January 2023

Genome-Wide Association Studies Of Depression And Alzheimer's: Identifying Pleiotropic Snps

Yara Ahmed Kamel Abdelsalam
yara.abdelsalam@yale.edu

GENOME-WIDE ASSOCIATION STUDIES OF DEPRESSION AND ALZHEIMER'S: IDENTIFYING PLEIOTROPIC SNPS

By

Yara A. Abdelsalam

A Thesis Presented to

The Faculty of the Department of Chronic Disease Epidemiology

In Candidacy for the Degree of Master of Public Health

Under the Supervision of

Dr. Andrew DeWan

and

Dr. Cecilia Dao

Yale School of Public Health

Yale University

May 2023

Abstract:

Background: Alzheimer's disease (AD) has been rising globally, making it important to understand risk factors. Depression has been identified as a risk factor in epidemiological studies and meta-analysis. The possible mechanism could be pleiotropic. This study sought to identify the pleiotropic SNPs associated with depression and AD so that can understand the relationship between the two diseases more evident.

Methods: The sample was from the UK biobank population, using the White European subpopulation. Depression was defined using self-reported depression status and electronic medical records. Proxy AD was defined by biological mother and father's AD status. Two univariate association analysis was conducted using a linear mixed model in REGENIE. A p-value comparison was conducted for both analyses. Linkage disequilibrium (LD) clumping was conducted using a four MB region surrounding the index variant and an r2 threshold of 0.2.

Results: The study had 406,7394 individuals in the depression sample and 409,704 individuals in the AD sample. The univariate analysis for depression identified five genome-wide significant SNPs, and the univariate analysis for AD identified 1,319 genome-wide significant SNPs. The p-value comparison found that 36 genome-wide significant SNPs for AD were genome-wide nominal for depression, with SNPs chromosome 8 having p-values for depression less than 0.001. The LD clumping identified a region in chromosome 19 associated with genes *APOE* and *TOMM40*. The leading SNP was rs2075650 and associated with the two phenotypes.

Conclusion: The study found SNPs associated with depression and AD. Furthermore, SNP rs2075650 on chromosome 19 seems promising in understanding the relationship between depression and AD. Further studies must be conducted to understand the association that can lead to prevention methods.

Acknowledgments:

I thank Dr. Andrew T. DeWan for his guidance and mentorship throughout this project. I also thank Dr. Paulina Tolosa-Tort for her support and mentorship throughout this project. I want to thank the other members of Dr. DeWan's lab for their support and help throughout this project. Finally, I would like to thank the individuals who participated in UKB, as their participation has made this project possible.

Table of Contents:

List of Figures:

List of Tables:

Introduction:

Alzheimer's Disease and Depression:

Dementia is a neurodegenerative disease that is associated with aging. The crucial feature of dementia is cognitive impairment. Dementia also has neuropsychiatric symptoms, including depression. (Cerejeira et al., 2012). One of the most common types of dementia is Alzheimer's disease (AD). The apolipoprotein E gene (APOE) region has been associated with an increased risk of Alzheimer's (Saunders et al., 1993). In 2019, it was estimated that there were 57.4 million cases of dementia globally, and it is expected to rise to 152.8 million cases by 2050 (Nichols et al., 2022). Over the past two decades, depression has been seen as a risk factor for Alzheimer's (Ownby et al., 2006). This includes depression that had appeared in earlier life and is not part of the prodromal stage of dementia.

Depression is one of the most common psychiatric disorders worldwide, affecting 241 million in 2017 (Santomauro et al., 2021). It is characterized by persistent sadness, anhedonia, and other symptoms that last more than two weeks (APA, 2013). Symptoms of depression also affect cognitive function, such as concentration, memory, and executive function (APA, 2013). Depression is considered a polygenic phenotype, and some studies estimate its heritability to be between 28% to 44% (Fernandez-Pujals et al., 2015). Furthermore, the multiple subtypes of depression differing in length, severity, frequency, and other features can cause phenotypic heterogeneity that makes the genetic epidemiology of depression a challenging topic to study (Jermy et al., 2021).

Association Between Depression and Alzheimer's Disease:

Depression has been identified as a risk factor for dementia in multiple association studies. Participants with depression had an adjusted hazard ratio (HR) of 1.83 (95% CI, 1.80-1.87) in a Danish population-based cohort study (Katon et al., 2015). A 2019 study using a Swedish population-based cohort that participants with depression 20 years or more before a

dementia diagnosis had an adjusted odds ratio (OR) of 1.60 (95% CI, 1.27–1.98) (Holmquist et al., 2020). Participants with depression 5–9.9 years before a dementia diagnosis had an adjusted OR of 1.85 (95% CI, 1.68–2.05) (Holmquist et al., 2020). In a meta-analysis conducted in 2006, the pooled OR was 2.03 (95% CI, 1.73-2.38) for case-control studies, while the pooled OR was 2.02 (95% CI, 1.80-2.26) for cohort studies (Ownby et al., 2006). A recent meta-analysis found a pooled OR of 1.91 (95% CI, 1.72-2.12) across 36 studies (Kuring et al., 2020). This indicates an association between depression and dementia, with plausible underlying mechanisms.

Mechanisms for Depression as a Risk Factor for Alzheimer's Disease:

There have been a few hypothesized underlying mechanisms between depression as a risk factor for dementia. One mechanism is that depression disrupts the hypothalamus-pituitary-adrenal (HPA) axis, increasing glucocorticoids. This increase is associated with damage in the hippocampus, which can lead to loss of hippocampal atrophy and cognitive decline (Byers & Yaffe, 2011; Dafsari & Jessen, 2020). Glucocorticoid is released as a stress response during depression and is also associated with amyloid plaque formation in people with AD (Byers & Yaffe, 2011; Dafsari & Jessen, 2020). Amyloid plaques are protein deposits in the brain's gray matter that disrupt neuron function (National Institutes on Aging, 2017). Another mechanism is the imbalance of neurotransmitters noradrenaline and serotonin, which is caused by depression and can lead to amyloid plaques seen among people with AD (Dafsari & Jessen, 2020). Given that these mechanisms are biological, depression and AD plausibly have a genetic association.

Review of the literature:

Recent genetic epidemiology studies have looked at the relationship between depression and dementia. A higher polygenic risk score (PRS) for major depressive disorder (MDD) was significantly associated with slower reactions and scoring lower on verbal-numerical reasoning, both signs of cognitive decline (Hagenaars et al., 2016). Higher PRS for MDD was also

associated with a family history of dementia in the Generation Scotland's Scottish Family Health Study (GS:SFHS). Still, it could not replicate those findings in the UKB (Gibson et al., 2017). A later study found that risk genes associated with MDD were linked to AD development, which can indicate mediation (Ni et al., 2018). A recent study reported a small, positive genetic correlation between depression and AD (Harerimana et al., 2022).

Pleiotropy:

Pleiotropy is when a gene affects more than one phenotype. There are multiple forms of pleiotropy: biological and mediated, which can be distinguished using genetic epidemiologic data through mediation analysis (Salinas et al., 2017). Biological pleiotropy is when the gene affects both phenotypes directly, while mediated pleiotropy is when a gene affects one phenotype through a second phenotype. To study pleiotropy, a mediation analysis of cross-phenotype associations can be conducted. There is plausible shared genetic variation at the SNP level, and we suspect part of the effect of AD is mediated by depression that might be explained by biological or mediated pleiotropy. Thus, this thesis aims to conduct two separate GWAS for the depression and AD phenotypes; and use the summary statistics to determine common SNPs among the two phenotypes that indicate possible biological or mediated pleiotropy.

