


METHODOLOGY

Open Access



A novel systems biology approach to evaluate mouse models of late-onset Alzheimer's disease

Christoph Preuss^{1†}, Ravi Pandey^{1†}, Erin Piazza², Alexander Fine¹, Asli Uyar¹, Thanneer Perumal², Dylan Garceau¹, Kevin P. Kotredes¹, Harriet Williams¹, Lara M. Mangravite³, Bruce T. Lamb⁴, Adrian L. Oblak⁴, Gareth R. Howell¹, Michael Sasner¹, Benjamin A. Logsdon³, the MODEL-AD Consortium and Gregory W. Carter^{1*} 

Abstract

Background: Late-onset Alzheimer's disease (LOAD) is the most common form of dementia worldwide. To date, animal models of Alzheimer's have focused on rare familial mutations, due to a lack of frank neuropathology from models based on common disease genes. Recent multi-cohort studies of postmortem human brain transcriptomes have identified a set of 30 gene co-expression modules associated with LOAD, providing a molecular catalog of relevant endophenotypes.

Results: This resource enables precise gene-based alignment between new animal models and human molecular signatures of disease. Here, we describe a new resource to efficiently screen mouse models for LOAD relevance. A new NanoString nCounter[®] Mouse AD panel was designed to correlate key human disease processes and pathways with mRNA from mouse brains. Analysis of the 5xFAD mouse, a widely used amyloid pathology model, and three mouse models based on LOAD genetics carrying APOE4 and TREM2*R47H alleles demonstrated overlaps with distinct human AD modules that, in turn, were functionally enriched in key disease-associated pathways. Comprehensive comparison with full transcriptome data from same-sample RNA-Seq showed strong correlation between gene expression changes independent of experimental platform.

Conclusions: Taken together, we show that the nCounter Mouse AD panel offers a rapid, cost-effective and highly reproducible approach to assess disease relevance of potential LOAD mouse models.

Background

Late-onset Alzheimer's disease (LOAD) is the most common cause of dementia worldwide [1]. LOAD presents as a heterogeneous disease with highly variable outcomes. Recent efforts have been made to molecularly characterize LOAD using large cohorts of post-mortem human brain transcriptomic data [2]. Systems-level analysis of these large human data sets has revealed key drivers and molecular pathways that reflect specific changes resulting from

disease [2, 3]. These studies have been primarily driven by gene co-expression analyses that reduce transcriptomes to modules representing specific disease processes or cell types across heterogeneous tissue samples [2, 4–6]. Similar approaches have been used to characterize mouse models of neurodegenerative disease [7]. Detailed cross-species analysis reveals a translational gap between animal models and human disease, as no existing models fully recapitulate pathologies associated with LOAD [8–11]. New platforms to rapidly assess the translational relevance of new animal models of LOAD will allow efficient identification of the most promising preclinical models.

* Correspondence: gregory.carter@jax.org

†Christoph Preuss and Ravi Pandey contributed equally to this work.

¹The Jackson Laboratory, Bar Harbor, ME 04609, USA

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



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

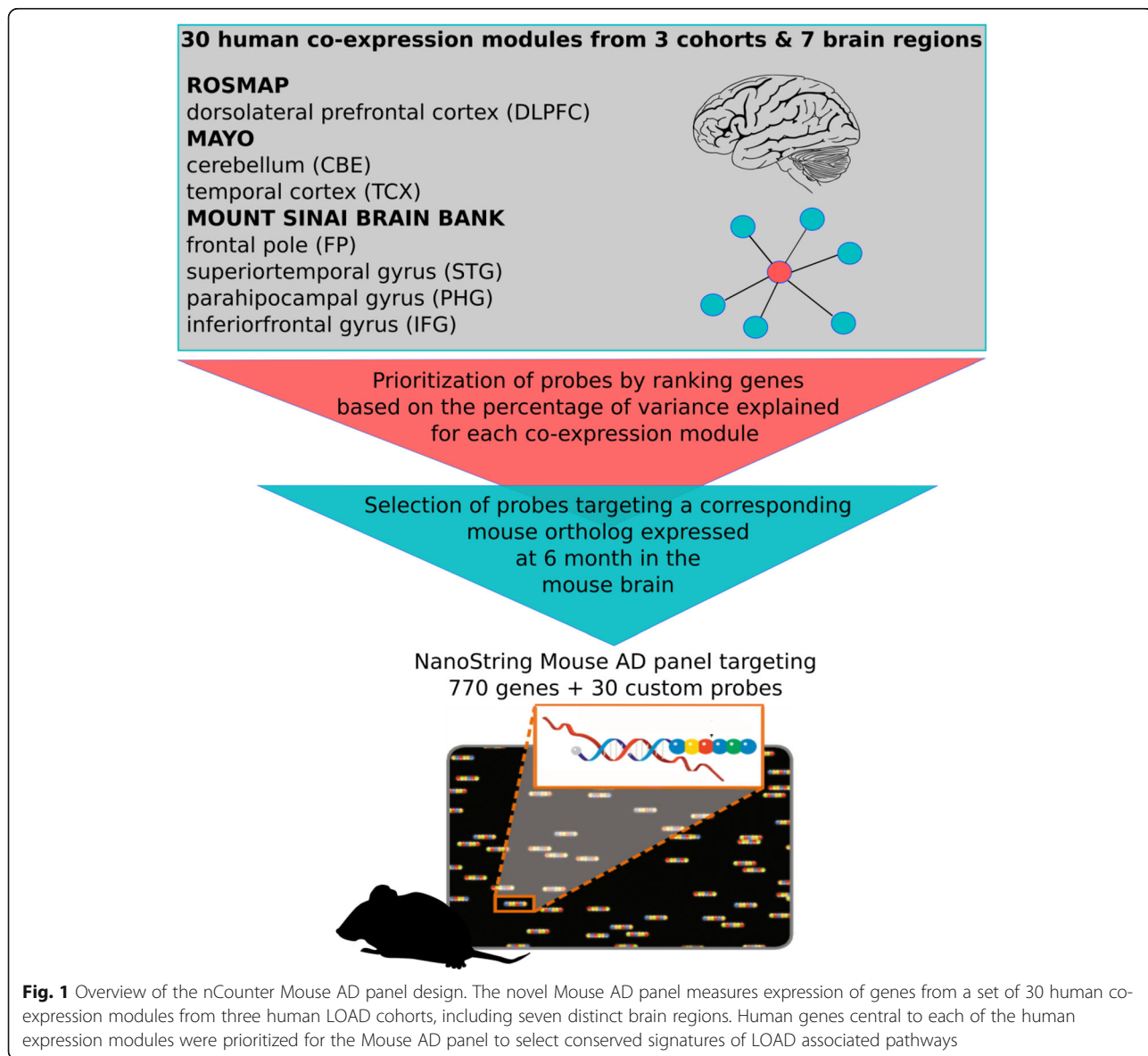
In this study, we describe a novel gene expression panel to assess LOAD-relevance of mouse models based on expression of key genes in the brain. We used a recent human molecular disease catalog based on harmonized co-expression data from three independent post mortem brain cohorts (ROSMAP, Mayo, Mount Sinai Brain bank) [12–14] and seven brain regions that define 30 human co-expression modules and five consensus clusters derived from the overlap of those modules [9]. These modules were used to design a mouse gene expression panel to assess the molecular overlap between human disease states and mouse models. This nCounter Mouse AD panel was piloted with samples from three novel mouse models of LOAD. Same-sample comparison between NanoString and RNA-Seq data demonstrated high per-gene correlation and overall concordance when

separately compared to human disease co-expression modules. Taken together, the rapid screening of mouse models in the course of different life stages will allow better characterization of models based on alignment with specific human molecular pathologies.

Results

Human-mouse co-expression module conservation and probe coverage across 30 LOAD associated modules

An overview of the Mouse AD panel design for translating the 30 human AMP-AD co-expression modules from three cohorts and seven brain regions is depicted in Fig. 1. Mouse to human gene prioritization resulted in the selection of 760 key mouse genes targeting a subset of highly co-expressed human genes plus 10 housekeeping genes, which explained a significant proportion of the observed variance



across the 30 human AMP-AD modules (Methods). Co-expression modules were grouped into functionally distinct consensus clusters as previously described by Wan, et al. (see also Table S1) [9]. These consensus clusters contain expression modules from different brain regions and independent studies that share a high overlap in gene content and similar expression characteristics. Consensus clusters were annotated based on Reactome pathway enrichment analysis for the corresponding genes within each functionally distinct cluster (Methods, Table S1). Since consensus clusters showed an enrichment of multiple biological pathways, the highest rank and non-overlapping Reactome pathway was used to refer to each cluster (Table S2). In order to assess the conservation of sequence and gene expression levels between human and mouse genes for each of the 30 human co-expression modules, dN/dS values were correlated with the overall overlap in expression in brains from six-month-old C57BL/6J (B6) mice (Fig. 2a). The fraction of orthologous genes expressed in the mouse brain, based on the presence or absence of transcripts at detectable levels, was very highly correlated with the overall module conservation ($p < 2.2 \times 10^{-16}$, Pearson's correlation coefficient: -0.96). Module conservation was based on the median dN/dS statistics measuring the rate of divergence in the coding sequence for all genes within a given module between both species (Figure S1). Notably, human co-expression modules of Consensus Cluster C, associated with the neuronal system and neurotransmission, showed the lowest degree of sequence divergence with a high proportion of human genes (64–72%) expressed in six-month-old B6 mice. In contrast to the highly conserved neuronal modules, immune modules of Consensus Cluster B contained genes that recently diverged on the sequence level and acquired a higher number of destabilizing missense variants. These modules showed the highest median dN/dS values and the lowest fraction of genes (27–46%) expressed in the mouse brain across all tested modules. The remaining human co-expression modules, associated with different functional categories (Fig. 2a, Table S1), had intermediate overlap in expression levels between human and mice. Each of the 30 human co-expression modules was covered with an average of 148 NanoString mouse probes (SD = 50 probes), where a single mouse probe can map to multiple human modules from different study cohorts and across several brain regions. Overall, mouse probe coverage for human co-expression modules ranged between 4 and 19%, depending on the size and level of conservation of the targeted human module (Fig. 2b and c, Tables S2 and S3). For three of the largest human co-expression modules harboring over 4000 transcripts, the probe coverage was slightly below the targeted 5% coverage threshold. However, these large modules are predominantly associated with neuronal function and show a high degree

of expression and sequence conservation between human and mouse (Figs. 2a). Immune modules, containing genes that recently diverged on the coding sequence level, are well covered with a median coverage of 10% (Fig. 2c). A complete annotation of mouse probes to human transcripts for each human co-expression module is provided in Table S3. In addition, we compared our novel panel to the existing nCounter Mouse Neuropathology panel designed to assess expression changes in multiple neurodegenerative diseases. We observed an overlap of 105 probes (7%) between both panels, highlighting that most of our selected probe content is novel and specific to LOAD associated disease processes and pathways.

