**Understanding the etiology of Alzheimer’s disease using genetically predicted gene expression and brain feature models**

 A Thesis Submitted to

The University of Chicago Biological Sciences Collegiate Division

Towards a Degree of Bachelor of Science in Biological Sciences with Research Honors

Evan Wu

Under the Supervision of Hae Kyung Im, Ph.D.



The University of Chicago

Chicago, Illinois

May 2022

**Table of Contents**

[Acknowledgements](#acknowledgements)……………………………………………………………………………..3

[Abstract](#abstract)…………………………………………………………………………………………..5

[Introduction](#intro)

[*Overview of Alzheimer’s disease*](#intro1)……………………………………………………...6

[*Genetic studies of Alzheimer’s disease*](#intro2)………………………………………………7

[*Regulation of gene expression in Alzheimer’s disease*](#intro3)……………………………..9

[*Structural brain features associated with Alzheimer’s disease*](#intro4)…………………...11

[Results](#results)

[*MultiXcan identifies 7 novel genes associated with Alzheimer’s disease*](#results1)……….12

[*Novel genes are implicated in known AD tissues and pathways*](#results2)…………………16

[*S-BrainXcan implicates white matter degeneration in AD*](#results3)………………………...19

[*MultiXcan detects genes responsible for significant IDP associations*](#results4)…………..26

[Discussion](#discussion)………………………………………………………………………………………27

[Methods](#methods)…………………………………………………………………………………………29

[Supplementary Information](#supplement)…………………………………………………………………...31

[References](#references)……………………………………………………………………………………..42

**Acknowledgements**

I would like to thank my principal investigator Dr. Hae Kyung Im introducing me to the world of statistical genetics and being an incredible mentor, giving me the tools and guidance to be able to succeed not only on this thesis but also in my future research endeavors. Thank you for always having the patience and availability to answer any questions and help solve the frequent bugs and errors that come up in our day-to-day work. You are also an incredible resource of biological and statistical knowledge and general wisdom, and I hope to continue learning for you in the time we still have together. The experiences I have had working in your lab have allowed me to grow in so many ways, and I am immensely grateful for all that you have done for me, the lab, and the statistical genetics community over the past few years.

I would also like to thank the rest of the current and previous lab members: Yanyu, Festus, Natasha, Sabrina, and others, for always being insightful, helpful, and welcoming in our correspondences. My work would have been many times harder without their assistance, and I am grateful that I got to work with such intelligent and kind people.

I would like to thank my thesis committee members: Dr. Andy Dahl, Dr. Guimin Gao, and Dr. Xuanyao Liu, for overseeing my progress in this study and providing the expertise and insight necessary to make this thesis scientifically rigorous. I would also like to thank them for their role in our statistical genetics mega-group, which has enriched my experience through the sharing resources and knowledge across the genetic medicine department. I really appreciate the group’s welcoming and accommodating atmosphere that has allowed for the inclusion of a diverse range of people interested in statistical genetic methods, which is a direct result of your incredible attitudes towards science and collaboration. Thank you for advising me through this now year-long project.

I would like to thank the rest of the honors program: Dr. D. Allan Drummond and my peers in providing criticism, guidance, and community over this past year.

Finally, I would like to thank my friends and family members who have supported me through the good times and the bad at the University of Chicago and have allowed me to reach this point. I would not be here without all of your contributions.

**Abstract**

Treatment of Alzheimer’s disease is one of the foremost public health concerns in the modern era. The genetic component of Alzheimer’s remains under-explained and is a promising direction of inquiry to further understand the complex mechanisms underlying the disease and develop effective treatment methods. In light of this, we applied recent genetic association techniques, PrediXcan and BrainXcan, on the most recent Alzheimer’s disease GWAS data, which leverage newly available large-scale transcriptome and brain imaging data to enhance the detection of trait-associated genes and brain features. We detected seven novel genes associated with Alzheimer’s disease and found suggestive signals that structural white matter atrophy predisposes disease development. While these are exciting steps towards elucidating the causes of Alzheimer’s disease, additional mechanistic studies will be needed to fully understand the biological signals captured by these results.

**Introduction**

*Overview of Alzheimer’s disease*

Alzheimer’s disease (AD) is a neurodegenerative disease and the most common form of dementia, which is a syndrome characterized by deteriorating cognitive functions. Dementia affects around 55 million people worldwide and is projected to affect up to 78 million people in 2030, with Alzheimer’s disease making up around 60-70% of cases [1]. Currently, there are no treatments that can cure or halt the progression of Alzheimer’s disease. Modern medical practice instead focuses on early diagnosis, risk reduction, and symptomatic relief for at-risk and diseased patients [1, 2], but an AD diagnosis is always fatal within 3-9 years [3]. AD also has an inordinately large disease burden: it is the 6th leading cause of death among adults in the US and total societal costs are projected to reach up to $2.8 trillion annually by 2040 [1]. As the exact etiology of AD remains unclear and no promising cures have been found after over a century of research on the disease, it is now more urgent than ever to identify novel insights in AD pathology that can aid the development of effective, curative therapies.

AD pathology is thought to begin up to 20 years in advance of symptom manifestation, the earliest observable stage of which is called mild cognitive impairment (MCI), where one or a few cognitive domains are impaired with little functional deterioration [2, 3]. The disease then increases in severity with degeneration of short-term memory, speech, spatial processing, and executive functions, eventually leading to near total loss of independence and neuropsychiatric changes like delusions, hallucinations, and emotional dyscontrol [1, 2]. While age is the most important risk factor for both dementia and AD, with disease incidence doubling every 5 years in adults >65, neither are regarded as natural outcomes of aging [1,3].

On the molecular level, AD is characterized by the presence of abnormal protein aggregates: extracellular amyloid-β (Aβ) plaques and intracellular tau neurofibrillary tangles [2, 3]. The historical amyloid cascade hypothesis posited that the accumulation of Aβ is the causal mechanism behind AD pathology by initiating a long-term cascade of inflammation, synaptic dysfunction, and neuronal death that eventually becomes recognizable as dementia [4, 5]. Eventually, the hypothesis came into question due to the presence of Aβ plaques not being sufficient to induce AD, and the failure of multiple clinical trials with drugs that remove Aβ plaques to improve or cure AD [4]. Further research into the etiology of AD has led to a multitude of hypotheses and a modern understanding of the disease as extremely complex and multifaceted. Along with Aβ and tau, other biological processes that have been linked to AD include lysosome and endosome dysfunction, aberrant cell cycle control, vascular injury, neuroinflammation, mitochondrial dysfunction, cholesterol metabolism, and synaptic dysfunction [2, 3, 4, 5]. Much of our understanding of these pathways came from investigation into the genetic basis of AD.

*Genetic studies of Alzheimer’s disease*

The advent of genome-wide studies over the past 15 years has greatly informed our understanding of the genetic basis of complex diseases. The simple genome-wide association study (GWAS) compares the genotypes of individuals with varying disease phenotype across millions of genetic markers to identify variants or single nucleotide polymorphisms (SNPs) that are significantly associated with disease [6]. The GWAS methodology has been hugely successful, with over 10,000 significant SNPs identified for hundreds of complex traits that are highly replicable and biologically relevant, leading to advancements in our knowledge and treatment of many previously intractable diseases [6]. With sample sizes for modern studies reaching up to a million individuals and continued technological improvements, GWAS continues to be a powerful tool to identify the genetic architecture of complex diseases.

Alzheimer’s disease has been separated into two distinct forms based on differing genetic architecture. Early-onset AD (EOAD) represents 1-5% of overall cases and is characterized by onset of symptoms prior to 65 years of age and a heritability of over 90%, which suggests rare variants with a strong effect are responsible for these cases [5]. In particular, autosomal dominant mutations in *APP*, *PSEN1*, and *PSEN2* are thought to cause familial forms of EOAD [7, 8]. On the other hand, late-onset AD (LOAD) is the predominant form of AD accounting for >95% of cases. The heritability of LOAD is estimated to be around 60-80% [5, 7, 8], but the disease has a complex and polygenic genetic component. Previous GWAS of LOAD have identified over 40 independent loci associated with disease, including highly significant and replicated risk genes like *APOE*, *CLU*, *CR1*, and *PICALM*, [5, 7, 9, 10]. However, most of the known SNPs are common variants with low effect sizes, and in total account for only 30% of the estimated AD genetic variation [5, 10]. The most recent AD GWAS’s with greater sample sizes continue to identify new common variants reaching genome-wide significance [11, 12], suggesting that disease arises from a cumulative or network effect between individually low-impact SNPs. This is supported by the polygenicity of AD: polygenic enrichment in AD is highly significant (*p* = 4.9 × 10−26) and using polygenic risk scores allows for over 80% accuracy in predicting AD for individuals diagnosed or with high genetic risk of disease [5]. The estimated number of causal SNPs responsible for AD is contentious, with one study predicting <100 variants and another predicting 100-10,000 variants [12]. Nevertheless, the current number of causal SNPs identified falls greatly below both figures, and future GWAS’s will continue to contribute to our understanding of AD genetic architecture.

However, there are several caveats to GWAS that have limited the usefulness of results obtained from these studies. First, GWAS design limits most of the SNPs used to be common biallelic variants for cost and convenience at the scale of sample sizes up to a million. This makes GWAS liable to miss rare causal variants without prior enrichment for such variants in samples or imputation panels [5, 6, 13]. In addition, because of the large number of variants procured, linkage disequilibrium (LD) concealing the true causal variants, and the unknown biological effects of non-coding variants, translating significant GWAS hits into a targetable functional study is prohibitively difficult [6, 13]. In light of these shortcomings, we propose the use of cutting-edge functional genomic techniques that use additional large-scale -omics data to enhance detection and contextualization of genetic associations and glean new insights into the etiology of AD.

*Regulation of gene expression in Alzheimer’s disease*

PrediXcan is one method within a family of analyses now known as transcriptome-wide association studies (TWAS) that bridges the functional gap between GWAS results and biology by considering the effect of genomic variation on gene expression. Changes in gene expression due to the effect of regulatory variants have been shown to robustly explain a large portion of phenotype variability for many diseases [14] and are readily interpretable as gene-level biological interactions. Additionally, the availability of large-scale expression quantitative trait locus (eQTL) data like the Genotype-Tissue Expression (GTEx) consortium, which has both genotype and expression level data for almost 1000 individuals in 54 tissue types in the current V8 release [15], forms the basis of PrediXcan and like approaches. By aggregating sets of biologically informed SNPs and incorporating their observed transcriptomic effects in a reference dataset, PrediXcan allows for “imputation” of the unobserved transcriptome in a sample of interest, which is associated with a complex trait to probe for genetic effects that arise from differential gene expression. The approach has been shown to have multiple advantages that allow it to complement results from GWAS. PrediXcan results are interpretable as gene-level variation associated with disease, and there is increased power to detect trait-associated loci due to aggregation of multiple SNP signals, smaller multiple testing burden, and the ability to probe different tissue contexts based on the eQTL dataset [14].

While PrediXcan operates on individual-level genotype data, the related method S-PrediXcan is generally more convenient as it can produce the same results from publicly available GWAS summary statistics without significant inaccuracies or loss of power [16]. A few recent studies have performed PrediXcan and/or other TWAS methods on AD with promising results, recapitulating prior GWAS hits and identifying novel AD targets [17, 18, 19]. Given these successes, we performed S-PrediXcan on AD using summary statistics from the newest available GWAS with unprecedented sample sizes [11, 12] to try to identify novel AD-associated genes, and investigated the support for targets identified based on implications from previous statistical or mechanistic studies on AD.

*Structural brain features associated with Alzheimer’s disease*

In addition to eQTL data, an exciting mode of genetic functional -omic studies is now emerging through the application of large-scale magnetic resonance imaging (MRI) data to identify macroscopic changes that occur in organs affected by disease. MRI is a noninvasive imaging technique that uses nuclear magnetic resonance to produce detailed cross-sectional images of internal structures, including the brain [20]. Historically, brain MRI studies of AD relied on case-control studies of a few individuals due to their cost, limiting the power of detection and reliability of these studies due to low statistical power and lack of standardization, in addition to concerns of reverse causality [20, 21]. Despite these limitations, AD has been robustly associated with medial temporal atrophy in structural MRI studies [22, 23] and reduced white matter integrity in functional MRI studies [23, 24], giving insight into the actual changes in the brain caused by and/or causing disease.

The ongoing UK Biobank study is producing quality-checked and processed brain, heart, and body MRI imaging data from 100,000 participants with structural and functional annotations [25], an unprecedented sample size of ready-to-use neuroimaging data that comes with deep genotype and phenotype data for these individuals, with data released for 25,000 individuals so far. BrainXcan is a recent imaging-wide association platform that extends the principles of PrediXcan to brain features or image-derived phenotypes (IDPs), regressing genetically inferred brain IDPs based on UK Biobank reference data against the phenotype of interest [26]. This analysis returns significant brain features associated with disease and has the potential to provide valuable insights into disease susceptibility and progression. Brain MRI-derived phenotypes have been shown to be heritable [27], but they have a more downstream and complex relationship to the underlying genetic architecture compared to the transcriptome. BrainXcan includes a Mendelian randomization module to test the direction of causality between significant IDPs and the complex trait [26]. Analogous to PrediXcan and S-PrediXcan, S-BrainXcan is the BrainXcan method that operates on GWAS summary statistics instead of individual data. We therefore applied S-BrainXcan to the most recent AD GWASs to discover novel insights into brain structure and functions altered in AD pathology on an unprecedented scale for AD MRI studies.

