

Curriculum Vitae

Name: Elizabeth Marella Humphries

Contact Information: emarellahumphries@gmail.com

Degree and Date to be Conferred: PhD, December 2021

Collegiate Institutions Attended:

- University of Maryland Baltimore, 2014-2021
Degree awarded: PhD, Epidemiology, December 2021
- University of Alaska Fairbanks, 2005-2008
Degree awarded: M.S. Biology, December 2008
- University of Maryland Baltimore County, 2000-2004
Degrees awarded: B.S. Biological Sciences, B.A. Chemistry, December 2004

Major: Molecular Epidemiology

Professional Publications:

Kvarta M.D., et al. 2021. "Multiple dimensions of stress vs. genetic effects on depression." *Trans. Psychiatry* 11: 254. doi: 10.1038/s41398-021-01369-9.

Kalra, G., et al. 2020. "Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty." *PLOS Genetics* <https://doi.org/10.1371/journal.pgen.1009025>

Li, L., N. Liu, J.-M. Yu, E.M. Humphries, S. Li, B. Lindsay, Z.-J. Duan, and O. Colin Stine. 2016. "Etiology of diarrheal disease and evaluation of viral-bacterial co-infection in children under 5 years old in China: a matched case-control study." *Clin. Microbiol. Infect.* 22:381.e9-381.e16.

Humphries, E.M. and K. Winker. 2011. "Discord reigns among nuclear, mitochondrial, and phenotypic estimates of divergence in nine lineages of trans-Beringian birds." *Molecular Ecology* 20:573–558.

Humphries, E.M. and K. Winker. "Working through polytomies: Auklets revisited". 2010. *Molecular Phylogenetics and Evolution* 54: 88-96.

Humphries, E.M., J.L. Peters, J.E. Jónsson, R. Stone, A.D. Afton, and K.E. Omland. 2009. "Genetic differentiation between sympatric and allopatric wintering populations of Snow Geese." *The Wilson Journal of Ornithology* 121:730-738.

Peters, J.L., Y.N. Zhuravlev, I. Fefelov, E.M. Humphries, and K.E. Omland. 2008. "Multilocus phylogeography of a holarctic duck: colonization of North America from Eurasia by Gadwall (*Anas strepera*).” *Evolution* 62: 1469-1483.

Gaigalas, A.K., L. Wang, K.D. Cook, and E. Humphries. 2004. "Photodegradation of fluorescein in solutions containing n-propyl gallate.” *Journal of Physical Chemistry* 108: 4378-4384.

Professional Positions Held:

- Adjunct Professor (course: Biology I), Reading Area Community College, Reading PA, Aug 2012-Aug 2014
- Adjunct Professor (courses: Biology for Non-Majors, Human Health and Disease), Harford Community College, Bel Air MD, Jan 2012-Jun 2012
- Intern, U.S. Senator Mark Begich's office (policy focus: diseases, health, STEM education, GM salmon), Washington DC, Sept 2011-Dec 2011
- Instructor, Alaska BIOPrep
 - Genetics, UAF Alaska Summer Research Academy (ASRA) 2010
 - Introduction to Molecular Biology, Rural Alaska Honors Institute (RAHI), Summer 2009
- Curatorial Assistant, UA Museum Dept. Ornithology, Fall 2007-Summer 2008
- Lab Technician, Omland Lab (UMBC), Jan 2003-May 2005
- Tutor, Chemistry Tutorial Center (UMBC), Feb 2001-May 2004; Student Director Fall 2004
- Instructor, UMBC Summer Chemistry Workshop, Summer 2002, 2003
- Lab Technician, Chemistry, National Institute of Standards and Technology, Summer 2002
- Field Assistant, Patuxent Wildlife Research Center, Summer 2001
- Lab Technician, Ottinger Lab (Univ. MD College Park), Sept 1999-May 2000

Community Activities or Special Awards:

- Panelist, Virtual STEM roundtable for female HS students, UMB, Fall 2020
- Graduate student representative to Dept of Epidemiology's Graduate Research Committee, Fall 2018 – Winter 2019
- Nominee, Outstanding Graduate Teaching Assistant (2008-2009)
- Organizer, Applying to Graduate School workshop series, UAF Dept of Wildlife and Biology, Spring 2008
- EPSCoR Fellow, University of Alaska Fairbanks (2005-2007)
- Phi Beta Kappa member, Eta Chapter of Maryland, UMBC (elected May 2005)
- Outstanding Undergraduate Research in Biology award, UMBC (May 2004)
- Meyerhoff Scholar, UMBC, M12 Cohort

Abstract

Title of Dissertation: Genetics of mood disorder diagnosis, behavioral endophenotypes, and cognition in the Old Order Amish founder population

Elizabeth M. Humphries, Doctor of Philosophy, 2021.

Dissertation Directed by: Seth A. Ament, Assistant Professor, Department of Psychiatry and

Institute of Genome Sciences

Mood disorders, including major depression and bipolar disorder, represent a substantial public health burden. Genome-wide association studies (GWAS) of mood disorders in large case-control cohorts have identified numerous loci that together account for part of the risk, but pathophysiological mechanisms have been elusive. Individual common variants confer very small risk, and samples have been underpowered to detect rare variants with larger effects. Extended pedigrees from population isolates could address these challenges, as some risk-conferring alleles may be highly enriched, relative to the broader population. Here I report a GWAS of mood disorders in an Old Order Amish (OOA) founder population. By integrating genetic and phenotypic data from three independently collected OOA cohorts enriched for mood disorders with OOA population controls (total $n=1,672$), I identified four genome-wide significant risk loci, 2 of which are novel. The observed effect sizes of the associated loci are substantially greater than those identified in the broader population, and the risk-associated haplotypes harbor otherwise rare coding variants enriched in OOA. At one locus, I identified a population-enriched, non-synonymous variant in the *CUX1* alternative splicing product associated with >3-fold relative risk. Quantitative behavioral and neurocognitive traits in a subset of 314 subjects revealed effects of risk variants on sub-clinical depressive symptoms and

working memory. These findings provide insight into the genetic architecture of mood disorders and provide a substrate for mechanistic and clinical studies.

Genetics of Mood Disorder Diagnosis, Behavioral Endophenotypes, and Cognition in the Old
Order Amish Founder Population

by
Elizabeth Marella Humphries

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Acknowledgements

My deepest thanks go to the Old Order Amish and Plain Mennonite families and communities who participated in these studies, even though the results would not directly benefit them or their loved ones. I would also like to thank all the researchers who contributed to this project: Dr L. Elliot Hong and Dr. Peter Kochunov from the Maryland Psychiatric Research Center, as well as their staff; Dr Francis McMahon, Dr Kwangmi Ahn, and Dr Fabiana Lopes and their staff at the National Institutes of Mental Health; Dr Maja Bucan, Dr Rachel Kember, and their associates from the University of Pennsylvania; Dr Braxton Mitchell, Dr Toni Pollin, and their associates from the Program for Personalized and Genomic Medicine at the University of Maryland Baltimore; Dr Peter Zandi and Dr Fernando Goes from Johns Hopkins University; Dr Alan Schuldiner and Dr Cris Van Hout from the Regeneron Genetics Center; and all the staff at the Amish Research Clinic in Lancaster County, PA.

Table of Contents

Chapter 1: Introduction	1
Chapter 2: Background Literature	4
Depressive Disorders	5
Bipolar Disorders	6
Research Domain Criteria and mood disorders	7
Genetics and neurobiology of mood disorders	8
The Plain people	10
History of genetics research in the Plain people	12
Mood disorder genetics research in the Plain people	15
Chapter 3: Methods	17
Cohorts	17
Preliminary data processing, quality control, and imputation	20
Assessing population stratification and sample overlap	21
Assessing imputation accuracy	22
Affection model status	23
GWA analysis	24
LD clumping	25
Pseudoreplication	25
Polygenic risk score analysis	25
Annotation of loci and variants	26
MAGMA	27
Gene interaction network analysis	28
Effects of risk variants on quantitative and neurocognitive phenotypes	29
Data availability	30
Chapter 4: Results	31
An expanded genetic cohort for the investigation of mood disorders in the OOA	31
Identification of genome-wide significant risk loci for mood disorders in the OOA	36
Pseudoreplication and overlap with known risk genes and risk loci	40
Gene networks associated with mood disorders in the OOA	45
Polygenic risk scores explain a small proportion of the risk for mood disorders in the OOA	48

Quantitative behavioral and neurocognitive phenotypes	50
An Amish-enriched missense variant in the <i>CUX1</i> Alternative Splicing Product	53
Chapter 5: Discussion	58
Appendix	63
References	99

List of Tables

Table 4.1: Imputation accuracy and missingness	33
Table 4.2: OOA cohorts samples by diagnosis and cohort.....	37
Table 4.3: GWAS results and MAF for significant loci	41
Table 4.4: MAGMA results	46
Table 4.5: Heritability and correlation of behavioral and cognitive phenotypes	52
Table 4.6: PheWAS results for risk loci	56
Table A1: Full GWAS results for SNPs in LD with lead SNPs	63
Table A2: Full pseudoreplication results	66
Table A3: GWAS results for suggestive loci	69
Table A4: Overlap of previously identified risk loci	72
Table A5: MAGMA results for top 250 genes in gene-level analysis	73
Table A6: GO term analysis results	80
Table A7: Full pheWAS results for behavioral traits	83
Table A8: Full pheWAS results for cognitive traits, part 1	86
Table A9: Full pheWAS results for cognitive traits, part 2	89
Table A10: 2x2 tables of genotype by affection status	92

List of Figures

Figure 4.1: Population structure among cohort	35
Figure 4.2: GWAS Q-Q and Manhattan plots	39
Figure 4.3: Region plots for significant risk loci	43
Figure 4.4: MAGMA results	47
Figure 4.5: PRS broken down by affection status	50
Figure 4.6: PheWAS results	55
Figure 4.7: Expression of <i>CUX1/CASP</i> across brain regions	57
Figure A1: Scatterplots of first 10 PCS among Lancaster OOA samples	93
Figure A2: Hierarchical clustering of 3q28/29 carriers	94
Figure A3: Pedigrees of ASMAD families with 5q13 carriers	95
Figure A4: Hierarchical clustering of 5q13 carriers	96
Figure A5: Pedigrees of ASMAD families with 7q22 carriers.....	97
Figure A6: Hierarchical clustering of 7q22 carriers	98

List of Abbreviations

ACP	Amish Connectome Project cohort
AMBiGen	Amish and Mennonite Bipolar Genetics cohort
ASD	Autism Spectrum Disorder
ASMAD	Amish Study of Major Affective Disorders cohort
BD	Bipolar Depression
BDI	Beck Depression Index
BSDS	Bipolar Spectrum Diagnostic Scale
DSTT	MD Depression State-Trait Scale (Trait)
GWAS	Genome-wide association study
MDD	Major Depressive Disorder
MDDR	Recurrent Major Depressive Disorder
OOA	Old Order Amish
PheWAS	Phenome-wide association study
PM	Plain Mennonite
PRS	Polygenic risk score
TOPMed	Trans-Omics of Precision Medicine cohort
SABP	Schizoaffective Disorder
SCZ	Schizophrenia
SNP	Single nucleotide polymorphism

Chapter 1: Introduction

Mood disorders, including major depressive disorder (MDD) and bipolar disorder (BD), affect more than 300 million people worldwide¹. Episodes of MDD are characterized by an inability to feel pleasure along with changes in sleep, energy, and appetite, while BD is characterized by episodes of the extreme mood states of mania and depression². Both conditions are major risk factors for suicide³. Many individuals with mood disorders also experience cognitive deficits, including diminished ability to sustain attention, decreased executive brain functions, and difficulty with emotional processing^{2,4,5}. While some patients' symptoms resolve with mood-stabilizing and antidepressant pharmacotherapy, ~50% do not respond adequately to any existing treatment. Consequently, mood disorders are a major cause of disability in the world.

Both disorders are strongly heritable. In MDD, first-degree relatives of probands have nearly three-fold relative risk, and 30-50% of the risk is estimated to be heritable^{6,7}. In BD, first-degree relatives of probands have an eight to ten-fold elevated risk for the disease, and 60-80% of risk is estimated to be heritable^{8,9}. First-degree relatives of BD probands also have a three-fold relative risk of developing MDD, suggesting shared genetics^{9,10}. Genome-wide association studies (GWAS) in large case-control cohorts have revealed 64 genome-wide significant risk loci for BD and 178 for MDD and confirmed a strong genetic correlation¹¹⁻¹⁶. However, the effect sizes of individual risk variants are extremely small, and collectively these risk loci explain between 10 and 20% of the observed heritability¹¹⁻²⁶. Moreover, the causal mechanisms at most of these loci remain speculative, since very few of the variants on risk-associated haplotypes have been functionally characterized. Thus, the genetic causes and biological mechanisms of

mood disorders remain poorly understood. Identifying genetic risk factors for mood disorders is a promising path toward novel therapeutics and diagnostics.

More than 50 years ago, McKusick first articulated the unique advantages of the Lancaster Old Order Amish (OOA) population for genetic epidemiology²⁷, and genetic studies in this population have subsequently led to the discovery of risk variants and pathophysiological mechanisms for numerous complex traits as well as more than 60 Mendelian traits²⁸⁻³². The OOA are a closed founder population of ~40,000 individuals living primarily in Lancaster County, Pennsylvania. Nearly all present-day OOA trace their ancestry to a small number of founding individuals, who emigrated from the Alsace region at the border of Switzerland, France, and Germany, starting in the early 1700s³³. The Lancaster OOA are conservative Anabaptists who belong to a religious sect of the Protestant Reformation³³. In addition to the Lancaster OOA, Anabaptist populations include Amish living in Ohio and Indiana, as well as Plain Mennonite individuals. These groups all share a similar population history, with a small founder population emigrating to the Americas in the 1700 and 1800s, followed by a rapid population expansion³³⁻³⁵. As in other population isolates^{28-32,36-39}, the population bottleneck led to the enrichment of many functional alleles that are rare in the broader population. Some of these alleles may have larger effects on disease risk than common variants typically identified through GWAS in the broader population. In addition to the highly enriched carrier frequency, the Plain people have large families and very low migration, and long-standing partnerships with the genetics research community greatly facilitate recruitment^{22,23,28-31,40-44}. The Plain population is also unique due to the relative uniformity in education, lifestyle, and socioeconomic status, and a reduced influence

of alcohol and illicit drugs -- all characteristics that should provide higher fidelity in neuropsychiatric phenotyping.

Genetic studies of mood disorders in the Plain people were initiated in the 1970s, primarily within the Amish Study of Major Affective Disorders. Initial studies in this cohort identified suggestive linkage peaks, while more recent genome sequencing studies revealed polygenic effects of single-nucleotide variants and copy number variants^{22,23,40-44}. However, previous studies were limited by their small sample sizes^{22,23,39-44}. Here, I report a GWAS of mood disorders in an expanded Old Order Amish cohort, revealing the first genome-wide significant risk loci for mood disorders in this population. I combined data from across 3 distinct Plain populations previously phenotyped for MDD and BD and for whom genome-wide genotyping was available; these data were also combined with an unphenotyped cohort that served as population controls. I used the combined dataset to perform an association analysis. I also used subclinical phenotype data from one of the cohorts to examine the association between the deeper phenotypic measures and the risk loci identified in the larger association analysis.

Chapter 2: Background Literature

Mood disorders are a class of psychiatric disorders characterized by symptoms of depression, mania, and other emotional disturbances². In the DSM-5, severe mood disorders included depressive disorders (disorders characterized by episodes of anhedonia and dysphoria) and bipolar disorders (disorders characterized by alternation between episodes of extreme mood states of mania/hypomania and depression)². A subset of patients with bipolar disorders and depressive disorders also experiences psychosis². Many individuals with mood disorders also experience cognitive deficits, including diminished ability to sustain attention, decrease in executive brain functions, and difficulty with emotional processing^{2,4,5}. Mood disorders, including major depressive disorder (MDD) and bipolar disorder (BD) affect more than 300 million people worldwide¹. In the United States, mood disorders affect ~9.5% of Americans 18 and older each year, or 20.9 million people⁴⁵, and an estimated 21.4% of Americans will be diagnosed with a mood disorder at some point in their life⁴⁵. While some patients' symptoms resolve with mood-stabilizing and antidepressant pharmacotherapy, ~50% do not respond adequately to any existing pharmacological treatment. Consequently, mood disorders are among the leading causes of disability in the United States and are associated with a substantial increased risk for suicide^{3,45}. Insufficient understanding of their disease mechanisms in the developing and adult brain hinders the development of more effective treatments. Identifying genetic risk factors for mood disorders is a promising path toward novel therapeutics and diagnostics.

Depressive disorders. The DSM-5 currently lists 8 separate depressive disorders, including major depressive disorder (MDD), persistent depressive disorder (dysthymia), disruptive mood dysregulation disorder, premenstrual dysphoric disorder, substance/medication-induced depressive disorder, depressive disorder due to another medical condition, other specified depressive disorder, and unspecified depressive disorder^{2,46,47}. Although all these disorders are characterized by the presence of sad, empty, or irritable moods (as well as sleep changes and cognitive deficits), they differ in their duration, intensity, timing, or suspected etiology⁴⁶. MDD is the primary mood disorder among these DSM diagnoses and is one of the leading causes of disability in people over the age of 5⁴⁵. Individuals with MDD have experienced episodes of depressed mood (anhedonia), loss of interest in pleasure (dysphoria), or both; these episodes must last for at least 2 weeks, though most last for 6 months^{2,46,47}. Although a diagnosis can be made based on a single depressive episode, 50% of individuals with MDD will have a recurrence of symptoms within the first year of their initial diagnosis, while 85% will experience a second episode within their lifetime⁴⁵.

Approximately 7.1% of American adults have had at least one major depressive episode⁴³. More females are affected than males (8.7% versus 5.3%)⁴⁵. Sixty-four percent of individuals with MDD (or 4.5% of Americans) reported severe impairment in their daily life during episodes⁴⁵. MDD is also more common among teenagers, affecting 13.3% of children between ages 12 and 18⁴⁵. Twenty percent of teenage girls are affected, compared to just 6.8% of teenage boys⁴⁵. Seventy-one percent of affected teenagers report severe impairment in their daily life during episodes⁴⁵. Additionally, 1.9 million children between the ages of 3 and 17 have been diagnosed with depression⁴⁸.

Bipolar disorders. The DSM-5 currently recognizes 7 types of bipolar disorders (BD), including bipolar I, bipolar II, cyclothymic disorder, substance/medication-induced bipolar and related disorder, bipolar and related disorder due to another medical condition, other specified bipolar and related disorder, and unspecified bipolar and related disorder^{2,46,47}. These disorders differ in the type of mania experienced as well as the duration, timing, and etiology of episodes⁴⁷. The most common bipolar subtypes are bipolar I (BDI) and bipolar II (BDII)⁴⁷. Individuals with BDI have experienced at least 1 manic episode (a period of at least 1 week in which the individual experiences extreme changes in behavior), while an individual with BDII has experienced at least 1 depressive episode and 1 hypomanic episode^{2,46,47}. (Hypomania is similar to mania, though the behavioral changes are less severe and may not last for a week⁴⁶. A hypomanic individual is still able to function, unlike individuals experiencing a manic episode⁴⁶).

Individuals with BD I or II may be misdiagnosed as having MDD if they experience depressive episodes first⁴⁶. Bipolar disorders are generally considered “adult-onset” disorders; although children as young as 13 can be diagnosed, most people experience their first symptoms in their early 20s⁴⁶. The DSM-5 recommends that children under the age of 12 with behavioral dyscontrol and persistent irritability be diagnosed with disruptive mood dysregulation disorder, a depressive disorder, as opposed to a bipolar disorder⁴⁶. (This diagnosis is placed under the depressive disorders instead of the bipolar disorders because children who are diagnosed with it usually go on to develop MDD and related disorders, not BD⁴⁶.)

Approximately 2.8% of American adults have been diagnosed with BD⁴⁵. Males and females are affected equally (2.9% versus 2.8%)⁴⁵. Eighty-three percent of individuals with BD

(or 2.3% of Americans) reported severe impairment in their daily life during episodes⁴⁵. BD has similar prevalence among teenagers, affecting 2.9% of children between ages 12 and 18, although more teenage girls are affected than teenage boys (3.3% versus 2.6%)⁴⁸. Ninety percent of affected teenagers report severe impairment in their daily life during episodes⁴⁸.

Research Domain Criteria and mood disorders. In the past decade, researchers have begun classifying psychiatric disorders based on research domains, or dimensions of observable behavioral symptoms and neurobiological measures⁴⁹. Research domains are a way to synthesize the behavioral and cognitive processes that, if perturbed, lead to psychiatric illness. These domains include negative valence (responses to aversive situations such as fear, anxiety, or loss), positive valence (responses to positive motivational situations such as reward seeking and reward or habit learning), cognitive processes (such as attention or perception), systems for social processes (processes involved in responses to interpersonal settings, such as perception and interpretation of others' actions), arousal/regulatory systems (processes that activate and regulate brain activity and behavior, such as sleep regulation and energy), and sensorimotor systems (systems responsible for control and execution of motor behaviors and their refinement during development)^{49,50}. MDD spans at least two research domains (negative valence and positive valence)⁵¹. BD may span both the negative valence and positive valence domains in addition to the arousal/regulatory systems domain⁵².

The research domain criteria approach to psychiatric disorders highlights the shared etiology between disorders that are commonly comorbid or have some overlapping symptoms⁵⁰. Mood disorders and psychotic disorders (particularly schizophrenia and schizoaffective disorder)

appear to share aspects of neurobiological origin⁴⁶. Some individuals with BD may experience psychosis in addition to their mood disorder symptoms, while individuals with schizoaffective disorders experience mood symptoms (either depression or mania) in addition to schizophrenia symptoms (such as hallucinations or delusions)^{2,46}. Additionally, ~40% of schizophrenia patients also exhibit depressive symptoms, and individuals with a comorbid diagnosis of depression and schizophrenia are at greater risk of negative outcomes such as relapsing, developing substance abuse disorders, and suicidal ideation⁵³.

Genetics and neurobiology of mood disorders. Mood disorders are thought to be polygenic in origin. In MDD, first-degree relatives of probands have nearly 3-fold relative risk of developing MDD, and 30-50% of the risk appears to be heritable^{6,7}. In BD, first-degree relatives of probands have an 8 to 10-fold elevated risk for the disease (almost ten times the risk of the general population), and 60-80% of risk is estimated to be due to heritable factors^{8,9}. First-degree relatives of BD probands also have a 3-fold relative risk of developing MDD, suggesting potentially some shared genetic origin of MDD from BD^{9,10}. They also have a 6-fold relative risk of developing schizophrenia¹⁰. Conversely, first-degree relatives of individuals with schizophrenia have a 2-fold increased risk of developing a mood disorder¹⁰.

Researchers began investigating the genes responsible for developing BD using linkage methods in the 1980s⁵⁴. Initial studies reported linkage between BD and regions on 17 different chromosomes⁵⁴. However, these findings have failed to be unambiguously replicated⁵⁴. Fewer studies have examined linkage between genomic regions and MDD, but they too have only resulted in ambiguous or weak results⁵⁴. More recently, consortia have created large case-control

cohorts for genome-wide association studies (GWAS), using both whole genome sequencing and whole exome sequencing. So far, these studies have revealed 64 genome-wide significant risk loci for BD and 178 for MDD^{14,15}. However, the effect sizes of individual risk variants are extremely small, and collectively these risk loci explain 10-20% of the observed heritability¹¹⁻²⁶. Moreover, the causal variants and target genes at most of these loci remain speculative, since very few of the risk haplotypes include protein-coding variants or functionally characterized non-coding variants. Mood disorders appear to be highly polygenic, with much of the risk conferred by large numbers of loci each contributing a small fraction of increased disease risk⁵⁵⁻⁵⁸. Additionally, the genetic architecture of mood disorders is likely similar to that of autism, in which rare variants explain a large portion of the heritability^{59,60}. One outcome of genetic evaluation to understand the basis of BD and MDD is that it is now believed that family-based studies allow for an enrichment of potential rare risk genes among the study sample, improving the study's power to detect moderate effect sizes despite a modest sample size³⁶.

Several risk genes that have been identified (including *CACNA1C* and *ANKK3*) suggest that mood disorders are at least partly the result of dysfunctional ion channels or ion channel-associated proteins¹⁷. *CACNA1C* encodes a core subunit of the L-type voltage-gated calcium channel, while *ANKK3* encodes ankyrin G, a protein involved in the coupling of voltage-gated sodium channels with the cytoskeleton¹⁷. *ANKK3* also appears to play a role in dendrites and glia¹⁷. In addition to the individual risk gene results, Nurnberger et al. identified six pathways with a replicable association to BD using four independent GWAS; these pathways include NMDA receptors and glutamate signaling in addition to calcium signaling, providing additional evidence that calcium dysregulation and dysfunctional ion channels are involved in the

development of BD⁶¹. MDD has a similar pathophysiology; recent research suggests MDD-associated genes are involved in calcium signaling (*CACNA1E*, *CACNA2D1*), dopaminergic and glutamate neurotransmission (*DRD2*, *GRIK5*, *GRM5*, *SORCS3*), and presynaptic vesicle trafficking (*PCLO*)^{11,12,15}. An additional risk gene, *NEGR1*, is involved in neuronal spine plasticity and is a member of the *Negr1-Fgrf2* pathway; Carboni et al. demonstrated that *NEGR1* is downregulated by escitalopram, fluoxetine, and nortriptyline (common antidepressants) in Flinders sensitive line rats (a depression rat model)⁶². A gene-set analysis also suggests MDD-associated risk genes are involved in synaptic structure and activity, somatosensory pyramidal neurons, and response to stimuli¹². Unfortunately, many of these studies fail to replicate the findings of each other.

The Plain people. The Plain people are a group of conservative Anabaptists, a religious sect that arose in 16th century Europe during the early years of the Protestant Reformation³³. The early Anabaptists were a small group of reformers who felt the changes adopted by most Protestant groups were not far-reaching enough and broke off to form their own sect³³. Anabaptists today are descended from three major groups: the Swiss Brethren (which includes both Old Order Amish and Swiss Mennonite groups), the Hutterite Brethren of Austria, and the Dutch and Prussian Mennonites³³. Plain Anabaptist sects structure their life around a literal interpretation of the Sermon on the Mount (Matthew 5-7), which precludes violence, taking oaths, and participating in military action or civil government³³. Plain people also eschew most modern technology and live much as their ancestors did in the 18th century to preserve family and their community as the center of their lives³³.

Two of the major Plain groups in the United States are the Old Order Amish (OOA) and the Plain Mennonites (PM), which include the descendants of both the Swiss Brethren and the Dutch and Prussian Mennonites. The Swiss Brethren are a branch of Anabaptists that originated in Zurich in the early 1500s and spread through Switzerland and into the nearby areas of Alsace and the Palatinate in eastern Germany³⁵. In the late 1600s, the Swiss Brethren split after a disagreement between Jakob Ammann and Hans Reist over the concept of shunning³³. Ammann's followers eventually became known as the OOA⁴³. Reist's followers adopted the name Mennonite³³. The OOA emigrated to North America in two waves; in the first wave (1727-1770), several hundred OOA individuals relocated to areas around Lancaster County, Pennsylvania^{33,34}. A much larger group of OOA (estimated around 3000 individuals) emigrated from Europe between 1815 and 1860 and settled in Ohio and Indiana³⁴. The Swiss Brethren PM began arriving in North America in 1683, when a group of 100 immigrants arrived in Germantown, Pennsylvania³⁴. Between 1683 to 1883 approximately 8000 Mennonites immigrated to North America, settling first in areas around Lancaster County, Pennsylvania, then in Ohio, Indiana, and Illinois, and finally in Kansas and South Dakota³⁴. Their descendants continued to spread throughout the United States. Additionally, the Dutch and Prussian Mennonite branch originated in the Netherlands in the early 1500s³⁵. A group of Dutch Mennonites emigrated first to the Vistula Delta in Prussia, then to the Russian Empire (present-day Ukraine as well as the Volga region, Orenburg Governorate, and Western Siberia)³⁵. By the late 1800s, religious persecution led these Mennonites to start emigrating to the Americas³⁵. Today, Dutch and Prussian-ancestry Mennonites are often referred to as Russian or colony Mennonites⁴⁵.

History of genetics research in the Plain people. The Plain people present unique advantages for genetics studies^{37,38}, first articulated by McKusick, Egeland, and Hostetler over 50 years ago²⁷. Their small founding population and subsequent population bottleneck, as well as their continued genetic isolation from other populations, led to the enrichment of many functional alleles that are extremely rare in the broader European population, some of which may have larger effects on complex disease risk than common variants typically identified through GWAS in the broader population^{30,31}. Besides the highly enriched carrier frequency, the OOA have large families and very low migration, and long-standing partnerships with the genetics research community greatly facilitate recruitment.

