

RESEARCH

Open Access



# Multi-omics and pathway analyses of genome-wide associations implicate regulation and immunity in verbal declarative memory performance

Hao Mei<sup>1,2\*</sup>, Jeannette Simino<sup>1,2\*</sup>, Lianna Li<sup>3</sup>, Fan Jiang<sup>4</sup>, Joshua C. Bis<sup>5</sup>, Gail Davies<sup>6</sup>, W David Hill<sup>6</sup>, Charley Xia<sup>6</sup>, Vilmundur Gudnason<sup>7,8</sup>, Qiong Yang<sup>9,10</sup>, Jari Lahti<sup>11,12</sup>, Jennifer A. Smith<sup>13</sup>, Mirna Kirin<sup>14</sup>, Philip De Jager<sup>15,16,17</sup>, Nicola J. Armstrong<sup>18</sup>, Mohsen Ghanbari<sup>19</sup>, Ivana Kolcic<sup>20</sup>, Christopher Moran<sup>21,22</sup>, Alexander Teumer<sup>23</sup>, Murali Sargurupremraj<sup>24</sup>, Shamsed Mahmud<sup>1</sup>, Myriam Fornage<sup>25,26</sup>, Wei Zhao<sup>13</sup>, Claudia L. Satizabal<sup>27</sup>, Ozren Polasek<sup>20,28</sup>, Katri Räikkönen<sup>12</sup>, David C. Liewald<sup>6</sup>, Georg Homuth<sup>29</sup>, Michele Callisaya<sup>22,30</sup>, Karen A. Mather<sup>31,32</sup>, B. Gwen Windham<sup>2,33</sup>, Tatijana Zemunik<sup>20</sup>, Aarno Palotie<sup>34,35</sup>, Alison Pattie<sup>6</sup>, Sandra van der Auwera<sup>36</sup>, Anbupalam Thalamuthu<sup>31,32</sup>, David S. Knopman<sup>37</sup>, Igor Rudan<sup>38</sup>, John M. Starr<sup>6,39</sup>, Katharina Wittfeld<sup>36,40</sup>, Nicole A. Kochan<sup>31</sup>, Michael E. Griswold<sup>2,41</sup>, Veronique Vitart<sup>42</sup>, Henry Brodaty<sup>31,43</sup>, Rebecca Gottesman<sup>44</sup>, Simon R. Cox<sup>6</sup>, Bruce M. Psaty<sup>5,45,46,47</sup>, Eric Boerwinkle<sup>26,48</sup>, Daniel I. Chasman<sup>49,50</sup>, Francine Grodstein<sup>51,52</sup>, Perminder S. Sachdev<sup>31,53</sup>, Velandai Srikanth<sup>21,22,30</sup>, Caroline Hayward<sup>42</sup>, James F. Wilson<sup>38,42</sup>, Johan G. Eriksson<sup>54,55,56</sup>, Sharon L. R. Kardina<sup>13</sup>, Hans J. Grabe<sup>36,40</sup>, David A. Bennett<sup>57,58</sup>, M. Arfan Ikram<sup>19</sup>, Ian J. Deary<sup>6</sup>, Cornelia M. van Duijn<sup>59</sup>, Lenore Launer<sup>60</sup>, Annette L. Fitzpatrick<sup>45,61</sup>, Sudha Seshadri<sup>10,62,63</sup>, Jan Bressler<sup>26</sup>, Stephanie Debette<sup>24,63,64</sup> and Thomas H. Mosley Jr<sup>2,41,65</sup>

## Abstract

**Background** Uncovering the functional relevance underlying verbal declarative memory (VDM) genome-wide association study (GWAS) results may facilitate the development of interventions to reduce age-related memory decline and dementia.

**Methods** We performed multi-omics and pathway enrichment analyses of paragraph (PAR-dr) and word list (WL-dr) delayed recall GWAS from 29,076 older non-demented individuals of European descent. We assessed the relationship between single-variant associations and expression quantitative trait loci (eQTLs) in 44 tissues and methylation quantitative trait loci (meQTLs) in the hippocampus. We determined the relationship between gene associations and transcript levels in 53 tissues, annotation as immune genes, and regulation by transcription factors (TFs) and microRNAs. To identify significant pathways, gene set enrichment was tested in each cohort and meta-analyzed across cohorts. Analyses of differential expression in brain tissues were conducted for pathway component genes.

\*Correspondence:

Hao Mei

hmei@umc.edu

Jeannette Simino

jsimino@umc.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Results** The single-variant associations of VDM showed significant linkage disequilibrium (LD) with eQTLs across all tissues and meQTLs within the hippocampus. Stronger WL-dr gene associations correlated with reduced expression in four brain tissues, including the hippocampus. More robust PAR-dr and/or WL-dr gene associations were intricately linked with immunity and were influenced by 31 TFs and 2 microRNAs. Six pathways, including type I diabetes, exhibited significant associations with both PAR-dr and WL-dr. These pathways included fifteen MHC genes intricately linked to VDM performance, showing diverse expression patterns based on cognitive status in brain tissues.

**Conclusions** VDM genetic associations influence expression regulation via eQTLs and meQTLs. The involvement of TFs, microRNAs, MHC genes, and immune-related pathways contributes to VDM performance in older individuals.

**Keywords** Genome-wide association study, Memory, Expression, Immunity, Multi-omics, Delayed recall

## Background

Delayed verbal declarative memory (VDM) performance, commonly measured by paragraph and word list delayed recall tests, is an important predictor of Alzheimer's disease (AD) [1]. Genome-wide association studies (GWAS) have leveraged VDM performance (heritability  $\approx 30\text{--}52\%$  [2, 3]) to identify variants influencing brain aging and AD susceptibility. The largest such GWAS, led by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Cognitive Working Group, identified three significant chromosomal regions (near *APOE*, *HS3ST4*, and *SPOCK3*) in a sample of 29,076 older non-demented participants of European descent [2]. A genetic risk score combining fifty-eight independent suggestive variants was associated with AD pathology (neurofibrillary tangle density and amyloid plaque burden) in autopsy samples [2], demonstrating that genetic studies of VDM can provide insight into the molecular contributors to AD pathobiology.

GWAS often implicate non-coding regions suspected to influence regulation [4], lack power to detect the small effect sizes bestowed by most genetic variants [5], are encumbered by the heterogeneity of genetic effects across studies [6], and have severe multiple testing corrections [5, 7, 8]. The integration of additional biological resources and aggregation of effects across genes and pathways can address these limitations and facilitate the interpretation of GWAS results [9] to understand biological functions [4]. Both multi-omics and pathway analyses can integrate GWAS findings with functional information from publicly available databases to gain insight into complex trait pathobiology [9] and provide context to interpret genotype–phenotype relationships [4].

DeBette et al. identified a VDM-associated genetic variant in proximity to genes linked to immune responses [2]. Additionally, they found that variants associated with suggestive memory risks correlate with gene expressions in human hippocampus samples. Building upon these findings, our study endeavors to expand beyond the limitations of prior research

by delving into the potential functions associated with VDM-related genetic variants. To achieve this, we employed multi-omics analyses to explore the intricate relationship between VDM-associated genetic variants and expression quantitative trait loci (eQTLs), methylation quantitative trait loci (meQTLs), and gene expressions across diverse tissue types. Our investigation also meticulously examined how the associations of genetic variants with VDM are intertwined with the regulatory activities of transcription factors (TFs) and microRNAs, along with immune gene functions. Additionally, we undertook the task of evaluating the genetic pathways that underlie the associations related to paragraph delayed recall (PAR-dr) and word list delayed recall (WL-dr) [4], along with exploring links between pathway gene expressions and cognitive status in brain tissues.

## Methods

### Participating cohorts and phenotypes

This study utilized data from twenty-seven cohorts comprising individuals of Caucasian descent, divided into 19 for the initial discovery phase and 8 for replication. The dataset included HapMap-imputed genome-wide single-nucleotide polymorphism (SNP) data, and at least one test of PAR-dr or WL-dr. Consequently, we conducted the analyses in this study using summarized data from a prior GWAS meta-analysis that specifically focused on PAR-dr and WL-dr within these cohorts [2]. Detailed information about these cohorts can be found in the [Supplemental Text](#) and Tables S1 and S2.

Participants provided written informed consent and all studies were approved by their respective institutional review boards. The nineteen discovery cohorts (8 for PAR-dr and 15 for WL-dr) collectively represented 29,076 ( $N_{\text{PAR-dr}} = 6674$ ;  $N_{\text{WL-dr}} = 24,604$ ) dementia- and stroke-free Caucasian participants aged 45 years or older (Figure S1). The eight replication cohorts represented approximately 8000 ( $N_{\text{PAR-dr}} = 8009$ ;  $N_{\text{WL-dr}} = 7518$ ) stroke-free Caucasian participants aged 65 years or older; dementia assessment was not universally available in the replication cohorts and

seven of the eight replication cohorts were restricted to women with some college education.

For the PAR-dr tests, participants were verbally presented one or two stories and asked to recall as many paragraph elements as possible after a 20- or 30-min delay and an interceding immediate recall task. For the WL-dr tests, participants were verbally or visually presented a list of semantically related or unrelated words (10–16 words over 1–5 exposure trials) and asked to recall as many words as possible after a 3- to 30-min delay and an interceding immediate recall task. The outcomes were the total number of items recalled during the delayed recall tasks.

### Cohort-specific genetic associations

#### Single-variant associations

Separate GWAS analyses were performed for PAR-dr and WL-dr within each cohort; the cohort-specific summary results for each trait were obtained from the CHARGE consortium. Within each cohort, a linear regression model of the number of story elements or words recalled was fit onto the number of minor alleles at each SNP while adjusting for age and sex, as well as study site, familial structure, and population substructure if necessary [2]. Subsequently, single-variant associations from each participating cohort were gathered for further analysis.

