Research Article

MOLECULAR BIOLOGY & GENETICS

Complement C7 is a novel risk gene for Alzheimer's disease in Han Chinese

Deng-Feng Zhang[1,2,14], Yu Fan[1,2,14], Min Xu[1,3,14], Guihong Wang[4], Dong Wang[1], Jin Li[5,6], Li-Li Kong[1,3], Hejiang Zhou[1], Rongcan Luo[1,3], Rui Bi[1], Yong Wu[1,3], Guo-Dong Li[1,3], for the Alzheimer's Disease Neuroimaging Initiative (ADNI) [*], Ming Li[1,7], Xiong-Jian Luo[1,2], Hong-Yan Jiang[8], Liwen Tan[9], Chunjiu Zhong[10], Yiru Fang [11], Chen Zhang[11], Nengyin Sheng[2,12], Tianzi Jiang[5,6,7], Yong-Gang Yao[1,3,7,13, #]

[1] Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

[2] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

[3] Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China

[4] Center for Neurodegenerative Diseases, Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100050, China

[5] Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[6] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[7] CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China

[8] Department of Psychiatry, the First Affiliated Hospital of Kunming Medical University, Kunming 650032, China

[9] Mental Health Institute of the Second Xiangya Hospital, Central South University, Changsha 410011, China

[10] Department of Neurology, Zhongshan Hospital, Fudan University, Shanghai, 200032, China

[11] Division of Mood Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China

[12] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

[13] KIZ–CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China

[*] Data used in preparation of this article were partly obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in the analysis or the writing of this report. A complete listing of the ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

[14] These authors contributed equally to this work.

Correspondence: Prof. Yong-Gang Yao, Kunming Institute of Zoology, Chinese Academy of Sciences, Tel/Fax: 86-871-65180085; e-mail: yaoyg@mail.kiz.ac.cn.

Running title: Rare variants in *C7* and Alzheimer's risk

Abstract

Alzheimer's disease is the most common neurodegenerative disease and has a high level of genetic heritability and population heterogeneity. In this study, we performed the whole exome sequencing of Han Chinese patients with familial and/or early onset Alzheimer's disease, followed by independent validation, imaging analysis, and function characterization. We identified an exome-wide significant rare missense variant rs3792646 (p.K420Q) in the *C7* gene in the discovery stage ($P = 1.09\times10^{-6}$, odds ratio = 7.853) and confirmed the association in different cohorts and combined sample (1,615 cases and 2,780 controls, $P_{combined} = 2.99\times10^{-7}$, odds ratio = 1.930). The risk-allele was associated with decreased hippocampal volume and poorer working memory performance in early adulthood, thus resulting in an earlier age of disease onset. Overexpression of the mutant p.K420Q disturbed cell viability, immune activation, and β-amyloid processing. Electrophysiological analyses showed that the mutant p.K420Q impairs the inhibitory effect of wild-type C7 on the excitatory synaptic transmission in pyramidal neurons. These findings suggested that *C7* is a novel risk gene for Alzheimer's disease in Han Chinese.

Key words: Alzheimer’s disease; whole exome sequencing; C7; neuroimaging; complement system

Abbreviations:

AAO = age at onset; ADNI = Alzheimer’s Disease Neuroimaging Initiative; ADSP = Alzheimer’s Disease Sequencing Project; Aβ = β-amyloid; EPSC = evoked excitatory postsynaptic current; ExAC = Exome Aggregation Consortium; GWAS = genome-wide association study; HM = human microglia; IPSC = inhibitory post-synaptic current; MAF = minor allele frequency; OR = odds ratio; PC = principal component; SNP = single nucleotide polymorphism; WES = whole exome sequencing

Introduction

Alzheimer's disease is the most common neurodegenerative disease in the elderly and is becoming a serious global health problem [1-3]. It is characterized by cognitive impairment resulting from extracellular β-amyloid (Aβ) plaques, intracellular neurofibrillary tangles (hyperphosphorylated tau), and cerebral atrophy [1-3]. Both genetic and environmental factors contribute to the onset and development of the disease, and its heritability was reported to be up to 0.79 [4-6]. Previous linkage analyses have revealed genes involved in the production of Aβ plaques, namely *APP* (β-amyloid precursor protein), *PSEN1* (Presenilin-1) and *PSEN2* (Presenilin-1), as the causal genes for early-onset familial Alzheimer's disease [5, 7-15]. However, mutations of these genes are mainly associated with autosomal dominant type, and account only for less than 5% of total cases [5, 16]. In fact, it is believed that in most cases the disease is polygenic and there are other causal and/or susceptibility genes remaining to be discovered [4, 17]. Recent genome-wide association studies (GWASs) have reported two dozen Alzheimer's susceptibility genes in populations of European ancestry, including *APOE*, *BIN1*, *CLU*, and *RIN3* [18, 19]. Nevertheless, most of the GWAS loci are non-coding common variants / SNPs (single nucleotide polymorphisms) with unknown function and show small to moderate effect sizes (odds ratio [OR] < 1.2). Since these GWAS hits can only explain about

16% of the total phenotypic variance [17], the missing heritability remains to be explained by other underlying variants (especially functionally causative variants) [20].

Recent advances in the next-generation sequencing technologies offer powerful tools for the discovery of rare causal variants with larger effect sizes in Alzheimer's disease [21], and previous studies have identified *UNC5C* [22], *TREM2* [23, 24], *PLD3* [25, 26] , *PLCG2* and *ABI3* [27] to be the top candidate genes harboring such variants. However, despite the success in applying the next-generation sequencing technologies, population heterogeneity has limited the success in characterizing the genetic basis of Alzheimer's disease [28, 29]. For example, many of top hits in the European populations identified by GWAS or next-generation sequencing technologies cannot be validated in Chinese populations [26, 30-32]. The investigation for genetic susceptibility of Alzheimer's disease at the whole genomic level in Han Chinese, the largest ethnic population in the world with the greatest number of Alzheimer's disease sufferers [33, 34], is therefore urgently needed. To this end, we have performed the whole exome sequencing (WES) of patients with Alzheimer's disease in Han Chinese to identify novel susceptibility genes.

Results

Identification of *C7* as a novel Alzheimer's risk gene in Han Chinese

We took an extreme phenotype sampling strategy for WES to increase the likelihood of identifying true disease-related variants [35, 36], followed by independent validations and functional characterization (Fig. 1A). 107 unrelated patients with an early age at onset [AAO] of Alzheimer's disease (AAO<= 55) and/or a positive familial history were selected from over 1,000 genetically unrelated patients from East and Southwest China [26, 32, 37-41]. 160 in-house non-dementia individuals [42], together with the whole genome data of Han Chinese in Beijing (N=103) and Southern Han Chinese (N=105) from the 1000 Genomes Project phase 3 [43], were combined as the initial population control (N = 368) based on the fact that principal component analysis showed no apparent population stratification between the studied subjects and the reference Chinese populations from the 1000 Genomes Project phase 3 [43] (Supplementary Fig. 1). Nonsense, frameshift, splice site variants, and missense variants that were predicted to be damaging by at least one of the five algorithms (PolyPhen2 HumDiv and HumVar [44], LRT [45], MutationTaster [46], and SIFT [47, 48]), were defined as functional. As we aimed at identifying novel rare coding variants that were associated with Alzheimer's disease or enriched in patients, we filtered out the common variants and obtained 23,373 rare or low-frequency coding variants with a minor allele frequency (MAF) < 5% in the 368 pooled population controls [42, 43]. The Bonferroni-correction-based threshold for the exome-wide significance was thus set as $P < 2.139\times10^{-6}$ (0.05/23373).

We were able to successfully validate the previously reported association of *APOE* ε4 with Alzheimer's disease [18] (rs429358, $P = 3.41\times10^{-9}$, odds ratio [OR] = 3.59) in our initial WES screening stage (Fig. 1B and Supplementary Table 1), suggesting the reliability of the current extreme phenotype sampling approach. One rare variant in the complement *C7* gene, rs3792646, leading to the missense mutant p.K420Q with a predicted damaging effect (CADD (Combined Annotation-Dependent Depletion) score = 20.3) [49], was the

only exome-wide significant hit showing association with susceptibility of Alzheimer's disease ($P = 1.09\times10^{-6}$, OR = 7.853, Fisher's exact test; Fig. 1B, Table 1 and Supplementary Table 1) with the exception of *APOE* rs429358. When the 107 patients (MAF = 0.08) were compared with 4,327 East Asians from the Exome Aggregation Consortium (ExAC) [50] (MAF = 0.036; $P = 2.95\times10^{-3}$, OR = 2.292, Fisher's exact test) or 11,670 Chinese individuals in the CONVERGE data [51] (MAF = 0.029; $P = 2.200\times10^{-4}$, OR = 2.884, Fisher's exact test), the positive association between rs3792646 and Alzheimer's disease survived (Table 1). To avoid false positive result, we performed the logistic regression analyses with adjustment of first three principal components (PC1, PC2, and PC3) (Supplementary Fig. 1), APOE ε4 status, and sex as the covariate(s) separately or together. The association of *APOE* ε4 with Alzheimer's disease remained to be significant at the exome-wide level after adjustment with PC1-PC3 ($P_{adj\text{-}PC} = 1.95\times10^{-7}$, $OR_{adj\text{-}PC} = 3.20$). The *C7* variant rs3792646 remained to be one of the top hits after adjustment using different covariates ($P_{adj\text{-}PC}= 9.508\times10^{-5}$, $P_{adj\text{-}sex}= 7.30\times10^{-6}$, $P_{adj\text{-}APOE}= 5.0\times10^{-4}$, and $P_{adj\text{-}PC\text{-}sex\text{-}APOE}= 9.90\times10^{-4}$), although no exome-wide significant hits were observed partially because of small sample size (Supplementary Table 1). We listed the top 100 rare functional variants showing suggestive associations with Alzheimer's disease (Fisher's exact test, $P < 0.01$) in the Supplementary Table 1. The summary statistics could be freely accessed through the webserver AlzData (http://www.alzdata.org/exome.html) [52].

In addition to the single-site evidence, the gene-level association based on the burden test showed that *C7* had an enrichment of rare missense variants in Alzheimer's patients compared with controls ($P = 2.28\times10^{-4}$; Fig. 1C). The SNP-set (Sequence) Kernel Association Tests (SKAT) [53] yielded even stronger association ($P = 5.83\times10^{-7}$) of combined effect of rare *C7* variants. When rs3792646 was excluded from the burden test, the significance of enrichment disappeared, suggesting that the signal might be driven by rs3792646. As there was a variant rs2271708 (p.C128R) overrepresenting in controls (Fig. 1C), we recalculated the burden test excluding both rs3792646 and rs2271708 and observed a marginally significant enrichment of rare variants in cases ($P = 0.02$), suggesting the existence of multiple effect alleles in *C7*.

Validation of the association of rs3792646 with Alzheimer's disease in Han Chinese

To validate the association between *C7* rs3792646 and early-onset and familial Alzheimer's disease identified during the discovery WES screening (stage 1), we sequenced this SNP in an independent Han Chinese sample with early-onset and/or familial Alzheimer's disease from Beijing (stage 2, N = 103 cases). The association of rs3792646 with Alzheimer's disease could be well-validated ($P= 6.10\times10^{-4}$, OR = 5.133, Table 1). Combing these two samples of patients with early-onset and/or familial Alzheimer's disease together, we observed a stronger association of rs3792646 with Alzheimer's disease ($P= 3.73\times10^{-7}$, OR = 6.500).

We then attempted to validate the association between rs3792646 and Alzheimer's disease in Chinese cohorts of sporadic patients (stage 3): the East China cohort contains 584 sporadic cases and 274 geographically matched controls; the Southwest China cohort contains 581 sporadic cases and 2190 geographically matched controls. We also analyzed a patient sample from Hunan, Southcentral China (N = 235 sporadic cases). Positive associations were observed in the East China cohort ($P = 3.53\times10^{-3}$, OR = 3.025, Fisher's exact test) and the Southwest China cohort ($P = 1.17\times10^{-2}$, OR = 1.594, Fisher's exact test). In the sample from Hunan Province, Southcentral

China no association with Alzheimer's disease was observed (P = 0.22, OR = 1.394, Fisher's exact test), but the risk effect remained in this relatively small sample. When we combined the subjects from Southwest and Southcentral China together, a positive association was observed ($P = 1.08\times10^{-2}$, OR = 1.532, Fisher's exact test). Though the associations from single validation cohorts did not reach the exome-wide significance, combining all samples from stage 1 to stage 3 together resulted in an exome-wide significant association between rs3792646 and Alzheimer's risk with a considerably large effect size ($P_{combined} = 2.99\times10^{-7}$, OR = 1.930).

Notably, we observed positive associations in both early-onset (AAO =< 65 years old; $P = 3.10\times10^{-4}$, OR = 2.066) and late-onset subjects (AAO > 65 years old; $P = 8.11\times10^{-6}$, OR = 1.883), with a stronger effect size in the early-onset patients (Table 1). When the patients were divided into different groups according to the *APOE* ε4 status, we observed positive associations of rs3792646 with Alzheimer's risk in both *APOE* ε4 carriers and non-carriers (Supplementary Table 2), and a stronger association was found in the *APOE* ε4 carriers ($P_{combined} = 1.43\times10^{-5}$, OR = 3.651) than non-carriers ($P_{combined} = 1.22\times10^{-3}$, OR = 1.770) (Supplementary Table 2).

Association of rs3792646 with Alzheimer's disease might be Chinese-specific

While we have confirmed the association between *C7* rs3792646 with Alzheimer's disease in Han Chinese, it is unclear whether it is Chinese-specific or not. We therefore re-analyzed the whole genome sequencing data of 812 individuals of European ancestry from the ADNI dataset [54]. There were six *C7* mutation carriers (including 3 rare damaging missense variants; Supplementary Table 3) in 296 patients and two carriers in 281 controls in the ADNI cohort (gene-based P = 0.29, OR = 2.886), suggesting a higher frequency of *C7* mutations in European patients [54], albeit the pattern might be different from that in Han Chinese. Among them, rs3792646-C (p.K420Q) and chr5:40936541 C>T (p.C128R) occurred one and five times respectively in 296 patients with Alzheimer's disease or late-stage mild cognitive impairment; whereas in the 281 controls, no individual harbored p.K420Q, and only one individual with p.C128R was found. Though there was a seemingly trend of *C7* mutation in patients, the enrichment was not significant (p.K420Q, P = 0.33; p.C128R, P = 0.20). We also retrieved the summary statistics of the International Genomics of Alzheimer's Project [18], a large GWAS meta-analysis of Alzheimer's disease (17,008 cases versus 37,154 controls), to investigate the association between *C7* variants and Alzheimer's disease in Europeans. No nominally significant *C7* SNPs were observed. In the recently released Alzheimer's Disease Sequencing Project (ADSP) cohort [55], there were also no significant exonic variants in *C7* showing an association with Alzheimer's disease. Only one p.K420Q carrier was found in the ADSP cohort [55] that contains 10,570 individuals of European ancestry. These results were consistent with the low allele frequency of rs3792646 in the non-Chinese populations from ExAC [50]: 0.0020 in 5,041 Ashkenazi Jewish population, 0.0020 in 14,972 South Asians, 0.00093 in 16,649 Latinos, 0.00034 in 62,858 non-Finnish Europeans, 0.00013 in 11,946 Africans, and 0.000039 in Finnish population (http://gnomad.broadinstitute.org/variant/5-40955653-A-C, accessed on Feb 8, 2018), suggesting that this variant is most likely Chinese or East Asian specific.

Association of rs3792646 with Alzheimer's-related endo-phenotypes and preclinical impairments

In addition to its effect on disease risk, we investigated whether rs3792646 affects the age of disease onset in our combined Han Chinese samples with available AAO information. The survival test showed that carriers of the risk allele rs3792646-C had a significant (Log-rank test, $P = 2.04\times10^{-2}$) earlier onset age (51 years) than carriers of rs3792646-AA (55 years) in Han Chinese patients with an AAO < 60 years (Fig. 2A). No significant difference of the AAO was observed in patients with late-onset sporadic Alzheimer's disease.

In order to discern whether the risk allele rs3792646-C would have a potential effect on brain structure and function of susceptible individuals in early adulthood, we took advantage of the imaging data that were previously collected in 360 healthy university students [32, 56] and analyzed the association of rs3792646 with brain structural changes and working memory performance. Intriguingly, the rs3792646-C carriers (genotypes CC and AC) showed significantly lower right hippocampus volume ($P = 0.02$) and worse working memory performance ($P = 0.03$) compared with the AA carriers (Fig. 2B-C). These observations indicated that the *C7* variant rs3792646 might affect the brain function of at-risk Han Chinese individuals several decades before disease onset.

