A candidate regulatory variant at the *TREM* gene cluster associates with decreased Alzheimer's disease risk, and increased *TREML1* and *TREM2* brain gene expression

Minerva M. Carrasquillo[1], Mariet Allen[1], Jeremy D. Burgess[1], Xue Wang[2], Samantha L. Strickland[1], Shivani Aryal[1], Joanna Siuda[1], Michaela L. Kachadoorian[1], Christopher Medway[1,3], Curtis S. Younkin[1], Asha Nair[4], Chen Wang[4], Pritha Chanana[4], Daniel Serie[2], Thuy Nguyen[1], Sarah Lincoln[1], Kimberly G. Malphrus[1], Kevin Morgan[3], Todd E. Golde[5], Nathan D. Price[6], Charles C. White [7,8], Philip L De Jager[7,8,9], David A. Bennett[10], Yan W. Asmann[2], Julia E. Crook[2], Ronald C. Petersen[11], Neill R. Graff-Radford[12], Dennis W. Dickson[1], Steven G. Younkin[1], Nilüfer Ertekin-Taner[1,12*]

[1] Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL 32224, USA.

[2] Department of Health Sciences Research, Mayo Clinic, Florida, Jacksonville, FL 32224, USA.

[3] Human Genetics Group, University of Nottingham, Nottingham, UK NG7 2UH.

[4] Department of Health Sciences Research, Mayo Clinic Rochester, MN 55905, USA

[5] Department of Neuroscience, University of Florida, Gainesville, FL, 32610 USA.

[6] Institute for Systems Biology, Seattle, WA, 98109 USA.

[7] Ann Romney Center for Neurologic Diseases and Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA, 02115 USA.

[8] Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, 02142 USA.

[9] Harvard Medical School, Boston, MA, 02115 USA.

[10] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612 USA.

[11] Department of Neurology, Mayo Clinic Minnesota, Rochester, MN 55905 USA.

[12] Department of Neurology, Mayo Clinic Florida, Jacksonville, FL 32224, USA.

* To whom correspondence should be addressed at: Departments of Neurology and Neuroscience, Mayo Clinic Florida, Birdsall Building 327, 4500 San Pablo Rd, Jacksonville, FL 32224; E-mail: taner.nilufer@mayo.edu; Tel: 904-953-7103; Fax: 904-953-7370

Abstract

INTRODUCTION: We hypothesized that common Alzheimer's disease (AD)-associated variants within the triggering receptor expressed on myeloid (*TREM*) gene cluster influence disease through gene expression.

METHODS: Expression microarrays on temporal cortex and cerebellum from ~400 neuropathologically diagnosed AD and non-AD subjects, and two independent RNAseq replication cohorts were used for expression quantitative trait locus (eQTL) analysis.

RESULTS: *TREML1* and *TREM2* have reliably detectable expression. A variant within a DNase hypersensitive site 5' of *TREM2*, rs9357347-C, associates with reduced AD-risk and increased *TREML1* and *TREM2* levels. Meta-analysis on eQTL results from three independent datasets (n=1,006) confirmed these associations ($p=3.4x10^{-2}$ and $3.5x10^{-3}$, respectively).

DISCUSSION: Our findings point to rs9357347 as a functional regulatory variant that contributes to a protective effect observed at the *TREM* locus in the International Genomics of Alzheimer's Project (IGAP) GWAS meta-analysis, and suggest concomitant increase of *TREML1* and *TREM2* brain levels as a potential mechanism for protection from AD.

Keywords: Alzheimer's disease, eQTL, *TREM2*, *TREML1*, regulatory variant

1. Introduction

Whole genome and exome sequencing are used as complementary approaches to uncover novel loci that can be missed by GWAS, and enabled the discovery of rare, missense alleles within *TREM2* that have a relatively large effect size on AD-risk [1, 2]. *TREM2* is a member of the triggering receptor expressed on myeloid (TREM) family, known to play a key role in modulating inflammation in the innate immune response [3]. This finding provided strong supportive evidence for the importance of inflammation in the etiology of AD, but the specific role played by *TREM2* in AD pathophysiology remains unclear [4].

Since the first two reports [1, 2], the risk effect of the most significant *TREM2* rare missense variant p.R47H (a.k.a. rs75932628) has been replicated in multiple Caucasian series [5-9], including a large meta-analysis of 24,086 AD cases and 148,993 controls [10]. *TREM2* resides within the *TREM* gene cluster on chromosome 6p21.1 (**Fig. 1**), which also includes the protein coding genes *TREM1, TREML1, TREML2, TREML4* that could be additional plausible AD-risk genes.

A missense variant in *TREML2,* p.S144G (a.k.a. rs3747742), that is not in linkage disequilibrium (LD) with *TREM2* p.R47H, was reported to associate with reduced AD-risk [11]. *TREML2* p.S144G is in tight LD with the intergenic variant, rs9381040, that demonstrated the most significant association at the *TREM* locus in the IGAP AD-risk GWAS meta-analysis ($p=6x10^{-04}$) [12]. The authors concluded that *TREML2* p.S144G is the functional variant that accounted for the IGAP *TREM* locus signal, even though the significance of the AD-risk association with the intergenic rs9381040 is greater than that observed with p.S144G. Further, *TREML2* p.S144G does not have a predicted functional consequence (PolyPhen2 score=benign)

or demonstrated functional outcome, suggesting that the IGAP signal at the *TREM* locus may be due to other functional variants.

Some variants at the *TREM* locus have been reported to show association with AD endophenotypes [11, 13, 14]. Cerebrospinal fluid (CSF) levels of AD biomarkers, tau and ptau, associate with three variants at the *TREM* locus that are not in LD with each other: *TREM2* p.R47H (rs75932628), rs6916710 located in intron 2 of *TREML2*, and rs6922617 located downstream from *NCR2* and outside the *TREM* cluster. Of these variants, only *TREM2* p.R47H was associated with AD-risk [13]. More recently, a variant upstream of *TREM2* (rs7759295) and a variant in intron 3 of *TREM1* (rs6910730) were reported to be independently associated with increased AD pathology burden and increased rate of cognitive decline [14]. However, neither of these two variants shows association with AD-risk in the IGAP meta-analysis ($p>0.05$) [15]. Thus, other than *TREM2* p.R47H, none of the *TREM* locus variants previously reported to associate with AD endophenotypes show association with AD-risk. Functional AD-risk variants that influence AD endophenotypes are expected to show association both with these endophenotypes and risk of AD. Therefore, it is possible that these latter variants are not the functional variants *per se*, but merely markers of other un-tested functional variants.

Collectively, these prior findings suggest that besides the *TREM2* rare missense variants, there may be common variants at the *TREM* locus that influence AD-risk and/or its endophenotypes. We hypothesized that some of the common AD-risk variants at the *TREM* locus confer disease risk via regulation of transcript levels of coding genes at the *TREM* gene cluster. In this study, we characterized the brain expression levels of the *TREM* family genes using microarray expression data; validated expression levels by RNA sequencing (RNAseq); performed genetic associations with *TREM* locus genes reliably detected in cerebellum and

temporal cortex with single nucleotide polymorphisms (SNP) that were also tested in the IGAP AD-risk GWAS meta-analysis; and annotated these variants for their effects on *TREM* gene expression levels and regulatory potential. Further, we obtained results for the top putative regulatory SNP from two other, independent cohorts with brain RNAseq data and performed meta-analysis of all three cohorts.

2. Materials and Methods

2.1 Variant selection

We restricted our analysis to variants located within 100kb of any coding *TREM* family gene at the chromosome 6p21.1 *TREM* gene cluster (**Fig. 1**). Variants were further selected based on the statistical significance of their AD-risk association in the IGAP stage 1 meta-analysis [12] (**Supplementary Methods**), where only those variants with p-values ≤ 0.0015 were kept. This p-value cut-off was arbitrarily chosen to select those variants that existed in both the IGAP stage 1 AD GWAS and our discovery eQTL cohort, Mayo Clinic Whole Genome-DASL dataset, and that could be genotyped, if needed, in the replication eQTL cohorts, using cost-effective medium-throughput assays. Variants were further prioritized by their Regulome score. Regulome scores were obtained from the Regulome database, which annotates variants with regulatory information from 962 different datasets and a variety of sources, including ENCODE [16]. Regulome scores are on a scale from 1 to 6, and these numerical categories are sub-classified with letters based on the number of lines of evidence of functional consequence. A value of 1a is assigned to the variant with the most evidence of regulatory potential, while a score of 6 has the least [16].

2.2 Mayo Clinic Whole Genome-DASL dataset (Discovery eQTL cohort)

We utilized Illumina (Whole Genome-DASL=WG-DASL, Illumina, San Diego, CA) microarray gene expression data from our published human brain expression genome-wide association study (Mayo Clinic eGWAS) [17] conducted on brain tissue from autopsied AD patients (197 cerebellum, 202 temporal cortex) and non-AD subjects (177 cerebellum, 197 temporal cortex) (**Table 1**). All AD subjects had neuropathologic diagnosis of definite AD [2]. The non-AD subjects did not fulfill neuropathologic criteria for definite AD, but many had other unrelated pathologies. Expression measures were generated as described previously [17]. A description of this cohort and generation of expression measures is provided in the **Supplemetary Methods**.

2.3 RNAseq datasets (Replication eQTL cohorts)

Temporal cortex RNAseq data from two RNAseq cohorts: "Mayo Clinic RNASeq" and "ROS/MAP RNAseq" were employed for replication of the associations that were detected with the WG-DASL gene expression measurements. The Mayo Clinic RNASeq dataset is comprised of 84 LOAD and 48 non-AD brains from the Mayo Clinic Brain Bank that were not part of the Mayo Clinic WG-DASL cohort but whose neuropathological diagnosis followed the same criteria. The ROS/MAP RNAseq dataset is comprised of RNAseq data from 288 AD and 206 non-AD samples that are part of the ROS/MAP cohort (**Table 1**) previously described [18, 19]. Methodological details for the RNAseq data generation are provided in the **Supplementary Methods**.

2.4. Statistical Analysis

Normalized transcript expression levels, on a log2 scale, were tested for associations with *TREM* locus genotypes in each of the three datasets (Mayo WG-DASL, Mayo Clinic RNAseq and ROS/MAP RNAseq) via multivariable linear regression analyses implemented in PLINK [20]. An additive model was applied adjusting for age-at-death, sex, diagnosis, RNA Integrity Number (RIN) and adjusted RIN squared $(\text{RIN-RINmean})^2$ in all expression analyses, and *APOE* ε4 dosage and PCR plate in Mayo WG-DASL only, and flowcell in the Mayo Clinic RNAseq dataset only. The eQTL analysis in the discovery, WG-DASL dataset, included *APOE* ε4 dose as a covariate given the strong effect of this allele on AD. However, since a significant association was not detected with this covariate in the rs9357347 eQTL analyses in the discovery set, *APOE* ε4 dose was not included in the eQTL analyses implemented on the replication cohorts. For comparison, we have performed the eQTL analyses in all three datasets with and without adjustment for *APOE* ε4 dose and do not observe a substantial difference in the association results between these two models.

Meta-analyses were performed on eQTL results from the three independent datasets. For these analyses, METAL [21] was implemented using weighted average of z-scores from the individual study p-values, weighted according their sample size.

3. Results

In the WG-DASL gene expression data from the temporal cortex (n=399) and cerebellum (n=374) of neuropathologically diagnosed AD and non-AD subjects (**Table 1**), we observed that of the 5 *TREM* locus coding genes, only *TREML1* and *TREM2* were reliably detected (**Table S1 and Fig. 2)**. *TREML1* was detected in both the temporal cortex and cerebellum, while *TREM2* was reliably detected only in the temporal cortex. We validated *TREML1* and *TREM2* WG-

DASL temporal cortex gene expression measurements, using RNAseq data generated from a subset of 93 autopsied AD subjects who also had microarray data. There was highly significant correlation between WG-DASL and RNAseq measurements for both *TREML1* (r_s=0.65, $p<10^{-40}$) and *TREM2* (r_s=0.80, $p<10^{-40}$) (**Fig. S1**).

Variants located within 100kb of the 5' or 3'end of any *TREM* coding gene that demonstrated association with AD-risk in the IGAP stage I meta-analysis (17,800 AD vs. 37,154 controls, p≤0.0015), were evaluated for their association with *TREML1* expression in the temporal cortex and cerebellum, and with *TREM2* expression in the temporal cortex. Of the 1,002 variants tested at this locus in the IGAP stage 1 meta-analysis, 28 had p-values ≤ 0.0015, and 16 of these have been genotyped in the autopsied samples in the Mayo Clinic brain expression genome-wide association study (Mayo eGWAS). We also assessed 5 other variants at this locus previously reported to be associated with either reduced AD-risk (rs3747742) [11], increased AD pathology burden and cognitive decline (rs6910730, rs7759295) [14], or decreased CSF tau levels (rs6916710, rs6922617) [13]. **Table 2** shows the association of *TREML1* and *TREM2* gene expression with these 21 variants. In 399 combined AD and non-AD temporal cortex samples tested for the 16 IGAP variants, 5 SNPs showed association (uncorrected p<0.05) with increased levels of both *TREML1* and *TREM2* (rs9381040, rs2093395, rs9357347, rs9394778, rs9296359), and a sixth variant (rs9394767) was significantly associated with increased *TREML1* levels only. As shown in **Fig. 3**, four of the six variants that associate with increased levels of *TREML1* and *TREM2* are in a single LD block (block 2: rs9357347, rs9381040, rs2093395 and rs9394767) and in tight linkage disequilibrium with each other (r^2≥0.90). Of these variants, rs9381040 has the most significant AD-risk association in the IGAP stage 1 meta-analysis (**Table 2**). This IGAP "hit" is located 5.5kb downstream from *TREML2*

and 23.7kb upstream from *TREM2* and is associated with *TREML1* and *TREM2* expression (p=0.0083, beta=0.086 and p=0.048, beta=0.091, respectively). Given that the expression measures were on a log2 scale, these changes in expression are equivalent to *TREML1* and *TREM2* fold-changes of 1.06 and 1.07, for each copy of the minor allele, respectively. Notably, the minor allele of the IGAP "hit" rs9381040 is associated with both decreased AD-risk and increased *TREML1* and *TREM2* levels. However, based on data from the Roadmap Epigenomics Consortium [22], rs9381040 lacks evidence of regulatory potential in brain regions relevant to AD.

The variant with the most significant association with brain *TREML1* expression, which also associates with *TREM2* levels, is rs9357347 in block 2 (**Fig. 3**). This SNP is located 6.9kb downstream from *TREML2* and 19.6kb upstream from *TREM2* and is in tight LD with the IGAP "hit" rs9381040 (D'=0.99, r^2=0.96). As expected, the minor allele of rs9357347 is associated with reduced AD-risk (OR=0.95, 95% CI=0.91-0.98, p=0.001) in the IGAP meta-analysis [12] and with increased *TREML1* and *TREM2* expression in the temporal cortex in the Mayo Clinic WG-DASL eQTL analysis (p=0.0063, beta=0.088 and p=0.046, beta=0.090, respectively) (**Table 2 and Fig. S2**). These beta coefficients can be interpreted as an estimated 1.06-fold change of both *TREML1* and *TREM2*, per rs9357347 minor allele, in this temporal cortex dataset. Unlike the IGAP "hit" (rs9381040), rs9357347 lies within sequence subject to histone modifications and within a DNAse hypersensitive site detected by the Roadmap Epigenomics Consortium [22] in brain regions relevant to AD pathology such as the hippocampus. Furthermore, this variant is predicted to affect transcription factor binding (SP1 and PPAR) as catalogued in HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) [23]. Consequently, it has a compelling Regulome score of 2b (http://www.regulomedb.org/) due to

the evidence of its regulatory potential [16] (**Table 2**). Indeed, of all the variants with an AD-risk p-value<0.0015 in the IGAP meta-analysis, and p-values<0.05 in our WG-DASL eQTL analysis of temporal cortex *TREML1* and *TREM2* gene expression levels, rs9357347 had the greatest regulatory potential as determined by their Regulome scores (**Fig. S3 and Fig. S4**).

The other two variants with gene expression associations in the temporal cortex are in a different LD block (block 4: rs9394778 and rs9296359) and in tight LD with each other (r^2 = 0.67). These SNPs are more significantly associated with *TREM2* than with *TREML1* expression; however, neither has compelling evidence of regulatory potential as both have Regulome scores of 6 (**Table 2**). In the 374 AD and non-AD subjects with cerebellum expression measures, none of the 16 IGAP AD-risk associated variants that were tested, associate with *TREML1* gene expression (p>0.05).

We determined the extent of linkage disequilibrium (LD) between the likely regulatory variant rs9357347, the IGAP "hit" rs9381040 and the significant *TREM2* rare missense AD-risk variants p.D87N (rs142232675) and p.R47H (rs75932628) [1]. As shown in **Fig. 3**, these two *TREM2* rare missense AD-risk variants are not in LD with either rs9357347 or rs9381040. This suggests that the protective effect of the regulatory rs9357347 and the IGAP "hit" are independent of the rare, missense *TREM2* variants.

We next evaluated LD amongst variants tested at this locus, including common *TREM* locus variants previously reported to have associations with AD-risk (rs3747742) [11], increased AD pathology burden and cognitive decline (rs7759295 and rs6910730) [14], or with lower CSF ptau (rs6922617 and rs6916710) [13]. The missense *TREML2* variant rs3747742 (p.S144G) is in LD with the regulatory variant implicated in our study, rs9357347. As reported, rs3747742 is

also in LD with rs9381040 (IGAP hit); and as expected associates with reduced AD-risk (p=0.009), however with slightly lesser significance than the AD-risk association of the regulatory rs9357347 (p=0.001) or the IGAP "hit" rs9381040 (0.0006). Further, the association of rs3747742 with *TREML1* expression is not as significant as that of rs9357347. In addition, rs3747742 has no association with brain *TREM2* levels, and has a weak Regulome score of 6 (**Table 2**).

Of the four common *TREM* locus variants that associate with AD endophenotypes, only rs6916710 is in tight LD with the regulatory rs9357347 (D'=0.91, r^2=0.62). However, rs6916710, does not show significant association with AD-risk in the IGAP meta-analysis (p=0.103) nor with *TREML1* or *TREM2* gene expression levels (**Table 2**).

None of the other three common *TREM* locus variants with reported AD-endophenotype associations are in tight LD with the regulatory rs9357347 or any of the other *TREM* locus variants that are associated with AD-risk. Only rs7759295 showed association with *TREML1* gene expression (uncorrected p=0.04), but neither this nor any of the other AD-endophenotype-associated SNPs have evidence of AD-risk association or Regulome scores that are indicative of likely regulatory function (**Fig. 3 and Table 2**).

Utilizing publicly available RNAseq data from two independent cohorts (**Table 1**) that do not overlap with the samples included in the WG-DASL eQTL analysis, we sought replication of the rs9357347 association with *TREML1* and *TREM2*. Although in the ROS/MAP RNAseq dataset a significant association was only detected with the levels of *TREM2* (**Table 3**), meta-analysis from the three independent study p-values (Mayo WG-DASL, Mayo RNAseq and ROS/MAP RNAseq) yielded significant results (*TREML1* p=3.4x10^{-2}; *TREM2* p=3.5x10^{-3}),

confirming the association of the rs9357347 minor allele with increased *TREML1* and *TREM2* gene expression. The evidence of association with *TREM2* expression was greater upon meta-analysis compared to the association observed in our discovery dataset; whereas the evidence of association with *TREML1* expression was slightly greater in our discovery dataset compared to the meta-analysis.

4. Discussion

In this study, we first sought to characterize the brain expression of *TREM* locus genes based on the premise that those *TREM* cluster genes that are expressed in the brain are likely to be candidate AD-risk genes. We determined that besides *TREM2*, only *TREML1* has reliable expression in the brain regions we studied. Whereas *TREML1* is expressed in both cerebellum and temporal cortex of all subjects, *TREM2* is expressed in 98% of temporal cortex and 41% of cerebellum samples. This suggests that cerebellar levels of *TREM2* are lower than those for temporal cortex, consistent with previous reports showing higher gene levels in the temporal cortex than cerebellum [24] and higher protein levels correlating with AD neuropathology [25]. In contrast, *TREM1*, *TREML2* and *TREML4* are expressed in only 0%-17% of the subjects. While lack of reliable brain expression of these genes does not definitively rule them out as plausible AD-risk genes, our findings provide the strongest evidence for *TREML1*, besides *TREM2*, as most likely *TREM* locus genes for further studies in AD.

Consequently, we focused our studies on *TREML1* and *TREM2*; and utilized their brain expression levels as endophenotypes to identify putative regulatory variants that modify risk for AD. Focusing on brain *TREML1* and *TREM2* expression associations with variants at the *TREM* locus that also show evidence of AD-risk association in the publicly available IGAP meta-analysis, we identified a putative regulatory variant, rs9357347, located between *TREM2* and

TREML2. The minor allele of this variant is associated with both decreased AD-risk and with increased *TREML1* and *TREM2* brain expression in the temporal cortex. The direction of effect of this variant on AD-risk and brain expression levels of these genes appears to be biologically congruent based on the known functions of these genes.

TREML1, which is also known as TREM-like transcript 1 (TLT-1), is a myeloid receptor expressed exclusively in the α-granules of platelets and megakaryocytes [26]. Identification of higher levels of soluble TREML1 (sTLT-1) in septic patients vs. controls and development of hemorrhage in mice lacking *Treml1* when exposed to inflammatory injury led to the conclusion that TREML1 functions to maintain vascular integrity during inflammation [27]. Further, TREML1 was shown to dampen leukocyte activation during sepsis, and inhibited pro-inflammatory activation of TREM1 by competing with its ligand [28]. These studies strongly support a role for TREML1 in promoting vascular homeostasis and limiting inflammation.

Functional, *in-vitro* studies of *TREM2* rare, missense mutations revealed reduced TREM2 function as a consequence of decreased maturation and ectodomain shedding, also supported by findings of decreased soluble TREM2 levels in the cerebrospinal (CSF) levels of patients with these mutations [13, 29]. TREM2 deficiency also led to increased amyloid pathology and neuronal loss in the 5XFAD mouse model of AD [30]. Interestingly, TREM2 deficiency in an ischemic mouse model resulted in reduced phagocytosis and resorption of infarcted brain tissue, and worse neurological recovery [31]. Collectively, these findings support a neuroprotective role for TREM2 in various neuronal injury models. There are, however, studies with contradictory results for TREM2. In a different mouse model of AD (APP/PS1), knock-out of *Trem2*, resulted in reduction of macrophages infiltrating from the periphery, along with less brain inflammation and reduced amyloid and tau pathology [32]. These opposite findings of *Trem2* knock-out could

be due to differences in the mouse models of Alzheimer's disease tested, different *Trem2* knockout mouse lines, and analyses performed at different time points (early stages versus later stages of Alzheimer's disease).

Given these collective data, a regulatory variant that enhances levels of *TREML1* in platelets, and levels of *TREM2* in brain resident microglia could conceivably promote vascular homeostasis and limit inflammatory damage to neurons in AD and potentially other nervous system diseases. Indeed, rs9357347 has compelling evidence of regulatory potential as it is located in a known DNase hypersensitive site and affects histone modification in the hippocampus and transcription factor binding, according to the evidence compiled in the Regulome database and HaploReg [16, 23]. Interestingly, rs9357347 is predicted to affect transcription factor binding (SP1 and PPAR) as catalogued in HaploReg. These two transcription factors are known be important in regulating key players in the inflammatory response and lipid metabolism [33, 34]. Further, rs9357347 shows the most significant association with *TREML1* gene expression amongst variants at the *TREM* locus with IGAP meta-analysis AD-risk p-values≤0.0015, in addition to its association with brain *TREM2* levels.

The regulatory rs9357347 SNP is in the same haplotype block as the variant with the most significant AD-risk association at the *TREM* locus in the IGAP meta-analysis, rs9381040, which is an intergenic variant downstream of *TREML2*. Though this IGAP *TREM* locus "hit" SNP has greater evidence of AD-risk association than rs9357347, there is no evidence of regulatory potential for rs9381040 in brain regions relevant to AD.

While the fold change estimates in gene expression associated with rs9357347-C are modest at 6-7%, the biological impact of the increase attributed to each copy of the minor allele,

can be significant and may provide sufficient protection from disease in some individuals, particularly when considered over a lifetime. Furthermore, these estimates are based on RNA isolated from tissue samples and not microglial cells where both *TREM2* and *TREML1* are predominantly expressed [35], and where expression levels of these genes may be impacted to a greater extent by regulatory variants. Additional studies will be needed to determine the impact of such expression changes on the biology of microglial cell function.

The *TREML2* p.S144G variant [11], which associates with reduced AD-risk, is also in LD with both rs9357347 and rs9381040. Though proposed to be the functional variant that accounts for the IGAP signal at this locus, *TREML2* p.S144G is not predicted to have a functional consequence based on PolyPhen2 nor does it have evidence of regulatory potential. Further, *TREML2* expression is too low to be reliably measured in brain tissue (TCX and CER). This raises the possibility that the association with *TREML2* p.S144G is due to its LD with a functional variant(s) that influences the function or level of a nearby *TREM* gene(s), such as *TREML1* or *TREM2*. Alternatively, the protective effect of p.S144G could be mediated directly through the function of TREML2 in a cell with abundant expression, such as macrophages, in which TREML2 is known to be upregulated in response to inflammation, [36]. It is also possible that significant rs9357347 eQTL associations would be detected with *TREML2* or other *TREM* locus transcripts in tissues were these genes are more abundantly expressed.

Our findings therefore challenge the conclusion that p.S144G is the only functional variant accounting for the protective effect detected in the IGAP meta-analysis at this locus, and propose rs9357347 as an alternative functional variant with regulatory effects. In reality, both variants could have functional consequences and contribute to the IGAP signal. It should be emphasized that, as demonstrated in our LD analysis, *TREM2* p.R47H is not in LD with these

two variants, and thus affects AD-risk independently. Both rs9357347 and p.S144G should be tested for their functional potential and influence on outcomes of inflammation and neuroprotection. It remains possible that rs9357347 is in LD with an untested true functional variant with effects on transcription and AD-risk. It is likewise possible that while rs9357347 is associated with both AD-risk and gene expression levels, these joint effects are coincidental due to LD, rather than being related. These possibilities need to be explored through sequencing of the entire *TREM* locus, or via targeted sequencing of LD block 2 where rs9357347 resides. Thus, our findings provide a testable hypothesis for a strong candidate functional variant, specific transcription factors and their effects on *TREML1* and *TREM2* levels.

Furthermore, our investigation of variants previously shown to associate with AD-related endophenotypes [13-15] suggests that these are unlikely to be functional AD-risk variants *per se*, though it remains possible that they are markers of functional variants at the *TREM* locus.

In summary, we characterized expression of *TREM* genes in cerebellum and temporal cortex and determined *TREML1* and *TREM2* to be the only reliably expressed *TREM* genes in these brain regions. We identified rs9357347 as a putative regulatory variant that is associated with protection from AD and with increased *TREML1* and *TREM2* brain levels, and nominate rs9357347 as one of the functional variants that accounts for the IGAP AD-risk signal. Additional studies are needed to validate the function of this variant, and to explore the possibility of the presence of other variants at this locus that could contribute to associations observed with rs9357347. Importantly, these findings suggest a potential link between *TREML1* and *TREM2*, as well as vascular homeostasis and neuroinflammation as related mediators of neuronal protection and injury in AD and possibly other central nervous system diseases.

Acknowledgements

We thank the patients and their families for their participation, without whom these studies would not have been possible, and the clinicians, technicians, and administrative staff who helped in the implementation of this study.

This work was supported by the Alzheimer's Association [MNIRGD 2013 award to M.M.C]; Mayo Alzheimer's Disease Research Center [P50 AG0016574 to D.W.D, N.E.T, N.R.G.-R., R.C.P. and S.G.Y.]; National Institute on Aging [R01 AG025711, AG017216, AG003949 to D.W.D.; R01 AG032990 to N.E.T.; R01 AG018023 to N.R.G.-R. and S.G.Y.; and U01 AG046139 to N.E.T., T.E.G, N.P. and S.G.Y.]; National Institute of Neurological Disorders and Stroke [R01 NS080820 to N.E.T].

Conflict of Interest Statement

Dr. Petersen has been a consultant to Genentech, Inc. Merck, Inc. and Roche, Inc. and has served on a data safety monitoring committee for Pfizer and Janssen Alzheimer Immunotherapy. Dr. Graff-Radford has multicenter treatment study grants from Lilly, TauRx and consulted for Cytox. Dr. Ertekin-Taner consulted for Cytox.

