

RESEARCH

Open Access

# Association of *MAPT* haplotypes with Alzheimer's disease risk and *MAPT* brain gene expression levels

Mariet Allen<sup>1</sup>, Michaela Kachadoorian<sup>1</sup>, Zachary Quicksall<sup>1</sup>, Fanggeng Zou<sup>1</sup>, High Seng Chai<sup>2</sup>, Curtis Younkin<sup>1</sup>, Julia E Crook<sup>3</sup>, V Shane Pankratz<sup>2</sup>, Minerva M Carrasquillo<sup>1</sup>, Siddharth Krishnan<sup>1</sup>, Thuy Nguyen<sup>1</sup>, Li Ma<sup>1</sup>, Kimberly Malphrus<sup>1</sup>, Sarah Lincoln<sup>1</sup>, Gina Bisceglia<sup>1</sup>, Christopher P Kolbert<sup>4</sup>, Jin Jen<sup>4</sup>, Shubhabrata Mukherjee<sup>5</sup>, John K Kauwe<sup>6</sup>, Paul K Crane<sup>5</sup>, Jonathan L Haines<sup>7,8</sup>, Richard Mayeux<sup>9</sup>, Margaret A Pericak-Vance<sup>10</sup>, Lindsay A Farrer<sup>11</sup>, Gerard D Schellenberg<sup>12</sup>, Alzheimer's Disease Genetics Consortium (ADGC), Joseph E Parisi<sup>13</sup>, Ronald C Petersen<sup>14</sup>, Neill R Graff-Radford<sup>15</sup>, Dennis W Dickson<sup>1</sup>, Steven G Younkin<sup>1</sup> and Nilüfer Ertekin-Taner<sup>1,15\*</sup>

## Abstract

**Introduction:** *MAPT* encodes for tau, the predominant component of neurofibrillary tangles that are neuropathological hallmarks of Alzheimer's disease (AD). Genetic association of *MAPT* variants with late-onset AD (LOAD) risk has been inconsistent, although insufficient power and incomplete assessment of *MAPT* haplotypes may account for this.

**Methods:** We examined the association of *MAPT* haplotypes with LOAD risk in more than 20,000 subjects ( $n$ -cases = 9,814,  $n$ -controls = 11,550) from Mayo Clinic ( $n$ -cases = 2,052,  $n$ -controls = 3,406) and the Alzheimer's Disease Genetics Consortium (ADGC,  $n$ -cases = 7,762,  $n$ -controls = 8,144). We also assessed associations with brain *MAPT* gene expression levels measured in the cerebellum ( $n$  = 197) and temporal cortex ( $n$  = 202) of LOAD subjects. Six single nucleotide polymorphisms (SNPs) which tag *MAPT* haplotypes with frequencies greater than 1% were evaluated.

**Results:** H2-haplotype tagging rs8070723-G allele associated with reduced risk of LOAD (odds ratio, OR = 0.90, 95% confidence interval, CI = 0.85-0.95,  $p$  = 5.2E-05) with consistent results in the Mayo (OR = 0.81,  $p$  = 7.0E-04) and ADGC (OR = 0.89,  $p$  = 1.26E-04) cohorts. rs3785883-A allele was also nominally significantly associated with LOAD risk (OR = 1.06, 95% CI = 1.01-1.13,  $p$  = 0.034). Haplotype analysis revealed significant global association with LOAD risk in the combined cohort ( $p$  = 0.033), with significant association of the H2 haplotype with reduced risk of LOAD as expected ( $p$  = 1.53E-04) and suggestive association with additional haplotypes. *MAPT* SNPs and haplotypes also associated with brain *MAPT* levels in the cerebellum and temporal cortex of AD subjects with the strongest associations observed for the H2 haplotype and reduced brain *MAPT* levels ( $\beta$  = -0.16 to -0.20,  $p$  = 1.0E-03 to 3.0E-03).

**Conclusions:** These results confirm the previously reported *MAPT* H2 associations with LOAD risk in two large series, that this haplotype has the strongest effect on brain *MAPT* expression amongst those tested and identify additional haplotypes with suggestive associations, which require replication in independent series. These biologically congruent results provide compelling evidence to screen the *MAPT* region for regulatory variants which confer LOAD risk by influencing its brain gene expression.

\* Correspondence: [taner.nilufer@mayo.edu](mailto:taner.nilufer@mayo.edu)

<sup>1</sup>Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL 32224, USA

<sup>15</sup>Department of Neurology, Mayo Clinic Florida, 4500 San Pablo Road, Birdsall 3, Jacksonville, FL 32224, USA

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

## Introduction

Alzheimer's disease (AD), the most prevalent cause of dementia, is defined by two neuropathological hallmarks: senile plaques primarily composed of extracellular amyloid-beta ( $A\beta$ ) deposits and intracellular neurofibrillary tangles (NFTs) comprised of hyper-phosphorylated tau protein. *MAPT* (micro-tubule associated protein) encodes tau and resides within a ~900 kilobase (kb) inversion polymorphism (reviewed [1]) that generates a ~1.3 megabase (Mb) region of linkage disequilibrium (LD) defined by two extended haplotypes, referred to as H1 and H2. Variants have evolved that occur on only the H1 haplotype resulting in multiple sub-haplotypes.

Both common and rare genetic variation in *MAPT* have been strongly implicated in primary tauopathies. Rare missense and exon 10 splicing mutations, which lead to increased levels of tau isoforms with four microtubule binding domains (aka 4-repeat or 4R tau) lead to familial frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) [2,3], whereas the common *MAPT* H1 haplotype strongly associates with increased risk of progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) [4-8]. A recent genome-wide association study (GWAS) of PSP risk identified *MAPT* as the strongest locus, with risk alleles at rs8070723 which tags the H1 haplotype and also for rs242557, which partially tags the H1c subhaplotype [8].

Despite having tauopathy as a defining lesion, reports of association between AD and genetic variants at the *MAPT* locus are inconsistent. While *MAPT* H1 [9] haplotype or H1c subhaplotype [10-13] showed association with AD risk in some studies, others failed to detect association with H1 [10,13,14], H1c [15] or other *MAPT* variants [16]. The sample size for most of these studies range from a few hundred to a few thousand; and the largest published study of ~17,000 subjects only evaluated the H1/H2 haplotypes but none of the H1-subhaplotypes [9].

In addition to investigations of *MAPT* variants with risk for tauopathies, some studies also assessed their role in gene expression. *MAPT* exons 2, 3, 4a, 6 8 and 10 are known to be alternatively spliced [1], there are FTDP-17 splicing mutations which increase 4R tau [2,3] and 4R tau is increased in affected brain regions in PSP and CBD [17,18]. Allele-specific gene expression studies in human brains and neuronal cell lines identified higher levels of exon 10 containing transcripts but not total *MAPT* associated with the H1-haplotype [19] and higher levels of exon 2- and 3-containing transcripts associated with the H2 haplotype [20]. *MAPT* H1c-subhaplotype was associated with higher total and 4R *MAPT* levels in human brains [11]. A study of exon levels in multiple brain regions from humans identified higher expression levels of exon 3 associated with the H2 haplotype, but no association of *MAPT* levels with the H1c-subhaplotype

[21]. We have previously reported association of *MAPT* H1-tagging and rs242557 SNPs with increased brain *MAPT* levels in ~400 brains from a combined cohort of subjects with AD and other brain pathologies [22]. Collectively, these findings suggest that the disease risk conferred by some *MAPT* variants could be due to higher total or 4R tau levels and/or that the protective effect of *MAPT* H2-haplotype might be secondary to an increase in N-terminal exon-containing *MAPT* transcripts. While these studies are informative, to date, there has not been a systematic and well-powered analysis of *MAPT* subhaplotypes for association with *MAPT* brain expression levels.

Herein, we present a comprehensive assessment of *MAPT* variants that tag all *MAPT* subhaplotypes of frequency >1% in the largest to date *MAPT* association study of 9,814 LOAD cases vs. 11,550 controls. Further, we evaluate association of these *MAPT* variants in two brain regions: the cerebellum, which is predominantly unaffected in AD and the typically affected temporal cortex from ~200 autopsied LOAD subjects. Our well-powered and complementary investigation of disease risk and gene expression provides compelling evidence for a role of transcriptional regulatory variants of *MAPT* in conferring LOAD risk.

## Methods

### Subjects and samples

#### *Mayo clinic cohort*

We evaluated LOAD risk association with *MAPT* variants in 2,052 LOAD cases vs. 3,406 controls from Mayo Clinic. These elderly European-American subjects were from two clinical case-control series recruited at the Mayo Clinic in Rochester, MN (RS series: 615 LOAD cases, 2,425 controls) and Jacksonville, FL (JS series: 886 LOAD cases, 981 controls), as well as 551 autopsy-confirmed LOAD subjects from the Brain Bank at Mayo Clinic Florida (Additional file 1: Table S1). All clinical subjects were evaluated by a Mayo Clinic neurologist and autopsied subjects were diagnosed by our neuropathologist (DWD). All clinical LOAD cases had probable or possible AD and all pathologic LOAD cases had definite AD according to NINCDS-ADRDA criteria [23]. All controls had a clinical dementia rating score of 0. All LOAD subjects had an age at disease diagnosis (clinical), death (autopsied) and controls at their most recent visit  $\geq 60$  years. A subset of the Mayo Clinic cohort was included in the Mayo LOAD GWAS [24] (Additional file 1: Table S1) and gene expression GWAS (eGWAS) [22] from the temporal cortex (TCX,  $n = 202$ ) and cerebellum (CER,  $n = 197$ ). This study was approved by the Mayo Clinic institutional review board and appropriate informed consent was obtained from all individuals.

### **ADGC cohort**

We utilized genetic data and covariate information on the European-American subjects from the Alzheimer's Disease Genetics Consortium (ADGC) cohort. These subjects were collected from multiple research centers and designated into the following 14 series: Adult Changes in Thought (ACT)/Electronic Medical Records and Genetics (eMERGE), National Institute on Aging (NIA) Alzheimer Disease Centers (ADCs), Alzheimer Disease Neuroimaging Initiative (ADNI), Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GenADA), University of Miami/Vanderbilt University/Mt. Sinai School of Medicine (UM/VU/MSSM), MIRAGE Study, Oregon Health and Science University (OHSU), NIA-LOAD, Translational Genomics Research Institute series 2 (TGEN2), Rush University Religious Orders Study/Memory and Aging Project (ROSMAP), University of Pittsburgh (UP), and Washington University (WU). Detailed descriptions of these cohorts are provided elsewhere [25,26].