**Results**

*MultiXcan identifies 7 novel genes associated with Alzheimer’s disease*

Alzheimer’s disease is a disorder that primarily affects the brain, so *a priori* we expect relevant gene regulatory systems to be dysfunctional in brain tissues. As such, we first performed S-PrediXcan on the Schwartzentruber et al. [11] AD GWAS summary statistics using each of the 13 GTEx brain tissue models available. We then aggregated these results using S-MultiXcan, a multivariate regression approach that treats each S-PrediXcan result as a correlated experiment, meta-analyzing across tissues to increase power to detect significant associations [28]. While the individual brain tissue runs identified an average of around 39 significant genes (Supplementary Table 1), the S-MultiXcan run identified a total of 90 significant genes after Bonferroni correction, highlighting this increase in statistical power. The most significant gene identified was *APOE* (*p* ≈ 0, *z*­mean = 20.349), which is expected as *APOE* is the most significant AD GWAS hit.Outside the *APOE* region on chromosome 19, the most significant gene was *BIN1* (*p* = 5.959 × 10-36, *z­*mean = -4.796), which is also a highly significant locus found in previous AD GWAS studies [5]. We confirmed that newly detected genes going from single-tissue to aggregate analysis were unlikely to be driven by false positives by plotting the z-scores of all genes from each tissue against the mean z-score from S-MultiXcan. These plots showed close adherence to a one-to-one relationship in all tissues, indicating general agreement between the analyses (Supplementary Figure 1A). There was also no significant genomic inflation, as seen by the QQ plot of the association p-values (Supplementary Figure 1B). The hippocampus is robustly implicated in AD, especially in earlier stages of disease [22, 23], making it the foremost candidate for a causal tissue. To make sure that no potentially causal signals from this region were missed by meta-analysis with other brain tissues, we examined the significant genes within the hippocampus and found no unique signals compared to the S-MultiXcan analysis.

One of the findings of studies using large-scale eQTL data like GTEx is that cis-eQTLs are generally shared and increasing eQTL sample size greatly improves the number of associations returned for tests like PrediXcan [28, 29, 30]. In light of this, we ran S-MultiXcan on the Schwartzentruber et al. data using all 49 GTEx V8 tissue models to further increase our power to detect associations. During this process, the summary statistics for the now most recent AD GWAS performed by Wightman et al. using over 1 million individuals [12] was released, so we also performed S-MultiXcan with all available tissues on this dataset. The two GWASs overlap heavily as both use IGAP and UK Biobank samples, though there are differences in AD-by-proxy sample selection from the UK Biobank. S-MultiXcan identified 134 significant genes for the Schwartzentruber run, and 109 significant genes for the Wightman run after Bonferroni correction (Figure 1). In both analyses, the most significant gene identified was *NECTIN2* or *PVRL2* in the *APOE* region (Schwartzentruber: *p* ≈ 0, *z­*mean = -4.796; Wightman: *p* ≈ 0, *z*­mean = -6.159), and *BIN1* outside the *APOE* region (Schwartzentruber: *p* = 1.672 × 10-37, *z*mean = -3.559; Wightman: *p* = 4.935 × 10-37, *z*mean = -3.136), which is consistent with a previous TWAS study performing S-MultiXcan on AD [17]. As before, we found no signs of disagreement between individual tissue and S-MultiXcan analyses nor significant genomic inflation (Supplementary Tables 2 and 3, Supplementary Figures 2 and 3). Only a single gene, *SIGLEC11*,was significant in the brain-only Schwartzentruber S-MultiXcan run but not in the all-tissue S-MultiXcan run. The full set of significant genes identified by S-MultiXcan of Schwartzentruber with brain regions, Schwartzentruber with all tissues, and Wightman with all tissues are listed in the supplement due to their length (Supplementary Tables 4-6).

Significant genes detected by TWAS approaches may be false hits due to co-regulation or pleiotropy of nearby genes, correlated predicted expression, and bias from non-trait-related tissues [31]. For example, the multitude of highly significant signals we found within the *APOE* region are likely not independent of a causal signal from *APOE*. Because of this, we grouped genes into 37 approximately independent LD blocks [32] and set the criterion for novel genes as not containing a locus or gene that was previously identified in an AD GWAS or TWAS study [5, 11, 12, 13, 17, 18, 19]. This does not mean that the genes that share an LD block with a previously identified locus (>100 genes) are not causal. Many have been implicated in AD-related pathways through mechanistic studies, and the polygenic nature of AD means we do not necessarily expect a sparse causal signal. To cast as wide of a net as possible, we took the union of genes in all three runs (166 genes total) and performed a comprehensive literature review on these genes to assess additional evidence supporting a gene’s role in AD pathology (Supplementary Table 7). Nevertheless, given the large number of these genes and the difficulty in disentangling their potential effects from established AD GWAS risk loci, we focus here on the seven novel genes: *RP11-138I18.2*, *AC012370.3*, *SIGLEC11*, *LRP4*, *ZNF652*, *SUSD3*, and *PITPNA*.

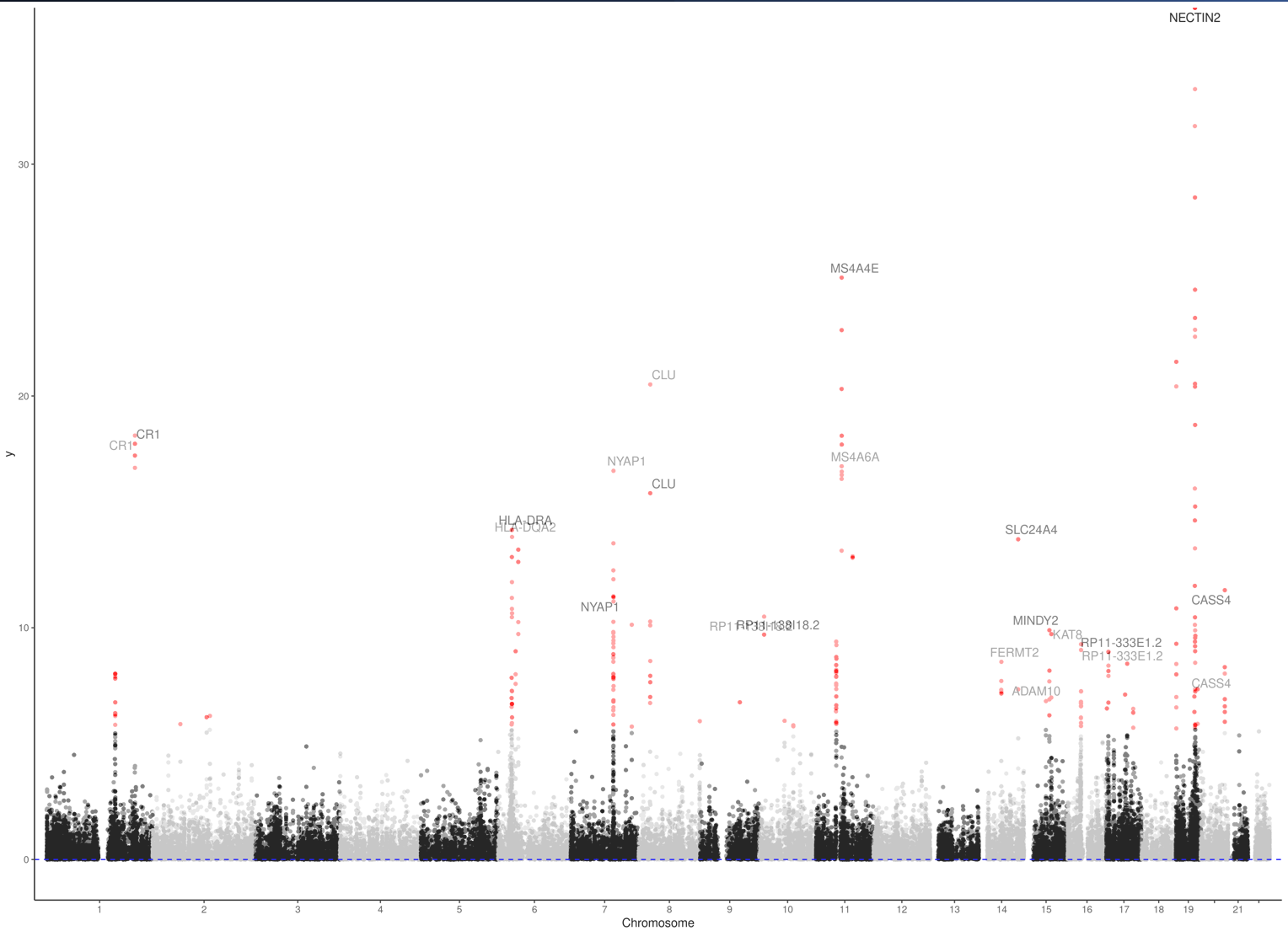


Figure 1: Manhattan plot of significant genes identified from the all-tissue MultiXcan runs. The two runs using Schwartzentruber et al. and Wightman et al. data are overlayed, with significant genes colored in red. Each gene locus grouped by independent LD blocks is labelled with the most significant gene. The *APOE* locus and *BIN1* are not resolved as the y-axis is limited to -log1035 to show less significant loci.

*Novel genes are implicated in known AD tissues and pathways*

*RP11-138I18.2* was the only novel gene significant in all three runs. This gene along with *AC012370.3* are both long noncoding RNAs with unknown function, though such genes are known to regulate the expression of other genes, alternative splicing, mRNA stability, and protein translation, and a few in particular are theorized to play a role in AD [33]. Further mechanistic studies, particularly into *RP11-138I18.2* as it is the most well-replicated gene from this analysis, will be required to understand what role these types of genes might play in AD.

*SIGLEC11* was the only gene that was significant in the brain-tissue-only Schwarzentruber run, but not in the all-tissue Schwartzentruber run. It encodes for the sialic acid-binding, immunoglobulin-like lectin receptor-11, which recognizes sialylated glycoproteins and glycolipids and is thought to mediate immunosuppressive signals in microglia. The presence of sialylated glycoproteins and gangliosides in extracellular Aβ may therefore act through Siglec-11 to facilitate immune avoidance of these plaques [34]. The gene was also identified in a recent unpublished AD GWAS by the International Genomics of Alzheimer’s Project using rare variant imputation [35]. Despite these suggestive lines of evidence, *SIGLEC11* has yet to be definitively linked to AD pathogenesis through mechanistic studies.

*LRP4* encodes for the low-density lipoprotein receptor-related protein 4, a receptor for Agrin which was known for its role in formation and maintenance of the neuromuscular junction and has recently been found to be a general synaptic organizer, being implicated in peripheral nerve regeneration, central nervous system development, cognitive function and plasticity, and adult hippocampal neurogenesis across various mechanistic studies [36]. In particular, a recent study showed LRP4 protein is specifically expressed in astrocytes and promotes Aβ uptake. Depletion of *LRP4* in an AD mouse model led to deficits in neurotransmission, hippocampal and prefrontal cortex synchrony, and cognition [37]. To our knowledge, statistical genetic studies of AD have not previously implicated this locus, so this study provides additional independent evidence for a causal role of *LRP4*.

*ZNF652* produces the zinc finger protein 652 transcription factor, which previous studies have implicated primarily in different cancers, especially breast cancer, suggesting a role in tumorigenesis [38]. While no mechanistic studies linking *ZNF652* to AD have been performed, one study performing independent component analysis in AD gene expression microarray data found that ZNF652 along with other metal protein genes are upregulated [39], and another study examining genetic regulation of microglia-specific gene expression changes in AD identified *ZNF652* as a causal gene [40]. As *ZNF652* appears to regulate cell cycle pathways in breast tissue, it could play a similar role in microglial or neuronal cells in AD, which requires further studies to validate.

*SUSD3* or sushi domain containing 3 is a relatively uncharacterized gene but has been implicated in breast cancer with a role in cancer cell proliferation and cell adhesion and migration mechanisms [41]. However, a recent study investigating microglial gene expression in AD showed *SUSD3* has an expression profile significantly related to the microglial marker *TMEM119*, signifying enrichment in microglia and a potential role in AD [42]. In addition, an imaging-wide association analysis, which is a similar approach to TWAS but integrating GWAS data with imaging endophenotypes, found *SUSD3* to be significantly associated with AD [43]. Despite these discoveries by statistical studies, the biological relation of *SUSD3* to AD remains unclear.

*PITPNA* encodes for the phosphatidylinositol transfer protein alpha, a lipid-binding protein that is involved in synaptic vesicle recycling. In the triple-transgenic Alzheimer mouse, this gene is significantly downregulated, which suggests that there is degradation of synaptic integrity in this AD mouse model [44]. However, a problem with AD mouse models is that they do not exhibit neurodegeneration. A more recent study using AD-like vervet monkeys which display similar pathological symptoms to AD also found *PITPNA* to be significantly downregulated, which makes the previously detected association more robust [45].

In summary, of the seven novel genes we identified through S-MultiXcan, two genes were noncoding and uncharacterized, while the other five coding genes showed support from the literature to be playing some role in AD pathology. Of the coding genes, *LRP4* was the only one that showed direct support from mechanistic studies implicating it in AD, while the others were generally supported through identification by independent statistical methods, especially profiling gene expression in AD-related tissues. As such, these novel genes represent an exciting step into further understanding the genetic and biological underpinnings of AD.

*S-BrainXcan implicates white matter degeneration in AD*

We next performed S-BrainXcan using the Schwartzentruber and Wightman GWAS summary statistics to understand what physiological features in the brain are genetically predicted to be associated with disease. We examined 261 total IDPs from two different MRI modalities: 109 T1-weighted MRI features that quantify cortical, subcortical, and cerebellar brain region volumes and 152 diffusion MRI (dMRI) features that measure various properties of white matter microstructure [26]. These measurements included intracellular volume fraction (ICVF), isotropic volume fraction (ISOVF), orientation dispersion (OD), and fractional anisotropy (FA), and were interpreted following recommendations from [47]. ICVF measures the water fraction within neurites and is a proxy for neurite density, so lower ICVF would signify that the white matter bundles in the region are less dense and possibly damaged by disease. ISOVF measures the free water fraction and can be thought of as the fraction of extracellular or cerebrospinal fluid, as such increased ISOVF is analogous to lower ICVF. OD measures the alignment of neurite fibers along a single axis. White matter tracts are highly organized bundles of neurons, thus increased OD in these regions would indicate dispersed directions of diffusion and degradation of their orderly structure. FA measures the restriction of diffusion in all directions, and is similar to OD but less specific; decreased FA in white matter would indicate white matter degeneration. These markers in the degenerative direction were detected across the brain in various studies examining dMRI in AD [47]. We also included as IDPs the principal components of larger gray matter regions as proxy for brain-wide volume changes e.g., PC-Cortical-1 as volume of the cerebral cortex, and principal components of dMRI measurements as proxies for brain-wide measurements e.g., PC-TBSS-ICVF-1 as neurite density across the brain [26].