A pair of studies in Indiana are likely the first examples of genetics research in the OOA⁶³. In the early 1960s, a group of physicians at the Caylor-Nickel Hospital identified 7 cases of an autosomal recessive limb-girdle muscular dystrophy among Amish families in Adams County, IN^{63,64}. The manuscript detailing their preliminary results is the first scientific research article about genetic diseases in the Plain people^{63,64}. Around the same time, David H. Martin (the chief health commissioner of Elkhart County, IN), and the Department of Human Genetics at the University of Michigan began testing an inexpensive kit to detect PKU in home birthed infants and uncovered a cluster of PKU cases in related Amish households^{63,65}.

A year later, Victor A. McKusick became interested in the possibilities of the OOA population for genetics research after meeting both John A. Hostetler, a professor of anthropology whose book *Amish Society* was reviewed for publication by McKusick, and Harold E. Cross, a second-year medical student at John Hopkins who introduced McKusick to Martin's

work on PKU in Amish families^{63,65}. These three, along with David Krusen (a Lancaster County family physician who had identified a cluster of Lancaster OOA patients with acondroplasia-like characteristics) and Janice A. Egeland (who had been doing field work for several years among the OOA in central PA and had mapped numerous OOA pedigrees) were instrumental in initiating the research partnership between Johns Hopkins' Moore Clinic and the OOA community⁶³. Employees of the Moore Clinic (including Egeland and two Amish women, Mary Ann Riehl and Sara E. Fisher) started interviewing members of the Lancaster County OOA parishes and collecting blood samples for analysis in 1962⁶³. Two years later, McKusick and Cross started surveying the OOA population in Holmes County, OH, focusing specifically on neurological disorders⁶³.

Genetics research in the OOA and other Plain populations has included numerous studies identifying genetic causes of both rare and common disorders. The first few publications from the Moore Clinic collaborations identified new diseases in the Lancaster County OOA (such as Ellis-van Creveld syndrome and cartilage hypoplasia), extended existing OOA genealogies, and presented a comprehensive demographic study of the Holmes County OOA^{28,29,67,68}. By 1978, researchers from Johns Hopkins, the University of Michigan, and other research institutions had identified 16 new genetic disorders among the OOA, in addition to increasing understanding about 18 previously known genetic disorders and contributing to work on blood groups, immunology, chromosomal variation, and complex, common diseases like hypertension and diabetes⁶⁹.

Although McKusick and many of his collaborators ended their genetics research in the OOA by the mid-1970s, many other researchers and institutions have continued to work with

OOA communities, as well as other Plain populations such as the Plain Mennonites and Hutterites. The Biochemical Genetics Laboratory in London, Ontario maintains a genetic disorder database for the OOA, Mennonite, and Hutterite populations as a reference for researchers and physicians who are identifying and treating genetic disorders in these groups⁷⁰. So far, more than 100 genetic disorders (including more than 60 Mendelian disorders) have been described and investigated in the Plain populations⁷⁰. In Lancaster County, the OOA community and researchers have opened specialized clinics like the Clinic for Special Children (focusing on diagnosing and treating rare disorders) and the Amish Research Clinic, operated in conjunction with the University of Maryland School of Medicine. The collaborations between the Lancaster OOA community and the University of Maryland have resulted in major contributions in the genetics of many common complex diseases. University of Maryland researchers identified a loss of function variant in *APOC3* among the OOA associated with a cardioprotective phenotype, as well as the *CYP2C19*2* genotype, which is associated with a lack of response to clopidogrel (Plavix)^{30,71}. The *CYP2C19*2* genotype association was later replicated in the general US population and the FDA now advises patients are first checked for this genotype before beginning a clopidogrel regimen⁷¹. Other research on diabetes and insulin phenotypes led to the discovery of a 19-bp deletion in *LIPC* that causes insulin resistance and impairment in lipolysis⁷². Additionally, many novel genetic associations for neuropsychiatric and neurodegenerative diseases have been identified in the OOA, including Alzheimer's, Parkinson's, brain aging, and Autism Spectrum Disorder⁷³⁻⁷⁸. One well-established autism risk gene, *CNTNAP2*, was first reported in the OOA and later replicated as a risk gene for multiple psychiatric disorders in the general population⁷⁸⁻⁸³.

Mood disorder genetics research in the Plain people. Certain characteristics of the OOA population may improve power for psychiatric genetics studies by providing higher fidelity in neuropsychiatric phenotyping. These include the relative uniformity in education, lifestyle, and socioeconomic status, and a reduced influence of illicit drugs. Genetic studies of mood disorders in the OOA were initiated in the 1970s by Egeland and collaborators, initially identifying a genomic region on chromosome 11 linked to a diagnosis of BD, though this signal disappeared as the pedigree was expanded^{40,41}. Subsequent linkage research also identified possible BD-associated regions on chromosomes 6, 13, and 15, as well as regions on both the p and q arms of chromosome 4, suggesting that BD in the OOA has a complex inheritance^{42,43}. In the 2010s, researchers generated whole genome genotypes, whole genome sequences, and exome sequencing data for the individuals enrolled in Egeland's cohort and confirmed the complex polygenic architecture of BD in the OOA^{23,84}. Exome sequencing also identified a missense variant in *KCNH7* that was associated with increased risk for bipolar disorder and results in decreased function of neuronal HERG3/Kv11.3 potassium channel, specifically altering the channel's steady state voltage dependence and activation kinetics²². *KCNH7* has also been identified as a risk gene for BD in a GWAS of a Taiwanese cohort, and a study among schizophrenia patients of Han Chinese ancestry found an association with variation in the *KCNH7* gene and response to risperidone treatment^{85,86}. Additionally, there is a trend of increasing numbers of CNVs in known Mendelian disease loci among OOA individuals diagnosed with BD⁴⁴. However, previous studies were limited by their small sample sizes and

provided inconsistent results^{22,23,39-44,85}, and most of the genetic risk architecture remain unknown.

Chapter 3: Methods

Cohorts

Amish Connectome Project (ACP)

The Amish Connectome Project (ACP) was initiated in 2015 (PIs: L. Elliot Hong and Peter Kochunov, University of Maryland School of Medicine). The ACP study recruits individuals from Pennsylvania and Maryland who belong to Anabaptist families with at least two affected probands with a severe mental illness. Individuals were excluded from the study if they had a history of major medical or neurological conditions such as intellectual disability or epilepsy, as well as individuals who were currently experiencing unstable or major medical issues.

Participants were administered the Structured Clinical Interview for DSM-IV or V (SCID)⁸⁷ by an experienced clinician, followed by a consensus-based psychiatric diagnosis determination. Several cognitive traits were assessed, including digit sequencing (verbal working memory), digit symbol coding (processing speed, visuospatial memory), spatial span (visuospatial working memory), and the Wechsler Abbreviated Scale of Intelligence⁸⁸ (WASI) matrix reasoning and vocabulary subtests (IQ and cognitive ability). In addition, all participants completed three self-report questionnaires on mood symptoms, including the Beck Depression Index⁸⁹ (assessing current and past depression symptoms), the Maryland Trait Depression scale⁹⁰ (which asks about lifetime depression symptoms), and the Bipolar Spectrum Diagnostic Scale⁹¹ that assess different aspects of depression and bipolar symptoms. SNP array genotypes are available in 428 participants of this sample. Genotyping was performed at the Human Genetics Branch of the National Institute of Mental Health Intramural Research Program using the Infinium Global

Screening Array (Illumina, Carlsbad, CA). I also had access to whole genome sequence for 214 ACP individuals, generated by the Regeneron Genetics Center in Tarrytown, NY.

Amish and Mennonite Bipolar Genetics Project (AMBiGen)

The AMBiGen study (Francis J. McMahon, PI; National Institute of Mental Health Intramural Research Program) recruits pedigrees through probands with bipolar or schizoaffective disorder. Individuals were recruited via screening admissions to local clinics and through publicity programs and underwent evaluations using the Diagnostic Interview for Genetics Studies^{92,93}. Researchers also collected information from available medical records and interviews with relatives. A best-estimate diagnosis was made by two or three independent, non-interviewing clinicians. SNP array genotypes were generated for 856 AMBiGen participants. 309 of these subjects were genotyped on the Illumina Global Screening Array through a contract with the Regeneron Genetics Center. These 309 samples initially included 624,445 SNPs; I removed 110,726 SNPs during QC based on minor allele frequency, HWE p-value, or missingness, for a final total of 309 individuals and 565,439 SNPs uploaded to the Michigan Imputation Server⁹⁴. The remaining 547 were genotyped with several different Illumina arrays. This group initially contained 547 individuals and 513,719 SNPs. I used the intersect of SNPs present on these arrays, consisting of 133,322 SNPs. A final total of 547 individuals and 133,322 SNPs in dataset 2 were used for imputation.

Amish Study of Major Affective Disorders (ASMAD)

The ASMAD cohort was ascertained by Janice Egeland (University of Miami) and her colleagues between the 1970s and 1990s⁴⁰⁻⁴⁴. Study participants were identified using both a community-wide epidemiological study and searches of Amish patient records from local psychiatric hospitals. Both affected individuals and their relatives were enrolled. A diagnosis for each study participant was made on the basis of both clinic/hospital records and interviews using the Schedule for Affective Disorders and Schizophrenia-Lifetime version⁹⁵. The final diagnosis was given by a psychiatric panel that was blinded to patient identify, pedigree, and previous diagnoses and treatments, following the clinical criteria outlined in the Research Diagnostic Criteria or the DSM-IV⁹⁶. Whole genome genotyping of 394 individuals from the ASMAD cohort was led by Maja Bucan (University of Pennsylvania) at the Center for Applied Genomics (Children's Hospital of Pennsylvania, Philadelphia, PA) using Illumina Omni 2.5 M SNP arrays. Full details on the genotyping methods and quality control have been described previously^{23,84}.

Amish Cohort from the Trans-Omics for Precision Medicine program (Amish TOPMed)

The NHLBI Trans-Omics for Precision Medicine (TOPMed) study on the Genetics of Cardiometabolic Health in the Amish (phs000956; Braxton Mitchell, PI; University of Maryland School of Medicine) includes individuals collected as part of several studies focused on wellness and cardiometabolic health in the Lancaster OOA population^{30,31,97}. Although individuals in this cohort did not undergo a diagnostic interview for mental illness, most of the underlying studies included questions about current and past medication use. ARC staff provided information about the use of psychotropic medication to the ACP for recruitment into that study, and individuals

who enrolled in ACP were not included in the current analysis. These factors, along with the general depletion of severe mental illness cases from adult wellness cohorts, suggest that the TOPMed cohort has a relatively low incidence of severe mental illness. I used variant calls derived from joint calling of whole genome sequences (~30x average coverage) within the Freeze 5b TOPMed cohort (Phases 1 and 2, n = 65,000). Variant calls from 1028 OOA individuals were available as part of this dataset⁹⁷ and were used in the study as a reference.

Preliminary data processing, quality control, and imputation

I used imputed genome sequences from the ASMAD cohort^{23,84} and applied an identical pipeline for data processing, quality control, and imputation in the ACP and AMBiGen samples. Before imputation, I removed SNPs missing from more than 2% of individuals, as well as those with a minor allele frequency less than 0.2% and HWE p-value less than 1×10^{-6} using the `--geno`, `--maf`, and `--HWE` commands in PLINK v1.9^{98,99}. Individuals missing more than 5% of SNPs or heterozygosity greater than 3 standard deviations from the mean were also removed (`--missing` and `--het` commands, respectively). I then generated frequency files in PLINK v1.9 with the `--freq` command and checked the frequencies against the Haplotype Reference Consortium and 1000 Genomes using perl commands¹⁰⁰ provided by the Wellcome Sanger Institute, as well as separating the cleaned files by chromosome.

I next imputed whole genomes on the Michigan Imputation Server⁹⁴ using the TOPMed Freeze5 reference panel, which includes WGS from 1,028 OOA subjects, who were also used as population controls. I used the GRC38/hg38 build with a European population, no rsq filtering, and the Eagle v2.4 phasing, using the quality control and imputation mode (which applies the

Michigan Imputation Server's QC checks to the data before imputation). I removed all non-polymorphic sites from both the imputed and directly sequenced genomes, then renamed all remaining sites by chromosome, position, reference allele, and alternate allele using bcftools annotate. Finally, I merged the polymorphic-subsetted datasets using PLINK v1.9^{98,99} with the `–merge-list` command. The imputed ACP dataset contained 428 individuals and 230,053,813 SNPs. The imputed AMBiGen dataset included 309 individuals genotyped on the Global Screening Array, for whom I was able to impute 230,030,597 SNPs, and 547 individuals genotyped with other Illumina arrays for whom I was able to impute 229,903,075 SNPs. The imputed ASMAD dataset contained 387 samples and 230,332,237 SNPs. I used only SNPs with a very high imputation r^2 (≥ 0.6).

Assessing population stratification and sample overlap

I calculated principal components for the genomes using the `–pca` command in PLINK v1.9^{98,99}, after removing SNPs missing from more than 5% of the entire sample and with a minor allele frequency less than 1% (`--geno` and `–maf`). This analysis was performed using the imputed genomes for the ACP, AMBiGen, and ASMAD cohorts and the WGS from the TOPMed cohort for the principal components analysis as there were only 598 polymorphic SNPs in common between the four genotyping panels. I visualized the population structure using a scatterplot of the first two principal components and assigned cluster identity. For the OOA-specific GWAS, I removed all individuals that did not belong to the Lancaster OOA population and recalculated principal components specific to Lancaster OOA individuals for use as covariates in the GWA analysis.

I calculated IBD statistics on all samples with the PLINK v1.9 –genome command. I used the proportion of IBD values to identify individuals who were enrolled in more than one cohort. Samples with a proportion value > 0.8 were assumed to be from the same individuals, and duplicate samples were removed. ACP samples were given top preference, followed by AMBiGen, ASMAD, and TOPMed. The order of preference was determined by depth of phenotyping, as well as by time since ascertainment. These analyses resulted in the removal of 127 OOA samples: 26 individuals in ASMAD, 13 individuals from AMBiGen, and 93 individuals from TOPMed, including seven individuals that were part of three cohorts.

Assessing imputation accuracy

For 93 individuals whose genomes were sequenced in the TOPMed cohort I also had imputed genotypes from one of the three mood disorder cohorts. I used these duplicate samples to assess the accuracy of the imputation. The TOPMed WGS were included in the imputation panel, so results may not be representative of all OOA individuals. For each sample pair, I calculated the proportion of SNPs that were incorrectly imputed as well as those SNPs that were not imputed at all and coded as “missing” in the final alignment. I also divided these mismatches into categories based on MAF ($MAF < 0.01$, $0.01 < MAF < 0.05$, and $MAF > 0.05$). Analyses were done in R version 3.6.2¹⁰¹.

I also had access to WGS for 214 individuals enrolled in ACP. For each sample pair, I calculated the proportion of SNPs, divided by MAF ($MAF < 0.01$, $0.01 < MAF < 0.05$, and $MAF > 0.05$), for which our imputed sequences mismatched the WGS. I also verified the genotypes in the ACP cohort for five SNPs: rs192622352, rs569742752, rs117752843,

rs118010189, and rs7185072. Additionally, I was able to verify the genotypes of these five SNPs in 80 ASMAD individuals for whom the University of Pennsylvania had WGS.

Affection status model

The primary phenotype was diagnosis of a bipolar spectrum, including individuals with primary diagnoses of Bipolar Disorder Type I (n=86), Bipolar Disorder Type II (n=17), Bipolar Disorder Not Otherwise Specified (n=10), or Recurrent Major Depressive Disorder (n=73). I did not include Single Episode Major Depressive Disorder in this phenotype because the heritability of this disorder is much lower than the heritability of Recurrent Major Depressive Disorder¹⁰².

Individuals from the AMBiGen, ASMAD, and ACP cohorts (the cohorts ascertained on mood disorders) were coded as unaffected if they had no Axis I or Axis II diagnosis (n=449). All individuals from the TOPMed general population cohort were coded as unaffected (n=938). In the primary analysis, individuals with other Axis I or Axis II diagnoses were coded as unknown, and I considered these diagnoses in alternative affection status models, as follows. The cohort included ten individuals with a primary diagnosis of schizophrenia, all of whom were first- and second-degree relatives of mood disorder cases. While SCZ is not classically identified as a mood disorder, there is substantial genetic, clinical, and brain pathophysiological overlaps between SCZ and mood disorders especially bipolar disorders^{103,104}. Therefore, I studied an alternative affection status model in which individuals with SCZ were coded as affected. In addition, I tested models in which Single-Episode Major Depressive Disorder and Persistent Depressive Disorder were coded as affected. Individuals from these cohorts with a psychiatric diagnosis other than the diagnoses above or who did not undergo a psychiatric evaluation were

always coded as unknown (n=62). I also considered a model in which only individuals with recurrent MDD were coded as affected, as well as a model in which only individuals with a BD diagnosis were coded as affected.

GWA analyses

I tested associations of genotyped and imputed variants with mood disorders, as defined above, in a sample in of 1,672 Lancaster OOA individuals using a mixed-effect linear regression model implemented with EMMAX^{105,106}. Covariates included an empirical kinship matrix and twenty principal components, which account for family structure and more distant relatedness, respectively. Robust power is critical to the success of cohort studies such as this one, as underpowered GWAS are susceptible to false positives. Formal power analysis indicates that I have 32% power to detect alleles with an OR of 1.5 or higher, 52% power for an OR of 1.7 or higher, and 78% power for an OR of 2.0 or higher under an additive disease model for an allele present in at least 10% of individuals. These effect sizes are consistent with effect sizes for Amish-enriched missense and loss-of-function variants, and potentially for specific non-coding haplotypes, especially in the context of the homogeneous Amish environment. However, these power calculations, assuming unrelated individuals, may be conservative in the context of a population isolate. A simulation study³⁶ for an extended pedigree of 148 people suggested fewer than 35 copies of the minor allele need to be sampled to detect genetic associations of moderate-effect risk alleles with 80% power.

LD clumping

I identified LD-independent lead SNPs and sets of genetically correlated SNPs in the Old Order Amish using the `-clump` command in PLINK v1.9^{98,99}, setting the significance threshold for lead SNPs to 1×10^{-5} and the secondary significance threshold for clumped SNPs to 0.05. I also set the linkage disequilibrium (LD) threshold to 0.6 and the physical distance threshold to 1000kb. I also allowed for non-index SNPs to appear in multiple loci. After I generated the list of loci, I used PLINK v1.9 to identify SNPs and indels in strong LD with the list of genome-wide significant SNPs ($r^2 > 0.6$) within 1000kb in the Lancaster OOA population sample using the `-ld-window` command. The physical distance threshold was set to a larger value than is typical of GWAS in the broader population due to the longer haplotypes in this founder population.

Pseudoreplication

I performed a pseudo-replication analysis using a leave-one-out strategy to verify that the results are not dependent on samples from a single cohort. I removed one cohort at a time (ACP, AMBiGen, ASMAD, and TOPMed) from the sample and reran EMMA¹⁰⁵. I recalculated PCA coordinates for each pseudo-replication dataset and used the first 20 recalculated coordinates as covariates in the model. I also reran the analysis on the ACP and ASMAD cohorts independently using the first 20 recalculated PCs as covariates.

Polygenic risk score (PRS) analysis

I calculated polygenic risk scores for each OOA individual, using results from large-scale GWAS of BD^{13,14}, SCZ²⁵, and MDD¹¹ using PRSice-2¹⁰⁷. SNPs with a minor allele frequency

less than 0.05, HWE p-value less than 1×10^{-6} , or missing in more than 10% of individuals were removed from the dataset before analysis, as well as individuals missing more than 10% of SNPs. Any SNPs removed due to these filters were removed from the entire analysis, so every individual has the same number of SNPs used to calculate their score. I calculated PRS using the full Lancaster OOA dataset, as well as for each mood disorder cohort (ACP, AMBiGen, and ASMAD) dataset separately.

Annotation of loci and variants

I assessed overlap of risk loci identified in the OOA with loci identified in published large-scale neuropsychiatric GWAS. I used the BEDtools v2.27.1¹⁰⁸ intersect command to calculate the overlap between the risk-associated loci (defined as all SNPs in LD with the 41 lead SNPs) and risk-associated SNPs identified in previous GWAS of mood disorders and related neuropsychiatric traits: MDD¹², BD^{13,14}, SCZ²⁴, and educational attainment¹⁰⁹. I used the authors' definitions of risk loci for MDD, BD, and SCZ. For the educational attainment dataset, bounds of risk loci were not described in the original publication, so I set bounds 250kb upstream and downstream of the lead SNPs. I also tested whether the lead SNPs identified in the Lancaster OOA sample were in LD with risk-associated SNPs identified in previous neuropsychiatric GWAS using the PLINK's `-ld` command and recorded the r^2 value.

I further annotated proximal candidate genes at risk loci using gene sets related to autism spectrum disorders (ASD), BD, and SCZ. Within the bounds of each risk locus, I identified differentially expressed genes in the prefrontal cortex of ASD, BD, and SCZ cases vs. controls from PsychENCODE¹¹⁰. I also identified genes from exome sequencing studies, including genes

with a gene burden p-value $< 2.5 \times 10^{-6}$ from SCHEMA¹¹¹ (SCZ-associated genes), and ASD-associated genes from Satterstrom *et al.*¹¹² and SFARI Gene¹¹³.

MAGMA

Gene-based p-values were computed from GWAS summary statistics using MAGMA¹¹⁴. SNPs were annotated to ENSEMBL genes, including a 10 kb window up- and downstream of each gene's genomic coordinates. Gene p-values were computed using the lowest SNP p-value as the test statistic (snp-wise=top,1), and gene-gene correlations were computed using our imputed OOA genotype matrix. MAGMA gene set enrichment analysis was performed with default parameters. I studied 21 gene sets with prior evidence for association with neuropsychiatry^{115,116}. Briefly, these gene sets were derived from the following sources: genes identified through GWAS of MDD¹², BD¹³, SCZ¹¹⁷, and neuroticism¹¹⁸; genes identified through genetic association studies of rare variants, including exome and genome sequencing studies of SCZ¹¹¹, autism spectrum disorders (ASD)¹¹², or other developmental disorders¹¹⁹, as well as genes intolerant to loss-of-function mutations¹²⁰; genes that are differentially expressed in the prefrontal cortex of individuals with BD, SCZ, or ASD¹¹⁰; genes that have been identified as targets of the RNA binding proteins FMRP, RBFOX2, RBFOX1/3, and CELF4, of the chromatin remodeling genes CHD8, and of the microRNA miR-137^{116,25}; and genes localized to synapses from SynptomeDB¹²¹.

Gene interaction network analysis

Human protein-protein interactions were downloaded from the STRING database¹²² (<https://stringdb-static.org/download/protein.links.detailed.v11.5/9606.protein.links.detailed.v11.5.txt.gz>). I defined OOA risk genes as those with a MAGMA gene p-value < 0.01 . I used the same 21 neuropsychiatry-related gene sets as in MAGMA gene set enrichment analysis. To assess interactions between established gene sets and OOA risk genes, I counted the number of protein-protein interactions that directly link OOA risk genes to genes in each of the 21 established neuropsychiatry gene sets. I tested whether the number of interactions was greater than expected by chance by two approaches. First, I computed Fisher's exact tests. Second, I repeatedly permuted the edges of the network, holding each node's degree constant, and compared the number of OOA-known edges in observed vs. permuted data. Edge permutations confirmed that for comparisons in which I found a significant enrichment by Fisher's exact test I always observed fewer interactions in permuted data ($n=100$ permutations). Odds ratios from Fisher's exact test are reported in this thesis, as they provide a more precise measure of the likelihood.

To prioritize specific OOA risk genes, I ranked them by their centrality within a gene interaction network centered on known neuropsychiatry risk genes. I defined a set of 684 core neuropsychiatry genes with evidence from at least three independent approaches from our 21 gene sets, as follows: Genes implicated by studies of rare variants, defined as the union of genes associated with disease in exome sequencing studies of SCZ and ASD. Genes implicated by gene expression profiling were defined as the union of genes that were significantly down- or up-regulated in prefrontal cortex of BD, SCZ, or ASD cases using data from psychENCODE. Genes

implicated by gene network analyses were defined as the union of genes that are targets of CELF4, FMRP, RBFOX1/3, RBFOX2, CHD8, and miR-137. Synaptic genes were defined from the SynptomeDB database. I excluded genes derived from GWAS, as these genes are potentially non-independent from association signals in our OOA dataset. I extracted all protein-protein interactions from the STRING database for which at least one node was one of these 684 genes. In practice, the large number of interactions in the STRING database means that nearly all genes are represented in this network, but only the subset of their interactions that involve neuropsychiatry-related genes. I used the `eigen_centrality()` function from the `igraph` R package¹²³ to calculate the centrality of each node, including OOA risk genes that have not previously been implicated in neuropsychiatric disorders. I computed ranks for the OOA risk genes, separately, based on eigen-centrality, as well as based on their MAGMA p-values. The final ranking is the median rank from these two metrics. I tested for functional enrichments within the top 250 genes from this analysis using DAVID¹²⁴.

Effects of risk variants on quantitative behavioral and neurocognitive phenotypes

I tested for the associations of genome-wide significant and suggestive SNPs identified by the GWA analysis with quantitative behavioral and cognitive traits in 314 OOA participants in the ACP study, including 84 cases. I studied self-reports of current depression symptoms from the Beck Depression Index⁸⁹, lifetime depression symptoms from the Maryland Trait and State Depression scale⁹⁰, and lifetime history of bipolarity from the Bipolar Spectrum Diagnostic Scale⁹¹. I also used scores from several cognitive tasks, including digit sequencing (verbal working memory), digit symbol coding (processing speed, visuospatial memory), spatial span

(visuospatial working memory), and the Wechsler Abbreviated Scale of Intelligence⁸⁸ (WASI) matrix reasoning and vocabulary subtests (IQ and cognitive ability); these tests were administered by trained clinical staff with the Maryland Psychiatric Research Center. I assessed normality, as well as associations of each trait with age and sex. The scores from Beck Depression Index, Bipolar Spectrum Diagnostic Scale, and spatial span were transformed using a square root transformation to improve normality. The other five traits displayed non-linear associations with age. For those traits, I applied a loess regression model (using the loess function in R v. 3.6.2¹⁰¹, span = 0.5) and performed genetic association tests on residuals. Covariates in the EMMAX model for these three traits included sex, age, and an empirically constructed kinship matrix. The heritability of each trait was calculated using SOLAR-Eclipse¹²⁵. I constructed mixed-effect linear regression models for each genotype-phenotype pair using EMMAX¹⁰⁵.

Data availability. Genotypic and phenotypic data from the Amish TOPMed study and from AMBiGen are available through the National Institute of Health Database of Genotypes and Phenotypes (phs000956.v1.p1, phs000899.v1.p1). Genotypic and phenotypic data from ACP are available through the NIMH Data Archive (Study #2902). Genotypic and phenotypic data from ASMAD are available through the Coriell Institute for Medical Research.

Chapter 4: Results

An expanded genetic cohort for the investigation of mood disorders in the Old Order Amish founder population

Previous studies of mood disorders in the Old Order Amish have primarily used samples from the Amish Study of Major Affective Disorders (ASMAD), which were collected by Egeland and colleagues in the 1970s to the 1990s⁴⁰⁻⁴⁴. Initially promising results in this cohort suggested linkage to the proximal arm of chromosome 11⁴⁰. More recently, genome sequencing and whole-genome genotyping in this cohort suggested a polygenic genetic architecture, consistent with the broader population^{22,23,84}. However, all these previous studies were limited by their very small sample size.

To address this, collaborators generated whole-genome genotyping data from two ongoing studies of Old Order Amish and Plain Mennonite families with mood disorders, the Amish Connectome Project (ACP; n=428) and Amish and Mennonite Bipolar Genetics study (AMBiGen; n=856). ACP enrolled families primarily in the Lancaster, PA area, including both OOA and Mennonites. AMBiGen enrolled OOA and Mennonite families from geographically diverse regions, including Pennsylvania, Ohio, Indiana, and central and South America. I then integrated these data with existing whole-genome genotyping data from ASMAD (n=387) and whole-genome sequencing from the Trans-Omics of Precision Medicine consortium^{23,97} (TOPMed; n=1,028), both of which recruited primarily OOA in the Lancaster area.

A uniform set of genotype calls were obtained for the four cohorts by imputing the genomes of individuals in the cohorts with whole-genome genotyping data using the TOPMed Freeze 5b reference panel, which includes the Amish TOPMed cohort among ~65,000 multi-

ancestry individuals¹²⁶. This strategy enabled imputation of 6.6 million polymorphic single-nucleotide polymorphisms (SNPs). Of these, 1,048,076 had an OOA-specific minor allele frequency (MAF) less than 0.01 (16%). An additional 1,744,218 SNPs (26%) had a MAF between 0.01 and 0.05.