The ADGC cohort included subjects from the Mayo Clinic. To avoid any overlap, all subjects from Mayo Clinic were removed from the ADGC cohort. Standard quality control (QC) measures were applied to the ADGC dataset [27] with the following cutoffs, 95% call rate per person, 1% minor allele frequency (MAF) and 95% call rate for SNP, Hardy-Weinberg equilibrium (HWE)  $p > 1E-06$  in controls. Additionally, directly observed (not imputed) SNPs from subjects across all series were evaluated for relatedness by using KING (Kinship-based Inference for Gwas)-Robust [27] and a single representative was chosen for each pair of individuals who were third degree relatives or closer. Similarly, one representative was chosen for each family for the MIRAGE and NIA-LOAD family based studies. All cohort genotypes were imputed to a common set of >2 million SNPs (HapMap2) by the ADGC, as described [26]. The 7,762 LOAD cases and 8,144 controls from the ADGC (Additional file 1: Table S1), which remained after the QC, were utilized for the *MAPT* variant associations.

### **RNA isolation and gene expression measurements**

All samples utilized in the brain gene expression analyses in this study are a subset of the previously published Mayo Clinic expression GWAS (eGWAS) [22]. In the current study brain gene expression levels of autopsied LOAD subjects measured from the cerebellum ( $n = 197$ ) and temporal cortex ( $n = 202$ ) were used. RNA extraction and gene expression measurements were previously reported [22]. Briefly, total RNA was isolated from frozen post-mortem brain tissue using the Ambion RNAqueous kit according to the manufacturer's instructions. The quantity and quality of the RNA were evaluated using the Agilent 2100 Bioanalyzer and RNA 6000 Nano Chip.

Whole Genome DASL assay (WG-DASL, Illumina, San Diego, CA) was used to measure transcript levels. This platform is designed for gene expression measurements for partially degraded RNA such as is typically isolated from frozen human brains. Details of gene expression measurements, data processing and QC were already published [22]. Briefly, 15 replicate samples measured on 5–6 different plates and on 2–3 different days were included in the study for QC and also for intra-class coefficient (ICC) [28] estimations. Raw probe level mRNA expression data were exported from GenomeStudio software (Illumina Inc.) for preprocessing with background correction, variance stabilizing transformation, quantile normalization and probe filtering using the lumi package of BioConductor [29,30]. Probes with detectable signal in >75% of the samples were used in subsequent analyses. We also annotated all of the probes by comparing their positions according to NCBI Ref Seq, Build 36.3 to those of all variants within dbSNP131 and identified the list of probes which have  $\geq 1$  variants within their sequence.

### **Genotyping**

Six *MAPT* locus haplotype tagging (ht) SNPs were selected for genotyping in the Mayo Clinic cohort (Additional file 1: Figure S1, Tables S1 and S2). SNP rs8070723 was used as a proxy for the H1/H2 haplotypes defining *del-1n9*. The remaining 5 SNPs have been previously described to tag the majority of H1 sub-haplotypes [6]. Genotypes for three SNPs (rs1467967, rs242557 and rs8070723) for a subset of the samples were obtained from the Mayo Clinic LOAD GWAS (Additional file 1: Table S1). The remaining genotypes for these and all genotypes for an additional three SNPs (rs3785883, rs2471738 and rs7521) were obtained using Applied Biosystems® Taqman genotyping assays. The genotypes for these six SNPs were extracted from the ADGC GWAS data [27] using PLINK [31].

### **Statistical analysis**

#### ***MAPT* single SNP association analysis with LOAD risk**

All six htSNPs were tested for association with disease risk in the combined Mayo Clinic cohort, as well as individually in the JS and RS series. The same SNPs were also tested in the ADGC cohort, as well as in the ADGC + Mayo combined cohorts. All SNPs were tested for deviations from Hardy-Weinberg equilibrium (HWE) [31] in controls.

Single SNP associations with disease risk were tested assuming an additive model, using multivariable logistic regression implemented in PLINK [31] including the following covariates: Age (defined for Mayo Clinic cohort as age at diagnosis/death/last diagnosis for clinical LOAD/autopsied LOAD/controls), sex, *APOE*  $\epsilon 4$  dosage and series. The analyses in the ADGC-only cohort included these covariates and also ten principal components

obtained from EIGENSTRAT [32]. Mayo Clinic-only and Mayo + ADGC analyses did not include principal components, as they were not available for many of the Mayo subjects.

#### ***MAPT haplotype association analysis with LOAD risk***

PLINK was used to estimate haplotype frequencies using the sliding window specification with a window size of six to encompass all six of the htSNPs. Haplotype associations with LOAD risk in the Mayo Clinic series were performed with both PLINK and haplo.score [33], which revealed identical results for the single haplotype analyses. According to the score statistic approach, all possible haplotypes consistent with the observed marker genotypes are obtained, maximum likelihood estimates of the haplotype frequencies, as well as the posterior probabilities of the pairs of haplotypes for each subject are computed. These posterior probabilities are then used to compute the score statistics for the association of (ambiguous) haplotypes with LOAD risk using multivariable logistic regression analysis with inclusion of the same covariates as discussed above. Only those haplotypes with frequencies >1% in the cohorts that they were tested in were included in the association analyses.

#### ***MAPT variant association analysis with gene expression levels***

Each of the *MAPT* htSNPs and the estimated haplotypes were also tested for association with gene expression levels of *MAPT* in the TCX and CER of LOAD subjects, as measured using three probes: ILMN\_1710903, ILMN\_2310814 and ILMN\_2298727. These LOAD subjects were also participants in our previously published eGWAS [22]. Association analysis was carried out in PLINK using linear regression approach, whereby preprocessed probe transcript levels for the three probes in each brain region (TCX and CER) were assessed as six individual quantitative phenotypes. Covariates included in the models were age at death, sex, *APOE*  $\epsilon$ 4 dosage, PCR plate, RNA integrity number (RIN) and adjusted RIN<sup>2</sup>, as described previously [22,34]. Only those haplotypes with frequencies >1% in the autopsy series that they were tested in were included in the association analyses.

## **Results**

### **Association of *MAPT* single SNPs with LOAD risk**

Six *MAPT* htSNPs were tested for association with LOAD risk in the Mayo Clinic and ADGC cohorts both individually and combined (Table 1). All SNPs had genotyping call rates  $\geq$ 90% in the Mayo Cohort (~90-97%), ~83-100% in the ADGC cohort and ~85-100% in the combined cohort (Additional file 1: Table S2). *MAPT* rs242557 had the lowest call rate of 83% in the ADGC cohort, with all other SNPs having call rates of  $\geq$ 89%. All SNPs passed the HWE cutoff of  $p > 1E-06$  in controls, although

rs242557 had HWE  $p < 0.05$  in the Mayo Clinic, but not the ADGC controls.

There was highly significant association of H2-tagging rs8070723-G allele with reduced risk of LOAD in the Mayo Clinic cohort (odds ratio = OR = 0.81,  $p = 7.0E-4$ ) with remarkably similar OR estimates in the JS and RS series (Additional file 1: Table S3) and in the independent ADGC cohort (OR = 0.89,  $p = 1.3E-4$ ) (Table 1). The association in the combined Mayo + ADGC cohort for this variant was highly significant (OR = 0.90,  $p = 5.3E-5$ ) and would withstand Bonferroni correction for the six tested variants but not achieve significance at a genome-wide level.

In addition, rs3785883-A allele was nominally significant in the combined Mayo + ADGC cohort (OR = 1.06,  $p = 0.034$ ) with very similar OR estimates in the Mayo Clinic (OR = 1.07, 95% confidence interval = 95% CI = 0.95-1.22) and ADGC (OR = 1.08, 95% CI = 1.01-1.15) cohorts. *MAPT* SNP rs1467967-G allele had suggestive LOAD risk association in the combined cohort (OR = 1.05, 95% CI = 1.0-1.10,  $p = 0.062$ ) with risky OR estimates in both cohorts. *MAPT* SNP rs242557, previously implicated in AD [11-13] is not associated with LOAD in the combined ADGC + Mayo cohort of 18,342 subjects ( $p = 0.974$ ), or the individual Mayo Clinic or ADGC cohorts.

### **Association of *MAPT* haplotypes with LOAD risk**

In the Mayo Clinic cohort of ~5,000 subjects, we identified 19 *MAPT* haplotypes with a frequency >1%. In this cohort, rs8070723-G allele tagged the H2 haplotype, present in 21.5% of the subjects, perfectly. Eighteen sub-haplotypes were identified on the H1 background. Three *MAPT* haplotypes were nominally significantly associated with LOAD risk (Table 2) and a global test for haplotypic association was also significant ( $p = 0.012$ ).

As expected, the *MAPT* H2 haplotype was significantly associated with decreased risk for LOAD in the Mayo Clinic cohort (OR = 0.80,  $p = 4.1E-04$ ). Additionally the most common sub-haplotype on the H1 background, H1b (frequency = 17.3%), was nominally significantly associated with increased risk for LOAD (OR = 1.15,  $p = 0.046$ ); as was a less frequent H1 sub-haplotype J (frequency = 1.2%, OR = 1.88,  $p = 0.031$ ), while three other H1 sub-haplotypes were marginally associated, also with increased LOAD risk (L, X and Y).

In the ADGC cohort, the *MAPT* H2 haplotype, was present in 22% of the subjects. On the H1 background, 19 sub-haplotypes were identified with a frequency of  $\geq$ 1%. As with the Mayo Clinic cohort, H2 haplotype was significantly associated with reduced risk of LOAD in the ADGC cohort (OR = 0.90,  $p = 6.29E-04$ ). None of the H1-subhaplotypes had significant association with LOAD risk in this cohort.

