Alzheimer's Research & Therapy

An Analysis of Genetically Regulated Gene Expression across Multiple Tissues Implicates Novel Gene Candidates in Alzheimer's Disease

--Manuscript Draft--

Manuscript Number:	AZRT-D-19-00239R1
Full Title:	An Analysis of Genetically Regulated Gene Expression across Multiple Tissues Implicates Novel Gene Candidates in Alzheimer's Disease
Article Type:	Research
Funding Information:	
Abstract:	Introduction: Genome-wide association studies (GWAS) have successfully identified multiple independent genetic loci that harbour variants associated with Alzheimer's disease, but the exact causal genes and biological pathways are largely unknown. Methods: To prioritise likely causal genes associated with Alzheimer's disease, we used S-PrediXcan to integrate expression quantitative trait loci (eQTL) from the Genotype-Tissue Expression (GTEx) study and CommonMind Consortium (CMC) with Alzheimer's disease GWAS summary statistics. We meta-analysed the GTEx results using S-MultiXcan, prioritised disease-implicated loci using a computational fine-mapping approach, and performed a biological pathway analysis on the gene-based results. Results: We identified 126 tissue-specific gene-based associations across 48 GTEx tissues, targeting 50 unique genes. Meta-analysis of the tissue-specific associations identified 73 genes whose expression was associated with Alzheimer's disease. Additional analyses in dorsolateral prefrontal cortex from the CMC identified 12 significant associations, 8 of which also had a significant association in GTEx tissues. Fine-mapping of causal gene sets prioritised gene candidates in 10 Alzheimer's disease loci with strong evidence for causality. Biological pathway analyses of the meta-analysed GTEx data and CMC data identified a significant enrichment of Alzheimer's disease association signals in plasma lipoprotein clearance, in addition to multiple immune-related pathways. Conclusions: The integration of GWAS and gene expression data across multiple tissues improves power to identify and prioritise candidate genes and biological pathways for Alzheimer's disease, and may identify accessible surrogate tissue for follow-up functional genetic studies.
Corresponding Author:	Zachary Gerring, Ph.D. QIMR Berghofer Medical Research Institute Brisbane, QLD AUSTRALIA
Corresponding Author E-Mail:	zachary.gerring@qimrberghofer.edu.au
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	QIMR Berghofer Medical Research Institute
Corresponding Author's Secondary Institution:	
First Author:	Zachary Gerring, Ph.D.
First Author Secondary Information:	
Order of Authors:	Zachary Gerring, Ph.D.
	Michelle K Lupton, PhD
	Daniel Edey, BSc
	Eric R Gamazon, PhD
	Eske M Derks, PhD
Order of Authors Secondary Information:	

Response to Reviewers:	Response to reviewers' comments Reviewer #1: 1.From the work presented here, there is not compelling evidence that the results can address the concerns previously expressed on performing TWAS in tissues that differ from the known tissue type of interest. In this proposed work, the authors state that they find 126 significant associations for 50 distinct genes across 48 tissues after multiple testing correction, 22 of which (for 12 distinct genes) remain after excluding the APOE locus. The authors argue that the best tissue overlap is found with the skin, which not coincidentally has the highest number of samples. Wainberg et al (Nature Genetics doi: 10.1038/s41588-019-0385-z) specifically warn against performing TWAS in tissues without clear relevance to the trait, showing how known causal genes drop out while new genes appear, without any evidence for a causal association. Wainberg et al go on to demonstrate how using reference gene expression panels from tissues that are less related to the trait introduces bias in TWAS. We have included a scatter plot of GTEx tissue sample against the number of associations (Supplementary Figure 1). There is certainly a positive correlation between tissue sample size and the number of significant associations with Alzheimer's disease, however the relatively large number of associations found in skin is an outlier. Furthermore, genes with a significant association in both skin and (biologically relevant) brain tissues had concordant effect directions. We accept these observations do not imply causality but rather reflect potential shared regulatory mechanisms that may provide useful surrogate information for less accessible brain tissues. We have further addressed the concerns raised by the reviewer and Wainberg et al. in response 2 and response 3 below, where we performed fine mapping of causal gene sets (FOCUS) and transcriptomic imputation analyses with a large brain expression dataset. 2.There is also considerable concern about identifying multiple causal genes at the same locus. For these results to be reasonably considered, I recommend that the authors perform additional fine mapping or conditional analysis that will help to prioritize genes for each identified locus. An example of such an approach is here: FOCUS fine-mapping methods (http://github.com/bogdanlab/focus). We applied FOCUS to fine-map each Alzheimer's disease risk locus. We used as input the IGAP GWAS summary statistics and expression weights from 48 GTEx (version 7) tissues, METSIM (adipose tissue), NTR (whole blood), YFS (whole blood), and CMC (dorsolateral prefrontal cortex). We excluded chromosome 19 as the APOE region cannot be disentangled with current computational fine-mapping approaches. FOCUS prioritised candidate causal genes in 10 loci, including several candidates not yet implicated in Alzheimer's pathophysiology (methods: lines x-y; results: lines 247-258). 3.To address if these results provide reasonable targets that are relevant in the brain, these results should be validated in either an independent dataset for the relevant tissue, or using a targeted approach with just the brain samples from GTEx. One possibility would be to validate their findings using the newly available CommonMind Consortium which has the requisite expression and WGS sequencing data available. We performed S-PrediXcan using expression weights generated using gene expression and SNP genotype data from the CommonMind Consortium (CMC) (methods: lines 141-154, results lines 231-244). There was substantial overlap with the meta-analysed GTEx results, as well as some novel gene candidates. Reviewer #2 Gerring et al., present an analysis of publicly available genome-wide genotyping using publicly available models of tissue expression data. They then used the results of the imputed transcriptome-wide association study to perform gene set enrichment analysis. The authors acknowledge that sample size is likely guiding much of their findings in peripheral versus brain tissue, but it is interesting that variants in and around APOE are associated with different expression in skin. However, there is not a clear

	picture of how to use these new findings since the organ of interest in Alzheimer's disease is generally thought to be the brain so expression differences in the skin (or other organ) seem likely due to pleiotropy. Still there is nothing that seems technically wrong with the analysis and it seems like it ought to be published. The authors are upfront that other people have published similar analyses (with one using the same GTEx data); however, there are some differences in approach. Minor points: 1.Line 98: there were 88-173 samples from 15 brain regions per table s1 Thank you, we have amended the text to reflect the number of regions in Table S1. 2.Figure 1 initially confused me since the location of zero is not in the center of the z-score scale is there a bias to positive z-scores overall? Thank you. We have centred the z-score scale in Figure 1. There was a slight bias towards positive z-scores; we have noted this bias in the results (lines 182-183). 3.Does the lipid pathway remain significant if APOE and the locus around it (APOC1, etc.) are removed? We ran the pathway analyses with and without the APOE locus, using both the GTEx and CMC data. In GTEx, the lipid pathway is no longer significant (P=0.5216) after the APOE region was removed. In the CMC analyses, plasma lipoprotein clearance pathway remained significant (P=5.88 × 10-56) with the removal of APOE, along with immune system pathways (Supplementary Table 9 and Supplementary Table 10). 4.Lines 263-267: the link between the accumulation of those neuropathologic changes and lipid clearance is not as settled in the Alzheimer's disease field as the authors seem to suggest - I would suggest presenting this as an observed association. We have changed the text to emphasise the observed (rather than casual) association between plasma lipoprotein clearance and genetically-regulated gene expression in Alzheimer's disease. We also emphasise the potential biological consequences of this association have yet to be causally linked with Alzheimer's disease. 5.The last paragraph seems out of step with the results. For instance, the major pathway identified was lipid transport, which was also seen in the UTMOST paper; therefore, it is unclear why the findings in microglia are relevant here. Also, there are really few large-scale attempts to identify Alzheimer's disease eQTLs given the rarity of brain material for sequencing. We have updated the final paragraph to more broadly discuss the potential utility of studying immune-specific effects of susceptibility variants in Alzheimer's disease, given we found strong enrichment of immune-related pathways in dorsolateral prefrontal cortex from the CMC. We suggest immune dysfunction may arise from microglial cells, the chief immune cells of the central nervous system, in line with a recent Alzheimer's disease GWAS (10.1038/s41588-018-0311-9). Thank you for your helpful comments on our manuscript. Sincerely, Zachary Gerring, PhD Postdoctoral Research Fellow, Translational Neurogenomics QIMR Berghofer Medical Research Institute
Additional Information:	
Question	**Response**
Is this study a clinical trial?<hr><i>A clinical trial is defined	No

by the World Health Organisation as 'any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes'.</i>	