The Schwartzentruber run identified eight significant IDP associations after multiple testing correction. Based on previous MRI studies, we expect to see significant IDP associations indicating grey and white matter atrophy, particularly reduced volumes in the medial temporal lobe and hippocampus. However, the results from this run do not appear to agree with this intuition. From the T1 MRI IDPs, a greater in volume of the right hemisphere of the pallidum is the only IDP significantly associated with AD (Figure 2A), which is not consistent with previously observed neurodegeneration in this region in AD patients [46]. Similarly, for dMRI IDPs only one out of the seven IDPs is suggestive of neurodegeneration (Figure 2B): lower ICVF in the left superior fronto-occipital fasciculus. The other significant IDPs included lower OD in the inferior cerebellar peduncle and higher FA in the right external capsule and right cerebral peduncle, both of which are suggestive of retained tract integrity. We also see a significant positive association between the principal component for ICVF and AD, which would suggest that overall neurite density is increased throughout the brain [47], inconsistent with reduced overall neurite density observed in neurodegenerative diseases. Mendelian randomization (MR) analysis was significant in the IDP affecting disease direction for PC-ICVF (*p* = 0.00288), but unavailable or inconclusive for all other significant IDPs (Supplementary Figure 4). Among the inconclusive cases, there was no trend in lower p-values for a particular direction across the MR analyses, so it is not clear whether these associations as a whole are disease-causing or consequences of disease.

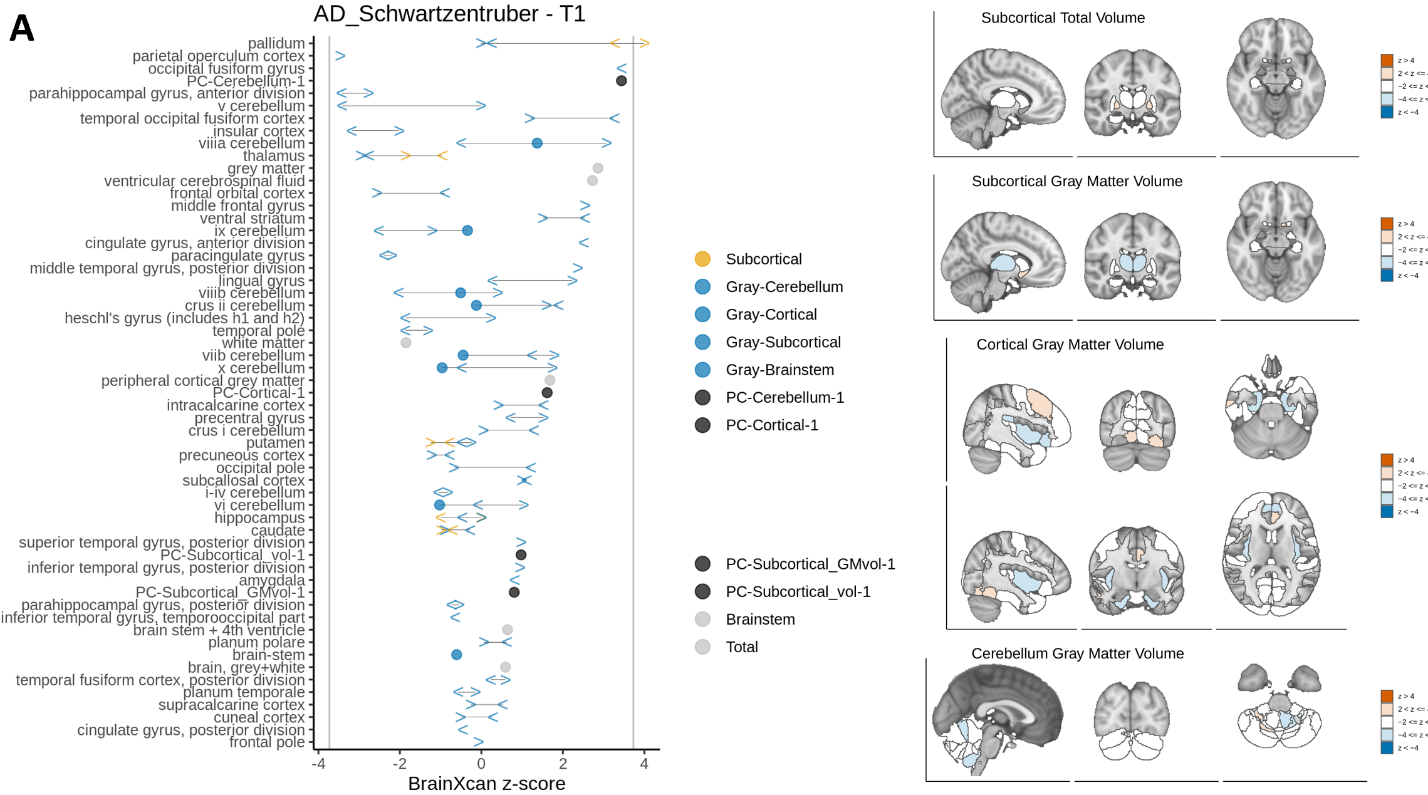
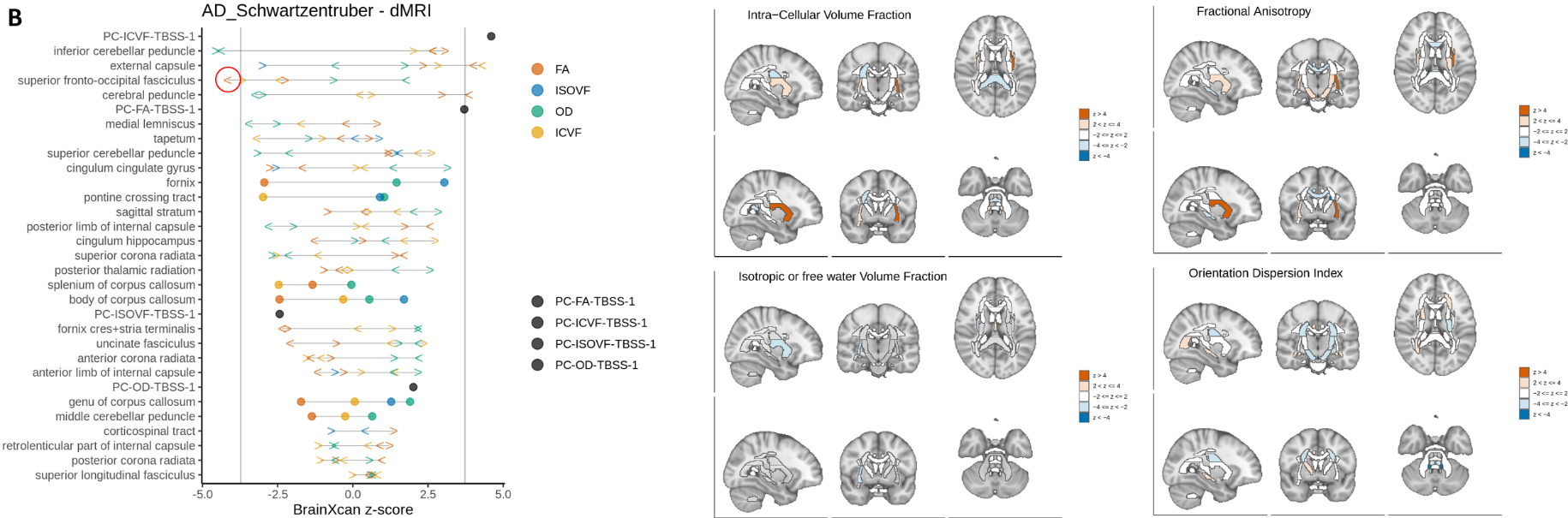


Figure 2: S-BrainXcan of Schwartzentruber et al. GWAS summary statistics showing z-scores of (A) T1 MRI and (B) dMRI IDPs and their respective visualizations in the brain. Gray vertical lines represent the significance threshold; IDPs with |z-score| > |threshold| are considered significant. Red circles indicate IDP associations that are consistent with neurodegeneration. Below are visualizations of significant IDPs and their direction of effect.

The Wightman run identified a total of 20 significant IDP associations (Figure 3) after multiple testing correction. There are seven significant T1 MRI IDPs which, similar to the Schwartzentruber run, do not agree with neurodegeneration within these regions as all of them are positive associations between regional volume and disease. These were higher volumes in the left temporal occipital fusiform cortex, left caudate, pallidum, ventricular cerebrospinal fluid, left occipital fusiform gyrus, and left cingulate gyrus (anterior division). On the other hand, of the thirteen significant dMRI IDPs, seven are consistent with signs that we would expect of white matter degeneration. These included lower ICVF in the superior fronto-occipital fasciculus; higher ISOVF in the fornix; higher OD in the genu of the corpus callosum and the PC; and lower FA in the left superior fronto-occipital fasciculus, fornix, and medial lemniscus. The six significant IDPs not supporting white matter atrophy are lower OD in the right tapetum, right superior corona radiata, and left external capsule; increased FA in the left external capsule; and increased ICVF in the right cerebral peduncle. Three IDPs were significant and had the same direction of effect between the two runs with Schwartzentruber and Wightman data: higher subcortical volume in the pallidum, higher FA in the external capsule, and lower FA in the left superior fronto-occipital fasciculus. Causal direction inference from MR was mostly unavailable or inconclusive, with only increased volume in the left temporal occipital fusiform cortex being marginally significant (*p* = 0.056) for the disease causing IDP direction (Supplementary Figure 5).

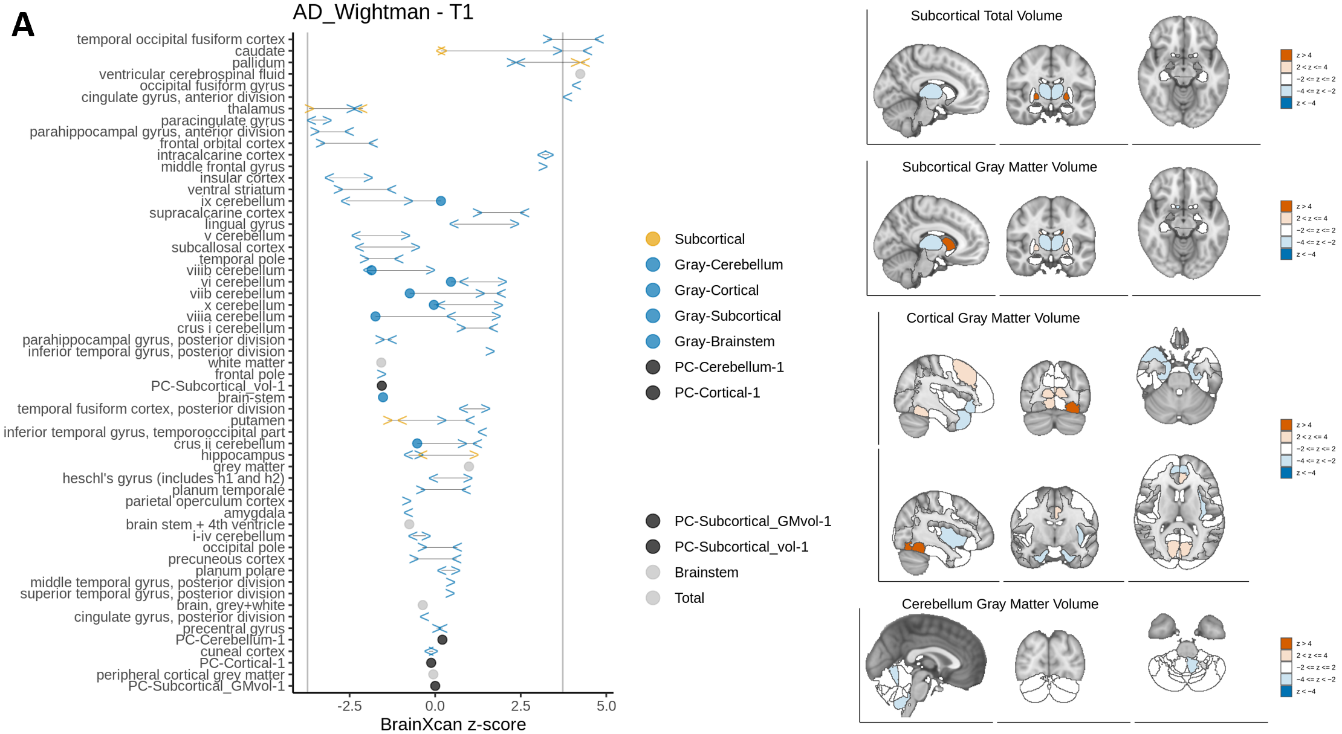
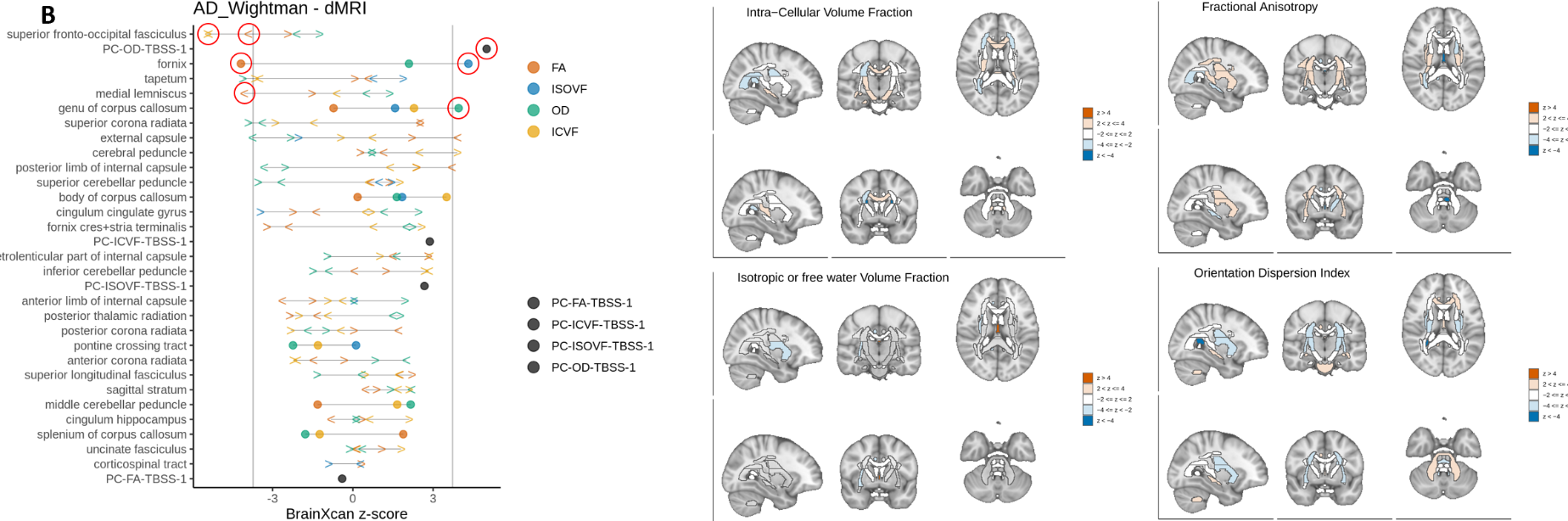


Figure 3: S-BrainXcan of Wightman et al. GWAS summary statistics showing z-scores of (A) T1 MRI and (B) dMRI IDPs and their respective visualizations in the brain. Gray vertical lines represent the significance threshold; IDPs with |z-score| > |threshold| are considered significant. Red circles indicate IDP associations that are consistent with neurodegeneration.