**Table 1 MAPT single SNPs association results with LOAD risk in the Mayo, ADGC and combined Mayo + ADGC cohorts**

SNP	A1	Mayo cohort						ADGC cohort						Mayo + ADGC					
		N	MAF_A	MAF_U	OR	95% CI	P	N	MAF_A	MAF_U	OR	95% CI	P	N	MAF_A	MAF_U	OR	95% CI	P
rs1467967	G	<b>4,986</b>	<b>0.335</b>	<b>0.330</b>	<b>1.10</b>	<b>0.99 - 1.22</b>	<b>0.079</b>	14,365	0.329	0.327	1.05	0.99 - 1.11	0.102	<b>19,351</b>	<b>0.330</b>	<b>0.327</b>	<b>1.05</b>	<b>1.00 - 1.10</b>	<b>0.062</b>
rs242557	A	4,935	0.377	0.379	1.00	0.90 - 1.11	0.988	13,407	0.358	0.356	1.01	0.95 - 1.07	0.721	18,342	0.363	0.364	1.00	0.95 - 1.05	0.974
rs3785883	A	5,247	0.182	0.183	1.07	0.95 - 1.22	0.274	<b>15,187</b>	<b>0.178</b>	<b>0.171</b>	<b>1.08</b>	<b>1.01 - 1.15</b>	<b>0.032</b>	<b>20,434</b>	<b>0.179</b>	<b>0.175</b>	<b>1.06</b>	<b>1.01 - 1.13</b>	<b>0.034</b>
rs2471738	T	5,282	0.223	0.209	1.07	0.95 - 1.21	0.250	14,181	0.207	0.201	1.05	0.98 - 1.12	0.194	19,463	0.209	0.203	1.05	0.99 - 1.11	0.109
rs8070723	G	<b>5,129</b>	<b>0.209</b>	<b>0.222</b>	<b>0.81</b>	<b>0.71 - 0.91</b>	<b>7.0E-04</b>	<b>15,895</b>	<b>0.211</b>	<b>0.221</b>	<b>0.89</b>	<b>0.84 - 0.94</b>	<b>1.3E-04</b>	<b>21,024</b>	<b>0.211</b>	<b>0.221</b>	<b>0.90</b>	<b>0.85 - 0.95</b>	<b>5.3E-05</b>
rs7521	A	5,171	0.458	0.457	1.08	0.97 - 1.19	0.152	15,656	0.466	0.468	1.02	0.97 - 1.07	0.529	20,827	0.465	0.465	1.02	0.97 - 1.06	0.437

Results of multivariable logistic regression analyses are shown. CHR = chromosome. A1 = Minor Allele, N = number of subjects with genotype calls, A = Affected (LOAD subjects), U = Unaffected (Control), MAF = Minor Allele Frequency, OR = Odds Ratio, 95% CI = 95% Confidence intervals, P = p-value.

Boldface values within the tables indicate significant or suggestive associations with a p-value <0.10.

**Table 2 MAPT haplotype association results with LOAD risk in the Mayo, ADGC and combined Mayo + ADGC cohorts**

Haplotype	Alleles	Mayo cohort					ADGC cohort					Mayo + ADGC cohort				
		F_All	F_A	F_U	OR	P	F_All	F_A	F_U	OR	P	F_All	F_A	F_U	OR	P
A (H2a)	AGGCGG	<b>0.215</b>	<b>0.205</b>	<b>0.221</b>	<b>0.80</b>	<b>4.1E-04</b>	<b>0.220</b>	<b>0.225</b>	<b>0.236</b>	<b>0.90</b>	<b>6.29E-04</b>	<b>0.221</b>	<b>0.228</b>	<b>0.240</b>	<b>0.90</b>	<b>1.53E-04</b>
B (H1b)	GGGCAA	0.173	0.180	0.169	1.15	<b>0.046</b>	0.190	0.194	0.197	1.05	0.208	<b>0.185</b>	<b>0.197</b>	<b>0.196</b>	<b>1.05</b>	<b>0.089</b>
C (H1c)	AAGTAG	0.118	0.118	0.118	0.91	0.277	0.127	0.131	0.131	1.03	0.545	0.124	0.132	0.131	1.00	0.997
D (H1d)	AAGCAA	0.076	0.071	0.078	0.99	0.905	0.075	0.074	0.079	0.93	0.195	<b>0.074</b>	<b>0.077</b>	<b>0.081</b>	<b>0.91</b>	<b>0.074</b>
E (H1e)	AGGCAA	0.074	0.075	0.073	1.12	0.308	0.077	0.082	0.081	1.00	0.961	0.077	0.083	0.081	1.02	0.618
G	GAACAA	0.017	0.014	0.018	0.91	0.692	0.012	0.013	0.013	0.84	0.256	0.014	0.014	0.015	0.85	0.176
H	AGACAA	0.044	0.038	0.048	0.91	0.506	0.042	0.045	0.041	1.08	0.309	0.042	0.045	0.045	1.03	0.641
I	GAGCAA	0.037	0.041	0.035	1.06	0.732	0.034	0.035	0.035	1.03	0.718	0.035	0.037	0.036	1.06	0.422
J	AGGCAG	0.012	0.015	0.010	1.88	<b>0.031</b>	0.010	0.011	0.010	1.20	0.292	<b>0.011</b>	<b>0.012</b>	<b>0.010</b>	<b>1.32</b>	<b>0.049</b>
L	AGACAG	0.029	0.032	0.027	1.37	<b>0.059</b>	0.032	0.034	0.033	1.06	0.483	0.031	0.034	0.032	1.10	0.187
M	GAGCAG	0.025	0.020	0.027	0.78	0.215	0.021	0.022	0.020	1.09	0.459	0.022	0.023	0.023	1.00	0.978
N	GGACAG	NA					0.011	0.011	0.011	1.12	0.527	NA				
O	AAACAA	0.016	0.016	0.017	0.90	0.696	0.018	0.020	0.018	1.18	0.192	0.018	0.020	0.019	1.09	0.457
P	GGGTAG	0.014	0.014	0.014	1.31	0.301	0.013	0.014	0.014	1.07	0.628	0.014	0.015	0.014	1.06	0.626
R	AGGTAG	0.017	0.017	0.017	1.01	0.952	0.012	0.013	0.011	1.26	0.162	0.013	0.014	0.013	1.21	0.144
U	AAGCAG	0.025	0.028	0.024	1.14	0.517	0.025	0.026	0.025	1.05	0.667	0.025	0.027	0.025	1.04	0.642
V	GGATAG	0.011	0.012	0.010	1.41	0.233	0.010	0.011	0.010	1.14	0.463	0.011	0.012	0.011	1.20	0.210
W	GGGCAG	0.012	<0.010	0.013	0.92	0.783	NA					NA				
X	GAATAG	0.016	0.018	0.014	1.59	<b>0.054</b>	0.013	0.014	0.014	1.03	0.817	0.014	0.015	0.014	1.16	0.223
Y*	AAATAG	0.015	0.017	0.014	1.64	<b>0.056</b>	0.013	0.014	0.013	1.19	0.306	0.013	0.015	0.013	1.22	0.147
Z*	GAGTAG	NA					0.011	0.011	0.010	1.10	0.590	NA				
<b>Global p value</b>		<b>0.0123</b>					<b>0.375</b>					<b>0.0329</b>				

Results of multivariable logistic regression analyses for MAPT haplotypes with frequencies >1% are shown. Haplotype nomenclature is assigned as previously reported [6,35]. Alleles for the SNPs defining the haplotypes are given in the 5' to 3' order as follows: rs1467967, rs242557, rs3785883, rs2471738, rs8070723, rs7521. Haplotypes not previously observed are designated by an asterisk (\*). F\_All = haplotype frequency in all subjects; F\_A = in affected (LOAD) and F\_U = unaffected (Control) subjects. OR = Odds Ratio, P = p-value.

Boldface values within the tables indicate significant or suggestive associations with a p-value <0.10.

In the combined Mayo + ADGC cohort, there was significant global haplotypic association ( $p = 0.033$ ). *MAPT* H2 haplotype had highly significant association with reduced risk of LOAD (OR = 0.90,  $p = 1.53E-04$ ). *MAPT* J subhaplotype had nominally significant association with LOAD risk in the combined cohort (OR = 1.32,  $p = 0.049$ ) with suggestive association observed for H1b and increased LOAD risk (OR = 1.05,  $p = 0.089$ ) and for H1d and reduced LOAD risk (OR = 0.91,  $p = 0.074$ ). H1c subhaplotype did not achieve significance in the Mayo Clinic, ADGC or Mayo + ADGC cohorts.

#### Association of *MAPT* single SNPs and haplotypes with gene expression levels

In our published eGWAS [22], there were three probes on the WG-DASL platform that were used to measure *MAPT* levels: ILMN\_1710903 and ILMN\_2310814 that anneal to different regions of the *MAPT* 3'UTR and ILMN\_2298727 that anneals to Exon 4a (Additional file 1: Figure S1). Given that the inclusion of exon 4a in tau transcripts in the central nervous system was not reported previously, we generated a quantitative PCR assay against this exon, and were able to successfully measure it in the human brain (data not shown). All three probes passed our QC threshold of detectability in >75% of subjects, with ILMN\_1710903 and ILMN\_2310814 detected in 100% of all AD brains tested in both the cerebellum (CER) and temporal cortex (TCX) and with ILMN\_2298727 detectable in 98.0% of AD CER and 83.7% of AD TCX tissue. We previously estimated intraclass coefficients [28] for all gene expression probes, which represent the percentage of variance in expression between samples over total variance and which reflect the genetic component that contributes to variability in gene expression. We determined that both ILMN\_2298727 and ILMN\_1710903 had high ICC estimates of 87%, whereas ILMN\_2310814 had a low ICC estimate of 18%. The variances of gene expression estimated from all subjects in our eGWAS of cerebellar tissue ( $n = 374$ ) [22] revealed consistent findings for these three *MAPT* probes, with both ILMN\_2298727 (0.24) and ILMN\_1710903 (0.12) having variance estimates that are ~ an order of magnitude greater than that of ILMN\_2310814 (0.03). We thus conclude that ILMN\_2310814 is unlikely to be an informative probe.

We previously annotated all our probes for variants in their sequence [22], given the concern that such variants may result in differential binding of probes with artifactual variance in the expression levels, and therefore could result in false positive associations with genetic variants in LD with probe sequence variants [36,37]. Our annotation detected two variants within the probe sequence of ILMN\_1710903 (rs67759530, rs66561280) that were also polymorphic in our autopsied AD series. ILMN\_2310814 did not have any variants within its probe sequence.

ILMN\_2298727 annotation identified rs73314997 within its sequence, although this variant was essentially monomorphic in our eGWAS subjects [22]. Thus, of the three *MAPT* probes assessed in our gene expression analyses, ILMN\_2310814 is unlikely to be informative and ILMN\_1710903 may be prone to artifactual results. We therefore focused on ILMN\_2298727 in our *MAPT* expression analyses (Tables 3 and 4), although we show results from all 3 *MAPT* probes for completeness.