An Analysis of Genetically Regulated Gene Expression across Multiple Tissues Implicates Novel Gene Candidates in Alzheimer's Disease

Zachary F Gerring[a] PhD, Michelle K Lupton[b] PhD, Daniel Edey BSc (Hons)[a], Eric R Gamazon[c] PhD, Eske M Derks[a] PhD

[a] Translational Neurogenomics Laboratory; QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston 4006, Brisbane, Queensland, Australia

b Genetic Epidemiology Laboratory; QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston 4006, Brisbane, Queensland, Australia

[c] Division of Genetic Medicine, Department of Medicine, Vanderbilt University, 1211 21st Ave S, Nashville, TN 37212, USA

Corresponding author
Zachary F Gerring, Ph.D
Translational Neurogenomics Laboratory
300 Herston Road
Brisbane City QLD 4006
zachary.gerring@qimrberghofer.edu.au

ABSTRACT

Introduction: Genome-wide association studies (GWAS) have successfully identified multiple independent genetic loci that harbour variants associated with Alzheimer's disease ~~(AD)~~, but the exact causal genes and biological pathways are largely unknown.

Methods: To prioritise likely causal genes associated with ~~AD~~Alzheimer's disease, we used S-PrediXcan to integrate~~d gene expression~~expression quantitative trait loci (eQTL) from the Genotype-Tissue Expression (GTEx) study and CommonMind Consortium (CMC) ~~and~~ with ~~AD~~ Alzheimer's disease GWAS summary statistics~~ across 48 tissues using S-PrediXcan~~. We meta-analysed the~~ single tissue~~ GTEx results using S-~~TissueXcan~~MultiXcan, prioritised disease-implicated loci using a computational fine-mapping approach, ~~–~~and performed a biological pathway analysis on the gene-based results.

Results: We identified 126 tissue-specific gene-based associations across 48 GTEx tissues, targeting 50 unique genes. Meta-analysis of the tissue-specific associations identified 73 genes whose expression was associated with Alzheimer's disease. Additional analyses in dorsolateral prefrontal cortex from the CMC identified 12 significant associations, 8 of which also had a significant association in GTEx tissues. ~~S-PrediXcan identified 126 significant associations targeting 50 genes after multiple testing correction, 22 of which (targeting 12 genes) remained significant after removal of the APOE region.~~ Fine-mapping of causal gene sets prioritised gene candidates in 10 Alzheimer's disease loci with strong evidence for causality. Biological pathway analyses of the meta-analysed GTEx data and CMC data identified a significant enrichment of Alzheimer's disease association signals in ~~Meta-analysis of S-PrediXcan results identified 73 genes that were significantly enriched in~~ plasma lipoprotein clearance, in addition to multiple immune-related pathways. ~~;~~

Conclusions: The integration of GWAS and gene expression data across multiple tissues improves power to identify and prioritise candidate genes and biological pathways for

Formatted: Space After: 0 pt

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font:

~~AD~~Alzheimer's disease, and may identify accessible surrogate tissue for follow-up functional genetic studies ~~of AD~~.

Keywords: Alzheimer's disease; Gene expression; Genome-wide Association Study; Genetics; Genetic Epidemiology; Computational Biology

1. BACKGROUND

An estimated 5.5 million Americans ~~are~~ were living with Alzheimer's disease ~~(AD)~~ in 2017, with a prevalence of 10% for people over the age of 65 years [1]. In the absence of a significant medical breakthrough the number of people living with ~~AD~~ Alzheimer's disease is estimated to reach 13.8 million in the US alone by 2050 [1]. ~~AD~~ Alzheimer's disease is ~~officially~~ the sixth leading cause of death in the US, but this is likely to be underestimation as complications of the disease, such as pneumonia, are often recorded as the primary cause of death. Alzheimer's disease~~AD~~ is characterised by neuronal death and key neuropathological changes, including the deposition of β-amyloid and hyperphosphorylated tau tangles. Genome wide association studies (GWAS) ~~for AD~~ have ~~been successful in~~ identif~~ying~~ied genetic risk factors for ~~AD~~ Alzheimer's disease and ~~providing~~ provided novel insights into disease aetiology. A GWAS meta-analysis of 74,046 individuals (25580 cases and 48466 controls) identified 19 genetic risk loci [2], which has since increased to some 24 loci with ~~larger~~ ~~the additional samples~~ ~~cases~~larger sample sizes [3]. Biological pathway analyses of these data implicate the immune system and lipid metabolism as well as tau binding and amyloid precursor protein metabolism [2], although a disease mechanism of action ~~,~~has yet to be established.

In GWAS, ~~-~~significant associations are reported for a~~n~~ ~~index~~ single nucleotide polymorphism (SNP) with the lowest *P* value, but the signal could be ~~led~~ explained by ~~any~~ one (or more) variant within the linkage disequilibrium block where that SNP resides. Furthermore, GWAS loci may contain multiple genes or regions that affect~~ing~~ the expression of ~~alternative~~ other genes. Additional ~~analysis~~ analyses ~~is~~ are required to elucidate the biological mechanisms that underlie statistical associations between genetic variants and disease risk. One method is to identify ~~the genomic regions~~loci where SNP variation is associated with differences in gene expression, called expression quantitative trait loci (eQTL~~s~~). Genome-wide gene expression

data has been successfully integrated with SNP genotype data to prioritise risk genes and reveal possible mechanisms underlying susceptibility to a range of psychiatric disorders [4–7]. This approach may be performed in cases and controls for whom both gene expression and SNP genotype data are available. However, these data sets are likely to have limited sample size and suffer from confounding from reverse causality as variation in gene expression may be influenced by disease status or drug treatment.

An alternative method is to integrate GWAS findings with independent gene expression data provided by large international consortia, such as the multi-tissue Genotype-Tissue Expression (GTEx) project [8] and the CommonMind Consortium (CMC). GTEx (version 7) contains SNP genotype data linked to gene expression across 53 tissues from 714 donors, including 13 brain ~~tissues~~ regions~~from 216 donors~~, and the CMC contains gene expression data from the dorsolateral prefrontal cortex of 986646 donors. ~~This~~ These data represent~~s~~ a valuable resource with which to quantify the association between genetically regulated expression in multiple tissues and the phenotype of interest. Association testing can be carried out using a gene-based approach implemented by ~~transcriptome-wide association study (TWAS)~~transcriptomic imputation approaches ~~-~~[5,9,10] which reduce the high level of multiple testing from single-variant tests, and increase power to identify trait associated loci from both a strong functional SNP signal, or from a combination of modest signals. The application of ~~TWAS~~ transcriptomic imputation ~~approaches~~ using GWAS summary statistics without the need for individual level data allows ~~this~~ these method~~s~~ to be applied to large scale GWAS meta-analyses results. Here, we apply a ~~TWAS~~ transcriptomic imputation approach called S-PrediXcan to Alzheimer's disease GWAS summary statistics in order to explore the genetic component of gene expression associated with the disorder.~~,~~ We then use these data in a fine-mapping approach to ~~identify~~ prioritise candidate causal genes with disease implicated loci, and ~~prioritise~~ identify

biologically informative surrogate tissues that might be used to ~~identify~~ characterise Alzheimer's disease pathways and processes.

