

## RESEARCH ARTICLE

# In vivo validation of late-onset Alzheimer's disease genetic risk factors

Michael Sasner<sup>1</sup> | Christoph Preuss<sup>1</sup> | Ravi S. Pandey<sup>2</sup> | Asli Uyar<sup>2</sup> |  
Dylan Garceau<sup>1</sup> | Kevin P. Kotredes<sup>1</sup> | Harriet Williams<sup>1</sup> | Adrian L. Oblak<sup>3</sup> |  
Peter Bor-Chian Lin<sup>3</sup> | Bridget Perkins<sup>3</sup> | Disha Soni<sup>3</sup> | Cindy Ingraham<sup>3</sup> |  
Audrey Lee-Gosselin<sup>3</sup> | Bruce T. Lamb<sup>3</sup> | Gareth R. Howell<sup>1</sup> | Gregory W. Carter<sup>1,2</sup>

<sup>1</sup>The Jackson Laboratory, Bar Harbor, Maine, USA

<sup>2</sup>The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA

<sup>3</sup>Stark Neurosciences Research Institute, School of Medicine, Indiana University, Indianapolis, Indiana, USA

**Correspondence**

Gregory W. Carter, The Jackson Laboratory, 600 Main St, Bar Harbor, ME 04609, USA.  
Email: [gregory.carter@jax.org](mailto:gregory.carter@jax.org)

**Funding information**

National Institute on Aging, Grant/Award Number: U54AG054345

**Abstract**

**INTRODUCTION:** Genome-wide association studies have identified over 70 genetic loci associated with late-onset Alzheimer's disease (LOAD), but few candidate polymorphisms have been functionally assessed for disease relevance and mechanism of action.

**METHODS:** Candidate genetic risk variants were informatically prioritized and individually engineered into a LOAD-sensitized mouse model that carries the AD risk variants *APOE*  $\epsilon 4/\epsilon 4$  and *Trem2*\*R47H. The potential disease relevance of each model was assessed by comparing brain transcriptomes measured with the Nanostring Mouse AD Panel at 4 and 12 months of age with human study cohorts.

**RESULTS:** We created new models for 11 coding and loss-of-function risk variants. Transcriptomic effects from multiple genetic variants recapitulated a variety of human gene expression patterns observed in LOAD study cohorts. Specific models matched to emerging molecular LOAD subtypes.

**DISCUSSION:** These results provide an initial functionalization of 11 candidate risk variants and identify potential preclinical models for testing targeted therapeutics.

**KEYWORDS**

*Abca7*, Alzheimer's disease, animal models, *APOE4*, *Plcg2*, *Mthfr*, preclinical, *Sor11*, transcriptomic analysis, *Trem2*

**Highlights**

- A novel approach to validate genetic risk factors for late-onset AD (LOAD) is presented.
- LOAD risk variants were knocked in to conserved mouse loci.
- Variant effects were assayed by transcriptional analysis.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

- Risk variants in *Abca7*, *Mthfr*, *Plcg2*, and *Sorl1* loci modeled molecular signatures of clinical disease.
- This approach should generate more translationally relevant animal models.

## 1 | BACKGROUND

Alzheimer's disease (AD) is the most common cause of dementia, with a growing clinical, financial, and social impact. An increasing body of evidence highlights the importance of genetic risk in AD.<sup>1–3</sup> While a small percentage of AD cases are linked to causative, familial mutations in the amyloid precursor protein (APP) processing pathway, the vast majority of cases are late-onset AD (LOAD), have heterogeneous symptoms and etiology, and are associated with polygenic risk from a combination of low-risk, relatively common variants.<sup>4–6</sup> Genome-wide association studies (GWAS) have identified numerous LOAD risk variants, but few have been experimentally validated, and physiological mechanisms have not been elucidated, even for the single strongest risk variant, the  $\epsilon 4$  allele of the APOE gene.<sup>4,7</sup> This is but one example<sup>8</sup> of the general problem of how to progress from the identification of genetic variants to the functional impact of variants to getting to physiological disease mechanisms.<sup>9</sup> Here we present a novel approach to assay the impact of individual polygenic risk factors using an *in vivo* approach.

While numerous potential therapeutics have shown promising results in transgenic mouse models of familial AD, few have advanced in clinical trials. This may result from numerous causes, but it is clear that one reason may be the lack of translational animal models available for preclinical studies.<sup>10–12</sup> Almost all existing rodent models are based on causative mutations in proteins in the APP processing pathway expressed in neurons. Most AD genetic risk resides in genes mainly expressed in microglia and other non-neuronal cell types, as recently reviewed,<sup>5,13,14</sup> indicating that complex cellular interactions play a causative role in disease etiology. While *in vitro* systems have been shown to have value, more relevant *in vivo* models are necessary to understand these cell–cell interactions.<sup>15</sup> In particular, animal models are required to study the early and progressive stages of pathology, which are not accessible in clinical studies but are critical to understand disease mechanisms so as to better target novel therapeutic approaches.

The Model Organism Development and Evaluation for Late-onset Alzheimer's Disease (MODEL-AD) Consortium was established to create and characterize translationally relevant mouse models of LOAD and to set up protocols for preclinical testing in these new models.<sup>16</sup> In this study we provide an overview of novel mouse models expressing human risk variants. Variants were introduced using a knock-in approach to avoid known issues with transgenic models.<sup>11,17–19</sup> To potentially enhance disease-relevant outcomes, variants were created on a more LOAD-susceptible genetic background expressing humanized APOE with the  $\epsilon 4$  variant and the R47H mutation in *Trem2*, two of the strongest genetic risk factors for LOAD.<sup>20</sup> The effects of each vari-

ant were assessed by gene expression changes in aging male and female brains using a newly developed transcriptomics panel,<sup>21</sup> representing key LOAD-associated changes in clinical AD samples.<sup>22</sup> This allowed us to functionalize GWAS variants with small but significant increases in disease risk and avoided a reliance on amyloid deposition or cognitive assays, which have not proven to translate to clinical studies.

## 2 | METHODS

### 2.1 | Late-onset AD risk variant prioritization

Prioritization and construction of the APOE and TREM2 variants in the LOAD1 strain were previously discussed.<sup>20</sup> Late-onset variants were selected based on human genetic association, predicted pathogenicity, conservation with mouse homolog, and allele frequency. We further prioritized based on diversity in predicted function to maximize our exploration of potential LOAD biology. Determining specific variants was primarily limited by the rarity of strong coding candidates (eg, non-synonymous, stop-gain) and strict mouse sequence homology that required the same single nucleotide polymorphism (SNP) be engineered into mice. This led to a mix of variants at high-confidence GWAS loci, functional candidates, and exploratory variants. Exome sequencing from the Alzheimer's Disease Sequencing Project (ADSP) was initially used to identify specific variants at loci,<sup>23</sup> buttressed by summary data at the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) (<https://www.niagads.org/genomics/app>). All variants are annotated as "ADSP Variants" that passed NIAGADS quality control checks (<https://www.niagads.org/genomics/app>).

ABCA7\*A1527G (rs3752246) is the most common of multiple predicted loss-of-function variants associated with increased LOAD risk at the ABCA7 locus.<sup>24,25</sup> The SORL1\*A528T (rs2298813) variant is among candidates in the SORL1 gene and likely involved in retromer function<sup>26</sup>; deficits in retromer-dependent endosomal recycling have been implicated as causal in AD.<sup>27–29</sup> The SNX1\*D465N (rs1802376) variant locus is associated with AD,<sup>24</sup> and SNX1 is involved in retromer function relevant to LOAD.<sup>30</sup> PLCG2\*M28L (rs61749044) has been associated with LOAD (<https://www.biorxiv.org/content/10.1101/2020.05.19.104216v1>),<sup>24,31</sup> and Plcg2 is a key protein in microglial activation in response to AD pathology.<sup>32</sup> The SHC2\*V433M (rs61749990) variant was identified in ADSP exomes and has been associated with neurodegeneration and neuron loss.<sup>33,34</sup> SLC6A17\*P61P (rs41281364) reduces gene expression in the brain ([gtexportal.org/home/gene/SLC6A17](http://gtexportal.org/home/gene/SLC6A17)), and its reduction is also associated with

LOAD (agora.adknowledgeportal.org/genes/ENSG00000197106). Rare variants have been associated with neurological phenotypes.<sup>35,36</sup> The CLASP2\*L163P (rs61738888) variant has been associated with neurodegeneration from meta-analysis.<sup>37</sup> The MTMR4\*V297G (rs2302189) variant has been linked to cognitive function.<sup>38,39</sup> Predicted CEACAM1 loss-of-function variants had a high disease burden in ADSP exome sequencing data (SKAT-O Bonferroni-adjusted  $p = 7.47 \times 10^{-7}$ ), and the gene was associated with AD-related traits in a model of mouse genetic variability.<sup>40</sup> The common MTHFR\*677C > T (rs1801133) has been associated with increased risk for LOAD and other age-related disorders.<sup>41,42</sup> To explore a copy-number variant linked to vascular function, we used an existing MEOX2 knock-out based on an association with AD<sup>43</sup> that may be related to the gene's role in neurovascular health.<sup>44</sup> This variant was assessed in a heterozygous state due to the non-viability of the homozygote.

## 2.2 | Model development

All experiments were approved by the Animal Care and Use Committee at The Jackson Laboratory. Mice were bred in the mouse facility at The Jackson Laboratory and maintained in a 12/12-h light/dark cycle, consisting of 12 h-ON 7 am-7 pm, followed by 12 h-OFF. Room temperatures are maintained at 18°C to 24°C (65°F to 75°F) with 40% to 60% humidity. All mice were housed in positive, individually ventilated cages (PIV). Standard autoclaved 6% fat diet (Purina Lab Diet 5K52) was available to the mice ad libitum, as was water with acidity regulated from pH 2.5 to 3.0.

Novel mouse alleles were generated using direct delivery of CRISPR-Cas9 reagents to LOAD1 (JAX No. 28709)<sup>20</sup> mouse zygotes. Analysis of genomic DNA sequence surrounding the target region, using the Benchling ([www.benchling.com](http://www.benchling.com)) guide RNA design tool, identified appropriate gRNA sequences with a suitable target endonuclease site.