The effect of *C7* variants on Alzheimer-related endo-phenotypes were further investigated using the ADNI data [54]. We observed a lower hippocampus volume in the only one p.K420Q carrier in the ADNI cohort [54] (Supplementary Fig. 2). Though the association between the disease and *C7* SNPs was not established in population of European origin, two *C7* variants (Supplementary Table 3 and Supplementary Fig. 2) did affect the cerebrospinal fluid Aβ and p-tau levels in the ADNI cohort [54]. In particular, carriers of p.C128R had a higher phosphorylated tau level in the cerebrospinal fluid (Supplementary Fig. 2), supporting the risk-promoting effect of *C7* variants in Alzheimer's disease.

Upregulation of *C7* mRNA expression in brain tissues of Alzheimer's disease

C7 is a component of the terminal complement cascade and physically interacts with the GWAS hit Clusterin (CLU) [57]. To characterize the involvement of *C7* in Alzheimer's disease, we analyzed the mRNA expression pattern of the complement cascades in frontal cortex tissues from patients and controls based on dataset GSE33000 [58]. All initial (e.g. *C1QA*, $P = 1.8\times10^{-18}$) and central (e.g. *C3*, $P = 4.01\times10^{-9}$) components of the complement cascades were significantly upregulated, whereas of the terminal complement components only *C7* was significantly upregulated in patients ($P = 3.21\times10^{-15}$, log2 Fold Change = 0.242) (Supplementary Table 4 and Fig. 3A). Consistently, we observed an early increase and a strong positive correlation of *C1q* and *C3* mRNA expression level with the severity of pathological changes (Aβ plaques and tau tangles) in hippocampus of Alzheimer's disease mouse models based on the Mouseac database (www.mouseac.org) [59] (Supplementary Fig. 3; *C7* was unfortunately not included in this dataset). The increase of *C7* mRNA level in patients could be mimicked by the significantly increased level of *C7* mRNA in U251 cells in response to Aβ treatment (Fig. 3B). All these results were consistent with the recent reports that the initial complement components play an essential role in early synapse loss during the course of the development of Alzheimer's disease [60, 61].

Overexpression of C7 mutant p.K420Q disturbed global gene expression pattern and affected cellular function

Previous studies showed that the complement components mainly function in glia [61] and astrocyte can produce C7 and other complement components [62]. We thus conducted cellular analyses using the U251 glioma cell line (of astrocyte origin) and the human microglia (HM) cell line, to understand the potential biological and physiological significance of the identified risk gene *C7*. The U251 cells were engineered to stably express mutant APP K670N/M671L (U251-APP) so that they would produce $A\beta_{42}$ under Doxorubicin induction [32, 37]. We performed RNA-seq of U251-APP cells overexpressing wild type and mutant C7 p.K420Q to determine the potential effect of the mutant. Consistent with the expression pattern of the complements in brain tissues of Alzheimer's patients (Supplementary Table 4), we observed no significant alterations in the mRNA levels of the terminal components (e.g. *C6*, *C8* and *C9*) in cells overexpressing wild type or mutant C7. The mRNA expression levels of initial components (e.g. *C1R*, *C1S* and *C3*) and regulatory factors (e.g. *C1INH* and *CFH*) of the complement cascade were significantly increased in cells overexpressing wild type or mutant C7 relative to cells transfected with empty vector (Fig. 3C). While the mRNA levels of these initial components and regulatory factors did not differ between cells overexpressing wild type C7 and mutant C7, most of the other genes (591/653) up-regulated in cells overexpressing wild type C7 were downregulated in cells overexpressing mutant p.K420Q. These altered genes were significantly (P_{adj} < 0.05) enriched in interferon-mediated signaling pathways (Fig. 3D), among which there were three GWAS reported Alzheimer's risk genes *BIN1*, *RIN3*, *ZCWPW1* [18], as well as several important immune genes such as *OASL*, *IL6* and complement components (Fig. 3C). Intriguingly, the differentially expressed genes in response to C7 wild type (enrichment $P = 2.57\times10^{-6}$) or mutant (enrichment $P = 2.87\times10^{-10}$) overexpression were significantly enriched in a C7-involved co-expression network / module (Fig. 3E) that was recently recognized to be dysregulated in brains of Alzheimer's patients [52].

We further characterized the downstream effect of overexpression of the C7 mutant p.K420Q on Aβ internalization and cell apoptosis, and observed significant impact on the internalization of fluorescently-labeled $A\beta_{42}$ in HM cells (Supplementary Fig. 4). Additionally, HM cells overexpressing mutant p.K420Q showed increased apoptosis in response to TNF-α treatment compared with cells overexpressing wild type C7 (Supplementary Fig. 5).

Overexpression of C7 mutant p.K420Q affects excitatory synaptic transmission

Besides functions in immune activation, Aβ internalization and cell apoptosis, the complement system also plays a role in neuronal activity [60, 61, 63-67]. The biolistic transfection on rat hippocampal slice cultures and the accompanied dual-whole cell recording analyses offered a convenient study system to characterize the physiological function of target gene(s) in neurons [68]. We used this strategy to investigate the effect of C7 and its p.K420Q mutant on synaptic transmission in excitatory neurons. We found that overexpression of wild type C7 in CA1 pyramidal neurons decreased both the AMPAR and NMDAR-mediated synaptic transmission compared with the respective neighboring control neurons (Fig. 4A1 and 4B1), but these inhibitory effects were compromised by the p.K420Q mutation (Fig. 4A2, 4A3, 4B2 and 4B3). However, overexpression of C7 or its mutant had no effect on the ratio of AMPAR and NMDAR-mediated EPSCs (Fig. 4C1 and C2), suggesting a general postsynaptic role of C7 in excitatory synaptic transmission. The paired-pulse ratio, which is the parameter for presynaptic release probability, was not affected by C7 or its mutant overexpression (Fig.

4D1 and D2). This observation indicated that the regulatory function of C7 is postsynaptic specific. Moreover, neither wild type C7 nor C7 mutant p.K420Q had any effect on the GABA receptor-mediated inhibitory postsynaptic transmission (Fig. 4E1 and E2), indicating that the C7-mediated effect is specific to excitatory synapses. Taken together, C7 likely inhibits the excitatory synaptic transmission in pyramidal neurons while the mutant p.K420Q impairs this negative regulation. Note that complement factors can be produced and secreted locally in the brain; it is surprising that neighboring non-transfected cells are not regulated by the over-expressed C7. It remained to be answered that where endogenous and over-expressed C7 are located and what are their extracellular levels in the growing medium.

Discussion

To date, most of the Alzheimer's risk genes identified by GWAS and the next-generation sequencing technologies have been found in populations of European ancestry [5, 18, 23, 24, 28, 69]. For East Asians, there has been only one GWAS in a Japanese population, with no genome-wide significant loci (excluding *APOE*) being reported [70]. Given the increasing burden of Alzheimer's disease and the population heterogeneity, there is an urgent need to investigate the genetic basis of the disease in the Han Chinese, the largest single population with the greatest number of Alzheimer's patients worldwide [33, 34]. In this study, we have used the WES to identify potential risk gene(s) of Alzheimer's disease in Han Chinese. By recruiting relatively homogeneous set of patients with features attributable to genetic factors (familial and extreme early-onset), we have countered the limitation of small sample size and discovered a novel exome-wide significant variant rs3792646 (p.K420Q) in the *C7* gene (Fig. 1). Importantly, this risk variant has a comparable effect size with the well-known hits *TREM2* p.R47H [23, 24] and *PLD3* p.V232M [25] identified in populations of European origin. Intriguingly, the effect size of rs3792646 (OR = 3.651, Supplementary Table 2) was dramatically increased in the *APOE* ε4 carriers, suggesting a potential interaction between this rare missense variant and *APOE* ε4.

The complement system has complex roles in Alzheimer's disease, including Aβ clearance, microglia activation, neuro-inflammation, apoptosis, and neuron death [60, 61, 63-67]. It has been a controversial topic as to whether the complement system was a driven factor or a byproduct [60]. Recent studies reported that the initial component *C1q* and the central component *C3* contributed to early synapse loss in response to Aβ and/or viral infection in Alzheimer's disease [60, 61]. Our current results indicated that *C7*, a canonical terminal component in the complement cascade, might be also involved in the early pathological stage of Alzheimer's disease, together with the other initial components. Previous results have suggested that *C7* plays a major role in the formation of the membrane attack complex, and it serves as a membrane anchor [71]. *C7* deficiency contributes to the susceptibility to a variety of immune and infectious disease, such as Meningococcal infection [72-75], and rare damaging variants of other complement components were reported to be enriched in age-related macular degeneration [76]. It is also known that both infection and metabolite (e.g. Aβ) accumulation would activate the complement cascade [60, 61]. While no report had linked *C7* with neurodegenerative disorders so far, our results indicated that *C7* might function in the early activation phase, rather than in the terminal membrane attack complex as previously reported [71]. Moreover, the *C7* risk allele affected brain morphological structure and impaired the working memory in young adults and disease-related endo-phenotypes in patients (Fig. 2). These results were further supported by the observation that overexpression of

mutant C7 affected the global gene expression pattern (Fig. 3), Aβ internalization and apoptosis (Supplementary Fig. 4-5), which would play an active role in pathogenesis of Alzheimer's disease. By the use of an electrophysiological assay with rat hippocampal slice cultures and dual-whole cell recordings, we showed that overexpression of C7 mutant p.K420Q affected excitatory synaptic transmission of neurons (Fig. 4). All these lines of evidence suggested a putative role of *C7* and its variant in the development of Alzheimer's disease, though the exact mechanism remains to be elucidated. Considering the complex roles of the complement system in Alzheimer's disease, it is still unclear whether there is link or interaction between C7-induced changes in glial activity and changes in synaptic function. The exact mechanism of complement genes in the disease remains to be elucidated.

Consistent with the functional assays, *in silico* prediction by four algorithms showed that the mutation p.K420Q was deleterious. Nevertheless, we should note that rs3792646 was also present in the general population (with allele frequency ranging from 0.0004 to 0.03), leading to an argument against its pathogenic status, similar to the case of NR1H3 p.R415Q in multiple sclerosis [77], although the situation in Alzheimer's disease might be a bit different, partly due to its late age-of-onset. We have previously shown that a common missense variant in another complement gene, *CFH*, conferred genetic risk to Alzheimer's disease, whereas this variant underwent pathogen-driven selection so that it was retained in the population due to the trade-off effect [32]. It is reasonable to speculate that mutant C7 might also have been positively selected during evolution and this has led to the observed differences in allele frequencies and distinct disease susceptibility patterns.

The current study has some limitations. First, although we observed an exome-wide significant association of rs3792646 with Alzheimer's disease in the WES discovery stage and validated the association in independent cohorts, it should be noted that the association of rs3792646 in the initial screening stage did not reach the exome-wide significance in the logistic regression analysis. This might be caused, at least partially, by small sample size in this stage (Table 1). As the association was initially recognized in early-onset patients that were selected by extreme phenotype sampling in the discovery stage, whereas most of the replication samples were sporadic late-onset patients, independent replications in larger cohorts with early-onset Alzheimer's disease are needed to further confirm the association. Second, the risk allele rs3792646-C was mainly found in Asian populations and was infrequent in European populations. In the ADNI samples [54] and ADSP samples [55] of European ancestry, we observed only one risk-allele carrier out of 812 individuals, and one carrier out of 10,570 individuals, respectively, suggesting a Chinese-specific effect of rs3792646. However, the results need further validation and should be interpreted with caution, as the sample size of cases of European ancestry was still limited. Nonetheless, our functional characterizations indicated the C7 mutant p.K420Q affected the expression of the interferon-mediated signaling pathways, Aβ internalization and apoptosis at the cellular level and excitatory synaptic transmission of neurons, which reinforce the conclusion of *C7* as a risk gene for Alzheimer's disease.

During the preparation of this manuscript, we noticed a recent publication about whole genome sequencing-based GWAS in Chinese population [78]. These authors identified two common variants *GCH1* (rs72713460) and *KCNJ15* (rs928771) showing nominal associations with Alzheimer's disease in Chinese. We checked these two risk variants in our WES data, but failed to find any association between these genes/ variants and Alzheimer's disease in our samples. This might be caused by different strategies that were used in

Zhou et al.'s study (low-coverage whole genome sequencing for sporadic patients) [78] and our study (WES for relatively homogeneous set of patients with extreme phenotype). Evidently, large sample sizes are needed for further validation of these risk genes in our current study and the study by Zhou et al. [78].

Conclusion

In summary, we have identified a rare damaging variant rs3792646 (p.K420Q) in *C7* conferring risk to Alzheimer's disease through the exome-wide screening in Chinese with early onset Alzheimer's disease or with familial history. Although *TREM2* p.R47H and *PLD3* p.V232M are extremely rare or absent in Chinese [26, 31], we have shown here the Han Chinese population harbors another risk factor, *C7* p.K420Q, with a comparable effect size. The *C7* risk allele was most likely specific to Han Chinese. This variant could potentially contribute to the risk of Alzheimer's disease via disrupting immune activation and β-amyloid processing, and was associated with changes of brain structure and function even decades before disease onset. Our results strongly suggest the active roles of *C7*, together with other complement components such as the GWAS hits *CR1* and *CLU* [18, 57] and *CFH* [32], in Alzheimer's disease. Further validation and functional investigations are needed to characterize the mechanisms underlying the risk for Alzheimer's disease conferred by these molecules.

Materials and Methods

Subjects: extreme phenotype sampling for exome sequencing

We took an extreme phenotype sampling strategy in the WES stage to increase the likelihood of identifying true disease-related variants [35, 36]. The criteria of extreme phenotypes were set as follows [79]: 1) AAO of Alzheimer's disease <= 55, and/or 2) with a positive familial history. In our collection of over 1,000 genetically unrelated patients from East and Southwest China [26, 32, 37-41], 107 unrelated patients (46.7% females; age 64.6 ± 10.29 years; AAO 56.0 ± 9.83; *APOE* ε4, 38.5%) met the criteria and were subjected to WES. For familial Alzheimer's disease, only the probands were included in the study and no family members were recruited. Detailed clinical records including age, sex, education, occupation, AAO, familial history, diseases history, diagnostic imaging tests, and neuropsychological assessment, were collected for each participant. All patients were diagnosed by at least two clinical psychiatrists using the revised National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association criteria [80, 81] and Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, as described in our previous studies [26, 32, 37-41]. 160 in-house control individuals (40.6% females, age 52.6 ± 16.5 years; *APOE* ε4, 15%) [42] showing no signs of memory loss and without any familial history of neurodegenerative disorders were compared with the patients with Alzheimer's disease. Sample collection complied with the Declaration of Helsinki, with written informed consents being obtained from each participant or their guardians. This study was approved by the Institutional Review Board of Kunming Institute of Zoology, Chinese Academy of Sciences.

Whole exome sequencing and data processing

The coding region (untranslated regions and exons, namely exome) of the whole genome of cases and in-house controls were captured using the SeqCap EZ Exome Kit v3.0 (#06465692001, Roche, Basel, Switzerland). The total size of the regions covered by 2.1 million long oligonucleotide probes was 64 Mb, achieving the most comprehensive coverage of coding regions in the genome. All the genome coordinates were based on human genome build GRCh37 (hg19, http://asia.ensembl.org/info/website/tutorials/grch37.html). Processed final libraries were pooled and sequenced on an Illumina HiSeq2500 or 4000 (150-bp paired-end, Illumina, San Diego, CA, USA).

Low-quality raw reads were removed using Trimmomatic-0.32 [82] with the parameters as "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36". Quality-filtered reads were then aligned to the NCBI human genome reference assembly (build GRCh37) using Burrows-Wheeler Aligner [83]. Picard Tools (http://broadinstitute.github.io/picard/) were used to flag duplicate reads. Variant call was performed through the canonical pipeline recommended by the Best Practice Variant Detection with the GATK (Genome Analysis Toolkit) [84]. Variant Quality Score Recalibration from GATK package was used to filter spurious variants resulted from sequencing errors and mapping artifacts. ANNOVAR was used to annotate variants into different functional categories according to their locations and expected effects on encoded gene products [85].