Prioritized subset of key genes shows a higher degree of sequence conservation and expression level across modules

In order to assess the level of sequence divergence and expression for the prioritized subset of genes on the novel panel, the selected subset of genes was compared to all genes across the 30 human co-expression modules. The 760 key genes, explaining a significant proportion of the observed variance in each human module, showed an overall lower level of sequence divergence (median dN/dS values) when compared to all other genes in the modules (Fig. 3, Figure S1). Furthermore, the selected key genes on the Mouse AD panel also displayed a higher average level of gene expression in brains of six-month-old B6 mice compared to the remaining genes for each of the 30 modules (Fig. 3). This highlights that our formal prioritization procedure resulted in the selection of a subset of highly expressed key genes, which are also more conserved between human and mouse facilitating the translation of co-expression profiles across species.

5xFAD mouse co-expression modules correlate with AMP-AD modules enriched for inflammatory and stress response pathways

To validate our novel Mouse AD panel, a time-course analysis was performed to correlate human AMP-AD co-expression modules with the 5xFAD mouse model carrying a transgenic insertion with five familial mutations in *APP* and *PSEN1*. The 5xFAD strain is a commonly used model of neurodegenerative disease for which neuropathology, histological as well as behavioral data are readily available (Fig. 4a). We analyzed mouse NanoString data from brain hemispheres from 1 to 12 months old mice in order to identify an overlap with human post-mortem co-expression modules in the course of amyloid deposition and aging. 5xFAD mice have been shown to accumulate high levels of intraneuronal A β 42 at around 1.5 months of age and extracellular amyloid at around 2 months of age in the subiculum and cortex

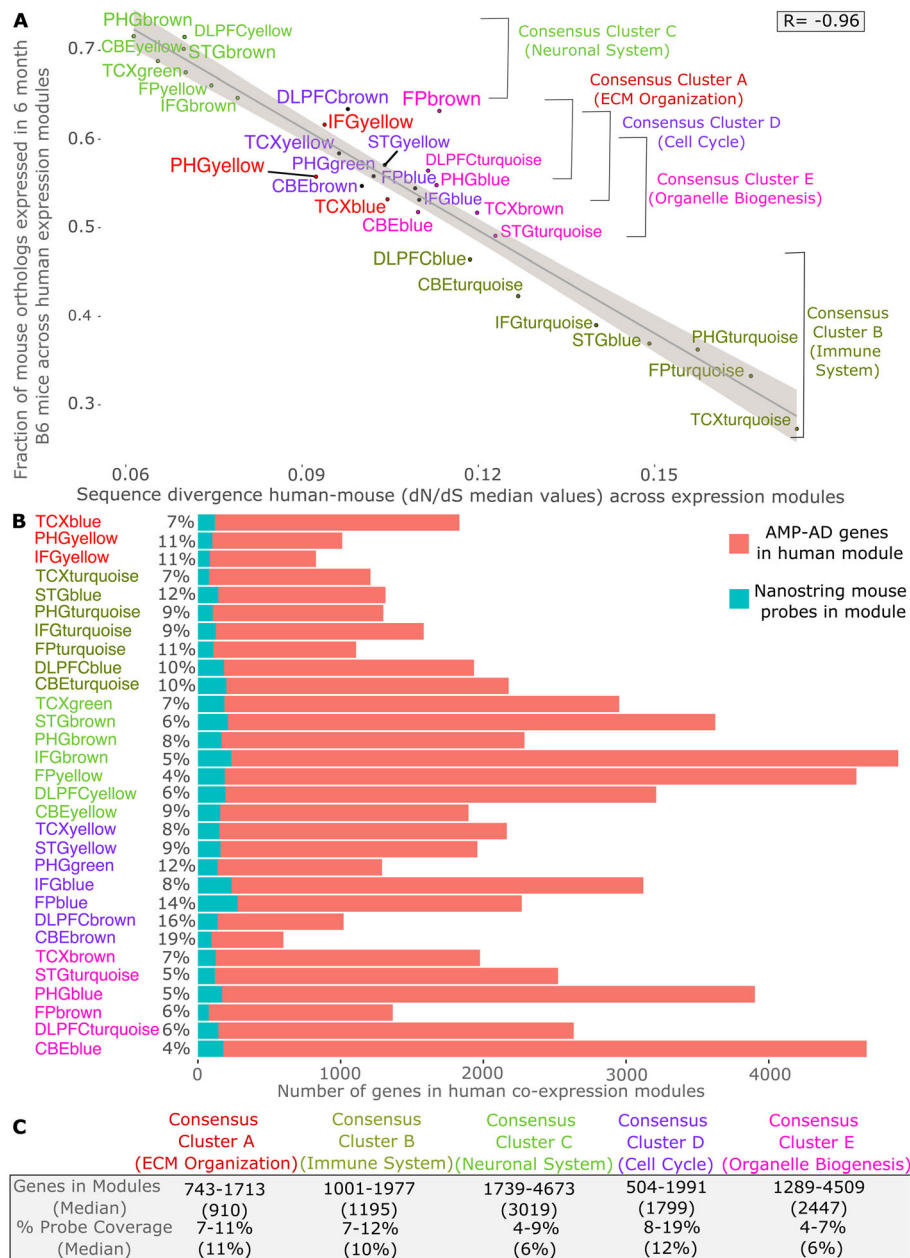
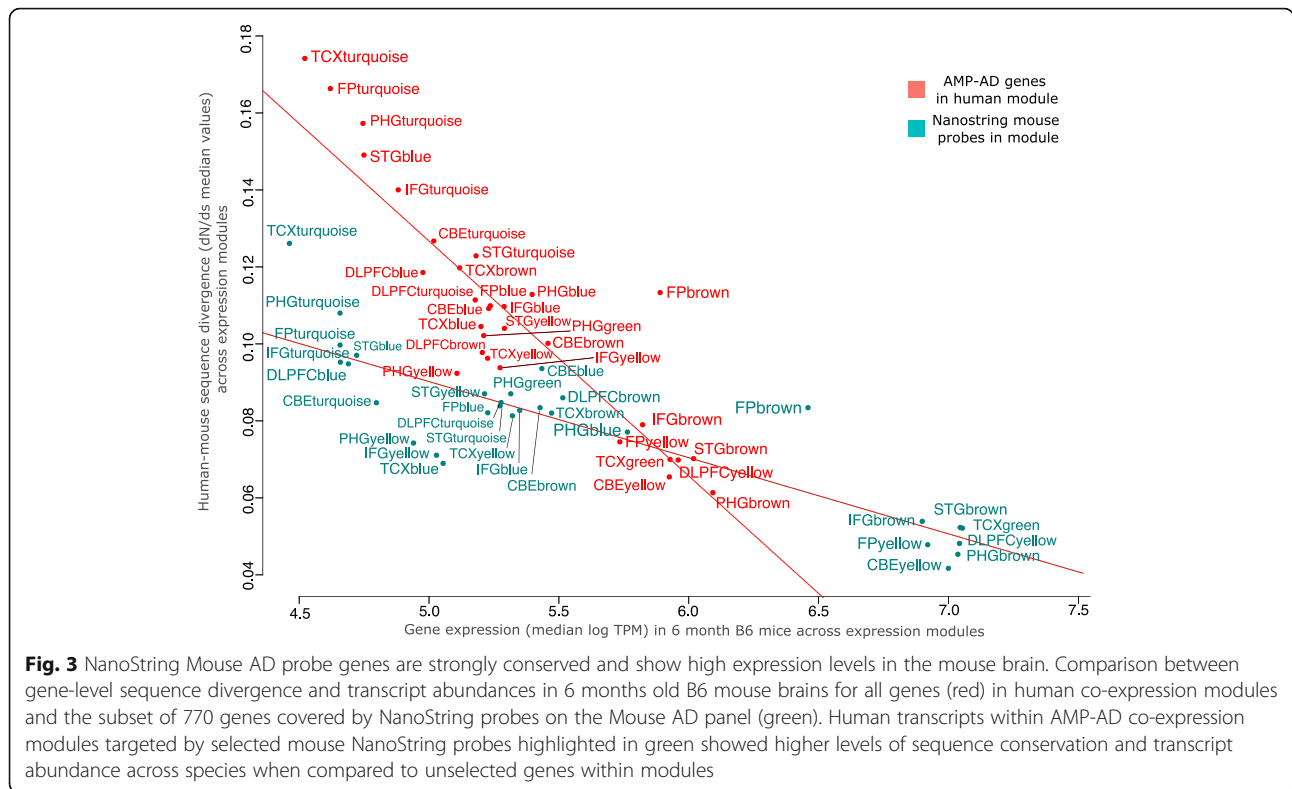