References

[1] R. Guerreiro, A. Wojtas, J. Bras, M. Carrasquillo, E. Rogaeva, E. Majounie, et al., TREM2 variants in Alzheimer's disease. The New England journal of medicine 2013;368:117-127.

[2] T. Jonsson, H. Stefansson, S. Steinberg, I. Jonsdottir, P.V. Jonsson, J. Snaedal, et al., Variant of TREM2 associated with the risk of Alzheimer's disease. The New England journal of medicine 2013;368:107-116.

[3] J.W. Ford,D.W. McVicar, TREM and TREM-like receptors in inflammation and disease. Current opinion in immunology 2009;21:38-46.

[4] R.E. Tanzi, TREM2 and Risk of Alzheimer's Disease--Friend or Foe? The New England journal of medicine 2015;372:2564-2565.

[5] C. Pottier, D. Wallon, S. Rousseau, A. Rovelet-Lecrux, A.C. Richard, A. Rollin-Sillaire, et al., TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. Journal of Alzheimer's disease : JAD 2013;35:45-49.

[6] B.A. Benitez, B. Cooper, P. Pastor, S.C. Jin, E. Lorenzo, S. Cervantes, et al., TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiology of aging 2013;34:1711 e1715-1717.

[7] J.D. Gonzalez Murcia, C. Schmutz, C. Munger, A. Perkes, A. Gustin, M. Peterson, et al., Assessment of TREM2 rs75932628 association with Alzheimer's disease in a population-based sample: the Cache County Study. Neurobiology of aging 2013;34:2889 e2811-2883.

[8] A. Ruiz, O. Dols-Icardo, M.J. Bullido, P. Pastor, E. Rodriguez-Rodriguez, A. Lopez de Munain, et al., Assessing the role of the TREM2 p.R47H variant as a risk factor for Alzheimer's disease and frontotemporal dementia. Neurobiology of aging 2014;35:444 e441-444.

[9] B.V. Hooli, A.R. Parrado, K. Mullin, W.K. Yip, T. Liu, J.T. Roehr, et al., The rare TREM2 R47H variant exerts only a modest effect on Alzheimer disease risk. Neurology 2014;83:1353-1358.

[10] C.M. Lill, A. Rengmark, L. Pihlstrom, I. Fogh, A. Shatunov, P.M. Sleiman, et al., The role of TREM2 R47H as a risk factor for Alzheimer's disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association 2015;

[11] B.A. Benitez, S.C. Jin, R. Guerreiro, R. Graham, J. Lord, D. Harold, et al., Missense variant in TREML2 protects against Alzheimer's disease. Neurobiology of aging 2014;35:1510 e1519-1526.

[12] J.C. Lambert, C.A. Ibrahim-Verbaas, D. Harold, A.C. Naj, R. Sims, C. Bellenguez, et al., Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics 2013;45:1452-1458.

[13] C. Cruchaga, J.S. Kauwe, O. Harari, S.C. Jin, Y. Cai, C.M. Karch, et al., GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. Neuron 2013;78:256-268.

[14] J.M. Replogle, G. Chan, C.C. White, T. Raj, P.A. Winn, D.A. Evans, et al., A TREM1 variant alters the accumulation of Alzheimer-related amyloid pathology. Annals of neurology 2015;77:469-477.

[15] G. Liu, No association of TREM1 rs6910730 and TREM2 rs7759295 with Alzheimer's disease. Annals of Neurology 2015;epub ahead opf print:

[16] A.P. Boyle, E.L. Hong, M. Hariharan, Y. Cheng, M.A. Schaub, M. Kasowski, et al., Annotation of functional variation in personal genomes using RegulomeDB. Genome research 2012;22:1790-1797.

[17] F. Zou, H.S. Chai, C.S. Younkin, M. Allen, J. Crook, V.S. Pankratz, et al., Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS genetics 2012;8:e1002707.

[18] D.A. Bennett, J.A. Schneider, Z. Arvanitakis,R.S. Wilson, Overview and findings from the religious orders study. Current Alzheimer research 2012;9:628-645.

[19] D.A. Bennett, J.A. Schneider, A.S. Buchman, C. Mendes de Leon, J.L. Bienias,R.S. Wilson, The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. Neuroepidemiology 2005;25:163-175.

[20] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 2007;81:559-575.
[21] C.J. Willer, Y. Li,G.R. Abecasis, METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-2191.
[22] B.E. Bernstein, J.A. Stamatoyannopoulos, J.F. Costello, B. Ren, A. Milosavljevic, A. Meissner, et al., The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology 2010;28:1045-1048.
[23] L.D. Ward,M. Kellis, HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic acids research 2016;44:D877-881.
[24] P. Forabosco, A. Ramasamy, D. Trabzuni, R. Walker, C. Smith, J. Bras, et al., Insights into TREM2 biology by network analysis of human brain gene expression data. Neurobiology of aging 2013;34:2699-2714.
[25] L.F. Lue, C.T. Schmitz, G. Serrano, L.I. Sue, T.G. Beach,D.G. Walker, TREM2 Protein Expression Changes Correlate with Alzheimer's Disease Neurodegenerative Pathologies in Post-Mortem Temporal Cortices. Brain pathology 2015;25:469-480.
[26] A.V. Washington, R.L. Schubert, L. Quigley, T. Disipio, R. Feltz, E.H. Cho, et al., A TREM family member, TLT-1, is found exclusively in the alpha-granules of megakaryocytes and platelets. Blood 2004;104:1042-1047.
[27] A.V. Washington, S. Gibot, I. Acevedo, J. Gattis, L. Quigley, R. Feltz, et al., TREM-like transcript-1 protects against inflammation-associated hemorrhage by facilitating platelet aggregation in mice and humans. The Journal of clinical investigation 2009;119:1489-1501.
[28] M. Derive, Y. Bouazza, N. Sennoun, S. Marchionni, L. Quigley, V. Washington, et al., Soluble TREM-like transcript-1 regulates leukocyte activation and controls microbial sepsis. Journal of immunology 2012;188:5585-5592.
[29] G. Kleinberger, Y. Yamanishi, M. Suarez-Calvet, E. Czirr, E. Lohmann, E. Cuyvers, et al., TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. Science translational medicine 2014;6:243ra286.
[30] Y. Wang, M. Cella, K. Mallinson, J.D. Ulrich, K.L. Young, M.L. Robinette, et al., TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. Cell 2015;160:1061-1071.
[31] M. Kawabori, R. Kacimi, T. Kauppinen, C. Calosing, J.Y. Kim, C.L. Hsieh, et al., Triggering receptor expressed on myeloid cells 2 (TREM2) deficiency attenuates phagocytic activities of microglia and exacerbates ischemic damage in experimental stroke. The Journal of neuroscience : the official journal of the Society for Neuroscience 2015;35:3384-3396.
[32] T.R. Jay, C.M. Miller, P.J. Cheng, L.C. Graham, S. Bemiller, M.L. Broihier, et al., TREM2 deficiency eliminates TREM2+ inflammatory macrophages and ameliorates pathology in Alzheimer's disease mouse models. The Journal of experimental medicine 2015;212:287-295.
[33] A.Y. Wen, K.M. Sakamoto,L.S. Miller, The role of the transcription factor CREB in immune function. Journal of immunology 2010;185:6413-6419.
[34] J.D. Brown,J. Plutzky, Peroxisome proliferator-activated receptors as transcriptional nodal points and therapeutic targets. Circulation 2007;115:518-533.
[35] Y. Zhang, K. Chen, S.A. Sloan, M.L. Bennett, A.R. Scholze, S. O'Keeffe, et al., An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 2014;34:11929-11947.
[36] R.G. King, B.R. Herrin,L.B. Justement, Trem-like transcript 2 is expressed on cells of the myeloid/granuloid and B lymphoid lineage and is up-regulated in response to inflammation. Journal of immunology 2006;176:6012-6021.

Figure Legends

Fig. 1. *TREM* gene cluster on Chr 6p21.1. The chromosomal positions are based on the human genome assembly from February 2009 (GRCh37/hg19). There are seven RefSeq genes at the *TREM* locus (*TREM1*, *TREML1*, *TREM2*, *TREML2*, *TREML3P*, *TREML4* and *TREML5P*); however, *TREML3P* and *TREML5P* are non-coding pseudogenes. The transcript figures are taken from the UCSC Genome Browser.

Fig. 2. Location of *TREML1* and *TREM2* WG-DASL probes.

The location of the (A) *TREML1* and (B) *TREM2* WG-DASL probes (highlighted in light blue) are shown relative to their Refseq transcripts. The chromosomal positions are based on the human genome assembly from February 2009 (GRCh37/hg19). As shown, both of these probes are complementary to all RefSeq transcripts for the respective gene. The transcript figures are taken from the UCSC Genome Browser.

Fig. 3. LD Plot of *TREM* locus variants.

LD plot of *TREM* locus variants where haplotype blocks were determined with the solid spine definition; square colors correspond to D' (tight LD=warmer colors, weak LD=cooler colors) and r^2 values are shown within the squares (**Supplementary Methods**). Red circles: The rare *TREM2* AD-risk missense variants rs142232675 (p.D87N) and rs75932628 (p.R47H) [1]. Blue circles: Variants that associate with increased AD pathology burden and cognitive decline (rs7759295 and rs6910730) [14], or with lower CSF ptau (rs6922617 and rs6916710) [13]. Green circles: The variant with the most significant AD-risk association in the IGAP meta-analysis (rs9381040); rs9357347, which has the most significant*TREML1* gene expression

association, also shows association with *TREM2* gene expression, IGAP AD-risk association and the best Regulome score within all tested SNPs; and rs9296359 which has the most significant association with *TREM2* expression. RefSeq gene transcripts are shown above the LD plot relative to the variant position according to the February 2009 human genome assembly (GRCh37hg19) across the targeted genomic region (*TREM* gene +/-100 kb: chr6:41016999-41354457).

Tables

Table 1. Description of samples included in the discovery and replication cohorts utilized for eQTL analysis.

	Mayo Clinic WG-DASL				Mayo Clinic RNAseq		ROS/MAP RNAseq	
	CER		TCX		TCX		PFCX	
	AD	Non-AD	AD	Non-AD	AD	non-AD	AD	non-AD
N	197	177	202	197	84	48	288	206
Mean age +/- SD	73.6 ± 5.6	71.7 ± 5.5	73.6 ± 5.5	71.6 ± 5.6	83.2 ± 8.7	85.7 ± 8.3	89.8 ± 5.8	86.5 ± 7.2
Female, N (%)	101 (51%)	63 (36%)	108 (53%)	78 (40%)	48 (57%)	26 (54%)	186 (65%)	121 (59%)
% *APOE* ε4+	64%	25%	61%	25%	51%	17%	34%	12%

Samples included in the Mayo Clinic eGWAS (discovery cohort), with cerebellar (CER) and temporal cortex (TCX) gene expression measurements from Illumina WG-DASL arrays have been previously described [17]. Samples in the Mayo Clinic RNAseq cohort (replication cohort #1) had temporal cortex gene expression measurements, and did not overlap with the Mayo eGWAS (WG-DASL) cohort. The ROS/MAP RNAseq cohort (replication cohort #2) had dorsolateral prefrontal cortex (PFCX) gene expression measurements, and did not overlap with the Mayo eGWAS (WG-DASL), or with the Mayo Clinic RNAseq cohort. The RNAseq data for these two cohorts is available at the Sage Synapse, AMP AD Knowledge Portal (https://www.synapse.org/#!Synapse:syn2580853/wiki/66722), under synapse IDs syn3388564 (ROS/MAP RNAseq) and syn3163039 (Mayo RNAseq).

Table 2. Association of variants at the *TREM* locus with AD-risk and *TREM* WG-DASL brain gene expression levels.

Chr	SNP	Position hg19	AD-Risk (IGAP Stage1 Meta-analysis)				Brain eQTL (Mayo Clinic eGWAS)						Regulome Score	HapMap CEU MAF
			Effect Allele	Non Effect Allele	OR (95% CI)	P-value	TREML1 CER BETA	TREML1 CER P	TREML1 TCX BETA	TREML1 TCX P	TREM2 TCX BETA	TREM2 TCX P		
6	**rs9381040**	**41,154,650**	**T**	**C**	**0.94 (0.91 - 0.98)**	**5.97E-04**	**0.021**	**2.30E-01**	**0.086**	***8.30E-03***	**0.091**	***4.80E-02***	**NA**	**26.70%**
6	rs2093395	41,155,026	C	G	0.94 (0.91 - 0.98)	6.40E-04	0.021	2.30E-01	0.086	*8.30E-03*	0.091	*4.80E-02*	6	27.90%
6	rs2038568	41,158,132	C	G	1.14 (1.05 - 1.23)	7.93E-04	0.018	7.80E-01	-0.08	3.90E-01	-0.186	1.60E-01	5	8.30%
6	rs12194214	41,028,574	C	A	1.16 (1.06 - 1.26)	8.36E-04	-0.081	9.40E-02	-0.104	2.70E-01	-0.129	3.20E-01	6	4.20%
6	rs9462675	41,153,238	A	G	1.15 (1.06 - 1.25)	9.54E-04	-0.015	7.50E-01	-0.114	1.60E-01	-0.207	6.50E-02	5	3.60%
6	rs6933067	41,133,522	C	T	1.15 (1.06 - 1.25)	1.07E-03	-0.013	7.60E-01	-0.098	2.10E-01	-0.134	2.10E-01	7	3.50%
6	**rs9357347**	**41,150,591**	**C**	**A**	**0.95 (0.91 - 0.98)**	**1.10E-03**	**0.013**	**4.60E-01**	**0.088**	***6.30E-03***	**0.09**	***4.60E-02***	**2b**	**28.10%**
6	rs9394767	41,159,905	G	A	0.95 (0.91 - 0.98)	1.14E-03	0.011	5.70E-01	0.096	*6.50E-03*	0.083	1.00E-01	5	28.80%
6	rs1542638	41,286,604	G	A	1.06 (1.02 - 1.09)	1.14E-03	-0.022	2.20E-01	-0.035	2.90E-01	-0.064	1.60E-01	4	28.30%
6	rs9471491	41,153,622	A	C	1.15 (1.05 - 1.26)	1.31E-03	-0.015	7.50E-01	-0.114	1.60E-01	-0.207	6.50E-02	7	3.50%
6	rs9471495	41,157,372	A	C	1.15 (1.05 - 1.25)	1.40E-03	0.014	8.30E-01	-0.099	2.90E-01	-0.235	7.20E-02	7	3.50%
6	rs9462677	41,158,856	A	T	1.15 (1.05 - 1.25)	1.41E-03	0.016	8.10E-01	-0.099	2.90E-01	-0.235	7.30E-02	7	4.30%
6	rs9394778	41,215,058	A	G	0.95 (0.92 - 0.98)	1.44E-03	0.015	3.30E-01	0.065	*2.70E-02*	0.099	*1.50E-02*	6	39.80%
6	rs9471494	41,157,344	G	C	1.15 (1.05 - 1.25)	1.46E-03	0.01	8.70E-01	-0.102	2.60E-01	-0.221	8.20E-02	6	4.50%
6	rs6912013	41,061,593	C	T	1.15 (1.05 - 1.25)	1.48E-03	-0.076	1.20E-01	-0.104	2.70E-01	-0.124	3.40E-01	5	2.70%
6	rs9296359	41,205,690	A	G	0.95 (0.92 - 0.98)	1.48E-03	0.017	2.80E-01	0.066	*2.40E-02*	0.116	*4.60E-03*	6	27.40%
6	rs3747742*	41,162,518	C	T	0.96 (0.92 - 0.99)	8.56E-03	0.018	2.90E-01	0.072	*2.30E-02*	0.064	1.50E-01	6	28.30%
6	rs6916710*	41,164,788	T	C	0.97 (0.94 - 1.01)	1.03E-01	0.013	4.30E-01	0.054	7.70E-02	0.072	9.20E-02	7	38.40%
6	rs7759295*	41,135,850	T	C	0.98 (0.93 - 1.03)	3.66E-01	-0.023	3.50E-01	0.094	*4.00E-02*	-0.008	9.00E-01	6	13.30%
6	rs6910730*	41,246,633	G	A	0.99 (0.94 - 1.04)	6.86E-01	-0.046	8.50E-02	-0.079	1.20E-01	-0.032	6.50E-01	4	8.40%
6	rs6922617*	41,336,101	A	G	0.99 (0.93 - 1.05)	6.98E-01	-0.033	2.60E-01	-0.098	7.40E-02	0.011	8.90E-01	7	8.50%

Shown are variants located within 100kb of a *TREM* gene that had an AD-risk $p \leq 0.0015$ in the IGAP stage 1 meta-analysis (top 16 rows), as well as 5 common *TREM* locus variants with previous reports of AD-risk or endophenotype association (bottom 5 rows, SNP marked with an *). AD-risk association results are from the publicly available IGAP meta-analysis stage 1. Brain gene expression associations are from the Mayo Clinic eGWAS and based on cerebellar (CER) and temporal cortex (TCX) gene expression measurements with Illumina WG-DASL arrays with *TREML1* probe ILMN_1690783 and *TREM2* probe ILMN_1701248. Variants showing association with gene expression (uncorrected $p<0.05$) are underlined and in italic font. The variant with the most significant AD-risk association in the IGAP meta-analysis (**rs9381040**), and the variant with the most significant gene expression association and best Regulome score (**rs9357347**) are in bold font. OR (95% CI): odds ratio and 95% confidence interval. Given that the eGWAS expression measures were on a log2 scale, fold-change for the Mayo eGWAS beta coefficients = 2^{beta}.

Table 3. Meta-analysis of rs9357347 eQTL results from three independent datasets.

Dataset	Sample size	MAF	*TREML1*			*TREM2*		
			beta	SE	p	beta	SE	p
Mayo WG-DASL	380	0.307	0.088	0.032	6.28E-03	0.090	0.045	4.61E-02
Mayo Clinic RNAseq	132	0.311	-0.030	0.108	7.82E-01	0.084	0.128	5.13E-01
ROS/MAP RNAseq	494	0.281	0.089	0.114	4.36E-01	0.124	0.060	3.77E-02
Meta-analysis	1006		**+-+**		**3.36E-02**	**+++**		**3.54E-03**

Meta-analysis of rs9357347 eQTL results from temporal cortex (Mayo WG-DASL and Mayo Clinic RNAseq) and dorsolateral prefrontal cortex samples (ROS/MAP). MAF = minor allele frequency. SE = standard error. Since in all three datasets the expression measures analyzed were on a log2 scale, fold-change for the beta coefficients = 2^{beta}. The meta-analysis was performed using METAL, with weighted average of z-scores from the individual study p-values, weighted according their sample size.

Carrasquillo_etal_TREM_SupplMaterial_08-18-2016_changes-accepted

Click here to download Supplementary files: Carrasquillo_etal_TREM_SupplMaterial_08-18-2016_changes-accepted.docx

Research in Context

Systematic review: We performed a comprehensive review of existing literature investigating the role of the *TREM* locus in AD. Although the involvement of *TREM* genes in AD pathophysiology and the underlying variants modifying AD-risk remain unclear, there have been several studies demonstrating association with AD risk and its endophenotypes.

Interpretation: We hypothesized that some variants at the *TREM* locus may modify AD-risk via regulation of *TREM* gene expression. We found a variant in a regulatory region (rs9357347-C) at the *TREM* locus that associates with reduced AD risk and higher *TREML1* and *TREM2* brain gene expression.

Future directions: Our findings nominate regulation of brain *TREML1* and *TREM2* as a potential mechanism for AD risk modification by *TREM* locus variants. In-depth sequencing of the *TREM* locus is needed to fully characterize regulatory variants at this locus that may modify AD-risk.

Highlights

- A regulatory SNP located 5’ *TREM2*, rs9357347-C, associates with reduced AD-risk.
- *TREM2* and *TREML1* are the only *TREM* cluster genes with reliable brain expression.
- Higher brain levels of *TREM2* and *TREML1* associate with rs9357347-C.
- rs9357347 is predicted to affect transcription factor binding (SP1 and PPAR).
- Increased gene expression of *TREML1* and *TREM2* may reduce AD-risk.

Carrasquillo_etal_TREM_Fig1_TREM_cluster_UCSC_081716

Click here to download high resolution image

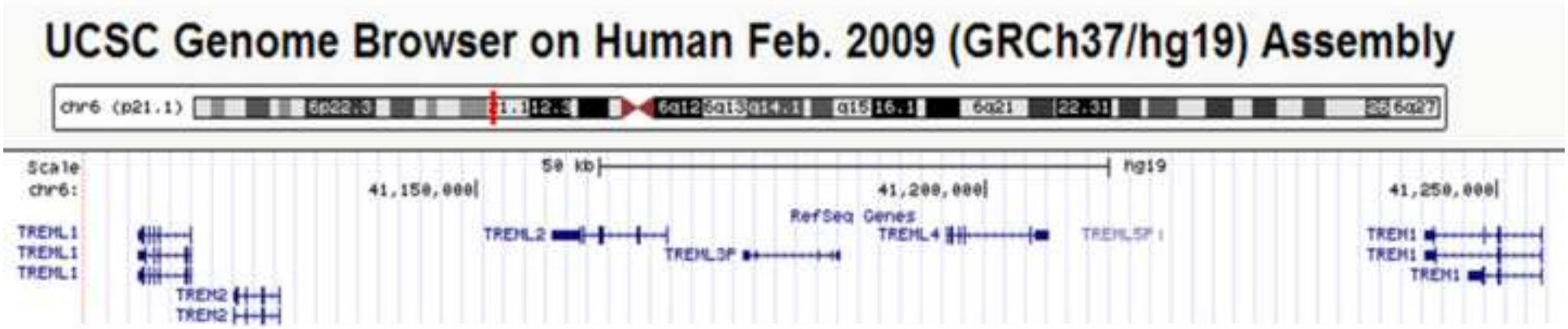


Figure2

Click here to download high resolution image

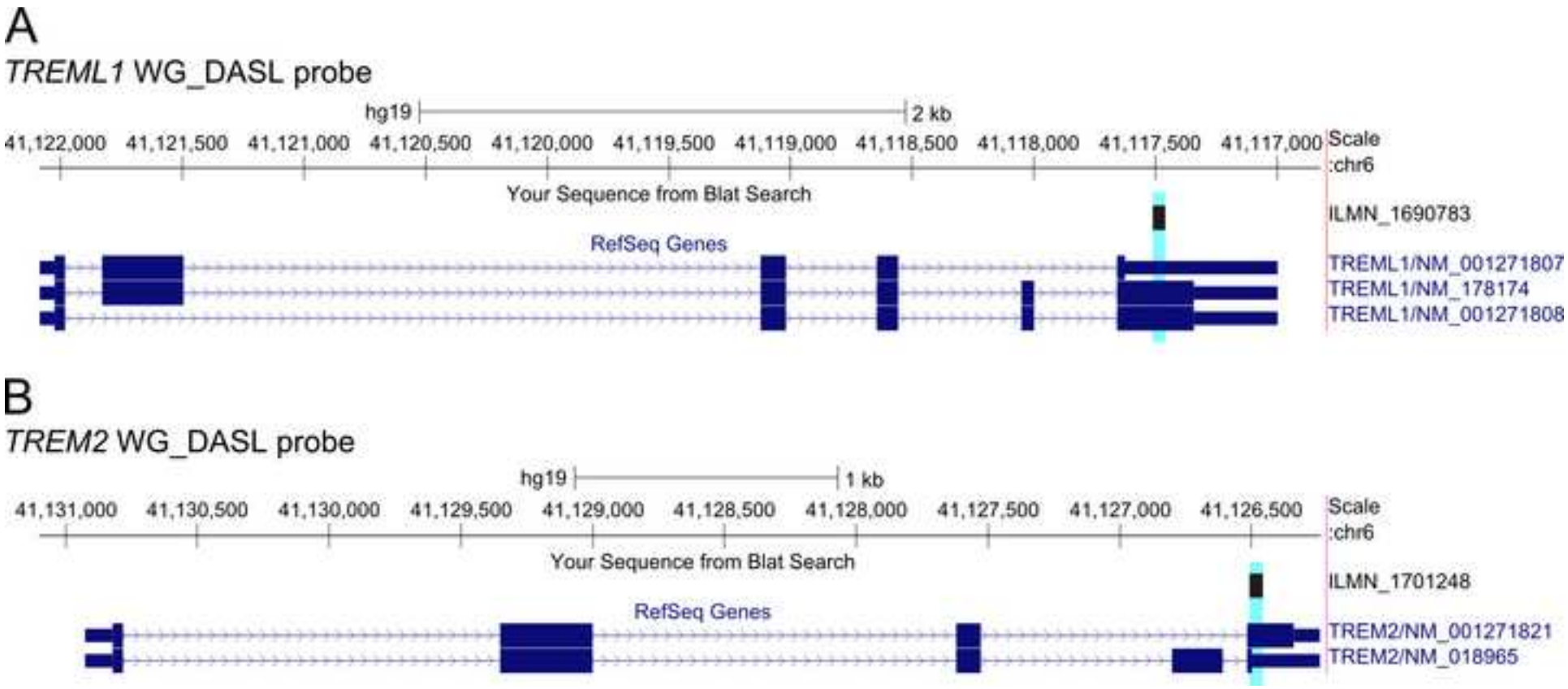
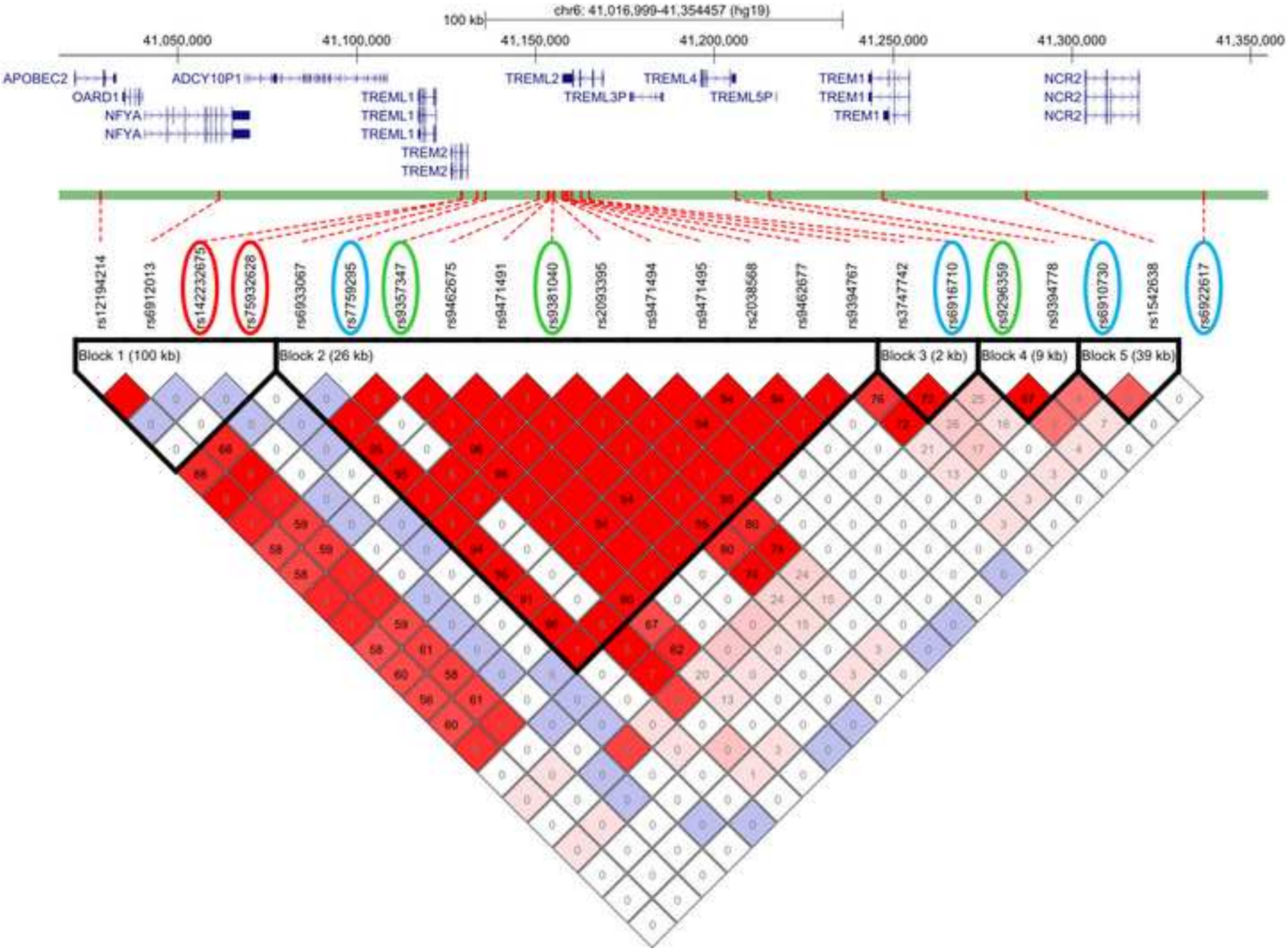


Figure3

Click here to download high resolution image



A regulatory variant at the *TREM* gene cluster associates with decreased Alzheimer's disease risk, and increased *TREML1* and *TREM2* brain gene expression

Carrasquillo et al., MS#: ADJ-D-16-00130

Responses to Reviewers

We provide in this document point-by-point responses to each of the comments by our Reviewers. The order follows that in the original "Comments" letter we received on 07-09-2016. The original text from the Editor or Reviewers is shown "*within quotation marks with italicized Calibri font*". Changes made to the manuscript to address reviewer's comments are highlighted in yellow, both on the manuscript and on the responses below. The location of changes made to the manuscript is indicated relative to the "**changes-accepted**" version of the manuscript files, and these locations are highlighted in blue to facilitate the review. We also added line numbers to the Main and Supplementary manuscript files. We are extremely grateful for the thoughtful and thorough reviews of our manuscript. We were delighted to receive many positive comments and believe that the changes that we made in response to each of the comments have further strengthened our manuscript.