Importance of Study:

Evidence suggests that earlier-life depression is associated with an increased risk of AD (Ownby et al., 2006). In a twin study using the Swedish Twin Registry, having depression increased twice the odds of having dementia (Yang et al., 2021). There were increased odds of all-cause dementia for mid-life depression and late-life depression (Yang et al., 2021). However, understanding the relationship between AD and depression is not robust. While some knowledge surrounds the plausible mechanisms for the association, the genetics behind the association are currently limited. With depression being a common disorder and AD being more prevalent due

to increasing life expectancy, this is a public health concern to investigate this relationship more. There is importance in determining if depression is a risk factor for AD for population health. It can lead to new preventive measures for AD, especially if there is biological or mediated pleiotropy. This means that prevention and treatment of depression can be methods of prevention for dementia.

Methods:

Study Population:

UK Biobank (UKB) is a prospective cohort study started in 2006. It recruited over 500,00 participants from the UK between 2006 to 2010 between the ages 40 and 69 years old (Sudlow et al., 2015). The study focused on individuals with a genetic ethnicity/ancestry as "White" due to this subpopulation's large sample size (n=410,835) in UKB.

Genotyping:

Data collection included the collection of biosamples, including blood, to generate genetic data (Collins, 2007). The blood samples are used to extract DNA for genotyping. Genotyping was performed by Affymetrix (currently Applied Biosystem) on two closely related purpose-designed arrays, the UK BiLEVE Axiom array and the UK Biobank Axiom array (Bycroft et al., 2017). Fifty thousand participants were genotyped using the UK BiLEVE Axiom array, while the remaining 450,000 were genotyped using the UK Biobank Axiom array (Bycroft et al., 2017).

Quality Control:

Quality control (QC) was conducted on the available genotype data (Bycroft et al., 2017). This involves both SNP and individual-level QC to ensure the quality of the genotype data. The QC for the SNPs first involves ensuring all the SNPs are autosomal variants covered by the UK BiLEVE Axiom array and the UK Biobank Axiom array. A QC is conducted on the batch-level to ensure no inconsistencies within batches. Genetic markers not considered SNPs are removed,

including insertion and deletions. Another stage of SNP QC was conducted (described below). Individual-level QC involved ensuring that there were no mistakes during genotyping. This included determining if reported sex and genetic sex matched and sex chromosomes that are not XX or XY. Individuals who had outliers in heterozygosity or missing rates were removed. Finally, individuals genetically identified as having a White European ancestry were included in the final dataset to decrease population stratification (Marees et al., 2018).

Phenotype:

Depression was defined as individuals who either: 1) answered "yes" to the data field "non-cancer illness code, self-reported" (field 20002); or had ICD-10 codes for depression (fields 41270, 41202, and 41204). The exclusion criteria included bipolar disorder, postnatal depression, depression with psychotic features, schizophrenia, and multiple personality disorder. Furthermore, exclusion criteria for controls included dysthymia and other mood disorders using ICD-10 codes. Due to the small sample size of AD cases in UKB, proxy AD was defined for this study. Proxy AD was defined by data fields "illness of father" (20110) and "illness of mother" (20107), with cases being defined as individuals who had a parent with AD; and controls with no parents with AD.

Genome-Wide Association Study:

QC for the genotypic data was conducted using PLINK (Purcell et al., 2007). This addressed any SNPs that are missing at a high rate (< 98%), with low minor allele frequency (MAF < 0.01), or that are not in Hardy-Weinberg equilibrium (HWE). SNP pruning was conducted using PLINK before principle component analysis (PCA). PCA was performed using EIGENSOFT to create 20 principle components (Patterson et al., 2006; Price et al., 2006).

Univariate association analysis was performed on the full set of imputed and directly genotyped variants using a linear mixed model implemented in REGENIE (Mbatchou et al.,

2021). REGENIE implements a multi-stage procedure that starts by fitting a null model with whole genome regression, closely related to the linear mixed model, using ridge regression on the directly genotyped variants (N=550,028) to estimate the polygenic effects parameter to account for relatedness and population stratification and subsequently using the full set of imputed variants to estimate the genetic prediction of each phenotype using a leave one chromosome out procedure.

Firth regression is used for binary traits to ensure that the results are well-calibrated in the presence of low-frequency variants and unbalanced case-control ratios, both of which are present in this analysis. REGENIE implements an approximate Firth regression to run this approach genome-wide. For the analysis of two phenotypes (depression and AD), age at recruitment (field 21022), genetic sex (field 22001), and ten principal components were included in the regression model.

Using Bonferroni corrections for multiple testing, the significance level of genome-wide significant association is $p < 5x10^{-8}$, and genome-wide suggestive association is $p < 5x10^{-6}$ and nominal association is $p < 0.05$. Manhattan and QQ plots were created using the *qqman* package in R (Turner, 2018). A p-value comparison was conducted using the univariate analysis results for the two analyses.

SNP Identification:

SNPs were searched on the GWAS Catalog to identify related traits and genes of previously identified SNPs (Buniello et al., 2019).

Linkage Disequilibrium Clumping:

Linkage disequilibrium (LD) clumping of the univariate association results was performed jointly for the two phenotypes in PLINK. A four MB region surrounding the index variant and an r2 threshold of 0.2 was used to identify overlapping genome-wide significant ($p \leq$

5 x 10-8) signals in the same LD block across the two phenotypes. LD clumping considers the correlation structure between variants and their p-values due to linkage disequilibrium to obtain a set of independently associated loci shared between the two phenotypes.

Results:

Sample:

The study included 406,7394 individuals in the depression sample (**Figure 1**) and 409,704 individuals in the AD sample. The demographics of the two samples are in **Table 1**. In the depression sample, 38,683 (9.5%) individuals had depression. The mean age of the depressed group was 56.0, which was slightly younger than the non-depressed group (56.9). As for sex, 35.8% of depressed participants were male, compared to 46.8% of the non-depressed participants being of the male sex. In the AD sample, 61,147 (13.3%) had a parental history of AD. The mean age of the cases was 58.7, which was older than the controls (56.4). As for sex, 43.5% of participants with a family history of AD were male, compared to 46.1% of control participants.

Genome-Wide Association Study:

Two separate genome-wide association studies were performed for depression (n= 406,7394) and AD (n=409,704). There were 14,476,405 SNPs for the depression sample and 14,477,147 SNPs for the AD sample. Five SNPs were genome-wide significant, with two on chromosome 11, one on chromosome 5, and two on chromosome 1 for the depression GWAS (**Table 1**). A total of 1,319 SNPs were genome-wide significant for the AD sample. The highest frequency of SNPs was found on chromosome 19, followed by chromosome 11 and chromosome 7 (**Table 3**). The Manhattan plots and QQ plots for the depression and AD GWAS are shown in **Figure 2**-**Figure 5**. The genomic inflation factor (λGC) is 1.119073 for the depression GWAS and 1.095683 for the AD GWAS, which is somewhat inflated but likely due to the polygenicity of the traits.

