

RESEARCH ARTICLE

Open Access

# Genetic perturbations of disease risk genes in mice capture transcriptomic signatures of late-onset Alzheimer's disease



Ravi S. Pandey<sup>1</sup>, Leah Graham<sup>2,3</sup>, Asli Uyar<sup>1</sup>, Christoph Preuss<sup>2</sup>, Gareth R. Howell<sup>2,3\*</sup> and Gregory W. Carter<sup>1,2,3\*</sup> 

## Abstract

**Background:** New genetic and genomic resources have identified multiple genetic risk factors for late-onset Alzheimer's disease (LOAD) and characterized this common dementia at the molecular level. Experimental studies in model organisms can validate these associations and elucidate the links between specific genetic factors and transcriptomic signatures. Animal models based on LOAD-associated genes can potentially connect common genetic variation with LOAD transcriptomes, thereby providing novel insights into basic biological mechanisms underlying the disease.

**Methods:** We performed RNA-Seq on whole brain samples from a panel of six-month-old female mice, each carrying one of the following mutations: homozygous deletions of *ApoE* and *Clu*; hemizygous deletions of *Bin1* and *Cd2ap*; and a transgenic *APOEε4*. Similar data from a transgenic *APP/PS1* model was included for comparison to early-onset variant effects. Weighted gene co-expression network analysis (WGCNA) was used to identify modules of correlated genes and each module was tested for differential expression by strain. We then compared mouse modules with human postmortem brain modules from the Accelerating Medicine's Partnership for AD (AMP-AD) to determine the LOAD-related processes affected by each genetic risk factor.

**Results:** Mouse modules were significantly enriched in multiple AD-related processes, including immune response, inflammation, lipid processing, endocytosis, and synaptic cell function. WGCNA modules were significantly associated with *ApoE*<sup>-/-</sup>, *APOEε4*, *Clu*<sup>-/-</sup>, and *APP/PS1* mouse models. *ApoE*<sup>-/-</sup>, *GFAP*-driven *APOEε4*, and *APP/PS1* driven modules overlapped with AMP-AD inflammation and microglial modules; *Clu*<sup>-/-</sup> driven modules overlapped with synaptic modules; and *APP/PS1* modules separately overlapped with lipid-processing and metabolism modules.

**Conclusions:** This study of genetic mouse models provides a basis to dissect the role of AD risk genes in relevant AD pathologies. We determined that different genetic perturbations affect different molecular mechanisms comprising AD, and mapped specific effects to each risk gene. Our approach provides a platform for further exploration into the causes and progression of AD by assessing animal models at different ages and/or with different combinations of LOAD risk variants.

**Keywords:** Alzheimer's disease, Animal models, Transcriptomic analysis

\* Correspondence: [gareth.howell@jax.org](mailto:gareth.howell@jax.org); [gregory.carter@jax.org](mailto:gregory.carter@jax.org)

<sup>2</sup>The Jackson Laboratory, Bar Harbor, ME, USA

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

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



## Background

Alzheimer's disease (AD) is the most common adult neurodegenerative disorder and accounts for around 60–80% of all dementia cases [1]. Neuropathologically, Alzheimer's disease is generally characterized by the presence of extracellular amyloid plaques composed of amyloid- $\beta$  (A $\beta$ ) surrounded by dystrophic neurites, neurofibrillary tangles (NFTs), and neuronal loss [2, 3]. Clinically, AD is classified into two subtypes: early onset with Mendelian inheritance, and late onset (or sporadic) AD [1, 4]. Early-onset Alzheimer's disease (EOAD) strikes prior to the age of 65 and accounts for approximately 5% of all AD cases, while the much more common late-onset Alzheimer's disease (LOAD) is diagnosed at later life stages (> 65 years) [2, 5]. In comparison to rare casual variants in three genes: *amyloid precursor protein* (*APP*), *presenilin 1* (*PSEN1*), and *presenilin 2* (*PSEN2*) that contribute to EOAD [1, 6, 7], the genetics factors influencing LOAD are complex due to the interplay of genetic and environmental factors that influence disease onset, progression and severity [8, 9]. Before the era of large-scale genome wide association studies, the e4 allele of the *apolipoprotein E* (*APOE*) gene was the only well-established major risk factor for LOAD, accounting for about 30% of genetic variance [10, 11]. *APOE $\epsilon$ 4* was inferred to have moderate penetrance [11] with homozygous carriers having a roughly five-times-increased risk compared to those who inherit only one e4 allele of *APOE* [1, 12].

Identification of new AD-related genes is important for better understanding of the molecular mechanisms leading to neurodegeneration [7]. Genome-wide association studies (GWAS) have identified dozens of additional genetic risk loci for LOAD, with candidate genes including clusterin (*CLU*), bridging integrator 1 (*BINI*), and CD2 associated protein (*CD2AP*) [1, 2, 7, 13]. These novel risk genes cluster in functional classes suggesting prominent roles in lipid processing, the immune system, and synaptic cell function such as endocytosis [1, 14]. Although these risk variants are often of small effect size, investigation of their functionality can reveal the biological basis of LOAD [1].

Despite recent advances in genetic and genomic resources to identify genetic risk factors, the disease mechanisms behind LOAD remain opaque. Most transgenic animal models are based on rare, early-onset AD genes which do not reflect the complete neuropathology or transcriptomic signatures of LOAD [15]. Although these transgenic mouse models were helpful to understand early molecular changes underlying A $\beta$  and tau pathology, the corresponding genetic factors only account for a small fraction of AD. Thus, animal models based on LOAD-associated genes are necessary to connect common genetic variation with LOAD transcriptomes.

To better understand the molecular mechanism underlying LOAD, we performed transcriptome profiling and analyses from brain hemispheres of 6 month old female mice carrying mutations in LOAD-relevant genes *ApoE*, *Clu*, *Bin1*, and *Cd2ap*. Weighted gene co-expression network analysis identified several mouse modules significantly driven by *ApoE*<sup>-/-</sup> and *Clu*<sup>-/-</sup> mouse strains. Moreover, we have compared mouse modules with human postmortem brain modules from the Accelerating Medicine's Partnership for AD (AMP-AD) to determine the AD relevance of risk genes. We observed enrichment of multiple AD-related pathways in these modules such as immune system, lipid metabolism, and neuronal system. This study of LOAD-relevant mice provides a basis to dissect the role of AD risk genes in AD pathologies.

## Methods

### Mouse strains and data generation

All mouse strains were obtained from The Jackson Laboratory and maintained in 12/12-h light/dark cycle (Table 1). All experiments were approved by the Animal Care and Use Committee at The Jackson Laboratory. RNA-Seq data were obtained from whole left hemisphere brain samples from a panel of six-month-old female mice carrying one of the following mutations in LOAD associated genes: homozygous deletion in *ApoE* and *Clu*; heterozygous deletion in *Cd2ap* and *Bin1*; and a transgenic *APOE $\epsilon$ 4* driven by a GFAP promoter on a *ApoE*<sup>-/-</sup> background (herein referred to as *ApoE*<sup>-/-</sup>, *Clu*<sup>-/-</sup>, *Cd2ap*<sup>+/-</sup>, *Bin1*<sup>+/-</sup> and *APOE $\epsilon$ 4*) (Table 1, [16–21]). There

**Table 1** Study population. Whole-brain left hemispheres were collected at 6 months of age from female mice

Model	Genetic Construct	JAX Strain name	JAX Stock #
B6	none; control	C57BL/6 J	000664
<i>APOE<math>\epsilon</math>4</i>	Homozygous transgene of human <i>APOE<math>\epsilon</math>4</i> allele with GFAP promoter	B6.Cg- <i>ApoE</i> <sup>tm1Un</sup> <i>Cdh18</i> <sup>Tg(GFAP-APOE<math>\epsilon</math>4)1Hol/J</sup>	004631
<i>ApoE</i> <sup>-/-</sup>	Homozygous gene knockout	B6.129P2- <i>ApoE</i> <sup>tm1Unc/J</sup>	002052
<i>Clu</i> <sup>-/-</sup>	Homozygous gene knockout	B6.Cg- <i>Clu</i> <sup>tm1Jakh/J</sup>	005642
<i>Bin1</i> <sup>+/-</sup>	Heterozygous gene knockout	B6.129S6- <i>Bin1</i> <sup>tm2Gcp/J</sup>	021145
<i>Cd2ap</i> <sup>+/-</sup>	Heterozygous gene knockout	B6.129X1- <i>Cd2ap</i> <sup>tm1Shaw/J</sup>	008907
<i>APP/PS1</i>	Homozygous transgenic	B6.Cg-Tg (APPswe,PSEN1dE9) 85Dbo/Mmjax	MMRC stock # 34,832-JAX

were six biological replicates for each late-onset model and control B6 mice. To minimize gene expression variation between mice, all mice in experimental cohorts were bred in the same mouse room and were aged together (to the extent possible). Cohorts were generated either by intercrossing heterozygous mice or in the case of *Bin1*<sup>+/-</sup> and *Cd2ap*<sup>+/-</sup> by crossing heterozygous mice to C57BL/6J (B6) mice, as homozygosity in these two genes is lethal. Data were also included from five whole left hemisphere brain samples from 6-month-old female mice from an early-onset AD model (APP/PS1, Table 1) [22] as well as seven additional B6 control replicates to account for batch effects.