Reviewers' comments:

"*This paper reports the results of a comprehensive and innovative study of common Alzheimer's disease (AD) risk variants at the TREM locus. Notably, the study tests the hypothesis that some of the common AD-risk variants at the TREM locus confer disease risk via regulation of transcript levels of some of the coding genes at the TREM gene cluster. This work is significant both for the TREM locus and potentially as a more general approach for other AD-risk loci. Strengths of the study include the strong replication/validation strategy and the focus on identification and characterization of functional regulatory genetic variants*."

We thank the reviewers for their many favorable comments, including those which indicated that the paper "*is well written*", "*has an interesting finding*", is "*based on well characterized IGAP and Mayo Clinic cohorts with adequate study sizes*", and that "*The expression association analysis was well designed and thoroughly executed*" with "*appropriate replication cohorts*". Importantly, one of the reviewers stated that "*The study contributes to the understanding of the broader role of TREM locus in AD*".

"*The reviewers have raised several major and numerous minor points that must be addressed for the paper to be considered for publication in the Journal. Many of the points are focused on clarification of the statistical analysis and interpretation of the results. These concerns must be addressed to improve*

the statistical rigor of the analysis and reflect the findings and conclusions reported in the paper."

We have addressed the concerns point by point below.

Reviewer #1:

Comment 1:

"*The paper has an interesting finding that a functional regulatory variant contributes to a protective effect observed in the TREM1 IGAP GWAS meta-analysis (even though not significant at 10-8 nominal level) and suggests concomitant increase in TREM1 and TREM2 brain levels. This study was based on well characterized IGAP and Mayo Clinic cohorts with adequate study sizes."*

We thank this Reviewer for their positive comments.

"However, the paper needs several clarifications.

Major Comments:

1. *Can the authors give a sense of variation explained by their variant in terms of AD pathology?*"

We sincerely appreciate this reviewer's thorough evaluation of our manuscript, and the rigor implemented in their assessment of our results.

To address this reviewer's first comment, we evaluated the effect of this variant on Braak stage using both ANOVA and linear regression models described on the Supplementary Material, page 8, 2nd paragraph, and in **Table S4**, as shown below:

"Given the association of rs9357347 with AD-risk, we tested the hypothesis that this variant could also show an association with Braak stage, as the latter is an important criterion for the neurophathological diagnosis of AD [14]. Implementing an ANOVA model in R that included age-at-death, sex and *APOE* ε4 dose (0, 1 or 2 alleles), in the two larger datasets (Mayo WG-DASL and ROS/MAP RNAseq), we determined that rs9357347 does not significantly contribute to the variance in Braak stage in either of these two cohorts (p=0.91 and p=0.27, respectively). In addition, we implemented linear regression analysis in R, again using the two larger datasets to estimate the effect of each copy of the rs9357347 minor allele on Braak stage, including age-at-death, sex and *APOE* ε4 dose in the model. As shown in **Table S4**, we did not detect a significant association of rs9357347 with Braak stage in either cohort."

Dataset	N	beta	SE	p-value
Mayo WG-DASL: Temporal Cortex	399	-0.139	0.160	0.387
ROS/MAP RNAseq: DFPC	492	0.053	0.081	0.515

Table S4. Association of rs9357347 with Braak stage. The two largest

cohorts were evaluated: Mayo WG-DASL and ROS/MAP RNAseq. The variant was tested for association with Braak stage using linear regression under an additive model and including age-at-death, sex and *APOE* ε4 dose as covariates. In this model, the beta coefficient is interpreted as the change in Braak score associated with each copy of the minor allele. N = sample size. SE= standard error. DFPC = dorsolateral prefrontal cortex.

Comment 2:

"*2. Authors use the word "nominally significant." This term is generally used for associations on the borderline. A p-value of 10-3 does not meet this threshold in a GWA study (Table 2). Can the authors use a different term, such an association of potential interest given your scientific hypothesis rather than this being somehow data driven.*"

We used the term "nominally significant" when referring to an uncorrected p-value < 0.05. We have now modified the language wherever the term "nominally significant" was used, as shown below:

Page 4, 3rd paragraph, 2nd sentence: "…the intergenic variant, rs9381040, that demonstrated the most significant association at the *TREM* locus in the IGAP AD-risk GWAS meta-analysis (uncorrected $p=6x10^{-04}$) [12]."

Page 9, 2nd paragraph, 5th sentence: "…5 SNPs showed association (uncorrected $p<0.05$) with increased levels of both *TREML1* and *TREM2*…"

Page 12, 3rd paragraph, 2nd sentence: "Only rs7759295 showed association with *TREML1* gene expression (uncorrected p=0.04),..."

Page 25, Table 2 legend, 4th sentence: "Variants showing association with gene expression (uncorrected $p<0.05$) are underlined and in italic font."

Comment 3:

"*3. What do the authors mean by non-AD subjects (abstract)? Do they mean neurological conditions other than AD, or subjects who did not have any AD pathology, as in not cognitive impaired? If the study had no controls, or not cognitive impaired subjects, does conditioning on the disease status cause any bias due to conditioning on a collider? For example, Cole et al., Int J Epidemiol. 2010. 39:417-420. This may not apply but is of interest given the conditioning statement.*"

In the post-mortem cohorts that were utilized in this study, non-ADs are those subjects whose neuropathological diagnoses did not meet criteria for definite AD, but they could have other unrelated neuropathologies. In our original submission, the diagnostic criteria for the Mayo Clinic WG-DASL eQTL dataset are indicated on the 1st paragraph on page 3 of the Supplementary Material and the reference is supplied in the Main Manuscript, page 7, 1st paragraph; that for the Mayo Clinic RNAseq dataset is mentioned on 2nd paragraph on page 4 of the Supplementary Material, and on the Main Manuscript, on page 7, 2nd paragraph; and references are supplied for the ROS/MAP cohort in the Main Manuscript, page 7, paragraph 2.

To clarify the diagnostic criteria in the main text, we now include in the Main Manuscript, page 7, 1st paragraph, the following text:

All AD subjects had neuropathological diagnosis of definite AD [2]. The non-AD subjects did not fulfill neuropathological criteria for definite AD, but many had other unrelated pathologies.

We also added the same references for the ROS/MAP dataset to the Supplementary Text, page 5, 3rd paragraph, for consistency.

As the primary goal of this study was to estimate the effect of genetic variants on gene expression, rather than their effect on disease status, we combined ADs and non-ADs in the linear regression analysis and included diagnosis as a covariate. The diagnosis covariate was coded as the presence or absence of AD. Our original submission described this analytic methodology in the Supplementary Material, "Mayo Clinic WG-DASL eQTL dataset" sub-section, formerly last paragraph, and in the now deleted "eQTL analysis of rs9357347 and *TREML1/TREM2* RNAseq gene expression" subsection, last sentence. The analytic methodology is instead included in the Main Manuscript in this revision for clarity. We now include a "2.4 Statistical Analysis" sub-section, in the Main Text, pages 7-8, shown below. In doing so, we also address Comment #2 of Reviewer #5.

2.4. Statistical Analysis

Normalized transcript expression levels, on a log2 scale, were tested for associations with *TREM* locus genotypes in each of the three datasets (Mayo WG-DASL, Mayo Clinic RNAseq and ROS/MAP RNAseq) via multivariable linear regression analyses implemented in PLINK [20]. An additive model was applied adjusting for age-at-death, sex, diagnosis, RNA Integrity Number (RIN) and adjusted RIN squared $(\text{RIN-RINmean})^2$ in all expression analyses, and *APOE* ε4 dosage and PCR plate in Mayo WG-DASL only, and flowcell in the Mayo Clinic RNAseq dataset only. The eQTL analysis in the discovery, WG-DASL dataset, included *APOE* ε4 dose as a covariate given the strong effect of this allele on AD. However, since a significant association was not detected with this covariate in the rs9357347 eQTL analyses in the discovery set, *APOE* ε4 dose was not included in the eQTL analyses implemented on the replication cohorts. For comparison, we have performed the eQTL analyses in all three datasets with and without adjustment for *APOE* ε4 dose and do not observe a substantial difference in the association results between these two models.

Meta-analyses were performed on eQTL results from the three independent datasets. For these analyses, METAL [21] was implemented using weighted average of z-scores from the individual study p-values, weighted according their sample size.

To address Comment #3 of our Reviewer #1, we also investigated the possibility of "collider conditioning bias" in our analyses. In the article by Cole et al. cited by this reviewer, the term "collider conditioning bias" refers to the bias that occurs when conditioning, adjusting or stratifying on a common effect of the "exposure" and "outcome" being measured. In our study,

the potential for collider conditioning bias arises if both the genotype and the expression levels are associated with disease status. "Collider conditioning bias" could be a potential issue in our study design, because we adjusted for disease status (disease status potentially being the "collider", i.e. a common effect of genotype and expression level). Therefore, **in addition** to analyzing the AD and non-ADs together (AD+nonAD) **and** adjusting for AD status, we have also now analyzed the combined set of AD+nonAD, **without** adjustment for disease status, in order to determine if the effect of genotype on expression disappears or remains in the latter analysis.

We include this new analysis, which revealed lack of evidence of "collider conditioning bias", as shown below and on page 7 of the Supplementary Material, and in **Table S2** on page 15:

Assessment of potential collider conditioning bias

"Since the primary goal of this study was to estimate the effect of genetic variants on gene expression, rather than their effect on disease status, we combined ADs and non-ADs in the linear regression analysis and included diagnosis as a covariate. The diagnosis covariate was coded as the presence or absence of AD. However, adjusting for diagnosis status could potentially introduce a collider conditioning bias if both the genotype and the expression levels are associated with disease status [13]. Therefore, we have also analyzed the combined set of AD+nonAD, without adjustment for disease status, in order to determine if the effect of genotype on expression disappears or remains in the latter analysis. **Table S2**, shows results for the two types of analyses that were performed in each of the three datasets (Mayo WG-DASL, Mayo RNAseq and ROS/MAP RNAseq): (1) AD and nonAD combined while adjusting for diagnosis, (2) AD and nonAD combined not adjusting for diagnosis. Overall, the results from analyses 1 and 2 are very similar, suggesting that there is no real impact of a collider effect."

			TREML1			***TREM2***		
Dataset	**Group**	**N**	**beta**	**SE**	**p-value**	**beta**	**SE**	**p-value**
Mayo WG-DASL: Temporal Cortex	All (W/Dx)[a]	380	0.088	0.032	6.28E-03	0.090	0.045	4.61E-02
	All (Wo/Dx)[b]	380	0.083	0.033	1.32E-02	0.088	0.045	5.33E-02
Mayo Clinic RNAseq: Temporal Cortex	All (W/Dx)[a]	132	-0.030	0.108	7.82E-01	0.084	0.128	5.13E-01
	All (Wo/Dx)[b]	132	-0.023	0.111	8.40E-01	0.102	0.145	4.86E-01
ROS/MAP RNAseq: DFPC	All (W/Dx)[a]	494	0.089	0.114	4.36E-01	0.124	0.060	3.77E-02
	All (Wo/Dx)[b]	494	0.089	0.114	4.35E-01	0.125	0.060	3.81E-02

Table S2. Analyses to assess the potential of introducing collider conditioning bias in the linear regression model due to adjustment for diagnosis. For each of the three datasets, linear regression analysis was run in a: AD and non-AD combined, with diagnosis included as a covariate; b: Analysis of AD and non-AD combined, without adjustment for diagnosis. N = sample size. SE= standard error. DFPC = dorsolateral prefrontal cortex. Given that all expression measures were on a log2 scale, fold-change for the beta coefficients = 2^{beta}.

Comment 4:

"4. Page 4, Paragraph 2, states that the meta-analysis was done in AD cases and controls. I am not sure if the authors are somehow use controls and non-AD in an exchangeable fashion."

The sentence referred to by our Reviewer states "Since the first two reports [1, 2], the risk effect of the most significant *TREM2* rare missense variant p.R47H (a.k.a. rs75932628) has been replicated in multiple Caucasian series [5-9], including a large meta-analysis of 24,086 AD cases and 148,993 controls [10]", where the meta-analysis refers to published case-control analysis.

In contrast, throughout the manuscript, we consistently used the term non-AD to refer to subjects included in our eQTL analyses whose neuropathological diagnosis was not consistent with AD. Hence, we did not use non-AD and control in an exchangeable fashion.

Comment 5:

"5. Since the cohorts for this study have other cognitive phenotypes, do these findings replicate with cognitive phenotypes?"

Cognitive phenotypes were available only for the ROS/MAP cohort. Measures from this cohort of global cognitive decline and global cognition at the last evaluation before death were tested for association with rs9357347. As shown below, and on the Supplementary Material, pages 8, last paragraph, and in **Table S5**, we did not detect a significant association of this variant with either of these two phenotypes:

"We also evaluated the association of rs9357347 with measures of global cognitive decline and global cognition at the last evaluation before death in the ROS/MAP cohort. In this dataset, global cognition is a variable for overall cognitive function measured by the raw scores from 19 different tests that are converted to z scores and averaged. Global cognitive decline is a longitudinal cognitive phenotype based on repeated measures of global cognition, as previously described [15, 16]. The analysis was performed using linear regression analysis implemented in R, under an additive model for rs9357347, and adjusting for age-at-death, sex and *APOE* ε4 dose. Neither global cognitive decline nor global cognition at last evaluation shows an association with rs9357347 in this cohort (**Table S5**)."

Phenotype	N	beta	SE	p-value
Global cognitive decline	470	-0.007	0.007	0.320
Global cognition at last visit	493	-0.058	0.071	0.418

Table S5. Association of rs9357347 with cognition. Measures of cognition that were available in the ROS/MAP cohort were tested for association with rs9357347 using linear regression under an additive model, including age-at-death, sex and *APOE* ε4 dose as covariates.

N = sample size. SE = standard error. Z scores of the cognitive scores were analyzed, thus the beta coefficients can be interpreted as changes in z-score associated with each copy of the minor allele.

Comment 6:

"6. Page 5, Paragraph 2, Line 27-35. I am not clear what the authors are stating here. Are they stating that their study did not replicate for the two variants, because of a variant that has not been defined functionally. Please clarify."

To contextualize this comment, we provide here the original version of this paragraph and show in bold font the line questioned by our Reviewer:

"Some variants at the *TREM* locus have been reported to associate with AD endophenotypes. Cerebrospinal fluid (CSF) levels of AD biomarkers, tau and ptau, associate with three variants at the *TREM* locus that are not in LD with each other: *TREM2* p.R47H (rs75932628), rs6916710 located in intron 2 of *TREML2*, and rs6922617 located downstream from *NCR2* and outside the *TREM* cluster. Of these variants, only *TREM2* p.R47H was associated with AD-risk [13]. More recently, a variant upstream of *TREM2* (rs7759295) and a variant in intron 3 of *TREM1* (rs6910730) were reported to independently associate with increased AD pathology burden and increased rate of cognitive decline [14]. **However, neither of these two variants shows association with AD-risk in the IGAP meta-analysis (p>0.05) [15]. Thus, it is possible that the effect observed with these variants on AD endophenotypes is due to their LD with an as yet defined functional variant(s) that influences AD-risk at the *TREM* locus.**"

In this above former version of this paragraph, we summarized previously reported associations of variants at the *TREM* locus with AD endophenotypes. On the last 2 sentences of this paragraph we intended to explain that since these variants do not any show association with AD-risk in the IGAP meta-analysis (p>0.05), it is possible that these variants themselves are not affecting AD endophenotypes, but are instead reflecting the association of a nearby functional variant, that is in LD with them and which, if tested, would show association with both AD-risk and endophenotypes. In order to clarify this point, we have modified this paragraph on page 5 as follows:

"Some variants at the *TREM* locus have been reported to show association with AD endophenotypes [11, 13, 14]. Cerebrospinal fluid (CSF) levels of AD biomarkers, tau and ptau, associate with three variants at the *TREM* locus that are not in LD with each other: *TREM2* p.R47H (rs75932628), rs6916710 located in intron 2 of *TREML2*, and rs6922617 located downstream from *NCR2* and outside the *TREM* cluster. Of these variants, only *TREM2* p.R47H was associated with AD-risk [13]. More recently, a variant upstream of *TREM2* (rs7759295) and a variant in intron 3 of *TREM1* (rs6910730) were reported to be independently associated with increased AD pathology burden and increased rate of cognitive decline [14]. However, neither of these two variants shows association with AD-risk in the IGAP meta-analysis (p>0.05) [15]. Thus, other than *TREM2* p.R47H, none of the *TREM* locus variants previously reported to associate with AD endophenotypes show association with AD-risk. Functional ADrisk variants that influence AD endophenotypes are expected to show association both with these

endophenotypes and risk of AD. Therefore, it is possible that these latter variants are not the functional variants *per se*, but merely markers of other un-tested functional variants."

Comment 7:

"7. Variant Selection (Section 2.1). Please clarify what you mean by "strength". A p-value does not provide strength of the association, but merely an error rate beyond the nominal value."

We agree with this comment, and note that we used the word "strength" twice in the manuscript in relation to a p-value. In both instances it was used to denote the "strength of the evidence". Therefore, we have modified the sentences that originally used the word "strength" as follows:

Page 6, 2nd paragraph, 2nd sentence: "Variants were further selected based on the statistical significance of their AD-risk association in the IGAP stage 1 meta-analysis [12] (**Supplementary Methods**), where only those variants with p-values ≤ 0.0015 were kept."

Page 13, 1st paragraph, last sentence: "The evidence of association with *TREM2* expression was greater upon meta-analysis compared to the association observed in our discovery dataset; whereas the evidence of association with *TREML1* expression was slightly greater in our discovery dataset compared to the meta-analysis."

Comment 8:

"8. The Mayo clinic WG-DASL samples were from participants who were younger than Mayo clinic RNAseq and ROS/MAP RNAseq. Does age have any association with the expression levels? (Table 1)"

We now address this on the Supplementary Material, page 9, 2nd paragraph, as shown below:

"Association of age with *TREML1* and *TREM2* expression

As the WG-DASL cohort was overall younger than the two RNAseq cohorts, we assessed the association of the age covariate on *TREML1* and *TREM2* gene expression levels in the linear regression model described in the Material & Methods section 2.4. Age was not significantly associated with either *TREML1* or *TREM2* expression in the Mayo WG-DASL cohort ($p>0.05$). On the other hand, both in the Mayo RNAseq and ROS/MAP RNAseq cohorts, *TREML1* and *TREM2* expression levels appeared to be slightly increased with age, albeit the magnitude of the effect sizes were modest, with beta coefficients equivalent to approximately a 1.01 and 1.03-fold change in expression levels (Mayo RNAseq: *TREML1* $p=0.085$, beta=0.01; *TREM2* $p=0.026$, beta=0.02. ROS/MAP RNAseq: *TREML1* $p=2.0\times10^{-3}$, beta=0.04; *TREM2* $p=4.4\times10^{-5}$, beta=0.03). Since *TREML1* and *TREM2* gene expression levels appear be increased with age, it is possible that this might have led to a decrease in power to detect an association of rs9357347-C with increased levels of these genes in the two older cohorts."

Comment 9:

"9. Can you provide an interpretation for your "beta" coefficient on Page 7, last line of Paragraph 1 of the Results?"

We apologize for the confusion caused by the "beta" symbol that was mistakenly used in that sentence. The symbol should have been "r_s", representing a Spearman's rank correlation coefficient, as we now indicate it on page 9, 1st paragraph, last sentence:

"There was highly significant correlation between WG-DASL and RNAseq measurements for both *TREML1* (r_s=0.65, $p<10^{-40}$) and *TREM2* (r_s=0.80, $p<10^{-40}$) (**Fig. S1**)."

Comment 10:

"10. Page 7 last line. Is the nominal p-value "0.0015"? How did you arrive that this significance level?"

We now explain the rationale for this cut-off in the Materials and Methods, 2.1 Variant selection section, page 6, 3rd sentence:

"This p-value cut-off was arbitrarily chosen to select those variants that existed in both the IGAP stage 1 AD GWAS and in our discovery eQTL cohort, Mayo Clinic Whole Genome-DASL dataset, and that could be genotyped, if needed, in the replication eQTL cohorts, using cost-effective, medium-throughput assays."

Comment 11:

"11. Page 8, Line 26. The variant rs9381040 is not associated with TREM1 CER, but only with TREM1 TCX, and at the borderline for TREM2 TCX. The word "strongest" appears 32 different times throughout the manuscript."

We agree with this reviewer's assessment of the modest evidence of association for rs9381040 with *TREML1* and *TREM2* brain expression levels, despite this variant having the most significant p-value of AD risk association at the *TREM* locus in the IGAP meta-analysis. This is in contrast to rs9357347, which is in LD with rs9381040, and which has both evidence of association with AD risk and brain expression. Indeed, we emphasize this point in the Discussion section, page 15, 3rd paragraph as follows:

"The regulatory rs9357347 SNP is in the same haplotype block as the variant with the most significant AD-risk association at the *TREM* locus in the IGAP meta-analysis, rs9381040, which is an intergenic variant downstream of *TREML2*. Though this IGAP *TREM* locus "hit" SNP has greater evidence of AD-risk association than rs9357347, there is no evidence of regulatory potential for rs9381040 in brain regions relevant to AD."

We note that in many of the instances where the terms "strong" and "strongest" were used in our manuscript, they were in relation to the p-value of association for rs9381040 with AD-risk in the IGAP stage 1 meta-analysis and not its association with brain gene expression levels. Nevertheless, we do acknowledge that the words strong and strongest could be substituted with other adjectives. Therefore, we have edited this word on 25 occasions, as in the following examples:

Page 4, 3rd paragraph, 2nd sentence: "*TREML2* p.S144G is in tight LD with the intergenic variant, rs9381040, that demonstrated the most significant association at the *TREM* locus in the IGAP AD-risk GWAS meta-analysis ($p=6x10^{-04}$) [12]."

Page 9, 2nd paragraph, penultimate sentence: "Of these variants, rs9381040 has the most significant AD-risk association in the IGAP stage 1 meta-analysis (**Table 2**)."

Comment 12:

"12. For the sake of clarity, please provide an interpretation for "beta" throughout the manuscript."

We now provide an interpretation for the beta coefficients throughout the manuscript, as in the following examples:

Page 10, 2nd sentence was added as follows: "This IGAP "hit" is located 5.5kb downstream from *TREML2* and 23.7kb upstream from *TREM2* and is associated with *TREML1* and *TREM2* expression (p=0.0083, beta=0.086 and p=0.048, beta=0.091, respectively). Given that the expression measures were on a log2 scale, these changes in expression are equivalent to *TREML1* and *TREM2* fold changes of 1.06 and 1.07, for each copy of the minor allele, respectively. Notably, the minor allele of the IGAP "hit" rs9381040 is associated with both decreased AD-risk and increased *TREML1* and *TREM2* levels."

Page 10, 2nd paragraph, 4th sentence was added as follows: "…the minor allele of rs9357347 associates with reduced AD-risk (OR=0.95, 95% CI=0.91-0.98, p=0.001) and increased *TREML1* and *TREM2* expression in the temporal cortex (p=0.0063, beta=0.088 and p=0.046, beta=0.090, respectively) (**Table 2 and Fig. S2**). These beta coefficients can be interpreted as an estimated 1.06-fold change of both *TREML1* and *TREM2*, per rs9357347 minor allele, in this temporal cortex dataset."

Table 2: The beta coefficients shown for the IGAP AD-risk meta-analysis reflected effect size of the allele for AD risk association. These have now been replaced with odds ratios and 95% confidence intervals for easier interpretation of these results. The beta coefficients for the brain gene expression associations are retained for consistency with the text but described in the Table 2 and 3 legends as below.

Table 2 legend: "Given that the eGWAS expression measures were on a log2 scale, fold-change for the Mayo eGWAS beta coefficients = 2^{beta}."

Table 3 legend: "Since in all three datasets the expression measures analyzed were on a log2 scale, fold-change for the beta coefficients = 2^{beta}."

Supplementary Material, **Table S5**, legend: "Z scores of the cognitive scores were analyzed, thus these beta coefficients can be interpreted as changes in z-score associated with each copy of the minor allele."

We have also added the following text to the Discussion section on page 15, 4th paragraph:

"While the fold change estimates in gene expression associated with rs9357347-C are modest at 6-7%, the biological impact of the increase attributed to each copy of the minor allele, can be significant and may provide sufficient protection from disease in some individuals, particularly when considered over a lifetime. Furthermore, these estimates are based on RNA isolated from tissue samples and not microglial cells where both *TREM2* and *TREML1* are predominantly expressed [35], and where expression levels of these genes may be impacted to a greater extent by regulatory variants. Additional studies will be needed to determine the impact of such expression changes on the biology of microglial cell function."

Comment 13:

"13. The OR presented in first line of Page 9 does not correspond to a risk statement. Is the OR conditional or marginal?"

In this line the OR pertains to the test of association of rs9357347 with the diagnosis of AD in the IGAP stage 1 study, which was a meta-analysis of the four largest AD case-control GWAS. This study is mentioned in the Introduction in the former and current versions of our manuscript (page 4, 3rd paragraph, 2nd sentence). To clarify the sentence queried by our Reviewer, we modified it as follows (page 10, 2nd paragraph, 3rd sentence):

"As expected, the minor allele of rs9357347 is associated with reduced AD-risk (OR=0.95, 95% CI=0.91-0.98, p=0.001) in the IGAP meta-analysis [12] and with increased *TREML1* and *TREM2* expression in the temporal cortex in the Mayo Clinic WG-DASL eQTL analysis (p=0.0063, beta=0.088 and p=0.046, beta=0.090, respectively) (**Table 2 and Fig. S2**)."

Comment 14:

"14. For the sake of clarity, can the authors provide a better explanation for the high regulome scores of some variants in their study?"

We now include the following text on page 6, in the Materials and Methods, section 2.1:

"Variants were further prioritized by their Regulome score. Regulome scores were obtained from the Regulome database, which annotates variants with regulatory information from 962 different datasets and a variety of sources, including ENCODE [16]. Regulome scores are on a scale from 1 to 6, and these numerical categories are sub-classified with letters based on the number of lines of evidence of functional consequence. A value of 1a is assigned to the variant with the most evidence of regulatory potential, while a score of 6 has the least [16]."

Comment 15:

"15. Page 10 Paragraph 3. Did the authors perform a conditional analysis to examine the association of rs9357347 after conditioning on rs6916710? Was the signal partially explained by the correlated variant? If so, can it provide any additional mechanistic understanding?"

As stated on that paragraph, rs6916710 did not show evidence of association with either AD-risk or gene expression; therefore, conditional analyses were not performed with this variant. To address our Reviewer's comment, we have now performed this analysis. The table below shows

results for the test of association of rs9357347 with temporal cortex *TREML1* and *TREM2* levels, both with and without conditioning on rs6916710. In addition, we show the results obtained when we tested the association of rs6916710 with these genes' levels, both with and without conditioning on rs9357347.

As shown here, the p-values for the association of rs9357347 with both *TREML1* and *TREM2* levels become larger upon conditioning on rs6916710, although the association remains significant at p=0.02. Rs6916710 does not have significant association with either *TREML1* and *TREM2* levels, as already stated above and in the manuscript. The association of rs6916710 with levels of these genes remains non-significant, with p-values that become larger upon conditioning on rs9357347.

While conditioning the test of association of rs9357347 on rs6916710 leads to larger p-values for the former, given the lack of any gene expression or AD-risk association with rs6916710, and lack of evidence of its regulatory potential, rs6916710 is unlikely to be accounting for part of the association of rs9357347 with expression due to a mechanistic, biological effect. Rather, it is possible that there are additional functional, regulatory variants in the same LD block as rs9357347, which could in part be accounting for the associations with expression observed with rs9357347. Rs6916710 may be a **marker** for such additional functional, regulatory variants.