The AD summary statistics among the genome-wide significant SNPs of depression are shown in **Table 4**. SNPs rs12122377 and rs6667006 on chromosome 1 were nominally significant ($p<0.05$). **Table 5** shows the 36 SNPs with genome-wide nominal p-values for depression among the non-chromosome 19, genome-wide significant SNPs of AD. The top SNP is rs11166811 on chromosome 8, with a p-value of 3.09×10^{-5}. The second top SNP rs1519369 on chromosome 8 is associated with opioid use in multiple studies (Rosoff et al., 2021; Lin et al., 2022). The following 10 SNPs belong to chromosome 8, with p-values less than 1×10^{-5}. Expectedly, there was a strong association signal with AD for SNPs on chromosome 19, which contains the APOE region and a few other known genetic regions associated with AD. In addition, SNPs from chromosome 11 are also related to PICALM, a genetic region associated with AD (Ando et al., 2022). Two SNPs from chromosome 16 are genome-wide nominal for depression and genome-wide significant for AD.

Clumping:

In addition to simply comparing p-values for the same SNPs across the two phenotypes, LD clumping was also performed to identify genome-wide significant SNPs for each phenotype that may be in strong LD with each other. One region from the LD clumping analysis contained genome-wide SNPs significantly associated with the two phenotypes (

Table 6). The region is from chromosome 19, which spanned 27 kilobases with four SNPS. The leading SNP is rs2075650, found in the depression and AD univariate analyses. In the p-value comparison, the p-value for depression was 0.0121. The SNP was also identified to have an association with Alzheimer's disease and the gene *TOMM40*. Two other SNPs are from both phenotypes: SNPs rs769449 and rs157581. They were identified in our p-value comparison, with the p-value calculated from the depression univariate analysis being 0.0001 and 0.0078, respectively. SNP rs769449 is associated with the *APOE* gene, while SNP rs157581 is associated with the *TOMM40* gene.

Discussion:

In this paper, the results of the depression GWAS indicate potential SNPs associated with depression, especially on chromosome 11 and chromosome 1. These chromosomes have not been associated in any studies as major top hits for depression. However, SNPs on chromosome 1 have been associated with depression and other related traits, such as alcoholism and other psychiatric disorders (Xia et al., 2019). The results of the AD GWAS have SNPs strongly associated with this phenotype, including the APOE region on chromosome 19 and related genes, such as *TOMM40* and *NECTIN2*. When looking at the traits associated with the SNPs on chromosome 19, many were related to Alzheimer's and a family history of Alzheimer's, which matches our AD phenotype definition. For the p-value comparison, the SNPs on chromosome 8 are promising as they are significantly above the genome-wide nominal threshold for the p-values for the depression univariate analysis. Of the 12 SNPs from chromosome 8, only two have been identified, SNP rs1519369 and rs16908884. SNP rs16908884 has been associated with smoking initiation and gene *FAM135B*. The function of *FAM135B* is unknown; however, it has been associated with Alzheimer's disease carrier status, alcohol drinking, and smoking initiation.

Considering that the SNPs range approximately 42,400 base pairs, it is a potential region that should be studied more to understand the association between depression and AD.

The region on chromosome 19 from the LD clumping analysis is fascinating as it is a region that is associated with both gene *APOE* and gene *TOMM40*. SNP rs2075650 and gene *TOMM40 have* been associated with Alzheimer's disease. Furthermore, a study identified that it is associated with depression (McFarquhar et al., 2014). The SNP was associated with decreased executive function, decreased extraversion, and decreased positive memory bias (McFarquhar et al., 2014). Considering the preventative factors of Alzheimer's disease include mental stimulation and social connection, depression might inhibit these, leading to an increased risk of Alzheimer's disease. Another possible explanation is that activating the *TOMM40* gene can lead to changes in the brain, increasing the risk of Alzheimer's disease. Thus, this association should be looked at more and considered from a genetic and environmental perspective.

Some limitations of this study should be considered. First, the definition of our phenotypes has some downsides. For the depression phenotype, the inclusion of self-reported depression and electronic medical records of depression results in a phenotype that encompasses "broad depression" and "EMR-coded depression" that results in potential heterogeneity of the cases. Depression as a phenotype has also been reported to be a heterogenetic disease resulting from a GWAS, depending on how it is defined, possibly resulting in results attributed to traits and conditions similar to depression, such as smoking and insomnia (Cai et al., 2020). This raises concern as many of the traits associated with the SNPs found in the depression univariate analysis were associated with related traits, such as smoking, alcohol use, and drug dependence. AD also raises concerns as the phenotype was defined as a family history of AD as a proxy for

AD. A family history of AD does not necessarily indicate the future development of AD. Thus, this phenotype overestimates the sample of individuals developing AD.

The strength of this study is the large sample size of the phenotype and the GWAS. Furthermore, using REGENIE allows for applying linear-mixed methods for the GWAS. In the following steps, using multiple definitions of depression should be conducted to consider the abovementioned limitations. Furthermore, it has become common for GWAS to include analyses of multiple depression phenotypes. Studies should look at pleiotropy using standard analysis techniques, such as mediation analysis and Mendelian randomization, to understand the relationship between depression and dementia. Furthermore, conducting a gene-interaction study can help understand possible risk factors for individuals with depression. This can help create prevention programs for people with depression that can be tailored to them, thus leading to more effective prevention of Alzheimer's disease.

References:

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Ando, K., Nagaraj, S., Küçükali, F., de Fisenne, M. A., Kosa, A. C., Doeraene, E., Lopez Gutierrez, L., Brion, J. P., & Leroy, K. (2022). PICALM and Alzheimer's Disease: An Update and Perspectives. *Cells*, *11*(24), 3994. https://doi.org/10.3390/cells11243994

Biobank, U. K. (2007). UK Biobank: protocol for a large-scale prospective epidemiological resource. 1-112.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connel, J., Cortes, A., Welsh, S., McVean, G., Leslie, S., Donnelly, P., & Marchini, J. (2017). Genome-wide genetic data on~ 500,000 UK Biobank participants. BioRxiv, 166298. https://doi.org/10.1038/s41586-018-0579-z

Byers, A. L., & Yaffe, K. (2011). Depression and risk of developing dementia. *Nature reviews. Neurology*, *7*(6), 323–331. https://doi.org/10.1038/nrneurol.2011.60

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., Flicek, P., … Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research*, 47(D1), D1005–D1012. https://doi.org/10.1093/nar/gky1120

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., Clarke, T. K., Forstner, A. J., Grabe, H. J., Hamilton, S. P., Levinson, D. F., Lewis, C. M., Lewis, G., Martin, N. G., Milaneschi, Y., Mors, O., Müller-Myhsok, B., Penninx, B. W. J. H., Perlis, R. H., Pistis, G., … Flint, J. (2020). Minimal phenotyping yields genome-wide

association signals of low specificity for major depression. *Nature genetics*, *52*(4), 437–447. https://doi.org/10.1038/s41588-020-0594-5

Cerejeira, J., Lagarto, L., & Mukaetova-Ladinska, E. B. (2012). Behavioral and psychological symptoms of dementia. *Frontiers in neurology*, *3*, 73. https://doi.org/10.3389/fneur.2012.00073

Collins, R. (2007). UK biobank: protocol for a large-scale prospective epidemiological resource. https://policycommons.net/artifacts/1774896/uk-biobank/2506542/

Dafsari, F. S., & Jessen, F. (2020). Depression-an underrecognized target for prevention of dementia in Alzheimer's disease. *Translational psychiatry*, *10*(1), 160. https://doi.org/10.1038/s41398-020-0839-1

