Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

_____ _____

[Nadia Victoria Harerimana] Date

Integrating human brain transcriptomes and proteomes with genome-wide association data identifies risk genes shared between depression and Alzheimer's disease

By
Nadia Victoria Harerimana, Master of Science
Graduate Division of Biological and Biomedical Science
Genetic and Molecular Biology

Thomas S. Wingo, MD Advisor	Aliza P. Wingo, M.D, M.S Advisor
Karen N. Conneely, PhD Committee Member	David J. Cutler, PhD Committee Member
Victor Corces, PhD Committee Member	Nicholas Seyfried, PhD Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Integrating human brain transcriptomes and proteomes with genome-wide association data identifies risk genes shared between depression and Alzheimer's disease

By

Nadia Victoria Harerimana
Bachelor of Arts

Advisor (S): Thomas S. Wingo, MD and Aliza P. Wingo, MD

An abstract of
A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of

Master of Science

Graduate Division of Biological and Biomedical Science

Genetic and Molecular Biology

2021

Abstract

Integrating human brain transcriptomes and proteomes with genome-wide association data identifies risk genes shared between depression and Alzheimer's disease

By

Nadia Victoria Harerimana

Depression increases the risk for Alzheimer's disease (AD) in prospective epidemiological studies and recent evidence has shown a significant genetic correlation between depressive symptoms and AD. These lines of evidence suggest the possibility of shared pathophysiology and/or genetic liability between depression and AD. The first step toward deciphering the underlying mechanisms is to test the genetic correlation and explain the implications of a shared genetic basis between depression and AD. We show that there is a genetic basis to this association using data from the latest genome-wide association studies of depression and AD, respectively. Furthermore, we demonstrate that depression contributes to AD pathogenesis at the genetic level using Mendelian randomization. Lastly, we identified 46 brain transcripts and 8 brain proteins that underlie the contribution of depression to AD. These brain transcripts and proteins were significantly associated with AD pathologies, cognitive trajectory, and AD clinical diagnosis at FDR <0.05. Our study suggests that depression has a consistent causal role on AD, and we nominate 54 genes as potential mediators for future mechanistic studies to effectively treat these two illnesses

Integrating human brain transcriptomes and proteomes with genome-wide association data identifies risk genes shared between depression and Alzheimer's disease

By

Nadia Victoria Harerimana
Bachelor of Arts

Advisor (S): Thomas S. Wingo, MD and Aliza P. Wingo, MD

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of

Master of Science

Graduate Division of Biological and Biomedical Science

Genetic and Molecular Biology

2021

TABLE OF CONTENTS

Chapter 1: Depression and risk of developing Alzheimer's disease

Introduction

Aging populations worldwide have led to an increasing prevalence of depression and Alzheimer's disease (AD) [1, 2]. An estimate of 25 % of individuals with AD also has depression, with other depressive syndromes affecting an additional 20% to 30 % [3-5]. The exact cause of their co-occurrence is far from conclusive, but depression at any age (i.e., early-, mid-, or late-life depression) is thought to be associated with a two-fold increased risk of developing AD in prospective epidemiological studies [6, 7]. Notably, the genetic link between depression and AD has also been explored, with several studies suggesting that risk genes for depression may be involved in AD pathology [8-10]. These previous findings and substantial research efforts suggest that there are likely shared etiologies between depression and AD. Therefore, studies of both epidemiologic and genetic approaches are required to better understand the biological mechanisms that underlie these clinical entities.

Epidemiological Studies

Early-life depression

Many of the investigations into the association between depression and AD have come from the field of epidemiology. Several population-based studies have confirmed that early-life depression is associated with an increased risk of AD [11-13], supporting a risk factor hypothesis. For example, one case-control study of approximately 4000 participants found that depressive symptoms before the onset of AD are significantly associated with AD risk, even when the first depressive symptoms occurred more than 25 years before the beginning of cognitive decline symptoms [11]. There is also evidence supporting a strong link between the number of depressive episodes and AD, with a 14 % increase in risk for AD with each episode of depression [12]. This is further confirmed by a longitudinal study that demonstrated a 4-fold increased risk of the likelihood of developing AD among individuals with early-onset depression (age < 60 years) [13]. Together, these observations suggest potential biological mechanisms underlying this association. Thus, these underlying mechanisms need to be further investigated.

Mid- and late-life depression

Several epidemiological studies also support an association of mid- and late-life depression with the risk of developing AD. Evidence from longitudinal studies confirms an association between depressive symptoms during mid-life and late-life and the risk of AD, with the risk being more pronounced with late-life depressive symptoms [14, 15]. Consistent with these observations, two separate meta-analyses on population-based prospective studies among approximately 50,000 participants concluded that late-depression significantly increased the risk of dementia by two-fold [6, 7]. Furthermore, these meta-analyses confirmed that depression at any age (i.e., early-, mid-, or late-life depression) is associated with an increased risk for all types of dementia, even if the effect was not robust across all individual studies [7]. Despite these well-known associations, there are still limited insights into molecular underpinnings behind the detrimental effects of depression on AD risk. Additionally, these epidemiological findings suggest that other factors, such as genetic and molecular, could explain a greater proportion of the role of depression on the risk of AD development, which I present in chapter 2.

Genetic Studies

Candidate Genes Associations

Genetic studies have attempted to identify genetic factors conferring risk for both depression and AD. The majority of these studies have focused on candidate gene approach, prioritizing relevant genes to increase statistical power to detect associations [8-10]. For example, two candidate gene studies based on genetic polymorphisms associated with depression identified a handful of risk genes contributing to AD, such as HACE1, NEGR1, and RERE [8, 9]. These findings suggested that genetic factors conferring the risk of depression might affect AD development. However, these previous studies explicitly focused on candidate genes based on either their functional relevance, biological pathways, or chromosomal regions to the disease pathogenesis. Therefore, the most comprehensive approach to investigating the shared genetic architecture between depression and AD would be extending these single genes to the entire genome via genome-wide association studies (GWAS).

Genome-Wide Associations

Genome-wide association studies allow for a hypothesis-free approach that tests genetic variants spanning the entire genome but require large sample sizes and thorough phenotyping. These large-scale genetic studies have revolutionized the field of complex genetic diseases, providing numerous compelling associations that influence disease predispositions. For example, recent GWAS for depression [16-21] and AD [22-25] have used the alternative strategy of increasing sample size (i.e., minimal phenotyping) and yielded numerous replicated genetic variants, respectively. These successful GWAS findings have led to the discovery of novel genetic variants associated with each disease.

A particular noteworthy value of genome-wide associations has been in the numerous applications for GWAS data. For instance, GWAS results — in the form of either individual-level genotype data or summary-level statistics data — have enabled a wide range of applications, including estimation of genetic correlations between traits [26, 27], cross-trait analyses [27, 28], Mendelian randomization studies [29, 30] and integration with quantitative trait loci data [31, 32], among other applications. These applications have begun to show promise in their ability to provide new insights into disease biology beyond gene identification for many complex diseases.

For example, two studies have accessed the genetic correlation between depression and AD using GWAS summary data for both diseases [25, 33]. One group, Gibson and colleagues, tested the genetic correlation between depression and AD and did not find evidence for a shared genetic architecture [33]. The likely reason for null results may be explained by the limited power to detect significant associations; hence this previous result cannot be considered conclusive. A recent group led by Jansen et al. tested the genetic correlation between AD and depressive symptoms using more recent AD GWAS results and detected a significant genetic correlation, supporting roles for shared genetic risk factors [25]. The most noticeable differences between these two studies are the availability of full GWAS summary statistics and larger depression and AD GWAS sample sizes. This new evidence serves as an example for future studies to investigate whether the genetic architecture of major depression contribute to AD, which I present in chapter 2.

Functional Approaches for GWAS data

Linkage Disequilibrium Score Regression

Genetic correlation is a key population factor that explains the shared genetic architecture of complex traits/diseases and can prioritize potential causal relationships [34, 35]. There exists a method for estimating the genetic overlap between traits using GWAS summary data that is not biased by sample overlap, e.g., linkage disequilibrium score (LDSC) regression [26, 27]. With the advent of GWAS summary data, this method has been applied to great effect. For example, genetic correlations have been found in 276 pairs of phenotypes among 24 complex diseases and traits [27], highlighting the power of GWAS summary statistics. Therefore, this approach can be applied to quantify the degree of overlap for genetic risk of depression and AD, using newly available GWAS summary statistics, which may inform the search for the underlying shared biological mechanisms between these two diseases.