To assess imputation quality, I took advantage of sample overlap across the cohorts. Pairwise comparisons revealed 152 pairs of samples with nearly identical genomes (average identity by descent >98%) presumably corresponding to individuals who enrolled in more than one study, including 127 pairs among the Lancaster OOA. The strongest overlap was among the ACP, ASMAD, and TOPMed studies, which recruited participants primarily in Lancaster County. I focused on 93 individuals for whom I had both a directly sequenced genome from TOPMed and an imputed genome from ACP (n=53), ASMAD (n=36), or AMBiGen (n=4). Remarkably, I estimate an imputation accuracy of 99.99% +/- 0.06% among all the duplicate individuals (mean +/- standard deviation). This included >99.9% accuracy for each of the genotyping platforms used in ACP, ASMAD, and AMBiGen, compared to TOPMed WGS (Table 4.1). Imputation accuracy remained high for uncommon SNPs with minor allele frequencies less than 5% (accuracy > 99.99%, \pm 0.0014). The percentage of missing genotypes amongst these 6.6 million SNPs varied by genotyping technology; in the ASMAD cohort, 14.6% of the 6.6 million polymorphic SNPs used for the final analysis were removed by the Michigan Imputation Server filters as low accuracy calls (and thus were not imputed), whereas the percentage of unimputed SNPs was less than 1% for the remaining cohorts (Table 4.1). Since these same individuals are in the imputation panel, these results may

Table 4.1. Accuracy of imputation for 93 individuals enrolled in both a mood disorder cohort and Amish TOPMed. A. Mismatches between imputed sequence and WGS, broken down by cohort and MAF. B) Unimputed SNPs broken down by cohort and MAF.

A)

Cohort comparison	n	Mean accuracy (SD)	Mean # mismatch	Mean # mismatch, MAF<0.01	Mean % mismatch, MAF<0.01	Mean # mismatch, 0.01<MAF<0.05	Mean % mismatch, 0.01<MAF<0.05	Mean # mismatch, MAF>0.05	Mean % mismatch, MAF>0.05
ACP: TOPMed	53	0.9999 (0.00005)	142.91	8.04	0.0008 (0.006)	63.42	0.004 (0.002)	71.45	0.002 (0.0008)
AMBiGen: TOPMed (Illumina)	1	0.9999 (NA)	204	6	0.0006 (NA)	96	0.006 (NA)	102	0.003 (NA)
AMBiGen: TOPMed (Agilent)	3	0.9997 (0.00003)	402.67	25.33	0.002 (0.003)	176.00	0.01 (0.001)	25.33	0.005 (0.0008)
ASMAD: TOPMed	36	0.9999 (0.00004)	137.06	6.94	0.0007 (0.0008)	61.58	0.004 (0.001)	68.53	0.002 (0.0008)
ACP: ACP WGS	214	0.9994 (0.0046)	3656.52	152.86	0.0146 (0.0969)	1675.26	0.0961 (0.8262)	0.0961 (0.8262)	0.047 (0.400)

B)

Cohort comparison	% genome not imputed	% SNPs not imputed, MAF<0.01	% SNPs not imputed, 0.01<MAF<0.05	% SNPs not imputed, MAF>0.05
ACP: TOPMed	0.6	3.3	0.3	1.0
AMBiGen: TOPMed (Illumina)	0	0	0	0
AMBiGen: TOPMed (Agilent)	0	0	0	0
ASMAD: TOPMed	14.6	71.1	13.0	25.2
ACP: ACP WGS	0.3	0.87	0.1	0.3

not be representative of all OOA in this cohort. I also examined the accuracy of the imputation by comparing imputed genomes to directly sequenced genomes for 214 ACP individuals (only one of these individuals were included in the imputation panel, so this is an independent comparison). For these comparisons, I estimated an imputation accuracy of 99.85 ± 0.01 for SNPs with minor allele frequencies less than 1%, 99.04 ± 0.08 for those with minor allele frequencies between 1% and 5%, and 99.53 ± 0.04 for those SNPs with a minor allele frequencies greater than 5%. I also found perfect correlation between the imputed genotypes of five SNPs of interest (discussed later) and the WGS genotypes in the 214 ACP samples, as well as for 80 ASMAD individuals for whom WGS had previously been generated. All individuals predicted to be carriers for each SNP in the imputed sequences were confirmed as carriers in the WGS (positive predictive value=1), and all predicted non-carriers were also confirmed as non-carriers in the WGS (negative predictive value=1). The combination of WGS, whole-genome genotyping, and population-specific imputation enabled highly accurate imputation, including for millions of uncommon and rare SNPs.

I applied principal component analysis to examine population structure in our combined sample. The first principal component, separating the Lancaster OOA from all other subjects, explains 48.4% of the variance. Clustering the cohort using principal components thus enabled me to distinguish two populations (Fig. 4.1A). One group is primarily OOA individuals from Lancaster, PA and surrounding areas (n=1672). A second ancestry group encompasses PM individuals, Russian Mennonite individuals, as well as OOA individuals from the American Midwest (n=904). These ancestry groups align well with migration history²³⁻³⁵. The Lancaster OOA are the descendants of the first wave of OOA immigrants, ~300 individuals who arrived in

PA in 1727-1770, while the Midwestern OOA are the descendants of ~3000 OOA immigrants who arrived in North America during the second wave between 1815-1860³³. I did not see geographic division among the PM, who are the descendants of ~8000 PM immigrants who arrived in North America between 1683 to 1883, though increased sampling of this diverse population may yield additional population structure³⁴. The small cluster of Mennonites from Brazil are the descendants of an early group of Anabaptists from the Netherlands that initially emigrated to Prussia and Russia, later emigrating to the Americas starting in the late 1800s³⁵. I then recalculated principal components using only the 1672 individuals in the Lancaster OOA cluster and verified genotyping or sequencing technology was not causing stratification (Fig 4.1B).

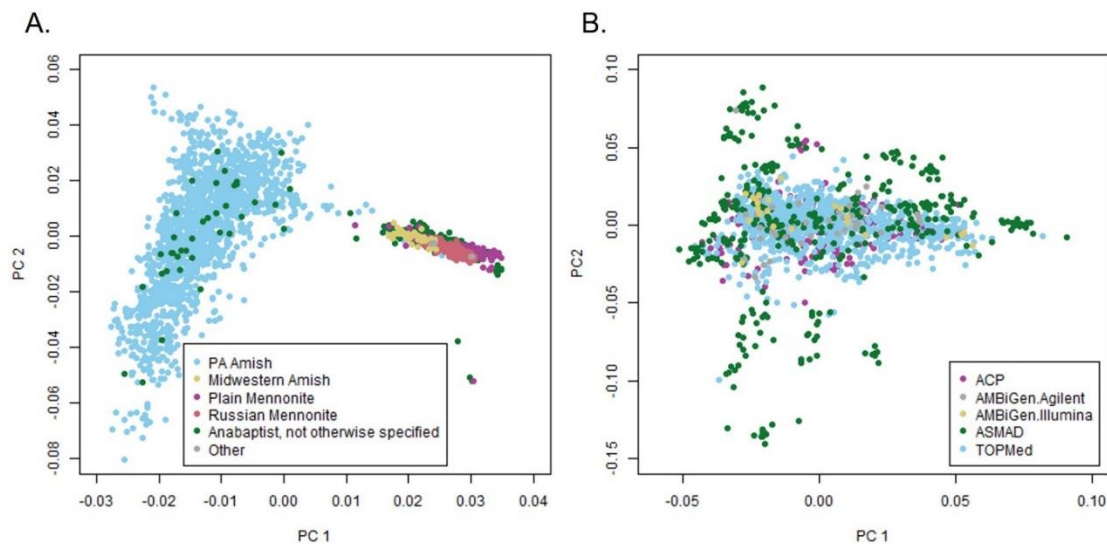


Figure 4.1. Population structure of Amish and Mennonite populations. A. Plot of the first two PCs (calculated using all samples) with samples colored by religious group. Amish samples are further subdivided by location (PA or non-PA B. Plot of the first two PCs (calculated using only the Lancaster OOA individuals) with samples colored by sequencing or genotyping technology.

Identification of genome-wide significant risk loci for mood disorders in the Old Order Amish

Next, I studied genetic associations with mood disorders. For these analyses, I used only the data from Lancaster OOA, as the other sub-populations were genetically distinct and had insufficient sample size to be analyzed independently. This cohort includes 1,672 individuals, of whom 196 were considered affected (Table 4.2). These include 86 individuals with a primary diagnosis of bipolar disorder type 1, 73 with recurrent major depression, 17 with bipolar disorder type 2, 10 with bipolar disorder not otherwise specified, and 10 individuals with schizophrenia or schizoaffective disorder.

Table 4.2. Sampling description of the Lancaster OOA cluster. Affected=broad mood disorder diagnosis; BDI=Bipolar I, BDII=Bipolar II, BD:NOS= Bipolar not otherwise specified (category dropped in DSM-V, thus not used in ACP or AMBiGen cohorts), MDDR=Major depressive disorder-recurring, SCZ= Schizoaffective or Schizophrenia. Unaffected=no mood disorder diagnosis on Axis I or II. Other diagnosis=diagnosed with something other than mood disorder (eg, anxiety disorder, single episode major depression). Missing=diagnosis not ascertained.

Cohort	Total	Affected					Unaffected	Other diagnosis	Missing
		BDI	BDII	BD:NOS	MDDR	SCZ			
Hong, ACP	314	14	3	-	58	9	214	14	2
McMahon, AMBiGen	62	20	4	-	3	1	8	1	25
Bucan, ASMAD	358	52	10	10	12	0	227	47	0
TOPMed Amish	938	-	-	-	-	-	938*	-	-
TOTAL	1672	86	17	10	73	10	1387	62	27

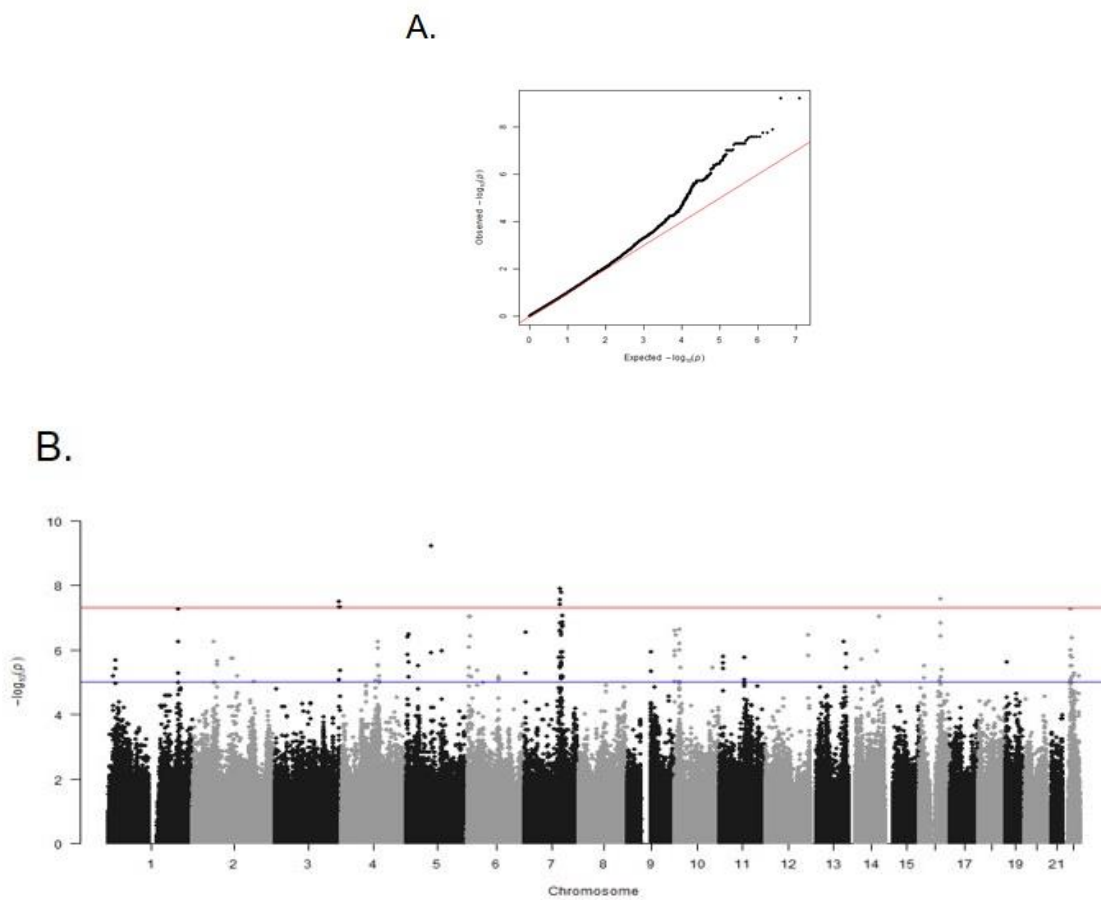
*healthy population control, presumed unaffected

I conducted a genome-wide association study to test associations of 6.6 million SNPs with mood disorders in the OOA using a linear mixed model implemented with EMMAX¹⁰⁵. I identified 25 risk-associated SNPs at a conventional genome-wide significance cutoff, p -value $< 5 \times 10^{-8}$ (Fig. 4.2B). The quantile-quantile plot of observed vs. expected p -values revealed no genomic inflation ($\lambda_{GC} = 0.8$; Fig. 4.2A). Clumping of these SNPs and genetically correlated SNPs supported four linkage disequilibrium (LD)-independent genome-wide significant risk loci, located at cytobands 3q28/29, 5q13, 7q22, and 16q21 (Table 4.3). The loci at 7q22 and 5q13 have been previously implicated in GWAS of psychiatric disorders^{11-15,24}, while the other 2 loci appear to be novel psychiatric risk loci. In addition to the four genome-wide significant risk loci, the Q-Q plot suggests an excess of alleles with low p -values, consistent with polygenicity, including 39 LD-independent loci below a suggestive threshold, $P < 5 \times 10^{-4}$ (Table A3). Notably, I found suggestive associations at a locus centered approximately two megabases from the original 11p15 locus identified by Egeland *et al*^{40,42} (lead SNP, rs529802370, $p=1.7e-06$). Given the sparsity of the markers used in the original linkage scan, this association signal is potentially consistent with the original result.

Risk-associated SNPs identified through GWAS in the broader population are generally common with very small effect sizes. In contrast, the lead SNPs at three of the four risk loci identified in the OOA, although uncommon, had moderate to large effects on risk. For the lead SNP at 3q28/29, rs192622352, 7 of 9 carriers (all heterozygous) were affected, compared to 190 of 1575 non-carriers (Fisher's exact test, not adjusted for relatedness: risk ratio [RR] = 6.4, 95% confidence interval [CI]: 4.4-9.4 EMMAX: $\beta = 0.66$, $P = 3.4 \times 10^{-8}$). For the lead SNP at 5q13, rs569742752, 8 of 10 carriers were affected (RR = 6.7, 95% CI: 4.8-9.3; $\beta = 0.74$, $P = 6.4 \times 10^{-8}$).

¹⁰). For the lead SNP at 7q22, rs117752843, 27 of 61 carriers were affected (RR = 4.0, 95% CI: 2.9-5.4; $\beta = 0.28$, $P = 1.3 \times 10^{-8}$). The exception is 16q21, for which the lead SNP, rs7185072, has a minor allele frequency of 31% in the sample. Even at this locus, the observed effect is substantial by GWAS standards (RR_{het} = 2.1, 85% CI: 1.6-2.8; $\beta = 0.08$, $P = 2.7 \times 10^{-8}$).

Figure 4.2. Discovery of risk loci for mood disorders in the Old Order Amish. A. Q-Q plot. B. Manhattan plot.



Two additional features of these risk loci are notable in comparison to risk loci identified in the broader population. The 3q28/29, 5q13, and 7q22 risk loci are all very broad, spanning 2.7 Mb, 6.1 Mb, and 10.5 Mb, respectively. (Again, the 16q21 locus is an exception, spanning just 9.2 kb). Yet each locus includes relatively few genetically correlated SNPs, with fewer than ten SNPs having an $r^2 > 0.6$ with the lead SNPs at 3q28/29, 5q13, and 16q21, and just 60 genetically correlated SNPs at the very broad 7q22 risk locus. Each locus likely comprises some long haplotypes that have not been broken up by recombination, as expected in this founder population³⁸.

Pseudoreplication and overlap with known risk genes and risk loci

I conducted several analyses to evaluate the robustness of these findings. First, I performed a leave-one-cohort-out pseudo-replication analysis within the OOA sample. All four lead SNPs remained nominally significant ($p < 0.05$) in all leave-one-out pseudo-replication analyses, indicating that the signal comes from more than one of the underlying cohorts (Table 4.3). This was also true of the suggestive signal at 11p15. Second, I considered an alternative affection status model in which I included Persistent Depressive Disorder and Single Episode MDD. All four lead SNPs remained either significant ($p < 5 \times 10^{-8}$) or suggestive ($p < 5 \times 10^{-6}$) when I broadened the affection status model in this way (Table 4.3). I saw similar results when I reran the GWAS for ACP and ASMAD separately, with the exception of the 3q28/29 locus, for which there is only a single carrier in ASMAD. Third, I verified the accuracy of the imputation of these four lead SNPs in 214 ACP and 80 ASMAD individuals for whom WGS has been generated; the imputed genotypes were 100% correct.

Table 4.3. A) MAF, OOA enrichment, and GWAS statistics for the lead SNPs identified in my analysis. B) Leave-one-out results for the four lead SNPs. C) MAF broken down by sequencing technology and case status for the four lead SNPs. D) GWAS results by phenotype definition

A)

SNP	Amish MAF	European MAF	Amish enrichment	Amish.GWAS BETA	Amish.GWAS P
chr3:191857829:C:A	0.0041866	0.02869	0.14592541	-0.65894	3.35E-08
chr5:76339511:G:A	0.00358852	0.001379	2.602262509	-0.73774	6.35E-10
chr7:103511937:C:T	0.0194378	0.005744	3.384018106	-0.27945	1.30E-08
chr16:62294564:T:C	0.311005	0.3712	0.837836746	-0.080107	2.70E-08

B)

SNP	No.ACP P	No.AMBiGen P	No.ASMAD P	No.TOPMed P	ACP.only P	ASMAD.only P
chr3:191857829:C:A	0.021076	9.28E-10	6.81E-08	0.0006	0.001	0.09
chr5:76339511:G:A	1.40E-09	3.17E-09	1.29E-08	0.004	0.014	0.031
chr7:103511937:C:T	7.15E-05	1.29E-08	5.14E-08	5.79E-07	0.0001	0.004
chr16:62294564:T:C	6.72E-06	3.91E-08	0.00016624	1.16E-06	0.009	6.30E-05

C)

SNP	MAF ACP (controls)	MAF ACP (cases)	MAF AMBiGen Agilent (controls)	MAF AMBiGen Agilent (cases)	MAF AMBiGen Illumina (cases)	MAF ASMAD (controls)	MAF ASMAD (cases)	MAF TOPMed (controls)
chr3:191857829:C:A	0.002	0.040	0.000	0.000	0.000	0.000	0.008	0.001
chr5:76339511:G:A	0.000	0.007	0.000	0.000	0.000	0.004	0.023	0.000
chr7:103511937:C:T	0.005	0.060	0.063	0.079	0.000	0.025	0.069	0.010
chr16:62294564:T:C	0.318	0.433	0.250	0.421	0.500	0.319	0.477	0.279

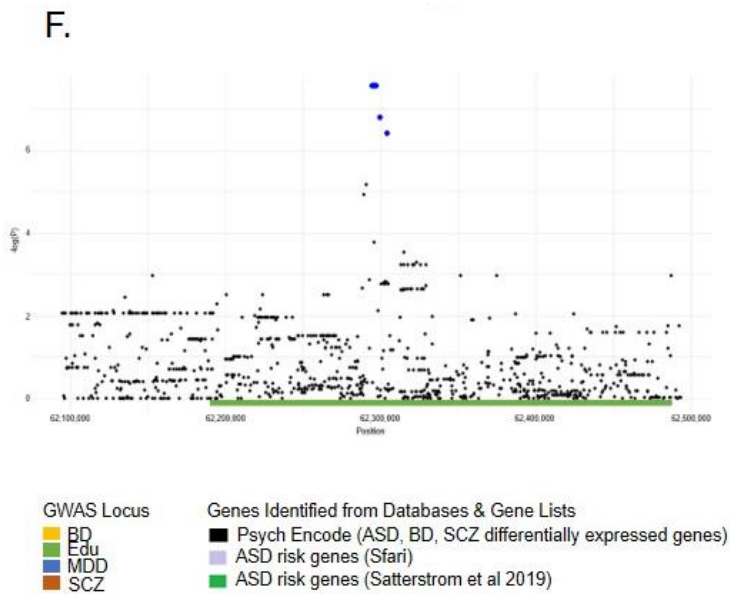
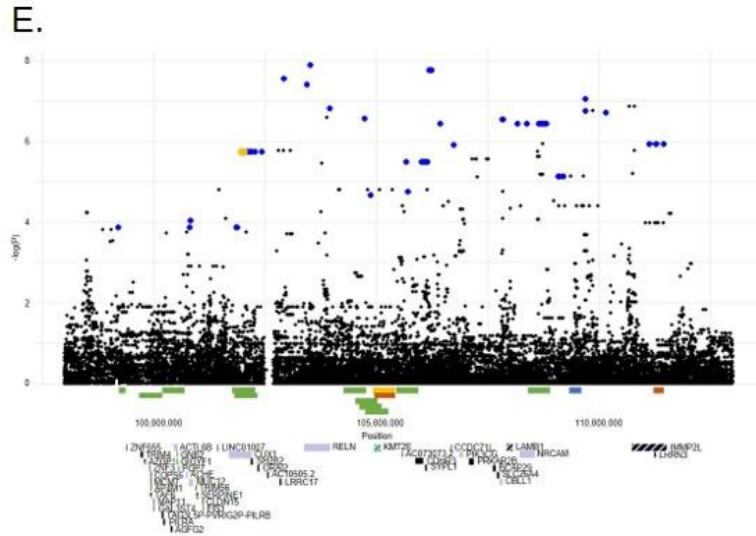
D)

SNP	RecurrentMDD P	BP P	No.SCZ P	Single.Episode.MDD P	All Depression Diagnoses P
chr3:191857829:C:A	6.773e-05	6.8237e-05	2.3284e-09	4.2677e-08	4.439e-07
chr5:76339511:G:A	2.911e-07	2.928e-07	1.2193e-06	4.729e-10	3.9813e-10
chr7:103511937:C:T	2.9391e-05	2.9435e-05	3.3571e-07	1.4328e-07	3.923e-07
chr16:62294564:T:C	6.0726e-06	6.1861e-06	2.7789e-07	2.089e-07	3.6031e-06

Next, I considered whether the risk-associated variants and loci identified in the OOA are associated with risk for mood disorders and related traits in the broader population (Fig. 4.3A-D). None of the lead SNPs or genetically correlated SNPs identified in the OOA had significant p-values in previous GWAS of mood disorders, though three of the four risk-associated SNPs and haplotypes in the OOA are uncommon in the broader population (Table 4.3). I do find evidence for overlap with known risk loci. The 5q13 risk locus overlaps genome-wide significant risk loci for BD^{13,14} and educational attainment¹⁰⁹. The 16q21 risk locus overlaps a previously identified risk locus for educational attainment¹⁰⁹. The 7q22 risk locus overlaps previously reported risk loci for MDD¹¹, BD^{13,14}, SCZ²⁶, and educational attainment¹⁰⁹. Though the 3q28/29 risk locus has not previously been identified in large-scale GWAS of BD^{13,14}, MDD¹¹, SCZ²⁶, or educational attainment¹⁰⁹, 3q29 microdeletions have been associated with increased risk for BD and SCZ^{127,125}.

I further annotated the risk loci based on their overlap with genes identified in exome sequencing studies of SCZ and autism spectrum disorder (ASD)¹¹¹⁻¹¹³, as well as genes dysregulated in the prefrontal cortex of SCZ, BD, or ASD cases vs. controls¹¹⁰. This analysis revealed numerous positional candidates, including *FGF12* at 3q28/29 and *HOMER1* and *SV2C* at 5q13. A total of 44 genes within the bounds of the 7q22 locus were highlighted in this analysis, including well-established ASD risk genes such as *RELN*, *KMT2E*, and *CUX1*.

Figure 4.3 continued. C-D. Region plots for genome-wide significant risk loci. Blue dots indicate the lead SNP and SNPs in LD with the lead SNP for each locus. Overlapping loci from previously-published GWAS are plotted directly underneath the x-axis. Genes identified in other large consortia and research (PsychENCODE, Satterstrom *et al.* 2019, and Sfari) are plotted below each region plot. C) Locus 7q22. The orange dot indicates the deleterious missense coding SNP in the CUX1/CASP gene. D) Locus 16q21



Gene networks associated with mood disorders in the OOA

I applied a network analysis approach to identify shared biological characteristics of the genes located at risk loci. As a starting point, I computed gene-based p-values from the GWAS summary statistics with MAGMA¹¹⁴. This analysis revealed three exome-wide significant genes: *ATP13A5* at 3q29 ($P = 1.6e-7$), *SV2C* at 5q13 ($P = 2.8e-7$), and *MB21D2* at 3q28 ($P = 6.5e-7$), along with 820 OOA risk genes reaching a nominal level of significance, $P < 0.01$ (Table A5). *ATP13A5* encodes ATPase 13A5, which is highly expressed in brain pericytes and is involved in the transport of diverse cargo across cellular membranes¹²⁹. *SV2C* encodes Synaptic Vesicle Glycoprotein 2C, which is expressed specifically on the vesicles of dopaminergic neurons and contributes to dopamine release¹³⁰. *MB21D2* encodes Mab-21 Domain Containing 2.

Next, I asked whether these OOA risk genes overlap genes and gene networks previously implicated in neuropsychiatric disorders, using 21 gene sets derived from psychiatric GWAS, exome sequencing, post-mortem prefrontal cortex gene expression, and analyses of disease-associated gene networks. I tested both for direct overlap of OOA risk genes within these gene sets, as well as guilt-by-association overlap, in which OOA risk genes interact with established neuropsychiatry genes via protein-protein interactions. OOA risk loci contained numerous established neuropsychiatry genes, including 44 genes within the bounds of the 7q22 locus, such as the well-established autism spectrum disorder risk genes *RELN*, *KMT2E*, and *CUX1*¹³¹. However, gene set enrichment analysis with MAGMA indicated that, overall, the OOA risk genes had only modest direct overlap with established neuropsychiatry genes.

By contrast, I found strong evidence that OOA risk genes interact with established neuropsychiatry genes (Table 4.4, Fig. 4.4). Specifically, I examined protein-protein interactions

in the STRING database that link OOA risk genes to genes from each of the established neuropsychiatry gene sets, and I tested for over-representation of these interactions using both hypergeometric distributions and permutation tests.

Table 4.4. OOA risk genes (820 genes; MAGMA, $P < 0.01$) have an elevated rate of STRING protein-protein interactions with gene sets derived from GWAS, rare variant studies, differential gene expression, and network analyses of psychiatric disorders.

