He et al. *Translational Psychiatry* (2021)11:146
https://doi.org/10.1038/s41398-021-01263-4

Translational Psychiatry

ARTICLE Open Access

Exome-wide age-of-onset analysis reveals exonic variants in *ERN1* and *SPPL2C* associated with Alzheimer's disease

Liang He[1], Yury Loika[1], Yongjin Park[2,3], Genotype Tissue Expression (GTEx) consortium, David A. Bennett[4], Manolis Kellis[2,3] and Alexander M. Kulminski[1], for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Despite recent discoveries in genome-wide association studies (GWAS) of genomic variants associated with Alzheimer's disease (AD), its underlying biological mechanisms are still elusive. The discovery of novel AD-associated genetic variants, particularly in coding regions and from *APOE ε4* non-carriers, is critical for understanding the pathology of AD. In this study, we carried out an exome-wide association analysis of age-of-onset of AD with ~20,000 subjects and placed more emphasis on *APOE ε4* non-carriers. Using Cox mixed-effects models, we find that age-of-onset shows a stronger genetic signal than AD case-control status, capturing many known variants with stronger significance, and also revealing new variants. We identified two novel variants, rs56201815, a rare synonymous variant in *ERN1*, and rs12373123, a common missense variant in *SPPL2C* in the *MAPT* region in *APOE ε4* non-carriers. Besides, a rare missense variant rs144292455 in *TACR3* showed the consistent direction of effect sizes across all studies with a suggestive significant level. In an attempt to unravel their regulatory and biological functions, we found that the minor allele of rs56201815 was associated with lower average FDG uptake across five brain regions in ADNI. Our eQTL analyses based on 6198 gene expression samples from ROSMAP and GTEx revealed that the minor allele of rs56201815 was potentially associated with elevated expression of *ERN1*, a key gene triggering unfolded protein response (UPR), in multiple brain regions, including the posterior cingulate cortex and nucleus accumbens. Our cell-type-specific eQTL analysis using ~80,000 single nuclei in the prefrontal cortex revealed that the protective minor allele of rs12373123 significantly increased the expression of *GRN* in microglia, and was associated with *MAPT* expression in astrocytes. These findings provide novel evidence supporting the hypothesis of the potential involvement of the UPR to ER stress in the pathological pathway of AD, and also give more insights into underlying regulatory mechanisms behind the pleiotropic effects of rs12373123 in multiple degenerative diseases including AD and Parkinson's disease.

Introduction

Late-onset sporadic Alzheimer's disease (AD) is a progressive neurodegenerative disorder accounting for 50–70% of all dementia cases in the elderly population[1]. Amyloid β-peptide (Aβ) is the primary component found in the neuritic plaques of AD patient brain, and multiple mutations in the *APP* gene and its related genes (*PSEN1* and *PSEN2*) promoting Aβ production have been identified in familial (early-onset) AD[2–6]. These observations support a causal role of Aβ deposition in the etiology of AD. Familial AD is, however, much rarer than sporadic AD, which is highly prevalent after age 65. Recent genome-wide association studies (GWAS) have identified a large number of genetic variants associated with the risk of late-onset

Correspondence: Liang He (liang.he@duke.edu) or Manolis Kellis (manoli@mit.edu) or Alexander M. Kulminski (alexander.kulminski@duke.edu)
[1]Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC, USA
[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA
Full list of author information is available at the end of the article
A list of authors and their affiliations appears at the end of the paper.
A full list of members and their affiliations appears in the Supplementary Information Text S1.

AD[7–13], most of which are located in genes exclusively expressed in microglia (e.g., *TREM2*). These insights suggest the involvement of microglia in the pathology of AD.

Despite recent progress in understanding the biological mechanisms underlying AD, the cellular and molecular activities and causation in the late-onset AD of most common variants discovered in GWAS, including those in *APOE*, remain unclear. Functional links between most of these AD-related loci and genes are still to be determined, although some microglia-related single nucleotide polymorphisms (SNPs) in, e.g., *CD33*, and the *MS4A* gene cluster, are shown to be mediated through *TREM2* (refs. [14,15]). The functional mechanisms of *TREM2* in Aβ uptake by microglia are also complicated, and contradictory biological consequences are observed in mouse models (see, e.g. ref. [16], for a review on this topic). Moreover, adding up the *APOE* variant and other nine identified top SNPs accounts for a small portion (5%) of variation of age-of-onset[17], suggesting that missing genetic mechanisms contribute to this complex disease. We expect that the discovery of additional AD-associated genetic variants will provide more insights into the understanding of AD pathology.

In this study, we performed an exome-wide association analysis of age-of-onset of AD, in which most genetic variants are rare or low frequency, using an Alzheimer's Disease Sequencing Project (ADSP) sample of 10,216 subjects in the discovery phase. Rare coding variants often show larger effect sizes, and their biological consequences are more explicable, but its association analysis is complicated by insufficient statistical power. Although the exome-wide association of AD has recently been explored using AD status[18–20], our rationale is that more AD-related rare variants can be identified using analysis of age-of-onset of AD with a Cox model given emerging evidence from a previous study showing its potential advantage in terms of statistical power[21]. We attempted to replicate significant findings in five other studies, with a meta-analysis sample size of about 20,000 subjects. To understand the biological consequences of the identified SNPs, we explored their influence on regulatory activities and gene expression at tissue and single-cell levels.

We further performed a separate exome-wide association analysis of the age-of-onset of AD by excluding the *APOE ε4* carriers. The overarching goal is to identify novel variants contributing to AD independently of the *APOE ε4* allele, the strongest single genetic risk factor for AD. Despite quarter Century research on the function of the *APOE* gene[22], the primary biological role of this gene in AD pathogenesis remains elusive as the gene and its protein are probably involved in many pathways related to Aβ deposition, Aβ clearance, tau pathology, and neuroinflammation[23]. Our analysis is designed to provide more insights into AD-related *APOE* biology.

Results

Description of the study sample in the discovery phase

In the discovery phase, we carried out an exome-wide association analysis of the age-of-onset of AD using a whole-exome sequencing (WES) sample from the ADSP[24]. We included 10,216 non-Hispanic white subjects (54.86% cases, 58.03% women) after filtering subjects with missing information about sex, AD status, or age-of-onset. The average age-of-onset of AD was 75.4 years (Table S1). We interrogated 108,509 biallelic SNPs with a missing rate <2% across the subjects and a minor allele count (MAC) >10. To identify genetic variants associated with the hazards of AD, we conducted three separate analyses. In the first and second analyses, we included all subjects and performed *ε4* allele (coded by the minor allele of rs429358) unconditional (first) and conditional (second) analyses as *APOE ε4* is a well-known strong predictor of AD. That is, we tested two models, differing as to whether the copy of the *APOE ε4* SNP rs429358 was included as a covariate. In the third analysis, we only included 7185 *APOE ε4* non-carriers. Despite this reduction of the sample size, we expect better statistical power by leveraging the age-of-onset analysis than logistic regression. In all analyses, we included as covariates sex and three principal components (PCs) (PC2, PC8, and PC10) that were significantly associated with AD ($p < 0.005$) among the top ten PCs. We built a genetic relatedness matrix (GRM) using the ADSP WES data and found that the ADSP sample contains a small number of family members or cryptic relatedness (120 subjects had a maximum genetic relatedness coefficient >0.25). All age-of-onset analyses were performed using Cox mixed-effects models implemented in the coxmeg R package[21] to correct for the relatedness of the subjects. We found that the genomic inflation was controlled in all three analyses ($\lambda = 1.028$, 1.073, and 1.023) (Fig. S1), comparable to those in ref. [18] using logistic regression models ($\lambda = 1.006–1.087$).

Exome-wide analysis of age-of-onset of AD in the discovery phase

In the first analysis (using all subjects without the adjustment for *APOE ε4*), we detected four independent signals passing the exome-wide threshold ($p = 5E{-}07$) (Fig. 1A, Table S2, and Model 1). The most significant SNP was the *APOE ε4*-coding variant rs429358, having a hazard ratio (HR) of 3.32 ($p = 4.39E{-}497$). The *p*-value is much more significant than that reported in the largest meta-analysis so far based on AD status ($p = 5.79E{-}276$)[10]. This result confirms previous findings[25–27] that *APOE ε4* is not only associated with AD status but also substantially decreases its age at onset (Fig. 2A). The three signals outside the *APOE* region were rs75932628 (the R47H mutation) in *TREM2* (HR = 2.76, $p = 8.16E{-}17$), rs7982 in *CLU* (HR = 0.890, $p = 1.1E{-}07$), and rs2405442

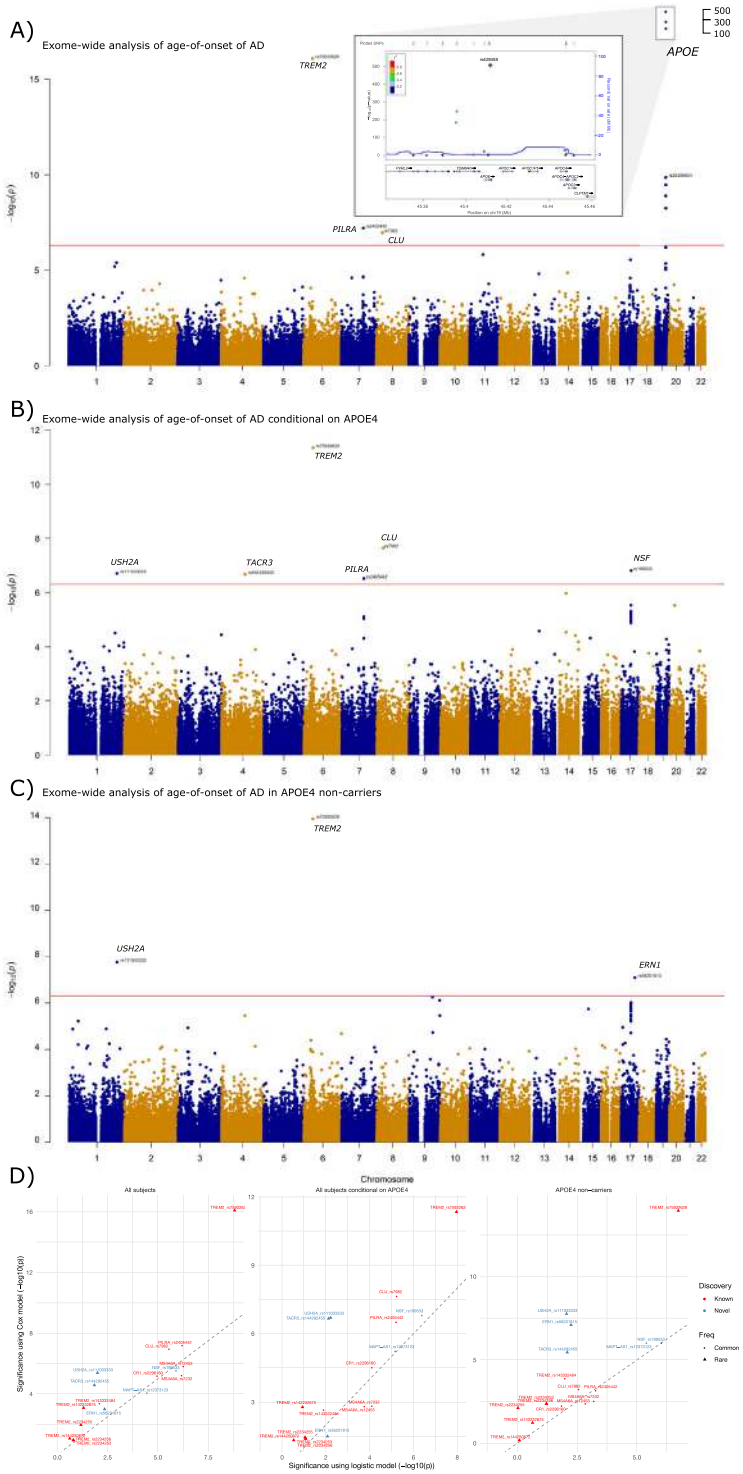

Fig. 1 Results of exome-wide association analyses of age-of-onset of AD in the ADSP sample. A Model 1: a Cox model with all subjects adjusted for three significant PCs and sex; **B** Model 2: a Cox model with all subjects adjusted for the copies of *APOE ε4*, three significant PCs and sex; **C** Model 3: a Cox model with *APOE ε4* non-carriers adjusted for three significant PCs and sex. Three top SNPs identified in the *APOE* region using Model 1 were highlighted in the regional plot due to their extremely significant *p*-values. The red horizontal line is a threshold based on the Bonferroni correction (0.05/100,000=5E−07). **D** Comparison of *p*-values between a Cox model and a logistic model for well-known AD-related SNPs and newly identified SNPs in this study in Model 1 (left), Model 2 (middle), and Model 3 (right). The same ADSP data and covariates were used to fit the Cox and logistic models.