For sample collection, mice were anesthetized with a lethal dose of ketamine/xylazine, transcardially perfused with 1X phosphate buffered saline (PBS), brains carefully dissected and hemisected in the midsagittal plane. The left hemisphere was snap frozen. RNA extraction was performed using TRIzol (Invitrogen, cat #: 15596026) according to manufacturer's instructions. Total RNA was purified from the aqueous layer using the QIAGEN miRNeasy mini extraction kit (QIAGEN) according to the manufacturer's instructions. RNA quality was assessed with the Bioanalyzer 2100 (Agilent Technologies). Poly(A) selected RNA-Seq sequencing libraries were generated using the TruSeq RNA Sample preparation kit v2 (Illumina) and quantified using qPCR (Kapa Biosystems). Using Truseq V4 SBS chemistry, all libraries were processed for 125 base pair (bp) paired-end sequencing on the Illumina HiSeq 2000 platform according to the manufacturer's instructions.

#### Quality control of RNA-Seq data

Sequence quality of reads was assessed using FastQC (v0.11.3, Babraham). Low-quality bases were trimmed from sequencing reads using Trimmomatic (v0.33) [23]. 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 were in range of 35–40 million reads.

#### Read alignments and gene expression

All RNA-Seq samples were mapped to the mouse genome (assembly 38) using ultrafast RNA-Seq aligner STAR (v2.5.3) [24]. First, a STAR index was built from mm10 reference sequence (Ensembl Genome Reference Consortium, build 38) for alignment, then STAR aligner output coordinate-sorted BAM files for each sample were mapped to mouse genome using this index. Gene expression was quantified in two ways, to enable multiple analytical methods: transcripts per million (TPM) using RSEM (v1.2.31) [25], and raw read counts using HTSeq-count (v0.8.0) [26].

#### Differential expression analysis

Differential expression in mouse models was assessed using Bioconductor package DESeq2 (v1.16.1) [27]. DESeq2 take raw read counts obtained from HTSeq-count as input and has its own normalization approach. The significance of differential expression was determined by the Benjamini-Hochberg corrected *p*-values. The threshold for significance was set to an adjusted *p* = 0.05. We included batch as a covariate in DESeq2 analysis to account for batch effect.

#### Principal component analysis and batch correction

We analyzed 48 RNA-Seq samples originating from three experimental batches: 1) all late-onset genetic models (*N* = 36); 2) one biological replicate of the *APP/PS1* strain with seven biological replicates of B6 control mice (*N* = 8); and 3) four additional biological replicates of *APP/PS1* (*N* = 4). First, we filtered out genes with TPM less than 10 for more than 90% of samples and then log-transformed to log<sub>2</sub>(TPM + 1) for downstream analysis. We then used the plotPCA function of Bioconductor package EDASeq [28] to observe the differences in distribution of samples due to batch effects. Finally, we implemented COMBAT [29] on above RNA-Seq datasets to remove known batch effects.

#### Network construction and mouse module detection

Modules (clusters) of correlated genes were identified using Weighted gene co-expression network analysis (WGCNA) implemented in R [30]. We used the step-by-step construction approach for network construction and module identification, which allows customization and alternate methods. The default unsigned network type was used, and a soft thresholding power of 8 was chosen to meet the scale-free topology criterion in the pickSoftThreshold function [31]. For module identification, WGCNA uses a topological overlap measure to compute network interconnectedness in conjunction with average linkage hierarchical clustering method. Modules correspond to branches of resulting clustering and are identified by cutting branches using dynamic tree cutting. To avoid small modules and ensure separation, we set the minimum module size to 30 genes and the minimum height for merging modules to 0.25. Each module is represented by the module eigengene (ME), defined as first principal component of the gene expression profiles of each module. Further, we have carried out one-way ANOVA (R function: aov) tests to determine differential expression between strains for each module eigengene. Modules with significant (*p* < 0.05) strain differences were analyzed for contributing strains using Tukey HSD (Tukey Honest Significant Differences, R function: TukeyHSD) for multiple pairwise-comparison between group means. The reported *p*-values were adjusted for multiple comparisons with Benjamini-Hochberg false discovery rate.

### Functional enrichment analysis

Functional annotations and enrichment analysis were performed using the R package clusterProfiler [32]. Gene Ontology terms and KEGG pathways enrichment analysis were performed using functions enrichGO and enrichKEGG, respectively, from the clusterProfiler package. The function compareCluster from this package was used to compare enriched functional categories of each gene module. The significance threshold for all enrichment analyses was set to 0.05 using Benjamini-Hochberg adjusted  $p$ -values.

### Calculation and significance of Jaccard indices

Jaccard indices were computed to find overlap strengths between mouse modules and AMP-AD human modules. The Jaccard index is measure of similarity between sample sets and defined as ratio of size of the intersection to the size of the union of two sample sets. Further, to test the significance of the Jaccard index for each pair of mouse-human module overlap, we performed permutation analysis by random sampling the equivalent number of genes in each mouse module from the union of all genes in the mouse modules. This was performed 10,000 times to generate null distributions of Jaccard index values. Cumulative  $p$ -values were then calculated empirically.

### Mouse-human orthologous genes

Mouse-human orthologous genes were identified using the genomic information on orthologous groups from the latest ENSEMBL build for the human genome version GRCh38. All orthologous gene relationships were retrieved from BioMart based on the Ensembl Compara Gene Tree comparison with the latest mouse genome build (biomart.org). Phylogenetic gene trees represent the evolutionary history of distinct gene families, which evolved from a common ancestor. Reconciliation of these gene trees against the mouse genome was used to distinguish duplication and speciation events across species, thus inferring distinct orthologue and paralogue gene pairs based on the method inferred by Cunningham et al. [33].

### Transcription factor analyses

Transcription factors in mouse module were identified using iRegulon (v1.3) [34] in Cytoscape (v3.2.0) [35] and the Enrichr webtool that contains ENCODE and ChEA consensus transcription factor annotations from Chip-X library [36].

### Human post-mortem brain cohorts and co-expression module identification

Whole-transcriptome data for human post-mortem brain tissue was obtained from the Accelerating Medicines

Partnership for Alzheimer Disease-(AMP-AD) consortium, which is a multi-cohort effort to harmonize genomics data from human LOAD patients. Harmonized co-expression modules from the AMP-AD data sets were obtained from Synapse (DOI: <https://doi.org/10.7303/syn11932957.1>). The human co-expression modules derive from three independent LOAD cohorts, including 700 samples from the ROS/MAP cohort, 300 samples from the Mount Sinai Brain bank and 270 samples from the Mayo cohort. A detailed description on post-mortem brain sample collection, tissue and RNA preparation, sequencing, and sample QC has been provided elsewhere [37–39]. As part of a transcriptome-wide meta-analysis to decipher the molecular architecture of LOAD, 30 co-expression modules from seven different brain regions across the three cohorts have been recently identified [40]. Briefly, Logsdon et al. identified 2978 co-expression modules using multiple techniques across the different regions after adjusting for co-variables and accounting for batch effects (<https://doi.org/10.7303/syn10309369.1>). A total of 660 co-expression modules were selected based on a specific enrichment in LOAD cases when compared to controls (<https://doi.org/10.7303/syn11914606>). Finally, multiple co-expression module algorithms were used to identify a set of 30 aggregate modules that were replicated by the independent methods [40].

### Correlation analysis

Standard gene set overlap tests are quick and easy, but do not account for direction of gene expression changes or coherence of changes across all genes in a module. To assess the directionality of genetic variants in model mice, we have computed the Pearson correlation across all genes in a given AMP-AD modules to determine human-mouse concordance.

To determine the effects of each genetic variant, we fit a multiple regression model as:

$$\log(expr) = \beta_0 + \sum_i \beta_i + \varepsilon$$

Where  $i$  denotes the genetic variants ( $ApoE^{-/-}$ ,  $APOE\epsilon4$ ,  $APP/PS1$ ,  $Bin1^{+/-}$ ,  $Cd2ap^{+/-}$ , and  $Clu^{-/-}$ ), and  $expr$  represents gene expression measured by RNA-Seq transcripts per million (TPM).

We have computed the Pearson correlation between log fold change gene expression in human AD cases versus controls ( $\text{Log}_2\text{FC}$  (AD/controls) and the effect of each mouse perturbation as determined by the linear model ( $\beta$ ) for the mouse orthologs genes within an AMP-AD module.  $\text{Log}_2\text{FC}$  values for human transcripts were obtained via the AMP-AD knowledge portal (<https://www.synapse.org/#!Synapse:syn11180450>).

Correlation coefficients were computed using `cor.test` function built in R as:

```
cor.test (log2FC (AD/control), β).
```

`cor.test` returns both the correlation coefficient and the significance level (*p*-value) of the correlation. Resulting *p*-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg (BH) procedure.

## Results

### Expression of target genes was modified by genetic perturbations

First, we have examined the relative expression (compared to control B6 mice) of LOAD associated genes to validate each strain. Expression of the mouse *ApoE* gene was downregulated in *ApoE*<sup>-/-</sup> mice ( $p < 1.00 \times 10^{-60}$ ) as well as in transgenic *APOEε4* ( $p < 1.00 \times 10^{-258}$ ) mice, which harbor human *APOE4* transcript driven by the GFAP promoter (Fig. 1a). Expression of *Clu* gene was also downregulated ( $p < 1.00 \times 10^{-30}$ ) in *Clu*<sup>-/-</sup> mice, while change in the expression of *Bin1* was significant but very small ( $\log_2FC = -0.3$ ;  $p = 8.72 \times 10^{-12}$ ) in *Bin1*<sup>+/-</sup> mice (Fig. 1a). The change in expression of *Cd2ap* gene was not significant ( $\log_2FC = -0.07$ ;  $p = 0.7$ ) in *Cd2ap*<sup>+/-</sup> mice (Fig. 1a). Overall, in each mouse strain, we observed significant downregulation in the expression of respective LOAD associated gene except in *Cd2ap*<sup>+/-</sup> models.