The results from the two runs are difficult to resolve. We appear to be detecting some shared signal at least in terms of direction of IDP effects, as correlation of the z-scores between the two runs is *r* = 0.665 (Supplementary Figure 6). However, the number of significant IDPs and their implicated changes in the brain differed greatly between the analyses. Nevertheless, as the Wightman GWAS study had substantially greater sample size and performing S-BrainXcan on it identified many more significant IDPs, we focus our interpretation on this run, temporarily putting aside the putative conflicts with the Schwartzentruber run.

The BrainXcan results for the Wightman run appear to support an association between AD and white matter atrophy as the majority of significant IDPs indicate neurodegeneration in these regions, which is consistent with previous MRI studies [24]. However, we did not detect significant associations with gray matter degeneration, especially hippocampal atrophy, which is the earliest structural change in AD [22] and is thought to cause sequential disruption of white matter bundles [48]. As BrainXcan predicts brain structure using the genomes and imaging features of healthy individuals as reference, we propose that reduced white matter integrity is a consequence of genetic risk that promotes AD pathogenesis (Figure 4). This hypothesis puts into question whether grey matter degeneration is actually upstream of disease initiation, suggests it may not be detectable using germline prediction of brain features. Given the distant relationship between the genome and brain structural features, and lack of evidence from our MR analysis to support the direction of causality these significant IDPs are acting in, our analysis at the moment is not well-powered to infer causality. These implications will require further mechanistic studies to explain and may be greatly informed by inclusion of a significant cohort of imaging data from patients currently with AD in the BrainXcan models.

Diagram

Description automatically generated

Figure 4: Evidence from BrainXcan for genetically predicted changes in brain structure and integrity implicate white matter degradation (colored red), which conflicts with the current model for the sequence of neurodegeneration in AD pathology (colored black).

*MultiXcan detects genes responsible for significant IDP associations*

To better understand the genetic architecture of IDPs, we performed MultiXcan on each IDP used in the S-BrainXcan analysis. We were especially interested to see if the genes that are associated with IDPs that were significant in S-BrainXcan were shared with AD-associated genes from S-MultiXcan. Across the 25 unique significant IDPs identified from S-BrainXcan of the Schwartzentruber and Wightman summary statistics, MultiXcan of only three IDPs resulted in a significant gene after Bonferroni correction: ISOVF in the fornix, total ventricular cerebrospinal fluid volume, and left cingulate gyrus volume (Table 1). The small number of significant genes detected is likely due to two factors: first our sample size is limited to the individuals in the UK Biobank with IDP phenotype data which is only around *n* = 25,000, and second IDPs are extremely polygenic and unlikely to be controlled by a single gene [26]. The lack of power in our genotype data is compounded by the polygenicity of our traits, which explains the sparsity of our results. Nevertheless, looking at the significant genes we found, one stands out in particular: *CHRNA5* was significant in two IDPs of different modality and brain region. *CHRNA5* encodes for a subunit of neuronal nicotinic acetylcholine receptors (nAChR), which is a ligand-gated ion channel that has been implicated in AD, schizophrenia, and epilepsy [49]. However, there does not seem to be much evidence for involvement of the *CHRNA5* gene in particular with AD. We did detect another nAChR subunit as a significant gene in our AD S-MultiXcan analysis: *CHRNA2*, which has support as an AD risk gene [50]. At the very least, this suggests that nAChR alterations affect both the relevant IDPs and AD, which warrants further investigation into the protein’s normal and pathological functioning. The other significant gene identified was *CELA3B*, a chymotrypsin-like elastase which plays a role in pancreatitis and is only known to be expressed in the pancreas [51]. It does not appear to be implicated as an AD risk gene, so the interpretation of this gene’s role in the brain is unclear.

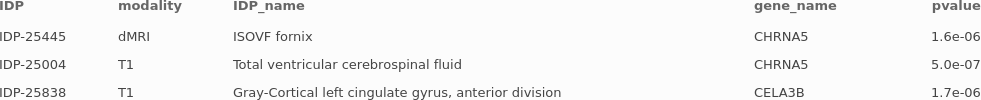


Table 1: Significant genes identified from MultiXcan of significant IDPs.

**Discussion**

In this study, we analyzed Alzheimer’s disease through the lens of genetically predicted gene expression and brain image-derived features. We discovered seven novel genes that have not previously been detected in AD GWAS or TWAS studies. These seven genes represent a promising starting point to further elucidate the biological pathways that are altered or disrupted in AD and are accompanied by over 100 additional genes whose signals are entangled with known AD risk loci, but nonetheless may be novel signals. Indeed, evidence from independent studies and methods suggests a putative role in AD for most of these genes. Previous TWAS studies have used fine-mapping of causal gene sets (FOCUS) to prioritize causal genes from a candidate gene set [49]. However, our focus on aggregate-level S-MultiXcan analyses makes FOCUS unsuitable as currently it is only applicable to single tissue TWAS outputs. In the future, developing a fine-mapping method that is applicable to aggregated TWAS outputs like MultiXcan will allow for statistical determination of causal genes.

From S-BrainXcan analysis, we found limited evidence that degradation of white matter microstructure is associated with AD primarily through using the Wightman data. Conflicts with the Schwartzentruber run may be due to the lack of statistical power from the GWAS studies or the BrainXcan models and may improve when more of the imaging data is released. We also observed putative contradictions with previous MRI studies, particularly a lack of hippocampal degeneration being associated with disease. This is most likely due to limitations of the BrainXcan methodology, as MRI case-control studies generally observed post-mortem or diagnosed patients as opposed to the predominantly healthy individuals included in the UK Biobank MRI data. This is a potential problem for most psychiatric disorders but is exacerbated for LOAD as disease begins in old age, which means expected structural changes like hippocampal atrophy are even less likely to be detected as these features may not be present even if we know an individual will develop AD later in life. As such, we favor interpreting S-BrainXcan results as genetic predisposition to brain features that could either initiate abnormal biological processes early on or make it easier for the core causal event to emerge or spread its influence. As white matter tracts generally function to connect and relay information between grey matter regions, dysfunction in these regions may suggest that brain dysconnectivity plays a role in AD pathology beyond general neurodegeneration. This hypothesis will require future work to fully test and may be aided by more powerful BrainXcan models once the full suite of IDP data is released, or longitudinal imaging studies on individuals with high genetic risk for AD.

Performing MutiXcan of significant IDPs from S-BrainXcan of AD revealed an interesting target, the nAChR protein that appears to be modified in some of the significant IDPs as well as AD through different *CHRNA* subunits. The sparsity of genes from this analysis and the polygenicity of IDPs suggests that increasing the size of the cohort with IDP phenotype data may allow for the identification of many more meaningful associations. Further characterization of the genetic architecture of both significant and non-significant IDPs can be achieved using gene set enrichment analysis or a hierarchical network-based approach that identifies significantly altered genes and IDP subnetworks. In summary, this study was able to confirm some of the results from previous studies and identify novel starting points in terms of genes and brain structural features for further mechanistic investigation into AD etiology and pathology.

**Methods**

*Summary statistics of most recent Alzheimer’s disease GWASs*

We downloaded publicly available GWAS summary statistics from the Schwartzentruber and Wightman studies to perform AD S-PrediXcan. The Schwartzentruber study was a meta-analysis using case-control samples from IGAP (21,982 cases and 41,944 controls) and familial AD status in the UK Biobank (898 cases, 52,791 proxy cases, and 355,900 controls), for a total of 473,515 samples across 10,687,126 variants [11]. The publicly available data from the Wightman study excluded 23andMe data, and similarly included both case-control and proxy data from IGAP, UK Biobank, DemGene, TwinGene, STSA, deCODE, Finngen, GR@CE, HUNT, BioVU, Gothenburg H70 Birth Cohort Studies and Clinical AD from Sweden, and ANMerge with a total of 1,126,563 samples (90,338 cases and 1,036,225 controls) over 12,688,340 variants [12]. Detailed processing and quality control procedures can be found in the respective studies [11, 12].

*S-PrediXcan and S-MultiXcan of Alzheimer’s disease*

S-PrediXcan requires z-scores from GWAS summary statistics and weights and SNP covariances generated from an eQTL study to estimate the gene expression weights and associate them with the phenotype of interest. We used MASHr tissue models (13 brain regions, 49 total) generated from the latest GTEx V8 prediction models to perform S-PrediXcan. Afterwards, we ran S-MultiXcan on individual tissue results. Runs with the Schwartzentruber and Wightman summary statistics incorporated over 90% of model SNPs, which we deemed sufficient to continue analysis without imputation. This allowed us to test 22,341 and 22,301 out of 22,535 genes, respectively. We adjusted the significance threshold of results using Bonferroni correction for both single-tissue and S-MultiXcan runs.

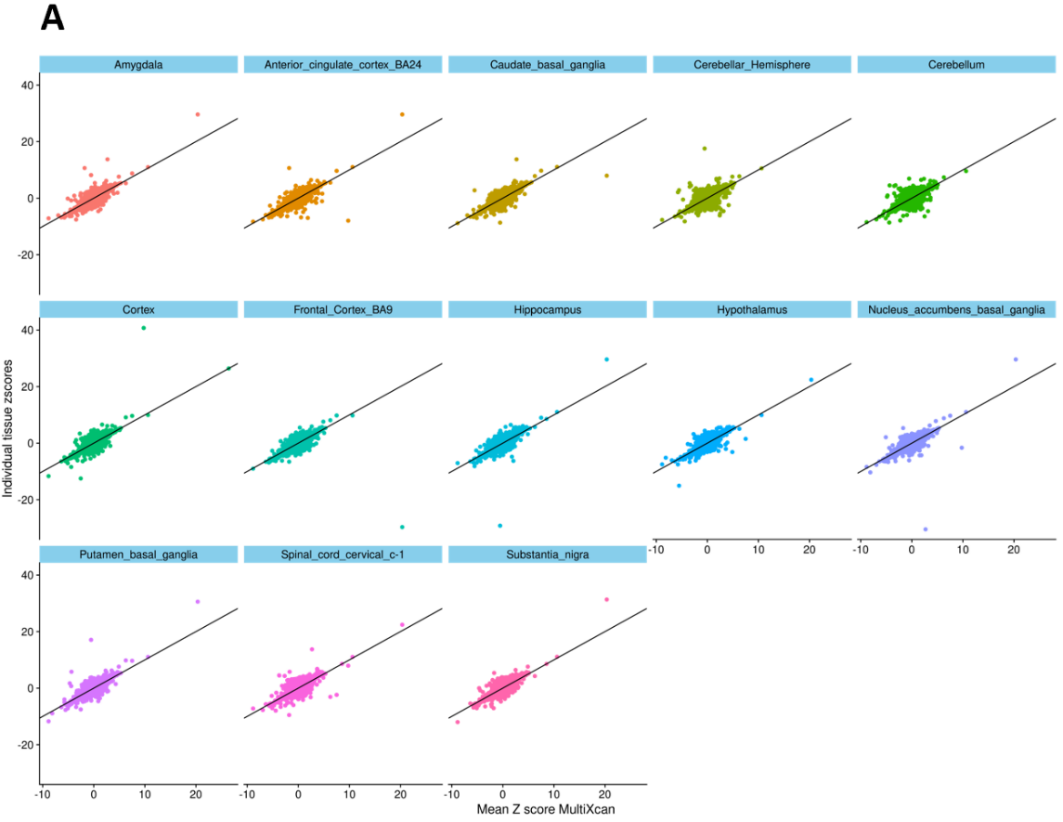
*S-BrainXcan of Alzheimer’s disease and Mendelian randomization*