Evaluation of the six *MAPT* SNPs revealed significant associations between ILMN\_2298727 and rs1467967, rs242557, rs8070723 and rs7521. The *MAPT* H2 haplotype tagging rs8070723 was associated with lower *MAPT* levels in both CER ( $\beta = -0.16$ ,  $p = 0.002$ ) and TCX ( $\beta = -0.20$ ,  $p = 4.9E-04$ ) of LOAD subjects (Table 3), as we previously reported in this cohort [22]. The other significant variants were associated with higher *MAPT* levels in both brain regions. Interestingly, the same variants showed associations in the same direction with the ILMN\_1710903 probe, although with higher levels of significance.

Fifteen *MAPT* haplotypes with frequencies >1% were identified in the autopsied LOAD subjects with complete genotypes for the 6 variants ( $n = 178$ ). There was globally significant haplotype association with the TCX gene expression levels measured with ILMN\_2298727 ( $p = 0.004$ ) (Table 4), that may be a reflection of the significant *MAPT* H2 association. *MAPT* H2 haplotype, as expected, was associated with lower CER ( $\beta = -0.16$ ,  $p = 0.003$ ) and TCX ( $\beta = -0.20$ ,  $p = 0.001$ ) *MAPT* levels. *MAPT* H1b was marginally associated with higher TCX levels ( $\beta = 0.13$ ,  $p = 0.058$ ), I with higher CER *MAPT* levels ( $\beta = 0.20$ ,  $p = 0.07$ ), and L with lower TCX *MAPT* levels ( $\beta = -0.33$ ,  $p = 0.009$ ). Significant associations with similar directions of effect were also observed with ILMN\_1710903 and *MAPT* H2, H1b and I haplotypes.

#### Discussion

In this largest to date evaluation of haplotypic variation at the *MAPT* locus in 9,814 LOAD cases and 11,550 controls, we find robust and replicable association of the *MAPT* H2 haplotype with reduced risk of LOAD or, equivalently, increased risk of LOAD with the *MAPT* H1 haplotype- in two independent cohorts from Mayo Clinic and ADGC, with similar effect size estimates. Most prior reports of haplotypic association identified LOAD risk conferred by *MAPT* H1c subhaplotype [10-12], which we were unable to replicate. One group identified an association between the *MAPT* H1 haplotype and an increased risk for amnesic mild cognitive impairment [38], which can be a prodrome to clinical AD. The only other study to evaluate *MAPT* in a large cohort (3,940 cases and 13,373 controls) also identified an association between the H2 haplotype and decreased LOAD risk [9]. In that study by

**Table 3** *MAPT* single SNPs association results with brain *MAPT* gene expression levels

SNP	Brain region	N	ILMN_1710903		ILMN_2298727		ILMN_2310814	
			BETA	P	BETA	P	BETA	P
rs1467967	CER	166	0.16	<b>9.3E-06</b>	0.11	<b>0.011</b>	0.00	0.859
	TCX	171	0.15	<b>2.6E-04</b>	0.12	<b>0.019</b>	-0.01	0.604
rs242557	CER	173	0.17	<b>1.2E-05</b>	0.08	<b>0.091</b>	0.01	0.584
	TCX	180	0.23	<b>3.4E-08</b>	0.10	<b>0.050</b>	-0.02	0.196
rs3785883	CER	175	0.14	<b>0.005</b>	0.00	0.977	-0.01	0.621
	TCX	181	0.03	0.548	-0.07	0.266	0.01	0.706
rs2471738	CER	176	0.07	0.101	0.02	0.665	0.02	0.386
	TCX	182	0.11	<b>0.016</b>	0.09	<b>0.089</b>	-0.01	0.397
rs8070723	CER	174	-0.44	<b>2.1E-30</b>	-0.16	<b>0.002</b>	0.01	0.666
	TCX	181	-0.48	<b>8.9E-31</b>	-0.20	<b>4.9E-04</b>	0.02	0.222
rs7521	CER	176	0.12	<b>2.8E-04</b>	0.08	<b>0.048</b>	0.00	0.897
	TCX	182	0.16	<b>1.1E-05</b>	0.08	<b>0.084</b>	0.00	0.976

Results of multivariable linear regression analyses are shown. Probes ILMN\_1710903 and ILMN\_2310814 anneal to the 3'UTR sequence and ILMN\_2298727 targets exon 4a. CER = cerebellum, TCX = temporal cortex. Beta: Coefficient of association with the minor allele. P = p-value. Boldface values within the tables indicate significant or suggestive associations with a p-value <0.10.

Gerrish et al. [9] the H2-haplotype tagging SNP had an OR estimate of 0.89 ( $p = 5.20E-04$ ), which is remarkably similar to the estimate of the H2-tagging SNP (OR = 0.90,  $p = 5.3E-05$ ) and H2 haplotype (OR = 0.90,  $p = 1.53E-04$ ) in our study. It should be noted that both the present study and Gerrish et al. included samples from the ADNI and TGen series. We confirmed that the *MAPT* H2 association retains its significance in the ADGC cohort even after removal of these two datasets (OR = 0.87,  $p = 6.1E-04$ ). Thus, there is evidence of *MAPT* H2 association with reduced risk of LOAD in two large and independent studies. Though robust, this LOAD risk association does not achieve genome wide significance in either study or a  $p$  value <  $1.0E-7$  in the recent meta-analysis of 74,046 individuals by the IGAP consortium [39]. It will be important to evaluate the IGAP dataset for availability of *MAPT* haplotype tagging variants and to pursue an in-depth analysis of haplotype association at this locus.

Although the *MAPT* H2 haplotypic association with LOAD was clearly the strongest of the *MAPT* haplotypes and one that we previously reported [22], we identified additional SNPs and haplotypes with nominal significance in our study. These weaker associations would not withstand multiple testing and could represent false positives and require replication in additional series. It should be noted that some of these variants, such as rs3785883, H1b, H1d and J showed consistent direction of effect in the Mayo Clinic and ADGC cohorts. *MAPT* rs3785883 minor allele was previously shown to associate with higher levels of CSF tau, phospho-tau and earlier age at onset [16]. Although this prior smaller study did not identify association with LOAD risk, the biological effect of this variant which associates with increased LOAD risk

in our study appears to be consistent between these two studies.

We and others previously reported association between *MAPT* haplotypes and brain *MAPT* levels [11,19-22]. In this study, we evaluated *MAPT* subhaplotypes for association with brain *MAPT* levels in two brain regions from LOAD subjects. The most robust gene expression association occurs with the H2 haplotype, as we had reported [22] ( $\beta_{CER} = -0.16$ ,  $p_{CER} = 0.003$ ;  $\beta_{TCX} = -0.20$ ,  $p_{TCX} = 0.001$ ), that also has the strongest association with LOAD risk in our study. We find that this haplotype with a protective effect on LOAD associates with lower brain *MAPT* levels. Given multiple *MAPT* alternatively spliced exons leading to multiple transcripts, each with potentially different effects on function [1,20], uncovering the precise regulatory change associated with genotypic variation in this region is critical. In our study, we mainly focus on results from one probe, ILMN\_2298727, that is both informative and does not have a variant in its sequence based on annotation and genotyping. This probe is expected to anneal to exon 4a, however the expression levels obtained from it can be a surrogate for total *MAPT* levels or levels of any of the alternatively spliced exon-containing transcripts that reside with exon 4a. Indeed, gene expression associations with this probe are consistent with those from ILMN\_1710903, which should recognize all transcripts, although ILMN\_1710903 is confounded by a confirmed variant within its sequence. Our findings are also congruous with prior reports of associations of H1 haplotype or H1c sub-haplotype with higher 4R [19] and or total *MAPT* [11] levels, as measured by alternative gene expression measurement methods.



**Table 4 MAPT haplotype association results with brain MAPT gene expression levels**

Haplotype	Alleles	Brain region	Freq	ILMN_1710903		ILMN_2298727		ILMN_2310814	
				BETA	P	BETA	P	BETA	P
A (H2a)	AGGCGG	CER	0.212	-0.45	<b>8.7E-33</b>	-0.16	<b>0.003</b>	0.00	0.832
		TCX	0.212	-0.49	<b>1.1E-31</b>	-0.20	<b>0.001</b>	0.02	0.308
B (H1b)	GGGCAA	CER	0.180	0.15	<b>0.003</b>	0.07	0.207	-0.01	0.691
		TCX	0.180	0.13	<b>0.026</b>	0.13	<b>0.058</b>	0.01	0.630
C (H1c)	AAGTAG	CER	0.104	0.11	0.057	0.02	0.807	0.02	0.338
		TCX	0.104	0.17	<b>0.008</b>	0.05	0.536	-0.02	0.309
D (H1d)	AAGCAA	CER	0.057	0.00	0.970	0.04	0.721	0.07	0.098
		TCX	0.057	0.30	<b>0.007</b>	0.14	0.314	-0.02	0.646
E (H1e)	AGGCAA	CER	0.079	0.09	0.273	0.04	0.621	0.02	0.554
		TCX	0.079	0.15	0.079	0.13	0.212	0.00	0.964
H	AGACAA	CER	0.025	0.05	0.748	-0.13	0.478	0.04	0.538
		TCX	0.025	0.13	0.424	-0.05	0.803	-0.05	0.353
I	GAGCAA	CER	0.059	0.24	<b>0.010</b>	0.20	<b>0.070</b>	-0.04	0.322
		TCX	0.059	0.07	0.422	-0.09	0.434	-0.01	0.766
L	AGACAG	CER	0.043	0.17	0.062	0.01	0.921	-0.02	0.634
		TCX	0.043	-0.01	0.954	-0.33	<b>0.009</b>	0.01	0.841
M	GAGCAG	CER	0.028	0.26	<b>0.043</b>	0.01	0.936	0.01	0.829
		TCX	0.028	0.16	0.298	-0.23	0.204	0.00	0.927
O	AAACAA	CER	0.037	0.19	0.104	0.00	0.985	-0.04	0.435
		TCX	0.037	0.20	0.070	0.21	0.110	0.00	0.933
P	GGGTAG	CER	0.019	0.04	0.765	0.07	0.637	0.02	0.731
		TCX	0.019	0.07	0.725	0.26	0.259	-0.03	0.613
T	AGATAG	CER	0.016	0.01	0.935	-0.24	0.212	-0.07	0.285
		TCX	0.016	0.00	0.994	0.47	0.075	0.02	0.729
U	AAGCAG	CER	0.029	0.15	0.251	0.04	0.786	-0.14	<b>0.008</b>
		TCX	0.029	0.35	<b>0.008</b>	0.25	0.112	-0.01	0.836
X	GAATAG	CER	0.025	0.10	0.406	0.19	0.168	0.06	0.248
		TCX	0.025	-0.04	0.756	0.11	0.510	0.02	0.584
Y*	AAATAG	CER	0.022	-0.02	0.884	-0.13	0.458	-0.02	0.761
		TCX	0.022	-0.03	0.841	-0.03	0.882	0.00	0.933
		Global p		<b>3.0E-42</b>		0.352		0.271	
		TCX		<b>3.2E-36</b>		<b>0.004</b>		0.999	

Haplotypes with frequencies > 1% are assessed with multivariable linear regression analysis. Alleles for the SNPs defining the haplotypes are given in the 5' to 3' order as follows: rs1467967, rs242557, rs3785883, rs2471738, rs8070723, rs7521. Haplotypes not previously observed are designated by an asterisk (\*). CER = cerebellum, TCX = temporal cortex. Beta: Coefficient of association with the minor allele. P = p-value. Boldface values within the tables indicate significant or suggestive associations with a p-value < 0.10.