2. MATERIALS AND METHODS

2.1. Alzheimer's disease GWAS summary statistics

Detailed methods, including a description of population cohorts, quality control of raw SNP genotype data, and association analyses for the Alzheimer's disease GWAS is described in detail elsewhere [2]. The Alzheimer's disease GWAS, performed by members of the International Genomics of Alzheimer's Project (IGAP), included an initial meta-analysis of 4 samples of European ancestry (17,008 cases and 37,154 controls) followed by an analysis of moderately associated SNPs ($P < 1 \times 10^{-3}$) in an independent sample of 8,572 cases and 11,312 controls of European ancestry. All cases received clinical confirmation of late-onset Alzheimer's disease. SNPs were imputed using the European population reference from the 1000 Genomes Project 2010 interim release based on the sequence data freeze from 4 August 2010 and phased haplotypes from December 2010) [11]. Logistic regression association tests were conducted for imputed marker dosages with age and sex as covariates, as well as principal components to control for possible population stratification. Summary statistics for 7,055,881 autosomal SNPs were made available by IGAP and were utilized in our study.

2.2. Identification of genes with differential expression levels between Alzheimer's disease cases and controls

We used S-PrediXcan to integrate eQTL information with ~~our~~ GWAS summary statistics to identify genes of which genetically predicted expression levels are associated with Alzheimer's disease status. S-PrediXcan estimates gene expression weights by training a linear prediction model in a reference sample with both gene expression and SNP genotype data. The weights are used to predict gene expression from GWAS summary statistics, while incorporating the

Formatted: Font: (Default) Times New Roman, 12 pt, Italic

variance and co-variance of SNPs from a~~n~~ linkage disequilibrium (LD) reference panel. We used expression weights for 48 tissues with S-PrediXcan expression weights from the GTEx Project (version 7) and dorsolateral prefrontal cortex from the CommonMind Consortium (CMC), and LD information from the 1000 Genomes Project Phase 3[12]. These data were processed with beta values and standard errors from the Alzheimer's disease GWAS to estimate the expression-GWAS association statistic. To increase power to identify genes whose expression is similarly differentially regulated across tissues, we meta-analysed the ~~(tissue-specific) statistics from~~GTEx S-PrediXcan results using the S-~~TissueXcan~~ MultiXcan algorithm[13]. We used Bonferroni correction to adjust for the number of tests performed within each tissue as well as across all tissues and genes (Table S1).

~~to adjust for both tissue-specific correction threshold for each interrogated tissue (Table S1) as well as a transcriptome-wide significance threshold of $P<2.68 \times 10^{-7}$, adjusting for all tissues and genes (i.e. N=186,230 gene-based tests in the GTEx).~~

2.3. Fine-mapping of causal gene sets (FOCUS)

S-PrediXcan and other transcriptomic approaches may yield false positive gene-trait associations due to correlation (LD) among SNPs used to generate the eQTL weights in the predication models[14]. We used FOCUS (fine-mapping of causal gene sets) to appropriately model the impact of gene-trait correlations on the S-PrediXcan expression weights and assign a causal probabilit~~ies~~y to each gene within Alzheimer's disease risk loci. We used a multi-tissue~~, multiple~~ eQTL reference panel database provided by the authors (https://github.com/bogdanlab/focus/) and LD information from the 1000 Genomes Project Phase 3[12] as reference genotypes. ~~We excluded c~~Chromosome 19 was removed due to the

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Normal, No bullets or numbering

Formatted: Font: Italic

Formatted: Normal, No bullets or numbering

Formatted: Font: Not Italic

Formatted: Default Paragraph Font, Font: (Default) Times New Roman, 12 pt

~~complexity of modelling association signals~~complex association signals within the APOE locus.

~~*Pathway analysis of gene-based analyses*~~

~~*2.3.*~~*2.4. Pathway analysis of gene-based analyses*

We performed a biological pathway analysis using generalised linear model regression, with the z-score from the GTEx S-~~TissueXcan~~ MultiXcan or CMC S-PrediXcan association data as the dependent variable and membership in Reactome pathways as a linear predictor. Pathways containing fewer than 10 *cis*-heritable genes (i.e. genes whose average expression across tissues is influenced by proximal [<1 Mb from the gene start or end] SNPs) were removed, resulting in 1318 biological pathways for pathway enrichment analysis. A Bonferroni corrected *P* value of $P = 3.79 \times 10^{-5}$ (adjusting for 1318 tested pathways) was used to correct for multiple testing.

3. RESULTS

3.1. A cross-tissue transcriptome-wide association study identifies peripheral tissues enriched with Alzheimer's disease association signals

Using S-PrediXcan, we identified 126 significant associations (Supplementary Table S2) targeting 50 unique genes (Supplementary Table S3) after multiple testing correction for all genes and tissues ($P<2.68 \times 10^{-7}$) ~~(Table 1; Supplementary Table S3)~~. Among significant associations, there was a slight bias towards positive z-scores (N=75 [60%]). The number of significant associations per tissue was largely a function of sample size, with skin (sun-exposed lower leg) (number of RNA-seq samples N=473) harbouring the ~~The tissue with the~~ largest number of associations ~~was skin (sun-exposed lower leg)~~ (n=9), followed by lung (n=8) (Table 2) (Supplementary Figure S1). For significant genes identified in multiple GTEx tissues, effect directions were largely consistent across tissues (Figure 1)~~, suggesting peripheral tissues may provide reliable surrogate information for brain-related processes~~. The most significant gene

association in GTEx data was for *APOE*; genetic variants associated with increased ~~liability to~~ Alzheimer's disease risk are predicted to downregulate expression levels of *APOE* in ~~3~~ three peripheral tissues, including sun-exposed skin (Z=19.50, $p=1.03 \times 10^{-84}$) and non-sun-exposed skin (--16.56, $p=1.27 \times 10^{-61}$) (Table 1; Supplementary Table S3) after multiple testing correction (Bonferroni correction for 186,230 tests [0.05/186,230] $P < 2.68 \times 10^{-7}$). Of note, although *APOE* is expressed more widely in brain compared to most other tissues (Supplementary Figure ~~S1~~S2), the eQTL associations with *APOE* are only found in non-brain tissues. While these associations are likely to be due, at least in part, to the increased sample size (and therefore statistical power) of ~~non-brain~~peripheral tissues, they highlight the importance of interrogating multiple (accessible) tissues in eQTL analyses of complex (brain-related) traits.

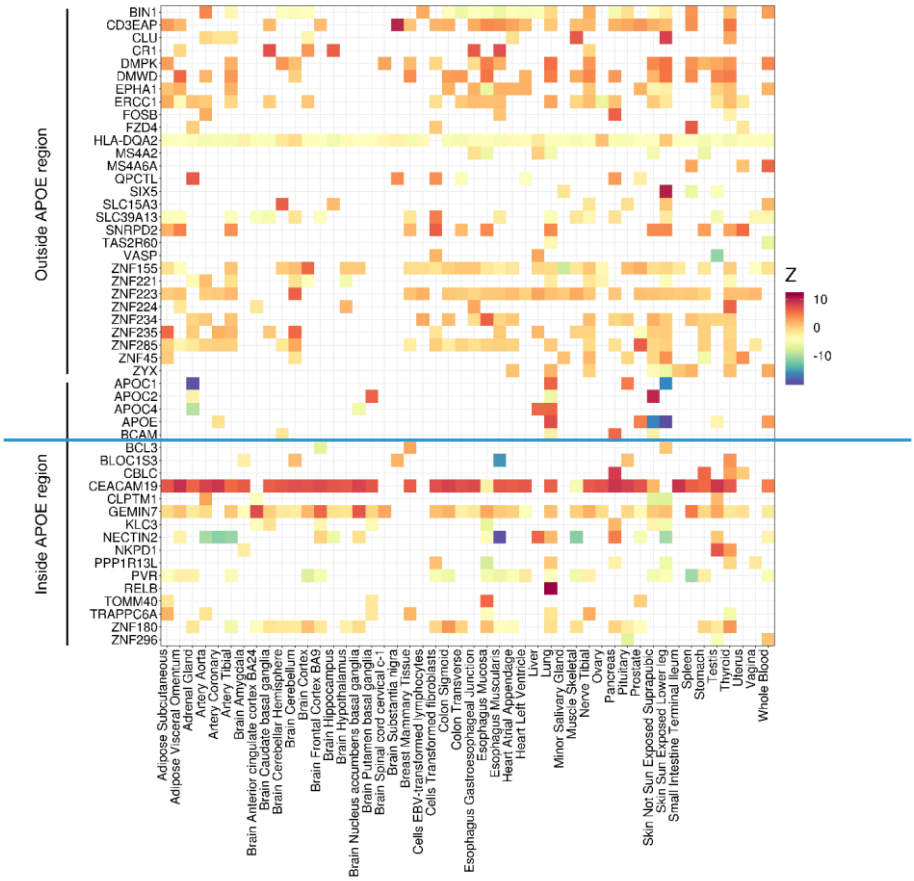
Table 1: Top 5 S-PrediXcan associations ~~inside and outside~~by APOE region