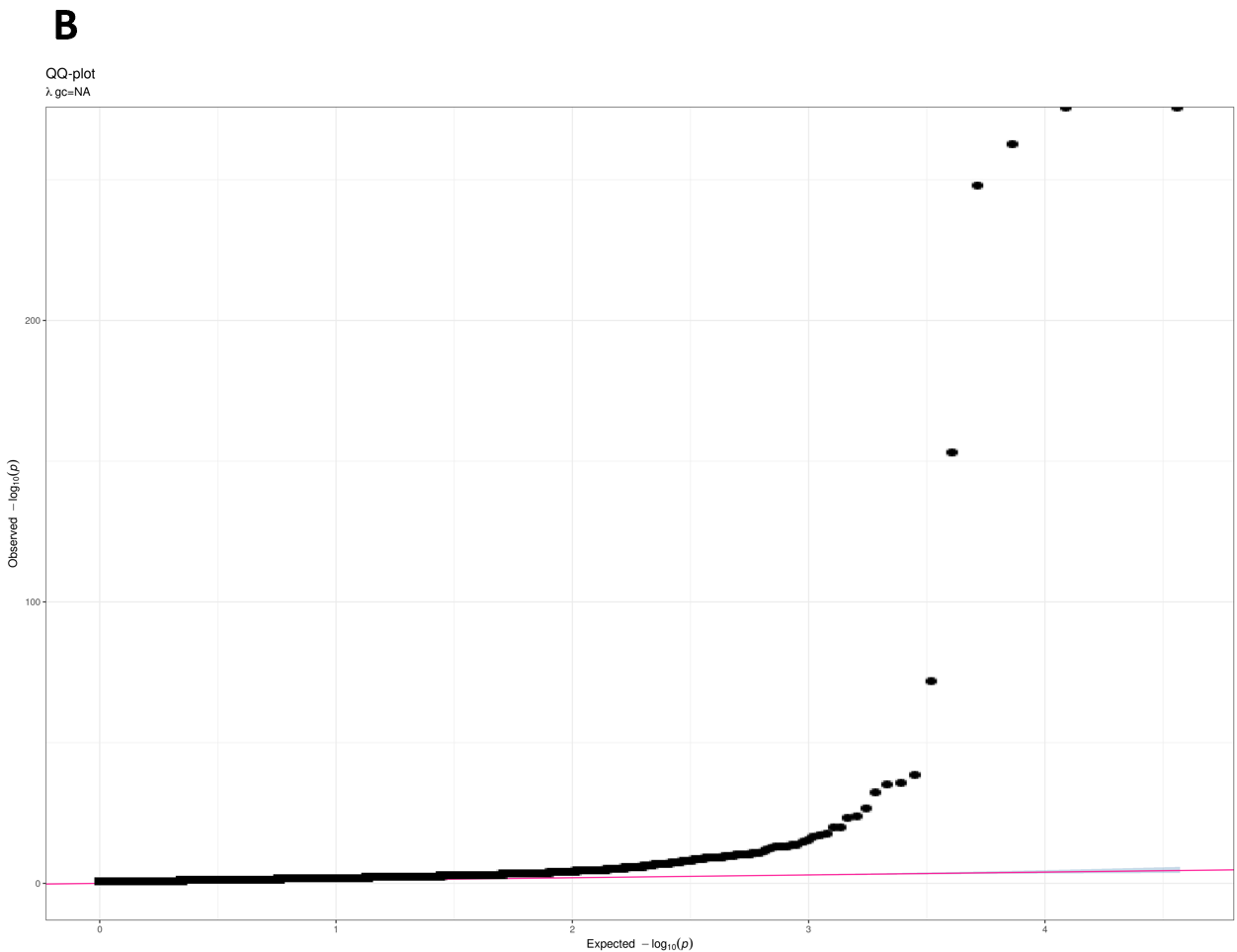
S-BrainXcan uses z-scores from GWAS summary statistics and data from the BrainXcan database including SNP covariances and prediction weights from either the elastic net or ridge regression model to estimate the brain IDP phenotypes and associate them with the phenotype of interest. We tested 327 IDPs with Spearman correlation ≥ 0.101, although plots focused on 261 IDPs (109 T1 and 152 dMRI) with significance level adjusted by Bonferroni correction. We also used BrainXcan’s MR module to determine the causal relationship between IDP and trait for the top 10 significant associations for each MRI modality if possible.

*MultiXcan of IDPs in the UK Biobank*

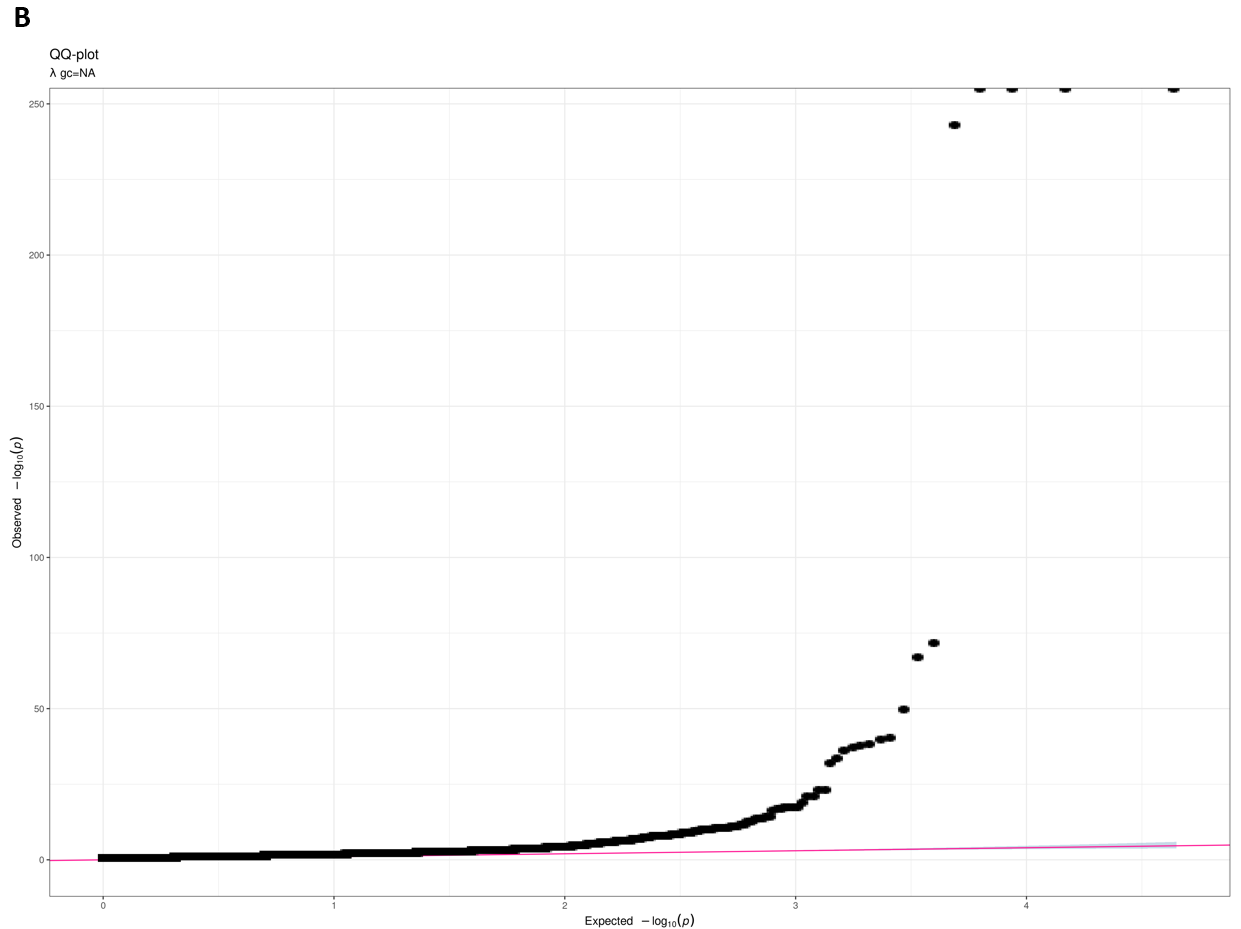
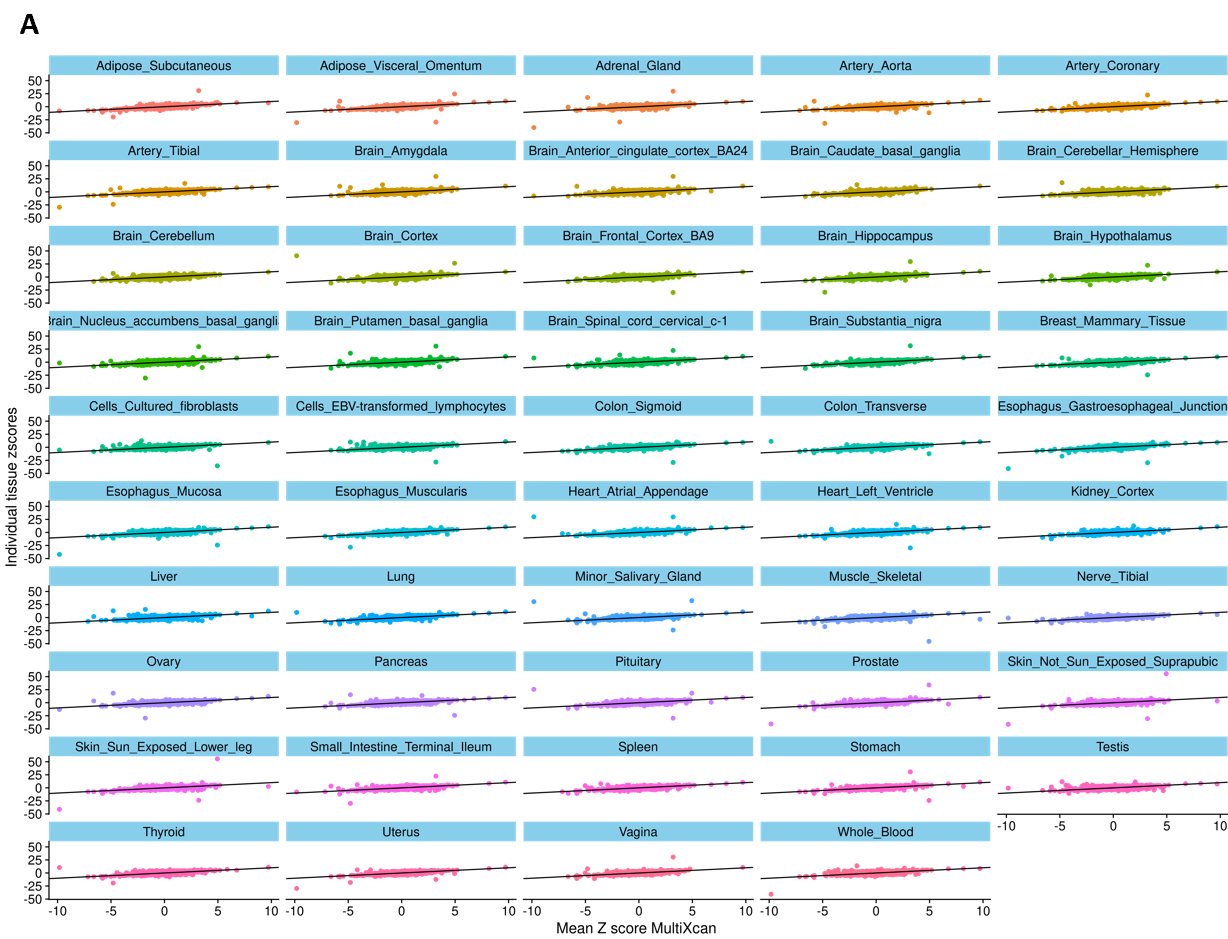
We queried genotypes of the 24409 individuals with IDP phenotype data from the UK Biobank for variants with MAF > 0.01 and predicted individual-level gene expression using the 49 GTEx V8 elastic net models, which used over 90% of model SNPs. We then performed MultiXcan for each available IDP on the predicted expression matrices with 21,211 genes included in each test, and Bonferroni correction.