We did not identify significant associations between the H1c subhaplotype and brain *MAPT* levels, though we did observe suggestive associations between both CER and TCX *MAPT* levels and rs242557, a variant that partially tags H1c. *MAPT* rs1467967 associated with significant *MAPT* elevations in both brain regions and a suggestive association with LOAD risk, which is biologically consistent. These and additional weaker gene

expression associations with variants such as H1b, I, L and rs7521 requires further replications.

In summary, our study provides evidence of robust LOAD risk and brain *MAPT* level associations with *MAPT* H2 haplotype and nominates additional variants and subhaplotypes for further investigations in LOAD. The overall genetic contribution of *MAPT* variants to LOAD risk appears to be modest, in contrast to primary

tauopathies, where the H1 haplotype, for example, has an estimated OR of 5.5 from the PSP GWAS [8]. This may be due to different sets of functional variants residing in the same haplotypic backbone and leading to different biological outcomes resulting either in a primary tauopathy vs. tau pathology in LOAD; a more complex genetic architecture in LOAD with contribution from multiple functional variants in different pathways; or a combination of both. Discovering the precise sets of *MAPT* functional variants; and assessing their biologic consequence, especially on transcriptional regulation, may be critical to deciphering the commonalities and distinctions in the etiology of LOAD vs. primary tauopathies. Our study highlights the importance of in-depth association of *MAPT* haplotypic variation in well-powered cohorts and nominates H2 and additional variants as LOAD risk factors with effects on gene expression. Larger scale *MAPT* haplotype LOAD risk association studies, variant discovery efforts targeting specific haplotypes and transcriptional studies that jointly evaluate haplotypes and specific transcripts are warranted.

## Conclusions

In summary, these findings confirm associations between *MAPT* H2 haplotype and both reduced risk of LOAD and lower *MAPT* transcript brain levels. In addition, we describe additional *MAPT* variants and subhaplotypes that associate with LOAD risk and/or brain *MAPT* levels, which require confirmation in additional series. These results highlight the importance of joint utilization of gene expression and disease risk phenotypes. Additionally, these biologically consistent findings should encourage screening efforts in the *MAPT* region for discovery of regulatory variants that confer LOAD risk via influencing brain levels of *MAPT* transcripts.

## Additional file

**Additional file 1:** This file includes **Table S1**. (Demographic information of the cohorts); **Table S2**. (Genotype counts, call rates and Hardy Weinberg results); **Table S3**. (*MAPT* single SNP association results with LOAD risk in the individual Mayo Clinic series. Results of multivariable logistic regression analysis); **Figure S1**. (*MAPT* Refseq mRNA isoforms and SNP annotation).

## Abbreviations

A $\beta$ : Amyloid-beta; AD: Alzheimer's disease; ADGC: Alzheimer's disease genetics consortium; CBD: Corticobasal degeneration; CSF: Cerebrospinal fluid; CER: Cerebellum; eGWAS: Gene expression genome-wide association study; GWAS: Genome-wide association study; HWE: Hardy-Weinberg equilibrium; ICC: Intraclass coefficient; JS: Mayo Clinic Jacksonville, FL series; LD: Linkage disequilibrium; LOAD: Late-onset Alzheimer's disease; MAPT: Microtubule-associated protein tau; NFT: Neurofibrillary tangles; OR: Odds ratio; PSP: Progressive supranuclear palsy; QC: Quality control; RS: Mayo Clinic Rochester, MN series; SNP: Single nucleotide polymorphism; TCX: Temporal cortex.

## Competing interests

R.C. Petersen, M.D., Ph.D. has been a consultant to GE Healthcare and Elan Pharmaceuticals, has served on a data safety monitoring board in clinical trials sponsored by Pfizer Incorporated and Janssen Alzheimer Immunotherapy and gave a CME lecture at Novartis Incorporated. N. Graff-Radford, M.D. has served as a consultant to Codman and received grant support from Elan Pharmaceutical Research, Pfizer Pharmaceuticals, Medivation, and Forrest.

## Authors' contributions

MA conceived and designed the experiments, performed the experiments, analyzed the data and wrote the paper, MK performed the experiments, ZQ performed the experiments, FZ performed the experiments, HSC analyzed the data, CSY analyzed the data, JC analyzed the data, VSP analyzed the data, MMC performed the experiments, MMC performed the experiments, SK performed the experiments, TN performed the experiments, LM performed the experiments, KGM performed the experiments, SL performed the experiments, GB performed the experiments, CPK performed the experiments, JJ performed the experiments, SM contributed reagents/materials/analysis tools, JKK contributed reagents/materials/analysis tools, PKC contributed reagents/materials/analysis tools, JLH contributed reagents/materials/analysis tools, RM contributed reagents/materials/analysis tools, MAPV contributed reagents/materials/analysis tools, LAF contributed reagents/materials/analysis tools, GDS contributed reagents/materials/analysis tools, ADGC contributed reagents/materials/analysis tools, JEP contributed reagents/materials/analysis tools, RCP contributed reagents/materials/analysis tools, NRGR contributed reagents/materials/analysis tools, DWD contributed reagents/materials/analysis tools, SGY contributed reagents/materials/analysis tools, NET conceived and designed the experiments, analyzed the data and wrote the paper. All authors read and approved the final manuscript.

## ADGC acknowledgements

The National Institutes of Health, National Institute on Aging (NIH-NIA) supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; NACC, U01 AG016976; NCRAD, U24 AG021886; NIA LOAD, U24 AG026395, U24 AG026390; Banner Sun Health Research Institute P30 AG019610; Boston University, P30 AG013846, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG017173, R01 AG025259, R01AG33193; Columbia University, P50 AG008702, R37 AG015473; Duke University, P30 AG028377, AG05128; Emory University, AG025688; Group Health Research Institute, UO1 AG06781, UO1 HG004610; Indiana University, P30 AG10133; Johns Hopkins University, P50 AG005146, R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG0016574, U01 AG006786; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, MO1RR00096, UL1 RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, 1R01AG035137; Northwestern University, P30 AG013854; Oregon Health & Science University, P30 AG008017, R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, R01 AG30146; TGen, R01 NS059873; University of Alabama at Birmingham, P50 AG016582, UL1RR02777; University of Arizona, R01 AG031581; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573, P50, P50 AG016575, P50 AG016576, P50 AG016577; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501, P01 AG019724; University of Kentucky, P30 AG028383, AG05144; University of Michigan, P50 AG008671; University of Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, AG019757; University of Washington, P50 AG005136; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681, P01 AG03991. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by NINDS grant # NS39764, NIMH MH60451 and by Glaxo Smith Kline. Genotyping of the TGEN2 cohort was supported by Kronos Science. The TGen series was also funded by NIA grant AG034504 to AJM, The Banner Alzheimer's Foundation, The Johnnie B. Byrd Sr. Alzheimer's Institute, the Medical Research Council, and the state of Arizona and also includes samples from the following sites: Newcastle Brain Tissue Resource (funding via the Medical Research Council, local NHS trusts and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council), South West Dementia Brain Bank (funding via numerous sources including the Higher Education Funding Council for

England (HEFCE), Alzheimer's Research Trust (ART), BRACE as well as North Bristol NHS Trust Research and Innovation Department and DeNDRoN), The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stichting Alzheimer Onderzoek), Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona. ADNI Funding for ADNI is through the Northern California Institute for Research and Education by grants from Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., Alzheimer's Association, Alzheimer's Drug Discovery Foundation, the Dana Foundation, and by the National Institute of Biomedical Imaging and Bioengineering and NIA grants U01 AG024904, RC2 AG036535, K01 AG030514. We thank Drs. D. Stephen Snyder and Marilyn Miller from NIA who are *ex-officio* ADGC members. Support was also from the Alzheimer's Association (LAF, IIRG-08-89720; MP-V, IIRG-05-14147) and the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program. P.S.G.-H. is supported by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health Research.

