

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. Carrasquillo¹, Mariet Allen¹, Jeremy D. Burgess¹, Xue Wang², Samantha L. Strickland¹, Shivani Aryal¹, Joanna Siuda¹, Michaela L. Kachadorian¹, Christopher Medway^{1,2}, Curtis S. Younkin¹, Asha Nair⁴, Chen Wang⁴, Pritha Chanana⁴, Daniel Serie², Thuy Nguyen¹, Sarah Lincoln¹, Kimberly G. Malphrus¹, Kevin Morgan³, Todd E. Golde⁵, Nathan D. Price⁶, Charles C. White^{7,8}, Philip L De Jager^{7,8,9}, David A. Bennett¹⁰, Yan W. Asmann², Julia E. Crook², Ronald C. Petersen¹¹, Neill R. Graff-Radford¹², Dennis W. Dickson¹, Steven G. Younkin¹, 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.

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

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

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

27

28 * To whom correspondence should be addressed at: Departments of Neurology and

29 Neuroscience, Mayo Clinic Florida, Birdsall Building 327, 4500 San Pablo Rd, Jacksonville, FL

30 32224; E-mail: taner.nilufer@mayo.edu; Tel: 904-953-7103; Fax: 904-953-7370

31

32

33 **Supplementary Methods**

34 **IGAP AD-risk meta-analysis**

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

47 **Mayo Clinic WG-DASL eQTL dataset**

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

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

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

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

72 **Mayo Clinic RNAseq dataset**

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

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

95 Genotypes for rs9357347 were obtained using a TaqMan® SNP genotyping assay,
96 C__2814743_10. Genotyping was performed at the Mayo Clinic in Jacksonville using an ABI
97 PRISM 7900HT Sequence Detection System with 384-Well Block Module from (Applied
98 Biosystems, Foster City, California). The genotype data was analyzed using the SDS software
99 version 2.2.3 (Applied Biosystems).

100

101 **Religious Orders Study and the Rush Memory and Aging Project (ROS/MAP)**

102 **RNAseq dataset**

103 RNA was isolated from frozen dorsolateral prefrontal cortex tissue of ROS/MAP subjects
104 [18, 19] using the miRNeasy Mini Kit and RNase-Free DNase Set (Qiagen, Germantown, MD).
105 RNA concentration and quality were determined using a Nanodrop (Thermo Fisher Scientific,
106 Wilmington, DE) and Bioanalyzer (Agilent Technologies, Santa Clara, CA), respectively. Only
107 samples with a RIN score >5 were used for library construction, which was assembled using
108 50ng/ul of RNA for the strand-specific dUTP method. The library was read using Illumina HiSeq
109 with 101 base pair paired-end reads and a goal coverage of >85 million paired-end reads. FPKM
110 (Fragments per Kilobase of Exon Per Million Fragments Mapped) were quantile normalized with
111 Combat correcting for batch. RNAseq data for this cohort is available at the Sage Synapse, AMP
112 AD Knowledge Portal (<https://www.synapse.org/#!/Synapse:syn2580853/wiki/66722>), under
113 synapse IDs syn3388564 (ROS/MAP RNAseq).

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

123

124 **Determination of linkage disequilibrium**

125 Linkage disequilibrium of variants at the *TREM* locus (**Fig. 3**) was evaluated in the Mayo
126 Clinic AD-risk GWAS HapMap2 imputed dataset (815 AD, 1218 controls) using Haploview 4.0
127 [9]. *TREM2* AD-risk missense variants rs142232675 (p.D87N) and rs75932628 (p.R47H) were
128 not present in the Mayo GWAS HapMap2 imputed dataset, but were directly genotyped in the
129 Mayo Clinic samples with TaqMan® assays, and were included to show their LD with variants
130 that associate with AD-risk in the IGAP meta-analysis. Sixteen variants located within 100kb of
131 a *TREM* gene and that had an AD-risk $p \leq 0.0015$ in the IGAP stage 1 meta-analysis are included,
132 in addition to two rare missense *TREM2* coding variants [rs142232675 (p.D87N), and
133 rs75932628 (p.R47H)] and 5 common *TREM* locus SNPs with prior reports of AD-risk
134 (rs3747742) [10] or endophenotype (rs7759295, rs6910730, rs6922617, rs6916710) [11, 12]
135 association, even though they did not meet the IGAP AD-risk association cutoff.