Gene ~~name~~	Chr	Most significant tissue	~~N~~ SNPs	~~N~~ ~~t~~Tissue~~ss~~	Z score	P value
Inside APOE						
APOE	19	Skin Sun Exposed ~~Lower~~ leg	10	3	-19.50	1.03×10^{-84}
NECTIN2	19	Oesophagus Muscularis	3	8	-19.28	8.32×10^{-83}
APOC1	19	Adrenal Gland	3	3	-19.13	1.48×10^{-81}
BLOC1S3	19	Oesophagus Muscularis	9	1	-15.63	4.29×10^{-55}
RELB	19	Lung	24	1	11.55	7.14×10^{-31}
Outside APOE						
VASP	19	Testis	53	1	-11.30	1.24×10^{-29}
SIX5	19	Skin Sun Exposed ~~Lower~~ leg	40	2	10.28	8.60×10^{-25}
CD3EAP	19	~~Brain~~ Substantia nigra	6	1	10.06	8.65×10^{-24}
ZNF155	19	Minor Salivary Gland	62	2	-8.45	3.02×10^{-17}
CLU	2	Skin Sun Exposed ~~Lower~~ leg	5	2	8.22	2.04×10^{-16}

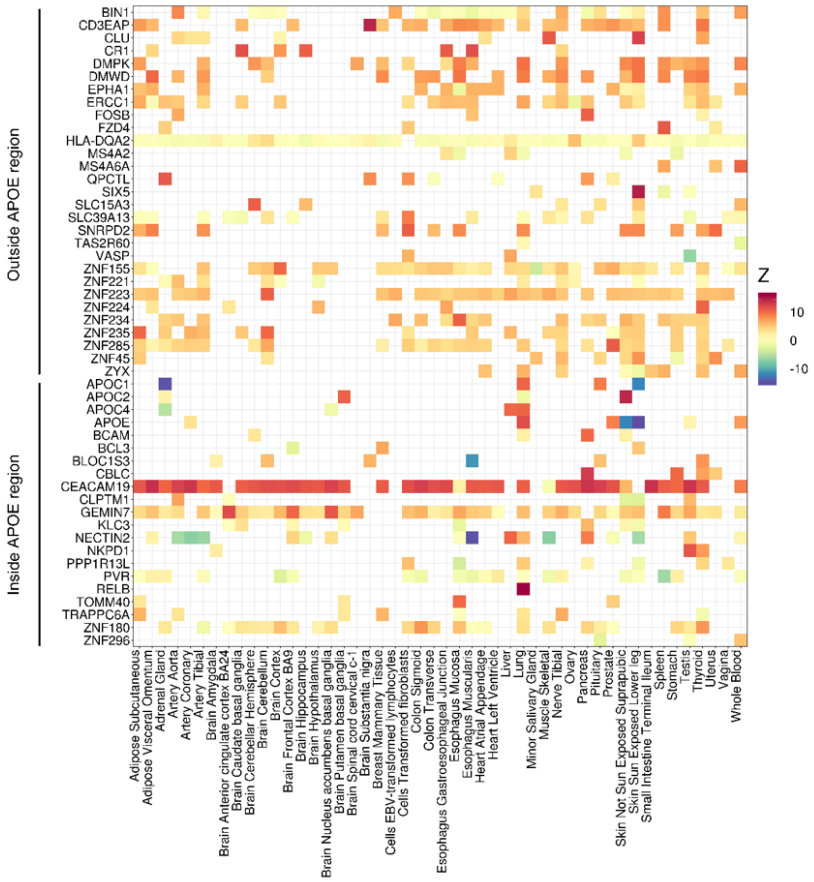
Notes: Chr=Chromosome; N SNPs is number of eQTLs included in the MetaXcan prediction model; N tissues=N tissues with P-value < 7.63×10^{-7}; Z score represents the strength of association between gene expression and disease risk. Positive values indicate that an increased level of gene expression is associated with increased disease risk while negative values indicate that a reduced level of gene expression increases disease risk.

Formatted: Font: 11 pt
Formatted Table
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Superscript
Formatted: Font: 11 pt
Formatted: Font: 11 pt

Figure 1: Heatmap of the Z score effect directions for significant genes identified in multiple tissues


Outside APOE region
Inside APOE region
BIN1
CD3EAP
CLU
CR1
DMPK
DMWD
EPHA1
ERCC1
FOSB
FZD4
HLA-DQA2
MS4A2
MS4A6A
QPCTL
SIX5
SLC15A3
SLC39A13
SNRPD2
TAS2R60
VASP
ZNF155
ZNF221
ZNF223
ZNF224
ZNF234
ZNF235
ZNF285
ZNF45
ZYX
APOC1
APOC2
APOC4
APOE
BCAM
BCL3
BLOC1S3
CBLC
CEACAM19
CLPTM1
GEMIN7
KLC3
NECTIN2
NKPD1
PPP1R13L
PVR
RELB
TOMM40
TRAPPC6A
ZNF180
ZNF296
Z
10
0
-10
Adipose Subcutaneous
Adipose Visceral Omentum
Adrenal Gland
Artery Aorta
Artery Coronary
Artery Tibial
Brain Amygdala
Brain Anterior cingulate cortex BA24
Brain Caudate basal ganglia
Brain Cerebellar Hemisphere
Brain Cerebellum
Brain Cortex
Brain Frontal Cortex BA9
Brain Hippocampus
Brain Hypothalamus
Brain Nucleus accumbens basal ganglia
Brain Putamen basal ganglia
Brain Spinal cord cervical c-1
Brain Substantia nigra
Breast Mammary Tissue
Cells EBV-transformed lymphocytes
Cells Transformed fibroblasts
Colon Sigmoid
Colon Transverse
Esophagus Gastroesophageal Junction
Esophagus Mucosa
Esophagus Muscularis
Heart Atrial Appendage
Heart Left Ventricle
Liver
Lung
Minor Salivary Gland
Muscle Skeletal
Nerve Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin Not Sun Exposed Suprapubic
Skin Sun Exposed Lower leg
Small Intestine Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood

221



We removed genes flanking the *APOE* region (+/- 500kb) due to its strong association with ~~AD~~Alzheimer's disease and identified ~~22~~ 29 significant associations ~~targeting 12 unique genes (Supplementary Table S3), 7 of which were not identified as candidate causal (i.e. nearest) gene in the Lambert *et al* GWAS~~. The most significant gene outside the *APOE* region was the vasodilator-stimulated phosphoprotein *VASP* ($Z = -11.30$, $P = 1.24 \times 10^{-24}$) in Testis (Table 1). The most significant association outside chromosome 19 was observed for the clusterin *CLU* in skin (sun exposed lower leg) (Table 1; Supplementary Table S3). Taken together with findings for *APOE*, these data suggest skin (together with other peripheral tissues) may be used

Formatted: Font: Italic

as an accessible surrogate tissue for peripheral biomarker discovery and molecular studies of causal disease processes.