#### ADGC affiliations

<sup>1</sup>Department of Neurology, Johns Hopkins University, Baltimore, Maryland, <sup>2</sup>Department of Neurology, University of Michigan, Ann Arbor, Michigan, <sup>3</sup>Geriatric Research, Education and Clinical Center (GRECC), VA Ann Arbor Healthcare System (VAAAHS), Ann Arbor, Michigan, <sup>4</sup>Michigan Alzheimer Disease Center, Ann Arbor, Michigan, <sup>5</sup>Department of Neurology, University of California Los Angeles, Los Angeles, California, <sup>6</sup>Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, <sup>7</sup>Geriatric Research, Education and Clinical Center (GRECC), University of Wisconsin, Madison, Wisconsin, <sup>8</sup>Department of Medicine, University of Wisconsin, Madison, Wisconsin, <sup>9</sup>Wisconsin Alzheimer's Institute, Madison, Wisconsin, <sup>10</sup>Department of Medicine (Genetics Program), Boston University, Boston, Massachusetts, <sup>11</sup>Department of Pharmacology and Neuroscience, University of North Texas Health Science Center, Fort Worth, Texas, <sup>12</sup>Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, <sup>13</sup>Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, <sup>14</sup>Department of Behavioral Sciences, Rush University Medical Center, Chicago, Illinois, <sup>15</sup>Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Phoenix, Arizona, <sup>16</sup>Departments of Psychiatry, Neurology, and Psychology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, <sup>17</sup>The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, <sup>18</sup>Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida, <sup>19</sup>National Alzheimer's Coordinating Center, University of Washington, Seattle, Washington, <sup>20</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, <sup>21</sup>Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, <sup>22</sup>Cognitive Neurology and Alzheimer's Disease Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois, <sup>23</sup>Department of Neurology, University of Washington, Seattle, Washington, <sup>24</sup>VA Puget Sound Health Care System/GRECC, Seattle, Washington, <sup>25</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, <sup>26</sup>Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, <sup>27</sup>Department of Neurology, Mayo Clinic, Rochester, Minnesota, <sup>28</sup>Swedish Medical Center, Seattle, Washington, <sup>29</sup>Department of Neurology, University of California San Francisco, San Francisco, California, <sup>30</sup>Department of Medicine, Duke University, Durham, North Carolina, <sup>31</sup>Department of Neuroscience, Mount Sinai School of Medicine, New York, New York, <sup>32</sup>Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, <sup>33</sup>Departments of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, <sup>34</sup>Department of Pathology and Immunology, Washington University, St. Louis, Missouri, <sup>35</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, <sup>36</sup>USF Health Byrd Alzheimer's Institute, University of South Florida, Tampa, Florida, <sup>37</sup>Fred Hutchinson Cancer Research Center, Seattle, Washington, <sup>38</sup>Department of Psychiatry and Behavioral Sciences,

Miller School of Medicine, University of Miami, Miami, Florida, <sup>39</sup>Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama, <sup>40</sup>Department of Neurology, University of Southern California, Los Angeles, California, <sup>41</sup>Department of Neurology, University of Alabama at Birmingham, Birmingham, Alabama, <sup>42</sup>Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona, <sup>43</sup>Department of Medicine, University of Washington, Seattle, Washington, <sup>44</sup>Department of Neurology, University of California Irvine, Irvine, California, <sup>45</sup>Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, Missouri, <sup>46</sup>Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology & Psychiatry, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, <sup>47</sup>Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, <sup>48</sup>Department of Neurology, University of California Davis, Sacramento, California, <sup>49</sup>University of Virginia School of Medicine, Charlottesville, Virginia, <sup>50</sup>Institute for Memory Impairments and Neurological Disorders, University of California Irvine, Irvine, California, <sup>51</sup>Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, Florida, <sup>52</sup>Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, <sup>53</sup>Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, <sup>54</sup>Department of Neurology, Indiana University, Indianapolis, Indiana, <sup>55</sup>Department of Psychiatry, New York University, New York, New York, <sup>56</sup>C.S. Kubik Laboratory for Neuropathology, Massachusetts General Hospital, Charlestown, Massachusetts, <sup>57</sup>Department of Neurosciences, University of California San Diego, La Jolla, California, <sup>58</sup>Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, <sup>59</sup>Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, <sup>60</sup>Emory Alzheimer's Disease Center, Emory University, Atlanta, Georgia, <sup>61</sup>Neurogenetics Program, University of California Los Angeles, Los Angeles, California, <sup>62</sup>Department of Pathology and Laboratory Medicine, Indiana University, Indianapolis, Indiana, <sup>63</sup>Department of Neurology, Emory University, Atlanta, Georgia, <sup>64</sup>Division of Genetics, Department of Medicine and Partners Center for Personalized Genetic Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, <sup>65</sup>Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, <sup>66</sup>Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, <sup>67</sup>Department of Pathology (Neuropathology), University of Pittsburgh, Pittsburgh, Pennsylvania, <sup>68</sup>Institute of Neurology, University College London, Queen Square, London, <sup>69</sup>Sanders-Brown Center on Aging, Department of Molecular and Biomedical Pharmacology, University of Kentucky, Lexington, Kentucky, <sup>70</sup>Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, New York, <sup>71</sup>Department of Pathology, Duke University, Durham, North Carolina, <sup>72</sup>Department of Genome Sciences, University of Washington, Seattle, Washington, <sup>73</sup>Department of Medicine (Medical Genetics), University of Washington, Seattle, Washington, <sup>74</sup>Sanders-Brown Center on Aging, Department of Neurology, University of Kentucky, Lexington, Kentucky, <sup>75</sup>Department of Pathology and Laboratory Medicine, University of California Davis, Sacramento, California, <sup>76</sup>Department of Biostatistics, Boston University, Boston, Massachusetts, <sup>77</sup>Department of Ophthalmology, Boston University, Boston, Massachusetts, <sup>78</sup>University of Pittsburgh Alzheimer's Disease Research Center, Pittsburgh, Pennsylvania, <sup>79</sup>Department of Neurology, Oregon Health & Science University, Portland, Oregon, <sup>80</sup>Department of Neurology, Portland Veterans Affairs Medical Center, Portland, Oregon, <sup>81</sup>Department of Pathology and Laboratory Medicine, University of California Irvine, Irvine, California, <sup>82</sup>Department of Neurology, Boston University, Boston, Massachusetts, <sup>83</sup>Department of Pathology, Boston University, Boston, Massachusetts, <sup>84</sup>Department of Neuropsychology, University of California San Francisco, San Francisco, California, <sup>85</sup>Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, Oregon, <sup>86</sup>Department of Epidemiology, University of Washington, Seattle, Washington, <sup>87</sup>Department of Neurobiology and Behavior, University of California Irvine, Irvine, California, <sup>88</sup>Group Health Research Institute, Group Health, Seattle, Washington, <sup>89</sup>Department of Pathology, University of Washington, Seattle, Washington, <sup>90</sup>Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington, <sup>91</sup>Department of Pathology, University of Michigan, Ann Arbor, Michigan, <sup>92</sup>Department of Psychiatry, Johns Hopkins University, Baltimore, Maryland, <sup>93</sup>Department of Preventive Medicine, University of Southern California, Los Angeles, California, <sup>94</sup>Department of Medicine - Pulmonary, New York University, New York, New York, <sup>95</sup>Department of Neurology, University of Miami, Miami, Florida, <sup>96</sup>Department of Pathology, University of

California San Diego, La Jolla, California,<sup>97</sup>School of Nursing Northwest Research Group on Aging, University of Washington, Seattle, Washington,<sup>98</sup>Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, Illinois,<sup>99</sup>Department of Pathology, University of Southern California, Los Angeles, California,<sup>100</sup>Department of Neurology, Washington University, St. Louis, Missouri,<sup>101</sup>Arizona Alzheimer's Consortium, Phoenix, Arizona,<sup>102</sup>Department of Psychiatry, University of Arizona, Phoenix, Arizona,<sup>103</sup>Banner Alzheimer's Institute, Phoenix, Arizona,<sup>104</sup>Alzheimer's Disease Center, New York University, New York, New York,<sup>105</sup>Gertrude H. Sergievsky Center, Columbia University, New York, New York,<sup>106</sup>Department of Neurology, Columbia University, New York, New York,<sup>107</sup>Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, Ontario,<sup>108</sup>Department of Neurology, University of Texas Southwestern, Dallas, Texas,<sup>109</sup>Department of Radiology and Imaging Sciences, Indiana University, Indianapolis, Indiana,<sup>110</sup>Department of Pathology (Neuropathology), Rush University Medical Center, Chicago, Illinois,<sup>111</sup>Department of Psychiatry, University of Southern California, Los Angeles, California,<sup>112</sup>Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge,<sup>113</sup>Center for Human Genetics and Research, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee,<sup>114</sup>Department of Pathology, Johns Hopkins University, Baltimore, Maryland,<sup>115</sup>Sanders-Brown Center on Aging, Department of Anatomy and Neurobiology, University of Kentucky, Lexington, Kentucky,<sup>116</sup>Department of Pathology & Laboratory Medicine, University of California Los Angeles, Los Angeles, California,<sup>117</sup>Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Pathology, Columbia University, New York, New York,<sup>118</sup>Department of Psychiatry, Northwestern University Feinberg School of Medicine, Chicago, Illinois,<sup>119</sup>Department of Psychiatry & Behavioral Sciences, Duke University, Durham, North Carolina,<sup>120</sup>Department of Pathology, Oregon Health & Science University, Portland, Oregon,<sup>121</sup>Evelyn F. McKnight Brain Institute, Department of Neurology, Miller School of Medicine, University of Miami, Miami, Florida.