### Transcriptional signatures from mice carrying different mutations in LOAD-relevant genes clustered into different groups by PCA

Principal component analysis (PCA) was performed on batch-corrected, log-transformed, and mean-centered TPM for 10,704 genes (Methods). The first principal component accounted for 13% of total variance and separated models of different types of AD: LOAD associated models and EOAD associated *APP/PS1* transgenic models cluster separately (Fig. 1b), and thus might be affecting different AD-related processes. In other hand, within LOAD associated models, samples from the *Clu*<sup>-/-</sup> mice grouped together and separately from all other LOAD associated models in the second principal component (10% of variance) (Fig. 1b). Across all strains, *APOEε4* transgenic and *ApoE*<sup>-/-</sup> mice were most similar to each other (Fig. 1b). Hemizygous *Bin1*<sup>+/-</sup>, and *Cd2ap*<sup>+/-</sup> mice grouped closely to each another, suggesting functional similarity, and were the mutant strains in closest proximity to control (B6) mice (Fig. 1b).

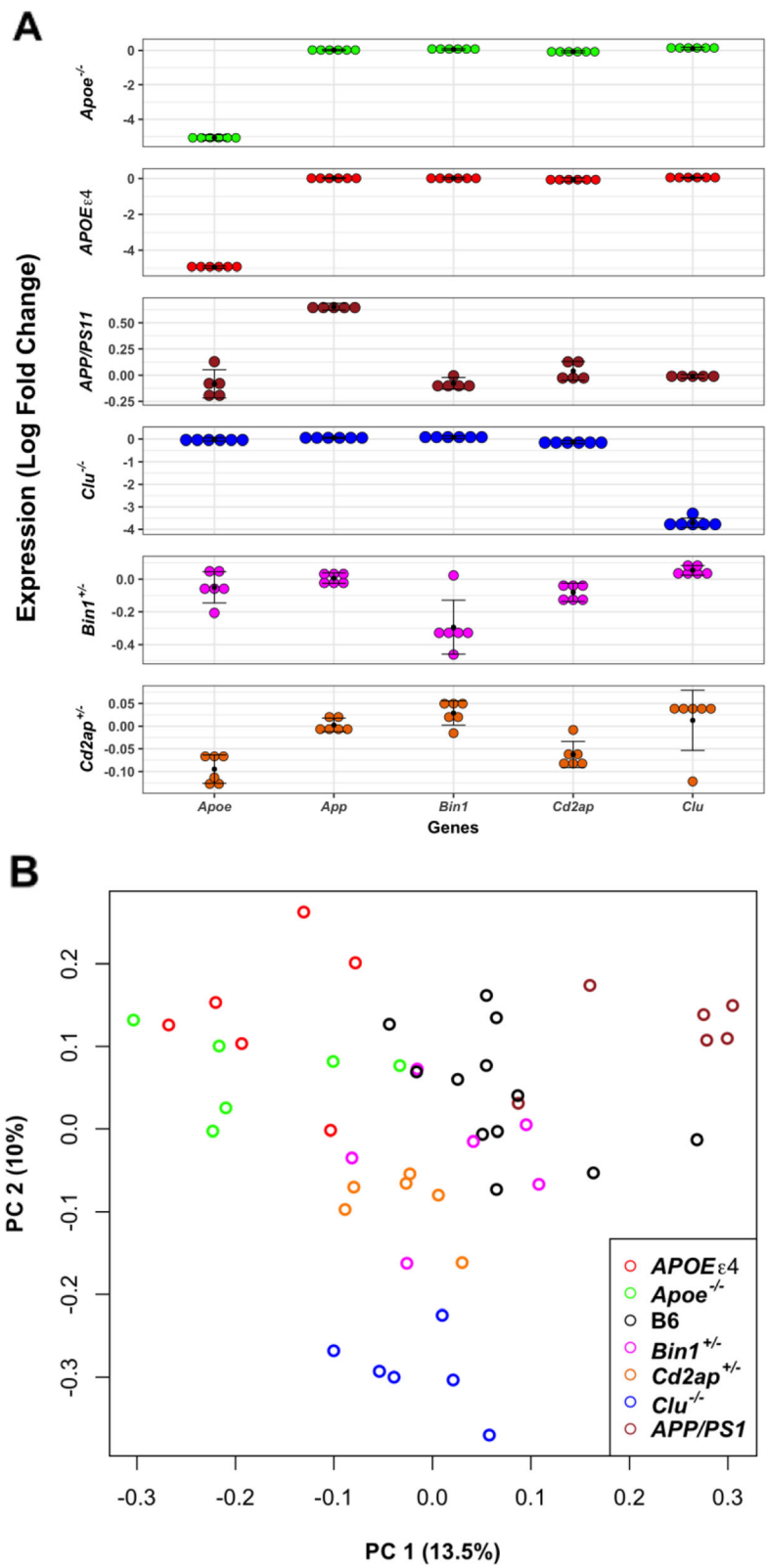
### Pathway analysis of differentially expressed genes identifies enrichment of different LOAD-related pathways in each mouse model

A total of 120 genes were significantly differentially expressed ( $p < 0.05$ ) in *APOEε4* transgenic mice, out of which 57 genes were upregulated and 63 genes were downregulated (Table 2;

Additional file 1: Table S1). We did not observe any pathway enrichment for differentially expressed genes in *APOEε4* transgenic mice. In *ApoE*<sup>-/-</sup> mice, 219 genes were identified significantly differentially expressed ( $p < 0.05$ ), 154 genes were upregulated and 65 genes were downregulated (Table 2; Additional file 1: Table S1). Inflammation/immune response related pathways were enriched in the upregulated list of DE genes in *ApoE*<sup>-/-</sup> mice (Additional file 2: Table S2), as well as osteoclast differentiation that is related to *TREM2* and *TYROBP*. We did not observe any enrichment for downregulated genes in *ApoE*<sup>-/-</sup> mice. In *Clu*<sup>-/-</sup> mice, a total of 1759 genes were identified significantly differentially expressed (762 genes were upregulated and 997 genes were downregulated) ( $p < 0.05$ ; Table 2; Additional file 1: Table S1). Pathway analysis of DE genes identified spliceosome, RNA transport, and ubiquitin mediated proteolysis as enriched pathways in downregulated genes of *Clu*<sup>-/-</sup> mice, while notch signaling as the enriched pathway in upregulated genes of *Clu*<sup>-/-</sup> mice (Additional file 2: Table S2). Only 16 and 34 genes were significantly differentially expressed ( $p < 0.05$ ) in *Bin1*<sup>+/-</sup> and *Cd2ap*<sup>+/-</sup> mice, respectively (Table 2; Additional file 1: Table S1). Pathway analysis identified endocytosis, phagosome, autoimmune, type I diabetes as enriched pathways in downregulated genes of *Cd2ap*<sup>+/-</sup> mice (Additional file 2: Table S2), while there was no pathway enrichment in upregulated genes of *Cd2ap*<sup>+/-</sup> mice. Downregulated genes of *Bin1*<sup>+/-</sup> mice were enriched in endocytosis and FC gamma R-mediated phagocytosis pathways (Additional file 2: Table S2). In the *APP/PS1* transgenic mice, 250 genes were differentially expressed (67 and 183 genes were up and downregulated, respectively) (Table 2). Pathway analysis of these DE genes identified ribosome, oxidative phosphorylation, and Alzheimer's disease as significantly enriched pathways (Additional file 2: Table S2).

### Co-expression network analysis identified mouse modules enriched for multiple LOAD-related pathways driven by *APOE* and *CLU* strains

Weighted gene co-expression network analysis (WGCNA) [30] identified 26 distinct modules of co-expressed genes (Fig. 2a, Additional file 3: Table S3). Further, we have carried out one-way ANOVA test followed by Tukey-HSD (see methods) to determine if there was differential expression between strains for each module eigengene. We identified that 13 out of 26 modules were significantly driven by one or more of *ApoE*<sup>-/-</sup>, *APOEε4*, *Clu*<sup>-/-</sup>, and *APP/PS1* models (Additional file 3: Table S3). Pathway enrichment analysis identified that multiple AD-related pathways were significantly enriched in these mouse modules. *ApoE*<sup>-/-</sup> mice were significantly associated with ivory module ( $N = 64$ ,  $p = 9.7 \times 10^{-6}$ ),



**Fig. 1** (See legend on next page.)

(See figure on previous page.)