136 **Supplementary Results**

137 **Assessment of potential collider conditioning bias**

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

150

151 **Effect of RIN on percent detection and rs9357347 eQTL association**

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

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

161

162 **Association of rs9357347 with Braak stage**

163 Given the association of rs9357347 with AD-risk, we tested the hypothesis that this
164 variant could also show an association with Braak stage, as the latter is an important criterion for
165 the neuropathological diagnosis of AD [14]. Implementing an ANOVA model in R that
166 included age-at-death, sex and *APOE* ϵ 4 dose (0, 1 or 2 alleles), in the two larger datasets (Mayo
167 WG-DASL and ROS/MAP RNAseq), we determined that rs9357347 does not significantly
168 contribute to the variance in Braak stage in either of these two cohorts ($p=0.91$ and $p=0.27$,
169 respectively). In addition, we implemented linear regression analysis in R, again using the two
170 larger datasets to estimate the effect of each copy of the rs9357347 minor allele on Braak stage,
171 including age-at-death, sex and *APOE* ϵ 4 dose in the model. As shown in **Table S4**, we did not
172 detect a significant association of rs9357347 with Braak stage in either cohort.

173

174 **Association of rs9357347 with cognition**

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

179 longitudinal cognitive phenotype based on repeated measures of global cognition, as previously
180 described [15, 16]. The analysis was performed using linear regression analysis implemented in
181 R, under an additive model for rs9357347, and adjusting for age-at-death, sex and *APOE* ϵ 4
182 dose. Neither global cognitive decline nor global cognition at last evaluation shows an
183 association with rs9357347 in this cohort (**Table S5**).

184

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

186 As the WG-DASL cohort was overall younger than the two RNAseq cohorts, we assessed
187 the association of the age covariate on *TREML1* and *TREM2* gene expression levels in the linear
188 regression model described in the Material & Methods section 2.4. Age was not significantly
189 associated with either *TREML1* or *TREM2* expression in the Mayo WG-DASL cohort ($p > 0.05$).
190 On the other hand, both in the Mayo RNAseq and ROS/MAP RNAseq cohorts, *TREML1* and
191 *TREM2* expression levels appeared to be slightly increased with age, albeit the magnitude of the
192 effect sizes were modest, with beta coefficients equivalent to approximately a 1.01 and 1.03-fold
193 change in expression levels (Mayo RNAseq: *TREML1* $p = 0.085$, $\beta = 0.01$; *TREM2* $p = 0.026$,
194 $\beta = 0.02$. ROS/MAP RNAseq: *TREML1* $p = 2.0 \times 10^{-3}$, $\beta = 0.04$; *TREM2* $p = 4.4 \times 10^{-5}$,
195 $\beta = 0.03$). Since *TREML1* and *TREM2* gene expression levels appear to be increased with age, it
196 is possible that this might have led to a decrease in power to detect an association of rs9357347-
197 C with increased levels of these genes in the two older cohorts.

198

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

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

209
210

211 **Acknowledgements for IGAP:**

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

223 NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01-AG-12100,
224 the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical
225 Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984,
226 U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC-10-196728.