#### ADGC co-authors

Marilyn S. Albert<sup>1</sup>, Roger L. Albin<sup>2-4</sup>, Liana G. Apostolova<sup>5</sup>, Steven E. Arnold<sup>6</sup>, Sanjay Asthana<sup>7-9</sup>, Craig S. Atwood<sup>7,9</sup>, Clinton T. Baldwin<sup>10</sup>, Robert Barber<sup>11</sup>, Michael M. Barmada<sup>12</sup>, Lisa L. Barnes<sup>13,14</sup>, Thomas G. Beach<sup>15</sup>, James T. Becker<sup>16</sup>, Gary W. Beecham<sup>17,18</sup>, Duane Beekly<sup>19</sup>, David A. Bennett<sup>13,20</sup>, Eileen H. Bigio<sup>21,22</sup>, Thomas D. Bird<sup>23,24</sup>, Deborah Blacke<sup>25,26</sup>, Bradley F. Boeve<sup>27</sup>, James D. Bowen<sup>28</sup>, Adam Boxer<sup>29</sup>, James R. Burke<sup>30</sup>, Joseph D. Buxbaum<sup>31-33</sup>, Nigel J. Cairns<sup>34</sup>, Laura B. Cantwell<sup>35</sup>, Chuanhai Cao<sup>36</sup>, Chris S. Carlson<sup>37</sup>, Cynthia M. Carlsson<sup>8</sup>, Regina M. Carney<sup>38</sup>, Steven L. Carroll<sup>39</sup>, Helena C. Chui<sup>40</sup>, David G. Clark<sup>41</sup>, Jason Corneveaux<sup>42</sup>, David H. Cribbs<sup>44</sup>, Elizabeth A. Crocco<sup>38</sup>, Carlos Cruchaga<sup>45</sup>, Philip L. De Jager<sup>46,47</sup>, Charles DeCarli<sup>48</sup>, Steven T. DeKosky<sup>49</sup>, F. Yesim Demirci<sup>12</sup>, Malcolm Dick<sup>50</sup>, Ranjan Duara<sup>51</sup>, Denis Evans<sup>52</sup>, Kelley M. Faber<sup>53</sup>, Kenneth B. Fallon<sup>39</sup>, Martin R. Farlow<sup>54</sup>, Steven Ferris<sup>55</sup>, Tatiana M. Foroud<sup>53</sup>, Matthew P. Frosch<sup>56</sup>, Douglas R. Galasko<sup>57</sup>, Mary Ganguli<sup>58</sup>, Marla Gearing<sup>59,60</sup>, Daniel H. Geschwind<sup>61</sup>, Bernardino Ghetti<sup>62</sup>, John R. Gilbert<sup>17,18</sup>, Jonathan D. Glass<sup>63</sup>, Alison M. Goate<sup>45</sup>, Robert C. Green<sup>64</sup>, John H. Growdon<sup>65</sup>, Hakon Hakonarson<sup>66</sup>, Ronald L. Hamilton<sup>67</sup>, Kara L. Hamilton-Nelson<sup>17</sup>, John Hardy<sup>68</sup>, Lindy E. Harrell<sup>41</sup>, Elizabeth Head<sup>69</sup>, Lawrence S. Honig<sup>70</sup>, Matthew J. Huentelman<sup>42</sup>, Christine M. Hulette<sup>71</sup>, Bradley T. Hyman<sup>65</sup>, Gail P. Jarvik<sup>72,73</sup>, Gregory A. Jicha<sup>74</sup>, Lee-Way Jin<sup>75</sup>, Gyungah Jun<sup>10,76,77</sup>, M. Ilyas Kambh<sup>12,78</sup>, Anna Karydas<sup>29</sup>, Jeffrey A. Kaye<sup>79,80</sup>, Ronald Kim<sup>81</sup>, Edward H. Koo<sup>57</sup>, Neil W. Kowal<sup>82,83</sup>, Joel H. Kramer<sup>84</sup>, Patricia Kramer<sup>79,85</sup>, Walter A. Kukull<sup>86</sup>, Brian W. Kunkle<sup>17</sup>, Frank M. LaFerla<sup>87</sup>, James J. Lah<sup>63</sup>, Eric B. Larson<sup>43,88</sup>, James B. Leverenz<sup>39</sup>, Allan I. Levey<sup>63</sup>, Ge Li<sup>90</sup>, Andrew P. Lieberman<sup>91</sup>, Chiao-Feng Lin<sup>35</sup>, Oscar L. Lopez<sup>78</sup>, Kathryn L. Lunetta<sup>76</sup>, Constantine G. Lyketsos<sup>92</sup>, Wendy J. Mack<sup>93</sup>, Daniel C. Marson<sup>41</sup>, Eden R. Martin<sup>17,18</sup>, Frank Martiniuk<sup>94</sup>, Deborah C. Mash<sup>95</sup>, Eliezer Masliah<sup>57,96</sup>, Wayne C. McCormick<sup>43</sup>, Susan M. McCurry<sup>97</sup>, Andrew N. McDavid<sup>37</sup>, Ann C. McKee<sup>82,83</sup>, Marsel Mesulam<sup>22,98</sup>, Bruce L. Miller<sup>29</sup>, Carol A. Miller<sup>99</sup>, Joshua W. Miller<sup>75</sup>, Thomas J. Montine<sup>89</sup>, John C. Morris<sup>34,100</sup>, Jill R. Murrell<sup>53,62</sup>, Amanda J. Myers<sup>38</sup>, Adam C. Naj<sup>35</sup>, John M. Olichney<sup>48</sup>, Amanda Partch<sup>35</sup>, Henry L. Paulson<sup>2</sup>, William Perry<sup>17</sup>, Elaine Peskind<sup>90</sup>, Aimee Pierce<sup>44</sup>, Wayne W. Poon<sup>50</sup>, Huntington Potter<sup>36</sup>, Joseph F. Quinn<sup>79</sup>, Ashok Raj<sup>36</sup>, Murray Raskind<sup>90</sup>, Eric M. Reiman<sup>42,101-103</sup>, Barry Reisberg<sup>55,104</sup>, Christiane Reitz<sup>70,105,106</sup>, John M. Ringman<sup>5</sup>, Erik D. Robertson<sup>41</sup>, Ekaterina Rogava<sup>107</sup>, Howard J. Rosen<sup>29</sup>, Roger N. Rosenberg<sup>108</sup>, Mark A. Sager<sup>95</sup>, Mary Sano<sup>32</sup>, Andrew J. Saykin<sup>53,109</sup>, Julie A. Schneider<sup>13,110</sup>, Lon S. Schneider<sup>40,111</sup>, William W. Seeley<sup>29</sup>, Amanda G. Smith<sup>36</sup>, Joshua A. Sonnen<sup>89</sup>, Salvatore Spina<sup>62</sup>, Peter St George-Hyslop<sup>107,112</sup>, Robert A. Stern<sup>82</sup>, Rudolph E. Tanzi<sup>65</sup>, Tricia A. Thornton-Wells<sup>113</sup>, John Q. Trojanowski<sup>35</sup>, Juan C. Troncoso<sup>114</sup>, Debby W. Tsuang<sup>24,90</sup>, Otto Valladares<sup>35</sup>, Vивиanna M. Van Deerlin<sup>35</sup>, Linda J. Van

Eldik<sup>115</sup>, Badri N. Vardarajan<sup>70,105,106</sup>, Harry V. Vinters<sup>5,116</sup>, Jean Paul Vonsattel<sup>117</sup>, Li-San Wang<sup>35</sup>, Sandra Weintraub<sup>22,118</sup>, Kathleen A. Welsh-Bohmer<sup>30,119</sup>, Jennifer Williamson<sup>70</sup>, Sarah Wishek<sup>17</sup>, Randall L. Woljter<sup>20</sup>, Clinton B. Wright<sup>121</sup>, Chang-En Yu<sup>43</sup>, Lei Yu<sup>13</sup>.

#### ADGC disclosure statement

T.D.B. received licensing fees from and is on the speaker's bureau of Athena Diagnostics, Inc. M.R.F. receives research funding from BristolMyersSquibb Company, Danone Research, Elan Pharmaceuticals, Inc., Eli Lilly and Company, Novartis Pharmaceuticals Corporation, OctaPharma AG, Pfizer Inc., and Sonexa Therapeutics, Inc; Receives honoraria as scientific consultant from Accera, Inc., Astellas Pharma US Inc., Baxter, Bayer Pharmaceuticals Corporation, BristolMyersSquibb, Eisai Medical Research, Inc., GE Healthcare, Medavante, Medivation, Inc., Merck & Co., Inc., Novartis Pharmaceuticals Corp., Pfizer, Inc., Prana Biotechnology Ltd., QR Pharma, Inc., The sanofi-aventis Group, and Toyama Chemical Co, Ltd; and is speaker for Eisai Medical Research, Inc., Forest Laboratories, Pfizer Inc. and Novartis Pharmaceuticals Corporation. A. M.G. has research funding from AstraZeneca, Pfizer and Genentech, and has received remuneration for giving talks at Pfizer and Genentech. R.C.P. is on the Safety Monitoring Committee of Pfizer, Inc. (Wyeth) and a consultant to the Safety Monitoring Committee at Janssen Alzheimer's Immunotherapy Program (Elan), to Elan Pharmaceuticals, and to GE Healthcare. R.E.T. is a consultant to Eisai, Japan in the area of Alzheimer's genetics and a shareholder in, and consultant to Pathway Genomics, Inc, San Diego, CA.

#### Acknowledgements

Support for this research was provided by the National Institutes of Health grants: National Institute on Aging (R01 AG032990 to NET and R01 AG018023 to NRG-R and SGY); National Institutes on Neurologic Diseases and Stroke (R01 NS080820 to NET), Mayo Alzheimer's Disease Research Center: (P50 AG0016574 to RCP, DWD, NRG-R, SGY, and NET); Mayo Alzheimer's Disease Patient Registry: (U01 AG006576 to RCP); National Institute on Aging (AG025711, AG017216, AG003949 to DWD). This project was also generously supported by the Robert and Clarice Smith and Abigail Van Buren Alzheimer's Disease Research Program (to RCP, DWD, NRG-R, and SGY), and by the Palumbo Professorship in Alzheimer's Disease Research (to SGY). MMC and NET are supported partly by GHR Foundation grants. Combining the ADGC data sets was performed by SM, and was supported by R01 AG 029672 (to PKC), U01 HG 006375 (to Eric Larson), and U01 AG 006781 (to Eric Larson). We thank the patients and their families for their participation, without whom these studies will not have been possible.

#### Author details

<sup>1</sup>Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL 32224, USA. <sup>2</sup>Department of Health Sciences Research, Mayo Clinic Minnesota, Rochester, MN 55905, USA. <sup>3</sup>Department of Health Sciences Research, Mayo Clinic Florida, Jacksonville, FL 32224, USA. <sup>4</sup>Medical Genome Facility, Mayo Clinic Minnesota, Rochester, MN 55905, USA. <sup>5</sup>Department of Medicine, University of Washington, Seattle 98104, WA, USA. <sup>6</sup>Departments of Biology, Neuroscience, Brigham Young University, Provo, UT 84602, USA. <sup>7</sup>Department of Molecular Physiology and Biophysics, and the Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA. <sup>8</sup>Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA. <sup>9</sup>Gertrude H. Sergievsky Center, Department of Neurology, and Taub Institute on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY, USA. <sup>10</sup>The John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA. <sup>11</sup>Departments of Biostatistics, Medicine (Genetics Program), Ophthalmology, Neurology, and Epidemiology, Boston University, Boston, MA, USA. <sup>12</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>13</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA. <sup>14</sup>Department of Neurology, Mayo Clinic Minnesota, Rochester, MN 55905, USA. <sup>15</sup>Department of Neurology, Mayo Clinic Florida, 4500 San Pablo Road, Birdsall 3, Jacksonville, FL 32224, USA.