**Fig. 1** Expression of LOAD associated genes in mice. **a** Expression of AD associated risk genes in LOAD-relevant mice and the *APP/PS1* transgenic model compared to B6 (control) mice. X-axis shows AD-associated risk genes and Y-axis represents average log fold change expression of above genes in genetically perturbed mice compare to controls. **b** Principal component analysis of batch corrected RNA-seq data from mouse strains. The *APOEε4* (red circle) and *Apoe* KO (green circle) samples are most similar to each other. Samples from mice carrying only one copy of either *Bin1* (magenta circle) or *Cd2ap* (orange circle) occupy similar regions, which might be due to their related functions. *APP/PS1* samples (brown circle) were separated from mice with late-onset perturbations by the first PC

while the skyblue3 ( $N = 80$ ,  $p = 4.6 \times 10^{-13}$ ) (Fig. 3; Fig. 4; Additional file 3: Table S3) module were significantly associated with both *Apoe*<sup>-/-</sup> and *APOEε4* strains. Pathway analysis identified that the ivory mouse module was enriched in inflammation and microglia related pathways such as osteoclast differentiation, staphylococcus aureus infection, phagosome, and endocytosis (Fig. 2b), implicating an important role of *Apoe* in inflammatory and microglia related functions [41–43]. Brown ( $N = 1778$ ,  $p = 3.1 \times 10^{-7}$ ), lightcyan1 ( $N = 1206$ ,  $p = 1.9 \times 10^{-5}$ ), black ( $N = 685$ ,  $p = 2.0 \times 10^{-2}$ ), plum1 ( $N = 80$ ,  $p = 1.0 \times 10^{-2}$ ), and brown4 ( $N = 55$ ,  $p = 0.04$ ) modules were significantly associated with *Clu*<sup>-/-</sup> (Fig. 3; Fig. 4; Additional file 3: Table S3). The steelblue module was driven by both *Clu*<sup>-/-</sup> ( $p = 5.02 \times 10^{-13}$ ) and *Cd2ap*<sup>+/-</sup> models ( $p = 9.5 \times 10^{-13}$ ) (Fig. 3; Fig. 4; Additional file 3: Table S3). These mouse modules were enriched in many different pathways particularly related to synaptic cell function, endocytosis, and RNA transport (Fig. 2b). This suggest the role of *Clu* gene in synaptic/neuronal related functions, which is in consistent with findings that reduced expression of *Clu* may results to aberrant synaptic development and neurodegeneration [44]. The darkorange2 ( $N = 61$ ,  $p = 1.0 \times 10^{-6}$ ), darkorange ( $N = 312$ ,  $p = 0.03$ ), orange ( $N = 142$ ,  $p = 4.64 \times 10^{-13}$ ), and lightgreen ( $N = 1456$ ,  $p = 1.0 \times 10^{-12}$ ) modules were found to be driven by *APP/PS1* (Fig. 3; Fig. 4; Additional file 3: Table S3). The lightyellow module ( $N = 163$ ) was observed to be associated with both *APP/PS1* ( $p = 8.7 \times 10^{-5}$ ) and *Clu*<sup>-/-</sup> mice ( $p = 1.4 \times 10^{-2}$ ), but more significantly with *APP/PS1* (Fig. 3; Fig. 4; Additional file 3: Table S3). *APP/PS1*-driven modules (lightyellow, lightgreen, darkorange2) were enriched in lipid-processing and metabolism related pathways (Fig. 2b). None of the modules were observed to be associated with *Bin1*<sup>+/-</sup> and *Cd2ap*<sup>+/-</sup> mice alone.

**Table 2** Differentially expressed genes by strain. Number of differentially expressed genes identified in each mouse strain compared to control mice (B6)

Mouse Model	Upregulated	Downregulated
<i>APOEε4</i>	57	63
<i>Apoe</i> <sup>-/-</sup>	154	65
<i>Clu</i> <sup>-/-</sup>	762	997
<i>Bin1</i> <sup>+/-</sup>	10	6
<i>Cd2ap</i> <sup>+/-</sup>	21	13
<i>APP/PS1</i>	67	183

### Comparison of mouse and AMP-AD modules

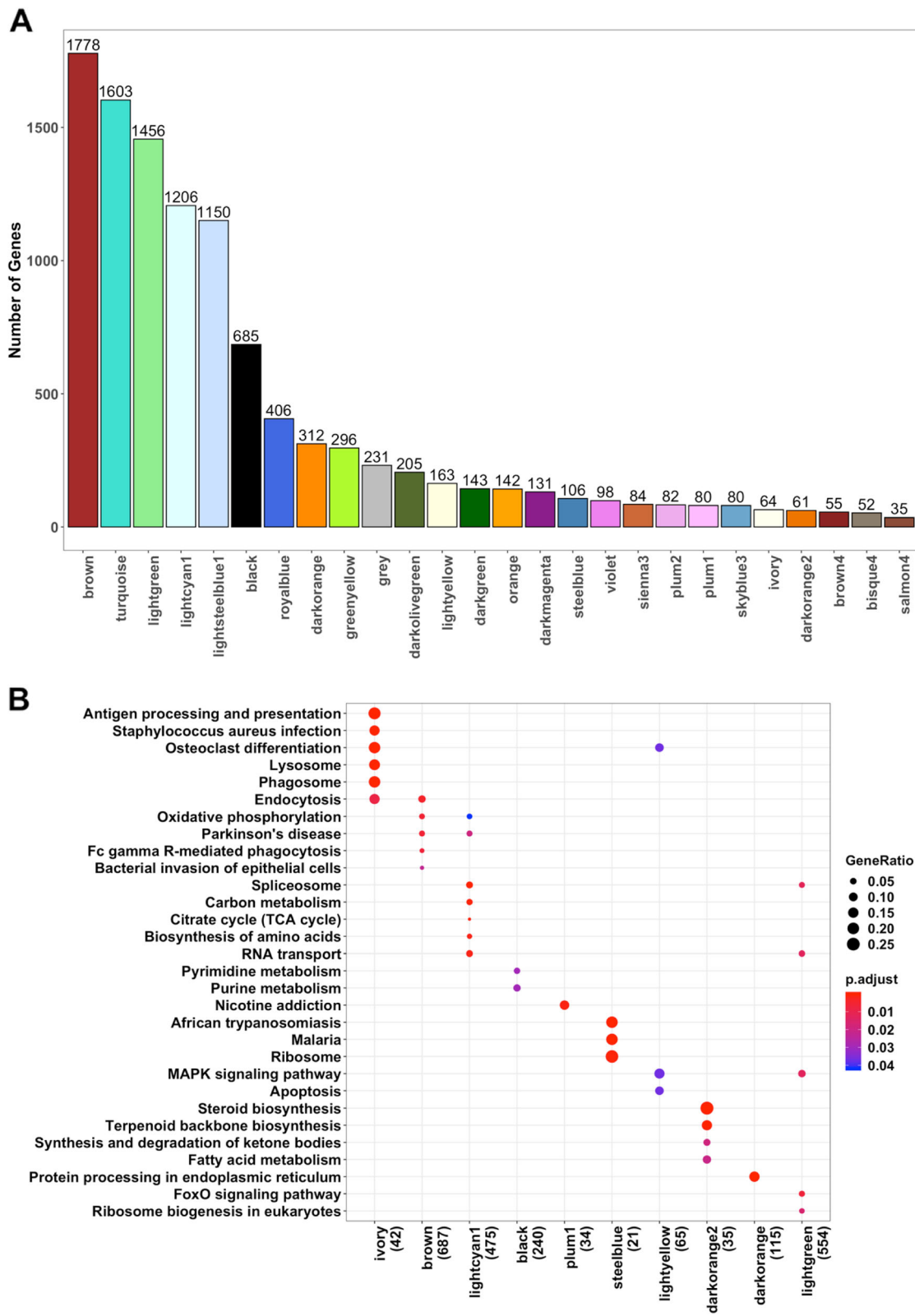
Finally, we compared mouse modules with the 30 human postmortem brain modules from the Accelerating Medicine’s Partnership for AD (AMP-AD). We computed Jaccard indices and its significance for each mouse - human module pair to identify which mouse module significantly overlap with human modules in order to identify AD-relevance of risk genes (Additional file 5: Table S5). Since each human module was derived from a specific brain region and study cohort, there are significant similarity between AMP-AD modules. Overlapping modules were therefore grouped into Consensus Clusters [40].

### *Apoe*-driven mouse module overlapped with AMP-AD inflammation and microglial consensus cluster

The ivory mouse module driven by *Apoe*<sup>-/-</sup> significantly overlapped with AMP-AD inflammation and microglia modules in Consensus Cluster B [40] (Fig. 4;  $p < 0.05$ ) and ranked among top ten mouse-human modules overlap (based on Jaccard indices) (Additional file 4: Table S4). These findings imply the significant role of *Apoe* in inflammation and microglia-related pathways. Furthermore, we identified that 22 genes were present in all AMP-AD microglial modules in Consensus Cluster B as well as in the *Apoe*<sup>-/-</sup>-driven ivory module (Fig. 5), as these genes were expressed from all human brain regions and therefore might be playing the important role in inflammation and microglia associated pathways. In order to identify transcriptional changes in these genes due to any AD-relevance genetic alteration, we assessed differential expression of these 22 genes in each mouse model (Additional file 1: Table S1). Nine out of these 22 genes (*TREM2*, *CSF1R*, *CIQA*, *CIQB*, *CIQC*, *PTGS1*, *AIF1*, *LAPTM5* and *LY86*) were significantly upregulated ( $p < 0.05$ ) in *Apoe*<sup>-/-</sup> mice and one gene (*TYROBP*) was significantly downregulated ( $p < 0.05$ ) in *Clu*<sup>-/-</sup> mice. Some of these genes (*TREM2*, *TYROBP*, *CIQA*, and *CSF1R*) have been associated with AD and reported to be potential drug targets (<https://agora.ampadportal.org/>). We did not find a significant overlap between the skyblue3 mouse module and any AMP-AD module.