Fig. 2 Human to mouse comparison and probe coverage summary statistics. **a** Human-mouse sequence divergence (median dN/dS values) is inversely correlated (Pearson’s correlation coefficient: -0.96) with the fraction of genes being expressed in B6 mouse brain for each of the human co-expression modules. **b** Coverage of the 770 selected mouse NanoString probes for the 30 human co-expression modules associated with five functional consensus clusters. The size and number of human co-expression modules differs for the three post-mortem brain cohorts (ROSMAP, Mayo, Mount Sinai Brain Bank) and across the seven included brain regions. **c** This results in a varying degree of probe coverage for each module with a number of disease associated consensus clusters (**a-e**), reflecting disease related pathways and processes

(Fig. 4a) [15]. Amyloid deposition in 5xFAD mice induced rising oxidative stress in surrounding cells and tissues and led to subsequent cell dysfunction and elevation of apoptosis markers. This is reflected in the transcriptional response of one to 2 months old 5xFAD mice which showed a significant positive correlation ($p < 0.05$,

Pearson’s correlation coefficient > 0.25) with multiple human AMP-AD modules enriched for pathways linked to the cellular stress response in Consensus Cluster E (Fig. 4a). The 5XFAD strain also exhibited neuroinflammation after 2 months of age, as previously demonstrated by an increase in astrogliosis and microgliosis after initial plaque



deposition [15]. This is in line with our 5xFAD data from the Mouse AD panel which showed an increased correlation between 3 and 4 months with several human modules in Consensus Cluster B enriched for immune related pathways. Furthermore, we observed an increased correlation with inflammatory AMP-AD co-expression modules after 4 months of age, where older mice at (6–12 months) showed a highly significant overlap with human immune modules in the cerebellum, superior temporal gyrus and inferior frontal gyrus ($p < 0.05$, Pearson’s correlation coefficient > 0.35). This clear shift in inflammatory gene expression signatures after 3 months of age is a hallmark of the 5xFAD model in response to aggressive amyloid deposition in the brain and corresponded with previous findings from a transcriptomic survey of 5xFAD mice [16]. Reactome pathway analysis in 5xFAD mice compared to age-matched B6 controls supported the results from our correlation analysis (Fig. 4b). At a young age (3 months), several stress and immune related pathways were enriched in 5xFAD mice when compared to B6 mice. These pathways, including the activation of FOXO transcription factors pathway, are well known to mediate a cellular stress response to A β 42 [17]. In older mice between the ages of 6 to 12 months, pathways linked to microglia and complement activation, such as the DAP12 signaling pathway, were enriched in 5xFAD mice which is in line with a previous study (Fig. 4b) [16]. Moreover, we observed a positive correlation with human AMP-AD

modules enriched for neuronal pathways in Consensus Cluster C. This transcriptional response occurred after early amyloid deposition in male mice at 3 months ($p < 0.05$, Pearson’s correlation coefficient > 0.15) and after the onset of neuronal loss in female mice at 12 months. Taken together, our novel approach identified several transcriptomic signatures in responses to amyloid deposition in the 5xFAD model that correlated with human post-mortem data from different brain regions. Despite the overlap with several key disease processes on the transcriptome level across species, the 5xFAD strain does not fully recapitulate LOAD pathologies. In addition, the highly penetrant nature of the early-onset familial variants expressed on the 5xFAD background made it difficult to identify disease related processes independent of amyloid pathology. For this reason, we used our approach to elucidate the role of additional AD risk variants in a set of novel mouse models .

Novel mouse models harboring LOAD associated risk variants correlate with distinct AMP-AD modules in a brain region- and pathway-specific manner

Three novel mouse models, harboring two LOAD risk alleles, (Table S4) were used to translate co-expression profiles between human and mouse brain transcriptome data using our novel Mouse AD panel. Transcriptome analysis was performed for the APOE4 KI mouse, carrying a humanized version of the strongest LOAD

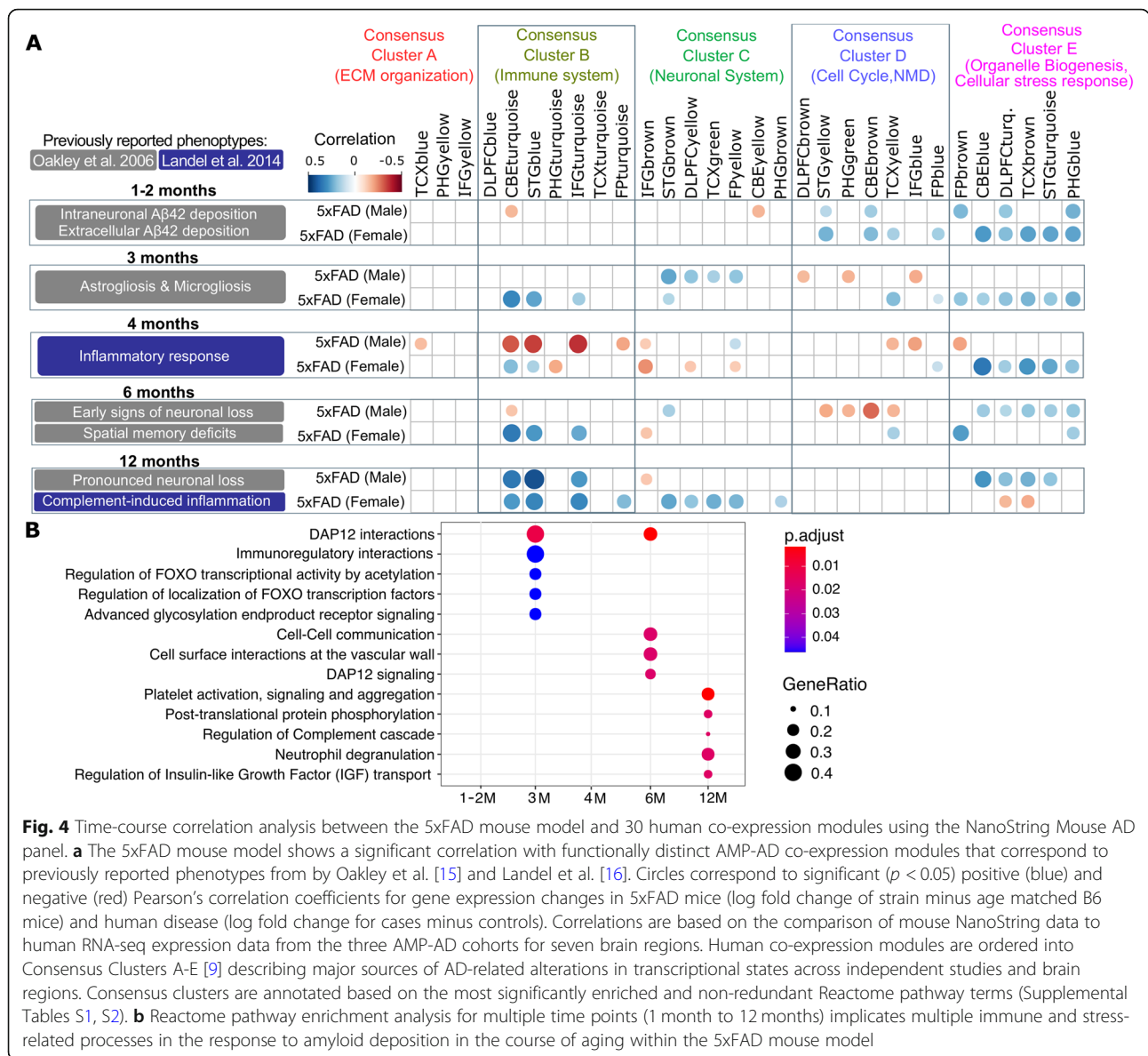


Fig. 4 Time-course correlation analysis between the 5xFAD mouse model and 30 human co-expression modules using the NanoString Mouse AD panel. **a** The 5xFAD mouse model shows a significant correlation with functionally distinct AMP-AD co-expression modules that correspond to previously reported phenotypes from by Oakley et al. [15] and Landel et al. [16]. Circles correspond to significant ($p < 0.05$) positive (blue) and negative (red) Pearson's correlation coefficients for gene expression changes in 5xFAD mice (log fold change of strain minus age matched B6 mice) and human disease (log fold change for cases minus controls). Correlations are based on the comparison of mouse NanoString data to human RNA-seq expression data from the three AMP-AD cohorts for seven brain regions. Human co-expression modules are ordered into Consensus Clusters A-E [9] describing major sources of AD-related alterations in transcriptional states across independent studies and brain regions. Consensus clusters are annotated based on the most significantly enriched and non-redundant Reactome pathway terms (Supplemental Tables S1, S2). **b** Reactome pathway enrichment analysis for multiple time points (1 month to 12 months) implicates multiple immune and stress-related processes in the response to amyloid deposition in the course of aging within the 5xFAD mouse model

associated risk allele (*APOE-ε4*) and the Trem2**R47H* mouse, which harbors a rare deleterious variant in *TREM2*. The rare *TREM2* R47H missense variant (rs75932628) has been previously associated with LOAD in multiple independent studies [15]. In our RNA-Seq data, we observed that the variant impairs expression in the Trem2**R47H* model and results in decreased expression of the major *Trem2* isoform (ENSMUST00000024791) compared to B6 mice (adjusted FDR $p = 4.26 \times 10^{-49}$, logFC = -1.25). The decreased expression of *Trem2* was also detected with the NanoString Mouse AD panel (adjusted FDR $p = 0.03$, logFC = -0.29). The difference in log fold change reduction was potentially due to mismatched dynamic range between platforms or differences between aligned isoforms of RNA-Seq reads and probe design in the Mouse AD Panel. In addition, a

