree (107), and more recently, neuroticism scores have successfully predicted MD, even using a European discovery neuroticism sample to predict into a Chinese validation sample (104). While virtually all polygenic prediction results explain minimal fractions of the total trait variance in any outcome, this is a still a fruitful area for meaningful contributions to be made by studies with insufficient sample sizes for genotypic heritability estimation.

X. Scope of dissertation project:

The overall story of MD and GAD seems to suggest they both developed in concert over the course of human evolution, and have been present in various guises throughout recorded history. While similarities are present between the two disorders in both specific and latent risk factors, including genetic and environmental influences, it is unclear to what extent the genetic elements are distinguishable between MD and GAD. In addition, the majority of psychosocial risk factors associated with the disorders act similarly on both (for instance, strong relationships are protective, stressful life events increase risk), though there is evidence of discriminatory effect for certain carefully measured environmental experiences, such as the specific consequences of certain types of stressful life events; events with elements of danger are especially anxiogenic, humiliation events are especially depressogenic, while loss events predispose to both disorders (63). Notably, neuroticism appears to be a relatively stable dispositional trait that constitutes a portion of the shared variance of MD and GAD.

To characterize the individual and communal influences on MD, GAD and neuroticism, the remaining chapters of this dissertation will address a comprehensive

range of psychosocial, genetic and genomic analyses conducted using the datasets available, which have overlapping strengths and limitations.

The China, Oxford, and VCU Experimental Research on Genetic Epidemiology (CONVERGE) Study is a clinically-ascertained genetic study designed to investigate Major Depression. It is composed of 5,303 recurrent MD cases and 5,337 screened controls, all women with full Han Chinese ancestry. There are extensive questionnaires in this study, but MD cases and controls received different assessments; the most important consequence of this design in terms of this project is that GAD is only assessed in MD cases. The Molecular Genetics of Schizophrenia (MGS) study is a case-control study of Schizophrenia, but has the useful quality of having full psychopathological assessment information on all their European-ancestry controls, which constitute the portion of the sample used for this study. The study does not include information about other sub-clinical phenotypes, however, such as social relationships, neuroticism or stressful life events.

Spit for Science (S4S), the Virginia Commonwealth University Student Survey, is a cohort sequential study of undergraduate students at VCU. Data in this project comes from cohorts 1, 2 and 3. There is a wide range of information on the participants, and there are individuals of both European and East Asian ancestry (among others), but since the population comprises emerging adults who are attending college, there is low prevalence of clinically significant disorder, and therefore the majority of relevant information is on risk factors and continuous correlates of MD and GAD.

For brief summary details on each sample, please refer to Table 6.

In Chapter 2, comparative associations between well-established risk factors for MD, GAD and neuroticism will be assessed to determine which are selectively predictive for one outcome over another, and which seem to have similar impact on multiple outcomes.

In Chapter 3, leveraging the three samples introduced above, results will be displayed from gene discovery efforts designed to determine if any single variant or genomic interval shows a statistically notable association with MD, GAD or neuroticism, and, subsequently, whether significant associations for one outcome are suggestively significant for the others as well.

In Chapter 4, the findings of biometrical genetic studies will be tested, attempting to estimate genotypic heritability of MD, GAD and neuroticism and genetic correlation between them. Polygenic scores will be used to assess the degree to which interprediction between outcomes is successful. A specialized method will assess the degree of heterogeneity of GAD-within-MD and MD-within-GAD to check for asymmetric relationships.

Identification of risk factors constituting the risk profiles of MD, GAD and neuroticism with increasing specificity enables a progressive account of the extent to which they overlap in each etiological domain, including both genetic and environmental.

Table 6. Sample characteristics of datasets

Sample	N	Phenotypic Assessment	Genotype or Sequencing platform	Ref.
CONVERGE	10,099	CIDI (108), VATSPSUD (109), BFI (110)	Illumina HiSeq	(111)
MGS	2,612	CIDI-SF (112)	Affymetrix 6.0	(113)
S4S	7,532	SCL-90 (114), BFI (110), LEC(115)	Affymetrix Biobank	(116)

CIDI: Composite International Diagnostic Interview; VATSPSUD: Virginia Adult Twin Study of Psychiatric and Substance Use Disorders; BFI: Big Five Inventory; CIDI-SF: Composite International Diagnostic Interview, Short Form; SCL-90: Symptom Checklist 90; LEC: Life Events Checklist.

Chapter 2: Exploration of psychosocial risk factors in MD, GAD, and neuroticism

I. **Methods:**

Samples: Psychosocial risk factors are available in CONVERGE and S4S samples, and include marital status, social support, and stressful life events in CONVERGE, as well as relationship status, relationship satisfaction, involvement in social activities, social support, and stressful life events in S4S. For S4S, there are 3016 individuals of European ancestry (EUR), 557 individuals of East Asian ancestry (EAS), 2929 males, and 4603 females in cohorts 1-3, which were the subset included in these analyses. S4S cohort 1 had responded to 3 years of survey follow-up, cohort 2 had responded to 2 years of survey follow-up, and cohort 3 had responded to 1 year of survey follow-up when data was assembled for these analyses.

Variables: Outcomes in CONVERGE include BFI neuroticism (110), which was standardized, MD status, and a dichotomous outcome of MD+GAD comorbidity versus MD controls – a ‘case-control’ test of MD+GAD status. GAD is not assessed in controls, but due to the substantial screening for MD in CONVERGE, and the high frequency of comorbidity between MD and GAD, the number of MD controls who have unassessed GAD is likely to be minimal, and if present, should only reduce power modestly.

Outcomes in S4S include an abbreviated 3-item BFI neuroticism score (110), and a sum score of 4 SCL-90 depressive symptoms (114), both of which were standardized. The psychometric properties of the shortened scales retain passable reliability compared to extended scales. Alpha levels of the shorter and longer (8-item) neuroticism scales are 0.695 and 0.813, respectively. Alpha levels of the shorter and longer (9-item) depressive symptom list are 0.917 and 0.94, respectively. Unfortunately, the shorter scales were the only assessment given to the majority of the sample, so choice is limited.

Predictors in CONVERGE include demographic information on marital status (choices included: 'Married', 'Separated', 'Divorced', 'Widowed', and 'Never married'), which was made into a marriage termination variable of 'Divorced/Separated' versus 'Married'; a social support scale (117), which was standardized; a stressful life events checklist, adapted from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD) including age of occurrence (109), which was divided into childhood (before age 17) and adulthood (after age 17, for this purpose) events, and standardized separately.

The stressful life events variable in CONVERGE was constructed from a count of the different events endorsed by each individual in their lifetime. The events included were: death of a spouse, child or sibling; divorce or long-term separation due to marital difficulties; unemployment for more than one month; fired from a job; major financial crisis; problems with police or required court appearance; serious illness; life-threatening accident; fire, flood or natural disaster; witnessed someone being badly injured or killed; rape; physical assault; physically abuse as a child; serious neglect as a

child; threatened with a weapon, held captive or kidnapped; and any other terrible experience.

Predictors in S4S include an abbreviated relationship satisfaction scale (118); a checklist of participation in group activities, including sports, fraternities/sororities, school-related, community, and church, which was standardized; a social support scale (119), which was standardized; and a checklist of stressful life experiences (115), which were summed for a lifetime count.

The stressful life events variable in S4S was constructed from a count of the different events endorsed by each individual in their lifetime, whether before or during college. The events included were: broken engagement or steady relationship; separation from other loved one or close friend; serious illness or injury; burglarized or robbed; trouble with police; laid off or fired from a job; major financial problems; serious housing problems; serious difficulties at school; someone emotionally close passed away; mother or father had a serious illness or injury; someone else close to you had a serious illness or injury.

Covariates for both CONVERGE and S4S were used to adjust for the effects of demographic factors, when appropriate. Age at assessment was included in all analyses, as well as sex when more than one sex was included in a test (CONVERGE is all female, for instance, so analyses with that sample did not include a sex covariate). SES was adjusted for using educational attainment as a proxy. In CONVERGE, this was directly assessed in the demographic interview, whereas in the S4S sample, given participants are still in college, the educational attainment of their parents was used instead. Finally, for those questions in S4S that were answered more than once in

follow-up surveys, number of occasions is included as a covariate as well, to control for, as an example, the increased exposure time a participant in S4S had to possibly experience a stressful life event in their third year of college compared with someone who only responded to the survey in their first year.

A brief aside is perhaps worthwhile to discuss the decision to consider neuroticism as an 'outcome' in these analyses. While the term 'outcome' is used throughout this manuscript, this should not be taken to imply that neuroticism is only subject to causal effects of other variables and has no causal action itself. 'Predictor' and 'outcome' are effectively useful shorthand terms to denote 'independent variable' and 'dependent variable' in analyses. This arrangement was selected because the primary interest is in the influence of predictive factors on neuroticism score, along with MD and GAD status. This is true throughout the manuscript, though in certain cases the implication is more pernicious. Neuroticism, as described earlier, is a dispositional trait of individuals that has high temporal stability over the life course. One of the key requirements for a causal relationship in the physical universe (so far as we know) is temporality – the cause must precede the effect. While some of the predictors, such as stressful life events, can be traced to a particular moment, others, such as neuroticism, persist diffusely throughout the lifespan, arising in concert with the rest of an individual's personality at some unknown (perhaps unknowable) moment. Temporality of neuroticism in general is not clear.

Moreover, prospective studies have shown that it acts as outcome and predictor both, in different ways. Neuroticism can change to some extent as a result of life experiences, for example, befriending new people tends to lower neuroticism over time,

whereas breaking up a longstanding relationship with a friend or relative increases neuroticism (120). Conversely, high neuroticism individuals have a higher likelihood of experiencing a variety of stressful life events (121). Therefore, while the terms 'predictor' and 'outcome' will be used throughout, it is with the understanding that causal links may very well be present in both directions, and the statistical correlation is the central interest, rather than directionality.

Analyses: Hierarchical regressions were conducted on each combination of predictor and outcome in CONVERGE and S4S. Initially, only predictor and outcome were included in the regression, followed by a step in which covariates were added, and finally multivariate analyses including all predictors in the sample to estimate overall fully-adjusted effects. Given past literature's suggestions of sex differences in the relationship between risk factors and MD, GAD and neuroticism (61, 122, 123), and potential differences in presentation between European and East Asian populations (59), the S4S analyses were also performed stratified by sex and ancestry. The ancestry groups were limited to EAS and EUR for (broadly) within-ancestry comparison of results between EAS and CONVERGE, and between EUR and MGS (in other chapters). These analyses were tests of theory derived from past literature, and concerned more with distinguishing effects between groups and outcomes than on significance, so multiple testing correction was not applied.

II. Results:

Results from the psychosocial predictors in CONVERGE can be found in Figures 1-4. In progressive models with added covariates and multiple predictors the effect

sizes are attenuated but still robustly significant. This attenuation is expected due to the correlation between variables; for instance, some of the harmful impact of stressful life events on these outcomes may be due to interpersonal conflict, and this same conflict may result in reduced quality of social relationships. The impact on the person's risk for depression will be the same whether this conflict is included in the life events or social predictors. When both are in the same model, this hypothetical event is effectively counted twice, which will lead to weaker relationships between both predictors and the outcome; similar overlap and partial mediation likely explains the other attenuated effect sizes observed in the multivariate results. Since they are likely the least biased, all betas (true regression coefficients) discussed in text are from the multivariate models.

Marital termination, childhood and adulthood stressful life events were all associated with increased neuroticism – multivariate model beta, (95% Confidence Interval [CI], p-value): 0.23, (0.16 – 0.31, 7.84e-10); 0.15, (0.13 – 0.17, 2.35e-48); 0.19, (0.17 – 0.21, 7.18e-69), respectively; likelihood for MD – beta, (95% CI, p): 0.71, (0.51 – 0.92, 6.41e-12); 0.37, (0.30 – 0.44, 3.27e-25); 0.27, (0.22 – 0.33, 1.46e-23); and comorbid MD+GAD – beta, (95% CI, p): 0.67, (0.39 – 0.94, 1.93e-06); 0.42, (0.33 – 0.50, 1.94e-20); 0.42, (0.34 – 0.49, 3.03e-28). Social support was associated with lower risk of MD and MD+GAD and with lower neuroticism – beta, (95% CI, p): -0.52, (-0.57 – -0.47, 7.89e-98); -0.59, (-0.66 – -0.51, 1.05e-48); -0.22, (-0.24 – -0.20, 2.28e-97). Each psychosocial predictor had a stronger association with risk of MD and MD+GAD than on neuroticism level, regardless of the direction of effect. Stressful life events in adulthood additionally had a stronger impact on risk of comorbid MD+GAD than on risk of MD – beta (95% CI): 0.42 (0.34 – 0.49) vs. (95% CI): 0.27 (0.22-0.33).

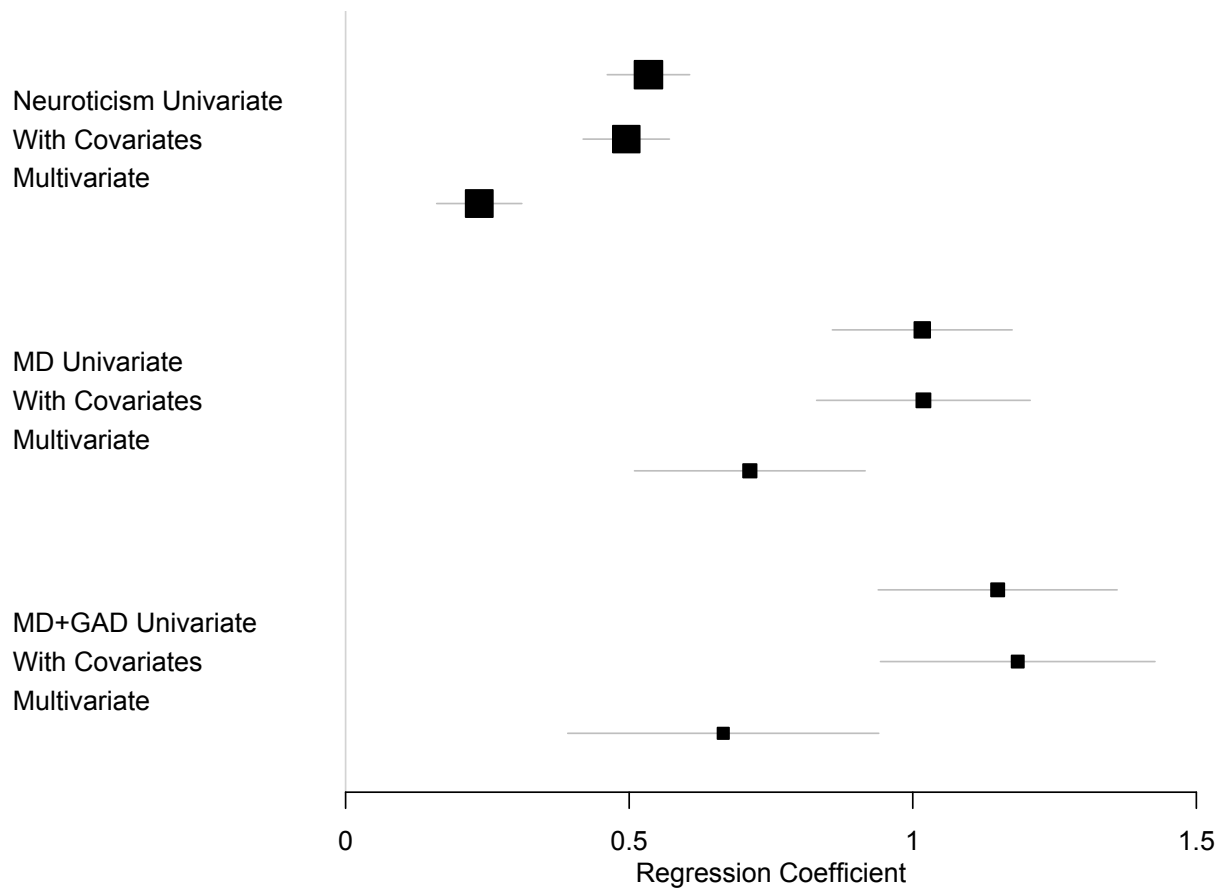


Figure 1. CONVERGE Marital Termination. Left labels indicate outcome and model. Covariates include age and educational attainment. Multivariate model additionally includes social support, childhood stressful life events and adulthood stressful life events.

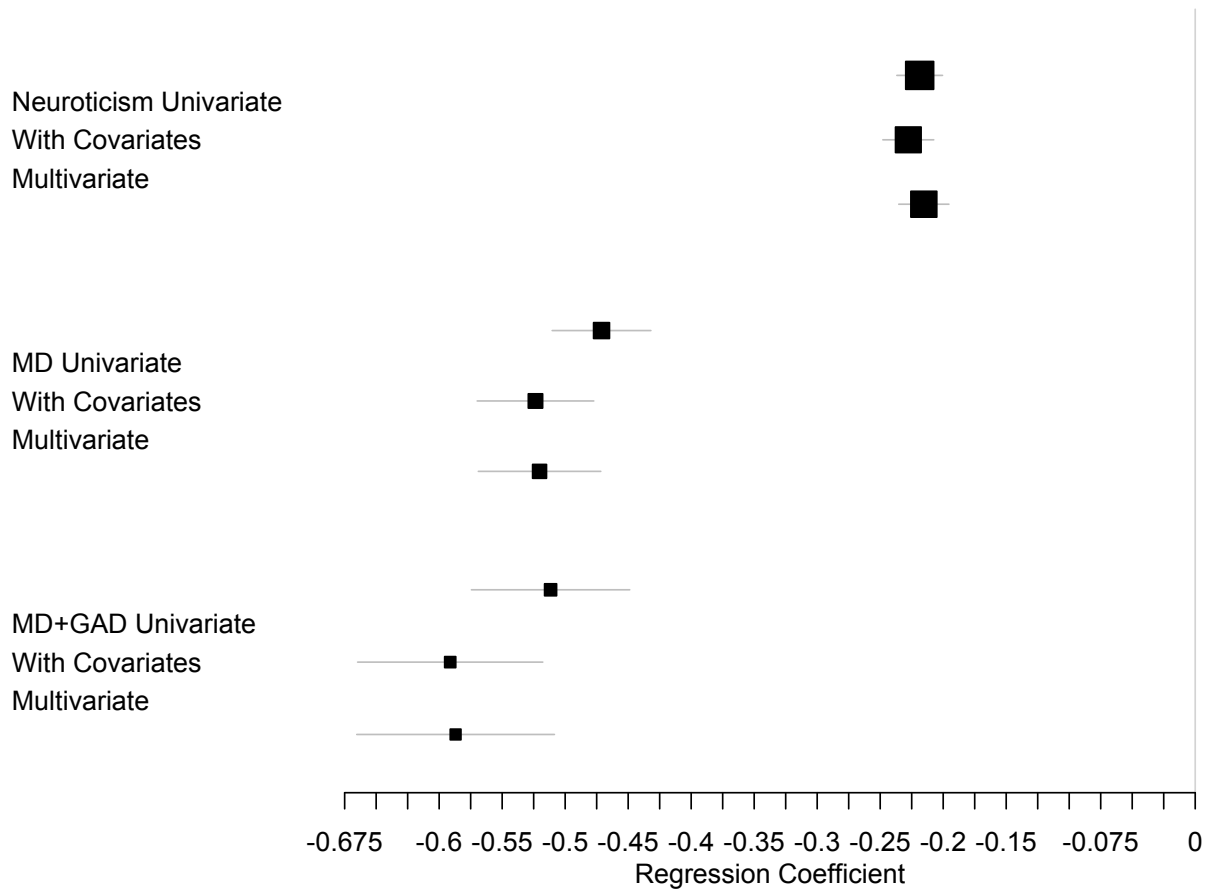


Figure 2. CONVERGE Social Support. Left labels indicate outcome and model.

Covariates include age and educational attainment. Multivariate model additionally includes marital termination, childhood stressful life events and adulthood stressful life events.

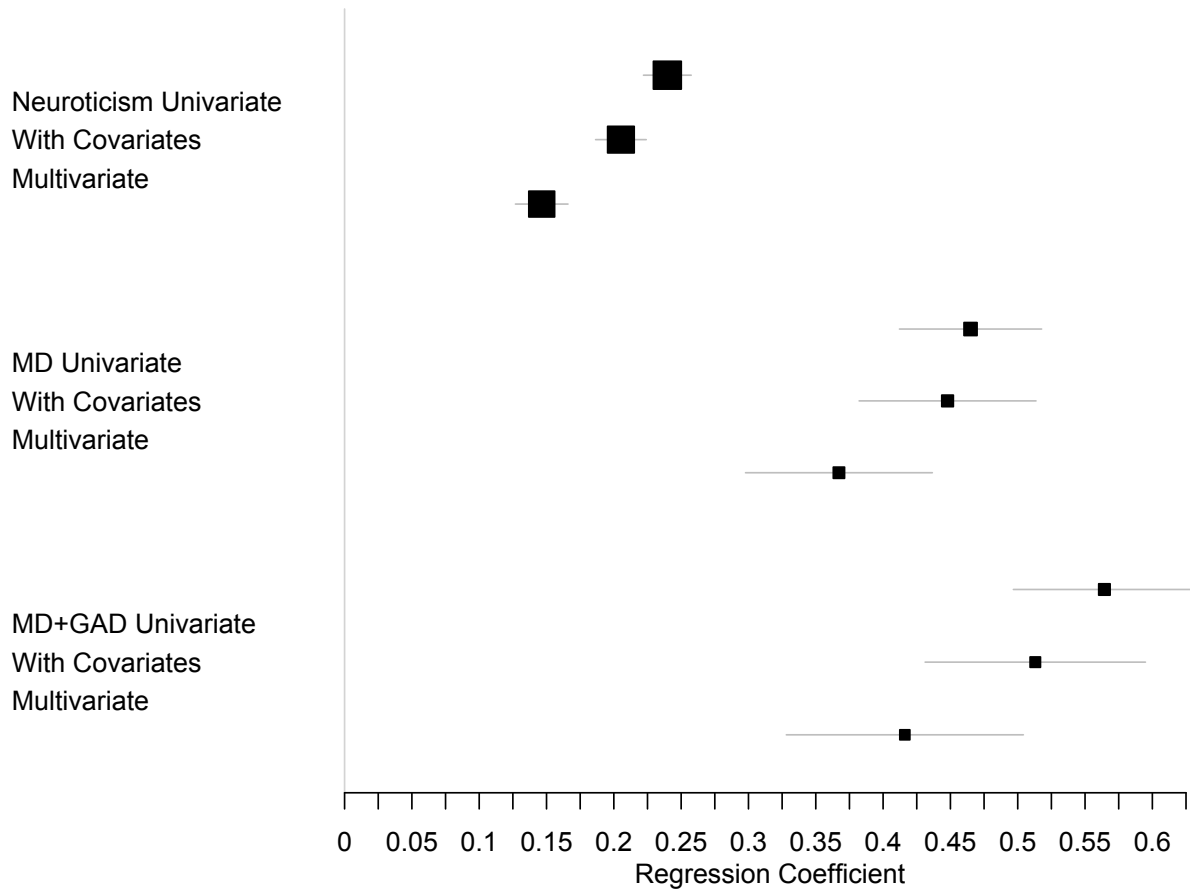
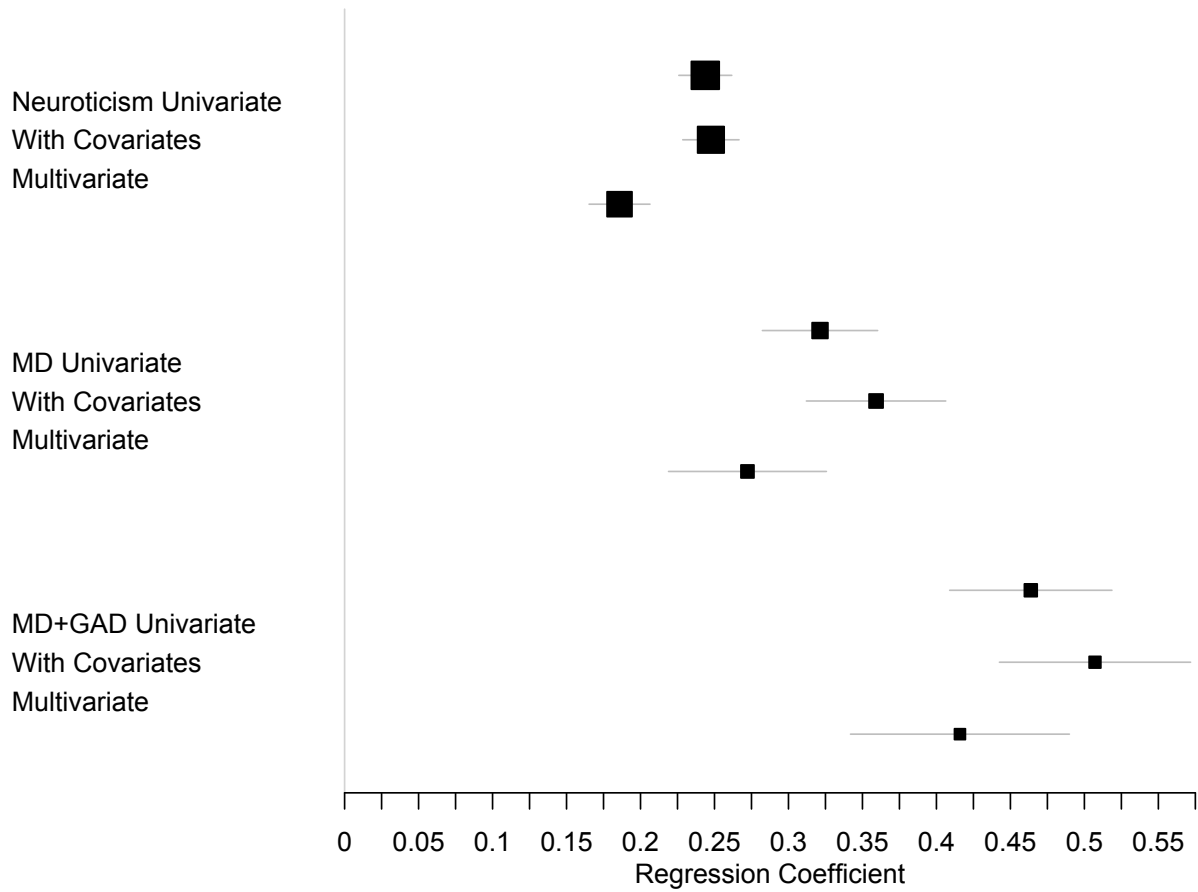


Figure 3. CONVERGE Childhood Stressful Life Events. Left labels indicate outcome and model. Covariates include age and educational attainment. Multivariate model additionally includes marital termination, social support and adulthood stressful life events.



Left **Figure 4. CONVERGE Adulthood Stressful Life Events.** Left labels indicate outcome and model. Covariates include age and educational attainment. Multivariate model additionally includes marital termination, social support and childhood stressful life events.

Results from the psychosocial predictors in S4S can be found in Figures 5-8. Due to the presence of males and females as well as both East Asian and European ancestries, stratified analyses to test differences in the effect of psychosocial variables by sex or ancestry were possible. These stratified analyses are indicated by different colors in Figures 5-8. Note that EAS individuals were represented in the sample much less than EUR individuals in terms of sample size, so statistical power for the EAS analyses is reduced. Commentary about results is confined to the multivariate models, to decrease bias by adjusting for all predictors and covariates simultaneously.

Relationship satisfaction had a small protective effect against depressive symptoms in the multivariate model – beta, (95% CI, p): -0.05 (-0.08 – -0.01, 0.01), which was also present in European individuals as a group, and males – beta, (95% CI, p): -0.06 (-0.12 – -0.00, 0.04); -0.10 (-0.17 – -0.04, 2.90e-3), respectively, but not in East Asian individuals or females – beta, (95% CI, p): -0.03 (-0.20 – 0.14, 0.74); -0.02 (-0.07 – 0.02, 0.34). The EAS effect estimate is comparable to the EUR (-0.03 vs. -0.06), however, so this may be a limitation of sample size. For Neuroticism, while the trend was for relationship satisfaction to be protective, the effect was not significant in the multivariate model – overall beta, (95% CI, p): -0.02 (-0.06 – 0.03, 0.47).

Involvement in social activities was also found to be protective against depressive symptoms – beta, (95% CI, p): -0.14 (-0.18 – -0.10, 8.47e-14), with all subgroups enjoying approximately the same level of protection – beta, (95% CI, p) for EAS, EUR, male and female subgroups, respectively: -0.15 (-0.30 – -0.00, 0.05); -0.11 (-0.17 – -0.06, 3.68e-5); -0.13 (-0.19 – -0.07, 3.74e-5); -0.14 (-0.19 – -0.10, 3.26e-10), though the precision of the EAS estimate is low. Social activities were also negatively

associated with neuroticism – overall beta, (95% CI, p): -0.10 (-0.14 – -0.06, 6.90e-6), in all subgroups but the EAS – beta, (95% CI, p) for EAS, EUR, male, and female, respectively: -0.15 (-0.31 – 0.01, 0.06); -0.09 (-0.15 – -0.03, 5.96e-3); -0.13 (-0.21 – -0.06, 4.6e-4); -0.08 (-0.14 – -0.03, 1.74e-3), again likely due to sample size. The influence of social activities was not notably different between depressive symptoms and neuroticism – betas: -0.14 vs. -0.10.

Social support had a stronger negative association with depressive symptoms than the other protective factors – beta, (95% CI, p): -0.27 (-0.31 – -0.23, 1.31e-35). There were differences in this effect between subgroups, however. In EUR, the effect was quite strong – beta, (95% CI, p): -0.36 (-0.43 – -0.30, 4.33e-27), but there was not a significant association in EAS – beta, (95% CI, p): -0.07 (-0.23 – 0.09, 0.40). The magnitude of association was also greater in females, than in males – beta, (95% CI, p) for female and male subgroups, respectively: -0.31 (-0.36 – -0.26, 5.35e-32); -0.18 (-0.25 – -0.11, 1.63e-6). Like relationship satisfaction, social support had a more modest negative association with neuroticism – overall beta, (95% CI, p): -0.13 (-0.17 – -0.08, 5.9e-7), substantially smaller in magnitude than with depressive symptoms – betas: -0.13 vs. -0.27. The EAS and male groups again had a reduced effect of social support on neuroticism, not reaching significance – beta, (95% CI, p) for EAS and male subgroups, respectively: -0.01 (-0.18 – 0.16, 0.91); -0.04 (-0.13 – 0.05, 0.39), whereas the effects in EUR and females were stronger – beta, (95% CI, p) for EUR and female subgroups: -0.19 (-0.26 – -0.11, 9.52e-7); -0.16 (-0.22 – -0.10, 7.00e-8).

Stressful life events were associated with increased depressive symptoms – beta (0.95% CI, p): 0.11 (0.09 – 0.12, 1.27e-37), and the magnitudes in the subgroups were

largely similar – beta (95% CI, p) for EAS, EUR, male and female subgroups, respectively: 0.12 (0.06 – 0.18, 1.25e-4); 0.09 (0.07 – 0.12, 3.50e-13); 0.09 (0.07 – 0.12, 3.83e-11); 0.12 (0.10 – 0.14, 1.05e-28). Stressful life events were also associated with neuroticism – beta (95% CI, p): 0.05 (0.03 – 0.07, 1.72e-8), but less strongly than they were with depressive symptoms – betas: 0.05 vs. 0.11, respectively. The effect of stressful life events on neuroticism in EAS was not significant, but the other subgroup displayed similar levels of effect – beta, (95% CI, p) for EAS, EUR, male and female subgroups, respectively: 0.01 (-0.05 – 0.08, 0.65); 0.05 (0.02 – 0.08, 3.31e-4); 0.05 (0.02 – 0.08, 3.06e-3); 0.06 (0.03 – 0.08, 9.94e-7).