Mendelian Randomization

Mendelian randomization (MR) is a powerful method that test whether an observational association between an exposure and an outcome is consistent with a causal effect, by using genetic markers as instrumental variables [36, 37]. The application of MR bypasses the challenge of unmeasured confounding and reverse causation in association studies, and is therefore seen as an appealing analysis to perform alongside with genome-wide genetic correlation studies in efforts to increase our understanding of the relationship between complex traits.

Several approaches for making causal inferences have been proposed. These include methods that expand the two-sample MR design to increase statistical power by exploiting independent summary-level data sources and accounting for linkage disequilibrium (LD) among the genetic variants [29]. For instance, the Genome-wide Complex Trait Analysis-Generalized Summary Mendelian Randomization (GCTA-GSMR) is a multivariate MR framework that provides a test to detect and eliminate genetic instruments that possess pleiotropic effects on both the exposure and the outcome. In addition, the GCTA-GSMR method includes a heterogeneity in dependent instrument (HEIDI) test that removes outliers, which may be associated with confounding factors [29]. Thus, this summary MR method can be used to interrogate questions such as: does depression increases the risk of AD or vice versa at the genomic level?

Quantitative Trait Mapping

Quantitative trait loci (QTL) mapping is an approach that links genotype and phenotype data (e.g., molecular data) in attempt to dissect the genetic basis of complex traits. Quantitative molecular traits such as transcriptome, proteome, or metabolome are often dysregulated in disease and can act as intermediate phenotypes, providing great insights into the functional consequences of genetic variation [38, 39]. Characterizing these molecular traits has been essential. For example, it has been shown that proximal-expression quantitative trait loci (eQTLs) (e.g., within 1MB region) tend to overlap with disease-associated variants identified by GWASs and are often used to detect candidate causal genes [40-42]. In the light of these eQTL associations, it has been shown that distal-eQTLs (e.g., outside of the 1MB region) not only explain a significant amount of genetic variation but also contain important biological interpretations [43, 44]. However, to date, there have been very few comprehensive investigations of whether genetic signals from GWAS are distal-eQTLs or -pQTLs in brain tissues. Thus, it is vital to understand how genetic variants that predispose to depression modulate brain transcript and protein expression levels and to put those changes in expression associated with depression in relation to AD risk.

My thesis work aims to increase our understanding of the underlying association between depression and AD. I use the abovementioned functional genomic approaches that take advantage of newly available GWAS summary data and brain transcriptomic and proteomic data. In chapter 2, I test whether a genetic variation predisposing to depression contribute to AD pathology and elucidate genes and pathways involved in this association, which may serve as potential targets for further mechanistic studies to support efforts in drug discovery. In chapter 3, I discuss the conclusions drawn from this work, including the relative importance of large-scale studies and brain omic profiles in providing a more holistic picture of the mechanisms behind the co-occurrence between these two illnesses.

Reference:

1. Nichols, E., et al., *Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016.* The Lancet Neurology, 2019. **18**(1): p. 88-106.
2. Vos, T., et al., *Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019.* The Lancet, 2020. **396**(10258): p. 1204-1222.
3. Olin, J.T., et al., *Provisional Diagnostic Criteria for Depression of Alzheimer Disease: Rationale and Background.* The American Journal of Geriatric Psychiatry, 2002. **10**(2): p. 129-141.
4. Olin, J.T., et al., *Provisional Diagnostic Criteria for Depression of Alzheimer Disease.* The American Journal of Geriatric Psychiatry, 2002. **10**(2): p. 125-128.
5. Ballard, C., et al., *The prevalence, associations and symptoms of depression amongst dementia sufferers.* J Affect Disord, 1996. **36**(3-4): p. 135-44.
6. Diniz, B.S., et al., *Late-life depression and risk of vascular dementia and Alzheimer's disease: systematic review and meta-analysis of community-based cohort studies.* Br J Psychiatry, 2013. **202**(5): p. 329-35.
7. Bellou, V., et al., *Systematic evaluation of the associations between environmental risk factors and dementia: An umbrella review of systematic reviews and meta-analyses.* Alzheimers Dement, 2017. **13**(4): p. 406-418.
8. Ni, H., et al., *The GWAS Risk Genes for Depression May Be Actively Involved in Alzheimer's Disease.* J Alzheimers Dis, 2018. **64**(4): p. 1149-1161.
9. Arlt, S., et al., *Genetic risk factors for depression in Alzheimer`s disease patients.* Curr Alzheimer Res, 2013. **10**(1): p. 72-81.
10. Kitzlerová, E., et al., *Interactions Among Polymorphisms of Susceptibility Loci for Alzheimer's Disease or Depressive Disorder.* Med Sci Monit, 2018. **24**: p. 2599-2619.
11. Green, R.C., et al., *Depression as a Risk Factor for Alzheimer Disease: The MIRAGE Study.* Archives of Neurology, 2003. **60**(5): p. 753-759.
12. Dotson, V.M., M.A. Beydoun, and A.B. Zonderman, *Recurrent depressive symptoms and the incidence of dementia and mild cognitive impairment.* Neurology, 2010. **75**(1): p. 27-34.

13. Geerlings, M.I., et al., *History of depression, depressive symptoms, and medial temporal lobe atrophy and the risk of Alzheimer disease.* Neurology, 2008. **70**(15): p. 1258-64.
14. Barnes, D.E., et al., *Depressive symptoms, vascular disease, and mild cognitive impairment: findings from the Cardiovascular Health Study.* Arch Gen Psychiatry, 2006. **63**(3): p. 273-9.
15. Saczynski, J.S., et al., *Depressive symptoms and risk of dementia: the Framingham Heart Study.* Neurology, 2010. **75**(1): p. 35-41.
16. Okbay, A., et al., *Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses.* Nat Genet, 2016. **48**(6): p. 624-33.
17. Hyde, C.L., et al., *Identification of 15 genetic loci associated with risk of major depression in individuals of European descent.* Nat Genet, 2016. **48**(9): p. 1031-6.
18. Wray, N.R., et al., *Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression.* Nat Genet, 2018. **50**(5): p. 668-681.
19. Howard, D.M., et al., *Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways.* Nat Commun, 2018. **9**(1): p. 1470.
20. Hall, L.S., et al., *Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank.* Transl Psychiatry, 2018. **8**(1): p. 9.
21. Howard, D.M., 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**(3): p. 343-352.
22. Lambert, J.C., et al., *Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.* Nat Genet, 2013. **45**(12): p. 1452-8.
23. Kunkle, B.W., et al., *Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing.* Nature Genetics, 2019. **51**(3): p. 414-430.
24. Schwartzentruber, J., et al., *Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes.* Nature Genetics, 2021.
25. Jansen, I.E., et al., *Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk.* Nature Genetics, 2019. **51**(3): p. 404-413.

26. Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.* Nat Genet, 2015. **47**(3): p. 291-5.
27. Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits.* Nat Genet, 2015. **47**(11): p. 1236-41.
28. Pickrell, J.K., et al., *Detection and interpretation of shared genetic influences on 42 human traits.* Nat Genet, 2016. **48**(7): p. 709-17.
29. Zhu, Z., et al., *Causal associations between risk factors and common diseases inferred from GWAS summary data.* Nature Communications, 2018. **9**(1): p. 224.
30. Burgess, S., et al., *A robust and efficient method for Mendelian randomization with hundreds of genetic variants.* Nature Communications, 2020. **11**(1): p. 376.
31. Wu, Y., et al., *Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits.* Nature Communications, 2018. **9**(1): p. 918.
32. Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.* Nat Genet, 2016. **48**(5): p. 481-7.
33. Gibson, J., et al., *Assessing the presence of shared genetic architecture between Alzheimer's disease and major depressive disorder using genome-wide association data.* Transl Psychiatry, 2017. **7**(4): p. e1094.
34. Lee, S.H., et al., *Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs.* Nat Genet, 2012. **44**(3): p. 247-50.
35. Mehta, D., et al., *Evidence for Genetic Overlap Between Schizophrenia and Age at First Birth in Women.* JAMA Psychiatry, 2016. **73**(5): p. 497-505.
36. Smith, G.D. and S. Ebrahim, *'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?* Int J Epidemiol, 2003. **32**(1): p. 1-22.
37. Grover, S., *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation. S. Burgess and S. G. Thompson (2015). London, UK: Chapman & Hall/CRC Press. 224 pages, ISBN: 9781466573178.* Biometrical Journal, 2017. **59**(5): p. 1086-1087.
38. Civelek, M. and A.J. Lusis, *Systems genetics approaches to understand complex traits.* Nature Reviews Genetics, 2014. **15**(1): p. 34-48.