Table ~~2~~2: Number of significant S-PrediXcan associations per tissue

Tissue	Tissue sample size (N)	Gene Associations (N)	Genes
Skin Sun Exposed Lower leg	414	~~9~~10	APOE, APOC1, NECTIN2, SIX5, CLU, CLPTM1, ZNF229, ZYX, PPP1R13L, KLC3
Lung	383	8	RELB, APOE, CEACAM19, APOC2, APOC1, APOC4, MS4A2, DMPK
Oesophagus Mucosa	358	~~7~~8	PPP1R13L, KLC3, EPHA1, ZNF234, MS4A2, RP11-385F7.1, TOMM40, PVR
Oesophagus Muscularis	335	6	NECTIN2, BLOC1S3, CR1, CEACAM19, BIN1, PVR
Skin Not Sun Exposed Suprapubic	335	~~5~~6	APOE, APOC2, ZNF229, CLPTM1, MS4A2, PVR
Adrenal Gland	175	4	APOC1, APOC4, QPCTL, CEACAM19
Brain Hippocampus	111	4	CEACAM19, CR1, NECTIN2, HLA-DQA2
Pancreas	220	4	CEACAM19, CBLC, FOSB, BCAM
Spleen	146	4	PVR, FZD4, CEACAM19, SIX5
Stomach	237	4	MS4A2, ZNF45, CBLC, CEACAM19

To improve power relative to the single-tissue analyses, we combined results from different single-tissue models into a single aggregate statistic using S-~~TissueXcan~~multiXcan. We identified 73 gene-level S-~~TissueXcan~~ MultiXcan associations after correction for multiple testing (Table 3, Supplementary Table S4), of which 36 were located outside the *APOE* region. The S-~~TissueXcan~~ MultiXcan analysis identified 27 additional significant genes not found in the single tissue analyses, 19 of which encoded genes outside the *APOE* region (Supplementary Table S4). The most significant S-~~TissueXcan~~ MultiXcan association was for *PVRL2* (also known as *NECTIN2*), located within the *APOE* region (oesophagus muscularis; Z_{mean} = -4.94, $P = 2.64 \times 10^{-131}$), followed by *APOE* (skin sun exposed lower leg; Z_{mean} = -3.58, $P = 4.25 \times 10^{-101}$). The most significant protein coding gene outside the *APOE* region was for Protein Tyrosine Phosphatase, Receptor Type H *PTPRH* (brain caudate basal ganglia); Z_{mean} = 0.35, *P*

Formatted: Font: 11 pt
Formatted Table
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

= 2.19 × 10^{-12}). A total of 7 genes were significant in the single-tissue analyses but not the S-~~TissueXcan~~ MultiXcan meta-analysis, due in part to heterogeneity in the effect directions of imputed gene expression across tissues ~~(Supplementary Table S5)~~.

Table 3: Top 5 S-~~TissueXcan~~ MultiXcan associations ~~inside and outside~~by APOE region

Gene ~~name~~	~~Most significant t~~Top tissue	N ~~N~~ ~~t~~Tissues	P ~~value~~	Z score Min	Max	Mean	SD
Inside APOE							
PVRL2	~~Oesophagus~~ Esophagus Muscularis	17	2.64×~~E~~10^{-131}	-19.28	5.78	-4.94	6.75
APOE	Skin Sun Exposed ~~Lower~~ leg	7	4.25×10~~E-~~$^{-101}$	-19.50	7.51	-3.58	10.50
APOC1	Adrenal Gland	4	4.05×10~~E~~$^{-92}$	-19.13	5.98	-6.24	13.43
BLOC1S3	~~Oesophagus~~ Esophagus Muscularis	6	9.00×10~~E~~$^{-75}$	-15.63	3.48	-1.78	7.07
APOC4	Adrenal Gland	4	1.40×10~~E~~$^{-39}$	-9.53	5.91	-0.92	7.96
Outside APOE							
SIX5	Skin Sun Exposed ~~Lower~~ leg	4	1.24×10~~E~~$^{-37}$	-6.18	10.3	-0.41	7.40
VASP	Testis	3	5.58×10~~E~~$^{-28}$	-11.3	2.61	-2.29	7.82
BIN1	~~Oesophagus~~ Esophagus Muscularis	23	3.58×10~~E~~$^{-16}$	-6.32	4.09	-2.21	3.26
CLU	Skin Sun Exposed ~~Lower~~ leg	8	7.51×10~~E~~$^{-14}$	-3.07	8.22	1.28	4.12
CR1	~~Oesophagus~~ Esophagus Muscularis	7	1.69×10~~E~~$^{-11}$	-0.38	7.33	4.299	3.45

Notes: N Tissues, number of tissues with significant gene-based association; Z score: Minimum, maximum, mean and standard deviation of the Alzheimer's disease association coefficient from S-~~TissueXcan~~MultiXcan.

3.2. A comparison of multi-tissue GTEx results with brain-specific eQTL database from the CommonMind consortium

We performed an S-PrediXcan analysis using expression weights for a single brain region (dorsolateral prefrontal cortex) collected by the CMC, and identified 12 significant (*P*<5.08 × 10^{-6}) gene-based associations (Supplementary Table S5). We compared these data with the meta-analysed results from 48 tissues in GTEx (Table 4). Of 12 significant gene-based associations in ~~GTEx~~CMC, 8 also showed a significant association in GTEx tissues (P<1.93 × 10^{-6}). The Z scores between CMC and GTEx were concordant where the mean absolute GTEx

Z score was ≥ 1, highlighting the consistency of the datasets. The top CMC association was TOMM40 (P=1.37 × 10^{-101}), located within the APOE gene cluster on Chromosome 19q13. The GTEx tissue with the strongest association for TOMM40 was esophagus mucosa (Z score=5.57, P=2.61 × 10^{-8}) (Supplementary Table S2), a tissue that contains over twice the number of samples as the largest brain tissue (N=407 versus N=173 in cerebellum). One TOMM40 association was observed in GTEx brain tissue (Putamen basal ganglia); the association was insignificant (P=6.73 × 10^{-2}) but the Z score direction of effect was consistent with CMC data (GTEx: Z=-1.83; CMC: -21.40).

Table 4: Significant associations in CMC dorsolateral prefrontal cortex and corresponding GTEx association statistics

CMC (DLPFC)			GTEx (48 tissues)			
Gene	**Z**	**P**	**Tissue**	**Z**	**Z SD**	**P**
TOMM40	-21.40	1.37×10^{-101}	Esophagus Mucosa	0.57	3.44	1.10×10^{-7}
ZNF222	13.93	4.11×10^{-44}	--	--	--	--
IRF2BP1	12.28	1.21×10^{-34}	Heart Atrial Appendage	-0.94	0.73	4.34×10^{-1}
EML2	7.98	1.50×10^{-15}	Cerebellar Hemisphere	-0.22	2.72	4.01×10^{-10}
CR1	7.91	2.63×10^{-15}	Esophagus Muscularis	4.30	3.45	1.69×10^{-11}
CLPTM1	-6.30	2.94×10^{-10}	Skin Sun Exposed leg	-2.98	4.89	3.35×10^{-23}
TRAPC6A	-6.29	3.22×10^{-10}	Thyroid	-1.02	2.84	2.26×10^{-13}
ZNF45	-6.12	9.18×10^{-10}	Stomach	0.19	2.72	2.47×10^{-23}
DMWD	5.91	3.48×10^{-9}	Adipose Visceral Omentum	3.45	1.45	3.08×10^{-8}
ZNF223	5.90	3.69×10^{-9}	Brain Cerebellum	1.03	1.17	5.88×10^{-4}
PVR	-4.82	1.41×10^{-6}	Spleen	-4.64	2.05	9.68×10^{-27}
AP2A2	-4.65	3.30×10^{-6}	Heart Atrial Appendage	-0.30	1.93	2.48×10^{-2}