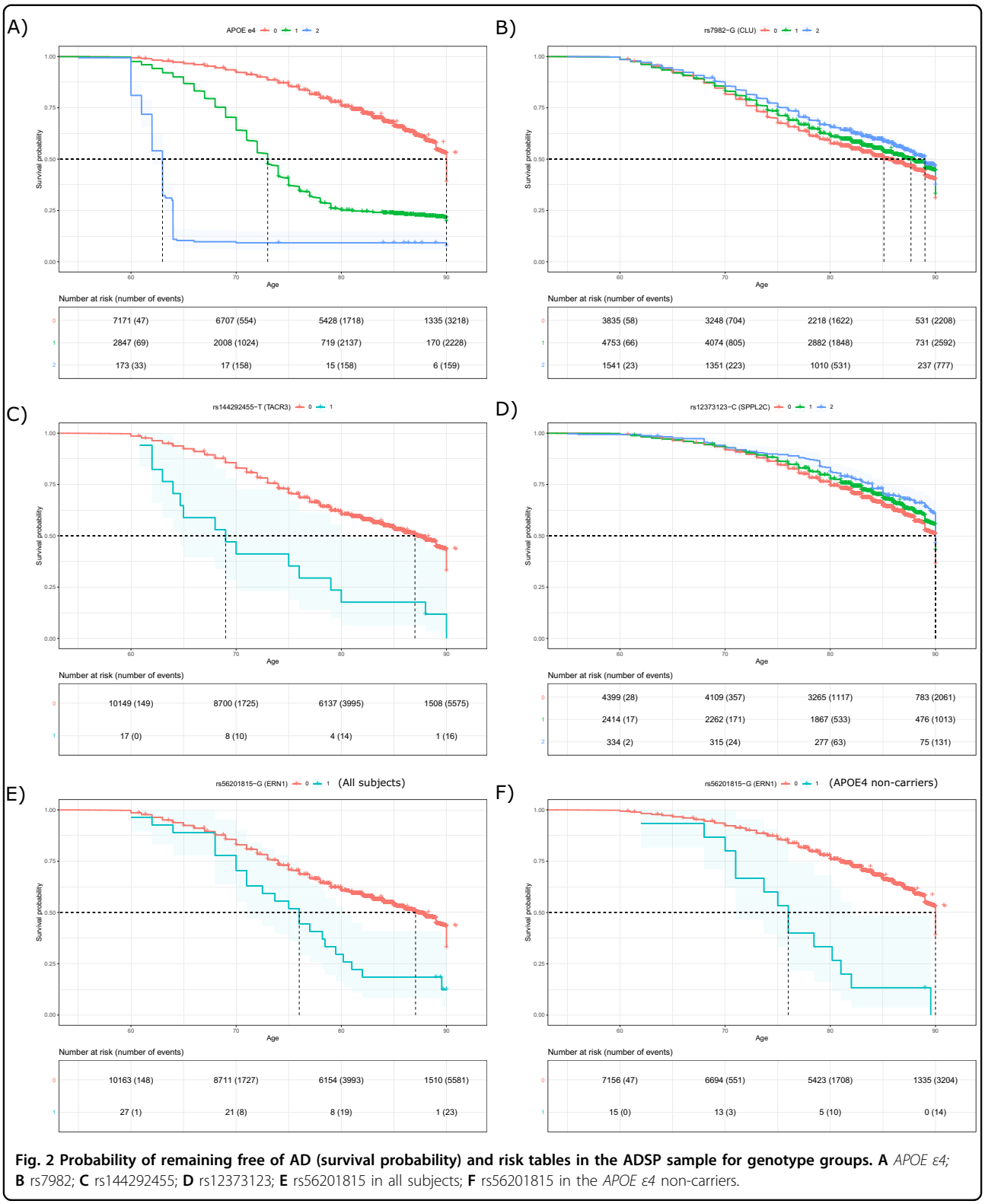

Fig. 2 Probability of remaining free of AD (survival probability) and risk tables in the ADSP sample for genotype groups. A *APOE ε4;* **B** rs7982; **C** rs144292455; **D** rs12373123; **E** rs56201815 in all subjects; **F** rs56201815 in the *APOE ε4* non-carriers.

in *PILRA* (HR = 0.879, $p = 6.35E{-}08$) (Fig. 1A, Table S2, and Model 1). The beneficial association of the missense variant rs7982 in *CLU* was not reported in the previous study of AD status using the same ADSP sample[18]. We observed that the minor allele carriers of rs7982 had lower hazards consistently across a wide age interval (Fig. 2B).

Although the R47H mutation in *TREM2* and rs2405442 in *PILRA* were identified in the previous analysis[18], our analysis achieved increased significance for the R47H mutation ($p = 8.16E{-}16$ vs. $4.8E{-}12$). In addition, we observed well-known AD-associated SNPs among the top hits, including rs12453 in *MS4A6A* ($p = 1.52E{-}06$), rs2296160 in *CR1* ($p = 6.50E{-}06$), and rs592297 in *PICALM* ($p = 5.26E{-}05$) (Table S2 and Model 1).

In the second analysis (using all subjects with the adjustment for *APOE ε4*), we identified six independent SNPs ($p < 5E{-}07$) (Fig. 1B, Table S2, and Model 2), including three aforementioned variants in *TREM2*, *CLU*, and *PILRA*. Three additional variants include rs144292455 in *TACR3* on 4q24 (HR = 5.15, $p = 2.16E{-}07$, MAC = 17), rs111033333 in *USH2A* on 1q41 (HR = 4.65, $p = 1.99E{-}07$, MAC = 19), and rs199533 in *NSF* on 17q21.31 (HR = 0.87, $p = 1.57E{-}07$, minor allele frequency (MAF) = 20.2%). The SNP rs199533 in *NSF* is previously reported in ref. [18] but does not reach the genome-wide significance in a follow-up meta-analysis incorporating replication studies[18]. The other two variants are novel. This analysis also identified two variants in *CST9* and *CDKL1* genes at the suggestive level of significance $p < 5E{-}06$ (Table 1).

In the third analysis (using only *APOE ε4* non-carriers), we identified three independent significant SNPs ($p < 5E{-}07$) (Fig. 1C, Table S2, and Model 3) including the R47H mutation in *TREM2* (HR = 2.99, $p = 1.11E{-}14$), and rs111033333 in *USH2A* (HR = 5.13, $p = 1.70E{-}08$) found in the second analysis. One novel SNP was the rare variant rs56201815 in *ERN1* within 17q23.3 locus (HR = 4.22, $p = 7.99E{-}08$, MAC = 29). The HR of the minor allele of this SNP was substantial and comparable to that of *APOE*, which is not surprising because rare coding variants tend to show more significant biological effects, and the MAF of this SNP in the ADSP sample is merely ~0.13%, much lower than that of the R47H mutation in *TREM2*.

We found that the *p*-values of the newly identified SNPs from the Cox models were more significant, particularly for the rare variants, than those from a logistic model using the same ADSP sample and covariates (Fig. 1D), explaining why these SNPs were not detected in the previous study. We compared the *p*-values of well-established AD-related coding-variants in the ADSP WES data between the two models. We found that the Cox model produced more significant *p*-values for almost all SNPs except for the two SNPs in *MS4A6A* (Fig. 1D).

Replication analyses confirm SNPs in *ERN1* and the *MAPT* region

The variants in *TREM2*, *CLU*, and *PILRA*, identified using the full sample in the first analysis, were reported by previous larger studies[10–12]. Accordingly, we focused on replication of the novel findings identified in the analyses

Table 1 Summary statistics of candidate SNPs associated with age-of-onset of AD identified from ADSP in the analysis using all subjects adjusted for *APOE ε4* and the analysis using *APOE ε4* non-carriers.