*Streptococcus pyogenes* Cas9 (SpCas9) V3 protein and gRNA were purchased as part of the Alt-R CRISPR-Cas9 system using the crRNA:tracrRNA duplex format as the gRNA species (IDT, USA). Alt-R CRISPR-Cas9 crRNAs (Product 1072532, IDT, USA) were synthesized using the gRNA sequences specified in the DESIGN section and hybridized with the Alt-R tracrRNA (Product 1072534, IDT, USA) as per the manufacturer's instructions. Plasmid or oligonucleotide constructs were synthesized by Genscript. See Table S1 for CRISPR reagents.

To prepare the gene editing reagent for electroporation, SpCas9:gRNA Ribonucleoprotein (RNP) complexes were formed by incubating AltR-SpCas9 V3 (Product 1081059, IDT, USA) and gRNA duplexes for 20 min at room temperature in embryo tested TE buffer (pH 7.5). The SpCas9 protein and gRNA duplex were at 833 ng/ul and 389 ng/ul, respectively, during complex formation. After RNP formation, the purified plasmid was added and the mixture spun at 14,000 rpm in a microcentrifuge. The supernatant was transferred to a clean tube and stored on ice until use in the embryo electroporation procedure. The final concentrations of the gRNA, SpCas9, and plas-

### RESEARCH IN CONTEXT

- 1. Systematic review:** The authors review the literature and associated public datasets to identify genetic risk factors for late-onset Alzheimer's disease (LOAD).
- 2. Interpretation:** Our findings support the use of mouse models to validate and prioritize disease risk variants identified by clinical studies and are an essential step toward the development of models of LOAD to be used in mechanistic studies and in therapeutic development efforts.
- 3. Future directions:** This manuscript establishes a process to validate and prioritize animal models expressing genetic risk factors for LOAD, which are currently lacking. Based on the relevance of transcriptomic signatures to those seen in clinical studies, we will combine alleles to create polygenic models that can serve as useful models of LOAD. Moving forward, we will do in-depth analysis of these novel models at extended ages using translationally relevant measures including: fluid biomarkers; transcriptomics, proteomics, and metabolomics; neuropathology; and in vivo imaging.

mid components in the electroporation mixture were 600, 500, and 20 ng/ul, respectively.

Founders were selected that were positive by short-range polymerase chain reaction (PCR) assays, had appropriate sequence across the homology arm junctions, were negative for the plasmid backbone, and had correct sequences of the inserted construct.

Allele-specific genotyping protocols for all models are available on JAX mice data sheets for each model.

Other models were obtained from the JAX mouse repository (Table 1).

## 2.3 | Brain harvest at 4 months of age

Anesthetized and subsequently perfused animals were decapitated and heads submerged quickly in cold 1X PBS. The brain was carefully removed from the skull, weighed, and divided midsagittally into left and right hemispheres using a brain matrix. The right hemisphere was quickly homogenized on ice and equally aliquoted into cryotubes for proteomic and transcriptomic analysis. Cryotubes were immediately snap-frozen on dry ice and stored long term at  $-80^{\circ}\text{C}$ .

## 2.4 | RNA sample extraction

Total RNA was extracted from snap-frozen right brain hemispheres using Trizol (Invitrogen, Carlsbad, CA). mRNA was purified from total RNA using biotin-tagged poly dT oligonucleotides and streptavidin-coated magnetic beads, and quality was assessed using

**TABLE 1** Listing of gene loci, human risk variants, and corresponding mouse alleles, allele type, and JAX ID of mouse models created.

Locus	Allele (Human)	Allele (Mouse)	SNP	Allele Type	JAX No.
Abca7	A1527G	A1541G	rs3752246	missense	30283
Ceacam1	LOF variants	KO	—	KO	30673
Clasp2	L163P	L163P	rs61738888	missense	31944
Meox2	LOF variants	HET KO	—	HET KO	33770
Mthfr	A222V (677C > T)	A262V	rs1801133	missense	30922
Mtmt4	V297G	V297G	rs2302189	missense	31950
Plcg2	M28L	M28L	rs61749044	missense	30674
Shc2	V577M	V433M	rs2298813	missense	31952
Slc6a17	P61P	P61P	rs41281364	silent mutation	31948
Snx1	D466N	D465N	rs1802376	missense	31942
Sor11	A528T	A528T	rs41281364	missense	31940
<b>Other models used</b>					
57BL/6J					664
5xFAD					8730
LOAD1					28709

Note: All models also contain a humanized APOE  $\epsilon$ 4 allele and a Trem2\**R47H* allele on the C57BL/6J background ("LOAD1"), which was used as a control.

an Agilent Technologies 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA).

RNA-Sequencing Assay Library Preparation Sequencing libraries were constructed using TruSeq DNA V2 (Illumina, San Diego, CA, USA) sample prep kits and quantified using qPCR (Kapa Biosystems, Wilmington, MA). The mRNA was fragmented, and double-stranded cDNA was generated by random priming. The ends of the fragmented DNA were converted into phosphorylated blunt ends. An "A" base was added to the 3' ends. Illumina-specific adapters were ligated to the DNA fragments. Using magnetic bead technology, the ligated fragments were size-selected, and then a final PCR was performed to enrich the adapter-modified DNA fragments since only the DNA fragments with adapters at both ends will amplify.

## 2.5 | RNA-Sequencing

Libraries were pooled and sequenced by the Genome Technologies core facility at The Jackson Laboratory. All samples were sequenced on Illumina HiSeq 4000 using HiSeq 3000/4000 SBS Kit reagents (Illumina), targeting 30 million read pairs per sample. Samples were split across multiple lanes when being run on the Illumina HiSeq; once the data were received, the samples were concatenated to have a single file for paired-end analysis.

## 2.6 | RNA-Sequencing data processing

Sequence quality of reads was assessed using FastQC (version 0.11.3, Babraham). Low-quality bases were trimmed from sequencing reads using Trimmomatic (version 0.33).<sup>45</sup> After trimming, reads of length

longer than 36 bases were retained. The average quality score was greater than 30 at each base position, and sequencing depth was in a range of 60 to 80 million reads. RNA-Seq sequencing reads from all samples were mapped to the mouse genome (version GRCm38.p6) using ultrafast RNA-Seq aligner STAR (version 2.5.3).<sup>46</sup> To measure human APOE gene expression, we created a chimeric mouse genome by concatenating the human APOE gene sequence (human chromosome 19:44905754-44909393) into the mouse genome (GRCm38.p6) as a separate chromosome (referred to as chromosome 21 in chimeric mouse genome). Subsequently, we added gene annotation of the human APOE gene into the mouse gene annotation file. Additionally, we have also introduced annotation for novel Trem2 isoform in mouse gene annotation file (GTF file), which is identical to the primary transcript but truncated exon2 by 119 bp from its start position.<sup>20</sup> Afterward, a STAR index was built for this chimeric mouse genome sequence for alignment, then STAR aligner output coordinate-sorted BAM files for each sample were mapped to the chimeric mouse genome using this index. Gene expression was quantified in two ways to enable multiple analytical methods: transcripts per million (TPM) using RSEM (version 1.2.31)<sup>47</sup> and raw read counts using HTSeq-count (version 0.8.0).<sup>48</sup>

Engineered variants were validated by inspecting RNA-Seq reads (Figure S1A), and targeted transcript expression was verified by RNA abundance (Figure S1B).

## 2.7 | NanoString transcriptomic analysis

The NanoString Mouse AD gene expression panel<sup>21</sup> was used for gene expression profiling on the nCounter platform (NanoString, Seattle,

WA, USA). Mouse NanoString gene expression data were collected from brain hemisphere homogenates at 4, 8, and 12 months of age for both sexes, from approximately six animals per group. The nSolver software was used for generating NanoString gene expression counts. Normalization was done by dividing counts within a lane by geometric mean of the designated housekeeping genes from the same lane. Next, normalized count values were log-transformed and corrected for potential batch effects using ComBat.<sup>49</sup>

Next, we determined the effects of each factor (sex and genetic variants) by fitting a multiple regression model using the `lm` function in R as<sup>50</sup>

$$\log(\text{expr}) = \beta_0 + \sum_i \beta_i + \varepsilon.$$

The sum is over sex (male), and all genetic variants (5xFAD, LOAD1, Abca7\*A1527G, Ceacam1KO, Mthfr\*677C > T, Shc2\*V433M, Slc6a17\*P61P, Clasp2\*L163P, Sorl1\*A528T, Meox2 KO [HET], Snx1\*D465N, Plcg2\*M28L, Mtmr4\*V297G) used in this study. The  $\log(\text{expr})$  represents log-transformed normalized count from the NanoString gene expression panel.<sup>21</sup> In this formulation, B6J was used as the control for the 5xFAD and LOAD1 mouse models, whereas LOAD1 served as controls for GWAS-based models in order to estimate the effects of individual variants. Separate models were run for each age cohort.

## 2.8 | Human AMP-AD gene co-expression modules

Data for 30 human brain co-expression modules from the Accelerating Medicines Partnership for Alzheimer's Disease (AMP-AD) studies were obtained from the Synapse data repository (<https://www.synapse.org/#!/Synapse:syn11932957/tables/>; SynapseID: syn11932957). Briefly, Wan et al. (2020)<sup>22</sup> identified 30 human brain co-expression modules based on meta-analysis of differential gene expression from seven distinct brain regions in *post mortem* samples obtained from three independent LOAD cohorts.<sup>51–53</sup> These 30 human AMP-AD modules were further classified into five distinct consensus clusters that describe the major functional alterations observed in human LOAD.<sup>21,22</sup> Module and consensus cluster annotations for each NanoString gene are listed in Table S2.

## 2.9 | Human AD subtypes

Milind et al.<sup>54</sup> integrated *post mortem* brain co-expression data from the frontal cortex, temporal cortex, and hippocampus brain regions and stratified patients into different molecular subtypes based on molecular profiles in three independent human LOAD cohorts (ROS/MAP, Mount Sinai Brain Bank, and Mayo Clinic).<sup>51–53</sup> Two distinct LOAD subtypes were identified in the ROSMAP cohort, three LOAD subtypes were identified in the Mayo cohort, and two distinct LOAD subtypes were identified in the MSBB cohort. Similar subtype results were observed in each cohort, with LOAD subtypes found to primarily differ in their inflammatory response based on differential expression

analysis.<sup>54</sup> Data for LOAD subtypes were obtained through AD Knowledge Portal<sup>55</sup> (<https://www.synapse.org/#!/Synapse:syn23660885>).