Received: 5 March 2014 Accepted: 28 May 2014

Published: 1 July 2014

## References

- Pittman AM, Fung HC, de Silva R: **Untangling the tau gene association with neurodegenerative disorders.** *Hum Mol Genet* 2006, **15**:R188–R195.
- Hutton M, Lendon AJ, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A, Hackett J, Adamson J, Lincoln S, Dickson D, Davies P, Petersen RC, Stevens M, de Graaf E, Wauters E, van Baren J, Hillebrand M, Joosse M, Kwon JM, Nowotny P, Che LK, Norton J, Morris JC, Reed LA, Trojanowski J, Basun H, et al: **Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.** *Nature* 1998, **393**:702–705.
- Spillantini MG, Murrell JR, Goedert M, Farlow MR, Klug A, Ghetti B: **Mutation in the tau gene in familial multiple system tauopathy with presenile dementia.** *Proc Natl Acad Sci U S A* 1998, **95**:7737–7741.
- Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, Hardy J, Lynch T, Bigio E, Hutton M: **Association of an extended haplotype in the tau gene with progressive supranuclear palsy.** *Hum Mol Genet* 1999, **8**:711–715.
- Houlden H, Baker M, Morris HR, MacDonald N, Pickering-Brown S, Adamson J, Lees AJ, Rossor MN, Quinn NP, Kertesz A, Khan MN, Hardy J, Lantos PL, St George-Hyslop P, Munoz DG, Mann D, Lang AE, Bergeron C, Bigio EH, Litvan I, Bhatia KP, Dickson D, Wood NW, Hutton M: **Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype.** *Neurology* 2001, **56**:1702–1706.
- Pittman AM, Myers AJ, Abou-Sleiman P, Fung HC, Kaleem M, Marlowe L, Duckworth J, Leung D, Williams D, Kilford L, Thomas N, Morris CM, Dickson D, Wood NW, Hardy J, Lees AJ, de Silva R: **Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration.** *J Med Genet* 2005, **42**:837–846.
- Rademakers R, Melquist S, Cruts M, Theuns J, Del-Favero J, Poorkaj P, Baker M, Sleegers K, Crook R, De Pooter T, Bel Kacem S, Adamson J, Van den Bossche D, Van den Broeck M, Gass J, Corsmit E, De Rijk P, Thomas N, Engelborghs S, Heckman M, Litvan I, Crook J, De Deyn PP, Dickson D, Schellenberg GD, Van Broeckhoven C, Hutton ML: **High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy.** *Hum Mol Genet* 2005, **14**:3281–3292.
- Hoglinger GU, Melhem NM, Dickson DW, Sleiman PM, Wang LS, Klei L, Rademakers R, de Silva R, Litvan I, Riley DE, van Swieten JC, Heutink P, Wszolek ZK, Uitti RJ, Vandrovova J, Hurtig HI, Gross RG, Maetzler W, Goldwurm S, Tolosa E, Borroni B, Pastor P, PSP Genetics Study Group, Cantwell LB, Han MR, Dillman A, van der Brug MP, Gibbs JR, Cookson MR, Hernandez DG, et al: **Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy.** *Nat Genet* 2011, **43**:699–705.
- Gerrish A, Russo G, Richards A, Moskvina V, Ivanov D, Harold D, Sims R, Abraham R, Hollingworth P, Chapman J, Hamshere M, Pahwa JS, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, et al: **The role of variation at AbetaPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease.** *J Alzheimers Dis* 2012, **28**:377–387.
- Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, Fung HC, Duckworth J, Leung D, Gibson A, Morris CM, de Silva R, Hardy J: **The H1c haplotype at the MAPT locus is associated with Alzheimer's disease.** *Hum Mol Genet* 2005, **14**:2399–2404.
- Myers AJ, Pittman AM, Zhao AS, Rohrer K, Kaleem M, Marlowe L, Lees A, Leung D, McKeith IG, Perry RH, Morris CM, Trojanowski JQ, Clark C, Karlawish J, Arnold S, Forman MS, Van Deerlin V, de Silva R, Hardy J: **The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts.** *Neurobiol Dis* 2007, **25**:561–570.
- Laws SM, Friedrich P, Diehl-Schmid J, Muller J, Eisele T, Bauml J, Forstl H, Kurz A, Riemenschneider M: **Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease.** *Mol Psychiatry* 2007, **12**:510–517.
- Abraham R, Sims R, Carroll L, Hollingworth P, O'Donovan MC, Williams J, Owen MJ: **An association study of common variation at the MAPT locus with late-onset Alzheimer's disease.** *Am J Med Genet B Neuropsychiatr Genet* 2009, **150B**:1152–1155.
- Russ C, Powell JF, Zhao J, Baker M, Hutton M, Crawford F, Mullan M, Roks G, Cruts M, Lovestone S: **The microtubule associated protein Tau gene and Alzheimer's disease—an association study and meta-analysis.** *Neurosci Lett* 2001, **314**:92–96.
- Mukherjee O, Kauwe JS, Mayo K, Morris JC, Goate AM: **Haplotype-based association analysis of the MAPT locus in late onset Alzheimer's disease.** *BMC Genet* 2007, **8**:3.
- Kauwe JS, Cruchaga C, Mayo K, Fenoglio C, Bertelsen S, Nowotny P, Galimberti D, Scarpini E, Morris JC, Fagan AM, Holtzman DM, Goate AM: **Variation in MAPT is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition.** *Proc Natl Acad Sci U S A* 2008, **105**:8050–8054.
- Buee L, Delacourte A: **Comparative biochemistry of tau in progressive supranuclear palsy, corticobasal degeneration, FTDP-17 and Pick's disease.** *Brain Pathol* 1999, **9**:681–693.
- Takanashi M, Mori H, Arima K, Mizuno Y, Hattori N: **Expression patterns of tau mRNA isoforms correlate with susceptible lesions in progressive supranuclear palsy and corticobasal degeneration.** *Brain Res Mol Brain Res* 2002, **104**:210–219.
- Caffrey TM, Joachim C, Paracchini S, Esiri MM, Wade-Martins R: **Haplotype-specific expression of exon 10 at the human MAPT locus.** *Hum Mol Genet* 2006, **15**:3529–3537.
- Caffrey TM, Joachim C, Wade-Martins R: **Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus.** *Neurobiol Aging* 2008, **29**:1923–1929.
- Trabzuni D, Wray S, Vandrovova J, Ramasamy A, Walker R, Smith C, Luk C, Gibbs JR, Dillman A, Hernandez DG, Arepalli S, Singleton AB, Cookson MR, Pittman AM, de Silva R, Weale ME, Hardy J, Ryten M: **MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies.** *Hum Mol Genet* 2012, **21**:4094–4103.
- Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, Carrasquillo MM, Rowley CN, Nair AA, Middha S, Maharjan S, Nguyen T, Ma L, Malphrus KG, Palusak R, Lincoln S, Bisceglia G, Georgescu C, Kouri N, Kolbert CP, Jen J, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD, Petersen RC, Graff-Radford NR, Dickson DW, Alzheimer's Disease Genetics Consortium, et al: **Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants.** *PLoS Genet* 2012, **8**:e1002707.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: **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.
- Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, Younkin SG, Younkin CS, Younkin LH, Bisceglia GD, Ertekin-Taner N, Crook JE, Dickson DW, Petersen RC, Graff-Radford NR, Younkin SG: **Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease.** *Nat Genet* 2009, **41**:192–198.
- Jun G, Naj AC, Beecham GW, Wang LS, Buros J, Gallins PJ, Buxbaum JD, Ertekin-Taner N, Fallin MD, Friedland R, Inzelberg R, Kramer P, Rogava E, St George-Hyslop P, Alzheimer's Disease Genetics C, Cantwell LB, Dombroski BA, Saykin AJ, Reiman EM, Bennett DA, Morris JC, Lunetta KL, Martin ER, Montine TJ, Goate AM, Blacker D, Tsuang DW, Beekly D, Cupples LA, Hakonarson H, et al: **Meta-analysis Confirms CR1, CLU, and PICALM as Alzheimer Disease Risk Loci and Reveals Interactions With APOE Genotypes.** *Arch Neurol* 2010, **67**:1473–1484.
- Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JS, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, et al: **Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease.** *Nat Genet* 2011, **43**:436–441.
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM: **Robust relationship inference in genome-wide association studies.** *Bioinformatics* 2010, **26**:2867–2873.
- Barry WT, Kernagis DN, Dressman HK, Griffis RJ, Hunter JD, Olson JA, Marks JR, Ginsburg GS, Marcom PK, Nevins JR, Geradts J, Datto MB: **Intratumor heterogeneity and precision of microarray-based predictors of breast cancer biology and clinical outcome.** *J Clin Oncol* 2010, **28**:2198–2206.

29. Du P, Kibbe WA, Lin SM: **lumi: a pipeline for processing Illumina microarray.** *Bioinformatics* 2008, **24**:1547–1548.
30. Lin SM, Du P, Huber W, Kibbe WA: **Model-based variance-stabilizing transformation for Illumina microarray data.** *Nucleic Acids Res* 2008, **36**:e11.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559–575.
32. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38**:904–909.
33. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA: **Score tests for association between traits and haplotypes when linkage phase is ambiguous.** *Am J Hum Genet* 2002, **70**:425–434.
34. Allen M, Zou F, Chai HS, Younkin CS, Miles R, Nair AA, Crook JE, Pankratz VS, Carrasquillo MM, Rowley CN, Nguyen T, Ma L, Malphrus KG, Bisceglia G, Ortolaza AI, Palusak R, Middha S, Maharjan S, Georgescu C, Schultz D, Rakhshan F, Kolbert CP, Jen J, Sando SB, Aasly JO, Barcikowska M, Uitti RJ, Wszolek ZK, Ross OA, Petersen RC, et al: **Glutathione S-transferase omega genes in Alzheimer and Parkinson disease risk, age-at-diagnosis and brain gene expression: an association study with mechanistic implications.** *Mol Neurodegener* 2012, **7**:13.
35. Roder HM, Hutton ML: **Microtubule-associated protein tau as a therapeutic target in neurodegenerative disease.** *Expert Opin Ther Targets* 2007, **11**:435–442.
36. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, Deloukas P, Dermitzakis ET: **Genome-wide associations of gene expression variation in humans.** *PLoS Genet* 2005, **1**:e78.
37. Doss S, Schadt EE, Drake TA, Lusis AJ: **Cis-acting expression quantitative trait loci in mice.** *Genome Res* 2005, **15**:681–691.
38. Di Maria E, Cammarata S, Parodi MI, Borghi R, Benussi L, Galli M, Galimberti D, Ghidoni R, Gonella D, Novello C, Pollero V, Perroni L, Odetti P, Scarpini E, Binetti G, Tabaton M: **The H1 haplotype of the tau gene (MAPT) is associated with mild cognitive impairment.** *J Alzheimers Dis* 2010, **19**:909–914.
39. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, Destefano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thornton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, Lin CF, Gerrish A, Schmidt H, Kunkle B, Dunstan ML, Ruiz A, Bihoreau MT, Choi SH, Reitz C, et al: **Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.** *Nat Genet* 2013, **45**:1452–1458.

doi:10.1186/alzrt268

**Cite this article as:** Allen et al.: Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels. *Alzheimer's Research & Therapy* 2014 **6**:39.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