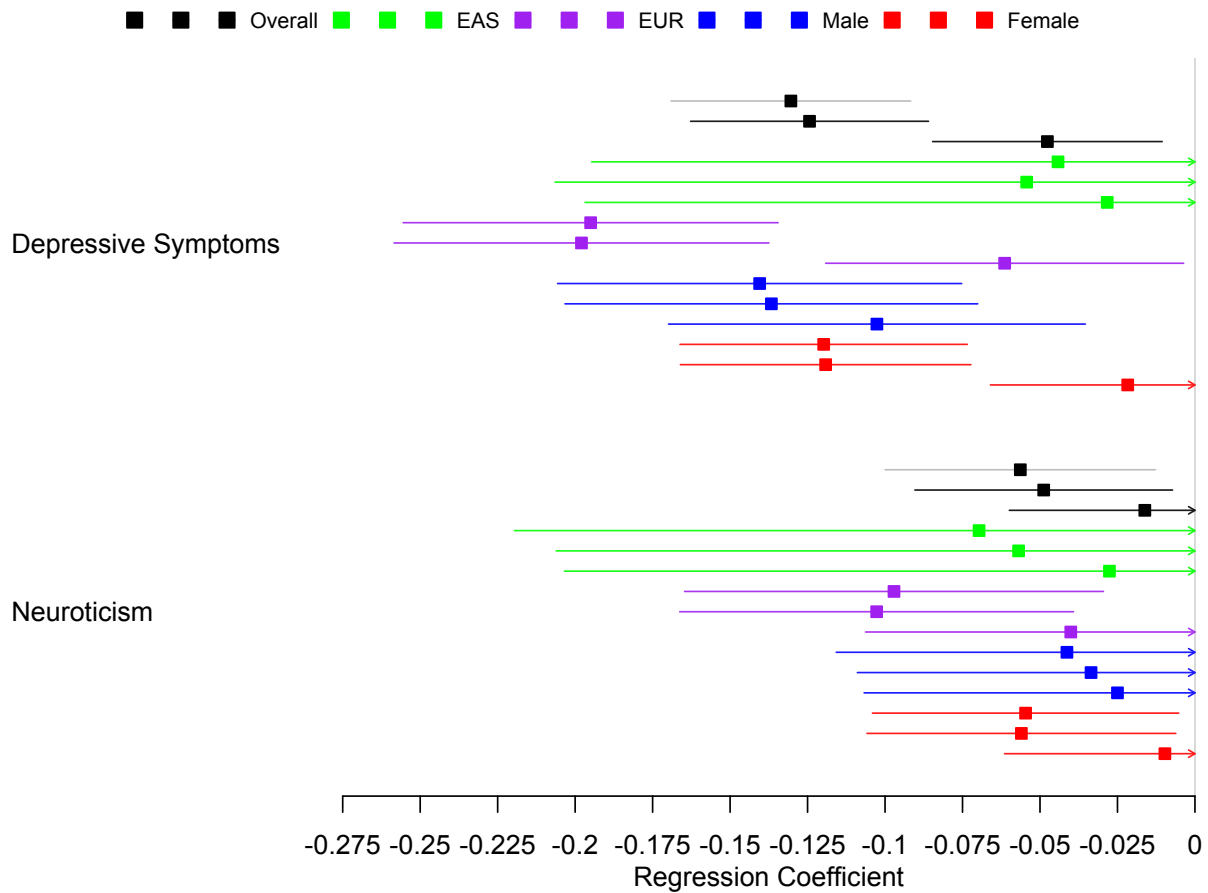


Figure 5. S4S Relationship Satisfaction. Left labels indicate outcome. Color indicates a subgroup of the sample, each with three consecutive models: univariate (on top), with covariates (next), and multivariate (bottom). Covariates include age, sex (when not stratified by sex), number of measurement occasions (when measured multiple times) and SES proxied by parental education level. Multivariate model additionally includes social activities, social support and stressful life events.

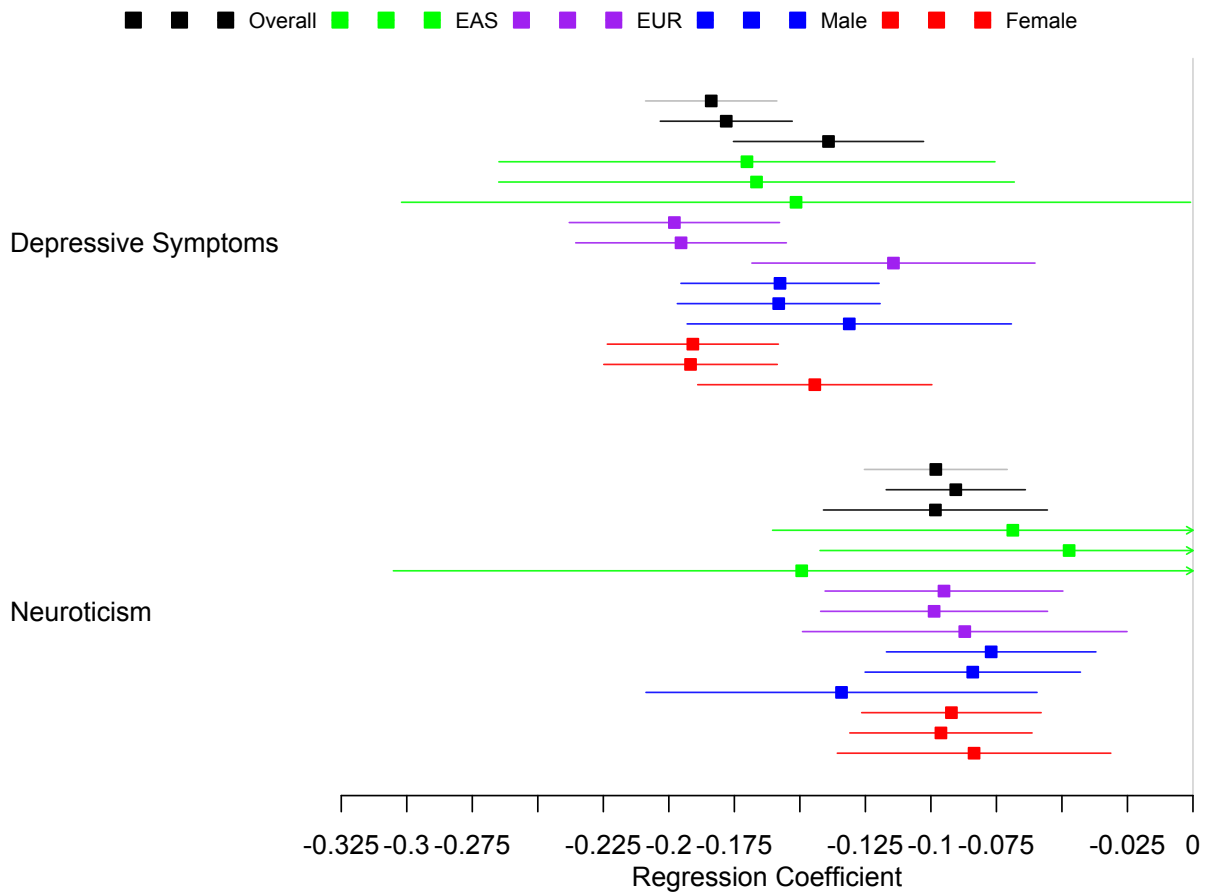


Figure 6. S4S Social Activities. Left labels indicate outcome. Color indicates a subgroup of the sample, each with three consecutive models: univariate (on top), with covariates (next), and multivariate (bottom). Covariates include age, sex (when not stratified by sex), number of measurement occasions (when measured multiple times) and SES proxied by parental education level. Multivariate model additionally includes relationship satisfaction, social support and stressful life events.

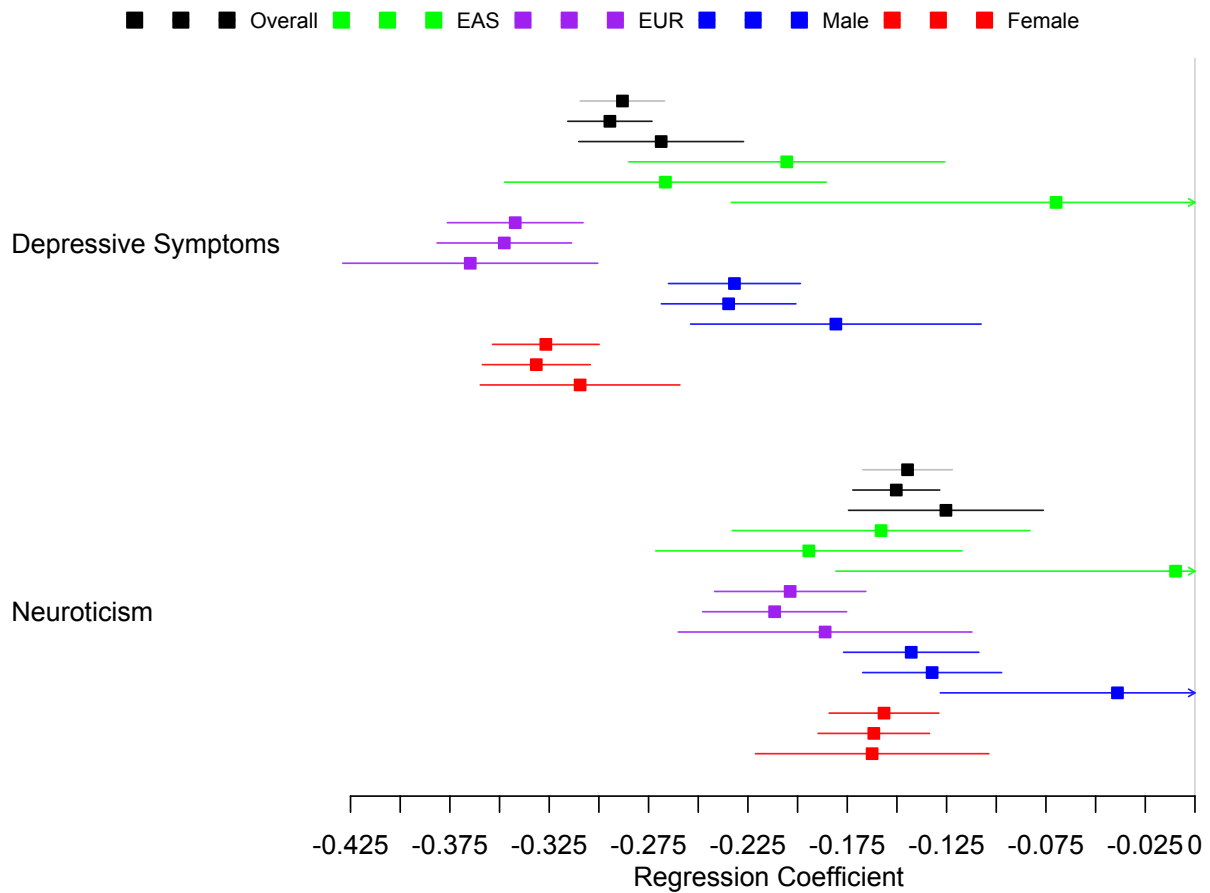


Figure 7. S4S Social Support. Left labels indicate outcome. Color indicates a subgroup of the sample, each with three consecutive models: univariate (on top), with covariates (next), and multivariate (bottom). Covariates include age, sex (when not stratified by sex), number of measurement occasions (when measured multiple times) and SES proxied by parental education level. Multivariate model additionally includes relationship satisfaction, social activities and stressful life events.

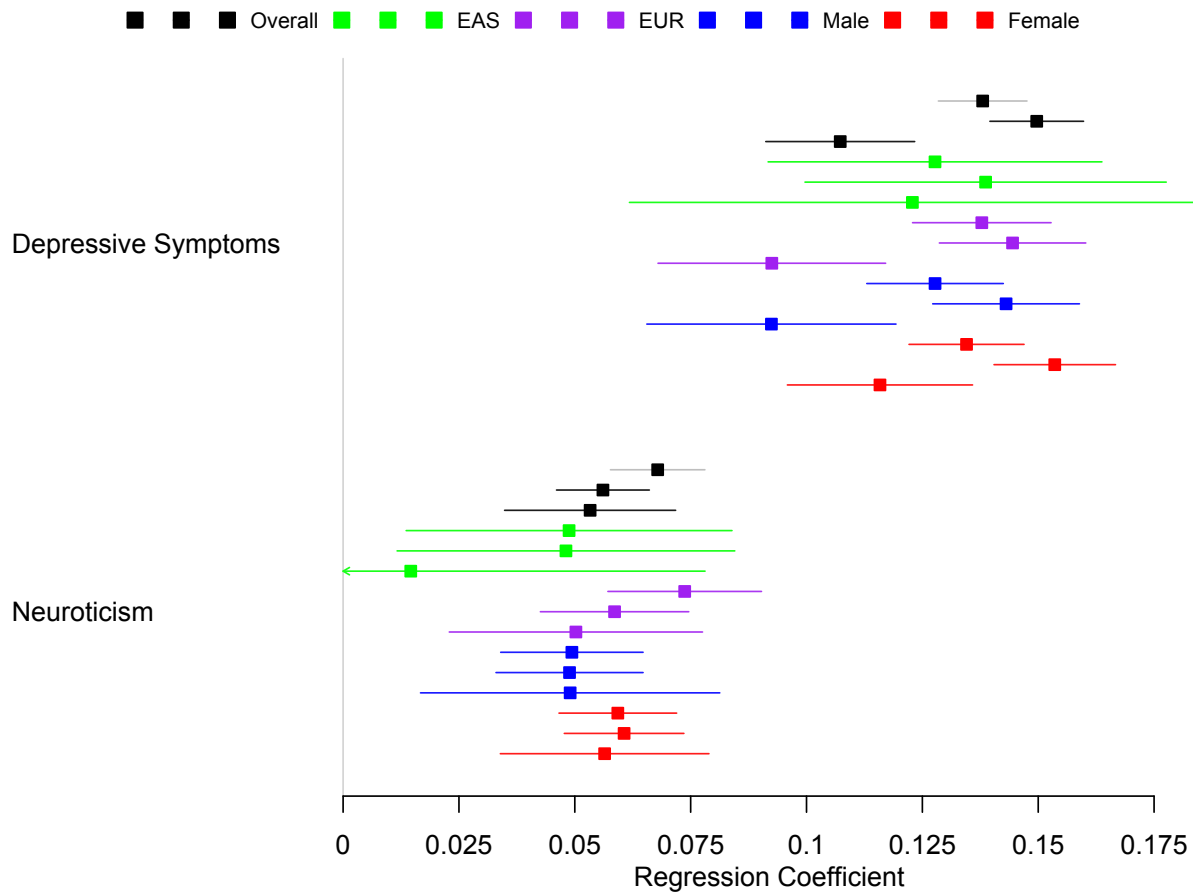


Figure 8. S4S Stressful Life Events. Left labels indicate outcome. Color indicates a subgroup of the sample, each with three consecutive models: univariate (on top), with covariates (next), and multivariate (bottom). Covariates include age, sex (when not stratified by sex), number of measurement occasions (when measured multiple times) and SES proxied by parental education level. Multivariate model additionally includes relationship satisfaction, social activities and social support.

III. Discussion:

As expected, the majority of the psychosocial factors displayed strong positive or negative associations with MD, MD+GAD, neuroticism and depressive symptoms outcomes. There were some notable differences between samples and subgroups, however. CONVERGE, likely in part due to the large sample size and detailed assessments, showed precise association estimates for the predictors, which were in every case of higher magnitude for the psychiatric outcomes of MD and MD+GAD than for neuroticism levels. This is consistent with the past biometrical literature suggesting that neuroticism constitutes a portion of the variance in risk of MD and GAD (51), showing associations with the same risk factors as the disorder outcomes do.

In S4S, the estimates were less precise, but analyses stratified by sex and ancestry were possible. Relationship satisfaction displayed no significant association with neuroticism, but did have a small protective effect against depressive symptoms, which is intuitive, given the preponderance of loneliness in many depressive presentations. Interestingly, the protective effect of relationship satisfaction on depressive symptoms was stronger in males than females (non-significant in females). This seems to corroborate work on sex differences in the effect of marriage as a protective factor – reduced risk of depression was found for married men, but not married women (122). Involvement in social activities had a similarly small protective effect against both depressive symptoms and neuroticism, and estimates were similar across subgroups, except for low precision in EAS.

Social support, however, had a stronger negative association with depressive symptoms, and a modest negative association with neuroticism. It is notable that the

effects were heightened for the subgroups of EUR and females for both outcomes (EAS effects for both outcomes were estimated lower and were non-significant, and effects in males were lower for depressive symptoms and non-significant for neuroticism). This extra protective effect of social support for females has been demonstrated in regards to major depression as well (123). As discussed in the methods section of this chapter, there is likely to be causal action in both directions between social support and the so-called outcomes of depressive symptoms and neuroticism. Internalizing presentations and neurotic personality features displayed by an individual can be difficult for friends and family, and may lead to erosion of social relationships and support networks. Conversely, strong support from loved ones and friends can have an ameliorative effect on depressive symptoms and neuroticism. The presence of the relationship (stronger in females than males) is the relevant finding from this evidence, rather than the directionality.

Finally, stressful life events were associated with increased depressive symptoms and neuroticism, though again the effect was stronger for depressive symptoms, and in this case there was little evidence of subgroup differences. Overall, these results suggest that neuroticism is associated with the same set of psychosocial correlates as MD and GAD. The S4S results indicate that there is evidence of sex differences in the association of psychosocial factors with depressive symptoms and (to a lesser degree) neuroticism. The variables that correspond with the quality of close personal relationships (romantic or platonic), such as relationship satisfaction and social support, seem to have differential impact on these outcomes. Relationship satisfaction shows a stronger negative relationship with depressive

symptoms for males, whereas social support is more strongly associated with lower depressive symptoms and neuroticism in females. Social support may be more strongly associated with protection against depressive symptoms and neuroticism for individuals of European ancestry than East Asian ancestry in the S4S sample. Note that the effect of social support on neuroticism in CONVERGE was very pronounced, however so the non-significance of the associations in EAS group may be an artifact of the small sample size. The two samples differ substantially, with CONVERGE being ascertained for major depression and S4S being recruited from college students, among other differences. This leads to different levels of endorsement of psychosocial risk factors, possibly indicating some confounding effects of sample differences – for instance, the most common level of social support endorsed in CONVERGE was moderate, whereas the most common level endorsed in S4S was the highest level of support. There is not sufficient information available in this project to disentangle these potential sources of bias, but the overall consistency of all risk factors acting in the expected direction (significantly so in most tests), is somewhat encouraging.

Chapter 3: Investigation into potential genetic variants related to MD, GAD, and neuroticism.

I. Methods:

Samples: All datasets were used in this set of analyses, including CONVERGE, MGS, and S4S. CONVERGE comprises information from Han Chinese females of ages 30-60. MGS comprises information from European individuals from ages 18-90. There are 1355 females and 1257 males used in this analysis from the MGS sample. S4S comprises information from individuals of multiple ancestries from ages 18-32. There are 3016 individuals of European ancestry (EUR), 557 individuals of East Asian ancestry (EAS), 1471 males, and 2069 females used in this analysis from the S4S sample cohorts 1-3, which were the subset included in these analyses. S4S cohort 1 had responded to 3 years of survey follow-up, cohort 2 had responded to 2 years of survey follow-up, and cohort 3 had responded to 1 year of survey follow-up when data was assembled for these analyses. Notably, imputation for all of these samples had been carried out by unrelated efforts prior to the beginning of this project, so QC focused primarily on downstream efforts to ensure all interpreted results were high quality.

Variables: Outcomes in CONVERGE include BFI neuroticism (110), which was standardized, MD status, and a dichotomous outcome of MD+GAD comorbidity versus MD controls – a ‘case-control’ test of MD+GAD status. Outcomes in MGS include GAD status, MD status, a similar MD+GAD comorbid outcome, and BFI neuroticism score (110). Outcomes in S4S include an abbreviated 3-item BFI neuroticism score (110), and a sum score of 4 SCL-90 depressive symptoms (114), both of which were standardized.

Covariates for CONVERGE, MGS and S4S were used to adjust for the effects of demographic factors, when appropriate. Age at assessment was included in all analyses, as well as sex (except in CONVERGE, as that sample is all female).

Analyses: Genetic association tests, with one exception, were conducted using BOLT-LMM (124), a linear mixed model method that achieves increased power over most other mixed model association tools while retaining the ability to account for population stratification and cryptic relatedness. The association test for depressive symptoms in the EAS subgroup of S4S was conducted in PLINK (125), with adjustment for 10 ancestry principal components in addition to the other covariates, since association analysis failed in BOLT-LMM due to the generation of an invalid (too low) heritability estimate, likely as a result of the small sample size of this subgroup. PLINK is a reliable analysis tool, and has relatively good power in a range of circumstances, though not as powerful as BOLT-LMM. That is to say, PLINK would be a tolerable alternative to all association analyses, but where BOLT-LMM’s assumptions are met and heritability estimation is possible, it is likely to provide more accurate results, with more evenly controlled Type 1 error. Mixed models, generally speaking, are combinations of fixed and random variables that partition the variance of an outcome

between them. A fixed effect is a variable whose impact is constant across individuals, such as the population effect on a trait of having one non-reference allele instead of the reference allele at a particular locus. A random effect is a variable that may have different impact for each individual, such as the overall effect of the polygenic background of a given person. With respect to genetic association analysis, mixed models generally refer to the inclusion of a matrix of genotypic relatedness as a random effect in the regression of the phenotype on the fixed effect of genotype at each locus. Mixed models have been used to prevent bias in the estimation of genetic effects due to selection in animal genetics (126), and more recently to control for both population and family structure in human genetics (127).

BOLT-LMM is a good choice among mixed model methods due in part to its efficiency; its direct use of genotypes to compute the random-effect variance components saves time, scaling linearly with both sample size and number of markers (compared to quadratic scaling with regards to one of the two variables implemented in other mixed model methods), and it is highly optimized for savings in both computational time and memory usage (124). It also has the capability to calculate random effects drawn from Gaussian mixture priors, as opposed to the 'infinitesimal' model (128) implicit in standard linear mixed models, when the circumstances suggest doing so will increase power. The infinitesimal model was predicted to be the highest powered in all analyses in this project, however. As suggested by the BOLT-LMM documentation (124), imputed dosage data were hard-called into binary PLINK genotypes, with variants with high-confidence genotype probability (the most likely genotype call had to be a probability of 0.9 or higher) included in the hard-called

genotype subset. SNPs were filtered in the process of association testing for no greater than 10% missingness among the sample individuals. These permissive filtering thresholds were intended to maximize the number of SNPs included in the mixed model. After associations were computed, results were further filtered so that all variants included in figures displaying association peaks, and interpreted in the text, had MAF >1%, and imputation information >0.5. Manhattan and quantile-quantile (q-q) plots were generated using the qqman package in R (129). Genomic inflation factors were calculated using the GenABEL package in R (130). LocusZoom plots were also produced for significant loci (131).

After association analyses, integration of the variant results by genomic intervals was performed with Multi-marker Analysis of Genomic Annotation, or MAGMA (132). MAGMA is a flexible gene and gene-set based analysis program that efficiently implements a number of analytic approaches to estimating the level of association between a specified region or regions and an outcome. Adjustment for multiple testing was computed using the Bonferroni correction (133) for the number of intervals included in each outcome and sample, to ensure a family-wise error rate of 0.05. To preserve the benefits of BOLT-LMM in these interval-based associations, first the p-value based method of mean variant association was used in MAGMA, which is comparable to other gene-based methods; this is broadly a test of the enrichment of association in variants in a given genomic region, which may be driven in part by LD. These results are then compared with the regression model using genotypes directly with MAGMA, though this neither uses a mixed model approach nor post-imputation dosage data, so is likely to have lower statistical power. It is, however, a very specific test, with a well-controlled

Type 1 error rate (132). To avoid theoretically-based assumptions about relevant genetic effects primarily stemming from e.g. coding regions, a comprehensive list of regions was pulled from GENCODE (134), 58,977 in all, which in addition to traditional primary protein-coding transcripts includes protein-coding genes with alternatively-spliced variants, non-coding loci, pseudogenes, as well as intergenic regions between them (to include any potentially unmapped regulatory sites of action).

II. Results:

The association results only resulted in genome-wide significant loci, with a p-value less than 5×10^{-8} (135), in a few of the outcomes tested. A subset of these loci evinced additional supportive signals consistent with a true genetic effect on a background haplotype. Manhattan and q-q plots were generated for each test, with LocusZoom (131) plots for all regions attaining genome-wide significance, using the appropriate reference panel (ASN or EUR) from 1000 Genomes (136) to determine whether these appear plausible (Figures 9-53).

For CONVERGE, the MD GWAS was unique in that the CONVERGE group had already published results (93), so to some extent the test was to validate the methodology, though slight differences including using BOLT-LMM as opposed to another mixed model technique, and including age as a covariate, did result in somewhat altered associations. The q-q plot displayed moderate inflation at the tail (Figure 10), and the test had a genomic inflation factor (λ) of 1.147, which when scaled to a theoretical test for 1000 cases and 1000 controls (λ_{1000}) was 1.028. The two published loci on chromosome 10 show strong peaks in the Manhattan plot (Figure 9),

though only one, the locus in the phospholysine phosphohistidine inorganic pyrophosphate phosphatase (*LHPP*) gene (rs35936514, $p=3.2e-08$), reaches significance in this test due to the different analytic approach (Figure 11). The *LHPP* gene hydrolyzes a range of phosphate molecules, and is expressed in the brain. The other locus from the original publication did reach significance without the adjustment for the age covariate: rs12415800 in the sirtuin1 or *SIRT1* gene, $p=3.1e-08$ in the unadjusted association test versus $8.4e-08$ in the adjusted test; rs35936514, the top variant in the *LHPP* gene, had a $p=7.0e-09$ in the unadjusted test vs. $3.2e-08$ in the adjusted test (unadjusted results not shown). The second variant reaching significance in the present analysis on chromosome 19 in the perilipin 4 (*PLIN4*) gene, (rs7257142, $p=1.9e-09$), was not supported by nearby variants of suggestive significance, and is not present in publically-available East Asian LD reference panels (Figure 12). It may have low LD, be a spurious signal or be a speciously imputed variant. The *PLIN4* gene may be related to adipocyte functioning, and is expressed during adipocyte differentiation, which would appear to be distal to the central pathology in MD. The results for comorbid MD+GAD in CONVERGE were less strong, which was as expected given the reduced number of cases in this analysis. There were no loci reaching significance (Figure 13). The q-q plot (Figure 14) suggests modest genomic inflation overall, and signal in the tail, with λ of 1.047 and λ_{1000} of 1.023. For CONVERGE Neuroticism, there were also no significant loci (Figure 15). This is likely due to insufficient signal to detect modest effects in this sample, as the q-q plot suggests perhaps the mixed model, in correcting for relatedness and population structure, deflated the strongest Neuroticism associations (Figure 16), though λ was 1.097 and λ_{1000} was 1.020.

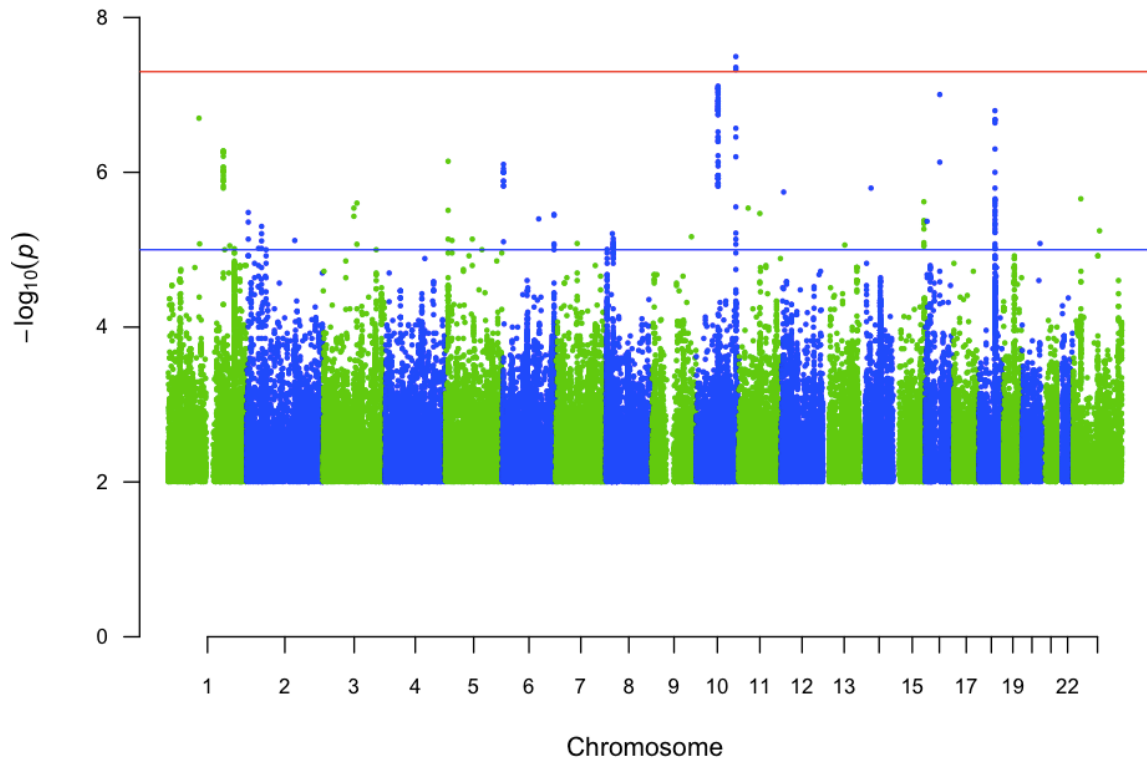


Figure 9. MD BOLT-LMM Results in CONVERGE. Manhattan plot displaying filtered adjusted association results, arranged by physical position.

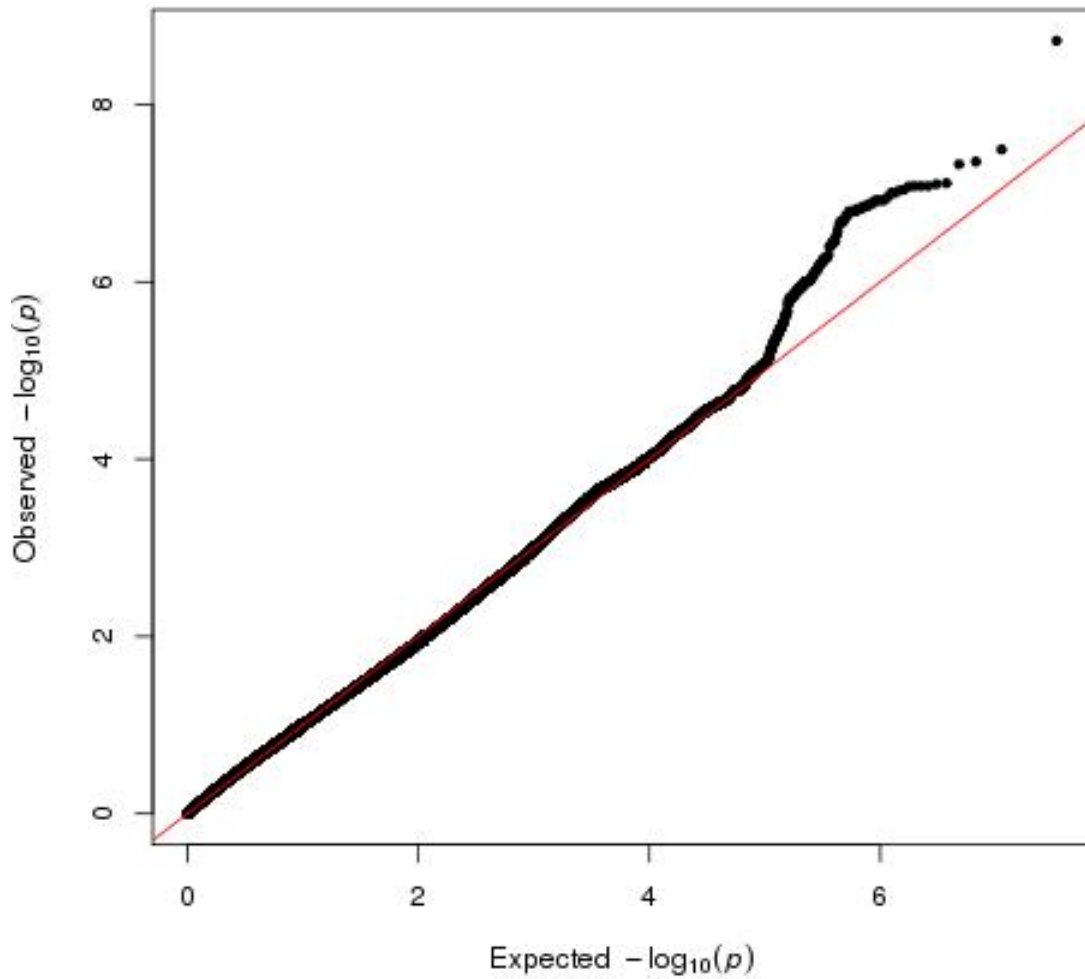


Figure 10. MD q-q Plot in CONVERGE. Quantile-quantile plot visualizing observed versus expected association strength.