## 2.10 | Mouse-human expression comparison

To assess the human disease relevance of LOAD risk variants in mice, we determined the extent to which changes due to genetic perturbations in mice matched those observed in human AD cases versus controls. For each mouse perturbation, we tested each of the 30 AMP-AD modules using mouse-human gene homologs and limited to the genes both present in the module and the NanoString Mouse AD Panel, which was designed to optimize coverage of these modules.<sup>21</sup> Pearson's correlations were computed for changes in gene expression (log-fold change) across all module genes for human AD cases versus controls<sup>22</sup> against the effect of each mouse perturbation ( $\beta$ ) as measured previously.<sup>21,50</sup> We used the `cor.test` function in R as follows:

$$\text{cor.test}(\text{Log}_2\text{FC}(\text{AD}/\text{control}), \beta),$$

from which we obtained the correlation coefficient and the significance level ( $p$ ) of the correlation for each perturbation-module pair.  $\text{Log}_2\text{FC}$  values for human transcripts were obtained through the AD Knowledge Portal<sup>55</sup> (<https://www.synapse.org/#!/Synapse:syn14237651>).

To determine the similarity of each mouse perturbation and the LOAD subtypes, we computed the Pearson's correlation between gene expression changes (log-fold change) in human AD subtype cases versus controls<sup>54</sup> and the effect of each mouse perturbation ( $\beta$ ) across genes on the NanoString panel<sup>21</sup> using `cor.test` function in R as follows:

$$\text{cor.test}(\text{Log}_2\text{FC}(\text{LOADSubtype}/\text{control}), \beta),$$

from which we obtained both the correlation coefficient and the significance level ( $p$ ) of the correlation. Here,  $\text{Log}_2\text{FC}(\text{LOAD Subtype}/\text{control})$  represented the log-fold change in gene expression in each subtype versus control, and the correlation spanned all homologous genes on the NanoString AD Mouse Panel.

We plotted the correlation results using the `ggplot2` package in R. Framed circles were used to denote significant ( $p < 0.05$ ) positive (blue) and negative (red) Pearson's correlation coefficients. The color intensity and size of the circles were sized proportional to Pearson's correlation coefficient.

## 2.11 | Functional enrichment analysis

Gene Set Enrichment Analysis (GSEA) was used based on the method proposed by Subramanian et al.<sup>56</sup> as implemented in the R Bioconductor package `clusterProfiler`<sup>57</sup> for the Reactome pathway library and Gene Ontology (GO) terms. Nanostring Mouse AD Panel genes<sup>21</sup> were ranked based on regression coefficients calculated for each factor, and GSEA was performed on this ranked dataset. The use of GSEA ensured that pathway effects were assessed relative to the genes on



the panel, as the panel was enriched for AD-relevant genes. Enrichment scores for all Reactome pathways and GO terms were computed to compare relative expression on the pathway level between each factor estimate from the regression models. We also performed GO term enrichment analyses using the “enrichGO” function in the clusterProfiler package.<sup>57</sup> The significance of pathways and GO terms was determined using false discovery rate (FDR) multiple testing correction (FDR-adjusted  $p < 0.05$ ).

### 3 | RESULTS

#### 3.1 | Validation of novel models

Sequence analysis demonstrated that the appropriate sequence variants had been established (Figure S1A). Quantification of transcript counts in homozygous LOAD models relative to littermate wild-type controls showed no significant differences in expression levels (Figure S1B).

#### 3.2 | LOAD-associated risk variants showed age-dependent concordance with distinct human co-expression modules

We assess the relevance of each LOAD risk variant to the molecular changes observed in human disease<sup>51–53,58</sup> by correlating the effect of each mouse perturbation (sex and genetic variants) with 30 human AMP-AD brain gene co-expression modules<sup>22</sup> using the NanoString Mouse AD Panel<sup>21</sup> (Figure 1). We analyzed mouse NanoString data from brain hemispheres at different ages (4 and 12 months) to assess the correlation with human *post mortem* co-expression modules as animals aged.

The amyloidogenic 5xFAD transgenic model exhibited significant positive correlations ( $p < 0.05$ ) with several human co-expression modules in Consensus Cluster B enriched for immune-system-related pathways at both 4 and 12 months but showed significant positive correlations ( $p < 0.05$ ) with neurodegeneration-associated human co-expression modules in Consensus Cluster C only at 12 months (Figure 2A–B). However, we did not observe significant positive correlations between effect of 5xFAD and human co-expression modules in Consensus Clusters A, D, and E, validating that the 5xFAD strain is primarily a model of amyloidosis and does not fully recapitulate LOAD changes.

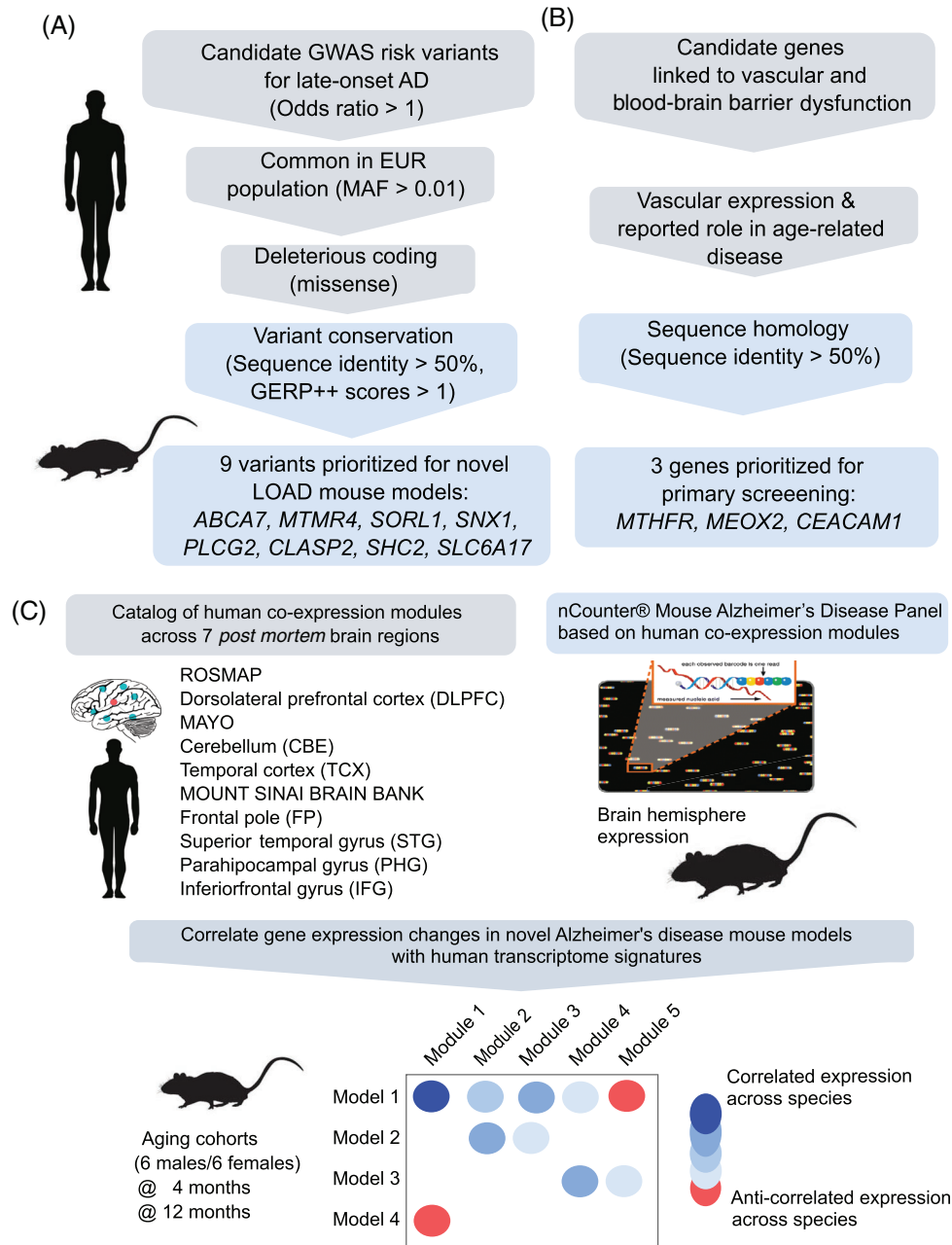
At 4 months, among all LOAD risk variants, only *Slc6a17\*P61P* showed significant positive correlations ( $p < 0.05$ ) with the immune-related modules (Figure 2A). The *Abca7\*A1527G*, *Sorl1\*A528T*, and *Mtmr4\*V297G* risk variants exhibited significant positive correlations ( $p < 0.05$ ) with extracellular matrix organization-related modules in Consensus Cluster A (Figure 2A). The *Ceacam1 KO*, *Plcg2\*M28L*, *Meox2 KO(HET)*, and *Mtmr4\*V297G* strains exhibited significant positive correlations ( $p < 0.05$ ) with cell-cycle- and myelination-associated modules in Consensus Cluster D and cellular stress-response-

associated modules in Consensus Cluster E (Figure 2A). *Abca7\*A1527G* and *Sorl1\*A528T* variants generated significant positive correlations ( $p < 0.05$ ) with cellular stress-response-associated modules in Consensus Cluster E.