#### Gene associations

We measured gene associations from independent SNPs in each cohort. GWAS SNPs ( $\approx 1.5$  to  $2.4$  million per GWAS) were mapped to genes ( $\approx 35,000$  to  $38,000$  including non-RNA coding genes) using 2 kb upstream/downstream boundaries of the transcription start/stop sites (Tables S1 and S2), referencing genome Build GRCh37. Within each gene, pairwise SNP correlation coefficients ( $r^2$ ) were calculated using VCFtools [10] and the European reference data from the 1000 Genomes project. Clumping was conducted to select independent SNPs through an iterative process; at each step, we selected the SNP with the strongest association and removed SNPs correlated ( $r^2 > 0.2$ ) to it.

We computed Simes' combination  $p$ -value of gene [11] as  $M = \min(k \bullet p_{(i)}/i)$ , where  $k$  was the number of total independent SNPs and  $p_{(i)}$  was the  $i$ th smallest  $p$ -value. Gene uniform-score ( $U$ -score) [12] was applied to measure gene association and it was calculated as  $= (\sum_{j=1}^L I(M_j < M) + 0.5 \sum_{j=1}^L I(M_j = M))/L$ , where  $M_j$  was the combination  $p$ -value of the  $j$ th gene and  $L$  is the total number of genes. Gene  $U$ -score ranges from zero to one, and it estimates the proportion of genes with a stronger association than the tested gene. Genes

with  $U$ -scores  $\leq 0.05$  were selected as phenotype-associated genes.

### Meta-analysis of genetic associations

#### Single-variant associations

We employed METAL [13] to conduct a sample-size weighted meta-analysis for each phenotype (PAR-dr and WL-dr) and genetic variant across the discovery cohorts alone and the discovery and replication cohorts together.

#### Gene associations

For each gene, we counted the number of cohorts with  $U$ -scores less than or equal to 0.05. Meta-analysis  $p$ -value of each gene (Gene\_p) was computed from binomial distribution and Bonferroni-corrected significance threshold was set as  $1E-06$  ( $0.05/50,000$  to adjust for 50,000 genes tested).

### Multi-omics function analyses

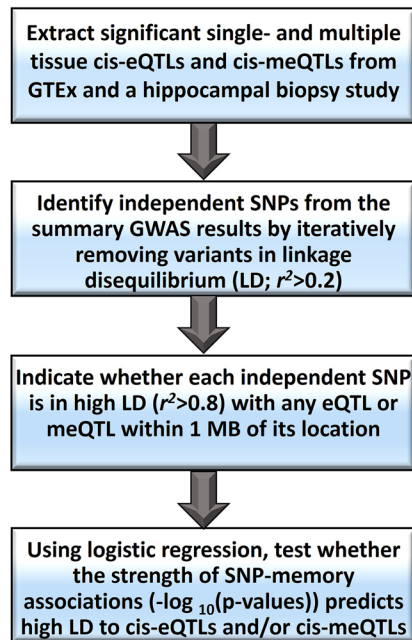
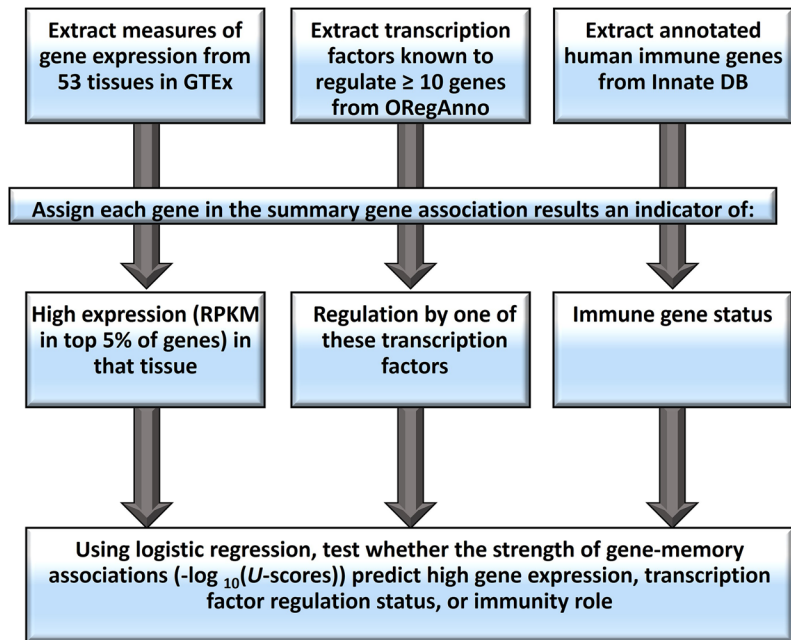
The overall design of the multi-omics function analyses for single-variant and gene associations is depicted in Fig. 1.

#### Functions of single-variant associations

We employed logistic regression to evaluate the relationship between VDM-associated genetic variants and eQTLs and meQTLs across different tissues. We extracted significant cis-eQTLs within  $\pm 1$  MB of transcription start sites from 44 different tissues of the GTEx Project [14]. We similarly extracted significant eQTLs and meQTLs from a genome-wide study of 110 human hippocampal biopsies [15]. We identified independent SNPs from meta-analysis of discovery cohorts and examined their LD status with eQTLs (and meQTLs for the hippocampal biopsy data) from each tissue alone and all tissues combined. The LD status indicated whether the SNP was in high LD ( $r^2 \geq 0.8$ ) with any eQTL or meQTL within 1 MB. We performed logistic regression of the LD status on the negative log base-10 of the single-variant association  $p$ -values in each tissue and all tissues combined. We conducted 10,000 permutations to adjust for multiple tests; permutation  $p$ -values  $\leq 0.05$  were considered significant.

#### Functions of gene associations

We utilized logistic regression to investigate potential links between VDM gene associations and gene expression, immune function, and transcription factor (TF) and microRNA regulation. We extracted GTEx gene expression, measured as reads per kilobase per million reads (RPKM), from 53 tissues via UCSC genome browser [16].

**SINGLE-VARIANT ASSOCIATIONS****GENE-BASED ASSOCIATIONS****Fig. 1** Design of the multi-omics analyses

A gene was highly expressed if its RPKM ranked in the top 5% of all genes for that tissue. We extracted 41 TFs and 52 microRNAs regulating at least ten genes from the Open Regulatory Annotation database (ORegAnno) [17]. TF regulation for a gene was identified if it was regulated by the TF/microRNA. Lastly, the immunity function of a gene was identified if it was annotated as a human immune gene in the InnateDB [18]. We fitted logistic models of status of gene expression, TF regulation, and immune function onto the  $-\log_{10}$   $U$ -scores for the gene association. An adjusted  $p$ -value  $\leq 0.05$  was considered significant, based on 1000 permutation tests.

**Pathway enrichment of genetic associations****Cohort-specific pathway associations**

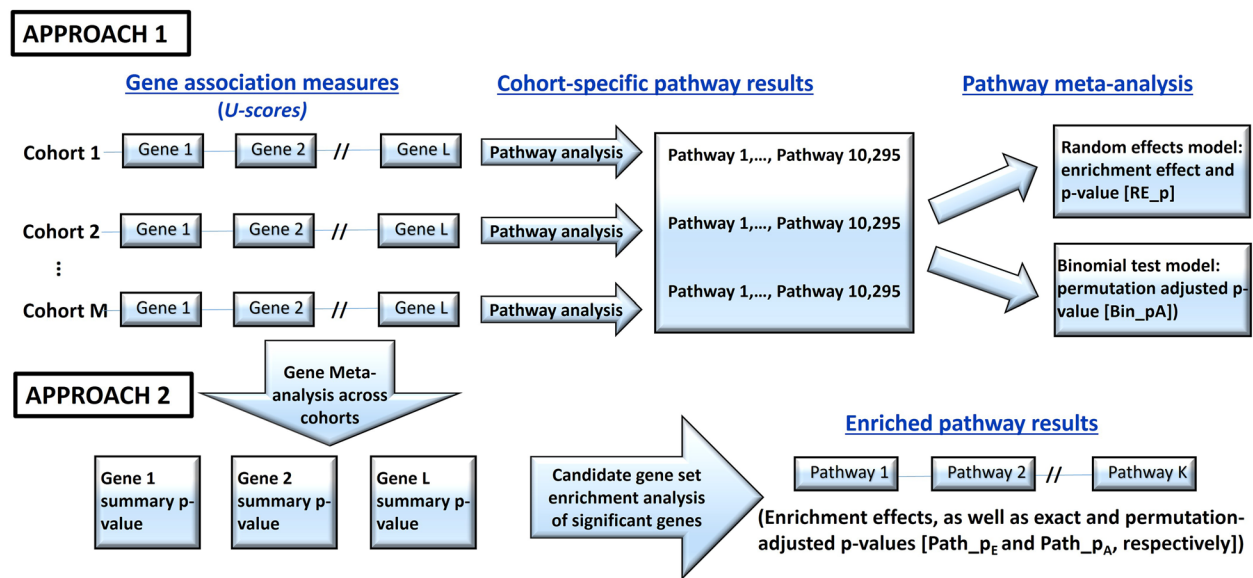
Gene set enrichment analyses were performed to examine VDM-associated pathways based on cohort-specific GWAS of PAR-dr and WL-dr. We employed the uniform-score gene-set analysis (USGSA) method [12] to test pathways enriched for genes with  $U$ -scores  $\leq 0.05$  among 10,295 curated gene sets from the MSigDB knowledge base [19] in every cohort. Pathway enrichment analysis was conducted using the R package of `snpGeneSets` [20]. For a MSigDB gene set ( $\Omega$ ) and a set of genes ( $\Phi$ ) with  $U$ -scores  $\leq 0.05$ , the probability that a component gene of  $\Omega(G_i)$  belongs to  $\Phi$  is defined as  $p_\Omega = \Pr(G_i \in \Phi | G_i \in \Omega)$  and estimated as  $\hat{p}_\Omega = \frac{\sum_i I(G_i \in \Omega \cap G_i \in \Phi)}{\sum_i I(G_i \in \Omega)}$ . In contrast,  $p_0 = 0.05$  is the null probability of a random gene ( $G_i$ )

belonging to  $\Phi$ . The pathway enrichment effect,  $E = \hat{p}_\Omega - p_0$ , shows the increased probability of a pathway component gene (versus a random gene) to have a  $U$ -score  $\leq 0.05$ , and the standard error (SE) is estimated as  $SE = \sqrt{p_0 \cdot (1 - p_0) / \sum_i I(G_i \in \Omega)}$ . The pathway exact  $p$ -value was calculated from the hypergeometric distribution; we adjusted for multiple testing and correlations due to genes belonging to multiple pathways by 10,000 permutations, yielding the adjusted  $p$ -value (path\_ $p_k$ ) in the  $k$ th cohort.