227

228 **References**

- 229 [1] J.C. Lambert, C.A. Ibrahim-Verbaas, D. Harold, A.C. Naj, R. Sims, C. Bellenguez, et al., Meta-analysis of
230 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*
231 2013;45:1452-1458.
- 232 [2] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan, Clinical diagnosis of
233 Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of
234 Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939-944.
- 235 [3] F. Zou, H.S. Chai, C.S. Younkin, M. Allen, J. Crook, V.S. Pankratz, et al., Brain expression genome-wide
236 association study (eGWAS) identifies human disease-associated variants. *PLoS genetics*
237 2012;8:e1002707.
- 238 [4] P. Du, W.A. Kibbe, S.M. Lin, lumi: a pipeline for processing Illumina microarray. *Bioinformatics*
239 2008;24:1547-1548.
- 240 [5] S.M. Lin, P. Du, W. Huber, W.A. Kibbe, Model-based variance-stabilizing transformation for Illumina
241 microarray data. *Nucleic acids research* 2008;36:e11.
- 242 [6] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq.
243 *Bioinformatics* 2009;25:1105-1111.
- 244 [7] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short
245 DNA sequences to the human genome. *Genome biology* 2009;10:R25.
- 246 [8] S. Anders, P.T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing
247 data. *Bioinformatics* 2015;31:166-169.
- 248 [9] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype
249 maps. *Bioinformatics* 2005;21:263-265.
- 250 [10] B.A. Benitez, B. Cooper, P. Pastor, S.C. Jin, E. Lorenzo, S. Cervantes, et al., TREM2 is associated with
251 the risk of Alzheimer's disease in Spanish population. *Neurobiology of aging* 2013;34:1711 e1715-1717.
- 252 [11] J.M. Replogle, G. Chan, C.C. White, T. Raj, P.A. Winn, D.A. Evans, et al., A TREM1 variant alters the
253 accumulation of Alzheimer-related amyloid pathology. *Annals of neurology* 2015;77:469-477.
- 254 [12] C. Cruchaga, J.S. Kauwe, O. Harari, S.C. Jin, Y. Cai, C.M. Karch, et al., GWAS of cerebrospinal fluid tau
255 levels identifies risk variants for Alzheimer's disease. *Neuron* 2013;78:256-268.
- 256 [13] S.R. Cole, R.W. Platt, E.F. Schisterman, H. Chu, D. Westreich, D. Richardson, et al., Illustrating bias
257 due to conditioning on a collider. *International journal of epidemiology* 2010;39:417-420.
- 258 [14] H. Braak, E. Braak, Neuropathological staging of Alzheimer-related changes. *Acta neuropathologica*
259 1991;82:239-259.
- 260 [15] P.A. Boyle, A.S. Buchman, R.S. Wilson, S.E. Leurgans, D.A. Bennett, Association of muscle strength
261 with the risk of Alzheimer disease and the rate of cognitive decline in community-dwelling older
262 persons. *Archives of neurology* 2009;66:1339-1344.
- 263 [16] R.S. Wilson, T.L. Beck, J.L. Bienias, D.A. Bennett, Terminal cognitive decline: accelerated loss of
264 cognition in the last years of life. *Psychosomatic medicine* 2007;69:131-137.

Supplementary Table

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

Symbol	Ensembl Gene ID	WG-DASL Probe ID	Mayo WG-DASL: Cerebellum ^a			Mayo WG-DASL: Temporal Cortex ^b			Mayo Clinic RNAseq: Temporal Cortex ^c	ROS/MAP RNAseq: DFPC ^d
			AD + non-AD	AD	nonAD	AD + non-AD	AD	Non-AD	AD + non-AD	AD + non-AD
<i>TREM1</i>	ENSG00000124731	ILMN_1688231	0.00	0.00	0.00	0.25	0.00	0.51	18.18	66.99
<i>TREML1</i>	ENSG00000161911	ILMN_1690783	100.00	100.00	100.00	100.00	100.00	100.00	100.00	97.84
<i>TREM2</i>	ENSG00000095970	ILMN_1701248	40.91	43.59	37.29	98.25	99.50	96.95	100.00	100.00
<i>TREML2</i>	ENSG00000112195	ILMN_1740864	17.38	17.44	17.51	6.27	10.40	2.03	8.33	24.75
<i>TREML4</i>	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; 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.

Dataset	Group	N	TREM1			TREM2		
			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} .

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
<i>TREML1</i>	RIN < 6.5	188	0.087	0.042	0.043
	RIN > 6.5	192	0.084	0.046	0.069
<i>TREM2</i>	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} .

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* ϵ 4 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.