We observed more significant correlations between LOAD risk variants and human AMP-AD modules at 12 months for most strains. The *Abca7\*A1527G* variant had the most pronounced correlations with LOAD expression changes, exhibiting significant positive correlations ( $p < 0.05$ ) with immune-related modules in Consensus Cluster B, cell-cycle- and myelination-associated modules in Consensus Cluster D, and cellular stress-response associated modules in Consensus Cluster E (Figure 2B). The *Mthfr\*677C > T* variant exhibited significant positive correlations ( $p < 0.05$ ) with cell-cycle- and myelination-associated modules in Consensus Cluster D and cellular stress-response-associated modules in Consensus Cluster E (Figure 2B). *Sorl1\*A528T* led to significant positive correlations ( $p < 0.05$ ) with several human co-expression modules in Consensus Cluster C enriched for neuron-related pathways (Figure 2B). The *Plcg2\*M28L* variant had significant positive correlations ( $p < 0.05$ ) with human co-expression modules in Consensus Cluster C enriched for neuron-related pathways and with cell-cycle- and myelination-associated modules in Consensus Cluster D (Figure 2B). *Ceacam1 KO*, *Slc6a17\*P61P*, and *Shc2\*V433M* exhibited significant positive correlations ( $p < 0.05$ ) with human co-expression modules in Consensus Cluster B enriched for transcripts associated with immune-related pathways in multiple brain regions, while *Clasp2\*L163P* and *Sorl1\*A528T* led to significant positive correlations ( $p < 0.05$ ) with the human co-expression module in Consensus Cluster B enriched for immune related pathways in cerebellum and frontal pole brain region, respectively (Figure 2B). The *Mtmr4\*V297G* variants exhibited significant positive correlations ( $p < 0.05$ ) with cell-cycle- and myelination-associated modules in Consensus Cluster D and cellular stress-response-associated modules in Consensus Cluster E (Figure 2B). *Snx1\*D465N* also exhibited significant positive correlation with cell-cycle- and myelination-associated modules in Consensus Cluster D (Figure 2B).

Overall, we observed LOAD risk variants in mice showed concordance with distinct human co-expression modules, reflecting a different transcriptional response resulting from each LOAD risk variant. The associations between LOAD risk variants and human gene co-expression modules increased with age. We note that models harboring LOAD risk variants exhibited significant positive correlation with human modules in Consensus Clusters A, D, and E, which were not captured by the 5xFAD strain, highlighting the importance of using LOAD risk variants to fully capture LOAD molecular pathologies.

We next assessed the similarities between variant effects in mice by comparing each model to all other models. To identify LOAD risk variants driving similar transcriptional responses in mice, we performed a correlation between regression coefficients calculated for each genetic variant at 4 and 12 months. At 4 months, the effects of the LOAD1 construct (APOE4 and TREM2\*R47H) were weakly and positively correlated with the effect of the 5xFAD transgene ( $p < 0.05$ ), but this correlation diminished at 12 months (Figure 3A,B). The effects of LOAD1 were also significantly positively correlated ( $p < 0.05$ ) with



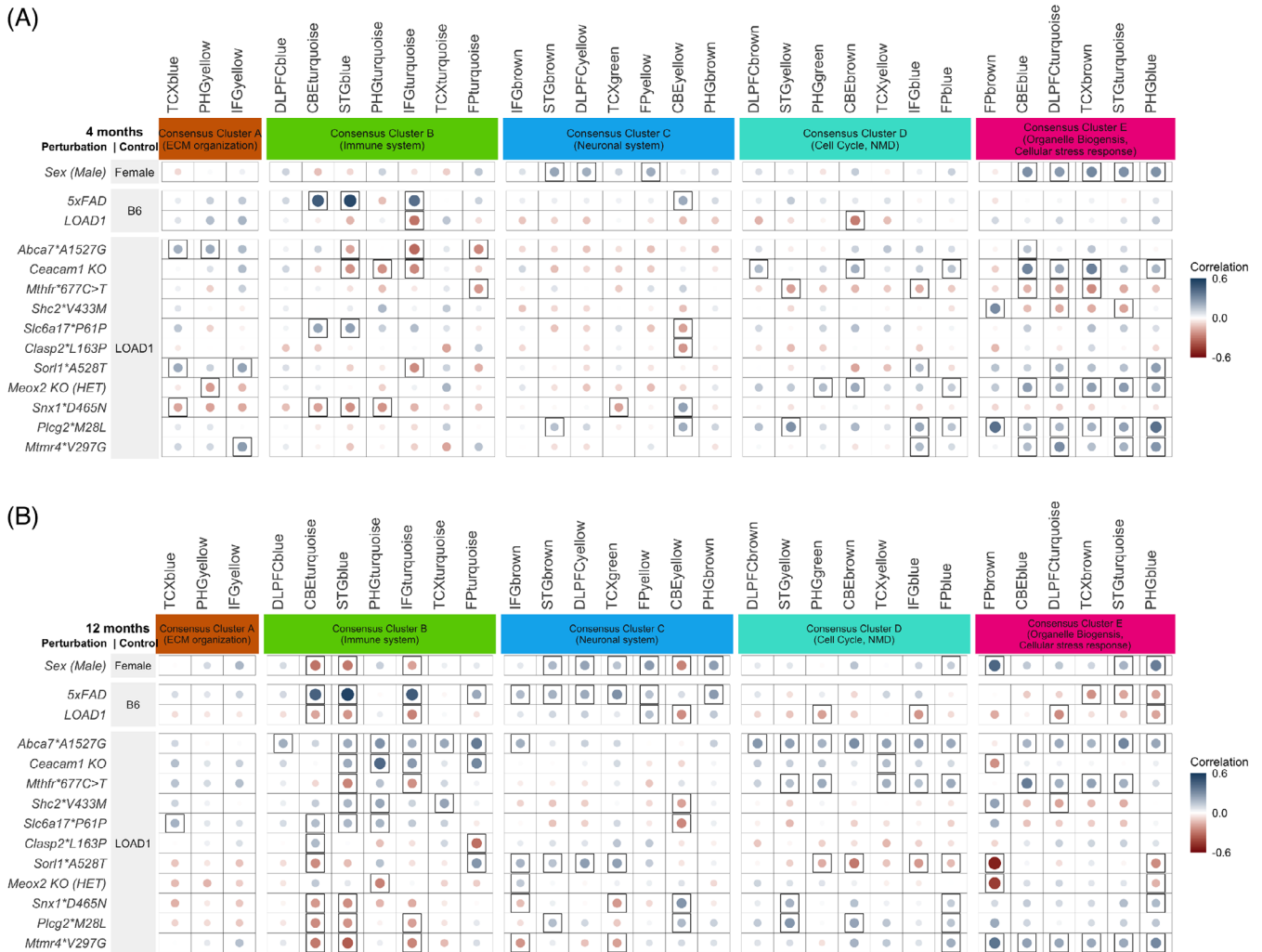
**FIGURE 1** Strategy to prioritize loci and LOAD risk variants. Summary of strategies for variant selection for (A) LOAD and (B) neurovascular risk factors. (C) Gene expression analysis comparing human and mouse gene expression data to identify human LOAD modules that are altered by genetically engineered variants in mice.

*Sorl1*\*A528T and *Mtmt4*\*V297G at 4 months, but this correlation diminished by 12 months (Figure 3A,B). The effects of the *Abca7*\*A1527G and *Ceacam1* KO variants were weakly correlated at 4 months ( $p < 0.05$ ), and this correlation increased at 12 months (Figure 3A,B). The effects of the *Shc2*\*V433M and *Slc6a17*\*P161P variants were also significantly positively correlated at 4 months ( $p < 0.05$ ) and became stronger with age (Figure 3A,B). Furthermore, the effects of the *Snx1*\*D465N, *Plcg2*\*M28L, and *Mtmt4*\*V297G risk variants were significantly positively correlated ( $p < 0.05$ ) at 12 months. Similarly, the effects of the *Sorl1*\*A528T and *Meox2* KO(HET) variants were significantly positively correlated ( $p < 0.05$ ) at 12 months (Figure 3A-B). In summary, we

observed that LOAD risk variants generally increased in similarity with age, supporting an age-dependent role for these genetic factors. However, not all strains converged on similar transcriptional responses, suggesting distinct mechanisms of influence on LOAD risk.

### 3.3 | Pathway alterations varied by LOAD genetic perturbation

To further elucidate the functional role of these LOAD risk variants in aged mice, we performed GSEA<sup>56</sup> for the Reactome pathway library