mouse model harboring both, the common and rare AD risk variants (*APOE4* KI/*Trem2***R47H*) was used to compare the transcriptional effects in mice carrying both variants to mice carrying only a single risk allele and B6 controls. Mouse transcriptome data for half brains was analyzed at different ages (2–14 months) to estimate the overlap with human post-mortem co-expression modules during aging. We observed specific overlaps with distinct disease processes and molecular pathways at different ages for the *APOE4* KI and Trem2**R47H* mouse models. At an early age (2–5 months), male *APOE4* KI and Trem2**R47H* mice showed strong negative correlations ($p < 0.05$, Pearson's correlation coefficient < -0.3) with human co-expression modules in Consensus Cluster E that are enriched for transcripts associated with organelle biogenesis and cellular stress response

pathways in multiple brain regions (Fig. 5a). Furthermore, Trem2**R47H* male mice showed a significantly negative association ($p < 0.05$, Pearson's correlation coefficient < -0.2) with immune related human modules in the superior temporal gyrus, the inferior frontal gyrus, cerebellum and prefrontal cortex (Fig. 5a). This effect becomes more pronounced later in development, between six and 14 months, when the negative correlation with human immune modules is also observed in Trem2**R47H* female mice. During mid-life, (6–9 months old age group), we observed an age-dependent effect for the APOE4 KI mouse in which human neuronal modules in Consensus Cluster C start to become positively correlated with the corresponding human expression modules (Fig. 5a). Interestingly, neuronal co-expression modules which are associated with synaptic signaling appear to be positively correlated with APOE4 KI, but not Trem2**R47H* mice in an age dependent manner. This up-regulation of genes associated with synaptic signaling and a decrease of transcripts enriched for cell cycle, RNA non-mediated decay, myelination and glial development in aged mice was consistent for multiple brain regions and across three independent human AD cohorts. When compared to APOE4 KI mice, Trem2**R47H* mice showed an age dependent decrease in genes associated with the immune response in several AMP-AD modules which is not observed for APOE4 KI mice (Fig. 5a). Notably, the APOE4 KI/Trem2**R47H* mice showed characteristics of both single variant mouse models. At an early age, an overlap with both neuronal and immune associated human modules is observed and becomes more pronounced during aging.

Differential expression and gene set enrichment analysis

In order to relate the human co-expression modules to disease associated genes and pathways, we performed differential expression (DE) analysis for the three novel mouse models and the 5xFAD mouse. Each mouse model was compared to the B6 control to assess the overall transcriptomic response and the differences in DE genes across models. Both the APOE4 KI and the APOE4 KI/Trem2**R47H* models showed a moderate number of DE genes compared to B6 mice (< 100) at mid-life (6–9 months) while the number of DE is smaller (< 20) late in life. Early in life (2–5 months), only one gene was found to be DE in both the humanized APOE4 KI and the APOE4 KI/Trem2**R47H* model (Tables S5, S6). We observed a significant decrease of the mouse *ApoE* gene (adjusted FDR $p = 1.78 \times 10^{-69}$, $\log_{2}FC > -3.5$) reflecting that it was replaced by the human version. The Trem2**R47H* model shows mostly down-regulated genes across all age groups, many of which are up-regulated in 5xFAD mice. While microglia related genes, including *Tyrobp*, *Trem2*, and complement components *C1qa*, *C1qb*, *C1qc* are highly up-regulated in 5xFAD mice,

these genes are down-regulated in the Trem2**R47H* model (Table S6). To elucidate the role of these immune related disease genes on the pathway level, gene set enrichment analysis (GSEA) was performed for the four mouse models and resulting pathways were compared to the human AMP-AD data. GSEA revealed multiple immune associated pathways up-regulated in 5xFAD mice when compared to B6 mice. The strongest association was observed in aged mice (10–14 months) where three immune related pathways (phagosome, Chagas disease, osteoclast differentiation) are significantly up-regulated (adjusted Benjamini-Hochberg $p < 0.05$) in the 5xFAD model. These pathways, which were also up-regulated in multiple brain regions from independent AMP-AD cohorts (cerebellum, superior temporal gyrus, temporal cortex), were down-regulated in the Trem2**R47H* model (Figure S2). Notably, this neuroprotective effect of the Trem2**R47H* allele was dampened in the presence of APOE4 on the APOE4 KI/Trem2**R47H* background, which did not show any significant associations with immune pathways (Figure S2). To follow up on this antagonistic effect, pathway enrichment analysis for differentially expressed genes between APOE4 KI and Trem2**R47H* mice was performed using both the Reactome [18] and WikiPathway knowledge pathway databases. Multiple pathways linked to immune function and specifically complement activation were significantly enriched (FDR adjusted $p < 0.05$) for genes showing opposite expression patterns in both mouse models late in life (10–14 months) (Fig. 5b). Among the genes that contribute significantly to this pathway enrichment were three members of the complement complex 1q (C1q), namely *C1qa*, *C1qb*, and *C1qc* that were also found to be up-regulated in the amyloidogenic 5xFAD model. These genes encode for the C1q complex and were significantly decreased in both the Trem2**R47H* and APOE4 KI/Trem2**R47H* model when compared to the APOE4 KI model (Fig. 5c, Table S7). This is in line with a recent study that linked the decrease in synaptic damage and vulnerability in a humanized Trem2**R47H* tauopathy mouse model to reduced C1q expression [19]. Moreover, C1q accumulation had been shown to be drastically increased in synapses of APOE4 KI mice, when compared to APOE3 KI mice suggesting an important role of C1q in enhancing synaptic vulnerability to classical-complement-cascade mediated neurodegeneration [20]. Furthermore, expression of the C1q complex is critical for microglia function. Immune pathways up-regulated in 5xFAD mice and down-regulated in our Trem2**R47H* model harbor multiple important genes (*Tyrobp*, *Trem2*, *Ctss*, and *ApoE*) linked to the activation of disease associated microglia (DAM). In order to further characterize the role of DAM genes, we compared DAM expression signatures based on recently published data single-cell transcriptomic data [21] (Figure S3). The expression

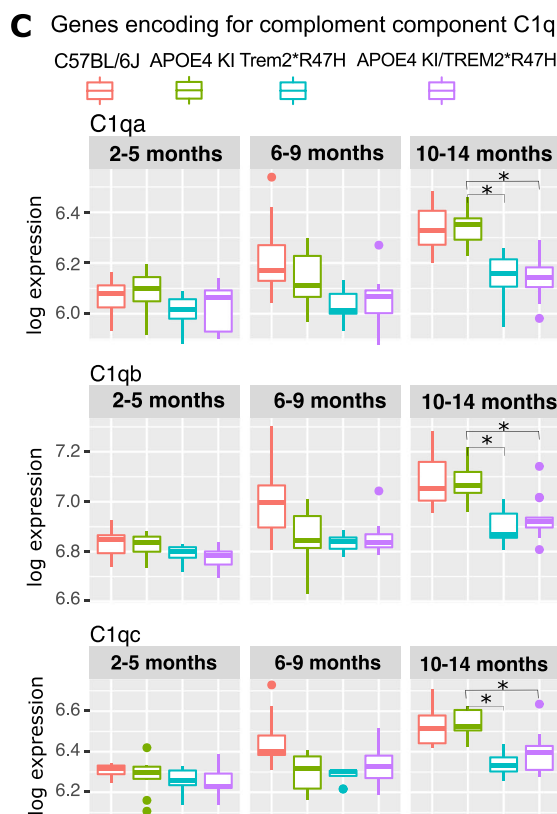
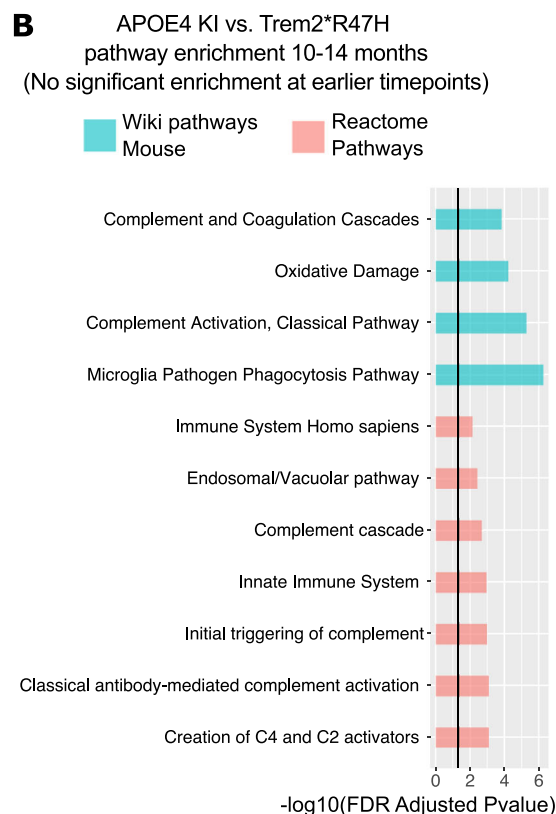
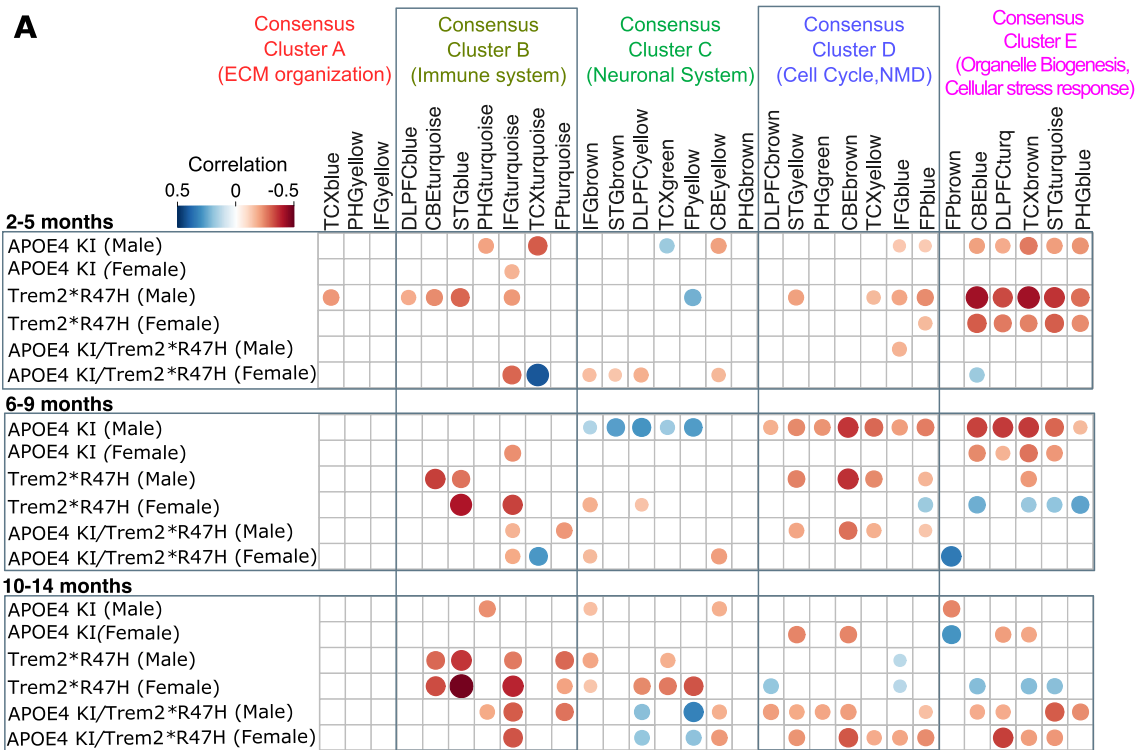