**Meta-analysis of pathway enrichment over cohorts (Approach 1)**

Two meta-analyses, random-effects (RE) model and the binomial test, were employed to estimate the effects of pathway enrichment across different cohorts and to ascertain whether the occurrence of VDM-associated pathways in the participating cohorts exhibited a non-random pattern (Figs. 2 and S2). Both meta-analyses were performed in the discovery cohorts alone, the replication cohorts alone, and all cohorts combined. The RE meta-analysis, performed using the R package `metafor` [21], incorporated the inverse variance of the effect estimate as a cohort weight. The RE model produced a summary enrichment effect estimate and a  $p$ -value (RE\_ $p$ ) of tested gene set over cohorts. The significance threshold for RE\_ $p$  in the meta-analysis of discovery cohorts alone and the discovery and replication cohorts combined was





**Fig. 2** Pictorial representation of the two approaches used to derive the overall pathway results from the cohort-specific genome-wide associations

4.86E−06 after Bonferroni correction (0.05/10,295). In the replication cohorts, a Bonferroni correction accounted for the number of pathways tested.

The binomial test was applied to count the number of cohorts with significant pathway enrichment and compute the exact  $p$ -value from binomial distribution (Supplemental Text). For the discovery cohorts alone and the discovery and replication cohorts combined, the binomial test was based on permutation-adjusted pathway  $p$ -values (path\_p<sub>k</sub>) from individual cohorts and  $p$ -value (Bin\_p<sub>A</sub>) ≤ 0.05 was considered significant. For replication cohorts alone, the  $p$ -value (Bin\_p) was based on pathway  $p$ -value from individual cohorts and Bonferroni adjustment was adopted.

#### Pathway enrichment of significant genes over cohorts (Approach 2)

Significant genes with meta-analysis  $p$ -values (i.e., Gene\_p ≤ 1E−06) were selected and tested for enrichment in a particular MSigDB gene set. The exact pathway  $p$ -value (Path\_p<sub>E</sub>) was calculated from the hypergeometric distribution; pathway  $p$ -value (Path\_p<sub>A</sub>) adjusted for multiple testing was obtained via 10,000 permutations with significance threshold of 0.05.

#### Differential expression (DE) analysis of significant pathway component genes

We performed DE analyses using significant component genes (Gene\_p ≤ 1E−06) from VDM-associated pathways. Three curated human (GDS4135 [22], GDS4231 [23], GDS4358 [24]) and rodent (GDS2082 [25], GDS2639

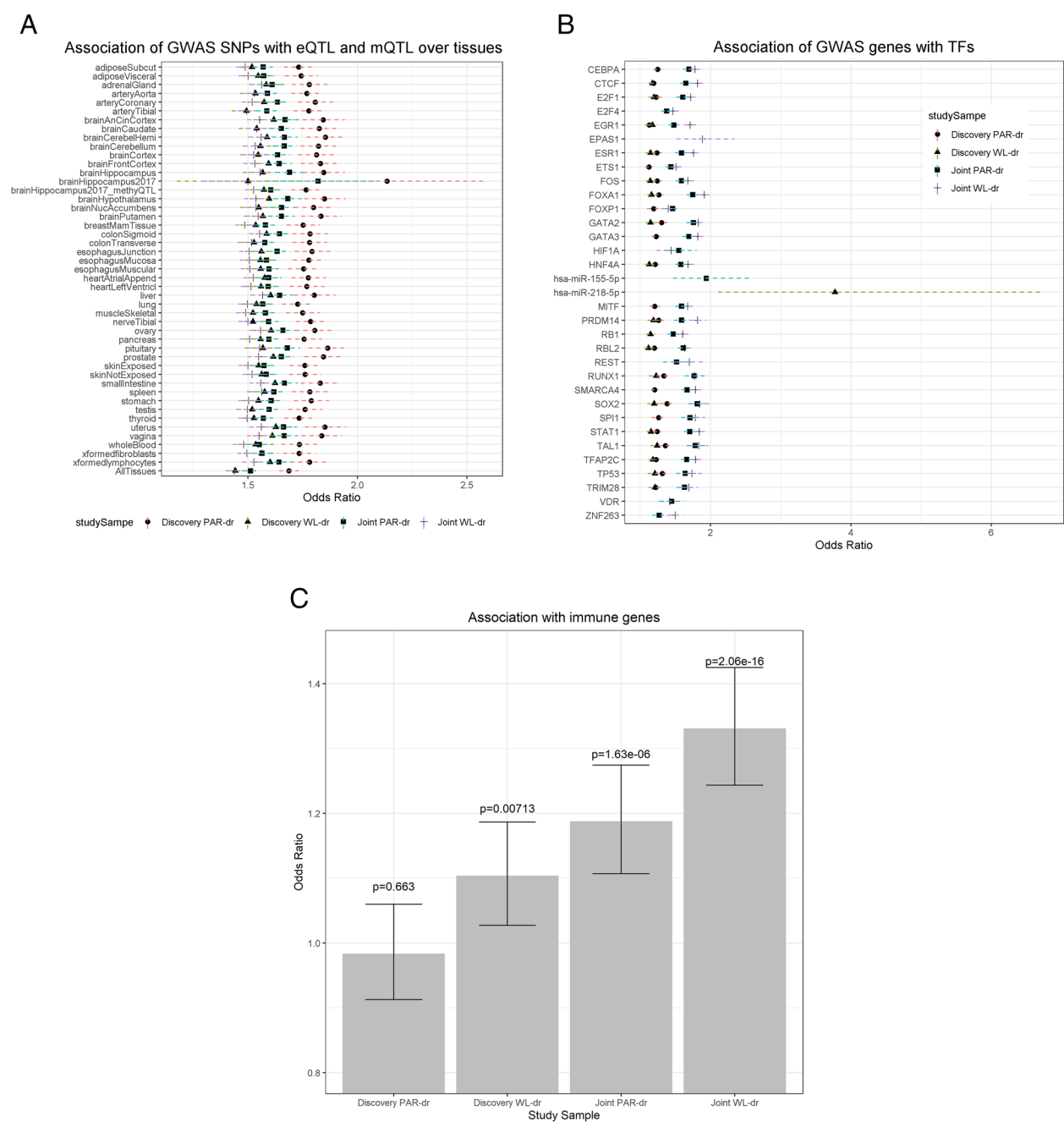
[26], GDS520 [27]) gene expression studies of cognitive traits were selected from the Gene Expression Omnibus [28]; descriptions of each study are provided in the Supplemental Text. The rodent studies used homologs (identified through the NCBI HomoloGene tool [29]) in hippocampal tissue.

For both human and rodent studies, the gene expression values were normalized by quantile normalization using the R package preprocessCore [30]. We used linear models from the R package limma [31] to analyze the DE of each gene across cognitive statuses; an  $F$  statistic and  $p$ -value were generated after moderating the test standard errors by empirical Bayesian modeling. The gene-set DE test was based on designed contrast tests for comparing expression levels by cognition status and utilized the mean-rank method [32] implemented in limma.  $P$ -values were obtained through permutation tests, with significance defined as  $p$ -values ≤ 0.05.

## Results

### Multi-omics function analysis of single-variant associations

Cross-cohort single-variant memory associations were related to markers of regulation (eQTLs and meQTLs) as shown in Fig. 3A and Table S3. Regardless of the tissue tested, variants highly associated with VDM phenotypes had significantly greater odds of being in high LD with eQTLs and meQTLs; the odds ratio (OR) estimates ranged from 1.43 ( $\beta = 0.36$ ) to 2.14 ( $\beta = 0.76$ ). Each power of 10 increase in association (e.g.  $p$ -value decreasing from 1E−05 to 1E−06) corresponded to at least a 1.43 increase in the odds of being in high LD with an eQTL



**Fig. 3** **A** Relationship between the strength ( $-\log_{10}(p\text{-values})$ ) of verbal declarative memory single-variant associations and being in high linkage disequilibrium ( $r^2 > 0.80$ ) with eQTLs and meQTLs across tissues. The shapes with dotted lines represent the odds ratios of being in linkage disequilibrium with an eQTL or meQTL given a one unit increase in SNP-memory association significance ( $p$ -value decreasing by a power of 10). The length of dotted line denotes the 95% confidence intervals of the odds ratios. **B** Relationship between the strength ( $-\log_{10}(U\text{-score})$ ) of verbal declarative memory gene associations and regulation by known transcription factors and microRNAs. The shapes with dotted lines represent the odds ratios of being regulated by a transcription factor or microRNA given a one unit increase in gene association significance ( $U$ -score decreasing by a power of 10). The length of dotted line denotes the 95% confidence intervals of the odds ratios. **C** Relationship between the strength ( $-\log_{10}(U\text{-score})$ ) of verbal declarative memory gene associations and annotation as an immunity gene. The heights of the bars represent the odds ratios of being an annotated immune gene given a one unit increase in gene association significance ( $U$ -score decreasing by a power of 10). The bars denote the 95% confidence intervals of the odds ratios

or meQTL. The OR of PAR-dr single-variant associations exceeded those of WL-dr. The largest OR (2.14; 95% CI [1.76, 2.60]) corresponded to the effect of PAR-dr single-variant associations on eQTLs from hippocampal biopsies in discovery cohorts, with an OR of 1.82 (95% CI [1.59, 2.09]) in the discovery and replication cohorts combined.

Multi-omics function analysis of gene associations

VDM gene associations were implicated in gene expression, regulation by TF/microRNA, and immunity function. As shown in Table 1, genes more strongly associated with WL-dr exhibited decreased odds of being highly expressed (RPKM in the top 5%) in four brain tissues, namely the anterior cingulate cortex, caudate, hippocampus, and pituitary gland. For the former three tissues, the negative association is significant in the discovery cohorts. For the pituitary gland, the negative association is significant in the joint discovery and replication cohorts. We failed to detect any significant relationship between PAR-dr gene associations and expression.