SNP	CHR	POS	Gene	MAF ADSP	ADSP		ROSMAP		CHS		LOADFS		GENADA		ADSP extension		Meta-analysis		Assumption test
					log(HR)	*p*	log(HR)	*p*	log(HR)	*p*	log(HR)	*p*	log(HR)	*p*	log(HR)	*p*	log(HR)	*p*	*p*
Analysis using all subjects adjusted for *APOE ε4*																			
rs111033333	1	216270469	USH2A	9.30E−04	1.54	1.99E−07	NA	NA	−1.42E+01	9.92E−01*	2.23E−01	6.94E−01	1.27E−03	9.98E−01	NA	NA	8.94E−01	5.22E−05	6.08E−01
rs144292455	4	104577415	TACR3	8.34E−04	1.64	2.16E−07	1.28	2.03E−01	−1.02E+01	9.95E−01*	3.46E−01	5.81E−01	5.06E−01	4.75E−01	0.34	7.35E−01	1.20E+00	9.92E−07	1.46E−01
rs61981931	14	50856882	CDKL1	4.91E−02	−0.26	1.09E−06	0.15	5.05E−01	2.55E−03	9.93E−01	−2.02E−01	4.58E−02	1.01E−01	3.64E−01	−0.46	5.98E−03	−1.93E−01	1.77E−06	2.45E−01
RS199533	**17**	**44828931**	**NSF**	**2.02E−01**	**−0.14**	**1.57E−07**	**−0.44**	**1.82E−03**	**−6.02E−02**	**6.70E−01**	**7.13E−02**	**1.89E−01**	**−6.85E−02**	**2.91E−01**	**−0.19**	**2.20E−02**	**−1.10E−01**	**3.77E−07**	**6.34E−01**
rs2983640	20	23586360	CST9	3.88E−01	−0.10	3.01E−06	−0.13	2.45E−01	5.68E−02	6.19E−01	−6.76E−03	8.78E−01	−4.41E−03	9.33E−01	NA	NA	−7.22E−02	6.25E−05	4.45E−01
Analysis using *APOE ε4* non-carriers																			
rs111033333	**1**	**216270469**	**USH2A**	**9.30E−04**	**1.63**	**1.70E−08**	**NA**	**NA**	**−1.21E+01**	**9.97E−01***	**NA**	**NA**	**4.30E−01**	**6.68E−01**	**NA**	**NA**	**1.54E+00**	**3.06E−08**	**5.38E−01**
rs144292455	4	104577415	TACR3	8.34E−04	1.30	3.49E−07	1.26	2.16E−01	−1.11E+01	9.97E−01*	8.73E−01	1.08E−01	NA	NA	0.38	7.02E−01	1.16E+00	6.80E−07	2.27E−01
rs142695278	9	10576791	CYLC2	9.31E−04	1.51	5.49E−07	NA	NA	−1.08E+01	9.96E−01*	−8.20E+00	8.18E−01*	2.00E−02	9.84E−01	NA	NA	1.39E+00	1.59E−06	7.62E−01
rs149524209	9	138712860	CAMSAP1	7.39E−04	1.57	7.59E−07	NA	NA	−1.29E+01	9.96E−01*	5.54E−01	6.19E−01	NA	NA	NA	NA	1.49E+00	9.96E−07	6.85E−01
rs79782048	**9**	**139409089**	**NOTCH1**	**9.80E−04**	**1.40**	**3.44E−06**	**NA**	**NA**	**−1.21E+01**	**9.97E−01***	**3.84E+00**	**1.41E−05**	**NA**	**NA**	**−0.67**	**5.05E−01**	**1.48E+00**	**6.87E−08**	**6.85E−01**
rs12373123	**17**	**43924073**	**SPPL2C/MAPT-AS1**	**2.16E−01**	**−0.15**	**9.67E−07**	**−0.38**	**7.95E−02**	**−5.18E−01**	**3.61E−02**	**1.19E−01**	**2.56E−01**	**−2.30E−01**	**3.45E−02**	**−0.22**	**1.01E−01**	**−1.50E−01**	**6.67E−08**	**7.99E−01**
rs2732703	17	44353222	ARL17B (intron)	2.03E−01	−0.14	1.94E−05	−0.39	1.30E−01	−6.07E−01	2.24E−02	1.51E−01	1.67E−01	−1.98E−01	7.82E−02	−0.31	5.52E−02	−1.36E−01	2.74E−06	NA
rs199533	**17**	**44828931**	**NSF**	**2.02E−01**	**−0.16**	**9.53E−07**	**−0.37**	**1.08E−01**	**−6.47E−01**	**1.42E−02**	**1.34E−01**	**2.24E−01**	**−1.72E−01**	**1.11E−01**	**−0.24**	**8.33E−02**	**−1.52E−01**	**1.20E−07**	**9.58E−01**
rs56201815	**17**	**62141416**	**ERN1**	**1.32E−03**	**1.44**	**7.99E−08**	**1.82**	**7.40E−02**	**2.33E+00**	**2.19E−02**	**2.44E+00**	**3.54E−03**	**NA**	**NA**	**2.66**	**8.92E−03**	**1.65E+00**	**1.85E−12**	**8.12E−01**

POS: coordinate of the SNPs in hg19; log(HR): logarithm of the hazard ratio. Assumption test: *p*-values for testing the assumption of proportional hazards in ADSP. The SNPs with a meta-analysis *p*-value <5E−07 are in boldface.
*The model did not converge in these analyses due to no minor allele carriers in the cases.

conditional on *APOE ε4*, and using the *ε4*-free sample. We attempted to replicate associations of ten candidate SNPs with a *p*-value <5E−06 in at least one of the models in the discovery phase (Table 1), including five common variants (MAF ≥5%) and five rare variants (MAF <1%). All these SNPs passed a test for the assumption of proportional hazards in the discovery phase (Table 1). We further included rs2732703, an intronic variant of *ARL17B* in the *MAPT* region reported being associated with AD in a previous study of *APOE ε4* non-carriers[28]. This SNP is in high linkage disequilibrium (LD) with our identified coding variants rs199533 ($r^2 = 0.90$) in *NSF* and rs12373123 ($r^2 = 0.93$) in *SPPL2C*. We examined these SNPs in non-Hispanic white populations of LOADFS (3473 subjects, 43.4% cases, imputed genotypes), CHS (3262 subjects, 6.2% cases, imputed genotypes), GenADA (1588 subjects, 50% cases, imputed genotypes), the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP) cohort (1195 subjects, 45% cases, whole-genome sequencing (WGS) genotypes[29]), and the ADSP extension study (1147 subjects, 45.8% cases, WGS genotypes) (Table S1). We removed ~400 subjects from the ROSMAP WGS cohort, 572 from CHS, 318 from LOADFS, who were already included in the ADSP sample, resulting in 681, 2690, 3155 non-Hispanic whites, respectively. The coxmeg R package[21] was used to analyze the LOADFS dataset with a GRM estimated from its genotype array, and the coxph function in the survival R package[30] was used to analyze the CHS, GenADA, ROSMAP, and ADSP extension datasets.

The meta-analysis of the summary statistics from the conditional model adjusted for *APOE ε4* showed that rs199533 in *NSF* reached the exome-wide significance of 5E−07 (meta-analysis $p = 3.77E{-}07$) (Table 1). Besides, rs144292455 in *TACR3* (MAF = 0.083% in ADSP) showed the consistent direction of effect sizes across all studies (The model did not converge in CHS as there was only one carrier.) with a *p*-value close to the exome-wide significance ($p = 9.92E{-}07$). Rs144292455 is a coding variant of *TACR3* resulting in a premature stop codon and, thus a shortened transcript. The minor allele of rs144292455 increased the risk of AD in ADSP (17 carriers, 16 cases), ROSMAP (2 carriers, 1 case), LOADFS (10 carriers, 4 cases), GenADA (2 carriers, 2 cases), and the ADSP extension study (2 carriers, 1 case). The vast majority of the minor allele carriers in ADSP (16 of 17; 3 of 16 also carry *APOE ε4* allele) had AD with an average age-of-onset of 71.03 (Fig. 2C). This age was substantially younger than the average age-of-onset of 75.4 years based on all AD cases. Two carriers in ROSMAP were both *APOE ε4* non-carriers and the AD case carried *APOE ε2/ε4* genotype.

In the analysis using *APOE ε4* non-carriers, three SNPs (rs56201815, rs12373123, and rs199533) showed exome-wide meta-analysis *p*-values ($p < 5E{-}07$) more significant than those from the ADSP sample alone. Association for rs111033333 in *USH2A* and rs79782048 in *NOTCH1* remained at the exome-wide significance. Replication of these two rare variants was, however, less robust because ≤1 minor allele carrier was observed in most of the replication cohorts and thus the significance of the meta-analysis *p*-value was dominantly attributed to the signal from the discovery phase. The novel AD-associated SNP rs56201815 (meta-analysis $p = 2.35E{-}12$) is a synonymous variant in *ERN1*. rs12373123, a missense variant of *SPPL2C* (Table 1), is located in a large LD block spanning the *MAPT region* and it is in complete LD with multiple synonymous, nonsense, or missense variants in *CRHR1* and *MAPT*. In *APOE ε4* non-carriers, the hazards of AD were consistently lower in the carriers of the minor allele of rs12373123 after age 70 (Fig. 2D). It had a more significant *p*-value (meta-analysis $p = 6.67E{-}08$) than the previously reported SNP rs2732703 (meta-analysis $p = 2.74E{-}06$) and rs199533 (meta-analysis $p = 1.11E{-}07$) among *APOE ε4* non-carriers, while rs199533 was more significant in the full sample. The minor allele of rs12373123 was consistently associated with decreased risk of AD in all studies except for LOADFS.

The minor allele of rs56201815 in *ERN1* increases the risk of AD and lowers glucose metabolism

Among the aforementioned replicated SNPs, rs56201815 in *ERN1* yielded the most significant meta-analysis *p*-value, and its minor allele (G) (MAF = 0.15% in a non-Finnish European sample)[31] increased the risk of AD consistently across all studies and independently of the *APOE ε4* allele. The HRs were nominally significant in LOADFS ($p = 3.54E{-}03$) and CHS ($p = 2.19E{-}02$). In GenADA, no carriers of the minor allele were observed. We analyzed the minor allele carriers in these studies in more detail. Twenty-seven (16 males) rs56201815-G carriers in ADSP (a total of 29 carriers in which two were excluded from the analyses because they transformed from control to mild cognitive impairment (MCI) during the follow-up in ADSP, and their AD status was unknown) were sampled from 11 cohorts including ACT, ADC, CHAP, MAYO, MIA, MIR, ROSMAP, VAN, ERF, FHS, and RS (Table 2). The genotypes of these rs56201815-G carriers passed the quality control and had high sequencing depth. Of them, 23 subjects were diagnosed with AD and their average age-of-onset (73.5 years) was lower than the average age-of-onset (75.4 years) of all AD cases in ADSP (Fig. 2E). Interestingly, three of the four rs56201815-G carriers in the control group carried *APOE ε4* allele that explained why this SNP was only identified in the analysis of *APOE ε4* non-carriers. Indeed, we observed that rs56201815-G had a stronger effect on the risk of AD in *APOE ε4* non-carriers (Fig. 2F and Table S2). In the ROSMAP WGS cohort (after excluding the

Table 2 Detailed information about rs56201815-G carriers in ADSP, ROSMAP, LOADFS, and CHS.

Project	Age at onset/ censoring	Total subjects	AD	Non-AD	Male (%)	*APOE ε4* carrier	Seq Quality/ IMP R2	Comments	Cohort
ADSP	60–70	6	6	0	50	4	PASS		ADC, MIR, VAN
	70–80	13	13	0	53.85	5			ADC, MAYO, MIA, MIR, ROS, VAN, ERF, FHS, RS
	>80	8	4	4	75	3			ACT, ADC, CHAP, ROS, RS
	NA	2	NA	NA	50	0		Control converted to MCI. Not included in the analysis due to unknown AD status.	ADC
ROSMAP	80–90	3*	2	1	0	1	99		
LOADFS	<70	2	1	1	100	1	0.804	The AD patient is not an *APOE ε4* carrier.	
	70–80	7	4	3	71.43	5		One of the *APOE ε4* non-carriers is an AD patient.	
	NA	1	1	0	100	0		Unknown exact age-at-onset of AD (age-at-onset<79). This subject had also dementia at age 73.	
CHS	70–80	2	0	2	50	0	0.950		
	80–90	5	1	4	60	1		The AD patient is an *APOE ε4* non-carrier.	
	90–100	2	0	2	50	2			
ADSP extension	60–70	1	1	0	100	0	99		ADC
	NA	1	NA	NA	100	0		This subject converted to dementia during follow-up.	ADC

Age at onset/censoring: age-of-onset if the subject had AD or age at the end of follow-up if the subject was a control. AD/Non-AD: number of AD/control subjects. Male: percentage of males. APOE e4 carrier: number of *APOE ε4* carriers. Seq Quality/IMP R2: sequencing quality of the minor allele carriers for ADSP WES, ADSP extension WGS, and ROSMAP WGS or imputation r^2-value for LOADFS and CHS. Comments: additional information about the subject. Cohort: the original cohort in ADSP.
*This number does not include the three subjects overlapping in ADSP.

duplicated subjects examined in the ADSP sample), we observed three rs56201815-G carriers, including one *APOE ε4* carrier (Table 2). Two of the three carriers were diagnosed with AD, which, albeit from a small sample size, is much higher than the incidence of 36.7% in the non-carriers. The genotypes of all carriers had high sequencing quality. In the LOADFS cohort, we observed ten rs56201815-G carriers (all with a dosage >0.98) (Table 2). Three out of the four *APOE ε4* non-carriers among these subjects had both AD and dementia (Table 2). This incidence (75%) was higher than that in rs56201815-G non-carriers (43%). In the CHS cohort, we observed nine rs56201815-G carriers (all with a dosage >0.98) (Table 2). One out of the six *APOE* ε4 non-carriers among these subjects (16.7%) had AD during the follow-up, higher than the incidence (6.16%) in rs56201815-G non-carriers. In the ADSP extension WGS study, we observed two rs56201815-G carriers in non-Hispanic whites, and both were *APOE ε4* non-carriers. One of the carriers was diagnosed with AD at age 69, and the other converted to dementia during the follow-up with unknown status of AD.