In order to achieve credible statistic power by increasing the control-case ratio, we pooled the exome data of the 160 in-house non-dementia individuals [42] with the whole genome data of Han Chinese in Beijing (N=103) and Southern Han Chinese (N=105) from the 1000 Genomes Project phase 3 [43] as the initial population control (N = 368). Principal component analysis was performed to ensure there was no apparent population stratification between the studied subjects and the reference Chinese populations from the 1000 Genomes Project phase 3 [43] by using the GCTA tool (http://cnsgenomics.com/software/gcta/#Overview). Based on the clustering pattern (Supplementary Fig. 1), there is no obvious population substructure among the East Asian populations, suggesting that it is reasonable to group the in-house controls with Han Chinese in Beijing and Southern Han Chinese from the 1000 Genomes Project phase 3 [43] as the general population control. Allele frequencies of exonic variants in patients were compared to that of the population controls by using the Fisher's exact test in the initial exome-wide case-control screening. To rule out the possibility of technical artifacts due to potential population substratification, we performed a logistic regression with the PC1-PC3 as the covariates using the open-source C/C++ toolset Plink/seq (https://atgu.mgh.harvard.edu/plinkseq/). We also included the *APOE* ε4 status and sex as covariates, besides PC1-PC3, in the logistic regression analysis.

We defined nonsense, frameshift, splice site variants, and missense variants as functional if these variants were predicted to be damaging by at least one of the five algorithms (PolyPhen2 HumDiv and HumVar [44], LRT [45], MutationTaster [46], and SIFT [47, 48]). Functional variants with a minor allele frequency (MAF) < 5% in the 368 pooled population controls [42, 43] were analyzed to identify the exome-wide significant rare variants. A total of 23,373 functional variants met this criterion, resulting in a threshold for the exome-wide significance of $P < 2.139\times10^{-6}$ (Bonferroni corrected: 0.05/23373). These exonic variants were directly compared to the population control by using the Fisher's exact test and logistic regression analysis. All damaging missense variants with a MAF < 5% in the control population were used for the gene-based burden testing [86] using PLINK/seq. The SKAT was also used to evaluate the combined effect of rare mutations using the SKAT R package [53]. Allele frequencies of the targeted loci in 4,327 East Asians from

ExAC [50] and in 11,670 Chinese samples from the CONVERGE Consortium (the largest Han Chinese low-coverage genome dataset so far) [51] were retrieved and used as the reference control for comparison with the Alzheimer's patients.

Independent validations in Chinese and European populations

The discovery WES screening (stage 1) revealed a significant association between *C7* rs3792646 and early-onset and familial Alzheimer's disease. To validate this association, we sequenced this SNP in an independent Han Chinese sample with early-onset and/or familial Alzheimer's disease from Beijing (stage 2, N = 103 cases; 60.2% females, age 61.2 ± 7.33 years, AAO 57.2 ± 9.12 years; *APOE* ε4, 24%). We attempted to confirm the association of *C7* rs3792646 with Alzheimer's disease in independent cohorts with sporadic Alzheimer's disease (stage 3): the East China cohort contains 584 sporadic cases (61.0% females, age 79.5 ± 8.45 years, AAO 73.2 ± 9.25 years; *APOE* ε4, 41%) and 274 geographically matched controls; the Southwest China cohort contains 581 sporadic cases (61.9% females, age 76.4 ± 10.04 years, AAO 74.7 ± 11.79 years; *APOE* ε4, 34%) and 2190 geographically matched controls. We also analyzed a patient sample from Hunan, Southcentral China (N = 235 sporadic cases; 63.9% females; age 79.1 ± 7.86 years; AAO 74.4 ± 7.72 years; *APOE* ε4, 35%). DNA fragments covering SNP rs3792646 were amplified using the primer pair 5'-TATAACGACATGTGCCCACCA-3' / 5'-GACTTCAGGAGCCCACAAGC-3' and sequenced using the primer 5'-GCCCTAAATATCCTTTGTGCT-3'.

Whole genome sequencing data and clinical phenotypes of 812 individuals of European ancestry (including 281 controls, 483 subjects with mild cognitive impairment, and 48 subject with Alzheimer's disease) were retrieved from the ADNI project (http://adni.loni.usc.edu/) [54] to explore rare *C7* variants in Europeans. Given the small sample size of Alzheimer's patients in ADNI data, patients with late-stage mild cognitive impairment were combined with Alzheimer's patients to achieve a better statistical power, resulting in 296 patients and 281 controls (named ADNI cohort in the subsequent discussions), whereas the remaining 235 subjects with early-stage or modest mild cognitive impairment were excluded from the analysis [54]. To validate the result in a larger European cohort, we obtained access to the whole-exome sequencing data of 5,815 Alzheimer's cases and 4,755 controls from the ADSP [55] through the dbGaP (Genotypes and Phenotypes database) under the study accession phs000572.v7.p4 (accessed in May 2018).

Statistic power and sample size calculations were performed using the Quanto software (version 1.2.4) [87] based on the observed parameters. For alleles with a MAF of 0.05 in the general populations (disease prevalence was set as 0.1), at least 279 pairs of case and control samples were needed to capture an odds ratio of 2.0 with a statistic power of 80% under an additive model. The current samples thus had sufficient power for validating associations with considerable effect sizes.

Brain structural changes and cognitive performance of at-risk individuals in early adulthood

We had previously recruited 360 young healthy adults (48% females, age 19.4 ± 1.1 years) to study the effects of potentially functional variants on morphological and functional changes of the brain [32, 56]. All these participants were university students without any

history of neuropsychiatric disorders or acquired brain injury. Their brain structure data were collected through structural MRI (magnetic resonance imaging) scans using an MR750 3.0 Tesla magnetic resonance scanner (GE Healthcare). Briefly, high-resolution 3D T1-weighted brain volume (BRAVO) sequence was performed with the following parameters: TR = 8.16 ms, TE = 3.18 ms, flip angle = 7°, FOV = 256 mm × 256 mm, voxel size = 1 × 1 × 1 mm^3, and 188 slices. The brain regions of interest were the hippocampus and entorhinal cortex, which were recognized as the most and the first affected brain regions by Alzheimer's disease, respectively [88, 89]. The magnetic resonance imaging data were analyzed with FreeSurfer software (version 5.3) [90] as previously described [32, 56, 91]. These young healthy donors also received a working memory test during their participation [92]. The working memory task was assessed with an N-back paradigm (2- and 3-back) [93]. In brief, participants were presented with a series of letters sequentially, and were asked to perform continuous judgments: whether the letter on the screen was the same as the one presented two letters earlier (2-back task) or the one presented three letters earlier (3-back task) [92]. We excluded the outliers in accuracy (more than mean + 2SD or lower than mean - 2SD) in the analysis of group differences in working memory performance. The Alzheimer's disease-related variant rs3792646 was genotyped in these healthy donors by sequencing as described above, and the effects of rs3792646 genotypes on morphological changes and working memory performance were assessed.

Effects of rs3792646 genotypes on Alzheimer-related endo-phenotypes

In order to further investigate the role of rs3792646 in the pathogenesis of Alzheimer’s disease, we obtained genetic, neuroimaging, and biomarker data of 812 individuals from the ADNI dataset [54]. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, as well as clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early stage of Alzheimer's disease [54]. The effects of disease-risk SNPs on endo-phenotypes, e.g. levels of tau, p-tau, and Aβ in the cerebrospinal fluid, cognitive score, and hippocampus volume, were analyzed using PLINK [94].

Cell culture and transfection

U251 glioma cells and HM cells were introduced from Kunming Cell Bank, Kunming Institute of Zoology, Chinese Academy of Sciences. The U251 cells were engineered to stably express mutant APP K670N/M671L (U251-APP) so that they would produce $A\beta_{42}$ under Doxorubicin induction [32, 37]. We overexpressed wild type and mutant p.K420Q of C7 in these two cell lines to characterize their potential roles. Briefly, U251-APP cells were cultured in Roswell RPMI-1640 medium (HyClone, #C11875500BT) supplemented with 10% fetal bovine serum (Gibco-BRL; #10099-141), 100 U/mL penicillin and 100 mg/mL streptomycin in 5% CO_2 at 37 °C. HM cells were maintained in Dulbecco’s modified Eagle’s medium (Gibco-BRL; #11965-092) supplemented with 10% fetal bovine serum (Gibco-BRL; #10099-141). Transfection of empty vector (pReceiver-M14 [CMV Promoter, 3×Flag], GeneCopoeia, Inc.) or C7 wild type and mutant p.K420Q expression vectors was performed using an electroporator (CUY21EDIT, Nepa gene Co., Japan) following the

manufacturer's instructions. In brief, cells were trypsinized and washed three times with Opti-MEM medium (Gibco-BRL). Around 1×10^6 cells were resuspended in 100 μl Opti-MEM medium, and electroporated with 10 μg plasmids. Transfected cells were seeded in pre-warmed growth medium for 72 h in 5% CO_2 at 37 °C before the harvest.

RNA-seq and mRNA expression profiling

We performed transcriptome sequencing for U251-APP cells overexpressing wild type or mutant C7 protein. After RNA quantification and qualification, 1.5 μg RNA per sample was used for the library preparation. Sequencing libraries were generated using NEBNext Ultra™ RNA Library Prep kit for Illumina (NEB, USA) following the manufacturer's recommendations. Index codes were added to attribute sequences to each sample. The processed final library was sequenced on an Illumina Hiseq 4000 platform and 150 bp paired-ends reads were generated. Sequenced reads were processed and differential gene expression analysis was performed according to standard protocols. In brief, the raw reads were trimmed to remove sequencing adapters and low quality reads. The clean reads were then aligned to the reference genome (hg19) using Tophat [95]. HTSeq-count [96] was then used to count aligned reads that mapped to the annotated human genes (gencode v19) [97]. Gene-level differential expression analyses were performed using DESeq2 [98]. Principal component analysis of gene expression levels was performed to remove outliers using the 'prcomp' function in the 'stats' package in R (http://www.R-project.org/). Hierarchical cluster analyses and heatmap analyses were performed using R-statistics. Gene Ontology biological processes enrichment analysis for differentially expressed genes was performed using DAVID online tools (https://david.ncifcrf.gov/) [99]. Global effect of C7 wild type and mutant overexpression was assessed using the co-expression network that was constructed based on expression profiles of brain tissues from individuals with Alzheimer's disease [52]. Network was visualized using the Cytoscape software [100].

We retrieved GSE33000 from GEO (Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/browse/), a microarray expression profile of frontal cortex from 309 Alzheimer's patients and 156 controls [58], to re-analyze the expression pattern of the complement components in brains of Alzheimer's patients. Differential gene expression analysis was performed using linear regression by *limma* package in R software as described elsewhere [52]. In addition to the differential expression analysis in brain tissues of patients, we also analyzed the expression alterations in brain tissues of mouse models [59]. In brief, the transgenic mouse models with human mutant genes responsible for familial type of Alzheimer's disease, which showed Alzheimer's pathological features such as amyloid plaques and neurofibrillary tangles, were used for the genome-wide gene profiling [59]. Expression profiling of hippocampus and cortex tissues were tested using the MouseRef8 v2 (Illumina) microarray platform. Microarray data was processed and shared by John Hardy and colleagues from the Mouse Dementia Network, available at Mouseac (www.mouseac.org) [59]. More details about this dataset were described in the original paper [59]. The number of Aβ plaques and the level of tau burden were quantified. Correlations between mRNA expression of genes of interest and the quantified indices of pathology were then measured based on the processed data, using the *Pearson's* correlation test.

$A\beta_{42}$ internalization and cell viability

HM cells were treated with 5 μM oligomeric, aggregated, and fibrillary fluorescently-labeled $A\beta_{42}$ (ChinaPeptides Co., Ltd.) for 2 h after transfection with the C7 wild type and mutant overexpression vectors for 24 h. Fluorescence (FITC) intensity in treated cells was measured by flow cytometry using LSR Fortessa (Becton Diskinson, USA) following the manufacturer's instruction. The FlowJo software was used for viewing and analyzing the flow cytometric data.

Cell viability induced by TNF-α (Sigma) was determined using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Sigma, #M2128) assay. HM cells were seeded in 96-well plates at a density of 5×10^3 cells per well after transfection for 24 h, followed by treatment with 2 μg/ml Actinomycin D (Merck Millipore, #129935) and 200 ng/ml TNF-α (peproTech, #300-01A) for 24 h. The MTT assay was then performed according to the manufacturer's instructions. Absorbance was measured at 490 nm using a Gen5 plate reader (Elx808, BioTek).

Electrophysiology in brain slice cultures

The electrophysiology in brain slice cultures was performed in accordance with the previously described protocol [68, 101]. Briefly, organotypic rat hippocampal slice cultures were made from postnatal day 6 to 8 wild type rat. The C7 wild type and mutant p.K420Q were subcloned into the pCAGGS vector harboring EGFP fluorescent protein. Biolistic transfections were carried out after culture for two days using a Helios Gene Gun (Bio-Rad) with 1 μm DNA-coated gold particles. Slices were maintained at 34 °C with media changes every other day. On day 6 after transfections, voltage-clamp dual whole-cell recordings for CA1 pyramidal neurons were taken from a fluorescent transfected neuron and a neighboring untransfected control neuron. During recording, slices were transferred to a perfusion stage on an Olympus BX51WI upright microscope and perfused at 2.5 mL/min with artificial cerebrospinal fluid bubbled with 95% O_2/5% CO_2. The artificial cerebrospinal fluid was composed of 119 mM NaCl, 2.5 mM KCl, 4 mM $CaCl_2$, 4 mM $MgSO_4$, 1 mM NaH_2PO_4, 26.2 mM $NaHCO_3$, and 11 mM Glucose. Series resistance was monitored on-line, and recordings in which series increased to >30 MOhm or varied by >50% between neurons were discarded. Dual whole-cell recordings measuring evoked excitatory postsynaptic currents (EPSCs) and inhibitory post-synaptic currents (IPSCs) were performed. When measuring EPSCs, 100 μM picrotoxin was added to block inhibitory currents and 4 μM 2-Chloroadenosine was used to control epileptiform activity. When measuring IPSCs, 10 μM NBQX (AMPAR antagonist) and 50 μM D-APV (NMDAR antagonist) were added to block AMPAR and NMDAR-mediated excitatory currents, respectively. Internal solution contained 135 mM $CsMeSO_4$, 8 mM NaCl, 10 mM HEPES, 0.3 mM EGTA, 5 mM QX314-Cl, 4 mM MgATP, 0.3 mM Na_3GTP, and 0.1 mM spermine. A bipolar stimulation electrode was placed in stratum radiatum, and responses were evoked at 0.2 Hz. Peak AMPAR responses were recorded at -70 mV, and NMDAR responses were recorded at +40 mV, with amplitudes measured 100 ms after stimulation to avoid contamination by AMPAR current. Paired-pulse ratio was determined by delivering two stimuli 40 ms apart and dividing the peak response to stimulus 2 by the peak response to stimulus 1. Peak GABA currents

were recorded at 0 mV. All the data were analyzed off-line with custom software (IGOR Pro). Responses were collected with a Multiclamp 700A amplifier (Axon Instruments), filtered at 2 kHz, and digitized at 10 kHz. All animal experiments were performed in accordance with established protocol approved by the Institutional Animal Care and Use Committee of Kunming Institute of Zoology, Chinese Academy of Sciences.

Statistical analysis

Statistical comparisons between two groups concerning relative cell numbers, imaging analysis, and cognitive test were conducted with two-tailed Student's *t* test. Significance of pathway or network enrichment was measured by using the Fisher's exact test. Significance of evoked dual whole-cell recordings compared to controls was determined using the two-tailed Wilcoxon signed-rank sum test. For experiments involving un-paired data, a Mann-Whitney U-test with Bonferroni correction for multiple comparisons was performed. Paired-pulse ratios were analyzed with the Student's *t* test. All statistical analyses were carried out with Igor Pro (Wavemetrics) and GraphPad Prism (GraphPad Software). We used the survival test to show the potential effect of *C7* variant rs3792646 on age at onset of Alzheimer's disease. In brief, Alzheimer's patients were grouped into the *C7* wild-type group and mutant carrier group, and the age at onset was set as the deaths/events. The survival proportion was assessed using the Log-rank (Mantel-Cox) test with the GraphPad Prism.

Data availability

The summary statistics of all 23,373 rare or low-frequency coding variants identified in stage 1 for initial discovery has been deposited into the AlzData webserver (http://www.alzdata.org/exome.html) [52] and could be freely accessible. RNA-seq data of U251-APP cells overexpressing wild type C7, mutant C7 (p.K420Q) and empty vector were deposited in the GEO database under accession number GSE101608.

Acknowledgements

We would like to thank the patients and their families for their participation in this project. We thank Ian Logan for helpful comments and language editing of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (31730037), the Strategic Priority Research Program (B) of the Chinese Academy of Sciences (CAS, XDBS1020200), and the Bureau of Frontier Sciences and Education, CAS (QYZDJ-SSW-SMC005). Data collection and sharing of the ADNI project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The ADSP data was obtained from dbGaP under accession phs000572.v7.p4. Full funding acknowledgements for ADSP are provided in the supplementary material.