Genes more strongly associated with VDM had significantly increased odds of being regulated by thirty-one TFs and two microRNAs (Fig. 3B and Table S4); thirty TFs were implicated for both PAR-dr and WL-dr using all cohorts. Their ORs ranged from 1.12 (95% CI [1.06, 1.18]) for *RBL2* to 3.78 (95% CI [2.10, 6.81]) for *hsa-miR-218-5p* (95% CI [2.10, 6.81]), both of which were observed in the discovery WL-dr. The ORs were larger analyzing all cohorts than discovery cohorts alone with one exception, WL-dr gene associations and *hsa-miR-218-5p*. Similarly, genes with stronger VDM associations had greater odds of being immune genes. Both PAR-dr (OR=1.19, 95% CI [1.11, 1.27]) and WL-dr (OR=1.33, 95% CI [1.24, 1.43]) gene associations were significantly related to immune gene functions when analyzing all cohorts (Fig. 3C).

Pathway enrichment analysis  
Meta-analysis of pathway enrichment over cohorts  
(Approach 1)

Six pathways, namely the set of genes upregulated with PSMD4 and the KEGG pathways of type I diabetes mellitus, graft-versus-host disease, allograft rejection, antigen processing and presentation, and viral myocarditis, were significantly ( $p$ -values:  $RE\_p \leq 4.86E-06$  or  $Bin\_p_A \leq 0.05$ ) associated with PAR-dr and WL-dr in discovery cohorts (Table 2). The enrichment effect sizes (12~28%) were similar for PAR-dr and WL-dr in discovery cohorts; forest plots of the enrichment effects for each pathway and trait are displayed in Figure S3.

The type I diabetes pathway association with WL-dr was replicated ( $p$ -value:  $Bin\_p = 0.006$ ) in independent cohorts. The PSMD4 targets exhibited marginal ( $p$ -value:  $RE\_p = 0.046$ ) replication for WL-dr. The meta-analytic effect sizes were small in the replication cohorts (−1~3%). All six pathways met significance criteria ( $p$ -value:  $RE\_p \leq 4.86E-06$  or  $Bin\_p_A \leq 0.05$ ) for both delayed recall assessments in the joint meta-analysis of discovery and replication cohorts. However, the  $p$ -values and effect sizes (ranged from 6 to 19%) were generally attenuated compared to the values from the discovery cohorts alone.

Pathway enrichment of significant genes over cohorts  
(Approach 2)

The meta-analysis of gene associations across discovery cohorts yielded 69 and 173 genes significantly associated with PAR-dr and WL-dr, respectively (Table S5,  $p$ -value:  $Gene\_p \leq 1E-06$ ); 66 genes were associated with both traits. Pathway enrichment analysis of significant genes identified the same six significant pathways ( $p$ -value:  $path\_p_A \leq 0.05$ ; Table 3) as the meta-analysis of cohort-specific pathway enrichments (approach 1). Pathway effect sizes for PAR-dr (7~16%) were half those for

Table 1 Significant tissue-specific correlation between GWAS associations and gene expression

Tissue	Summary gene result	WL-dr			PAR-dr		
		OR	95% CI	Permutation-adjusted $p$ -value	OR	95% CI	Permutation adjusted $p$ -value
Brain anterior cingulate cortex	Discovery	0.75	[0.64,0.87]	<b>0.005</b>	1.00	[0.88,1.15]	1.000
	Joint	0.86	[0.75,0.99]	0.399	0.96	[0.84,1.10]	1.00
Brain caudate	Discovery	0.78	[0.68,0.91]	<b>0.026</b>	1.03	[0.90,1.17]	1.000
	Joint	0.91	[0.79,1.04]	0.897	0.98	[0.86,1.12]	1.000
Brain hippocampus	Discovery	0.80	[0.69,0.92]	<b>0.046</b>	0.99	[0.87,1.13]	1.000
	Joint	0.89	[0.77,1.02]	0.725	0.93	[0.81,1.07]	0.997
Pituitary	Discovery	0.90	[0.78,1.03]	0.838	0.93	[0.81,1.07]	0.998
	Joint	0.78	[0.68,0.90]	<b>0.021</b>	0.85	[0.74,0.98]	0.350

**Table 2** Significant pathways identified by Approach 1 (meta-analysis of cohort-specific pathway enrichment effects and tests)

Gene set (size)	Meta discovery				WL-dr			
	PAR-dr							
	Effect	SE	RE_p	Bin_p <sub>A</sub>	Effect	SE	RE_p	Bin_p <sub>A</sub>
Type 1 diabetes (44)	20.94%	1.22%	3.32E−66	5.79E−03	22.67%	1.72%	7.59E−40	2.32E−10
PSMD4 targets (73)	14.33%	1.11%	7.24E−38	5.79E−03	15.40%	1.20%	8.44E−38	1.83E−07
Graft-versus-host disease (42)	24.26%	1.69%	1.60E−46	1.54E−05	25.60%	2.07%	4.17E−35	2.32E−10
Allograft rejection (38)	25.57%	1.34%	3.98E−81	4.01E−07	27.73%	1.99%	3.28E−44	7.42E−09
Antigen processing and presentation (89)	12.53%	0.88%	1.52E−46	5.79E−03	13.05%	1.31%	1.70E−23	3.52E−06
Viral myocarditis (73)	11.76%	0.94%	6.97E−36	>0.05	13.51%	1.04%	1.05E−38	5.28E−05
Gene set	Meta replication				WL-dr			
	PAR-dr							
	Effect	SE	RE_p	Bin_p <sup>‡</sup>	Effect	SE	RE_p	Bin_p <sup>‡</sup>
Type 1 diabetes	0.63%	1.47%	0.67	1.00	3.44%	1.83%	0.060	0.006
PSMD4 targets	0.39%	1.01%	0.70	1.00	1.88%	0.94%	0.046	0.34
Graft-versus-host disease	−0.48%	1.28%	0.71	1.00	2.64%	1.36%	0.053	0.34
Allograft rejection	0.28%	1.34%	0.83	1.00	1.42%	1.76%	0.42	0.34
Antigen processing and presentation	−1.25%	0.86%	0.15	1.00	2.19%	1.56%	0.16	0.34
Viral myocarditis	0.40%	1.13%	0.72	0.34	1.52%	1.44%	0.29	0.057
Gene set	Meta joint				WL-dr			
	PAR-dr							
	Effect	SE	RE_p	Bin_p <sub>A</sub>	Effect	SE	RE_p	Bin_p <sub>A</sub>
Type 1 diabetes	10.78%	2.77%	1.01E−04	0.04	15.98%	2.32%	6.23E−12	6.11E−08
PSMD4 targets	7.36%	1.94%	1.50E−04	0.04	10.69%	1.60%	2.70E−11	9.71E−06
Graft-versus-host disease	11.88%	3.35%	3.97E−04	8.57E−04	17.60%	2.72%	1.04E−10	6.11E−08
Allograft rejection	12.93%	3.38%	1.33E−04	8.09E−05	18.57%	3.02%	7.73E−10	8.39E−07
Antigen processing and presentation	5.64%	1.86%	2.43E−07	0.04	9.27%	1.48%	3.98E−10	9.40E−05
Viral myocarditis	6.08%	1.61%	1.54E−04	0.04	9.34%	1.47%	2.05E−10	7.53E−04

This table shows the significant pathways identified through the meta-analysis of cohort-specific pathway enrichment effects (random effects model) or tests (binomial tests). Results are shown for the meta-analysis of discovery cohorts alone, replication cohorts alone, and the discovery and replication cohorts combined. Gene set (Size) is the name of the gene set and the number of genes it included. Effect indicates the increased probability of a pathway component gene to have a significant measure of association ( $U\text{-score} \leq 0.05$ ) compared to a random gene. SE is the standard error of the effect; PAR-dr and WL-dr are the meta-analysis results for the paragraph and word list delayed recall assessments, respectively. RE\_p is the  $p$ -value from the meta-analysis using the random effects model. Bin\_p<sub>A</sub> is the meta-analysis permutation-adjusted  $p$ -value from the binomial test. Bin\_p is the meta-analysis exact  $p$ -value from the binomial test that does not adjust for multiple testing. <sup>‡</sup>Only six pathways were tested for replication, thus the exact path  $p$ -values (Bin\_p) were used instead of the permutation-adjusted adjusted  $p$ -values (Bin\_p<sub>A</sub>) when meta-analyzing pathway results across cohorts

**Table 3** Significant pathways identified by Approach 2 (candidate gene enrichment analyses of summary gene associations from discovery cohorts)

Gene set	PAR-dr				WL-dr		
	Size	Effect	SE	Path_p <sub>E</sub> <sup>‡</sup>	Effect	SE	Path_p <sub>E</sub> <sup>‡</sup>
Type 1 diabetes	44	13.43%	0.69%	4.98E−10	26.74%	1.10%	6.72E−18
PSMD4 targets	73	9.38%	0.53%	2.13E−10	18.64%	0.85%	2.56E−18
Graft-versus-host disease	42	14.08%	0.70%	3.71E−10	28.04%	1.13%	3.56E−18
Allograft rejection	38	15.58%	0.75%	1.97E−10	31.04%	1.18%	8.87E−19
Antigen processing and presentation	89	6.53%	0.49%	3.80E−08	12.97%	0.77%	6.14E−14
Viral myocarditis	73	8.01%	0.54%	1.14E−08	15.90%	0.85%	5.13E−15

This table displays pathways that were significantly enriched for memory-associated genes using Approach 2 and USGSA on the discovery cohorts. Candidate gene set enrichment analyses were run on 69 and 173 memory-associated genes for PAR-dr and WL-dr, respectively. Please refer to the legend of Table 1 for descriptions of Gene Set, Size, PAR-dr, and WL-dr. Effect is the increased probability of a memory-associated gene to be from the specified pathway versus a random gene. SE is the standard error of the effect estimate; Path\_p<sub>E</sub><sup>‡</sup> is the exact pathway  $p$ -value based on the hypergeometric distribution; Path\_p<sub>A</sub>, the adjusted pathway  $p$ -value based on 10,000 permutations, took values “<0.001” for all displayed gene sets and outcomes and was omitted from the table



WL-dr (13~31%). These six pathways harbored fifteen genes significantly associated with VDM in discovery cohorts (Table 4 and S6); eight and fifteen genes were significantly associated with PAR-dr and WL-dr, respectively. There were 75–100% of discovery cohorts showing the significant PAR-dr genes and 60–93% supporting the significant WL-dr genes ( $U$ -scores  $\leq 0.05$ ). All fifteen genes are members of the major histocompatibility complex (MHC), with eleven present in all six significant pathways. One gene, *HLA-DRA*, exhibited marginal evidence ( $p$ -value:  $\text{Gene\_}p = 0.006$ ) of replication for WL-dr with support from 38% of the replication cohorts.