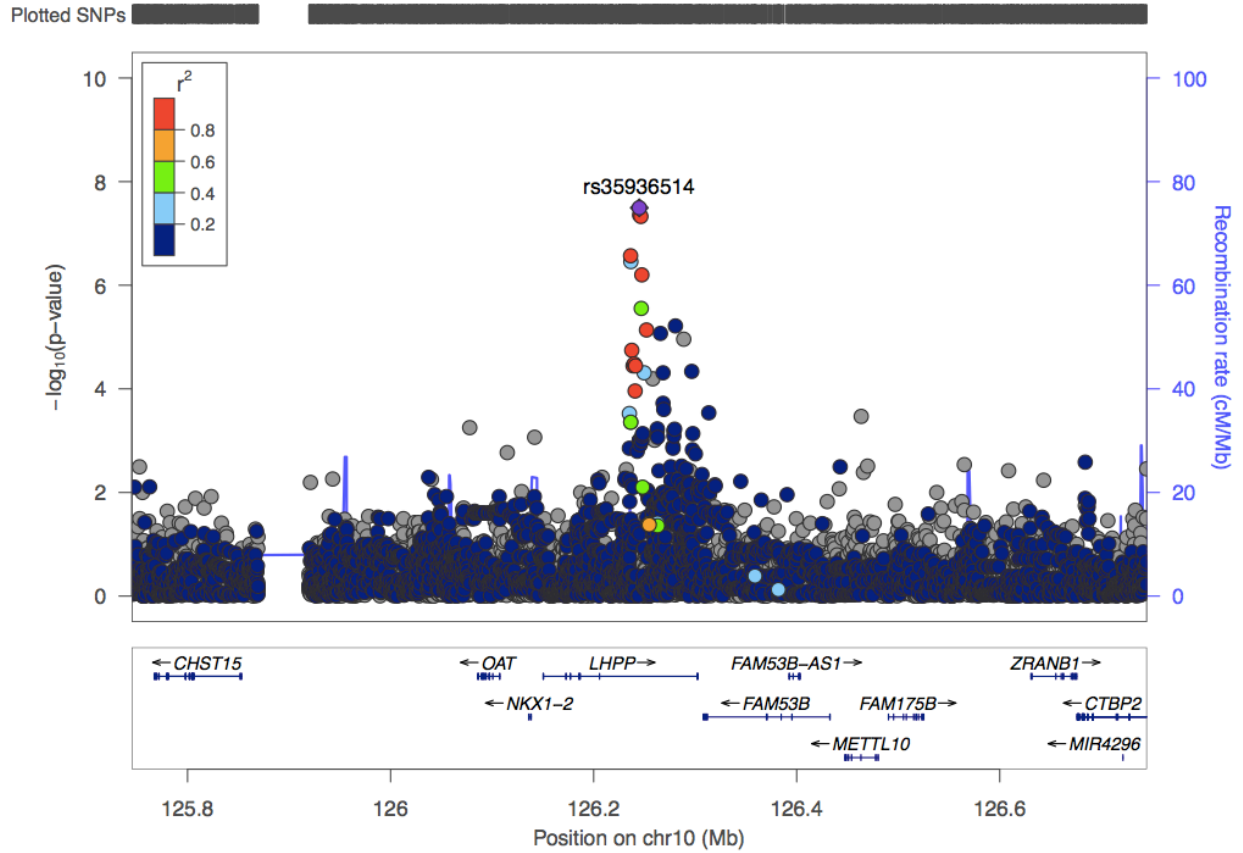


Figure 11. CONVERGE MD Chromosome 10 Locus. LocusZoom plot displaying strongest associated variant with coloring to indicate LD relationships with other nearby variants. Grey points indicate variants missing LD information with respect to the strongest variant. Genes in the region displayed for reference.

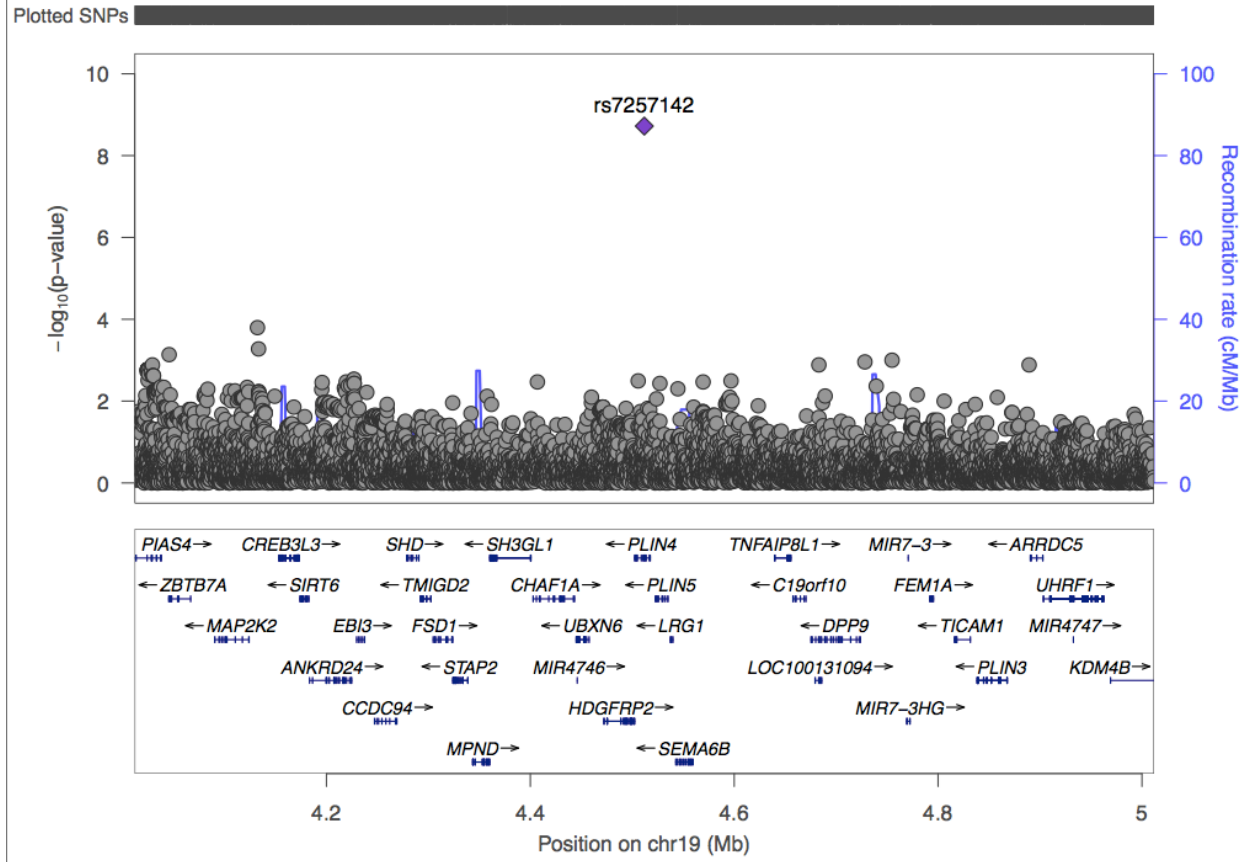


Figure 12. CONVERGE MD Chromosome 19 Locus. LocusZoom plot displaying strongest associated variant with coloring to indicate LD relationships with other nearby variants. Grey points indicate variants missing LD information with respect to the strongest variant. Genes in the region displayed for reference.

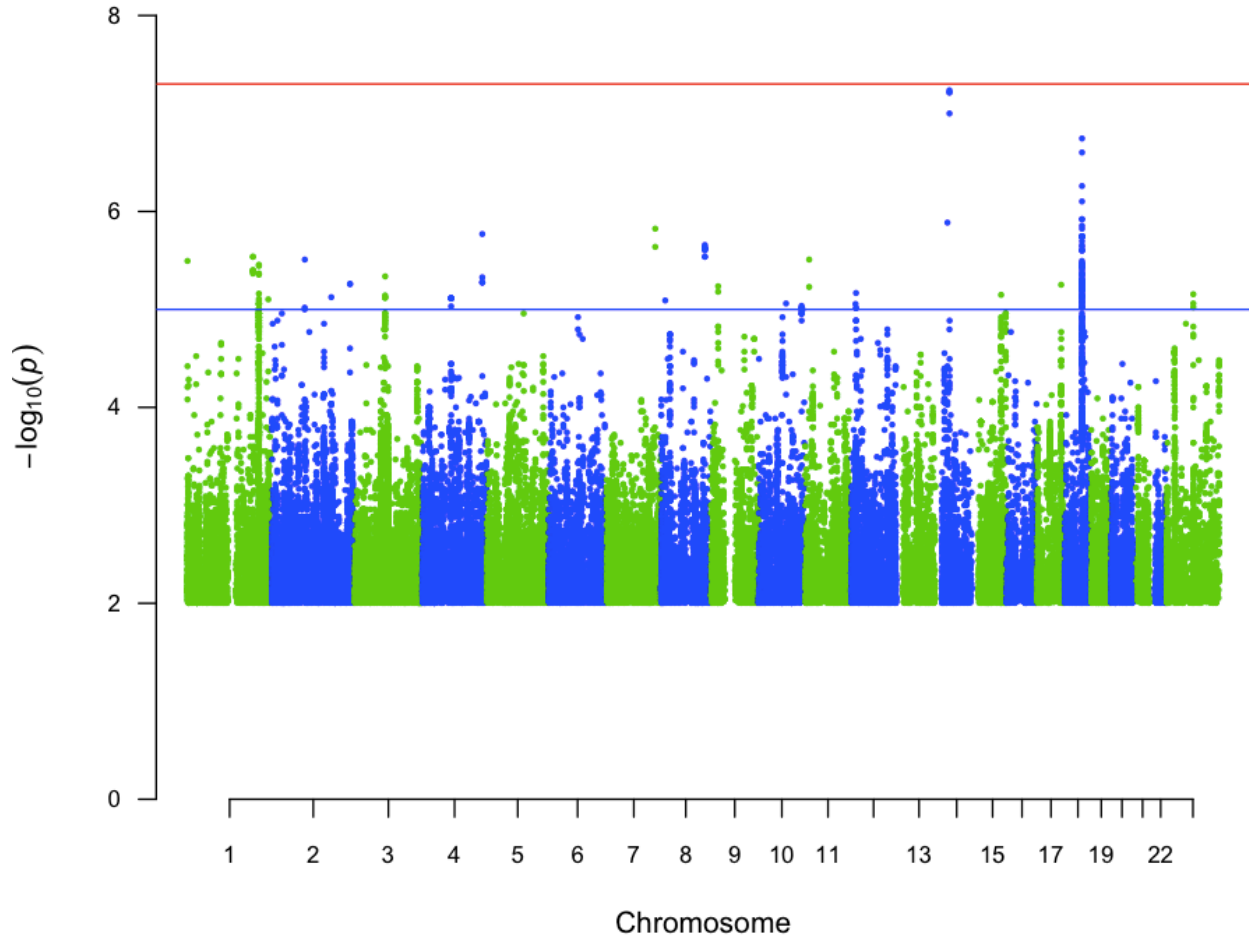


Figure 13. MD+GAD BOLT-LMM Results in CONVERGE. Manhattan plot displaying filtered adjusted association results, arranged by physical position.

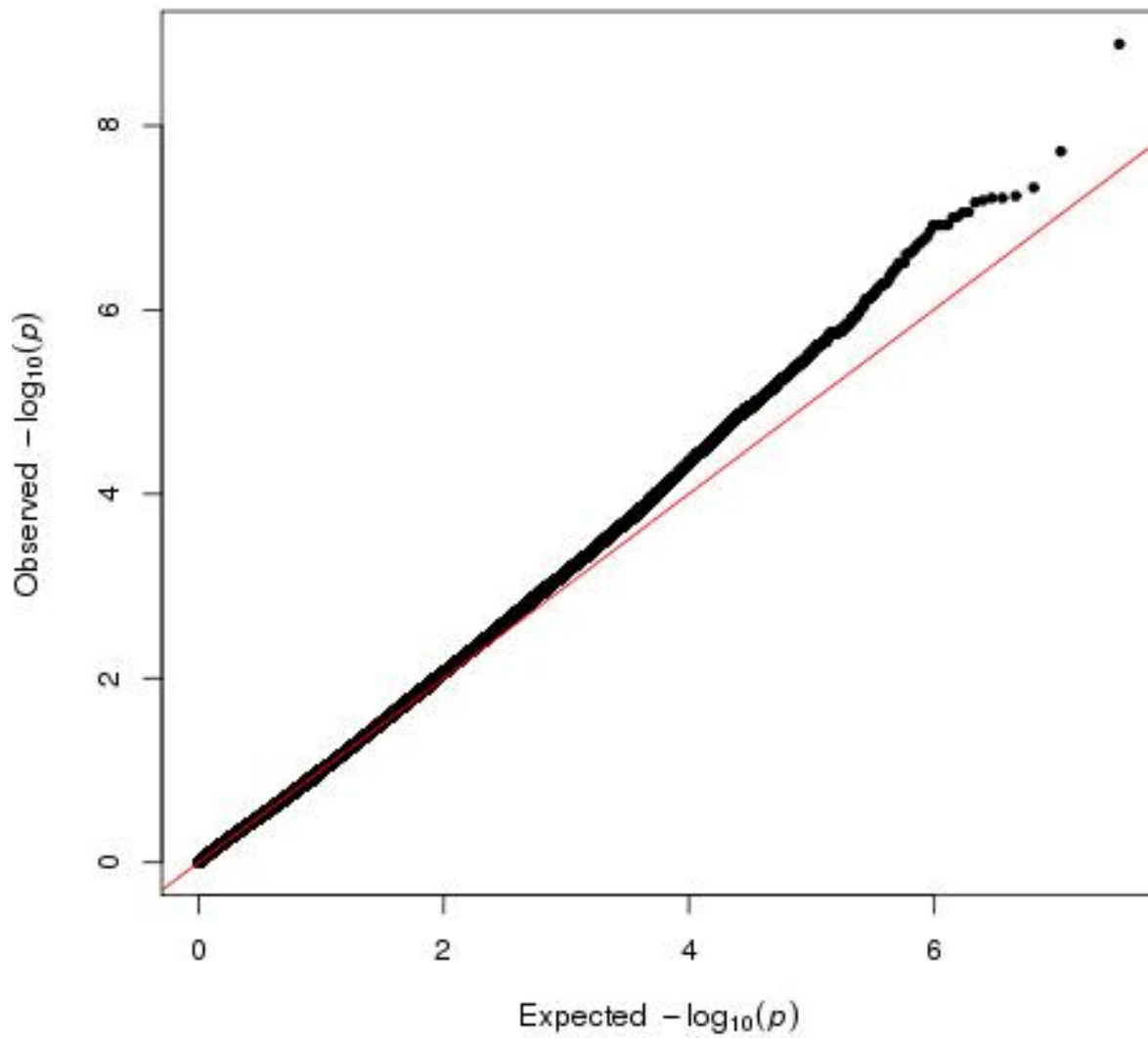


Figure 14. MD+GAD q-q Plot in CONVERGE. Quantile-quantile plot visualizing observed versus expected association strength.

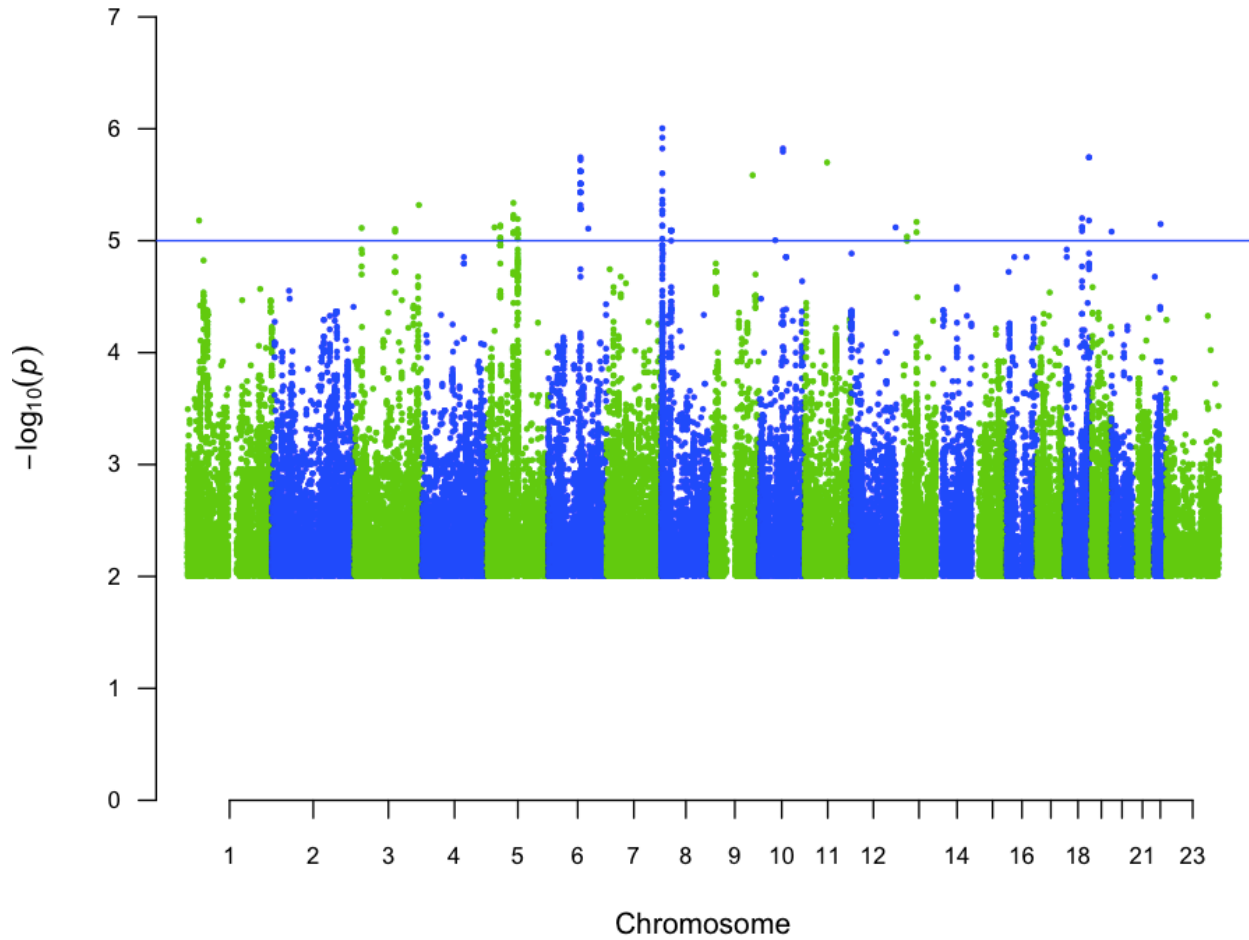


Figure 15. Neuroticism BOLT-LMM Results in CONVERGE. Manhattan plot displaying filtered adjusted association results, arranged by physical position.

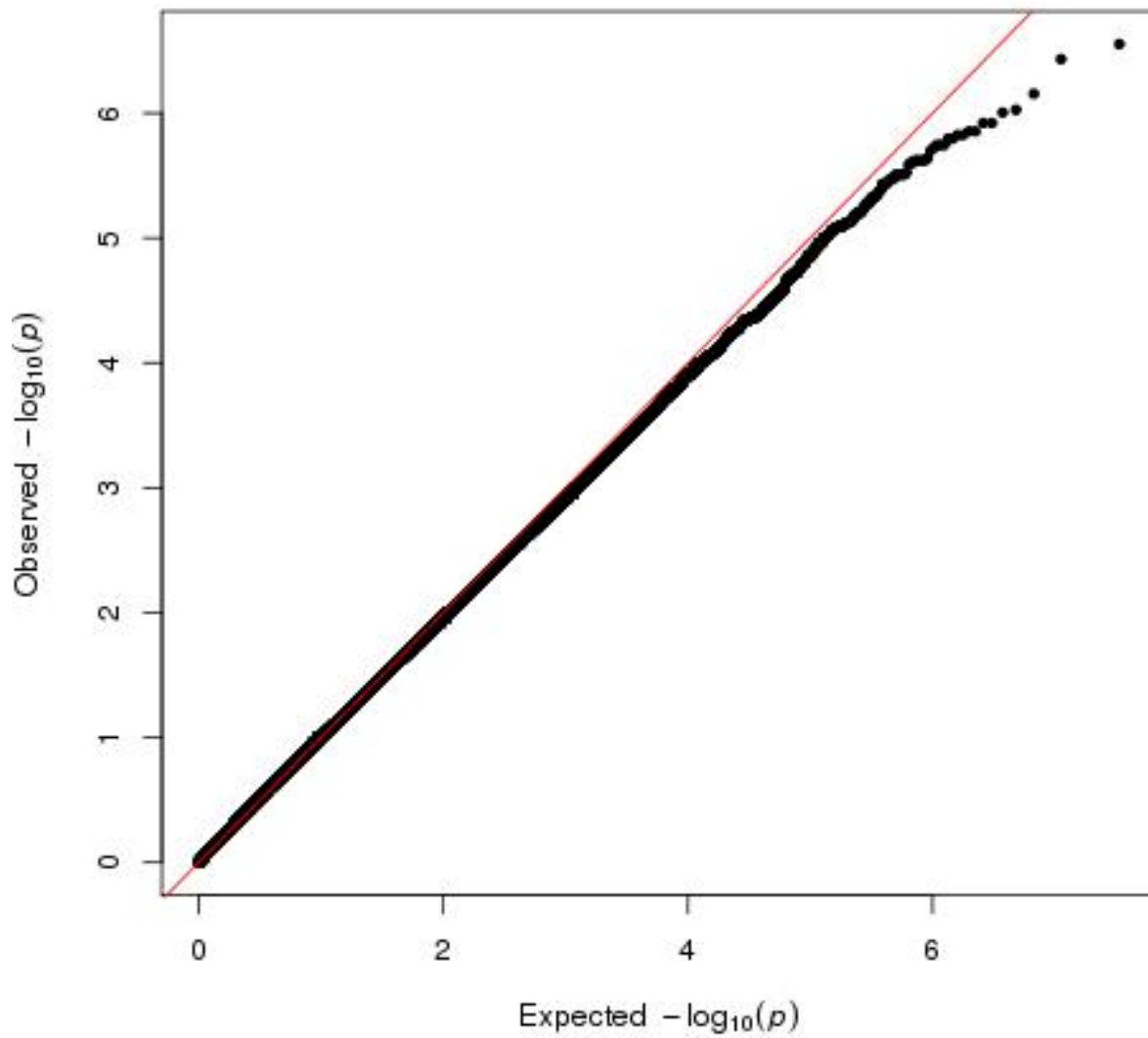


Figure 16. Neuroticism q-q Plot in CONVERGE. Quantile-quantile plot visualizing observed versus expected association strength.

Association results for the MGS sample were mostly, but not completely, null. MD in the MGS sample displayed no significant loci (Figure 17). Any potential signal did not survive correction (Figure 18), though overall there was some inflation, with a λ estimate of 1.147 and estimate of and λ_{1000} of 1.149.

There were also no significant loci for the comorbid MD+GAD outcome in MGS (Figure 19). Genomic inflation, in contrast to the MD outcome in MGS, appeared minimal with a λ and λ_{1000} estimate of 1 (the minimum in GenABEL), with signal at the high end balanced by deflation in the lower p-values (Figure 20).

The results for GAD in MGS were somewhat stronger, with two loci reaching significance (Figure 21). Genomic inflation appeared minimal overall with a λ and λ_{1000} estimate of 1 (the minimum in GenABEL), with signal at the high end balanced by deflation in the lower p-values (Figure 22). Both these loci have some supporting signals, though not enough to declare them genuine associations with complete certainty. The first locus (most significant variant: rs117420048, $p=1.7e-09$), in the intron of *LINC00824*, a long non-coding RNA (Figure 23), has the more supportive pattern, with a few strong highly-correlated variants near the top hit, and suggestive variants in weaker LD further away. The other locus on chromosome 13 (Figure 24), around *LINC00348*, another long non-coding RNA (most significant variant: rs146220192, $p=4.6e-08$), has a few supporting signals of suggestive significance, but only modest LD with the top variant.

The MGS associations for Neuroticism were, like CONVERGE, non-significant (Figure 25). Signals appeared to be over-corrected in the tail, similar to the pattern of neuroticism in CONVERGE (Figure 26). λ was 1.047 and λ_{1000} 1.101.

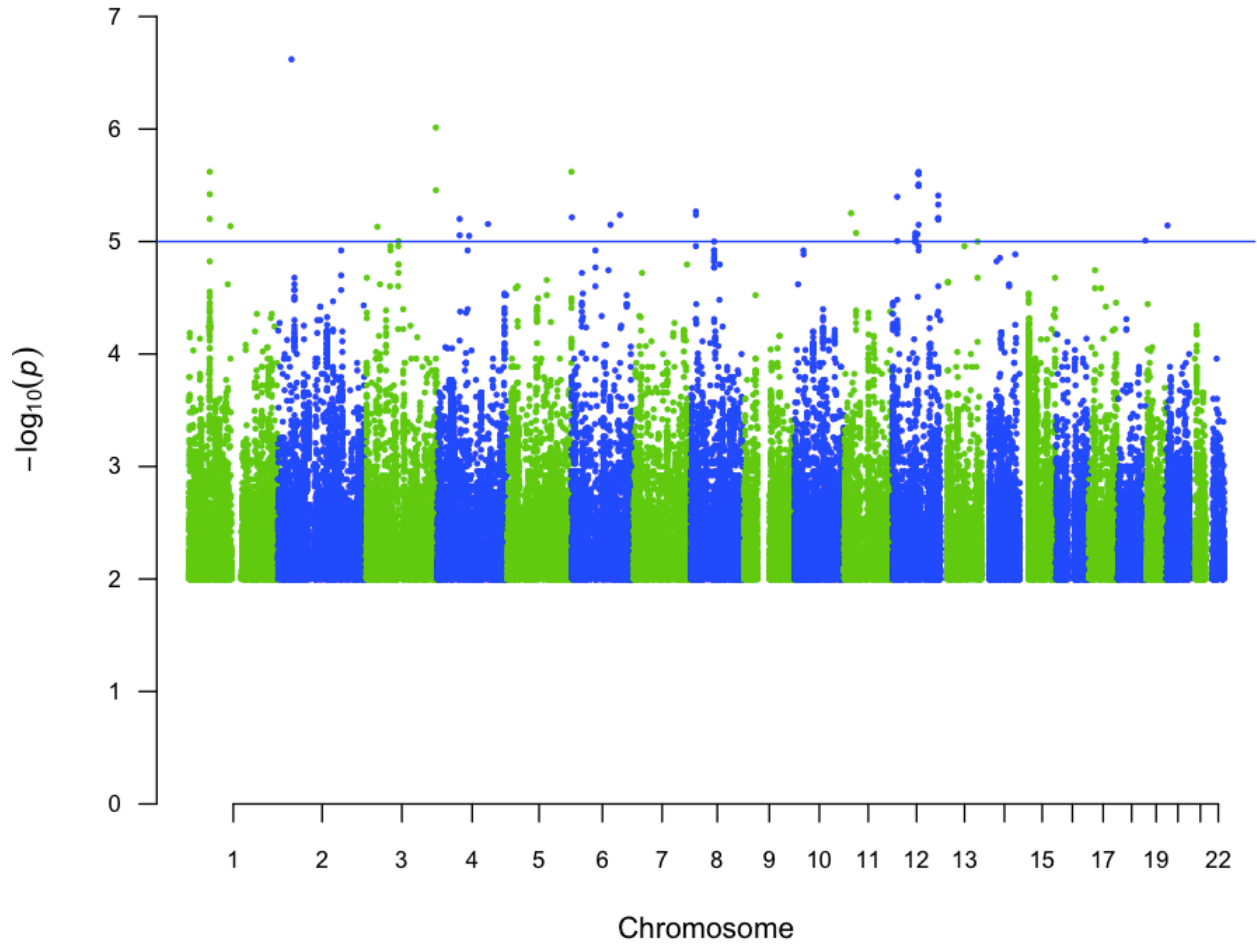


Figure 17. MD BOLT-LMM Results in MGS. Manhattan plot displaying filtered adjusted association results, arranged by physical position.

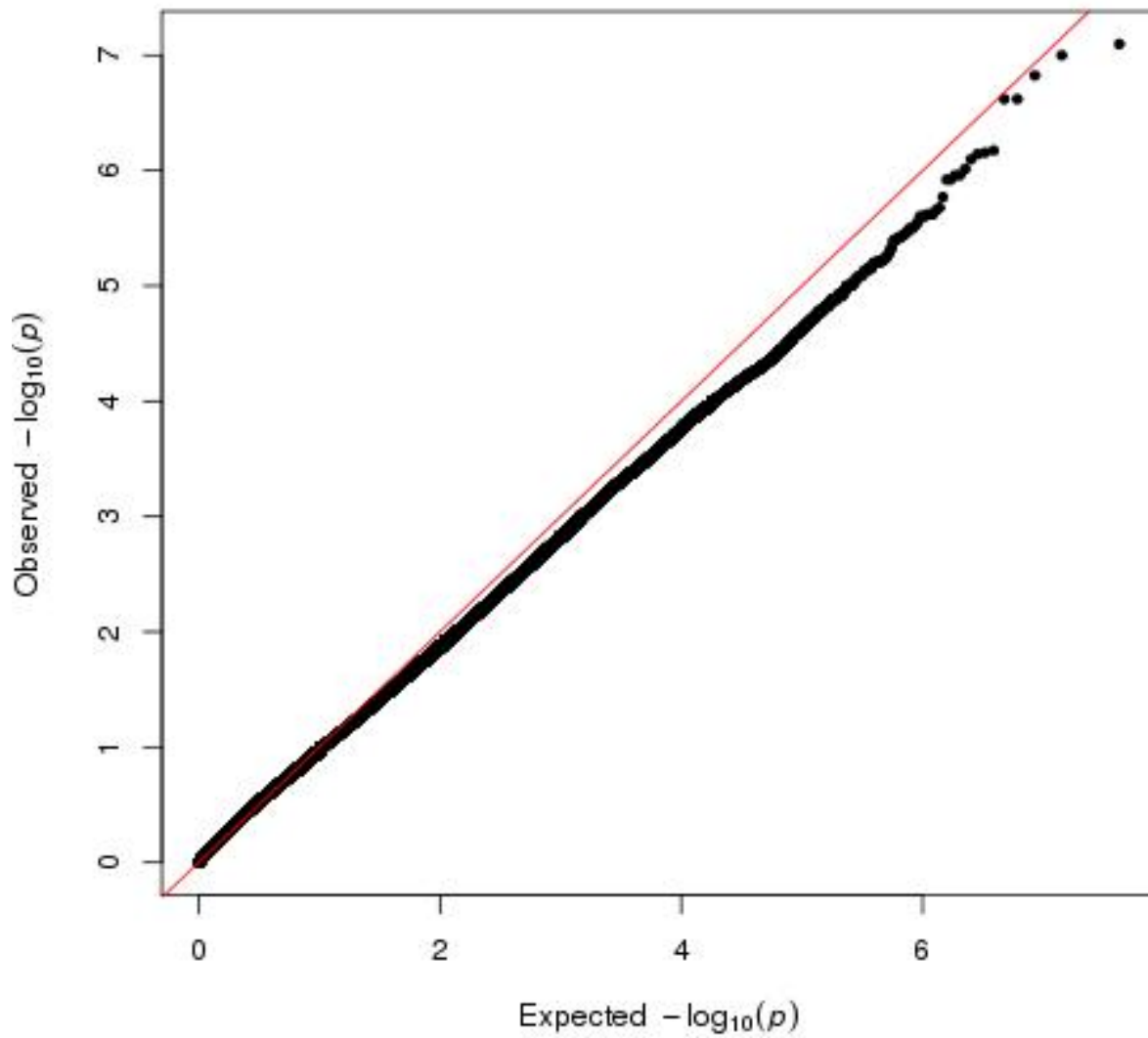


Figure 18. MD q-q Plot in MGS. Quantile-quantile plot visualizing observed versus expected association strength.

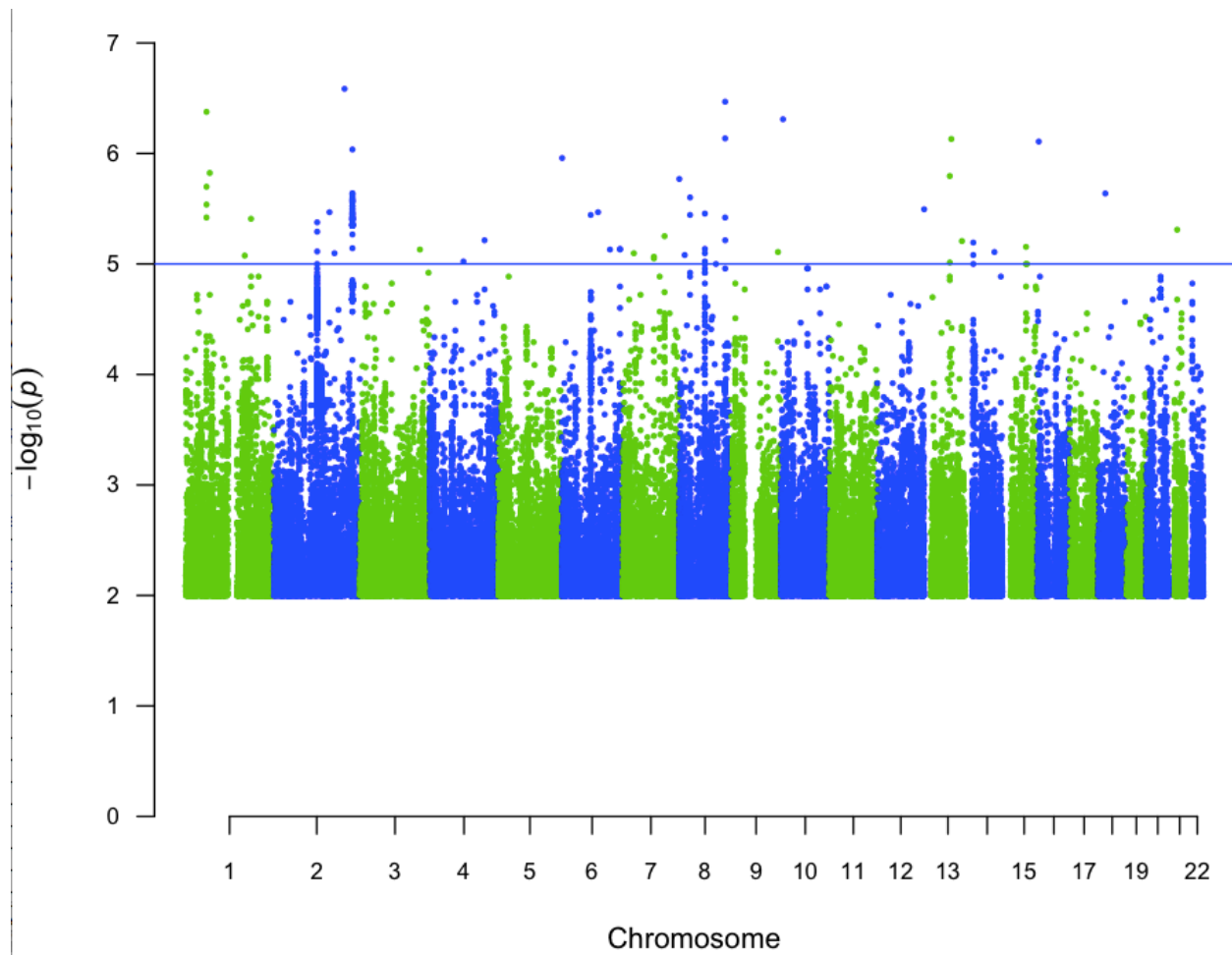


Figure 19. MD+GAD BOLT-LMM Results in MGS. Manhattan plot displaying filtered adjusted association results, arranged by physical position.