Gene set	Type	nGene	nEdge	OR	P
CELF4_binding_targets_ensg	Networks	2633	174778	1.064103	1.02E-116
targets_of_RBFOX2_WGSPD	Networks	2851	212276	1.053999	1.07E-97
targets_of_RBFOX1_3_WGSPD	Networks	3185	235084	1.049038	1.84E-87
targets_of_FMRP_WGSPD	Networks	1155	99834	1.059803	2.16E-64
targets_of_mir137_WGSPD	Networks	2844	174238	1.031929	1.46E-31
gnomAD_pLI.0.9	Rare Variants	2877	243258	1.021907	5.13E-20
PsychENCODE_down.regulated_Prefrontal Cortex_schizophrenia	Gene Expression	1840	101476	1.027954	2.41E-16
PsychENCODE_down.regulated_Prefrontal Cortex_bipolar	Gene Expression	311	17684	1.058622	1.75E-13
102_autism_risk_Satterstrom	Rare Variants	100	11242	1.07178	7.77E-13
PsychENCODE_up.regulated_Prefrontal Cortex_schizophrenia	Gene Expression	2107	128284	1.019616	1.13E-10
Depression_risk_MAGMA_p2.77e6	GWAS	244	15354	1.049876	3.23E-09
PsychENCODE_down.regulated_Prefrontal Cortex_autism	Gene Expression	724	46778	1.02728	1.54E-08
PsychENCODE_up.regulated_Prefrontal Cortex_autism	Gene Expression	663	44386	1.020629	2.06E-05
Neuroticism_risk_MAGMA_p2.77e6	GWAS	236	16586	1.031308	6.41E-05
SCZ_risk_SCHEMA_p.0.05	Rare Variants	1050	69582	1.013499	0.0004325
All_DDD_Genes	Rare Variants	2004	157388	1.007324	0.0046836
PsychENCODE_up.regulated_PFC_bipolar	Gene Expression	587	33624	1.006917	0.1126221
targets_of_CHD8_in_human_brain_WGSPD	Networks	2368	153708	1.000993	0.3634919
SynptomeDB	Networks	1743	164134	0.998966	0.6466566
BP_risk_MAGMA_p2.75e6	GWAS	138	8336	0.983685	0.9294734
SCZ_GWAS_all_risk_PsychENCODE	GWAS	759	45688	0.991991	0.9502678

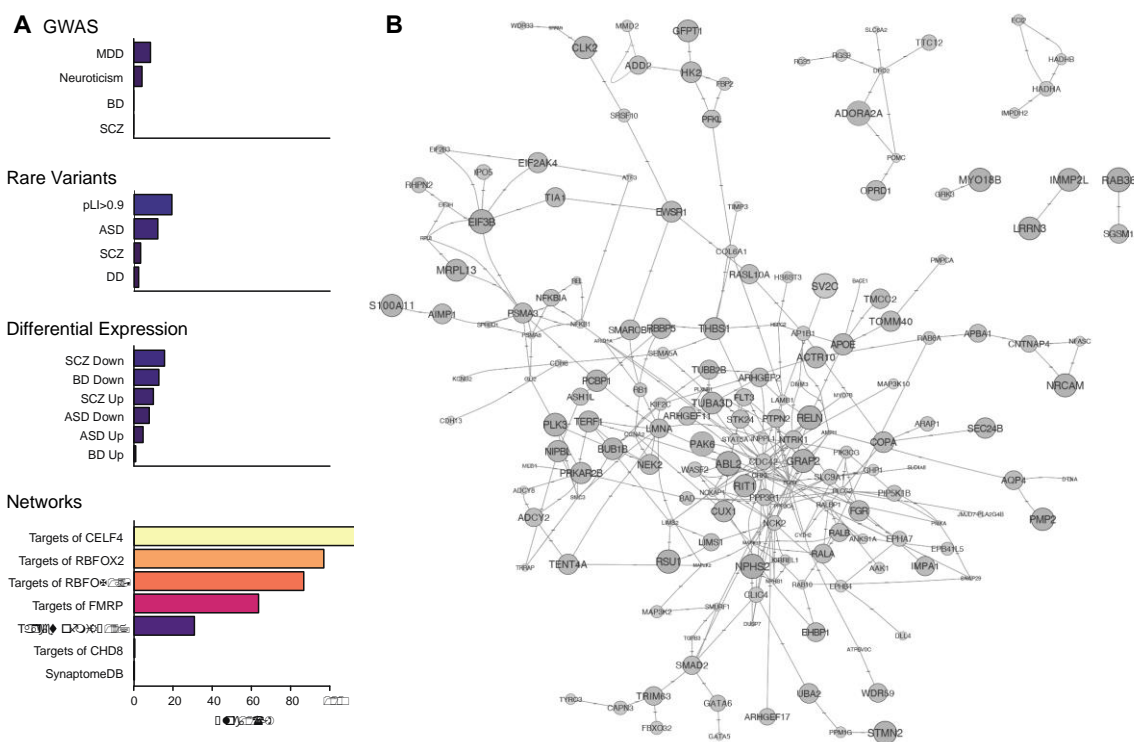


Figure 4.4. Risk loci for mood disorders in the OOA harbor novel risk genes within a polygenic gene network that is shared with neuropsychiatry genes discovered by independent approaches. A) OOA risk genes (820 genes; MAGMA, $P < 0.01$) have an elevated rate of STRING protein-protein interactions with gene sets derived from GWAS, rare variant studies, differential gene expression, and network analyses of psychiatric disorders. B) STRING protein-protein interactions among the top 250 genes at OOA risk loci prioritized by their centrality in a gene network centered on known neuropsychiatry-related genes and strength of statistical association with mood disorders. Node size and color correspond to MAGMA p-value.

The neuropsychiatry gene sets most strongly over-represented for interactions with OOA risk genes included target genes of the neuronal RNA-binding protein CELF4 (174,778 interactions, odds ratio (OR) = 1.06, $P = 1.0e-116$), genes down-regulated in prefrontal cortex from bipolar disorder cases (17,684 interactions, OR = 1.06, $P = 1.8e-13$), and autism spectrum disorder risk genes from exome sequencing studies (11,242 interactions, OR = 1.07, $P = 7.8e-13$).

I prioritized specific OOA risk genes based on the extent of their interaction with this shared neuropsychiatry gene network. I defined a core set of 684 neuropsychiatry-related genes with prior evidence from at least three independent approaches, and I identified their interaction partners in the STRING database. I ranked the OOA risk genes based on their eigen-centrality in this network. The top 250 genes, selected from among the 820 OOA risk genes with GWAS p-values < 0.01 , are shown in Fig. 4.4B. These genes were enriched for 13 Gene Ontology functional categories (FDR < 0.01), including genes localized to dendrites (18 genes, $P = 5.5e-6$) and genes involved in signal transduction (39 genes, $P = 3.1e-6$) and focal adhesion (18 genes, $P = 2.5e-5$). These results suggest that OOA risk loci harbor novel risk genes within a polygenic gene network that is shared with neuropsychiatry genes discovered by independent approaches. Full data can be found in Appendix (Table A6).

Polygenic risk scores from the broader European population explain a small proportion of risk for mood disorders in the Old Order Amish

I tested whether polygenic effects of common risk variants identified in the broader European population contribute to risk for mood disorders in the OOA, extending previous work in the ASMAD cohort^{23,84}. Using PRSice-2, I derived polygenic risk scores (PRSs) from published European ancestry GWAS of BD^{13,14}, MDD¹¹, and SCZ²⁶ and applied them in the OOA. In this framework, PRS are calculated at multiple significance thresholds in the training dataset and evaluated in the test dataset to identify an optimal set of SNPs with the greatest predictive power. At the optimal threshold, PRS for BD explained 2.4% of risk for mood disorders in the OOA ($P = 8.3 \times 10^{-6}$, number of SNPs used = 36,150); 42 of the individuals (21%) in the top quartile of

BD scores were affected, compared to just 16 (8%) of the individuals in the bottom quartile (OR = 3.0, 95% CI = 1.6 - 5.5, $P = 0.0005$). PRS for SCZ explained 1.8% of risk for mood disorders in the OOA ($P = 1.3 \times 10^{-4}$, number of SNPs = 36,603); 35 of the individuals (18%) in the top quartile of SCZ scores were affected, compared to 18 (9% of the individuals) in the bottom quartile (OR = 2.2, 95% CI = 1.2 – 4.0, $P = 0.01$). PRS for MDD explained 0.8% of the risk for mood disorders in the OOA (threshold=0.00015, $P = 0.01$, number of SNPs = 2,019); 39 of the individuals (20%) in the top quartile of BD scores were affected, compared to 19 (10% of the individuals) in the bottom quartile (OR = 2.1, 95% CI = 1.2 – 3.8, $P = 0.01$). Individuals with mood disorders had significantly higher PRS than their unaffected family members (BD, $P = 9.4 \times 10^{-6}$; SCZ, $P = 2.5 \times 10^{-5}$; MDD, $P = 0.02$). Interestingly, population controls generally had lower PRS than either affected or unaffected members of families in which the proband had a mood disorder. Therefore, polygenic inheritance of known polygenic risk factors from the broader European population explain a small but statistically significant percentage of risk for mood disorders in the Lancaster OOA (Figure 4.5). The difference in PRS between the family and population controls was also significantly different.

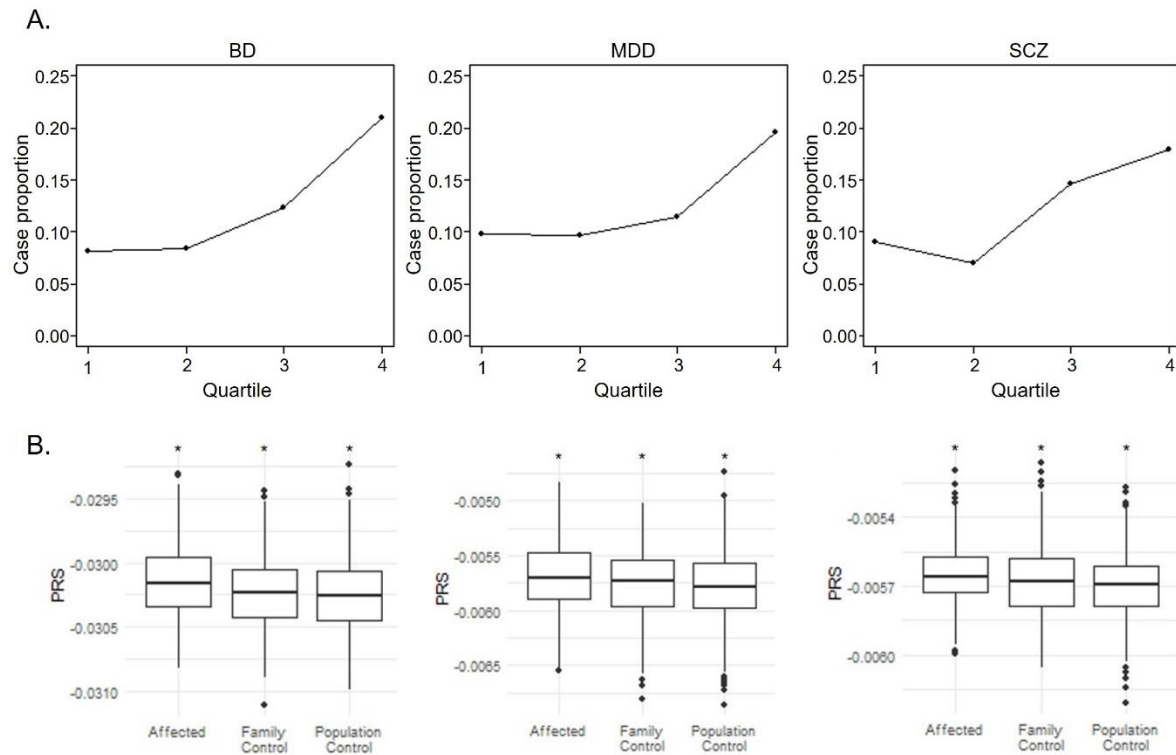


Figure 4.5. Polygenic risk scores from large-scale GWAS in the broader population predict a small proportion of risk for mood disorders in the OOA. A) Individuals were divided into quartiles based on PRS score. The proportion of individuals who are cases are higher in the top quartiles compared to the bottom quartiles. B) Average PRS divided by case status as well as by familial or population control.

Quantitative behavioral and neurocognitive phenotypes

The availability of deep phenotyping data in the ACP sample (n=314) presented an opportunity to evaluate the associations of mood disorder-associated SNPs with quantitative, sub-clinical cognitive and behavioral traits. These 314 individuals were also included in the primary GWAS of mood disorders. I analyzed three behavioral scales that assess both the current and lifelong history of classic behavioral symptoms of mood disorders: Beck Depression Index⁸⁹ (depression symptoms in the two weeks prior to testing), Maryland Depression Trait scale⁹⁰ (lifetime depression symptoms), and the modified Bipolar Spectrum Diagnostic Scale⁹¹ (BSDS; measuring

the polarity of the depressive and manic symptoms). I also analyzed data from five cognitive dimensions previously implicated in mood disorders, assessed using the WASI⁸⁸: Digit Sequencing (which primarily measures verbal working memory), Spatial Span (which measures visuospatial working memory), Digit Symbol Coding (which measures mainly processing speed), Matrix Reasoning and Vocabulary (which are part of the intelligence assessment). I found significant heritability in this sample for all eight traits ($p < 0.006$; Table 4.5), with broad-sense heritability estimates ranging from $h^2 = 0.25$ (MD Depression Trait, $n = 310$, $p = 0.003$) to $h^2 = 0.53$ (Vocabulary, $n = 160$, $p = 4 \times 10^{-9}$). I also confirmed a positive association between a mood disorder diagnosis and higher scores on each of the mood symptom scales (Table 4.6). I found negative associations between a mood disorder diagnosis and performance on cognitive tasks, strongest for digit sequencing and digit symbol coding. In general, the heritability of the cognitive traits was higher than the heritability of the mood symptom scales, while the correlations with mood disorder diagnoses were stronger for mood symptom scales. These results suggest the traits have a genetic basis in the OOA and are relevant to mood disorders.

Table 4.5. Heritability and summary statistics of the behavioral and neurocognitive traits used in the PheWAS analysis (all individuals from the ACP cohort)

PheWAS Trait	N	Heritability (SD)	Heritability p-value	Median Score (Min, Max)	Correlation with mood disorder diagnosis
Beck Depression Index (BDI)	310	0.41 (0.10)	2.2×10^{-6}	4 (0,33)	0.29
Bipolar Spectrum Diagnostic Scale (BSDS)	297	0.37 (0.12)	5.4×10^{-5}	5 (0,99)	0.43
MD Depression Trait	310	0.25 (0.10)	0.003	5 (58)	0.58
Digit Sequencing	294	0.48 (0.11)	3.6×10^{-7}	19 (3,28)	-0.11
Digit Symbol Coding	301	0.35 (0.10)	1.1×10^{-5}	60 (14, 111)	-0.16
Matrix Reasoning	160	0.45 (0.15)	6.3×10^{-5}	21 (1, 32)	-0.09
Spatial Span	304	0.43 (0.10)	2.5×10^{-7}	16 (6, 62)	0.02
Vocabulary	160	0.53 (0.12)	4.1×10^{-9}	45.5 (15, 77)	-0.08

I tested for association of each lead SNP from my GWAS of mood disorders with the eight behavioral and cognitive traits (Fig. 4.5, Table 4.7). Despite small samples, I found several nominally significant associations. The lead SNP at the 3q28/29 locus, rs192622352, was associated with a higher MD Depression Trait score: $n = 7$ carriers (all affected) and 303 non-carriers; $\beta = 15.3$, $P = 0.0005$. The lead SNP at the chr16 locus, rs7185072, was associated with an increased MD Depression Trait score: $n = 185$ carriers and 129 non-carriers; $\beta = 2.27$, $P = 0.01$. The lead SNP at the 5q13 locus (rs569742752) was associated with decreased performance on the digit symbol coding task: $n=2$ carriers (both affected) and 299 non-carriers, $\beta = -28.3$, $P = 0.006$. The lead SNP at the 7q22 locus, rs117752843, was associated with increased scores on both the MD Depression Trait and BSDS scales: $n = 12$ carriers (10 affected) and 302 non-carriers, MD Depression Trait, $\beta = 12.07$, $P = 0.0001$; BSDS, $\beta = 0.4$, $P = 0.04$. These associations persisted for all four loci when I compared the mean score between affected carriers and affected non-carriers ($P < 2 \times 10^{-10}$). Notably, $n=2$ carriers of rs117752843 who did not have

a DSM-5 diagnosis nonetheless had elevated depression and bipolarity symptoms compared to non-affected non-carriers ($P < 0.05$), suggesting effects on these mood-related traits that are independent of diagnosis. In summary, for all four lead SNPs I found nominally significant associations ($P < 0.05$) with either behavioral or cognitive phenotypes; two of these associations (3q28/29 and 7q22) remain significant after correction for multiple testing (adjusted $P < 0.002$). These findings also buttress the primary GWAS results, as 3 of the 4 SNPs were associated with a higher score on a scale frequently used as a screening tool for possible mood disorder diagnosis.

An Amish-enriched missense variant in the CUX1 Alternative Splicing Product associated with mood disorders

Association testing in founder populations has the potential to identify population-enriched, functional alleles with substantial effects on disease risk. I therefore searched among the SNPs at each locus for protein-coding variants in strong LD with our lead SNPs. This analysis revealed a missense variant in strong LD with the lead SNP ($r^2 = 0.90$) at the 7q22 risk locus, rs118010189; $P = 9.0e-6$; *CUX1* Lys500Gln. rs118010189 had a minor allele frequency of 0.017 in this Amish cohort, 2.3-fold enriched compared to the non-Finnish Europeans in the gnomAD database. rs118010189 is located in an alternatively spliced exon of the *CUX1* gene that is included only in the *CUX1* Alternative Splicing Product (*CASP*) isoform. The protein product of the canonical *CUX1* transcript is a homeodomain transcription factor. By contrast, *CASP*, which shares ~400 amino acids with other *CUX1* isoforms, lacks transcription factor activity and instead encodes a transmembrane protein involved in intra-Golgi retrograde transport¹³²⁻¹³⁵. The two protein

products share a promoter and several hundred amino acids, and the structure of this conjoined gene is conserved across most vertebrate species.

While *CUX1* has established roles in brain development, the functions of *CASP* in the brain have not been studied. I examined the expression of exons specific to each protein product in the developing and adult brain using RNA-seq data from BrainSpan (Fig. 4.6). Both *CUX1* and *CASP* exons were expressed broadly across cortical and sub-cortical brain regions at prenatal timepoints, becoming restricted to specific regions postnatally. Interestingly, the postnatal expression patterns differ. *CUX1* exons were highly expressed specifically in the primary visual cortex and in the cerebellum. By contrast, *CASP* exons were highly expressed only in the cerebellum and did not have substantial expression in the cortex. These results suggest differential use of *CUX1* and *CASP* isoforms in cortical vs. cerebellar cell types.

Figure 4.6. PheWAS results for four putative mood disorder risk loci. A. Z-score of beta coefficients for 3 behavioral traits and 5 cognitive traits. Blue dots indicate the z-score for the lead SNP at each locus. B. Plot of individual MD depression trait scores by genotype for the 3q28 locus. C. Plot of individual digit symbol coding scores by genotype for the 5q13 locus. D. Plot of individual MD depression trait scores and BSDS scores and BSDS scores by genotype for the 7q22 locus. E. Plot of individual MD depression trait scores by genotype for the 16q21 locus.

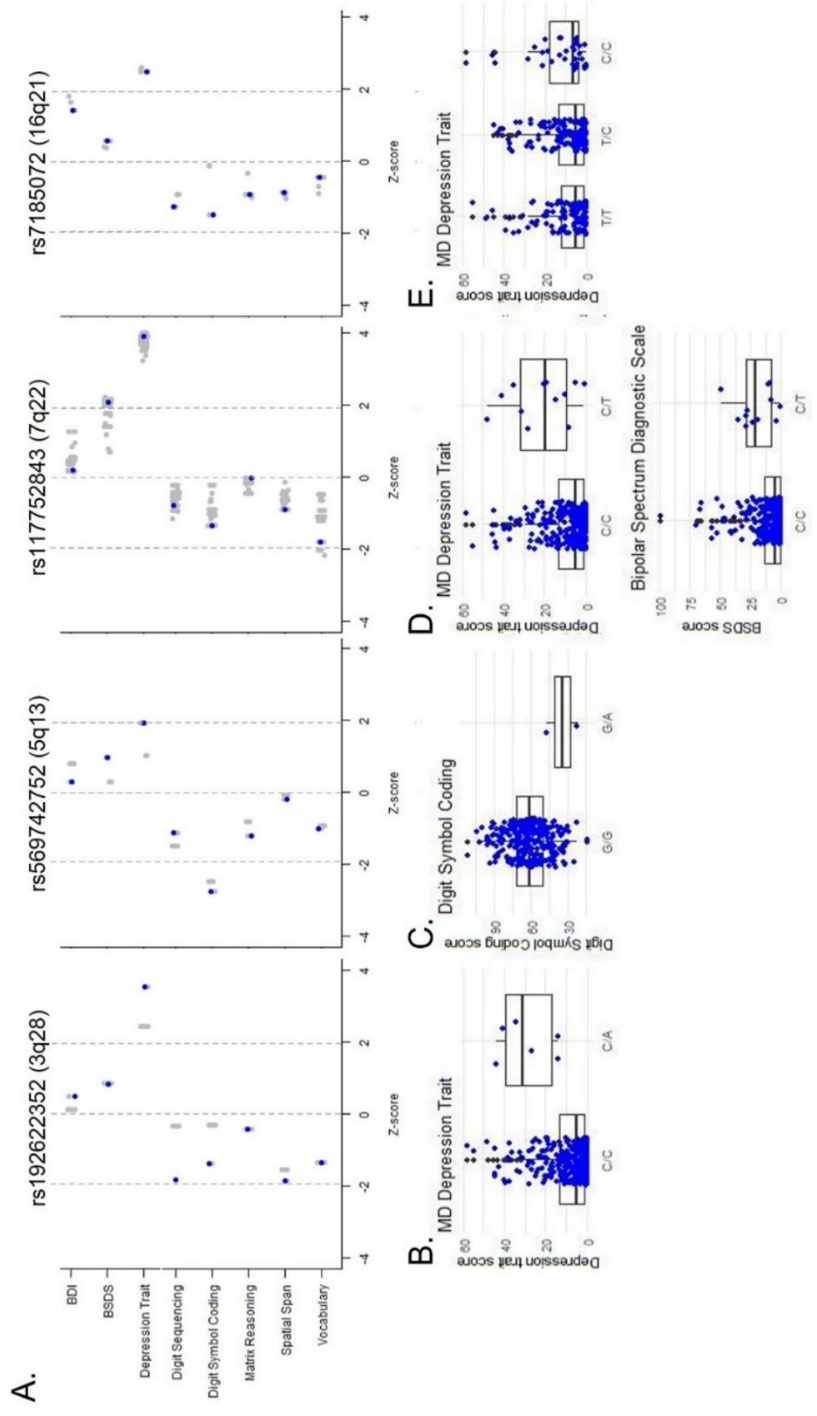


Table 4.6. PheWAS results for lead SNPs of risk loci associated with mood disorders in the OOA. A) Behavioral phenotypes B) Cognitive phenotypes

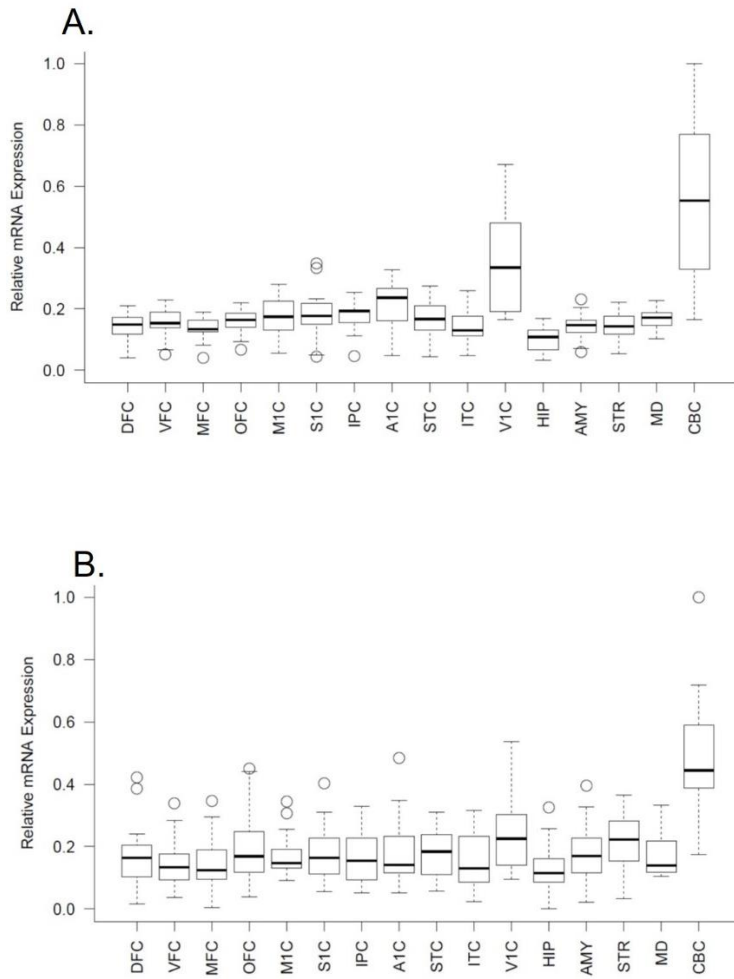
A)

SNP	BDI_p	BSDS_p	dstT_p
chr16:62294564:T:C	0.160	0.587	0.013
chr3:191857829:C:A	0.641	0.428	0.0005
chr5:76339511:G:A	0.795	0.348	0.057
chr7:103511937:C:T	0.851	0.037	0.0001

B)

SNP	DigitSequencing_p	DigitSymbol Coding_p	Matrix Reasoning_p	SpatialSpan_p	Vocabulary_p
chr16:62294564:T:C	0.2037	0.139	0.353	0.396	0.667
chr3:191857829:C:A	0.067	0.169	0.688	0.062	0.180
chr5:76339511:G:A	0.267	0.006	0.223	0.861	0.311
chr7:103511937:C:T	0.436	0.177	0.973	0.374	0.071

Figure 4.7. A. Expression of *CUX1* exons across the developing and adult brain using RNA-seq from BrainSpan. B. Expression of *CASP* exons across the developing and adult brain using RNA-seq from BrainSpan.



Chapter 5: Discussion

Here, I present a GWAS of mood disorders and related traits in the largest available sample of psychiatrically phenotyped members of an Old Order Amish founder population, revealing the first genome-wide significant risk loci for mood disorders in this population. Deep phenotyping revealed effects of these loci on quantitative behavioral and cognitive endophenotypes. Annotation of risk loci revealed a missense variant on a risk-associated haplotype at 7q22, CUX1 Lys500Gln. I discuss these findings in the broader context of the genetic architecture and causal mechanisms underlying psychiatric disorders.

It is perhaps remarkable that I was able to discover genome-wide significant risk loci despite having genome-wide data from just 1,672 individuals, of whom 196 were affected. The apparent effect sizes of the identified risk loci are substantially larger than those that have been identified in the broader population, which increases power to detect true associations. This result has several limitations. While some of the risk loci overlap known risk loci in the broader population, most are novel, and I lack an independent cohort by which to further evaluate the associations of specific OOA-enriched risk alleles. I cannot rule out winner's curse effects, which would predict that the true effect sizes are smaller than is observed in the current sample. It is also possible these loci, particularly the rare 3q28/29 and 5q13 variants, are false positive results, which is a concern given the small sample size of this GWA analysis. Nonetheless, there are several reasons to expect risk loci with larger effects when studying population isolates, including reduced genetic and environmental heterogeneity, population-enriched alleles with substantial effects, and additive or synergistic effects of co-inherited variants on the same risk-associated haplotype.

In support of the idea that the large effect sizes result from population-enriched alleles with substantial effects, all of the risk loci include OOA-enriched SNPs that are in strong LD with the lead SNPs. Causal, OOA-enriched alleles may also include unimputed variants in LD with the SNPs in the analysis, including indels and structural variants. The most promising directly-analyzed OOA-enriched SNP was rs118010189 (*CUX1* Lys500Gln) at the 7q22 risk locus. *CUX1*, including the *CASP* isoform directly impacted by rs118010189, has been previously implicated in other brain disorders. A GWAS in a small MDD cohort found associations between SNPs near *CUX1* and response to antidepressant treatment¹³⁶. In addition, a rare non-coding variant associated with increased activity of the shared *CUX1/CASP* promoter was identified in two consanguineous families with autism spectrum disorders¹³⁷. The *CUX1* transcription factor has been extensively studied in the context of neuronal development. In the cortex, *CUX1* is expressed specifically in upper-layer pyramidal neurons and has roles in dendritic branching, spine morphology, and synaptic function and is more highly expressed in caudal vs. rostral cortical areas, suggesting roles in cortical arealization¹³⁹⁻¹⁴². In the cerebellum, *CUX1* contributes to the development of granule neurons¹⁴². These established functions are consistent with the gene expression analyses. By contrast, roles of the *CASP* isoform in the brain are unknown. Biochemical evidence suggested that *CASP* and the *CUX1* transcription factor can form protein-protein interactions *in vitro*, raising the possibility of a functional interaction¹³⁵. If so, high expression of both *CUX1* and *CASP* in the cerebellum could point toward modulation of granular neuron development, an intriguing hypothesis to pursue in future research. Cerebellar circuits are increasingly recognized to play important roles in emotional control, with deficits in mood disorders^{137,143,144}.

Large effects of risk loci in the OOA could also result from additive or synergistic effects of co-inherited variants. Observing effects of this kind is more likely in population isolates, since genomic segments shared identical by descent are substantially longer than in the broader population³⁸. The risk loci that I identified, while broad, include a relatively small number of linked SNPs, suggesting that they are composed of relatively few ancestral haplotypes. The 7q22 risk locus provides a fascinating example. This locus spans more than ten megabases, yet I found just 60 SNPs in LD with the lead SNP. The region includes multiple genes with prior evidence for association with neuropsychiatric traits. The lead SNP in the OOA overlaps previously described risk loci for several neuropsychiatric traits¹⁴⁵⁻¹⁴⁷ and is located in an intron of *RELN*, an extracellular matrix protein important in neuronal migration and in the establishment of cortical layers during early development¹⁴⁸. Rare loss-of-function variants in *RELN* have been identified in probands with autism spectrum disorders and related neurodevelopmental disorders¹¹³. Interestingly, *CUX1* has been shown to be required for the expression of *RELN* in developing cortical neurons¹⁴⁹, raising the possibility that SNPs impacting the functions of *CUX1* and *RELN* could influence brain development and risk for mood disorders non-independently. However, *RELN* and *CUX1* are just two among 44 genes within the bounds of the risk locus with prior evidence for association with neuropsychiatric disorders based on rare variants or differential expression in the prefrontal cortex. *CUX1* Lys500Gln is the only protein-coding variant I identified at the locus, yet many genes in this region could be impacted by non-coding variants. Annotation of non-coding variants in this region identified many potential regulatory consequences, but in general these annotations do not provide high enough confidence to rule in or out specific causal genes (data not shown). Since many of the variants are in nearly perfect LD

with the lead SNP within the cohort, they are co-inherited in most of the carriers, increasing the likelihood of additive or synergistic effects. Strong LD in this population precludes statistical fine-mapping approaches that are used in GWAS of the broader population to fine map causal SNPs.