We added a sentence to the Discussion section, page 17, 3rd paragraph, 2nd sentence, to highlight the possibility of additional functional, regulatory variants in the same LD block as rs9357347:

"We identified rs9357347 as a putative regulatory variant that is associated with protection from AD and with increased *TREML1* and *TREM2* brain levels, and nominate rs9357347 as one of the functional variants that accounts for the IGAP AD-risk signal. Additional studies are needed to validate the function of this variant, and to explore the possibility of the presence of other variants at this locus that could contribute to associations observed with rs9357347."

Given the complexities discussed above and the lack of further mechanistic insight gained from the analyses we show below, we opted not to include the conditional analyses with rs6916710 in the manuscript, however, we can do so, if our Reviewer feels that we should.

	Rs9357347 temporal cortex gene expression association **without** conditioning on rs6916710	
Gene	**beta**	**p-value**
TREML1	0.088	6.28E-03
TREM2	0.090	4.61E-02
	Rs9357347 temporal cortex gene expression association conditioning on rs6916710	
Gene	**beta**	**p-value**
TREML1	0.120	2.09E-02
TREM2	0.058	4.36E-01

	Rs6916710 temporal cortex gene expression association **without** conditioning on rs9357347	
Gene	**beta**	**p-value**
TREML1	0.054	7.70E-02
TREM2	0.072	9.20E-02
	Rs6916710 temporal cortex gene expression association conditioning on rs9357347	
Gene	**beta**	**p-value**
TREML1	-0.040	4.53E-01
TREM2	0.040	6.15E-01

Comment 16:

"16. Can the authors provide an explanation as to why their findings did not replicate in either of the replication cohorts? Does age play a role in any of these expressions?"

In the course of the revision of our manuscript, we noticed that, unlike for the other two cohorts, the gene expression values from the ROS/MAP RNAseq dataset were not log2 transformed as they should have been. We repeated the analyses for ROS/MAP using the log2 transformed-FPKM gene expression values. We also repeated the meta-analyses. As shown in the updated **Table 3** and below, rs9357347 is associated with *TREM2* levels in both the Mayo Clinic WG-DASL and ROS/MAP cohorts, although a significant association with *TREML1* is observed only in the former. We discuss this point further in our response to Reviewer #2, Comment #2. Please also see our response to Reviewer #5, Comment #4.

To look into the influence of age on expression levels we performed additional analyses and included these results as outlined in our response to Comment #8 by our first Reviewer.

Table 3. Meta-analysis of rs9357347 eQTL results from three independent datasets.

Dataset	Sample size	MAF	*TREML1*			*TREM2*		
			beta	**SE**	**p**	**beta**	**SE**	**p**
Mayo WG-DASL	380	0.307	0.088	0.032	6.28E-03	0.090	0.045	4.61E-02
Mayo Clinic RNAseq	132	0.311	-0.030	0.108	7.82E-01	0.084	0.128	5.13E-01
ROS/MAP RNAseq	494	0.281	0.089	0.114	4.36E-01	0.124	0.060	3.77E-02
Meta-analysis	1006		**+-+**		**3.36E-02**	**+++**		**3.54E-03**

Meta-analysis of rs9357347 eQTL results from temporal cortex (Mayo WG-DASL and Mayo Clinic RNAseq) and dorsolateral prefrontal cortex samples (ROS/MAP). MAF = minor allele frequency. SE = standard error. Since in all three datasets the expression measures analyzed were on a log2 scale, fold-change for the beta coefficients = 2^{beta}. The meta-analysis was performed using METAL, with weighted average of z-scores from the individual study p-values, weighted according their sample size.

Minor Comment 1:

"1. In certain places "that" might be more appropriate than "which." "

The word "which" has been edited when appropriate, as follows:

Page 4, last paragraph, 2nd sentence: This sentence was simplified, and the phrase starting with the word "which" was deleted: "*TREML2* p.S144G is in tight LD with the intergenic variant, rs9381040, that demonstrated the most significant association at the *TREM* locus in the IGAP AD-risk GWAS meta-analysis ($p=6x10^{-04}$) [12].

Page 14, 1st paragraph, 1st sentence: This sentence was simplified, and the word "which" was deleted:"The minor allele of this variant is associated with both decreased AD-risk and with increased *TREML1* and *TREM2* brain expression in the temporal cortex."

Minor Comment 2:

"2. Some sentences are too long and harder to follow. A little simplification might help."

In addition to the sentences shown in the response above, we have also simplified the following sentences:

Page 7, 2nd paragraph, 1st sentence: "Temporal cortex RNAseq data from two RNAseq cohorts: "Mayo Clinic RNASeq" and "ROS/MAP RNAseq" were employed for replication of the associations that were detected with the WG-DASL gene expression measurements. The Mayo Clinic RNASeq dataset is comprised of 84 LOAD and 48 non-AD brains from the Mayo Clinic Brain Bank that were not part of the Mayo Clinic WG-DASL cohort but whose neuropathological diagnosis followed the same criteria. The ROS/MAP RNAseq dataset is comprised of RNAseq data from 288 AD and 206 non-AD samples that are part of the ROS/MAP cohort (**Table 1**) previously described [18, 19]."

Page 10, 2nd paragraph, 6th sentence:
"Unlike the IGAP "hit" (rs9381040), rs9357347 lies within sequence subject to histone modifications and within a DNAse hypersensitive site detected by the Roadmap Epigenomics Consortium [22] in brain regions relevant to AD pathology such as the hippocampus. Furthermore, this variant is predicted to affect transcription factor binding (SP1 and PPAR) as catalogued in HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) [23]."

Page 13, 2nd paragraph, 3rd sentence: "Whereas *TREML1* is expressed in both cerebellum and temporal cortex of all subjects, *TREM2* is expressed in 98% of temporal cortex and 41% of cerebellum samples."

Page 15, 2nd paragraph, 4th sentence: "Interestingly, rs9357347 is predicted to affect transcription factor binding (SP1 and PPAR) as catalogued in HaploReg. These two transcription factors are

known be important in regulating key players in the inflammatory response and lipid metabolism [33, 34].”

Page 16, 2nd paragraph, 4th sentence: “Further, *TREML2* expression is too low to be reliably measured in brain tissue (TCX and CER). This raises the possibility that the association with *TREML2* p.S144G is due to its LD with a functional variant(s) that influences the function or level of a nearby *TREM* gene(s), such as *TREML1* or *TREM2*.”

Page 17, 3rd paragraph, 1st sentence: “In summary, we characterized expression of *TREM* genes in cerebellum and temporal cortex and determined *TREML1* and *TREM2* to be the only reliably expressed *TREM* genes in these brain regions. We identified rs9357347 as a putative regulatory variant that associates with protection from AD and with increased *TREML1* and *TREM2* brain levels, and nominate rs9357347 as one of the functional variants that accounts for the IGAP AD-risk signal.”

Minor Comment 3:
“3. Page 5, Paragraph 2, Line 10, missing a reference for "… reported …"."

We have added the references to this sentence on page 5, 2nd paragraph, as shown below:

“Some variants at the *TREM* locus have been reported to associate with AD endophenotypes [11, 13, 14].”

Minor Comment 4:
“4. Page 5, Line 25, "associated" instead of "associate"."

This word has been replaced as suggested (now on page 5, line 80).

Minor Comment 5:
“5. Page 6, continuation of paragraph from Page 5. You are already making a conclusion in the introduction. I'd suggest that you remove the last line.”

We have deleted this sentence.

Minor Comment 6:
“6. Page 8 last line replace "associates" with "is associated"."

This word, now on page 10, 2nd paragraph, line 195, has been replaced, as suggested.

Reviewer #2:

"The manuscript entitled "A regulatory variant at the TREM gene cluster associates with decreased Alzheimer's disease risk, and increased TREML1 and TREM2 brain gene expression" by Carrasquillo et al. describes an association study between TREM2 SNPs and brain expression levels of genes positioned within the TREM cluster. The expression association analysis was well designed and thoroughly executed. The study includes appropriate replication cohorts. The study contributes to the understanding of the broader role of TREM locus in AD, beyond the rare coding mutations; the results are of interest to the community of investigators studying the genetic of LOAD."

We thank this Reviewer for their favorable comments.

"However, based on the results presented in this manuscript the authors cannot rule out the high possibility that other genetic variants, including structural variants that are in high LD with SNP rs9357347, are responsible for the observed statistical associations with TREM2 and TREM1L expression. There are several major concerns and revisions are needed accordingly:"

Comment 1:

"1) The conclusion should be phrased more carefully. The results reported in this work suggest associations, however, the actual variants underlying the observed associations remained to be determined. This will require experiments using appropriate biological systems, in which the candidate variant is the only different site.

The title should be revised as well accordingly for accuracy; at this stage SNP rs9357347 is a candidate regulatory variant."

We fully agree with this reviewer's comment. As suggested, the word "candidate" has been added to the title of the manuscript.

In addition, we added a paragraph (Discussion, page 15, last paragraph) to emphasize the need for experiments in appropriate systems as follows:

"While the fold change estimates in gene expression associated with rs9357347-C are modest at 6-7%, the biological impact of the increase attributed to each copy of the minor allele, can be significant and may provide sufficient protection from disease in some individuals, particularly when considered over a lifetime. Furthermore, these estimates are based on RNA isolated from tissue samples and not microglial cells where both *TREM2* and *TREML1* are predominantly expressed [35], and where expression levels of these genes may be impacted to a greater extent by regulatory variants. Additional studies will be needed to determine the impact of such expression changes on the biology of microglial cell function."

We have also provided additional language in the discussion section (page 17, 3rd paragraph, penultimate sentence):

"Additional studies are needed to validate the function of this variant, and to explore the possibility of the presence of other variants at this locus that could contribute to associations observed with rs9357347."

We note that our original manuscript had already raised these possibilities in the Discussion (please see the un-highlighted text in the paragraph shown under Comment #2 below). However, our additions in response to Comment #1 further enhance the cautionary language.

Comment 2:

"2) It is likely that SNP rs9357347 and the other 3 SNPs in LD block 2 tag the actual regulatory variant/s //haplotype. Deep sequencing analysis targeted for block 2 (specifically the region that overlaps with the Roadmap Epigenomics signals) is necessary to identify the actual risk/protective variant."

We are also in complete agreement with this assessment, and now propose this approach specifically in the discussion section (page 17, 1st paragraph, penultimate sentence), as shown below:

"Both rs9357347 and p.S144G should be tested for their functional potential and influence on outcomes of inflammation and neuroprotection. It remains possible that rs9357347 is in LD with an untested true functional variant with effects on transcription and AD-risk. It is likewise possible that while rs9357347 is associated with both AD-risk and gene expression levels, these joint effects are coincidental due to LD, rather than being related. These possibilities need to be explored through sequencing of the entire *TREM* locus, or via targeted sequencing of LD block 2 where rs9357347 resides. Thus, our findings provide a testable hypothesis for a strong candidate functional variant, specific transcription factors and their effects on *TREML1* and *TREM2* levels."

Comment 3:

"3) The authors should present the mRNA expression levels of TREM2 and TREM1L stratified by disease status using the study's cohorts. This will be also helpful to determine the direction of the change in expression in AD vs. control."

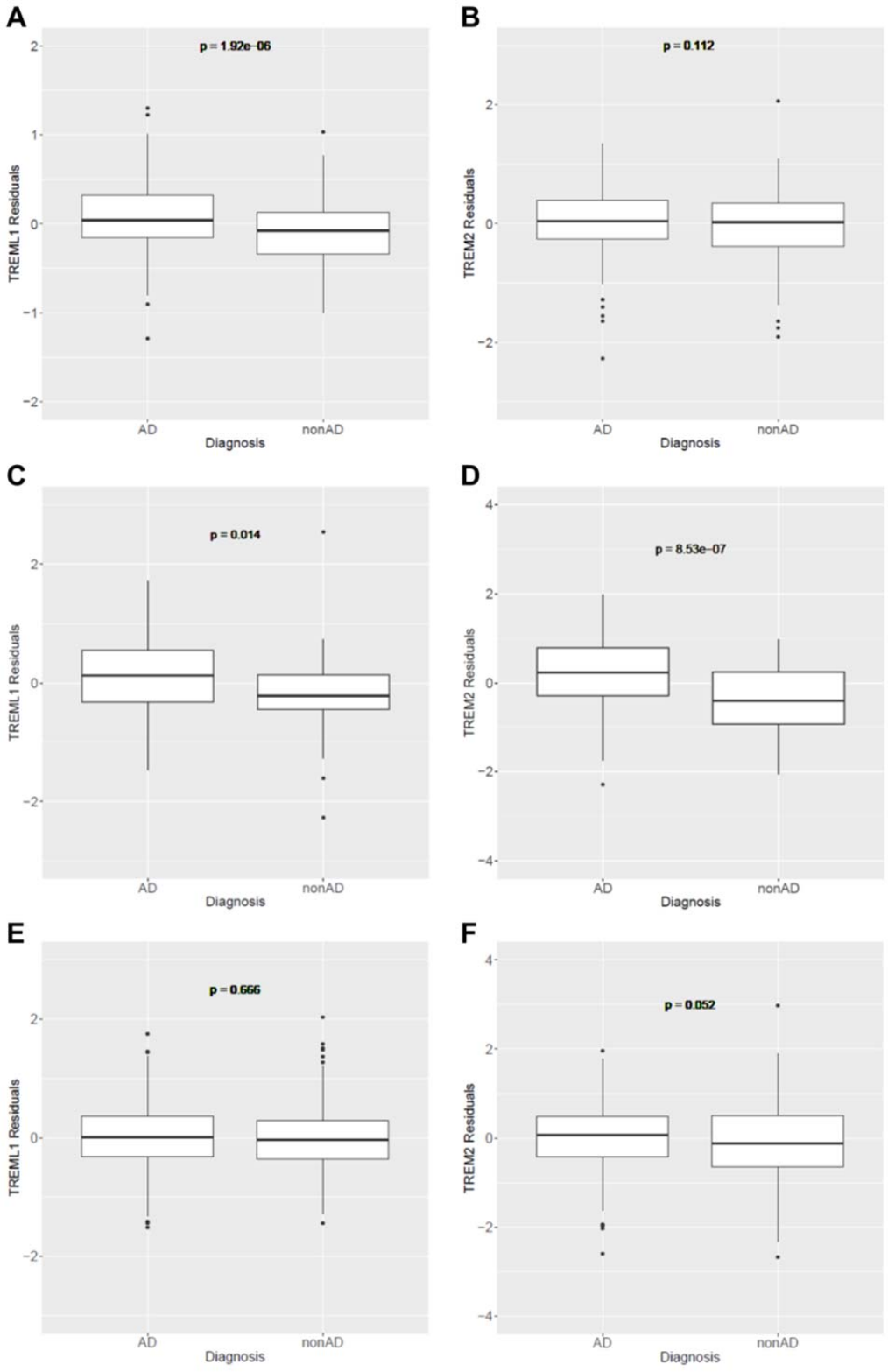
We are thankful for this suggestion. We now include box plots of *TREML1* and *TREM2* gene expression levels stratified by diagnosis for each of the 3 datasets in **Fig. S5**. The trends observed in these box plots are described in the Supplementary Material (page 9, 3rd paragraph) as shown below:

"**Association of diagnosis with *TREML1* and *TREM2* expression**

To assess if diagnosis is associated with *TREML1* and/or *TREM2* gene expression levels, linear regression analyses were performed in R in each of the three datasets, adjusting for all other covariates included in the eQTL analyses described in the Materials and Methods section 2.4, as well as rs9357347 minor allele dose. The box plots in **Fig. S5** show the direction of the change in expression between AD and nonAD subjects, and indicate the significance of the association for each test. We observe a consistent trend of higher *TREML1* and *TREM2* expression in AD versus

nonADs, although some of these associations do not reach statistical significance. The trend toward higher *TREML1* and *TREM2* expression in AD subjects could be a reflection of microglial activation and/or proliferation known to occur in AD brains."

Fig. S5. Box plots of gene expression residuals for *TREML1* and *TREM2* in AD and nonAD



subjects, for each of the three cohorts investigated.

A and B: Expression measure residuals for *TREML1* (A) and *TREM2* (B) in the Mayo WG-DASL dataset, adjusted for rs9357347 minor allele dose, age-at-death, sex, *APOE* ε4 dose, RIN, $(RIN-RINmean)^2$ and PCR plate. *TREML1* (C) and *TREM2* (D) in the Mayo Clinic RNAseq dataset, adjusted for rs9357347 minor allele dose, age-at-death, sex, *APOE* ε4 dose, RIN, $(RIN-RINmean)^2$ and flowcell. *TREML1* (E) and *TREM2* (F) in the ROS/MAP RNAseq dataset, adjusted for rs9357347 minor allele dose, age-at-death, sex, *APOE* ε4 dose, RIN, $(RIN-RINmean)^2$.

Comment 4:

"4) A description of the statistical analyses has to be included in the method section of the main manuscript (not the supplementary material). The results section relies completely on the statistical methods."

A description of the statistical analyses is now included in the Material and Methods, section 2.4, of the main manuscript on page 7, 3rd paragraph.

"2.4. Statistical Analysis

Normalized transcript expression levels, on a log2 scale, were tested for associations with *TREM* locus genotypes in each of the three datasets (Mayo WG-DASL, Mayo Clinic RNAseq and ROS/MAP RNAseq) via multivariable linear regression analyses implemented in PLINK [20]. An additive model was applied adjusting for age-at-death, sex, diagnosis, RNA Integrity Number (RIN) and adjusted RIN squared $(RIN-RINmean)^2$ in all expression analyses, and *APOE* ε4 dosage and PCR plate in Mayo WG-DASL only, and flowcell in the Mayo Clinic RNAseq dataset only. The eQTL analysis in the discovery, WG-DASL dataset, included *APOE* ε4 dose as a covariate given the strong effect of this allele on AD. However, since a significant association was not detected with this covariate in the rs9357347 eQTL analyses in the discovery set, *APOE* ε4 dose was not included in the eQTL analyses implemented on the replication cohorts. For comparison, we have performed the eQTL analyses in all three datasets with and without adjustment for *APOE* ε4 dose and do not observe a substantial difference in the association results between these two models.

Meta-analyses were performed on eQTL results from the three independent datasets. For these analyses, METAL [21] was implemented using weighted average of z-scores from the individual study p-values, weighted according their sample size."

Comment 5:

"5) Corrections for multiple tests should be applied for the nominal p values. (different tissues, several LD blocks, 2 genes)."

Correction for multiple tests was not applied to the eQTL results from the Mayo WG-DASL since this was our discovery cohort from which variants with eQTL p-values <0.05 and evidence of AD risk association were selected and further prioritized based on annotation of regulatory potential. Since only one variant, rs9357347, fulfilled these criteria, only this variant was evaluated in the replication eQTL cohorts. We have clarified the variant selection process by adding text to page 6, 2nd paragraph, as shown below. We also added, on page 7, the terms "Discovery eQTL cohort" and "Replication eQTL cohorts", following "Mayo Clinic Whole Genome-DASL dataset" and "RNAseq datasets" subtitles in the Methods section, respectively.

"We restricted our analysis to variants located within 100kb of any coding *TREM* family gene at the chromosome 6p21.1 *TREM* gene cluster (**Fig. 1**). Variants were further selected based on the statistical significance of their AD-risk association in the IGAP stage 1 meta-analysis [12] (**Supplementary Methods**), where only those variants with p-values ≤ 0.0015 were kept. This p-value cut-off was arbitrarily chosen to select those variants that existed in both the IGAP stage 1 AD GWAS and our discovery eQTL cohort, Mayo Clinic Whole Genome-DASL dataset, and that could be genotyped, if needed, in the replication eQTL cohorts, using cost-effective medium-throughput assays. Variants were further prioritized by their Regulome score. Regulome scores were obtained from the Regulome database, which annotates variants with regulatory information from 962 different datasets and a variety of sources, including ENCODE [16]. Regulome scores are on a scale from 1 to 6, and these numerical categories are sub-classified with letters based on the number of lines of evidence of functional consequence. A value of 1a is assigned to the variant with the most evidence of regulatory potential, while a score of 6 has the least [16]."

Comment 6:

"6) rs9357347 demonstrated only suggestive association in the replication cohort that has no overlap with the discovery cohort."

As also discussed in our response to Reviewer #1, Comment #16, in the course of the revision of our manuscript, we noticed that, unlike for the other two cohorts, the gene expression values from the ROS/MAP RNAseq dataset were not log2 transformed as they should have been. We repeated the analyses for ROS/MAP using the log2 transformed-FPKM gene expression values. We also repeated the meta-analyses. As shown in the updated **Table 3**, rs9357347 associates with *TREM2* levels in both the Mayo Clinic WG-DASL and ROS/MAP cohorts, although a significant association with *TREML1* is observed only in the former.

Although the association with *TREML1* does not reach significance in ROS/MAP, the beta coefficient overlaps with that in the WG-DASL cohort (**Table 3**). The lack of significant association in the Mayo RNAseq cohort is likely due to its relatively small size; yet in this cohort the effect size detected for *TREM2* is very similar to that in the WG-DASL cohort, as evidenced by the improved significance of the association with *TREM2* levels upon meta-analysis, as compared to this association in the discovery, WG-DASL cohort.

Please also see our response to Comment #8 by Reviewer #1, Comment#4 by Reviewer #5.

Reviewer #4:

"*This paper describes the identification of an intergenic variant (rs9357347) as a causative factor in expression regulation of transcripts TREM2 and TREML1. The functional variant was previously associated with reduced risk of Alzheimer disease within the IGAP AD-risk GWAS meta-analysis published by Lambert et al (2013). The authors argue that the identified variant underlies the genome-wide association signal within the TREM locus. Overall, the paper is well written. Methodology and results are well structured. The identification of several putative eQTL variants and selection of the putative causal variant are described in detail."*

We thank this Reviewer for their favorable comments.

"A drawback is the absence of significant expression regulation by this variant in the replication datasets, which should be addressed in more detail."

Please see the response to Reviewer #1 Comments #8 and #16; Reviewer #2 Comment #6 and Reviewer #5, Comment #4.

"*Second, absence of expression data for five out of seven genes within the TREM locus limits the interpretation of the identified variants as the true functional factor explaining the GWAS signal at this locus.*"

Please see the response to Comment #2 below.

Comment 1:

"*Main Comments:*

** The evidence for the regulatory variant 'nominated' in this paper is more limited than the text and title lead to believe. Association of rs9357347 with increased TREML1 and TREM2 expression levels reaches significance in the Mayo WG-DASL discovery dataset only. Given that associations fail to reach statistical significance in both the Mayo Clinic RNAseq and ROS/MAP RNAseq datasets, the meta-analysis association is driven by the Mayo WG-DASL only. Furthermore, the direction of effect in TREML1 expression is not equal among the datasets. Since TREML1 and TREM2 expression levels from the microarray could be correlated to expression levels in a 93 AD patient dataset (results; page 7; line 37-47, and figure S1) why would the association of rs9357347 with TREML1 and TREM2 expression levels remain specific to the microarray data? Please tone down and discuss.*"

We address the replication comment in our responses to Reviewer #1, Comment #16; Reviewer #2, Comments #2, and #6.

We discuss differences in the ages of the cohorts as a potential source of lack of replication in our response to Reviewer #1, Comment #8.

We address the toning down of the Discussion in our responses to Reviewer #2, Comments #1 and #2.

Comment 2:

"** Strongest association signal from the IGAP meta-analysis is found near the TREML2 gene. Analysis of TREML2 gene expression was excluded due failure to detect expression levels. Do the authors expect that TREML2 expression regulation would be relevant to the protective effects of TREM locus eQTL variants? Do the authors expect any eQTL effect of rs9357347 on TREML2 or additional TREM locus transcripts?*"

Table S1 now includes the percent detection of each of the five *TREM* coding genes, in each of the three expression datasets evaluated in this study. These data demonstrate the low percentage of subjects with detectable levels of *TREML2* across all 3 cohorts.

Table S1. Percent detection of *TREM* locus transcripts.

			Mayo WG-DASL: Cerebellum[a]			Mayo WG-DASL: Temporal Cortex[b]			Mayo Clinic RNAseq: Temporal Cortex[c]	ROS/MAP RNAseq: DFPC[d]
Symbol	**Ensembl Gene ID**	**WG-DASL Probe ID**	**AD + non-AD**	**AD**	**nonAD**	**AD + non-AD**	**AD**	**Non-AD**	**AD + non-AD**	**AD + non-AD**
TREM1	ENSG00000124731	ILMN_1688231	0.00	0.00	0.00	0.25	0.00	0.51	18.18	66.99
TREML1	ENSG00000161911	ILMN_1690783	100.00	100.00	100.00	100.00	100.00	100.00	100.00	97.84
TREM2	ENSG00000095970	ILMN_1701248	40.91	43.59	37.29	98.25	99.50	96.95	100.00	100.00
TREML2	ENSG00000112195	ILMN_1740864	17.38	17.44	17.51	6.27	10.40	2.03	8.33	24.75
TREML4	ENSG00000188056	ILMN_2205322	6.15	4.62	7.91	2.26	2.97	1.52	2.27	15.13

The percentage of samples with detectable expression of *TREM* family transcripts in each of the expression datasets studied. For the WG-DASL dataset (a,b) the corresponding WG-DASL probe is indicated. Only *TREML1* and *TREM2* expression are detectable above background in at least 50% of the Mayo WG-DASL samples tested (a,b), in at least one tissue; c: A detection threshold >-1, for cqn normalized expression levels was used to determine percent detection; d: percent detection was calculated as the proportion of subjects who express > 0 FPKM, DFPC = dorsolateral prefrontal cortex.

Genes that are lowly expressed, or not expressed in the brain are unlikely to directly impact the pathophysiology of AD. However, it is indeed possible that their expression in the periphery could affect the disease process, as had been stated in the discussion (page 16, 2nd paragraph, penultimate sentence) and shown below:

"Alternatively, the protective effect of p.S144G could be mediated directly through the function of TREML2 in a cell with abundant expression, such as macrophages, in which TREML2 is known to be upregulated in response to inflammation, [36]."

We now add (page 16, 2nd paragraph, last sentence) that: "It is also possible that significant rs9357347 eQTL associations would be detected with *TREML2* or other *TREM* locus transcripts in tissues where these genes are more abundantly expressed."

Also, it is important to acknowledge that moderate levels of expression in specific central nervous system cell types, may not be detectable in tissue samples composed of a heterogenous set of cell types, such as brain tissue used in our study. We now raise this point in the discussion section, page 16, 1st paragraph, last two sentences.

"Furthermore, these estimates are based on RNA isolated from tissue samples and not microglial cells where both *TREM2* and *TREML1* are predominantly expressed [35], and where expression levels of these genes may be impacted to a greater extent by regulatory variants. Additional studies will be needed to determine the impact of such expression changes on the biology of microglial cell function."

Comment 3:

"Lower bound cut-off for inclusion of RNA was set at RIN>5. Regardless of the correction for RIN value that was employed in the eQTL analysis of rs9357347, could the authors identify any group differences in TREM2 and TREML1 expression regulation or detection percentage when clustering samples based on RIN values?"

We have done additional analyses, which we depict in a new Figure (**Fig. S6**) and in a new Table (**Table S3**). In summary, RIN does not significantly impact the results of the eQTL analyses. We added the following new text to Supplementary Material, page 7, last paragraph.

"Effect of RIN on percent detection and rs9357347 eQTL association

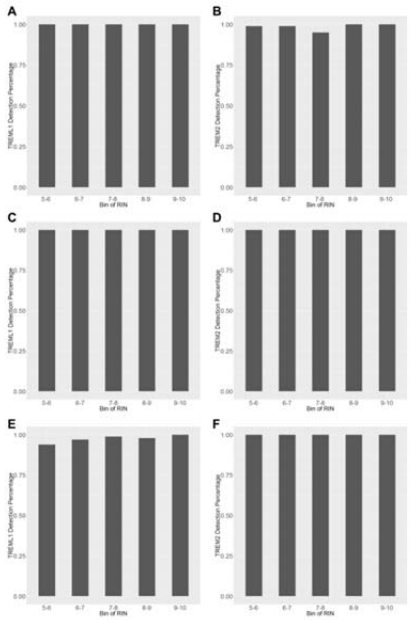
Fig. S6 shows *TREML1* and *TREM2* detection percentage stratified by RIN, and demonstrates that neither *TREML1* nor *TREM2* detection percentage is affected by RIN. **Table S3** shows results of the rs9357347 eQTL associations in the Mayo WG-DASL dataset when stratifying by samples above and below the median RIN of 6.5 (**Table S3**). These results indicate that RIN does not significantly impact the magnitude of the rs9357347 eQTL associations, as the estimates of the beta coefficients overlap with those observed in the analysis not stratified by RIN (**Table S2**). Although the significance of the association is lessened in the stratified analysis, this is likely due to the smaller sample size of the stratified groups compared to the sample size of the combined analysis."