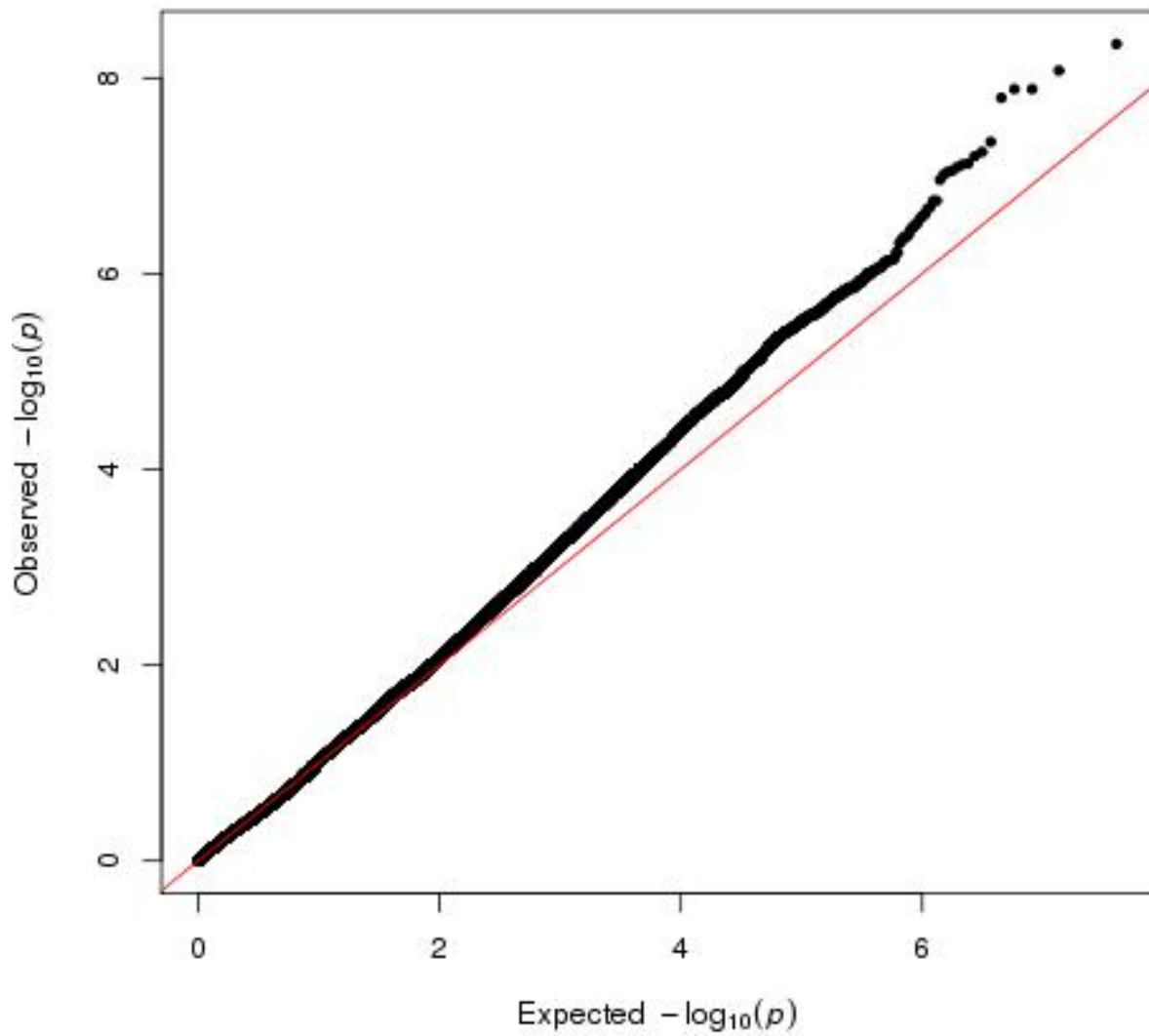


Figure 20. MD+GAD q-q Plot in MGS. Quantile-quantile plot visualizing observed versus expected association strength.

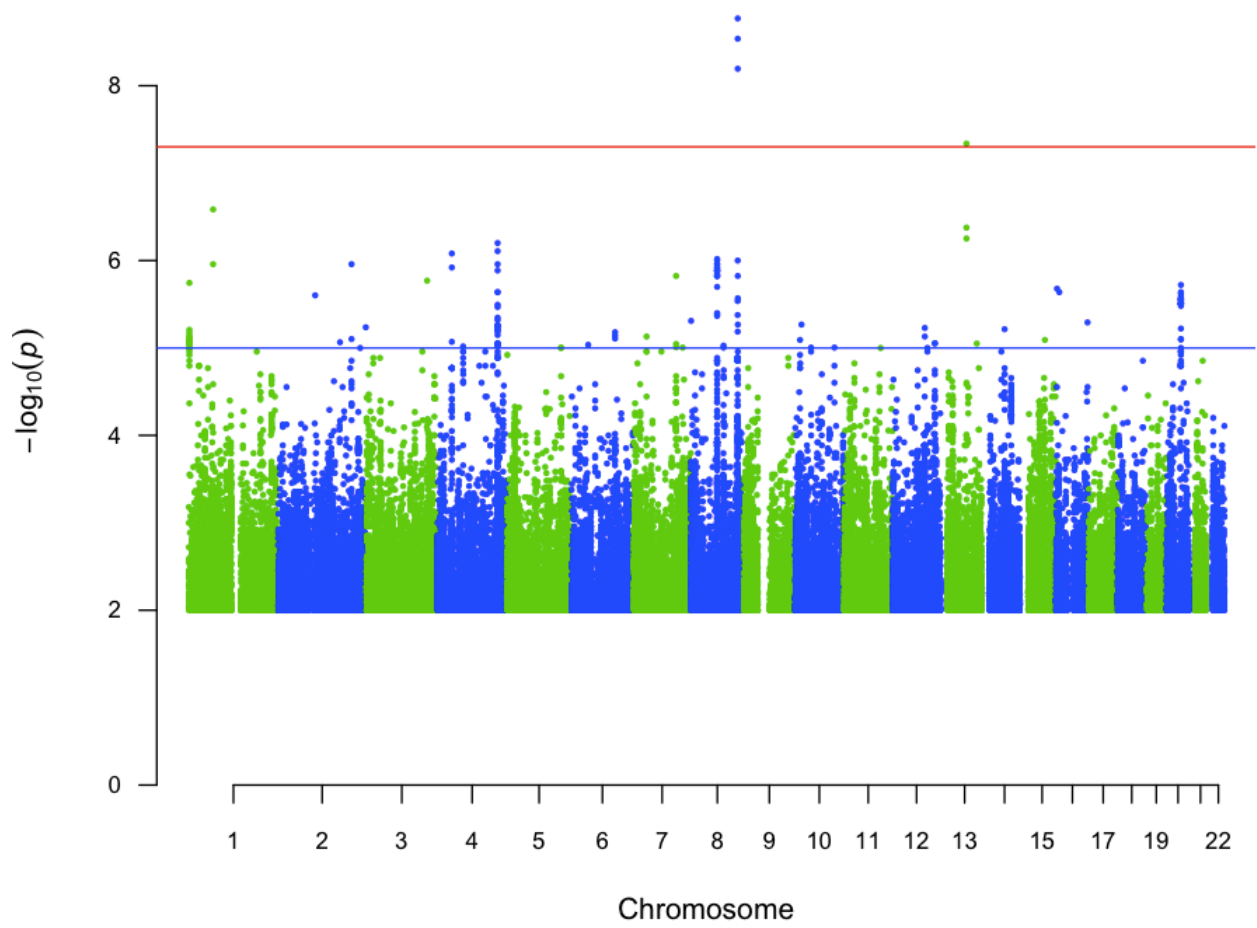


Figure 21. GAD BOLT-LMM Results in MGS. Manhattan plot displaying filtered adjusted association results, arranged by physical position.

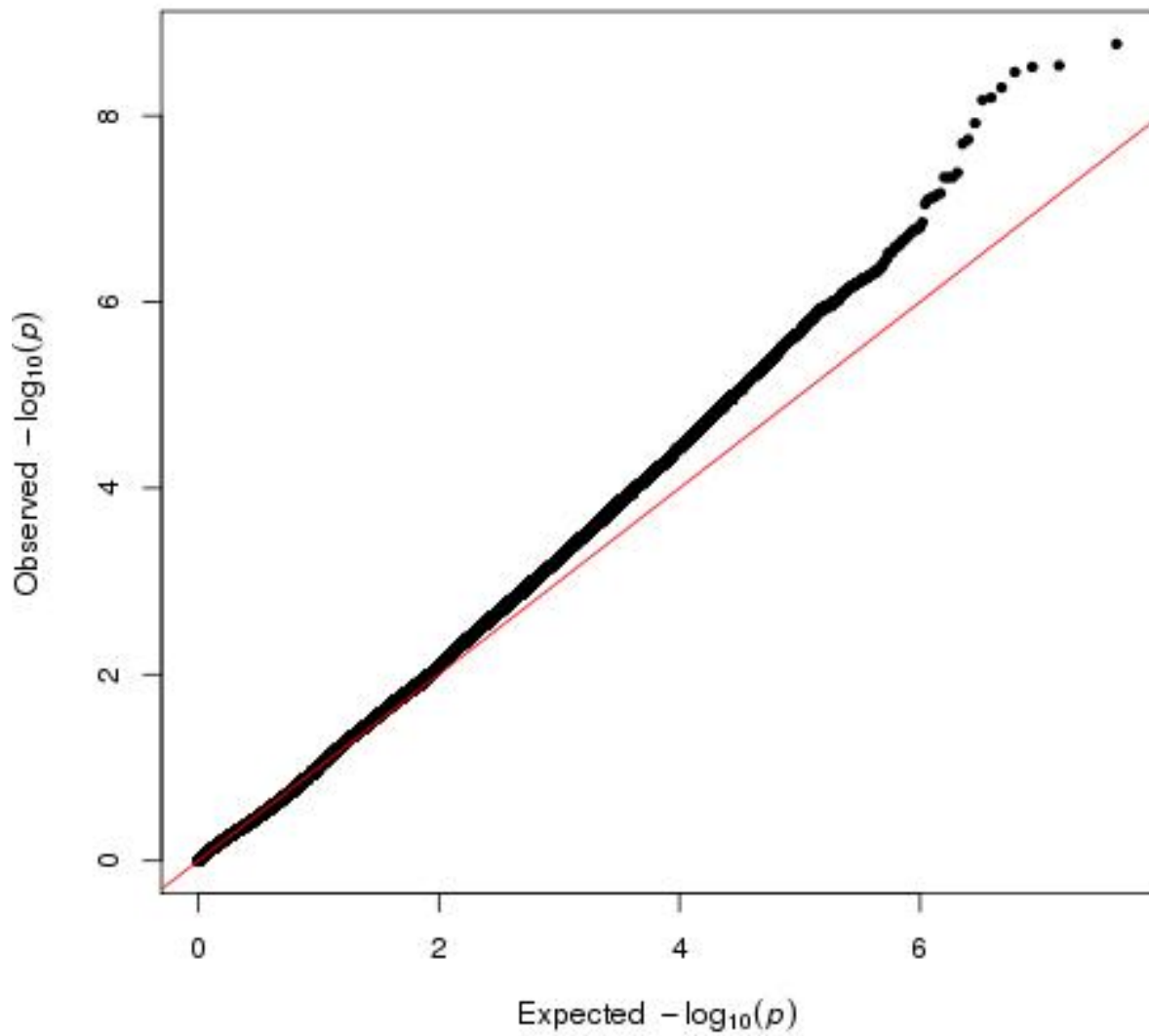


Figure 22. GAD q-q Plot in MGS. Quantile-quantile plot visualizing observed versus expected association strength.

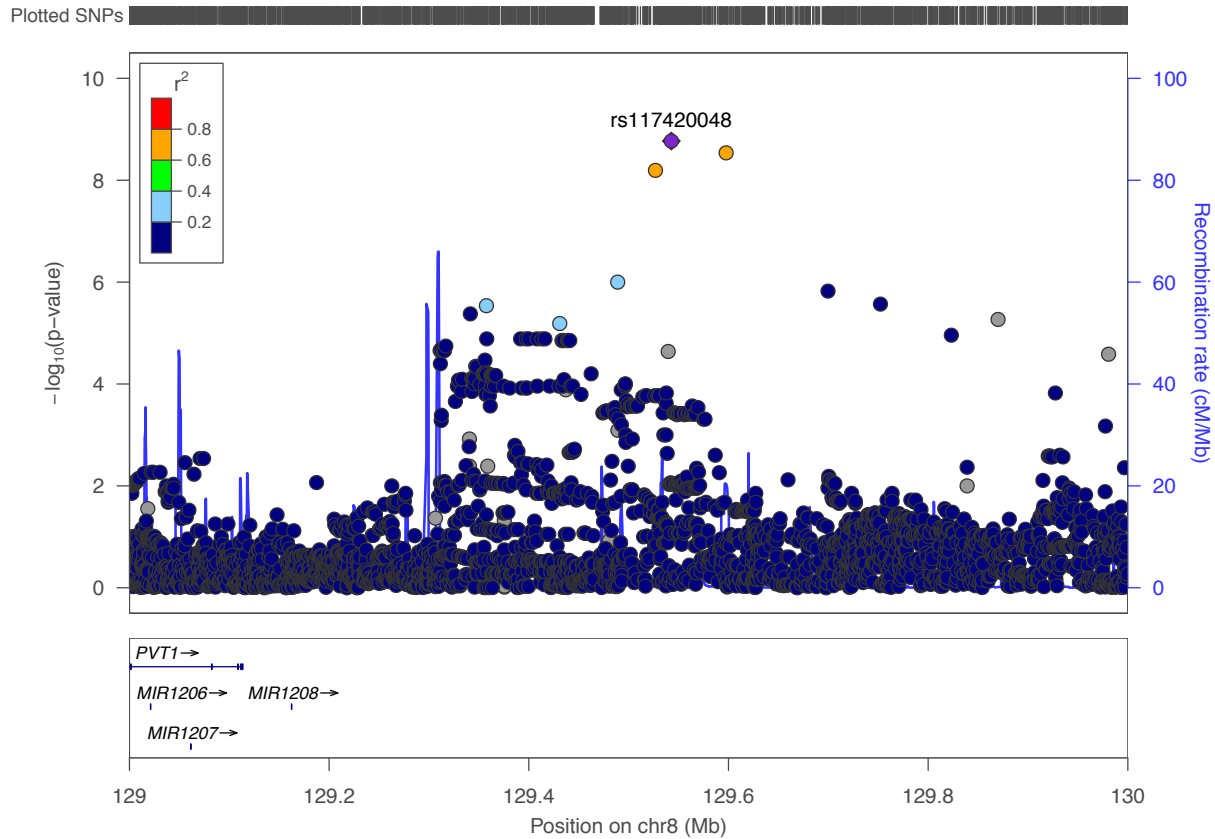


Figure 23. MGS GAD Chromosome 8. LocusZoom plot displaying strongest associated variant with coloring to indicate LD relationships with other nearby variants. Grey points indicate variants missing LD information with respect to the strongest variant. Genes in the region displayed for reference.

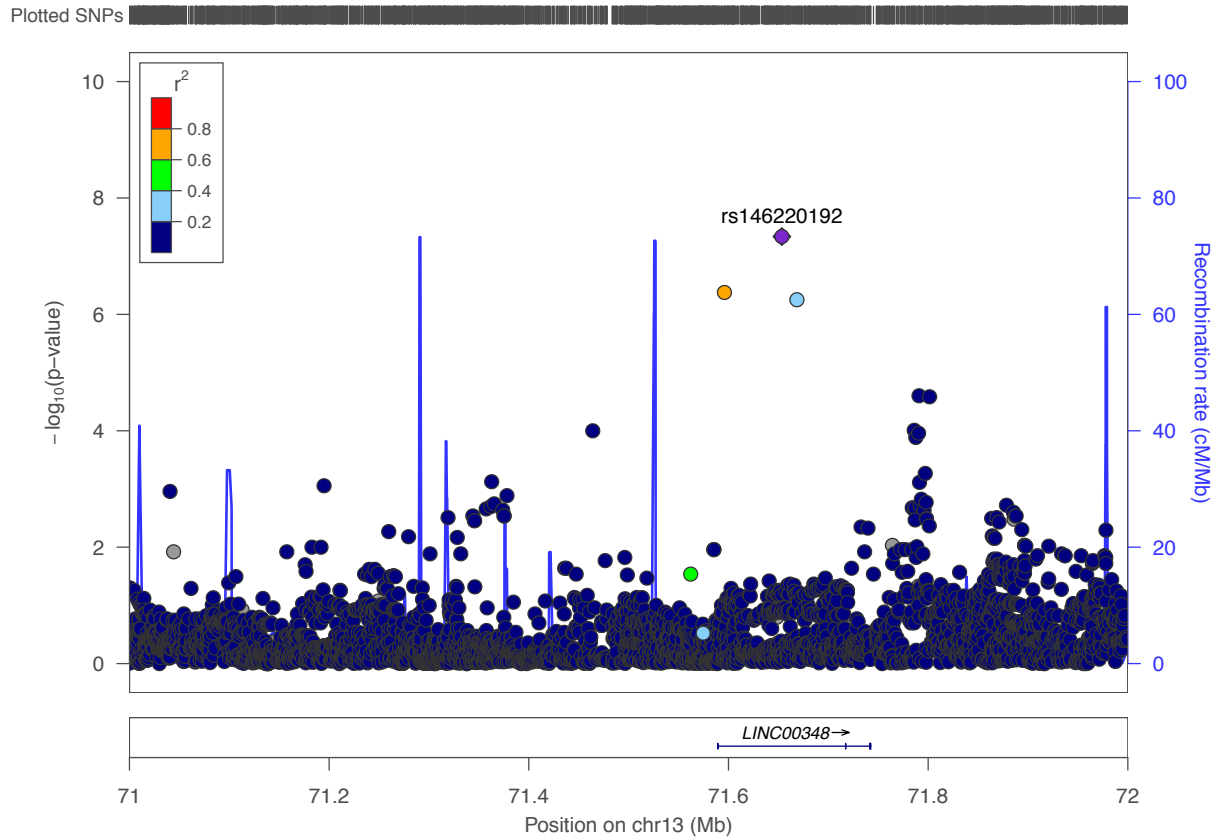


Figure 24. MGS GAD Chromosome 13 Locus. LocusZoom plot displaying strongest associated variant with coloring to indicate LD relationships with other nearby variants. Grey points indicate variants missing LD information with respect to the strongest variant. Genes in the region displayed for reference.

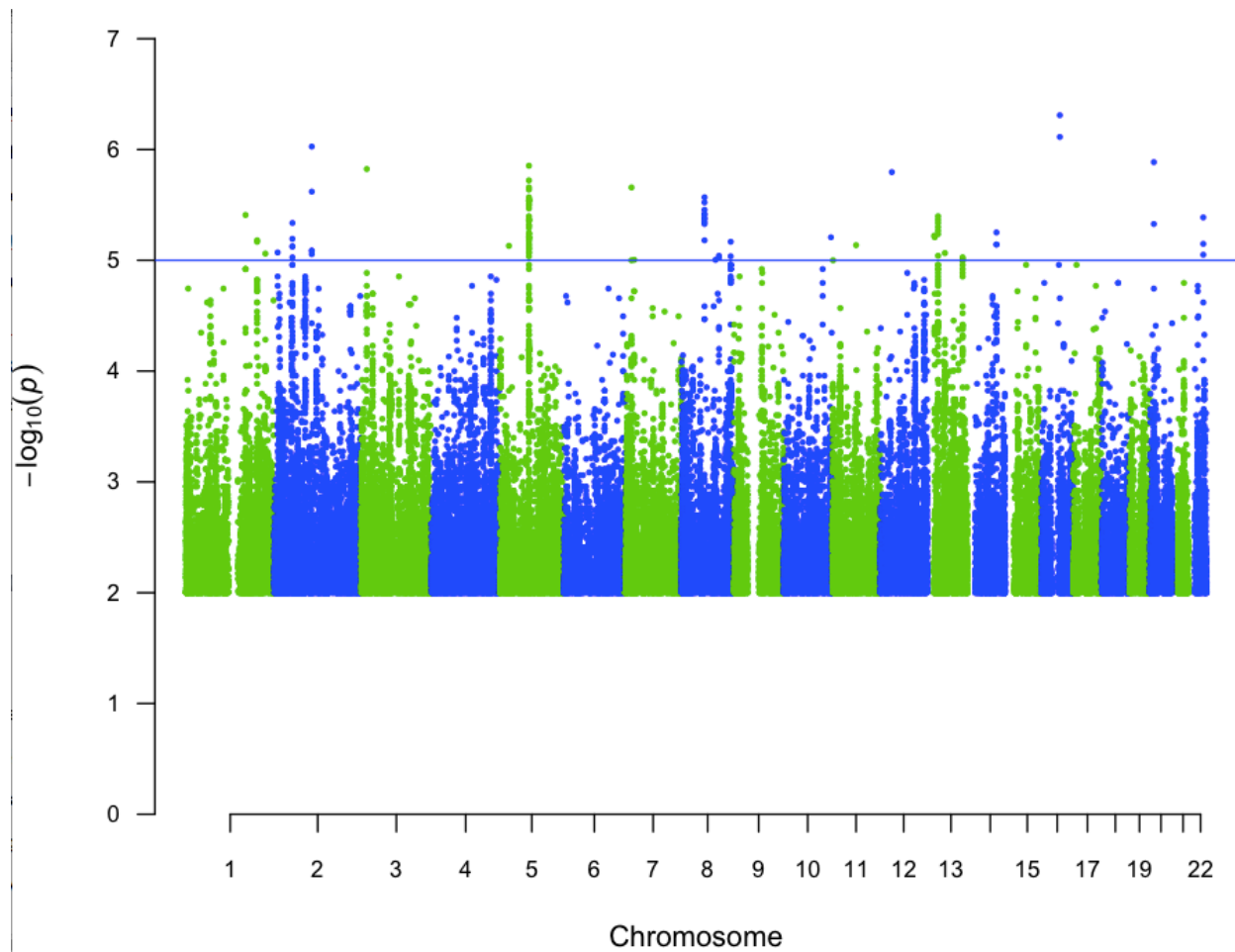


Figure 25. Neuroticism BOLT-LMM Results in MGS. Manhattan plot displaying filtered adjusted association results, arranged by physical position.

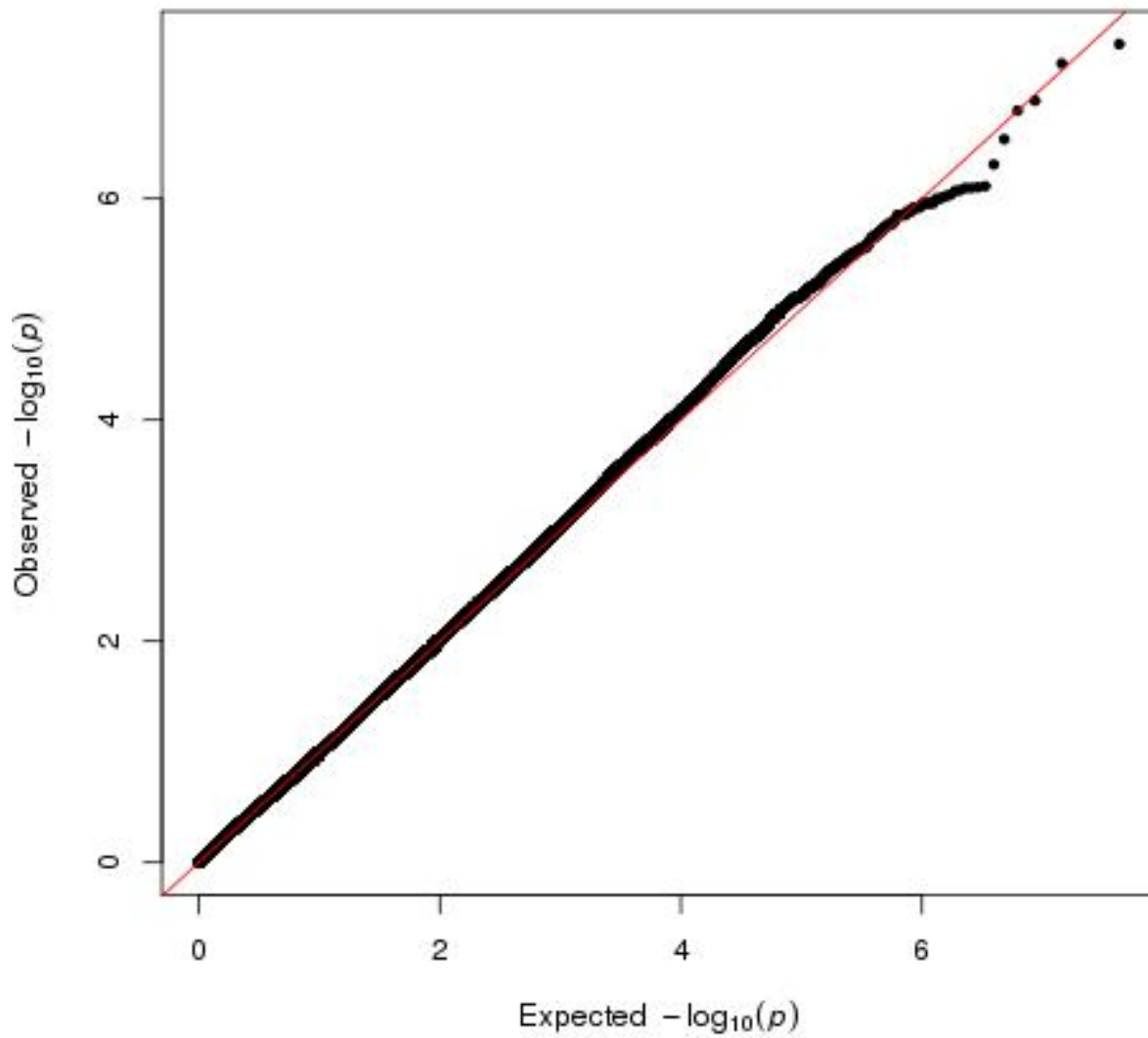


Figure 26. Neuroticism q-q Plot in MGS. Quantile-quantile plot visualizing observed versus expected association strength.

The results for the S4S sample included the fewest significant loci, with only one genomewide significant finding. The small sample size in the EAS group, and the reduced contrast between high and low levels of depressive symptoms and neuroticism in this healthy college-age population likely contributed to lack of power. Depressive symptoms results in the EAS subgroup had no loci reach significance (Figure 27). The q-q plot indicates that there may be uneven correction of the association results by PLINK using principal components (Figure 28), in addition to PLINK's reduced power compared with BOLT-LMM. Genomic inflation, however, was modest with a λ estimate of 1.004 and λ_{1000} estimate of 1.044. Depressive symptoms results in the EUR subgroup were somewhat stronger, but again none reached significance (Figure 29). The q-q plot suggests good correction (Figure 30), which is corroborated by the λ value of 1.047 and λ_{1000} of 1.021, but power appears to have been insufficient. Neuroticism results were quite weak in the EAS subgroup (Figure 31). As with Neuroticism results in other samples, the q-q plot suggests overcorrection by the mixed model approach (Figure 32), which is consistent with the λ and λ_{1000} estimate of 1 (the minimum in GenABEL). Neuroticism results in the EUR subgroup presented the only significant locus from S4S (Figure 33), on chromosome 4 (most significant variant: rs113888384, $p=1.4e-08$). This is notable due to the same pattern of overcorrection seen in the other Neuroticism association tests being evident in the q-q plot (Figure 34), though the λ estimate of 1.047 and λ_{1000} estimate of 1.021 suggest some signal. The association peak is in the collapsin response mediator protein 1 (CRMP1) gene (Figure 35). The gene is expressed exclusively in the nervous system, and is thought to be involved in neural development.

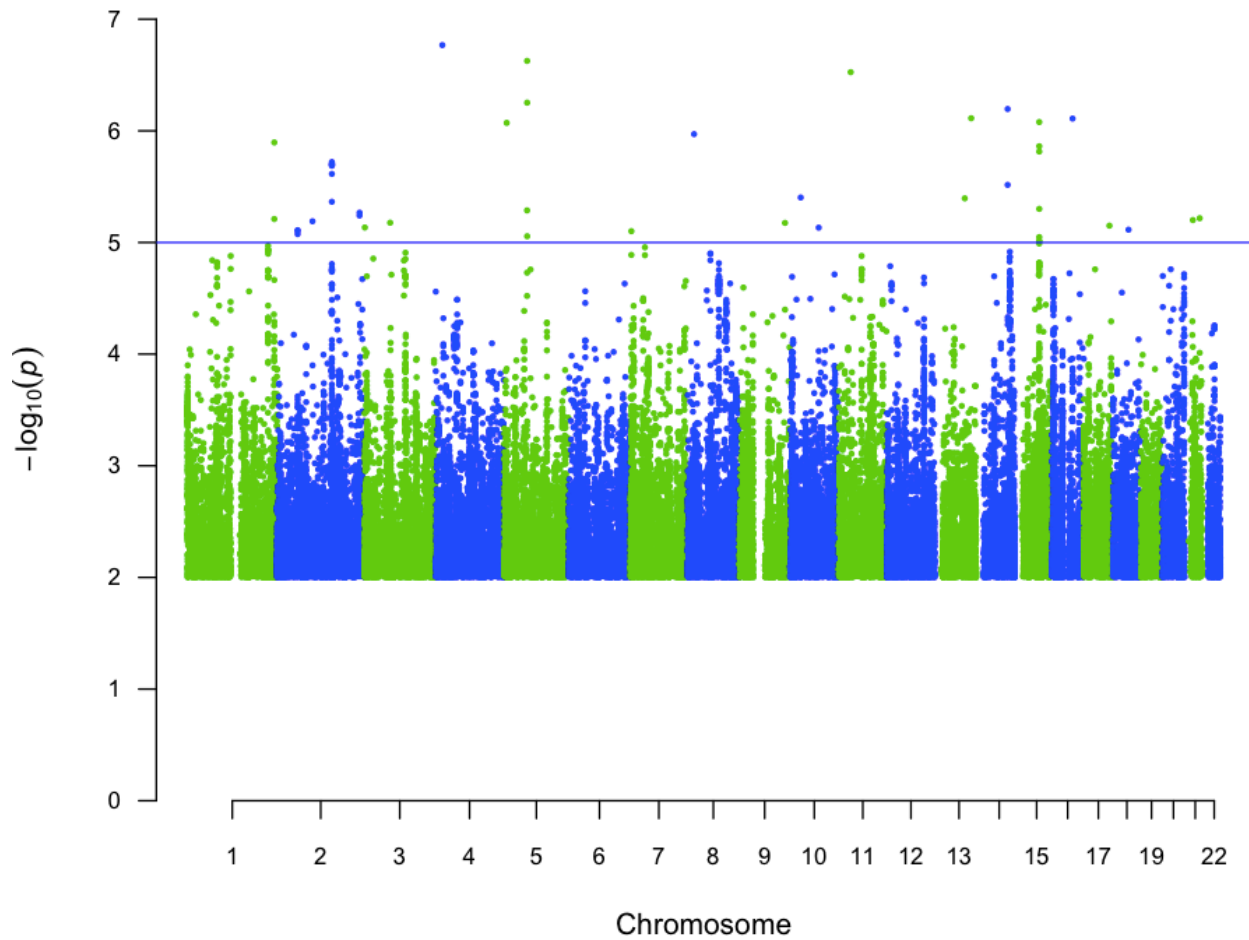


Figure 27. Depressive Symptoms PLINK Results in S4S (EAS). Manhattan plot displaying filtered adjusted association results, arranged by physical position.