DE analysis of significant pathway component genes

Fifteen significant genes from memory-associated pathways were differentially expressed by cognitive status in human brain tissue (Table 5); expression differed by Braak stage in astrocytes ( $p = 0.006$ ) for the first data set (GDS4135) and by human immunodeficiency virus (HIV) cognitive impairment status (impaired infected versus uninfected controls) in brain tissues ( $p = 3.28\text{E} - 08$ ) for the second data set (GDS4231). In basal ganglia of data set GDS4358, memory-associated pathway genes were differentially expressed across control, HIV-1 infected only (HIV-only), HIV-1 infected with substantial neurocognitive impairment (HIV-NCI), and HIV with neurocognitive impairment and HIV encephalitis

(HIV-NCI-HIVE) groups (Trend I test;  $p = 3.33\text{E} - 05$ ), as well as across the latter three groups after excluding control (Trend II test;  $p = 8.83\text{E} - 05$ ). DE was found in the white matter tissue samples when controls were included (Trend I test;  $p = 0.03$ ) but not when omitted (Trend II test;  $p = 0.50$ ). No DE was found in the frontal cortex.

We also examined the DE of homologous genes in three rodent studies of hippocampal tissue. Twelve and six homologous genes were available in the house mouse and Norway rat, respectively (Table S7). Mean-rank tests confirmed DE of these genes in the hippocampus of house mice with age-related spatial memory deficits compared to young mice ( $p = 0.03$ ) for the data set GDS2082, Norway rats with impaired versus normal cognition ( $p = 0.016$ ) for the data set GDS2639, and Norway rats with age-dependent cognitive decline at 4, 14, and 24 months for the data set GDS520 ( $p = 0.015$ ).

Discussion

Debette et al. conducted meta-analyses of PAR-dr and WL-dr GWAS data across cohorts participating in the CHARGE consortium. They identified a significant VDM-associated variant located near genes involved in the immune response and found a correlation between memory risk variants and gene expression in human hippocampal cells. They also conducted

Table 4 Significant component genes from verbal declarative memory-associated pathways

	Gene	GeneID	Meta discovery		Meta replication		Meta joint	
			PAR-dr (M/Gene_p)	WL-dr (M/Gene_p)	PAR-dr (M/Gene_p)	WL-dr (M/Gene_p)	PAR-dr (M/Gene_p)	WL-dr (M/Gene_p)
MHC Class I	<i>HLA-A</i> <sup>a</sup>	3105	5/1.54E−05	10/2.32E−10	1/0.34	1/0.34	6/8.09E−05	11/3.76E−09
	<i>HLA-B</i> <sup>a</sup>	3106	6/4.01E−07	13/1.17E−15	1/0.34	0/1.00	7/5.98E−6	13/8.69E−12
	<i>HLA-C</i> <sup>a</sup>	3107	8/3.91E−11	14/8.73E−18	0/1.00	0/1.00	8/3.50E−7	14/3.25E−13
	<i>HLA-E</i> <sup>a</sup>	3133	3/0.006	9/7.42E−09	0/1.00	0/1.00	3/0.043	9/8.39E−07
	<i>HLA-F</i> <sup>a</sup>	3134	3/0.006	9/7.42E−09	2/0.057	0/1.00	5/8.57E−04	9/8.39E−07
	<i>HLA-G</i> <sup>a</sup>	3135	6/4.01E−07	13/1.17E−15	0/1.00	0/1.00	6/8.09E−05	13/8.69E−12
	<i>HLA-H</i>	3136	3/0.006	12/9.64E−14	2/0.057	1/0.34	5/8.57E−04	13/8.69E−12
MHC Class II	<i>HLA-DMA</i> <sup>a</sup>	3108	6/4.01E−07	11/5.53E−12	0/1.00	1/0.34	6/8.09E−05	12/1.96E−10
	<i>HLA-DMB</i> <sup>a</sup>	3109	7/5.98E−09	12/9.64E−14	1/0.34	0/1.00	8/3.50E−7	12/1.96E−10
	<i>HLA-DPA1</i> <sup>a</sup>	3113	5/1.54E−05	11/5.53E−12	0/1.00	1/0.34	5/8.57E−04	12/1.96E−10
	<i>HLA-DPB1</i> <sup>a</sup>	3115	5/1.54E−05	11/5.53E−12	0/1.00	0/1.00	5/8.57E−04	11/3.76E−09
	<i>HLA-DPB2</i>	3116	7/5.98E−09	11/5.53E−12	0/1.00	0/1.00	7/5.98E−06	11/3.76E−09
	<i>HLA-DQA2</i>	3118	7/5.98E−09	10/2.32E−10	0/1.00	0/1.00	7/5.98E−06	10/6.11E−08
	<i>HLA-DQB2</i>	3120	7/5.98E−09	11/5.53E−12	0/1.00	0/1.00	7/5.98E−06	11/3.76E−09
	<i>HLA-DRA</i> <sup>a</sup>	3122	5/1.54E−05	13/1.17E−15	0/1.00	3/0.006	5/8.57E−04	16/2.67E−16

This table shows significant gene-based tests results for the component genes from the six memory-associated pathways. The binomial test was used to meta-analyze the gene associations ( $U$ -scores) across cohorts. Gene is the gene name or symbol; GeneID is the NCBI gene identifier; PAR-dr and WL-dr represent the results of the gene-based tests for the paragraph and word list delayed recall traits, respectively; M is the number of cohorts in which the gene had a  $U$ -score  $\leq 0.05$ ; Gene\_p is the meta-analysis  $p$ -value (significant for values  $\leq 1\text{E} - 06$ )

<sup>a</sup> Indicates a gene that is a component of all six memory-associated pathways

**Table 5** Differential expression analysis of significant component genes from verbal declarative memory-associated pathways

GEO_ID	PUBMED_ID	Organism	Tissue	N_genes	Contrast	p-value
GDS4135	21705112	Human	Astrocytes	13	Braak stage I-II, III-IV, V-VI	0.006
GDS4231	21909266	Human	Brain tissues	13	HIV cognitive impairment vs uninfected control	3.28E−08
GDS4358	23049970	Human	Basal ganglia	13	Trend I	3.33E−05
					Trend II	8.83E−05
			Frontal cortex	13	Trend I	0.74
					Trend II	0.99
			White matter	13	Trend I	0.03
					Trend II	0.50
GDS2082	15169854	House mouse	Hippocampus	12	15-month-old mice with age-related cognitive deficit vs 2-month-old normal mice	0.03
GDS2639	17376971	Norway rat	Hippocampus	6	Impaired vs. unimpaired cognition	0.016
GDS520	12736351	Norway rat	Hippocampus	5	Age 4, 14, and 24 months	0.015

This table contains results for the differential expression analysis (mean-rank test) of genes in memory-associated pathways by cognitive status in human and rodent samples. GEO\_ID is the curated Gene Expression Omnibus data set identifier; PUBMED\_ID is the PUBMED publication identifier. N\_genes is the number of component genes that have expression measured in the given dataset (refer to Supplementary Tables S4 and S5 for the gene lists). Contrast indicates the cognitive function groups across which we contrasted the differential expression of component genes. p-value is the p-value from the gene-set differential analysis. The trend I test assessed differential expression across control, HIV-only, HIV-NCI, and HIV-NCI-HIVE statuses. The trend II test compared expression across the HIV-only, HIV-NCI, and HIV-NCI-HIVE statuses

pathway analyses focused on molecules with physical contact [2]. In this study, we expanded beyond the confines of prior research and adopted a more comprehensive approach to investigate the potential functions of VDM-associated variants. Our investigation demonstrated that VDM-associated variants are in high linkage disequilibrium with eQTLs across all 44 tissues and meQTLs in the hippocampus. Our analyses indicated that VDM-associated genes have reduced odds of being highly expressed in four specific brain tissues. Furthermore, VDM-associated genes appeared to be regulated by thirty-one TFs and two microRNAs, while also being implicated in immune function. Our analyses highlighted six pathways, including one relevant to type I diabetes, significantly correlated with both PAR-dr and WL-dr. Remarkably, these pathways encompassed fifteen MHC genes intricately tied to VDM performance. These MHC genes exhibited differential expression by cognitive status in brain tissues.

This investigation showcased the ability of multi-omics and pathway analyses to attribute function to GWAS associations. Our findings implicate gene expression regulation and immunity as functions underlying VDM genetic associations in older non-demented individuals of European descent. The multi-omics analyses showed that PAR-dr and WL-dr single-variant associations exhibited LD with eQTLs in every tissue and meQTLs in the hippocampus, bolstering evidence that trait-associated variants are enriched in eQTLs [9, 33] and regions involved in expression regulation [34, 35]. The connection between VDM-associated variants and meQTLs in hippocampal tissue echoed the association

of Alzheimer’s neuropathology and disease with methylation changes in brain tissue (including the hippocampus) [36–38]. We observed a lack of tissue specificity in the eQTL analysis which is similar to other memory-related traits [39]. However, the strongest eQTL relationship was with PAR-dr genetic associations in the hippocampus, a brain region involved in the acquisition of new memories and verbal and narrative memory [40].