Supplementary data

Supplementary material is available online.

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

1. Querfurth HW, LaFerla FM; Alzheimer's disease. *N Engl J Med* 2010; **362**(4):329-44.
2. Ballard C, Gauthier S, Corbett A, et al.; Alzheimer's disease. *Lancet* 2011; **377**(9770):1019-31.
3. Alzheimer's association; 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* 2016; **12**(4):459-509.
4. Gatz M, Reynolds CA, Fratiglioni L, et al.; Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006; **63**(2):168-74.
5. Guerreiro R, Bras J, Hardy J; SnapShot: genetics of Alzheimer's disease. *Cell* 2013; **155**(4):968-968 e1.
6. Gatz M, Pedersen NL, Berg S, et al.; Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci* 1997; **52**(2):M117-25.
7. Schellenberg GD, Pericak-Vance MA, Wijsman EM, et al.; Linkage analysis of familial Alzheimer disease, using chromosome 21 markers. *Am J Hum Genet* 1991; **48**(3):563-83.
8. Kamino K, Orr HT, Payami H, et al.; Linkage and mutational analysis of familial Alzheimer disease kindreds for the APP gene region. *Am J Hum Genet* 1992; **51**(5):998-1014.
9. St George-Hyslop PH, Tanzi RE, Polinsky RJ, et al.; The genetic defect causing familial Alzheimer's disease maps on chromosome 21. *Science* 1987; **235**(4791):885-90.
10. Tanzi RE, Gusella JF, Watkins PC, et al.; Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. *Science* 1987; **235**(4791):880-4.
11. Schellenberg GD, Bird TD, Wijsman EM, et al.; Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. *Science* 1992; **258**(5082):668-71.
12. St George-Hyslop P, Haines J, Rogaev E, et al.; Genetic evidence for a novel familial Alzheimer's disease locus on chromosome 14. *Nat Genet* 1992; **2**(4):330-4.
13. Van Broeckhoven C, Backhovens H, Cruts M, et al.; Mapping of a gene predisposing to early-onset Alzheimer's disease to chromosome 14q24.3. *Nat Genet* 1992; **2**(4):335-9.
14. Levy-Lahad E, Wijsman EM, Nemens E, et al.; A familial Alzheimer's disease locus on chromosome 1. *Science* 1995; **269**(5226):970-3.
15. Rogaev EI, Sherrington R, Rogaeva EA, 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* 1995; **376**(6543):775-8.
16. Campion D, Dumanchin C, Hannequin D, et al.; Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *Am J Hum Genet* 1999; **65**(3):664-70.
17. Ridge PG, Mukherjee S, Crane PK, et al.; Alzheimer's disease: analyzing the missing heritability. *PLoS One* 2013; **8**(11):e79771.
18. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al.; Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12):1452-8.
19. Bertram L, McQueen MB, Mullin K, et al.; Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 2007; **39**(1):17-23.

20. Ridge PG, Hoyt KB, Boehme K, et al.; Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiology of Aging* 2016; **41**:200 e13-20.

21. Bertram L, Klein C; Probing the exome in Alzheimer disease and other neurodegenerative disorders. *JAMA Neurol* 2015; **72**(4):389-91.

22. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, et al.; A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat Med* 2014; **20**(12):1452-7.

23. Jonsson T, Stefansson H, Steinberg S, et al.; Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**(2):107-16.

24. Guerreiro R, Wojtas A, Bras J, et al.; TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013; **368**(2):117-27.

25. Cruchaga C, Karch CM, Jin SC, et al.; Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* 2014; **505**(7484):550-4.

26. Zhang DF, Fan Y, Wang D, et al.; PLD3 in Alzheimer's disease: a modest effect as revealed by updated association and expression analyses. *Mol Neurobiol* 2016; **53**(6):4034-4045.

27. Sims R, van der Lee SJ, Naj AC, et al.; Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet* 2017.

28. Lambert JC, Amouyel P; Genetic heterogeneity of Alzheimer's disease: complexity and advances. *Psychoneuroendocrinology* 2007; **32 Suppl 1**:S62-70.

29. McClellan J, King MC; Genetic heterogeneity in human disease. *Cell* 2010; **141**(2):210-7.

30. Tan L, Yu JT, Zhang W, et al.; Association of GWAS-linked loci with late-onset Alzheimer's disease in a northern Han Chinese population. *Alzheimers Dement* 2013; **9**(5):546-53.

31. Jiao B, Liu X, Tang B, et al.; Investigation of TREM2, PLD3, and UNC5C variants in patients with Alzheimer's disease from mainland China. *Neurobiol Aging* 2014; **35**(10):2422 e9-2422 e11.

32. Zhang DF, Li J, Wu H, et al.; CFH variants affect structural and functional brain changes and genetic risk of Alzheimer's disease. *Neuropsychopharmacology* 2016; **41**(4):1034-45.

33. Jia J, Wang F, Wei C, et al.; The prevalence of dementia in urban and rural areas of China. *Alzheimers Dement* 2014; **10**(1):1-9.

34. Chan KY, Wang W, Wu JJ, et al.; Epidemiology of Alzheimer's disease and other forms of dementia in China, 1990-2010: a systematic review and analysis. *Lancet* 2013; **381**(9882):2016-23.

35. Barnett IJ, Lee S, Lin X; Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet Epidemiol* 2013; **37**(2):142-51.

36. MacArthur DG, Manolio TA, Dimmock DP, et al.; Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014; **508**(7497):469-76.

37. Xiang Q, Bi R, Xu M, et al.; Rare genetic variants of the *Transthyretin* gene are associated with Alzheimer's disease in Han Chinese. *Mol Neurobiol* 2017; **54**(7):5192-5200.

38. Li GD, Bi R, Zhang DF, et al.; Female-specific effect of the BDNF gene on Alzheimer's disease. *Neurobiol Aging* 2017; **53**:192 e11-192 e19.

39. Bi R, Kong LL, Xu M, et al.; The *Arc* gene confers genetic susceptibility to Alzheimer's disease in Han Chinese. *Mol Neurobiol* 2017:doi: 10.1007/s12035-017-0397-6.

40. Bi R, Zhang W, Yu D, et al.; Mitochondrial DNA haplogroup B5 confers genetic susceptibility to Alzheimer's disease in Han Chinese. *Neurobiol Aging* 2015; **36**(3):1604 e7-16.

41. Bi R, Zhang W, Zhang DF, et al.; Genetic association of the cytochrome c oxidase-related genes with Alzheimer's disease in Han Chinese. *Neuropsychopharmacology* 2018; **DOI: 10.1038/s41386-018-0144-3**.

42. Wang D, Fan Y, Malhi M, et al.; Missense variants in HIF1A and LACC1 contribute to leprosy risk in Han Chinese. *Am J Hum Genet* 2018; **102**:794-805.

43. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al.; A global reference for human genetic variation. *Nature* 2015; **526**(7571):68-74.

44. Adzhubei IA, Schmidt S, Peshkin L, et al.; A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**(4):248-9.

45. Chun S, Fay JC; Identification of deleterious mutations within three human genomes. *Genome Res* 2009; **19**(9):1553-61.

46. Schwarz JM, Rodelsperger C, Schuelke M, et al.; MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; **7**(8):575-6.

47. Ng PC, Henikoff S; SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**(13):3812-4.

48. Kumar P, Henikoff S, Ng PC; Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**(7):1073-81.

49. Kircher M, Witten DM, Jain P, et al.; A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**(3):310-5.

50. Lek M, Karczewski KJ, Minikel EV, et al.; Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**(7616):285-91.

51. CONVERGE consortium; Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 2015; **523**(7562):588-91.

52. Xu M, Zhang DF, Luo R, et al.; A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes as important upstream regulators in Alzheimer's disease. *Alzheimers Dement* 2018; **14**:215-229.

53. Ionita-Laza I, Lee S, Makarov V, et al.; Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet* 2013; **92**(6):841-53.

54. Weiner MW, Aisen PS, Jack CR, Jr., et al.; The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement* 2010; **6**(3):202-211.e7.

55. Bis JC, Jian X, Kunkle BW, et al.; Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Molecular Psychiatry* 2018.

56. Li J, Cui Y, Wu K, et al.; The cortical surface area of the insula mediates the effect of DBH rs7040170 on novelty seeking. *Neuroimage* 2015; **117**:184-190.

57. Tschopp J, Chonn A, Hertig S, et al.; Clusterin, the human apolipoprotein and complement inhibitor, binds to complement C7, C8 beta, and the b domain of C9. *J Immunol* 1993; **151**(4):2159-65.

58. Narayanan M, Huynh JL, Wang K, et al.; Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. *Mol Syst Biol* 2014; **10**:743.

59. Matarin M, Salih DA, Yasvoina M, et al.; A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep* 2015; **10**(4):633-44.

60. Hong S, Beja-Glasser VF, Nfonoyim BM, et al.; Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* 2016; **352**(6286):712-6.

61. Vasek MJ, Garber C, Dorsey D, et al.; A complement-microglial axis drives synapse loss during virus-induced memory impairment. *Nature* 2016; **534**(7608):538-43.

62. Gasque P, Fontaine M, Morgan BP; Complement expression in human brain. Biosynthesis of terminal pathway components and regulators in human glial cells and cell lines. *J Immunol* 1995; **154**(9):4726-33.

63. Zanjani H, Finch CE, Kemper C, et al.; Complement activation in very early Alzheimer disease. *Alzheimer Dis Assoc Disord* 2005; **19**(2):55-66.

64. McGeer PL, McGeer EG; The possible role of complement activation in Alzheimer disease. *Trends Mol Med* 2002; **8**(11):519-23.

65. Webster S, Bonnell B, Rogers J; Charge-based binding of complement component C1q to the Alzheimer amyloid beta-peptide. *Am J Pathol* 1997; **150**(5):1531-6.

66. Walker DG, McGeer PL; Complement gene expression in human brain: comparison between normal and Alzheimer disease cases. *Brain Res Mol Brain Res* 1992; **14**(1-2):109-16.

67. Rogers J, Cooper NR, Webster S, et al.; Complement activation by beta-amyloid in Alzheimer disease. *Proc Natl Acad Sci U S A* 1992; **89**(21):10016-20.

68. Sheng N, Shi YS, Nicoll RA; Amino-terminal domains of kainate receptors determine the differential dependence on Neto auxiliary subunits for trafficking. *Proc Natl Acad Sci U S A* 2017; **114**(5):1159-1164.

69. Lambert JC, Grenier-Boley B, Bellenguez C, et al.; PLD3 and sporadic Alzheimer's disease risk. *Nature* 2015; **520**(7545):E1.

70. Hirano A, Ohara T, Takahashi A, et al.; A genome-wide association study of late-onset Alzheimer's disease in a Japanese population. *Psychiatr Genet* 2015; **25**(4):139-46.

71. DiScipio RG, Chakravarti DN, Muller-Eberhard HJ, et al.; The structure of human complement component C7 and the C5b-7 complex. *J Biol Chem* 1988; **263**(1):549-60.

72. Thomas AD, Orren A, Connaughton J, et al.; Characterization of a large genomic deletion in four Irish families with C7 deficiency. *Mol Immunol* 2012; **50**(1-2):57-9.

73. Barroso S, Lopez-Trascasa M, Merino D, et al.; C7 deficiency and meningococcal infection susceptibility in two spanish families. *Scand J Immunol* 2010; **72**(1):38-43.

74. Halle D, Elstein D, Geudalia D, et al.; High prevalence of complement C7 deficiency among healthy blood donors of Moroccan Jewish ancestry. *Am J Med Genet* 2001; **99**(4):325-7.

75. Segurado OG, Arnaiz-Villena AA, Iglesias-Casarrubios P, et al.; Combined total deficiency of C7 and C4B with systemic lupus erythematosus (SLE). *Clin Exp Immunol* 1992; **87**(3):410-4.

76. Seddon JM, Yu Y, Miller EC, et al.; Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat Genet* 2013; **45**(11):1366-70.

77. Minikel EV, MacArthur DG; Publicly Available Data Provide Evidence against NR1H3 R415Q Causing Multiple Sclerosis. *Neuron* 2016; **92**(2):336-338.

78. Zhou X, Chen Y, Mok KY, et al.; Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer's disease pathogenesis. *Proc Natl Acad Sci U S A* 2018.

79. Rovelet-Lecrux A, Legallic S, Wallon D, et al.; A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. *Eur J Hum Genet* 2012; **20**(6):613-7.

80. Jack CR, Jr., Albert MS, Knopman DS, et al.; Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; **7**(3):257-62.

81. Khachaturian ZS; Revised criteria for diagnosis of Alzheimer's disease: National Institute on Aging-Alzheimer's Association diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; **7**(3):253-6.

82. Bolger AM, Lohse M, Usadel B; Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**(15):2114-20.

83. Li H, Durbin R; Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**(14):1754-60.

84. McKenna A, Hanna M, Banks E, et al.; The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**(9):1297-303.

85. Wang K, Li M, Hakonarson H; ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**(16):e164.

86. Lee S, Abecasis GR, Boehnke M, et al.; Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014; **95**(1):5-23.

87. Gauderman WJ; Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med* 2002; **21**(1):35-50.

88. Harris JA, Devidze N, Verret L, et al.; Transsynaptic progression of amyloid-beta-induced neuronal dysfunction within the entorhinal-hippocampal network. *Neuron* 2010; **68**(3):428-41.

89. Khan UA, Liu L, Provenzano FA, et al.; Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nat Neurosci* 2014; **17**(2):304-11.

90. Dale AM, Fischl B, Sereno MI; Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 1999; **9**(2):179-94.

91. Zhang X, Yu JT, Li J, et al.; Bridging integrator 1 (BIN1) genotype effects on working memory, hippocampal volume, and functional connectivity in young healthy individuals. *Neuropsychopharmacology* 2015; **40**(7):1794-803.

92. Li J, Liu B, Chen C, et al.; RAB2A Polymorphism impacts prefrontal morphology, functional connectivity, and working memory. *Hum Brain Mapp* 2015; **36**(11):4372-82.

93. Kirchner WK; Age differences in short-term retention of rapidly changing information. *J Exp Psychol* 1958; **55**(4):352-8.

94. Purcell S, Neale B, Todd-Brown K, et al.; PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**(3):559-75.

95. Trapnell C, Pachter L, Salzberg SL; TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**(9):1105-11.

96. Anders S, Pyl PT, Huber W; HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31**(2):166-9.

97. Harrow J, Frankish A, Gonzalez JM, et al.; GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012; **22**(9):1760-74.

98. Love MI, Huber W, Anders S; Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**(12):550.

99. Huang DW, Sherman BT, Lempicki RA; Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**(1):44-57.

100. Shannon P, Markiel A, Ozier O, et al.; Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**(11):2498-504.

101. Schnell E, Sizemore M, Karimzadegan S, et al.; Direct interactions between PSD-95 and stargazin control synaptic AMPA receptor number. *Proc Natl Acad Sci U S A* 2002; **99**(21):13902-7.