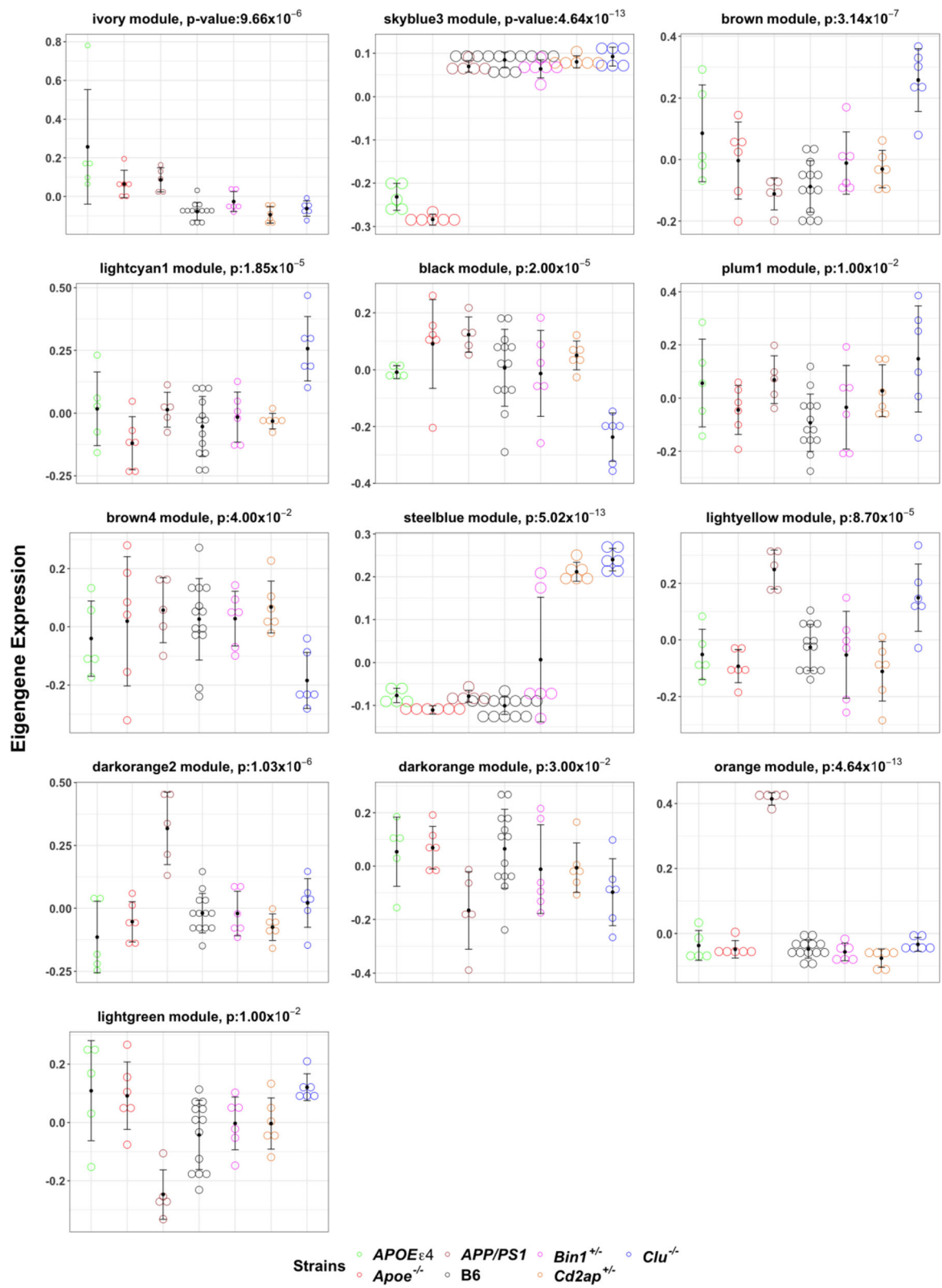
### *Clu*-driven modules overlapped with AMP-AD neuronal system consensus cluster

*Clu*<sup>-/-</sup>-driven mouse modules (brown, lightcyan1, and plum1) prominently overlapped with AMP-AD neuronal

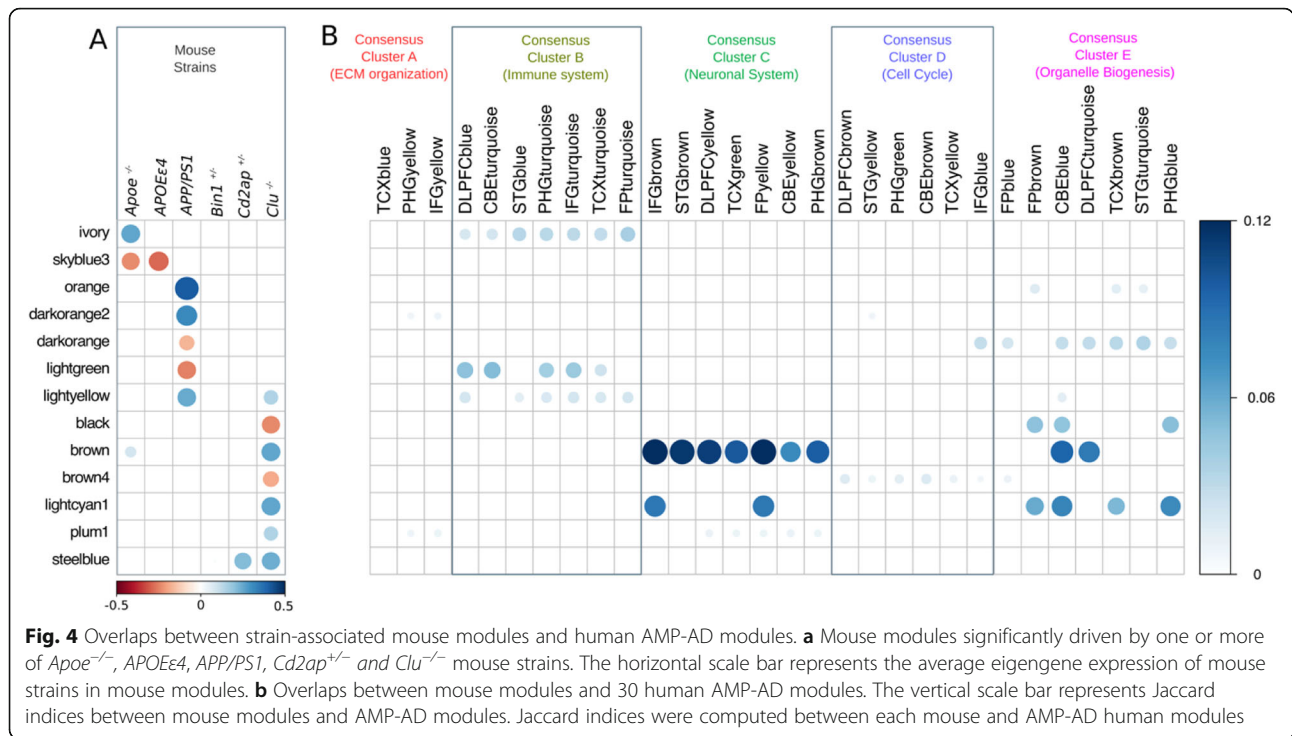


**Fig. 2** Mouse Modules Identified through WGCNA. **a** Twenty-six distinct mouse modules were identified from 10,704 mouse genes using WGCNA. Mouse modules of various sizes represented by different color names. **b** KEGG Pathway enrichment analysis ( $p < 0.05$ ) in mice using enrichKEGG function build under clusterprofiler R package





**Fig. 3** Mouse Modules Significantly driven by specific mouse strains. Expression of module eigengenes in mouse modules significantly driven by *Apoe*<sup>-/-</sup>, *APOE* $\epsilon$ 4, *Clu*<sup>-/-</sup> and *APP/PS1* mice (arbitrary units)



system modules in Consensus Cluster C [40], while black, lightcyan1, and brown modules overlapped with organelle biogenesis associated AMP-AD modules in Consensus Cluster E (Fig. 4;  $p < 0.05$ ). The *Clu*<sup>-/-</sup>-driven brown4 module showed association with cell cycle associated AMP-AD modules in Consensus Cluster D (Fig. 4;  $p < 0.05$ ). Also, we have observed that the top five mouse-human module overlaps (based on Jaccard indices) were between the brown module and AMP-AD neuronal system modules in Consensus Cluster C (Additional file 4: Table S4). Further, we also identified that 122 genes were common between the *Clu*<sup>-/-</sup>-driven brown mouse module and all AMP-AD neuronal system modules in Consensus Cluster C (Fig. 5b). We assessed these 122 genes for differential expression in each mouse strain (Additional file 1: Table S1) and found that 35 out of these 122 genes were differentially expressed (30 genes were upregulated and 5 genes were downregulated) only in *Clu*<sup>-/-</sup> mice, while three out of these 122 genes were differentially expressed only in *APP/PS1* transgenic mice (one gene was upregulated and two were downregulated). One of these 122 genes (*Syt7*) was upregulated in both *Clu*<sup>-/-</sup> mice and the *APP/PS1* transgenic mice. These findings support the likely role of *CLU* in neuronal function.