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Correlation analysis between three novel LOAD mouse models and human co-expression modules reveals age-related changes in immune function. **a** Correlation analysis highlights age-related changes in Trem2**R47H*, APOE4 KI and APOE4 KI/Trem2**R47H* mice. Circles correspond to significant ($p < 0.05$) positive (blue) and negative (red) Pearson's correlation coefficients for gene expression changes in mice associated with distinct human co-expression modules. This includes multiple modules linked to immune function (Consensus Cluster B) and stress response (Consensus Cluster E). **b** Pathway analysis for the Reactome and WikiPathway resources highlights a significant enrichment (FDR adjusted $p < 0.05$) of pathways involved in complement activation in both older (10–14 months) APOE4 KI and Trem2**R47H* mice. **c** Genes encoding for complement component C1q show an antagonistic transcriptional effect between the Trem2**R47H* and APOE4 KI/Trem2**R47H* mice when compared to the humanized APOE4 knock-in model (*denotes FDR adjusted $p < 0.05$)

signatures of 21 DAM associated genes on the NanoString panel supported an increased expression of DAM associated genes in the 5xFAD mice compared to B6 late in life (10–14 months) (Figure S3). This increased expression of DAM genes was absent in the APOE4 and Trem2 models when compared to age-matched B6 mice.

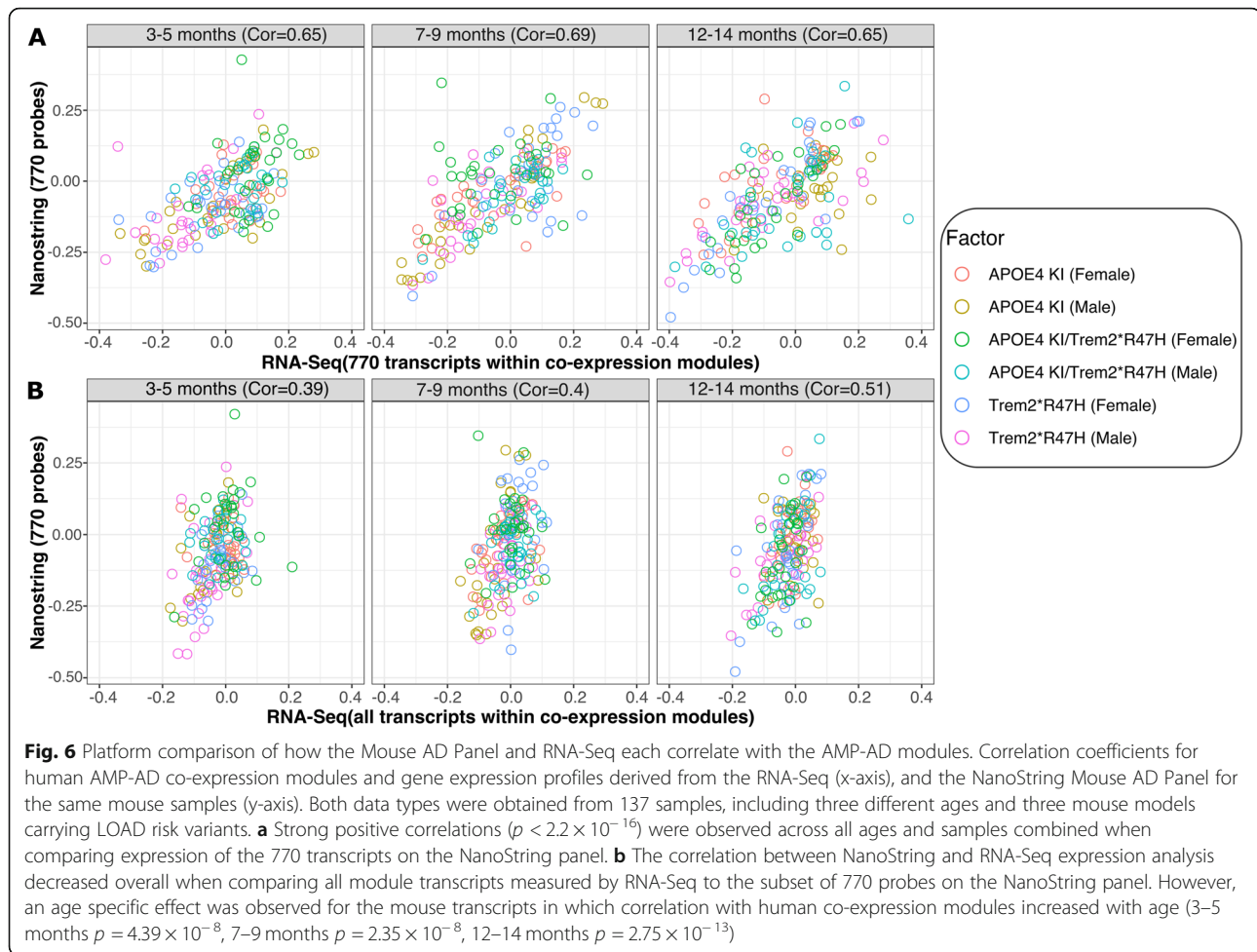
Comparison between nCounter mouse AD panel and RNA-Seq data

To assess the validity of the novel Mouse AD panel across transcriptomic platforms, we compared the results from the nCounter platform to RNA-Seq data for the same 137 mouse brain samples from three novel LOAD models carrying APOE4 and TREM2**R47H* alleles for which both RNA-Seq and NanoString data was available. A correlation analysis was performed to compare the expression of the 770 NanoString probes across co-expression modules with RNA-Seq transcript expression for all ages, highlighting the different LOAD mouse models as independent variables (Fig. 6). For the direct comparison, between the 770 NanoString probes with corresponding RNA-Seq transcripts, a similar range of correlation coefficients between human data and the three mouse models was observed (Fig. 6a). Overall, the correlation between the RNA-Seq and NanoString platforms were high across all age groups (Pearson's correlation coefficients: 0.65–0.69) when comparing the subset of 760 key transcripts and 10 housekeeping transcripts across platforms. This demonstrates that the novel NanoString panel, despite the limited number of key custom probes, can achieve similar results when compared to high-throughput RNA-Seq data. Furthermore, the alignment of human and mouse modules based on the expression of all genes within each modules showed a weaker range of correlations when compared to transcripts covered by the 770 NanoString probes (Fig. 6b). Overall, we found strong and significant correlations between the results. Notably, these correlations generally increased with mouse age, suggesting that the human relevance of the models is increasing with age and that this relevance is captured well by both the NanoString and RNA-Seq platforms (Fig. 6b). A mild correlation at around 3 months of age (Pearson's correlation coefficient: 0.39) increased to a moderate correlation at 12

months of age (Pearson's correlation coefficient: 0.51). Furthermore, we observed a high correlation of log count values for the majority of NanoString probes when compared to log TPM transcript ratios from RNA-Seq data. The majority of the 770 measured NanoString probes (716/770 probes, 93%) were positively correlated with RNA-Seq transcripts (Figure S4). In order to test whether noise introduced by highly variable transcripts affects the correlation between NanoString probes and RNA-Seq transcripts, Pearson's correlation coefficients and variance in RNA-Seq expression across 137 samples were compared. There was no significant trend indicating an effect of highly variable transcripts on the overall correlation coefficients between transcripts measured by RNA-Seq and NanoString (Figure S4).

Discussion

Here, we describe a novel systems biology approach to rapidly assess disease relevance for three novel mouse models carrying two human risk variants, strongly associated with LOAD. The nCounter Mouse AD gene expression panel was designed to align human brain transcriptome data covering 30 co-expression modules. Cross-species comparison of human and mouse revealed that immune associated co-expression modules which harbor genes that have recently diverged in sequence were more likely to be lowly expressed or absent at the transcript level in brains from 6 months old B6 mice. In contrast, neuronal modules containing genes with a lower degree of sequence divergence between both species were more likely to be highly and constitutively expressed in the mouse brain when compared to the remaining co-expression modules. This is in line with evidence from multiple studies highlighting that conserved neuronal process in the brain are under strong purifying selection while immune related genes are more likely to diverge in function and expression patterns across species [22, 23]. By using our prioritization approach, we selected for key mouse genes targeting a subset of highly co-expressed human genes. This subset of genes on the NanoString panel showed overall lower levels of sequence divergence compared to human genes and higher expression levels in the mouse brain,



reducing potential noise introduced by lowly expressed transcripts across expression modules.