39. Vandiedonck, C., *Genetic association of molecular traits: A help to identify causative variants in complex diseases.* Clin Genet, 2018. **93**(3): p. 520-532.
40. Gibson, G., J.E. Powell, and U.M. Marigorta, *Expression quantitative trait locus analysis for translational medicine.* Genome medicine, 2015. **7**(1): p. 60-60.
41. Nicolae, D.L., et al., *Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS.* PLoS Genet, 2010. **6**(4): p. e1000888.
42. Hormozdiari, F., et al., *Colocalization of GWAS and eQTL Signals Detects Target Genes.* Am J Hum Genet, 2016. **99**(6): p. 1245-1260.
43. Brynedal, B., et al., *Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of Transcriptional Co-regulation.* The American Journal of Human Genetics, 2017. **100**(4): p. 581-591.
44. Lloyd-Jones, L.R., et al., *The Genetic Architecture of Gene Expression in Peripheral Blood.* The American Journal of Human Genetics, 2017. **100**(2): p. 228-237.

Chapter 2: Depression contributes to Alzheimer's disease through shared genetic risk
This chapter will be submitted to Science Advances with the following authors:
Harerimana NV, Liu Y, Gerasimov ES, Duong DM, Dammer EB, Beach TG, Reiman EM, Schneider JA, Boyle P, De Jager PL, Bennett DA, Lah JJ, Levey AI, Seyfried NT, Wingo TS, Wingo AP.

Introduction

Depression and dementia are commonly comorbid in people over 65 years [1-3]. Each condition alone can impair global functioning, physical health, and quality of life of affected individuals [4]. Furthermore, depression at any age (i.e., early-, mid-, or late-life depression) has been found to be associated with an increased risk for AD in prospective epidemiological studies among approximately 50,000 participants [5-8]. Though depression might be a risk factor for AD or an early sign of AD, there are likely shared etiologies that need to be elucidated.

The genetic link between depression and AD has previously been explored with mixed results. A number of studies have attempted to identify links between AD and depression by examining candidate sites or regions with mixed results [9-11]. A recent systematic review argued that depression could be a modifiable target for the prevention of cognitive decline and dementia [12]. Expanding the focus to considering the entire genome by using genome-wide association study (GWAS) summary results for both diseases initially did not reveal a significant correlation [13]; however, using more recent AD GWAS results, a significant genetic correlation between AD and depressive symptoms was found [14]. The two most salient differences between early and recent studies are the availability of full GWAS summary statistics and larger depression and AD GWAS sample sizes. Thus, with new evidence supporting a genetic correlation, we undertook the present study to identify shared molecular changes between these two illnesses using functional genomic approaches that capitalize on newly available brain transcriptomic and proteomic data. We hypothesized that depression-associated genetic variants alter expression levels of transcripts and proteins in the brain and these alterations can influence AD risk.

In this study, we first confirmed the shared genetic basis between depression and AD using summary statistics from the most recent depression [15] and AD [14] GWAS. Next, using Mendelian randomization [16], we observed that depression genes contribute to AD but we did not find evidence that AD genes contribute to depression. To resolve the genetic signal to

specific proteins, we identified brain transcripts and proteins regulated by the depression-associated genetic variants using human post-mortem brain transcriptomic and proteomic profiles. Lastly, we found that these transcripts and proteins were associated with AD endophenotypes, including rate of decline of cognitive performance over time, beta-amyloid plaques, and neurofibrillary tangles, as well as AD clinical diagnosis. Together, our findings identify key brain transcripts and proteins that link these two diseases and suggest mechanisms underlying the association.

Methods

GWAS data sets

Alzheimer's disease GWAS summary statistics

Genetic associations with AD were obtained from the meta-analysis GWAS of clinically diagnosed AD and AD-by-proxy in individuals of European ancestry by Jansen et al [14] (71,880 cases and 383,378 controls, Table 1). More details of the participants were described in the paper [14]. In brief, the meta-analysis included three independent consortia, namely the Alzheimer's disease working group of the Psychiatric Genomics Consortium (N=17,477 individuals), the International Genomics of Alzheimer's Project (N=54,162 individuals), and the Alzheimer's Disease Sequencing Project (N=7,506 individuals). The GWAS of AD-by-proxy was performed in 376,113 individuals of European ancestry from the UK Biobank with one or both parents diagnosed with AD. The meta-analysis reported a genetic correlation of 0.81 between AD status and AD-by-proxy status[14]. We used the full summary statistics (*P*-values, odds ratios, standard errors, and effect alleles) which consisted of genotyped and imputed data on 9,862,738 genetic variants from 455,258 individuals of European ancestry.

Depression GWAS summary statistics

Genetic associations with depression were obtained from a larger meta-analysis GWAS of clinically diagnosed major depressive disorder and broader self-declared definitions of depression on individuals of European ancestry (246,363 cases and 561,190 controls, Table 1). More details of the participants were described in the paper [15]. In brief, the meta-analysis included of three independent cohorts, namely 23andMe, Inc (N=307,354 individuals), UK Biobank (N=361,315 individuals), Psychiatric Genomics Consortium (N=138,884 individuals).

The meta-analysis reported a genetic correlation of 0.86 between broader self-declared definitions of depression and clinically diagnosed major depressive disorder [15]. We used the full summary statistics (*P*-values, beta coefficients, standard errors, and effect alleles) which consisted of genotyped and imputed data on 8,098,588 genetic variants from 807,553 individuals of European ancestry.

Data sources

Religious Orders Study and Rush Memory and Aging Project (ROS/MAP)

The ROS/MAP are two prospective clinical-pathologic cohorts of aging and dementia [17, 18]. All ROS/MAP participants are *without* known dementia at enrollment, undergo annual clinical evaluations, and agree to brain donation. Both studies share clinical, neuropathological, and brain autopsy standards, allowing joint analyses of the data and were approved by an Institutional Review Board of Rush University Medical Center. All participants signed an informed consent, Anatomical Gift Act, and Repository Consent to allow their data to be shared. For this study, we used the transcriptomic and proteomic data derived from the dorsolateral prefrontal cortex (dPFC) of post-mortem brain samples.

Banner Sun Health Research Institute (Banner)

The Banner cohort is a longitudinal study of healthy aging, Alzheimer's disease, and Parkinson's disease [19]. Participants are enrolled as cognitively healthy volunteers residing in the retirement communities of metropolitan Phoenix, Arizona. All participants underwent a standardized general medical, neurological, and neuropsychological assessments during life [19] and had available brain proteomic data measured from the dPFC .

Pathological phenotypes

Methods of assessing brain autopsies and neuropathologic traits for each cohort have been extensively described in previous studies [19, 20]. For the ROS/MAP study, we included measures of AD pathology including beta amyloid and neurofibrillary tangle identified by immunohistochemistry assays from eight brain regions, quantified by image analysis, and then averaged to obtain a summary measure for each pathology. For the Banner study, we included plaque total, which is the average amyloid plaque in the cortex of the frontal lobe. We also

included tangle total, which is the average neurofibrillary tangle density in the cortex of the frontal lobe. Briefly, the amount of amyloid plaque and neurofibrillary tangle density are identified using the Campbell-Switzer silver stain, scored according to the Consortium to Establish a Registry for Alzheimer's Disease template [21], and then averaged to obtain a summary measure for each pathology. Here, we used the square roots of these measures to enhance their normal distribution. Information on the AD pathologies of the subjects is provided in Table 2.