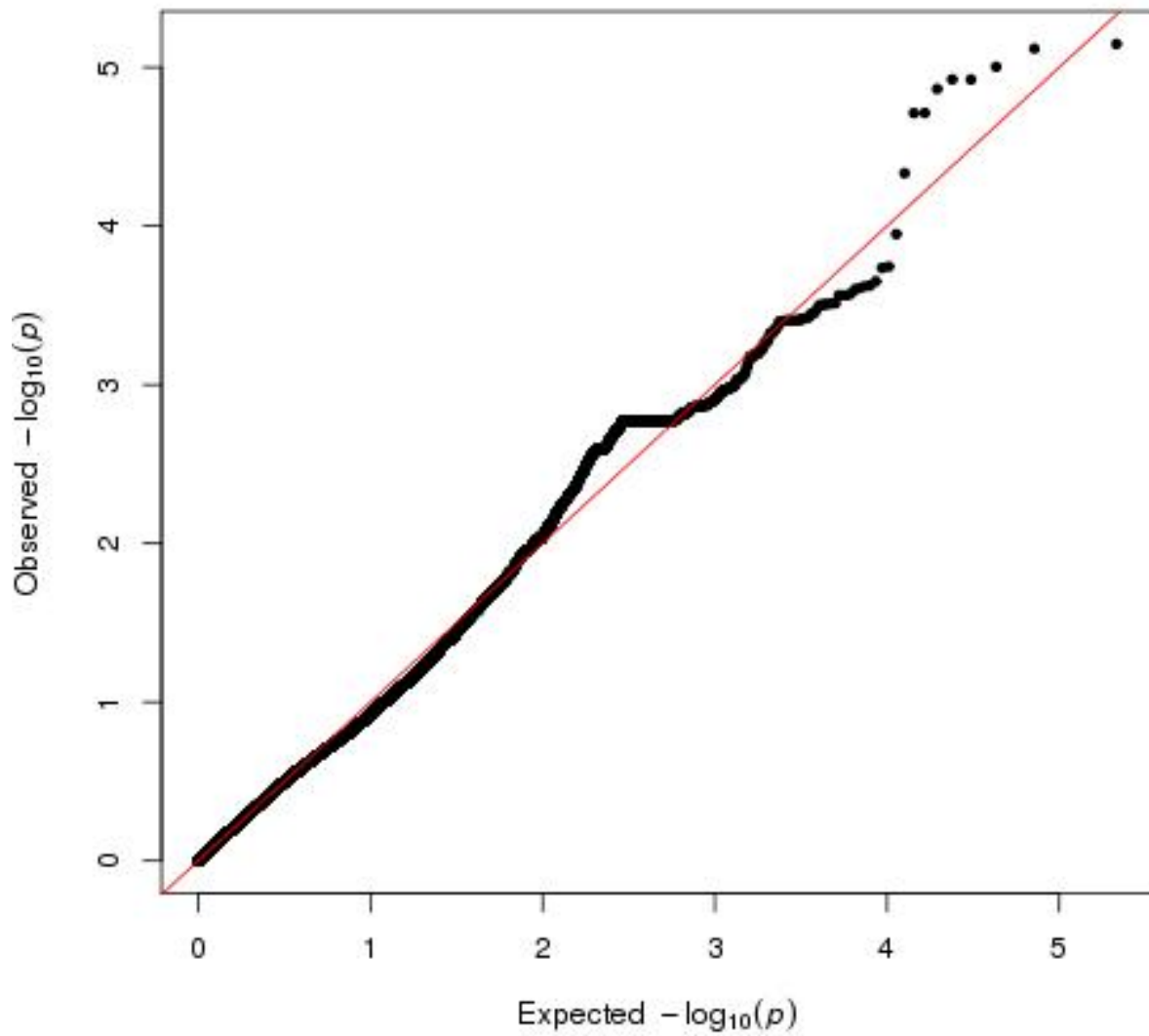


Figure 28. Depressive Symptoms q-q Plot in S4S (EAS). Quantile-quantile plot visualizing observed versus expected association strength.

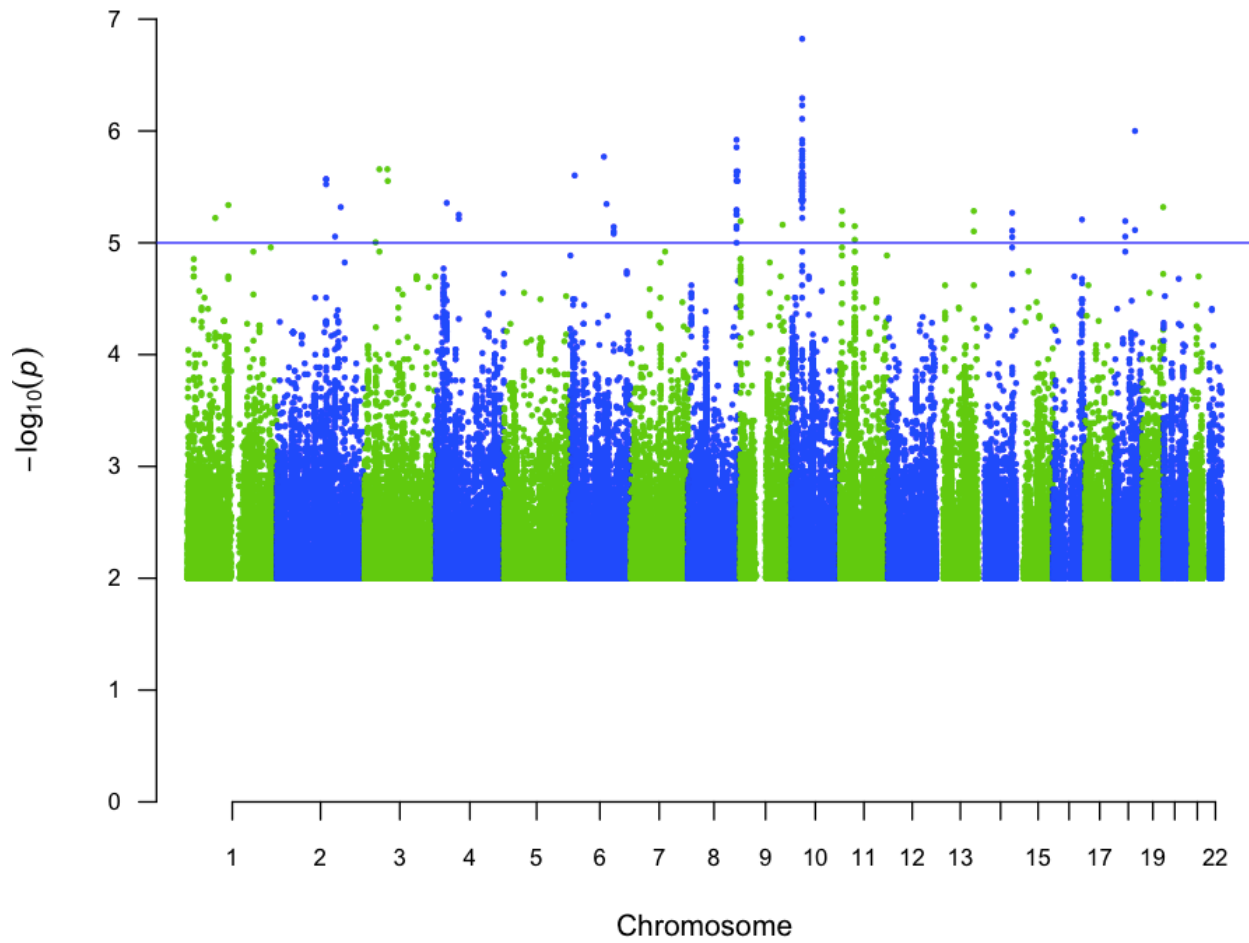


Figure 29. Depressive Symptoms BOLT-LMM Results in S4S (EUR). Manhattan plot displaying filtered adjusted association results, arranged by physical position.

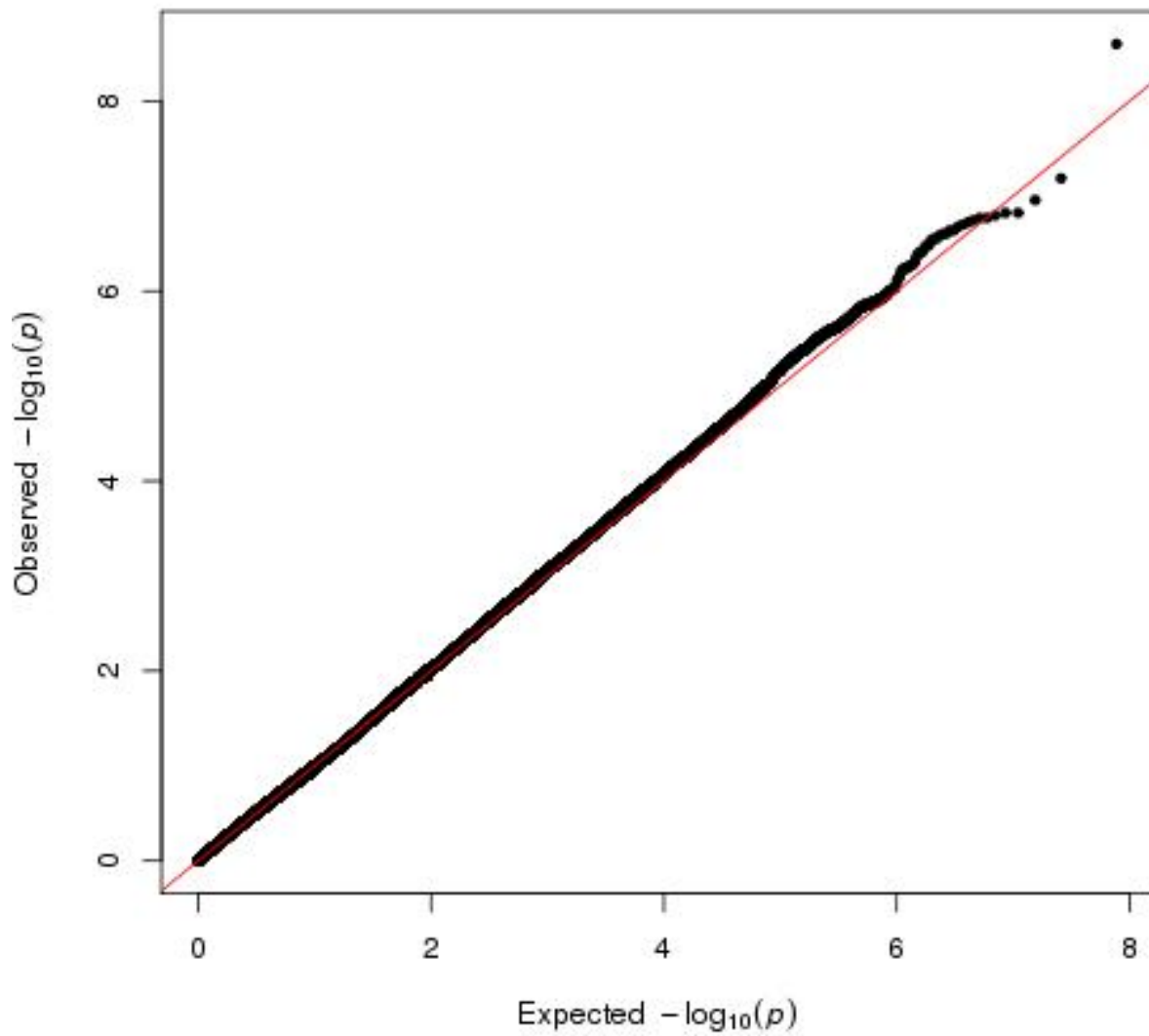


Figure 30. Depressive Symptoms q-q Plot in S4S (EUR). Quantile-quantile plot visualizing observed versus expected association strength.

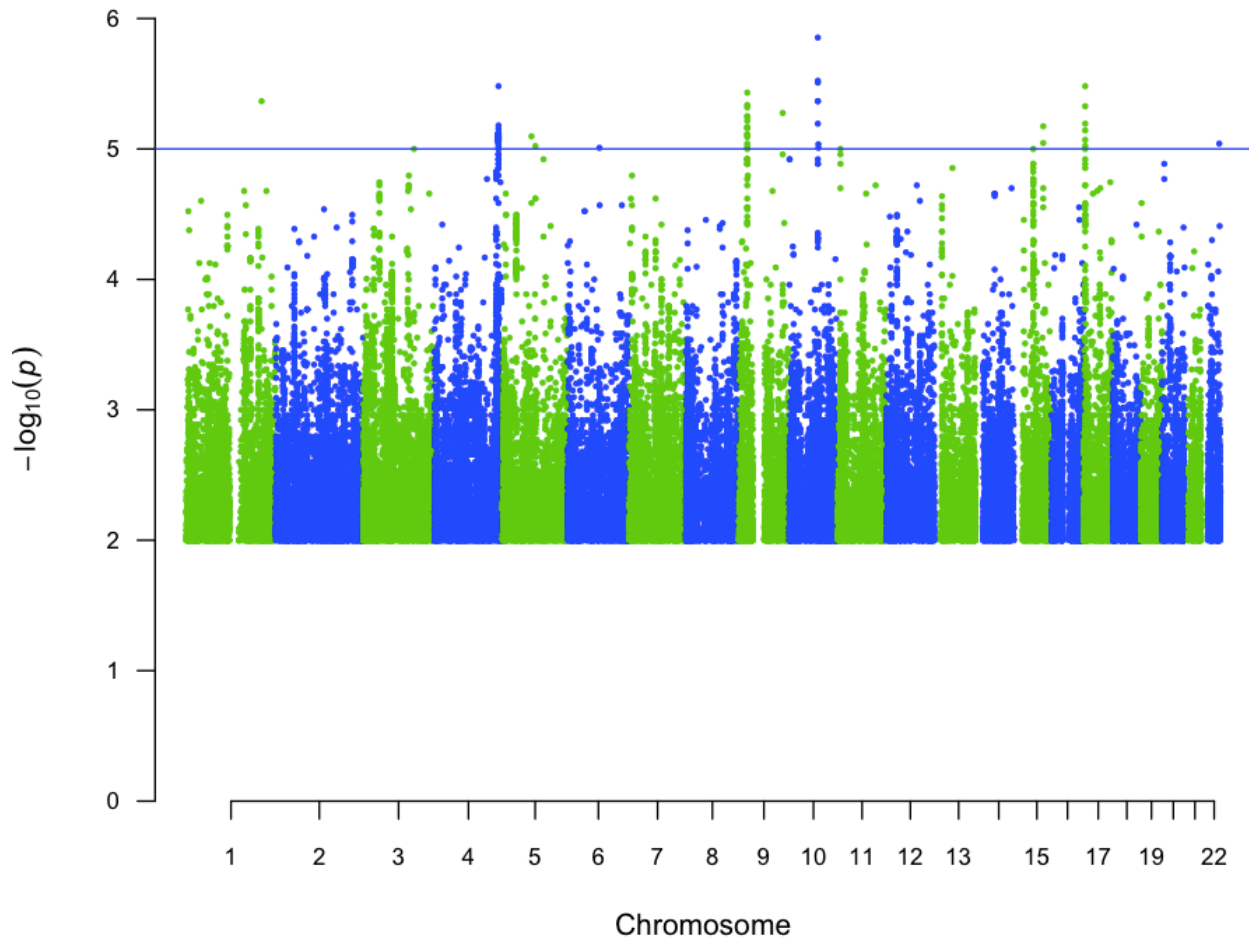


Figure 31. Neuroticism BOLT-LMM Results in S4S (EAS). Manhattan plot displaying filtered adjusted association results, arranged by physical position.

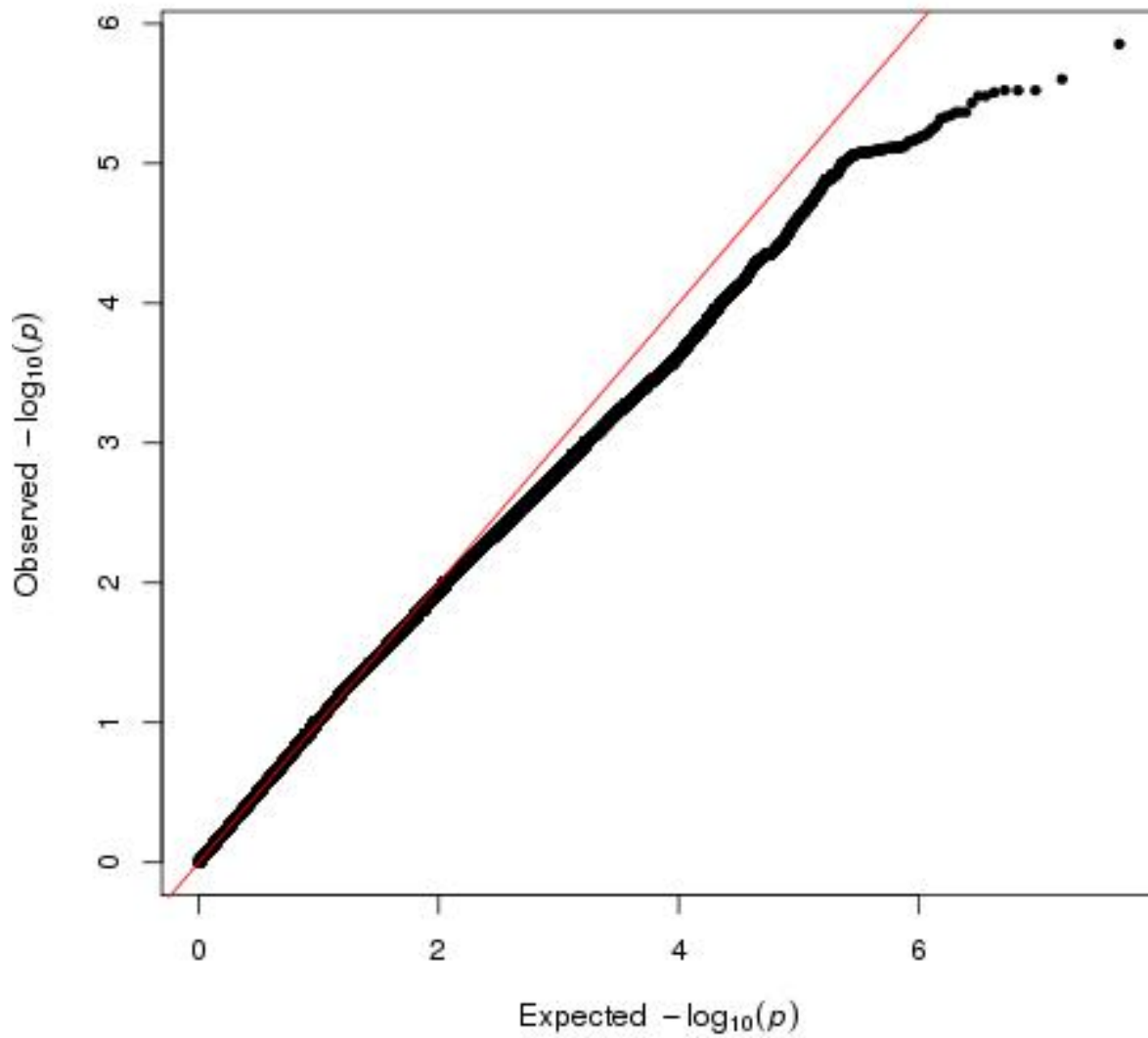


Figure 32. Neuroticism q-q Plot in S4S (EAS). Quantile-quantile plot visualizing observed versus expected association strength.

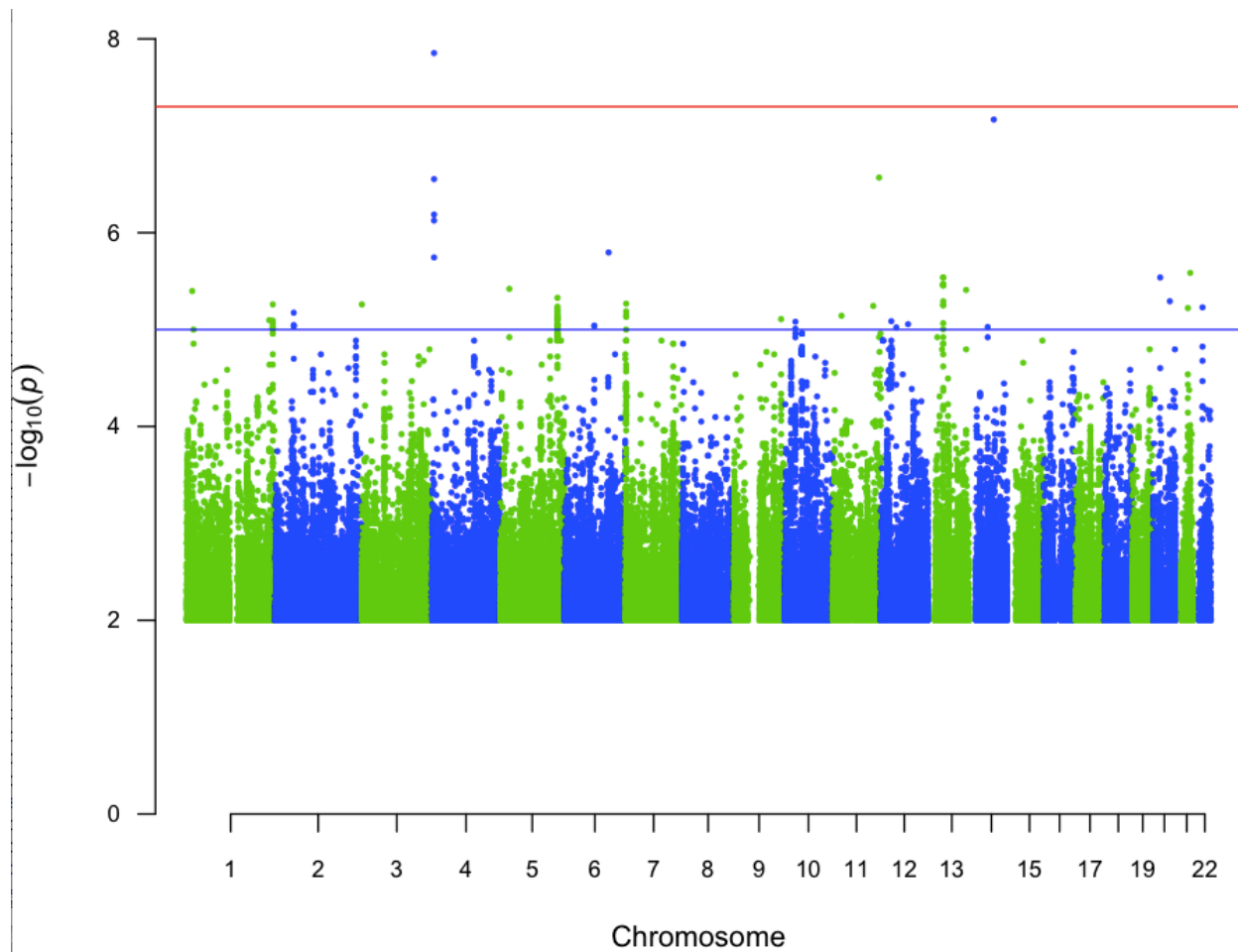


Figure 33. Neuroticism BOLT-LMM Results in S4S (EUR). Manhattan plot displaying filtered adjusted association results, arranged by physical position.

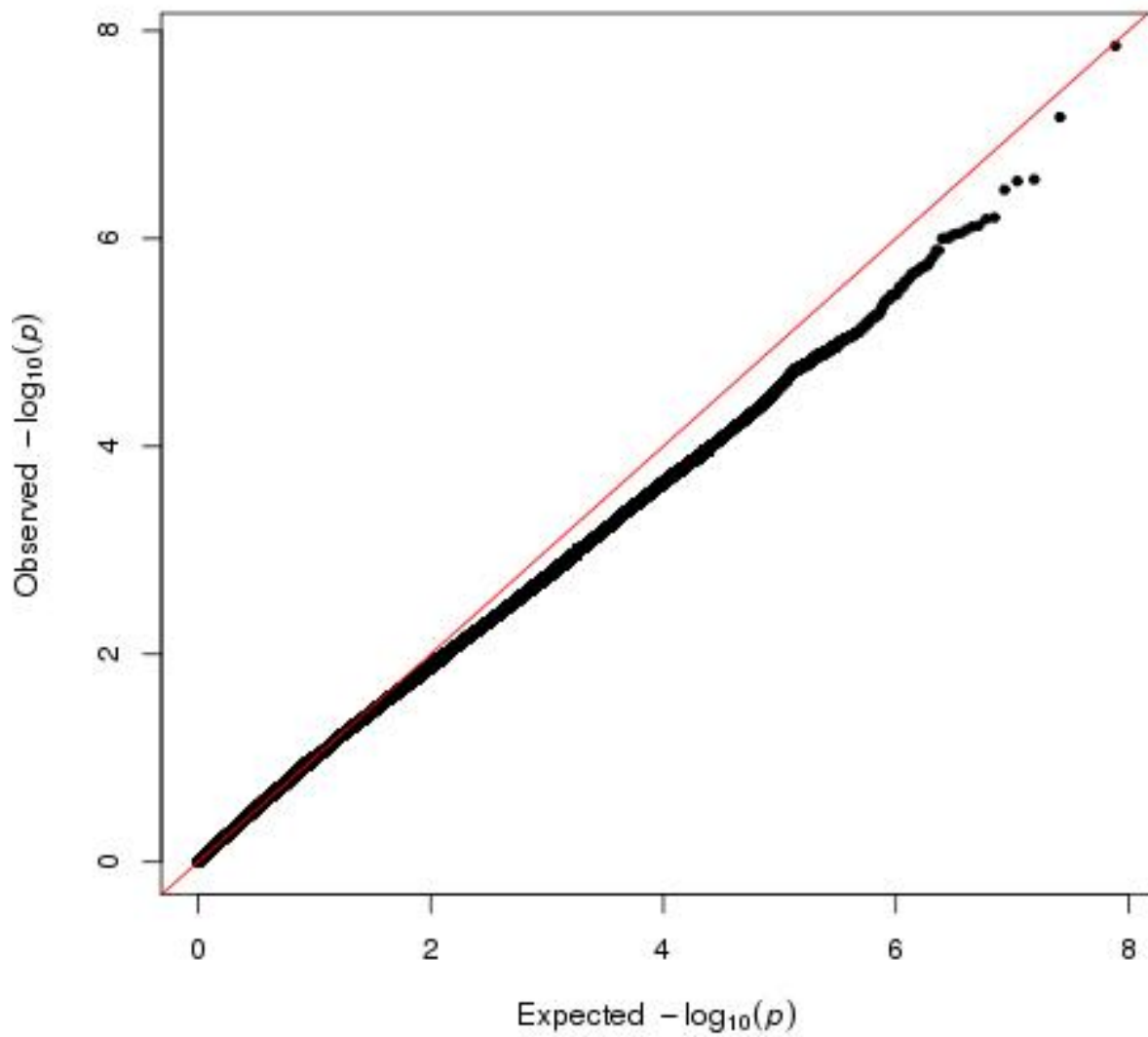


Figure 34. Neuroticism q-q Plot in S4S (EUR). Quantile-quantile plot visualizing observed versus expected association strength.

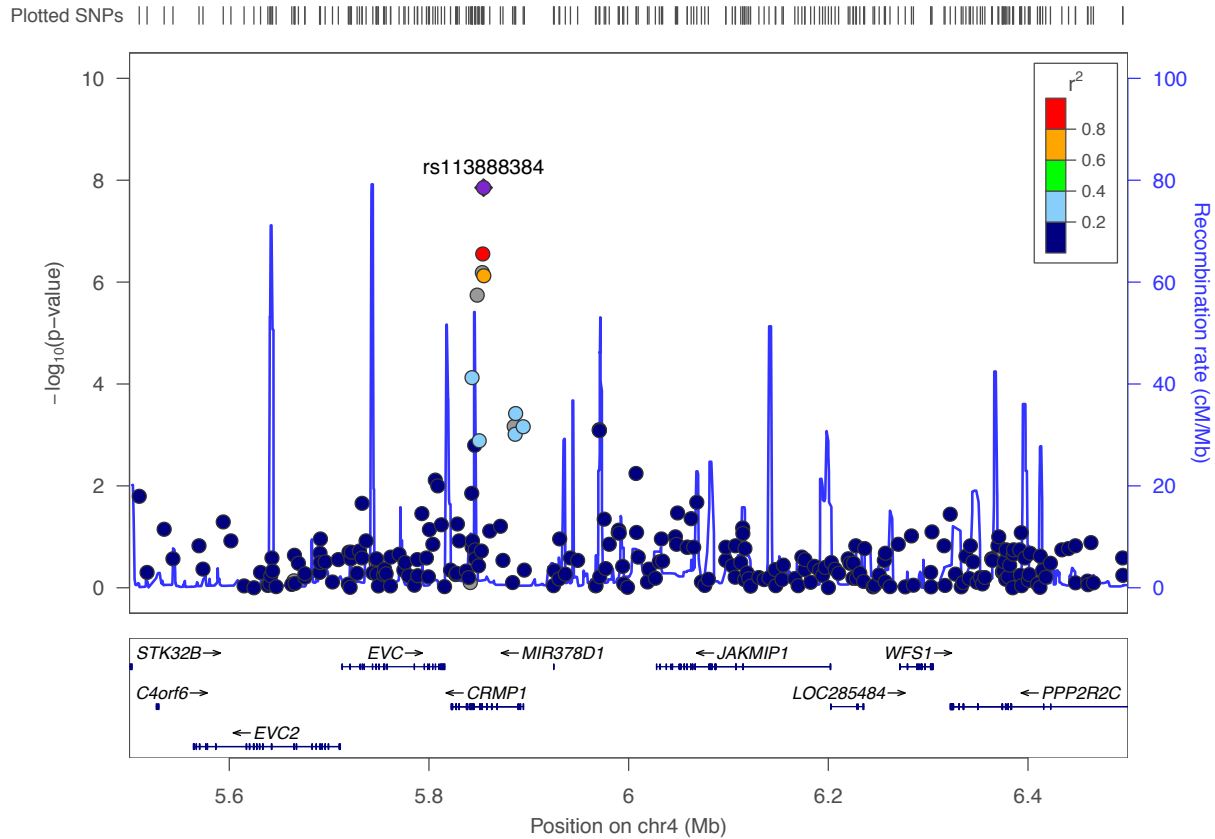


Figure 35. S4S (EUR) Neuroticism Chromosome 4 Locus. LocusZoom plot displaying strongest associated variant with coloring to indicate LD relationships with other nearby variants. Grey points indicate variants missing LD information with respect to the strongest variant. Genes in the region displayed for reference.

Interval-based results from MAGMA provide a single test statistic for all variants within a particular genomic region, allowing for biologically-relevant aggregation of signal. Intervals meeting genome-wide significance in the aggregation of BOLT-LMM association p-values are displayed in Tables 7-11, along with their p-values from the genotypic regression, a conservative test, due to low Type 1 error rate.

In CONVERGE, there were significant intervals in all outcomes. 17 intervals reached significance for MD in the p-value aggregation, with 8 reaching nominal significance in the conservative genotype regression, and two reaching genome-wide significance in both tests, an intergenic region, which does contain a number of potential regulatory regions, and a CpG island (genotypic regression p-value = $4.35e-07$), and the HECT and RLD domain containing E3 ubiquitin protein ligase 4 (*HERC4*) gene (genotypic regression p-value = $4.95e-07$), which is related to ubiquitin and is brain-expressed (Table 7).

12 intervals reached significance in the p-value aggregation analysis for MD+GAD in CONVERGE, with 2 reaching nominal significance in the genotypic regression test, *CDKN2A-AS1* and *SOX9-AS1* (genotypic regression p-values = $2.5e-04$ and 0.04, respectively), both long non-coding RNA (Table 8).

Table 7. Significant Gene-based MAGMA Results for MD in CONVERGE

Interval	BOLT-LMM-based P-value	MAGMA Regression P-value
KIAA0040	5.8167e-09	0.002968
SYT14	2.2852e-08	0.047366
HHAT	8.2152e-08	0.38987
ETV5-AS1	3.814e-07	0.00035098
ADH4	2.7816e-08	0.45402
SEMA5A	5.212e-07	0.18465
RANP1	1.2247e-08	0.51969
CTC-756D1.2	2.2481e-07	0.39747
DNAJC12	2.4477e-09	3.7593e-05
intergenic_chr10_00536*	1.096e-08	4.3519e-07
SIRT1	2.6752e-08	0.00020047
HERC4*	1.6725e-08	4.9483e-07
RP11-474D14.2	1.145e-07	1.1102e-06
ABCC8	3.0105e-124	0.35324
RP11-72M17.1	3.7146e-07	0.3233
RP11-580I1.2	2.5724e-72	0.78932
HIC2	3.4965e-53	0.59541

77,543 intervals were included in this test, corresponding to a Bonferroni significance threshold of $6.4e-07$ for a family-wise error rate of 5%.

Bold results are at least nominally significant in both tests.

* indicates genomic region met Bonferroni significance threshold in both tests.

Table 8. Significant Gene-based MAGMA Results for MD+GAD in CONVERGE

Interval	BOLT-LMM-based P-value	MAGMA Regression P-value
SYT14	2.4821e-08	0.14522
HHAT	6.1425e-08	0.37532
RANP1	2.5215e-09	0.6582
CDKN2A-AS1	3.313e-07	0.00024967
ABCC8	3.8574e-118	0.30306
OR4C11	6.6884e-10	0.071712
RP11-490P13.2	4.7155e-09	0.53378
intergenic_chr13_00386	9.9138e-08	0.12516
SQSTM1P1	3.3201e-08	0.083037
RP11-580I1.2	2.3201e-89	0.76914
SOX9-AS1	6.2512e-07	0.040681
HIC2	6.6655e-42	0.59986

77,543 intervals were included in this test, corresponding to a Bonferroni significance threshold of $6.4e-07$ for a family-wise error rate of 5%.
 Bold results are at least nominally significant in both tests.