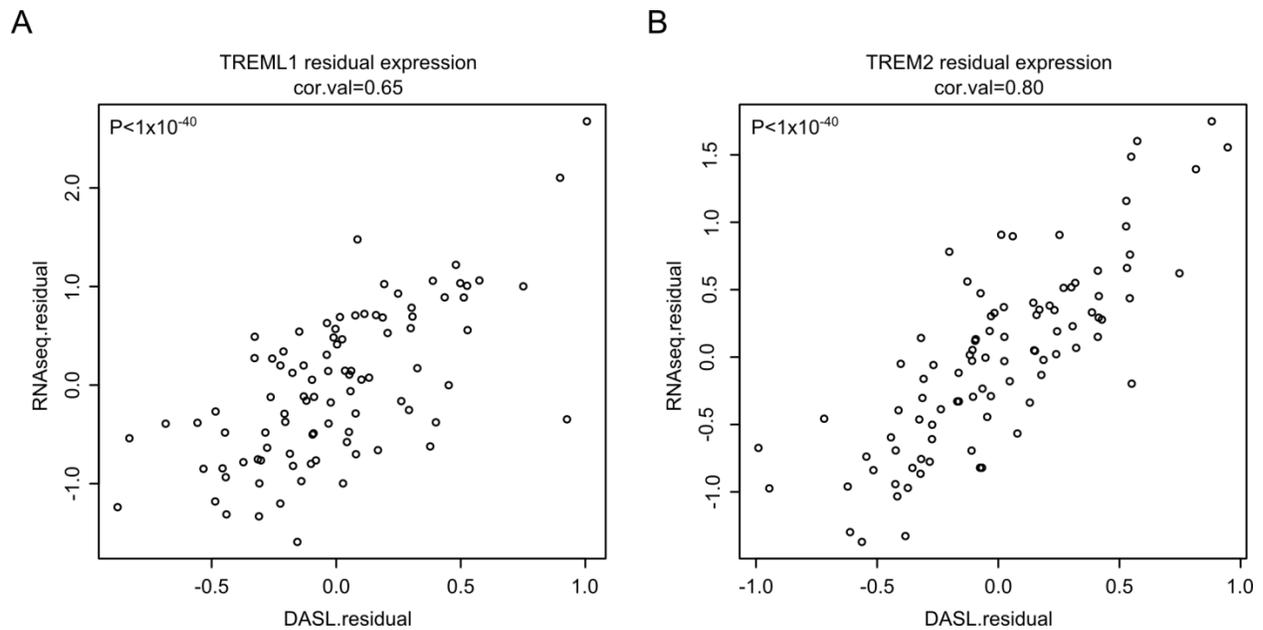
Table S5. Association of rs9357347 with cognition.

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* ϵ 4 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.

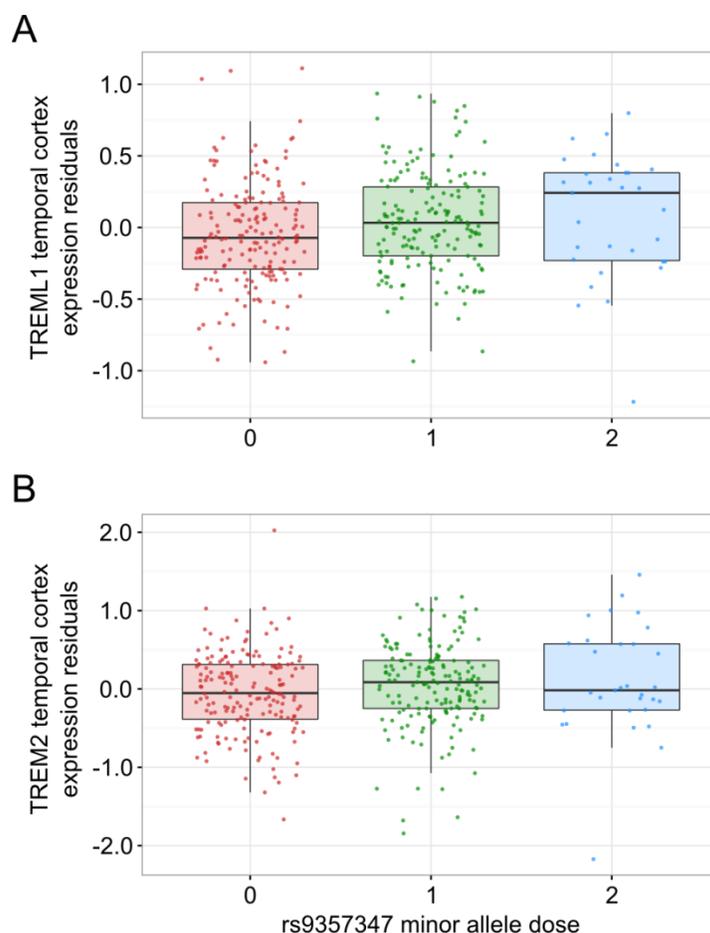
Supplementary Figures

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.