Neuropsychiatry phenotypes

Methods of assessing cognition for each cohort have been extensively summarized in previous studies [19, 22]. For the ROS/MAP study, depressive symptoms were assessed annually using the 10-item version of the Center for Epidemiological Studies Depression scale (CES-D) [23]. The CES-D score ranged between 0-10 with higher score indicating more depressive symptoms. Here we averaged the depression scores over the follow-up years to obtain an average score for each participant, which was then converted to a Z score to be used in the joint analysis including both ROS/MAP and Banner participants described below. Also, each participant underwent a full clinical evaluation, including a comprehensive cognitive assessment each year and a final clinical diagnosis of AD which follows the recommendation of the National Institute on Aging Reagan criteria [24]. We included two measures of cognition. The first is a person-specific rate of cognitive decline over time (aka cognitive trajectory) based on annual objective cognitive testing. The second is a clinical diagnosis of Alzheimer's disease (AD diagnosis) assessed by a neurologist specializing in dementia. For the Banner study, depressive symptoms was assessed annually using the 15-item version of the Hamilton Depression Rating Scale (HAMD)[25]. The score for HAMD ranged between 0-27 with higher score indicating more depressive symptoms. Additionally, participants underwent assessment by neurologists, psychiatrists, and neuropsychologists for a final clinicopathological diagnosis of cognition after death [19]. Participants also had rate of cognitive decline over time based on annual MMSE. Here, we standardized the score of cognitive trajectories to improve statistical comparisons. Information on the disease status of the subjects is provided in Table 2.

Brain Proteomic data

Protein abundance from the dPFC of the 400 ROS/MAP and 201 Banner samples was generated using tandem mass tag isobaric labelling by mass spectrometry methods for protein identification and quantification, as described by Wingo et al.,2020 [26]. Proteomic quality control and normalization were performed using previously described procedures [26]. Briefly, the workflow consisted of (i) proteins with missing values in more than 50% of the subjects were excluded, (ii) each protein abundance was divided by total protein abundance to remove effects of protein loading differences, (iii) $\log_2$ transformation of protein abundance and sample outlier detection. The residualized proteomic profiles were obtained via linear regression after removing the effects of technical and biological covariates. We retained quantified abundance of 8,356 proteins in 391 ROS/MAP subjects and 7,854 proteins in 196 Banner subjects (Table 2).

Brain Transcriptomic data

Gene expression was derived from the dPFC donated by 638 ROS/MAP participants, as described previously by De Jager et al., 2018 [27]. To account for differences between samples, experimental batch effects and unwanted RNA-sequencing specific technical effects, we performed library normalization and covariate adjustments using fixed/mixed-effects modeling. The reprocessing was done following the Sage-Bionetworks RNAseq normalization procedures [28]. Briefly, the residualized profiles were obtained after removing the effects of known biological and technical confounding factors. We retained quantified expression for 15,822 transcripts in 630 individuals (Table 2).

Genotyping data

Individuals from ROS/MAP were genotyped either using whole genome sequencing (WGS) or genome-wide genotyping by either Illumina OmniQuad Express or Affymetrix GeneChip 6.0 platforms, as described [27, 29]. Use of WGS was prioritized when multiple data sources were available. Individuals from Banner were genotyped using the Affymetrix Precision Medicine Array, as described here [30]. Quality control of WGS and each array-based genotyping source were performed separately using Plink [31] and included removing individuals with genotyping missing rate >5%, variants with Hardy Weinberg equilibrium p-value $< 10^{-5}$, variants with missing genotype rate >5%, variants with minor allele frequency <1%, and variants that are not single nucleotide polymorphisms (SNPs). KING was used to

remove randomly individuals estimated to be closer than second degree kinship [32]. Genotyping was imputed to the 1000 Genome Project Phase 3 [33] using the Michigan Imputation Server [34] and SNPs with imputation $R^2 > 0.3$ were retained for analysis. After quality control, 580 ROS/MAP subjects had complete genetic and transcriptomic data, 372 ROS/MAP subjects had complete genetic and proteomic data, and 97 Banner subjects had complete genetic, proteomic, and phenotypic data to be included in the analysis.

Statistical analysis

Linkage disequilibrium score (LDSC) regression analysis

Linkage disequilibrium score (LDSC) regression [35] was performed to estimate the genetic correlation (r_G) between depression and AD using data from the latest GWAS summary statistics. The genetic correlation was calculated on HapMap3 SNPs only (LD reference panel SNPs) to minimize potential bias by differences in LD structure [35]. Additionally, we removed SNPs with extremely large effect sizes ($X_1^2 > 80$), including the *APOE* region, since outliers can unduly influence LDSC regression [35].

Mendelian randomization (MR) analysis

Bidirectional MR was conducted to explore the causal relationship between depression and AD, using Generalized Summary data-based Mendelian Randomization (GSMR) approach [16]. To meet the assumptions of MR, we used clumping parameter settings of a *P*-value threshold of 1×10^{-8}, LD R^2 <0.05, removed SNPs that have large difference of allele frequency with the reference sample, and only considered the SNPs with the strongest effect on the exposures (*P*-value $< 5 \times 10^{-8}$) as the instrumental variables in the forward and reverse models, respectively. We used the HEIDI-outlier approach [16] to remove SNPs that have pleiotropic effects on both the exposure and the outcome. The retained SNPs were then tested for the association with the outcome for causal effect. Here, we conducted the forward GSMR analysis on 99 depression SNPs as the instrumental variables and the reverse GSMR analysis on 37 AD SNPs as the instrumental variables.

Surrogate Variable Analysis

We applied the Surrogate Variable Analysis (SVA) [36] approach to identify hidden-confounders in the expression data. Surrogate variables (SVs) were detected using the sva() function provided by the SVA package in R[37]. For both ROS/MAP transcriptomic and proteomic data, we included the first ten significant SVs as covariates for all relevant analyses

Identification of quantitative trait locus (QTL) analysis

Linear regression was used to identify association between genetic variation and transcript or protein expression levels in the brain. The expression levels were regressed against the genotype of each SNP, adjusting for depressive symptoms, AD status, and ten principal components of ancestry estimated using EIGENSTRAT [38]. We used the Benjamini-Hochberg approach to control the false discovery rate (FDR) and set a threshold of 5% to declare a QTL statistically significant. QTLs were defined as proximal if the SNP location was within 1Mb window of a gene's transcription start site (TSS), and those outside that window were defined as distal.

Meta-analysis

METAL [39] was used to perform meta-analysis using results from the ROS/MAP discovery and Banner replication analyses. Meta-analysis was carried out using effect size and standard error as input. We defined replication as associations with meta-analysis *p*-values smaller than those of both the discovery and replication datasets and with the same directions of associations.

Results

The shared genetic risk between depression and AD

To investigate whether depression and AD share genetic risk, we used LDSC regression, a method that requires only GWAS summary statistics and is not biased by sample overlap [35]. We used the depression GWAS by Howard et al. [15] and the AD GWAS by Jansen et al. [14] (Table 1). We found a significant positive genetic correlation between depression and AD, with r_g=0.17 (Z-score = 4.03, P = 5.54x10^{-5}), suggesting that these conditions have a shared genetic basis.

Evidence for a causal genetic effect of depression on AD

Genetic correlation may arise from pleiotropy (i.e., genes independently affecting both depression and AD) or from the causal effect of depression on AD or vice versa. To identify a potential causal effect of depression on AD, we performed Mendelian randomization using the GSMR[40]. We used the 115 SNPs found to be associated with depression at genome-wide significant level from the depression GWAS as the instruments, depression as the exposure, and AD as the outcome. We found a significant causal effect of depression on AD (effect size β = 0.029, standard error (SE) = 0.009, P_{GSMR} = 0.001, Figure 1A). Next, to test for the probability of a causal effect of AD on depression, we used the 61 AD GWAS-significant SNPs as the instruments, AD as the exposure, and depression as the outcome. We did not find a causal effect of AD on depression (β = -0.001, SE = 0.019, P_{GSMR} = 0.954; Figure 1B). Together, these findings suggest that the 115 depression-associated genetic variants predispose to depression, which in turn contributes to AD pathogenesis.

Depression-associated variants are associated with brain transcript and protein expression

To investigate how the 115 depression-associated SNPs underlie the potential causal effect of depression on AD, we performed brain expression quantitative trait locus (eQTL) and protein quantitative trait locus (pQTL) analyses.