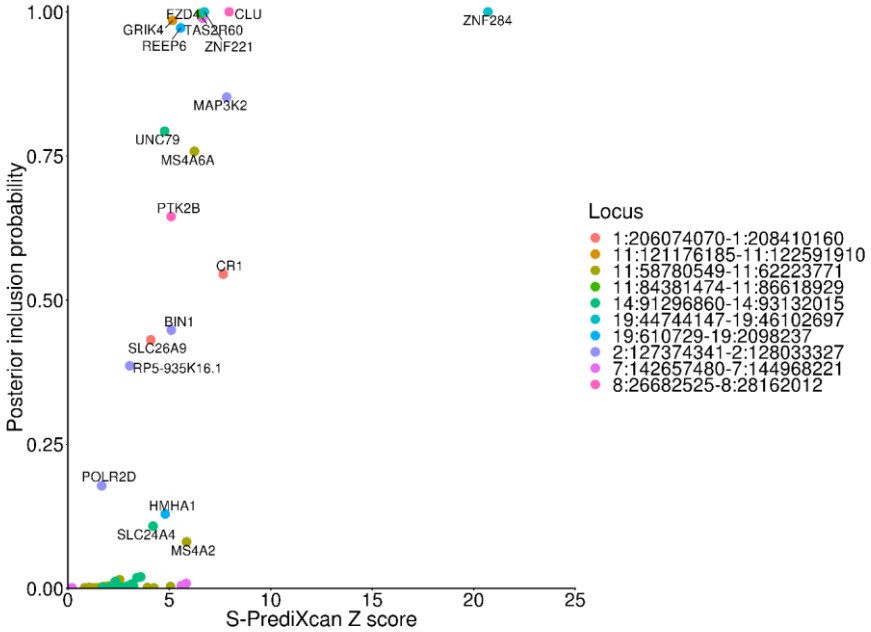
Notes: DLPFC, dorsolateral prefrontal cortex

3.3. Fine-mapping further prioritises genes within GWAS risk loci

We applied the fine-mapping of causal gene gets (FOCUS) algorithm to prioritise genes within GWAS risk loci. Genes with a higher posterior inclusion probability tended to have a higher S-PrediXcan Z score (Spearman correlation = 0.8269, P = 1.64 ×10^{-87}) (Figure 2). Candidate casual genes not nearest the GWAS index SNP included *GRIK4* (*SROL1* locus; S-PrediXcan Z score: -5.16; PIP: 0.985) and *UNC79* (*SLC24A4* locus: S-PrediXcan Z score: 4.77; PIP:

0.793) (Supplementary Table S6). Both *GRIK1* and *UNC79* are involved in ion transmembrane transport and were not prioritised as likely causal genes in a recent GWAS of Alzheimer's disease [3], highlighting the potential utility of FOCUS in gene prioritisation.

Figure 2: Marginal posterior inclusion probability of credible casual genes versus the S-PrediXcan Z score by chromosomal region



~~1.1.~~3.4. *A biological pathway analysis identifies altered expression of lipoprotein clearance pathways in Alzheimer's disease*

We tested for the enrichment of ~~S-TissueXcan~~ Alzheimer's disease associations in Reactome biological pathways by regressing gene pathway membership against the (signed) ~~z-~~Z-score from the ~~S-TissueXcan~~ S-~~PrediXcan~~MultiXcan analyses. This approach allowed us to assess the enrichment of ~~S-TissueXcan~~Alzheimer's disease associations within biological pathways, as well as the mean effect size and effect direction of gene expression within the enriched pathways. In the (multi-tissue) GTEx S-MultiXcan analysis, ~~One~~ one pathway ~~—~~ "plasma

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Not Italic
Formatted: Font: Bold
Formatted: Line spacing: single
Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55"

lipoprotein clearance" was significantly downregulated in Alzheimer's disease cases after correction for multiple testing (beta coefficient = -0.7861, P = 6.64 × 10^{-6}) (Table 45, Supplementary Table S6S7)), consistent with the known pathoaetiology of Alzheimer's disease. Plasma lipoprotein clearance was also significantly downregulated in cases using the CMC data (beta=-0.5646; P=8.31 × 10^{-26}). Furthermore, we identified the upregulation of multiple immune-related pathways, especially related to Toll Like Receptor (TLR) cascades (e.g. Toll Like Receptor TLR1:TLR2 Cascade; beta=0.3684, 1.32 × 10^{-44}) (Table 5, Supplementary Table S8), using the CMC data.

Table 5: Biological pathways associated with Alzheimer's disease association signals in dorsolateral prefrontal cortex from the CMC

Pathway ID	Pathway name	Coef	SE	P
S-MultiXcan				
R-HSA-8964043	Plasma lipoprotein clearance	-0.7861	0.1745	6.64×10^{-6}
CMC DLPFC				
R-HSA-168179	Toll Like Receptor TLR1:TLR2 Cascade	0.3684	0.0263	1.32×10^{-44}
R-HSA-167044	Signalling to RAS	0.6713	0.0493	2.78×10^{-42}
R-HSA-187687	Signalling to ERKs	0.5786	0.0438	6.52×10^{-40}
R-HSA-447115	Interleukin-12 family signalling	0.6109	0.0468	7.27×10^{-39}
R-HSA-354192	Integrin alphaIIb beta3 signalling	0.6337	0.0486	8.14×10^{-39}

Notes: Coef, beta coefficient from a logistic regression model testing the enrichment of genes associated with Alzheimer's disease in Reactome pathways.

2.4.DISCUSSION

We performed multi-tissue analysis of gene expression underlying Alzheimer's disease to identify and prioritise candidate causal genes and pathogenic tissues. Using the transcriptome-wide association study method S-PrediXcan and tissue-specific eQTL information from GTEx, we identified 50 unique candidate risk genes for Alzheimer's disease. A meta-analysis of these tissue-specific data found 73 genes associated with Alzheimer's disease. Because GTEx-derived brain tissues may lack sufficient power to identify robust association signals underlying

Formatted: Font: Italic

Formatted: Font: Times New Roman, 12 pt, Font color: Auto

Formatted Table

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Left

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

complex diseases, we ran S-PrediXcan using expression weights derived from 646 dorsolateral prefrontal cortex samples from the CommonMind Consortium. We identified 12 gene-based associations, 8 of which were also significant in the meta-analysed GTEx analysis. Fine-mapping of causal gene sets further prioritised novel gene candidates within 10 independent risk loci. Biological ~~P~~pathway analysis of the meta-analysed GTEx data and CMC data ~~the meta-analysed association signals found enrichment of~~found down-regulation of genes involved in plasma lipoprotein clearance. Furthermore, the CMC data strongly implicated upregulation of genes involved in immune-related pathways and processes, particularly toll-like receptor activity. ~~–~~These results highlight the utility of investigating multiple tissues underlying complex disorders, including peripheral tissues unrelated to the pathogenic tissue of interest (such as skin tissue for brain-related processes in Alzheimer's disease)[7]. Our results demonstrate a multi-tissue approach to gene discovery in Alzheimer's disease may not only identify candidate causal genes and pathways, but peripheral (i.e. accessible) surrogate tissues for diagnostic biomarkers and the discovery of causal mechanisms. ~~and gene expression data from 48 tissues from the GTEx project, we identified 50 unique genes across 45 tissues that reached transcriptome-wide significance. The largest number of significant associations were found in sun-exposed lower leg tissue, and included differentially expressed genes previously thought to be involved in AD pathophysiology (for example, *APOE*, *CLU*, and *PPP1R13L*). To increase power relative to the single-tissue analyses, we meta-analysed the single-tissue association signals using S-TissueXcan and found an additional 27 genes significantly associated with Alzheimer's disease (73 gene-level associations in total).~~

~~Pathway analysis of the meta-analysed association signals found enrichment of genes involved in plasma lipoprotein clearance. These results highlight the utility of investigating multiple tissues underlying complex disorders, including peripheral tissues unrelated to the pathogenic tissue of interest (such as skin tissue for brain-related processes in Alzheimer's disease). Our~~

~~results demonstrate a multi-tissue approach to gene discovery in Alzheimer's disease may not only identify candidate causal genes and pathways, but peripheral (i.e. accessible) surrogate tissues for diagnostic biomarkers and the discovery of causal mechanisms.~~

Formatted: Normal, Justified

Two recent studies performed transcriptome-wide association analyses of brain samples in Alzheimer's disease. Raj *et al.*[15] used TWAS FUSION[16] with eQTL data derived from 450 frontal cortex samples and genotype data from the Religious Order Study or the Memory and Aging Project (ROS/MAP), while Marioni *et al.*[17] applied Summary-data-based Mendelian Randomization (SMR)[18] to GWAS summary data from a meta-analysis of proxy Alzheimer's disease cases from the UK Biobank and IGAP meta-analysis summary data, and eQTL data from over 600 frontal cortex samples from the Common Mind Consortium. These analyses identified a total of 9 candidate genes whose expression in brain tissue was associated with Alzheimer's disease. We found a significant association with 4 of these candidate genes (*CR1*, *TOMM40*, *PVR*, *CLPTM1*) in at least one peripheral tissue. The effect direction of the beta coefficients in our study had the same effect directions for the candidate genes *CR1*, *PVR*, *CLPTM1*, and the strongest associations were found in peripheral tissues, including skin.