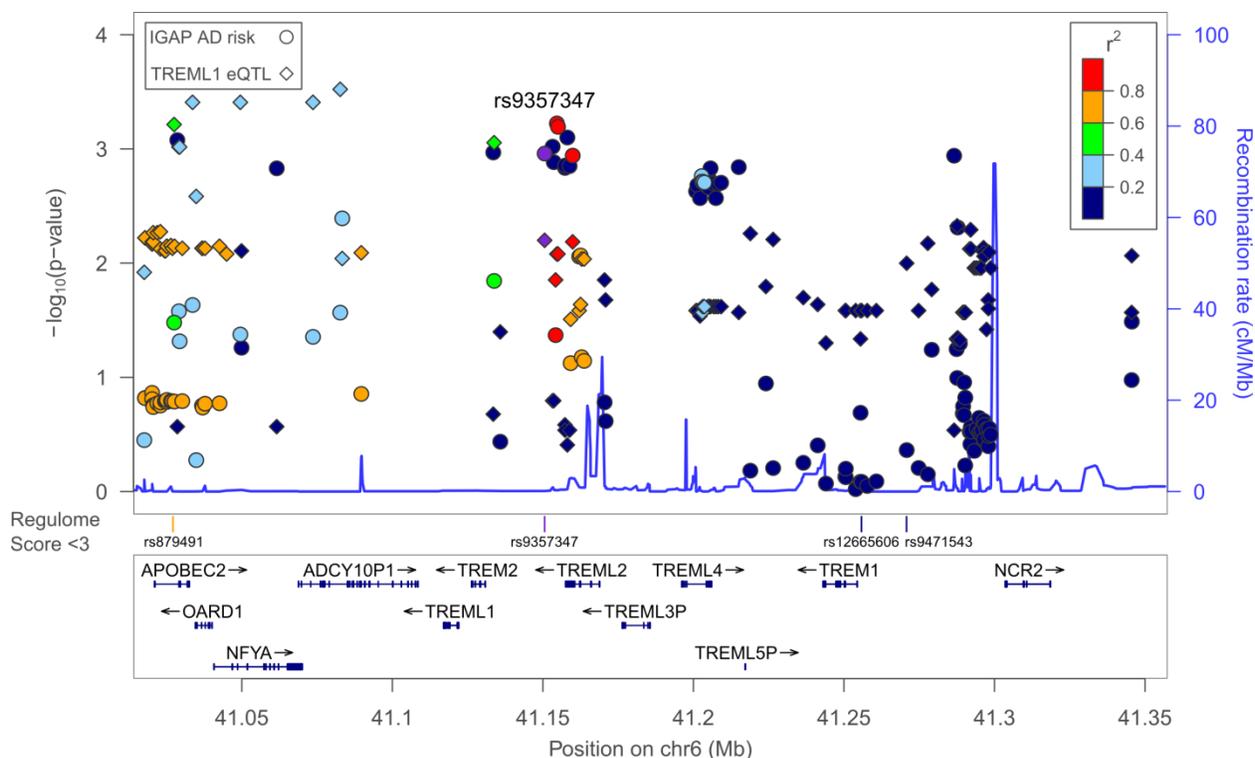
Fig. S2. Box plots of rs9357347 genotype associations with *TREML1* and *TREM2* brain gene expression levels.



Gene expression residuals obtained in R adjusted for covariates [*APOE* ϵ 4 dosage, age at death, diagnosis, sex, PCR plate, RIN, $(\text{RIN}-\text{RINmean})^2$] 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 $1.5 \times \text{IQR}$ below the 1st quartile and $1.5 \times \text{IQR}$ 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*.