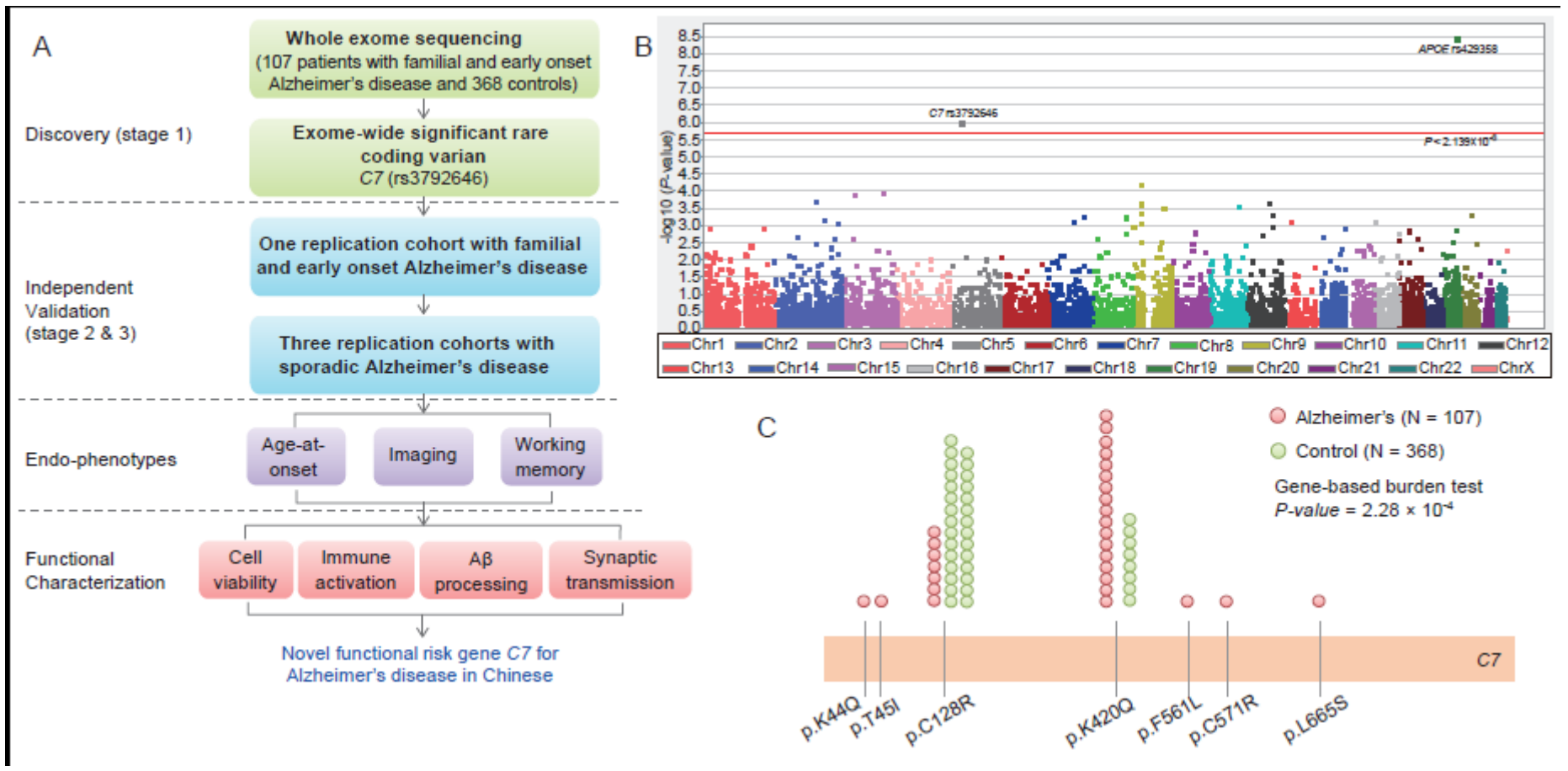


Figure 1. Identification of *C7* variant rs3792646 (p.K420Q) in Han Chinese patients with early-onset and/or familial Alzheimer's disease. (A) Workflow of the current study. (B) Manhattan plot of the exome-wide single site association in 107 cases and 368 population controls for rare and low frequency (MAF < 5%) coding (missense, nonsense, splice site) variants, with *APOE* rs429358 (which defines the ε4 allele) being a positive control. Red line, exome-wide significance. (C) Rare damaging variants of *C7* in Chinese Alzheimer's cases and controls. *P-value*, gene-based burden test.

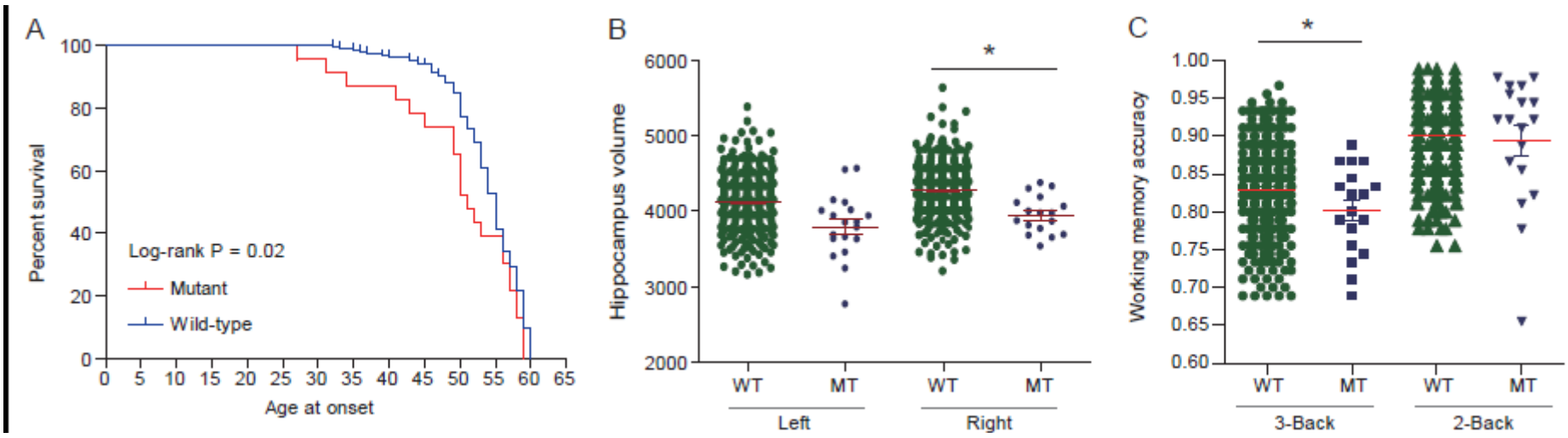


Figure 2. Clinical effects of *C7* rs3792646 (p.K420Q). (A) Effect of rs3792646 on age at onset (AAO) in patients with early onset Alzheimer's disease. Carriers of C7 mutant p.K420Q had a younger AAO relative to carriers of wild type C7. (B-C) Carriers of rs3792646-C (p.K420Q, MT [genotypes CC+AC]) have a decreased hippocampus volume and poor working memory relative to the wild type carriers (rs3792646-A, WT [genotype AA]). Working memory test was performed for 2-back task and 3-back task. Shown was mean ± SD. *, $P < 0.05$.

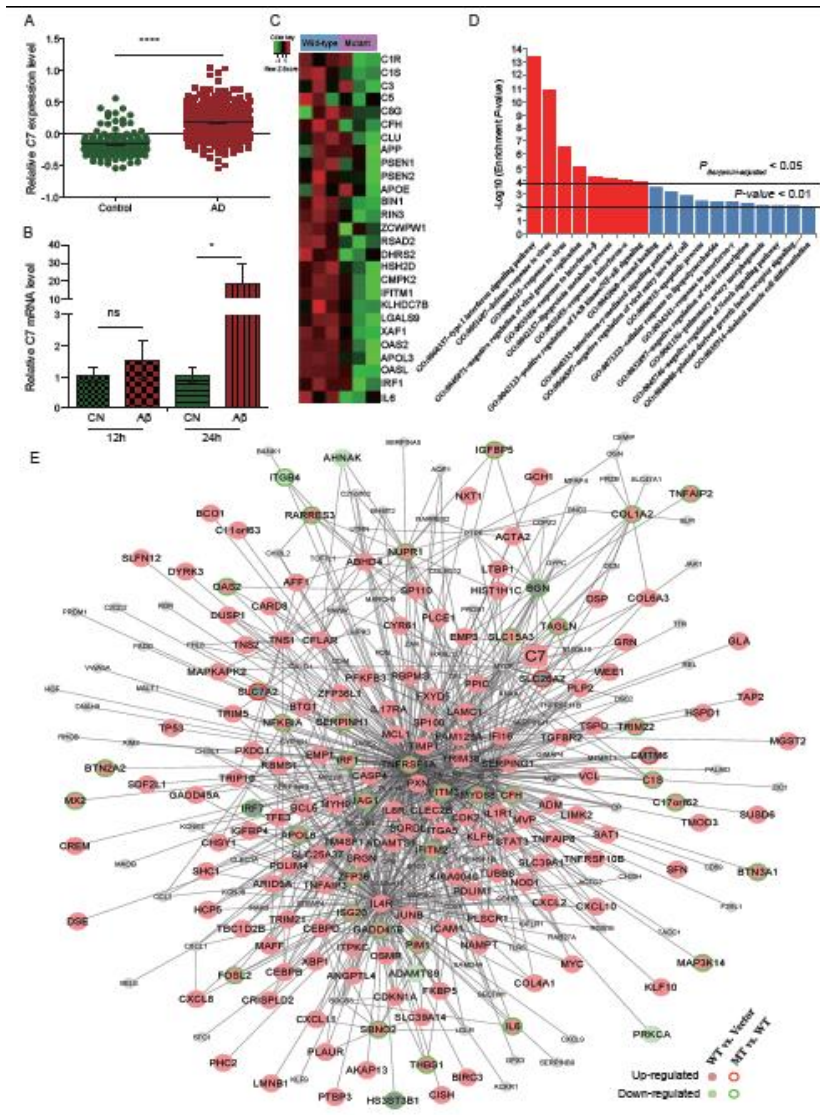


Figure 3. mRNA expression change of *C7* in brain of Alzheimer's patients and effects of overexpressing C7 wild type and mutant on global gene expression pattern. Increased level of *C7* mRNA in frontal cortex of patients and controls based on GSE33000 [58]. ****, $P < 0.0001$. Increased level of *C7* mRNA in Aβ treated cells. U251 cells were treated with 5 μM $A\beta_{42}$ for 12h or 24 h before harvest for real-time quantitative PCR. Aβ was dissolved in 5% DMSO; CN, 5% DMSO. *, $P < 0.05$; ns, not significant. (C) Heatmap of the complement genes, Alzheimer's core genes, and top differentially expressed genes (DEGs) up-regulated in cells overexpressing wild type C7 (Wild type) but downregulated in cells overexpressing mutant p.K420Q (Mutant). (D) Enrichment of genes up-regulated in cells overexpressing wild type C7 (WT, compared with vector) but downregulated in cells overexpressing mutant p.K420Q (MT, compared with WT) in Biological Processes determined by the Gene Ontology terms. DEGs were detected by RNA-seq of U251-APP cells overexpressing wild type and mutant C7. Red, Benjamini-adjusted enrichment-*P*-value <0.05; Blue, original enrichment-*P*-value <0.01. (E) Enrichment of DEGs in response to C7 overexpression in the C7-involved co-expression network. The C7-involved co-expression network (immune module) was dysregulated in Alzheimer's brain tissues according to our recent gene profiling analysis for patients [52]. Significance of the enrichment of DEGs in response to C7 wild type (WT, enrichment $P = 2.57\times10^{-6}$) or mutant (MT, enrichment $P = 2.87\times10^{-10}$) overexpression in the network was measured by Fisher's exact test.

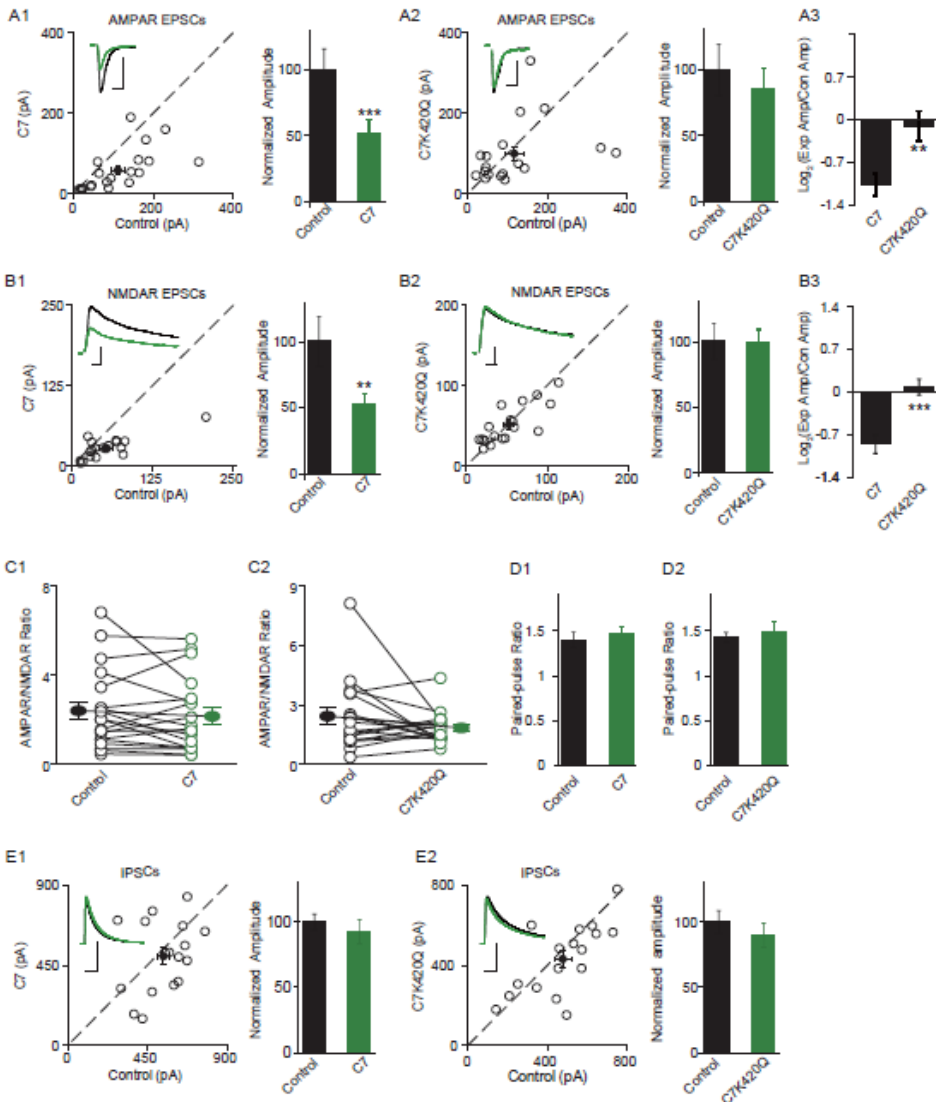


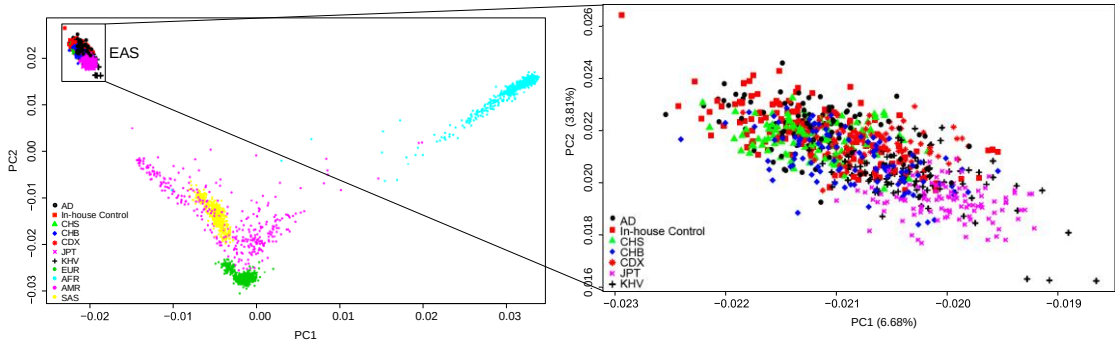
Figure 4. *C7* mutant p.K420Q reverses its physiological regulation of excitatory synaptic transmission. (A-B) Rat hippocampal slice cultures were biolistically transfected with wild type C7 or C7 p.K420Q. Simultaneous dual whole-cell recordings from a transfected CA1 pyramidal neuron (green trace) and a neighboring wild type one (black trace) were performed. The evoked AMPA (A1 and A2) and NMDA (B1 and B2) EPSCs were measured at -70 mV and +40 mV (the current amplitudes were measured 100 ms after stimulation) respectively. Open and filled circles represent amplitudes for single pairs and mean ± SEM, respectively. Insets show sample current traces from control (black) and experimental (green) cells. Bar graphs show normalized EPSC amplitudes (mean ± SEM) of -70 mV (A1, N=20, 51.90 ± 10.45% control, *** $P < 0.001$; A2, N=18, 85.43 ± 15.72% control, $P > 0.05$) and +40 mV (B1, N=19, 52.97 ± 7.33% control, ** $P < 0.005$; B2, 98.37 ± 10.84% control, $P > 0.05$) presented in scatter plots. The scale bars for representative EPSC trace were: 100 pA/25 ms (A1) and 50 pA/25 ms (A2, B1 and B2). All the statistical analyses are compared to respective control neurons with two-tailed Wilcoxon signed-rank sum test. (A3 and B3) Comparison of the logarithm of AMPA EPSC (A3: C7, -1.07 ± 0.19; C7 p.K420Q, -0.12 ± 0.24, ** $P < 0.01$) and NMDA EPSC (B3: C7, -0.86 ± 0.16, C7 p.K420Q, -0.07 ± 0.13, *** $P < 0.0005$) amplitude ratios between the experimental and respective control neurons (mean ± SEM) from wild type C7 and C7 p.K420Q transfections. All statistical analyses are tested using Mann-Whitney U-test. (C) AMPA/NMDA ratios recorded from wild type C7 ($P > 0.05$, N=19) or C7 p.K420Q ($P > 0.05$, N=18) overexpression neurons are not significantly different from respective wild type ones. Two-tailed Wilcoxon signed-rank sum test is used for statistical analyses. (D) No change in paired-pulse ratio, defined as second EPSC over first EPSC, from wild type C7 (control: 1.39 ± 0.09, C7: 1.46 ± 0.09; $P > 0.05$, N = 18) or C7 p.K420Q (control: 1.43 ± 0.05, C7 p.K420Q: 1.49 ± 0.11; $P > 0.05$, N = 18) transfections. (E) Wild type C7 and C7 mutant p.K420Q have no effect on inhibitory synaptic transmission. The same experiments as in Fig. A except that IPSCs were measured at 0 mV. Bar graphs show normalized IPSC amplitudes (mean ± SEM) (E1, N=17, 92.67 ± 9.08% control, $P > 0.05$; E2, N=17, 90.43 ± 8.99% control, $P > 0.05$) presented in scatter plots. The scale bars for representative IPSC trace were: 200 pA/25 ms (E1) and 300 pA/25 ms (E2). All the statistical analyses are compared to respective control neurons with two-tailed Wilcoxon signed-rank sum test.