Fernandez-Pujals, A. M., Adams, M. J., Thomson, P., McKechanie, A. G., Blackwood, D. H., Smith, B. H., Dominiczak, A. F., Morris, A. D., Matthews, K., Campbell, A., Linksted, P., Haley, C. S., Deary, I. J., Porteous, D. J., MacIntyre, D. J., & McIntosh, A. M. (2015). Epidemiology and Heritability of Major Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PloS one*, *10*(11), e0142197. https://doi.org/10.1371/journal.pone.0142197

Gibson, J., Russ, T. C., Adams, M. J., Clarke, T. K., Howard, D. M., Hall, L. S., Fernandez-Pujals, A. M., Wigmore, E. M., Hayward, C., Davies, G., Murray, A. D., Smith, B. H., Porteous, D. J., Deary, I. J., & McIntosh, A. M. (2017). Assessing the presence of shared genetic architecture between Alzheimer's disease and major depressive disorder using genome-wide association data. *Translational psychiatry*, *7*(4), e1094. https://doi.org/10.1038/tp.2017.49

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Cullen, B., Malik, R., METASTROKE Consortium, International Consortium for Blood Pressure GWAS, SpiroMeta Consortium, CHARGE Consortium Pulmonary Group, CHARGE Consortium Aging and Longevity Group, Worrall, B. B., Sudlow, C. L., Wardlaw, J. M., Gallacher, J., Pell, J., McIntosh, A. M., Smith, D. J., … Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular psychiatry*, *21*(11), 1624–1632. https://doi.org/10.1038/mp.2015.225

Harerimana, N. V., Liu, Y., Gerasimov, E. S., Duong, D., Beach, T. G., Reiman, E. M., Schneider, J. A., Boyle, P., Lori, A., Bennett, D. A., Lah, J. J., Levey, A. I., Seyfried, N. T., Wingo, T. S., & Wingo, A. P. (2022). Genetic Evidence Supporting a Causal Role of Depression in Alzheimer's Disease. *Biological psychiatry*, *92*(1), 25–33. https://doi.org/10.1016/j.biopsych.2021.11.025

Holmquist, S., Nordström, A., & Nordström, P. (2020). The association of depression with subsequent dementia diagnosis: A Swedish nationwide cohort study from 1964 to 2016. *PLoS medicine*, *17*(1), e1003016. https://doi.org/10.1371/journal.pmed.1003016

Jermy, B. S., Glanville, K. P., Coleman, J. R. I., Lewis, C. M., & Vassos, E. (2021). Exploring the genetic heterogeneity in major depression across diagnostic criteria. *Molecular psychiatry*, *26*(12), 7337–7345. https://doi.org/10.1038/s41380-021-01231-w

Katon, W., Pedersen, H. S., Ribe, A. R., Fenger-Grøn, M., Davydow, D., Waldorff, F. B., & Vestergaard, M. (2015). Effect of depression and diabetes mellitus on the risk for dementia: a national population-based cohort study. *JAMA psychiatry*, *72*(6), 612–619. https://doi.org/10.1001/jamapsychiatry.2015.0082

Kuring, J. K., Mathias, J. L., & Ward, L. (2020). Risk of Dementia in persons who have previously experienced clinically-significant Depression, Anxiety, or PTSD: A Systematic Review and Meta-Analysis. *Journal of affective disorders*, *274*, 247–261. https://doi.org/10.1016/j.jad.2020.05.020

Lin, L., Lin, J., Qiu, J., Wei, F., Bai, X., Ma, W., Zeng, J., & Lin, D. (2022). Gut microbiota alterations may increase the risk of prescription opioid use, but not vice versa: A two-sample bi-directional Mendelian randomization study. *Frontiers in microbiology*, *13*, 994170. https://doi.org/10.3389/fmicb.2022.994170

McFarquhar, M., Elliott, R., McKie, S., Thomas, E., Downey, D., Mekli, K., Toth, Z. G., Anderson, I. M., Deakin, J. F., & Juhasz, G. (2014). TOMM40 rs2075650 may represent a new candidate gene for vulnerability to major depressive disorder. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *39*(7), 1743–1753. https://doi.org/10.1038/npp.2014.22

Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International journal of methods in psychiatric research*, *27*(2), e1608. https://doi.org/10.1002/mpr.1608

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., Habegger, L., Ferreira, M., Baras, A., Reid, J., Abecasis, G., Maxwell, E., & Marchini, J. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature genetics*, *53*(7), 1097–1103. https://doi.org/10.1038/s41588-021-00870-7

McFarquhar, M., Elliott, R., McKie, S., Thomas, E., Downey, D., Mekli, K., Toth, Z. G., Anderson, I. M., Deakin, J. F., & Juhasz, G. (2014). TOMM40 rs2075650 may represent a new candidate gene for vulnerability to major depressive disorder. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *39*(7), 1743–1753. https://doi.org/10.1038/npp.2014.22

National Institutes on Aging. (2017, May 16). *What Happens to the Brain in Alzheimer's Disease?* National Institutes on Aging. https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease

Ni, G., Moser, G., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Wray, N. R., & Lee, S. H. (2018). Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *American journal of human genetics*, *102*(6), 1185–1194. https://doi.org/10.1016/j.ajhg.2018.03.021

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., ... & Liu, X. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105-e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Ownby, R. L., Crocco, E., Acevedo, A., John, V., & Loewenstein, D. (2006). Depression and risk for Alzheimer disease: systematic review, meta-analysis, and metaregression analysis. *Archives of general psychiatry*, *63*(5), 530–538. https://doi.org/10.1001/archpsyc.63.5.530

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, *2*(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Rosoff, D. B., Smith, G. D., & Lohoff, F. W. (2021). Prescription Opioid Use and Risk for Major Depressive Disorder and Anxiety and Stress-Related Disorders: A Multivariable Mendelian Randomization Analysis. *JAMA psychiatry*, *78*(2), 151–160. https://doi.org/10.1001/jamapsychiatry.2020.3554

Salinas, Y. D., Wang, Z., & DeWan, A. T. (2018). Statistical analysis of multiple phenotypes in genetic epidemiologic studies: from cross-phenotype associations to pleiotropy. *American journal of epidemiology*, *187*(4), 855-863. https://doi.org/10.1093/aje/kwx296

Salinas, Y. D., Wang, Z., & DeWan, A. T. (2021). Discovery and mediation analysis of cross-phenotype associations between asthma and body mass index in 12q13. 2. *American Journal of Epidemiology*, *190*(1), 85-94. https://doi.org/10.1093/aje/kwaa144

Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., ... & Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, *398*(10312), 1700-1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Saunders, A. M., Strittmatter, W. J., Schmechel, D., George-Hyslop, P. S., Pericak-Vance, M. A., Joo, S. H., Rosi, B. L., Gusella, J. F., Crapper-MacLachlan, D. R., Alberts, M.J., Hulette, C., Crain, B., Goldgaber, D., & Roses, A. D. (1993). Association of apolipoprotein E allele ϵ4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*, *43*(8), 1467-1467. https://doi.org/10.1212/WNL.43.8.1467

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*, 005165. https://doi.org/10.21105/joss.00731

WHO. (2022). *Depression*. https://www.who.int/health-topics/depression

Xia, L., Xia, K., Weinberger, D. R., & Zhang, F. (2019). Common genetic variants shared among five major psychiatric disorders: a large-scale genome-wide combined analysis. *Glob Clin Transl Res*, *1*(1), 21-30. https://doi.org/10.36316/gcatr.01.0003