We observed largely concordant effect directions in the S-PrediXcan association statistics (z scores) across brain and peripheral tissues, which can be expected given the observed high level of tissue-shared eQTL regulation at GWAS loci [19]. Furthermore, eQTL sharing among brain and skin—the peripheral tissue with the highest number of Alzheimer's disease associations—is higher than other peripheral tissues [20]. These results ~~highlight the utility of studying multiple (~~indicate accessible peripheral~~)~~ tissues, especially skin, may capture the genetic effects on gene expression underlying Alzheimer's disease and other brain-related traits. Future studies can therefore increase power to identify molecular effects in Alzheimer's

disease by studying eQTL effects in large peripheral tissue eQTL datasets, before the use fine-mapping techniques in disease-relevant brain tissue.~~in genetic studies of brain-related traits, where larger sample sizes may provide increased power to identify biologically meaningful associations.~~

Transcriptome imputation methods such as S-PrediXcan are prone to false positive associations due to linkage disequilibrium between SNPs used to build the expression weights, which induce spurious gene-trait associations within chromosomal regions. We used fine-mapping of causal gene sets to further prioritise genes within risk loci. We found the probability for each gene in a region to be causal was a largely a function of its S-PrediXcan Z score, where genes with larger Z scores had larger posterior inclusion probabilities as the causal gene. Nonetheless, we identified 6 genes that were not reported as the closest gene within ± 100 kb of the top SNP of known GWAS-defined associated genes at the time of publication of Lambert et al. [2], which represent novel, functionally relevant candidate causal genes in Alzheimer's disease. Among these novel candidates is *GRIK4* at the *SORL1* locus and *UNC79* at the *SLC24A4-RIN3* locus. Both *GRIK4* (glutamate ionotropic receptor kainate type subunit 4) and *UNC79* (unc-79 homolog, NALCN channel complex subunit) have biased expression in the brain and encode ion channel subunits, and it is conceivable their dysfunction may contribute to altered synaptic plasticity, learning and development in Alzheimer's disease[21].

~~Gene~~ Biological pathway analysis of genes ~~from~~ in both our meta-analysed ~~(S-TissueXcan)~~GTEx and CMC results ~~found one pathway~~found down-regulation of ~~—~~"plasma lipoprotein clearance" in Alzheimer's disease. ~~" significantly downregulated in Alzheimer's disease cases compared to controls.~~ These results are consistent with a recent meta-analysis of cross-tissue expression imputation of 44 GTEx tissues[22], which found the enrichment genes

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

whose expression was associated with Alzheimer's disease in gene ontology terms related to lipoprotein clearance. Lipoprotein clearance ~~has a well-established~~may play an important role in Alzheimer's disease pathogenesis through the association of *APOE* and several other genes that function in lipid or lipoprotein metabolism, including Clusterin (*CLU*) and ATP binding cassette (ABC) transporter A7 (*ABCA7*)[23]. Specifically, it has been hypothesised that dysfunctional lipoprotein clearance in the central nervous system ~~is thought to be involved in~~may facilitate the formation of two critical neuroanatomical features in Alzheimer's disease: amyloid plaques and neurofibrillary tangles. These neuroanatomical features may be indicated by global changes in gene (mRNA) and protein expression of lipid and lipoprotein-related genes in both brain tissue and peripheral blood [24]. The association of lipoprotein-related genes with Alzheimer's disease in skin and other non-brain tissues, together with concordant effect directions across tissues (including brain tissue), suggests peripheral tissues may provide a biologically valid substrate for the study of genetic factors and their impact on higher order molecular processes in Alzheimer's disease.

Pathway analysis of the CMC gene-based found the up-regulation of genes involved in immune-related processes, most notably toll-like receptor cascades. Toll-like receptors are involved in many physiological and pathological responses, and their activity is thought to play a role in several neurological disorders, including Alzheimer's disease [25,26]. The receptors are widely expressed on microglial cells—the chief immune cells of the central nervous system—and their activation is associated with Aβ plaque deposition [27] and enhanced neurodegeneration [28]. Although we cannot draw mechanistic conclusions, our results suggest a potential relationship between altered immune signalling, impaired plasma lipoprotein clearance, and Aβ plaque deposition in Alzheimer's disease.

Our multi-tissue transcriptome ~~-wide association~~ imputation approach has a number of advantages over traditional expression quantitative loci studies of complex diseases. First, ~~TWAS~~ transcriptome imputation methods allow the study of genetically regulated gene expression without directly measuring expression data from an appropriate cell type in diseased cases and health controls. Second, by ~~imputing~~ estimating the genetically regulated component of gene expression, ~~TWAS~~ transcriptome imputation methods remove the impact of unmeasured (i.e. uncontrolled) environmental factors on gene expression, thereby improving the interpretability of expression association signals. Third, transcriptome ~~TWAS~~ imputation aggregates SNP level associations to individual genes, reducing the multiple testing burden and increasing statistical power. A multi-tissue meta-analysis, such as ~~TissueXcan~~S-MultiXcan, further reduces the multiple testing burden by combining association statistics across all interrogated tissues. Fourth, TWAS methods utilise eQTL information from large eQTL databases with uniform sample collection and strict quality control protocols which improves the reliability of results and enables replication across disorders/traits.

A disadvantage of the use of datasets such as GTEx is that tissues are not homogeneous, and thus under represent certain cell populations. Many of the ~~AD~~Alzheimer's disease risk loci identified through GWAS are not highly expressed in whole brain tissues. Previous attempts to identify brain tissue eQTLs corresponding to ~~AD~~Alzheimer's disease GWAS loci have likely been affected by this ~~issue~~ cellular heterogeneity [29,30]. A large proportion of ~~AD~~Alzheimer's disease risk loci have been linked to immune function, and our results in (dorsolateral prefrontal cortex) brain tissue corroborate these findings. However, the study of immune function in the brain is complicated by ~~The~~ the heterogeneous ~~population of brain~~ cell populations, which ~~s~~ dilute~~s~~ ~~cell~~immune-specific signatures ~~of~~ from small populations of cells

such as microglia. Analyses of primary cell-type specific expression from the Immune Variation project have shown that ~~AD~~Alzheimer's disease risk alleles are enriched among monocyte-specific eQTLs~~, as opposed to T cell-specific eQTLs~~. More easily accessible ~~Monocytes~~ monocytes could be used as a proxy to examine ~~the~~ the (immune) cell-specific effects of ~~microglia~~ susceptibility variants in ~~AD~~Alzheimer's disease. ~~This implication of specific immune cell types points to the need to identify the cell-autonomous effects of disease susceptibility variants.~~

~~3.~~5. CONCLUSIONS

In summary, we performed a multi-tissue transcriptome-wide association study of Alzheimer's disease. We confirmed an association between DNA sequence variation and gene expression for known Alzheimer's disease candidate genes and identified multiple genes whose expression has not previously been associated with the disease. ~~Most~~ Many disease associations were observed in peripheral tissues, most notably skin tissue, rather than brain tissues, and the effect directions for the association statistics were largely consistent across tissues. This suggests accessible peripheral tissues such as skin may provide biologically meaningful surrogate information for brain-related processes. A meta-analysis of 48 GTEx tissues, including 13 brain tissues~~,~~ confirmed the association of candidate genes identified in single tissue analyses, in addition~~al~~ to several novel genes, most of which were also identified in an analysis of gene expression in dorsolateral prefromtal cortex. These results suggest gene expression data from peripheral tissues improves power to identify and prioritise candidate genes for brain-related traits. The use of skin tissue~~, where~~ the peripheral tissue with ~~-~~the largest number of associations with Alzheimer's disease~~ was observed,~~ represents a particularly useful avenue for future research, and might provide a useful surrogate for biomarker discovery for disease onset and progression.