Table S3. Association of the *TREM* locus candidate regulatory variant, rs9357347, with *TREML1* and *TREM2* gene expression stratified by RIN.

Gene Symbol	RIN group	N	beta	SE	p-value
TREML1	RIN < 6.5	188	0.087	0.042	0.043
	RIN > 6.5	192	0.084	0.046	0.069
TREM2	RIN < 6.5	188	0.110	0.063	0.084
	RIN > 6.5	192	0.075	0.065	0.250

Data shown for Mayo WG-DASL temporal cortex (AD+Non-AD) dataset. Samples were stratified into two groups representing those with a RIN below the median RIN of 6.5 and those with a RIN above 6.5. N = sample size. SE= standard error. Given that all expression measures were on a log2 scale, fold-change for the beta coefficients = 2^{beta}.

Fig. S6. Bar charts of percentage of subjects with detectable gene expression for *TREML1* and *TREM2* across groups of subjects defined by RIN value.



Subjects were binned according to RIN value and the proportion of subjects in each bin that met the detection threshold was calculated. A and B: Expression detection percentage for each RIN bin for *TREML1* (A) and *TREM2* (B) in the Mayo WG-DASL dataset. C and D: Expression detection percentage for each RIN bin for *TREML1* (A) and *TREM2* (B) in the Mayo Clinic RNAseq dataset. E and F: Expression detection percentage for each RIN bin for *TREML1* (A) and *TREM2* (B) in the ROS/MAP RNASseq dataset.

Minor Comment 1:

"*Please report RNA concentrations submitted for sequencing*."

These concentrations are now indicated as shown below:

Supplementary Material, page 4, 2nd paragraph, 3rd sentence for the Mayo RNAseq: "Samples were randomized prior to the transfer of 40 (TCX) or 50 (CER) ng/ul of RNA to the Mayo Clinic Medical Genome Facility Gene Expression and Sequencing Cores for library preparation and sequencing."

Supplementary Material, page 5, 3rd paragraph, 3rd sentence for the ROS/MAP RNAseq: "Only samples with a RIN score >5 were used for library construction, which was assembled using 50ng/ul of RNA for the strand-specific dUTP method."

Minor Comment 2:

"* Detection percentage for TREM locus transcripts is provided for the Illumina WGS-DASL microarray dataset only. Please provide the detection percentages for all seven TREM locus transcripts in both the discovery and replication datasets."

Table S1 now includes the percent detection of each of the five *TREM* coding genes, in each of the three expression datasets evaluated in this study. The WG-DASL array lacked probes for the two *TREM* pseudogenes; therefore they were not measured in the Mayo WG-DASL cohort. Expression levels were not available for the *TREM* pseudogenes in the ROS/MAP dataset. Based on the Mayo RNAseq dataset, the percent detection of *TREML3P* and *TREML5P* in temporal cortex are 15% and 5% respectively. This information is now included in the legend of **Table S1**.

Please also see responses to Main Comment #2 by Reviewer #4.

Minor Comment 3:

"* *Interpretation of pairwise LD between rare and common variants is difficult due to frequency inequalities. Conclusions on LD blocks (e.g. page 9, last paragraph) should be presented with more caution.*"

We note that the LD blocks and pairwise disequilibrium were assessed in our study using both D' and r^2. D' provides a more accurate assessment of LD between variants that have different allele frequencies than r^2, as D' is a relative measure of LD based of the maximum disequilibrium attainable given the allele frequencies. The use of both LD measures is stated throughout the manuscript and in Figure 3 (relevant section of legend pasted below).

"**Fig. 3. LD Plot of *TREM* locus variants.**

LD plot of *TREM* locus variants where haplotype blocks were determined with the solid spine definition; square colors correspond to D' (tight LD=warmer colors, weak LD=cooler colors) and r^2 values are shown within the squares (**Supplementary Methods**)."

Minor Comment 4:

"* Figures 1, 2 and 3 are a little rough and premature; the authors might reshape these figures to aid interpretation of gene and variant positions."

Please note that all figures were submitted as high resolution images, all of which are of high quality and can be downloaded by clicking on the link provided on the merged pdf of the manuscript documents. If there are specific suggestions regarding how these can be improved further, we will be happy to apply these.

Minor Comment 5:

* Exclusion of missense variant TREML2 p.S144G as a functional factor is motivated by the variant being labeled 'benign' in the PolyPhen2 prediction software. This conclusion might be too strong, especially since TREML2 expressions levels could not be evaluated.

We recognize that functional prediction algorithms like PolyPhen are fallible, and acknowledge that *TREML2* p.S144G may be functional and could influence AD-risk, as indicated on page 16, 2nd paragraph, last two sentences:

"Alternatively, the protective effect of p.S144G could be mediated directly through the function of TREML2 in a cell with abundant expression, such as macrophages, in which TREML2 is known to be upregulated in response to inflammation, [36]. It is also possible that significant rs9357347 eQTL associations would be detected with *TREML2* or other *TREM* locus transcripts in tissues were these genes are more abundantly expressed."

Reviewer #5:

"*Drs. Carrasquillo et.al. used both their own samples and data plus publically available RNA-Seq data to elucidate the association between the TREM gene cluster and AD and it's pathologies, primarily by exploring gene expression in the TREM region. They found that a protective variant in TREM2, shows a significant eQTL and this association might be driving or at least contributing to the protective effect seen in the TREM region.*"

Overall nicely done analyses, just a couple comments:"

We thank this reviewer for their positive comments.

Comment 1:

"(*1) The authors should explain the Regulome Score, how it is calculated and how to interpret it. I did not see that in either the main manuscript (methods or results) or supplementary.*"

We now provide a description of Regulome scores, and explain how they are calculated and how to interpret them on page 6, in the Materials and Methods, section 2.1:

"Variants were further prioritized by their Regulome score. Regulome scores were obtained from the Regulome database, which annotates variants with regulatory information from 962 different datasets and a variety of sources, including ENCODE [16]. Regulome scores are on a scale from 1 to 6, and these numerical categories are sub-classified with letters based on the number of lines of evidence of functional consequence. A value of 1a is assigned to the variant with the most evidence of regulatory potential, while a score of 6 has the least [16]."

Comment 2:

"(2) Since models used for the eQTL analyses should be mentioned in the main paper, specifically what covariates were adjusted for. This is important especially considering the differences in the 3 cohorts used for the meta-analyses. And with respect to the adjustments, what was the rationale for adjusting for APOE status. This should be explained."

A description of the statistical analyses is now included in the Material and Methods, section 2.4, of the Main Manuscript on page 7. Also, we now indicate the following in section 2.4:

"The eQTL analysis in the discovery, WG-DASL dataset, included *APOE* ε4 dose as a covariate given the strong effect of this allele on AD. However, since a significant association was not detected with this covariate in the rs9357347 eQTL analyses, *APOE* ε4 dose was not included in the eQTL analyses implemented on the replication cohorts. For comparison, we have performed the eQTL analyses in all three datasets with and without adjustment for *APOE* ε4 dose and do not observe a substantial difference in the estimates of the association between these two models."

Please also see our response to Reviewer #2, Comment #4.

Comment 3:

"(3) p=0.14 or p=0.11 is not "suggestive association". That is quite a stretch. Just focus on the meta analysis results and direction"

We have modified this sentence in light of the new results discussed in the response to comment #4 of this reviewer, and have replaced the phrase "suggestive association" on page 12, 4th paragraph, 2nd sentence, as shown below:

"Although in the ROS/MAP RNAseq dataset a significant association was only detected with the levels of *TREM2* (**Table 3**), meta-analysis from the three independent study p-values (Mayo WG-DASL, Mayo RNAseq and ROS/MAP RNAseq) yielded significant results (*TREML1* p=3.4x10^{-2}; *TREM2* p=3.5x10^{-3}), confirming the association of the rs9357347 minor allele with increased *TREML1* and *TREM2* gene expression."

Please also see our response to Reviewer #1, Comment #16 and Reviewer #2, Comment #6.

Comment 4:

"Table 3: It would be helpful to include both MAF for each cohort/gene and also SE of the Betas. The B=0.43 for the ROSMAP TREM2 is a bit out of place. The authors should comment on what might be driving the very different beta (different from the other 2 cohorts)."

We have added to **Table 3** (also shown below) the MAF in each cohort and the SE of the beta coefficients. We are grateful for this reviewer's comment regarding the larger beta reported for the ROS/MAP cohort. To address this concern, we plotted the expression values used as input for the linear regression analyses and realized that the input values for the ROS/MAP cohort had not been log2 transformed, but were rather the FPKM values. The ROS/MAP eQTL results presented now in **Table 3** and elsewhere in the manuscript were re-generated using the log2 transformed FPKM values. Upon this correction, the beta coefficients estimated in the ROS/MAP dataset overlap with the beta estimates in the discovery, WG-DASL cohort. In the corrected analysis of the ROS/MAP cohort, association of rs9357347 with *TREM2* levels is significant. The pertinent text in the manuscript is modified accordingly.

Please also see the response to Reviewer #1 Comments #8 and #16; Reviewer #2 Comment #6 and Reviewer #4, Comment #1.

Table 3. Meta-analysis of rs9357347 eQTL results from three independent datasets.

Dataset	Sample size	MAF	*TREML1*			*TREM2*		
			beta	SE	p	beta	SE	p
Mayo WG-DASL	380	0.307	0.088	0.032	6.28E-03	0.090	0.045	4.61E-02
Mayo Clinic RNAseq	132	0.311	-0.030	0.108	7.82E-01	0.084	0.128	5.13E-01
ROS/MAP RNAseq	494	0.281	0.089	0.114	4.36E-01	0.124	0.060	3.77E-02
Meta-analysis	1006		**+-+**		**3.36E-02**	**+++**		**3.54E-03**

Meta-analysis of rs9357347 eQTL results from temporal cortex (Mayo WG-DASL and Mayo Clinic RNAseq) and dorsolateral prefrontal cortex samples (ROS/MAP). MAF = minor allele frequency. SE = standard error. Since in all three datasets the expression measures analyzed were on a log2 scale, fold-change for the beta coefficients = 2^{beta}. The meta-analysis was performed using METAL, with weighted average of z-scores from the individual study p-values, weighted according their sample size.

A candidate regulatory variant at the *TREM* gene cluster associates with decreased Alzheimer's disease risk, and increased *TREML1* and *TREM2* brain gene expression

Minerva M. Carrasquillo[1], Mariet Allen[1], Jeremy D. Burgess[1], Xue Wang[2], Samantha L. Strickland[1], Shivani Aryal[1], Joanna Siuda[1], Michaela L. Kachadoorian[1], Christopher Medway[1,23], Curtis S. Younkin[1], Asha ~~Nair[3]~~Nair[4], Chen ~~Wang[3]~~Wang[4], Pritha ~~Chanana[3]~~Chanana[4], ~~Xue Wang[4],~~ Daniel ~~Serie[4]~~Serie[2], Thuy Nguyen[1], Sarah Lincoln[1], Kimberly G. Malphrus[1], Kevin Morgan[32], Todd E. Golde[5], Nathan D. Price[6], Charles C. White [7,8], Philip L De Jager[7,8,9], David A. Bennett[10], Yan W. ~~Asmann[4]~~Asmann[2], Julia E. ~~Crook[4]~~Crook[2], Ronald C. Petersen[11], Neill R. Graff-Radford[12], Dennis W. Dickson[1], Steven G. Younkin[1], Nilüfer Ertekin-Taner[1,12*]

[1] Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL 32224, USA.

~~[4]~~[2] Department of Health Sciences Research, Mayo Clinic, Florida, Jacksonville, FL 32224, USA.

~~[2]~~[3] Human Genetics Group, University of Nottingham, Nottingham, UK NG7 2UH.

~~[3]~~[4] Department of Health Sciences Research, Mayo Clinic Rochester, MN 55905, USA

~~[4] Department of Health Sciences Research, Mayo Clinic, Florida, Jacksonville, FL 32224, USA.~~

[5] Department of Neuroscience, University of Florida, Gainesville, FL, 32610 USA.

[6] Institute for Systems Biology, Seattle, WA, 98109 USA.

[7] Ann Romney Center for Neurologic Diseases and Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA, 02115 USA.

[8] Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, 02142 USA.

[9] Harvard Medical School, Boston, MA, 02115 USA.

Formatted: Highlight

[10] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612 USA.

[11] Department of Neurology, Mayo Clinic Minnesota, Rochester, MN 55905 USA.

[12] Department of Neurology, Mayo Clinic Florida, Jacksonville, FL 32224, USA.

* To whom correspondence should be addressed at: Departments of Neurology and Neuroscience, Mayo Clinic Florida, Birdsall Building 327, 4500 San Pablo Rd, Jacksonville, FL 32224; E-mail: taner.nilufer@mayo.edu; Tel: 904-953-7103; Fax: 904-953-7370

Abstract

INTRODUCTION: We hypothesized that common Alzheimer's disease (AD)-associated variants within the triggering receptor expressed on myeloid (*TREM*) gene cluster influence disease through gene expression.

METHODS: Expression microarrays on temporal cortex and cerebellum from ~400 neuropathologically diagnosed AD and non-AD subjects, and two independent RNAseq replication cohorts were used for expression quantitative trait locus (eQTL) analysis.

RESULTS: *TREML1* and *TREM2* have reliably detectable expression. A variant within a DNase hypersensitive site 5' of *TREM2*, rs9357347-C, associates with reduced AD-risk and increased *TREML1* and *TREM2* levels. Meta-analysis on eQTL results from three independent datasets (n=1,006) confirmed these associations ($p=9.3 3.4 \times 10^{-3 -2}$ and $9.3 \times 10 3.5 \times 10^{-43}$, respectively).

DISCUSSION: Our findings point to rs9357347 as a functional regulatory variant that contributes to a protective effect observed at the *TREM* locus in the International Genomics of Alzheimer's Project (IGAP) GWAS meta-analysis, and suggest concomitant increase of *TREML1* and *TREM2* brain levels as a potential mechanism for protection from AD.

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

Keywords: Alzheimer's disease, eQTL, *TREM2*, *TREML1*, regulatory variant

1. Introduction

Whole genome and exome sequencing are used as complementary approaches to uncover novel loci that can be missed by GWAS, and enabled the discovery of ~~strong, yet~~ rare, missense ~~coding AD-risk~~ alleles within *TREM2* that have a relatively large effect size on AD-risk [1, 2]. *TREM2* is a member of the triggering receptor expressed on myeloid (TREM) family, known to play a key role in modulating inflammation in the innate immune response [3]. This finding provided strong supportive evidence for the importance of inflammation in the etiology of AD, but the specific role played by *TREM2* in AD pathophysiology remains unclear [4].

Since the first two reports [1, 2], the risk effect of the most significant *TREM2* rare missense variant p.R47H (a.k.a. rs75932628) has been replicated in multiple Caucasian series [5-9], including a large meta-analysis of 24,086 AD cases and 148,993 controls [10]. *TREM2* resides within the *TREM* gene cluster on chromosome 6p21.1 (**Fig. 1**), which also includes the protein coding genes *TREM1*, *TREML1*, *TREML2*, *TREML4* that could be additional plausible AD-risk genes.

A missense variant in *TREML2*, p.S144G (a.k.a. rs3747742), that is not in linkage disequilibrium (LD) with *TREM2* p.R47H, was reported to associate with reduced AD-risk [11]. *TREML2* p.S144G is in ~~strong~~ tight LD with the ~~*TREM* locus~~ intergenic variant, rs9381040, ~~that showed nominally significant AD association~~ that demonstrated the most significant association at the *TREM* locus in the IGAP AD-risk GWAS meta-analysis ($p=6x10^{-04}$) [12] ~~and which was the strongest variant at this locus in the IGAP dataset [11]~~. The authors concluded that *TREML2* p.S144G is the functional variant that accounted for the IGAP *TREM* locus signal, ~~although~~even though the significance of the AD-risk association with the intergenic rs9381040 is ~~stronger~~

Formatted: Highlight
Formatted: Highlight
Formatted: Font: Not Italic, Highlight
Formatted: Highlight
Formatted: Font: Not Italic
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Font: Italic, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Superscript, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

greater than that observed with ~~for~~ p.S144G. Further, *TREML2* p.S144G does not have a predicted functional consequence (PolyPhen2 score=benign) or demonstrated functional outcome, suggesting that the IGAP signal at the *TREM* locus may be due to other functional variants.

Some variants at the *TREM* locus have been reported to show associate~~e~~ion with AD endophenotypes [11, 13, 14]. Cerebrospinal fluid (CSF) levels of AD biomarkers, tau and ptau, associate with three variants at the *TREM* locus that are not in LD with each other: *TREM2* p.R47H (rs75932628), rs6916710 located in intron 2 of *TREML2*, and rs6922617 located downstream from *NCR2* and outside the *TREM* cluster. Of these variants, only *TREM2* p.R47H was associated with AD-risk [13]. More recently, a variant upstream of *TREM2* (rs7759295) and a variant in intron 3 of *TREM1* (rs6910730) were reported to be independently associated with increased AD pathology burden and increased rate of cognitive decline [14]. However, neither of these two variants shows association with AD-risk in the IGAP meta-analysis (p>0.05) [15].

~~Thus, it is possible that the effect observed with these variants on AD endophenotypes is due to their LD with an as yet defined functional variant(s) that influences AD risk at the *TREM* locus.~~ Thus, other than *TREM2* p.R47H, none of the *TREM* locus variants previously reported to associate with AD endophenotypes show association with AD-risk. Functional AD~~-~~risk variants that influence AD endophenotypes are expected to show associatione both with these endophenotypes and risk of AD. Therefore, it is possible that these latter variants are not the functional variants *per se*, but merely markers of other un-tested functional variants.

Collectively, these prior findings suggest that besides the *TREM2* rare missense variants, there may be ~~additional~~ common variants at the *TREM* locus that influence AD-risk and/or its endophenotypes. We hypothesized that some of the common AD-risk variants at the *TREM* locus

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Italic

confer disease risk via regulation of transcript levels of coding genes at the *TREM* gene cluster. In this study, we characterized the brain expression levels of the *TREM* family genes using microarray expression data; validated expression levels by RNA sequencing (RNAseq); performed genetic associations with *TREM* locus genes reliably detected in cerebellum and temporal cortex with single nucleotide polymorphisms (SNP) that were also tested in the IGAP AD-risk GWAS meta-analysis; and annotated these variants for their effects on *TREM* gene expression levels and regulatory potential. Further, we obtained results for the top putative regulatory SNP from two other, independent cohorts with brain RNAseq data and performed meta-analysis of all three cohorts. Our findings suggest that the protective association at the *TREM* locus observed in the IGAP meta-analysis may be due, at least in part, to a common regulatory variant that influences brain levels of *TREM2* and *TREML1*.

Formatted: Highlight

2. Materials and Methods

2.1 Variant selection

We restricted our analysis to variants located within 100kb of any coding *TREM* family gene at the chromosome 6p21.1 *TREM* gene cluster (**Fig. 1**). Variants were further selected based on the strength statistical significance of their AD-risk association in the IGAP stage 1 meta-analysis [12] (**Supplementary Methods**), where only those variants with p-values ≤ 0.0015 were kept. This p-value cut-off was arbitrarily chosen to select those variants that existed in both the IGAP stage 1 AD GWAS and our discovery eQTL cohort, Mayo Clinic Whole Genome-DASL dataset, and that could be genotyped, if needed, in the replication eQTL cohorts, using cost-effective medium-throughput assays. Variants were further prioritized by their Regulome score. Regulome scores were obtained from the The regulatory potential of the tested

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

~~variants was assessed utilizing the~~ Regulome database, which annotates variants with regulatory information from 962 different datasets and a variety of sources, including ENCODE [16]. Regulome scores are on a scale from 1 to 6, and these numerical categories are sub-classified with letters based on the number of lines of evidence of functional consequence. A value of 1a is assigned to the variant with the most evidence of regulatory potential, while a score of 6 has the least [16].~~.~~

2.2 Mayo Clinic Whole Genome-DASL dataset (Discovery eQTL cohort)

We utilized Illumina (Whole Genome-DASL=WG-DASL, Illumina, San Diego, CA) microarray gene expression data from our published human brain expression genome-wide association study (Mayo Clinic eGWAS) [17] conducted on brain tissue from autopsied AD patients (197 cerebellum, 202 temporal cortex) and non-AD subjects (177 cerebellum, 197 temporal cortex) (**Table 1**). All AD subjects had neuropathologic diagnosis of definite AD [2]. The non-AD subjects did not fulfill neuropathologic criteria for definite AD, but many had other unrelated pathologies. Expression measures were generated as described previously [17]. A description of this cohort~~,~~ and generation of expression measures~~, and eQTL analysis is~~ is provided in the **Supplemetary Methods**.

2.3 RNAseq datasets (Replication eQTL cohorts)

Temporal cortex RNAseq data from two RNAseq cohorts: "Mayo Clinic RNASeq" and "ROS/MAP RNAseq" were employed for replication of the associations that were detected with the WG-DASL gene expression measurements. The Mayo Clinic RNASeq dataset is comprised of 84 LOAD and 48 non-AD brains from the Mayo Clinic Brain Bank that were not part of the Mayo Clinic WG-DASL cohort but whose neuropathological diagnosis followed the same

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Not Highlight
Formatted: Font: Font color: Auto
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

criteria. The ROS/MAP RNAseq dataset is comprised of ~~, and dorsolateral prefrontal cortex~~ RNAseq data from 288 AD and 206 non-AD samples that are part of the ROS/MAP cohort (**Table 1**) previously described [18, 19]. ~~were employed for replication of the associations that were detected with the WG-DASL gene expression measurements.~~ Methodological details for the RNAseq data generation ~~and eQTL analysis~~ are provided in the **Supplementary Methods**.

2.4. Statistical Analysis

Normalized transcript expression levels, on a log2 scale, were tested for associations with *TREM* locus genotypes in each of the three datasets (Mayo WG-DASL, Mayo Clinic RNAseq and ROS/MAP RNAseq) via multivariable linear regression analyses implemented in PLINK [20]. An additive model was applied adjusting for age-at-death, sex, diagnosis, RNA Integrity Number (RIN) and adjusted RIN squared $(RIN-RINmean)^2$, in all expression analyses, and *APOE* ε4 dosage and PCR plate in Mayo WG-DASL only, and flowcell in the Mayo Clinic RNAseq dataset only. The eQTL analysis in the discovery, WG-DASL dataset, included *APOE* ε4 dose as a covariate given the strong effect of this allele on AD. However, since a significant association was not detected with this covariate in the rs9357347 eQTL analyses in the discovery set, *APOE* ε4 dose was not included in the eQTL analyses implemented on the replication cohorts. For comparison, we have performed the eQTL analyses in all three datasets with and without adjustment for *APOE* ε4 dose and do not observe a substantial difference in the association results between these two models.

Meta-analyses were performed on eQTL results from the three independent datasets. For these analyses, METAL [21] was implemented using weighted average of z-scores from the individual study p-values, weighted according their sample size.

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

3. Results

In the WG-DASL gene expression data from the temporal cortex (n=399) and cerebellum (n=374) of neuropathologically diagnosed AD and non-AD subjects (**Table 1**), we observed that of the 5 *TREM* locus coding genes, only *TREML1* and *TREM2* were reliably detected (**Table S1 and Fig. 2)**. *TREML1* was detected in both the temporal cortex and cerebellum, while *TREM2* was reliably detected only in the temporal cortex. We validated *TREML1* and *TREM2* WG-DASL temporal cortex gene expression measurements, using RNAseq data generated from a subset of 93 autopsied AD subjects who also had microarray data. There was highly significant correlation between WG-DASL and RNAseq measurements for both *TREML1* (r_s=0.65, $p<10^{-40}$) and *TREM2* (r_s=0.80, $p<10^{-40}$) (**Fig. S1**).

Variants located within 100kb of the 5' or 3'end of any *TREM* coding gene that demonstrated association with AD-risk in the IGAP stage I meta-analysis (17,800 AD vs. 37,154 controls, p≤0.0015), were evaluated for their association with *TREML1* expression in the temporal cortex and cerebellum, and with *TREM2* expression in the temporal cortex. Of the 1,002 variants tested at this locus in the IGAP stage 1 meta-analysis, 28 had p-values ≤ 0.0015, and 16 of these have been genotyped in the autopsied samples in the Mayo Clinic brain expression genome-wide association study (Mayo eGWAS). We also assessed 5 other variants at this locus previously reported to be associated with either reduced AD-risk (rs3747742) [11], increased AD pathology burden and cognitive decline (rs6910730, rs7759295) [14], or decreased CSF tau levels (rs6916710, rs6922617) [13]. **Table 2** shows the association of *TREML1* and *TREM2* gene expression with these 21 variants. In 399 combined AD and non-AD temporal cortex samples tested for the 16 IGAP variants, 5 SNPs ~~achieved nominally significant~~

Formatted: Highlight
Formatted: Subscript, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

~~association~~ showed association (uncorrected p<0.05) with increased levels of both *TREML1* and *TREM2* (rs9381040, rs2093395, rs9357347, rs9394778, rs9296359), and a sixth variant (rs9394767) was significantly associated with increased *TREML1* levels only. As shown in **Fig. 3**, four of the six variants that associate with increased levels of *TREML1* and *TREM2* are in a single LD block (block 2: rs9357347, rs9381040, rs2093395 and rs9394767) and in tight linkage disequilibrium with each other ($r^2 \geq 0.90$). Of these variants, rs9381040 ~~is the strongest~~has the most significant ~~IGAP~~ AD-risk ~~associating~~ association in ~~SNP~~ the IGAP stage 1 meta-analysis~~at the *TREM* locus~~ (**Table 2**). This IGAP "hit" is located 5.5kb downstream from *TREML2* and 23.7kb upstream from *TREM2* and is associated with *TREML1* and *TREM2* expression (p=0.0083, beta=0.086 and p=0.048, beta=0.091, respectively). Given that the expression measures were on a log2 scale, these changes in expression are equivalent to *TREML1* and *TREM2* fold-changes of 1.06 and 1.07, for each copy of the minor allele, respectively. Notably, the minor allele of the IGAP "hit" rs9381040 is associated with both decreased AD-risk and increased *TREML1* and *TREM2* levels. However, based on data from the Roadmap Epigenomics Consortium [22], rs9381040 lacks evidence of regulatory potential in brain regions relevant to AD.

The variant with the ~~strongest effect~~most significant association ~~on~~ with brain *TREML1* expression, which also associates with *TREM2* levels, is rs9357347 in block 2 (**Fig. 3**). This SNP is located 6.9kb downstream from *TREML2* and 19.6kb upstream from *TREM2* and is in tight LD with the IGAP "hit" rs9381040 (D'=0.99, r^2=0.96). As expected, the minor allele of rs9357347 is ~~associates~~ associated with reduced AD-risk (OR=0.95, 95% CI=0.91-0.98, p=0.001) in the IGAP ~~GWAS~~meta-analysis [12] and with increased *TREML1* and *TREM2* expression in the temporal cortex in the Mayo Clinic WG-DASL eQTL analysis (p=0.0063,

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

beta=0.088 and p=0.046, beta=0.090, respectively) (**Table 2 and Fig. S2**). These beta coefficients can be interpreted as an estimated 1.06-fold change of both *TREML1* and *TREM2*, per rs9357347 minor allele, in this temporal cortex dataset. Unlike the IGAP "hit" (rs9381040), rs9357347 lies within sequence subject to histone modifications and within a DNAse hypersensitive site detected by the Roadmap Epigenomics Consortium [22] in brain regions relevant to AD pathology such as the hippocampus. Furthermore, this variant~~, and it~~ is predicted to affect transcription factor binding (SP1 and PPAR) as catalogued in HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) [23]. Consequently, it has a ~~strong~~ compelling Regulome score of 2b (http://www.regulomedb.org/) due to the evidence of its regulatory potential [16] (**Table 2**). Indeed, of all the variants with an AD-risk p-value<0.0015 in the IGAP meta-analysis, and p-values<0.05 in our WG-DASL eQTL analysis of temporal cortex *TREML1* and *TREM2* gene expression levels, rs9357347 had the greatest regulatory potential as determined by their Regulome scores (**Fig. S3 and Fig. S4**).