In 580 participants from the ROS/MAP cohorts with genotyping and brain transcript data (Table 2), we identified 80 transcripts associated with 31 depression SNPs at FDR <0.05 (Figure 2A). Among the 80 SNP-transcript pairs, 67% of the SNPs (21 out of 31 SNPs) had a proximal effect on 36 corresponding transcripts, and 35% (11 out of 31 SNPs) had a distal effect on 44

transcripts (Figure 2A). Overall, we found that 31 of the 115 depression-associated SNPs regulate expression of 75 brain transcripts (Figure 2A).

In 372 subjects from the ROS/MAP cohorts and 97 subjects from the Banner cohorts (total, n=469) with complete genotyping and proteomic data (Table 2), we identified 32 pQTLs at FDR <0.05 reflecting 28 proteins associated with 13 depression-associated SNPs (Figure 2). Among the 32 SNP-protein pairs, 69% of the SNPs (9 out of 13 SNPs) had a proximal effect on 9 corresponding proteins (RAB27B, B3GLCT, ACADS, GMPPB, GPX1, DDAH2, ACYP1, DCC, and VARS), and 46% (6 out of 13 SNPs) had a distal effect on 19 proteins (Figure 2). Further, we found that eight of the depression-associated sites were both eQTLs and pQTLs, with six sites regulating the expression of both transcript and protein of the same gene (Figure 2, bolded genes).In sum, we found that 13 of the 115 depression-associated SNPs regulate expression of 28 brain proteins and six of these depression-associated SNPs control expression of both brain transcripts and proteins (Figure 2).

Brain transcripts regulated by depression genetic variants are associated with AD features

Since we observed a causal effect of depression on AD at the genetic level, we examined whether the 75 brain transcripts regulated by the depression SNPs are associated with AD-related features and endophenotypes, including trajectory of cognitive performance over time, beta-amyloid plaques, neurofibrillary tangles, and clinical diagnosis of AD in 587 participants in the ROS/MAP studies with available brain transcriptomic data. We identified 34 transcripts associated with beta-amyloid and 29 transcripts associated with tau tangles after adjusting for sex, age at death, and depressive symptoms (FDR p <0.05, Figure 3). Additionally, we identified 25 transcripts associated with AD diagnosis and 30 transcripts associated with cognitive trajectory after adjusting for sex, age at death, and depressive symptoms (at FDR p <0.05, Figure 3). Notably, among these transcripts, we found that higher levels of SAP30L, B3GLCT, FAM168B, ZNF740, and RERE were associated with higher levels of beta-amyloid and tau tangles, greater probability of having AD diagnosis, and faster rate of decline of cognitive performance (more negative slope of cognitive trajectory, Figure 3). Furthermore, we found that lower levels of SGIP1, AMT, SST, SNX10, TRIM36, and SST transcripts were associated with higher levels of beta-amyloid and tau tangles, greater probability of having AD diagnosis, and faster rate of decline of cognitive performance (Figure 3). In sum, we found that 46 brain

transcripts (of 75 tested, or 61%) were significantly associated with at least one AD feature. These findings support the notion that the depression risk genes contribute to AD via regulating expression of their corresponding transcripts in the brain.

Brain proteins regulated by depression genes are associated with AD characteristics

Likewise, we examined associations between AD-associated features and the 28 brain proteins regulated by the depression SNPs in subjects with brain proteomic and phenotypic data. In the ROS/MAP discovery cohort, we identified two proteins (RAB27B and DDAH2) associated with beta-amyloid and five proteins (RAB27B, DDAH2, CACNG2, B3GLCT, and GPAA1) associated with tau tangles after adjusting for sex, age at death, and depressive symptoms (FDR p <0.05, Figure 3). Also, we identified four proteins (RAB27B, DDAH2, CACNG2, and DCC) associated with AD diagnosis and four proteins (RAB27B, DDAH2, CACNG2, and ACADS) associated with cognitive trajectory after adjusting for sex, age at death, and depressive symptoms (at FDR p <0.05, Figure 3). Among these proteins, we found that RAB27B and DDAH2 were significantly associated with all four AD features in consistent directions of association— higher abundance of RAB27B was associated higher levels of beta-amyloid (β = 0.76, FDR p = 3.80E-02), higher levels of tangles (β = 0.56, FDR p = 4.73E-02), greater probability of having an AD diagnosis (β = 1.85, adjusted p = 2.43E-03), and more negative slope of cognitive trajectory (i.e., faster cognitive decline; β = -1.15, adjusted p = 6.63E-07; Figure 3). We also found that lower abundance of DDAH2 was associated higher levels of beta-amyloid (β = 1.86, FDR p = 7.99E-03), higher levels of tangles (β = 2.22, FDR p = 7.69E-06), higher probability of having an AD diagnosis (β = 4.16, adjusted p = 1.07E-03), and faster cognitive decline (β = -1.78, FDR, p = 9.90E-05; Figure 3).

In the Banner replication cohort, three proteins (DDAH2, GMPPB, and GMPPA) were associated with beta-amyloid and eight proteins (DDAH2, GPAA1, CACNG2, GMPPB, and GMPPA, RAB27B, GPX1, and DCC) associated with tau tangles after adjusting for sex, age at death, and depressive symptoms (FDR p <0.05, Figure 3). Additionally, we identified five proteins (DDAH2, GPAA1, CACNG2, GMPPB, and GMPPA) associated with AD diagnosis and four proteins (DDAH2, GPAA1,GMPPB, and GMPPA) associated with cognitive trajectory after adjusting for sex, age at death, and depressive symptoms (at FDR p <0.05, Figure 3). Among these proteins, we found that DDAH2, GMPPB, and GMPPA were significantly

associated with all four AD features in consistent directions of association (Figure 3). Several factors may be behind the different proteins identified in the discovery and replication cohorts including the different sample sizes and different methods of assessing AD diagnosis (taking into account AD pathologies or not) and pathologies (immunohistochemistry versus silver staining). Given these differences, we performed a meta-analysis of these results as meta-analysis was designed to combine data from multiple independent studies [41].

There were 13 proteins profiled in both ROS/MAP and Banner cohorts (RAB27B, CACNG2, DDAH2, ACADS, B3GLCT, GMPPB, GMPPA, GPAA1, DCC, GPX1, ACYP1, VARS, and DPY30) and were included in the meta-analysis. Replication was defined as having a meta-analysis p-value smaller than those from both the discovery and replication analyses and having the same directions of association in both. Of the 13 proteins, three proteins (RAB27B, DDAH2,and CACNG2) replicated in beta-amyloid, six proteins replicated in tangles (RAB27B, DDAH2, CACNG2, B3GLCT, GPAA1, and GPAA1), six proteins replicated in AD diagnosis (RAB27B, DDAH2, CACNG2, GPAA1, DCC, and ACADS), and five proteins (RAB27B, DDAH2, CACNG2, GPAA1, and ACADS) replicated in cognitive trajectory (Figure 3). Following the meta-analysis, three brain proteins (RAB27B and DDAH2were notable for replicating in all four AD features in consistent directions of association in both the discovery and replication cohorts (Figure 3). Taken together, these results further support the evidence of a molecular link between depression and elevated risk for AD through protein regulation in the brain.

Discussion

Prospective epidemiological studies in approximately 50,000 participants found an association between depression and elevated risk for dementia [5-8]. More recently, well-powered studies using AD and depression GWAS results identified a genetic correlation between these two diseases [14]. Thus, the goal of our study was to elucidate the genetic and molecular basis underlying this association. First, we confirmed the genetic correlation between depression and AD. Next, we used two-sample MR to identify that the preponderance of genetic evidence is consistent with a causal role of depression to AD but not vice versa. To elucidate the genes and molecular mechanisms of this relationship, we identified brain transcripts and proteins regulated by the depression-predisposing genetic variants and found that a subset of these were associated with AD diagnosis, AD pathologies, and cognitive trajectory. These findings suggest 46 specific transcripts and 8 proteins that are likely important contributors to the risk of AD pathology from depression.