The ADNI project was not included in the replication analysis because the age-of-onset of AD was not available. Moreover, the vast majority of the ADNI WGS sample (738 subjects) was MCI or control subjects, and AD cases accounted for merely 5.8%. Instead, we investigated the association between rs56201815 and average FDG-PET intensity, one of the most accurate biomarkers to predict conversion from MCI to AD and to distinguish between control, early MCI (EMCI), late MCI (LMCI), and AD subjects[32–36], across five brain regions of interest (ROIs) (left/right angular gyrus, bilateral posterior cingulate gyrus, and left/right inferior temporal gyrus). We observed that the average FDG uptake of the five rs56201815-G carriers (two LMCI subjects, one EMCI subject, and two controls) adjusted for within-subject variability, age at measurement, sex, and diagnosis groups (control, EMCI, LMCI, and AD) was significantly lower than that of the homozygous subjects (Fig. 3A), suggesting that the rs56201815-G carriers had lower cerebral glucose metabolism and will more likely convert to advanced stages.

rs56201815 is a synonymous variant and potential brain-specific expression quantitative trait locus (eQTL) of *ERN1*

As rs56201815 in *ERN1* was the most significant SNP identified from the discovery and replication phases, we next sought to examine its biological and regulatory functions. rs56201815 is a synonymous coding variant, indicating that it unlikely alters the amino acid sequence of *ERN1*. However, rs56201815 is located in a CTCF binding site, an open chromatin region in multiple cell types, and an evolutionarily conserved region (Fig. 3B). Moreover, a recent mouse study reports that inhibition of *ERN1* expression reduces amyloid precursor protein (APP) in cortical and hippocampal areas, and restores the learning and memory capacity of AD mice[37]. We, therefore, hypothesized that rs56201815 is a cis-eQTL of *ERN1* in the brain, and the detrimental effect of rs56201815 on AD is mediated by upregulating the expression of *ERN1*. To test this hypothesis, we examined the effect of rs56201815 on the expression of *ERN1* using RNA-seq data in ROSMAP and GTEx, and microarray data in ADNI.

We collected 2213 RNA-seq samples from 838 subjects in the ROSMAP cohort in three brain regions including the dorsolateral prefrontal cortex (PFC), posterior cingulate cortex (PCC), and anterior caudate nucleus, among which four subjects were rs56201815-G carriers. Our differential expression (DE) analysis revealed that the minor allele of rs56201815 was associated with increased expression of *ERN1* (log(fold-change (FC)) = 0.204, $p = 0.0285$) in PCC (Fig. 3C). We then analyzed a WGS dataset of 838 healthy subjects from the GTEx project. The WGS data included two rs56201815-G carriers, one of which had RNA-seq data in nine brain tissues including the amygdala, anterior cingulate cortex (ACC), hypothalamus, caudate, nucleus accumbens, putamen, cerebellar hemisphere, cerebellum, and spinal cord. Despite the small sample size, our DE analyses indicated that rs56201815 was a potential eQTL of *ERN1* in several regions in the cerebrum, particularly the nucleus accumbens (log(FC) = 1.28, $p = 1E{-}4$), and the putamen (log(FC) = 0.734, $p = 0.05$) (Fig. 3D). In line with the result from the ROSMAP data in PCC, rs56201815-G was correlated, albeit not significant (log(FC) = 0.35, $p = 0.437$), with the expression in ACC, leading to a significant meta-analysis p-value of 0.0213 for cingulate cortex. In almost all regions in the cerebrum, the rs56201815-G carrier had uniformly higher expression of *ERN1* than the average (Figs. 3D and S2A).

We then investigated the effects of rs56201815 on *ERN1* expression in other brain regions, and in four non-brain tissues including the sigmoid colon, lung, spleen, and whole blood. The RNA-seq data in the sigmoid colon had two rs56201815-G carriers, and one rs56201815-G carrier was available in the other tissues. The DE results showed no evidence of an association between rs56201815 and the gene expression in any of these tissues (Fig. S2A). As the number of rs56201815-G carriers in the GTEx project is small, we further analyzed a peripheral whole blood sample from the ADNI project, comprising 733 subjects having both a WGS dataset and a microarray gene expression dataset, three of whom were rs56201815-G carriers with high sequencing quality. Our DE analyses of two probes in *ERN1* showed that the minor allele rs56201815-G was not associated with either probe (Fig. S2B).

These results suggested that rs56201815 was associated with elevated expression of *ERN1* in cerebral regions (most predominantly in PCC and several regions in the basal

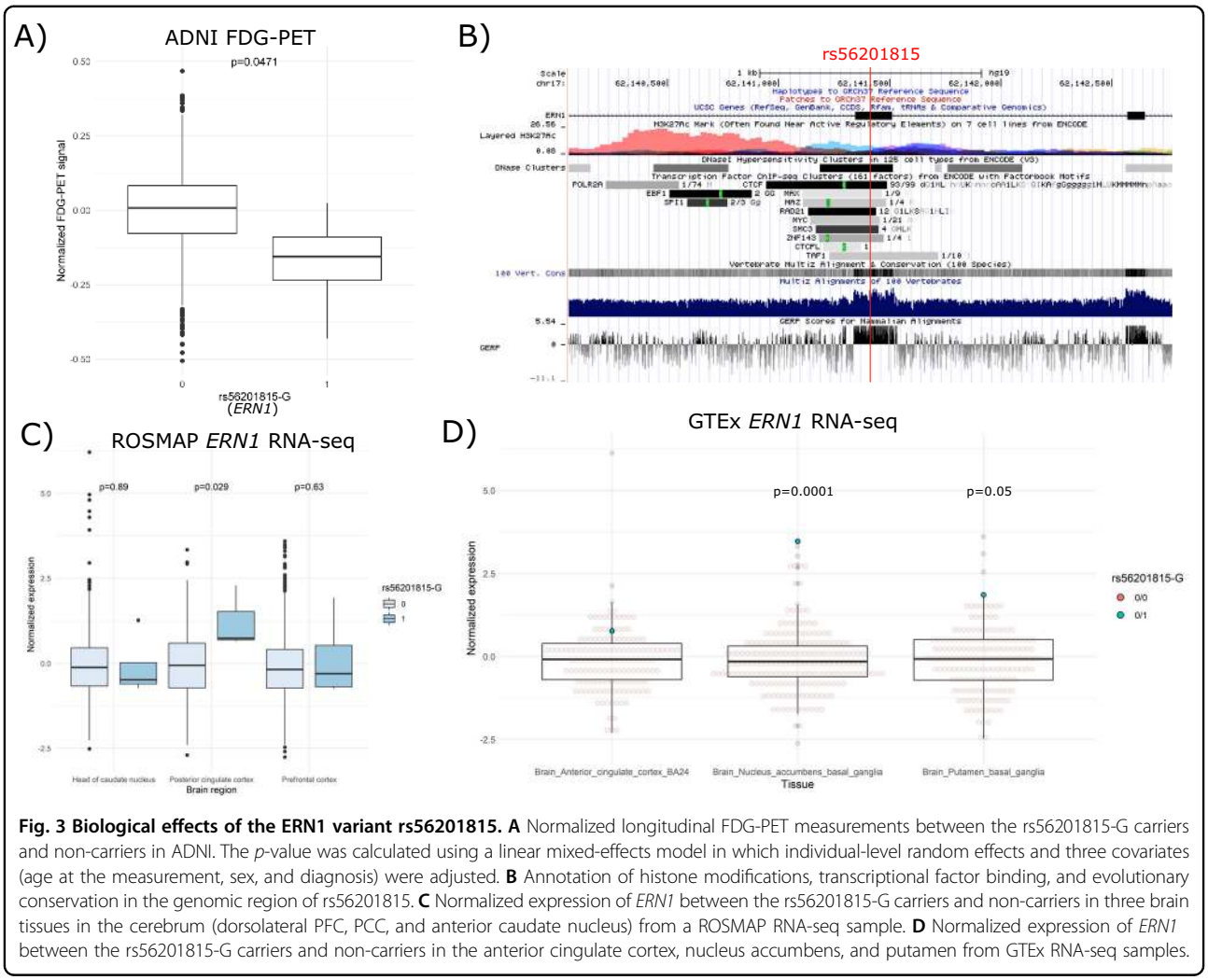

Fig. 3 Biological effects of the ERN1 variant rs56201815. A Normalized longitudinal FDG-PET measurements between the rs56201815-G carriers and non-carriers in ADNI. The *p*-value was calculated using a linear mixed-effects model in which individual-level random effects and three covariates (age at the measurement, sex, and diagnosis) were adjusted. **B** Annotation of histone modifications, transcriptional factor binding, and evolutionary conservation in the genomic region of rs56201815. **C** Normalized expression of *ERN1* between the rs56201815-G carriers and non-carriers in three brain tissues in the cerebrum (dorsolateral PFC, PCC, and anterior caudate nucleus) from a ROSMAP RNA-seq sample. **D** Normalized expression of *ERN1* between the rs56201815-G carriers and non-carriers in the anterior cingulate cortex, nucleus accumbens, and putamen from GTEx RNA-seq samples.

ganglia), but not likely in other tissues. To examine whether its regulatory effects in the brain are mediated by a change of chromatin activity, we further carried out association analyses of epigenetic markers including DNA methylation and histone modifications in PFC. We collected an Illumina 450k array DNA methylation dataset of 721 subjects (four rs56201815-G carriers) from a ROSMAP sample[38,39]. Among 11 probes located in the region of *ERN1*, we found no evidence of significant association after adjustment for multiple testing (Table S3). The most significant probe (chr17:62134117), also the probe closest to rs56201815, was located in an enhancer with a *p*-value of 0.012. For histone modifications, we interrogated histone 3 lysine 9 acetylation (H3K9ac) peaks using a ChIP-seq dataset of 632 subjects (four rs56201815-G carriers) from a ROSMAP sample[38,40]. We conducted differential analyses of 26,384 broad peaks adjusted for fraction of reads in peaks (FRiPs), GC bias, and ten remove unwanted variation (RUV) components. No significant association was found among nine broad peaks within a ±200 kb flanking region of *ERN1* after adjustment of multiple testing although eight peaks showed slightly increased intensity in the carriers (Table S4). The most significant association was in an enhancer at chr17:62,337,374-62,342,372 with a *p*-value of 0.043.