Formatted: Highlight

Formatted: Highlight

The other two variants with gene expression associations in the temporal cortex are in a different LD block (block 4: rs9394778 and rs9296359) and in ~~strong~~ tight LD with each other (r^2 = 0.67). These SNPs are more ~~strongly~~ significantly associated with *TREM2* than with *TREML1* expression; however, neither has ~~a strong~~compelling evidence of regulatory potential as both have Regulome scores of 6 (**Table 2**). In the 374 AD and non-AD subjects with cerebellum expression measures, none of the 16 IGAP AD-risk associated variants that were tested, associate with *TREML1* gene expression (p>0.05).

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

We determined the extent of linkage disequilibrium (LD) between the likely regulatory variant rs9357347, the IGAP "hit" rs9381040 and the significant *TREM2* rare missense AD-risk variants p.D87N (rs142232675) and p.R47H (rs75932628) [1]. As shown in **Fig. 3**, these two

TREM2 rare missense AD-risk variants are not in LD with either rs9357347 or rs9381040. This suggests that the protective effect of the regulatory rs9357347 and the IGAP "hit" are independent of the rare, missense *TREM2* variants.

We next evaluated LD amongst variants tested at this locus, including common *TREM* locus variants previously reported to have associations with AD-risk (rs3747742) [11], increased AD pathology burden and cognitive decline (rs7759295 and rs6910730) [14], or with lower CSF ptau (rs6922617 and rs6916710) [13]. The missense *TREML2* variant rs3747742 (p.S144G) is in LD with the regulatory variant implicated in our study, rs9357347. As reported, rs3747742 is also in LD with rs9381040 (IGAP hit); and as expected associates with reduced AD-risk (p=0.009), however with slightly lesser ~~strongly~~ significance than the AD-risk association of the regulatory rs9357347 (p=0.001) or the IGAP "hit" rs9381040 (0.0006). Further, the association of rs3747742 ~~has less strong association~~ with ~~brain~~ *TREML1* expression is not as significant as that of rs9357347.~~,~~ In addition, rs3747742 has no association with brain *TREM2* levels, and has a weak Regulome score of 6 (**Table 2**).

Of the four common *TREM* locus variants that associate with AD endophenotypes, only rs6916710 is in ~~strong~~ tight LD with the regulatory rs9357347 (D'=0.91, r^2=0.62). However, rs6916710, does not show significant association with AD-risk in the IGAP meta-analysis (p=0.103) nor with *TREML1* or *TREM2* gene expression levels (**Table 2**).

None of the other three common *TREM* locus variants with reported AD-endophenotype associations are in tight ~~strong~~ LD with the regulatory rs9357347 or any of the other *TREM* locus variants that are associated with AD-risk. Only rs7759295 showed association with ~~nominally~~ ~~significant~~ *TREML1* ~~brain~~ gene expression ~~association~~ (uncorrected p=0.04), but neither this nor

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

any of the other AD-endophenotype-associated SNPs have evidence of AD-risk association or ~~strong~~ Regulome scores that are indicative of likely regulatory function (**Fig. 3 and Table 2**).

Utilizing publicly available RNAseq data from two independent cohorts (**Table 1**) that do not overlap with the samples included in the WG-DASL eQTL analysis, we sought replication of the rs9357347 association with *TREML1* and *TREM2*. Although in the ROS/MAP RNAseq dataset a significant association was only detected with the levels of *TREM2* ~~the results were not replicated in the smaller of the two cohorts (Mayo RNAseq 84 AD and 48 non-AD: *TREML1* beta= 0.03, p=0.78; *TREM2* beta=0.08, p=0.51), rs9357347 demonstrated suggestive association with increased *TREML1* and *TREM2* gene expression in the larger cohort (ROS/MAP RNAseq 288 AD and 206 non-AD: *TREML1* beta= 0.03, p=0.14 ; *TREM2* beta=0.43, p=0.11)~~ (**Table 3**~~).~~ ~~Meta~~meta-analysis from the three independent study p-values (Mayo WG-DASL, Mayo RNAseq and ROS/MAP RNAseq) yielded ~~p-values that reached~~ significant resultsee (*TREML1* p=~~9.3~~3.4x10$^{-3.2}$; *TREM2* p=3.5~~9.3~~x10^{-3}), confirming the association of the rs9357347 minor allele with increased *TREML1* and *TREM2* gene expression. The ~~strength~~ evidence of association with *TREM2* expression was ~~stronger~~greater ~~for the *TREM2* association in this~~ upon meta-analysis compared to the association observed in our ~~the initial~~ discovery ~~results~~dataset~~,~~ ~~:~~ whereas ~~that for~~the evidence of association with *TREML1* expression was slightly ~~weaker~~greater in our discovery dataset compared to the meta-analysis.

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Superscript, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Superscript, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

4. Discussion

In this study, we first sought to characterize the brain expression of *TREM* locus genes based on the premise that those *TREM* cluster genes that are expressed in the brain are likely to be candidate AD-risk genes. We determined that besides *TREM2*, only *TREML1* has reliable expression in the brain regions we studied. Whereas *TREML1* is expressed in both cerebellum

and temporal cortex of all subjects, *TREM2* is expressed in ~~the temporal cortex of~~ 98% of temporal cortex ~~all subjects but only in~~and 41% of cerebellum ~~subjects~~ samples~~in the cerebellum~~. This suggests that cerebellar levels of *TREM2* are lower than those for temporal cortex, consistent with previous reports showing higher gene levels in the temporal cortex than cerebellum [24] and higher protein levels correlating with AD neuropathology [25]. In contrast, *TREM1*, *TREML2* and *TREML4* are expressed in only 0%-17% of the subjects. While lack of reliable brain expression of these genes does not definitively rule them out as plausible AD-risk genes, our findings provide the strongest evidence for *TREML1*, besides *TREM2*, as most likely *TREM* locus genes for further studies in AD.

Consequently, we focused our studies on *TREML1* and *TREM2*; and utilized their brain expression levels as endophenotypes to identify putative regulatory variants that modify risk for AD. Focusing on brain *TREML1* and *TREM2* expression associations with variants at the *TREM* locus that also show evidence of AD-risk association in the publicly available IGAP meta-analysis, we identified a putative regulatory variant, rs9357347, located between *TREM2* and *TREML2*~~,~~ T~~t~~he minor allele of this variant~~which associates~~ is associated with both decreased AD-risk and with increased *TREML1* and *TREM2* brain expression in the temporal cortex. The direction of effect of this variant on AD-risk and brain expression levels of these genes appears to be biologically congruent based on the known functions of these genes.

TREML1, which is also known as TREM-like transcript 1 (TLT-1), is a myeloid receptor expressed exclusively in the α-granules of platelets and megakaryocytes [26]. Identification of higher levels of soluble TREML1 (sTLT-1) in septic patients vs. controls and development of hemorrhage in mice lacking *Treml1* when exposed to inflammatory injury led to the conclusion that TREML1 functions to maintain vascular integrity during inflammation [27]. Further,

Formatted: Highlight

Formatted: Highlight

TREML1 was shown to dampen leukocyte activation during sepsis, and inhibited pro-inflammatory activation of TREM1 by competing with its ligand [28]. These studies strongly support a role for TREML1 in promoting vascular homeostasis and limiting inflammation.

Functional, *in-vitro* studies of *TREM2* rare, missense mutations revealed reduced TREM2 function as a consequence of decreased maturation and ectodomain shedding, also supported by findings of decreased soluble TREM2 levels in the cerebrospinal (CSF) levels of patients with these mutations [13, 29]. TREM2 deficiency also led to increased amyloid pathology and neuronal loss in the 5XFAD mouse model of AD [30]. Interestingly, TREM2 deficiency in an ischemic mouse model resulted in reduced phagocytosis and resorption of infarcted brain tissue, and worse neurological recovery [31]. Collectively, these findings support a neuroprotective role for TREM2 in various neuronal injury models. There are, however, studies with contradictory results for TREM2. In a different mouse model of AD (APP/PS1), knock-out of *Trem2*, resulted in reduction of macrophages infiltrating from the periphery, along with less brain inflammation and reduced amyloid and tau pathology [32]. These opposite findings of *Trem2* knock-out could be due to differences in the mouse models of Alzheimer's disease tested, different *Trem2* knockout mouse lines, and analyses performed at different time points (early stages versus later stages of Alzheimer's disease).

Given these collective data, a regulatory variant that enhances levels of *TREML1* in platelets, and levels of *TREM2* in brain resident microglia could conceivably promote vascular homeostasis and limit inflammatory damage to neurons in AD and potentially other nervous system diseases. Indeed, rs9357347 has ~~strong~~ compelling evidence of regulatory potential as it is located in a known DNase hypersensitive site and affects histone modification in the hippocampus and transcription factor binding, according to the evidence compiled in the

Formatted: Highlight

Regulome database and HaploReg [16, 23]. Interestingly, rs9357347 is predicted to affect transcription factor binding (SP1 and PPAR) as catalogued in HaploReg~~. and t~~These two transcription factors are known be important in regulating key players in the inflammatory response and lipid metabolism [33, 34]. Further, rs9357347 shows the ~~strongest~~ most significant association with *TREML1* gene expression amongst variants at the *TREM* locus with IGAP meta-analysis AD-risk p-values≤0.0015, in addition to its association with brain *TREM2* levels.

The regulatory rs9357347 SNP is in the same haplotype block as the ~~strongest AD-risk associating~~ variant with the most significant AD-risk association at the *TREM* locus in the IGAP meta-analysis, rs9381040, which is an intergenic variant downstream of *TREML2*. Though this IGAP *TREM* locus "hit" SNP has ~~stronger~~ greater evidence of AD-risk association than rs9357347, there is no evidence of regulatory potential for rs9381040 in brain regions relevant to AD.

While the fold change estimates in gene expression associated with rs9357347-C are modest at 6-7%, the biological impact of the increase attributed to each copy of the minor allele, can be significant and may provide sufficient protection from disease in some individuals, particularly when considered over a lifetime. Furthermore, these estimates are based on RNA isolated from tissue samples and not microglial cells where both *TREM2* and *TREML1* are predominantly expressed [35], and where expression levels of these genes may be impacted to a greater extent by regulatory variants. Additional studies will be needed to determine the impact of such expression changes on the biology of microglial cell function.

The *TREML2* p.S144G variant [11], which associates with reduced AD-risk, is also in LD with both rs9357347 and rs9381040. Though proposed to be the functional variant that accounts

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

for the IGAP signal at this locus, *TREML2* p.S144G is not predicted to have a functional consequence based on PolyPhen2 nor does it have evidence of regulatory potential. Further, *TREML2* expression is too low to be reliably measured in brain tissue (TCX and CER). This ~~,~~ rais~~ing~~es the possibility that the association with *TREML2* p.S144G is due to its LD with a functional variant(s) that influences the function or level of a nearby *TREM* gene(s), such as *TREML1* or *TREM2*. Alternatively, the protective effect of p.S144G could be mediated directly through the function of TREML2 in a cell with abundant expression, such as macrophages, in which TREML2 is known to be upregulated in response to inflammation, [36]. It is also possible that significant rs9357347 eQTL associations would be detected with *TREML2* or other *TREM* locus transcripts in tissues were these genes are more abundantly expressed.

Our findings therefore challenge the conclusion that p.S144G is the only functional variant accounting for the protective effect detected in the IGAP meta-analysis at this locus, and propose rs9357347 as an alternative functional variant with regulatory effects. In reality, both variants could have functional consequences and contribute to the IGAP signal. It should be emphasized that, as demonstrated in our LD analysis, *TREM2* p.R47H is not in LD with these two variants, and thus affects AD-risk independently. Both rs9357347 and p.S144G should be tested for their functional potential and influence on outcomes of inflammation and neuroprotection. It remains possible that rs9357347 is in LD with an untested true functional variant with effects on transcription and AD-risk. It is likewise possible that while rs9357347 is associated with both AD-risk and gene expression levels, these joint effects are coincidental due to LD, rather than being related. These possibilities need to be explored through sequencing of the entire *TREM* locus, or via targeted sequencing of LD block 2 where rs9357347 resides.

Formatted: Highlight

Formatted: Highlight

~~Nevertheless~~ Thus, our findings provide a testable hypothesis for a strong candidate functional variant, specific transcription factors and their effects on *TREML1* and *TREM2* levels.

Furthermore, our investigation of variants previously shown to associate with AD-related endophenotypes [13-15] suggests that these are unlikely to be functional AD-risk variants *per se*, though it remains possible that they are markers of functional variants at the *TREM* locus.

In summary, we characterized expression of *TREM* genes in cerebellum and temporal cortex~~,~~ and determined *TREML1* and *TREM2* to be the only reliably expressed *TREM* genes in these brain regions. ~~,~~We identified rs9357347 as a putative regulatory variant that is ~~associates~~ associated with protection from AD and with increased *TREML1* and *TREM2* brain levels, and nominate rs9357347 as one of the functional variants that accounts for the IGAP AD-risk signal. Additional studies are needed to validate the function of this variant, and to explore the possibility of the presence of other variants at this locus that could contribute to associations observed with rs9357347. Importantly, these findings suggest a potential link between *TREML1* and *TREM2*, as well as vascular homeostasis and neuroinflammation as related mediators of neuronal protection and injury in AD and possibly other central nervous system diseases.

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Acknowledgements

We thank the patients and their families for their participation, without whom these studies would not have been possible, and the clinicians, technicians, and administrative staff who helped in the implementation of this study.

This work was supported by the Alzheimer's Association [MNIRGD 2013 award to M.M.C]; Mayo Alzheimer's Disease Research Center [P50 AG0016574 to D.W.D, N.E.T,

N.R.G.-R., R.C.P. and S.G.Y.]; National Institute on Aging [R01 AG025711, AG017216, AG003949 to D.W.D.; R01 AG032990 to N.E.T.; R01 AG018023 to N.R.G.-R. and S.G.Y.; and U01 AG046139 to N.E.T., T.E.G, N.P. and S.G.Y.]; National Institute of Neurological Disorders and Stroke [R01 NS080820 to N.E.T].

Conflict of Interest Statement

Dr. Petersen has been a consultant to Genentech, Inc. Merck, Inc. and Roche, Inc. and has served on a data safety monitoring committee for Pfizer and Janssen Alzheimer Immunotherapy. Dr. Graff-Radford has multicenter treatment study grants from Lilly, TauRx and consulted for Cytox. Dr. Ertekin-Taner consulted for Cytox.

Formatted: Highlight

References

[1] R. Guerreiro, A. Wojtas, J. Bras, M. Carrasquillo, E. Rogaeva, E. Majounie, et al., TREM2 variants in Alzheimer's disease. The New England journal of medicine 2013;368:117-127.

[2] T. Jonsson, H. Stefansson, S. Steinberg, I. Jonsdottir, P.V. Jonsson, J. Snaedal, et al., Variant of TREM2 associated with the risk of Alzheimer's disease. The New England journal of medicine 2013;368:107-116.

[3] J.W. Ford,D.W. McVicar, TREM and TREM-like receptors in inflammation and disease. Current opinion in immunology 2009;21:38-46.

[4] R.E. Tanzi, TREM2 and Risk of Alzheimer's Disease--Friend or Foe? The New England journal of medicine 2015;372:2564-2565.

[5] C. Pottier, D. Wallon, S. Rousseau, A. Rovelet-Lecrux, A.C. Richard, A. Rollin-Sillaire, et al., TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. Journal of Alzheimer's disease : JAD 2013;35:45-49.

[6] B.A. Benitez, B. Cooper, P. Pastor, S.C. Jin, E. Lorenzo, S. Cervantes, et al., TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiology of aging 2013;34:1711 e1715-1717.

[7] J.D. Gonzalez Murcia, C. Schmutz, C. Munger, A. Perkes, A. Gustin, M. Peterson, et al., Assessment of TREM2 rs75932628 association with Alzheimer's disease in a population-based sample: the Cache County Study. Neurobiology of aging 2013;34:2889 e2811-2883.

[8] A. Ruiz, O. Dols-Icardo, M.J. Bullido, P. Pastor, E. Rodriguez-Rodriguez, A. Lopez de Munain, et al., Assessing the role of the TREM2 p.R47H variant as a risk factor for Alzheimer's disease and frontotemporal dementia. Neurobiology of aging 2014;35:444 e441-444.

[9] B.V. Hooli, A.R. Parrado, K. Mullin, W.K. Yip, T. Liu, J.T. Roehr, et al., The rare TREM2 R47H variant exerts only a modest effect on Alzheimer disease risk. Neurology 2014;83:1353-1358.

[10] C.M. Lill, A. Rengmark, L. Pihlstrom, I. Fogh, A. Shatunov, P.M. Sleiman, et al., The role of TREM2 R47H as a risk factor for Alzheimer's disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association 2015;

[11] B.A. Benitez, S.C. Jin, R. Guerreiro, R. Graham, J. Lord, D. Harold, et al., Missense variant in TREML2 protects against Alzheimer's disease. Neurobiology of aging 2014;35:1510 e1519-1526.

[12] J.C. Lambert, C.A. Ibrahim-Verbaas, D. Harold, A.C. Naj, R. Sims, C. Bellenguez, et al., Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics 2013;45:1452-1458.

[13] C. Cruchaga, J.S. Kauwe, O. Harari, S.C. Jin, Y. Cai, C.M. Karch, et al., GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. Neuron 2013;78:256-268.

[14] J.M. Replogle, G. Chan, C.C. White, T. Raj, P.A. Winn, D.A. Evans, et al., A TREM1 variant alters the accumulation of Alzheimer-related amyloid pathology. Annals of neurology 2015;77:469-477.

[15] G. Liu, No association of TREM1 rs6910730 and TREM2 rs7759295 with Alzheimer's disease. Annals of Neurology 2015;epub ahead opf print:

[16] A.P. Boyle, E.L. Hong, M. Hariharan, Y. Cheng, M.A. Schaub, M. Kasowski, et al., Annotation of functional variation in personal genomes using RegulomeDB. Genome research 2012;22:1790-1797.

[17] F. Zou, H.S. Chai, C.S. Younkin, M. Allen, J. Crook, V.S. Pankratz, et al., Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS genetics 2012;8:e1002707.

[18] D.A. Bennett, J.A. Schneider, Z. Arvanitakis,R.S. Wilson, Overview and findings from the religious orders study. Current Alzheimer research 2012;9:628-645.

[19] D.A. Bennett, J.A. Schneider, A.S. Buchman, C. Mendes de Leon, J.L. Bienias,R.S. Wilson, The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. Neuroepidemiology 2005;25:163-175.

448 [20] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, et al., PLINK: a tool set for
449 whole-genome association and population-based linkage analyses. American journal of human genetics
450 2007;81:559-575.
451 [21] C.J. Willer, Y. Li,G.R. Abecasis, METAL: fast and efficient meta-analysis of genomewide association
452 scans. Bioinformatics 2010;26:2190-2191.
453 [22] B.E. Bernstein, J.A. Stamatoyannopoulos, J.F. Costello, B. Ren, A. Milosavljevic, A. Meissner, et al.,
454 The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology 2010;28:1045-1048.
455 [23] L.D. Ward,M. Kellis, HaploReg v4: systematic mining of putative causal variants, cell types,
456 regulators and target genes for human complex traits and disease. Nucleic acids research 2016;44:D877-
457 881.
458 [24] P. Forabosco, A. Ramasamy, D. Trabzuni, R. Walker, C. Smith, J. Bras, et al., Insights into TREM2
459 biology by network analysis of human brain gene expression data. Neurobiology of aging 2013;34:2699-
460 2714.
461 [25] L.F. Lue, C.T. Schmitz, G. Serrano, L.I. Sue, T.G. Beach,D.G. Walker, TREM2 Protein Expression
462 Changes Correlate with Alzheimer's Disease Neurodegenerative Pathologies in Post-Mortem Temporal
463 Cortices. Brain pathology 2015;25:469-480.
464 [26] A.V. Washington, R.L. Schubert, L. Quigley, T. Disipio, R. Feltz, E.H. Cho, et al., A TREM family
465 member, TLT-1, is found exclusively in the alpha-granules of megakaryocytes and platelets. Blood
466 2004;104:1042-1047.
467 [27] A.V. Washington, S. Gibot, I. Acevedo, J. Gattis, L. Quigley, R. Feltz, et al., TREM-like transcript-1
468 protects against inflammation-associated hemorrhage by facilitating platelet aggregation in mice and
469 humans. The Journal of clinical investigation 2009;119:1489-1501.
470 [28] M. Derive, Y. Bouazza, N. Sennoun, S. Marchionni, L. Quigley, V. Washington, et al., Soluble TREM-
471 like transcript-1 regulates leukocyte activation and controls microbial sepsis. Journal of immunology
472 2012;188:5585-5592.
473 [29] G. Kleinberger, Y. Yamanishi, M. Suarez-Calvet, E. Czirr, E. Lohmann, E. Cuyvers, et al., TREM2
474 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. Science
475 translational medicine 2014;6:243ra286.
476 [30] Y. Wang, M. Cella, K. Mallinson, J.D. Ulrich, K.L. Young, M.L. Robinette, et al., TREM2 lipid sensing
477 sustains the microglial response in an Alzheimer's disease model. Cell 2015;160:1061-1071.
478 [31] M. Kawabori, R. Kacimi, T. Kauppinen, C. Calosing, J.Y. Kim, C.L. Hsieh, et al., Triggering receptor
479 expressed on myeloid cells 2 (TREM2) deficiency attenuates phagocytic activities of microglia and
480 exacerbates ischemic damage in experimental stroke. The Journal of neuroscience : the official journal
481 of the Society for Neuroscience 2015;35:3384-3396.
482 [32] T.R. Jay, C.M. Miller, P.J. Cheng, L.C. Graham, S. Bemiller, M.L. Broihier, et al., TREM2 deficiency
483 eliminates TREM2+ inflammatory macrophages and ameliorates pathology in Alzheimer's disease mouse
484 models. The Journal of experimental medicine 2015;212:287-295.
485 [33] A.Y. Wen, K.M. Sakamoto,L.S. Miller, The role of the transcription factor CREB in immune function.
486 Journal of immunology 2010;185:6413-6419.
487 [34] J.D. Brown,J. Plutzky, Peroxisome proliferator-activated receptors as transcriptional nodal points
488 and therapeutic targets. Circulation 2007;115:518-533.
489 [35] Y. Zhang, K. Chen, S.A. Sloan, M.L. Bennett, A.R. Scholze, S. O'Keeffe, et al., An RNA-sequencing
490 transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. The
491 Journal of neuroscience : the official journal of the Society for Neuroscience 2014;34:11929-11947.
492 [36] R.G. King, B.R. Herrin,L.B. Justement, Trem-like transcript 2 is expressed on cells of the
493 myeloid/granuloid and B lymphoid lineage and is up-regulated in response to inflammation. Journal of
494 immunology 2006;176:6012-6021.

495

496

497

Formatted: Line spacing: Double

Figure Legends

Fig. 1. *TREM* gene cluster on Chr 6p21.1. The chromosomal positions are based on the human genome assembly from February 2009 (GRCh37/hg19). There are seven RefSeq genes at the *TREM* locus (*TREM1*, *TREML1*, *TREM2*, *TREML2*, *TREML3P*, *TREML4* and *TREML5P*); however, *TREML3P* and *TREML5P* are non-coding pseudogenes. The transcript figures are taken from the UCSC Genome Browser.

Fig. 2. Location of *TREML1* and *TREM2* WG-DASL probes.

The location of the (A) *TREML1* and (B) *TREM2* WG-DASL probes (highlighted in light blue) are shown relative to their Refseq transcripts. The chromosomal positions are based on the human genome assembly from February 2009 (GRCh37/hg19). As shown, both of these probes are complementary to all RefSeq transcripts for the respective gene. The transcript figures are taken from the UCSC Genome Browser.

Fig. 3. LD Plot of *TREM* locus variants.

LD plot of *TREM* locus variants where haplotype blocks were determined with the solid spine definition; square colors correspond to D' (~~strong~~ tight LD=warmer colors, weak LD=cooler colors) and r^2 values are shown within the squares (**Supplementary Methods**). Red circles: The rare *TREM2* AD-risk missense variants rs142232675 (p.D87N) and rs75932628 (p.R47H) [1]. Blue circles: Variants that associate with increased AD pathology burden and cognitive decline (rs7759295 and rs6910730) [14], or with lower CSF ptau (rs6922617 and rs6916710) [13]. Green circles: The variant with the ~~strongest~~ most significant AD-risk association in the IGAP meta-analysis (rs9381040); rs9357347, which has the most significant~~strongest~~ ~~brain~~ *TREML1* gene expression association, also shows association with~~has brain~~ *TREM2* gene expression

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

~~association~~, IGAP AD-risk association and the best Regulome score within all tested SNPs; and rs9296359 which has the ~~strongest brain~~most significant association with *TREM2* expression ~~association~~. RefSeq gene transcripts are shown above the LD plot relative to the variant position according to the February 2009 human genome assembly (GRCh37hg19) across the targeted genomic region (*TREM* gene +/-100 kb: chr6:41016999-41354457).

Formatted: Highlight

Formatted: Font: Not Italic, Highlight

Formatted: Highlight

Tables

Table 1. Description of samples included in the discovery and replication cohorts utilized for eQTL analysis.

	Mayo Clinic WG-DASL				Mayo Clinic RNAseq		ROS/MAP RNAseq	
	CER		TCX		TCX		PFCX	
	AD	Non-AD	AD	Non-AD	AD	non-AD	AD	non-AD
N	197	177	202	197	84	48	288	206
Mean age +/- SD	73.6 ± 5.6	71.7 ± 5.5	73.6 ± 5.5	71.6 ± 5.6	83.2 ± 8.7	85.7 ± 8.3	89.8 ± 5.8	86.5 ± 7.2
Female, N (%)	101 (51%)	63 (36%)	108 (53%)	78 (40%)	48 (57%)	26 (54%)	186 (65%)	121 (59%)
% *APOE* ε4+	64%	25%	61%	25%	51%	17%	34%	12%

Samples included in the Mayo Clinic eGWAS (discovery cohort), with cerebellar (CER) and temporal cortex (TCX) gene expression measurements from Illumina WG-DASL arrays have been previously described [17]. Samples in the Mayo Clinic RNAseq cohort (replication cohort #1) had temporal cortex gene expression measurements, and did not overlap with the Mayo eGWAS (WG-DASL) cohort. The ROS/MAP RNAseq cohort (replication cohort #2) had dorsolateral prefrontal cortex (PFCX) gene expression measurements, and did not overlap with the Mayo eGWAS (WG-DASL), or with the Mayo Clinic RNAseq cohort. The RNAseq data for these two cohorts is available at the Sage Synapse, AMP AD Knowledge Portal (https://www.synapse.org/#!Synapse:syn2580853/wiki/66722), under synapse IDs syn3388564 (ROS/MAP RNAseq) and syn3163039 (Mayo RNAseq).

Table 2. Association of variants at the *TREM* locus with AD-risk and *TREM* WG-DASL brain gene expression levels.