For neuroticism in CONVERGE, 9 results were significant in the p-value aggregation, and 2 reached nominal significance in the genotypic regression test as well, *POLR1C* and *POLH* (genotypic regression p-values = 0.03 and 0.02, respectively), polymerase subunit genes (Table 9).

In MGS, only the GAD outcome displayed Bonferonni-significant findings in the p-value aggregation test, which were also nominally significant in the genotype regression test (Table 10). *LINC01346* and *RP11-1144P22.1* are long non-coding RNA, while *KLB* is involved in the synthesis of bile acid, and is expressed to some extent in the brain.

Finally, in S4S, only neuroticism in the EAS subgroup showed significant findings in the p-value aggregation test, but none reached significance in the genotypic regression (Table 11).

Table 9. Significant Gene-based MAGMA Results for Neuroticism in CONVERGE

Interval	BOLT-LMM-based P-value	MAGMA Regression P-value
ADGRV1	3.233e-07	0.1948
RANP1	2.7614e-14	0.071325
POLR1C	3.6846e-07	0.029229
POLH	4.8682e-08	0.022696
RP3-337H4.8	1.4238e-07	0.0041096
ABCC8	1.4917e-154	0.44783
IGHV1-12	1.1472e-08	0.54778
RP11-580I1.2	2.3842e-80	0.36532
HIC2	6.6655e-42	0.85815

77,543 intervals were included in this test, corresponding to a Bonferroni significance threshold of $6.4e-07$ for a family-wise error rate of 5%.
 Bold results are at least nominally significant in both tests.

Table 10. Significant Gene-based MAGMA Results for GAD in MGS

Interval	BOLT-LMM-based P-value	MAGMA Regression P-value
LINC01346	4.8836e-07	5.6434e-05
KLB	6.4351e-07	0.00042416
RP11-1144P22.1	1.5175e-07	0.00034395

73,602 intervals were included in this test, corresponding to a Bonferroni significance threshold of $6.8e-07$ for a family-wise error rate of 5%.

Bold results are at least nominally significant in both tests.

Table 11. Significant Gene-based MAGMA Results for Neuroticism in S4S (EAS)

Interval	BOLT-LMM-based P-value	MAGMA Regression P-value
CHIA	3.3799e-28	0.67696
SYT6	1.7558e-13	0.70484
GNB1L	1.6089e-11	0.36906

58,952 intervals were included in this test, corresponding to a Bonferroni significance threshold of $8.5e-07$ for a family-wise error rate of 5%.

III. Discussion:

The search for specific genetic loci with a probable impact on MD, GAD, and neuroticism is not an easy one. The vast majority of signals for complex traits have small effect sizes, as has been well-established in the GWAS literature (137). The outcomes studied here also have moderate heritability, suggesting that a substantial portion of their etiologies are environmental in origin. Nevertheless, there was some degree of promise in the findings, with 10 loci that may contain variants with some role in MD, GAD or neuroticism, and a number of genes whose variants, in aggregate, suggest a relationship with one of these outcomes as well.

Perhaps the most expected finding was the significance of previously published significant variants from MD in CONVERGE. Only one of the two loci previously found on chromosome 10 was significant in the covariate-adjusted analysis, however. The rs35936514 variant, in the *LHPP* gene, which despite having a function in phosphate processing (a process ostensibly somewhat distal to the etiology of MD), is expressed in the brain, and based upon follow-up work in the literature, the variant appears to have some effects on activity in certain brain regions (138). The other previously published variant, in the *SIRT1* gene, has a clear peak just before *LHPP* on chromosome 10 as (Figure 9), and indeed was significant when the association test was not adjusted for the age covariate. The only novel variant significant in the MD association analysis was rs7257142, in the *PLIN4* gene, on chromosome 19. The variant is not supported by other nearby suggestive signals, though LD is low in East Asian populations in genomic region around this locus. *PLIN4* is involved in adipocyte functioning, which is of unclear relevance to MD etiology; replication and further research will be needed to determine

whether the gene is likely to be a true risk locus for MD, and what role it may play in MD physiology.

MGS, despite the smaller sample size than CONVERGE, did produce a pair of significant loci as well. Only the GAD outcome produced significant results in this sample. The most encouraging peak was found for rs117420048, in the intron of *LINC00824*, a long non-coding RNA on chromosome 8. The other significant variant was rs146220192, in *LINC00348*, another long non-coding RNA on chromosome 13. The LD patterns of supporting signals in these loci are not as strong as the chromosome 10 CONVERGE MD locus, and the potential functional roles of long non-coding RNA as a class are highly varied, with established functions still missing for many of them (139). Thus, the etiological significance of these loci is unclear.

In S4S, only one locus reached significance, which was for neuroticism in the EUR subgroup. The top variant was rs113888384, in the *CRMP1* gene on chromosome 4, which is involved in the collapse of the growth cone in neural development. As this is believed to be an important part of axon guidance in the brain, the relevance to neuroticism as a dispositional trait corresponding to liability to internalizing psychopathology is clear; indeed, *CRMP1* has been implicated in connection with schizophrenia and major depression (140)

In comparing results to past work, such as the published variants listed in Table 5, the majority did not replicate. This is likely due more to the limited sample size of the present analysis than high rates of false positives in past published work, but further analyses will be required to determine which results are true positives. A total of four variants from a large meta-analysis (95) did replicate in at least one test, however. The

neuroticism variants of rs2150472 and rs12903563 replicated at suggestive significance in the CONVERGE neuroticism analysis (p -values of 4.7×10^{-3} and 1.4×10^{-3} , respectively) which, it is worth noting, would constitute a transethnic replication from European to East Asian ancestries, if true. The neuroticism variant of rs2572431 replicated at suggestive significance in the MGS neuroticism analysis ($p = 4.9 \times 10^{-3}$), and, surprisingly, the variant rs7973260 associated with depressive symptoms replicated with suggestive significance in the S4S EAS subgroup ($p = 8.9 \times 10^{-3}$), which is both East Asian in ancestry and of small sample size. These consistencies with past work are somewhat encouraging, as one would not expect any variants to replicate in samples the size of MGS and S4S, but the weak p -values indicate that these ostensible replications may simply be due to chance variation.

Interval-based results from MAGMA are similarly mixed, with some promising and some probably spuriously significant results. When the initial BOLT-LMM follow-up method of aggregating p -values returned a significant result after multiple testing correction, validation by looking at the regression-based p -value provides a reasonable check of whether the result is likely to be a true positive. The most significant intervals were found for CONVERGE MD, which is unsurprising given that this association analysis was the one with the highest power in the whole project. 8 of these regions retained nominal significance in the regression-based test. Some of these were obscure in function: *KIAA0040* is a gene encoding an uncharacterized protein and *ETV5-AS1* is a long non-coding RNA. Others were seemingly more related to MD, such as *SYT14*, a gene involved in synaptic transmission. Notably, there was a collection of significant intervals in the same genomic region of chromosome 10, which have some overlap:

DNAJC12 is a protein in the heat shock family which is highly brain-expressed, an intergenic region in the area containing regulatory markers but no transcripts, *HERC4* is a brain-expressed ubiquitin ligase, a catalyst for certain reactions involving ubiquitin; *RP11-474D14.2* is a pseudogene in an intron of *HERC4*, and *SIRT1*, of course, is the locus which was genome-wide significant in the previously published CONVERGE paper (93), but just failed to meet the significance threshold in the covariate-adjusted association analysis. Two of these, the intergenic region in this area, and *HERC4*, actually met genome-wide significance in both interval-based tests, suggesting this region likely has some role in MD physiology.

Tests in the other outcomes in CONVERGE produced a couple of significant intervals each. For the MD+GAD outcome, two long non-coding RNAs were at least nominally significant in both MAGMA test methods: *CDKN2A-AS1* and *SOX9-AS1*. In CONVERGE neuroticism, two genes involved in the production of polymerase units came up, *POLR1C* and *POLH*. These are genes whose role in psychopathology is difficult to discern; it's unclear what the role of these particular long non-coding RNA may be, given the highly variable and poorly understood functions of many long non-coding RNA, as mentioned above (139). Conversely, mutations in polymerase enzymes are known to be involved in certain Mendelian conditions, such as Treacher Collins syndrome (141) and Xeroderma Pigmentosum (142), but no evidence for a relationship with personality has been found previously.

Of the non-CONVERGE interval-based results, only the GAD outcome in MGS produced any significant regions, though all three met nominal significance thresholds in both MAGMA tests. All seem to lack obvious connection to GAD, with two more long

non-coding RNA in *LINC01346* and *RP11-1144P22.1*, and *KLB*, a gene involved in the synthesis of bile acid expressed in a variety of tissues including the brain.

The minimal overlap between the variant-based and interval-based tests is perhaps unexpected, though the goals of the two methods are somewhat different, if related. The variant-based tests explore the statistical effect of specific polymorphisms on the outcome of interest. The interval-based tests examine the degree to which variation in an entire genomic region is related to the outcome. Instances where there are polymorphisms of (relatively) high impact in a tightly-defined area, such as an exon may produce an individually-significant variant. Alternatively, the density of associated polymorphisms with sub-threshold effects in the region may suffice to produce a significant interval even without any variants of genome-wide significance. Of course, a sufficiently impactful area of the genome may produce both results – the *SIRT1* area on chromosome 10 produced a significant peak in the CONVERGE MD association test without age adjustment, plus multiple significant intervals in the same vicinity (including the *SIRT1* gene itself), two with genome-wide significance in both interval-based tests.

The specific genetic loci identified in this chapter are initial efforts in detailing the modest proportion of variance in MD, GAD and neuroticism attributable to genetic factors in biometrical studies (81, 84, 87). Some are truly promising, such as strong findings in CONVERGE (some of which have already been included in prior manuscripts), while others are potentially artefactual. Due primarily to the limited sample size in this analysis, few of the results are strong and unambiguous. It is wise, therefore, to not rely on the ultimate truth of any of these findings until they have been convincingly replicated in other future studies.

Chapter 4: Estimation of genetic variance, correlation, and cross-prediction in MD, GAD, and neuroticism.

I. Methods:

Samples: As with Chapter 3, all datasets were used in at least some of this set of analyses, including CONVERGE, MGS, and S4S. CONVERGE comprises information from Han Chinese females of ages 30-60. MGS comprises information from European individuals from ages 18-90. There are 1355 females and 1257 males used in this analysis from the MGS sample. S4S comprises information from individuals of multiple ancestries from ages 18-32. There are 3016 individuals of European ancestry (EUR), 557 individuals of East Asian ancestry (EAS), 1471 males, and 2069 females used in this analysis from the S4S sample cohorts 1-3, which were the subset included in these analyses. S4S cohort 1 had responded to 3 years of survey follow-up, cohort 2 had responded to 2 years of survey follow-up, and cohort 3 had responded to 1 year of survey follow-up when data was assembled for these analyses.

Variables: Most of the analyses in this chapter use genetic information to estimate overall quantitative values related to the main outcome phenotypes, singly or in pairs. These outcomes include MD, GAD, comorbid MD+GAD, neuroticism, and

depressive symptoms. For polygenic scoring, two sets of external summary statistics were used for the discovery phenotypes of depressive symptoms and neuroticism (95).

Analyses: Estimation of univariate genetic variance was undertaken with GCTA (99), and LDSC (100). GCTA is a method that uses a genetic relatedness matrix, a representation of overall identical-by-state (IBS) relatedness (whether a pair of individuals has zero, one, or two alleles at a particular polymorphic genomic locus in common), between each pair of individuals in the sample. LDSC uses linkage disequilibrium relationships between a variant and nearby regions to calculate a single quantitative index of how much LD a variant has; the 'LD score'. This LD score is compared with the level of association the locus has with the outcome, with respect to the intuition that, for a polygenic trait, a locus with more LD should have a larger effect. The differential between this expected relationship and the observed pattern of associations is used to estimate genetic variance. Both methods were used in order to provide convergent evidence of the precise heritability estimate for each outcome, and also due to differences in implementation potentially allowing one to succeed in an edge case where the other fails.

For bivariate estimation of genetic correlation, GCTA (143) and LDSC (144) were used once more. GCTA uses a bivariate linear mixed model, where polygenic effects on both outcomes are used for individuals to estimate an average information matrix, which is used to calculate the genetic correlation value. LDSC uses a similar intuition to the univariate case, but instead of single association results being compared to LD scores, it is products of the two z-scores, which likewise should, all else being equal, covary

with LD score, with larger z products than expected indicating increased genetic variance in those regions.

Polygenic scoring was undertaken using the LDpred (145) package in Python (www.python.org) was used for these analyses because of its ability to account for LD structure in population samples, and its use of all genetic variants (lack of an arbitrary p -value threshold for inclusion of the genetic variants in the polygenic score). LDpred allows for the modeling of LD based on LD in the discovery sample to weight the relative contributions of nearby variants to the outcome phenotype. LDpred uses postulated proportions of causal variants in the genome as Bayesian prior probabilities for GPS calculations, and a range of 8 different priors were used to construct scores; proportions of 1, 0.3, 0.1, 0.03, 0.01, 0.003, and 0.001, as well as the model of infinite variants of infinitesimally small effect(128). Using polygenic scores for depressive symptoms and neuroticism calculated from summary statistics derived from a large meta-analysis of samples from a number of different sources (95) as predictors, regressions were run using R (www.r-project.org) to determine prediction of the primary outcomes. These scores could not be used validly to predict into the MGS sample, however, since MGS was included, as part of the Psychiatric Genomics Consortium (146), in the meta-analysis that produced the summary statistics.

Finally, in order to directly compare the patterns of GAD comorbidity with respect to MD versus MD comorbidity with respect to GAD, the method of Breaking Up Heterogeneous Mixture Based On Cross-locus correlations (BUHMBOX) was used (147). BUHMBOX is designed to detect heterogeneity in the presence of pleiotropy, where heterogeneity refers to disease cases having cryptic subgroups that may differ in

their genetic architecture from the main group of cases, and pleiotropy refers to the association of some genetic variants with more than one outcome. In a sample of disease cases, such as those affected with MD, evaluating the correlations between loci for another disease, such as schizophrenia, allows a determination of whether there are likely to be subgroups of MD that have a similar genetic architecture to schizophrenia. This is one application BUHMBOX is used for in the initial paper (147), with a non-significant result, indicating that (in the tested sample), MD cases had no hidden subgroup of individuals with similar genetic architecture to schizophrenia cases, though power was limited.

II. Results:

Univariate heritability estimates were calculated using both GCTA and LDSC. Some estimates were not precise enough to be distinguished from 0. All estimates are presented in Table 12. GCTA and LDSC estimates roughly agree, though only estimates in CONVERGE reached significance. In CONVERGE, MD and comorbid MD+GAD were estimated in the range of 0.2 and 0.25, indicating modest heritability in both GCTA and LDSC. Neuroticism was estimated at a lower level still, about 0.1, and due to the larger standard errors in LDSC, only reached significance in GCTA. Estimates were not significant in MGS or S4S samples, likely due to small sample sizes resulting in larger standard errors. Clearly, a heritability estimate that is below 0 or greater than 1 is nonsense (though some estimates pass these thresholds in Table 12); due to a quirk of LDSC's estimation methodology, these limits are not firmly fixed, leading to occasional impossible estimates. Note, however, that in every case the

standard errors are large enough to include potentially sensible values, therefore this indicates low precision due to low power, rather than paradoxical heritability values.

Bivariate estimates were only estimable for CONVERGE outcomes, due to sample size and only between neuroticism and the other outcomes, MD and MD+GAD, since MD+GAD cases are a subset of MD cases. One of the central questions of the genetic correlation between MD and GAD (which only had suitable data in MGS) did not produce an estimate, due to insufficient sample size. Moreover, in CONVERGE, the correlation between MD and neuroticism also failed to estimate in GCTA, due to the information matrix not being invertible. It's possible that this was due to a separation in neuroticism values between MD cases and controls – only a handful of MD cases had the lowest values of neuroticism, and no MD controls had the highest values, leading to a kind of bimodal distribution conditional on MD status (see Figure 36). The GCTA estimate for genetic correlation between the MD+GAD comorbidity and neuroticism was 0.873 (SE 0.031, p-value for being non-zero: $1.74e-174$), suggesting high genetic overlap. The LDSC estimate for genetic correlation between MD+GAD and neuroticism was 1.134 (SE 0.298, p-value for being non-zero: $1.41e-04$), and the genetic correlation between MD and neuroticism (successfully estimated in LDSC) was 1.086 (0.214, p-value for being non-zero: $3.88e-07$). As with the univariate estimates, a genetic correlation greater than 1 is impossible, but LDSC's estimation procedures can generate these values when underpowered. It is clear that both MD and comorbid MD+GAD have substantial (non-zero) genetic correlation with neuroticism, but whether the true correlations are high (~ 0.8) or unity (~ 1.0) cannot be established in this analysis.

LDpred polygenic score regression coefficients and p-values can be found in Table 13. The eight prior causal allele proportions can be construed as hypotheses about the genetic architecture of the discovery phenotype. Note that the exact effect estimates are less pertinent than the direction of effect and significance, due to the different scales of the polygenic scores and outcomes. The strongest results show significant prediction at all or most prior levels. Both depressive symptoms and neuroticism polygenic scores displayed strong prediction of phenotypic MD, neuroticism, and depressive symptoms, in both CONVERGE and S4S (EUR subgroup). The MD+GAD outcome in CONVERGE was additionally predicted by the polygenic score for neuroticism but not by the depressive symptoms polygenic score. The EAS subgroup of S4S, however, had insufficient sample size for significant effects in any outcome.

BUHMBOX computed levels of positive correlations between MD-associated variants in GAD cases, and vice-versa, in the MGS sample, which was the only sample containing independent MD and GAD groups. There was moderate evidence for heterogeneity in GAD cases with respect to MD (p-value 0.009), and also for pleiotropy (Mendelian randomization p-value 3e-05). There was much stronger evidence for heterogeneity in MD cases with respect to GAD (p-value 4e-20), as well as pleiotropy (Mendelian randomization p-value 5e-19).

Table 12. Univariate heritability estimates from GCTA and LDSC

Sample & Outcome	GCTA Estimate (SE, P)	LDSC Estimate (SE, P)
CONVERGE MD	0.225 (0.029, 4.29e-15)	0.256 (0.047, 2.56e-08)
MGS MD	0.000 (0.186, 0.5)	-0.244 (0.248, 0.837)
CONVERGE MD+GAD	0.227 (0.051, 4.27e-06)	0.233 (0.062, 8.56e-05)
MGS MD+GAD	0.127 (0.244, 0.301)	0.161 (0.371, 0.332)
MGS GAD	0.000 (0.225, 0.5)	0.116 (0.359, 0.373)
CONVERGE Neuroticism	0.090 (0.028, 6.54e-04)	0.073 (0.046, 0.056)
MGS Neuroticism	0.000 (0.125, 0.5)	0.141 (0.193, 0.233)
S4S EAS Neuroticism	0.156 (0.403, 0.349)	-2.118 (1.498, 0.921)
S4S EUR Neuroticism	0.103 (0.128, 0.211)	-0.262 (0.228, 0.875)
S4S EAS Depressive Symptoms	0.000 (0.448, 0.5)	1.373 (1.408, 0.165)
S4S EUR Depressive Symptoms	0.142 (0.133, 0.143)	0.020 (0.226, 0.465)

Bold values indicate an estimate significantly different from 0.

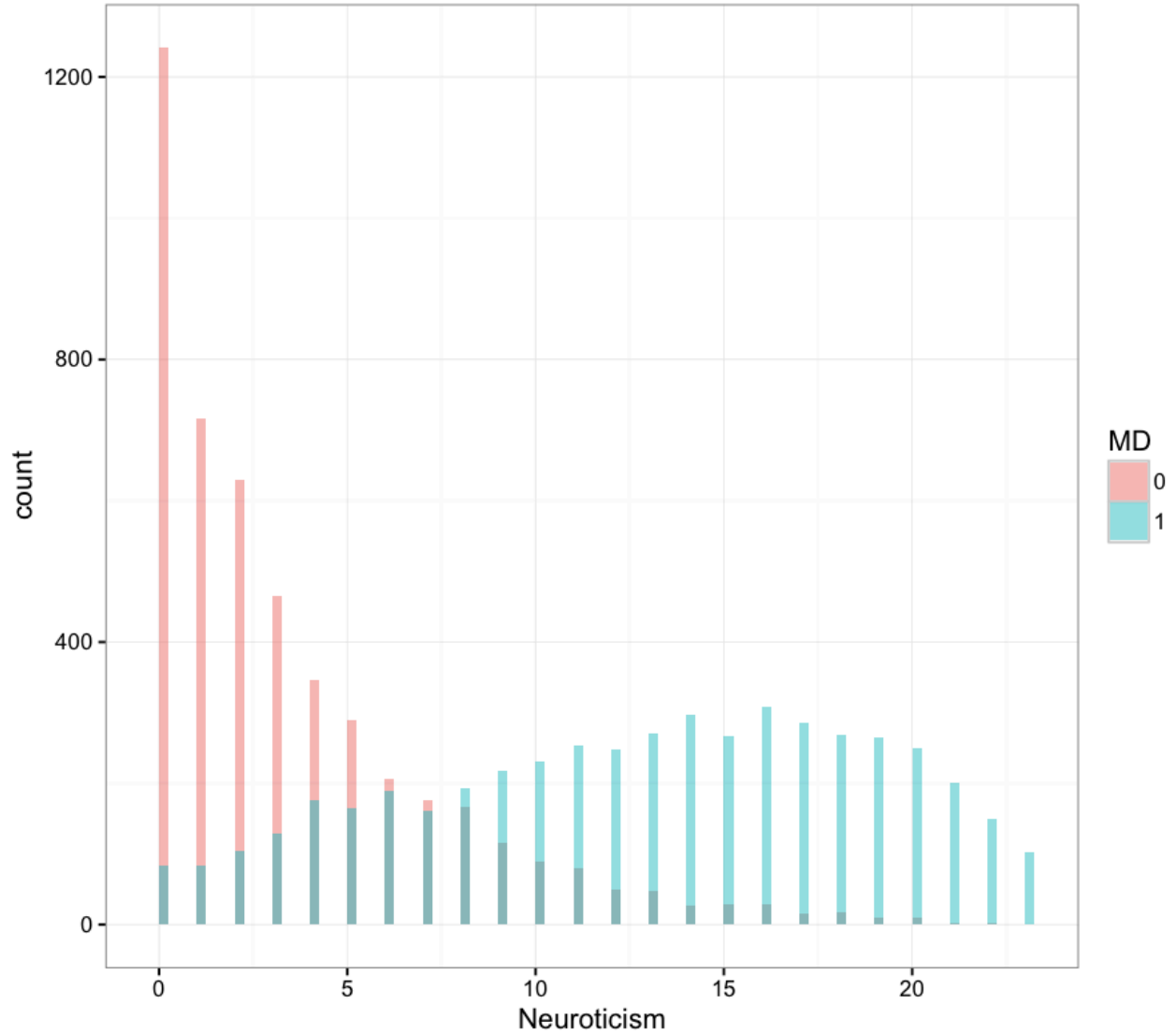


Figure 36. Stratified histogram of neuroticism values with respect to MD status.

Red bars indicate quantity of MD controls at each value; blue bars indicate quantity of cases.

Table 13. Associations between LDpred polygenic scores and outcomes at each prior level – proportion of causal alleles

Discovery Phenotype	Test Phenotype and Sample	inf.	1.0	0.3	0.1	0.03	0.01	0.003	0.001
Depressive Symptoms	CONVERGE MD	3.47 (0.006)	1.23 (0.004)	1.23 (0.004)	1.23 (0.004)	1.22 (0.004)	1.19 (0.006)	1.01 (0.016)	0.72 (0.040)
	CONVERGE MD+GAD	2.57 (0.203)	0.95 (0.160)	0.96 (0.158)	0.96 (0.159)	0.99 (0.150)	1.01 (0.144)	0.98 (0.142)	0.93 (0.101)
	CONVERGE Neuroticism	12.07 (0.004)	4.05 (0.004)	4.06 (0.004)	4.07 (0.004)	4.09 (0.004)	4.12 (0.004)	3.69 (0.008)	2.35 (0.046)
	S4S EAS Neuroticism	0.035 (0.988)	0.190 (0.808)	0.197 (0.801)	0.183 (0.815)	0.195 (0.805)	0.134 (0.866)	0.104 (0.893)	-0.052 (0.938)
	S4S EUR Neuroticism	1.788 (0.055)	0.868 (0.005)	0.870 (0.005)	0.872 (0.005)	0.886 (0.004)	0.913 (0.003)	0.902 (0.003)	0.799 (0.003)
	S4S EAS Dep. Symptoms	2.430 (0.340)	0.998 (0.238)	1.007 (0.234)	1.004 (0.236)	1.000 (0.241)	0.967 (0.262)	0.827 (0.325)	0.154 (0.833)
	S4S EUR Dep. Symptoms	2.905 (0.001)	1.081 (4e-04)	1.083 (4e-04)	1.087 (4e-04)	1.104 (3e-04)	1.133 (2e-04)	1.088 (3e-04)	0.831 (0.002)
Neuroticism	CONVERGE MD	1.80 (0.001)	0.74 (3e-04)	0.75 (3e-04)	0.75 (3e-04)	0.76 (3e-04)	0.80 (1e-04)	0.60 (0.002)	0.00 (0.366)
	CONVERGE MD+GAD	2.32 (0.009)	0.90 (0.006)	0.90 (0.006)	0.90 (0.006)	0.92 (0.006)	0.98 (0.004)	0.71 (0.024)	-0.01 (0.337)
	CONVERGE Neuroticism	8.24 (7e-06)	3.12 (5e-06)	3.13 (5e-06)	3.18 (4e-06)	3.23 (4e-06)	3.37 (2e-06)	2.80 (2e-05)	-0.01 (0.400)
	S4S EAS Neuroticism	-1.010 (0.350)	0.109 (0.781)	0.111 (0.777)	0.106 (0.789)	0.112 (0.780)	0.100 (0.806)	0.050 (0.895)	0.011 (0.433)
	S4S EUR Neuroticism	1.249 (1e-03)	0.574 (7e-05)	0.574 (7e-05)	0.579 (7e-05)	0.593 (5e-05)	0.598 (5e-05)	0.002 (0.759)	0.003 (0.386)
	S4S EAS Dep. Symptoms	0.968 (0.405)	0.636 (0.132)	0.639 (0.130)	0.646 (0.129)	0.693 (0.109)	0.751 (0.085)	0.657 (0.105)	-0.005 (0.737)
	S4S EUR Dep. Symptoms	1.623 (1e-05)	0.746 (e-07)	0.748 (1e-07)	0.754 (1e-07)	0.769 (8e-08)	0.767 (1e-07)	0.010 (0.149)	0.003 (0.259)

The columns headed 'inf' or with proportions correspond to the tested prior levels.
Regression coefficients are presented, with p-values in parentheses.
Bold entries are significant at p-value <0.05.

III. Discussion:

The aggregate use of genomic information allows singular values to address such complicated questions as the degree to which a trait's variance is attributable to additive genetic factors, or the level of genetic overlap between traits. The results in this chapter suggest that, broadly speaking, the constellation of genetic signatures of and relationships between MD, GAD, neuroticism and depressive symptoms seem to match expectations. Heritability estimates for all outcomes were either modestly heritable (0.1-0.25) or non-significant, in both GCTA and LDSC, which is consistent with most previous empirical heritability estimates. Generally, estimates are small for neuroticism and GAD-related outcomes in past work, though MD's heritability in this project was smaller than expected: ~0.24 to 0.41 (100), although again, the 0.41 estimate is somewhat of an outlier. These estimates, as with most traits, are substantially smaller than heritability estimates from the biometrical methods applied largely to twin samples. This frequent observation may be due to any combination of a number of contributory factors: effect sizes that are smaller than the minimum detectable threshold, or even smaller than can ever be detected with significance (101); non-additive interactions between genetic loci could be contributing to biometrical heritability estimates (102); structural variants that aren't captured by most genotyping methods; gene-environment interplay; genomic imprinting; heritability epigenetic modifications and 'entirely unforeseen sources' (148). Explorations of these hypotheses in model systems suggest that the largest contribution is due to multitudes of small-effect variants (149), with substantial contributions of epistatic interactions varying greatly between traits, most of which involves more than two loci. Human traits may contain more heritability difficult to

assess with methods like GCTA and LDSC, since interactions may be more prevalent (such as dominance interaction between alleles at a single locus), rare variants may contribute more to trait variance, and certain traits may incorporate more genetic networks than traits in model organisms (149).

Two outcomes and samples that produced genome-wide significant variants in Chapter 3 did not display significant heritability using GCTA and LDSC: the GAD outcome in the MGS sample, and neuroticism in the S4S EUR subgroup. This is an inconsistent set of results only if we assume both types of tests have perfect sensitivity and specificity. Unfortunately, this is far from true; both genome-wide association tests and genotype-based heritability estimation are prone to not only false positives but false negatives as well, and these trends are exacerbated by low power. In this case, while it's possible that there is truly no additive genetic variance in GAD or neuroticism, the preponderance of evidence from twin studies (84, 87) suggests that the more likely explanation is that the MGS and S4S samples were underpowered to detect modest additive genetic variance using GCTA and LDSC.

Estimation of genetic correlation was, disappointingly, only possible in one of three samples (due to sample size), and in two of three pairings (due to multicollinearity). Only the genetic correlation of neuroticism with the two dichotomous outcomes of MD and comorbid MD+GAD produced results. While estimates varied widely from 0.87 to over 1 (non-significantly), it is clear that neuroticism is closely genetically related to MD (and GAD), in accordance with the twin literature. Neuroticism having a genetic correlation of (approximately) unity with both MD and MD+GAD,

further, is consistent with the theory that neuroticism comprises one element of the etiological architecture that gives rise to MD and GAD.

Polygenic risk scoring was somewhat more successful than other approaches, producing mostly significant results in a majority of tests. Genetic variants involved in the phenotypes of neuroticism and depressive symptoms do a good job of predicting within-trait, between one another, and also the disease outcome of MD. Neuroticism, but not depressive symptoms, also has a substantial association with comorbid MD+GAD in CONVERGE. This suggests that depressive symptoms may be a trait that has a closer relationship with depression-specific genetic risk factors than with genetic risk factors for GAD, whereas genetic risk factors for neuroticism are related to both MD and GAD.

BUHMBOX results suggest that the heterogeneity of GAD within MD is greater than the heterogeneity of MD within GAD, in the MGS sample. The circumstances that may jointly explain this include comorbidity (which is clearly true in MGS, with 306 individuals reporting comorbid MD+GAD), molecular subtypes (there are alternative genetic architectures for MD, one or more of which also predispose to GAD), mediated pleiotropy or asymmetric causation (having the GAD phenotype is a contributory cause of MD onset, more so than the reverse) and ascertainment bias (147). Of these, comorbidity is clearly evident, while molecular subtypes and asymmetric causation are plausible. Ascertainment bias (specifically that MD cases with GAD were more likely to be included in the study) is unlikely, given, as the name suggests, the Molecular Genetics of Schizophrenia sample controls were ascertained for (lack of) schizophrenia, and not selected for MD status, with comorbid GAD or otherwise.

Chapter 5: Conclusions

MD and GAD are psychiatric disorders that contribute a great deal of misery to the world. MD has the 2nd biggest impact on years lived with disability worldwide, with anxiety disorders 7th (11). 60% of individuals suffering from MD report at least severe role impairment (10), while GAD sabotages global well-being and life satisfaction (15). In addition to their effects on quality of life, the two disorders have a pronounced effect on mortality as well, mediated by other forms of illness and injury. MD has perhaps the larger effect on mortality, increasing the risk of suicide by 11 times (19). GAD makes up for this with multiplex effects on morbidity, however, being associated with a wide variety of chronic debilitating physiological syndromes, such as pain conditions like arthritis, back pain, and migraine (20). Both MD and GAD contribute to heart problems, with depression increasing risk of myocardial infarction (16), and GAD predisposing to coronary artery disease (22).

Despite similar presentations being noted during much of recorded history, the study of both MD and GAD in psychiatric epidemiology has only been pursued systematically since the mid-20th century. During this time, a number of risk factors and related constructs have been identified. High-quality romantic relationships, such as

happy marriages, are protective against both MD (60) and GAD (72). Stressful life events are also powerful contributors to MD (62) and GAD (73). Perhaps most significantly, the personality trait of neuroticism has both an epidemiological (50) and a latent genetic link (51) to MD and GAD.

Latent genetic methods such as twin studies have concluded that MD, GAD and Neuroticism are moderately heritable (81, 84, 87), and that all three have high-to-complete overlap in additive genetic factors (51, 86, 89-91). Attempts to find specific genetic loci involved in the etiology of these outcomes have begun to bear fruit in recent years, with significant variants identified for MD (93, 98), Neuroticism (95), and GAD-related outcomes (94, 96, 97), though none (until now) for GAD itself. Aggregate genetic techniques have estimated heritability for these outcomes as lower than the twin estimates, ranging from around 0.1 for neuroticism (95, 104) to as high as 0.41 for MD liability (100). As mentioned in the introduction, however, the 0.41 estimate is much closer to the twin heritability value for MD than most genotype-based heritability methods generally produce, so this may be an unrealistic point of comparison.