Stronger WL-dr gene associations were connected to expression downregulation in four brain tissues (the anterior cingulate cortex, caudate, hippocampus, and pituitary gland), while stronger PAR-dr and/or WL-dr gene associations implicated regulation by thirty-one TFs and two microRNAs and classification as immune genes. Sequence variation in TFs and their binding site clusters, as well as microRNA expression levels (specifically hsa-miR-218–1-5p), have been associated with AD [41–43]. Similarly, the increased odds of immune function ascribed to genetic associations are supported by previous studies of AD [41, 44].

While Dettie et al. utilized summarized statistics to pinpoint VDM-associated pathways through a network of molecules with physical interactions [2], our study took a different approach. We leveraged the Molecular Signatures Database (MSigDB) to broaden our pathway analysis to include 10,295 curated gene sets. In this endeavor, we gathered individual GWAS results from each cohort and examined VDM-associated pathways within their respective contexts. We conducted meta-analysis using the random-effect model to gauge pathway enrichment effects across cohorts and employed binomial meta-analysis to assess if VDM-associated pathways within cohorts

exhibited non-random trends. To validate our findings, replication cohorts were examined alongside the original discovery cohorts in this study.

The pathway enrichment analysis identified six VDM-associated pathways (type 1 diabetes, graft-versus-host disease, allograft rejection, antigen processing and presentation, viral myocarditis, and targets of PSMD4 regulation) which were all interrelated within the framework of immunity. Antigen presentation, the process by which MHC proteins bind and transport ingested antigens to the surface of antigen presenting cells where they can be recognized by T-cells [45, 46], is critically involved in the early stages of type 1 diabetes (during the autoimmune destruction of pancreatic beta cells [47]), graft-versus-host disease (when T-cells from a foreign donor graft attack antigens expressed by the recipient [48]), allograft rejection (when T-cells from the recipient directly or indirectly attack antigens from transplanted tissue from a genetically non-identical human donor [49]), viral myocarditis (when viral antigens are presented to T-cells following an infection of cardiac myocytes [50]), and the induction of inflammatory cytokine production (several cytokines are members of the PSMD4 targets pathway [51, 52]).

The type I diabetes pathway association was replicated in independent cohorts and is biologically plausible. Insulin deficiency may reduce VDM performance through altered cerebral glucose metabolism, neurotransmitter expression/activity, neurotrophins, long-term potentiation, or inflammatory responses [47]. Increasing plasma insulin levels intravenously while preserving euglycemia aids VDM (story and word list recall) in both healthy adults and AD patients [47]. Similarly, acute and chronic intranasal insulin administration improved verbal memory in AD patients and healthy young adults, respectively [47]. In general, adults with type I diabetes perform worse on memory tests than non-diabetics [53]. AD patients (who often exhibit reduced VDM performance) have decreased hippocampal glucose consumption, hippocampal insulin receptor mRNA, and brain insulin receptor protein levels compared to age-matched controls [54, 55]. Gene expression studies also link diabetes with the AD pathway [56].

Our pathway enrichment findings may reflect a single pathway or MHC gene associations. The six VDM-associated pathways shared eleven MHC genes and collectively harbored fifteen MHC genes exhibiting differential expression by cognitive status in human and rodent brain tissues. MHC I proteins may be required for hippocampus-dependent memory [57]. An MHC II gene (*HLA-DRB1*) was associated with delayed verbal recall performance in older non-demented individuals [58] and

AD [59], while hippocampal MHC II protein levels were inversely associated with mini-mental state examination scores [60]. Several MHC genes associated with VDM in this investigation (including the marginally replicated *HLA-DRA*) have been associated with AD (*HLA-A*, *HLA-B*, *HLA-DRA* [61–63]) or showed increased hippocampal (*HLA-DMA*, *HLA-DMB*, *HLA-DPA1*, *HLA-DRA* [60]) or pre-frontal cortex (*HLA-A*, *HLA-C*, *HLA-E*, *HLA-F*, *HLA-G*, *HLA-DPB1* [60]) expression in mild AD dementia cases compared to non-demented controls. MHC genes may influence memory through their effects on synaptic plasticity, development, morphology, and function [57, 64–66].

This investigation had a few limitations, including the lack of stringent replication for the multi-omics and pathway analyses. The replication cohorts were mainly restricted to women with some college education and had different PAR-dr and WL-dr assessments compared to the discovery cohorts. Each cohort-specific GWAS used HapMap II CEU-imputed data, which has a sparser gene coverage than 1000 Genomes-imputed or whole genome/exome sequence data. Therefore, the findings may be less accurate due to the omission of rare genetic variation of large effect. The original cohort-specific findings assumed additive genetic effects, thus we possibly missed genes and pathways containing dominant or recessive variant effects.

In this research, we investigated the relationship between VDM-associated variants and eQTLs and meQTLs. Specifically, we leveraged logistic regression to evaluate the linear association between the negative logarithm of *p*-values and the logarithm of odds that variants are in high LD with eQTLs and meQTLs. However, one limitation is that our research cannot definitively establish whether the same variant is causally linked to VDM and the regulation of eQTLs and meQTLs. Therefore, it is worthwhile to explore the identification of VDM variants that may be responsible for both the GWAS signals and regulatory effects by employing techniques such as colocalization and fine-mapping approaches [67, 68]. Additionally, we selected a threshold of  $r^2=0.8$  to determine if a genetic variant is in high LD with eQTLs and meQTLs. However, selection using a different threshold may impact the findings, thus incorporating more sophisticated methods such as LD scoring may enhance the robustness of our tests.

Our study may also be hindered by different gene association measures, selection of gene boundaries for SNP mapping, the incompleteness of omics databases, and annotation biases [5]. Lastly, this investigation included participants of European ancestry, thus findings may not generalize to other racial or ethnic groups.

## Conclusions

In conclusion, our results add to the mounting evidence implicating expression regulation, immunity, and insulin deficiency in memory impairment. Future studies should attempt to dissect the molecular mechanisms underlying these relationships, so treatments can be developed to combat the increasing burden of cognitive decline and AD on society.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13195-023-01376-6>.

**Additional file 1: Supplemental Text. Table S1.** Sample size and the number of SNPs in the paragraph delayed recall GWAS from each discovery and replication cohort. **Table S2.** Sample size and the number of SNPs in the word list delayed recall GWAS from each discovery and replication cohort. **Table S3.** Tissue-specific relationships between delayed recall test (PAR-dr and WL-dr) summary SNP associations and eQTLs and meQTLs. **Table S4.** Relationship Between Delayed Recall Summary Gene Associations and Transcription Factor Genes. **Table S5.** Significant Genes Associated with Paragraph Delayed Recall (PAR-dr) and Word List Delayed Recall (WL-dr). **Table S6.** Significant component genes in the six memory-associated pathways. **Table S7.** Homologous genes in memory-associated pathways for differential expression analysis. **Figure S1.** GWAS cohorts and microarray expression datasets. **Figure S2.** Design of the pathway analyses. **Figure S3.** Forest plots of significant pathway enrichment effects and *p*-values from discovery cohorts (Approach 1).

## Acknowledgements

**Aging Gene-Environment Susceptibility-Reykjavik Study:** The research has been funded by National Institute on Aging contract N01-AG-12100 with contributions from National Eye Institute, National Institute on Deafness and Other Communication Disorders, and National Heart, Lung and Blood Institute; the National Institute on Aging Intramural Research Program; Hjartavernd (the Icelandic Heart Association); and the Althingi (the Icelandic Parliament). **The Atherosclerosis Risk in Communities Study:** The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. HHSN2682017000011, HHSN2682017000021, HHSN2682017000031, HHSN2682017000041, and HHSN2682017000051. Funding was also supported by R01HL70825, R01HL087641, R01HL059367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors thank the staff and participants of the ARIC study for their important contributions. **The Cardiovascular Health Study:** This Cardiovascular Health Study research was supported by National Heart, Lung and Blood Institute contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, and N01HC85086 and National Heart, Lung and Blood Institute Grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided through R01AG023629, R01AG20098, and R01AG05133 NIA. A full list of principal Cardiovascular Health Study investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported, in part, by the National Center for Advancing Translational Sciences, Clinical and Translational Sciences Institute Grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center Grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Croatian Cohorts-Split and Korčula:** The CROATIA-Korčula and CROATIA-Split studies were funded by Grants from the Medical Research Council (United Kingdom), European Commission Framework 6 project EUROSPAN (Contract No. LSHG-CT-2006-018947), and Republic of Croatia Ministry of Science, Education and Sports research Grants to IR (108-1080315-0302), the Croatian National Center of Research Excellence in Personalized Healthcare (grant number KK.01.1.1.01.0010), and the Center of Competence in Molecular Diagnostics (KK.01.2.2.03.0006). We acknowledge the invaluable contributions of the recruitment teams in Korčula and Split, the administrative teams in Croatia and Edinburgh, and the people of Korčula and Split. The single-nucleotide polymorphism genotyping for the CROATIA-Korčula cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany. The single-nucleotide polymorphism genotyping for the CROATIA-Split cohort was performed by AROS Applied Biotechnology, Aarhus, Denmark.

**Erasmus Rucphen Family Study:** This study is financially supported by the Netherlands Organization for Scientific Research, the Internationale Stichting Alzheimer Onderzoek, the Hersenstichting Nederland, and the Centre for Medical Systems Biology (\*1 and 2\*) in the framework of the Netherlands Genomics Initiative. We thank the participants from the Genetic Research in Isolated Populations, Erasmus Rucphen Family, who made this work possible. **Framingham Heart Study:** From the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc. for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the Single Nucleotide Polymorphism Health Association Resource project. This study was also supported by Grants from the National Institute of Neurological Disorders and Stroke (NS17950) and the National Institute on Aging (AG08122, AG16495, AG033193, AG031287).

**Genetic Epidemiology Network of Arteriopathy:** Support for the Genetic Epidemiology Network of Arteriopathy was provided by the National Heart, Lung and Blood Institute (U01 HL054464, U01 HL054457, U01 HL054481, R01 HL071917, R01 HL087660, and R01 HL119443) and the National Institute of Neurological Disorders and Stroke (R01 NS041558) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (Stephen T. Turner, Mariza de Andrade, Julie Cunningham) and was made possible by the University of Texas Health Sciences Center (Erick Boerwinkle and Megan L. Grove-Gaona). We also thank the families that participated in the GENOA study.