Cross-platform comparison between the novel Mouse AD panel and RNA-Seq data revealed a strong correlation between mouse gene expression changes independent of platform related effects. Notably, the correlation between nCounter probe and RNA-Seq transcript expression with human co-expression modules was highest in aged mice older than 12 months. This age-dependent overlap might be expected due to the late-onset nature of Alzheimer's disease resulting in an increased number of highly co-expressed genes in aged mice carrying human LOAD risk variants. In addition, the strongest correlation between human and mouse module signatures was observed when using the subset of 770 transcripts on the NanoString panel. This highlights that assessment of key genes in the brain, contributing highly to module expression, can improve the characterization of novel LOAD mouse models and their alignment with specific human co-expression modules.

The time-course analysis of the well-established 5xFAD mouse model using our novel panel revealed a

significant overlap in transcriptional signatures with several human co-expression modules from distinct brain regions during aging. Several hallmark features of the 5xFAD model, such as the increased cellular stress response after early amyloid deposition, correlated well with human AMP-AD modules enriched for stress responsive transcripts. In addition, an age-related increase in several key inflammatory pathways and processes following more severe amyloid pathology was observed in 5xFAD mice. These findings highlighted that aged 5xFAD mice closely resemble pro-inflammatory transcriptional signatures of post-mortem brain samples from deceased LOAD patients. Interestingly, novel LOAD mouse models showed concordance with distinct human co-expression modules, reflecting a different transcriptional response driven by the human *APOE* and *TREM2* associated LOAD risk variants. The strong negative correlation between the Trem2*R47H knock-in mice and immune related human co-expression highlights the important role of the LOAD associated *TREM2* R47H variant in Alzheimer's related immune processes. This effect, which was reproduced across co-expression modules from multiple

human brain regions (cerebellum, frontal cortex, temporal gyrus, frontal gyrus, frontal pole), was inverted in the presence of a high amyloid burden in the 5xFAD model. Immune related pathways containing genes linked to microglia activation were significantly increased in 5xFAD mice and decreased in Trem2**R47H* mice. Interestingly, this inflammatory response was partially restored in the presence of *APOE4* allele on the *APOE4* KI/Trem2**R47H* background suggesting an interaction between the two LOAD risk alleles. Pathway and differential gene expression analysis revealed antagonistic expression signatures between *APOE4* KI and Trem2**R47H* mice linked to the activation of the classical complement component through C1q members (*C1qa*, *C1qb*, *C1qc*). C1q protein accumulates at senescent synapses in the course of normal brain aging making them more vulnerable to complement mediated neurodegeneration. A recent study showed that expressing the *R47H* risk allele in a humanized Trem2 mouse model lowers the expression of C1q at synapses which in turn protects them from damage in the setting of a tauopathy mouse model [19]. This is in line with multiple studies which showed that the profound loss of synapses at the early stages of AD can be prevented by blocking activation of the complement cascade through C1q depletion in the mouse [24, 25]. In contrast, the *APOE4* isoform increases C1q accumulation at synapses making them more vulnerable to degeneration when compared to the *APOE2* and *APOE3* isoforms in a set of *APOE* KI models [20]. However, C1q accumulation at synapses alone might not be sufficient to trigger synaptic loss in the aging brain. Other factors mediated by the Trem2 *R47H* and *APOE4* risk alleles may activate the neuroinflammatory cascade that leads to age-related neurodegeneration which will require further studies in these novel LOAD models. Furthermore, a strong negative correlation between co-expression modules associated with cell cycle and DNA repair was observed for the mouse *APOE4* KI model. This overlap with human late-onset co-expression signatures early in life was observed for a number of different brain regions and is absent in Trem2**R47H* knock-in mice. Furthermore, aged *APOE4* KI mice show a strong overlap with several human neuronal co-expression modules enriched for genes that play an important role in synaptic signaling and myelination. Although, *APOE4* KI mice lack a clear neurodegenerative phenotype, this age dependent shift in co-expression patterns associated with core LOAD pathologies points to an increased susceptibility to cognitive decline in aged mice. This is in line with several studies which have shown that cognitive deficits in *APOE4* transgenic mice develop late in life [26, 27]. Taken together, these results suggest that correlating gene expression signatures in LOAD and FAD mouse models to disease-associated AMP-AD modules can identify transcriptional disruptions relevant to human

disease, even when the models are insufficiently advanced to exhibit full LOAD pathology. Assessing the effects of individual and combinations of LOAD variants in mouse models in this way can potentially separate the causal co-expression modules that drive LOAD pathology through genetic risk factors from modules that respond to established pathology. Furthermore, staging such in vivo models over a lifetime can determine the order of events, including microglia activation and, ultimately, neuronal loss observed in LOAD patients. The Mouse AD Panel described here provides an efficient platform to detect these events.

Limitations of the approach

Albeit being an excellent resource for characterizing molecular pathways and key drivers of disease, co-expression modules based on human post-mortem brain data have several limitations. As end stage measures, they might not reflect changes that occur early in disease pathogenesis. For this reason, we were unable to devise age-specific mouse panels that could be most informative at specific disease stages. However, as mouse models improve and are characterized at multiple ages, early transcriptomic indicators of LOAD might motivate additional panels corresponding to stages of pathogenesis. In addition, although a high concordance was observed across brain regions for the 30 modules, they might not cover individual or region-specific differences in patients in response to amyloid and tau pathology [9]. Furthermore, we used brain homogenates from our mouse models for the transcript comparison with different human brain regions in this study. Dissection of mouse brain regions to match the human studies might further improve the observed co-expression module correlations.

Methods

AMP-AD post-mortem brain cohorts and gene co-expression modules

Data on the 30 human AMP-AD co-expression modules was obtained from the Synapse data repository (<https://www.synapse.org/#!Synapse:syn11932957/tables/>; SynapseID: syn11932957). The modules derive from three independent LOAD cohorts, including 700 samples from the ROSMAP cohort, 300 samples from the Mount Sinai Brain bank and 270 samples from the Mayo cohort. Details on post-mortem brain sample collection, tissue and RNA preparation, sequencing, and sample QC can be found in previously published work related to each cohort [12–14]. A detailed description on how co-expression modules were identified can be found in the recent study that identified the harmonized human co-expression modules as part of transcriptome wide AD meta-analysis [9]. Briefly, Wan, et al. performed library normalization and covariate adjustments for each human study separately using fixed/mixed effects

modeling to account for batch effects. Among the 2978 AMP-AD modules identified across all tissues (DOI: <https://doi.org/10.7303/syn10309369.1>), 660 modules were selected by Wan, et al. which showed an enrichment for at least one AD-specific differential expressed gene set from the meta-analysis (DOI: <https://doi.org/10.7303/syn11914606>) in cases compared to controls. Lastly, the edge betweenness graph clustering method was applied to identify 30 aggregate modules that are not only differentially expressed but are also replicated across multiple independent co-expression module algorithms [9]. Among the 30 aggregate co-expression modules, five consensus clusters have been described by Wan, et al. [9]. These consensus clusters consist of a subset of modules which are associated with similar AD related changes across the multiple studies and brain regions. Here, we used Reactome pathway (<https://reactome.org/>) enrichment analysis to identify specific biological themes across these five consensus clusters. A hypergeometric model, implemented in the clusterProfiler R package [28], was used to assess whether the number of selected genes associated within each set of AMP-AD modules defining a consensus cluster was larger than expected. All *p*-values were calculated based the hypergeometric model [29]. Pathways were ranked based on their Bonferroni corrected *p*-values to account for multiple testing. Finally, consensus clusters were annotated based on the highest ranked and non-overlapping term for each functionally distinct cluster.

Selection of NanoString probes for the nCounter mouse AD panel

Since NanoString gene expression panels are comprised of 770 probes with the option to customize 30 additional probes, we developed a formal prioritization procedure to identify the most representative genes and ensure broadest coverage across all modules (Fig. 1). Expression and transcript annotations for the 30 human co-expression modules were obtained via the AMP-AD knowledge portal (www.synapse.org/#!Synapse:syn11870970/tables/). To prioritize probe targets for the novel Mouse AD panel, human genes were ranked within each of the human AMP-AD co-expression modules based on their percentage of variation explaining the overall module behavior. First, we calculated a gene ranking score by multiplying correlations of transcripts with the percentage of variation explained by the first five principal components within each of the aggregated human AMP-AD modules. Secondly, the sums of the resulting gene scores for the first five principal components were calculated and converted to absolute values in order to rank highly positive or negative correlated transcripts within each human co-expression module. As a next step, only human transcripts with corresponding one-to-one mouse orthologous genes that are expressed in whole-brain tissue samples from six-month-old B6 mice were retained

for downstream prioritization. While this filter risks excluding very few genes (6/760, <1%) that may only be expressed at an advanced age, we maintained high representation of each human LOAD module. Disease-relevant effects are therefore robustly captured even if a few specific genes are omitted, as no module was determined to have more than five unexpressed genes in six-month-old mice. Furthermore, we included information on drug targets for LOAD from the AMP-AD Agora platform (agora.ampadportal.org), as nominated by members of the AMP-AD consortium (<https://doi.org/10.7303/syn2580853>). A total of 30 AMP-AD drug discovery targets that were highly ranked in our gene ranking approach and nominated by multiple AMP-AD groups were included on the panel (Table S3). Finally, ten housekeeping genes (*AARS*, *ASB7*, *CCDC127*, *CNOT10*, *CSNK2A2*, *FAM104A*, *LARS*, *MTO1*, *SUPT7L*, *TADA2B*) were included on the panel as internal standard references for probe normalization. This resulted in a total of 770 proposed NanoString probes, targeting the top 5% of ranked genes for each human AMP-AD expression module.

nCounter mouse AD panel probe design

The probe design process breaks a transcript's sequence down into 100 nucleotide (nt) windows to profile for probe characteristics, with the final goal of choosing the optimal pair of adjacent probes to profile any given target. Each window is profiled for intrinsic sequence makeup – non-canonical bases, G/C content, inverted and direct repeat regions, runs of poly-nucleotides, as well as the predicted melting temperature (*T_m*) for each potential probe-to-target interaction. The window is then divided in half to generate a probe pair, wherein each probe is thermodynamically tuned to determine the optimal probe length (ranging in size from 35 to 50 nt) within the 100 nt target region. Next, a cross-hybridization score is calculated for each probe region, using BLAST [30] to identify potential off-target interactions. In addition to a cross-hybridization score, a splice isoform coverage score was generated to identify transcripts that are isoforms of the gene intended to be targeted by the probe in question. Once all of this information is compiled, the final probe is then selected by identifying the candidate with the optimal splice form coverage, cross-hybridization score, and thermodynamic profile.