Rs12373123 is a neural cell type-specific eQTL of *MAPT* and *GRN*

Previous studies show that rs12373123 is a cis-eQTL of multiple nearby genes (e.g., *MAPT*, *CRHR1*, and *LRRC37A*) in multiple tissues including the brain[28,41–43], and shows chromatin interactions with these genes (Fig. 4A). But it is not clear which cell type and genes mediate its effect on AD. We then explored the regulatory effects of rs12373123 at a cell-type level using a single-nucleus RNA-seq (snRNA-seq) dataset. Cell type-specific analysis can also reduce the potential confounding effects originating from unobserved heterogeneous cell type proportion across subjects in the tissue-level analysis, and therefore produces more accurate and refined estimates. We performed cell type-specific eQTL analyses using 44 subjects having both genotype data

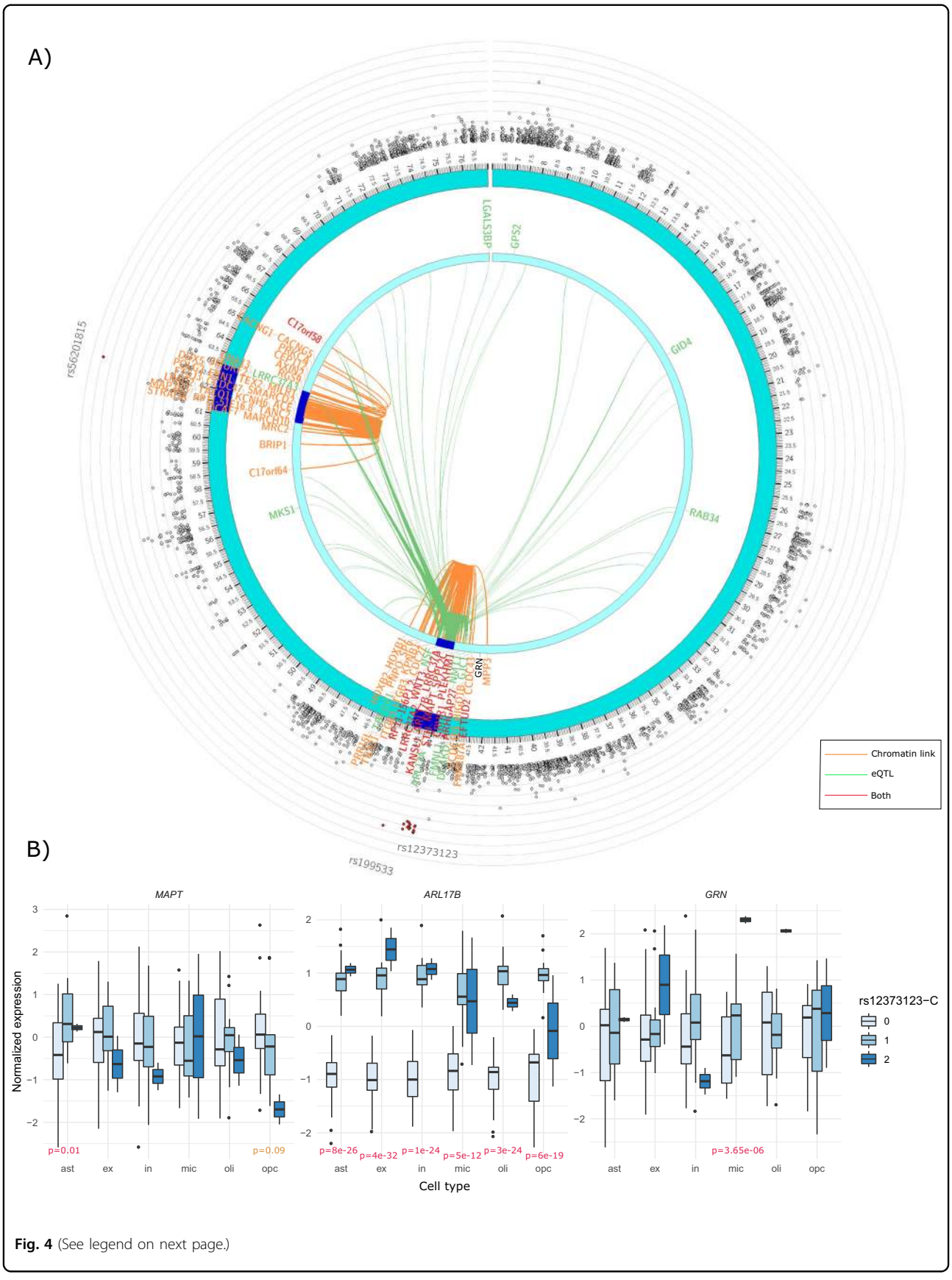

Fig. 4 (See legend on next page.)

(see figure on previous page)
Fig. 4 Local regulatory effects of rs12373123. A Chromatin interaction (orange links) and tissue-specific eQTLs (green links) for rs56201815 and rs12373123 on chromosome 17 identified from the exome-wide association analysis of age-of-onset of AD in *APOE ε4* non-carriers in ADSP. A gene that is in chromatin interaction or an eGene with these SNPs is highlighted in orange or green, respectively. A gene highlighted in red indicates both features. **B** Normalized cell type-specific (astrocytes, excitatory neurons, inhibitory neurons, microglia, OPCs, and oligodendrocytes) expression of *MAPT*, *ARL17B*, and *GRN* across the genotype groups of rs12373123 from 44 subjects (including 13 rs12373123-T/C carriers and 2 rs12373123-C/C carriers) in the snRNA-seq data in the prefrontal cortex. All cells in each cell type from each subject were first pooled, and the gene expression was aggregated by subjects. The gene expression was then adjusted for age, sex, and AD status.

(39 subjects from WGS and five subjects from a SNP array) and snRNA-seq data from ~80,000 cells in PFC from a ROSMAP sample. We classified cells into excitatory neurons, inhibitory neurons, astrocytes, microglia, oligodendrocytes, and oligodendrocyte progenitor cells (OPCs) based on previous clustering results[44]. We then aggregated cells within each cell type and each subject.

In each cell type, we interrogated 11 protein-coding genes (10 genes within a ±500 kb flanking region and *GRN*, a nearby gene linked to frontotemporal lobar degeneration (FTD), a type of dementia). The cell type-specific eQTL analyses revealed that one or more copies of rs12373123-C were associated with elevated expression of *ARL17B* in all six brain cell types ($p < 1E{-}11$) (Fig. 4B and Table S5). rs12373123 was also an eQTL of *LRRC37A2, LRRC37A3*, and *KANSL1* in most cell types except for microglia (Fig. S3 and Table S5). The protective allele rs12373123-C was associated with elevated *MAPT* expression in astrocytes ($p = 0.01$) while a decreasing trend in OPCs ($p = 0.09$) (Fig. 4B and Table S5). We further found that rs12373123-C, particularly its homozygous protective genotype, was significantly associated with increased expression of *GRN* in microglia ($p = 3.65E{-}06$) (Fig. 4B and Table S5), which is a protective gene against dementia and is important for lysosome homeostasis in the brain[45,46].

We also assessed the cell type-specific association between rs56201815 and the expression of *ERN1*. We observed that *ERN1* was ubiquitously expressed in all brain cell types, most abundantly in microglia, followed by astrocytes and OPCs. As there was only one rs56201815-G carrier among the 39 WGS subjects, and, unfortunately, its total sequencing depth was much lower than that of the other subjects (~10% of the average library size), we investigated three major abundant cell types (excitatory neurons, astrocytes, and oligodendrocytes), for which the carrier had a library size >50,000. We observed that rs56201815-G was slightly correlated with increased expression of *ERN1* in excitatory neurons, but not significant (Fig. S4).

Gene-set analysis identifies astrocyte, microglia, and amyloid-beta-related pathways

As aggregating signals within a gene can often increase the statistical power, in particular, for detecting rare coding variants, we carried out gene-based analyses using the summary statistics of all examined SNPs estimated from the ADSP sample. Our gene-based analyses using MAGMA[47] showed that *TREM2* was the most significant gene associated with AD in all individuals ($p = 5.0E{-}10$) and *APOE ε4* non-carriers ($p = 1.62E{-}10$) (Fig. S5A), consistent with previous results[18]. Indeed, all six exonic SNPs (rs2234256, rs2234255, rs2234253, rs142232675, rs143332484, rs75932628) in *TREM2* were at least nominally associated with AD (Table S2). Its significance in *APOE ε4* non-carriers was higher, suggesting that the effects of *TREM2* on AD were independent of *APOE*. Besides, multiple genes in the *MAPT* region including *MAPT, KANSL1, NSF*, and *SPPL2C* were associated with the risk of AD in both analyses (Fig. S5A, B). We also observed that *CLU, PILRA, EXO5*, and *ERN1* were among the top associated genes.

Our gene-set analysis using FUMA[48] based on the summary statistics from the exome-wide association analysis conditional on *APOE ε4* revealed that Gene Ontology (GO) gene sets related to the regulation of astrocytes, amyloid-beta, endoplasmic reticulum (ER) stress, and unfolded protein response (UPR) were among the top enriched gene sets associated with AD (Fig. 5A). In contrast, the gene sets related to astrocyte activation, microglia migration, and lipoprotein metabolic process were among the top in the gene-set analysis using *APOE ε4* non-carriers (Fig. 5B). Our cell-type association analysis using FUMA[49] (Watanabe et al., 2019) showed that microglia were associated with AD among nine major cell types in the brain ($p < 0.05$) in the analysis of *APOE ε4* non-carriers (Fig. 5D). No cell type was associated with AD based on the summary statistics from the association analysis conditional on *APOE ε4* (Fig. 5C).