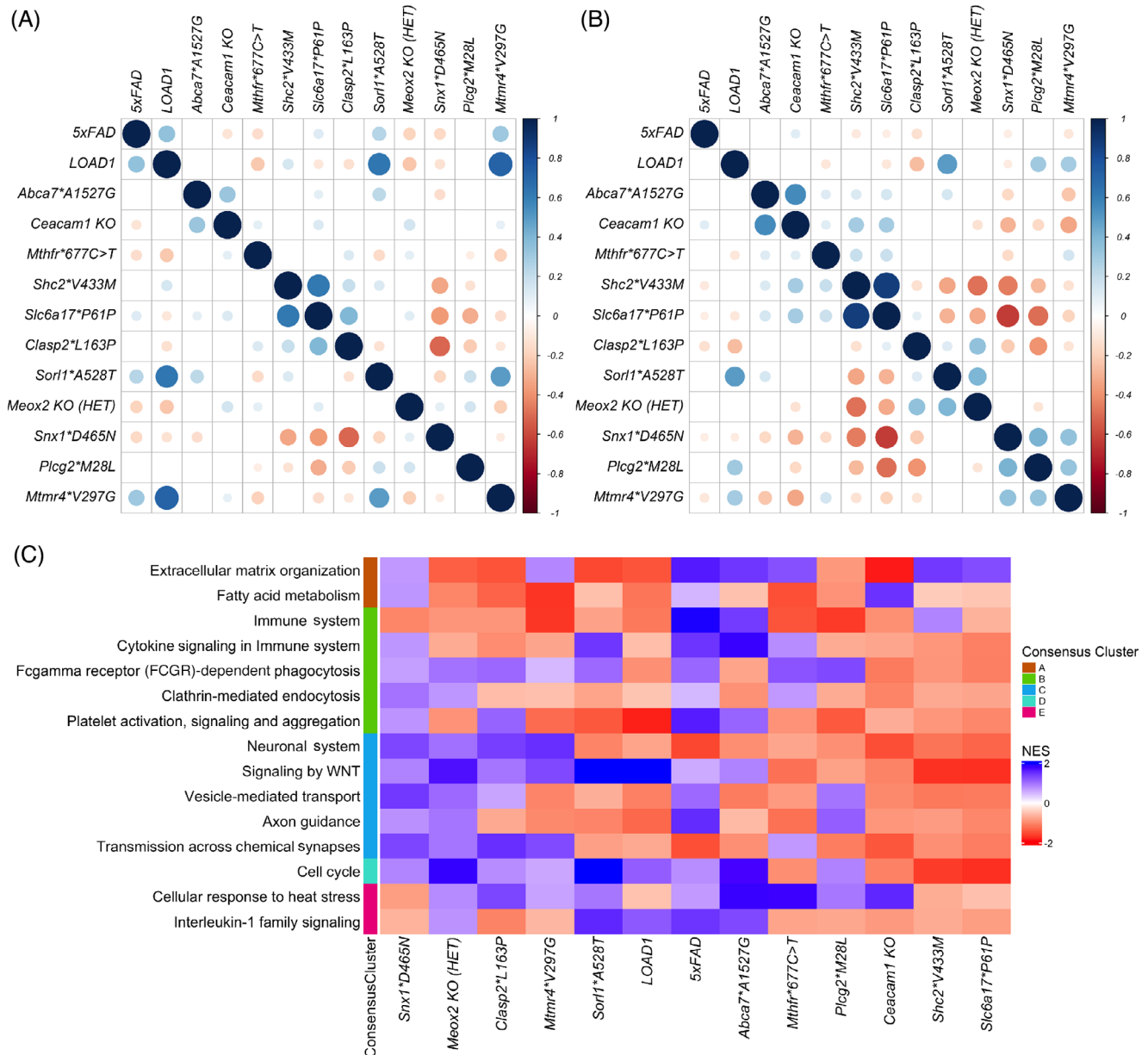


**FIGURE 2** Correlation between LOAD-associated risk variants and 30 human AMP-AD brain co-expression modules using the NanoString Mouse AD panel. (A) Correlation between the effect of each mouse perturbation relative to the LOAD1 background in 4-month-old mice and 30 human co-expression modules,<sup>22</sup> also including the early-onset transgenic model 5XFAD and the LOAD1 background relative to C57BL/6J. The 30 human co-expression modules were grouped into five consensus clusters with similar gene content across the multiple studies and brain regions.<sup>22</sup> Framed circles correspond to significant ( $p < 0.05$ ) positive (blue) and negative (red) Pearson's correlation coefficients, with size and color intensity proportional to the correlation. The effects of multiple LOAD risk variants in mice were positively correlated ( $p < 0.05$ ) with cell cycle and myelination-associated modules in Consensus Cluster D and cellular stress-response-associated modules in Consensus Cluster E. (B) Correlation between effect of each mouse perturbation at 12 months and the 30 human co-expression modules. LOAD risk variants showed significant correlation with functionally distinct AMP-AD co-expression modules. The effects of *Abca7*\*A1527G, *Shc2*\*V433M, *Ceacam1* KO, and *Slc6a17*\*P61P in aged mice correlated with the immune modules in Consensus Cluster B, while the effects of *Sorl1*\*A528T and *Plcg2*\*M28L correlated with the neuronal modules in Consensus Cluster C.

for all 12-month samples. GSEA revealed upregulation of immune system pathway in the presence of *Abca7*\*A1527G (NES = 1.42,  $p = 0.01$ ) and cytokine signaling in the immune system pathway in the presence of *Abca7*\*A1527G (NES = 1.83,  $p = 0.006$ ) and *Sorl1*\*A528T (NES = 1.45,  $p = 0.04$ ) (Figure 3C, Table S3), while neuron-associated pathways were downregulated in the presence of risk variants such as *Ceacam1* KO (NES = -1.51,  $p = 0.04$ ), *Shc2*\*V433M (NES = -1.72,  $p = 0.004$ ), and *Slc6a17*\*P61P (NES = -1.74,  $p = 0.005$ ) (Figure 3C, Table S3). The extracellular matrix organization pathway was downregulated in risk variants such as *Sorl1*\*A528T (NES = -1.56,  $p = 0.01$ )

and *Ceacam1* KO (NES = -1.85,  $p = 0.007$ ) but possibly upregulated in the presence of risk variants such as *Abca7*\*A1527G (NES = 1.45,  $p = 0.07$ ) and *Mthfr*\*677C > T (NES = 1.25,  $p = 0.16$ ) (Figure 3C, Table S3). The cell cycle pathway was downregulated in the presence of *Shc2*\*V433M (NES = -1.65,  $p = 0.01$ ) and *Slc6a17*\*P61P (NES = -1.73,  $p = 0.004$ ) but upregulated in the presence of other risk variants such as *Abca7*\*A1527G (NES = 1.75,  $p = 0.01$ ), *Meox2* KO(HET) (NES = 1.85,  $p = 0.006$ ), and *Sorl1*\*A528T (NES = 2.61,  $p = 0.002$ ) (Figure 3C, Table S3). Cellular response to the heat stress pathway were upregulated in the presence of risk variants such as



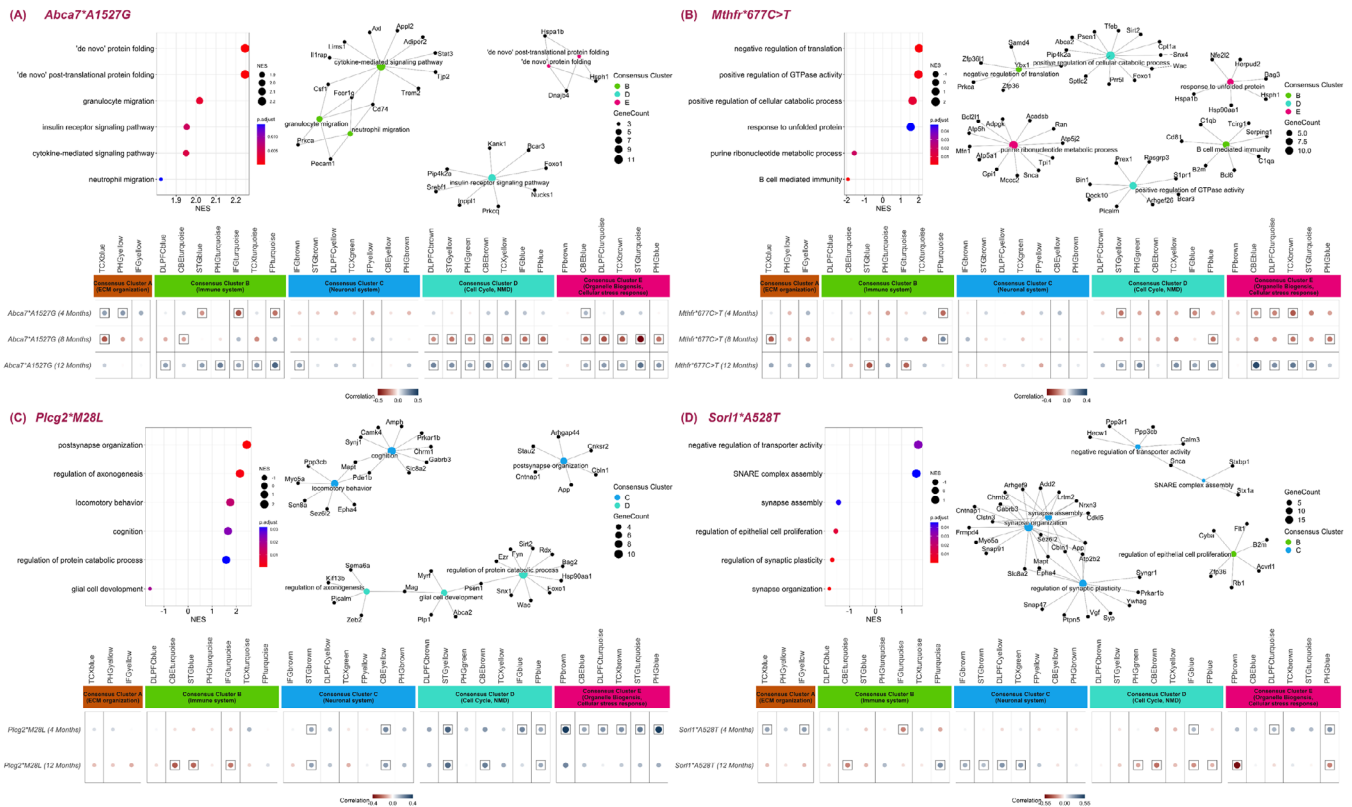


**FIGURE 3** Correlation between effect of genetic variants and Gene Set Enrichment Analysis (GSEA). (A) Correlation between regression coefficients calculated for each genetic variant at 4 months. Color intensity and size of circles are proportional to Pearson correlation coefficient, with insignificant correlations ( $p > 0.05$ ) left blank. (B) Correlation between regression coefficients calculated for each genetic variant at 12 months. The effects of *Snx1*\*D465N, *Plcg2*\*M28L, and *Mtmr4*\*V297G risk variants in mice showed significantly positive correlation ( $p < 0.05$ ) at 12 months. (C) GSEA results of selected AD-associated pathways from Reactome library in presence of each LOAD risk variant in mice. Enriched pathways are grouped by their overlap with functional annotations of human AMP-AD Consensus Clusters. Immune-related pathways had increased expression in the presence of multiple risk variants such as *Abca7*\*A1527G, *Mthfr*\*677C > T, and *Snx1*\*D465N, while neuron-associated pathways had reduced expression in the presence of risk variants such as *Abca7*\*A1527G, *Mthfr*\*677C > T, *Sorl1*\*A528T, *Plcg2*\*M28L, *Ceacam1* KO, *Shc2*\*V433M, and *Slc6a17*\*P61P.

*Abca7*\*A1527G (NES = 1.83,  $p = 0.01$ ), *Mthfr*\*677C > T (NES = 1.81,  $p = 0.009$ ), and *Ceacam1* KO (NES = 1.57,  $p = 0.04$ ) (Figure 3C, Table S3). Overall, we observed that multiple AD-associated pathways were upregulated in the presence of some LOAD risk variants but downregulated in the presence of another set of risk variants. This suggests that distinct risk variants perturb distinct molecular changes associated with LOAD in aging mice.