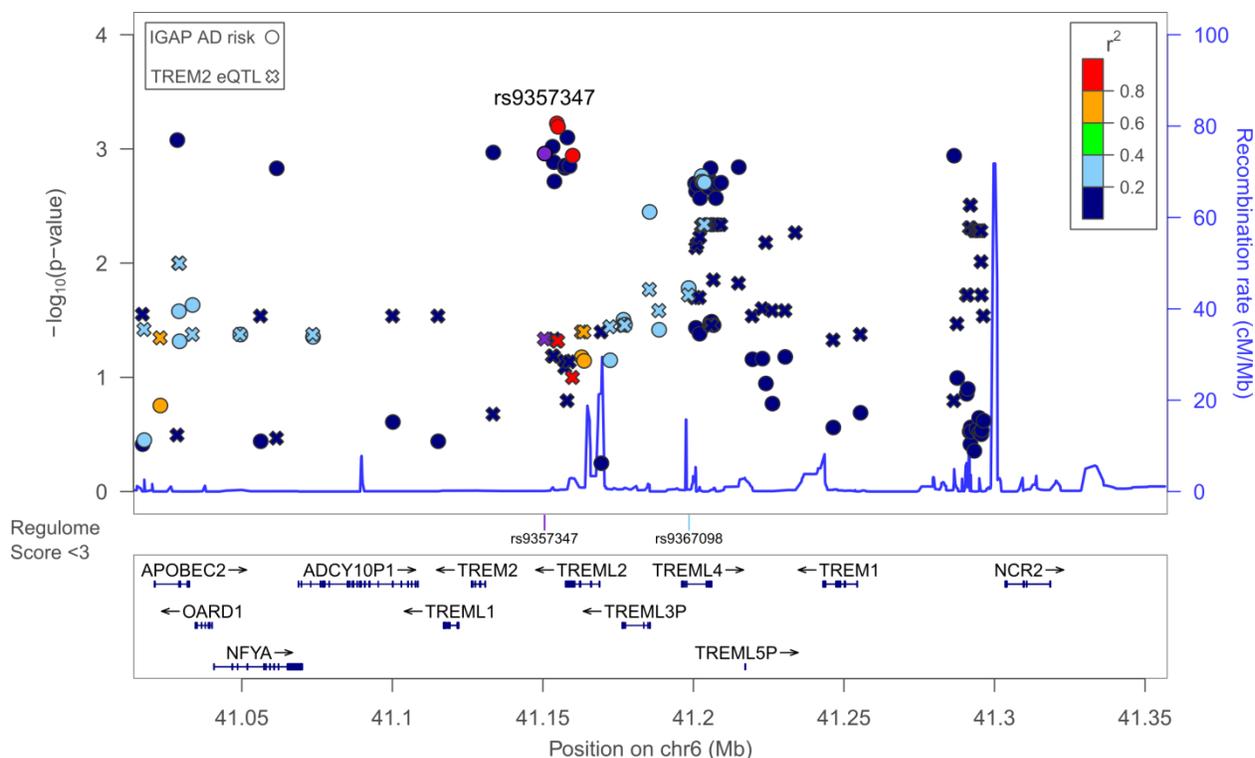
Fig. S3. Regional association plot showing AD-risk and *TREML1* eQTL p-values at the *TREM* locus.



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 microarrays are shown, with p-values indicated by the scale on the left y-axis as $-\log_{10}(\text{p-value})$. 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