Among the eight proteins that appear involved in both depression and AD risk, RAB27B, DDAH2, and CACNG2 are notable for their consistent association with all four AD features in both cohorts. RAB27B belongs to a protein family, Rab GTPases, mainly expressed in neurons, where they function to regulate different types of membrane trafficking, and have been implicated in neurodegenerative disorders [42, 43]. Of note, our finding of higher RAB27B expression being associated with AD characteristics is consistent with another study that found upregulation of RAB27B transcript and protein levels in postmortem cholinergic basal forebrain neurons of participants with mild cognitive impairment and AD [44]. DDAH2 is a member of the dimethylarginine dimethylaminohydrolase (DDAH) family of enzymes, where they maintain homeostatic control of nitric oxide across different tissues [45]. Immunostaining results showed that DDAH2 is specifically elevated in neurons demonstrating oxidative stress in AD patients, whereas it was undetectable in the neurons of age-matched healthy controls [46]. Consistent with such an observation, we found that higher DDAH2 expression was significantly associated with AD characteristics.

Our findings, which rely on the largest and most recent GWAS of AD and depression, are consistent with prior work using similarly powered GWAS summary results. The reasons for prior null results are likely due power limited power [13] and should not be viewed as

inconsistent with our findings. It is also consistent with recent work showing that individuals with higher polygenetic risk scores for major depression predicted the conversion from amnestic mild cognitive impairment to Alzheimer's disease [47]. And, our finding are also consistent a recent study demonstrating evidence for pleiotropy between depression and AD [11].

Our findings should be interpreted in light of the study's limitations. First, the MR assumptions were carefully examined, but the causative associations identified in this study are not definitive. Second, we analyzed 8606 proteins, which are not complete proteomic profiles, and thus deeper proteomic sequencing coverage can advance investigation of mechanisms underlying the associations between depression and AD. Third, this is an association study, and thus, further mechanistic studies in model systems are needed to validate our findings. Our study has several notable strengths. First, it is the first study to use deep human brain proteomic data to elucidate the molecular links between depression and AD. Second, we used data from the largest and latest GWAS available. Third, we performed both proximal and distal eQTL and pQTL analyses. Fourth, we used a discovery and replication design when examining the associations between brain proteins and AD features to increase the level of confidence in our findings. Fifth, we noted that some of these genes, to varying extents, correspond to well-known therapeutics within neurological and non-neurological diseases [48] and others have been nominated as promising therapies for future AD drug development [49].

In conclusion, we demonstrated that there is a genetic basis for the association between depression and elevated AD risk. Furthermore, we showed that depression genetic risk contributes to AD risk. Finally, we identified brain transcripts and proteins that likely underlie the contribution of depression to AD, which we nominate as promising candidates for further mechanistic studies in depression and AD.

Tables

Table 1: Summary of genome-wide association studies (GWAS) datasets

GWAS	Cohort	Disease	Sample size	Definition of cases	Case ascertainment
Howard, D. M., et al. 2019	23andMe+UKB+PGC	Depression	807,553	Broad definitions of depression: MDD and self-reported clinical diagnosis of depression	**23andMe cases** were ascertained based on self-report of depression phenotypes using web-based surveys . **UKB cases** were ascertained using a structured methodological review by MDD assessment experts. **PGC cases** were ascertained based on clinically-derived phenotypes for MDD using the international consensus criteria (DSM-IV, ICD-9, or ICD-10)
Jansen, I. E., et al. 2019	PGC-ALZ + IGAP + ADSP + UKB	AD	455,258	AD/AD-by-proxy	**PGC_ALZ, IGAP, and ADSP cases** were diagnosed by physician examination based on the NINCDS-ADRDA criteria, the international consensus criteria (ICD-10), the DSM criteria or autopsy confirmation. **UKB AD-by-proxy cases** were ascertained based on self-reported diagnoses of parents with AD. A sum of 47,793 individuals with one or both parents affected

Table 2 : Demographics of ROS/MAP and Banner participants

Note: Variables are shown as mean [range] or number (percentage)

Characteristic	ROS/MAP samples with transcriptomic data	ROS/MAP samples with proteomic data	Banner samples with proteomic data
Sample Size (complete genotype data)	587 (580)	381 (372)	125(97)
Female sex (%)	64%	70%	48%
Age at death [mean] (range)	86[67.4-108.3]	89[65.9-106.5]	86[73.0-103.0]
Education [mean] (range)	16[3.0-28.0]	16[5.0-28.0]	14[8.0-20.0]
Post-mortem interval [mean] (range)	7.3[1.0-40.8]	8.1[2.3-61.5]	2.95[1.25-5.5]
Depressive symptoms [mean] (range)	1.4[0.0-8.0]	1.3[0.0-7.1]	4.0[0.0-27.0]
Traditional AD Pathology			
Beta-amyloid burden [mean] (range)	4.20[0.00-19.92]	4.70[0.00-19.92]	9.46[0.0-15.0]
Tau tangle density [mean] (range)	6.30[0.00-78.52]	4.85[0.006-30.458]	8.86[0.75-15.00]
Clinical diagnosis of cognitive status at death			
Normal cognition (NCI) (%)	33%	41%	37%
Mild cognitive impairment (MCI) (%)	28%	26%	13%
Alzheimer's disease (AD) (%)	42%	32%	50%
Slope of cognitive trajectory [mean] (range)	0.00[-4.26-1.66]	0.078[-4.17-1.88]	-0.44[-4.94-0.22]

Figures

Figure 1. Generalized summary data-based Mendelian randomization (GSMR) results of (**A**) depression liability on AD. The forward GSMR analysis included 115 SNPs associated with depression at a genome-wide significance level (i.e., $p < 5e\text{-}8$). (**B**) AD liability on depression. The reverse GSMR analysis on 61 AD SNPs (at $p < 5e\text{-}8$) as the instrumental variables. Bonferroni-corrected significance threshold for 2 tests: $p < 0.05/2$. Bold represents a significant *p*-value. Abbreviations: AD : Alzheimer's disease; SNP: number of single nucleotide polymorphisms included in each GSMR analysis; β: effect size; p_{GSMR} : *p*-value for the causal estimates.

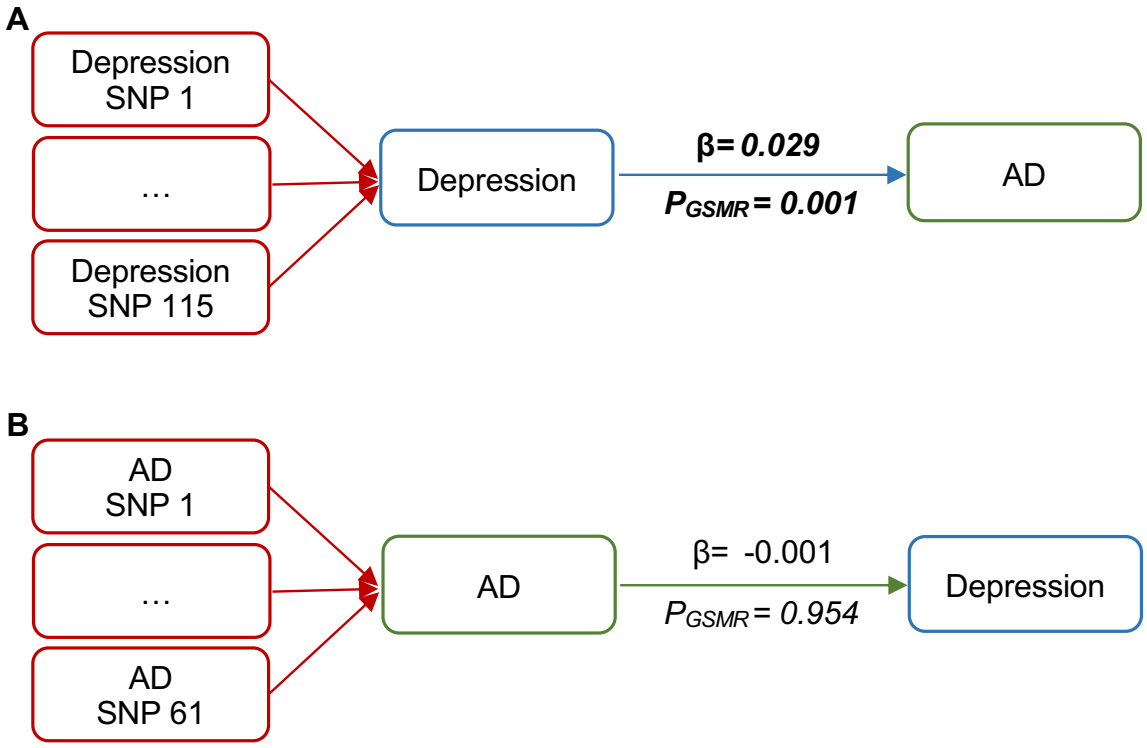


Figure 2. Identified brain proximal- and distal-QTLs from the 115 depression-associated SNPs. (**A**) 80 eQTLs were identified (green dotes). (**B**) 32 pQTLs were found (blue dotes). We labeled the gene names of the corresponding transcripts and proteins for the top eQTLs and pQTLs. Bold represents sites regulating the expression of both transcript and protein of the same gene.