Yang, W., Li, X., Pan, K. Y., Yang, R., Song, R., Qi, X., Pedersen, N. L., & Xu, W. (2021). Association of life-course depression with the risk of dementia in late life: A nationwide twin study. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *17*(8), 1383–1390. https://doi.org/10.1002/alz.12303

Figures:

Figure 1: Sample for Depression Phenotype

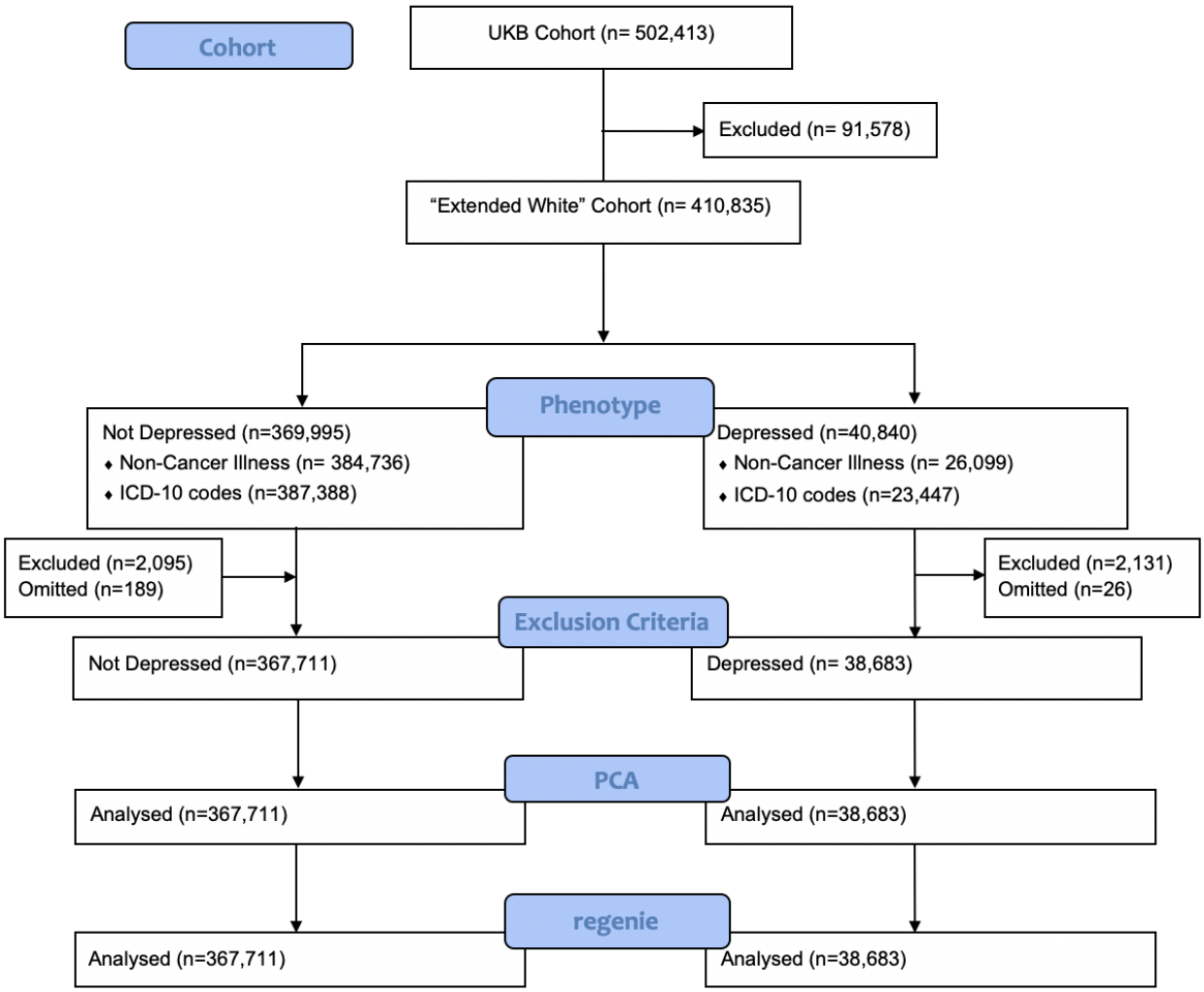

Figure 1 shows the process of creating our depression phenotype. This involved including individuals identified as European white using principal component analysis (PCA). The depression cases and controls were defined using self-reported depression and EMR-coded depression. Individuals diagnosed with bipolar disorder, schizophrenia, multiple personality disorder, depression with psychotic symptoms, and post-natal depression were excluded from the sample. The final sample included 406,7394 individuals that were used for the analysis.