The present investigation builds upon this framework of knowledge about the etiological architecture of GAD, neuroticism and MD. Analyses to clarify the relative strength of psychosocial risk factors were detailed in Chapter 2. Attempts to find specific genetic loci involved in these outcomes were reported in Chapter 3. Multiple methods were used to quantify the degree of genetic variance in each outcome and also relationships between them in Chapter 4.

The results in Chapter 2, from the CONVERGE and S4S samples, suggest that virtually all tested risk factors were associated with the MD, GAD and related outcomes, with broadly aversive predictors increasing risk, and pleasant ones reducing risk. There was some evidence of differences in strength of association between subgroups in the S4S sample, with relationship satisfaction perhaps providing more protective effect against depressive symptoms to males than females (Figure 5). Social support had a reverse effect, with a more pronounced negative association in females than males with both depressive symptoms and neuroticism, with the association with neuroticism not even reaching significance in males (Figure 7). There was also a tendency for protective effects to be stronger in the EUR ancestry subgroup than the EAS subgroup, though this may be a function of sample size (Figure 7). Overall, the psychosocial factors had a weaker effect on a one standard deviation change in neuroticism than on risk of MD, MD+GAD, and depressive symptoms. This is consistent with neuroticism comprising a portion of the liability to MD (and GAD); in that a difference in psychosocial risk (approximately) corresponding to a two standard deviation difference in neuroticism is comparable to the increase in psychosocial risk that would predict an MD case based on liability, rather than an MD control.

The results from Chapter 3 include a number of ostensibly significant findings, using Bonferroni significance thresholds. The genome-wide association tests produced a total of 5 significant loci that may represent true effects among all the samples and outcomes. For the CONVERGE sample, two significant loci were detected for the MD outcome (Figure 9). The variant rs35936514 was significant for MD (Figure 11), a result previously identified in the CONVERGE study (93). The variant rs7257142 had not

previously been associated with MD (Figure 12), in the *PLIN4* gene, which is related to adipocyte functioning. The MGS sample performed surprisingly strongly, producing 2 loci of interest. Both of these loci were found for the GAD outcome (Figure 21). The first variant, rs117420048, is in the intron of *LINC00824*, a long non-coding RNA on chromosome 8 (Figure 23). The other, rs146220192, is found around *LINC00348*, another long non-coding RNA on chromosome 13 (Figure 24). In all S4S outcomes, the only significant variant occurred for neuroticism in the EUR subgroup (Figure 33), which was rs113888384, in the *CRMP1* gene (Figure 35), which is involved in neural development, and has been previously implicated in relation to major depression and schizophrenia (140). Gene-based results from MAGMA also produced a number of associated genes that appeared potentially genuine, mostly in the CONVERGE sample (Tables 7-9), though also three for the GAD outcome in MGS (Table 10). None of the loci or genes that reached significance were found for more than one outcome, though this is not surprising. Even replication within extremely well-matched phenotypes across samples is generally rare, in all but the largest meta-analyses to date, so with the moderate to low sample sizes in this project, finding common signals for related outcomes would be an optimistic expectation.

In Chapter 4, a number of methods were deployed to compare the overall degree of heritability in the outcomes of MD, GAD, neuroticism, and depressive symptoms. GCTA and LDSC univariate heritability estimates were only significant in CONVERGE (Table 12), suggesting modest heritability for MD and MD+GAD (0.22-0.25), and low heritability for neuroticism (0.09, only significant in GCTA). Bivariate genetic correlation between outcomes likewise was only successfully estimated in CONVERGE, resulting

in high-complete overlap in genetic factors neuroticism and both MD and MD+GAD outcomes. Polygenic risk score prediction of outcomes using LDpred (Table 13) provided additional evidence that neuroticism has some shared genetic factors with MD and MD+GAD. The polygenic score for depressive symptoms predicted only itself, neuroticism and MD (not MD+GAD). Finally, BUHMBOX suggested increased heterogeneity in MD for GAD than in GAD for MD. In other words, there was evidence for subgroups of MD in which GAD was prominent, and also asymmetric GAD → MD causation, more than the reverse. These results suggest that while there is high overlap in genetic factors between MD, GAD, and neuroticism, there are complexities nevertheless. For one, the low empirical heritability estimates suggest these outcomes are mostly comprised of environmental effects as well as perhaps some portion of non-additive/non-common genetic effects. Additionally, the failure of the polygenic score for depressive symptoms to predict the comorbid MD+GAD outcome, and the asymmetry evident in the BUHMBOX results suggests that GAD may have a more prominent phenotypic effect on MD risk than the MD phenotype does on GAD risk.

In order to understand what the gestalt import of these results is, it would perhaps be helpful to explore a few different possible states of affairs of how MD, GAD and neuroticism interrelate. Any future work on interventions in this area, such as treatments to help those with MD and GAD, would be more likely to succeed if designed with an understanding of the nature of their etiology and comorbidity. Thus, investigation into competing models of comorbidity is a basic foundational effort that may inform future more applied research. By comparing levels of evidence for different models of

comorbidity, one is essentially testing an array of hypotheses, some compatible and some mutually exclusive. Ten discrete models will be considered, which have been derived mathematically by authoritative prior work (150). Primarily, they will be judged on their consistency with the evidence produced in this project as theories about the relationship between MD and GAD, with the role of neuroticism to be considered later. The basic models include chance comorbidity, sampling bias, population stratification, symptom overlap, alternate forms, multiformity, separate comorbidity outcome, correlated liabilities, directional causation, and reciprocal causation.

Comorbidity due to chance results from essentially accidental overlap in statistically independent disorders; if there is no correlation, a portion of cases for one disorder will stochastically happen to also be cases for the other. This is not likely to constitute the relationship between MD and GAD, based on the wealth of statistical and genetic work suggesting co-occurrence at much higher than chance rates.

Sampling bias may contribute to the appearance of comorbidity, when samples are (for instance) clinically ascertained and those affected with multiple disorders may be more likely to seek or receive treatment. While comorbid MD+GAD does indeed result in more severe presentations (25, 26), of the samples used in this project only CONVERGE was clinically ascertained for the presence of a psychiatric disorder. It is unlikely that ascertainment for MD increased the proportion of comorbid individuals in CONVERGE, however, due to the proportion of MD cases with comorbid GAD in CONVERGE being approximately 25%, scarcely half the proportion of MD cases with GAD in MGS at about 48%. The individuals used from the MGS sample, notably, were selected for not having schizophrenia.

Population stratification may be another source of spurious comorbidity. If there are separate sets of risk factors for two disorders, but both sets happen to have increased prevalence in the same segment of the population, the disorders will appear to be comorbid. While stratification could be conceived along virtually any sort of classifying factor that describes people, many of the common sources, including age (life stage), sex, ancestry, and SES (proxied via educational attainment) were included in analyses whenever possible in this project. These covariates and stratification variables often had an effect on the risk of MD and GAD, but in no case did they explain the entirety of the relationship between MD and GAD. This suggests that population stratification (while it cannot be completely ruled out), is not the central force driving the connection between MD and GAD.

Symptom overlap could result in comorbidity when some of the same criteria are used as decision factors for the diagnosis of two diseases. There are three criteria in the DSM diagnoses of major depressive episode (Table 1) and GAD (Table 3) that could be construed as overlapping: “Insomnia or Hypersomnia nearly every day [for a two-week period]” vs. “Sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep) [more days than not for at least 6 months]”; “Fatigue or loss of energy nearly every day [for a two-week period]” vs. “Being easily fatigued [more days than not for at least 6 months]”; and “Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others) [for a two-week period]” vs. “Difficulty concentrating or mind going blank [more days than not for at least 6 months].” While the duration and frequency of these symptoms is quite dissimilar between MD and GAD criteria, the descriptions are very close.

Symptom level data was not available for all subgroups and samples, which means that for some comorbid individuals the contribution of this symptom overlap could not be assessed. It is possible that these similarities in diagnostic criteria contribute to MD and GAD comorbidity, not only in this project, but in existing literature as well.

Alternate forms is a comorbidity model in which there is a single liability distribution associated with both disorders, suggesting that the expression of, for instance, MD, GAD, or both MD+GAD would depend on factors that do not affect overall liability. There is some support for this model in the literature and in the similar impact of all the psychosocial risk factors on both MD and MD+GAD (except perhaps adult stressful life events). However, it is inconsistent with comorbid cases suffering increased symptoms (due, for instance, to having higher liability load, on average), as has been mentioned previously. It is also incompatible with evidence of directional phenotypic causation, such as the BUHMBOX results suggesting more pronounced GAD→MD causation.

The multiformity model of comorbidity posits that one disorder, say MD, is responsible for generating the symptoms of another, GAD, despite being unrelated to the liability for GAD at all. There are multiple subtypes of this model, depending on when the chance of displaying the second constellation of symptoms arises, such as a single threshold for disease status and chance of displaying secondary symptoms, or two thresholds: the higher corresponding to the secondary presentation (and resulting in higher liability in comorbid cases). A gradual increase in risk of GAD symptoms as a function of MD liability is also possible. A model in which GAD is only observed as a consequence of MD liability is also conceivable. The types of analyses contained in this

project are not well-suited for choosing between these models; a statistical effect on risk for GAD may indicate an impact on GAD liability, but could instead have an effect on MD liability with multiformity for GAD. The only subset of multiformity models that can be confidently excluded are those that suggest that one disorder encompasses another – the presence of MD cases without GAD, and GAD cases without MD precludes any sort of multiformity incompatible with cases of both disorders in absence of the other.

Models that posit a comorbid condition as a separate outcome qualitatively dissimilar from the separate disorders, are, in comparison, quite easy to test. If the liability for a comorbid condition (MD+GAD) is independent of liabilities for the individual outcomes (MD and GAD), one would not necessarily expect the same risk factors to increase risk similarly across multiple outcomes. It's possible that one (or more) will, but this would be random, due to their statistical independence. The similarity of effect of all psychosocial risk factors between MD and MD+GAD (with the possible exception of adult stressful life events, which had a somewhat stronger association to MD+GAD than MD) would argue that a separate MD+GAD liability distribution is unlikely. Also, the genetic correlation estimate between MD and neuroticism was near unity, as was the correlation between MD+GAD and neuroticism – an unlikely scenario for independent outcomes.

The model of correlated liabilities suggests each disorder has its own liability distribution, but an increase in liability on one continuum also constitutes an (often diminished) increase in liability on the other. This model is consistent with a high degree of similarity in risk factors between MD and GAD, as well as with deviations from perfect

matching. This model appears to be perfectly consistent with all findings except suggestions of unequal heterogeneity from BUHMBOX.

Directional causation places some hierarchy on the disorders – their liabilities may be separate, with the exception that one disorder is a risk factor for the other. That is, the phenotypic state of meeting criteria for one disorder directly increases the risk of also having the other. Key to this model is that this relationship is one way. In terms of genetic risk factors, this could also be conceived as mediated pleiotropy. The most direct support for this model in this project is the GAD→MD causation suggested by BUHMBOX. It is also supported by the failure of the depressive symptoms polygenic score (a broad proxy for MD polygenic score) to predict the comorbid MD+GAD outcome as well as it predicted the MD outcome. However, the fact that most risk factors and genetic correlations behaved identically between MD and MD+GAD would suggest that the MD liability may not be meaningfully increased by the addition of GAD. This is also the opposite conclusion from one of the best fitting models in the biometrical comorbidity paper (150), which found rather that MD tended to cause GAD.

Finally reciprocal causation is similar to two directional causation models put together – it could indeed be symmetric causation, but the two paths need not be equal in magnitude. This suggests that there is no direct relationship between the liabilities to MD and GAD, but that any statistical link found between them is due to the mediation of phenotypic causation. The appearance of MD risk factors having an association with GAD is only due to the MD phenotype tending to contribute to onset of a GAD phenotype as well. This is another model that the present analyses are not well-suited to disprove. The plausibility of this model depends on there being sub-threshold

phenotypic effects, however – due to the patterns of familial aggregation between mood and anxiety disorders (such as MD and GAD). If the only connection between MD and GAD was the impact of full diagnosis of one disorder on the other, there should be no cross-disorder familial aggregation (the risk for MD should not contribute to the risk for GAD in the absence of MD), but this is well-supported (151). Allowing for phenotypic effects of liability levels that fall below the diagnostic threshold would enable this type of model to be consistent with the majority of available evidence, however.

Overall, it seems as though there are a number of comorbidity models that are consistent with the observed pattern of results in this project. Various types of multiformity, correlated liabilities, and reciprocal causation appear to be the most plausible basic models. Of course, there is no theoretical reason that the true arrangement of the etiological architecture of MD and GAD couldn't include aspects of multiple of these disorders. The results from the tests of neuroticism, initially included in this project as a predictor, then an ancillary continuous outcome, and now as an interesting trait in its own right, actually provide a hypothesis. Neuroticism appears to function, as far as can be evaluated using the battery of tests in this manuscript, as a portion of the liability distribution that is shared between MD and GAD. That is, for the range of predictive factors that can be construed as the determinants of neuroticism, the model of comorbidity, with MD, GAD and neuroticism that fits best is actually alternate forms, in the following sense: neuroticism may only contribute a minority of the overall variance in MD or GAD, but there is no neuroticism variance not shared with the other outcomes. The 'liability' to neuroticism composes a portion of the liability to MD and GAD both. The remainder, however, appears to be consistent with a correlated liabilities

model, in that genetic and environmental correlation estimates from biometrical methods (the present project could not directly estimate an MD-GAD genetic correlation) suggest high similarity. There is also evidence, however, of both MD→GAD causation (150), and GAD→MD causation (BUHMBOX mediated pleiotropy, and asymmetric performance of depressive symptoms polygenic score).

Thus, the most plausible model from all available data appears to be a core of common etiology shared entirely between MD and GAD, of which neuroticism is an important component. The remainder of the risk structure for MD and GAD are correlated liability distributions, in which the phenotype (perhaps even sub-threshold) of each disorder acts as a risk factor for the other, though the relative magnitude of these effects is as yet unknown.

This project has a number of important limitations that are vital to be aware of when evaluating its conclusions. First, the sample size, as has been mentioned throughout, and therefore statistical power, varied widely in adequacy from perfectly appropriate for psychosocial tests in CONVERGE, to insufficient for estimation for most bivariate aggregate genetic methods. Second, replication (and meta-analysis) between samples was largely impossible, due to the overlapping ancestry and data compositions – all findings should therefore be viewed as preliminary until they can be validated with further research. Third, there were methodological differences between samples that could only be highlighted rather than controlled, and therefore there may be bias due to different ascertainment, assessment, and quality control procedures in the component datasets beyond presently tractable adjustment strategies.

Finally, some of the central questions of the study – to what degree do the genetic and psychosocial risk dimensions of MD and GAD overlap, could not be asked with the available data, and conclusions are therefore based on indirect evidence and past work – future work will be vital to determining whether these conclusions hold.

The dual faces of misery present many obstacles to understanding, due to their inscrutable origins and pernicious, prevalent influence. We should not, however, relinquish hope. In the enigmatic strength of their divided unity, there is a kind of vulnerability; it is highly likely that when we finally unmask the foundation of one disorder, the principal essence of the other will be revealed as well.

List of References

List of References

1. S. B-C. The evolution of a theory of mind. In: M. C. Corballis SEGL, editor. The descent of mind: Psychological perspectives on hominid evolution. New York: Oxford University Press; 1999. p. 261-77.
2. MacLean EL. Unraveling the evolution of uniquely human cognition. Proc Natl Acad Sci U S A. 2016;113(23):6348-54. doi: 10.1073/pnas.1521270113. PubMed PMID: 27274041; PubMed Central PMCID: PMC4988573.
3. Lev-Ran A. Human obesity: an evolutionary approach to understanding our bulging waistline. Diabetes Metab Res Rev. 2001;17(5):347-62. PubMed PMID: 11747140.
4. Nolan J. "The Bicameral Mind". In: Jonathan Nolan LJ, editor. Westworld: HBO; 2016.
5. Andics A, Gacsi M, Farago T, Kis A, Miklosi A. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr Biol. 2014;24(5):574-8. doi: 10.1016/j.cub.2014.01.058. PubMed PMID: 24560578.
6. Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci U S A. 2003;100(3):1056-61. doi: 10.1073/pnas.0334222100. PubMed PMID: 12552136; PubMed Central PMCID: PMC298725.
7. Baumeister RF. The Cultural Animal : Human Nature, Meaning, and Social Life. Oxford, UK: Oxford University Press; 2005.
8. De Waal F, Waal F. The Ape And The Sushi Master: Cultural Reflections Of A Primatologist: Basic Books; 2008.
9. Damasio A. Descartes' Error: Emotion, Reason, and the Human Brain: Penguin Publishing Group; 2005.
10. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS, National Comorbidity Survey R. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA. 2003;289(23):3095-105. doi: 10.1001/jama.289.23.3095. PubMed PMID: 12813115.

11. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basanez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabe E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne SD, Dherani M, Diaz-Torne C, Dolk H, Dorsey ER, Driscoll T, Duber H, Ebel B, Edmond K, Elbaz A, Ali SE, Erskine H, Erwin PJ, Espindola P, Ewoigbokhan SE, Farzadfar F, Feigin V, Felson DT, Ferrari A, Ferri CP, Fevre EM, Finucane MM, Flaxman S, Flood L, Foreman K, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabbe BJ, Gabriel SE, Gakidou E, Ganatra HA, Garcia B, Gaspari F, Gillum RF, Gmel G, Gosselin R, Grainger R, Groeger J, Guillemin F, Gunnell D, Gupta R, Haagsma J, Hagan H, Halasa YA, Hall W, Haring D, Haro JM, Harrison JE, Havmoeller R, Hay RJ, Higashi H, Hill C, Hoen B, Hoffman H, Hotez PJ, Hoy D, Huang JJ, Ibeanusi SE, Jacobsen KH, James SL, Jarvis D, Jasrasaria R, Jayaraman S, Johns N, Jonas JB, Karthikeyan G, Kassebaum N, Kawakami N, Keren A, Khoo JP, King CH, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Laloo R, Laslett LL, Lathlean T, Leasher JL, Lee YY, Leigh J, Lim SS, Limb E, Lin JK, Lipnick M, Lipshultz SE, Liu W, Loane M, Ohno SL, Lyons R, Ma J, Mabweijano J, MacIntyre MF, Malekzadeh R, Mallinger L, Manivannan S, Marcenes W, March L, Margolis DJ, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGill N, McGrath J, Medina-Mora ME, Meltzer M, Mensah GA, Merriman TR, Meyer AC, Miglioli V, Miller M, Miller TR, Mitchell PB, Mocumbi AO, Moffitt TE, Mokdad AA, Monasta L, Montico M, Moradi-Lakeh M, Moran A, Morawska L, Mori R, Murdoch ME, Mwaniki MK, Naidoo K, Nair MN, Naldi L, Narayan KM, Nelson PK, Nelson RG, Nevitt MC, Newton CR, Nolte S, Norman P, Norman R, O'Donnell M, O'Hanlon S, Olives C, Omer SB, Ortblad K, Osborne R, Ozgediz D, Page A, Pahari B, Pandian JD, Rivero AP, Patten SB, Pearce N, Padilla RP, Perez-Ruiz F, Perico N, Pesudovs K, Phillips D, Phillips MR, Pierce K, Pion S, Polanczyk GV, Polinder S, Pope CA, 3rd, Popova S, Porrini E, Pourmalek F, Prince M, Pullan RL, Ramaiah KD, Ranganathan D, Razavi H, Regan M, Rehm JT, Rein DB, Remuzzi G, Richardson K, Rivara FP, Roberts T, Robinson C, De Leon FR, Ronfani L, Room R, Rosenfeld LC, Rushton L, Sacco RL, Saha S, Sampson U, Sanchez-Riera L, Sanman E, Schwebel DC, Scott JG, Segui-Gomez M, Shahraz S, Shepard DS, Shin H, Shivakoti R, Singh D, Singh GM, Singh JA, Singleton J, Sleet DA, Sliwa K, Smith E, Smith JL, Stapelberg NJ, Steer A, Steiner T, Stolk WA, Stovner LJ, Sudfeld C, Syed S, Tamburlini G, Tavakkoli M, Taylor HR, Taylor JA, Taylor WJ, Thomas B, Thomson WM, Thurston GD, Tleyjeh IM, Tonelli M, Towbin JA, Truelsen T, Tsilimbaris MK, Ubeda C, Undurraga EA, van der Werf MJ, van Os J, Vavilala MS, Venketasubramanian N, Wang M, Wang W, Watt K, Weatherall DJ, Weinstock MA, Weintraub R, Weisskopf MG, Weissman MM, White RA, Whiteford H, Wiersma ST, Wilkinson JD, Williams HC, Williams SR, Witt E, Wolfe F, Woolf AD, Wulf S, Yeh PH, Zaidi AK, Zheng ZJ, Zonies D, Lopez AD,

Murray CJ, AlMazroa MA, Memish ZA. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2163-96. doi: 10.1016/S0140-6736(12)61729-2. PubMed PMID: 23245607.

12. Kessler RC. The epidemiology of pure and comorbid generalized anxiety disorder: a review and evaluation of recent research. *Acta psychiatrica Scandinavica Supplementum*. 2000(406):7-13. Epub 2000/12/29. PubMed PMID: 11131470.

13. Greenberg PE, Kessler RC, Birnbaum HG, Leong SA, Lowe SW, Berglund PA, Corey-Lisle PK. The economic burden of depression in the United States: how did it change between 1990 and 2000? *The Journal of clinical psychiatry*. 2003;64(12):1465-75. PubMed PMID: 14728109.

14. Greenberg PE, Sisitsky T, Kessler RC, Finkelstein SN, Berndt ER, Davidson JR, Ballenger JC, Fyer AJ. The economic burden of anxiety disorders in the 1990s. *The Journal of clinical psychiatry*. 1999;60(7):427-35. PubMed PMID: 10453795.

15. Stein MB, Heimberg RG. Well-being and life satisfaction in generalized anxiety disorder: comparison to major depressive disorder in a community sample. *Journal of affective disorders*. 2004;79(1-3):161-6. doi: 10.1016/S0165-0327(02)00457-3. PubMed PMID: 15023490.

16. Barefoot JC, Schroll M. Symptoms of depression, acute myocardial infarction, and total mortality in a community sample. *Circulation*. 1996;93(11):1976-80. PubMed PMID: 8640971.

17. Jonas BS, Mussolino ME. Symptoms of depression as a prospective risk factor for stroke. *Psychosomatic medicine*. 2000;62(4):463-71. PubMed PMID: 10949089.

18. Jonas BS, Franks P, Ingram DD. Are symptoms of anxiety and depression risk factors for hypertension? Longitudinal evidence from the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study. *Arch Fam Med*. 1997;6(1):43-9. PubMed PMID: 9003169.

19. Kessler RC, Borges G, Walters EE. Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. *Archives of general psychiatry*. 1999;56(7):617-26. PubMed PMID: 10401507.

20. McWilliams LA, Goodwin RD, Cox BJ. Depression and anxiety associated with three pain conditions: results from a nationally representative sample. *Pain*. 2004;111(1-2):77-83. doi: 10.1016/j.pain.2004.06.002. PubMed PMID: 15327811.

21. Goodwin RD, Stein MB. Generalized anxiety disorder and peptic ulcer disease among adults in the United States. *Psychosomatic medicine*. 2002;64(6):862-6. PubMed PMID: 12461190.

22. Barger SD, Sydean SJ. Does generalized anxiety disorder predict coronary heart disease risk factors independently of major depressive disorder? *Journal of affective disorders*. 2005;88(1):87-91. doi: 10.1016/j.jad.2005.05.012. PubMed PMID: 16009431.
23. Grigsby AB, Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. Prevalence of anxiety in adults with diabetes: a systematic review. *J Psychosom Res*. 2002;53(6):1053-60. PubMed PMID: 12479986.
24. Muller JE, Koen L, Stein DJ. Anxiety and medical disorders. *Curr Psychiatry Rep*. 2005;7(4):245-51. PubMed PMID: 16098277.
25. Hoffman DL, Dukes EM, Wittchen HU. Human and economic burden of generalized anxiety disorder. *Depression and anxiety*. 2008;25(1):72-90. doi: 10.1002/da.20257. PubMed PMID: 17146763.
26. Wittchen HU, Kessler RC, Beesdo K, Krause P, Hofler M, Hoyer J. Generalized anxiety and depression in primary care: prevalence, recognition, and management. *The Journal of clinical psychiatry*. 2002;63 Suppl 8:24-34. PubMed PMID: 12044105.
27. Reynolds EH, Wilson JV. Depression and anxiety in Babylon. *J R Soc Med*. 2013;106(12):478-81. doi: 10.1177/0141076813486262. PubMed PMID: 23759889.
28. Radden J. Is This Dame Melancholy?: Equating Today's Depression and Past Melancholia. *Philosophy, Psychiatry, & Psychology*. 2003;10(1).
29. Burton R. *The Anatomy of Melancholy: What it Is, with All the Kinds, Causes, Symptoms, Prognostics, and Several Cures of It. In Three Partitions. With Their Several Sections, Members, and Subsections, Philosophically, Medically, Historically Opened and Cut Up*: Sheldon; 1862.
30. Davison K. Historical aspects of mood disorders. *Psychiatry*. 2006;5(4):115-8. doi: <http://dx.doi.org/10.1383/psyt.2006.5.4.115>.
31. American Psychiatric Association. Task Force on N, Statistics. *Diagnostic and Statistical Manual of Mental Disorders*: American Psychiatric Association; 1994.
32. American Psychiatric A, American Psychiatric Association. Task Force on D-I. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*: American Psychiatric Association; 2000.
33. World Health O. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*: Geneva: World Health Organization; 1992.
34. Cannon WB. *Bodily Changes in Pain, Hunger, Fear and Rage: An Account of Recent Researches Into the Function of Emotional Excitement*: D. Appleton; 1920.

35. Lee WE, Wadsworth ME, Hotopf M. The protective role of trait anxiety: a longitudinal cohort study. *Psychological medicine*. 2006;36(3):345-51. doi: 10.1017/S0033291705006847. PubMed PMID: 16393365.
36. Feinberg M, Willer R, Keltner D. Flustered and faithful: embarrassment as a signal of prosociality. *Journal of personality and social psychology*. 2012;102(1):81-97. doi: 10.1037/a0025403. PubMed PMID: 21928915.
37. Brosschot JF, Verkuil B, Thayer JF. The default response to uncertainty and the importance of perceived safety in anxiety and stress: An evolution-theoretical perspective. *J Anxiety Disord*. 2016;41:22-34. doi: 10.1016/j.janxdis.2016.04.012. PubMed PMID: 27259803.
38. Coplan JD, Hodulik S, Mathew SJ, Mao X, Hof PR, Gorman JM, Shungu DC. The Relationship between Intelligence and Anxiety: An Association with Subcortical White Matter Metabolism. *Front Evol Neurosci*. 2011;3:8. doi: 10.3389/fnevo.2011.00008. PubMed PMID: 22347183; PubMed Central PMCID: PMC3269637.
39. Hippocrates. *Epidemics 2, 4-7*. Cambridge, MA: Harvard University Press; 1994.
40. Shorter E. *A Historical Dictionary of Psychiatry*: Oxford University Press, USA; 2005.
41. Arbiser S, Schneider J. *On Freud's "Inhibitions, Symptoms and Anxiety"*: Karnac Books; 2013.
42. Kraepelin E. *Psychiatrie: ein Lehrbuch für Studierende und Ärzte*: Barth; 1909.
43. Crocq MA. A history of anxiety: from Hippocrates to DSM. *Dialogues Clin Neurosci*. 2015;17(3):319-25. PubMed PMID: 26487812; PubMed Central PMCID: PMC4610616.
44. Klein DF. Delineation of Two Drug-Responsive Anxiety Syndromes. *Psychopharmacologia*. 1964;5:397-408. PubMed PMID: 14194683.
45. Thompson ER. Development and Validation of an International English Big-Five Mini-Markers. *Personality and Individual Differences*. 2008;45(6):542-8. doi: <http://dx.doi.org/10.1016/j.paid.2008.06.013>.
46. Lindberg DC. *The Beginnings of Western Science: The European Scientific Tradition in Philosophical, Religious, and Institutional Context, Prehistory to A.D. 1450*, Second Edition: University of Chicago Press; 2010.
47. Eysenck HJ. *Dimensions of Personality*: Routledge & Kegan Paul; 1950.
48. Eysenck HJ, Eysenck SBG. *Psychoticism as a Dimension of Personality*: Hodder and Stoughton; 1976.
49. McCrae RR, John OP. An introduction to the five-factor model and its applications. *J Pers*. 1992;60(2):175-215. PubMed PMID: 1635039.

50. Ormel J, Jeronimus BF, Kotov R, Riese H, Bos EH, Hankin B, Rosmalen JG, Oldehinkel AJ. Neuroticism and common mental disorders: meaning and utility of a complex relationship. *Clinical psychology review*. 2013;33(5):686-97. Epub 2013/05/25. doi: 10.1016/j.cpr.2013.04.003. PubMed PMID: 23702592.
51. Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS. A population-based twin study of the relationship between neuroticism and internalizing disorders. *The American journal of psychiatry*. 2006;163(5):857-64. Epub 2006/05/02. doi: 10.1176/appi.ajp.163.5.857. PubMed PMID: 16648327.
52. Eysenck HJ. Neuroticism in twins. *Eugen Rev*. 1951;43(2):79-82. PubMed PMID: 21260594; PubMed Central PMCID: PMCPMC2973302.
53. Eysenck HJ, Prell DB. The inheritance of neuroticism: an experimental study. *J Ment Sci*. 1951;97(408):441-65. PubMed PMID: 14861606.
54. Andrade L, Caraveo-Anduaga JJ, Berglund P, Bijl RV, De Graaf R, Vollebergh W, Dragomirecka E, Kohn R, Keller M, Kessler RC, Kawakami N, Kilic C, Offord D, Ustun TB, Wittchen HU. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *International journal of methods in psychiatric research*. 2003;12(1):3-21. Epub 2003/06/28. PubMed PMID: 12830306.
55. Weissman MM, Klerman GL. Sex differences and the epidemiology of depression. *Archives of general psychiatry*. 1977;34(1):98-111. PubMed PMID: 319772.
56. Platt J, Prins S, Bates L, Keyes K. Unequal depression for equal work? How the wage gap explains gendered disparities in mood disorders. *Social Science & Medicine*. 2016;149:1-8. doi: <http://dx.doi.org/10.1016/j.socscimed.2015.11.056>.
57. Lee S, Tsang A, Huang YQ, He YL, Liu ZR, Zhang MY, Shen YC, Kessler RC. The epidemiology of depression in metropolitan China. *Psychological medicine*. 2009;39(5):735-47. doi: 10.1017/S0033291708004091. PubMed PMID: 18713484; PubMed Central PMCID: PMCPMC2681248.
58. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, de Graaf R, Demyttenaere K, Hu C, Iwata N, Karam AN, Kaur J, Kostyuchenko S, Lepine JP, Levinson D, Matschinger H, Mora ME, Browne MO, Posada-Villa J, Viana MC, Williams DR, Kessler RC. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med*. 2011;9:90. doi: 10.1186/1741-7015-9-90. PubMed PMID: 21791035; PubMed Central PMCID: PMCPMC3163615.
59. Parker G, Gladstone G, Chee KT. Depression in the planet's largest ethnic group: the Chinese. *The American journal of psychiatry*. 2001;158(6):857-64. doi: 10.1176/appi.ajp.158.6.857. PubMed PMID: 11384889.