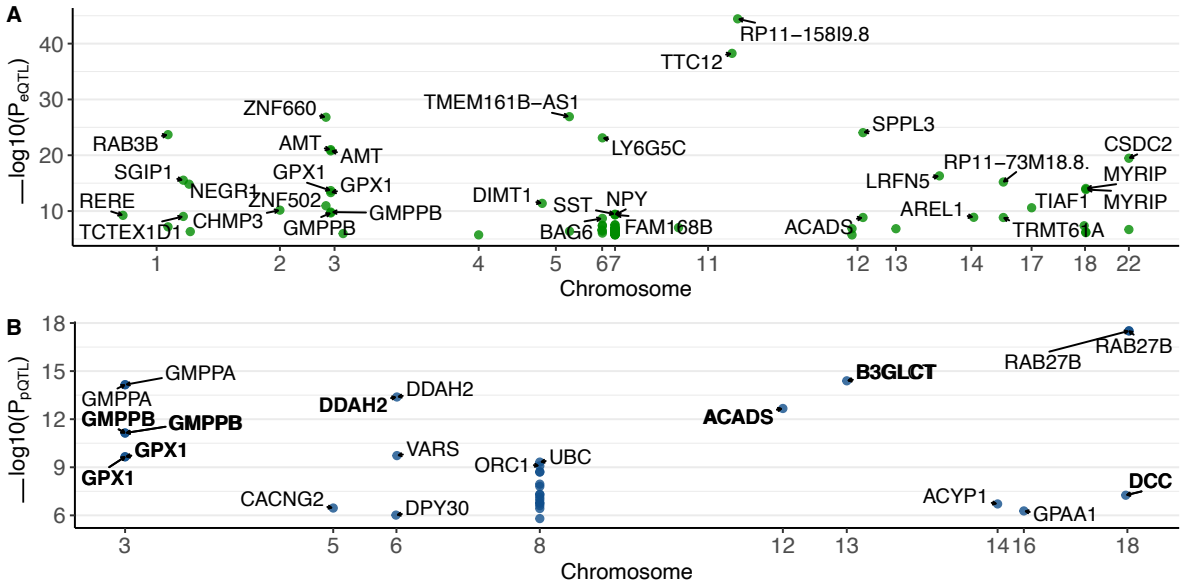
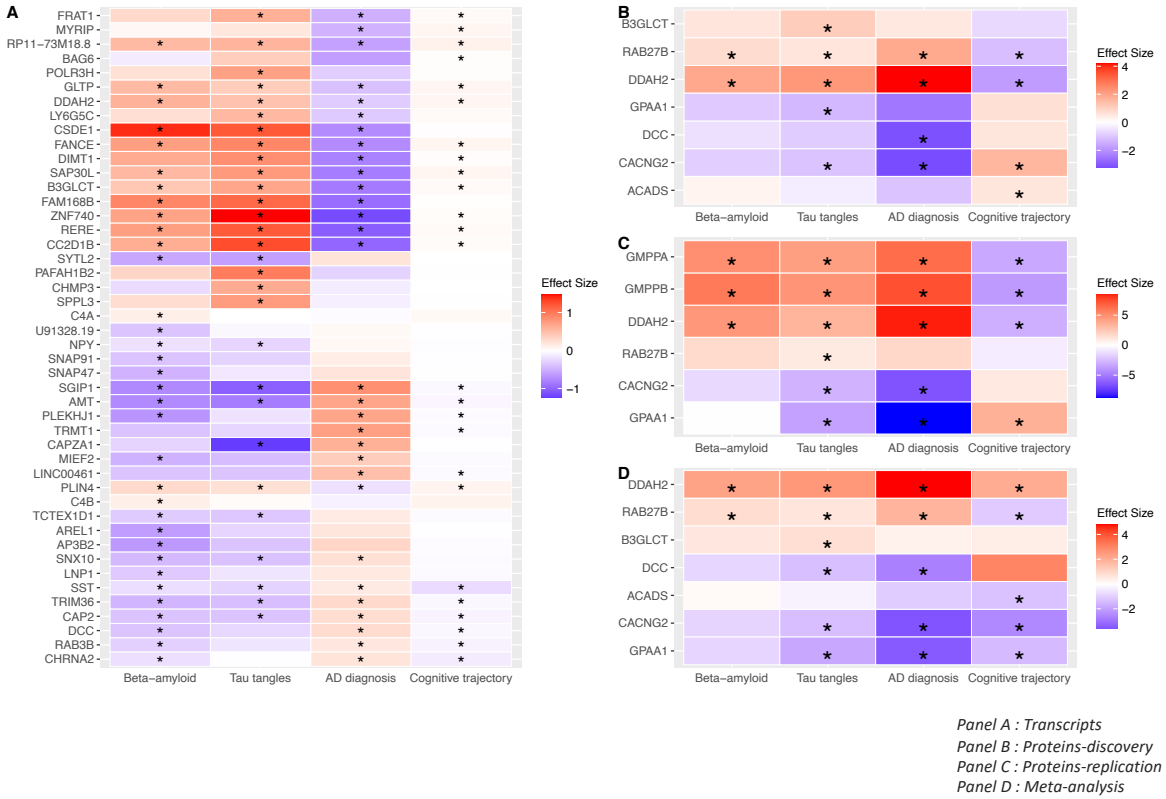


Figure 3. Heatmaps for associations between AD features and the brain transcripts and proteins regulated by the depression-associated SNPs (**A**) Heatmap of the 46 transcripts (*y*-axis) associated with AD-related traits (*x*-axis) in ROS/MAP samples. (**B**) Heatmap of seven proteins (*y*-axis) associated with AD-related traits (*x*-axis) in ROS/MAP samples. (**C**) Heatmap of six proteins (*y*-axis) associated with AD-related traits (*x*-axis) in Banner samples. The asterisk depicts the FDR $p < 0.05$. (**D**) Heatmap of seven proteins (*y*-axis) that replicated in both the discovery and replication analyses after a meta-analysis. The asterisk represents replication in the meta-analysis. The color reflects the direction of the association of expression levels with each AD-related trait.



Panel A : Transcripts
Panel B : Proteins-discovery
Panel C : Proteins-replication
Panel D : Meta-analysis

Reference:

1. Wilson, R.S., et al., *Depressive symptoms, cognitive decline, and risk of AD in older persons.* Neurology, 2002. **59**(3): p. 364.
2. Steffens, D.C., et al., *Prevalence of depression among older Americans: the Aging, Demographics and Memory Study.* Int Psychogeriatr, 2009. **21**(5): p. 879-88.
3. Saczynski, J.S., et al., *Depressive symptoms and risk of dementia: the Framingham Heart Study.* Neurology, 2010. **75**(1): p. 35-41.
4. *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.* Lancet, 2018. **392**(10159): p. 1789-1858.
5. Bellou, V., et al., *Systematic evaluation of the associations between environmental risk factors and dementia: An umbrella review of systematic reviews and meta-analyses.* Alzheimers Dement, 2017. **13**(4): p. 406-418.
6. Diniz, B.S., et al., *Late-life depression and risk of vascular dementia and Alzheimer's disease: systematic review and meta-analysis of community-based cohort studies.* Br J Psychiatry, 2013. **202**(5): p. 329-35.
7. Mourao, R.J., et al., *Depressive symptoms increase the risk of progression to dementia in subjects with mild cognitive impairment: systematic review and meta-analysis.* Int J Geriatr Psychiatry, 2016. **31**(8): p. 905-11.
8. Barnes, D.E., et al., *Depressive symptoms, vascular disease, and mild cognitive impairment: findings from the Cardiovascular Health Study.* Arch Gen Psychiatry, 2006. **63**(3): p. 273-9.
9. Ni, H., et al., *The GWAS Risk Genes for Depression May Be Actively Involved in Alzheimer's Disease.* J Alzheimers Dis, 2018. **64**(4): p. 1149-1161.
10. Arlt, S., et al., *Genetic risk factors for depression in Alzheimer`s disease patients.* Curr Alzheimer Res, 2013. **10**(1): p. 72-81.
11. Lutz, M.W., et al., *Shared genetic etiology underlying Alzheimer's disease and major depressive disorder.* Translational Psychiatry, 2020. **10**(1): p. 88.
12. Dafsari, F.S. and F. Jessen, *Depression—an underrecognized target for prevention of dementia in Alzheimer's disease.* Translational Psychiatry, 2020. **10**(1): p. 160.