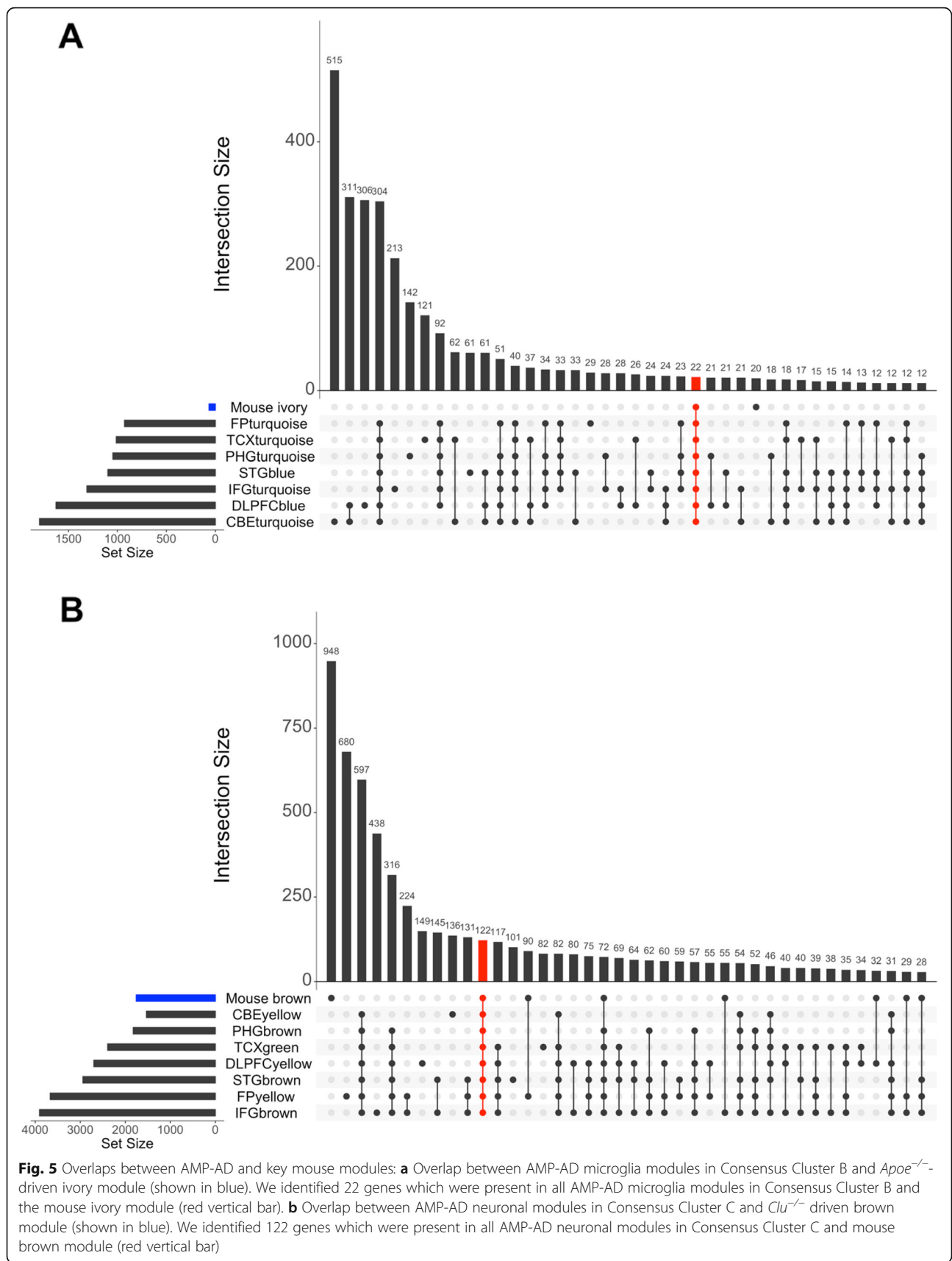
**APP/PS1-driven modules overlapped with inflammation, lipid-processing, and metabolism AMP-AD modules**

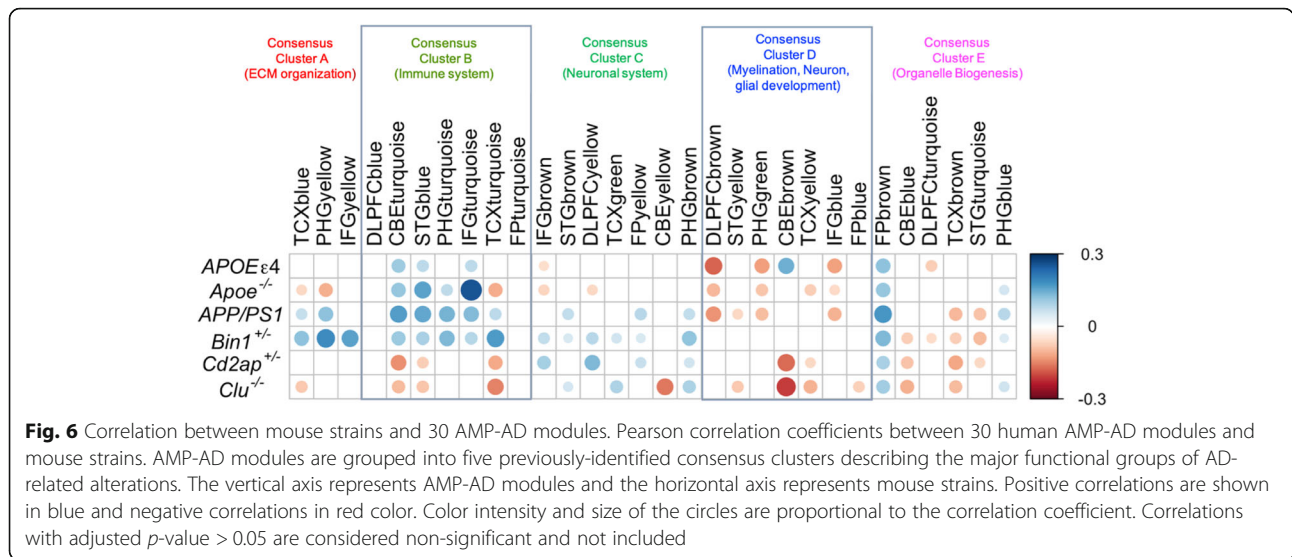
The *APP/PS1*-driven orange and darkorange modules overlapped with lipid processing and metabolism associated

AMP-AD modules in Consensus Cluster E, the lightgreen module overlapped with immune system modules in Consensus Cluster B, and the lightyellow module overlapped with both microglia and organelle biogenesis related AMP-AD modules in Consensus Clusters B and E, respectively (Fig. 4;  $p < 0.05$ ). We found significant overlap for the darkorange2 mouse module with AMP-AD modules in Consensus Cluster E, which are in turn enriched in organelle biogenesis related pathways (Fig. 4;  $p < 0.05$ ).

**Correlation analysis provides directional coherence between mouse models and AMP-AD consensus clusters**

The gene set overlap analysis identified mouse modules that are significantly overlapped with AMP-AD modules, but it does not assess directional coherence between AMP-AD modules and the effects of genetic perturbations in mice. To address this issue, we computed the Pearson correlation between log fold change gene expression in human AD cases versus controls ( $\text{Log}_2\text{FC}$ ) and the effect of each mouse perturbation on mouse orthologs as determined by the linear model ( $\beta$ ) for the genes within an AMP-AD module. *Apoe*<sup>-/-</sup> and *APOEε4* mice showed significant positive correlation ( $r = 0.1-0.3$ ,  $p < 0.05$ ) with immune associated AMP-AD modules in Consensus Cluster B and significant negative correlation ( $r = -0.05$ ,  $p < 0.05$ ) with AMP-AD neuronal modules in Consensus Cluster C (Fig. 6). Furthermore, *Clu*<sup>-/-</sup> and *Cd2ap*<sup>+/-</sup> mice showed significantly positive association ( $r = 0.1$ ,  $p < 0.05$ ) with AMP-AD neuronal modules in Consensus Cluster C and negative correlation ( $r = -0.15$ ,





$p < 0.05$ ) with AMP-AD immune related modules in Consensus Cluster B (Fig. 6). *Bin1*<sup>-/-</sup> and *APP/PS1* mice showed significant positive correlation ( $r = 0.1–0.2$ ,  $p < 0.05$ ) with AMP-AD immune response associated modules in Consensus Cluster B as well as AMP-AD neuronal modules in Consensus Cluster C. The cell cycle and RNA non-mediated decay pathways enriched AMP-AD modules in Consensus Cluster D were significantly negatively correlated ( $r = -0.2$ ,  $p < 0.05$ ) with *Apoe*<sup>-/-</sup>, *APOEε4*, *Clu*<sup>-/-</sup>, *Cd2ap*<sup>+/-</sup>, and *APP/PS1* mice, but *Bin1*<sup>+/-</sup> mice showed significant positive correlation ( $r = 0.11$ ,  $p > 0.05$ ) with AMP-AD cell cycle module in the cerebellum (Fig. 6). Most of the AMP-AD modules in Consensus Cluster E that is enriched for organelle biogenesis associated pathways showed significant negative correlation ( $r = -0.1$ ,  $p < 0.05$ ) with all strains except the *Apoe*<sup>-/-</sup> models ( $r = 0.12$ ,  $p < 0.05$ ), while the AMP-AD modules of Consensus Cluster E in the frontal pole (FPbrown) and parahippocampal gyrus (PHGblue) showed significant positive association ( $r = 0.05–0.2$ ,  $p < 0.05$ ) with all strains (Fig. 6).