In-silico panel QC for intramolecular interactions

To ensure that there are no potential intramolecular probe-probe interactions that could cause elevated background for any individual probe pair, a stringent intermolecular screen is run on every collection of probes assembled into a panel. A sensitive algorithm was used that calculates both the *T_m* and the free energy potential of interactions between every possible pair of probes in

the project. If two probes conflict in a way that would likely cause background based on this calculation, an alternative probe is selected for one of the targets and the screening is re-run until there are no known conflicts.

Mouse models

All experiments involving mice (Supplemental Table S5) 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. All mice were bred and housed in a 12/12 h light/dark cycle. All experiments were performed on a unified genetic background (C57BL/6 J).

Whole-genome sequencing

Whole-genome sequencing was performed by Novogene in Beijing, China for three founders of the APOE4 KI/Trem2^{*R47H} strain to exclude potential off-target effects in loci that were not targeted. Briefly, DNA was extracted from spleen and library preparation was performed using the KAPA HyperPrep sample preparation kit (KAPA Biosystems, Wilmington, MA, USA). Libraries were analyzed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), with the DNA 2100 kit. Sequencing was performed on a HiSeq X sequencer according to the manufacturer's guidelines using 2x150bp paired-end reads. Reads were quality trimmed and filtered using the NGS QC toolkit. The resulting high-quality reads were aligned to the mm10 release of the mouse reference genome using BWA v0.5.10 [31]. SNPs and indels were called using the GATK tool suite v2.7 [32]. Finally, all variants were annotated using the SnpEff software [33]. All variants with a potential deleterious effect (missense, nonsense, splice-site, frame-shift) on the protein level are listed in Table S8 for three founders of the APOE4 KI/Trem2^{*R47H}. Figure S5 gives an overview of deleterious variants shared across these three founders. Only three genes of which two are annotated as predicted genes (*Gm11168*, *Gm10717*) and one as an olfactory receptor (*Vmn2r115*) were identified that share private, deleterious variants in our models. Expression levels of these genes in the brain were below our cut-off criteria of 1 transcript per million reads.

Mouse brain sample collection

Upon arrival at the terminal endpoint for each aged mouse cohort, individual animals were weighed prior to intraperitoneal administration of ketamine (100 mg/kg) and xylazine (10 mg/kg). First confirming deep anesthetization via toe pinch, an incision was made along the midline to expose the thorax and abdomen followed by removal of the lateral borders of the diaphragm and ribcage revealed the heart. A small cut was placed in the right atrium to relieve pressure

from the vascular system before transcardially perfusing the animal with 1XPBS via injection into the left ventricle. With the vascular system cleared, the entire brain was carefully removed and weighed before hemisecting along the midsagittal plane. Hemispheres were immediately placed in a cryovial and snap-frozen on dry ice. Brain samples were stored at -80 °C until RNA extraction was performed.

RNA sample preparation

RNA was isolated from tissue using the MagMAX mirVana Total RNA Isolation Kit (ThermoFisher) and the KingFisher Flex purification system (ThermoFisher, Waltham, MA). Brain hemispheres were thawed to 0 °C and were lysed and homogenized in TRIzol Reagent (ThermoFisher). After the addition of chloroform, the RNA-containing aqueous layer was removed for RNA isolation according to the manufacturer's protocol, beginning with the RNA bead binding step. RNA concentration and quality were assessed using the Nanodrop 2000 spectrophotometer (Thermo Scientific) and the RNA Total RNA Nano assay (Agilent Technologies, Santa Clara, CA).

RNAseq library preparation and data collection

Sequencing libraries were constructed using TruSeq DNA V2 (Illumina, San Diego, CA) 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 adaptors 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 adaptors at both ends will amplify. Libraries were pooled and sequenced by the Genome Technologies core facility at The Jackson Laboratory. 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 was received the samples were concatenated to have a single file for paired-end analysis.

NanoString gene expression panel and data collection

The NanoString Mouse AD gene expression panel was used for gene expression profiling on the nCounter platform (NanoString, Seattle, WA) as described by the manufacturer. nSolver software was used for analysis of NanoString gene expression values. Normalized log intensity and raw gene expression data can be accessed via NCBI GEO (Accession Number: GSE141509).

Normalization of NanoString data

Normalization was done by dividing counts within a lane by geometric mean of the housekeeping genes from the same lane. For the downstream analysis, counts were log-transformed from normalized count values.

Mouse-human expression comparison

First, we performed differential gene expression analysis for each mouse model and sex using the voom-limma [34] package in R. Secondly, we computed correlation between changes in expression (log fold change) for each gene in a given module with each mouse model, sex and age. Correlation coefficients were computed using `cor.test` function built in R as:

$$\text{cor.test}(\text{LogFC}(h), \text{LogFC}(m)) \quad (1)$$

where $\text{LogFC}(h)$ is the log fold change in transcript expression of human AD patients compared to control patients and $\text{LogFC}(m)$ is the log fold change in expression of mouse transcripts compare to control mouse models. LogFC values for human transcripts were obtained via the AMP-AD knowledge portal (<https://www.synapse.org/#!Synapse:syn11180450>).

Differential expression, Gene set and pathway enrichment analysis

Differential gene expression analysis was performed using the limma package in the R software environment [34] for all analyzed mouse strains. Each model at each individual time point was compared to age-matched B6 wild type mice. In order to identify significantly enriched pathways across mouse models and human brain regions, gene set enrichment analysis was used based on the method proposed by Subramanian, et. al [35] as implanted in the clusterProfiler package for the KEGG pathway library. Briefly, human data with expression log fold changes for the seven AMP-AD brain regions were downloaded from Synapse (<https://www.synapse.org/#!Synapse:syn14237651>). We filtered to orthologous genes on the NanoString Mouse AD panel and KEGG pathway enrichment was performed for each brain region independently to identify significantly up and down-regulated gene sets. For the mouse data, differential expression analysis between each mouse model and B6 controls was performed to obtain a list of fold changes highlighting genes that are either up or down-regulated in the presence of the genetic risk variant. Enrichment scores for all significantly associated KEGG pathways were computed to compare relative expression on the pathway level between post-mortem brain samples and the four mouse models. Biological pathway enrichment analysis was performed using the clusterProfiler [28] package within the R software environment for the Reactome [18] and

WikiPathways ([wikipathways.org](http://www.wikipathways.org)) knowledge bases. Pathways were determined to be significant after multiple testing correction (FDR adjusted $p < 0.05$).

Quality control of RNA-Seq data and read alignment

Sequence quality of reads was assessed using FastQC (v0.11.3, Babraham). Low-quality bases were trimmed from sequencing reads using Trimmomatic (v0.33) [36]. After trimming, reads of length longer than 36 bases were retained. The average quality score at each base position was greater than 30 and sequencing depth were in range of 60–120 million reads. All RNA-Seq samples were mapped to the mouse genome (mm10 reference, build 38, ENSEMBL) using ultrafast RNA-Seq aligner STAR [37] (v2.5.3). The genes annotated for mm10 (GRCm38) were quantified in two ways: Transcripts per million (TPM) using RSEM (v1.2.31) and raw read counts using HTSeq-count (v0.8.0).

Mouse-human co-expression module conservation

Genomic information on orthologous groups was obtained via the latest ENSEMBL build for human genome version GRCh38. All orthologous relationships were downloaded via BioMart [38] ([biomart.org](http://www.ebi.ac.uk/biomart)). dN/dS statistics were retrieved for all orthologous gene pairs with a one-to-one relationship between human and mouse. dN/dS values are calculated as the ratio of nonsynonymous substitutions to the number of synonymous substitutions in protein coding genes. The dN/dS values in ENSEMBL were calculated based on the latest version of the codeml (<http://abacus.gene.ucl.ac.uk/software/paml.html>) package using standard parameters (ensembl.org/info/genome/compara/homology_method.html) [39].

Conclusions

Taken together, we show that the novel nCounter Mouse AD gene expression panel offers a rapid and cost-effective approach to assess disease relevance of novel LOAD mouse models. Furthermore, this approach based on gene co-expression signatures offers a high level of reproducibility and will supplement methods solely based on differential expression analysis. Ultimately, this will help us to better understand the relevance of novel LOAD mouse models in regard to specific pathways and processes contributing to late-onset Alzheimer's disease.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13024-020-00412-5>.

Additional file 1.