13. Gibson, J., et al., *Assessing the presence of shared genetic architecture between Alzheimer's disease and major depressive disorder using genome-wide association data.* Transl Psychiatry, 2017. **7**(4): p. e1094.
14. Jansen, I.E., et al., *Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk.* Nature Genetics, 2019. **51**(3): p. 404-413.
15. Howard, D.M., 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**(3): p. 343-352.
16. Zhu, Z., et al., *Causal associations between risk factors and common diseases inferred from GWAS summary data.* Nature Communications, 2018. **9**(1): p. 224.
17. Bennett, D.A., et al., *Overview and findings from the rush Memory and Aging Project.* Curr Alzheimer Res, 2012. **9**(6): p. 646-63.
18. Bennett, D.A., et al., *Religious Orders Study and Rush Memory and Aging Project.* J Alzheimers Dis, 2018. **64**(s1): p. S161-s189.
19. Beach, T.G., et al., *Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program.* Neuropathology, 2015. **35**(4): p. 354-89.
20. Schneider, J.A., et al., *The neuropathology of probable Alzheimer disease and mild cognitive impairment.* Annals of neurology, 2009. **66**(2): p. 200-208.
21. Creavin, S.T., et al., *Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations.* Cochrane Database Syst Rev, 2016(1): p. Cd011145.
22. Wilson, R.S., et al., *Conscientiousness, dementia related pathology, and trajectories of cognitive aging.* Psychology and aging, 2015. **30**(1): p. 74-82.
23. Kohout, F.J., et al., *Two Shorter Forms of the CES-D Depression Symptoms Index.* Journal of Aging and Health, 1993. **5**(2): p. 179-193.
24. Hyman, B.T. and J.Q. Trojanowski, *Consensus recommendations for the postmortem diagnosis of Alzheimer disease from the National Institute on Aging and the Reagan Institute Working Group on diagnostic criteria for the neuropathological assessment of Alzheimer disease.* J Neuropathol Exp Neurol, 1997. **56**(10): p. 1095-7.

25. Hamilton, M., *The Hamilton Rating Scale for Depression*, in *Assessment of Depression*, N. Sartorius and T.A. Ban, Editors. 1986, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 143-152.
26. Wingo, A.P., et al., *Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain.* Nature Neuroscience, 2020. **23**(6): p. 696-700.
27. De Jager, P.L., et al., *A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research.* Sci Data, 2018. **5**: p. 180142.
28. Logsdon, B.A., et al., *Meta-analysis of the human brain transcriptome identifies heterogeneity across human AD coexpression modules robust to sample collection and methodological approach.* bioRxiv, 2019: p. 510420.
29. De Jager, P.L., et al., *A genome-wide scan for common variants affecting the rate of age-related cognitive decline.* Neurobiol Aging, 2012. **33**(5): p. 1017.e1-15.
30. Wingo, A.P., et al., *Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis.* Nature Genetics. **in press**.
31. Purcell, S., et al., *PLINK: a toolset for whole-genome association and population-based linkage analysis.* American Journal of Human Genetics, 2007. **81**: p. 559-575.
32. Manichaikul, A., et al., *Robust relationship inference in genome-wide association studies.* Bioinformatics, 2010. **26**(22): p. 2867-73.
33. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
34. Das, S., et al., *Next-generation genotype imputation service and methods.* Nat Genet, 2016. **48**(10): p. 1284-1287.
35. Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.* Nat Genet, 2015. **47**(3): p. 291-5.
36. Leek, J.T. and J.D. Storey, *Capturing heterogeneity in gene expression studies by surrogate variable analysis.* PLoS Genet, 2007. **3**(9): p. 1724-35.
37. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
38. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies.* Nature Genetics, 2006. **38**(8): p. 904-909.

39. Willer, C.J., Y. Li, and G.R. Abecasis, *METAL: fast and efficient meta-analysis of genomewide association scans.* Bioinformatics, 2010. **26**(17): p. 2190-1.
40. Yang, J., et al., *GCTA: A Tool for Genome-wide Complex Trait Analysis.* The American Journal of Human Genetics, 2011. **88**(1): p. 76-82.
41. Haidich, A.B., *Meta-analysis in medical research.* Hippokratia, 2010. **14**(Suppl 1): p. 29-37.
42. Chan, C.C., et al., *Systematic discovery of Rab GTPases with synaptic functions in Drosophila.* Curr Biol, 2011. **21**(20): p. 1704-15.
43. Yu, E., et al., *Role of Rab27 in synaptic transmission at the squid giant synapse.* Proceedings of the National Academy of Sciences, 2008. **105**(41): p. 16003.
44. Ginsberg, S.D., et al., *Upregulation of select rab GTPases in cholinergic basal forebrain neurons in mild cognitive impairment and Alzheimer's disease.* Journal of chemical neuroanatomy, 2011. **42**(2): p. 102-110.
45. Leiper, J.M., et al., *Identification of two human dimethylarginine dimethylaminohydrolases with distinct tissue distributions and homology with microbial arginine deiminases.* Biochem J, 1999. **343 Pt 1**(Pt 1): p. 209-14.
46. Smith, M.A., et al., *Dimethylargininase, a nitric oxide regulatory protein, in Alzheimer disease.* Free Radical Biology and Medicine, 1998. **25**(8): p. 898-902.
47. Xu, J., et al., *Neurobiological substrates underlying the effect of genomic risk for depression on the conversion of amnestic mild cognitive impairment.* Brain : a journal of neurology, 2018. **141**(12): p. 3457-3471.
48. Corsello, S.M., et al., *The Drug Repurposing Hub: a next-generation drug library and information resource.* Nat Med, 2017. **23**(4): p. 405-408.
49. Greenwood, A.K., et al., *Agora: An open platform for exploration of Alzheimer's disease evidence.* Alzheimer's & Dementia, 2020. **16**(S2): p. e046129.

Chapter 3: Conclusions and recommendations for future studies

The overarching objective of this study is to identify the shared molecular underpinnings between depression and AD. In Chapter 2, I used largest GWAS results of AD and depression to test whether genetic variation predisposing to depression contributes to AD. There are three key ways that this study advances the current knowledge of the links between these two conditions. First, it confirmed the significant genetic correlation between depression and AD. Second, it demonstrated that depression has genetic evidence consistent with a causal role on AD, but not vice versa. Third, it identified candidate molecular targets that underlie the contribution of depression to AD. These transcripts and proteins are promising candidates for mechanistic studies to understand further the mechanisms by which depression contributes to AD for prevention and early treatment of AD.

The results and methods presented here lay the groundwork for future investigations of the molecular mechanisms behind how depression modifies AD risk. The conclusions drawn rely heavily on data from GWAS results and community-based cohorts of primarily European ancestry individuals. Future investigations that include diverse population samples are needed to understanding whether these conclusions are shared or unique to different races/ethnicities.

The approach of integrating human brain transcriptomic and proteomic data with GWAS signals can be applied to numerous other “omics” expression phenotypes such as microRNA, DNA methylation, or histone modification, which may also be relevant in the context of depression and AD pathology. Furthermore, single-cell omic platforms might identify and prioritize the specific cellular mechanisms that are the primary drivers of depression and AD pathology. These investigations should be further explored in future studies.

Overall, this thesis outlines the potential utility of omics-based profiles to improve brain health research. This study adds to the body of knowledge surrounding the contribution of depression to the development of AD. This study suggests a highly credible hypothesis of the genetic and molecular mechanisms linking depression to AD. Future work connecting other neuropsychiatric traits and AD is needed to further characterize modifiable targets for prevention and early AD treatment.