Discussion

In this study, we interrogated the associations between 108,509 exome-wide SNPs and age-of-onset of late-onset AD using Cox models with a sample consisting of ~20,000 AD patients and controls. We also attempted to identify SNPs contributing to earlier onset in *APOE ε4* non-carriers alone. Most of these SNPs are rare variants. Our results not only confirm previously reported AD-related SNPs with much higher significance but also reveal novel

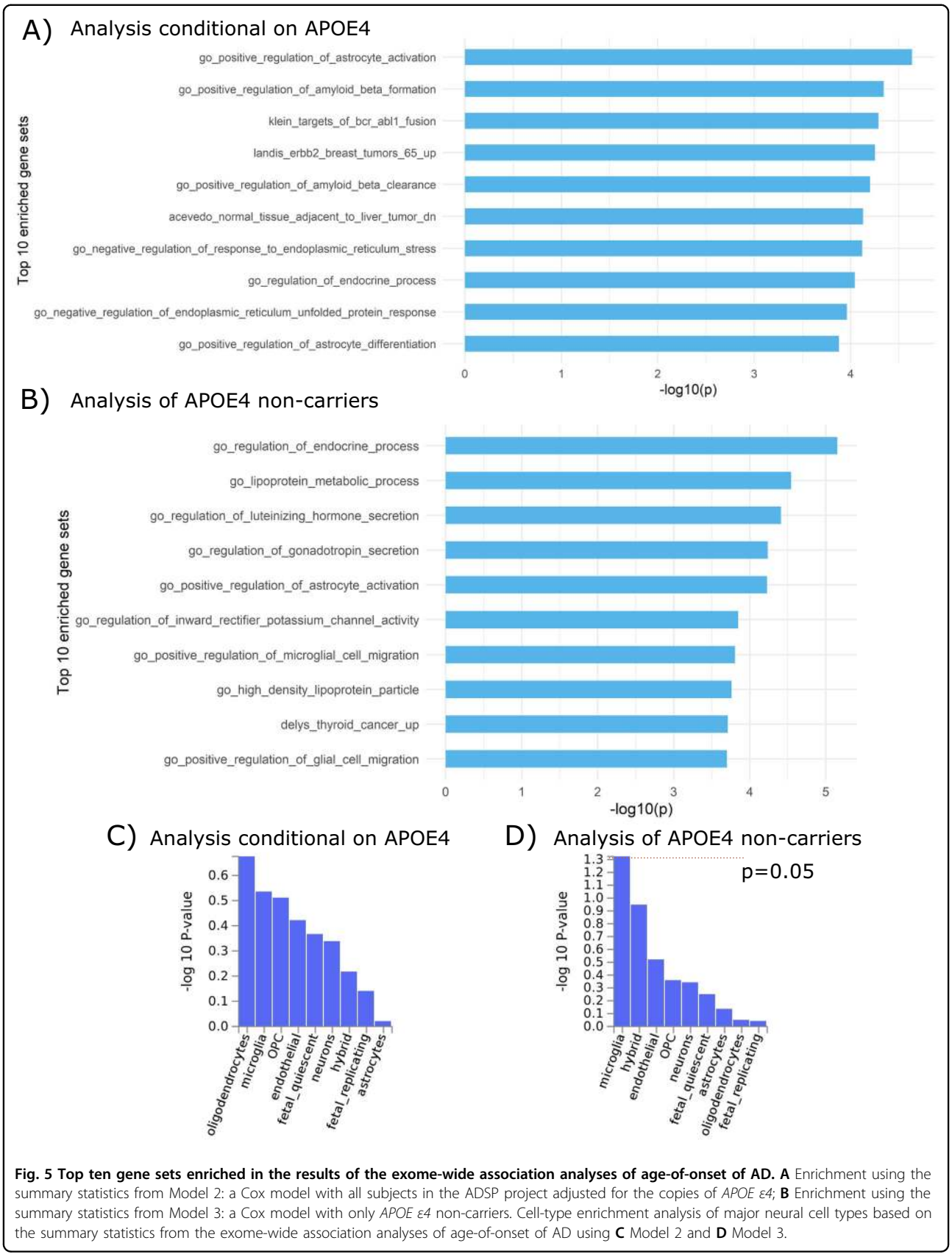

Fig. 5 Top ten gene sets enriched in the results of the exome-wide association analyses of age-of-onset of AD. A Enrichment using the summary statistics from Model 2: a Cox model with all subjects in the ADSP project adjusted for the copies of *APOE ε4*; **B** Enrichment using the summary statistics from Model 3: a Cox model with only *APOE ε4* non-carriers. Cell-type enrichment analysis of major neural cell types based on the summary statistics from the exome-wide association analyses of age-of-onset of AD using **C** Model 2 and **D** Model 3.

genetic variants associated with age-of-onset of AD, particularly in *APOE ε4* non-carriers.

One of our major findings is a synonymous rare variant, rs56201815, in *ERN1* (also known as *IRE1*). Our results showed that the minor allele of this SNP was associated with a dramatically higher risk of AD, particularly in *APOE ε4* non-carriers. Its large effect size, unanimously replicated in three other cohorts, is not surprising as its MAF in the population is only ~10% of the rare variant rs75932628 in *TREM2* according to ExAC (https://gnomad.broadinstitute.org/). *ERN1* encodes a key protein, containing a serine/threonine-protein kinase domain and a ribonuclease (RNase) domain, involved in UPR to ER stress by activating its downstream target *XBP1* (refs. [50,51]). Interestingly, a recent experimental study shows that the proportion of activated *ERN1* in postmortem brain tissue is associated with a Braak stage of advanced AD patients[37]. Deactivation of the RNase domain of *ERN1* in neurons reduces all hallmarks of AD including amyloid-beta load, cognitive impairment, and astrogliosis in 5xFAD mice[37]. Moreover, the ablation of eIF2α kinase PERK, one of the three major UPR genes, also prevents defects in synaptic plasticity and spatial memory in AD mice[52]. Our findings show that the minor allele of rs56201815, increasing mRNA expression of *ERN1* in multiple brain regions, also significantly increases the risk of AD, which corroborate these experimental results and provide more evidence that responses to ER stress are probably involved in the causal pathway of AD.

Aging is the most important risk factor for late-onset AD, indicating that certain risk factors during the aging process might be implicated and required in the pathogenesis of AD. The UPR is one of the mechanisms disrupted during aging, resulting in augmented susceptibility to ER stress and the accumulation of unfolded protein[53]. Previous studies show that aging leads to deficits in the systems involved in the defense against unfolded proteins in the rat hippocampus[54]. Persistent ER stress in the central nervous system during aging can initiate apoptosis of neurons and can trigger the innate immune response in microglia[55,56]. Combined with the fact that many AD-related genes identified by GWAS are expressed exclusively in microglia, our findings indicate that the interaction between the UPR and innate immune system might play a critical role in biological mechanisms underlying AD.

As rs56201815, the variant rs12373123 in the *MAPT* region was also identified in *APOE ε4* non-carriers. The minor allele of rs12373123 was associated with reduced susceptibility to AD in ADSP, ROSMAP, CHS, and GenADA. This SNP is located in an LD block spanning >400 kb, and is in high LD with a large number of SNPs including multiple missense variants in *MAPT*, *SPPL2C*, *CRHR1*, and *KANSL1*. Previous GWAS show that rs12373123 and two nearby missense SNPs (rs12185268 and rs12373124) in complete LD with rs12373123 exhibit pleiotropic associations with numerous diseases and traits including intracranial volume[57], corticobasal degeneration[58], Parkinson's disease (PD)[59–62], primary biliary cirrhosis[63], red blood cell count[64], and androgenetic alopecia[65]. On the other hand, the major allele, more predisposed to degenerative diseases, is significantly associated with increased bone mineral density[66,67]. Because SNPs contributing to age-related degenerative diseases are generally not subject to evolutionary selection[68,69], its major allele is probably selected by evolution due to its beneficial effect on bone mineral density. The results of our age-of-onset analyses indicate that this pleiotropic region might also be implicated in late-onset AD, especially in *APOE ε4* non-carriers. Our cell type-specific analyses reveal that rs12373123 is a cis-eQTL in different brain cells of multiple critical genes implicated in PD and FTD (e.g., *MAPT* and *GRN*), elucidating the regulatory mechanisms underlying its pleiotropy. Due to the involvement of tau protein in the etiology of AD and PD, the effect of rs12373123 on these diseases might be mediated by *MAPT*. Indeed, rs12373123 is in high LD with multiple missense SNPs (e.g., rs62056781 and rs74496580) in *MAPT*, and we found in the snRNA-seq data that rs12373123 is also an eQTL of *MAPT* in astrocytes. Our finding also suggests that the effects of rs12373123 can be mediated by increasing the expression of *GRN* in microglia, which is a key gene protective against FTD.

Also, our results demonstrated advantages in the statistical power of using a Cox model for age-of-onset traits than a logistic model for binary outcomes in the study of AD. The power gain in terms of *p*-values is evident for many well-known AD-related SNPs in e.g., *TREM2* and *CLU*, which all achieved more significant *p*-values than a previous study using the same cohort[18]. Despite a smaller sample size, the *p*-value from the Cox model for detecting *APOE ε4*, the recognized true positive signal, is much more significant than a recent large-scale meta-analysis of AD status[10] and a previous analysis using a linear model of log-transformed age-of-onset[26]. Moreover, our age-of-onset analysis showed promising results for identifying rare variants compared to logistic regression. An advantage of a Cox model over Poisson regression or logistic regression is that it implicitly accounts for age-varying hazards, a characteristic in many age-related diseases, e.g., AD[70]. Our results in AD suggest that Cox models can have a power advantage for exploring rare variant association in other age-related diseases.

Although our identified SNPs were validated in multiple independent cohorts, we acknowledge some limitations. The definitions and criteria of diagnosis of AD can vary across these cohorts. AD has a certain similarity in the clinical and biological manifestation of other common neurodegenerative diseases such as FTD, which makes the

clinical diagnosis of AD more complicated. Also, one of our findings rs56201815 in *ERN1* is a rare variant (MAF = ~0.13%), which had slightly lower imputation quality compared to common variants. Although this SNP showed solid associations in our meta-analyses, as the sample sizes of our WGS replication cohorts are small for rare variants, more GWAS using large-scale WGS or WES data are preferable to further validate this SNP and other candidate SNPs identified in the discovery phase.

In conclusion, we identified two novel SNPs in *ERN1* and *SPPL2C/MAPT-AS1* that exhibit strong associations with the age-of-onset of AD. We also explored their regulatory consequences at the tissue and single-cell levels in the brain. These findings support the hypothesis of the potential involvement of the UPR to ER stress and tau protein in the pathological pathway of AD, contributing to the understanding of the biological mechanisms underlying AD. Our findings are useful for guiding follow-up studies and provide more insight into the molecular mechanisms and implications of the relevant genes in AD.

Methods

Phenotypes in age-of-onset GWAS

A total of 10,913 European-American participants used in the discovery phase of the exome-wide age-of-onset association analyses of AD were collected from the ADSP project. These subjects were sampled from 24 cohorts, among which >3000 subjects were sampled from the ADC project (Table S6). The AD status of individuals used in the analyses was defined by clinical assessment based on NINCDS-ADRDA criteria of AD. All controls were cognitively normal individuals aged 60+. Details about study design and sample selection were described in ref. [71]. The AD status variable in the ADSP dataset was constructed based on information on prevalent and incident AD status from the updated dataset (Version 7 with the release date on June 09, 2016) if available. Otherwise, information on prevalent and incident AD status as given in Version 5 (release date on July 13, 2015) was used. More specifically, a subject was treated as AD if either prevalent or incident AD status during the ADSP follow-up was observed. The age-of-onset variable was based on the same datasets as the AD status. In both versions (Version 5 and 7), all data for age-of-onset, which we received from dbGaP, were censored by age 90.

Five cohorts (ROSMAP, LOADFS, CHS, GenADA, the ADSP extension study) were included in the replication phase of the age-of-onset GWAS. To be consistent with the AD status in ADSP, AD status in ROSMAP was based on the clinical diagnosis of AD at the last visit. For AD cases, the age at first Alzheimer's dementia diagnosis variable was used as age-of-onset, which was also censored by age 90 if it was 90+. For controls, age-of-onset was calculated as age at the last visit or age at death if age at the last visit was not available. In LOADFS, some subjects had missing information about the age-of-onset of AD. For these subjects, we treated them as censored and set its age-of-onset as the age at the recruitment. In CHS and GenADA, the AD status and age-of-onset variables in phenotype files provided in dbGaP were used. In the ADSP extension study, the "AD" and "Age" variables in phenotype files were used as the AD status and the age-of-onset. We included definitive AD and control subjects, and subjects diagnosed with probable AD, possible AD, family AD, non-family AD, or unknown were not included in the analysis.