60. Bachrach LL, National Institute of Mental Health. Marital Status and Mental Disorder: An Analytical Review: Alcohol, Drug Abuse, and Mental Health Administration; 1975.
61. Whitton SW, Weitbrecht EM, Kuryluk AD, Bruner MR. Committed dating relationships and mental health among college students. *J Am Coll Health*. 2013;61(3):176-83. doi: 10.1080/07448481.2013.773903. PubMed PMID: 25158015.
62. Kessler RC. The effects of stressful life events on depression. *Annual review of psychology*. 1997;48:191-214. Epub 1997/01/01. doi: 10.1146/annurev.psych.48.1.191. PubMed PMID: 9046559.
63. Kendler KS, Hettema JM, Butera F, Gardner CO, Prescott CA. Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. *Archives of general psychiatry*. 2003;60(8):789-96. Epub 2003/08/13. doi: 10.1001/archpsyc.60.8.789. PubMed PMID: 12912762.
64. Benjet C, Bromet E, Karam EG, Kessler RC, McLaughlin KA, Ruscio AM, Shahly V, Stein DJ, Petukhova M, Hill E, Alonso J, Atwoli L, Bunting B, Bruffaerts R, Caldas-de-Almeida JM, de Girolamo G, Florescu S, Gureje O, Huang Y, Lepine JP, Kawakami N, Kovess-Masfety V, Medina-Mora ME, Navarro-Mateu F, Piazza M, Posada-Villa J, Scott KM, Shalev A, Slade T, ten Have M, Torres Y, Viana MC, Zarkov Z, Koenen KC. The epidemiology of traumatic event exposure worldwide: results from the World Mental Health Survey Consortium. *Psychological medicine*. 2016;46(2):327-43. doi: 10.1017/S0033291715001981. PubMed PMID: 26511595; PubMed Central PMCID: PMC4869975.
65. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of general psychiatry*. 2005;62(6):629-40. doi: 10.1001/archpsyc.62.6.629. PubMed PMID: 15939840.
66. Boden JM, Fergusson DM. Alcohol and depression. *Addiction*. 2011;106(5):906-14. doi: 10.1111/j.1360-0443.2010.03351.x. PubMed PMID: 21382111.
67. Marshall BD, Werb D. Health outcomes associated with methamphetamine use among young people: a systematic review. *Addiction*. 2010;105(6):991-1002. doi: 10.1111/j.1360-0443.2010.02932.x. PubMed PMID: 20659059.
68. Kessler RC, Brandenburg N, Lane M, Roy-Byrne P, Stang PD, Stein DJ, Wittchen H-U. Rethinking the duration requirement for generalized anxiety disorder: evidence from the National Comorbidity Survey Replication. *Psychological medicine*. 2005;35(7):1073-82. doi: 10.1017/S0033291705004538.
69. Remes O, Brayne C, van der Linde R, Lafortune L. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav*. 2016;6(7):e00497. doi: 10.1002/brb3.497. PubMed PMID: 27458547; PubMed Central PMCID: PMC4951626.

70. Ma X, Xiang YT, Cai ZJ, Lu JY, Li SR, Xiang YQ, Guo HL, Hou YZ, Li ZB, Li ZJ, Tao YF, Dang WM, Wu XM, Deng J, Lai KY, Ungvari GS. Generalized anxiety disorder in China: prevalence, sociodemographic correlates, comorbidity, and suicide attempts. *Perspect Psychiatr Care*. 2009;45(2):119-27. doi: 10.1111/j.1744-6163.2009.00212.x. PubMed PMID: 19366422.
71. Kleinman A. *Social Origins of Distress and Disease: Depression, Neurasthenia, and Pain in Modern China*: Yale University Press; 1988.
72. Priest JB. Anxiety disorders and the quality of relationships with friends, relatives, and romantic partners. *Journal of clinical psychology*. 2013;69(1):78-88. doi: 10.1002/jclp.21925. PubMed PMID: 23033057.
73. Faravelli C, Lo Sauro C, Lelli L, Pietrini F, Lazzeretti L, Godini L, Benni L, Fioravanti G, Talamba GA, Castellini G, Ricca V. The role of life events and HPA axis in anxiety disorders: a review. *Curr Pharm Des*. 2012;18(35):5663-74. PubMed PMID: 22632471.
74. Morissette SB, Tull MT, Gulliver SB, Kamholz BW, Zimering RT. Anxiety, anxiety disorders, tobacco use, and nicotine: a critical review of interrelationships. *Psychological bulletin*. 2007;133(2):245-72. doi: 10.1037/0033-2909.133.2.245. PubMed PMID: 17338599.
75. Bruce MS, Lader M. Caffeine abstinence in the management of anxiety disorders. *Psychological medicine*. 1989;19(1):211-4. PubMed PMID: 2727208.
76. Zubin J, Spring B. Vulnerability--a new view of schizophrenia. *Journal of abnormal psychology*. 1977;86(2):103-26. PubMed PMID: 858828.
77. Shipley BA, Weiss A, Der G, Taylor MD, Deary IJ. Neuroticism, extraversion, and mortality in the UK Health and Lifestyle Survey: a 21-year prospective cohort study. *Psychosomatic medicine*. 2007;69(9):923-31. doi: 10.1097/PSY.0b013e31815abf83. PubMed PMID: 17991814.
78. Eysenck HJ, Eysenck MW. *Personality and Individual Differences: A Natural Science Approach*: Plenum Press; 1985.
79. Zinbarg RE, Mineka S, Bobova L, Craske MG, Vrshek-Schallhorn S, Griffith JW, Wolitzky-Taylor K, Waters AM, Sumner JA, Anand D. Testing a Hierarchical Model of Neuroticism and Its Cognitive Facets: Latent Structure and Prospective Prediction of First Onsets of Anxiety and Unipolar Mood Disorders During 3 Years in Late Adolescence. *Clinical Psychological Science*. 2016;4(5):805-24.
80. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics*. 1965;29(1):51-76. doi: 10.1111/j.1469-1809.1965.tb00500.x.

81. Kendler KS, Gatz M, Gardner CO, Pedersen NL. A Swedish national twin study of lifetime major depression. *The American journal of psychiatry*. 2006;163(1):109-14. Epub 2006/01/05. doi: 10.1176/appi.ajp.163.1.109. PubMed PMID: 16390897.
82. Guffanti G, Gameroff MJ, Warner V, Talati A, Glatt CE, Wickramaratne P, Weissman MM. Heritability of major depressive and comorbid anxiety disorders in multi-generational families at high risk for depression. *Am J Med Genet B Neuropsychiatr Genet*. 2016;171(8):1072-9. doi: 10.1002/ajmg.b.32477. PubMed PMID: 27452917.
83. Chen J, Li X, Natsuaki MN, Leve LD, Harold GT. Genetic and environmental influences on depressive symptoms in Chinese adolescents. *Behavior genetics*. 2014;44(1):36-44. doi: 10.1007/s10519-013-9632-8. PubMed PMID: 24311200; PubMed Central PMCID: PMC5110008.
84. Hettema JM, Neale MC, Kendler KS. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *The American journal of psychiatry*. 2001;158(10):1568-78. Epub 2001/10/02. PubMed PMID: 11578982.
85. Hettema JM, Prescott CA, Kendler KS. A population-based twin study of generalized anxiety disorder in men and women. *The Journal of nervous and mental disease*. 2001;189(7):413-20. PubMed PMID: 11504317.
86. Roy MA, Neale MC, Pedersen NL, Mathe AA, Kendler KS. A twin study of generalized anxiety disorder and major depression. *Psychological medicine*. 1995;25(5):1037-49. PubMed PMID: 8588001.
87. Bouchard TJ, Jr., Loehlin JC. Genes, evolution, and personality. *Behavior genetics*. 2001;31(3):243-73. PubMed PMID: 11699599.
88. Vink JM, Bartels M, van Beijsterveldt TC, van Dongen J, van Beek JH, Distel MA, de Moor MH, Smit DJ, Minica CC, Ligthart L, Geels LM, Abdellaoui A, Middeldorp CM, Hottenga JJ, Willemsen G, de Geus EJ, Boomsma DI. Sex differences in genetic architecture of complex phenotypes? *PLoS One*. 2012;7(12):e47371. doi: 10.1371/journal.pone.0047371. PubMed PMID: 23272036; PubMed Central PMCID: PMC3525575.
89. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Archives of general psychiatry*. 1992;49(9):716-22. Epub 1992/09/01. PubMed PMID: 1514877.
90. Hettema JM, Kuhn JW, Prescott CA, Kendler KS. The impact of generalized anxiety disorder and stressful life events on risk for major depressive episodes. *Psychological medicine*. 2006;36(6):789-95. doi: 10.1017/S0033291706007367. PubMed PMID: 16515735.
91. Kendler KS, Gardner CO, Gatz M, Pedersen NL. The sources of co-morbidity between major depression and generalized anxiety disorder in a Swedish national twin sample.

Psychological medicine. 2007;37(3):453-62. Epub 2006/11/24. doi: 10.1017/S0033291706009135. PubMed PMID: 17121688.

92. Kendler KS. Major depression and generalised anxiety disorder. Same genes, (partly)different environments--revisited. The British journal of psychiatry Supplement. 1996(30):68-75. PubMed PMID: 8864151.

93. Consortium C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 2015;523(7562):588-91. doi: 10.1038/nature14659. PubMed PMID: 26176920; PubMed Central PMCID: PMC4522619.

94. Dunn EC, Sofer T, Gallo LC, Gogarten SM, Kerr KF, Chen CY, Stein MB, Ursano RJ, Guo X, Jia Y, Qi Q, Rotter JI, Argos M, Cai J, Penedo FJ, Perreira K, Wassertheil-Smoller S, Smoller JW. Genome-wide association study of generalized anxiety symptoms in the Hispanic Community Health Study/Study of Latinos. Am J Med Genet B Neuropsychiatr Genet. 2016. doi: 10.1002/ajmg.b.32448. PubMed PMID: 27159506.

95. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, Meddens SF, Linner RK, Rietveld CA, Derringer J, Gratten J, Lee JJ, Liu JZ, de Vlaming R, Ahluwalia TS, Buchwald J, Cavadino A, Frazier-Wood AC, Furlotte NA, Garfield V, Geisel MH, Gonzalez JR, Haitjema S, Karlsson R, van der Laan SW, Ladwig KH, Lahti J, van der Lee SJ, Lind PA, Liu T, Matteson L, Mihailov E, Miller MB, Minica CC, Nolte IM, Mook-Kanamori D, van der Most PJ, Oldmeadow C, Qian Y, Raitakari O, Rawal R, Realo A, Rueedi R, Schmidt B, Smith AV, Stergiakouli E, Tanaka T, Taylor K, Wedenoja J, Wellmann J, Westra HJ, Willems SM, Zhao W, LifeLines Cohort S, Amin N, Bakshi A, Boyle PA, Cherney S, Cox SR, Davies G, Davis OS, Ding J, Direk N, Eibich P, Emeny RT, Fatemifar G, Faul JD, Ferrucci L, Forstner A, Gieger C, Gupta R, Harris TB, Harris JM, Holliday EG, Hottenga JJ, De Jager PL, Kaakinen MA, Kajantie E, Karhunen V, Kolcic I, Kumari M, Launer LJ, Franke L, Li-Gao R, Koini M, Loukola A, Marques-Vidal P, Montgomery GW, Mosing MA, Paternoster L, Pattie A, Petrovic KE, Pulkki-Raback L, Quaye L, Raikkonen K, Rudan I, Scott RJ, Smith JA, Sutin AR, Trzaskowski M, Vinkhuyzen AE, Yu L, Zabaneh D, Attia JR, Bennett DA, Berger K, Bertram L, Boomsma DI, Snieder H, Chang SC, Cucca F, Deary IJ, van Duijn CM, Eriksson JG, Bultmann U, de Geus EJ, Groenen PJ, Gudnason V, Hansen T, Hartman CA, Haworth CM, Hayward C, Heath AC, Hinds DA, Hypponen E, Iacono WG, Jarvelin MR, Jockel KH, Kaprio J, Kardia SL, Keltikangas-Jarvinen L, Kraft P, Kubzansky LD, Lehtimaki T, Magnusson PK, Martin NG, McGue M, Metspalu A, Mills M, de Mutsert R, Oldehinkel AJ, Pasterkamp G, Pedersen NL, Plomin R, Polasek O, Power C, Rich SS, Rosendaal FR, den Ruijter HM, Schlessinger D, Schmidt H, Svento R, Schmidt R, Alizadeh BZ, Sorensen TI, Spector TD, Steptoe A, Terracciano A, Thurik AR, Timpson NJ, Tiemeier H, Uitterlinden AG, Vollenweider P, Wagner GG, Weir DR, Yang J, Conley DC, Smith GD, Hofman A, Johannesson M, Laibson DI, Medland SE, Meyer MN, Pickrell JK, Esko T, Krueger RF, Beauchamp JP, Koellinger PD, Benjamin DJ, Bartels M, Cesarini D. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nature genetics. 2016. doi: 10.1038/ng.3552. PubMed PMID: 27089181.

96. Davies MN, Verdi S, Burri A, Trzaskowski M, Lee M, Hetttema JM, Jansen R, Boomsma DI, Spector TD. Generalised Anxiety Disorder--A Twin Study of Genetic Architecture, Genome-Wide Association and Differential Gene Expression. *PLoS One*. 2015;10(8):e0134865. doi: 10.1371/journal.pone.0134865. PubMed PMID: 26274327; PubMed Central PMCID: PMC4537268.
97. Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, Bigdeli T, Aggen SH, Adkins D, Wolen A, Fanous A, Keller MC, Castelao E, Kutalik Z, der Auwera SV, Homuth G, Nauck M, Teumer A, Milaneschi Y, Hottenga JJ, Direk N, Hofman A, Uitterlinden A, Mulder CL, Henders AK, Medland SE, Gordon S, Heath AC, Madden PA, Pergadia ML, van der Most PJ, Nolte IM, van Oort FV, Hartman CA, Oldehinkel AJ, Preisig M, Grabe HJ, Middeldorp CM, Penninx BW, Boomsma D, Martin NG, Montgomery G, Maher BS, van den Oord EJ, Wray NR, Tiemeier H, Hetttema JM. Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular psychiatry*. 2016. doi: 10.1038/mp.2016.11. PubMed PMID: 26857599.
98. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, Tung JY, Hinds DA, Perlis RH, Winslow AR. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics*. 2016;48(9):1031-6. doi: 10.1038/ng.3623. PubMed PMID: 27479909.
99. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics*. 2011;88(1):76-82. Epub 2010/12/21. doi: 10.1016/j.ajhg.2010.11.011. PubMed PMID: 21167468; PubMed Central PMCID: PMC3014363.
100. Bulik-Sullivan B, Loh P-R, Finucane H, Ripke S, Yang J, Psychiatric Genomics Consortium SWG, Patterson N, Daly MJ, Price AL, Neale BM. LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies 2014.
101. Rockman MV. The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*. 2012;66(1):1-17. doi: 10.1111/j.1558-5646.2011.01486.x.
102. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*. 2012;109(4):1193-8. doi: 10.1073/pnas.1119675109. PubMed PMID: 22223662; PubMed Central PMCID: PMC3268279.
103. Purcell S. Variance components models for gene-environment interaction in twin analysis. *Twin Res*. 2002;5(6):554-71. doi: 10.1375/136905202762342026. PubMed PMID: 12573187.
104. Docherty AR, Moscati A, Peterson R, Edwards AC, Adkins DE, Bacanu SA, Bigdeli TB, Webb BT, Flint J, Kendler KS. SNP-based heritability estimates of the personality dimensions and polygenic prediction of both neuroticism and major depression: findings from CONVERGE. *Transl Psychiatry*. 2016;6(10):e926. doi: 10.1038/tp.2016.177. PubMed PMID: 27779626.

105. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-52. doi: 10.1038/nature08185. PubMed PMID: 19571811; PubMed Central PMCID: PMC3912837.
106. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, Hayeck T, Won HH, Schizophrenia Working Group of the Psychiatric Genomics Consortium DB, Risk of Inherited Variants in Breast Cancer s, Kathiresan S, Pato M, Pato C, Tamimi R, Stahl E, Zaitlen N, Pasaniuc B, Belbin G, Kenny EE, Schierup MH, De Jager P, Patsopoulos NA, McCarroll S, Daly M, Purcell S, Chasman D, Neale B, Goddard M, Visscher PM, Kraft P, Patterson N, Price AL. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American journal of human genetics*. 2015;97(4):576-92. doi: 10.1016/j.ajhg.2015.09.001. PubMed PMID: 26430803; PubMed Central PMCID: PMC3912837.
107. Demirkan A, Penninx BW, Hek K, Wray NR, Amin N, Aulchenko YS, van Dyck R, de Geus EJ, Hofman A, Uitterlinden AG, Hottenga JJ, Nolen WA, Oostra BA, Sullivan PF, Willemsen G, Zitman FG, Tiemeier H, Janssens AC, Boomsma DI, van Duijn CM, Middeldorp CM. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Molecular psychiatry*. 2011;16(7):773-83. doi: 10.1038/mp.2010.65. PubMed PMID: 20567237; PubMed Central PMCID: PMC3142964.
108. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, et al. The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of general psychiatry*. 1988;45(12):1069-77. Epub 1988/12/01. PubMed PMID: 2848472.
109. Kendler KSP, C.A. Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders. 1st ed. New York: Guildford Press; 2006.
110. John OP, Srivastava S. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*. 1999;2(1999):102-38.
111. Flint J, Chen Y, Shi S, Kendler KS, consortium C. Epilogue: Lessons from the CONVERGE study of major depressive disorder in China. *Journal of affective disorders*. 2012;140(1):1-5. Epub 2011/10/01. doi: 10.1016/j.jad.2011.09.006. PubMed PMID: 21955396; PubMed Central PMCID: PMC3387371.
112. Kessler RC, Andrews, G., Mroczek, D., Üstün, T.B., & Wittchen, H-U. The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). *International journal of methods in psychiatric research*. 1998;7:171-85.

113. Sanders AR, Levinson DF, Duan J, Dennis JM, Li R, Kendler KS, Rice JP, Shi J, Mowry BJ, Amin F, Silverman JM, Buccola NG, Byerley WF, Black DW, Freedman R, Cloninger CR, Gejman PV. The Internet-based MGS2 control sample: self report of mental illness. *The American journal of psychiatry*. 2010;167(7):854-65. Epub 2010/06/03. doi: 10.1176/appi.ajp.2010.09071050. PubMed PMID: 20516154.
114. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. *Psychopharmacol Bull*. 1973;9(1):13-28. PubMed PMID: 4682398.
115. Gray MJ, Litz BT, Hsu JL, Lombardo TW. Psychometric properties of the life events checklist. *Assessment*. 2004;11(4):330-41. doi: 10.1177/1073191104269954. PubMed PMID: 15486169.
116. Dick DM, Nasim A, Edwards AC, Salvatore JE, Cho SB, Adkins A, Meyers J, Yan J, Cooke M, Clifford J, Goyal N, Halberstadt L, Ailstock K, Neale Z, Opalesky J, Hancock L, Donovan KK, Sun C, Riley B, Kendler KS. Spit for Science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. *Front Genet*. 2014;5:47. doi: 10.3389/fgene.2014.00047. PubMed PMID: 24639683; PubMed Central PMCID: PMC3944794.
117. Schuster TL, Kessler RC, Aseltine RH, Jr. Supportive interactions, negative interactions, and depressed mood. *Am J Community Psychol*. 1990;18(3):423-38. PubMed PMID: 2264558.
118. Hendrick SS. A Generic Measure of Relationship Satisfaction. *Journal of Marriage and Family*. 1988;50(1):93-8. doi: 10.2307/352430.
119. Hays RD. User's manual for the medical outcomes study (MOS) core measures of health-related quality of life / Ron D. Hayes, Cathy Donald Sherbourne, Rebecca M. Mazel. Sherbourne CD, Mazel RM, editors. Santa Monica, CA: Rand Corp; 1995.
120. Jeronimus BF, Ormel J, Aleman A, Penninx BW, Riese H. Negative and positive life events are associated with small but lasting change in neuroticism. *Psychological medicine*. 2013;43(11):2403-15. doi: 10.1017/S0033291713000159. PubMed PMID: 23410535.
121. Kendler KS, Gardner CO, Prescott CA. Personality and the experience of environmental adversity. *Psychological medicine*. 2003;33(7):1193-202. PubMed PMID: 14580074.
122. Scott KM, Wells JE, Angermeyer M, Brugha TS, Bromet E, Demyttenaere K, de Girolamo G, Gureje O, Haro JM, Jin R, Karam AN, Kovess V, Lara C, Levinson D, Ormel J, Posada-Villa J, Sampson N, Takeshima T, Zhang M, Kessler RC. Gender and the relationship between marital status and first onset of mood, anxiety and substance use disorders. *Psychological medicine*. 2010;40(9):1495-505. doi: 10.1017/S0033291709991942. PubMed PMID: 19939327; PubMed Central PMCID: PMC3944794.

123. Kendler KS, Myers J, Prescott CA. Sex differences in the relationship between social support and risk for major depression: a longitudinal study of opposite-sex twin pairs. *The American journal of psychiatry*. 2005;162(2):250-6. doi: 10.1176/appi.ajp.162.2.250. PubMed PMID: 15677587.
124. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsdottir BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics*. 2015;47(3):284-90. doi: 10.1038/ng.3190. PubMed PMID: 25642633; PubMed Central PMCID: PMC4342297.
125. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007;81(3):559-75. Epub 2007/08/19. doi: 10.1086/519795. PubMed PMID: 17701901; PubMed Central PMCID: PMC1950838.
126. Kennedy BW, Quinton M, van Arendonk JA. Estimation of effects of single genes on quantitative traits. *J Anim Sci*. 1992;70(7):2000-12. PubMed PMID: 1644672.
127. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*. 2006;38(2):203-8. doi: 10.1038/ng1702. PubMed PMID: 16380716.
128. Fisher RA. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*. 1919;52(02):399-433.
129. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*. 2014.
130. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007;23(10):1294-6. doi: 10.1093/bioinformatics/btm108. PubMed PMID: 17384015.
131. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-7. doi: 10.1093/bioinformatics/btq419. PubMed PMID: 20634204; PubMed Central PMCID: PMC4342297.
132. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219. doi: 10.1371/journal.pcbi.1004219. PubMed PMID: 25885710; PubMed Central PMCID: PMC4401657.

133. Dunn OJ. Multiple Comparisons Among Means. *Journal of the American Statistical Association*. 1961;56(293):52-64. doi: 10.2307/2282330.
134. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R, Hubbard TJ. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research*. 2012;22(9):1760-74. doi: 10.1101/gr.135350.111. PubMed PMID: 22955987; PubMed Central PMCID: PMC3431492.
135. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;273(5281):1516-7. Epub 1996/09/13. PubMed PMID: 8801636.
136. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi: 10.1038/nature15393. PubMed PMID: 26432245; PubMed Central PMCID: PMC4750478.
137. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell*. 2010;141(2):210-7. doi: 10.1016/j.cell.2010.03.032. PubMed PMID: 20403315.
138. Cui L, Gong X, Tang Y, Kong L, Chang M, Geng H, Xu K, Wang F. Relationship between the LHPP Gene Polymorphism and Resting-State Brain Activity in Major Depressive Disorder. *Neural Plast*. 2016;2016:9162590. doi: 10.1155/2016/9162590. PubMed PMID: 27843651; PubMed Central PMCID: PMC45097818.
139. St Laurent G, Wahlestedt C, Kapranov P. The Landscape of long noncoding RNA classification. *Trends Genet*. 2015;31(5):239-51. doi: 10.1016/j.tig.2015.03.007. PubMed PMID: 25869999; PubMed Central PMCID: PMC4417002.
140. Bader V, Tomppo L, Trossbach SV, Bradshaw NJ, Prikulis I, Leliveld SR, Lin CY, Ishizuka K, Sawa A, Ramos A, Rosa I, Garcia A, Requena JR, Hipolito M, Rai N, Nwulia E, Henning U, Ferrea S, Luckhaus C, Ekelund J, Veijola J, Jarvelin MR, Hennah W, Korth C. Proteomic, genomic and translational approaches identify CRMP1 for a role in schizophrenia and its underlying traits. *Human molecular genetics*. 2012;21(20):4406-18. doi: 10.1093/hmg/dd273. PubMed PMID: 22798627; PubMed Central PMCID: PMC3529585.
141. Noack Watt KE, Achilleos A, Neben CL, Merrill AE, Trainor PA. The Roles of RNA Polymerase I and III Subunits Polr1c and Polr1d in Craniofacial Development and in Zebrafish Models of Treacher Collins Syndrome. *PLoS genetics*. 2016;12(7):e1006187. doi: 10.1371/journal.pgen.1006187. PubMed PMID: 27448281; PubMed Central PMCID: PMC4957770.

142. Opletalova K, Bourillon A, Yang W, Pouvelle C, Armier J, Despras E, Ludovic M, Mateus C, Robert C, Kannouche P, Soufir N, Sarasin A. Correlation of phenotype/genotype in a cohort of 23 xeroderma pigmentosum-variant patients reveals 12 new disease-causing POLH mutations. *Hum Mutat.* 2014;35(1):117-28. doi: 10.1002/humu.22462. PubMed PMID: 24130121.

143. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics.* 2012;28(19):2540-2. Epub 2012/07/31. doi: 10.1093/bioinformatics/bts474. PubMed PMID: 22843982; PubMed Central PMCID: PMC3463125.

144. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Consortium R, Genomics Consortium P, of the Wellcome Trust Consortium 3 GCFAN, Perry JRB, Patterson N, Robinson E, Daly MJ, Price AL, Neale BM. An Atlas of Genetic Correlations across Human Diseases and Traits 2015 2015-01-01 00:00:00.

145. Vilhjalmsjon BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, Hayeck T, Won HH, Kathiresan S, Pato M, Pato C, Tamimi R, Stahl E, Zaitlen N, Pasaniuc B, Belbin G, Kenny EE, Schierup MH, De Jager P, Patsopoulos NA, McCarroll S, Daly M, Purcell S, Chasman D, Neale B, Goddard M, Visscher PM, Kraft P, Patterson N, Price AL. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American journal of human genetics.* 2015;97(4):576-92. Epub 2015/10/03. doi: 10.1016/j.ajhg.2015.09.001. PubMed PMID: 26430803; PubMed Central PMCID: PMC4596916.

146. Cross-Disorder Group of the Psychiatric Genomics C, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, Absher D, Agartz I, Akil H, Amin F, Andreassen OA, Anjorin A, Anney R, Anttila V, Arking DE, Asherson P, Azevedo MH, Backlund L, Badner JA, Bailey AJ, Banaschewski T, Barchas JD, Barnes MR, Barrett TB, Bass N, Battaglia A, Bauer M, Bayes M, Bellivier F, Bergen SE, Berrettini W, Betancur C, Bettecken T, Biederman J, Binder EB, Black DW, Blackwood DH, Bloss CS, Boehnke M, Boomsma DI, Breen G, Breuer R, Bruggeman R, Cormican P, Buccola NG, Buitelaar JK, Bunney WE, Buxbaum JD, Byerley WF, Byrne EM, Caesar S, Cahn W, Cantor RM, Casas M, Chakravarti A, Chambert K, Choudhury K, Cichon S, Cloninger CR, Collier DA, Cook EH, Coon H, Cormand B, Corvin A, Coryell WH, Craig DW, Craig IW, Crosbie J, Cuccaro ML, Curtis D, Czamara D, Datta S, Dawson G, Day R, De Geus EJ, Degenhardt F, Djurovic S, Donohoe GJ, Doyle AE, Duan J, Dudbridge F, Duketis E, Eibstein RP, Edenberg HJ, Elia J, Ennis S, Etain B, Fanous A, Farmer AE, Ferrier IN, Flickinger M, Fombonne E, Foroud T, Frank J, Franke B, Fraser C, Freedman R, Freimer NB, Freitag CM, Friedl M, Frisen L, Gallagher L, Gejman PV, Georgieva L, Gershon ES, Geschwind DH, Giegling I, Gill M, Gordon SD, Gordon-Smith K, Green EK, Greenwood TA, Grice DE, Gross M, Grozeva D, Guan W, Gurling H, De Haan L, Haines JL, Hakonarson H, Hallmayer J, Hamilton SP, Hamshere ML, Hansen TF, Hartmann AM, Hautzinger M, Heath AC, Henders AK, Herms S, Hickie IB, Hipolito M, Hoefels S, Holmans PA, Holsboer F, Hoogendijk WJ, Hottenga JJ, Hultman CM, Hus V, Ingason A, Ising M, Jamain S, Jones EG, Jones I, Jones L, Tzeng JY, Kahler AK, Kahn RS, Kandaswamy R,

Keller MC, Kennedy JL, Kenny E, Kent L, Kim Y, Kirov GK, Klauck SM, Klei L, Knowles JA, Kohli MA, Koller DL, Konte B, Korszun A, Krabbendam L, Krasucki R, Kuntsi J, Kwan P, Landen M, Langstrom N, Lathrop M, Lawrence J, Lawson WB, Leboyer M, Ledbetter DH, Lee PH, Lencz T, Lesch KP, Levinson DF, Lewis CM, Li J, Lichtenstein P, Lieberman JA, Lin DY, Linszen DH, Liu C, Lohoff FW, Loo SK, Lord C, Lowe JK, Lucae S, MacIntyre DJ, Madden PA, Maestrini E, Magnusson PK, Mahon PB, Maier W, Malhotra AK, Mane SM, Martin CL, Martin NG, Mattheisen M, Matthews K, Mattingsdal M, McCarroll SA, McGhee KA, McGough JJ, McGrath PJ, McGuffin P, McInnis MG, McIntosh A, McKinney R, McLean AW, McMahon FJ, McMahon WM, McQuillin A, Medeiros H, Medland SE, Meier S, Melle I, Meng F, Meyer J, Middeldorp CM, Middleton L, Milanova V, Miranda A, Monaco AP, Montgomery GW, Moran JL, Moreno-De-Luca D, Morken G, Morris DW, Morrow EM, Moskvina V, Muglia P, Muhleisen TW, Muir WJ, Muller-Myhsok B, Murtha M, Myers RM, Myin-Germeys I, Neale MC, Nelson SF, Nievergelt CM, Nikolov I, Nimgaonkar V, Nolen WA, Nothen MM, Nurnberger JI, Nwulia EA, Nyholt DR, O'Dushlaine C, Oades RD, Olincy A, Oliveira G, Olsen L, Ophoff RA, Osby U, Owen MJ, Palotie A, Parr JR, Paterson AD, Pato CN, Pato MT, Penninx BW, Pergadia ML, Pericak-Vance MA, Pickard BS, Pimm J, Piven J, Posthuma D, Potash JB, Poustka F, Propping P, Puri V, Quested DJ, Quinn EM, Ramos-Quiroga JA, Rasmussen HB, Raychaudhuri S, Rehnstrom K, Reif A, Ribases M, Rice JP, Rietschel M, Roeder K, Roeyers H, Rossin L, Rothenberger A, Rouleau G, Ruderfer D, Rujescu D, Sanders AR, Sanders SJ, Santangelo SL, Sergeant JA, Schachar R, Schalling M, Schatzberg AF, Scheftner WA, Schellenberg GD, Scherer SW, Schork NJ, Schulze TG, Schumacher J, Schwarz M, Scolnick E, Scott LJ, Shi J, Shilling PD, Shyn SI, Silverman JM, Slager SL, Smalley SL, Smit JH, Smith EN, Sonuga-Barke EJ, St Clair D, State M, Steffens M, Steinhausen HC, Strauss JS, Strohmaier J, Stroup TS, Sutcliffe JS, Szatmari P, Szelinger S, Thirumalai S, Thompson RC, Todorov AA, Tozzi F, Treutlein J, Uhr M, van den Oord EJ, Van Grootheest G, Van Os J, Vicente AM, Vieland VJ, Vincent JB, Visscher PM, Walsh CA, Wassink TH, Watson SJ, Weissman MM, Werge T, Wienker TF, Wijsman EM, Willemsen G, Williams N, Willsey AJ, Witt SH, Xu W, Young AH, Yu TW, Zammit S, Zandi PP, Zhang P, Zitman FG, Zollner S, Devlin B, Kelsoe JR, Sklar P, Daly MJ, O'Donovan MC, Craddock N, Sullivan PF, Smoller JW, Kendler KS, Wray NR, International Inflammatory Bowel Disease Genetics C. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics*. 2013;45(9):984-94. doi: 10.1038/ng.2711. PubMed PMID: 23933821; PubMed Central PMCID: PMC3800159.

147. Han B, Pouget JG, Slowikowski K, Stahl E, Lee CH, Diogo D, Hu X, Park YR, Kim E, Gregersen PK, Dahlqvist SR, Worthington J, Martin J, Eyre S, Klareskog L, Huizinga T, Chen WM, Onengut-Gumuscu S, Rich SS, Major Depressive Disorder Working Group of the Psychiatric Genomics C, Wray NR, Raychaudhuri S. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nature genetics*. 2016;48(7):803-10. doi: 10.1038/ng.3572. PubMed PMID: 27182969; PubMed Central PMCID: PMC4925284.

148. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. 2010;11(6):446-50. doi: 10.1038/nrg2809. PubMed PMID: 20479774; PubMed Central PMCID: PMC2942068.

149. Bloom JS, Ehrenreich IM, Loo WT, Lite TL, Kruglyak L. Finding the sources of missing heritability in a yeast cross. *Nature*. 2013;494(7436):234-7. doi: 10.1038/nature11867. PubMed PMID: 23376951; PubMed Central PMCID: PMC4001867.
150. Neale MC, Kendler KS. Models of comorbidity for multifactorial disorders. *American journal of human genetics*. 1995;57(4):935-53. Epub 1995/10/01. PubMed PMID: 7573055; PubMed Central PMCID: PMC1801512.
151. Merikangas KR, Mehta RL, Molnar BE, Walters EE, Swendsen JD, Aguilar-Gaziola S, Bijl R, Borges G, Caraveo-Anduaga JJ, DeWit DJ, Kolody B, Vega WA, Wittchen HU, Kessler RC. Comorbidity of substance use disorders with mood and anxiety disorders: results of the International Consortium in Psychiatric Epidemiology. *Addict Behav*. 1998;23(6):893-907. PubMed PMID: 9801724.

Vita

Arden Anthony Mecca Moscati was born on December 19th, 1986 in Silver Spring, Maryland, and is an American citizen. He graduated from Quince Orchard High School, Gaithersburg, Maryland in 2005. He received his Bachelor of Arts in Psychology and Philosophy from Brasenose College, Oxford University, Oxford, United Kingdom in 2008. He received his Master of Arts (Oxon.) from Oxford University in 2013.