**Helsinki Birth Cohort Study:** We thank all study participants as well as everybody involved in the Helsinki Birth Cohort Study. The Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Juho Vainio Foundation, Signe and Ane Gyllenberg Foundation, University of Helsinki, Ministry of Education, Jalmari ja Rauha Ahokas foundation, Emil Aaltonen Foundation, and Yrjö Jahnsson foundation.

**Lothian Birth Cohort 1921 and Lothian Birth Cohort 1936:** We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the Biotechnology and Biological Sciences Research Council, The Royal Society, and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Research Into Ageing (continues as part of Age United Kingdom The Disconnected Mind project). Genotyping of the cohorts was funded by the United Kingdom Biotechnology and Biological Sciences Research Council. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, and Medical Research Council is gratefully acknowledged. W.D.H., C.X., and D.C.M.L. are supported by a Career Development Award from the Medical Research Council (MRC) [MR/T030852/1] for the project titled "From genetic sequence to phenotypic consequence: genetic and environmental links between cognitive ability, socioeconomic position, and health". **Orkney Complex Disease Study:** The Orkney Complex Disease Study (ORCADES) was supported by the Chief Scientist Office of the Scottish Government



(CZB/4/276, CZB/4/710), a Royal Society URF to J.F.W., the MRC Human Genetics Unit quinquennial program "QTL in Health and Disease", Arthritis Research UK and the European Union framework program 6 EUROSPLAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Edinburgh Clinical Research Facility, University of Edinburgh. We would like to acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney.

*The Religious Order Study and Rush Memory and Aging Project:* The Religious Order Study and Rush Memory and Aging Project Study are supported in part by National Institute on Aging Grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, K08AG34290, and K25AG41906.

*The Rotterdam Study:* The generation and management of genome-wide association study genotype data for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr.175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, and Marjolein Peters for their help in creating the genome-wide association study database and Karol Estrada, Yuri Aulchenko, and Carolina Medina-Gomez for their support in creation and analysis of imputed data. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam; Netherlands Organization for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly; the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission (DG XII), and the Municipality of Rotterdam. The Rotterdam Scan Study is supported by the Netherlands Organization of Scientific Research project nrs. 918-46-615, 904-61-096, 904-61-133, 948-00-010, and 916-13-054 (ZonMW) and International Parkinson Fonds. Dr. Ikram was supported by a ZonMW Veni Grant: 916.13.054. The authors are grateful to the study participants, the staff from the Rotterdam Study, and the participating general practitioners and pharmacists.

*Study of Health in Pomerania:* Study of Health in Pomerania is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (Grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs, and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (Grant no. 03ZIK012) and a joint Grant from Siemens Healthineers, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH. This work was also funded by the German Research Foundation (DFG: GR 1912/5-1).

*The Tasmanian Study of Gait and Cognition* is supported by Project Grants from the National Health and Medical Research Council (NHMRC IDs 403000, 491109, 606543) and a Grant from the Wicking Dementia Education and Research Centre, Hobart. Velandai Srikanth is supported by a National Health and Medical Research Council/National Heart Foundation Career Development Fellowship (ID 606544). Matthew Brown is supported by a National Health and Medical Research Council Principal Research Fellowship.

*Nurses' Health Study:* This study was supported by research Grants CA87969, CA49449, HL34594, U01HG004399, DK058845, CA65725, CA67262, CA50385, 5U01CA098233, EY09611, EY015473, HG004728, HL35464, CA55075, CA134958, and DK070756 from the National Institutes of Health. The genotyping was partly supported by an unrestricted Grant from Merck Research Laboratories.

*Sydney Memory and Ageing Study:* We thank the participants and their informants for their time and generosity in contributing to this research. We also acknowledge the MAS research team: <https://cheba.unsw.edu.au/research-projects/sydney-memory-and-ageing-study>. DNA was extracted by Genetic Repositories Australia, an Enabling Facility, supported by National Health & Medical Research Council Grant 401184. Genome-wide genotyping was performed by the Ramaciotti Centre, University of New South Wales. The Sydney Memory and Ageing Study has been funded by three National Health & Medical Research Council (NHMRC) Program Grants (ID No. ID350833, ID568969, and APP1093083).

*Women's Genome Health Study:* The Women's Genome Health Study is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913), with funding for genotyping provided by Amgen.

#### Authors' contributions

H.M., J.S., L.L., F.J. and T.H.M. conceptualized and designed study. H.M. and L.L. analyzed the data, H.M., L.L. and J.S. drafted the manuscript. J.C.B., G.D., W.D.H., C.X., V.G., Q.Y., J.L., J.A.S., M.K., P.D.J., N.J.A., M.G., I.K., C.M., A.T., M.S., S.M., M.F., W.Z., C.L.S., O.P., K.R., D.C.L., G.H., M.C., K.A.M., B.G.W., T.Z., A.P., A.P., S.V.D.A., A.T., D.S.K., I.R., J.M.S., K.W., N.A.K., M.E.G., V.V., H.B., R.G., S.R.C., B.M.P., E.B., D.I.C., F.G., P.S.S., V.S., C.H., J.F.W., J.G.E., S.L.K., H.J.G., D.A.B., M.A.I., I.J.D., C.M.V.D., L.L., A.L.F., S.S., J.B., S.D. and T.H.M. contributed to the acquisition of the data. All authors read and approved the final manuscript.

#### Funding

This work was accomplished as a part of the CHARGE project (CHARGE consortium: omics discovery for CVD and aging phenotypes. NIH R01HL105756), and was funded by grant U01-HL096917 from the National Institutes of Health (NIH) / NIH Heart, Lung and Blood Institute <http://www.nhlbi.nih.gov/>. Hao Mei and Jeannette Simino are also partially funded by National Institute of General Medical Sciences grant 1P20GM144041 which established the Molecular Center of Health and Disease at the University of Mississippi Medical Center. Lianna Li is funded by National Science Foundation grant 2100805. Stéphanie Debetto is a recipient of a Chaire d'Excellence Grant from the French national research agency (Agence Nationale de la Recherche). Hans Grabe has received travel grants and speaker's honoraria from Fresenius Medical Care, Neuraxpharm, Servier, and Janssen Cilag as well as research funding from Fresenius Medical Care.

#### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on request.

#### Declarations

##### Ethics approval and consent to participate

This study utilized summary results specific to each cohort from a previously published genome-wide association study (GWAS) on verbal declarative memory. Each cohort received approval from the relevant institutional review boards. All participants provided written informed consent for study participation, cognitive testing, and use of their genetic information for research. In this study, we rely on publicly accessible, extensive datasets rather than individual-level data. As a result, ethical approval was not pursued.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>Department of Data Science, John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, MS, USA. <sup>2</sup>Gertrude C. Ford Memory Impairment and Neurodegenerative Dementia (MIND) Center, University of Mississippi Medical Center, Jackson, MS, USA. <sup>3</sup>Department of Biology, Tougaloo College, Jackson, MS, USA. <sup>4</sup>Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China. <sup>5</sup>Department of Medicine, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. <sup>6</sup>Department of Psychology, Lothian Birth Cohorts Group, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. <sup>7</sup>Icelandic Heart Association, Kopavogur, Iceland. <sup>8</sup>Faculty of Medicine, University of Iceland, Reykjavik, Iceland. <sup>9</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. <sup>10</sup>The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA. <sup>11</sup>Turku Institute for Advanced Research, University of Turku, Turku, Finland. <sup>12</sup>Department of Psychology and Logopedics, Faculty of Medicine,



University of Helsinki, Helsinki, Finland. <sup>13</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. <sup>14</sup>Work completed while at The University of Edinburgh, Edinburgh, UK. <sup>15</sup>Taub Institute for Research On Alzheimer's Disease and the Aging Brain, Columbia Irving University Medical Center, New York, NY, USA. <sup>16</sup>Center for Translational and Computational Neuro-Immunology, Columbia University Medical Center, New York, NY, USA. <sup>17</sup>Department of Neurology, Columbia University Medical Center, New York, NY, USA. <sup>18</sup>Mathematics and Statistics, Curtin University, Perth, Australia. <sup>19</sup>Department of Epidemiology, Erasmus Medical Center University Medical Center, Rotterdam, The Netherlands. <sup>20</sup>School of Medicine, University of Split, Split, Croatia. <sup>21</sup>Department of Geriatric Medicine, Frankston Hospital, Peninsula Health, Melbourne, Australia. <sup>22</sup>Peninsula Clinical School, Central Clinical School, Monash University, Melbourne, Australia. <sup>23</sup>Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. <sup>24</sup>Inserm, Bordeaux Population Health Research Center, Team VINTAGE, UMR 1219, University of Bordeaux, Bordeaux, France. <sup>25</sup>The Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA. <sup>26</sup>Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA. <sup>27</sup>The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. <sup>28</sup>Algebra University College, Ilica 242, Zagreb, Croatia. <sup>29</sup>Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. <sup>30</sup>Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia. <sup>31</sup>Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, Australia. <sup>32</sup>Neuroscience Research Australia, Sydney, Australia. <sup>33</sup>Department of Medicine, Division of Geriatrics, School of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. <sup>34</sup>Department of Medicine, Department of Neurology and Department of Psychiatry, Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. <sup>35</sup>The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>36</sup>Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. <sup>37</sup>Department of Neurology, Mayo Clinic, Rochester, MN, USA. <sup>38</sup>Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK. <sup>39</sup>Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK. <sup>40</sup>German Center for Neurodegenerative Diseases (DZNE), Site Rostock/ Greifswald, Rostock, Germany. <sup>41</sup>Department of Medicine, School of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. <sup>42</sup>Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. <sup>43</sup>Dementia Centre for Research Collaboration, University of New South Wales, Sydney, NSW, Australia. <sup>44</sup>Stroke, Cognition, and Neuroepidemiology (SCAN) Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA. <sup>45</sup>Department of Epidemiology, University of Washington, Seattle, WA, USA. <sup>46</sup>Department of Health Services, University of Washington, Seattle, WA, USA. <sup>47</sup>Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. <sup>48</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. <sup>49</sup>Harvard Medical School, Boston, MA, USA. <sup>50</sup>Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. <sup>51</sup>Department of Medicine, Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>52</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA. <sup>53</sup>Neuropsychiatric Institute, Prince of Wales Hospital, Sydney, Australia. <sup>54</sup>Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>55</sup>Department of Public Health Solutions, Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland. <sup>56</sup>Folkhälsan Research Centre, Helsinki, Finland. <sup>57</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. <sup>58</sup>Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA. <sup>59</sup>Nuffield Department of Population Health, Medical Sciences Division, University of Oxford, Oxford, UK. <sup>60</sup>Laboratory of Epidemiology and Population Sciences, National Institute On Aging, Bethesda, MD, USA. <sup>61</sup>Department of Family Medicine, University of Washington, Seattle, WA, USA. <sup>62</sup>Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA. <sup>63</sup>Department of Neurology, Boston University School of Medicine, Boston, MA, USA. <sup>64</sup>Department of Neurology, CHU de Bordeaux, Bordeaux, France. <sup>65</sup>Department of Neurology, University of Mississippi Medical Center, Jackson, MS, USA.