The findings build on >40 years of research on mood disorders in the OOA population. Egeland and colleagues initiated the ASMAD study in the 1970s and reported a linkage scan in an 81-member extended pedigree in 1987⁴⁰. This initial study suggested linkage of bipolar disorder to markers on the proximal arm of chromosome 11, providing great excitement at the time. However, this signal could not be replicated in an expanded Amish cohort or in other populations, part of the broader failure of linkage analysis to deliver robust genetic findings in psychiatric disorders. More recently, genome sequences of 394 OOA individuals from the ASMAD cohort were ascertained through a combination of direct genome sequencing and family-based imputation. Analyses of these data identified suggestive linkage and association signals involving single-nucleotide variants²³ and copy number variants⁴⁴ and demonstrated shared polygenic risk between OOA and the broader population^{23,84}. However, these prior studies of the ASMAD cohort did not detect strong statistical evidence implicating a particular set of risk loci or common pathways, most likely because the studies were underpowered due to heterogeneity and polygenicity.

If replicated, these findings would implicate genes that play a major role in risk for mood disorders. The discovery of genome-wide significant risk loci with relatively large effects represents an important step forward, enabling a variety of future studies. One goal of future studies will be to better understand the clinical phenotypes of OOA risk allele carriers. I have

already begun this through the analysis of behavioral and cognitive endophenotypes in the ACP cohort. Additional phenotypes of interest will include assessments of brain structure and function, which are being collected through ACP. Studies in patient-derived and genome-edited induced pluripotent stem cells also provide exciting opportunities to test the functions of these and other risk variants.

Appendix

Table A1. Risk loci associated with mood disorders in the OOA, as well as their relative enrichment.

*indicates lead SNP. All other SNPs are in LD with the lead SNP.

SNP	Amish MAF	European MAF	Amish enrichment	Amish.GWAS BETA	Amish.GWAS P	Amish_lead R2
chr16:62294564:T:C*	0.311005	0.3712	0.837836746	-0.080107	2.70E-08	1
chr16:62294615:A:G	0.311005	0.3714	0.837385568	-0.080107	2.70E-08	1
chr16:62295890:A:T	0.311005	0.371	0.83828841	-0.080107	2.70E-08	1
chr16:62296173:T:C	0.311005	0.3711	0.838062517	-0.080107	2.70E-08	1
chr16:62296780:C:T	0.311005	0.3811	0.816071897	-0.080107	2.70E-08	1
chr16:62299596:T:C	0.385766	0.3774	1.022167462	-0.071648	1.55E-07	0.704158
chr16:62303850:G:A	0.38128	0.3668	1.039476554	-0.069825	3.81E-07	0.686347
chr3:190970484:G:A	0.00358852	0.04793	0.074870019	-0.74435	4.98E-08	0.856628
chr3:191487002:T:C	0.00358852	0.01166	0.307763293	-0.74435	4.98E-08	0.856628
chr3:191544286:C:T	0.00358852	0.01919	0.186999479	-0.74435	4.98E-08	0.856628
chr3:191594537:C:T	0.00358852	0.01919	0.186999479	-0.74435	4.98E-08	0.856628
chr3:191609545:C:T	0.00358852	0.0228	0.157391228	-0.74435	4.98E-08	0.856628
chr3:191850272:C:T	0.00358852	0.001334	2.690044978	-0.74435	4.98E-08	0.856628
chr3:191857829:C:A*	0.0041866	0.02869	0.14592541	-0.65894	3.35E-08	1
chr3:191891184:G:A	0.00358852	0.006772	0.529905493	-0.74435	4.98E-08	0.856628
chr3:191892660:A:G	0.00358852	0.0008369	4.287871908	-0.74435	4.98E-08	0.856628
chr3:191954871:C:G	0.00358852	0.0008213	4.369316937	-0.74435	4.98E-08	0.856628
chr3:192130608:G:A	0.00358852	0.0007278	4.930640286	-0.74435	4.98E-08	0.856628
chr3:192135614:A:G	0.00448565	0.09576	0.046842627	-0.47004	2.76E-05	0.684272
chr3:192882063:C:T	0.00358852	0.006365	0.563789474	-0.74435	4.98E-08	0.856628
chr3:192926052:A:G	0.00388756	0.01423	0.273194659	-0.58034	4.42E-06	0.790338
chr3:193299693:G:A	0.00358852	6.20E-05	57.91672046	-0.74435	4.98E-08	0.856628
chr5:74704443:A:G	0.00448565	0.006674	0.672108181	-0.49075	1.28E-06	0.79928
chr5:74836279:A:G	0.00448565	0.004895	0.916373851	-0.49075	1.28E-06	0.79928
chr5:74860694:G:C	0.00448565	0.004833	0.928129526	-0.49075	1.28E-06	0.79928
chr5:76339511:G:A*	0.00358852	0.001379	2.602262509	-0.73774	6.35E-10	1
chr5:76357798:G:A	0.00358852	0.001673	2.144961148	-0.73774	6.35E-10	1
chr5:76506247:G:A	0.00257732	9.29E-05	27.73399333	-0.73429	6.39E-05	0.699225
chr7:100749379:A:G	0.0155502	0.05	0.311004	-0.20028	0.00018163	0.765761
chr7:100808738:G:A	0.0152512	0.01133	1.346090026	-0.20642	0.00013373	0.78128
chr7:100833004:G:A	0.0170455	0.0192	0.887786458	-0.20021	9.14E-05	0.696332
chr7:101858974:A:G	0.0152512	0.00954	1.598658281	-0.20642	0.00013373	0.78128
chr7:101868021:A:G	0.0152512	0.009371	1.627489062	-0.20642	0.00013373	0.78128
chr7:102076433:A:G	0.0155502	0.01083	1.435844875	-0.20028	0.00018163	0.765761
chr7:102093647:G:T	0.0179426	0.01408	1.274332386	-0.24123	1.80E-06	0.89055

Table A1 continued.

SNP	Amish MAF	European MAF	Amish enrichment	Amish.GWAS BETA	Amish.GWAS P	Amish_lead R2
chr7:102116443:G:C	0.0179426	0.01056	1.699109848	-0.24123	1.80E-06	0.89055
chr7:102117558:G:A	0.0179426	0.01059	1.694296506	-0.24123	1.80E-06	0.89055
chr7:102142747:C:A	0.0179426	0.01713	1.047437245	-0.24123	1.80E-06	0.89055
chr7:102150913:T:C	0.0179426	0.02311	0.776399827	-0.24123	1.80E-06	0.89055
chr7:102172436:G:C	0.0179426	0.01085	1.653695853	-0.24123	1.80E-06	0.89055
chr7:102194960:G:C	0.0179426	0.01308	1.37175841	-0.24123	1.80E-06	0.89055
chr7:102278021:A:C	0.0179426	0.008708	2.060473128	-0.24123	1.80E-06	0.89055
chr7:102431830:T:C	0.0179426	0.007541	2.37933961	-0.24123	1.80E-06	0.89055
chr7:102812541:G:A	0.0257177	0.04142	0.620900531	-0.20465	1.72E-06	0.750973
chr7:102917271:C:T	0.0224282	0.01853	1.210372369	-0.25419	2.86E-08	0.864024
chr7:102917556:A:C	0.0257177	0.04132	0.622403195	-0.20465	1.72E-06	0.750973
chr7:103072929:C:T	0.0257177	0.03863	0.66574424	-0.20465	1.72E-06	0.750973
chr7:103439754:G:A	0.0254187	0.02019	1.25897474	-0.23817	3.91E-08	0.760042
chr7:103511937:C:T*	0.0194378	0.005744	3.384018106	-0.27945	1.30E-08	1
chr7:103884660:C:T	0.0191388	0.01554	1.231583012	-0.25762	2.62E-07	0.953177
chr7:103942612:T:C	0.0191388	0.01577	1.213620799	-0.26194	1.55E-07	0.92252
chr7:104731826:T:C	0.0194378	0.01456	1.335013736	-0.25481	2.73E-07	0.907996
chr7:104859444:G:A	0.0188397	0.009684	1.945446097	-0.21161	2.11E-05	0.761019
chr7:105621726:C:A	0.0212321	0.04215	0.503727165	-0.19051	5.70E-05	0.647925
chr7:105673301:C:T	0.0176435	0.005003	3.52658405	-0.24127	3.23E-06	0.784721
chr7:105695684:C:A	0.0188397	0.01328	1.418652108	-0.21464	1.78E-05	0.733333
chr7:105707364:A:C	0.0188397	0.0134	1.405947761	-0.21464	1.78E-05	0.733333
chr7:106011511:C:A	0.0176435	0.00251	7.029282869	-0.24127	3.23E-06	0.784721
chr7:106057497:G:A	0.0176435	0.01381	1.277588704	-0.24127	3.23E-06	0.784721
chr7:106067427:G:T	0.0176435	0.002354	7.495114698	-0.24127	3.23E-06	0.784721
chr7:106067431:A:G	0.0176435	0.002354	7.495114698	-0.24127	3.23E-06	0.784721
chr7:106077945:G:A	0.0176435	0.002434	7.248767461	-0.24127	3.23E-06	0.784721
chr7:106083023:T:C	0.0176435	3.10E-05	569.3288157	-0.24127	3.23E-06	0.784721
chr7:106151467:G:A	0.0176435	0.006303	2.799222592	-0.24127	3.23E-06	0.784721
chr7:106187521:A:C	0.0188397	0.005792	3.252710635	-0.28057	1.72E-08	0.733333
chr7:106247727:C:T	0.0188397	0.005811	3.242075374	-0.28057	1.72E-08	0.733333
chr7:106437895:C:T	0.0167464	0.002587	6.473289525	-0.26991	3.61E-07	0.738103
chr7:106739799:C:G	0.0179426	0.004292	4.180475303	-0.24911	1.23E-06	0.687196
chr7:108163502:C:T	0.0167464	0.003561	4.702723954	-0.26991	3.61E-07	0.738103
chr7:108367937:C:T	0.0167464	0.003052	5.487024902	-0.26991	3.61E-07	0.738103
chr7:108637911:C:T	0.0170455	0.004634	3.678355632	-0.26991	3.61E-07	0.724707
chr7:108692182:C:T	0.0170455	0.004774	3.570485966	-0.26991	3.61E-07	0.724707
chr7:108724884:C:T	0.0170455	0.004003	4.258181364	-0.26991	3.61E-07	0.724707

Table A1 continued.

SNP	Amish MAF	European MAF	Amish enrichment	Amish.GWAS BETA	Amish.GWAS P	Amish_lead R2
chr7:108776155:G:A	0.0170455	0.003548	4.804255919	-0.26991	3.61E-07	0.724707
chr7:108820968:C:A	0.0170455	0.003548	4.804255919	-0.26991	3.61E-07	0.724707
chr7:109680191:T:C	0.0164474	0.00693	2.373362193	-0.2843	9.03E-08	0.722589
chr7:109680628:G:A	0.0167464	0.03178	0.526947766	-0.27431	1.74E-07	0.709226
chr7:109691709:G:C	0.0167464	0.02277	0.735458937	-0.27431	1.74E-07	0.709226
chr7:109702206:A:G	0.0167464	0.02725	0.614546789	-0.27431	1.74E-07	0.709226
chr7:109868482:A:G	0.0167464	0.02704	0.619319527	-0.27431	1.74E-07	0.709226
chr7:110148126:C:T	0.0164474	0.00477	3.448092243	-0.27339	1.96E-07	0.693759
chr7:111118542:G:A	0.0155502	0.01792	0.867756696	-0.26439	1.17E-06	0.647447
chr7:111167898:A:G	0.0155502	0.02919	0.532723535	-0.26439	1.17E-06	0.647447
chr7:111265572:C:G	0.0155502	0.0122	1.274606557	-0.26439	1.17E-06	0.647447
chr7:111321049:G:T	0.0155502	0.02865	0.542764398	-0.26439	1.17E-06	0.647447
chr7:111429406:T:A	0.0155502	0.02788	0.557754663	-0.26439	1.17E-06	0.647447
chr7:111429408:T:C	0.0155502	0.0287	0.541818815	-0.26439	1.17E-06	0.647447
chr7:111440525:C:T	0.0155502	0.01853	0.839190502	-0.26439	1.17E-06	0.647447
chr7:99213680:A:G	0.0152512	0.004554	3.34896794	-0.20642	0.00013373	0.78128

Table A2. Beta and p-values from leave-one-out EMMAX analysis. * indicates lead SNP. All other SNPs are in LD with the lead SNP.

SNP	No.ACP BETA	No.ACP P	No.AMBiGen BETA	No.AMBiGen P	No.ASMAD BETA	No.ASMAD P
chr16:62294564:T:C*	-0.059537	6.72E-06	-0.075253	3.91E-08	-0.052926	0.00016624
chr16:62294615:A:G	-0.059537	6.72E-06	-0.075253	3.91E-08	-0.052926	0.00016624
chr16:62295890:A:T	-0.059537	6.72E-06	-0.075253	3.91E-08	-0.052926	0.00016624
chr16:62296173:T:C	-0.059537	6.72E-06	-0.075253	3.91E-08	-0.052926	0.00016624
chr16:62296780:C:T	-0.059537	6.72E-06	-0.075253	3.91E-08	-0.052926	0.00016624
chr16:62299596:T:C	-0.053964	1.65E-05	-0.071022	4.05E-08	-0.045644	0.0005829
chr16:62303850:G:A	-0.055386	1.18E-05	-0.069059	1.15E-07	-0.043151	0.0012563
chr3:190970484:G:A	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191487002:T:C	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191544286:C:T	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191594537:C:T	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191609545:C:T	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191850272:C:T	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191857829:C:A*	-0.44915	0.021076	-0.68093	9.28E-10	-0.62057	6.81E-08
chr3:191891184:G:A	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191892660:A:G	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:191954871:C:G	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:192130608:G:A	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:192135614:A:G	-0.20879	0.13673	-0.4919	2.66E-06	-0.40156	0.00014576
chr3:192882063:C:T	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr3:192926052:A:G	-0.31165	0.10948	-0.64805	4.61E-08	-0.4902	5.96E-05
chr3:193299693:G:A	-0.9434	0.00054561	-0.76644	1.72E-09	-0.69799	2.00E-07
chr5:74704443:A:G	-0.49267	6.05E-07	-0.45051	5.29E-06	-0.40493	0.0004696
chr5:74836279:A:G	-0.49267	6.05E-07	-0.45051	5.29E-06	-0.40493	0.0004696
chr5:74860694:G:C	-0.49267	6.05E-07	-0.45051	5.29E-06	-0.40493	0.0004696
chr5:76339511:G:A*	-0.69561	1.40E-09	-0.7003	3.17E-09	-0.93469	1.29E-08
chr5:76357798:G:A	-0.69561	1.40E-09	-0.7003	3.17E-09	-0.93469	1.29E-08
chr5:76506247:G:A	-0.74342	2.62E-06	-0.7228	3.17E-05	6.69E+14	1
chr7:100749379:A:G	-0.082804	0.10076	-0.19855	0.00010916	-0.22903	1.61E-05
chr7:100808738:G:A	-0.085995	0.092682	-0.20533	7.55E-05	-0.23645	1.08E-05
chr7:100833004:G:A	-0.097769	0.042293	-0.19893	5.00E-05	-0.21403	2.54E-05
chr7:101858974:A:G	-0.085995	0.092682	-0.20533	7.55E-05	-0.23645	1.08E-05
chr7:101868021:A:G	-0.085995	0.092682	-0.20533	7.55E-05	-0.23645	1.08E-05
chr7:102076433:A:G	-0.082804	0.10076	-0.19855	0.00010916	-0.22903	1.61E-05
chr7:102093647:G:T	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102116443:G:C	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102117558:G:A	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102142747:C:A	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102150913:T:C	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102172436:G:C	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102194960:G:C	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102278021:A:C	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06

Table A2 continued.

SNP	No.ACP BETA	No.ACP P	No.AMBiGen BETA	No.AMBiGen P	No.ASMAD BETA	No.ASMAD P
chr7:102431830:T:C	-0.12843	0.0071534	-0.23918	7.39E-07	-0.252	1.47E-06
chr7:102812541:G:A	-0.12782	0.0014321	-0.19666	1.38E-06	-0.22422	7.61E-07
chr7:102917271:C:T	-0.15385	0.00042076	-0.24724	1.66E-08	-0.26909	2.88E-08
chr7:102917556:A:C	-0.12782	0.0014321	-0.19666	1.38E-06	-0.22422	7.61E-07
chr7:103072929:C:T	-0.12782	0.0014321	-0.19666	1.38E-06	-0.22422	7.61E-07
chr7:103439754:G:A	-0.1656	7.10E-05	-0.2295	3.44E-08	-0.23698	1.39E-07
chr7:103511937:C:T*	-0.18187	7.15E-05	-0.26832	1.29E-08	-0.28417	5.14E-08
chr7:103884660:C:T	-0.1776	0.00010814	-0.24552	3.16E-07	-0.24734	4.21E-06
chr7:103942612:T:C	-0.18538	4.89E-05	-0.24929	1.95E-07	-0.23838	7.11E-06
chr7:104731826:T:C	-0.18094	6.08E-05	-0.2412	3.77E-07	-0.22668	1.63E-05
chr7:104859444:G:A	-0.12802	0.0045021	-0.19628	3.78E-05	-0.20864	3.82E-05
chr7:105621726:C:A	-0.11362	0.0080827	-0.17531	0.0001054	-0.17356	0.00020932
chr7:105673301:C:T	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:105695684:C:A	-0.13367	0.0032128	-0.20227	2.44E-05	-0.19192	0.00014216
chr7:105707364:A:C	-0.13367	0.0032128	-0.20227	2.44E-05	-0.19192	0.00014216
chr7:106011511:C:A	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:106057497:G:A	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:106067427:G:T	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:106067431:A:G	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:106077945:G:A	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:106083023:T:C	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:106151467:G:A	-0.152	0.0013378	-0.22664	5.22E-06	-0.22668	1.63E-05
chr7:106187521:A:C	-0.20918	3.47E-06	-0.26921	1.64E-08	-0.22668	1.63E-05
chr7:106247727:C:T	-0.20918	3.47E-06	-0.26921	1.64E-08	-0.22668	1.63E-05
chr7:106437895:C:T	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:106739799:C:G	-0.14669	0.001634	-0.23215	2.46E-06	-0.24773	2.02E-06
chr7:108163502:C:T	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:108367937:C:T	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:108637911:C:T	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:108692182:C:T	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:108724884:C:T	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:108776155:G:A	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:108820968:C:A	-0.16268	0.00078761	-0.25438	6.15E-07	-0.2709	6.09E-07
chr7:109680191:T:C	-0.16268	0.00078761	-0.26734	1.78E-07	-0.29088	8.89E-08
chr7:109680628:G:A	-0.16268	0.00078761	-0.25881	3.03E-07	-0.27542	2.48E-07
chr7:109691709:G:C	-0.16268	0.00078761	-0.25881	3.03E-07	-0.27542	2.48E-07
chr7:109702206:A:G	-0.16268	0.00078761	-0.25881	3.03E-07	-0.27542	2.48E-07
chr7:109868482:A:G	-0.16268	0.00078761	-0.25881	3.03E-07	-0.27542	2.48E-07
chr7:110148126:C:T	-0.16268	0.00078761	-0.25655	3.91E-07	-0.27925	1.87E-07
chr7:111118542:G:A	-0.17623	0.00039449	-0.24644	2.65E-06	-0.26354	3.08E-06
chr7:111167898:A:G	-0.17623	0.00039449	-0.24644	2.65E-06	-0.26354	3.08E-06
chr7:111265572:C:G	-0.17623	0.00039449	-0.24644	2.65E-06	-0.26354	3.08E-06
chr7:111321049:G:T	-0.17623	0.00039449	-0.24644	2.65E-06	-0.26354	3.08E-06

Table A2 continued.

SNP	No.ACP BETA	No.ACP P	No.AMBiGen BETA	No.AMBiGen P	No.ASMAD BETA	No.ASMAD P
chr7:111429406:T:A	-0.17623	0.00039449	-0.24644	2.65E-06	-0.26354	3.08E-06
chr7:111429408:T:C	-0.17623	0.00039449	-0.24644	2.65E-06	-0.26354	3.08E-06
chr7:111440525:C:T	-0.17623	0.00039449	-0.24644	2.65E-06	-0.26354	3.08E-06
chr7:99213680:A:G	-0.085995	0.092682	-0.20533	7.55E-05	-0.23645	1.08E-05

Table A3. Genome-wide suggestive loci ($1 \times 10^{-5} < P < 5 \times 10^{-8}$, as calculated by EMMAX analysis).

SNP	BP	Amish_MAF	Amish.GWAS_BETA	Amish.GWAS_P
chr1:14857974:C:G	14857974	0.00449	-0.48464	6.403E-06
chr1:205390367:A:C	205390367	0.00368	-0.60139	5.545E-06
chr1:205497755:C:T	205497755	0.00331	-0.72545	5.610E-07
chr1:207003360:C:T	207003360	0.00598	-0.50331	5.450E-08
chr1:21626391:G:A	21626391	0.07805	-0.11609	2.161E-06
chr10:112195650:G:T	112195650	0.00179	-0.69274	3.693E-06
chr10:1243234:C:G	1243234	0.00295	-0.93656	2.549E-07
chr10:15869221:T:C	15869221	0.01944	-0.25144	2.415E-07
chr10:1628314:C:T	1628314	0.00239	-0.88662	1.072E-06
chr10:16344936:C:T	16344936	0.04187	-0.14723	9.720E-06
chr10:1662446:C:T	1662446	0.00295	-0.93656	2.549E-07
chr10:16715524:C:T	16715524	0.02123	-0.21233	3.642E-06
chr10:1801644:T:A	1801644	0.00331	-0.79672	9.854E-06
chr10:2490488:T:G	2490488	0.00295	-0.91333	3.667E-07
chr10:853363:A:T	853363	0.00295	-0.93656	2.549E-07
chr11:11133798:T:C	11133798	0.00837	-0.36541	1.591E-06
chr11:11456463:C:T	11456463	0.00987	-0.32478	3.920E-06
chr11:73351751:T:A	73351751	0.00897	-0.37165	1.701E-06
chr11:73661008:A:G	73661008	0.00897	-0.37165	1.701E-06
chr11:73785643:G:C	73785643	0.00927	-0.33757	8.632E-06
chr12:128587647:A:G	128587647	0.00209	-0.86448	1.491E-06
chr12:128772119:C:T	128772119	0.00258	-0.91333	3.667E-07
chr12:128889154:G:A	128889154	0.00239	-0.86448	1.491E-06
chr13:105322359:C:T	105322359	0.08493	-0.11448	1.335E-06
chr13:98669926:T:C	98669926	0.00807	-0.36978	5.633E-07
chr14:37761704:T:C	37761704	0.00837	-0.35291	1.952E-06
chr14:82577537:C:A	82577537	0.00748	-0.42117	1.132E-06
chr14:88094103:T:C	88094103	0.01286	-0.31694	9.525E-08
chr16:13789100:A:T	13789100	0.00598	-0.36170	3.141E-06
chr16:60376930:A:G	60376930	0.31250	0.06298	9.401E-06
chr16:62291043:A:G	62291043	0.47668	0.05915	6.836E-06
chr16:66178602:G:A	66178602	0.22189	-0.07519	4.213E-06
chr19:3220859:C:T	3220859	0.00090	-0.93541	2.366E-06
chr19:3251228:C:T	3251228	0.00090	-0.93541	2.366E-06
chr2:114908053:A:G	114908053	0.00190	-0.89113	1.920E-06
chr2:116316924:A:G	116316924	0.00190	-0.89113	1.920E-06
chr2:116636367:T:C	116636367	0.00190	-0.89113	1.920E-06

Table A3 continued.

SNP	BP	Amish_MAF	Amish.GWAS_BETA	Amish.GWAS_P
chr2:116810063:G:A	116810063	0.00190	-0.89113	1.920E-06
chr2:117135130:G:A	117135130	0.00190	-0.89113	1.920E-06
chr2:117334422:A:C	117334422	0.00190	-0.89113	1.920E-06
chr2:117404230:T:C	117404230	0.00190	-0.89113	1.920E-06
chr2:117473409:C:G	117473409	0.00190	-0.89113	1.920E-06
chr2:118100588:C:T	118100588	0.00190	-0.89113	1.920E-06
chr2:132868533:G:T	132868533	0.01047	-0.29142	6.729E-06
chr2:181519355:A:G	181519355	0.00628	-0.36293	9.785E-06
chr2:62496849:G:T	62496849	0.00359	-0.63792	5.838E-07
chr2:70591867:A:G	70591867	0.02931	-0.18402	2.353E-06
chr22:23523967:A:G	23523967	0.02841	-0.19779	1.014E-06
chr22:23928876:C:G	23928876	0.04396	-0.17352	5.689E-08
chr22:25877559:C:T	25877559	0.09151	-0.10557	3.170E-06
chr22:27559449:G:A	27559449	0.03768	-0.18072	4.308E-07
chr22:27574607:T:G	27574607	0.04845	-0.14065	7.133E-06
chr22:29273723:C:G	29273723	0.09510	-0.10777	1.753E-06
chr22:30585521:A:G	30585521	0.03439	-0.16331	6.756E-06
chr22:33521682:T:G	33521682	0.02243	-0.20839	5.530E-06
chr22:49933448:C:T	49933448	0.00221	-0.70194	6.411E-06
chr22:50356210:T:G	50356210	0.00221	-0.70194	6.411E-06
chr3:190197177:A:G	190197177	0.00568	-0.45551	8.484E-06
chr3:190384284:T:C	190384284	0.00568	-0.45551	8.484E-06
chr3:192926052:A:G	192926052	0.00389	-0.58034	4.416E-06
chr4:101391127:G:C	101391127	0.03349	-0.16389	9.168E-06
chr4:107665709:A:C	107665709	0.33463	-0.06633	3.074E-06
chr4:108412912:A:T	108412912	0.21621	-0.08074	5.632E-07
chr4:109315949:G:T	109315949	0.40431	0.06240	9.260E-06
chr4:109558594:C:T	109558594	0.11423	-0.09995	6.450E-06
chr5:107761197:T:C	107761197	0.00598	-0.49317	1.126E-06
chr5:36402675:T:C	36402675	0.07955	-0.11586	3.238E-06
chr5:5276372:T:C	5276372	0.00179	-0.76394	1.433E-06
chr5:5414590:G:A	5414590	0.00221	-0.80192	4.168E-07
chr5:6231816:A:T	6231816	0.00179	-0.76394	1.433E-06
chr5:6648928:C:T	6648928	0.00552	-0.48291	7.076E-06
chr5:7166723:C:A	7166723	0.00479	-0.60030	3.364E-07
chr5:7971663:T:C	7971663	0.00515	-0.52719	2.415E-06
chr6:13138144:G:A	13138144	0.00419	0.52817	6.863E-06

Table A3 continued.

SNP	BP	Amish_MAF	Amish.GWAS_BETA	Amish.GWAS_P
chr6:13839974:T:C	13839974	0.00419	0.52817	6.863E-06
chr6:30114592:T:C	30114592	0.00778	-0.29982	4.318E-06
chr6:5007869:T:C	5007869	0.00331	-0.76416	8.662E-07
chr6:5046241:G:A	5046241	0.00748	-0.33464	6.507E-06
chr6:5596605:A:G	5596605	0.00269	-0.72017	3.547E-06
chr6:6642148:G:A	6642148	0.00295	-0.88492	9.788E-08
chr6:7652539:G:A	7652539	0.00239	-0.84305	3.823E-07
chr6:7674439:C:T	7674439	0.00295	-0.88492	9.788E-08
chr6:8113459:A:G	8113459	0.00295	-0.88492	9.788E-08
chr6:8396640:C:T	8396640	0.00295	-0.88492	9.788E-08
chr6:8863092:C:T	8863092	0.00269	-0.84305	3.823E-07
chr6:93338892:T:C	93338892	0.38816	0.06271	6.836E-06
chr7:102812541:G:A	102812541	0.02572	-0.20465	1.716E-06
chr7:103884660:C:T	103884660	0.01914	-0.25762	2.624E-07
chr7:106057497:G:A	106057497	0.01764	-0.24127	3.228E-06
chr7:107146977:A:G	107146977	0.02183	-0.21737	2.783E-06
chr7:108629794:G:A	108629794	0.02303	-0.21600	1.806E-06
chr7:108736632:G:A	108736632	0.02183	-0.21921	1.162E-06
chr7:109089075:G:A	109089075	0.02572	-0.19092	7.418E-06
chr7:109680628:G:A	109680628	0.01675	-0.27431	1.742E-07
chr7:110680523:A:T	110680523	0.01764	-0.26913	1.422E-07
chr7:110801574:T:A	110801574	0.02213	-0.21354	1.750E-06
chr7:111167898:A:G	111167898	0.01555	-0.26439	1.172E-06
chr7:1560546:G:A	1560546	0.00515	-0.52946	5.395E-06
chr7:1952387:T:G	1952387	0.00515	-0.52946	5.395E-06
chr7:4176913:C:T	4176913	0.00628	-0.47566	2.835E-07
chr9:70634994:A:G	70634994	0.04067	-0.16532	1.146E-06

Table A4. Locus overlap analysis using previously published GWAS results.