### 3.4 | Age-dependent pathway effects driving AMP-AD module correlations in ABCA7, MTHFR, SORL1, and PLCG2 mouse models

In our mouse-human correlation analysis, the effects of multiple LOAD variants (*Abca7*\*A1527G, *Mthfr*\*677C > T, *Sorl1*\*A528T, and *Plcg2*\*M28L) correlated with human AMP-AD co-expression modules



**FIGURE 4** Identification of specific AD-associated processes in LOAD risk variants exhibiting transcriptomic changes similar to human LOAD in age-dependent manner. For four new mouse strains the following are displayed: the six top enriched GO terms identified by GSEA and GO enrichment analysis of genes with common directional changes with human AD modules (top left); gene module networks with common directional changes with human AMP-AD modules, where node colors correspond to human AMP-AD Consensus Clusters A (orange), B (green), C (blue), D (turquoise), or E (pink) (top right); and the effects of each variant at multiple ages correlated across LOAD effects in 30 AMP-AD modules, following the legend of Figure 3. Results for (A) *Abca7*\*A1527G model, (B) *Mthfr*\*677C > T model, (C) *Plcg2*\*M28L model, and (D) *Sor11*\*A528T model. All results are relative to LOAD1 genetic background for all strains.

in age-dependent and pathway-specific manners. To further identify the AD-relevant biological processes associated with these selected LOAD risk variants (*Abca7*\*A1527G, *Mthfr*\*677C > T, *Sor11*\*A528T, and *Plcg2*\*M28L) we adopted two approaches. First, we performed GSEA<sup>56</sup> on the NanoString Mouse AD Panel genes ranked based on regression coefficients calculated for each factor at 12 months and identified significantly enriched GO terms (padj < .05). Next, we isolated the homologous genes exhibiting directional coherence between the effects of genetic risk variants (*Abca7*\*A1527G, *Mthfr*\*677C > T, *Sor11*\*A528T, and *Plcg2*\*M28L) at 12 months and changes in human cases versus controls and performed GO enrichment analysis to find processes underlying the module-level correlations. These subsets represent the pathways that<sup>1</sup> are altered in each mouse model and<sup>2</sup> quantitatively underlie the mouse-human module associations. GO terms common to both enrichment tests were then annotated to the modules in which they appeared.

The *Abca7*\*A1527G variant showed significant negative correlations ( $p < 0.05$ ) with immune-related modules in Consensus Cluster B, cell-cycle- and myelination-associated modules in Consensus Cluster D, and cellular stress-response-associated modules in Consensus Cluster E (Figure 4A) at 4 months. However, at 12 months these effects were

reversed and the variant exhibited significant positive correlations ( $p < 0.05$ ) with several immune-related modules in Consensus Cluster B, cell-cycle-associated and myelination-associated modules in Consensus Cluster D, and cellular stress-response-associated modules in Consensus Cluster E (Figure 4A). Biological processes such as “de novo” protein folding, “de novo” post-translational protein folding, granulocyte migration, cytokine-mediated signaling pathway, insulin receptor signaling pathway, and neutrophil migration increased expression in the presence of *Abca7*\*A1527G (Figure 4A, Table S4). The correlation between the *Abca7*\*A1527G variant and the immune-associated human co-expression modules (Consensus Cluster B) (Figure 4A, Table S5) was exhibited by genes enriched for granulocyte migration, cytokine-mediated signaling pathway, and neutrophil migration (including *Pecam1*, *Cd74*, *Trem2*, *Trem1*, *Csf1*, *Il1rap*, and *Ceacam1*) (Table S6). As key correlating genes between *Abca7*\*A1527G and Consensus Cluster E modules (Figure 4A, Table S5), we found genes enriched in “de novo” protein folding and “de novo” post-translational protein folding (eg, *Hspa2*, *Hspa1b*, and *Dnajb4*) (Table S6). Insulin receptor signaling was enriched in genes (*Foxo1*, *Prkacq*, and *Bcar3*) (Table S6), driving the correlation between *Abca7*\*A1527G and modules in Consensus Cluster D (Figure 4A, Table S5).

A similar reversal of effects with age was observed for *MTHFR*. The *Mthfr*\*677C > T variants exhibited significant negative correlations ( $p < 0.05$ ), with several cell-cycle- and myelination-associated modules in Consensus Cluster D and cellular stress-response-associated modules in Consensus Cluster E (Figure 4B) at 4 months. At 12 months, these correlations were positive (Figure 4B). GSEA of the *Mthfr*\*677C > T variant identified significant enrichments of response to unfolded protein, positive regulation of cellular catabolic process, negative regulation translation, positive regulation of GTPase activity, B-cell-mediated immunity, and purine ribonucleotide metabolic process (Figure 4B, Table S4). B-cell-mediated immunity and negative regulation translation biological processes were also enriched in genes (including *C1qa*, *C1qb*, *Cd81*, and *Zfp36*) (Table S6) with directional coherence for *Mthfr*\*677C > T and LOAD effects in Consensus Cluster B (Figure 4B, Table S5). Correlations between the *Mthfr*\*677C > T variant and Consensus Cluster D changes (Figure 4B, Table S5) were exhibited by genes enriched for positive regulation of cellular catabolic process and positive regulation of GTPase activity (including *Bin1*, *Picalm*, *Dock10*, and *Psen1*) (Table S6). Biological processes such as response to unfolded protein and purine ribonucleotide metabolic process were enriched in genes (eg, *Hspa1b*, *Hsph1*, *Hsp90aa1*, *Snca*, and *Atp5h*) (Table S6) underlying the correlations between *Mthfr*\*677C > T and Consensus Cluster E effects (Figure 4B, Table S5).

The *Plcg2*\*M28L variant caused significant positive correlations ( $p < 0.05$ ) with neuron-related modules in Consensus Cluster C and cell-cycle-associated modules in Consensus Cluster D at both 4 and 12 months (Figure 4C). Enriched biological processes included postsynapse organization, regulation of axonogenesis, cognition, locomotory behavior, glial cell development, and regulation of protein catabolic process (Figure 4C, Table S4). Biological processes such as postsynapse organization, cognition, and locomotory behavior were enriched in genes (*Mapt*, *Gabbr3*, *App*, *Ppp3cb*, and *Slc8a2*) (Table S6) with directional coherence for *Plcg2*\*M28L human AD changes in Consensus Cluster C (Figure 4C, Table S5). Biological processes such as regulation of axonogenesis, glial cell development, and regulation of protein catabolic process were enriched in genes (*Snx1*, *Picalm*, *Psen1*, *Mag*, *Foxo1*, and *Kif13b*) (Table S6) and drove the correlations between *Plcg2*\*M28L and Consensus Cluster D effects (Figure 4C, Table S5).

Aged *Sorl1*\*A528T mice (12 months) showed positive correlations ( $p < .05$ ) with neuron-associated modules in Consensus Cluster C that were not apparent at 4 months of age (Figure 4D). Enriched processes included the downregulation of synapse organization, synapse assembly, regulation of synaptic plasticity, and regulation of epithelial cell proliferation and the increased expression of negative regulation of transporter activity and SNARE complex assembly genes. These processes drove the correlation between the *SORL1* variant and LOAD effects in Consensus Cluster C modules (Figure 4D, Table S5), where GSEA for genes with directional coherence generated synapse organization, synapse assembly, regulation of synaptic plasticity, upregulation of negative regulation of transporter activity, and soluble N-ethylmaleimid-sensitive factor attachment receptor (SNARE) complex assembly (including the genes *Mapt*, *App*, *Gabbr3*, *Calm3*, *Snca*, *Cdkl5*, *Vgf*, and *Ywhg*) (Table S6).

Overall, we found that late-onset genetic factors in mice generally led to both more abundant changes with age and increasingly disease-relevant pathway changes with age.

### 3.5 | Alignment of mouse models with AD subtypes

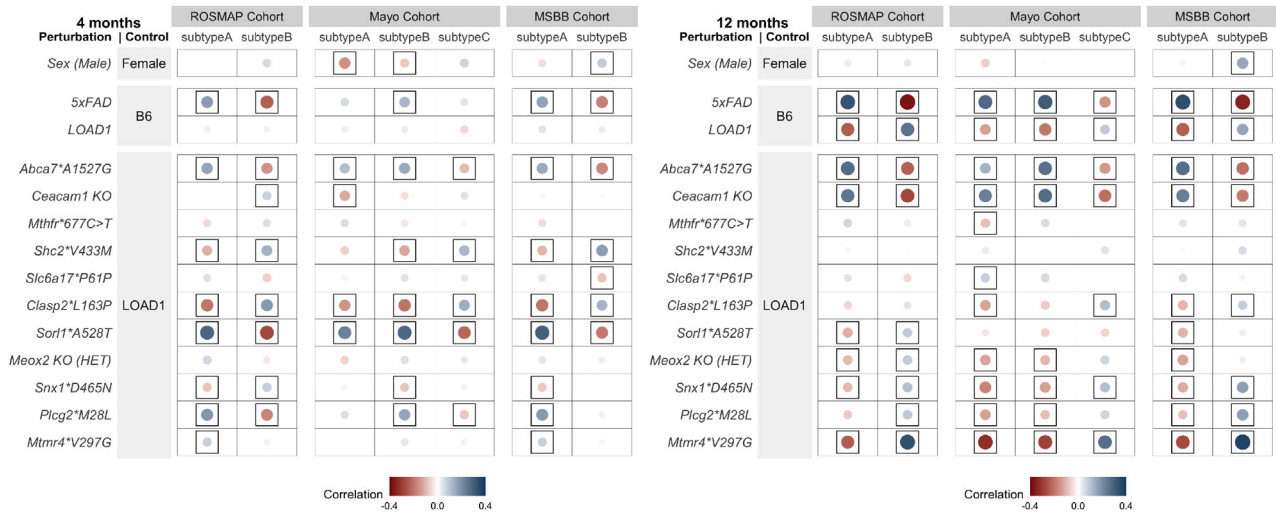
*Post mortem* transcriptomics from AMP-AD and similar studies have enabled the partitioning of AD cases into potential disease subtypes. These studies have often stratified AD subjects into inflammatory and non-inflammatory subtypes.<sup>54,59,60</sup> To determine whether our mouse models preferentially resembled putative AD subtypes, we correlated the effect of each variant with inflammatory and non-inflammatory subtypes associated with LOAD<sup>54</sup> in the ROSMAP, MSBB, and Mayo cohorts.<sup>51-53</sup>

We found that at 4 months of age, variants did stratify by human subtypes. The effects of *Abca7*\*A1527G, *Sorl1*\*A528T, and *Plcg2*\*M28L were positively correlated ( $p < 0.05$ ) with the inflammatory subtypes across all three cohorts, while *Mtmr4*\*V297G was positively correlated ( $p < 0.05$ ) with ROSMAP and MSBB inflammatory subtypes (Figure 5). In contrast, *Shc2*\*V433M and *Clasp2*\*L163P exhibited significant positive correlations ( $p < 0.05$ ) with non-inflammatory subtypes across all three cohorts (Figure 5).