#### **Apoe-associated modules are enriched in SPI1 regulatory targets**

Transcription regulation play an important role in the initiation and progression of AD [45]. Our results provide evidence of the AD relevance of risk genes, but it is also important to identify the regulatory elements and transcriptional factors that regulate the expression of these genes for molecular dissection of disease etiology [45, 46]. Recent study have shown that *APOEε4* genotype suppress transcription of autophagy mRNA's by competing with transcription factor EB for binding to coordinated lysosomal expression and regulation(-CLEAR) DNA motifs [47]. TFs were identified for

each module with high normalized enrichment scores (NES  $\geq 4$ ) from iRegulon (Methods), which correspond to an estimated false discovery rate of less than 0.01 [34] (Additional file 5: Table S5). The SPI1 transcription factor was enriched for regulatory targets in the *Apoe*<sup>-/-</sup> driven ivory and skyblue3 modules (Table S6). It has been previously reported that SPI1 responds to inflammatory signals and regulates genes that can contribute to neurodegeneration in AD [48]. We also observed that transcription factors from ELF, ETS, TCF, PEA3, GABP, and ERF sub-family of the E26 transformation-specific (ETS) family were enriched in the *Clu*<sup>-/-</sup>-driven modules (Additional file 5: Table S5). ETS-domain proteins play a role in the regulation of neuronal functions [49]. ETS family members ELK1 and ETS1 have been reported to expressed in neuronal cells and activate transcription of early onset AD candidate gene *PSEN1* [45, 46]. This transcription factor analysis was based solely on bioinformatics and general data resources, and therefore require experimental validation in specific AD-related contexts. Nevertheless, understanding the role of these and other transcription factors in regulating AD associated genes can provide a molecular basis for potential therapeutic development.

#### **Conclusions**

In this study, we have performed transcriptomic analysis of mouse strains carrying different mutations in genes linked to AD by GWAS to better understand the genetics and basic biological mechanisms underlying LOAD. We have also performed a comprehensive comparison at the transcriptomic level between mouse strains and human postmortem brain data from LOAD patients. This study of LOAD-relevant mouse models provides a basis to dissect the role of AD risk genes in relevant AD pathologies. We determined that different genetic perturbations affect

different molecular mechanisms underlying AD, and mapped specific effects to each risk gene. In our study, we observed that *ApoE*<sup>-/-</sup> and *Clu*<sup>-/-</sup> mice at the relatively early age of 6 months show transcriptomic patterns similar to human AD cases. Pathway analysis suggested that *ApoE*<sup>-/-</sup> driven mouse modules specifically affect inflammation/microglia related pathways, while *Clu*<sup>-/-</sup> driven mouse modules have affected neurosignaling, lipid transport, and endocytosis related pathways. These findings suggest that *APOE* and *CLU* risk genes are associated with distinct AD-related pathways. We have also identified that 22 genes were co-expressed in the *ApoE*<sup>-/-</sup>-driven ivory mouse module and in AMP-AD modules from all human brain regions in Consensus Cluster B that were enriched in inflammation and microglia associated pathways. Further, some of these genes (*Tyrobp*, *Trem2*, and *Csf1r*) were differentially expressed in *ApoE*<sup>-/-</sup> mice. Previous studies have already implicated the role of *TREM2* in AD susceptibility due to association of heterozygous rare variants in *TREM2* with elevated risk of AD [50] and higher cortical *TREM2* RNA expression with increased amyloid pathology [51]. *TYROBP* has been also previously reported as key regulator of immune/microglia associated pathways, which is strongly associated with LOAD pathology [14]. These genes have been also proposed as potential drug targets (<https://agora.ampad-portal.org/>) and our findings supports the role of these genes with pathophysiology of LOAD.

Correlation analysis also identified that mice carrying different mutations capture distinct transcriptional signatures of human LOAD. Moreover, we have observed contrasting correlations of *APOEε4*, *ApoE*<sup>-/-</sup>, and *Clu*<sup>-/-</sup> mice with AMP-AD modules, implicating that these genetic perturbations might affect LOAD risk through different physiological pathways. It has been speculated that absence of both *ApoE* and *Clu* resulted in accelerated disease onset, and more extensive amyloid deposition in the *PDAPP* transgenic mice brain [52]. Furthermore, *APOE* and *CLU* proteins interact with amyloid-beta (Aβ) and regulates its clearance from brain. In particular, the presence of *CLU* and the *APOEε2* allele promotes Aβ clearance from brain, whereas *APOEε4* reduces the clearance process [44]. These observations also suggest a protective role of *CLU* [44, 53, 54], consistent with our transcriptome-based anti-correlation of *Clu*<sup>-/-</sup> mice LOAD modules (Fig. 6). Understanding of the complex interaction between these genes is essential to interpret molecular mechanisms underlying AD. Hence, it would be interesting to analyze mice models carrying different combinations of genetic variants.

We did not observe any striking responses in brain gene expression patterns in *APOEε4*, *Bin1*<sup>+/-</sup>, and *Cd2ap*<sup>+/-</sup> mice based on the small subset of differentially expressed genes, as opposed to effects observed in the *Clu*<sup>-/-</sup> and *ApoE*<sup>-/-</sup> models (Table 2). Nor did we observe any mouse

modules significantly driven by these perturbations alone. We note that these models were limited to heterozygous mutations in *Bin1* and *Cd2ap* and astrocyte-specific expression of *APOEε4*. The latter limitation may be insufficient to capture the role of *APOE* variants in microglia and disease risk [55]. However, our human-mouse comparison revealed significant correlation of these mouse models with multiple human-derived AMP-AD co-expression modules. We interpret this as these models expression global changes relevant to human cases, while few individual gene expression changes are large enough to be captured by differential expression analysis. This may suggest region-specific and/or cell-specific signals that are diluted by our bulk whole-brain analysis. We have observed that *Bin1*<sup>+/-</sup> models were significantly associated with multiple AMP-AD co-expression modules, which in turn were enriched in immune response, inflammation, and synaptic functioning pathways, which is in concordance with other studies [56, 57]. Furthermore, *Cd2ap*<sup>+/-</sup> mice captured similar human AD signatures as *Clu*<sup>-/-</sup> mice, it may be due to their involvement in similar pathways like blood-brain carrier, and loss of function in *Cd2ap* may contribute to genetic risk of AD by facilitating age related blood-brain barrier breakdown [58]. In-depth investigation of the functional variants of these high-risk AD genes will be essential to evaluate their role in LOAD onset and progression.

The molecular mechanisms of AD driven by rare mutations in *APP*, *PSEN1*, and *PSEN2* are relatively well understood, but the functional impact of LOAD associated risk factors still remain unclear. Although early-onset models have provided critical insights into amyloid accumulation, pathology, and clearance, they do not reflect the full transcriptomic signatures and complete neuropathology of LOAD. Indeed, the primary transcriptomic signatures from mice carrying major early-onset and late-onset genetic factors are distinct (Fig. 1b), although our functional analysis in the context of human disease modules also detected some common neuroimmune effects (Fig. 6). Many of these differences are likely due to the presence of amyloid deposition in *APP/PS1* mice that drives gene expression signatures [22]. In this context, the common neuroimmune response suggests similar signatures arising in the absence of amyloid. It therefore remains unclear whether the relatively uncommon EOAD cases and the more common late-onset AD cases proceed through similar disease mechanisms. Understanding these distinctions motivates the development and characterization of new models for the late onset of AD. In this study, we have analyzed mice carrying alterations in LOAD candidate genes and found that different AD risk genes are associated with different AD-related pathways. Our approach provides a platform for further exploration into the causes and progression of

LOAD by assessing animal models at different ages and/or with different combinations of LOAD risk variants. This study highlighted that implementing state-of-the-art approaches to generate and characterize LOAD-associated mouse models might be helpful to identify variants and pathways to understand complete AD mechanisms and ultimately develop effective therapies for AD.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13024-019-0351-3>.

**Additional file 1: Table S1.** Differentially expressed genes in LOAD mouse strains. The attached table depicts the differentially expressed genes in LOAD mouse strains compared to C7BL/6 J mice.

**Additional file 2: Table S2.** KEGG pathway annotation in LOAD mouse strains. The attached table depicts the KEGG pathway annotations for the up and down-regulated genes in LOAD mouse strains compared to C7BL/6 J mice.

**Additional file 3: Table S3.** Mouse modules of co-expressed genes. Summaries of the 26 mouse modules of co-expressed genes. In sheet 1, genes in each mouse modules were listed. Sheet 2 depicts mouse modules were observed to be significantly ( $p < 0.05$ ) driven by at-least one of the mouse strains. Sheets 3 and 4 illustrate enriched KEGG pathways and enriched GO terms in each mouse modules.

**Additional file 4: Table S4.** Jaccard indices between Mouse and AMP-AD modules. The attached table contains Jaccard indices and its significance ( $p$ -value) for each mouse-human module pair.

**Additional file 5: Table S5.** Transcriptional factor annotations in LOAD mouse modules. The attached table illustrate transcriptional factor enriched in each 26 mouse modules of co-expressed genes.