Genotyping, imputation, and quality control

In the discovery study, WES genotypes of bi-allelic SNPs mapped to hg19 from 10,913 ADSP participants were called using the quality-controlled Atlas-only pipeline at Baylor College of Medicine (We did not use the data from the GATK pipeline at the Broad institute due to known quality issues (https://www.niagads.org/adsp/data-notices)). More details about the production of the WES data in ADSP can be found in ref. [18]. Variants with a missing rate >2% or MAC ≤10 were excluded from the age-of-onset association analyses. After the filtering, 110,450 and 98,334 variants remained in the analysis using all subjects and *APOE ε4* non-carriers, respectively. In the replication study, VCF files of recalibrated WGS data from 1196 participants in ROSMAP were downloaded from the synapse website (https://www.synapse.org/). A total of 681 subjects were included in the replication phase after removing 16 discordant WGS samples, 17 duplicates, and 477 subjects overlapping the ADSP sample. WGS project level genotype VCF files (hg38) called by GATK in the ADSP extension study were downloaded from NIAGADS (https://dss.niagads.org/datasets/ng00067/), from which the genotypes of 1147 non-Hispanic whites were extracted. Genotyping of 3043 participants in CHS was performed using an Illumina HumanCNV370v1 array (~370 K SNPs). Genotyping of 3456 non-Hispanic Caucasian participants in NIA-LOADFS was performed using a Human610-Quad Illumina array (~600 K SNPs). Genotyping of 1588 non-Hispanic Caucasian participants in GenADA was performed using two Affymetrix 250K arrays (a total of ~500 K SNPs). More information about these cohorts can be found in refs. [72–74]. We phased and imputed the genotypes in the three array-based cohorts using the TOPMED imputation server[75] with the TOPMed reference panel (Version R2 on GRC38)[76].

Exome-wide age-of-onset association analysis

The association analyses of the age-of-onset of AD in the discovery phase of ADSP was conducted using a Cox mixed-effects model implemented in the coxmeg R package[21], which accounted for the clustering structure using a GRM. A dense GRM was first estimated from the original

WES data based on the GCTA model[77] implemented in the SNPRelate R package[78]. In the discovery phase of ADSP, we built a sparse GRM by setting any entry below 0.03 to zero. We evaluated ten top PCs (PC1 to PC10) calculated from the dense GRM, and included the only significant PC2, PC8, and PC10 in the analyses. We first estimated a variance component in the null model, which was then used to estimated HRs and *p*-values for all SNPs. We performed two analyses, (a) including all subjects with the three PCs, sex and the number of copies of *APOE ε4* included as covariates, (b) including only *APOE ε4* non-carriers with the three PCs and sex included as covariates. We found that the estimated variance component was zero in the analysis (b), suggesting no evidence of random effects, and therefore we instead used a simple Cox model. The threshold to declare significant associations was calculated as 0.05 divided by the total number of tested SNPs. For comparison with the analysis of AD status, we performed association analysis by fitting a logistic regression using the glm R function adjusting for the same covariates with the same sample.

We performed age-of-onset association analyses in LOADFS, CHS, ROSMAP, GenADA, and the ADSP extension study for the top SNPs passing the suggestive threshold ($p < 5E{-}06$) in the discovery phase. The same model and estimation procedures as in ADSP were used in LOADFS, which is also a family-based cohort. In LOADFS, the GRM was estimated from the genotype array data. The association analyses were conducted in the other four cohorts (i.e., CHS, ROSMAP, GenADA and the ADSP extension study) using a Cox model implemented in the survival R package[30] because these cohorts consisted of unrelated subjects. We also included sex and the number of copies of *APOE ε4* as covariates. Meta-analysis effect sizes and standard errors were computed using the summary statistics from all six studies based on the following fixed-effects model, $\beta = \sum_i \beta_i w_i / \sum_i w_i$ and $sd(\beta) = 1/\sqrt{\sum_i w_i}$, where w_i is the weight for the study i. To compare age-of-onset analysis with case-control analysis, we also performed association analyses of AD status in ADSP using logistic regression.

Gene-based association analysis

The gene-based analysis was performed based on the summary statistics obtained from the age-of-onset association analyses. We only included SNPs with MAC >10 and a missing rate <2% in the gene-based analyses. Each SNP was first annotated to a gene using its SNP ID according to a gene location file obtained in the MAGMA website (https://ctg.cncr.nl/software/magma). We only included SNPs within the boundary of a gene body. Gene-based *p*-values were then computed using MAGMA (v1.08b) with a SNP-wise mean model[47]. LD between the SNPs was estimated using the raw WES data in ADSP.

Gene-set and cell-type association analysis

The gene-set analysis was performed for curated gene sets and GO terms using the procedure SNP2GENE in FUMA[48] based on the summary statistics obtained from the age-of-onset association analyses. The 1000 Genomes Project (phase 3) for the European population was used as a reference panel in the analysis. The cell-type association analysis was also performed using FUMA[79] following the SNP2GENE procedure. We selected a human brain single-cell RNA-seq dataset provided in ref. [80] as a reference for cell type-specific gene expression.

Analysis of FDG-PET data

The longitudinal FDG-PET average intensity scores across five ROIs (left/right angular gyrus, bilateral posterior cingulate gyrus, and left/right inferior temporal gyrus) for 738 subjects in ADNI having the WGS data were downloaded from the ADNI website (https://ida.loni.usc.edu). Details about sample preparation and data generation were described in refs. [33,34]. The association analysis between average FDG-ROI and the genotype of rs56201815 was performed by fitting a linear mixed-effects model using lme4 R package[81] including a random effect accounting for within-subject variability and three covariates (age, sex, and diagnosis group).

Analysis of tissue-specific RNA-seq and microarray data

BAM files of aligned reads from a total of 2213 RNA-seq samples in three brain regions (dorsolateral PFC, PCC, and anterior caudate nucleus) in the ROSMAP project were downloaded from the synapse website (https://www.synapse.org/). Raw counts of 57,905 coding and non-coding genes were called using featureCounts[82] according to the GENCODE annotations GRCh37(r87). Samples with the RNA integrity number (RIN) < 5 were excluded before the analysis. We first removed low-expressed genes (those genes for which fewer than three individuals had counts-per-million >1) before normalization. We then normalized the RNA-seq raw counts using the trimmed mean of *M*-values (TMM) normalization method[83]. In the analysis of PFC, 761 non-Hispanic Caucasian subjects (including four rs56201815-G carriers) having both gene expression and genotype of rs56201815 from the WGS data with RIN ≥4.5 were included. Differential eQTL analysis was performed using edgeR[84,85] adjusted for RIN, age at death, sex, AD status, and RNA extraction methods (polyA selection or rRNA depletion). In the analysis of PCC and anterior caudate nucleus, 371 (including three rs56201815-G carriers) and 585 (including four rs56201815-G carriers) non-Hispanic Caucasian subjects having both genotypes and gene expression with RIN ≥4.5 and rRNA depletion were included, respectively. To minimize technical noise resulted from sample preparation, we did not include polyA selection samples

(accounting for merely 10% and 15% of all samples) because different RNA extraction methods have a large impact on measured expression in postmortem samples[86], and the samples of all rs56201815-G carriers were generated using rRNA depletion. Differential eQTL analysis was performed using edgeR adjusted for RIN, age at death, sex, and AD status.

The raw count data of 3252 RNA-seq samples in nine brain tissues (i.e., amygdala, ACC, hypothalamus, caudate (basal ganglia), nucleus accumbens (basal ganglia), putamen (basal ganglia), cerebellar hemisphere, cerebellum, and spinal cord (cervical c1)) and four non-brain tissues (i.e., sigmoid colon, lung, spleen, and whole blood) from the GTEx project (version 8) were downloaded from the GTEx portal (https://gtexportal.org/home/datasets). Gene-level quantification was conducted by RSEM[87]. All GTEx raw count data were normalized using the same pipeline as in the analysis of ROSMAP. Differential eQTL analysis was then performed using edgeR with age, sex, and RIN as adjusted covariates.

The gene expression microarray data in peripheral blood from 742 ADNI subjects were profiled using the Affymetrix Human Genome U219 Array. Raw expression values were pre-processed using the robust multiarray average normalization method. More details about sample collection and data pre-processing can be found in ref. [88]. Differential gene expression analyses were performed using linear regression adjusted for RIN and plate number.

Analysis of DNA methylation data

The DNA methylation data in PFC were collected from 740 individuals in ROSMAP using the Illumina HumanMethylation450 BeadChip. Eighteen samples lying beyond ±3 standard deviations for the top three PCs were removed as outliers. We converted methylation beta-value to *M*-value using a logistic transformation. Differential methylation analysis was carried out using a linear regression adjusted for the top ten PCs.

Analysis of H3K9ac ChIP-seq data

H3K9ac ChIP-seq raw count data were downloaded from the synapse website (https://www.synapse.org/). This dataset is previously described in detail in ref. [40]. Briefly, the sample comprising 26,384 H3K9ac peaks (nine peaks in the *ERN1* region) across the genome was collected from dorsolateral PFC of 669 subjects from the ROSMAP project, among which 625 subjects had also the WGS genotype data of rs56201815. The raw count data were normalized using the TMM method[83]. Estimation of common and tagwise dispersions and the analysis of differential peaks for rs56201815 were carried out using edgeR[84,85] adjusted for FRiPs and GC bias. A sensitivity analysis was performed by further adjusting for ten RUV components estimated using RUVSeq[89].

Analysis of snRNA-seq data

We collected snRNA-seq raw count data generated by ref. [44] using the 10X Genomics Cell ranger pipeline in human PFC from 48 subjects (50% AD cases) including 17,926 genes profiled in 75,060 nuclei. We assigned cell identity and divided all cells into six subtypes (excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, microglia, and OPCs) according to the previous clustering results[44] using the scanpy package[90]. The clustering of the cells is described in more detail in ref. [44]. We excluded endothelial cells or pericytes because of the lack of abundant cell counts in these two cell types.

To perform cell type-specific eQTL analysis, we first merged cells in each cell type and in each subject to obtain a raw count matrix of 17,926 genes and 39 subjects (six subjects were excluded due to lack of WGS data). We then followed the preprocessing and normalization procedures in the previous eQTL analysis of the bulk RNA-seq data. Differential eQTL analyses were then performed using edgeR[84,85] with age, sex, and AD status as covariates. RIN was not available for most of the subjects.

Functional annotation

The epigenetic and regulatory annotation of the identified SNPs and its nearby SNPs in high LD ($r^2 > 0.8$) was performed using Haploreg v4 (ref. [91]), in which its tissue-specific epigenetic markers (H3K27ac), regulatory regions (enhancers and promoters), motif changes, and eQTL information were annotated based on the ENCODE[92], Roadmap[93], and GTEx[42] projects. GWAS catalog[93] and GRASP[94] were used to annotate whether a SNP is an existing QTL.