Received: 3 March 2023 Accepted: 26 December 2023  
Published online: 20 January 2024

## References

- Dubois B, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol*. 2007;6(8):734–46.
- DeBette S, et al. Genome-wide studies of verbal declarative memory in nondemented older people: the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium. *Biol Psychiatry*. 2015;77(8):749–63.
- Glahn DC, et al. Genetic basis of neurocognitive decline and reduced white-matter integrity in normal human brain aging. *Proc Natl Acad Sci U S A*. 2013;110(47):19006–11.
- Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18(1):83.
- Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet*. 2010;11(12):843–54.
- Papassotiropoulos A, de Quervain DJ. Genetics of human episodic memory: dealing with complexity. *Trends Cogn Sci*. 2011;15(9):381–7.
- Mooney MA, Wilmot B. Gene set analysis: A step-by-step guide. *Am J Med Genet B Neuropsychiatr Genet*. 2015;168(7):517–27.
- Jin L, et al. Pathway-based analysis tools for complex diseases: a review. *Genomics Proteomics Bioinformatics*. 2014;12(5):210–20.
- Akiyama M. Multi-omics study for interpretation of genome-wide association study. *J Hum Genet*. 2021;66(1):3–10.
- Danecek P, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
- Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. 1986;73(3):751–4.
- Mei H, et al. The uniform-score gene set analysis for identifying common pathways associated with different diabetes traits. *BMC Genomics*. 2015;16(1):336.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190–1.
- Gamazon ER, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet*. 2018;50(7):956–67.
- Schulz H, et al. Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nat Commun*. 2017;8(1):1511.
- Kent WJ, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996–1006.
- Lesurf R, et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res*. 2016;44(D1):D126–32.
- Breuer K, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res*. 2013;41(Database issue):D1228–33.
- Liberzon A. A description of the Molecular Signatures Database (MSigDB) Web site. *Methods Mol Biol*. 2014;1150:153–60.
- Mei H, et al. snpGeneSets: An R Package for Genome-Wide Study Annotation. *G3 (Bethesda)*. 2016;6(12):4087–95.
- Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1–48.
- Simpson JE, et al. Microarray analysis of the astrocyte transcriptome in the aging brain: relationship to Alzheimer's pathology and APOE genotype. *Neurobiol Aging*. 2011;32(10):1795–807.
- Borjabad A, et al. Significant effects of antiretroviral therapy on global gene expression in brain tissues of patients with HIV-1-associated neurocognitive disorders. *PLoS Pathog*. 2011;7(9):e1002213.
- Gelman BB, et al. The National NeuroAIDS Tissue Consortium brain gene array: two types of HIV-associated neurocognitive impairment. *PLoS ONE*. 2012;7(9):e46178.
- Verbitsky M, et al. Altered hippocampal transcript profile accompanies an age-related spatial memory deficit in mice. *Learn Mem*. 2004;11(3):253–60.
- Rowe WB, et al. Hippocampal expression analyses reveal selective association of immediate-early, neuroenergetic, and myelinogenic pathways with cognitive impairment in aged rats. *J Neurosci*. 2007;27(12):3098–110.

27. Blalock EM, et al. Gene microarrays in hippocampal aging: statistical profiling identifies novel processes correlated with cognitive impairment. *J Neurosci*. 2003;23(9):3807–19.
28. Barrett T, et al. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res*. 2011;39(Database issue):D1005–10.
29. NCBI HomoloGene. <https://www.ncbi.nlm.nih.gov/homologene>.
30. Bolstad BM, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185–93.
31. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:3.
32. Michaud J, et al. Integrative analysis of RUNX1 downstream pathways and target genes. *BMC Genomics*. 2008;9:363.
33. Nicolae DL, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6(4):e1000888.
34. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*. 2015;16(4):197–212.
35. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190–5.
36. De Jager PL, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci*. 2014;17(9):1156–63.
37. Chouliaras L, et al. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol Aging*. 2013;34(9):2091–9.
38. Semick SA, et al. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathol*. 2019;137(4):557–69.
39. Liu G, et al. rs4147929 variant minor allele increases ABCA7 gene expression and ABCA7 shows increased gene expression in Alzheimer's disease patients compared with controls. *Acta Neuropathol*. 2020;139(5):937–40.
40. Burgess N, Maguire EA, O'Keefe J. The human hippocampus and spatial and episodic memory. *Neuron*. 2002;35(4):625–41.
41. Bis JC, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Mol Psychiatry*. 2020;25(8):1859–1875.
42. Wu HZY, et al. Differential blood miRNA expression in brain amyloid imaging-defined Alzheimer's disease and controls. *Alzheimers Res Ther*. 2020;12(1):59.
43. Urfer-Buchwalder A, Urfer R. Identification of a Nuclear Respiratory Factor 1 Recognition Motif in the Apolipoprotein E Variant APOE4 linked to Alzheimer's Disease. *Sci Rep*. 2017;7:40668.
44. Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry*. 2015;77(1):43–51.
45. KEGG\_APP: Antigen processing and presentation - Homo sapiens (human). [http://www.genome.jp/dbget-bin/www\\_bget?hsa04612](http://www.genome.jp/dbget-bin/www_bget?hsa04612).
46. Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol*. 2013;31:443–73.
47. Watson GS, Craft S. Insulin resistance, inflammation, and cognition in Alzheimer's Disease: lessons for multiple sclerosis. *J Neurol Sci*. 2006;245(1–2):21–33.
48. KEGG\_GVHD: Graft-versus-host disease - Homo sapiens (human). [http://www.genome.jp/dbget-bin/www\\_bget?hsa05332](http://www.genome.jp/dbget-bin/www_bget?hsa05332).
49. KEGG\_AR: Allograft rejection - Homo sapiens (human). [http://www.genome.jp/dbget-bin/www\\_bget?hsa05330](http://www.genome.jp/dbget-bin/www_bget?hsa05330).
50. KEGG\_VM: Viral myocarditis - Homo sapiens. [http://www.genome.jp/dbget-bin/www\\_bget?hsa05416](http://www.genome.jp/dbget-bin/www_bget?hsa05416).
51. Gaurnier-Hausser A, et al. The novel angiogenic inhibitor, angiocidin, induces differentiation of monocytes to macrophages. *Cancer Res*. 2008;68(14):5905–14.
52. PSMD4\_TARGETS: Inflammatory cytokines, chemokines and their cognate receptors up-regulated in THP-1 cells (monocyte) after treatment with PSMD4 [GeneID=5710]. [http://software.broadinstitute.org/gsea/msigdb/cards/GAURNIER\\_PSMD4\\_TARGETS](http://software.broadinstitute.org/gsea/msigdb/cards/GAURNIER_PSMD4_TARGETS).
53. Tonoli C, et al. Type 1 diabetes-associated cognitive decline: a meta-analysis and update of the current literature. *J Diabetes*. 2014;6(6):499–513.
54. Kandimalla R, Thirumala V, Reddy PH. Is Alzheimer's disease a Type 3 Diabetes? A critical appraisal. *Biochim Biophys Acta*. 2017;1863(5):1078–89.
55. Steen E, et al. Impaired insulin and insulin-like growth factor expression and signaling mechanisms in Alzheimer's disease—is this type 3 diabetes? *J Alzheimers Dis*. 2005;7(1):63–80.
56. Mei H, et al. Tissue non-specific genes and pathways associated with diabetes: an expression meta-analysis. *Genes (Basel)*. 2017;8(1):44.
57. Nelson PA, et al. MHC class I immune proteins are critical for hippocampus-dependent memory and gate NMDAR-dependent hippocampal long-term depression. *Learn Mem*. 2013;20(9):505–17.
58. Payton A, et al. Influence and interactions of cathepsin D, HLA-DRB1 and APOE on cognitive abilities in an older non-demented population. *Genes Brain Behav*. 2006;5(Suppl 1):23–31.
59. Lambert JC, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452–8.
60. Parachikova A, et al. Inflammatory changes parallel the early stages of Alzheimer disease. *Neurobiol Aging*. 2007;28(12):1821–33.
61. Candore G, et al. Major histocompatibility complex and sporadic Alzheimer's disease: a critical reappraisal. *Exp Gerontol*. 2004;39(4):645–52.
62. Swaminathan S, et al. Analysis of copy number variation in Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *PLoS ONE*. 2012;7(12):e50640.
63. Yokoyama JS, et al. Association between genetic traits for immune-mediated diseases and Alzheimer disease. *JAMA Neurol*. 2016;73(6):691–7.
64. McGuffin P, Power RA. Schizophrenia as a human leukocyte antigen-associated disease revisited. *Am J Psychiatry*. 2013;170(8):821–3.
65. Huh GS, et al. Functional requirement for class I MHC in CNS development and plasticity. *Science*. 2000;290(5499):2155–9.
66. Goddard CA, Butts DA, Shatz CJ. Regulation of CNS synapses by neuronal MHC class I. *Proc Natl Acad Sci U S A*. 2007;104(16):6828–33.
67. Giambartolomei C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383.
68. Benner C, et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*. 2016;32(10):1493–501.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.