Figure 2: Depression GWAS Manhattan Plot

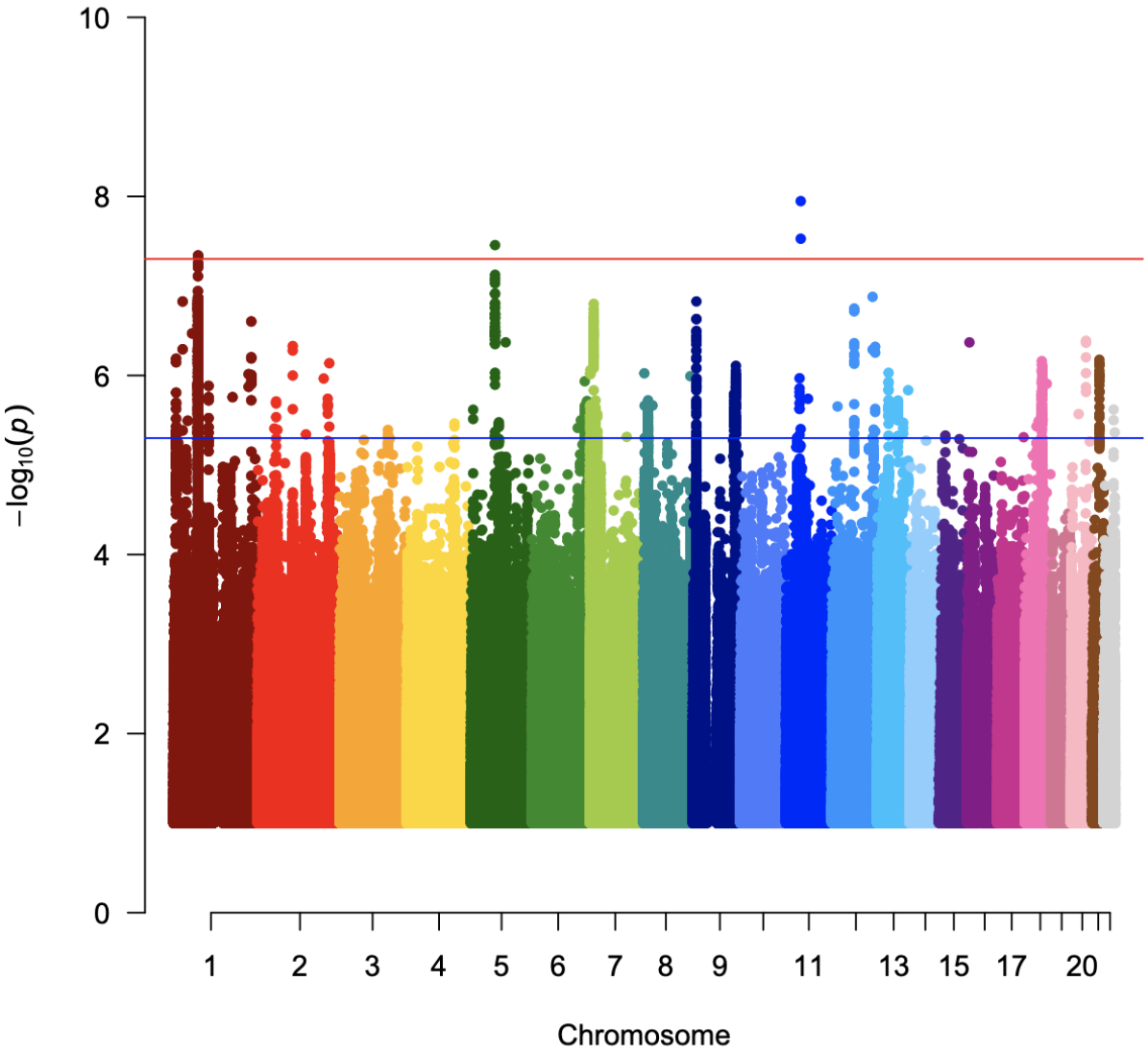

Figure 2 is a Manhattan plot of the depression GWAS. The x-axis shows the 22 autosomal chromosomes in order. The y-axis is -$\log_{10}$ of the *p*-values of SNPs. The red is the genome-wide significance level (10^{-8}). The blue line is the suggestive significance level (10^{-6}).

Figure 3: Depression GWAS QQ Plot

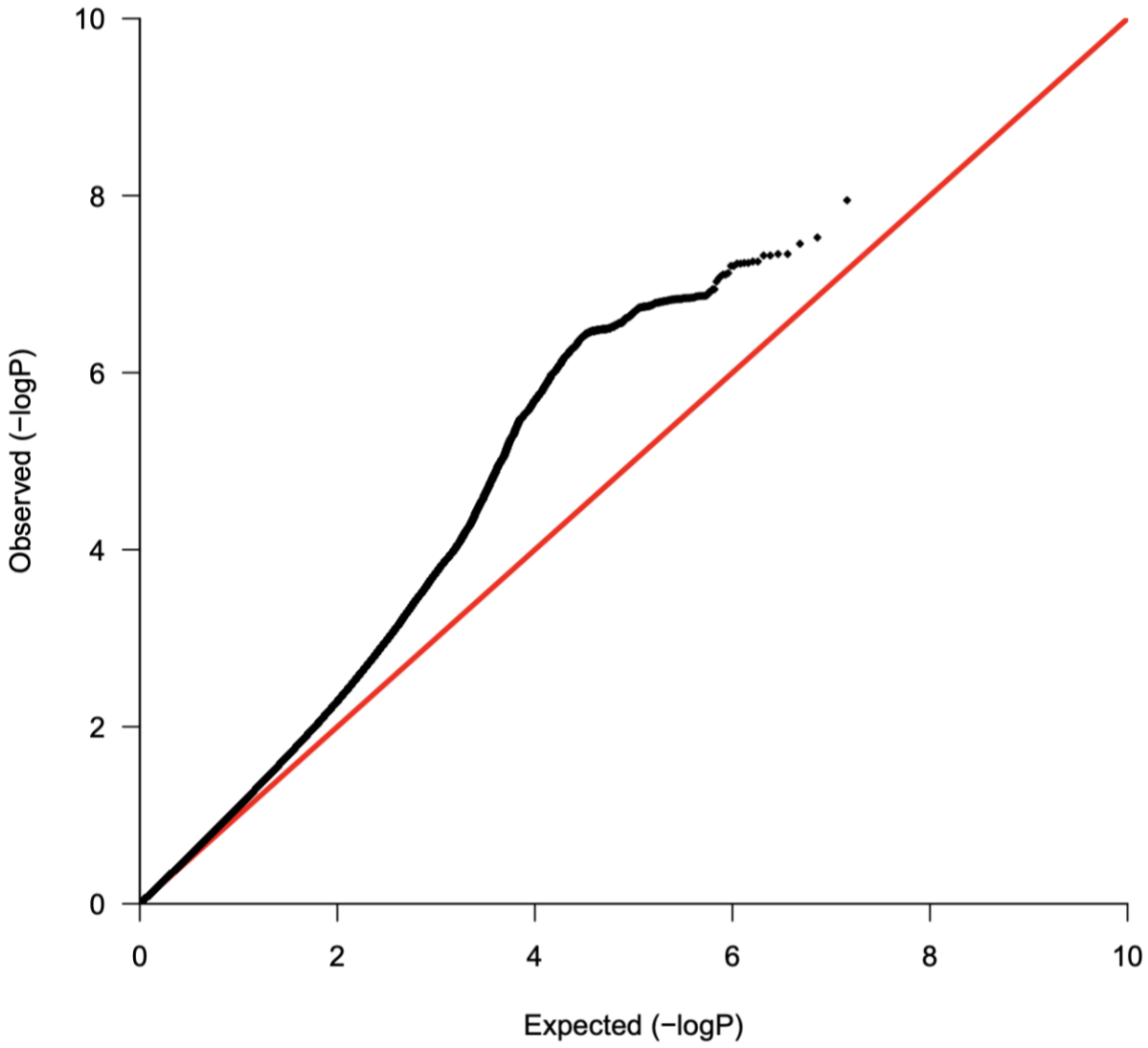

Figure 3 is the QQ plot of the depression GWAS. The x-axis shows the $-\log_{10}$ of *p*-values from the expected chi-square distribution, and the y-axis shows the $-\log_{10}$ of *p*-values from the observed chi-square distribution. The red line is the expected trend of the SNPs ordered from the highest *p*-value to the lowest *p*-value under the null hypothesis.