At 12 months, the correlations between *Abca7*\*A1527G effects and the inflammatory subtypes across all three cohorts increased ( $p < 0.05$ ), and the *Ceacam1* KO variant had become positively correlated ( $p < 0.05$ ) with the inflammatory subtypes across all three cohorts (Figure 5). On the other hand, *LOAD1*, *Meox2* KO (HET), and *Snx1*\*D465N were positively correlated ( $p < 0.05$ ) with non-inflammatory subtypes across all three cohorts (Figure 5). Three strains, *Sorl1*\*A528T, *Plcg2*\*M28L, and *Mtmr4*\*V297G, which were positively correlated ( $p < 0.05$ ) with inflammatory subtypes at 4 months, transitioned to correlation ( $p < 0.05$ ) with non-inflammatory subtypes at 12 months (Figure 5). These results are in concordance with our findings that *Abca7*\*A1527G was significantly correlated with immune-related human modules and were enriched for immune-associated biological processes (Figure 4A), while the *Sorl1*\*A528T and *Plcg2*\*M28L variants were significantly correlated with neuron-related human modules and enriched for neuron-associated biological processes (Figure 4C,D). Overall, these findings suggest that different mouse strains may provide better models for distinct AD subtypes and that risk for these subtypes may be influenced by distinct AD genetic factors.

## 4 | DISCUSSION

In this study, we performed gene expression screening of new knock-in mouse models harboring candidate genetic variants for LOAD. Our ultimate goal is to provide the research community and therapeutic development programs with improved preclinical models of LOAD, suitable for preclinical testing of therapeutics that target molecular



**FIGURE 5** Correlation between effect of each mouse perturbation and molecular subtypes of LOAD. Two molecular LOAD subtypes inferred in ROSMAP cohort, three subtypes in Mayo cohort, and two subtypes in Mount Sinai Brain Bank (MSBB) cohort.<sup>54</sup> Framed circles correspond to significant ( $p < 0.05$ ) positive (blue) and negative (red) Pearson's correlation coefficients across all genes on the NanoString panel, with color intensity and circle size proportional to the correlation. (B) At 4 months, the *Abca7*\*A1527G and *Sorl1*\*A528T variants represent inflammatory subtypes of LOAD (Subtypes A) in each cohort, while *Shc2*\*V433M and *Clasp2*\*L163P variants mimic the non-inflammatory subtypes of LOAD (Subtypes B). (C) At 12 months, the *Abca7*\*A1527G and *Ceacam1* KO variants recapitulate inflammatory subtypes of LOAD (Subtypes A), while the *Snx1*\*D465N, *Mttr4*\*V297G, and *LOAD1* variants model non-inflammatory subtypes of LOAD (Subtypes B).

processes contributing to LOAD origins and progression. By basing these models on human genetics, we also provide a preliminary functional characterization of possible disease-relevant effects from the candidate genetic variants.

Notable results include the finding that many AD-related pathways, modules, and processes are affected by the introduction of late-onset variants. However, the changes were not consistent across strains, suggesting that different genetic loci contributed to distinct AD-related dysfunction (Figures 2 and 4). For example, we determined that the *SORL1* risk factor impinged primarily on AD-relevant synaptic gene expression, while the *ABCA7* variant broadly affected non-neuronal gene expression, including immune, protein folding, and metabolic pathways. Meanwhile, the *PLCG2* variant primarily affected genes that were annotated to behavior, synapses, and glial cells and similarly changed in human LOAD. We noted that a transgenic model harboring familial AD mutations in *App* and *Psen1* exhibited different gene expression changes focused on an acute inflammatory response. This can be contrasted with the *LOAD1* model that exhibited reduced immune activity and more subtle overall results, augmented by our panel of late-onset GWAS variants. We observed effects on multiple non-immune modules resulting from these variants, providing utility to study pathways not well modeled by the amyloid strains. Finally, the limited effects of variants like *Clasp2*\*L163P suggest that the specific variant is not disease-associated, its AD-related effects are not visible in the transcriptome, and/or it does not trigger changes until later age. This diversity of effects across mouse strains provides specific models to study different aspects of AD biology and paves the way for precision preclinical testing of candidate therapeutics that target these pathways.

Preliminary analysis further suggested that the different loci contributed in an age-dependent manner (Figures 2 and 4) and model putative disease subtypes (Figure 5). However, validation of such partitioning of genetic risk is difficult in human studies due to *post mortem* tissue sampling and limited cohort size for multiomic data.<sup>54</sup> We also found that the gene expression effects of LOAD variant knock-ins generally increased in terms of magnitude and disease relevance as mice aged from 4 to 12 months (Figures 2 and 4). This finding supports the notion that LOAD genetic factors become more relevant in an aging brain as they contribute to late-life disease risk.

We note that genetic variants from frequently associated loci tended to produce the most consistent AD-relevant phenotypes (eg, *SORL1*, *ABCA7*, *PLCG2*), although many of the more exploratory variants also generated AD-like expression signatures across multiple modules in aging mice (eg, *CEACAM1*, *MTMR4*) (Figure 2). Recent advances in variant inference and functional prediction, including many non-coding variants and major GWAS loci, will enable the next round of models to address additional GWAS loci without candidate coding variants, such as the *EPHA1* locus.<sup>25</sup> Furthermore, many AD-associated loci suffered from insufficient homology in mice (eg, *MS4A4A/MS4A6E*, *INPP5D*, *CR1*), which will be addressed by ongoing efforts to humanize these relevant regions of the mouse genome (Benzow, et al., this issue).

This study had several caveats that need to be noted. Most importantly, aging is the strongest risk factor for LOAD,<sup>61</sup> and it needs to be recognized that mice at 12 months of age are roughly equivalent to humans at 38 to 47 years of age. Therefore, our transcriptomic comparison to *post mortem* AMP-AD clinical samples, while practical, is unrealistic, and we are now testing those models that best approximated human transcriptional changes at 12 months to at least 24



months of age<sup>31,62</sup> (Oblak, et al., this issue). Likewise, recent studies (as well as our pilot data) have shown that proteomics is a more reliable means to correlate models to disease than transcriptomics<sup>63,64</sup> (Oblak, et al., this issue), so we will be using proteomic analysis on prioritized models.

The Trem2<sup>R47H</sup> allele in the LOAD1 base model used here has been shown to cause an ~twofold decrease in Trem2 expression.<sup>65</sup> However, our analysis technique factors out allele effects individually, so we are confident in our results. We have since created a new model (JAX No. 33781) that we have shown has normal Trem2 transcript levels and that will replace the allele used here in future projects.

In this study, we have focused on introducing coding variants on a LOAD1 background,<sup>20</sup> aged the mice to middle age (12 months), and characterized the animals using a gene expression panel developed for rapid comparison to recent human study results.<sup>21</sup> In future work, we will extend our approach to model candidate non-coding variants at LOAD genetic loci without strong candidate coding SNPs, humanizing loci and regulatory regions when necessary (Benzow, et al., this issue). We will breed the most promising variants presented here – *Abca7*<sup>A1527G</sup>, *Sorl1*<sup>A528T</sup>, *Mthfr*<sup>677C > T</sup>, and *Plcg2*<sup>M28L</sup> – to a genetic background with humanized A $\beta$  peptide (the LOAD2 strain) and age cohorts beyond 18 months to assess additional disease-related progression with advanced age. These select strains will be assessed in depth with multiple genome-scale omics measures (RNA-Seq, tandem mass tag proteomics, metabolomics), plasma biomarkers, in vivo imaging, neuropathology, and behavioral metrics. Each assay will be optimized for translational value. We will also introduce modifiable risk factors through unhealthy diets and exposure to common environmental toxicants. At the same time, all models are distributed without use restrictions to enable all researchers to obtain, study, and modify these models as desired.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Genome Technology core and Candice Baker and Kim Martens in the Genetic Engineering Technologies Service at The Jackson Laboratory for expert assistance with the work described in this publication. The results published here are based, in whole or in part, on data obtained from the AD Knowledge Portal (<https://adknowledgeportal.org>). Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by National Institute on Aging (NIA) grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG36836 (RNAseq), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at [www.radc.rush.edu](http://www.radc.rush.edu). Mount Sinai Brain Bank data were generated from *post mortem* brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine. The Mayo RNAseq study data were led by Nilüfer Ertekin-Taner, Mayo Clinic, Jacksonville, Florida, as part of the multi-PI U01 AG046139 (MPIs Golde, Ertekin-Taner, Younkin, Price). Samples were provided from the following sources: The Mayo

Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, National Institute of Neurological Disorders and Stroke (NINDS) grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data include samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program was supported by the NINDS (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the NIA (P30 AG19610 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium), and the Michael J. Fox Foundation for Parkinson's Research. The IU/JAX/PITT MODEL-AD Center was supported through funding by NIA grant U54AG054345.

## CONFLICT OF INTEREST STATEMENT

The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## CONSENT STATEMENT

No consent was required as all human subject data were reused under controlled access with an active AD Knowledge Portal Data Use Certificate (version 7.3). All data were anonymized by original sources with no possibility of deanonymization.

## DATA AVAILABILITY STATEMENT

The MODEL-AD datasets are available via the AD Knowledge Portal (<https://adknowledgeportal.org>). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses, and tools are shared early in the research cycle without a publication embargo on secondary use. Data are available for general research use according to the following requirements for data access and data attribution (<https://adknowledgeportal.org/DataAccess/Instructions>). All mouse models are available from the Jackson Laboratory mouse repository.

## ETHICS STATEMENT

The animal study was reviewed and approved by the Jackson Laboratory Animal Use Committee.