Table 1. Identification and validation of the association between *C7* variant rs3792646 and Alzheimer's disease in Han Chinese

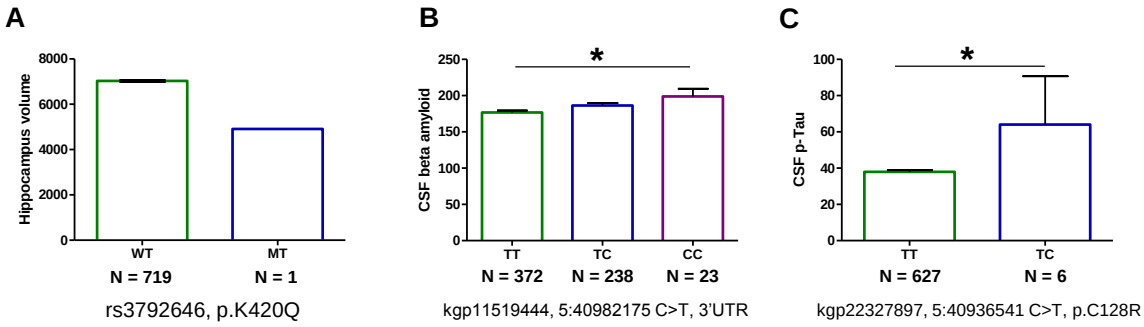
Stage	Type	Region	Alzheimer		Control		*P-value*	OR	95% CI
			Sample size	C/A allele	Sample size	C/A allele			
1	Early-onset/familial	WES[1]	107	17/197	368	8/728	**1.09×10^{-6}**	7.853	3.340-18.464
		ExAC	-	-	4327	295/7836	2.95×10^{-3}	2.292	1.378-3.813
		CONVERGE	-	-	11670	677/22623	2.20×10^{-4}	2.884	1.747-4.761
2	Early-onset/familial	North-Beijing	103	11/195	368	8/728	6.10×10^{-4}	5.133	2.037-12.937
Combined 1 & 2	Early-onset/familial		210	28/392	368	8/728	**3.73×10^{-7}**	6.500	2.934-14.398
3	Sporadic	East	584	44/1124	274	7/541	3.53×10^{-3}	3.025	1.354-6.761
	Sporadic	Southwest	581	45/1117	2190	108/4272	1.17×10^{-2}	1.594	1.119-2.270
	Sporadic	Southcentral	235	16/454	2190	108/4272	0.218	1.394	0.817-2.378
	Sporadic	Pooled South	818	61/1575	2190	108/4272	1.08×10^{-2}	1.532	1.113-2.108
	Sporadic early-onset		248	21/475	2464	115/4813	1.51×10^{-2}	1.85	1.151-2.974
Combined 1-3	All early-onset		421	37/805	2832	123/5541	3.10×10^{-4}	2.066	1.419-3.008
	All late-onset		1194	96/2292	2832	123/5541	8.11×10^{-6}	1.883	1.434-2.472
	All cases		1615	133/3097	2832	123/5541	**2.99×10^{-7}**	1.93	1.503-2.479
European	Sporadic	ADNI	296	1/591	281	0/562	NA	NA	NA
	Sporadic	ADSP	5815	0/11630	4755	1/9509	NA	NA	NA

Note: Same control sample (N=368) was used in stage 1 and stage 2. In stage 3, same control sample (N=2,190) was used for comparing with cases from Southwest China and Southcentral China, respectively. The ADNI European sample was taken from the ADNI dataset [54]; the ADSP European sample was taken from the Alzheimer's Disease Sequencing Project (ADSP) through the dbGaP (phs000572.v7.p4). Data of 4,327 East Asians from the Exome Aggregation Consortium (ExAC) [50] and data of 11,670 Chinese individuals in the CONVERGE Consortium [51] were retrieved as the reference controls. C/A allele, risk allele / reference allele; *P*-value, Fisher exact test; OR, Odds ratio of effect (minor) allele; CI, confidence interval; NA, not applicable. A total of 23,373 functional variants with low allele frequency (MAF <5%) were used in the analysis, with a threshold for the exome-wide significance of $P < 2.139\times10^{-6}$ (Bonferroni corrected: 0.05/23373). The exome-wide significant *P-values* were marked in bold.

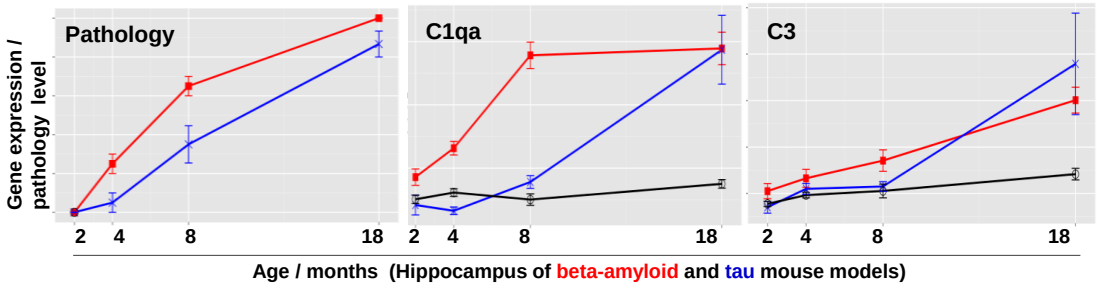
[1] Logistic regression analysis was also performed for stage 1 sample. Suggestive associations of rs3792646 with Alzheimer's disease were observed after adjustment with different covariates: PC1, PC2, and PC3 adjusted $P=9.51\times10^{-5}$, OR= 5.731; *APOE ε4* adjusted $P=5.36\times10^{-4}$, OR= 5.107; Sex adjusted $P=7.29\times10^{-6}$, OR=8.716; PCs, *APOE* ε4, and sex adjusted $P=9.90\times10^{-4}$, OR= 5.382.



Supplementary Figure 1. Principal component analysis (PCA) showed no population stratification between the studied subjects and the reference Chinese populations. Han Chinese in Beijing (CHB) and Southern Han Chinese (CHS) from the 1000 Genome Project phase 3 [1] were used as the reference Chinese populations. A total of 112,075 common SNPs shared by subjects undergoing whole exome sequencing in this study and the 1000 genomes were used in the PCA by using the GCTA tool (http://cnsgenomics.com/software/gcta/#Overview). The PCA distinguished clearly the East Asian populations from the populations outside of East Asia, suggesting that these selected SNPs contain ancestry-informative markers. Based on the clustering pattern, there is no obvious population substructure among the East Asian populations. This result suggested that it is reasonable to group our in-house controls (N=160) [2] with the two Han Chinese populations from the 1000 Genome Project (CHB, N=103; CHS, N=105) [1] as the general population control (N = 368). Abbreviations in the principal component map: AD, Han Chinese with Alzheimer's disease; In-house Control, non-dementia Han Chinese individuals; CHS, Southern Han Chinese; CHB, Han Chinese in Beijing, China; CDX, Chinese Dai in Xishuangbanna, China; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; EAS, East Asian; EUR, European; AFR, African; AMR, Ad Mixed American; SAS, South Asian.

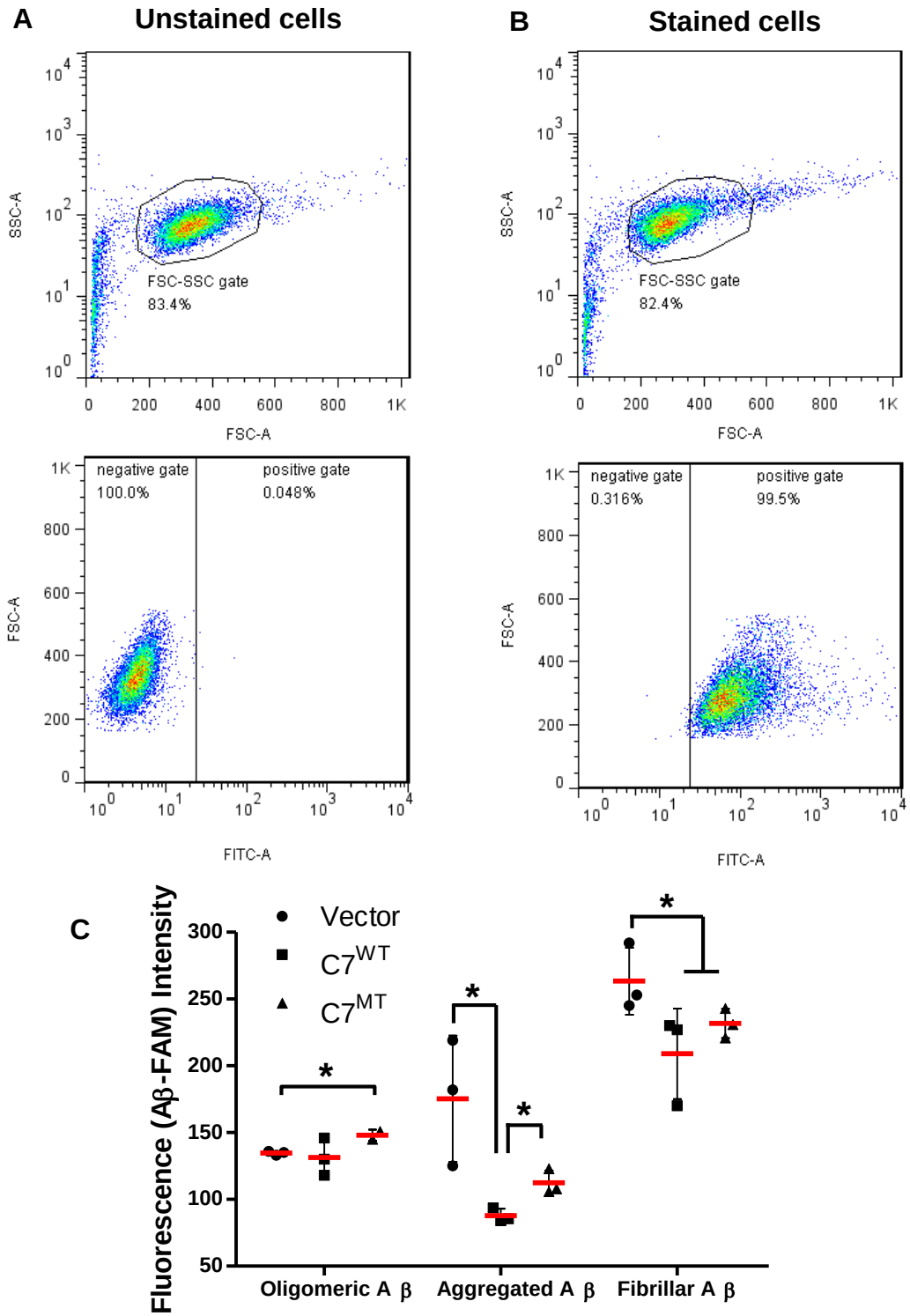


Supplementary Figure 2. **Effects of *C7* variants on endo-phenotypes in ADNI samples.** (**A**) Decreased hippocampus volume of the rs3792646-C (p.K420Q) carrier (MT, genotype AC) compared with wild type carriers (WT, genotype AA) in the ADNI sample containing 812 individuals [3]. The effect of AD-risk SNP on hippocampus volume was analyzed by using PLINK [4]. (**B-C**) Two rare variants of *C7* affected the cerebrospinal fluid (CSF) Aβ and p-tau levels. *, $P < 0.05$, linear regression analysis; values were shown as mean ± SD.



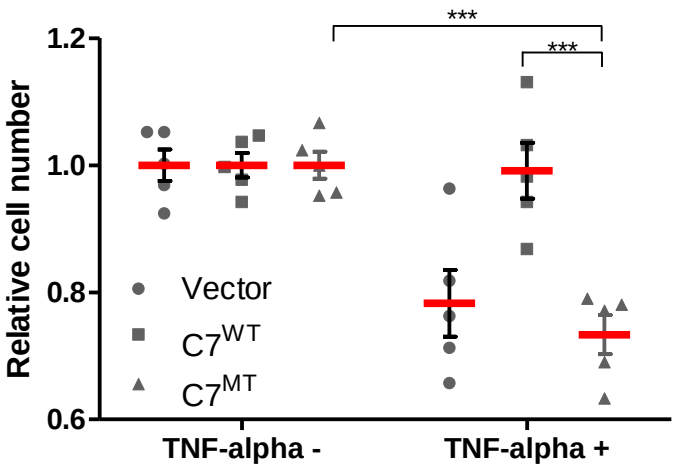
Supplementary Figure 3. Expression change of *C1q* and *C3* mRNA levels in hippocampus tissues of AD mouse models. *C1q* and *C3* mRNA expression levels increase along with the severity level of pathology (Aβ plaques [red] and tau tangles [blue]) [5]. Expression data and pathological features of wild type and AD mouse models were downloaded from the Mouseac database (http://www.mouseac.org) [5]. Red line, transgenic mice with homozygous human mutant APP (K670N/M671L) and PSEN1 (M146V, TPM), HO _TASTPM; Blue line, transgenic mice with human mutant MAPT (P301L), TAU; Black line, wide type mice. Data shown were mean ± SD.

.



Supplementary Figure 4. Overexpression of C7 mutant p.K420Q affected the internalization of Aβ in human microglia (HM) cells. HM cells were treated with 5 μM oligomeric, aggregated, and fibrillary fluorescently-labeled $A\beta_{42}$ for 24 h after transfection of expression vectors for the C7 wild type ($C7^{WT}$) and mutant p.K420Q ($C7^{MT}$) and empty vector (Vector), respectively. Cells were harvested 2 h after $A\beta_{42}$

treatment (in triples). Fluorescence intensity was measured by flow cytometry based on 10,000 cells. The FlowJo software was used for viewing and analyzing flow cytometric data. Starting cell population was determined by forward and side scatter gating, to remove debris, cell fragments, and pyknotic cells. The events with very low FSC and SSC, as well as those with very high FSC and SSC are eliminated and the major (>80%) density of events is captured by this gate (**A**-**B**, top panel). After identification of the cell population of interest, unstained cells (without $A\beta_{42}$ treatment) were used as negative controls in setting the voltages and negative gates (**A**-**B**, bottom panel), to determine the level of background fluorescence or autofluorescence. (**C**) The mean fluorescence intensity (mean ± SD) of stained cell population for each group was compared by Student's *t* test. *, two-tailed *P*-value < 0.05.



Supplementary Figure 5. Overexpression of C7 mutant p.K420Q promoted cell apoptosis. Cell viability induced by TNF-α was determined by using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Sigma, #M2128) assay. HM cells were seeded in each well of 96-well plates at a density of 5×10^3 cells per well after transfection of expression vectors for the C7 wild type ($C7^{WT}$), mutant p.K420Q ($C7^{MT}$) and empty vector (Vector) for 24 h, respectively, then were treated with 2 μg/mL Actinomycin D (Merck Millipore, #129935) and 200 ng/ml TNF-α (peproTech, #300-01A). After 24 h incubation, the MTT assay was performed according to the manufacture's instruction. Absorbance measurements were obtained by using a Gen5 plate reader (Elx808, BioTek) at 490 nm. Shown (mean ± SD) were relative values of absorbance normalized to the corresponding treatment without TNF-α treatment. Difference measured by Student's *t* test. ***, two-tailed *P*-value < 0.001.