**Supplementary Information**

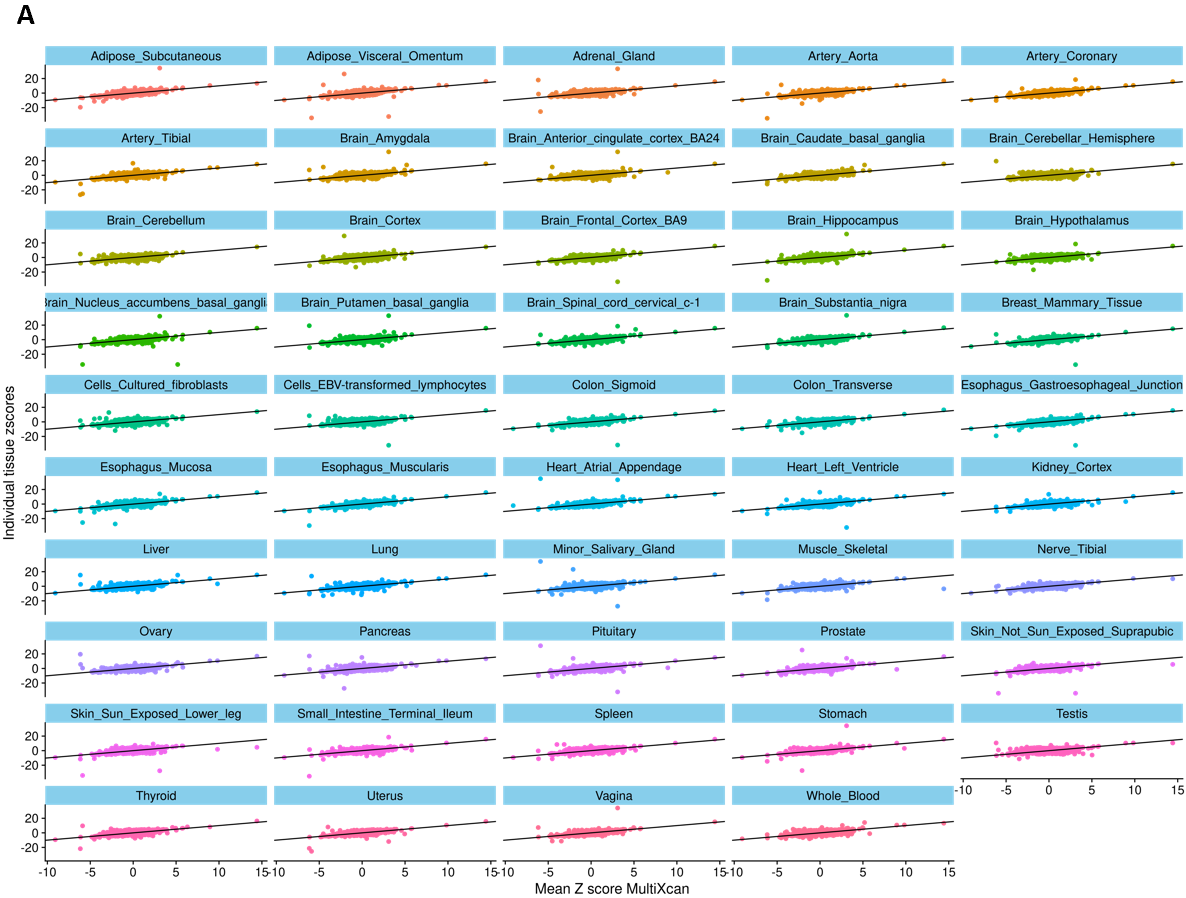
****

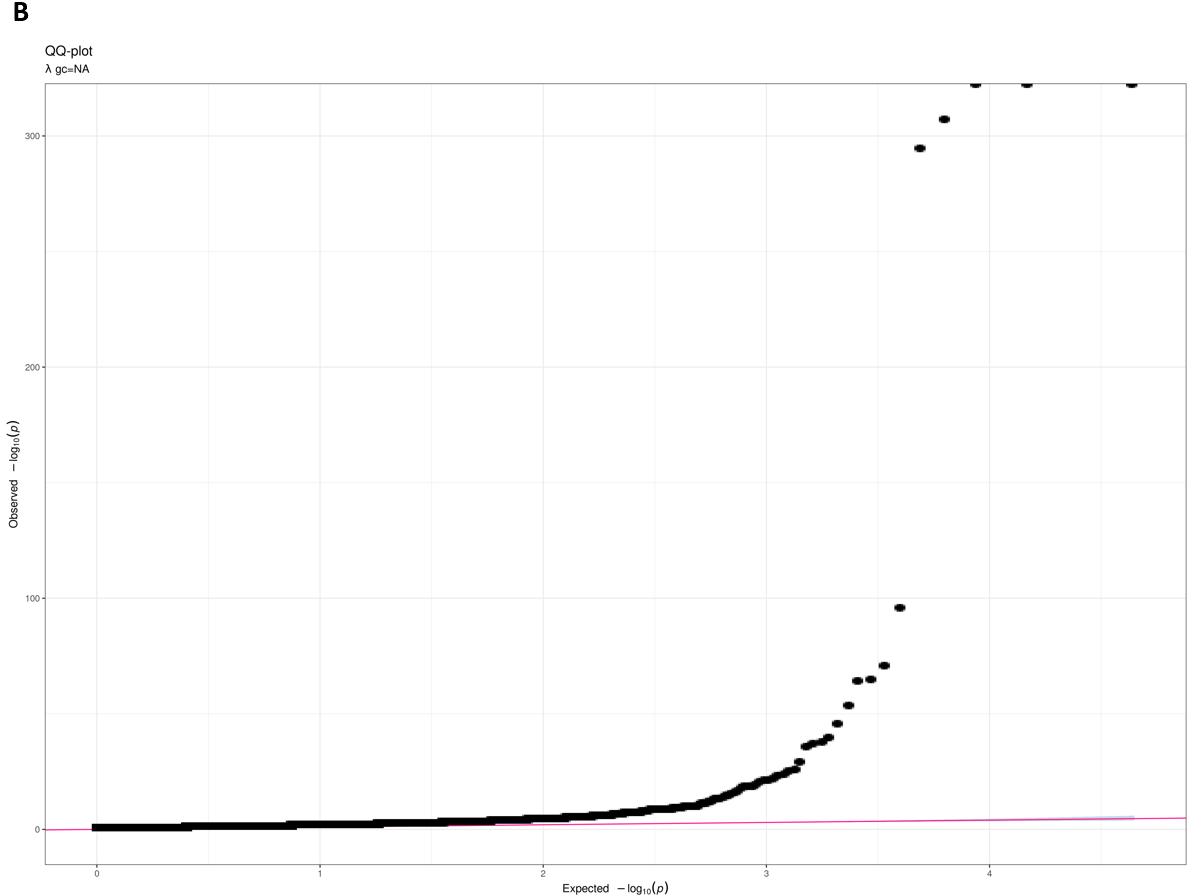
****

Supplementary Figure 1: (A) Z-scores of genes from single tissue S-PrediXcan plotted against mean z-score from brain-tissue-only S-MultiXcan using Schwartzentruber summary statistics. (B) QQ plot of Schwartzentruber brain-tissue-only S-MultiXcan p-values.



Supplementary Figure 2: (A) Z-scores of genes from single tissue S-PrediXcan plotted against mean z-score from all tissues S-MultiXcan using Schwartzentruber summary statistics. (B) QQ plot of Schwartzentruber all tissue S-MultiXcan p-values.

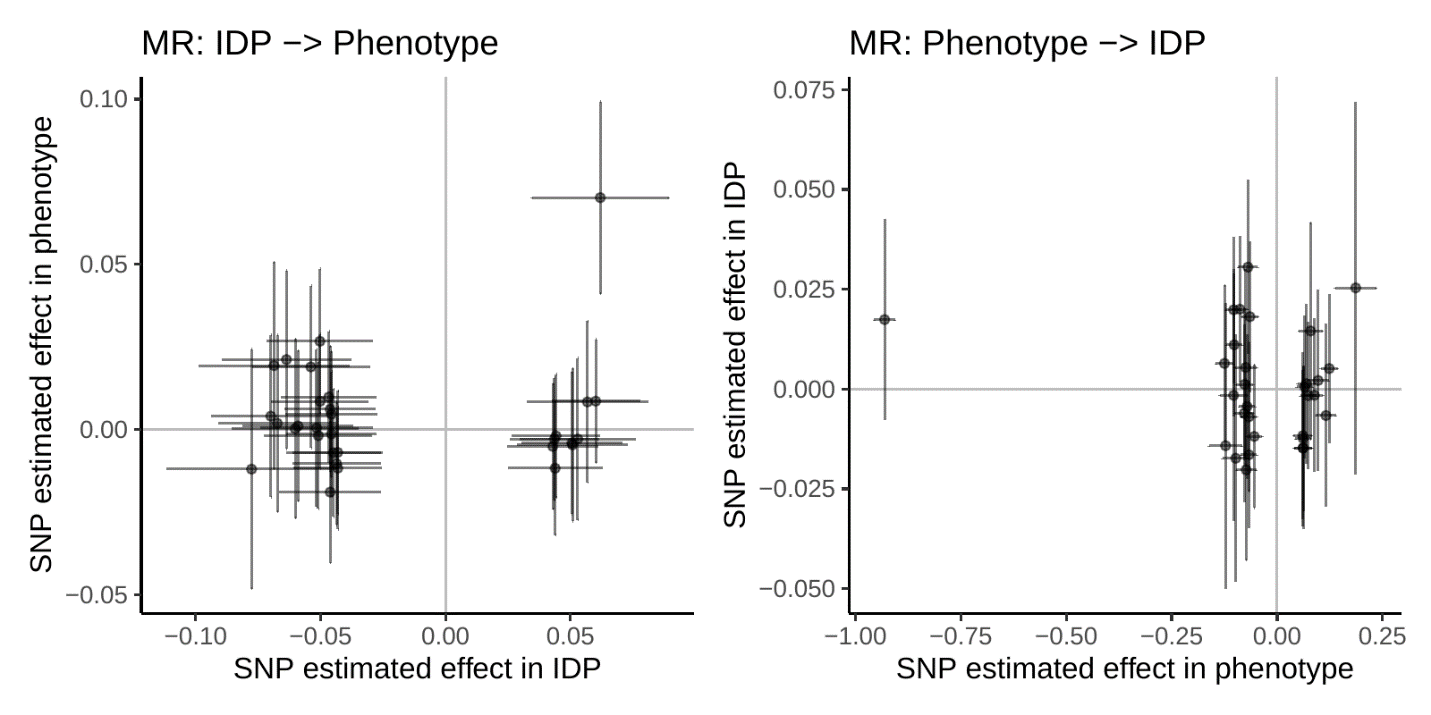




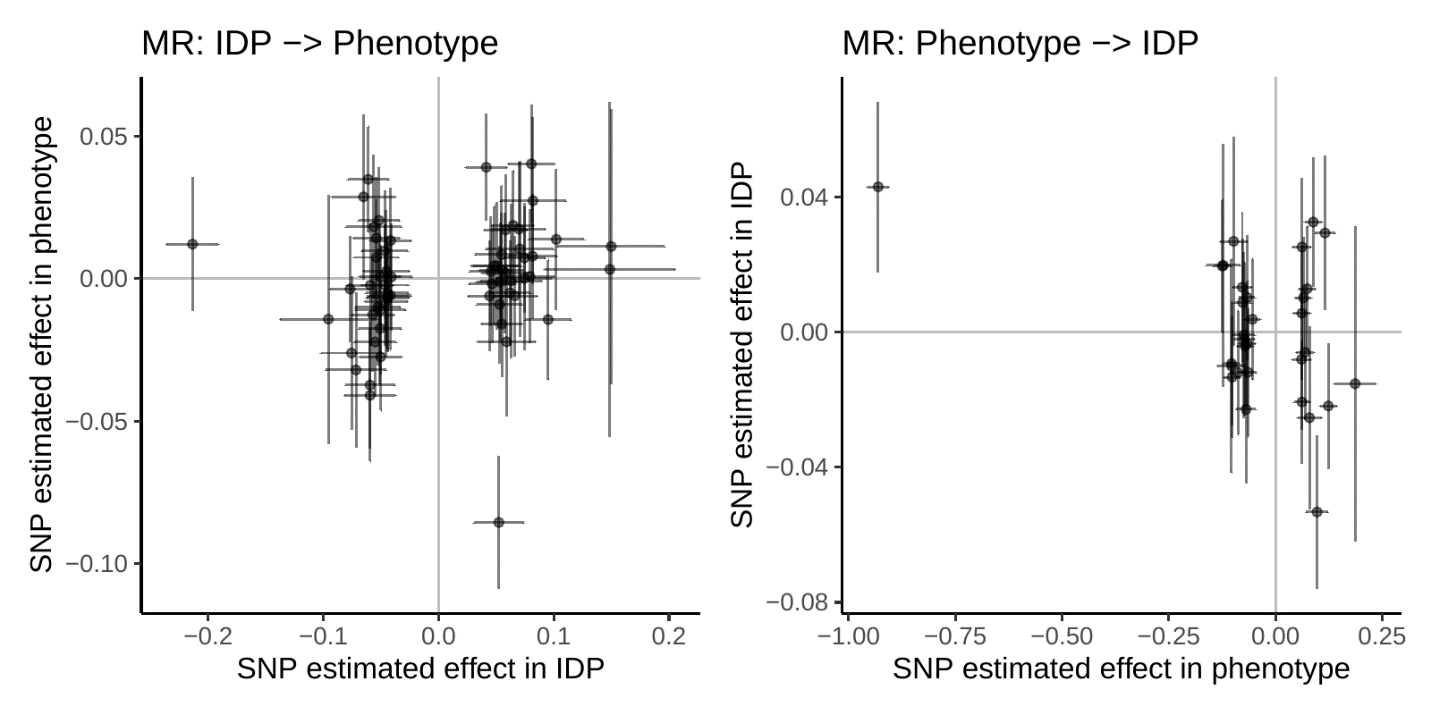
Supplementary Figure 3: Z-scores of genes from single tissue S-PrediXcan plotted against mean z-score from all tissue S-MultiXcan using Wightman summary statistics. (B) QQ plot of Wightman all tissue S-

MultiXcan p-values.