Chr	SNP	Position hg19	AD-Risk (IGAP Stage1 Meta-analysis)					Brain eQTL (Mayo Clinic eGWAS)						Regulome Score	HapMap CEU MAF
			Effect Allele	Non Effect Allele	Beta	SE	Pvalue	TREML1 CER	TREML1 CER	TREML1 TCX	TREML1 TCX	TREM2 TCX	TREM2 TCX		
6	rs9381040	41,154,650	T	C	-0.0588	0.0171	5.97E-04	0.021	2.3E-01	0.086	*8.3E-03*	0.091	*4.8E-02*	NA	26.70%
6	rs2093395	41,155,026	C	G	-0.0586	0.0172	6.40E-04	0.021	2.3E-01	0.086	*8.3E-03*	0.091	*4.8E-02*	6	27.90%
6	rs2038568	41,158,132	C	G	0.1272	0.0379	7.93E-04	0.018	7.8E-01	-0.080	3.9E-01	-0.186	1.6E-01	5	8.30%
6	rs12194214	41,028,574	C	A	0.1443	0.0432	8.36E-04	-0.081	9.4E-02	-0.104	2.7E-01	-0.129	3.2E-01	6	4.20%
6	rs9462675	41,153,238	A	G	0.1374	0.0416	9.54E-04	-0.015	7.5E-01	-0.114	1.6E-01	-0.207	6.5E-02	5	3.60%
6	rs6933067	41,133,522	C	T	0.1409	0.0431	1.07E-03	-0.013	7.6E-01	-0.098	2.1E-01	-0.134	2.1E-01	7	3.50%
6	rs9357347	41,150,591	C	A	-0.0554	0.017	1.10E-03	0.013	4.6E-01	0.088	*6.3E-03*	0.090	*4.6E-02*	2b	28.10%
6	rs9394767	41,159,905	G	A	-0.0559	0.0172	1.14E-03	0.011	5.7E-01	0.096	*6.5E-03*	0.083	1.0E-01	5	28.80%
6	rs1542638	41,286,604	G	A	0.0544	0.0167	1.14E-03	-0.022	2.2E-01	-0.035	2.9E-01	-0.064	1.6E-01	4	28.30%
6	rs9471491	41,153,622	A	C	0.1401	0.0436	1.31E-03	-0.015	7.5E-01	-0.114	1.6E-01	-0.207	6.5E-02	7	3.50%
6	rs9471495	41,157,372	A	C	0.1364	0.0427	1.40E-03	0.014	8.3E-01	-0.099	2.9E-01	-0.235	7.2E-02	7	3.50%
6	rs9462677	41,158,856	A	T	0.1388	0.0435	1.41E-03	0.016	8.1E-01	-0.099	2.9E-01	-0.235	7.3E-02	7	4.30%
6	rs9394778	41,215,058	A	G	-0.0494	0.0155	1.44E-03	0.015	3.3E-01	0.065	*2.7E-02*	0.099	*1.5E-02*	6	39.80%
6	rs9471494	41,157,344	G	C	0.1359	0.0427	1.46E-03	0.010	8.7E-01	-0.102	2.6E-01	-0.221	8.2E-02	6	4.50%
6	rs6912013	41,061,593	C	T	0.1377	0.0433	1.48E-03	-0.076	1.2E-01	-0.104	2.7E-01	-0.124	3.4E-01	5	2.70%
6	rs9296359	41,205,690	A	G	-0.0518	0.0163	1.48E-03	0.017	2.8E-01	0.066	*2.4E-02*	0.116	*4.6E-03*	6	27.40%
6	rs3747742*	41,162,518	C	T	-0.0455	0.0173	8.56E-03	0.018	2.9E-01	0.072	*2.3E-02*	0.064	1.5E-01	6	28.30%
6	rs6916710*	41,164,788	T	C	-0.0265	0.0163	1.03E-01	0.013	4.3E-01	0.054	7.7E-02	0.072	9.2E-02	7	38.40%
6	rs7759295*	41,135,850	T	C	-0.0228	0.0252	3.66E-01	-0.023	3.5E-01	0.094	*4.0E-02*	-0.008	9.0E-01	6	13.30%
6	rs6910730*	41,246,633	G	A	-0.0103	0.0254	6.86E-01	-0.046	8.5E-02	-0.079	1.2E-01	-0.032	6.5E-01	4	8.40%
6	rs6922617*	41,336,101	A	G	-0.0121	0.0312	6.98E-01	-0.033	2.6E-01	-0.098	7.4E-02	0.011	8.9E-01	7	8.50%

Chr	SNP	Position hg19	AD-Risk (IGAP Stage1 Meta-analysis)				Brain eQTL (Mayo Clinic eGWAS)						Regulome Score	HapMap CEU MAF
			Effect Allele	Non Effect Allele	OR (95% CI)	P-value	TREML1 CER BETA	TREML1 CER P	TREML1 TCX BETA	TREML1 TCX P	TREM2 TCX BETA	TREM2 TCX P		
6	rs9381040	41,154,650	T	C	0.94 (0.91 - 0.98)	5.97E-04	0.021	2.30E-01	0.086	*8.30E-03*	0.091	*4.80E-02*	NA	26.70%

Formatted: Left, Indent: First line: 0"
Formatted: Left
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Left, Line spacing: Double
Formatted: Tab stops: 0.8", Left
Formatted Table
Formatted: Highlight
Formatted: Highlight

6	rs2093395	41,155,026	C	G	0.94 (0.91 - 0.98)	6.40E-04	0.021	2.30E-01	0.086	*8.30E-03*	0.091	*4.80E-02*	6	27.90%
6	rs2038568	41,158,132	C	G	1.14 (1.05 - 1.23)	7.93E-04	0.018	7.80E-01	-0.08	3.90E-01	-0.186	1.60E-01	5	8.30%
6	rs12194214	41,028,574	C	A	1.16 (1.06 - 1.26)	8.36E-04	-0.081	9.40E-02	-0.104	2.70E-01	-0.129	3.20E-01	6	4.20%
6	rs9462675	41,153,238	A	G	1.15 (1.06 - 1.25)	9.54E-04	-0.015	7.50E-01	-0.114	1.60E-01	-0.207	6.50E-02	5	3.60%
6	rs6933067	41,133,522	C	T	1.15 (1.06 - 1.25)	1.07E-03	-0.013	7.60E-01	-0.098	2.10E-01	-0.134	2.10E-01	7	3.50%
6	**rs9357347**	**41,150,591**	**C**	**A**	**0.95 (0.91 - 0.98)**	**1.10E-03**	**0.013**	**4.60E-01**	**0.088**	***6.30E-03***	**0.09**	***4.60E-02***	**2b**	**28.10%**
6	rs9394767	41,159,905	G	A	0.95 (0.91 - 0.98)	1.14E-03	0.011	5.70E-01	0.096	*6.50E-03*	0.083	1.00E-01	5	28.80%
6	rs1542638	41,286,604	G	A	1.06 (1.02 - 1.09)	1.14E-03	-0.022	2.20E-01	-0.035	2.90E-01	-0.064	1.60E-01	4	28.30%
6	rs9471491	41,153,622	A	C	1.15 (1.05 - 1.26)	1.31E-03	-0.015	7.50E-01	-0.114	1.60E-01	-0.207	6.50E-02	7	3.50%
6	rs9471495	41,157,372	A	C	1.15 (1.05 - 1.25)	1.40E-03	0.014	8.30E-01	-0.099	2.90E-01	-0.235	7.20E-02	7	3.50%
6	rs9462677	41,158,856	A	T	1.15 (1.05 - 1.25)	1.41E-03	0.016	8.10E-01	-0.099	2.90E-01	-0.235	7.30E-02	7	4.30%
6	rs9394778	41,215,058	A	G	0.95 (0.92 - 0.98)	1.44E-03	0.015	3.30E-01	0.065	*2.70E-02*	0.099	*1.50E-02*	6	39.80%
6	rs9471494	41,157,344	G	C	1.15 (1.05 - 1.25)	1.46E-03	0.01	8.70E-01	-0.102	2.60E-01	-0.221	8.20E-02	6	4.50%
6	rs6912013	41,061,593	C	T	1.15 (1.05 - 1.25)	1.48E-03	-0.076	1.20E-01	-0.104	2.70E-01	-0.124	3.40E-01	5	2.70%
6	rs9296359	41,205,690	A	G	0.95 (0.92 - 0.98)	1.48E-03	0.017	2.80E-01	0.066	*2.40E-02*	0.116	*4.60E-03*	6	27.40%
6	rs3747742*	41,162,518	C	T	0.96 (0.92 - 0.99)	8.56E-03	0.018	2.90E-01	0.072	*2.30E-02*	0.064	1.50E-01	6	28.30%
6	rs6916710*	41,164,788	T	C	0.97 (0.94 - 1.01)	1.03E-01	0.013	4.30E-01	0.054	7.70E-02	0.072	9.20E-02	7	38.40%
6	rs7759295*	41,135,850	T	C	0.98 (0.93 - 1.03)	3.66E-01	-0.023	3.50E-01	0.094	*4.00E-02*	-0.008	9.00E-01	6	13.30%
6	rs6910730*	41,246,633	G	A	0.99 (0.94 - 1.04)	6.86E-01	-0.046	8.50E-02	-0.079	1.20E-01	-0.032	6.50E-01	4	8.40%
6	rs6922617*	41,336,101	A	G	0.99 (0.93 - 1.05)	6.98E-01	-0.033	2.60E-01	-0.098	7.40E-02	0.011	8.90E-01	7	8.50%

Shown are ~~V~~variants located within 100kb of a *TREM* gene ~~and~~ that had an AD-risk $p \leq 0.0015$ in the IGAP stage 1 meta-analysis (top 16 rows~~);~~), as well as ~~and~~ 5 common *TREM* locus variants with previous reports of AD-risk or endophenotype association (bottom 5 rows, SNP marked with an *). AD-risk association results are from the publicly available IGAP meta-analysis stage 1. Brain gene expression associations are from the Mayo Clinic eGWAS and based on cerebellar (CER) and temporal cortex (TCX) gene expression measurements with Illumina WG-DASL arrays with *TREML1* probe ILMN_1690783 and *TREM2* probe ILMN_1701248. Variants showing association with ~~nominally significant brain~~gene expression ~~associations~~ (uncorrected $p<0.05$) are underlined and in

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

italic font. The variant with the ~~strongest~~ most significant AD-risk association in the IGAP meta-analysis (**rs9381040**), and the variant with the ~~strongest~~ most significant gene expression association and best Regulome score (**rs9357347**) are in bold font. OR (95% CI): odds ratio and 95% confidence interval. Given that the eGWAS expression measures were on a log2 scale, fold-change for the Mayo eGWAS beta coefficients = 2^{beta}.

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Not Superscript/ Subscript, Highlight

Table 3. Meta-analysis of rs9357347 eQTL results from three independent datasets.

Dataset	Sample size	TREML1 beta	TREML1 p		TREM2 beta	TREM2 p
Mayo WG-DASL	380	0.088	$6.3x10^{-3}$		0.09	$4.6x10^{-2}$
Mayo Clinic RNAseq	132	-0.03	0.78		0.08	0.51
ROS/MAP RNAseq	494	0.03	0.14	-	0.43	0.11
Meta-analysis	1006	+-+	**$9.3x10^{-3}$**	-	+++	**$9.3x10^{-4}$**

Dataset	Sample size	MAF	TREML1 beta	TREML1 SE	TREML1 p	TREM2 beta	TREM2 SE	TREM2 p
Mayo WG-DASL	380	0.307	0.088	0.032	6.28E-03	0.090	0.045	4.61E-02
Mayo Clinic RNAseq	132	0.311	-0.030	0.108	7.82E-01	0.084	0.128	5.13E-01
ROS/MAP RNAseq	494	0.281	0.089	0.114	4.36E-01	0.124	0.060	3.77E-02
Meta-analysis	1006	-	+-+	-	**3.36E-02**	+++	-	**3.54E-03**

Meta-analysis of rs9357347 eQTL results from temporal cortex (Mayo WG-DASL and Mayo Clinic RNAseq) and dorsolateral prefrontal cortex samples (ROS/MAP). MAF = minor allele frequency. SE = standard error. Since in all three datasets the expression measures analyzed were on a log2 scale, fold-change for the beta coefficients = 2^{beta}. The meta-analysis was performed using METAL, with weighted average of z-scores from the individual study p-values, weighted according their sample size.

Formatted: Left, Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Left, Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Left, Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Left, Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Left, Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Left, Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Line spacing: Multiple 1.15 li

Formatted Table

Formatted: Line spacing: Multiple 1.15 li

Formatted: Highlight

Supplementary Material

A candidate regulatory variant at the *TREM* gene cluster associates with decreased Alzheimer's disease risk, and increased *TREML1* and *TREM2* brain gene expression

Minerva M. Carrasquillo[1], Mariet Allen[1], Jeremy D. Burgess[1], Xue Wang[2], Samantha L. Strickland[1], Shivani Aryal[1], Joanna Siuda[1], Michaela L. Kachadoorian[1], Christopher Medway[1,2], Curtis S. Younkin[1], Asha Nair[4], Chen Wang[4], Pritha Chanana[4], Daniel Serie[2], Thuy Nguyen[1], Sarah Lincoln[1], Kimberly G. Malphrus[1], Kevin Morgan[3], Todd E. Golde[5], Nathan D. Price[6], Charles C. White [7,8], Philip L De Jager[7,8,9], David A. Bennett[10], Yan W. Asmann[2], Julia E. Crook[2], Ronald C. Petersen[11], Neill R. Graff-Radford[12], Dennis W. Dickson[1], Steven G. Younkin[1], Nilüfer Ertekin-Taner[1,12*]

[1] Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL 32224, USA.

[2] Department of Health Sciences Research, Mayo Clinic, Florida, Jacksonville, FL 32224, USA.

[3] Human Genetics Group, University of Nottingham, Nottingham, UK NG7 2UH.

[4] Department of Health Sciences Research, Mayo Clinic Rochester, MN 55905, USA

[5] Department of Neuroscience, University of Florida, Gainesville, FL, 32610 USA.

[6] Institute for Systems Biology, Seattle, WA, 98109 USA.

[7] Ann Romney Center for Neurologic Diseases and Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA, 02115 USA.

[8] Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, 02142 USA.

[9] Harvard Medical School, Boston, MA, 02115 USA.

Formatted: Width: 8.5", Height: 11", Numbering: Continuous

Deleted: [3]

Deleted: [3]

Deleted: [3]

Deleted: , Xue Wang[4]

Deleted: [4]

Deleted: [2]

Deleted: [4]

Deleted: [4]

Moved (insertion) [1]

Deleted: [4]

Deleted: ¶ [2]

Deleted: [3]

Moved up [1]: [4] Department of Health Sciences Research, Mayo Clinic, Florida, Jacksonville, FL 32224, USA.¶

[10] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612 USA.

[11] Department of Neurology, Mayo Clinic Minnesota, Rochester, MN 55905 USA.

[12] Department of Neurology, Mayo Clinic Florida, Jacksonville, FL 32224, USA.

* To whom correspondence should be addressed at: Departments of Neurology and Neuroscience, Mayo Clinic Florida, Birdsall Building 327, 4500 San Pablo Rd, Jacksonville, FL 32224; E-mail: taner.nilufer@mayo.edu; Tel: 904-953-7103; Fax: 904-953-7370

Supplementary Methods

IGAP AD-risk meta-analysis

AD-risk association results shown in this study were obtained from the International Genomics of Alzheimer's Project (IGAP) stage 1 AD-risk GWAS meta-analysis [1]. The IGAP AD-risk meta-analysis is a large two-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 7,055,881 single nucleotide polymorphisms (SNPs) to meta-analyze four previously-published GWAS datasets consisting of 17,008 Alzheimer's disease cases and 37,154 controls (The European Alzheimer's disease Initiative – EADI, the Alzheimer Disease Genetics Consortium – ADGC, The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium – CHARGE, The Genetic and Environmental Risk in AD consortium – GERAD). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and 11,312 controls. Finally, a meta-analysis was performed combining results from stages 1 & 2.

Mayo Clinic WG-DASL eQTL dataset

Total RNA, utilized in the array-based Illumina WG-DASL discovery cohort (**Table 1**) was isolated from frozen brain tissue using the Ambion RNAqueous kit and assessed for RNA quality and quantity using the Agilent RNA 6000 Nano Chip and Agilent 2100 Bioanalyzer. Only samples with an RNA integrity number (RIN) score ≥5 were used. All subjects were from the Mayo Clinic Brain Bank and underwent neuropathological evaluation by DWD. All ADs had a Braak score of ≥4.0 and non-ADs a Braak score of ≤2.5. Many of the non-ADs had unrelated pathologies. All ADs had a definite diagnosis according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria [2].

Expression measures were generated as described previously [3]. Briefly, samples were randomized across plates and chips prior to array processing. Internal replicates were included for quality control purposes. PCR and array processing was conducted at the Mayo Clinic Medical Genome Facility Gene Expression Core in accordance with the manufactures' protocols. Raw probe data was exported from GenomeStudio (Illumina Inc) and the lumi package of Bioconductor [4, 5] was used for background subtraction, variance stabilizing transformation and quantile normalization.

Although there are seven RefSeq genes at the *TREM* locus (*TREM1*, *TREML1*, *TREM2*, *TREML2*, *TREML3P, TREML4* and *TREML5P*) (**Fig. 1**), *TREML3P* and *TREML5P* are non-coding pseudogenes for which there are no probes on the WG-DASL array. Only transcripts whose expression was detected above background in ≥ 50% of the samples tested (**Table S1)** were evaluated for their associations with *TREM* locus variants (**Table 2**). The location of the WG-DASL probes relative to the transcripts is shown in **Fig. 2**. The probes were determined to

be complementary to sequences lacking known polymorphisms based on the human genome assembly from March 2006 (NCBI36/hg18).

Mayo Clinic RNAseq dataset

Temporal cortex RNAseq data from 84 LOAD and 48 non-AD brains from the Mayo Clinic Brain Bank that were not part of the Mayo Clinic WG-DASL eGWAS cohort but whose neuropathological diagnosis followed the same criteria (**Table 1**), were employed for replication of the associations that were detected with the WG-DASL gene expression measurements. Total RNA, utilized for the RNAseq replication cohort was extracted using Trizol® reagent and cleaned using Qiagen RNeasy columns with DNase treatment. Samples were randomized prior to the transfer of 40 (TCX) or 50 (CER) ng/ul of RNA to the Mayo Clinic Medical Genome Facility Gene Expression and Sequencing Cores for library preparation and sequencing. The TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA) was used for library preparation. The library concentration, size distribution and RIN were measured on an Agilent Technologies 2100 Bioanalyzer. Only samples with a RIN score >5 were used. Sequencing was performed on the Illumina HiSeq2000 using 101 base-pair (bp), paired end sequencing, with triplicate multiplexing of barcoded samples (3 samples per flowcell lane). Base-calling was performed using Illumina's RTA 1.18.61 or RTA 1.17.21.3. FASTQ sequence reads were aligned to the human reference genome using TopHat 2.0.12 [6] and Bowtie 1.1.0 [7], and Subread 1.4.4 was used for gene counting [8]. FastQC was used for quality control (QC) of raw sequence reads, and RSeQC was used for QC of mapped reads. Raw read counts were normalized using Conditional Quantile Normalization (CQN) via the Bioconductor package; accounting for sequencing depth, gene length, and GC content. RNAseq data for this cohort is available at the Sage Synapse, AMP AD Knowledge Portal

Deleted: Normalized transcript brain expression levels were tested for associations with *TREM* locus genotypes (Mayo Clinic eGWAS) [3] via multivariable linear regression analyses implemented in PLINK [6], using an additive model and adjusting for *APOE* ε4 dosage, age at death, diagnosis, sex, PCR plate, RNA Integrity Number (RIN), adjusted RIN squared $(RIN-RINmean)^2$ as covariates. ¶

Formatted: Highlight

(https://www.synapse.org/#!Synapse:syn2580853/wiki/66722), under synapase ID syn3163039 (Mayo RNAseq).

Genotypes for rs9357347 were obtained using a TaqMan® SNP genotyping assay, C___2814743_10. Genotyping was performed at the Mayo Clinic in Jacksonville using an ABI PRISM 7900HT Sequence Detection System with 384-Well Block Module from (Applied Biosystems, Foster City, California). The genotype data was analyzed using the SDS software version 2.2.3 (Applied Biosystems).

Religious Orders Study and the Rush Memory and Aging Project (ROS/MAP) RNAseq dataset

RNA was isolated from frozen dorsolateral prefrontal cortex tissue of ROS/MAP subjects [18, 19] using the miRNeasy Mini Kit and RNase-Free DNase Set (Qiagen, Germantown, MD). RNA concentration and quality were determined using a Nanodrop (Thermo Fisher Scientific, Wilmington, DE) and Bioanalyzer (Agilent Technologies, Santa Clara, CA), respectively. Only samples with a RIN score >5 were used for library construction, which was assembled using 50ng/ul of RNA for the strand-specific dUTP method. The library was read using Illumina HiSeq with 101 base pair paired-end reads and a goal coverage of >85 million paired-end reads. FPKM (Fragments per Kilobase of Exon Per Million Fragments Mapped) were quantile normalized with Combat correcting for batch. RNAseq data for this cohort is available at the Sage Synapse, AMP AD Knowledge Portal (https://www.synapse.org/#!Synapse:syn2580853/wiki/66722), under synapase IDs syn3388564 (ROS/MAP RNAseq).

Formatted: Highlight

Formatted: Highlight

Genotypes for rs9357437 were obtained from three subsets of subjects. Genotypes for the first two subsets were generated in 2009 on the Affymetrix Genechip 6.0 platform (Affymetrix, Inc, Santa Clara, CA, USA) at the Broad Institute's Center for Genotyping or the Translational Genomics Research Institute. The third subset was genotyped in 2012 on the Illumina HumanOmniExpress platform (Illumina, Inc, San Diego, CA, USA) at the Children's Hospital of Philadelphia. All three data sets underwent the same quality control (QC) analysis (genotype call rate > 95%, Hardy Weinberg Equilibrium > 0.001). Using Beagle software (version: 3.3.2), dosage data was imputed for all genotyped samples who passed QC using the 1000 Genomes Project (2011, Phase 1b) as a reference

Deleted: . ¶
¶
eQTL analysis of rs9357347 and *TREML1/TREM2* RNAseq gene expression ¶
For the Mayo RNAseq dataset, rs9357347 genotypes used in the gene expression association analysis were obtained using a TaqMan® SNP genotyping assay, C___2814743_10. Genotyping was performed at the Mayo Clinic in Jacksonville using an ABI PRISM 7900HT Sequence Detection System with 384-Well Block Module from (Applied Biosystems, Foster City, California). The genotype data was analyzed using the SDS software version 2.2.3 (Applied Biosystems). For the ROS/MAP RNAseq dataset, genotyping was done in three subsets. Genotypes for the first two subsets were generated in 2009 on the Affymetrix Genechip 6.0 platform (Affymetrix, Inc, Santa Clara, CA, USA) at the Broad Institute's Center for Genotyping or the Translational Genomics Research Institute. The third subset was genotyped in 2012 on the Illumina HumanOmniExpress platform (Illumina, Inc, San Diego, CA, USA) at the Children's Hospital of Philadelphia. All three data sets underwent the same quality control (QC) analysis (genotype call rate > 95%, Hardy Weinberg Equilibrium > 0.001). Using Beagle software (version: 3.3.2), dosage data was imputed for all genotyped samples who passed QC using the 1000 Genomes Project (2011, Phase 1b) as a reference. For both the Mayo RNAseq and the ROS/MAP RNAseq dataset, normalized gene counts were used as the gene expression phenotype in linear regression analysis using an additive model and adjusting for age at death, sex, diagnosis, RIN, (RIN-RINmean)^2 and flowcell as covariates in PLINK.¶
¶
Meta-analysis of rs9357347 eQTL results¶
Meta-analyses were performed on eQTL results from three independent cohorts: Mayo WG-DASL, Mayo RNAseq and ROS/MAP RNAseq, for which there was no sample overlap. For these analyses METAL [10] was implemented using weighted average of z-scores from the individual study p-values, weighted according their sample size.¶
¶
¶

Determination of linkage disequilibrium

Linkage disequilibrium of variants at the *TREM* locus (**Fig. 3**) was evaluated in the Mayo Clinic AD-risk GWAS HapMap2 imputed dataset (815 AD, 1218 controls) using Haploview 4.0 [9]. *TREM2* AD-risk missense variants rs142232675 (p.D87N) and rs75932628 (p.R47H) were not present in the Mayo GWAS HapMap2 imputed dataset, but were directly genotyped in the Mayo Clinic samples with TaqMan® assays, and were included to show their LD with variants that associate with AD-risk in the IGAP meta-analysis. Sixteen variants located within 100kb of a *TREM* gene and that had an AD-risk p≤0.0015 in the IGAP stage 1 meta-analysis are included, in addition to two rare missense *TREM2* coding variants [rs142232675 (p.D87N), and rs75932628 (p.R47H)] and 5 common *TREM* locus SNPs with prior reports of AD-risk (rs3747742) [10] or endophenotype (rs7759295, rs6910730, rs6922617, rs6916710) [11, 12] association, even though they did not meet the IGAP AD-risk association cutoff.

Supplementary Results

Assessment of potential collider conditioning bias

Since the primary goal of this study was to estimate the effect of genetic variants on gene expression, rather than their effect on disease status, we combined ADs and non-ADs in the linear regression analysis and included diagnosis as a covariate. The diagnosis covariate was coded as the presence or absence of AD. However, adjusting for diagnosis status could potentially introduce a collider conditioning bias if both the genotype and the expression levels are associated with disease status [13]. Therefore, we have also analyzed the combined set of AD+nonAD, without adjustment for disease status, in order to determine if the effect of genotype on expression disappears or remains in the latter analysis. **Table S2**, shows results for the two types of analyses that were performed in each of the three datasets (Mayo WG-DASL, Mayo RNAseq and ROS/MAP RNAseq): (1) AD and nonAD combined while adjusting for diagnosis, (2) AD and nonAD combined not adjusting for diagnosis. Overall, the results from analyses 1 and 2 are very similar, suggesting that there is no real impact of a collider effect.

Effect of RIN on percent detection and rs9357347 eQTL association

Fig. S6 shows *TREML1* and *TREM2* detection percentage stratified by RIN, and demonstrates that neither *TREML1* nor *TREM2* detection percentage is affected by RIN. **Table S3** shows results of the rs9357347 eQTL associations in the Mayo WG-DASL dataset when stratifying by samples above and below the median RIN of 6.5 (**Table S3**). These results indicate that RIN does not significantly impact the magnitude of the rs9357347 eQTL associations, as the

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

estimates of the beta coefficients overlap with those observed in the analysis not stratified by RIN (**Table S2**). Although the significance of the association is lessened in the stratified analysis, this is likely due to the smaller sample size of the stratified groups compared to the sample size of the combined analysis.

Association of rs9357347 with Braak stage

Given the association of rs9357347 with AD-risk, we tested the hypothesis that this variant could also show an association with Braak stage, as the latter is an important criterion for the neurophathological diagnosis of AD [14]. Implementing an ANOVA model in R that included age-at-death, sex and *APOE* ε4 dose (0, 1 or 2 alleles), in the two larger datasets (Mayo WG-DASL and ROS/MAP RNAseq), we determined that rs9357347 does not significantly contribute to the variance in Braak stage in either of these two cohorts (p=0.91 and p=0.27, respectively). In addition, we implemented linear regression analysis in R, again using the two larger datasets to estimate the effect of each copy of the rs9357347 minor allele on Braak stage, including age-at-death, sex and *APOE* ε4 dose in the model. As shown in **Table S4**, we did not detect a significant association of rs9357347 with Braak stage in either cohort.

Association of rs9357347 with cognition

We also evaluated the association of rs9357347 with measures of global cognitive decline and global cognition at the last evaluation before death in the ROS/MAP cohort. In this dataset, global cognition is a variable for overall cognitive function measured by the raw scores from 19 different tests that are converted to z scores and averaged. Global cognitive decline is a

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

longitudinal cognitive phenotype based on repeated measures of global cognition, as previously described [15, 16]. The analysis was performed using linear regression analysis implemented in R, under an additive model for rs9357347, and adjusting for age-at-death, sex and *APOE* ε4 dose. Neither global cognitive decline nor global cognition at last evaluation shows an association with rs9357347 in this cohort (**Table S5**).

Association of age with *TREML1* and *TREM2* expression

As the WG-DASL cohort was overall younger than the two RNAseq cohorts, we assessed the association of the age covariate on *TREML1* and *TREM2* gene expression levels in the linear regression model described in the Material & Methods section 2.4. Age was not significantly associated with either *TREML1* or *TREM2* expression in the Mayo WG-DASL cohort (p>0.05). On the other hand, both in the Mayo RNAseq and ROS/MAP RNAseq cohorts, *TREML1* and *TREM2* expression levels appeared to be slightly increased with age, albeit the magnitude of the effect sizes were modest, with beta coefficients equivalent to approximately a 1.01 and 1.03-fold change in expression levels (Mayo RNAseq: *TREML1* p=0.085, beta=0.01; *TREM2* p=0.026, beta=0.02. ROS/MAP RNAseq: *TREML1* p=2.0×10^{-3}, beta=0.04; *TREM2* p=4.4×10^{-5}, beta=0.03). Since *TREML1* and *TREM2* gene expression levels appear be increased with age, it is possible that this might have led to a decrease in power to detect an association of rs9357347-C with increased levels of these genes in the two older cohorts.

Association of diagnosis with *TREML1* and *TREM2* expression

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

To assess if diagnosis is associated with *TREML1* and/or *TREM2* gene expression levels, linear regression analyses were performed in R in each of the three datasets, adjusting for all other covariates included in the eQTL analyses described in the Materials and Methods section 2.4, as well as rs9357347 minor allele dose. The box plots in **Fig. S5** show the direction of the change in expression between AD and nonAD subjects, and indicate the significance of the association for each test. We observe a consistent trend of higher *TREML1* and *TREM2* expression in AD versus nonADs, although some of these associations do not reach statistical significance. The trend toward higher *TREML1* and *TREM2* expression in AD subjects could be a reflection of microglial activation and/or proliferation known to occur in AD brains.

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Font: Bold, Highlight
Formatted: Highlight
Formatted: Font: Bold, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

Acknowledgements for IGAP:

We thank IGAP for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i–Select chips were funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100,

Deleted: was

the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.

Deleted: ¶

Deleted: ¶
¶
¶
¶
¶

References

[1] J.C. Lambert, C.A. Ibrahim-Verbaas, D. Harold, A.C. Naj, R. Sims, C. Bellenguez, et al., Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics 2013;45:1452-1458.