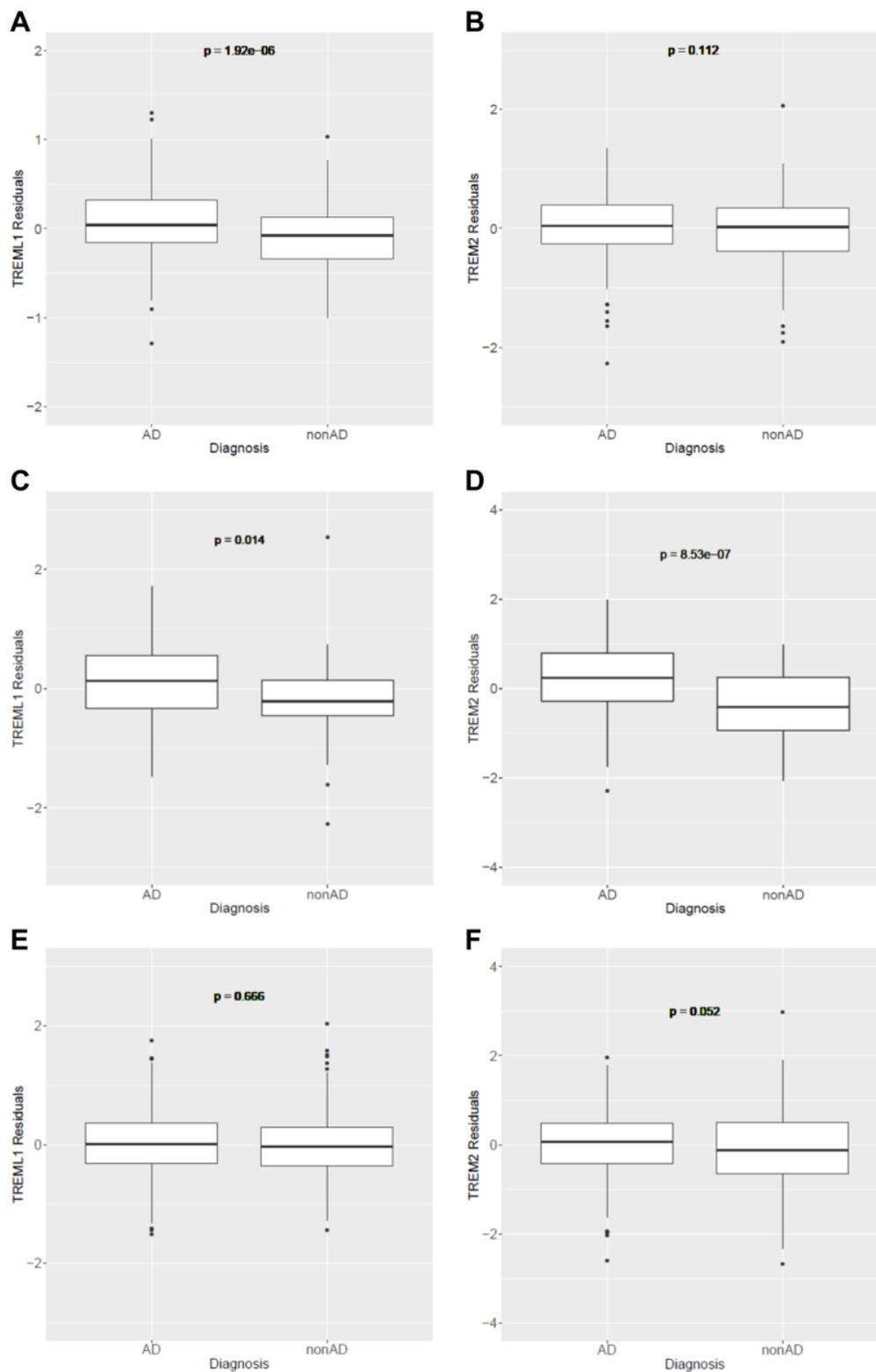
plot. Gene locations across the targeted genomic region (*TREM* gene +/-100 kb: chr6:41016999-41354457) are shown below the plot relative to the variant positions according to the February 2009 human genome assembly (GRCh37hg19). The regional association plot was generated using LocusZoom (<http://locuszoom.sph.umich.edu/locuszoom/>).