**A**

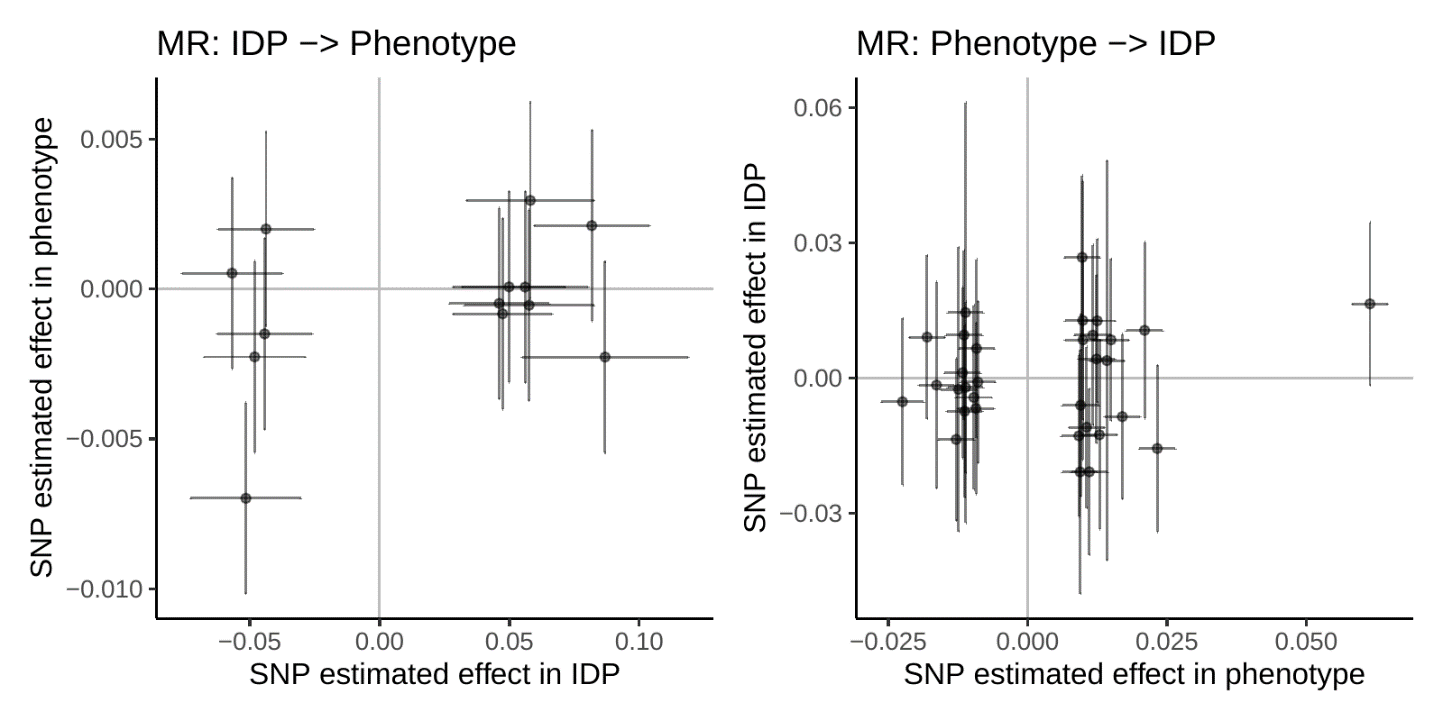


**B**

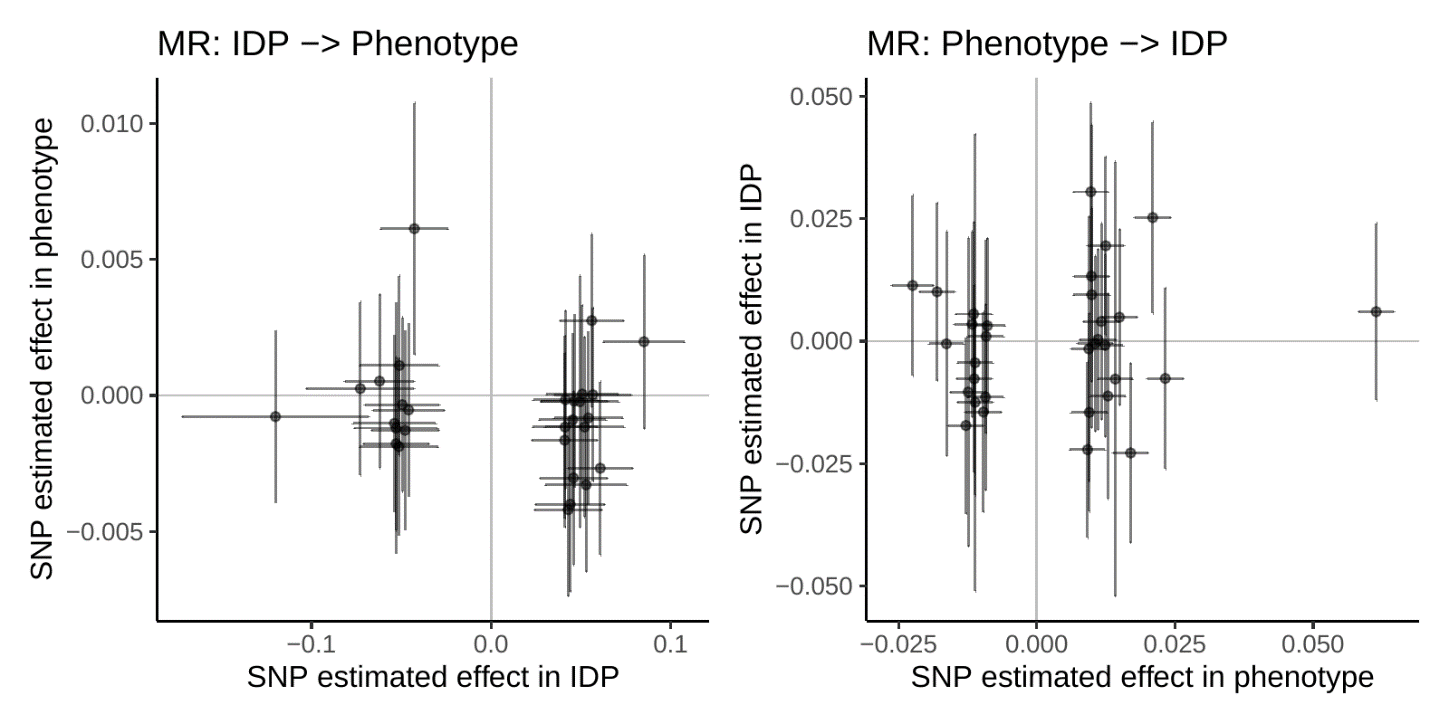


Supplementary Figure 4: MR plots from the Schwartzentruber S-BrainXcan run for (A) IDP-25402 mean OD in the right inferior cerebellar peduncle, IDP 🡪 trait: *p* = 0.5973, trait 🡪 IDP: *p* = 0.1817; (B) PC-ICVF-TBSS-1, IDP 🡪 trait: *p* = 0.00288, trait 🡪 IDP: *p* = 0.9981.