[2] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price,E.M. Stadlan, Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;34:939-944.

[3] F. Zou, H.S. Chai, C.S. Younkin, M. Allen, J. Crook, V.S. Pankratz, et al., Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS genetics 2012;8:e1002707.

[4] P. Du, W.A. Kibbe,S.M. Lin, lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547-1548.

[5] S.M. Lin, P. Du, W. Huber,W.A. Kibbe, Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic acids research 2008;36:e11.

[6] C. Trapnell, L. Pachter,S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009;25:1105-1111.

[7] B. Langmead, C. Trapnell, M. Pop,S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 2009;10:R25.

[8] S. Anders, P.T. Pyl,W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-169.

[9] J.C. Barrett, B. Fry, J. Maller,M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-265.

[10] B.A. Benitez, B. Cooper, P. Pastor, S.C. Jin, E. Lorenzo, S. Cervantes, et al., TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiology of aging 2013;34:1711 e1715-1717.

[11] J.M. Replogle, G. Chan, C.C. White, T. Raj, P.A. Winn, D.A. Evans, et al., A TREM1 variant alters the accumulation of Alzheimer-related amyloid pathology. Annals of neurology 2015;77:469-477.

[12] C. Cruchaga, J.S. Kauwe, O. Harari, S.C. Jin, Y. Cai, C.M. Karch, et al., GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. Neuron 2013;78:256-268.

[13] S.R. Cole, R.W. Platt, E.F. Schisterman, H. Chu, D. Westreich, D. Richardson, et al., Illustrating bias due to conditioning on a collider. International journal of epidemiology 2010;39:417-420.

[14] H. Braak,E. Braak, Neuropathological stageing of Alzheimer-related changes. Acta neuropathologica 1991;82:239-259.

[15] P.A. Boyle, A.S. Buchman, R.S. Wilson, S.E. Leurgans,D.A. Bennett, Association of muscle strength with the risk of Alzheimer disease and the rate of cognitive decline in community-dwelling older persons. Archives of neurology 2009;66:1339-1344.

[16] R.S. Wilson, T.L. Beck, J.L. Bienias,D.A. Bennett, Terminal cognitive decline: accelerated loss of cognition in the last years of life. Psychosomatic medicine 2007;69:131-137.

Supplementary Table

Table S1. Percent detection of *TREM* locus transcripts.

			Mayo WG-DASL: Cerebellum[a]			Mayo WG-DASL: Temporal Cortex[b]			Mayo Clinic RNAseq: Temporal Cortex[c]	ROS/MAP RNAseq: DFPC[d]
Symbol	Ensembl Gene ID	WG-DASL Probe ID	AD + non-AD	AD	nonAD	AD + non-AD	AD	Non-AD	AD + non-AD	AD + non-AD
TREM1	ENSG00000124731	ILMN_1688231	0.00	0.00	0.00	0.25	0.00	0.51	18.18	66.99
TREML1	ENSG00000161911	ILMN_1690783	100.00	100.00	100.00	100.00	100.00	100.00	100.00	97.84
TREM2	ENSG00000095970	ILMN_1701248	40.91	43.59	37.29	98.25	99.50	96.95	100.00	100.00
TREML2	ENSG00000112195	ILMN_1740864	17.38	17.44	17.51	6.27	10.40	2.03	8.33	24.75
TREML4	ENSG00000188056	ILMN_2205322	6.15	4.62	7.91	2.26	2.97	1.52	2.27	15.13

The percentage of samples with detectable expression of *TREM* family transcripts in each of the expression datasets studied. For the WG-DASL dataset (a,b) the corresponding WG-DASL probe is indicated. Only *TREML1* and *TREM2* expression are detectable above background in at least 50% of the Mayo WG-DASL samples tested (a,b), in at least one tissue; c: A detection threshold >-1, for cqn normalized expression levels was used to determine percent detection; d: percent detection was calculated as the proportion of subjects who express > 0 FPKM, DFPC = dorsolateral prefrontal cortex. The WG-DASL array lacked probes for the two *TREM* pseudogenes:

Deleted: ¶
¶

Formatted: Highlight

Formatted: Left

Deleted: with Illumina WG-DASL probes

Deleted: ¶
¶
¶
... [1]

Formatted: Line spacing: Multiple 1.15 li

Formatted: Highlight

Formatted: Font: Italic, Highlight

Formatted: Highlight

Formatted: Font: Italic, Highlight

Formatted: Highlight

Formatted: Font: Italic, Highlight

Formatted: Highlight

Deleted: ¶
The percentage of samples with detectable expression of *TREM* family transcripts and their corresponding WG-DASL probes. Only *TREML1* and *TREM2* expression is detectable above background in at least 50% of the samples tested.¶
Page Break
¶

therefore they were not measured in the Mayo WG-DASL cohort. Expression levels were not available for the *TREM* pseudogenes in the ROS/MAP dataset. Based on the Mayo RNAseq dataset, the percent detection of *TREML3P* and *TREML5P* in temporal cortex are 15% and 5% respectively.

Table S2. Analyses to assess the potential of introducing collider conditioning bias in the linear regression model due to adjustment for diagnosis.

			TREML1			*TREM2*		
Dataset	**Group**	**N**	**beta**	**SE**	**p-value**	**beta**	**SE**	**p-value**
Mayo WG-DASL: Temporal Cortex	All (W/Dx)[a]	380	0.088	0.032	6.28E-03	0.090	0.045	4.61E-02
	All (Wo/Dx)[b]	380	0.083	0.033	1.32E-02	0.088	0.045	5.33E-02
Mayo Clinic RNAseq: Temporal Cortex	All (W/Dx)[a]	132	-0.030	0.108	7.82E-01	0.084	0.128	5.13E-01
	All (Wo/Dx)[b]	132	-0.023	0.111	8.40E-01	0.102	0.145	4.86E-01
ROS/MAP RNAseq: DFPC	All (W/Dx)[a]	494	0.089	0.114	4.36E-01	0.124	0.060	3.77E-02
	All (Wo/Dx)[b]	494	0.089	0.114	4.35E-01	0.125	0.060	3.81E-02

For each of the three datasets, linear regression analysis was run in a: AD and non-AD combined, with diagnosis included as a covariate; b: Analysis of AD and non-AD combined, without adjustment for diagnosis. N = sample size. SE= standard error. DFPC = dorsolateral prefrontal cortex. Given that all expression measures were on a log2 scale, fold-change for the beta coefficients = 2^{beta}

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Superscript, Highlight
Formatted: Font: Not Bold, Highlight

Table S3. Association of the *TREM* locus candidate regulatory variant, rs9357347, with *TREML1* and *TREM2* gene expression stratified by RIN.

Gene Symbol	RIN group	N	beta	SE	p-value
TREML1	RIN < 6.5	188	0.087	0.042	0.043
	RIN > 6.5	192	0.084	0.046	0.069
TREM2	RIN < 6.5	188	0.110	0.063	0.084
	RIN > 6.5	192	0.075	0.065	0.250

Data shown for Mayo WG-DASL temporal cortex (AD+Non-AD) dataset. Samples were stratified into two groups representing those with a RIN below the median RIN of 6.5 and those with a RIN above 6.5. N = sample size. SE= standard error. Given that all expression measures were on a log2 scale, fold-change for the beta coefficients = 2^{beta}

Formatted: Font: Bold, Highlight
Formatted: Font: Bold, Italic, Highlight
Formatted: Font: Bold, Highlight
Formatted: Font: Bold, Italic, Highlight
Formatted: Font: Bold, Highlight
Formatted: Font: Bold, Italic, Highlight
Formatted: Font: Bold, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Font: Not Bold

Table S4. Association of rs9357347 with Braak stage.

Dataset	N	beta	SE	p-value
Mayo WG-DASL: Temporal Cortex	399	-0.139	0.160	0.387
ROS/MAP RNAseq: DFPC	492	0.053	0.081	0.515

The two largest cohorts were evaluated: Mayo WG-DASL and ROS/MAP RNAseq. The variant was tested for association with Braak stage using linear regression under an additive model and including age-at-death, sex and *APOE* ε4 dose as covariates. In this model, the beta coefficient is interpreted as the change in Braak score associated with each copy of the minor allele. N = sample size. SE= standard error. DFPC = dorsolateral prefrontal cortex.

Formatted: Font: 12 pt, Highlight
Formatted: Space After: 0 pt
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Font: 12 pt, Not Bold, Highlight
Formatted: Font: 12 pt, Not Bold, Italic, Highlight
Formatted: Font: 12 pt, Not Bold, Highlight
Formatted: Highlight
Formatted: Font: 12 pt, Not Bold, Highlight
Formatted: Font: (Default) Times New Roman, 12 pt, Highlight
Formatted: Font: 12 pt, Not Bold, Highlight
Formatted: Font: 12 pt, Not Bold

Table S5. Association of rs9357347 with cognition.

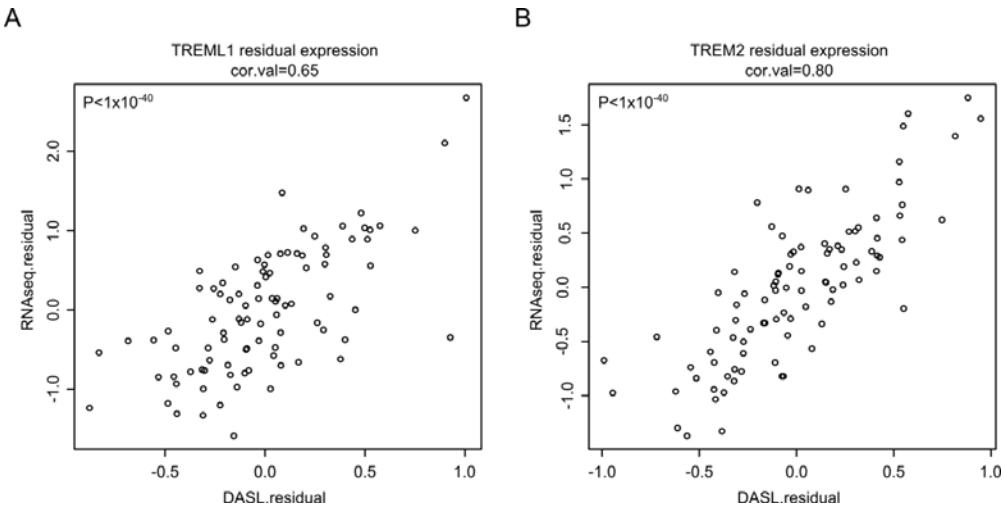
Phenotype	N	beta	SE	p-value
Global cognitive decline	470	-0.007	0.007	0.320
Global cognition at last visit	493	-0.058	0.071	0.418

Measures of cognition that were available in the ROS/MAP cohort were tested for association with rs9357347 using linear regression under an additive model, including age-at-death, sex and *APOE* ε4 dose as covariates. N = sample size. SE = standard error. Z scores of the cognitive scores were analyzed, thus these beta coefficients can be interpreted as changes in z-score associated with each copy of the minor allele.

Formatted: Font: 12 pt, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: 12 pt, Not Bold

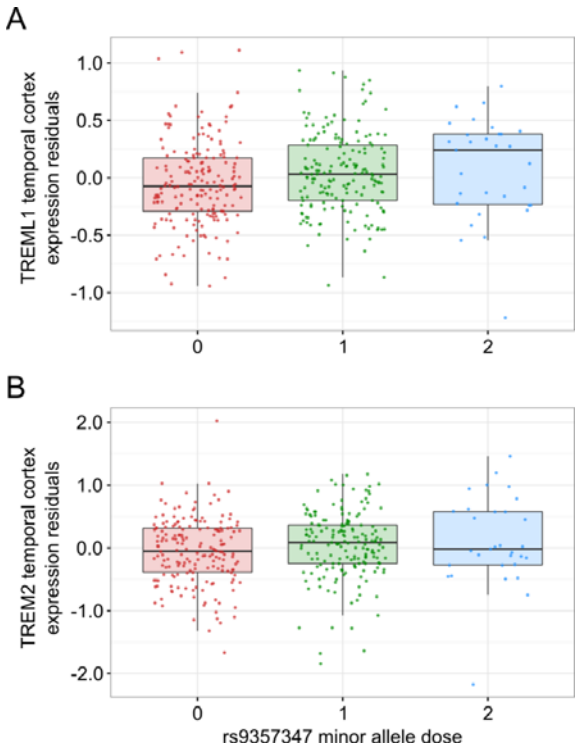
Supplementary Figures

Fig. S1. Spearman correlation plots for *TREML1* and *TREM2* brain expression levels measured by WG-DASL vs. RNAseq approaches.



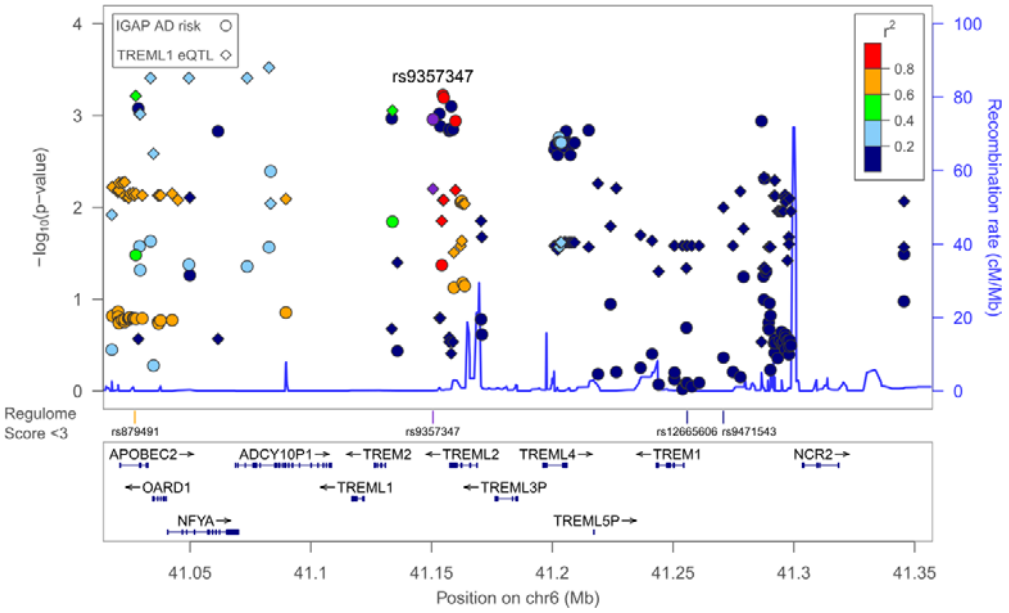
(A) *TREML1* and (B) *TREM2* temporal cortex gene expression residuals (adjusted for covariates) are plotted for RNAseq vs. WG-DASL values. The Spearman correlation coefficient is shown above the scatter plot and the correlation p-value is shown inside the plot. The RNAseq and DASL residuals show a highly significant positive correlation.

Fig. S2. Box plots of rs9357347 genotype associations with *TREML1* and *TREM2* brain gene expression levels.



Gene expression residuals obtained in R adjusted for covariates [*APOE* ε4 dosage, age at death, diagnosis, sex, PCR plate, RIN, $(RIN-RINmean)^2$] were plotted for each rs9357347 genotype. Each circle represents an individual gene expression residual; the horizontal line within the box is the median; the box represents the interquartile range (IQR); the whiskers represent the range of the data points within 1.5 × IQR below the 1st quartile and 1.5 × IQR above the 3rd quartile (anything outside of this range is called an outlier). The x-axis indicates the number of minor alleles. The minor (C) allele of rs9357347 is associated with increased brain expression of both *TREML1* and *TREM2*.

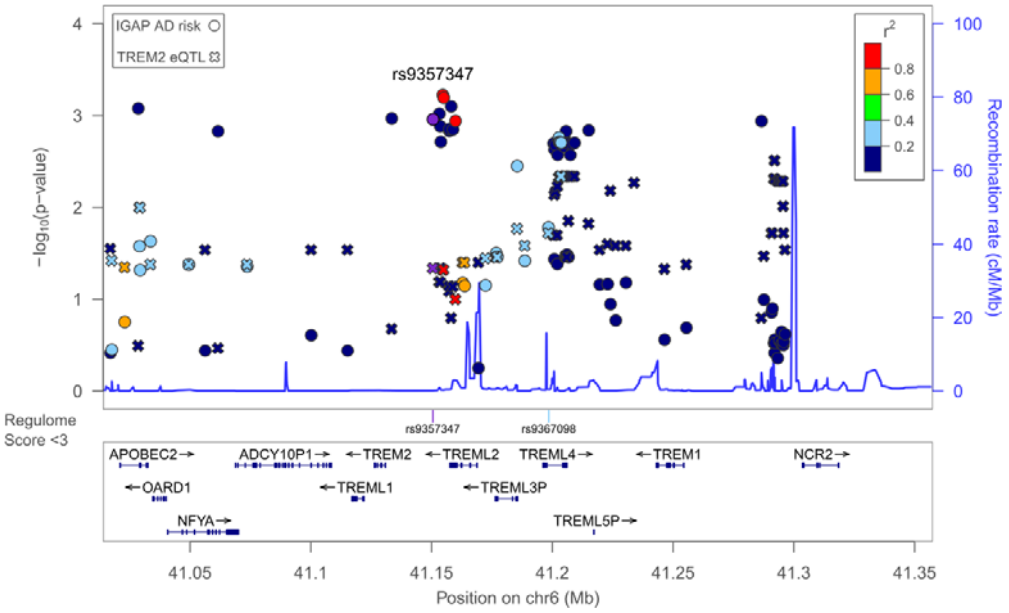
Fig. S3. Regional association plot showing AD-risk and *TREML1* eQTL p-values at the *TREM* locus.



On this plot each variant is depicted as both, a circle denoting its IGAP stage 1 meta-analysis p-value, and a diamond denoting its *TREML1* eQTL p-value. All variants at the *TREM* locus that either achieved a p-value ≤ 0.0015 in the IGAP stage1 meta-analysis or that had a *TREML1* p-value < 0.05 in our eQTL analysis of temporal cortex gene expression measured by WG-DASL mircroarrays are shown, with p-values indicated by the scale on the left y-axis as $-\log_{10}$(p-value). The putative regulatory variant, rs9357347, is represented by the purple circle/diamond. The colors of all other circles and diamonds correspond to the colors on the r^2 scale shown at the top right corner of the plot, and denote the LD of each variant with rs9357347. The values on the y-axis on the right side of the plot correspond to the recombination rates across this region as shown by the blue line. Variants that have Regulome scores <3 are shown directly below the

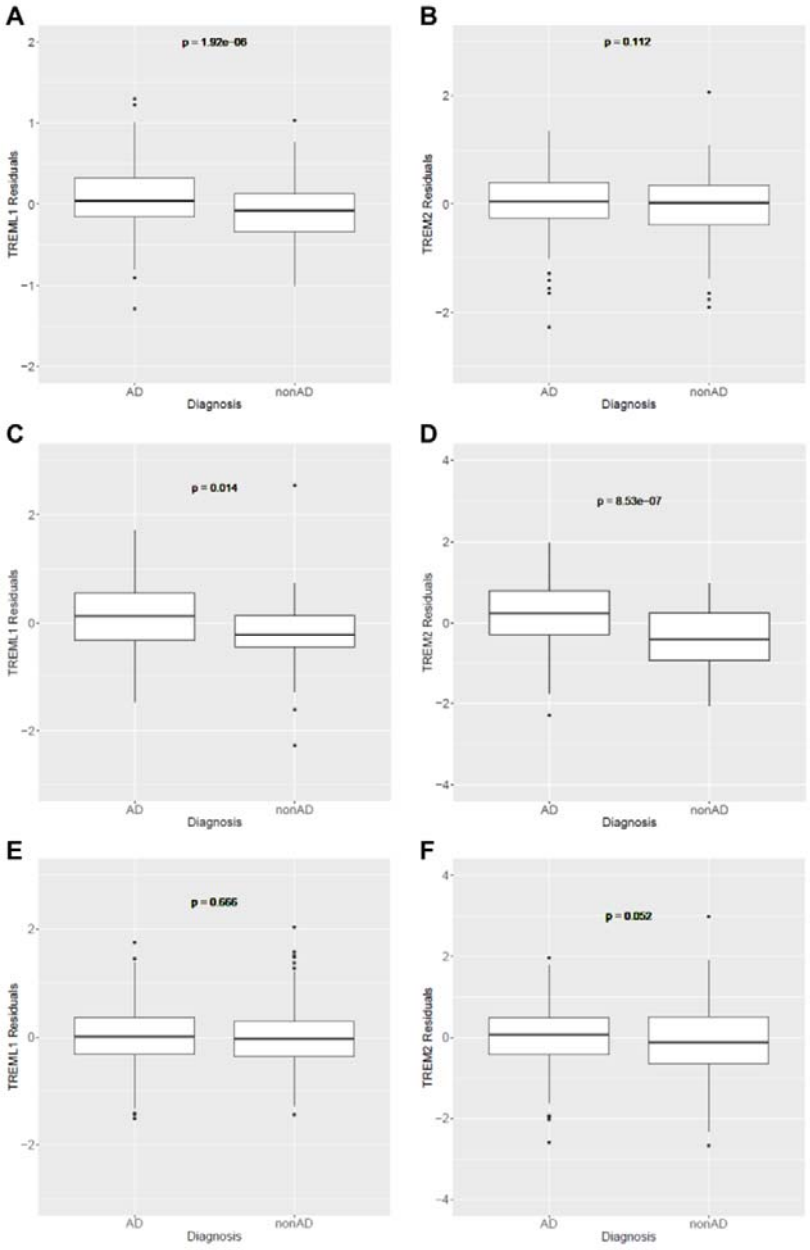
plot. Gene locations across the targeted genomic region (*TREM* gene +/-100 kb: chr6:41016999-41354457) are shown below the plot relative to the variant positions according to the February 2009 human genome assembly (GRCh37hg19). The regional association plot was generated using LocusZoom (http://locuszoom.sph.umich.edu/locuszoom/).

Fig. S4. Regional association plot showing AD-risk and *TREM2* eQTL p-values at the *TREM* locus.



On this plot each variant is depicted as both, a circle that denotes its IGAP stage 1 meta-analysis p-value, and an "X" denoting its *TREM2* eQTL p-value. All variants at the *TREM* locus that either achieved a p-value ≤ 0.0015 in the IGAP stage1 meta-analysis or that had a *TREM2* p-value < 0.05 in our eQTL analysis of temporal cortex gene expression measured by WG-DASL mircroarrays are shown, with p-values indicated by the scale on the left y-axis as $-\log_{10}$(p-value). The putative regulatory variant, rs9357347, is represented by the purple circle/X. The color of all other circles and Xs correspond to the colors on the r^2 scale shown at the top right corner of the plot, and denote the LD of each variant with rs9357347. All other symbols are described in **Fig. S3**.

Fig. S5. Box plots of gene expression residuals for *TREML1* and *TREM2* in AD and nonAD subjects, for each of the three cohorts investigated.



Formatted: Highlight
Formatted: Highlight
Formatted: Font: Bold, Highlight
Formatted: Font: Bold, Italic, Highlight
Formatted: Font: Bold, Highlight
Formatted: Font: Bold, Italic, Highlight
Formatted: Font: Bold, Highlight
Formatted: Highlight

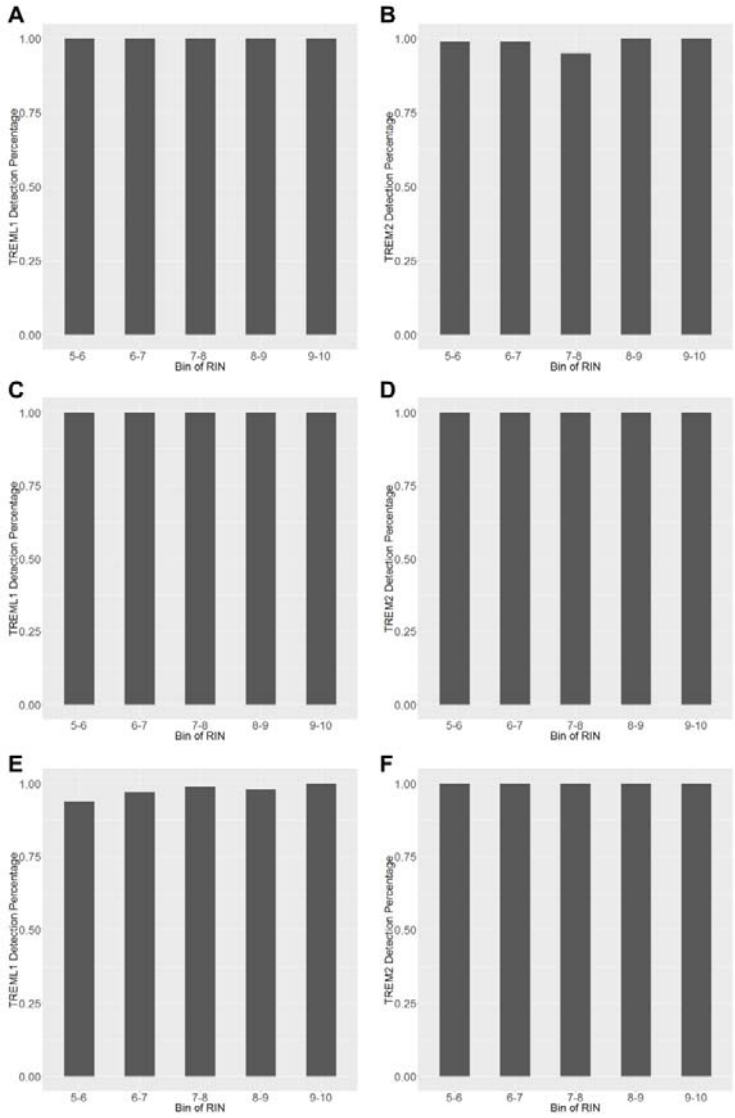
A and B: Expression measure residuals for *TREML1* (A) and *TREM2* (B) in the Mayo WG-DASL dataset, adjusted for rs9357347 minor allele dose, age-at-death, sex, *APOE* ε4 dose, RIN, $(RIN-RINmean)^2$ and PCR plate. *TREML1* (C) and *TREM2* (D) in the Mayo Clinic RNAseq dataset, adjusted for rs9357347 minor allele dose, age-at-death, sex, *APOE* ε4 dose, RIN, $(RIN-RINmean)^2$ and flowcell. *TREML1* (E) and *TREM2* (F) in the ROS/MAP RNAseq dataset, adjusted for rs9357347 minor allele dose, age-at-death, sex, *APOE* ε4 dose, RIN, $(RIN-RINmean)^2$.

Deleted: ¶

Formatted: Font: 14 pt, Highlight

Formatted: Space Before: 12 pt, Line spacing: Double

Fig. S6. Bar charts of percentage of subjects with detectable gene expression for *TREML1* and *TREM2* across groups of subjects defined by RIN value.



Deleted: ¶
¶
¶
¶

Formatted: Font: Bold, Highlight

Formatted: Line spacing: Double

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Bold, Highlight

Formatted: Font: Bold, Italic, Highlight

Formatted: Font: Bold, Highlight

Formatted: Font: Bold, Italic, Highlight

Formatted: Font: Bold, Highlight

Formatted: Highlight

Formatted: Highlight

Subjects were binned according to RIN value and the proportion of subjects in each bin that met the detection threshold was calculated. A and B: Expression detection percentage for each RIN bin for *TREML1* (A) and *TREM2* (B) in the Mayo WG-DASL dataset. C and D: Expression detection percentage for each RIN bin for *TREML1* (A) and *TREM2* (B) in the Mayo Clinic RNAseq dataset. E and F: Expression detection percentage for each RIN bin for *TREML1* (A) and *TREM2* (B) in the ROS/MAP RNASseq dataset.

Formatted: Font: Not Bold, Highlight
Formatted: Line spacing: Double
Formatted: Font: Not Bold, Italic, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Italic, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Italic, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Italic, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Italic, Highlight
Formatted: Font: Not Bold, Highlight
Formatted: Font: Not Bold, Highlight

		Percent Detection in Cerebellum			Percent Detection in Temporal Cortex		
Symbol	**Probe ID**	**AD + non-AD**	**AD**	**nonAD**	**AD + non-AD**	**AD**	**Non-AD**
TREM1	ILMN_1688231	0	0	0	0.25	0	0.51
TREML1	ILMN_1690783	100	100	100	100	100	100
TREM2	ILMN_1701248	40.91	43.59	37.29	98.25	99.5	96.95
TREML2	ILMN_1740864	17.38	17.44	17.51	6.27	10.4	2.03
TREML4	ILMN_2205322	6.15	4.62	7.91	2.26	2.97	1.52