Acknowledgements

This manuscript was prepared using limited access datasets obtained through dbGaP (accession numbers: phs000168.v2.p2 (LOADFS), phs000572.v8.p4 (ADSP), phs000287.v5.p1 (CHS), phs000219.v1.p1 (GenADA), phs00424.v8.p1 (GTEx), NG00067 (the ADSP extension study)).

This research was supported by Grants from the National Institutes on Aging R01 AG047310, R01 AG061853, AG065477, and AG070488 to A.M.K. and RF1 AG054012, R01 AG058002, R01 AG062335, RF1 AG062377, U01 NS110453 to M.K., P30 AG10161, R01 AG15819, R01 AG17917, R01 AG36042, and R01 AG61356 to D.A.B.

The funders had no role in study design, data collection, and analysis, decision to publish, or manuscript preparation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

We are grateful to the participants in the Religious Order Study, the Rush Memory and Aging Project. Data can be requested at www.radc.rush.edu. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

See also further acknowledgements in Supplementary materials Text S1.

Author details
[1]Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA. [4]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

Author contributions
L.H. conceived the study. L.H. and Y.L. imputed the genotype data. L.H. performed the age-of-onset association analyses, gene-based analyses, microarray, RNA-seq, snRNA-seq, and ChIP-seq analyses. Y.P. analyzed the DNA methylation data. D.A.B. generated WGS, DNA methylation, H3K9Ac, and RNA-seq data in ROSMAP. A.K. and M.K. contributed to acquiring the data, and discussing of final results. All authors contributed to the writing of the manuscript.

Conflict of interest
The authors declare no competing interests.

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

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-021-01263-4.

Received: 7 July 2020 Revised: 7 January 2021 Accepted: 3 February 2021
Published online: 26 February 2021

References

1. Winblad, B. et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurol.* **15**, 455–532 (2016).
2. Goate, A. et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704–706 (1991).
3. Levy-Lahad, E. et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973–977 (1995).
4. Mullan, M. et al. A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N-terminus of beta-amyloid. *Nat. Genet.* **1**, 345–347 (1992).
5. Rogaev, E. I. et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376**, 775–778 (1995).
6. Sherrington, R. et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754–760 (1995).
7. Guerreiro, R. et al. TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* **368**, 117–127 (2013).
8. Harold, D. et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* **41**, 1088–1093 (2009).
9. Hollingworth, P. et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* **43**, 429–435 (2011).
10. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404 (2019).
11. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
12. Lambert, J. -C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452 (2013).
13. Naj, A. C. et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* **43**, 436–441 (2011).
14. Deming, Y. et al. The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci. Transl. Med.* **11**, eaau2291 (2019).
15. Griciuc, A. et al. TREM2 acts downstream of CD33 in modulating microglial pathology in Alzheimer's disease. *Neuron* **103**, 820–835.e7 (2019).
16. Gratuze, M., Leyns, C. E. G. & Holtzman, D. M. New insights into the role of TREM2 in Alzheimer's disease. *Mol. Neurodegener.* **13**, 66 (2018).
17. Raghavan, N. & Tosto, G. Genetics of Alzheimer's disease: the importance of polygenic and epistatic components. *Curr. Neurol. Neurosci. Rep.* **17**, 78 (2017).
18. Bis, J. C. et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Mol. Psychiatry* https://doi.org/10.1038/s41380-018-0112-7 (2018).
19. Cruchaga, C. et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**, 550–554 (2014).
20. Raghavan, N. S. et al. Whole-exome sequencing in 20,197 persons for rare variants in Alzheimer's disease. *Ann. Clin. Transl. Neurol.* **5**, 832–842 (2018).
21. He, L. & Kulminski, A. M. Fast algorithms for conducting large-scale GWAS of age-at-onset traits using cox mixed-effects models. *Genetics* **215**, 41–58 (2020).
22. Belloy, M. E., Napolioni, V. & Greicius, M. D. A quarter century of APOE and Alzheimer's disease: progress to date and the path forward. *Neuron* **101**, 820–838 (2019).
23. Yamazaki, Y. et al. and Alzheimer disease: pathobiology and targeting strategies. *Nat. Rev. Neurol.* **15**, 501–518 (2019).
24. Crane, P. K., Foroud, T., Montine, T. J. & Larson, E. B. Alzheimer's Disease Sequencing Project discovery and replication criteria for cases and controls: data from a community-based prospective cohort study with autopsy follow-up. *Alzheimers Dement.* **13**, 1410–1413 (2017).
25. Blacker, D. et al. ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. *Neurology* **48**, 139–147 (1997).
26. Naj, A. C. et al. Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. *JAMA Neurol.* **71**, 1394–1404 (2014).
27. Sando, S. B. et al. APOE epsilon 4 lowers age at onset and is a high risk factor for Alzheimer's disease; a case control study from central Norway. *BMC Neurol.* **8**, 9 (2008).
28. Jun, G. et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol. Psychiatry* **21**, 108–117 (2016).
29. Bennett, D. A. et al. Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).
30. Therneau, T. M. & Lumley, T. Package 'survival'. *R Topics Documented* 128. https://cran.r-project.org/web/packages/survival/survival.pdf (2015).
31. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
32. Caminiti, S. P. et al. FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort. *NeuroImage Clin.* **18**, 167–177 (2018).
33. Landau, S. M. et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol. Aging* **32**, 1207–1218 (2011).
34. Landau, S. M. et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* **75**, 230–238 (2010).
35. Nozadi, S. H., Kadoury, S. & The Alzheimer's Disease Neuroimaging Initiative. Classification of Alzheimer's and MCI patients from semantically parcelled PET images: a comparison between AV45 and FDG-PET. *Int. J. Biomed. Imaging* **2018** (2018).
36. Shivamurthy, V. K. N., Tahari, A. K., Marcus, C., Subramaniam, R. M. & Brain, F. D. G. PET and the diagnosis of dementia. *Am. J. Roentgenol.* **204**, W76–W85 (2014).
37. Duran-Aniotz, C. et al. IRE1 signaling exacerbates Alzheimer's disease pathogenesis. *Acta Neuropathol. (Berl.)* **134**, 489–506 (2017).
38. De Jager, P. L. et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).
39. De Jager, P. L. et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
40. Klein, H.-U. et al. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.* **22**, 37–46 (2019).
41. Gibbs, J. R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
42. GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
43. Zou, F. et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.* **8**, e1002707 (2012).
44. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
45. Arrant, A. E., Filiano, A. J., Unger, D. E., Young, A. H. & Roberson, E. D. Restoring neuronal progranulin reverses deficits in a mouse model of frontotemporal dementia. *Brain* **140**, 1447–1465 (2017).

46. Holler, C. J., Taylor, G., Deng, Q. & Kukar, T. Intracellular proteolysis of progranulin generates stable, lysosomal granulins that are haploinsufficient in patients with frontotemporal dementia caused by GRN mutations. *eNeuro* **4** (2017).
47. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
48. Watanabe, K., Taskesen, E., Bochoven, Avan & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1–11 (2017).
49. Watanabe, K., Umićević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P. & Posthuma, D. Genetic mapping of cell type specificity for complex traits. Nat. Commun. 10, 3222 (2019).
50. Calfon, M. et al. IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* **415**, 92–96 (2002).
51. Lee, K. et al. IRE1-mediated unconventional mRNA splicing and S2P-mediated ATF6 cleavage merge to regulate XBP1 in signaling the unfolded protein response. *Genes Dev.* **16**, 452–466 (2002).
52. Ma, T. et al. Suppression of eIF2α kinases alleviates Alzheimer's disease-related plasticity and memory deficits. *Nat. Neurosci.* **16**, 1299–1305 (2013).
53. Naidoo, N., Ferber, M., Master, M., Zhu, Y. & Pack, A. I. Aging impairs the unfolded protein response to sleep deprivation and leads to proapoptotic signaling. *J. Neurosci.* **28**, 6539–6548 (2008).
54. Paz Gavilán, M. et al. Cellular environment facilitates protein accumulation in aged rat hippocampus. *Neurobiol. Aging* **27**, 973–982 (2006).
55. Sprenkle, N. T., Sims, S. G., Sánchez, C. L. & Meares, G. P. Endoplasmic reticulum stress and inflammation in the central nervous system. *Mol. Neurodegener.* **12**, 42 (2017).
56. Zhang, K. & Kaufman, R. J. From endoplasmic-reticulum stress to the inflammatory response. *Nature* **454**, 455–462 (2008).
57. Ikram, M. A. et al. Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat. Genet.* **44**, 539–544 (2012).
58. Kouri, N. et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nat. Commun.* **6**, 7247 (2015).
59. Do, C. B. et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.* **7**, e1002141 (2011).
60. Hamza, T. H. et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet.* **42**, 781–785 (2010).
61. Lill, C. M. et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. *PLoS Genet.* **8**, e1002548 (2012).
62. Pankratz, N. et al. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann. Neurol.* **71**, 370–384 (2012).
63. Liu, J. Z. et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **44**, 1137–1141 (2012).
64. van der Harst, P. et al. Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369–375 (2012).
65. Li, R. et al. Six novel susceptibility Loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS Genet.* **8**, e1002746 (2012).
66. Estrada, K. et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet.* **44**, 491–501 (2012).
67. Morris, J. A. et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266 (2019).
68. Kulminski, A. M. Unraveling genetic origin of aging-related traits: evolving concepts. *Rejuvenation Res.* **16**, 304–312 (2013).
69. Nesse, R. M., Ganten, D., Gregory, T. R. & Omenn, G. S. Evolutionary molecular medicine. *J. Mol. Med. (Berl.)* **90**, 509–522 (2012).
70. Hebert, L. E. et al. Age-specific incidence of Alzheimer's disease in a community population. *JAMA* **273**, 1354–1359 (1995).
71. Beecham, G. W. et al. The Alzheimer's Disease Sequencing Project: study design and sample selection. *Neurol. Genet.* **3**, e194 (2017).
72. Gottdiener, J. S. et al. Predictors of congestive heart failure in the elderly: the cardiovascular health study. *J. Am. Coll. Cardiol.* **35**, 1628–1637 (2000).
73. Li, H. et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch. Neurol.* **65**, 45–53 (2008).
74. Wijsman, E. M. et al. Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. *PLoS Genet.* **7**, e1001308 (2011).
75. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
76. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2019).
77. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
78. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
79. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
80. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, https://doi.org/10.18637/jss.v067.i01 (2014).
81. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
82. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
83. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
84. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
85. Sigurgeirsson, B., Emanuelsson, O. & Lundeberg, J. Sequencing degraded RNA addressed by 3′ tag counting. *PLoS ONE* **9**, e91851 (2014).
86. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
87. Saykin, A. J. et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. *Alzheimers Dement.* **11**, 792–814 (2015).
88. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
89. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
90. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877–D881 (2016).
91. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
92. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
93. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
94. Eicher, J. D. et al. GRASP v2.0: an update on the genome-wide repository of associations between SNPs and phenotypes. *Nucleic Acids Res.* **43**, D799–D804 (2015).

Author details

Genotype Tissue Expression (GTEx) consortium

Yongjin Park[2,3], Manolis Kellis[2,3]