Supplementary Table 1. Top 100 hits showing suggestive significant associations ($P < 0.01$) with Alzheimer's disease in the whole exome sequencing stage

Chr	Position	SNP_ID	Allele	Gene	Function	AC/AN_AD	AC/AN_Ctrl	Fisher *P*	Fisher OR	PC adj *P*	PC adj OR	Sex adj *P*	Sex adj OR	APOE adj *P*	APOE adj OR	PC Sex APOE adj *P*	PC Sex APOE adj OR
chr19	45411941	rs429358	T/C	APOE	p.130C>R	51/208	61/736	3.4E-09	3.59	2.0E-07	3.20	4.2E-06	2.87	1.9E-02	4.59	NA	0.11
chr5	40955653	rs3792646	A/C	C7	p.420K>Q	17/214	8/736	1.1E-06	7.85	9.5E-05	5.73	7.3E-06	8.72	5.4E-04	5.11	9.9E-04	5.38
chr2	85570857	rs4832169	G/A	RETSAT	p.533A>V	26/214	5/160	2.0E-03	4.29	1.3E-01	3.54	1.1E-04	7.55	9.3E-03	4.03	9.7E-02	9.72
chr2	152566961	rs36105240	T/C	NEB	p.305D>G	13/214	9/736	1.9E-04	5.22	3.6E-03	3.78	1.5E-04	5.85	6.2E-04	4.87	1.9E-03	5.10
chr11	102584176	rs34009635	A/G	MMP8	p.436V>A	15/212	13/736	2.5E-04	4.23	4.5E-04	4.31	1.5E-04	4.97	1.1E-04	5.07	1.2E-03	5.04
chr2	182780126	rs78774163	G/A	SSFA2	p.587D>N	22/214	28/736	6.9E-04	2.90	5.8E-03	2.44	1.6E-04	3.34	5.7E-04	3.01	8.0E-03	2.82
chr3	38167080	rs117916664	T/C	ACAA1	p.299N>S	14/214	15/736	2.3E-03	3.36	1.4E-02	2.77	2.8E-04	4.49	4.0E-04	4.22	5.3E-03	3.86
chr12	93171826	rs77729665	G/A	EEA1	p.1262R>W	20/212	23/736	4.9E-04	3.23	9.9E-04	3.14	4.4E-04	3.47	3.2E-03	2.84	3.9E-03	3.47
chr2	231406070	rs17275036	G/A	SP100	p.796A>T	16/214	11/526	8.7E-04	3.78	1.1E-02	3.16	5.3E-04	4.35	1.2E-03	3.96	2.4E-02	3.34
chr12	96371782	rs183059673	C/T	HAL	p.532V>I	9/212	3/526	1.1E-03	7.73	4.4E-03	7.75	6.3E-04	10.81	1.2E-02	5.99	1.3E-03	15.46
chr9	104188842	rs3739721	C/G	ALDOB	p.207E>Q	9/212	9/736	8.6E-03	3.58	3.8E-02	3.00	6.7E-04	5.33	1.5E-02	3.65	8.7E-03	5.38
chr7	94940782	rs13306698	T/C	PON1	p.160R>G	25/212	36/736	7.1E-04	2.60	1.0E-03	2.64	8.5E-04	2.82	1.0E-03	2.70	5.2E-03	2.84
chr16	4908667	rs3747614	A/G	UBN1	spliceSite	17/212	19/736	7.7E-04	3.29	2.3E-04	4.27	1.1E-03	3.54	1.8E-02	2.61	1.4E-03	5.35
chr13	25831336	rs75005059	T/C	MTMR6	p.365M>V	11/212	8/736	7.5E-04	4.98	7.3E-03	4.22	1.2E-03	4.97	3.3E-03	4.36	6.8E-03	5.26
chr17	80040034	rs2228306	A/G	FASN	p.2005V>A	8/212	6/736	4.7E-03	4.77	1.4E-02	4.38	1.5E-03	6.19	8.9E-03	4.88	8.3E-03	6.56
chr12	56351128	rs2071024	G/T	SILV	p.320P>H	14/212	14/736	1.8E-03	3.65	5.2E-03	3.14	1.7E-03	3.66	1.1E-02	2.89	1.9E-01	1.97
chr2	231368897	rs6705605	G/T	SP100	spliceSite	16/214	14/526	6.2E-03	2.96	4.1E-02	2.37	1.9E-03	3.44	3.6E-03	3.20	5.6E-02	2.58
chr15	101606145	rs35128996	C/T	LRRK1	p.1835L>F	12/210	14/736	6.5E-03	3.13	3.4E-03	3.78	2.2E-03	3.80	2.3E-02	2.78	3.6E-03	5.17
chr16	89261471	rs149307887	G/A	CDH15	p.785G>R	16/212	19/736	1.6E-03	3.08	5.1E-02	2.12	2.3E-03	3.31	1.2E-02	2.82	2.8E-01	1.69
chr14	50605490	.	G/T	SOS2	p.933T>K	17/212	3/160	9.8E-03	4.56	4.7E-02	6.82	2.4E-03	7.21	5.1E-03	6.43	4.3E-03	31.57
chr20	61167658	rs145416632	C/G	C20orf166	p.43P>R	10/212	9/736	3.5E-03	4.00	3.1E-03	4.47	2.5E-03	4.61	3.4E-04	5.79	1.2E-03	6.49

chr9	104161450	rs61755098	G/A	ZNF189	p.4P>L	19/208	27/736	2.9E-03	2.64	3.4E-02	2.02	2.6E-03	2.81	8.4E-03	2.55	3.1E-02	2.46
chr16	30123523	rs138146407	C/A	GDPD3	p.168R>L	8/212	7/736	8.2E-03	4.08	3.1E-02	3.42	2.8E-03	5.20	3.8E-02	3.42	1.7E-02	4.82
chr2	71649966	rs61739715	A/G	ZNF638	p.1108I>V	19/214	30/736	8.0E-03	2.29	6.4E-02	1.91	3.0E-03	2.81	1.9E-03	2.89	6.4E-02	2.22
chr16	84256422	rs4782905	C/T	KCNG4	p.321E>K	15/212	21/736	7.5E-03	2.59	9.2E-02	1.89	3.0E-03	2.91	2.0E-03	3.04	1.3E-01	1.94
chrX	2867424	rs138149353	G/C	ARSE	p.259H>D	12/212	13/736	5.3E-03	3.34	3.3E-02	2.38	3.3E-03	3.09	3.5E-03	3.09	8.3E-03	3.70
chr19	10426524	rs79442975	G/A	FDX1L	spliceSite	15/206	22/736	8.0E-03	2.55	2.9E-02	2.23	3.4E-03	2.91	2.0E-02	2.38	1.1E-01	2.03
chr1	144931251	rs142679243	C/T	PDE4DIP	p.153S>N	7/214	5/736	7.5E-03	4.94	6.3E-02	3.31	3.6E-03	6.09	3.6E-02	4.06	6.2E-02	4.26
chr20	44047974	rs80158178	G/A	PIGT	p.178R>Q	12/212	9/736	5.0E-04	4.85	5.9E-03	3.98	3.7E-03	4.18	2.3E-02	3.13	5.6E-02	3.41
chr19	55795872	.	A/C	BRSK1	p.21H>P	22/126	4/120	2.9E-04	6.13	2.1E-01	2.47	4.0E-03	3.77	1.1E-02	3.37	7.9E-02	10.02
chr17	1944781	rs200231675	G/C	DPH1	p.370V>L	7/202	2/526	2.5E-03	9.41	1.2E-02	9.63	4.1E-03	10.76	1.0E-02	8.79	1.6E-02	18.83
chr2	152420386	rs147159176	C/T	NEB	p.4476R>H	9/214	7/736	3.3E-03	4.57	3.3E-03	5.09	4.2E-03	4.87	3.8E-03	5.10	1.3E-02	5.25
chr8	28321247	rs3735726	C/T	FBXO16	p.75R>Q	8/212	7/736	8.2E-03	4.08	1.5E-02	4.23	4.2E-03	4.87	5.5E-03	4.62	1.3E-02	5.23
chr19	12059699	rs75607624	T/G	ZNF700	p.287F>C	7/212	5/736	7.2E-03	4.99	1.1E-02	5.44	4.2E-03	5.92	1.4E-03	7.07	2.2E-03	11.21
chr1	15428057	rs140076587	G/T	KAZ	p.522E>D	10/214	7/736	1.2E-03	5.11	5.3E-02	2.73	4.5E-03	4.53	2.8E-03	4.90	1.5E-01	2.42
chr10	75138691	rs3750575	C/T	ANXA7	p.419R>Q	15/212	19/736	5.1E-03	2.87	7.4E-02	1.86	4.5E-03	2.80	2.5E-01	1.58	8.0E-01	1.13
chr11	130750661	rs147010503	G/A	SNX19	p.872R>C	11/212	11/736	3.7E-03	3.61	7.7E-03	3.18	5.7E-03	3.83	3.4E-03	3.67	1.6E-02	3.55
chr1	167780071	rs117021474	C/T	ADCY10	p.1521C>Y	7/214	4/736	3.8E-03	6.19	9.5E-03	4.93	6.3E-03	6.52	2.3E-02	4.96	5.3E-02	5.00
chr14	94546058	rs142609376	T/C	DDX24	p.11K>E	8/212	6/736	4.7E-03	4.77	1.2E-02	4.41	6.7E-03	5.00	7.5E-03	4.84	5.9E-03	6.14
chr15	63893706	rs181302627	C/T	FBXL22	p.183P>S	9/206	9/736	7.4E-03	3.69	2.8E-03	4.91	6.9E-03	4.12	1.9E-02	3.41	1.8E-02	4.69
chr9	88257811	rs143779850	T/A	AGTPBP1	p.371E>D	8/212	7/736	8.2E-03	4.08	8.6E-02	2.64	8.1E-03	4.06	6.4E-02	2.96	1.4E-01	2.90
chr3	129811029	rs191831656	C/T	ALG1L2	p.73R>W	8/214	6/736	5.0E-03	4.72	1.4E-02	3.76	8.7E-03	4.07	1.6E-02	3.53	1.1E-02	4.25
chr8	120629807	rs148588719	A/C	ENPP2	spliceSite	17/212	18/736	5.7E-04	3.48	3.4E-03	2.93	9.6E-03	2.96	1.4E-03	3.50	4.9E-02	2.53
chr17	37829778	rs142596676	T/C	PGAP3	p.228N>S	11/212	9/736	1.3E-03	4.42	1.2E-02	3.44	9.8E-03	3.85	4.9E-02	3.01	5.4E-02	3.38
chr14	89171861	rs144622692	T/C	EML5	p.633I>V	6/212	2/576	6.0E-03	8.36	3.5E-02	5.77	1.0E-02	8.96	1.3E-02	8.64	1.2E-01	4.48
chr2	209204243	rs148994064	G/A	PIKFYVE	p.1592G>R	8/212	4/576	4.3E-03	5.61	3.8E-03	6.86	1.0E-02	5.49	1.0E-02	5.47	4.2E-02	5.42
chr19	36674347	rs79279971	T/C	ZNF565	p.174K>R	9/212	7/736	3.1E-03	4.62	2.6E-02	3.43	1.2E-02	4.22	6.8E-03	4.63	9.0E-02	3.11

chr8	120592406	rs2289886	T/C	ENPP2	p.629N>S	16/212	19/736	1.6E-03	3.08	1.0E-02	2.67	1.3E-02	2.81	2.1E-03	3.36	5.5E-02	2.45
chr15	63014548	rs10775181	A/G	TLN2	spliceSite	200/212	723/736	5.3E-03	0.30	1.1E-03	0.22	1.3E-02	0.31	6.3E-03	0.30	2.6E-02	0.26
chr10	85944516	rs76221724	G/T	C10orf99	p.80Q>H	13/212	13/736	1.6E-03	3.63	8.1E-04	4.29	1.3E-02	3.21	1.4E-02	2.96	7.7E-02	2.70
chr14	21790040	rs10151259	G/T	RPGRIP1	p.547A>S	6/212	1/576	2.0E-03	16.75	5.7E-03	21.68	1.5E-02	15.42	1.2E-02	16.92	9.5E-03	23.27
chr15	80450501	rs151264725	G/T	FAH	p.61V>F	11/212	11/736	3.7E-03	3.61	1.3E-02	3.09	1.6E-02	3.10	1.6E-02	3.12	5.0E-02	3.02
chr3	170715865	rs140138702	G/C	SLC2A2	p.468L>V	10/214	10/736	5.9E-03	3.56	6.3E-03	4.05	1.6E-02	3.43	9.6E-04	4.87	1.9E-02	4.30
chr10	121565909	rs3736822	A/G	INPP5F	p.453I>V	10/212	10/736	5.6E-03	3.59	1.5E-03	5.23	1.6E-02	3.43	2.6E-02	3.11	1.1E-02	5.63
chr1	55642119	rs117816458	T/A	USP24	spliceSite	5/214	1/526	9.0E-03	12.56	3.1E-02	11.18	1.9E-02	14.12	9.1E-03	18.23	1.8E-02	18.82
chr19	49362376	.	C/G	PLEKHA4	p.238R>P	15/198	1/160	1.3E-03	13.03	7.5E-02	18.63	1.9E-02	11.16	3.1E-02	9.37	6.9E-02	11.49
chr1	94048138	rs138527879	C/T	BCAR3	p.469R>Q	6/214	2/526	8.9E-03	7.56	2.2E-02	7.78	2.1E-02	7.66	1.6E-02	7.69	1.7E-01	3.66
chr10	17659338	rs368871717	T/C	PTPLA	p.1M>V	6/106	3/404	3.5E-03	8.02	1.5E-02	4.86	2.2E-02	4.04	1.2E-01	2.82	5.3E-02	7.35
chr1	976598	rs200607541	C/T	AGRN	p.258T>I	7/124	7/508	9.6E-03	4.28	1.8E-02	3.38	2.3E-02	3.01	2.6E-02	3.01	7.9E-03	4.73
chr8	52321722	rs200216958	A/C	PXDNL	p.821L>R	5/208	1/576	6.0E-03	14.16	6.2E-02	8.00	2.3E-02	12.07	2.6E-02	11.89	2.2E-02	15.25
chr9	4662580	rs190018180	G/A	PPAPDC2	p.69G>S	11/206	9/736	1.1E-03	4.56	4.9E-04	6.72	3.2E-02	3.21	5.9E-03	4.02	1.2E-02	5.79
chr2	209215654	rs137922460	C/T	PIKFYVE	p.1865A>V	8/214	7/736	8.6E-03	4.04	1.8E-02	3.91	3.3E-02	3.60	3.8E-03	5.10	2.7E-01	2.38
chr1	36552858	rs142743253	G/T	TEKT2	p.267K>N	8/214	7/736	8.6E-03	4.04	5.2E-03	4.90	3.3E-02	3.60	7.9E-03	4.68	1.4E-02	5.82
chr15	42138159	rs144874529	A/G	JMJD7-PLA2G4B	p.736Y>C	5/200	1/576	5.2E-03	14.74	6.9E-02	6.97	3.4E-02	10.99	9.3E-02	7.29	7.3E-02	7.80
chr5	60050522	rs116939630	G/A	ELOVL7	p.259R>C	2/212	35/736	8.2E-03	0.19	2.4E-02	0.18	4.1E-02	0.12	2.5E-02	0.10	1.0E+00	0.00
chr8	30701641	rs142485241	C/G	TEX15	p.1631Q>H	2/212	35/736	8.2E-03	0.19	6.0E-02	0.25	4.2E-02	0.13	6.9E-02	0.26	1.3E-01	0.20
chr11	68854029	rs78034812	C/T	TPCN2	p.681S>L	18/212	28/736	9.8E-03	2.35	2.1E-02	2.19	4.2E-02	2.08	7.0E-03	2.50	1.0E-01	2.05
chr11	119156193	rs2227988	C/T	CBL	p.620L>F	2/212	35/736	8.2E-03	0.19	8.1E-02	0.27	4.5E-02	0.13	5.1E-02	0.24	2.2E-01	0.28
chr11	58125774	rs55810057	A/G	OR5B17	p.257Y>H	7/212	5/736	7.2E-03	4.99	9.9E-03	5.46	4.6E-02	3.93	3.7E-03	6.21	5.2E-03	9.84
chr1	16725271	rs117944955	G/A	SPATA21	p.467R>*	2/214	35/736	8.1E-03	0.19	1.1E-01	0.31	4.9E-02	0.14	5.6E-02	0.14	1.0E+00	0.00
chr9	99413954	rs144710877	T/G	C9orf21	p.101Y>S	1/212	31/736	4.3E-03	0.11	2.8E-02	0.10	6.0E-02	0.15	7.6E-02	0.16	7.4E-02	0.15
chr15	89453152	rs143117049	T/C	MFGE8	p.26I>V	1/212	31/736	4.3E-03	0.11	7.7E-02	0.16	6.1E-02	0.15	3.9E-02	0.12	2.2E-01	0.28
chr2	54023133	rs192316706	A/G	ERLEC1	p.111S>G	1/214	30/736	7.1E-03	0.11	1.5E-01	0.23	6.5E-02	0.15	1.0E+00	0.00	1.0E+00	0.00