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

DECLARATIONS

1. Ethics approval and consent to participate

Not applicable

2. Consent for publication

Not applicable

3. Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

4. *Competing interests*

ERG receives an honorarium from the journal Circulation Research of the American Heart Association, as a member of the Editorial Board. He also performs consulting on pharmacogenetic analysis with the City of Hope / Beckman Research Institute. ~~The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.~~

5. Funding

The i–Select chips were funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Centre and Erasmus University. ADGC was supported

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Auto

Formatted: Justified, Space After: 0 pt, Pattern: Clear (White)

Formatted: Font: (Default) Times New Roman, 12 pt, Not Italic, Font color: Auto

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Auto

by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.

6. *Authors' contributions*

Conceptualization: ZFG; Formal analysis: ZFG, DE; Methodology: ZFG, ERG, EMD; Supervision: ERG, EMD, MLK; Writing – original draft: ZFG; Writing – review & editing: ZFG, EMD, DE, MLK

7. *Acknowledgements*

We thank the International Genomics of Alzheimer's Project (IGAP) for providing ~~AD~~Alzheimer's disease meta-analysis summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. ERG is grateful to the President and Fellows of Clare Hall, University of Cambridge for the stimulating intellectual environment during his fellowship in the college. ERG is supported by a NIH Genomic Innovator Award (R35HG010718). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Formatted: Font: (Default) Times New Roman, 12 pt, Font color: Auto

Formatted: Line spacing: Double, Pattern: Clear

REFERENCES

1. 2017 Alzheimer's disease facts and figures. Alzheimer's Dement [Internet]. 2017;13:325–73. Available from: http://www.sciencedirect.com/science/article/pii/S1552526017300511

2. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45:1452–8.

3. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019;51:414–30.

4. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. Elsevier; 2017;101:5–22.

5. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;advance on.

6. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat Neurosci. 2018;

7. Gamazon ER, Segrè A V, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet [Internet]. 2018;50:956–67. Available from: https://doi.org/10.1038/s41588-018-0154-4

8. eGTEx Project. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2017;

9. Barbeira A, Dickinson SP, Torres JM, Torstenson ES, Zheng J, Wheeler HE, et al. Integrating tissue specific mechanisms into GWAS summary results. bioRxiv. 2017;

10. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. Am J Hum Genet. 2017;100:473–87.

11. Delaneau O, Marchini J, Consortium T 1000 GP. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;5:3934. Available from: http://dx.doi.org/10.1038/ncomms4934

12. Delaneau O, Marchini J, Consortium T 1000 GP. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;5:3934.

13. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9:1825.

14. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, et al. Probabilistic fine-mapping of transcriptome-wide association studies. Nat Genet. 2019;51:675–82.

15. Raj T, Li YI, Wong G, Humphrey J, Wang M, Ramdhani S, et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet. 2018;50:1584–92.

16. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52.

17. Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. Transl Psychiatry. Springer US; 2018;8:8–14.

18. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48:481–7.

19. Gamazon ER, Segrè A V, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet. 2018;

20. Consortium Gte, Aguet F, Brown AA, Castel SE, Davis JR, He Y, et al. Genetic effects on gene expression across human tissues. Nature [Internet]. The Author(s); 2017;550:204. Available from: https://doi.org/10.1038/nature24277

21. Maragakis NJ, Rothstein JD. Glutamate Transporters in Neurologic Disease. JAMA Neurol. 2001;58:365–70.

22. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. bioRxiv. 2019;286013.

23. Stukas S, Kulic I, Shahab Z, Cheryl L. W. Lipids and Lipoproteins in Alzheimer's Disease. In: Zerr I, editor. Alzheimer's Dis. Rijeka: IntechOpen; 2015. p. 99–139.

24. Allen M, Zou F, Chai HS, Younkin CS, Crook J, Pankratz VS, et al. Novel late-onset Alzheimer disease loci variants associate with brain gene expression. Neurology [Internet]. 2012;79:221 LP – 228. Available from: http://n.neurology.org/content/79/3/221.abstract

25. Su F, Bai F, Zhou H, Zhang Z. Microglial toll-like receptors and Alzheimer's disease. Brain Behav Immun [Internet]. 2016;52:187–98. Available from: http://www.sciencedirect.com/science/article/pii/S0889159115300350

26. Chakrabarty P, Li A, Ladd TB, Strickland MR, Koller EJ, Burgess JD, et al. TLR5 decoy receptor as a novel anti-amyloid therapeutic for Alzheimer's disease. J Exp Med [Internet]. 2018;215:2247 LP – 2264. Available from: http://jem.rupress.org/content/215/9/2247.abstract

27. Walter S, Letiembre M, Liu Y, Heine H, Penke B, Hao W, et al. Role of the Toll-Like Receptor 4 in Neuroinflammation in Alzheimer's Disease. Cell Physiol Biochem [Internet]. 2007;20:947–56. Available from: https://www.karger.com/DOI/10.1159/000110455

28. Lehmann SM, Krüger C, Park B, Derkow K, Rosenberger K, Baumgart J, et al. An unconventional role for miRNA: let-7 activates Toll-like receptor 7 and causes neurodegeneration. Nat Neurosci [Internet]. 2012;15:827–35. Available from: https://doi.org/10.1038/nn.3113

29. Pimenova AA, Raj T, Goate AM. Untangling Genetic Risk for Alzheimer's Disease. Biol Psychiatry [Internet]. 2018;83:300–10. Available from: http://www.sciencedirect.com/science/article/pii/S0006322317315871

30. Efthymiou AG, Goate AM. Late onset Alzheimer's disease genetics implicates microglial pathways

in disease risk. Mol Neurodegener [Internet]. BioMed Central; 2017;12:43. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28549481

SUPPLEMENTARY INFORMATION

Supplementary Table 1: Bonferroni corrected P value thresholds after adjusting for the number of genes expressed in a given tissue (i.e. 0.05/number of genes in tissue).

Supplementary Table 2: Single-tissue S-PrediXcan results across 48 GTEx tissues

Supplementary Table 3: Single-tissue S-PrediXcan results across 48 GTEx collapsed by gene

Supplementary Table 4: Significant S-TissueXcan results

Supplementary Table 5: Single-tissue S-PrediXcan results from dorsolateral prefrontal cortex (CMC) ~~Genes identified as significantly associated with Alzheimer's disease using S-PrediXcan but not S-TissueXcan~~

Supplementary Table 6: Fine-mapping of causal gene sets (FOCUS) of Alzheimer's disease GWAS summary statistics

Supplementary Table 7: Pathway analysis of GTEx S-TissueXcan results

Supplementary Table 8: Pathway analysis of CMC S-PrediXcan results

Supplementary Table 9: Pathway analysis of GTEx S-TissueXcan results, no APOE region

Supplementary Table 10: Pathway analysis of CMC S-PrediXcan results, no APOE region

Supplementary Figure 1: Number of significant S-PrediXcan associations against GTEx tissue sample size

Supplementary Figure 2~~1~~: Expression of APOE (ENSG00000130203.5) across GTEx tissues

Formatted: Font: Bold

Figure S1


Click here to access/download
Supplementary Material
Figure_S1.tiff

Figure S2

Click here to access/download
Supplementary Material
Figure_S2.png

Table S1


Click here to access/download
Supplementary Material
Table_S1.xlsx

Table S2

Click here to access/download
Supplementary Material
Table_S2.xls

Table S3

Click here to access/download
Supplementary Material
Table_S3.xlsx

Table S5

Click here to access/download
Supplementary Material
Table_S5.xlsm

Table S6

Table S7

Click here to access/download
Supplementary Material
Table_S7.xlsx

Table S8

Table S9

Table S10