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



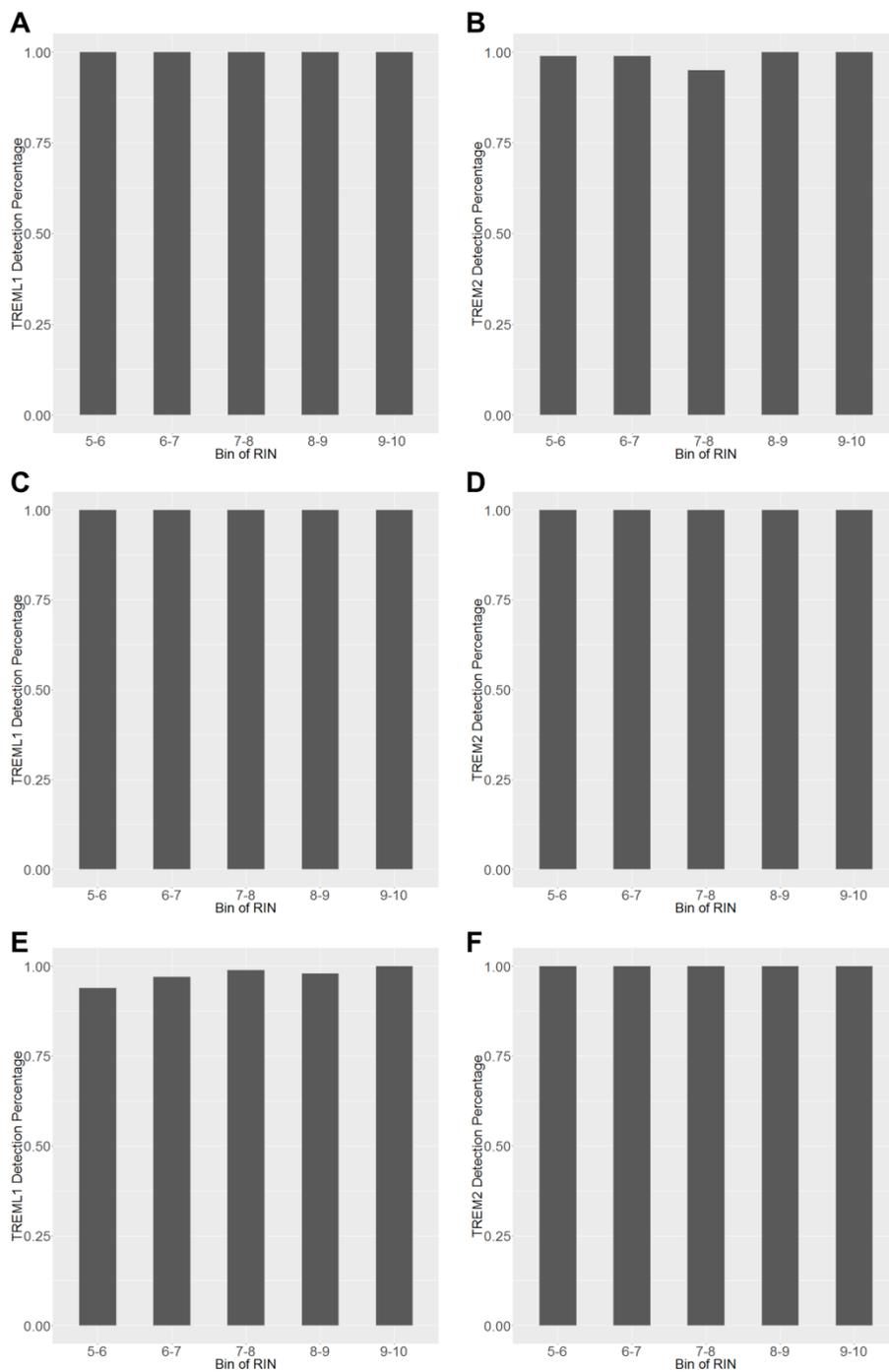
On this plot each variant is depicted as both, a circle that denotes its IGAP stage 1 meta-analysis p-value, and an “X” denoting its *TREM2* eQTL p-value. All variants at the *TREM* locus that either achieved a p-value ≤ 0.0015 in the IGAP stage1 meta-analysis or that had a *TREM2* p-value < 0.05 in our eQTL analysis of temporal cortex gene expression measured by WG-DASL microarrays are shown, with p-values indicated by the scale on the left y-axis as $-\log_{10}(p\text{-value})$. 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.



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* ϵ 4 dose, RIN, $(\text{RIN}-\text{RINmean})^2$ 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* ϵ 4 dose, RIN, $(\text{RIN}-\text{RINmean})^2$ and flowcell. *TREML1* (E) and *TREM2* (F) in the ROS/MAP RNAseq dataset, adjusted for rs9357347 minor allele dose, age-at-death, sex, *APOE* ϵ 4 dose, RIN, $(\text{RIN}-\text{RINmean})^2$.

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 RNAseq dataset.