chr8	25230071	rs143521106	C/G	DOCK5	spliceSite	1/212	30/736	7.0E-03	0.11	1.0E-01	0.18	6.5E-02	0.15	1.0E+00	0.00	1.0E+00	0.00
chr1	155217643	rs2072648	C/T	FAM189B	p.646R>H	1/212	30/736	7.0E-03	0.11	1.4E-02	0.08	6.7E-02	0.15	6.3E-02	0.15	3.6E-02	0.11
chr19	36336398	rs114615449	C/G	NPHS1	p.601G>A	1/208	29/736	7.0E-03	0.12	3.0E-02	0.10	6.7E-02	0.15	4.1E-02	0.12	6.8E-02	0.14
chr1	11008844	rs117528334	C/G	C1orf127	p.283A>L	1/214	30/736	7.1E-03	0.11	7.0E-02	0.15	7.4E-02	0.16	6.1E-02	0.15	2.2E-01	0.25
chr8	19362969	rs140161612	G/A	CSGALNACT1	p.126S>L	11/212	10/736	2.3E-03	3.97	1.4E-02	3.23	1.5E-01	2.14	4.0E-03	3.90	2.0E-01	2.09
chr7	44180676	rs117394324	G/A	MYL7	spliceSite	7/210	5/736	6.9E-03	5.04	5.5E-02	3.45	1.6E-01	2.84	5.8E-02	3.57	4.0E-01	1.93
chr17	45786519	rs369896558	G/A	TBKBP1	p.474A>T	6/114	5/512	6.8E-03	5.63	3.8E-03	4.75	8.6E-01	1.18	2.5E-02	3.38	2.2E-01	3.78
chr7	56128100	rs192970041	A/G	CCT6A	p.402I>V	1/212	30/736	7.0E-03	0.11	9.7E-02	0.18	1.0E+00	0.00	4.6E-02	0.13	1.0E+00	0.00
chr12	80169728	.	A/C	PPP1R12A	p.938I>R	15/210	0/160	2.2E-04	25.45	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32488848	.	T/G	DDX58	p.279E>D	17/212	0/160	6.2E-05	28.73	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32480272	.	A/T	DDX58	p.573N>K	15/212	0/160	2.2E-04	25.19	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32488815	.	C/A	DDX58	p.290K>N	15/210	0/160	2.2E-04	25.45	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32480253	.	C/T	DDX58	p.580D>K	14/212	0/160	4.3E-04	23.45	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32708645	.	C/A	RP11-555J4.2	spliceSite	11/166	0/158	8.5E-04	23.44	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32467869	.	A/C	DDX58	p.692N>K	13/212	0/160	8.2E-04	21.72	NA	NA	NA	NA	NA	NA	NA	NA
chr14	102973415	rs199786978	G/C	ANKRD9	p.271A>G	11/188	0/160	1.2E-03	20.80	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32457310	.	A/T	DDX58	p.863F>Y	11/212	0/160	3.1E-03	18.32	NA	NA	NA	NA	NA	NA	NA	NA
chr9	32466410	.	A/T	DDX58	p.739F>I	11/212	0/160	3.1E-03	18.32	NA	NA	NA	NA	NA	NA	NA	NA
chr1	147092352	rs200395772	G/T	BCL9	p.797L>F	10/214	0/160	6.1E-03	16.48	NA	NA	NA	NA	NA	NA	NA	NA
chr16	20359959	.	G/T	UMOD	p.222R>S	8/196	0/160	9.4E-03	14.47	NA	NA	NA	NA	NA	NA	NA	NA
chr1	36298037	.	G/A	EIF2C4	spliceSite	9/208	0/160	5.9E-03	15.29	NA	NA	NA	NA	NA	NA	NA	NA
chr2	216237006	rs76749241	C/T	FN1	p.1998V>I	6/214	0/366	2.4E-03	22.85	NA	NA	NA	NA	NA	NA	NA	NA
chr17	66890377	rs530154281	A/T	ABCA8	p.951N>K	6/212	0/366	2.3E-03	23.07	NA	NA	NA	NA	NA	NA	NA	NA
chr5	134210196	.	G/T	TXNDC15	p.27G>*	8/198	0/160	9.7E-03	14.32	NA	NA	NA	NA	NA	NA	NA	NA
chr1	2560896	.	T/C	MMEL1	p.1M>V	10/210	0/160	6.0E-03	16.81	NA	NA	NA	NA	NA	NA	NA	NA
chr19	407689	.	A/G	C2CD4C	p.225S>P	9/194	0/148	6.1E-03	15.21	NA	NA	NA	NA	NA	NA	NA	NA

chr17	65989048	rs28368756	T/C	C17orf58	p.72E>G	208/212	736/736	2.4E-03	0.03	NA	NA	NA	NA	NA	NA	NA	NA

Note: Summary statistics of the exome-wide variants were available at the AlzData webserver (http://www.alzdata.org/exome.html), which was established in our previous study [6].

Chr, chromosomal number

Position, chromosomal location of target variant according to hg19 (http://asia.ensembl.org/info/website/tutorials/grch37.html)

SNP_ID, rs# in dbSNP (https://www.ncbi.nlm.nih.gov/snp/)

Allele, reference allele / alterative allele

Gene, gene containing the target variant

Function, consequence of the target variant on protein coding

AC/AN_AD, allele count / total number of alleles in patients with Alzheimer's disease

AC/AN_Ctrl, allele count / total number of alleles in healthy controls. The exome data of 160 in-house non-dementia individuals [2] were pooled with the whole genome data of Han Chinese in Beijing (N=103) and Southern Han Chinese (N=105) from the 1000 Genome Project phase 3 [1] as the initial population control (N = 368). The total number of alleles might be different for some variants, as the call rate for each variant varies due to different sequencing platform.

Fisher *P*, *P*-value of the Fisher's exact test for allele frequency difference between cases and controls

Fisher OR, odds ratio of the Fisher's exact test for the alternative allele relative to the reference allele

PC adj *P*, adjusted *P*-value based on the top three principal components as estimated in Supplementary Figure 1

PC adj OR, adjusted odds ratio for the alternative allele relative to the reference allele based on the top three principal components as estimated in Supplementary Figure 1

Sex adj *P*, adjusted *P*-value by sex

Sex adj OR, adjusted odds ratio for the alternative allele relative to the reference allele by sex

APOE adj *P*, adjusted *P*-value by APOE ε4 status

APOE adj OR, adjusted odds ratio for the alternative allele relative to the reference allele by APOE ε4 status

PC Sex APOE adj *P*, adjusted *P*-value with the top three principal components, sex, and APOE4 ε4 status as covariates

PC Sex APOE adj OR, adjusted odds ratio for the alternative allele relative to the reference allele, with the top three principal components, sex, and APOE4 ε4 status as covariates

Supplementary Table 2. Association of *C7* variant rs3792646 with Alzheimer's disease stratified by *APOE* ε4 status

Sample [1]	Allele	APOE ε4+					APOE ε4-				
		Alzheimer	Control	*P*-value[2]	OR	95% CI	Alzheimer	Control	*P*-value[2]	OR	95% CI
WES	C	10	1	**0.00072**	16.143	2.022-128.850	6	7	**0.01385**	4.286	1.416-12.972
	A	70	113				123	615			
North-Beijing	C	4	1	**0.03052**	9.826	1.069-90.291	3	7	**0.04973**	4.624	1.164-18.371
	A	46	113				57	615			
East	C	16	3	0.77767	1.34	0.38281-4.692	21	4	**0.00375**	4.291	1.462-12.595
	A	386	97				531	434			
Southwest	C	10	9	**0.01475**	3.214	1.291-8.000	17	76	0.11594	1.54	0.90061-2.632
	A	252	729				401	2760			
Southcentral	C	6	9	**0.03739**	3.115	1.093-8.879	10	76	0.579	1.219	0.62355-2.382
	A	156	729				298	2760			
Combined	C	46	13	**1.435E-05**	3.651	1.960-6.803	57	87	**0.00122**	1.77	1.260-2.485
	A	910	939				1410	3809			

Note: A allele, reference allele; C allele, risk allele; *P*-value, Fisher's exact test for allele frequency difference between cases and controls; OR, Odds ratio of the Fisher's exact test for the alternative allele relative to the reference allele; 95% CI, 95% confidence interval

[1] Only individuals with genotyping information for both rs3792646 and the APOE ε4 status were included in the analyses

[2] *P-values* < 0.05 were marked in bold.

Supplementary Table 3. No association of *C7* variants with Alzheimer's disease in the ADNI European individuals

Chr: Position	Allele	SNP	F_A	F_U	*P*	OR	Annotation	Mutation	Prediction
5:40909694	T/C	kgp22074974	0.000	0.004	0.146	0.000	5'utr	-	-
5:40936541	C/T	kgp22327897	0.007	0.002	0.198	3.816	Coding	p.C128R	Probably damaging
5:40945397	A/G	kgp22664496	0.002	0.004	0.533	0.474	Coding	p.R222H	Possibly damaging
5:40955561	C/G	rs1063499	0.394	0.427	0.248	0.871	Coding	p.S389T	Probably damaging
5:40955653	C/A	rs3792646	0.002	0.000	0.330	NA	Coding	p.K420Q	Damaging
5:40964852	C/A	rs13157656	0.233	0.230	0.886	1.020	Coding	p.T587P	Benign
5:40981927	G/A	kgp3954039	0.019	0.027	0.354	0.690	3'utr	-	-
5:40982031	C/T	rs10473230	0.159	0.155	0.853	1.031	3'utr	-	-
5:40982175	C/T	kgp11519444	0.213	0.226	0.590	0.926	3'utr	-	-
5:40982620	T/G	kgp6970181	0.210	0.210	0.983	0.997	3'utr	-	-
5:40982622	T/G	rs1061443	0.208	0.210	0.927	0.987	3'utr	-	-
5:40982780	G/A	rs8264	0.159	0.155	0.853	1.031	3'utr	-	-
5:40982977	G/A	kgp9618283	0.208	0.210	0.927	0.987	3'utr	-	-

Note: Data were retrieved from the ADNI WGS phase (http://adni.loni.usc.edu/) [3]. Allele, reference allele/alternative allele; SNP: rs number in dbSNP dataset and the original SNP number (labeled with "kgp") in the ADNI dataset; F_A, allele frequency in patients (N = 296); F_U, allele frequency in controls (N = 281); *P*, Fisher's exact test *P*-value, OR, odds ratio; utr, untranslated region. Data were processed by using PLINK [4]. Variants annotation was performed by the web tool SNPnexus (http://snp-nexus.org/index.html). For missense substitution, we provided the predicted effect on protein function (possibly damaging, probably damaging, and benign) based on the PolyPhen program [7].

Supplementary Table 4. mRNA expression pattern of the complement components in frontal cortex tissues of patients with Alzheimer's disease compared with controls

Catalog	Gene	*P-value*	Log_2FC
Initial	***C1QA***	$\mathbf{1.80\times10^{-18}}$	0.313052
	C1QB	$\mathbf{1.07\times10^{-15}}$	0.236765
	C1QC	$\mathbf{1.90\times10^{-13}}$	0.258328
Central	***C2***	$\mathbf{8.45\times10^{-04}}$	0.068271
	C3	$\mathbf{4.01\times10^{-09}}$	0.19697
Terminal	*C5*	8.26×10^{-1}	-0.00303
	C6	5.44×10^{-1}	0.011006
	C7	$\mathbf{3.21\times10^{-15}}$	0.242259
	C8B	4.11×10^{-1}	0.017471
	C8G	4.37×10^{-2}	0.017017
Regulator	***C1QBP***	$\mathbf{1.95\times10^{-10}}$	-0.06962
	CR2	3.29×10^{-2}	0.018832
	C3AR1	$\mathbf{6.39\times10^{-17}}$	0.250126
	C4BPA	8.83×10^{-2}	0.069149
	C4BPB	8.33×10^{-1}	0.003635
	CFHR3	8.81×10^{-3}	-0.0173
	CFHR4	9.32×10^{-1}	0.000676
	CFHR5	$\mathbf{7.18\times10^{-5}}$	-0.05121
	CFD	$\mathbf{4.10\times10^{-6}}$	0.063789
	CFP	$\mathbf{2.51\times10^{-3}}$	0.030998

Note: Data were retrieved from GSE33000, which is the largest individual data set of frontal cortex tissues from patients with Alzheimer's disease and controls [8]. Differential expression *P*-values were calculated by using the *limma* package of R. Log_2FC, log of fold change for target genes in patients compared with controls. Full profiles are available at our newly established webserver AlzData (www.alzdata.org) [6]. *P-values* less than the threshold of Bonferroni correction for 20 genes ($P_{corrected} = 2.50\times10^{-3}$) are marked in bold.

Acknowledgment statement for the ADSP

The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Schellenberg, Farrer, Pericak-Vance, Mayeux, and Haines); U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance and U01 AG052409 to Drs. Seshadri and Fornage. Data generation and harmonization in the Follow-up Phases is supported by U54AG052427 (to Drs. Schellenberg and Wang).

The ADGC cohorts include: Adult Changes in Thought (ACT), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP) and the Washington University Sequencing Project (WUSP), the Columbia University Hispanic- Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA), the University of Toronto (UT), and Genetic Differences (GD).
The CHARGE cohorts are supported in part by National Heart, Lung, and Blood Institute (NHLBI) infrastructure grant HL105756 (Psaty), RC2HL102419 (Boerwinkle) and the neurology working group is supported by the National Institute on Aging (NIA) R01 grant AG033193. The CHARGE cohorts participating in the ADSP include the following: Austrian Stroke Prevention Study (ASPS), ASPS-Family study, and the Prospective Dementia Registry-Austria (ASPS/PRODEM-Aus), the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). ASPS is funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180 and the Medical University of Graz. The ASPS-Fam is funded by the Austrian Science Fund (FWF) project I904),the EU Joint Programme - Neurodegenerative Disease Research (JPND) in frame of the BRIDGET project (Austria, Ministry of Science) and the Medical University of Graz and the Steiermärkische Krankenanstalten Gesellschaft. PRODEM-Austria is supported by the Austrian Research Promotion agency (FFG) (Project No. 827462) and by the Austrian National Bank (Anniversary Fund, project 15435. ARIC research is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Neurocognitive data in ARIC is collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the NIH (NHLBI, NINDS, NIA and NIDCD), and with previous brain MRI examinations funded by R01-HL70825 from the

NHLBI. CHS research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the NHLBI with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629, R01AG15928, and R01AG20098 from the NIA. FHS research is supported by NHLBI contracts N01-HC-25195 and HHSN2682015000001I. This study was also supported by additional grants from the NIA (R01s AG054076, AG049607 and AG033040 and NINDS (R01 NS017950). The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by a joint grant from the Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the municipality of Rotterdam. Genetic data sets are also supported by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project 050-060-810. All studies are grateful to their participants, faculty and staff. The content of these manuscripts is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the U.S. Department of Health and Human Services.

The four LSACs are: the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), The American Genome Center at the Uniformed Services University of the Health Sciences (U01AG057659), and the Washington University Genome Institute (U54HG003079).

Biological samples and associated phenotypic data used in primary data analyses were stored at Study Investigators institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated Phenotypic Data used in primary and secondary data analyses were provided by Study Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania, funded by NIA, and at the Database for Genotypes and Phenotypes (dbGaP) funded by NIH. This research was supported in part by the Intramural Research Program of the National Institutes of health, National Library of Medicine. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by NIA, and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

References

1. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al.; A global reference for human genetic variation. *Nature* 2015; **526**(7571):68-74.
2. Wang D, Fan Y, Malhi M, et al.; Missense variants in HIF1A and LACC1 contribute to leprosy risk in Han Chinese. *Am J Hum Genet* 2018; **102**:794-805.
3. Weiner MW, Aisen PS, Jack CR, Jr., et al.; The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement* 2010; **6**(3):202-211.e7.
4. Purcell S, Neale B, Todd-Brown K, et al.; PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**(3):559-75.
5. Matarin M, Salih DA, Yasvoina M, et al.; A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep* 2015; **10**(4):633-44.
6. Xu M, Zhang DF, Luo R, et al.; A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes as important upstream regulators in Alzheimer's disease. *Alzheimers Dement* 2018; **14**:215-229.
7. Adzhubei IA, Schmidt S, Peshkin L, et al.; A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**(4):248-9.
8. Narayanan M, Huynh JL, Wang K, et al.; Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. *Mol Syst Biol* 2014; **10**:743.