## REFERENCES

1. De Jager PL, Shulman JM, Chibnik LB, et al. A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiol Aging*. 2012;33(5):1017 e1-1017 e15.
2. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414-430.



3. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45(12):1452-1458.
4. Frisoni GB, Altomare D, Thal DR, et al. The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. *Nat Rev Neurosci.* 2021;23(1):53-66.
5. Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry.* 2015;77(1):43-51.
6. Visscher PM, Yengo L, Cox NJ, Wray NR. Discovery and implications of polygenicity of common diseases. *Science.* 2021;373(6562):1468-1473.
7. Martens YA, Zhao N, Liu CC, et al. ApoE Cascade Hypothesis in the pathogenesis of Alzheimer's disease and related dementias. *Neuron.* 2022;110(8):1304-1317.
8. Khani M, Gibbons E, Bras J, Guerreiro R. Challenge accepted: uncovering the role of rare genetic variants in Alzheimer's disease. *Mol Neurodegener.* 2022;17(1):3.
9. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. *Science.* 2021;373(6562):1464-1468.
10. King A. The search for better animal models of Alzheimer's disease. *Nature.* 2018;559(7715):S13-S15.
11. Onos KD, Sukoff Rizzo SJ, Howell GR, Sasner M. Toward more predictive genetic mouse models of Alzheimer's disease. *Brain Res Bull.* 2016;122:1-11.
12. Scarsee-Levie K, Sanchez PE, Lewcock JW. Leveraging preclinical models for the development of Alzheimer disease therapeutics. *Nat Rev Drug Discov.* 2020;19(7):447-462.
13. Ertekin-Taner N. Genetics of Alzheimer's disease: a centennial review. *Neurol Clin.* 2007;25(3):611-667. v.
14. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med.* 2016;18(5):421-430.
15. Sierksma A, Escott-Price V, De Strooper B. Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets. *Science.* 2020;370(6512):61-66.
16. Oblak AL, Forner S, Territo PR, et al. Model Organism Development and Evaluation for Late-Onset Alzheimer's disease: mODEL-AD. *Alzheimers Dement (N Y).* 2020;6(1):e12110.
17. Gulbranson DR, Ho K, Yu GQ, et al. Phenotypic differences between the Alzheimer's disease-related hAPP-J20 model and heterozygous Zbtb20 knockout mice. *eNeuro.* 2021;8(3):ENEURO.0089-21.2021..
18. Jankowsky JL, Zheng H. Practical considerations for choosing a mouse model of Alzheimer's disease. *Mol Neurodegener.* 2017;12(1):89.
19. Sasaguri H, Nilsson P, Hashimoto S, et al. APP mouse models for Alzheimer's disease preclinical studies. *EMBO J.* 2017;36(17):2473-2487.
20. Kotredes KP, Oblak A, Pandey RS, et al. Uncovering disease mechanisms in a novel mouse model expressing humanized APOEepsilon4 and Trem2\*<sup>R47H</sup>. *Front Aging Neurosci.* 2021;13:735524.
21. Preuss C, Pandey R, Piazza E, et al. A novel systems biology approach to evaluate mouse models of late-onset Alzheimer's disease. *Mol Neurodegener.* 2020;15(1):67.
22. Wan YW, Al-Ouran R, Mangleburg CG, et al. Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models. *Cell Rep.* 2020;32(2):107908.
23. Bis JC, Jian X, Kunkle BW, 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.
24. Bellenguez C, Kucukali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet.* 2022;54(4):412-436.
25. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet.* 2011;43(5):436-441.
26. Lee JH, Barral S, Reitz C. The neuronal sortilin-related receptor gene SORL1 and late-onset Alzheimer's disease. *Curr Neurol Neurosci Rep.* 2008;8(5):384-391.
27. Qureshi YH, Berman DE, Marsh SE, et al. The neuronal retromer can regulate both neuronal and microglial phenotypes of Alzheimer's disease. *Cell Rep.* 2022;38(3):110262.
28. Simoes S, Guo J, Buitrago L, et al. Alzheimer's vulnerable brain region relies on a distinct retromer core dedicated to endosomal recycling. *Cell Rep.* 2021;37(13):110182.
29. Young JE, Holstege H, Andersen OM, Petsko GA, Small SA. On the causal role of retromer-dependent endosomal recycling in Alzheimer's disease. *Nat Cell Biol.* 2023;25(10):1394-1397.
30. Vardarajan BN, Bruesegem SY, Harbour ME, et al. Identification of Alzheimer disease-associated variants in genes that regulate retromer function. *Neurobiol Aging.* 2012;33(9):2231 e15-2231 e30.
31. Oblak AL, Kotredes KP, Pandey RS, et al. Plcg2(M28L) interacts with high fat/high sugar diet to accelerate Alzheimer's disease-relevant phenotypes in mice. *Front Aging Neurosci.* 2022;14:886575.
32. Tsai AP, Dong C, Lin PB, et al. PLCG2 is associated with the inflammatory response and is induced by amyloid plaques in Alzheimer's disease. *Genome Med.* 2022;14(1):17.
33. Ferguson MC, Garland EM, Hedges L, et al. SHC2 gene copy number in multiple system atrophy (MSA). *Clin Auton Res.* 2014;24(1):25-30.
34. Sakai R, Henderson JT, O'Bryan JP, Elia AJ, Saxton TM, Pawson T. The mammalian ShcB and ShcC phosphotyrosine docking proteins function in the maturation of sensory and sympathetic neurons. *Neuron.* 2000;28(3):819-833.
35. Hagglund MG, Hellsten SV, Bagchi S, et al. Characterization of the transporterBOAT3 (Slc6a17) in the rodent central nervous system. *BMC Neurosci.* 2013;14:54.
36. Iqbal Z, Willemsen MH, Papon MA, et al. Homozygous SLC6A17 mutations cause autosomal-recessive intellectual disability with progressive tremor, speech impairment, and behavioral problems. *Am J Hum Genet.* 2015;96(3):386-396.
37. Li MD, Burns TC, Morgan AA, Khatri P. Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. *Acta Neuropathol Commun.* 2014;2:93.
38. Davies G, Armstrong N, Bis JC, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N = 53949). *Mol Psychiatry.* 2015;20(2):183-192.
39. Demange PA, Malanchini M, Mallard TT, et al. Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nat Genet.* 2021;53(1):35-44.
40. O'Connell KMS, Ouellette AR, Neuner SM, Dunn AR, Kaczorowski CC. Genetic background modifies CNS-mediated sensorimotor decline in the AD-BXD mouse model of genetic diversity in Alzheimer's disease. *Genes Brain Behav.* 2019;18(8):e12603.
41. Hu Q, Teng W, Li J, Hao F, Wang N. Homocysteine and Alzheimer's disease: evidence for a causal link from mendelian randomization. *J Alzheimers Dis.* 2016;52(2):747-756.
42. You M, Zhou X, Yin W, et al. The influence of MTHFR polymorphism on gray matter volume in patients with amnesic mild cognitive impairment. *Front Neurosci.* 2021;15:778123.
43. Rovelet-Lecrux A, Legallic S, Wallon D, et al. A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. *Eur J Hum Genet.* 2012;20(6):613-617.
44. Soto I, Grabowska WA, Onos KD, et al. Meox2 haploinsufficiency increases neuronal cell loss in a mouse model of Alzheimer's disease. *Neurobiol Aging.* 2016;42:50-60.
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120.
46. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2012;29(1):15-21.

47. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):323.
48. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-169.
49. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127.
50. Pandey RS, Graham L, Uyar A, Preuss C, Howell GR, Carter GW. Genetic perturbations of disease risk genes in mice capture transcriptomic signatures of late-onset Alzheimer's disease. *Mol Neurodegener*. 2019;14(1):50.
51. Allen M, Carrasquillo MM, Funk C, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data*. 2016;3:160089.
52. De Jager PL, Ma Y, McCabe C, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data*. 2018;5:180142.
53. Wang M, Beckmann ND, Roussos P, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data*. 2018;5:180185.
54. Milind N, Preuss C, Haber A, et al. Transcriptomic stratification of late-onset Alzheimer's cases reveals novel genetic modifiers of disease pathology. *PLoS Genet*. 2020;16(6):e1008775.
55. Greenwood AK, Montgomery KS, Kauer N, et al. The AD knowledge portal: a repository for multi-omic data on Alzheimer's disease and aging. *Curr Protoc Hum Genet*. 2020;108(1):e105.
56. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.
57. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS : A J Integrative Biology*. 2012;16(5):284-287.
58. Mostafavi S, Gaiteri C, Sullivan SE, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci*. 2018;21(6):811-819.
59. Neff RA, Wang M, Vatansever S, et al. Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. *Sci Adv*. 2021;7(2):eabb5398.
60. Zheng C, Xu R. Molecular subtyping of Alzheimer's disease with consensus non-negative matrix factorization. *PLoS One*. 2021;16(5):e0250278.
61. Lau V, Ramer L, Tremblay ME. An aging, pathology burden, and glial senescence build-up hypothesis for late onset Alzheimer's disease. *Nat Commun*. 2023;14(1):1670.
62. Reagan AM, Christensen KE, Graham LC, et al. The 677C >T variant in methylenetetrahydrofolate reductase causes morphological and functional cerebrovascular deficits in mice. *J Cereb Blood Flow Metab*. 2022;42(12):2333-2350.
63. Nelson RS, Dammer EB, Santiago JV, Seyfried NT, Rangaraju S. Brain cell type-specific nuclear proteomics is imperative to resolve neurodegenerative disease mechanisms. *Front Neurosci*. 2022;16:902146.
64. Rayaprolu S, Higginbotham L, Bagchi P, et al. Systems-based proteomics to resolve the biology of Alzheimer's disease beyond amyloid and tau. *Neuropsychopharmacology*. 2021;46(1):98-115.
65. Xiang X, Piers TM, Wefers B, et al. The Trem2 R47H Alzheimer's risk variant impairs splicing and reduces Trem2 mRNA and protein in mice but not in humans. *Mol Neurodegener*. 2018;13(1):49.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Sasner M, Preuss C, Pandey RS, et al. In vivo validation of late-onset Alzheimer's disease genetic risk factors. *Alzheimer's Dement*. 2024;20:4970-4984.  
<https://doi.org/10.1002/alz.13840>