Figure 4: Proxy AD GWAS Manhattan Plot

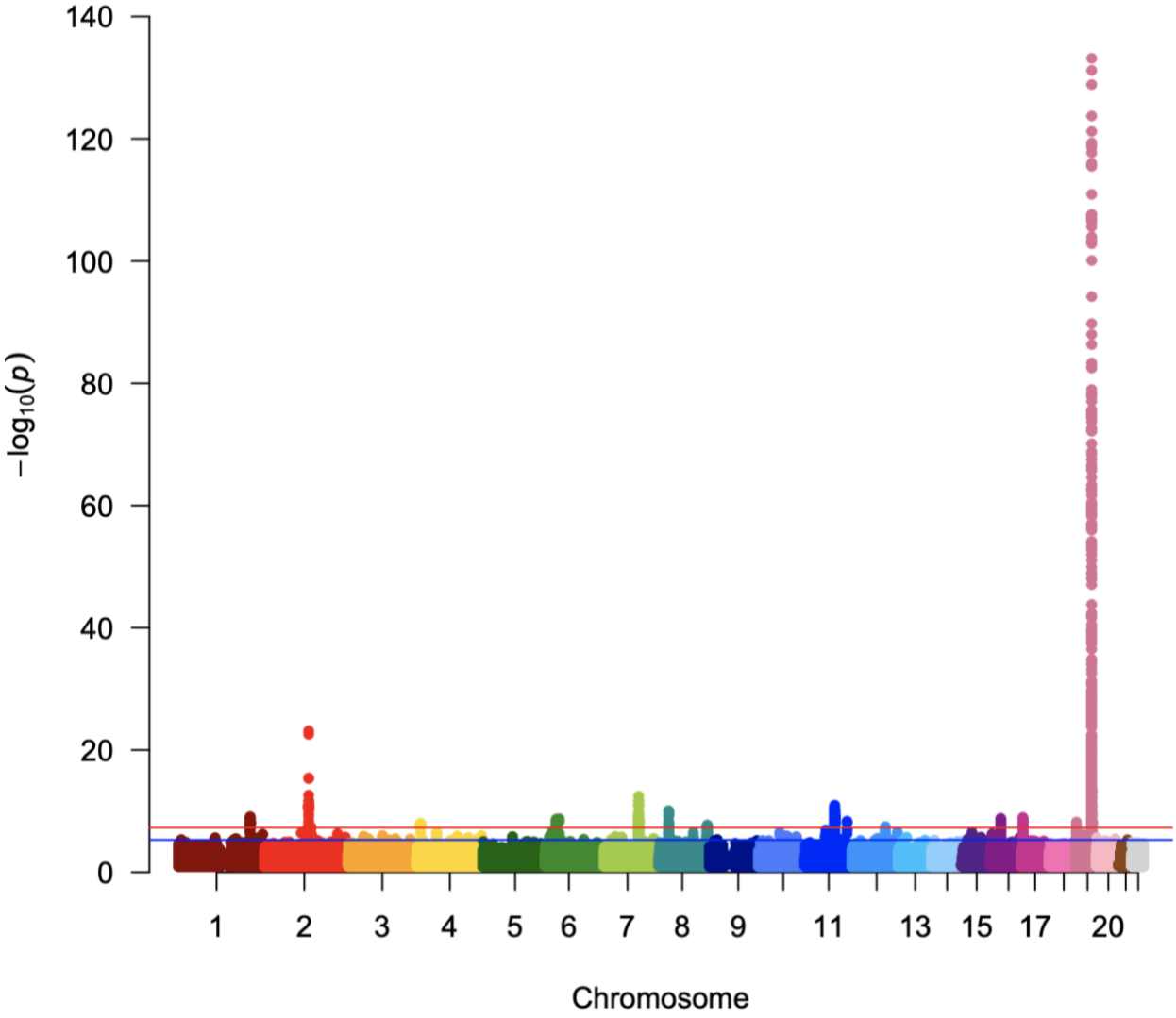

Figure 4 is the Manhattan plot of the proxy Alzheimer's disease GWAS. The x-axis shows the 22 autosomal chromosomes in order. The y-axis is -$\log_{10}$ of the *p*-values of SNPs. The red is the genome-wide significance level (10^{-8}). The blue line is the suggestive significance level (10^{-6}).

Figure 5: Proxy AD GWAS QQ Plot

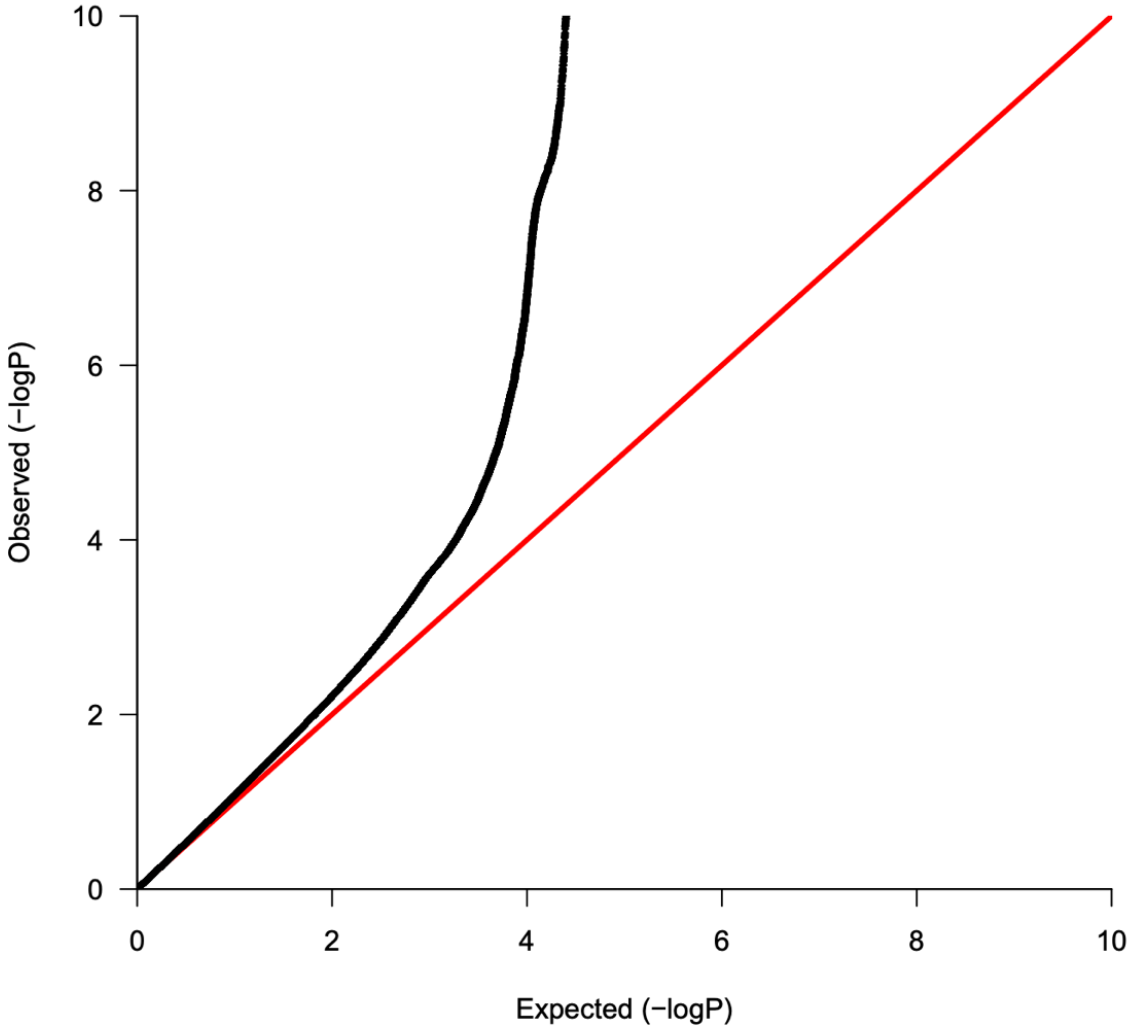

Figure 5 is the QQ plot of the proxy Alzheimer's disease GWAS. The x-axis shows the -log10 of p-values from the expected chi-square distribution, and the y-axis shows the -log10 of p-values from the observed chi-square distribution. The red line is the expected trend of the SNPs ordered from the highest *p*-value to the lowest *p*-value under the null hypothesis.