Chr	Amish start	Amish end	Locus	Study start	Study end	Study.leadSNP	Study
chr16	62293007	62303850	16q21	61909736	62409736	chr16:62159736:G:T	Educational attainment
chr5	67547732	86309654	5q13	68018555	68518555	chr5:68268555:G:A	Educational attainment
chr5	67547732	86309654	5q13	68259445	68759445	chr5:68509445:C:T	Educational attainment
chr5	67547732	86309654	5q13	72633747	73133747	chr5:72883747:G:C	Educational attainment
chr5	67547732	86309654	5q13	75419729	75919729	chr5:75669729:G:A	Educational attainment
chr5	67547732	86309654	5q13	81433516	81673688	chr5:81500549:C:T	Bipolar Disorder
chr5	67547732	86309654	5q13	81565047	82065047	chr5:81815047:C:T	Educational attainment
chr7	99213680	111440525	7q22	98908974	99408974	chr7:99158974:C:T	Educational attainment
chr7	99213680	111440525	7q22	99718384	100218384	chr7:99968384:G:A	Educational attainment
chr7	99213680	111440525	7q22	100229650	100729650	chr7:100479650:G:A	Educational attainment
chr7	99213680	111440525	7q22	101802537	102302537	chr7:102052537:G:C	Educational attainment
chr7	99213680	111440525	7q22	101862242	102362242	chr7:102112242:C:T	Educational attainment
chr7	99213680	111440525	7q22	104298635	104798635	chr7:104548635:C:T	Educational attainment
chr7	99213680	111440525	7q22	104540442	105040442	chr7:104790442:G:C	Educational attainment
chr7	99213680	111440525	7q22	104630081	105130081	chr7:104880081:T:A	Educational attainment
chr7	99213680	111440525	7q22	104776887	105276887	chr7:105026887:G:A	Educational attainment
chr7	99213680	111440525	7q22	104935691	105407711	chr7:105407711:C:T	Bipolar Disorder
chr7	99213680	111440525	7q22	104957617	105422617	chr7:104957617:C:A	Schizophrenia
chr7	99213680	111440525	7q22	105451161	105951161	chr7:105701161:G:A	Educational attainment
chr7	99213680	111440525	7q22	108393162	108893162	chr7:108643162:G:A	Educational attainment
chr7	99213680	111440525	7q22	109331603	109589327	chr7:109459862:G:A	Major Depressive Disorder
chr7	99213680	111440525	7q22	111203759	111565859	chr7:111203759:C:T	Schizophrenia

Table A5. Top 250 genes (by p-value) identified in the MAGMA gene-level analysis.

HGNC symbol	CHR	START	STOP	NSNPS	P	eigen-centrality
ATP13A5	3	193269789	193383820	220	1.63E-07	0.038451457
SV2C	5	76078383	76358939	475	2.83E-07	0.054147694
MB21D2	3	192791815	192922856	184	6.46E-07	0.002960792
ICE1	5	5415664	5495220	64	3.59E-06	0.004453917
LRRRC17	7	102908000	102950111	54	4.30E-06	0.02423863
EEF1AKNMT	1	171776660	171819023	145	4.98E-06	0.009639465
FAM163A	1	179738163	179821198	264	9.49E-06	0.005386533
TDRD5	1	179586613	179696272	290	1.11E-05	0.034807872
NPHS2	1	179545539	179580952	118	1.24E-05	0.04138139
KIAA0040	1	175151986	175197999	124	1.54E-05	0.004675967
TMEM184A	7	1537235	1565821	49	1.63E-05	0.005193958
MRPS28	8	79913717	80035289	126	1.68E-05	0.014667477
SDK1	7	3296252	4274000	3669	1.77E-05	0.048025897
ABL2	1	179094330	179234684	195	1.82E-05	0.306908274
TMEM17	2	62495218	62516894	71	1.86E-05	0.002954023
RAP1GAP	1	21591215	21674363	246	2.09E-05	0.09076395
RAB36	22	23140326	23169350	33	2.10E-05	0.040182376
SLC26A5	7	103347730	103451207	228	2.47E-05	0.030270191
NAMPT	7	106243298	106291326	79	2.83E-05	0.038524055
EIF3B	7	2349086	2385745	81	2.97E-05	0.066728048
TRIM31	6	30097897	30118090	106	3.22E-05	0.006791017
MINDY3	10	15773170	15865507	117	3.32E-05	0.012553809
PRRC2C	1	171480551	171598511	180	3.32E-05	0.018099989
FBXL13	7	102808230	103079843	389	3.47E-05	0.023553422
RELN	7	103466784	103994658	1724	3.91E-05	0.15416986
CELF5	19	3219661	3302076	226	4.13E-05	0.103973556
RSPH14	22	23054415	23150021	150	4.42E-05	0.013120363
ADAMTS16	5	5135330	5325304	430	4.98E-05	0.006702686
INSIG2	2	118083452	118115997	83	5.43E-05	0.021987059
NA	8	79913860	80085775	205	5.68E-05	NA
ALPL	1	21504397	21583410	315	6.43E-05	0.024444032
RAB29	1	205762986	205780482	41	7.48E-05	0.031101461
NA	7	1810804	2247215	1524	0.00010877	NA
FAM171A1	10	15206643	15376289	531	0.00011196	0.009206096
FAM169A	5	74772574	74871966	187	0.00013334	0.004862044
USP47	11	11836423	11966887	215	0.00013452	0.046873717
TNR	1	175310194	175748616	1076	0.0001382	0.280129989
CLDN16	3	190317541	190417143	261	0.00014703	0.008172373
SRD5A1	5	6628427	6679386	168	0.00015875	0.015179791

Table A5 continued.

HGNC symbol	CHR	START	STOP	NSNPS	P	eigen-centrality
RSU1	10	16585611	16822463	754	0.00016534	0.039288437
NUCKS1	1	205707822	205755182	84	0.00016651	0.01199234
HEXB	5	74635023	74727647	208	0.00016784	0.010698272
PAPSS1	4	107585276	107725234	296	0.00016904	0.032235497
MYO18B	22	25737144	26036041	445	0.00017494	0.072637571
MAD1L1	7	1810793	2238243	1505	0.00017708	0.032706825
LY86	6	6583108	6659983	258	0.00018176	0.012007322
GRAP2	22	39896084	39978721	55	0.00019513	0.070087931
NRCAM	7	108142623	108461717	763	0.00019989	0.145742494
SLC15A1	13	98678801	98757672	197	0.00021264	0.020627737
NA	10	801914	936705	324	0.00021446	NA
ARMC10	7	103069881	103104759	27	0.00021792	0.004884697
STMN2	8	79605814	79671175	187	0.00021944	0.176834831
DYNLT4	1	44800913	44811675	25	0.00022346	0.00173235
BTBD19	1	44803482	44820585	37	0.00022426	0.005014238
RPP40	6	4989717	5009063	83	0.00024034	0.005738306
PLK3	1	44795377	44810990	33	0.0002411	0.064748178
DPP10	2	114437299	115850752	2929	0.00024379	0.085536797
NCLN	19	3180563	3214575	113	0.00025044	0.016806325
BMP6	6	7721099	7886728	422	0.0002508	0.030392916
S1PR4	19	3167346	3185332	48	0.00026245	0.020400106
IMMP2L	7	110658051	111567517	1510	0.00026422	0.030426846
LRRN3	7	111086006	111130454	53	0.00028598	0.045066398
MRPL53	2	74466982	74477687	14	0.00028908	0.013875664
CCDC178	18	32932402	33446101	596	0.0002919	0.003459221
ANKRD30BL	2	132142591	132262969	80	0.00030022	0.02372877
ORAI2	7	102428106	102461825	75	0.00031929	0.021909347
IGLL1	22	23568125	23585302	81	0.0003203	0.003448393
FAM185A	7	102743971	102814225	138	0.00032383	0.001422226
MSTO1	1	155605205	155619967	11	0.00033095	0.010934855
PMP2	8	81435326	81452439	49	0.00037955	0.110727919
NA	2	74456057	74470410	18	0.00039044	NA
OSTC	4	108645585	108672820	20	0.00039767	0.00855751
CLK2	1	155257868	155283491	33	0.00040037	0.056376281
SLC30A1	1	211566568	211584161	33	0.00041349	0.027743752
FARP1	13	98137562	98460176	936	0.00042372	0.10716734
FLG2	1	152343735	152365006	29	0.00042746	0.004017458
CCDC142	2	74466986	74488408	25	0.00044347	NA
S100A11	1	152027506	152052907	30	0.00044949	0.037153312

Table A5 continued.

HGNC symbol	CHR	START	STOP	NSNPS	P	eigen-centrality
RPTN	1	152148595	152164228	11	0.0004564	0.001550924
KHDC4	1	155908045	155939413	20	0.00048751	0.004351094
RIT1	1	155892808	155916404	24	0.00048792	0.093936732
GALNT18	11	11265877	11627005	1335	0.00049979	0.008762108
EIF3J	15	44532125	44568029	47	0.00050314	0.040124799
PEX19	1	160271812	160291348	17	0.00051107	0.02503865
MZT2A	2	131459900	131497743	137	0.00054452	0.0027587
THBS1	15	39576079	39604466	63	0.0005476	0.0656968
MRPL13	8	120375761	120450402	52	0.00055056	0.070577734
MON1B	16	77185835	77207398	68	0.00055563	0.014734328
LRRC71	1	156915632	156938094	57	0.00055798	0.001971988
TUBA3D	2	131471119	131487934	77	0.00056581	0.090788703
ADARB2	10	1172313	1742525	1877	0.00057143	0.068943388
NA	16	75709131	75731490	61	0.00058991	NA
NAXE	1	156586762	156599299	32	0.00059518	0.024799733
NCKAP5	2	132666788	133573463	2033	0.00061124	0.008398788
SULT1C3	2	108242195	108270351	63	0.00061804	0.001928435
GFPT1	2	69314769	69392254	79	0.00062604	0.065908964
GPATCH4	1	156589487	156606496	59	0.00062763	0.015771224
NA	16	75655227	75682009	79	0.00063598	NA
CRTC2	1	153942669	153963615	27	0.00065185	0.025251772
TOMM20L	14	58390928	58413702	13	0.00065808	0.023078184
BUB1B	15	40156023	40226136	52	0.00065811	0.053774278
CYP2B6	19	40986282	41023398	134	0.00066188	0.022517671
SEC16B	1	177918956	177989303	237	0.00067723	0.029111151
DERL3	22	23829503	23844128	69	0.00068655	0.011134676
PCBP1	2	70082477	70094203	22	0.00069022	0.083934411
OR1J1	9	122471958	122482926	49	0.00070128	0.003248709
MRRF	9	122259603	122336337	68	0.0007094	0.028526772
TTC24	1	156574727	156591770	37	0.00071305	0.028259909
SYCE1L	16	77194408	77218215	103	0.00071667	0.003385587
ACTR10	14	58195080	58240636	45	0.00073319	0.112948762
APOE	19	44900791	44914393	29	0.00074191	0.147911041
NIPBL	5	36871769	37071413	246	0.00074342	0.048053814
OR6K3	1	158711327	158725720	40	0.00077083	0.003095601
PRKAR2B	7	107039705	107166811	229	0.00077161	0.357741504
TRIM13	13	49990888	50025481	54	0.00078202	0.006188343
CUX1	7	101810904	102288958	868	0.00080921	0.064640229
INSRR	1	156835063	156864117	37	0.00081941	0.085643715

Table A5 continued.

HGNC symbol	CHR	START	STOP	NSNPS	P	eigen-centrality
SOAT1	1	179288714	179363680	218	0.00084006	0.031511702
COG5	7	107196555	107569514	724	0.00085057	0.019976099
HCN3	1	155272463	155294848	41	0.00085543	0.061636428
ZNF624	17	16615734	16658856	25	0.00085935	0.006487381
FDX1	11	110424948	110469884	138	0.0008597	0.023916504
THEM4	1	151865866	151914637	50	0.00088426	0.015196531
NA	14	37590847	38046442	771	0.00088435	NA
GMCL1	2	69824660	69886384	104	0.00089052	0.005011672
TERF1	8	73003864	73053123	85	0.00090203	0.040741894
MCUB	4	109555209	109693719	102	0.00093129	0.011856183
CTDSPL2	15	44422622	44534038	153	0.00094047	0.013623047
EWSR1	22	29263009	29305525	53	0.00094306	0.071581201
FLYWCH1	16	2906937	2956208	184	0.00095076	0.00663259
COPA	1	160283594	160348273	75	0.00096991	0.04801438
HK2	2	74828981	74898359	104	0.00098536	0.083779219
UGGT1	2	128086200	128200677	177	0.0009915	0.021361223
NA	22	25346418	25410377	32	0.0010009	NA
C9orf153	9	86215265	86264657	103	0.0010028	0.002785047
	1	160211800	160290130	55	0.0010178	0.027119018
TLX3	5	171304248	171317139	16	0.0010199	0.021546156
PHF11	13	49490610	49533981	120	0.0010298	0.012361633
RBBP5	1	205081142	205127015	111	0.0010492	0.039304921
SFT2D3	2	127696497	127710242	36	0.0010606	0.001532034
SEC24B	4	109428772	109545896	236	0.0010752	0.042586454
TNFRSF11B	8	118918557	118956885	126	0.001084	0.022577947
MED30	8	117515713	117545262	74	0.0010896	0.013247146
FGR	1	27607064	27640185	20	0.0011169	0.062894209
DNMT3L	21	44241339	44267216	16	0.0011192	0.013055096
UBQLN4	1	156030299	156058798	36	0.001127	0.111043735
METTL21C	13	102680744	102699504	74	0.0011284	0.009389003
TOMM40	19	44885569	44908689	51	0.0011373	0.073799309
NA	15	40213500	40269890	92	0.0011394	NA
RGPD4	2	107821937	107895841	40	0.0011545	0.022404518
TENT4A	5	6708432	6762044	206	0.0011576	0.039361376
CHST6	16	75467052	75500445	105	0.0011653	0.002401228
LHFPL3	7	104323700	104912232	1524	0.001167	0.052636284
MXD1	2	69892688	69947945	113	0.0011696	0.010733975
EIF2AK4	15	39929115	40040591	270	0.0011719	0.059422597
NA	4	106352485	106373825	51	0.0011824	NA

Table A5 continued.

HGNC symbol	CHR	START	STOP	NSNPS	P	eigen-centrality
SERPINF1	17	1757029	1782565	70	0.0011919	0.024422187
AZI2	3	28310003	28354050	50	0.0012066	0.003660344
ZNRF1	16	74994024	75115994	194	0.0012275	0.013952348
NEK2	1	211653657	211680630	71	0.0013152	0.103911455
YY1AP1	1	155654443	155694000	42	0.0013482	0.003575092
PSMA3	14	58239843	58277012	38	0.0013568	0.091838803
KRT7	12	52227520	52257186	65	0.0013632	0.02134723
SMARCB1	22	23781931	23843008	225	0.0013718	0.074982026
NIPAL4	5	157455019	157479717	77	0.0013742	0.01663481
AIMP1	4	106310544	106354456	94	0.0013811	0.061004347
OPRD1	1	28807142	28876267	164	0.0013995	0.052461226
SPRR2E	1	153088135	153111184	72	0.0014056	0.002233061
RLF	1	40156387	40245921	111	0.0014058	0.006669842
MT1M	16	56627659	56638981	35	0.00141	0.006785995
FCRL6	1	159795511	159821257	95	0.0014127	0.003062205
GOLM2	15	44283719	44420758	175	0.0014515	0.004602782
UBA2	19	34423352	34476251	75	0.0014907	0.061712707
ALDH9A1	1	165657216	165703863	117	0.0014996	0.035480463
TMCC2	1	205222946	205290632	152	0.0015003	0.041258858
PAK6	15	40212428	40282487	114	0.0015072	0.075949477
NA	16	56633666	56645087	48	0.0015105	NA
SULT1C2	2	108283639	108314915	75	0.001542	0.005545684
IMPA1	8	81651914	81691331	99	0.0015538	0.036294767
TIA1	2	70204444	70253660	101	0.001578	0.106671704
SLAMF8	1	159821811	159842492	52	0.0016014	0.010339994
NA	1	159829474	159878053	101	0.0016125	NA
GPR176	15	39794032	39925892	220	0.0016325	0.015683598
RASL10A	22	29307933	29324679	18	0.0016373	0.070553908
NA	2	126893864	126907097	34	0.0016691	NA
TERF2IP	16	75642773	75766872	353	0.001674	0.013004827
PLA2G12A	4	109704989	109735070	13	0.0016773	0.009401599
WNT11	11	76181325	76215736	67	0.0016992	0.048300951
NFKBID	19	35882653	35907303	78	0.0017069	0.023073643
RORC	1	151801071	151836845	73	0.0017078	0.012967745
SLC45A2	5	33939623	33989693	57	0.0017341	0.020074911
TUBB2B	6	3219277	3236730	20	0.0017381	0.490763286
B3GLCT	13	31194975	31337276	232	0.0017654	0.015083223
RALA	7	39618565	39713120	145	0.0017679	0.123923443
SLC5A7	2	107981523	108018994	83	0.0017765	0.04372234

Table A5 continued.

HGNC symbol	CHR	START	STOP	NSNPS	P	eigen-centrality
ELK4	1	205592556	205636962	115	0.0017769	0.017295948
NA	3	105650461	105874552	509	0.0017807	NA
EHBP1	2	62668851	63051487	422	0.0018013	0.077959764
PNO1	2	68152888	68181238	48	0.0018197	0.030554024
SGSM1	22	24801169	24932578	445	0.0018232	0.064685919
AQP4	18	26847043	26870771	69	0.0018596	0.17827284
AIRE	21	44280838	44303648	57	0.0018691	0.04023267
LMNA	1	156077573	156145089	157	0.0018743	0.125500376
FNDC3A	13	48970912	49214779	239	0.0018966	0.00890326
ZC2HC1A	8	78661089	78724765	87	0.0019003	0.012058283
WDR59	16	74866362	75005173	372	0.0019052	0.097202039
ARHGEF2	1	155941851	156012070	69	0.0019241	0.255330931
TRIM63	1	26046304	26073436	45	0.0019291	0.047223509
DCLRE1C	10	14892359	14959432	139	0.0019428	0.01110051
LPAR6	13	48384567	48449704	74	0.0019716	0.010012295
ENPP2	8	119552086	119678453	256	0.0019832	0.047273053
FSIP1	15	39583357	39787841	565	0.0020103	0.003858282
RALGPS2	1	178720147	178926841	420	0.0020427	0.038625881
ARHGEF11	1	156929840	157050370	240	0.0020466	0.060184259
MED4	13	48048323	48100131	109	0.0020546	0.026331565
SMAD2	18	47803957	47936146	132	0.0020645	0.092999449
TGFA	2	70442284	70559193	182	0.002073	0.031612263
ADCY2	5	7391138	7835081	897	0.0021171	0.370823027
ELOVL2	6	10975759	11049305	142	0.0021336	0.026198571
ADD2	2	70602618	70773225	110	0.0021544	0.092038366
FGF14	13	101705804	102407457	1693	0.0021565	0.310746627
GLG1	16	74442427	74612144	327	0.0021702	0.017832494
STK24	13	98440185	98582940	339	0.0021887	0.06583262
IFI16	1	158994968	159060155	178	0.0021932	0.021449704
RALB	2	120235064	120299710	53	0.0022067	0.101915718
NA	7	129121518	129135793	27	0.0022313	NA
NTRK1	1	156810640	156886850	127	0.0022714	0.183303964
PTPN2	18	12780478	12934643	355	0.0022719	0.052535592
SPART	13	36296638	36375180	174	0.0022754	0.023425326
CCDC190	1	162819458	162873815	172	0.0022871	2.44E-05
SLC30A8	8	116945273	117181714	440	0.0023025	0.023007617
CCDC169	13	36217008	36302840	204	0.0023067	0.000548593
KIRREL2	19	35850861	35872109	36	0.0023393	0.019837347
APBA1	9	69422532	69677371	460	0.0023446	0.118231251

Table A5 continued.

HGNC symbol	CHR	START	STOP	NSNPS	P	eigen-centrality
PFKL	21	44295051	44332376	114	0.0023744	0.077390035
KRT80	12	52163996	52197014	95	0.0023747	0.00484006
APCDD1	18	10449635	10494949	120	0.0024015	0.009457716
BLOC1S6	15	45582214	45620945	51	0.0024215	0.012352277
ZC3H12C	11	110088392	110176841	135	0.002425	0.00787054
C7orf33	7	148585766	148620860	81	0.0024363	0.000901469
UHRF1	19	4898080	4967154	198	0.0024404	0.023643852
NUF2	1	163261576	163360764	330	0.0024439	0.025598789
SQOR	15	45626148	45696281	279	0.0024721	0.026040107
CCDC121	2	27620639	27634012	17	0.0024794	0.001127419
CAB39L	13	49303650	49449064	303	0.0024857	0.019276075
SMYD4	17	1774485	1835634	141	0.0024923	0.017497673
GLT1D1	12	128848427	128989968	473	0.002502	0.002456109
OST4	2	27065472	27076654	12	0.0025103	0.00119002
MFSD4A	1	205563885	205607918	117	0.0025266	0.007561925

Table A6. Results of GO term analysis.

Category	Term	Count	P	Genes
GOTERM_BP_DIRECT	GO:0018108~peptidyl-tyrosine phosphorylation	14	2.92E-07	ENSG00000126561, ENSG00000176444, ENSG00000092445, ENSG00000198400, ENSG00000135333, ENSG00000196411, ENSG00000027644, ENSG00000189056, ENSG00000146648, ENSG00000257923, ENSG00000126934, ENSG00000000938, ENSG00000122025, ENSG00000143322
GOTERM_MF_DIRECT	GO:0004672~protein kinase activity	20	5.77E-07	ENSG00000105851, ENSG00000176444, ENSG00000156970, ENSG00000130758, ENSG00000198355, ENSG00000115977, ENSG00000132356, ENSG00000156711, ENSG00000136643, ENSG00000117650, ENSG00000146648, ENSG00000136297, ENSG00000126934, ENSG00000128829, ENSG00000071909, ENSG00000181409, ENSG00000100077, ENSG00000169967, ENSG00000143322, ENSG00000102572
GOTERM_BP_DIRECT	GO:0046777~protein autophosphorylation	14	1.12E-06	ENSG00000176444, ENSG00000092445, ENSG00000198400, ENSG00000130758, ENSG00000198355, ENSG00000115977, ENSG00000196411, ENSG00000027644, ENSG00000117650, ENSG00000146648, ENSG00000128829, ENSG00000000938, ENSG00000122025, ENSG00000102572
GOTERM_BP_DIRECT	GO:0007165~signal transduction	39	1.20E-06	ENSG00000131941, ENSG00000148484, ENSG00000076864, ENSG00000241973, ENSG00000173482, ENSG00000130758, ENSG00000105835, ENSG00000171680, ENSG00000071051, ENSG00000115138, ENSG00000164050, ENSG00000132356, ENSG00000155897, ENSG00000006451, ENSG00000146648, ENSG00000128917, ENSG00000158022, ENSG00000160310, ENSG0000017797, ENSG00000102466, ENSG00000092529, ENSG00000134769, ENSG00000092445, ENSG00000157927, ENSG00000163191, ENSG00000114279, ENSG00000133731, ENSG00000136643, ENSG00000164022, ENSG00000143622, ENSG00000144118, ENSG00000159461, ENSG00000136044, ENSG00000108055, ENSG00000109320, ENSG00000100077, ENSG00000143322, ENSG00000186635, ENSG00000102572
GOTERM_CC_DIRECT	GO:0030425~dendrite	18	3.10E-06	ENSG00000149295, ENSG00000126243, ENSG00000130203, ENSG00000076864, ENSG00000018189, ENSG00000128271, ENSG00000198400, ENSG00000152767, ENSG00000135333, ENSG00000152910, ENSG00000115665, ENSG00000078295, ENSG00000189056, ENSG00000173846, ENSG00000159461, ENSG00000182674, ENSG00000188529, ENSG00000143630

Table A6 continued.

Category	Term	Count	P	Genes
GOTERM_BP_DIRECT	GO:0051968~positive regulation of synaptic transmission, glutamatergic	6	5.53E-06	ENSG00000189056, ENSG00000146648, ENSG00000198400, ENSG00000128271, ENSG00000116147, ENSG00000162105
GOTERM_CC_DIRECT	GO:0048471~perinuclear region of cytoplasm	24	1.18E-05	ENSG00000160803, ENSG00000090020, ENSG00000160789, ENSG00000131941, ENSG00000104435, ENSG00000156970, ENSG00000173482, ENSG00000091136, ENSG00000171680, ENSG00000140945, ENSG00000138802, ENSG00000005249, ENSG00000169756, ENSG00000197959, ENSG00000146648, ENSG00000159461, ENSG00000136297, ENSG00000126934, ENSG00000100427, ENSG00000107282, ENSG00000117280, ENSG00000181409, ENSG00000183853, ENSG00000169504
GOTERM_MF_DIRECT	GO:0005524~ATP binding	42	1.42E-05	ENSG00000105851, ENSG00000241973, ENSG00000198400, ENSG00000130758, ENSG00000115977, ENSG00000159399, ENSG00000137843, ENSG00000132356, ENSG00000155897, ENSG00000117650, ENSG00000146648, ENSG00000126934, ENSG00000071909, ENSG00000142945, ENSG00000169967, ENSG00000131730, ENSG00000176444, ENSG00000156970, ENSG00000141959, ENSG00000092445, ENSG00000126261, ENSG00000198355, ENSG00000135333, ENSG00000196411, ENSG00000027644, ENSG00000133454, ENSG00000156711, ENSG00000078295, ENSG00000136643, ENSG00000173846, ENSG00000169994, ENSG00000128829, ENSG00000000938, ENSG00000108055, ENSG00000181409, ENSG00000107242, ENSG00000187522, ENSG00000100077, ENSG00000122025, ENSG00000143322, ENSG00000187527, ENSG00000102572
GOTERM_MF_DIRECT	GO:0004674~protein serine/threonine kinase activity	18	1.83E-05	ENSG00000105851, ENSG00000176444, ENSG00000156970, ENSG00000130758, ENSG00000198355, ENSG00000115977, ENSG00000137843, ENSG00000132356, ENSG00000156711, ENSG00000136643, ENSG00000117650, ENSG00000173846, ENSG00000126934, ENSG00000128829, ENSG00000071909, ENSG00000181409, ENSG00000169967, ENSG00000102572
GOTERM_CC_DIRECT	GO:0005925~focal adhesion	18	2.33E-05	ENSG00000090020, ENSG00000148484, ENSG00000241973, ENSG00000115109, ENSG00000140945, ENSG00000072163, ENSG00000169756, ENSG00000006451, ENSG00000116584, ENSG00000084733, ENSG00000061676, ENSG00000146648, ENSG00000126934, ENSG00000185883, ENSG00000187446, ENSG00000070831, ENSG00000163531, ENSG00000142798

Table A6 continued.