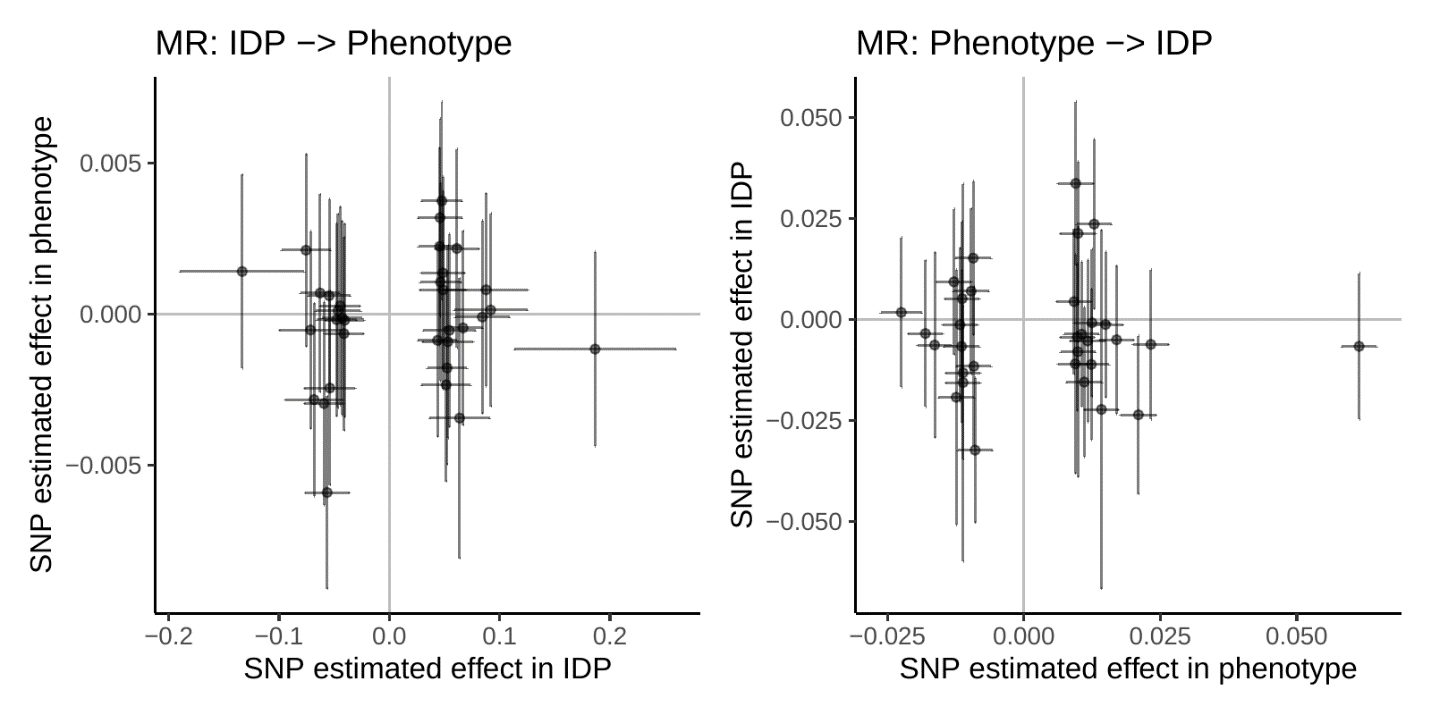
**A**



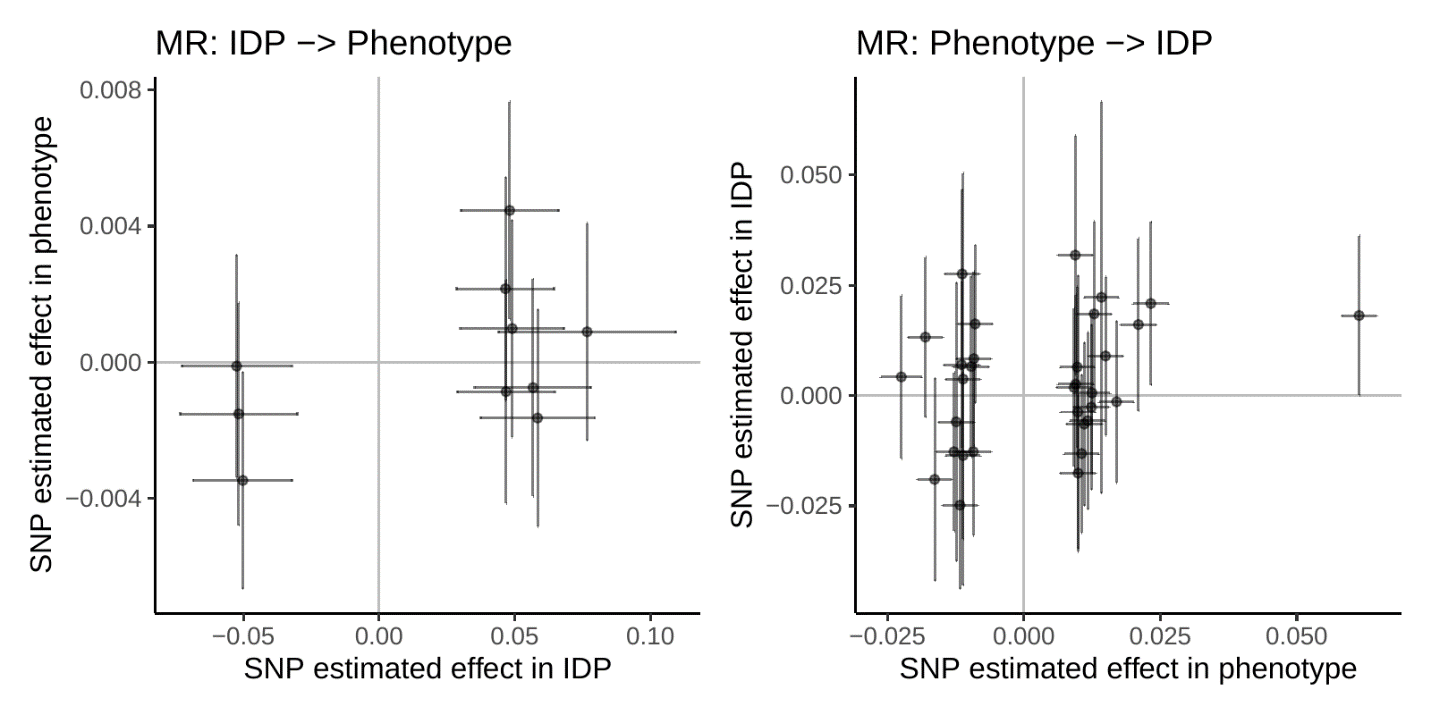
**B**

****

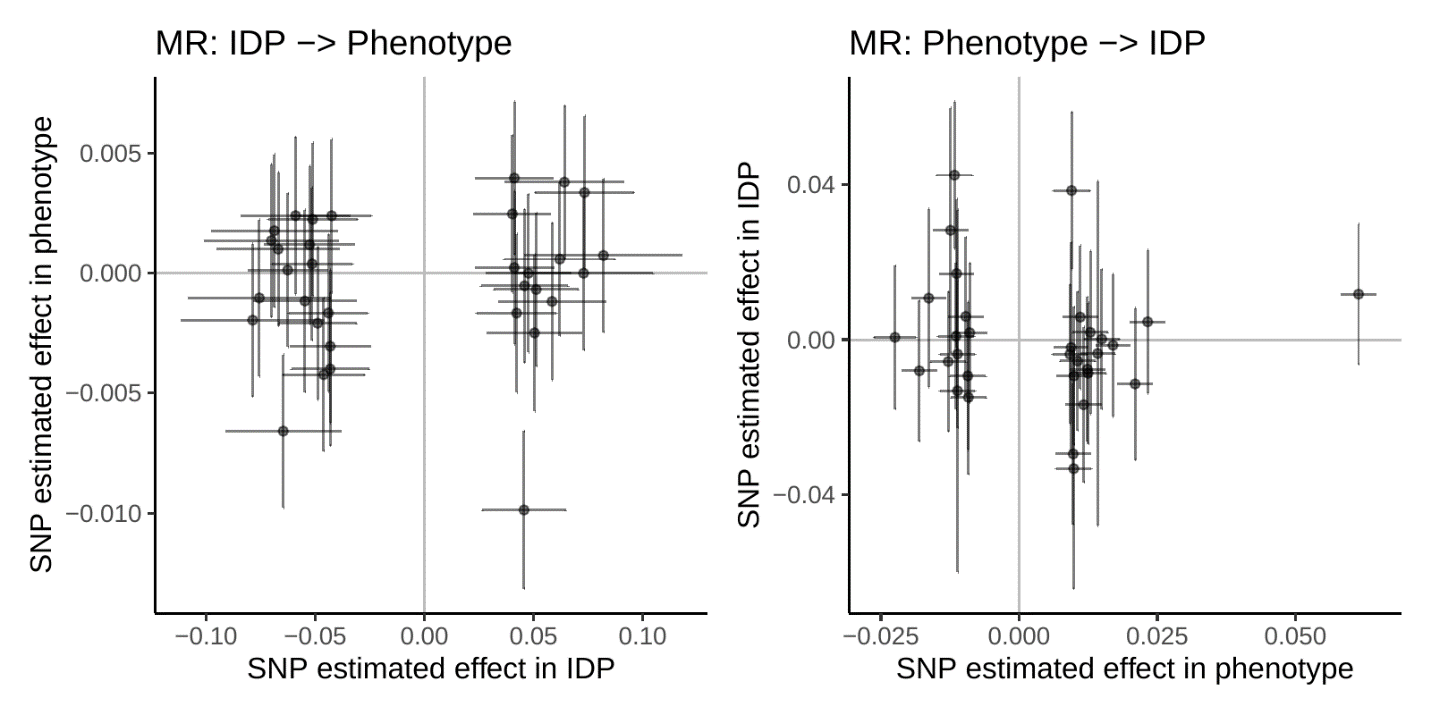
**C**

****

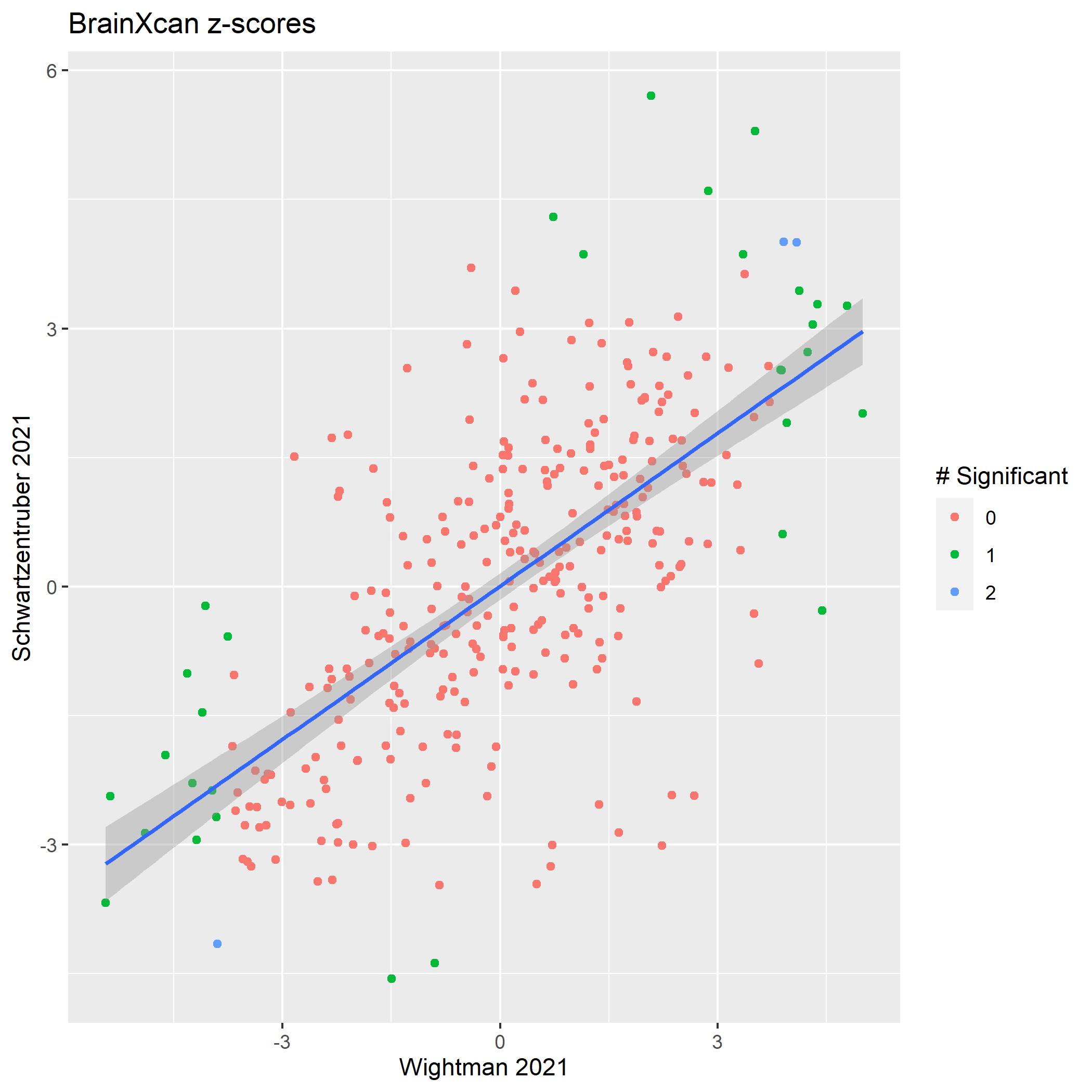
**D**

****

**E**

****

Supplementary Figure 5: MR plots from the Wightman S-BrainXcan run for (A) IDP-25386 mean ICVF in the right superior fronto-occipital fasciculus, IDP 🡪 trait: *p* = 0.7341, trait 🡪 IDP: *p* = 0.8535; (B) IDP-25387 mean ICVF in the left superior fronto-occipital fasciculus, IDP 🡪 trait: *p* = 0.4545, trait 🡪 IDP: *p* = 0.8287; (C) PC-OD-TBSS-1 IDP 🡪 trait: *p* = 0.6949, trait 🡪 IDP: *p* = 0.5697; (D) IDP-25858 volume of grey matter in left temporal occipital fusiform cortex, IDP 🡪 trait: *p* = 0.2961, trait 🡪 IDP: *p* = 0.05605; (E) IDP-25880 volume of grey matter in left caudate, IDP 🡪 trait: *p* = 0.6363, trait 🡪 IDP: *p* = 0.2507.

****

Supplementary Figure 6: Plot of S-BrainXcan IDP z-scores between the Schwartzentruber and Wightman runs. The trendline has a slope of 0.593 with error 0.0778 and intercept of 0.00218. The color of the points indicates whether an IDP was significant in zero, one, or both S-BrainXcan runs for red, green, and blue, respectively.



Supplementary Table 1: Number of significant genes per tissue from the brain-region-only Schwartzentruber run and total number of genes used in each tissue model.



Supplementary Table 2: Number of significant genes per tissue from the all-tissue Schwartzentruber run and total number of genes used in each tissue model.



Supplementary Table 3: Number of significant genes per tissue from the all-tissue Wightman run and total number of genes used in each tissue model.

Additional supplementary information (Supplementary tables 4-7) and all data used in this study: <https://uchicago.app.box.com/s/4ka9mtpdxlejgl2rhjrvcqv7eh2o75bt>

Supplementary Tables 4-6: Significant genes from S-MultiXcan of Schwartzentruber brain-only, Schwartzentruber all-tissue, and Wightman all-tissue, respectively.

Supplementary Table 7: Literature support for each gene in the union gene set from all three S-MultiXcan runs. Genes are split into alternating white and grey blocks based on independent LD blocks.

**References**

1. World Health Organization. (2021) Dementia Fact Sheet. *WHO*. <https://www.who.int/en/news-room/fact-sheets/detail/dementia> (accessed May 4, 2021).

2. Knopman DS, et al. (2021) Alzheimer disease. *Nat Rev Dis Primers* 7: 33.

3. Querfurth HW and Laferla FM. (2010) Alzheimer’s Disease. *N Engl J Med* 364(6): 588.

4. Herrup K. (2015) The case for rejecting the amyloid cascade hypothesis. *Nat Neurosci* 18: 794-799.

5. Sims R, Hill M, Williams J. (2020) The multiplex model of the genetics of Alzheimer’s disease. *Nat Neurosci* 23: 311322.

6. Visscher PM, et al. (2017) 10 years of GWAS discovery: Biology, Function, and Translation. *Am J Human Genet* 101(1): 5-22.

7. Seto M, Weiner RL, Dumitrescu L, and Hohman TJ. (2021). Protective genes and pathways in Alzheimer’s disease: moving towards precision interventions. *Mol Neurodegeneration* 16(29).

8. Ramanan VK and Saykin AJ. (2013) Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer’s disease, Parkinson’s disease, and related disorders. *Am J Neurodegener Dis* 2(3): 145175.

9. Tábuas-Pereira M, et al. (2020) Alzheimer’s Disease Genetics: Review of Novel Loci Associated with Disease. *Curr Genet Med Rep* 8: 1-16.

10. de Rojas I, et al. (2021) Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat Commun* 12(1): 3417.

11. Schwartzentruber J, et al. (2021) Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer’s disease risk genes. *Nat Genet* 53: 392402.

12. Wightman DA, et al. (2021) A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease. *Nat Genet* 53: 12761282.

13. Andrews SJ, Fulton-Howard B, and Goate A. (2020) Interpretation of risk loci from genome-wide association studies of Alzheimer’s disease. *Lancet Neurol* 19(4): 326-335.

14. Gamazon ER, et al. (2015) A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 47(9): 10911098.

15. GTEx Consortium. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45: 580-585.

16. Barbeira AN, et al. (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics.

17. Gerring ZF, Lupton MK, Edey D, Gamazon ER, and Derks EM. (2020) An analysis of genetically regulated gene expression across multiple tissues implicates novel gene candidates in Alzheimer’s disease. *Alzheimer’s Research & Therapy* 12: 43.

18. Chen TH and Boughal H. (2021) A penalized structural equation modeling method accounting for secondary phenotypes for variable selection on genetically regulated expression from PrediXcan for Alzheimer’s disease. *Biometrics* 77: 362371.

19. Chen HH, et al. (2020) Tissue-specific genetically regulated expression in late-onset Alzheimer’s disease implicates risk genes within known and 30 novel loci. *Alzheimer’s & Dementia* 16(3): e039475.

20. Poldrack RA, et al. (2017) Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat Rev Neurosci* 18: 115-126.

21. Marek S, et al. (2020) Towards reproducible brain-wide association studies. *bioRxiv*.

22. Leandrou S, et al. (2018) Quantitative MRI brain studies in mild cognitive impairment and Alzheimer’s disease: A methodological review. *IEEE Rev Biomed Eng* 11: 97-111.

23. Chandra A, Dervenoulas G, Politis M, and Alzheimer’s Disease Neuroimaging Initiative. (2019) Magnetic resonance imaging in Alzheimer’s disease and mild cognitive impairment. *J Neurol* 266(6): 1293-1302.

24. Harrison JR, et al. (2020) Imaging Alzheimer’s genetic risk using diffusion MRI: A systematic review. *Neuroimage Clin* 27: 102359.

25. Elliott L, et al. (2018) Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562: 210216.

26. Liang Y, et al. (2021) BrainXcan identifies brain features associated with behavioral and psychiatric traits using large scale genetic and imaging data. *bioRxiv*.

27. Alfaro-Almagro F, et al. (2018) Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* 166: 400424.

28. Barbeira AN, et al. (2019) Integrating predicted transcriptome from multiple tissues improves association detection. *PLOS Genet* 15(1): e1007889.

29. GTEx Consortium. (2017) Genetic effects on gene expression across human tissues. *Nature* 550(7675): 204-213.

30. Gamazon ER, et al. (2018) Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* 50: 956-967.

31. Wainberg M, et al. (2019) Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* 51: 592-599.

32. Berisa T and Pickrell JK. (2016) Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32(2): 283-285.

33. Li D, et al. (2021) Insights into lncRNAs in Alzheimer’s disease mechanisms. *RNA Biol* 18(7): 1037-1047.

34. Salminen A and Kaarniranta K. (2009) Siglec receptors and hiding plaques in Alzheimer’s disease. *J Mol Med* 87: 697.

35. Naj AC, et al. (2021) Genome-Wide Meta-Analysis of Late-Onset Alzheimer’s Disease Using Rare Variant Imputation in 65,602 Subjects Identifies Novel Rare Variant Locus *NCK2*: The International Genomics of Alzheimer’s Project (IGAP). *bioRxiv*.

36. DePew AT and Mosca TJ. (2021) Conservation and Innovation: Versatile Roles for LRP4 in Nervous System Development. *J Dev Biol* 9(1): 9.

37. Zhang H, et al. (2021) A Role of Low-Density Lipoprotein Receptor-Related Protein 4 (LRP4) in Astrocytic Aβ Clearance. *J Neurosci* 40(28): 5347-5361.

38. Kumar R, et al. (2011) Genome-wide mapping of ZNF652 promoter binding sites in breast cancer cells. *J Cell Biochem* 112(10): 2742-2747.

39. Kong W, Mou X, and Yang B. (2009) Study DNA Microarray Gene Expression Data of Alzheimer’s Disease by Independent Component Analysis. *2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*: 44-47.

40. Kosoy R, et al. (2021) Genetics of the human microglia regulome refines Alzheimer’s disease risk loci. *bioRxiv*.

41. Moy I, et al. (2015) Estrogen-dependent sushi domain containing 3 regulates cytoskeletal organization and migration in breast cancer cells. *Oncogene* 34: 323-333.

42. Bonham LW, Sirkis DW, and Yokoyama JS. (2019) The Transcriptional Landscape of Microglial Genes in Aging and Neurodegenerative Disease. *Front Immunol* 10: 1170.

43. Xu Z, Wu C, and Pan W. (2017) Imaging-wide association study: Integrating imaging endophenotypes in GWAS. *NeuroImage* 159: 159-169.

44. Ciavardelli D, et al. (2010) Alterations of brain and cerebellar proteomes linked to Aβ and tau pathology in a female triple-transgenic murine model of Alzheimer’s disease. *Cell Death & Disease* 1: e90.

45. Wang X, et al. (2021) Isoform-specific dysregulation of AMP-activated protein kinase signaling in a non-human primate model of Alzheimer’s disease. *Neurobiol Disease* 158: 105463.

46. Lehéricy S, Hirsch EC, Hersh LB, Agid Y. (1991) Cholinergic neuronal loss in the globus pallidus of Alzheimer’s disease patients. *Neurosci Letters* 123(2): 152-155.

47. Kamiya K, Hori M, and Aoki S. (2020) Noddi in clinical research. *J Neurosci Methods* 346: 108908.

48. Villain N, et al. (2010) Sequential relationships between grey matter and white matter atrophy and brain metabolic abnormalities in early Alzheimer’s disease. *Brain* 133(11): 3301-3314.

49. Duga S, et al. (2001) Characterization of the genomic structure of the human neuronal nicotinic acetylcholine receptor *CHRNA5/A3/B4* gene cluster and identification of novel intragenic polymorphisms. *J Hum Genet* 46: 640-648.

50. Chesi A, et al. (2006) O3-03-04: A HIGH RESOLUTION CAPTURE-C PROMOTER INTERACTOME IMPLICATES CAUSAL GENES AT ALZHEIMERS DISEASE GWAS LOCI. *Alzheimer’s and Dementia* 14(7S\_Part\_19): P1016.

51. Moore PC, et al. (2019) Elastase 3B mutation links to familial pancreatitis with diabetes and pancreatic adenocarcinoma. *J Clin Investig* 129(11): 4676-4681.

52. Mancuso N, et al. (2019) Probabilistic fine-mapping of transcriptome-wide association studies. *Nat Genet* 51: 675-682.