## Abbreviations

AD: Alzheimer's disease; AMP-AD: Accelerating Medicines Partnership for Alzheimer's Disease; B6: C57BL/6 J; LOAD: Late-onset Alzheimer's disease; RNA-Seq: RNA sequencing; ROSMAP: Religious Orders Study/Memory and Aging Project

## Acknowledgements

We thank the many institutions and their staff that provided support for this study and who were involved in this collaboration. We would like to acknowledge Ben Logsdon for curating human brain data.

## Authors' contributions

RSP performed the gene expression analyses in human and mouse brain tissue. AU and CP contributed to the analyses. LG and GRH performed mouse experiments. GWC and GRH supervised and designed the project. RSP, GWC, and CP wrote the manuscript. All authors read and approved the final manuscript.

## Funding

This study was supported by the National Institutes of Health grants AG054345 and AG055104.

## Availability of data and materials

The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (<https://doi.org/10.7303/syn2580853>). ROSMAP Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute. Mayo RNA-Seq Study data were provided by the following sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilufer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from

the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center, and 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, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's Disease CoreCenter), 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. MSBB data were generated from postmortem 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. Mouse RNA-Seq data from the MODEL-AD consortium is available through Synapse via the AMP-AD knowledge portal ([www.synapse.org/#!Synapse:syn15811463](http://www.synapse.org/#!Synapse:syn15811463)).

## Ethics approval

All experiments involving mice were conducted in accordance with policies and procedures described in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and were approved by the Institutional Animal Care and Use Committee at The Jackson Laboratory.

## Consent for publication

All authors have approved of the manuscript and agree with its submission.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. <sup>2</sup>The Jackson Laboratory, Bar Harbor, ME, USA. <sup>3</sup>Sackler School of graduate Biomedical Sciences, Tufts University, Boston, MA, USA.

Received: 30 August 2019 Accepted: 11 December 2019

Published online: 26 December 2019

## References

- Bettens K, Sleegers K, Van Broeckhoven C. Genetic insights in Alzheimer's disease. *Lancet Neurol.* 2013;12(1):92–104.
- Tanzi RE. The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med.* 2012;2(10):a006296.
- Bouter Y, et al. Deciphering the molecular profile of plaques, Memory Decline and Neuron Loss in Two Mouse Models for Alzheimer's Disease by Deep Sequencing. *Front Aging Neurosci.* 2014;6:75.
- Rossor MN, et al. Neurochemical characteristics of early and late onset types of Alzheimer's disease. *Br Med J (Clin Res Ed).* 1984;288(6422):961–4.
- Jackson HM, et al. Clustering of transcriptional profiles identifies changes to insulin signaling as an early event in a mouse model of Alzheimer's disease. *BMC Genomics.* 2013;14:831.
- Bertram L, Tanzi RE. Chapter 3 - The Genetics of Alzheimer's Disease. In: Teplow DB, editor. *Progress in Molecular Biology and Translational Science*; 2012, Academic Press. p. 79–100.
- Bagyinszky E, et al. The genetics of Alzheimer's disease. *Clin Interv Aging.* 2014;9:535–51.
- Ertekin-Taner N. Genetics of Alzheimer's Disease: A Centennial Review. *Neurol Clin.* 2007;25(3):611.
- Bertram L, Lill CM, Tanzi RE. The genetics of Alzheimer disease: Back to the future. *Neuron.* 2010;68(2):270–81.
- Corder EH, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science.* 1993;261(5123):921.
- Genin E, et al. APOE and ALZHEIMER disease: a major gene with semi-dominant inheritance. *Mol Psychiatry.* 2011;16(9):903–7.
- Alzheimer's Association. 2018 Alzheimer's disease facts and figures. *Alzheimer's Dement.* 2018;14(3):367–429.
- Lambert J-C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45(12):1452–8.

14. Zhang B, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013;153(3):707–20.
15. Wan Y-W, et al. Functional dissection of Alzheimer's disease brain gene expression signatures in humans and mouse models. *bioRxiv*. 2019;1:506873.
16. Fadale DJ, et al. Mutant presenilins specifically elevate the levels of the 42 residue  $\beta$ -amyloid peptide in vivo: evidence for augmentation of a 42-specific  $\gamma$  secretase. *Hum Mol Genet*. 2003;13(2):159–70.
17. Hong T, et al. Cardiac BIN1 folds T-tubule membrane, controlling ion flux and limiting arrhythmia. *Nat Med*. 2014;20(6):624–32.
18. McLaughlin L, et al. Apolipoprotein J/clusterin limits the severity of murine autoimmune myocarditis. *J Clin Invest*. 2000;106(9):1105–13.
19. Piedrahita JA, et al. Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells. *Proc Natl Acad Sci*. 1992;89(10):4471.
20. Shih N-Y, et al. Congenital Nephrotic syndrome in mice lacking CD2-associated protein. *Science*. 1999;286(5438):312.
21. Sun Y, et al. Glial Fibrillary acidic protein–Apolipoprotein E (apoE) transgenic mice: astrocyte-specific expression and differing biological effects of astrocyte-secreted apoE3 and apoE4 lipoproteins. *J Neurosci*. 1998;18(9):3261.
22. Chintapaludi SR, et al. Staging Alzheimer's disease in the brain and retina of B6.APP/PS1 mice by transcriptional profiling. *bioRxiv*. 2019;1:741421.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
24. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
25. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):323.
26. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–9.
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
28. Risso D, et al. GC-content normalization for RNA-Seq data. *BMC Bioinformatics*. 2011;12:480.
29. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–27.
30. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559.
31. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis, in *Statistical Applications in Genetics and Molecular Biology*; 2005.
32. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
33. Cunningham F, et al. Ensembl 2015. *Nucleic Acids Res*. 2015;43(Database issue):D662–9.
34. Janky Rs, et al. iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput Biol*. 2014;10(7):e1003731.
35. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
36. Chen EY, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
37. Allen M, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Scientific Data*. 2016;3:160089.
38. De Jager PL, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Scientific Data*. 2018;5:180142.
39. Wang M, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Scientific data*. 2018;5:180185.
40. Logsdon B, et al. Meta-analysis of the human brain transcriptome identifies heterogeneity across human AD coexpression modules robust to sample collection and methodological approach. *bioRxiv*. 2019;1:510420.
41. Kim J, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. *Neuron*. 2009;63(3):287–303.
42. Rodriguez GA, et al. Human APOE4 increases microglia reactivity at A $\beta$  plaques in a mouse model of A $\beta$  deposition. *J Neuroinflammation*. 2014;11:111.
43. Ulrich JD, et al. ApoE facilitates the microglial response to amyloid plaque pathology. *J Exp Med*. 2018;215(4):1047.
44. Nelson AR, Sagare AP, Zlokovic BV. Role of clusterin in the brain vascular clearance of amyloid- $\beta$ . *Proc Natl Acad Sci*. 2017;114(33):8681.
45. Chen X-F, et al. Transcriptional regulation and its misregulation in Alzheimer's disease. *Mol Brain*. 2013;6:44.
46. Theuns J, Van Broeckhoven C. Transcriptional regulation of Alzheimer's disease genes: implications for susceptibility. *Hum Mol Genet*. 2000;9(16):2383–94.
47. Parcon PA, et al. Apolipoprotein E4 inhibits autophagy gene products through direct, specific binding to CLEAR motifs. *Alzheimer's Dementia*. 2018;14(2):230–42.
48. Citron BA, et al. Transcription factor Sp1 inhibition, memory, and cytokines in a mouse model of Alzheimer's disease. *Am J Neurodegenerative Dis*. 2015;4(2):40–8.
49. Sharrocks AD. The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol*. 2001;2:827.
50. Guerreiro R, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med*. 2013;368(2):117–27.
51. Chan G, et al. Modulation of TREM2 by CD33: a protein QTL study integrates Alzheimer loci in human monocytes. *Nat Neurosci*. 2015;18(11):1556–8.
52. DeMattos RB, et al. ApoE and Clusterin cooperatively suppress A $\beta$  levels and deposition: evidence that ApoE regulates extracellular A $\beta$  metabolism in vivo. *Neuron*. 2004;41(2):193–202.
53. Roussotte FF, et al. Combined Effects of Alzheimer Risk Variants in the  $\epsilon$  &  $\kappa$ ;  $\epsilon$  &  $\eta$ ;  $\kappa$  &  $\eta$ ;  $\epsilon$  &  $\zeta$ ;  $\kappa$  &  $\zeta$ ; ApoE &  $\epsilon$ ;  $\kappa$ ;  $\eta$ ;  $\zeta$ ; Genes on Ventricular Expansion Patterns in the Elderly. *J Neurosci*. 2014;34(19):6537.
54. Calero M, et al. Apolipoprotein J (clusterin) and Alzheimer's disease. *Microsc Res Tech*. 2000;50(4):305–15.
55. Krasemann S, et al. The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. *Immunity*. 2017;47(3):566–581.e9.
56. Tan M-S, Yu J-T, Tan L. Bridging integrator 1 (BIN1): form, function, and Alzheimer's disease. *Trends Mol Med*. 2013;19(10):594–603.
57. Karch CM, et al. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One*. 2012;7(11):e50976.
58. Cochran JN, et al. The Alzheimer's disease risk factor CD2AP maintains blood-brain barrier integrity. *Hum Mol Genet*. 2015;24(23):6667–74.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