Category	Term	Count	P	Genes
GOTERM_CC_DIRECT	GO:0043025~neuronal cell body	16	2.52E-05	ENSG00000130203, ENSG00000104435, ENSG00000076864, ENSG00000128271, ENSG00000198400, ENSG00000135333, ENSG00000198742, ENSG00000005249, ENSG00000119699, ENSG00000116584, ENSG00000173846, ENSG00000159461, ENSG00000070831, ENSG00000188529, ENSG00000143630, ENSG00000162105
GOTERM_MF_DIRECT	GO:0004713~protein tyrosine kinase activity	10	9.31E-05	ENSG00000146648, ENSG00000126561, ENSG00000257923, ENSG00000176444, ENSG00000126934, ENSG00000092445, ENSG00000130758, ENSG00000000938, ENSG00000135333, ENSG00000143322
GOTERM_CC_DIRECT	GO:0045121~membrane raft	12	1.12E-04	ENSG00000090020, ENSG00000146648, ENSG00000100427, ENSG00000104537, ENSG00000116329, ENSG00000103546, ENSG00000116218, ENSG00000186318, ENSG00000183853, ENSG00000116147, ENSG00000005249, ENSG00000078295

Table A7. PheWAS results for risk loci associated with mood disorders using the behavioral endophenotypes. *indicates lead SNP

SNP	BDI_beta	BDI_p	BSDS_beta	BSDS_p	MD Depression Trait_beta	MD Depression Trait_p
chr16:62294564:T:C*	-0.0602	0.1602	-0.0306	0.5867	-2.2685	0.0135
chr16:62294615:A:G	-0.0602	0.1602	-0.0306	0.5867	-2.2685	0.0135
chr16:62295890:A:T	-0.0602	0.1602	-0.0306	0.5867	-2.2685	0.0135
chr16:62296173:T:C	-0.0602	0.1602	-0.0306	0.5867	-2.2685	0.0135
chr16:62296780:C:T	-0.0602	0.1602	-0.0306	0.5867	-2.2685	0.0135
chr16:62299596:T:C	-0.0724	0.0741	-0.0204	0.7014	-2.2423	0.0096
chr16:62303850:G:A	-0.0657	0.1007	-0.0189	0.7200	-2.1697	0.0110
chr3:190970484:G:A	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191487002:T:C	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191544286:C:T	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191594537:C:T	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191609545:C:T	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191850272:C:T	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191857829:C:A*	-0.0896	0.6410	-0.2020	0.4278	-15.3009	0.0005
chr3:191891184:G:A	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191892660:A:G	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:191954871:C:G	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:192130608:G:A	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:192135614:A:G	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:192882063:C:T	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:192926052:A:G	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr3:193299693:G:A	-0.0234	0.9100	-0.2247	0.4149	-11.6005	0.0159
chr5:74704443:A:G	0.2225	0.4412	-0.1023	0.7869	-6.2586	0.3134
chr5:74836279:A:G	0.2225	0.4412	-0.1023	0.7869	-6.2586	0.3134
chr5:74860694:G:C	0.2225	0.4412	-0.1023	0.7869	-6.2586	0.3134
chr5:76339511:G:A*	0.0920	0.7947	-0.4344	0.3481	-14.4121	0.0571
chr5:76357798:G:A	0.0920	0.7947	-0.4344	0.3481	-14.4121	0.0571
chr5:76506247:G:A	NA	NA	NA	NA	NA	NA
chr7:100749379:A:G	0.0652	0.6700	-0.3556	0.0758	-12.8496	0.0001
chr7:100808738:G:A	0.0652	0.6700	-0.3556	0.0758	-12.8496	0.0001
chr7:100833004:G:A	0.0211	0.8859	-0.2880	0.1339	-11.3718	0.0003
chr7:101858974:A:G	0.0652	0.6700	-0.3556	0.0758	-12.8496	0.0001
chr7:101868021:A:G	0.0652	0.6700	-0.3556	0.0758	-12.8496	0.0001
chr7:102076433:A:G	0.0652	0.6700	-0.3556	0.0758	-12.8496	0.0001
chr7:102093647:G:T	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102116443:G:C	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102117558:G:A	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001

Table A7 continued.

SNP	BDI_beta	BDI_p	BSDS_beta	BSDS_p	MD Depression Trait_beta	MD Depression Trait_p
chr7:102142747:C:A	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102150913:T:C	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102172436:G:C	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102194960:G:C	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102278021:A:C	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102431830:T:C	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:102812541:G:A	-0.0505	0.6694	-0.1048	0.4993	-9.3899	0.0002
chr7:102917271:C:T	-0.0147	0.9084	-0.1301	0.4378	-8.7996	0.0013
chr7:102917556:A:C	-0.0505	0.6694	-0.1048	0.4993	-9.3899	0.0002
chr7:103072929:C:T	-0.0505	0.6694	-0.1048	0.4993	-9.3899	0.0002
chr7:103439754:G:A	-0.0173	0.8752	-0.1691	0.2399	-8.9447	0.0001
chr7:103511937:C:T*	0.0275	0.8514	-0.3989	0.0373	-12.0766	0.0001
chr7:103884660:C:T	0.0529	0.7294	-0.3989	0.0464	-11.6246	0.0004
chr7:103942612:T:C	0.0529	0.7294	-0.3989	0.0464	-11.6246	0.0004
chr7:104731826:T:C	0.0529	0.7294	-0.3989	0.0464	-11.6246	0.0004
chr7:104859444:G:A	0.0529	0.7294	-0.3989	0.0464	-11.6246	0.0004
chr7:105621726:C:A	-0.0194	0.8904	-0.3140	0.0888	-10.0901	0.0008
chr7:105673301:C:T	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:105695684:C:A	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:105707364:A:C	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106011511:C:A	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106057497:G:A	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106067427:G:T	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106067431:A:G	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106077945:G:A	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106083023:T:C	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106151467:G:A	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106187521:A:C	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106247727:C:T	0.0222	0.8798	-0.3953	0.0392	-11.4144	0.0003
chr7:106437895:C:T	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:106739799:C:G	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:108163502:C:T	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:108367937:C:T	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:108637911:C:T	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:108692182:C:T	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:108724884:C:T	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:108776155:G:A	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002
chr7:108820968:C:A	0.0716	0.6398	-0.4347	0.0296	-12.0304	0.0002

Table A7 continued.

SNP	BDI_beta	BDI_p	BSDS_beta	BSDS_p	MD Depression Trait_beta	MD Depression Trait_p
chr7:109680191:T:C	0.1343	0.4017	-0.3726	0.0760	-13.6417	0.0001
chr7:109680628:G:A	0.0813	0.5955	-0.2797	0.1637	-12.3326	0.0002
chr7:109691709:G:C	0.0813	0.5955	-0.2797	0.1637	-12.3326	0.0002
chr7:109702206:A:G	0.0813	0.5955	-0.2797	0.1637	-12.3326	0.0002
chr7:109868482:A:G	0.0813	0.5955	-0.2797	0.1637	-12.3326	0.0002
chr7:110148126:C:T	0.1515	0.3451	-0.4677	0.0261	-13.4358	0.0001
chr7:111118542:G:A	0.2143	0.2050	-0.4620	0.0371	-12.6149	0.0005
chr7:111167898:A:G	0.2143	0.2050	-0.4620	0.0371	-12.6149	0.0005
chr7:111265572:C:G	0.2143	0.2050	-0.4620	0.0371	-12.6149	0.0005
chr7:111321049:G:T	0.2143	0.2050	-0.4620	0.0371	-12.6149	0.0005
chr7:111429406:T:A	0.2143	0.2050	-0.4620	0.0371	-12.6149	0.0005
chr7:111429408:T:C	0.2143	0.2050	-0.4620	0.0371	-12.6149	0.0005
chr7:111440525:C:T	0.2143	0.2050	-0.4620	0.0371	-12.6149	0.0005
chr7:99213680:A:G	0.0652	0.6700	-0.3556	0.0758	-12.8496	0.0001

Table A8. PheWAS results for risk loci associated with mood disorders using the digit sequencing, digit symbol coding, and spatial span scores. *indicates lead SNP

SNP	DigitSequencing beta	DigitSequencing p	DigitSymbol Coding beta	DigitSymbol Coding p	SpatialSpan beta	SpatialSpan p
chr16:62294564:T:C*	-0.4721	0.2033	1.8928	0.1393	-0.0333	0.3962
chr16:62294615:A:G	-0.4721	0.2033	1.8928	0.1393	-0.0333	0.3962
chr16:62295890:A:T	-0.4721	0.2033	1.8928	0.1393	-0.0333	0.3962
chr16:62296173:T:C	-0.4721	0.2033	1.8928	0.1393	-0.0333	0.3962
chr16:62296780:C:T	-0.4721	0.2033	1.8928	0.1393	-0.0333	0.3962
chr16:62299596:T:C	-0.3213	0.3605	0.1703	0.8883	-0.0366	0.3290
chr16:62303850:G:A	-0.3210	0.3542	-0.1115	0.9256	-0.0384	0.2996
chr3:190970484:G:A	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191487002:T:C	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191544286:C:T	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191594537:C:T	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191609545:C:T	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191850272:C:T	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191857829:C:A*	3.1906	0.0671	8.2169	0.1686	0.3244	0.0623
chr3:191891184:G:A	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191892660:A:G	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:191954871:C:G	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:192130608:G:A	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:192135614:A:G	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:192882063:C:T	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:192926052:A:G	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr3:193299693:G:A	0.6256	0.7433	1.9958	0.7600	0.2918	0.1207
chr5:74704443:A:G	3.6319	0.1388	20.8108	0.0128	0.0225	0.9314
chr5:74836279:A:G	3.6319	0.1388	20.8108	0.0128	0.0225	0.9314
chr5:74860694:G:C	3.6319	0.1388	20.8108	0.0128	0.0225	0.9314
chr5:76339511:G:A*	3.3587	0.2633	28.1471	0.0059	-0.0560	0.8611
chr5:76357798:G:A	3.3587	0.2633	28.1471	0.0059	-0.0560	0.8611
chr5:76506247:G:A	NA	NA	NA	NA	NA	NA
chr7:100749379:A:G	-0.6250	0.6465	4.6330	0.3197	0.1232	0.3740
chr7:100808738:G:A	-0.6250	0.6465	4.6330	0.3197	0.1232	0.3740
chr7:100833004:G:A	-1.2162	0.3499	2.6731	0.5481	0.1096	0.4097
chr7:101858974:A:G	-0.6250	0.6465	4.6330	0.3197	0.1232	0.3740
chr7:101868021:A:G	-0.6250	0.6465	4.6330	0.3197	0.1232	0.3740
chr7:102076433:A:G	-0.6250	0.6465	4.6330	0.3197	0.1232	0.3740
chr7:102093647:G:T	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102116443:G:C	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102117558:G:A	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740

Table A8 continued.

SNP	DigitSequencing beta	DigitSequencing p	DigitSymbol Coding beta	DigitSymbol Coding p	SpatialSpan beta	SpatialSpan p
chr7:102142747:C:A	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102150913:T:C	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102172436:G:C	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102194960:G:C	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102278021:A:C	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102431830:T:C	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:102812541:G:A	-0.6929	0.5128	0.7254	0.8412	0.0704	0.5334
chr7:102917271:C:T	-0.9611	0.4071	1.7756	0.6543	0.0658	0.5947
chr7:102917556:A:C	-0.6929	0.5128	0.7254	0.8412	0.0704	0.5334
chr7:103072929:C:T	-0.6929	0.5128	0.7254	0.8412	0.0704	0.5334
chr7:103439754:G:A	-0.5249	0.5929	2.1971	0.5124	0.0363	0.7343
chr7:103511937:C:T*	-1.0127	0.4364	6.0003	0.1769	0.1232	0.3740
chr7:103884660:C:T	-1.2372	0.3638	4.1099	0.3776	0.1043	0.4727
chr7:103942612:T:C	-1.2372	0.3638	4.1099	0.3776	0.1043	0.4727
chr7:104731826:T:C	-1.2372	0.3638	4.1099	0.3776	0.1043	0.4727
chr7:104859444:G:A	-1.2372	0.3638	4.1099	0.3776	0.1043	0.4727
chr7:105621726:C:A	-1.4418	0.2478	2.7772	0.5153	0.0168	0.8993
chr7:105673301:C:T	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:105695684:C:A	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:105707364:A:C	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106011511:C:A	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106057497:G:A	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106067427:G:T	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106067431:A:G	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106077945:G:A	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106083023:T:C	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106151467:G:A	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106187521:A:C	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106247727:C:T	-0.7455	0.5672	3.9197	0.3786	0.0627	0.6510
chr7:106437895:C:T	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:106739799:C:G	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:108163502:C:T	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:108367937:C:T	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:108637911:C:T	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:108692182:C:T	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:108724884:C:T	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:108776155:G:A	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509
chr7:108820968:C:A	-0.5593	0.6816	6.1873	0.1837	0.0865	0.5509

Table A8 continued.

SNP	DigitSequencing beta	DigitSequencing p	DigitSymbol Coding beta	DigitSymbol Coding p	SpatialSpan beta	SpatialSpan p
chr7:109680191:T:C	-0.9323	0.5159	5.5755	0.2554	0.0580	0.7040
chr7:109680628:G:A	-0.6842	0.6163	6.2424	0.1806	0.0623	0.6678
chr7:109691709:G:C	-0.6842	0.6163	6.2424	0.1806	0.0623	0.6678
chr7:109702206:A:G	-0.6842	0.6163	6.2424	0.1806	0.0623	0.6678
chr7:109868482:A:G	-0.6842	0.6163	6.2424	0.1806	0.0623	0.6678
chr7:110148126:C:T	-0.4590	0.7498	5.1064	0.2990	0.0715	0.6407
chr7:111118542:G:A	-0.3415	0.8231	5.5570	0.2868	0.0875	0.5901
chr7:111167898:A:G	-0.3415	0.8231	5.5570	0.2868	0.0875	0.5901
chr7:111265572:C:G	-0.3415	0.8231	5.5570	0.2868	0.0875	0.5901
chr7:111321049:G:T	-0.3415	0.8231	5.5570	0.2868	0.0875	0.5901
chr7:111429406:T:A	-0.3415	0.8231	5.5570	0.2868	0.0875	0.5901
chr7:111429408:T:C	-0.3415	0.8231	5.5570	0.2868	0.0875	0.5901
chr7:111440525:C:T	-0.3415	0.8231	5.5570	0.2868	0.0875	0.5901
chr7:99213680:A:G	-0.6250	0.6465	4.6330	0.3197	0.1232	0.3740

Table A9. PheWAS results for risk loci associated with mood disorders, matrix reasoning and vocabulary scores. *indicates lead SNP

SNP	MatrixReasoning beta	MatrixReasoning p	Vocabulary beta	Vocabulary p
chr16:62294564:T:C*	-0.8063	0.3527	-0.6001	0.6667
chr16:62294615:A:G	-0.8063	0.3527	-0.6001	0.6667
chr16:62295890:A:T	-0.8063	0.3527	-0.6001	0.6667
chr16:62296173:T:C	-0.8063	0.3527	-0.6001	0.6667
chr16:62296780:C:T	-0.8063	0.3527	-0.6001	0.6667
chr16:62299596:T:C	-0.8005	0.3207	-0.9060	0.4837
chr16:62303850:G:A	-0.2522	0.7496	-1.1293	0.3723
chr3:190970484:G:A	-1.2282	0.6883	-6.5601	0.1799
chr3:191487002:T:C	-1.2282	0.6883	-6.5601	0.1799
chr3:191544286:C:T	-1.2282	0.6883	-6.5601	0.1799
chr3:191594537:C:T	-1.2282	0.6883	-6.5601	0.1799
chr3:191609545:C:T	-1.2282	0.6883	-6.5601	0.1799
chr3:191850272:C:T	-1.2282	0.6883	-6.5601	0.1799
chr3:191857829:C:A*	-1.2282	0.6883	-6.5601	0.1799
chr3:191891184:G:A	-1.2282	0.6883	-6.5601	0.1799
chr3:191892660:A:G	-1.2282	0.6883	-6.5601	0.1799
chr3:191954871:C:G	-1.2282	0.6883	-6.5601	0.1799
chr3:192130608:G:A	-1.2282	0.6883	-6.5601	0.1799
chr3:192135614:A:G	-1.2282	0.6883	-6.5601	0.1799
chr3:192882063:C:T	-1.2282	0.6883	-6.5601	0.1799
chr3:192926052:A:G	-1.2282	0.6883	-6.5601	0.1799
chr3:193299693:G:A	-1.2282	0.6883	-6.5601	0.1799
chr5:74704443:A:G	3.1834	0.4169	5.8637	0.3509
chr5:74836279:A:G	3.1834	0.4169	5.8637	0.3509
chr5:74860694:G:C	3.1834	0.4169	5.8637	0.3509
chr5:76339511:G:A*	5.8204	0.2234	7.7661	0.3114
chr5:76357798:G:A	5.8204	0.2234	7.7661	0.3114
chr5:76506247:G:A	NA	NA	NA	NA
chr7:100749379:A:G	-0.3395	0.8964	-5.0387	0.2270
chr7:100808738:G:A	-0.3395	0.8964	-5.0387	0.2270
chr7:100833004:G:A	-0.3395	0.8964	-5.0387	0.2270
chr7:101858974:A:G	-0.3395	0.8964	-5.0387	0.2270
chr7:101868021:A:G	-0.3395	0.8964	-5.0387	0.2270
chr7:102076433:A:G	-0.3395	0.8964	-5.0387	0.2270
chr7:102093647:G:T	-0.0813	0.9735	-7.0342	0.0711
chr7:102116443:G:C	-0.0813	0.9735	-7.0342	0.0711

Table A9 continued.

SNP	MatrixReasoning beta	MatrixReasoning p	Vocabulary beta	Vocabulary p
chr7:102117558:G:A	-0.0813	0.9735	-7.0342	0.0711
chr7:102142747:C:A	-0.0813	0.9735	-7.0342	0.0711
chr7:102150913:T:C	-0.0813	0.9735	-7.0342	0.0711
chr7:102172436:G:C	-0.0813	0.9735	-7.0342	0.0711
chr7:102194960:G:C	-0.0813	0.9735	-7.0342	0.0711
chr7:102278021:A:C	-0.0813	0.9735	-7.0342	0.0711
chr7:102431830:T:C	-0.0813	0.9735	-7.0342	0.0711
chr7:102812541:G:A	-0.2701	0.8979	-6.7781	0.0432
chr7:102917271:C:T	-0.2701	0.8979	-6.7781	0.0432
chr7:102917556:A:C	-0.2701	0.8979	-6.7781	0.0432
chr7:103072929:C:T	-0.2701	0.8979	-6.7781	0.0432
chr7:103439754:G:A	-0.3828	0.8619	-7.6338	0.0293
chr7:103511937:C:T*	-0.0813	0.9735	-7.0342	0.0711
chr7:103884660:C:T	-0.0813	0.9735	-7.0342	0.0711
chr7:103942612:T:C	-0.0813	0.9735	-7.0342	0.0711
chr7:104731826:T:C	-0.0813	0.9735	-7.0342	0.0711
chr7:104859444:G:A	-0.0813	0.9735	-7.0342	0.0711
chr7:105621726:C:A	-0.3232	0.8887	-4.0867	0.2691
chr7:105673301:C:T	-0.3232	0.8887	-4.0867	0.2691
chr7:105695684:C:A	-0.3232	0.8887	-4.0867	0.2691
chr7:105707364:A:C	-0.3232	0.8887	-4.0867	0.2691
chr7:106011511:C:A	-0.3232	0.8887	-4.0867	0.2691
chr7:106057497:G:A	-0.3232	0.8887	-4.0867	0.2691
chr7:106067427:G:T	-0.3232	0.8887	-4.0867	0.2691
chr7:106067431:A:G	-0.3232	0.8887	-4.0867	0.2691
chr7:106077945:G:A	-0.3232	0.8887	-4.0867	0.2691
chr7:106083023:T:C	-0.3232	0.8887	-4.0867	0.2691
chr7:106151467:G:A	-0.3232	0.8887	-4.0867	0.2691
chr7:106187521:A:C	-0.3232	0.8887	-4.0867	0.2691
chr7:106247727:C:T	-0.3232	0.8887	-4.0867	0.2691
chr7:106437895:C:T	0.4002	0.8698	-1.7840	0.6487
chr7:106739799:C:G	0.4002	0.8698	-1.7840	0.6487
chr7:108163502:C:T	0.4002	0.8698	-1.7840	0.6487
chr7:108367937:C:T	0.4002	0.8698	-1.7840	0.6487
chr7:108637911:C:T	0.4002	0.8698	-1.7840	0.6487
chr7:108692182:C:T	0.4002	0.8698	-1.7840	0.6487
chr7:108724884:C:T	0.4002	0.8698	-1.7840	0.6487
chr7:108776155:G:A	0.4002	0.8698	-1.7840	0.6487

Table A9 continued.

SNP	MatrixReasoning beta	MatrixReasoning p	Vocabulary beta	Vocabulary p
chr7:108820968:C:A	0.4002	0.8698	-1.7840	0.6487
chr7:109680191:T:C	-1.1401	0.6613	-2.6831	0.5200
chr7:109680628:G:A	-1.1401	0.6613	-2.6831	0.5200
chr7:109691709:G:C	-1.1401	0.6613	-2.6831	0.5200
chr7:109702206:A:G	-1.1401	0.6613	-2.6831	0.5200
chr7:109868482:A:G	-1.1401	0.6613	-2.6831	0.5200
chr7:110148126:C:T	-0.8108	0.7559	-2.4083	0.5645
chr7:111118542:G:A	-0.3488	0.9014	-4.1346	0.3592
chr7:111167898:A:G	-0.3488	0.9014	-4.1346	0.3592
chr7:111265572:C:G	-0.3488	0.9014	-4.1346	0.3592
chr7:111321049:G:T	-0.3488	0.9014	-4.1346	0.3592
chr7:111429406:T:A	-0.3488	0.9014	-4.1346	0.3592
chr7:111429408:T:C	-0.3488	0.9014	-4.1346	0.3592
chr7:111440525:C:T	-0.3488	0.9014	-4.1346	0.3592
chr7:99213680:A:G	-0.3395	0.8964	-5.0387	0.2270

Table A10. Number of carriers by disease status for the four risk loci. A) Locus 3q28/29 B) Locus 5q13
 C) Locus 7q22 D) Locus 16q11

A)

	Unaffected	Affected
C/C	1385	190
C/A	2	7

B)

	Unaffected	Affected
G/G	1385	189
G/A	2	8

C)

	Unaffected	Affected
C/C	1353	170
C/T	34	27

D)

	Unaffected	Affected
T/T	686	64
T/C	594	92
C/C	107	41

Figure A1. Scatterplots of the first 10 PCs for the Lancaster OOA cluster, coded by array technology. None of the PCs appear to correspond to array technology.

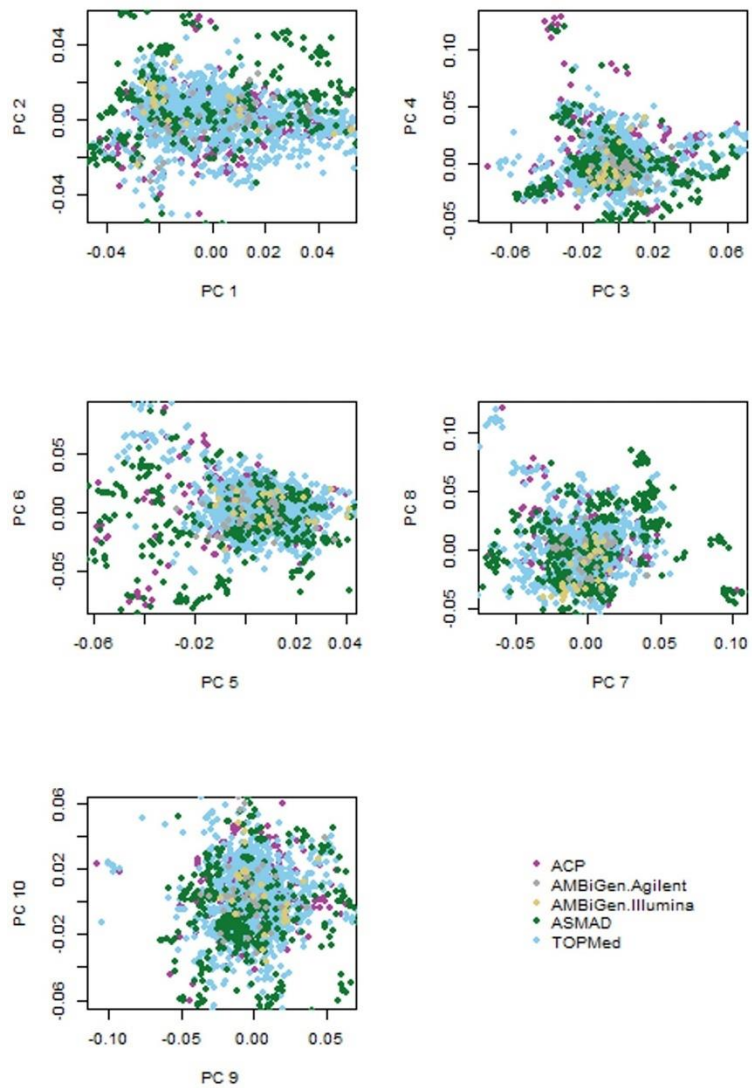


Figure A2. Hierarchical clustering diagram OOA carriers of the 3q28/29 haplotype. Clustering was done using the empirically calculated kinship matrix for all 1672 Lancaster OOA individuals. a=ACP, b=AMBiGen, c=ASMAD, d=TOPMed

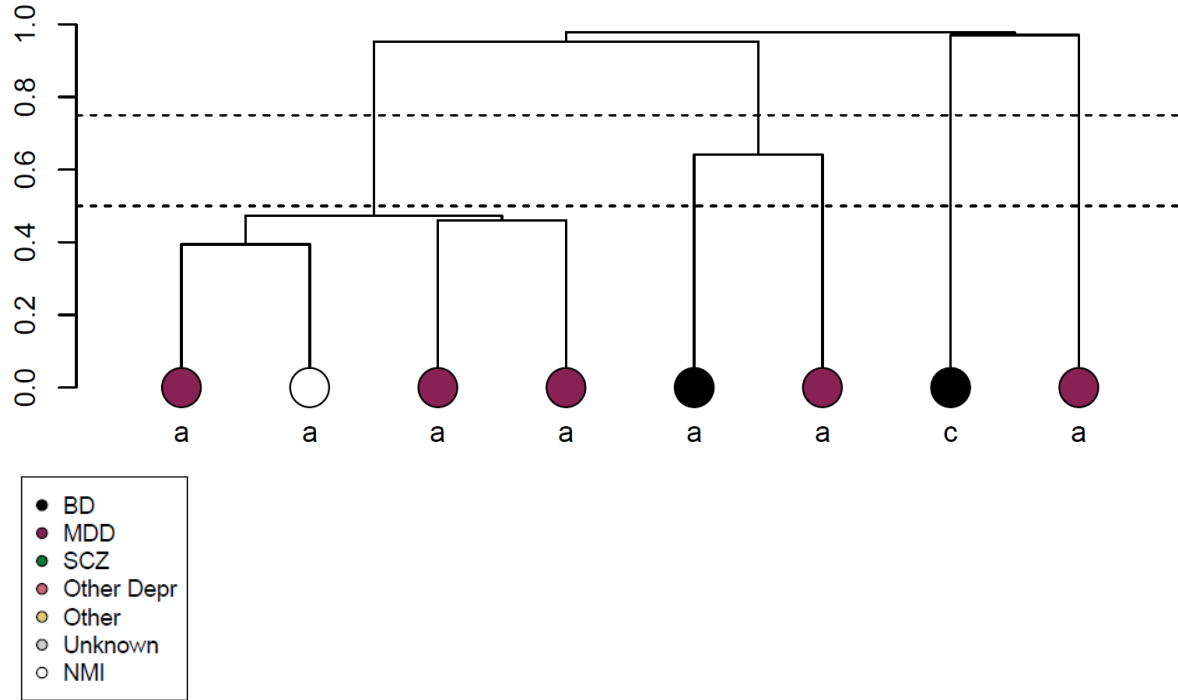


Figure A3. Pedigree of two ASMAD families with the 5q13 locus. The genotype for the lead 5q13 SNP is below each individual; missing genotypes indicate an ungenotyped individual. Question marks indicate an unknown mood disorder phenotype.