Tables:

Table 1: Demographics of Depression and Proxy AD Samples

Characteristics	Depression (n=38,683)	No Depression (n=367,711)	Proxy Alzheimer's (n= 54,376)	No Proxy Alzheimer's (n= 355,328)
Age (mean)	56.0 (±8.0)	56.9 (±8.0)	58.7 (±6.8)	56.4 (±8.2)
Sex				
Male	13,854 (35.8%)	171,983 (46.8%)	23,536 (43.3%)	163,701 (46.1%)
Female	24,829 (64.2%)	195,728 (53.2%)	30,840 (56.7%)	191627 (53.9%)

Table 2: Genome-Wide Significant SNPs from Depression GWAS

CHROM	SNP	*P*-value	Beta (SE)
11	rs77341890	1.13×10^{-08}	0.101 (0.0179)
11	rs79130449	2.97×10^{-08}	0.102 (0.0186)
5	rs10942927	3.50×10^{-08}	-0.0434 (0.0079)
1	rs12122377	4.57×10^{-08}	-0.0417 (0.0076)
1	rs6667006	4.75×10^{-08}	-0.0416 (0.0076)

Table 3: Frequencies of SNPS for each autosomal chromosome for Proxy AD GWAS

CHROM	SNP Frequency
1	50
2	51
3	0
4	19
5	0
6	76
7	116
8	36
9	0
10	0
11	159
12	1
13	0
14	0
15	0
16	55
17	34
18	0
19	721
20	0
21	0
22	0

Table 4: Genome-significant SNPs for Depression GWAS and associated p-values from Proxy AD GWAS

CHROM	SNP	*P*-value of Depression	Beta (SE) of Depression	*P*-value of Proxy AD	Beta (SE) of Proxy AD
1	rs12122377	1.13×10^{-08}	0.101 (0.0179)	0.0308	0.0143 (0.0066)
1	rs6667006	2.97×10^{-08}	0.102 (0.0186)	0.0288	0.0145 (0.0066)
5	rs10942927	3.50×10^{-08}	-0.0434 (0.0079)	0.457	-0.0051 (0.0069)
11	rs77341890	4.57×10^{-08}	-0.0417 (0.0076)	0.596	0.0080 (0.0150)
11	rs79130449	4.75×10^{-08}	-0.0416 (0.0076)	0.504	0.0103 (0.0155)

Table 5: Non-APOE Genome-significant SNPs for Proxy AD GWAS and associated p-values from Depression GWAS

CHROM	SNP	*P*-value of Proxy AD	Beta (SE) of Proxy AD	*P*-value of Depression	Beta (SE) of Depression
8	rs11166811	3.86×10^{-08}	0.0384 (0.0070)	3.09×10^{-05}	-0.0338 (0.0081)
8	rs1519369	2.03×10^{-08}	-0.0397 (0.0071)	3.54×10^{-05}	0.0339 (0.0082)
8	rs10087496	3.53×10^{-08}	0.0385 (0.0070)	3.61×10^{-05}	-0.0335 (0.0081)
8	rs10112372	1.96×10^{-08}	0.0392 (0.0070)	4.23×10^{-05}	-0.0331 (0.0081)
8	rs34625086	3.03×10^{-08}	-0.0387 (0.0070)	4.42×10^{-05}	0.0331 (0.0081)
8	rs12546144	$3.13x\ 10^{-08}$	0.0387 (0.0070)	4.46×10^{-05}	-0.0331 (0.0081)
8	rs4410912	$4.51x\ 10^{-08}$	-0.0383 (0.0070)	4.89×10^{-05}	0.0330 (0.0081)
8	rs11782838	2.56×10^{-08}	-0.0388 (0.0070)	5.78×10^{-05}	0.0325 (0.0081)
8	rs996955	2.83×10^{-08}	-0.0387 (0.0070)	6.04×10^{-05}	0.0324 (0.0081)
8	rs908843	3.77×10^{-08}	-0.0383 (0.0070)	6.27×10^{-05}	0.0323 (0.0081)
8	rs139920711	4.29×10^{-08}	-0.0391 (0.0071)	7.44×10^{-05}	0.0327 (0.0083)
8	rs16908884	2.87×10^{-08}	-0.0387 (0.0070)	8.87×10^{-05}	0.0317 (0.0081)
11	rs3915642	1.70×10^{-08}	-0.0376 (0.0067)	0.0038	-0.0222 (0.0077)
11	rs2888903	2.52×10^{-08}	-0.0371 (0.0067)	0.0041	-0.0221 (0.0077)
11	rs10751134	7.39×10^{-09}	-0.0389 (0.0067)	0.0055	-0.0215 (0.0078)
11	11:85841508_TGTGAAAA_T	7.60×10^{-10}	-0.0414 (0.0067)	0.0067	-0.0210 (0.0077)
11	rs4944560	2.66×10^{-09}	-0.0396 (0.0066)	0.0095	-0.0199 (0.0077)

11	11:85843982_GA_G	5.71×10^{-10}	-0.0416 (0.0067)	0.0096	-0.0200 (0.0077)
11	rs7114401	4.76×10^{-10}	-0.0417 (0.0067)	0.0097	-0.0200 (0.0077)
11	rs5793182	2.30×10^{-10}	-0.0423 (0.0067)	0.0113	-0.0195 (0.0077)
11	rs7952506	1.79×10^{-09}	-0.0398 (0.0067)	0.0113	-0.0193 (0.0076)
11	rs764392357	1.01×10^{-09}	-0.0405 (0.0066)	0.0116	-0.0193 (0.0076)
11	11:85855076_AC_A	1.01×10^{-09}	-0.0405 (0.0066)	0.0116	-0.0193 (0.0076)
11	rs774124637	1.01×10^{-09}	-0.0405 (0.0066)	0.0116	-0.0193 (0.0076)
11	rs3851178	2.12×10^{-09}	-0.0397 (0.0066)	0.0118	-0.0192 (0.0076)
11	rs11234554	5.51×10^{-09}	-0.0387 (0.0066)	0.0121	-0.0192 (0.0076)
11	rs10792831	1.23×10^{-09}	-0.0402 (0.0066)	0.0122	-0.0191 (0.0076)
11	rs10898439	3.03×10^{-09}	-0.0393 (0.0066)	0.0122	-0.0191 (0.0076)
11	rs10898438	1.82×10^{-09}	-0.0398 (0.0066)	0.0127	-0.0190 (0.0076)
11	rs9787874	3.03×10^{-09}	-0.0393 (0.0066)	0.0128	-0.0190 (0.0076)
11	rs11234555	8.49×10^{-09}	-0.0383 (0.0066)	0.0140	-0.0188 (0.0077)
11	rs10898440	4.03×10^{-10}	-0.0416 (0.0067)	0.0154	-0.0186 (0.0077)
11	rs3844143	3.66×10^{-10}	-0.0414 (0.0066)	0.0290	-0.0166 (0.0076)
11	rs11234556	3.34×10^{-10}	-0.0415 (0.0066)	0.0298	-0.0165 (0.0076)
16	rs62057090	4.28×10^{-08}	-0.0387 (0.0070)	0.0399	-0.0167 (0.0081)
16	rs375452507	1.53×10^{-09}	0.0413 (0.0068)	0.0408	0.0161 (0.0079)

Table 6: Regions from LD Clumping containing SNPs from both phenotype

CHROM	BP Range (Span)	SNP	*P*-value of Depression	Beta (SE) of Depression	*P*-value of Proxy AD	Beta (SE) of Proxy AD
19	45395619-45422946 (27kb)	rs2075650	0.012	-0.0270 (0.0108)	$<10^{-100}$	-0.405 (0.0084)
		rs157581	7.78×10^{-03}	-0.0246 (0.0092)	$<10^{-100}$	-0.3435 (0.0074)
		rs769449	1.14×10^{-04}	-0.0439 (0.0113)	$<10^{-100}$	-0.4880 (0.0087)
		rs4420638	-	-	$<10^{-100}$	0.4215 (0.0076)