Abbreviations

AD: Alzheimer's disease; LOAD: Late-onset Alzheimer's disease; B6: C57BL/6 J mice; RNA-Seq: RNA sequencing; AMP-AD: Accelerating Medicines

Partnership for Alzheimer's Disease; ROSMAP: Religious Orders Study/Memory and Aging Project; DAM: Disease associated microglia (DAM)

Acknowledgments

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 Jamie Kuhar for her critically reading of the manuscript.

Authors' contributions

CP designed the novel transcriptome panel and performed bioinformatics analyses. RP, AF, AU, TP performed the gene-expression analyses in human and mouse brain tissue. EP designed the NanoString probes and guided the creation of the novel NanoString panel. BAL and LM curated human brain data. DG, GRH and MS performed mouse experiments. GWC and MS supervised and designed the project. CP, GWC and RP wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the National Institutes of Health grant U54 AG 054345.

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 (doi:<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 MedicalCenter 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:syn17095980).

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

Not applicable

Author details

¹The Jackson Laboratory, Bar Harbor, ME 04609, USA. ²NanoString Technologies, Seattle, WA 98109, USA. ³Sage Bionetworks, Seattle, WA 98121, USA. ⁴Stark Neurosciences Research Institute, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

Received: 5 December 2019 Accepted: 17 October 2020

Published online: 10 November 2020

References

1. Mayeux R, Stern Y. Epidemiology of Alzheimer disease. *Cold Spring Harb Perspect Med* [Internet]. 2012 Aug 1 [cited 2019 Jan 14];2(8). Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22908189>.
2. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* [Internet]. 2013;153(3):707–20 [cited 2019 Jan 14] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23622250>.
3. Heuer SE, Neuner SM, Hadad N, O'Connell KMS, Williams RW, Philip VM, Gaiteri C, Kaczorowski CC. Identifying the molecular systems that influence cognitive resilience to Alzheimer's disease in genetically diverse mice. *Learn Mem*. 2020 27(9):355–371. Available from: <https://doi.org/10.1101/lm.051839.120>.
4. Seyfried NT, Dammer EB, Swarup V, Nandakumar D, Duong DM, Yin L, et al. A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. *Cell Syst* [Internet]. 2017;4(1):60–72 e4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27989508>.
5. Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med* [Internet]. 2013;5, 48(5) [cited 2019 Jan 17] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23705665>.
6. Milind N, Preuss C, Haber A, Ananda G, Mukherjee S, John C, et al. Transcriptomic stratification of late-onset Alzheimer's cases reveals novel genetic modifiers of disease pathology. *Barsh GS. PLOS Genet* [Internet]. 2020;16(6):e1008775. [cited 2020 Sep 28] Available from: <https://doi.org/10.1371/journal.pgen.1008775>.
7. Matarin M, Salih DA, Yasvoina M, Cummings DM, Guelfi S, Liu W, et al. A Genome-wide Gene-Expression Analysis and Database in Transgenic Mice during Development of Amyloid or Tau Pathology. *Cell Rep* [Internet]. 2015 Feb 3 [cited 2019 Jan 14];10(4):633–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25620700>.
8. 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 [cited 2019 Jan 14] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26708939>.
9. Wan Y-W, Al-Ouran R, Mangleburg CG, Perumal TM, Lee TV, Allison K, et al. Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models. *Cell Rep*. 2020;32(2):107908. Available from: <https://doi.org/10.1016/j.celrep.2020.107908>.
10. Onos KD, Uyar A, Keezer KJ, Jackson HM, Preuss C, Acklin CJ, et al. Enhancing face validity of mouse models of Alzheimer's disease with natural genetic variation. *Frankel WN, editor. PLOS Genet* [Internet]. 2019 May 31 [cited 2020 Sep 22];15(5):e1008155. Available from: <https://dx.plos.org/10.1371/journal.pgen.1008155>.
11. 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* [Internet]. 2019 Dec 26 [cited 2020 Sep 24];14(1):50. Available from: <https://moleculareneurodegeneration.biomedcentral.com/articles/10.1186/s13024-019-0351-3>.
12. Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data*. 2016 3:160089. Available from: <https://doi.org/10.1038/sdata.2016.89>.
13. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data*. 2018;5:180142. Available from: <https://doi.org/10.1038/sdata.2018.142>.
14. Wang M, Beckmann ND, Roussos P, Wang E, Zhou X, Wang Q, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data*. 2018 Sep;5:180185.
15. Oakley H, Cole SL, Logan S, Maus E, Shao P, Craft J, et al. Intraneuronal β -amyloid aggregates, Neurodegeneration, and Neuron loss in transgenic Mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J Neurosci* [Internet]. 2006;26(40):10129–40 Available from: <http://www.jneurosci.org/content/26/40/10129.abstract>.
16. Landel V, Baranger K, Virard I, Llorid B, Khrestchatsky M, Rivera S, et al. Temporal gene profiling of the 5XFAD transgenic mouse model highlights

- the importance of microglial activation in Alzheimer's disease. *Mol Neurodegener* [Internet]. 2014 Sep 11 [cited 2020 Sep 22];9(1):33. Available from: [/pmc/articles/PMC4237952/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/25605792/).
17. Maiese K. Forkhead transcription factors: new considerations for alzheimer's disease and dementia. *J Transl Sci* [Internet]. 2016 [cited 2020 Sep 22];2(4): 241–7. Available from: [/pmc/articles/PMC4932907/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/25605792/).
 18. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2016; 44(D1):D481–7.
 19. Gratzke M, Leyns CE, Sauerbeck AD, St-Pierre MK, Xiong M, Kim N, et al. Impact of TREM2R47H variant on tau pathology-induced gliosis and neurodegeneration. *J Clin Invest*. 2020;130(9):4954–68. Available from: <https://doi.org/10.1172/JCI138179>.
 20. Chung WS, Verghese PB, Chakraborty C, Joung J, Hyman BT, Ulrich JD, et al. Novel allele-dependent role for APOE in controlling the rate of synapse pruning by astrocytes. *Proc Natl Acad Sci U S A* [Internet]. 2016 Sep 6 [cited 2020 Sep 21];113(36):10186–91. Available from: [/pmc/articles/PMC5018780/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/27033548/).
 21. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* [Internet]. 2017;169(7):1276–90 [cited 2019 Nov 27]e17. Available from: [http://www.ncbi.nlm.nih.gov/pubmed/28602351](https://www.ncbi.nlm.nih.gov/pubmed/28602351).
 22. Geschwind DH, Rakic P. Cortical evolution: judge the brain by its cover. *Neuron* [Internet]. 2013;80(3):633–47 [cited 2019 Apr 10] Available from: [http://www.ncbi.nlm.nih.gov/pubmed/24183016](https://www.ncbi.nlm.nih.gov/pubmed/24183016).
 23. Bauernfeind AL, Soderblom EJ, Turner ME, Moseley MA, Ely JJ, Hof PR, et al. Evolutionary Divergence of Gene and Protein Expression in the Brains of Humans and Chimpanzees. *Genome Biol Evol* [Internet]. 2015;7(8):2276–88 [cited 2019 Apr 10] Available from: [http://www.ncbi.nlm.nih.gov/pubmed/26163674](https://www.ncbi.nlm.nih.gov/pubmed/26163674).
 24. Fonseca MI, Zhou J, Botto M, Tenner AJ. Absence of C1q leads to less neuropathology in transgenic mouse models of Alzheimer's disease. *J Neurosci* [Internet]. 2004 Jul 21 [cited 2020 Sep 21];24(29):6457–65. Available from: <https://www.jneurosci.org/content/24/29/6457>.
 25. Hong S, Beja-Glasser VF, Nfonoyim BM, Frouin A, Li S, Ramakrishnan S, et al. Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* (80-) [Internet]. 2016 May 6 [cited 2020 Sep 21];352(6286): 712–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/27033548/>.
 26. Rijpmma A, Jansen D, Arnoldussen IAC, Fang XT, Wiesmann M, Mutsaers MPC, et al. Sex Differences in Presynaptic Density and Neurogenesis in Middle-Aged ApoE4 and ApoE Knockout Mice. *J Neurodegener Dis* [Internet]. 2013; 2013:531326 [cited 2019 Jan 17] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26316992>.
 27. Siegel JA, Haley GE, Raber J. Apolipoprotein E isoform-dependent effects on anxiety and cognition in female TR mice. *Neurobiol Aging* [Internet]. 2012; 33(2):345–58 [cited 2019 Jan 17] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20400205>.
 28. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* [Internet]. 2012 May [cited 2019 Jun 3];16(5):284–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22455463>.
 29. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, et al. GO::TermFinder—open source software for accessing Gene Ontology terms information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* [Internet]. 2004 Dec 12 [cited 2019 Jun 3];20(18):3710–5. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bth456>.
 30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* [Internet]. 1990;215(3):403–10 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2231712>.
 31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* [Internet]. 2009 Jul 15 [cited 2013 Aug 6]; 25(14):1754–60. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2705234&tool=pmcentrez&rendertype=abstract>.
 32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20(9):1297–303.
 33. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* (Austin) [Internet]. 2012;6(2):80–92 Available from: <http://www.tandfonline.com/doi/abs/10.4161/fly.19695>.
 34. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* [Internet]. 2015;43(7):e47 [cited 2019 Aug 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25605792>.
 35. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* [Internet]. 2005;102(43):15545–50 [cited 2014 Jul 10] Available from: <http://www.pnas.org/content/102/43/15545.abstract>.
 36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* [Internet]. 2014;30(15):2114–20 [cited 2019 Jan 17] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24695404>.
 37. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]. 2013;29(1):15–21 [cited 2019 Jan 17] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23104886>.
 38. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* [Internet]. 2005;21(16):3439–40. Available from: <https://doi.org/10.1093/bioinformatics/bti525>.
 39. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* [Internet]. 2009;19(2):327–35 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19029536>.

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