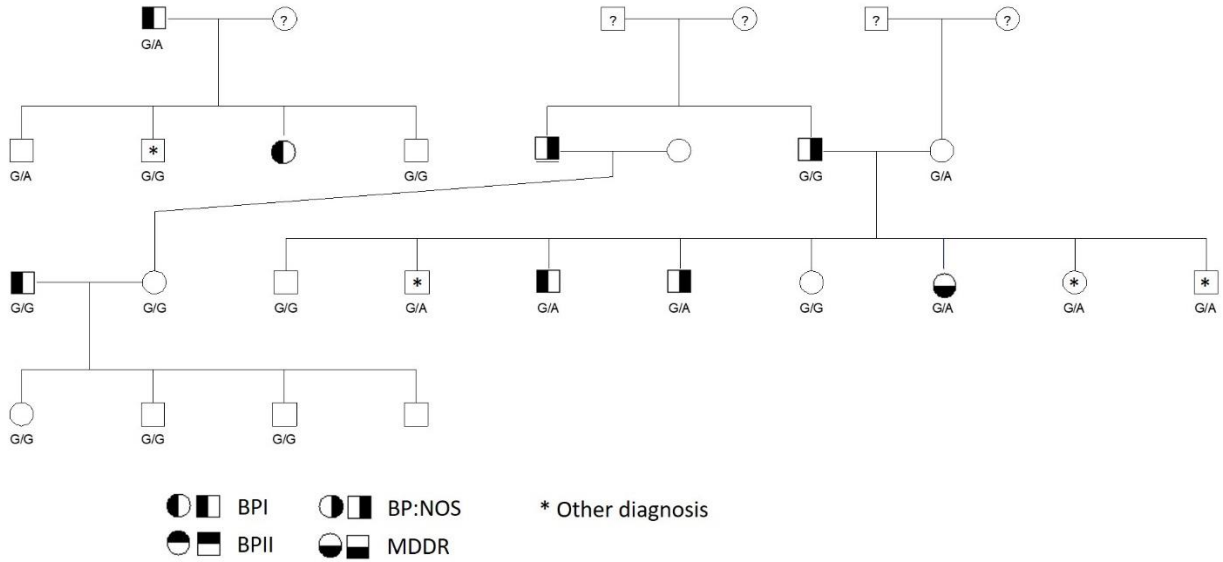


Figure A4. Hierarchical clustering diagram OOA carriers of the 5q13 haplotype. Clustering was done using the empirically calculated kinship matrix for all 1672 Lancaster OOA individuals. a=ACP, b=AMBiGen, c=ASMAD, d=TOPMed

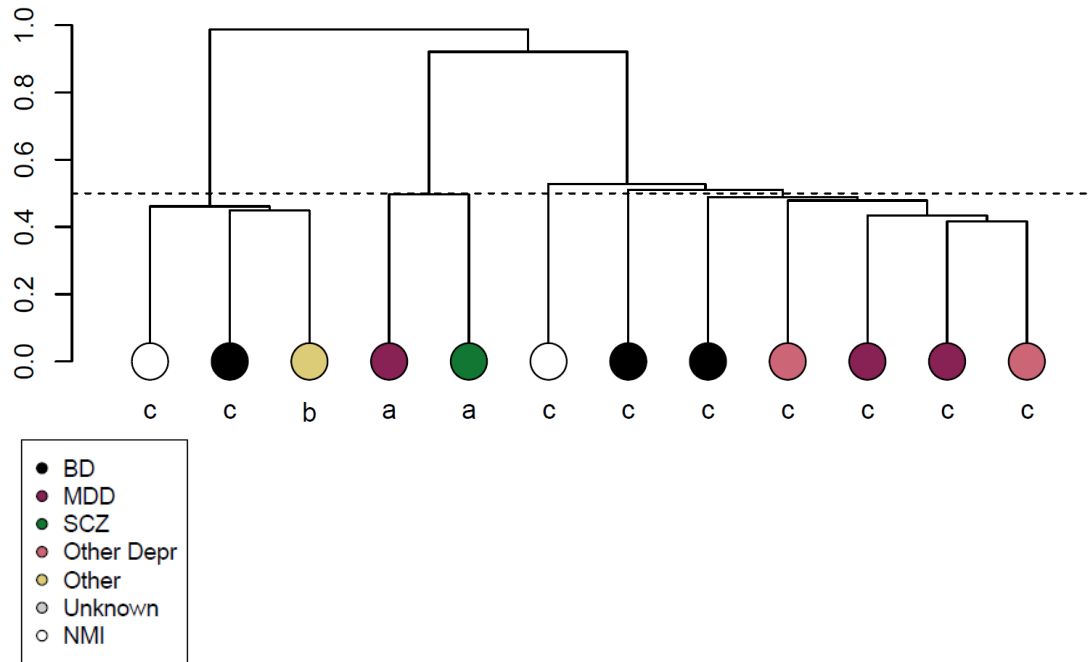
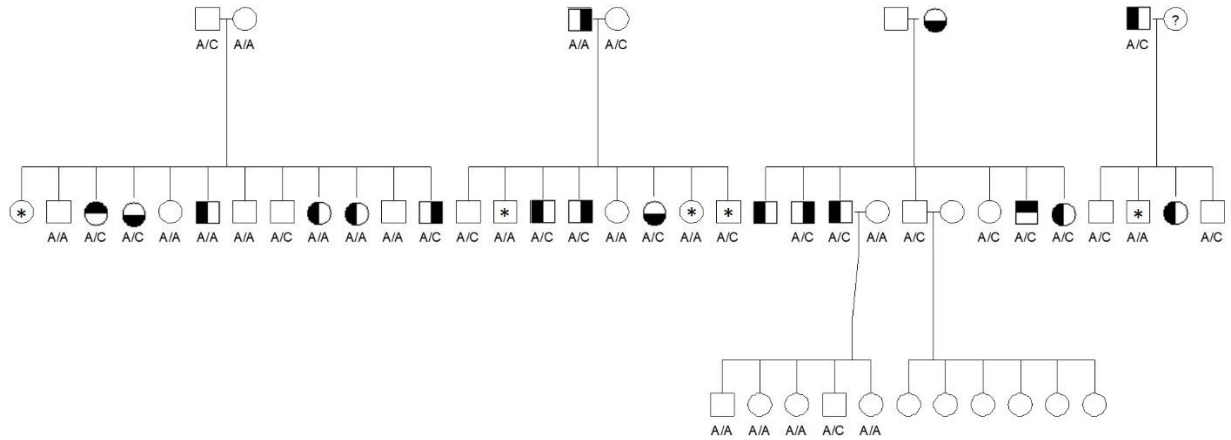
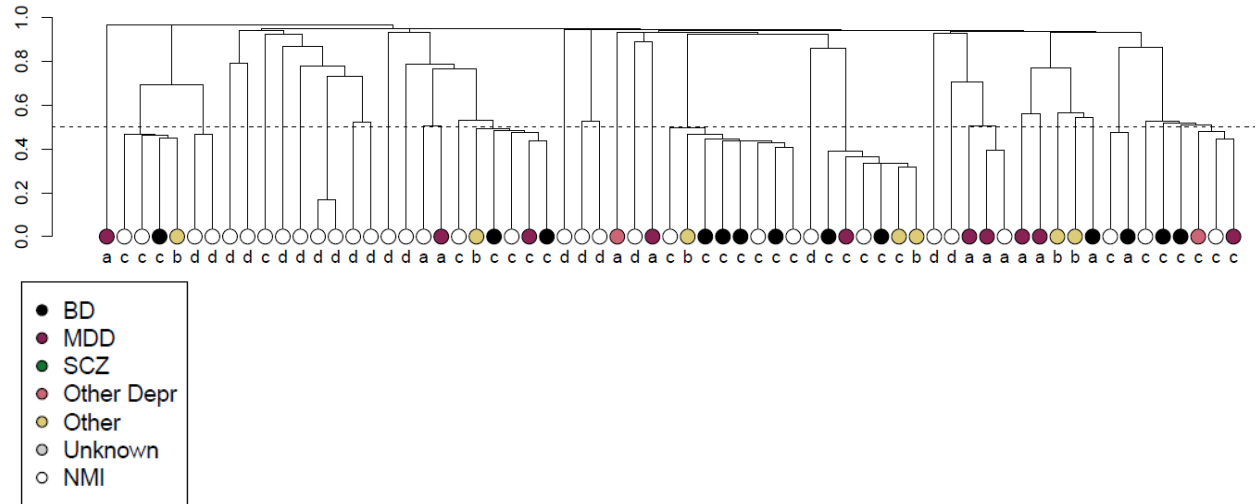


Figure A5. Pedigree of four ASMAD families with the 7q22 locus. The genotype for the lead 5q13 SNP is below each individual; missing genotypes indicate an ungenotyped individual. Question marks indicate an unknown mood disorder phenotype.



BPI
 BP:NOS
 * Other diagnosis
 BPII
 MDDR

Figure A6. Hierarchical clustering diagram OOA carriers of the 7q22 haplotype. Clustering was done using the empirically calculated kinship matrix for all 1672 Lancaster OOA individuals. a=ACP, b=AMBiGen, c=ASMAD, d=TOPMed



References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* (2018).
2. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing (2013).
3. Kessler RC, Borges G, Walters EE. Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. *Arch Gen Psychiatry* **56**, 617-626 (1999).
4. Solé B, *et al.* Cognitive Impairment in bipolar disorder: treatment and prevention strategies. *Int J Neuropsychopharmacology* **20**, 670–680 (2017).
5. Clark L, Sahakian BL. Cognitive neuroscience and brain imaging in bipolar disorder. *Dialogues Clin Neuroscience* **10**, 153-163 (2008).
6. Sullivan PF, M Neale, KS Kendler. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry*. **157**,1552–1562 (2000).
7. Kendler K., M Gatz, CO Gardner, NL Pedersen. A Swedish national twin study of lifetime major depression. *Am. J. Psychiatry*. **163**,109–114 (2006).
8. Smoller JW, CT Finn. Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet*. **123C**,48–58 (2003).
9. Barnett JH, JW Smoller. The genetics of bipolar disorder. *Neuroscience*. **164**,331-343 (2009). doi:10.1016/j.neuroscience.2009.03.080
10. Rasic D, *et al.* Risk of mental illness in offspring of parents with schizophrenia, bipolar

- disorder, and major depressive disorder: a meta-analysis of family high-risk studies. *Schizophr Bull.* **40**, 28-38 (2014).
11. Wray NR, Ripke S, *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genetics* **50**, 668-681 (2018).
 12. Howard DM *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* 2019, 22:343-352.
 13. Stahl EA, *et al.* Genome-wide association study identifies 30 Loci Associated with Bipolar Disorder. *Nat Genetics* **51**, 793-803 (2019).
 14. Mullins N, *et al.* Genome-wide association study of over 40,000 bipolar disorder cases provides novel biological insights.
<https://www.medrxiv.org/content/10.1101/2020.09.17.20187054v1>
 15. Levey DF, *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat Neurosci* **24**, 954-936 (2021).
 16. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469-1482.e11 (2019).
 17. Ferreira MAR, *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet.* **40**, 1056-8 (2008).
 18. Lee SH, *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet.* **45**, 984-94 (2013).

19. Ament SA, *et al.* Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc Natl Acad Sci USA*. **112**, 3576-81 (2015).
20. Cruceanu C, *et al.* Family-based exome-sequencing approach identifies rare susceptibility variants for lithium-responsive bipolar disorder. *Genome* **56**, 634–40 (2013).
21. Goes FS, *et al.* Exome Sequencing of Familial Bipolar Disorder. *JAMA psychiatry* **73**, 590–7 (2016).
22. Strauss KA, *et al.* A population-based study of KCNH7 p.Arg394His and bipolar spectrum disorder. *Hum. Mol. Genet.* (2014). doi:10.1093/hmg/ddu335
23. Georgi B, *et al.* Genomic view of bipolar disorder revealed by whole genome sequencing in a genetic isolate. *PLoS Genet.* **10**, e1004229 (2014).
24. Purcell SM, *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 506, 185-90 (2014).
25. Genovese G, *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci.* 19, 1433-1441 (2016).
26. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
27. McKusick VA, Hostetler JA, Egeland JA. Genetic Studies of the Amish: Background and Potentialities. *Bulletin of the Johns Hopkins Hospital* **115**, 203–222 (1964).
28. McKusick VA, *et al.* Dwarfism in the Amish I. The Ellis-van Creveld Syndrome. *Bull. Johns Hopkins Hosp.* **115**, 306–36 (1964).
29. McKusick VA, Eldridge R, Hostetler JA, Ruangwit U, Egeland JA. Dwarfism in the Amish, Part II: Cartilage-Hair Hypoplasia. *Bulletin of the Johns Hopkins Hospital* **116**,

- 285–326 (1965).
30. Pollin TI, *et al.* A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection (2008) *Science* **322**, 1702-1705 (2008).
 31. Shen H, *et al.* Apolipoprotein B (APOB) R3500Q is common in the Old Order Amish and is a major cause of increased low density lipoprotein cholesterol concentrations and coronary artery calcification. *Arch. Int. Med.* **170**, 1850-1855 (2010).
 32. Strauss KA, Puffenberger EG. Genetics, medicine, and the Plain people. *Ann Rev Genomics Hum Genet* **10**, 513-536 (2009).
 33. Hostetler JA. Amish Society, fourth ed. Johns Hopkins University Press (1993).
 34. Krahn C , Bender HS, Friesen JJ. "Migrations." *Global Anabaptist Mennonite Encyclopedia Online*. 1989. Web. 13 Nov 2020. <https://gameo.org/index.php?title=Migrations&oldid=143668>.
 35. Smith CA. *The Mennonites: A Brief History of their Origins and Later Development in Both Europe and America*. Hard Press Publishing (2012).
 36. Glahn DC, *et al.* Rediscovering the values of families for psychiatric genetics research. *Mol Psychiatry* (2018). doi: 10.1038/s41380-018-0073
 37. Lopes FL, *et al.* Finding Rare, Disease-Associated Variants in Isolated Groups: Potential Advantages of Mennonite Populations. *Hum. Biol.* **88**, 109–120 (2016).
 38. Hou L, *et al.* Amish revisited: next generation sequencing studies of psychiatric disorders among the Plain people. *Trends Genet.* **29**, 412–418 (2013).
 39. Stefansson H, *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet.* **71**, 877-92 (2002).

40. Egeland JA, *et al.* Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* **325**, 783-787 (1987).
41. Kelsoe JR, *et al.* Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature* **342**, 238-243 (1989).
42. Ginns EI, *et al.* A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. *Nature Genetics* **12**, 431-435 (1996).
43. Ginns EI, *et al.* A genome-wide search for chromosomal loci linker to mental health wellness in relatives at high risk for bipolar affective disorder among the Old Order Amish. *PNAS* **95**,15531-15536 (1998).
44. Kember RL, *et al.* Copy number variants encompassing Mendelian disease genes in a large multigenerational family segregating bipolar disorder. *BMC Genet* **15**, 16-27 (2015).
45. Harvard Medical School, 2007. National Comorbidity Survey (NCS). (2017, August 21). Retrieved from <https://www.hcp.med.harvard.edu/ncs/index.php>.
46. Kupfer DJ. Depression and the new DSM-5 classification. *Medicographia* **34**, 521-525 (2014).
47. American Psychiatric Association. [Depressive disorders: DSM-5® selections](#). Arlington, VA: American Psychiatric Publishing (2016).
48. Centers for Disease Control “Data and Statistics on Children’s Mental Health”, 2018
49. Cuthbert BN. Research Domain Criteria: toward future psychiatric nosologies. *Dialogues Clin Neurosci.* 2015, 17:89-97.

50. <https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/index.shtml>
51. Woody ML, Gibb BE. Integrating NIMH Research Domain Criteria (RDoC) into depression research. *Current Opinion in Psychology* 2015, 4:6-12.
52. Cosgrove VE, Kelsoe JR, Suppes T. Toward a valid animal model of bipolar disorder: how the Research Domain Criteria help bridge the clinical-basic science divide. *Biological Psychiatry* 2016, 79:62-70.
53. Upthegrove R, S Marwaha, M Birchwood. Depression and schizophrenia: cause, consequence, or trans-diagnostic issue? *Schizophr Bul.* **43**, 240-244 (2017).
54. Merikangas KR, *et al.* Future of genetics of mood disorders research. *Biol Psych* 2002, 52:457-477.
55. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–86 (2017).
56. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med.* **363**, 166–76 (2010).
57. Manolio T, Collins F, Cox N, Goldstein D, Hindorff L, Hunter D, *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
58. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb.* **52**, 399–433 (1918).
59. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–15 (2014).
60. Sanders SJ. First glimpses of the neurobiology of autism spectrum disorder. *Curr Opin Genet Dev.* **33**, 80–92 (2015).

61. Nurnberger JI, DL Koller, J Jung, *et al.* Identification of pathways for Bipolar Disorder. *JAMA Psychiatry* **71**, 657-664 (2014).
62. Carboni L, Pischedda F, *et al.* Depression-associated gene *Negr1-Fgfr2* pathway is altered by antidepressant treatment. *Cell* 2020, 9:1818. doi:10.3390/cells9081818
63. Nolt SM. The emergence of Amish genetic studies: a brief history of collaboration and reciprocity. *Journal of Plain Anabaptist Communities* **1** (2020).
DOI: <http://dx.doi.org/10.18061/jpac.v1i1.7659>
64. Jackson CE, Carey JH. Progressive Muscular Dystrophy: Autosomal Recessive Type. *Pediatrics* **28**, 77–84 (1961).
65. Martin PH. PKU: A Casefinding Project. *Children* **11**, 123–124 (1964).
66. Francomano CA. “Victor A. McKusick and Medical Genetics among the Amish.” In *Victor McKusick and the History of Medical Genetics*, edited by Krishna R. Dronamraju and Clair A. Francomano, 119–130. Springer (2012).
67. McKusick VA, Hostetler JA, Egeland JA, Eldridge R. The Distribution of Certain Genes in the Old Order Amish. *Cold Spring Harbor Symposia on Quantitative Biology* **29**, 99–114 (1964). <https://doi.org/10.1101/SQB.1964.029.01.015>
68. McKusick VA, HE Cross. Ataxia-telangiectasia and Swiss-type agammaglobulinemia: Two genetic disorders of the immune mechanism in related Amish sibships. *JAMA* **195**, 739-745 (1966).
69. Strauss KA, Puffenberger EG. Genetics, medicine, and the Plain people. *Ann Rev Genomics Hum Genet* **10**, 513-536 (2009).
70. Payne M, Rupar CA, Siu GM, Siu VM. Amish, Mennonite, and Hutterite Genetic

- Disorder Database. *Paediatr Child Health* **16**, e23-e24 (2011).
71. Shuldiner AR, O'Connell JR, Bliden KP, Gandhi A, Ryan K, Horenstein RB, *et al.* Association of Cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* **302**, 849 (2009).
72. Albert JS, Yerges-Armstrong LM, Horenstein RB, Pollin TI, Sreenivasan UT, Chai S, *et al.* Null mutation in hormone-sensitive lipase gene and risk of type 2 diabetes. *N Engl J Med* **370**, 2307–2315 (2014).
73. Cummings AC, Jiang L, Velez Edwards DR, McCauley JL, Laux R, McFarland LL, *et al.* Genome-wide association and linkage study in the Amish detects a novel candidate late-onset Alzheimer disease gene. *Ann Hum Genet* **76**, 342–351 (2012).
74. Edwards DRV, Gilbert JR, Hicks JE, Myers JL, Jiang L, Cummings AC, *et al.* Linkage and association of successful aging to the 6q25 region in large Amish kindreds. *Age (Dordr)* **35**, 1467–1477 (2013).
75. Davis MF, Cummings AC, D'Aoust LN, Jiang L, Velez Edwards DR, Laux R, *et al.* Parkinson disease loci in the mid-western Amish. *Hum Genet* **132**, 1213–1221 (2013).
76. Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, *et al.* Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* **382**, 951–962 (2013).
77. Telatar M, Teraoka S, Wang Z, Chun HH, Liang T, Castellvi-Bel S, *et al.* Ataxia-Telangiectasia: Identification and detection of founder-effect mutations in the ATM gene

- in ethnic populations. *Am J Hum Genet* **62**:86–97 (1998).
78. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, *et al.* Recessive symptomatic focal epilepsy and mutant Contactin-Associated Protein-like 2. *N Engl J Med* **354**, 1370–1377 (2006).
79. Alarcón M, Abrahams BS, Stone JL, Duvall JA, Perederiy J V., Bomar JM, *et al.* Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet* **82**, 150–159 (2008).
80. Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BHW, *et al.* CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry*. **13**, 261–266 (2008).
81. Bakkaloglu B, O’Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, *et al.* Molecular cytogenetic analysis and resequencing of Contactin Associated Protein-Like 2 in Autism Spectrum Disorders. *Am J Hum Genet* **82**, 165–173 (2008).
82. Rossi E, Verri AP, Patricelli MG, Destefani V, Ricca I, Vetro A, *et al.* A 12Mb deletion at 7q33-q35 associated with autism spectrum disorders and primary amenorrhea. *Eur J Med Genet* **51**, 631–638 (2008).
83. Belloso JM, Bache I, Guitart M, Caballin MR, Halgren C, Kirchhoff M, *et al.* Disruption of the CNTNAP2 gene in a t(7;15) translocation family without symptoms of Gilles de la Tourette syndrome. *Eur J Hum Genet* **15**, 711–713 (2007).
84. Kember RL, *et al.* Genetic pleiotropy between mood disorders, metabolic, and endocrine

- traits in a multigenerational pedigree. *Transl Psychiatry* **8**, 218 doi: 10.1038/s41398-018-0226-3 (2018).
85. Kuo P.H., Chuang L.C., Liu J.R., Liu C.M., Huang M.C., Lin S.K., Sunny Sun H., Hsieh M.H., Hung H., Lu R.B. Identification of novel loci for bipolar I disorder in a multi-stage genome-wide association study. *Prog. Neuropsychopharm. Biol. Psychiat.* 2014;51:58–64.
86. Wang X, *et al.* Association of *KCNH7* polymorphisms and individual responses to risperidone treatment in schizophrenia. *Front Psychiatry* (2019).
<https://doi.org/10.3389/fpsy.2019.00633>
87. First MB, Williams JBW, Karg RS, Spitzer RL. “Structured clinical interview for DSM-5 – Research Version (SCID-5-RV). Arlington, VA, American Psychiatric Association (2015).
88. Wechsler Abbreviated Scale of Intelligence. Pearson Assessments.
89. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation (1996)
90. Chiappelli J, KL Nugent, K Thangavelu, K Searcy, LE Hong. Assessment of trait and state aspects of depression in schizophrenia. *Schizophr Bull* **40**, 132-142 (2014).
91. Bruce HA, *et al.* Clinical and genetic validity of quantitative bipolarity. *Trans Psychiatry* **9**, 228 (2019).
92. DIGS Version 4.0. <https://www.nimhgenetics.org/resources/clinical-instruments/digs/list-of-digs>

93. Nurnberger Jr JI, *et al.* Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* **51**, 849-859 (1994).
94. Das S, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
95. Endicott J, Spitzer RL. "A diagnostic interview: the schedule for affective disorders and schizophrenia." *Arch. Gen. Psychiatry* **35**, 837-844 (1978).
96. *Diagnostic and statistical manual of mental disorders: DSM-IV*. Washington, DC: American Psychiatric Association (1994).
97. Kessler MD, *et al.* De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. *Proc Natl Acad Sci USA* **117**, 2560-2569 (2020).
98. Chang CC, *et al.* "Second-generation PLINK: rising to the challenge of larger and richer datasets." *GigaScience* **4** (2015)
99. PLINK v1.9. Authors: Shaun Purcell, Christopher Chang. URL: www.cog-genomics.org/plink/1.9
100. <https://imputation.sanger.ac.uk/>
101. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2019).
102. Fernandez-Pujals AM, *et al.* Epidemiology and heritability of Major Depressive Disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).

103. Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study
104. Hulshoff Pol HE, van Baal GC, Schnack HG, Brans RG, van der Schot AC, Brouwer RM, van Haren NE, Lepage C, Collins DL, Evans AC, Boomsma DI, Nolen W, Kahn RS. Overlapping and segregating structural brain abnormalities in twins with schizophrenia or bipolar disorder. *Archives of general psychiatry*. 2012;69:349-359.
105. Kang HM, *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348-354 (2010).
106. Hellevik O. Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity* **43**, 59-74 (2007).
107. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for Biobank-scale data. *GigaScience* **8** (2019)
108. Quinlan AR. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protocols Bioinformatics* **47** (2014).
109. Lee JJ, Wedow R, Okbay A, *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112–1121 (2018).
110. Gandal MJ, *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**:eaat8127 (2018).
111. Singh T, BM Neale, MJ Daly. Exome sequencing identified rare coding variants in 10 genes which confer substantial risk for schizophrenia.

<https://doi.org/10.1101/2020.09.18.20192815>

112. Satterstrom FK, *et al.* Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568-584, e23 (2020).
113. Abrahams BS, *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **36**, doi: 10.1186/2040-2392-4-36 (2013).
114. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. 2015;11:e1004219.
115. Hasin N, Riggs LM, Shekhtman T, Ashworth J, Lease R, Oshone RT, *et al.* A rare variant in D-amino acid oxidase implicates NMDA receptor signaling and cerebellar gene networks in risk for bipolar disorder. *MedRxiv*. 2021:2021.06.02.21258261.
116. Casella AM, Colantuoni C, Ament SA. Regulome-wide association study identifies enhancer properties associated with risk for schizophrenia. *BioRxiv*. 2021:2021.06.14.448418.
117. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* (80-). 2018;362:eaat8464.
118. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat Genet* **50**, 6–11 (2018).
119. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, Van Kogelenberg M, *et al.* Genetic diagnosis of developmental disorders in the DDD study: A scalable analysis of

- genome-wide research data. *Lancet* **385**,1305–1314 (2015).
120. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
 121. Pirooznia M, Wang T, Avramopoulos D, Valle D, Thomas G, Hugarir RL, et al. SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* **28**, 897–899 (2012).
 122. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
 123. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, *Complex Syst* (2006).
 124. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
 125. Ganjgahi H, et al. Fast and powerful heritability inference for family-based neuroimaging studies. *Neuroimage* **115**, 256-268 (2015).
 126. Taliun D, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
 127. Mulle JG. The 3q29 deletion confers >40-fold increase in risk for schizophrenia. *Mol Psychiatry* **20**, 1028-1029 (2015).
 128. Clayton-Smith J, Giblin C, Smith RA, Dunn C, Willatt L. Familial 3q29 microdeletion

- syndrome providing further evidence of involvement of the 3q29 region in bipolar disorder. *Clin Dysmorphol* **19**, 128-132 (2010).
129. Guo X, Ge T, Xia S, Wu H, Colt M, Xie X, et al. *Atp13a5* Marker Reveals Pericytes of the Central Nervous System in Mice. *SSRN Electron J.* (2021).
<https://doi.org/10.2139/SSRN.3881359>.
 130. Dunn AR, Stout KA, Ozawa M, Lohr KM, Hoffman CA, Bernstein AI, et al. Synaptic vesicle glycoprotein 2C (SV2C) modulates dopamine release and is disrupted in Parkinson disease. *Proc Natl Acad Sci* **114**, E2253–E2262 (2017).
 131. Larsen E, Menashe I, Ziats MN, Poreanu W, Packer A, Banerjee-Basu S. A systematic variant annotation approach for ranking genes associated with autism spectrum disorders. *Mol Autism* **7**, 44 (2016).
 132. Ramdzan ZM, Nepveu A. CUX1, a haploinsufficient tumour suppressor gene overexpressed in advanced cancers. *Nature Reviews Cancer* **14**, 673-682 (2014).
 133. Gillingham AK, Pfeifer AC, Munro S. CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a Golgi membrane protein related to Giantin. *Mol Biol Cell* **13**, 3761-3774 (2002).
 134. Osterrieder A, et al. Stacks off tracks: a role for the golgin AtCASP in plant endoplasmic reticulum-Golgi apparatus tethering. *Journal of Experimental Botany* **68**, 3339-3350 (2017).
 135. Lievens PM, Tufarelli C, Donady JJ, Stagg A, Neufeld EJ. CASP, a novel, highly conserved alternative-splicing product of the CDP/cut/cux gene, lacks cut-repeat and homeo DNA-binding domains, and interacts with full-length CDP in vitro. *Gene* **197**, 73-

- 81 (1997).
136. Sasayama D, *et al.* Possible association of *CUX1* gene polymorphisms with antidepressant response in major depressive disorder. *The Pharmacogenomics Journal* **13**, 354-358 (2013).
 137. Doan RM *et al.* Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341-354.e12 (2016).
 138. Cubelos B, *et al.* *Cux1* and *Cux2* regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. *Neuron* **66**, 523-535 (2010).
 139. Ning L, Zhao C-T, Wang Y, Yuan X-B. The transcription factor *Cux1* regulates dendritic morphology of cortical pyramidal neurons. *PLOS One* <https://doi.org/10.1371/journal.pone.0010596> (2010).
 140. Cubelos B, Briz CG, Esteban-Ortega GM, Nieto M. *Cux1* and *Cux2* selectively target basal and apical dendritic compartments of layer II-III cortical neurons. *Developmental Neurobiology* **75**, 163-172 (2014).
 141. Rodríguez-Tornos FM, *et al.* *Cux1* enables interhemispheric connections of layer II/III neurons by regulating Kv1-dependent firing. *Neuron* **89**, 494-506 (2015).
 142. Topka S, Glassmann A, Weisheit G, Schüller U, Schilling K. The transcription factor *Cux1* in cerebellar granule cell development and medulloblastoma pathogenesis. *Cerebellum* **13**, 698-712 (2014).
 143. Schmahmann JD, Caplan D. Cognition, emotion and the cerebellum. *Brain* **2**, 290-292 (2006).
 144. Lupo M, Siliciano L, Leggio M. From cerebellar alterations to mood disorders: a

- systematic review. *Neuroscience & Behavioral Reviews* **103**, 21-28 (2019).
145. Goes FS, *et al.* Sex-specific association of the Reelin gene with bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 549-553 (2010).
146. Fatemi SH, Earle JA, McMenemy T. Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression. *Mol Psych* **5**, 654-663 (2000).
147. Teixeira CM, *et al.* Overexpression of Reelin prevents the manifestation of behavioral phenotypes related to schizophrenia and bipolar disorder. *Neuropsychopharmacology* **36**, 2395-2405 (2011).
148. Lakatosova S, Ostatnikova D. Reelin and its complex involvement in brain development and function. *Int J Biochem Cell Biol* **44**, 1501-1504 (2012).
149. Ferrere A, Vitalis T, Gingras H, Gaspar P, Cases O. Expression of Cux-1 and Cux-2 in the developing somatosensory cortex of normal and barrel-defective mice. *The Anatomical Record Part A* **288A**, 158-165 (2006).