LETTER

doi:10.1038/nature12825

Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease

Carlos Cruchaga[1,2], Celeste M. Karch[1,2]*, Sheng Chih Jin[1]*, Bruno A. Benitez[1], Yefei Cai[1], Rita Guerreiro[3,4], Oscar Harari[1], Joanne Norton[1], John Budde[1], Sarah Bertelsen[1], Amanda T. Jeng[1], Breanna Cooper[1], Tara Skorupa[1], David Carrell[1], Denise Levitch[1], Simon Hsu[1], Jiyoon Choi[1], Mina Ryten[3], UK Brain Expression Consortium (UKBEC)†, Celeste Sassi[3,4], Jose Bras[3], J. Raphael Gibbs[3,4], Dena G. Hernandez[3,4], Michelle K. Lupton[5,6], John Powell[5], Paola Forabosco[7], Perry G. Ridge[8], Christopher D. Corcoran[9,10], JoAnn T. Tschanz[10,11], Maria C. Norton[10,11,12], Ronald G. Munger[12,13], Cameron Schmutz[8], Maegan Leary[8], F. Yesim Demirci[14], Mikhil N. Bamne[14], Xingbin Wang[14], Oscar L. Lopez[15,16], Mary Ganguli[17], Christopher Medway[18], James Turton[18], Jenny Lord[18], Anne Braae[18], Imelda Barber[18], Kristelle Brown[18], The Alzheimer's Research UK (ARUK) Consortium†, Pau Pastor[19,20,21], Oswaldo Lorenzo-Betancor[19], Zoran Brkanac[22], Erick Scott[23], Eric Topol[23], Kevin Morgan[18], Ekaterina Rogaeva[24], Andrew B. Singleton[4], John Hardy[3], M. Ilyas Kamboh[14,15,16], Peter St George-Hyslop[24,25], Nigel Cairns[2,26], John C. Morris[26,27,28], John S. K. Kauwe[8] & Alison M. Goate[1,2,27,28,29]

Genome-wide association studies (GWAS) have identified several risk variants for late-onset Alzheimer's disease (LOAD)[1,2]. These common variants have replicable but small effects on LOAD risk and generally do not have obvious functional effects. Low-frequency coding variants, not detected by GWAS, are predicted to include functional variants with larger effects on risk. To identify low-frequency coding variants with large effects on LOAD risk, we carried out whole-exome sequencing (WES) in 14 large LOAD families and follow-up analyses of the candidate variants in several large LOAD case–control data sets. A rare variant in *PLD3* (phospholipase D3; Val232Met) segregated with disease status in two independent families and doubled risk for Alzheimer's disease in seven independent case–control series with a total of more than 11,000 cases and controls of European descent. Gene-based burden analyses in 4,387 cases and controls of European descent and 302 African American cases and controls, with complete sequence data for *PLD3*, reveal that several variants in this gene increase risk for Alzheimer's disease in both populations. *PLD3* is highly expressed in brain regions that are vulnerable to Alzheimer's disease pathology, including hippocampus and cortex, and is expressed at significantly lower levels in neurons from Alzheimer's disease brains compared to control brains. Overexpression of PLD3 leads to a significant decrease in intracellular amyloid-β precursor protein (APP) and extracellular Aβ42 and Aβ40 (the 42- and 40-residue isoforms of the amyloid-β peptide), and knockdown of PLD3 leads to a significant increase in extracellular Aβ42 and Aβ40. Together, our genetic and functional data indicate that carriers of *PLD3* coding variants have a twofold increased risk for LOAD and that PLD3 influences APP processing. This study provides an example of how densely affected families may help to identify rare variants with large effects on risk for disease or other complex traits.

The identification of pathogenic mutations in *APP*, presenilin 1 (*PSEN1*) and *PSEN2*, and the association of apolipoprotein E (*APOE*) genotype with disease risk led to a better understanding of the pathobiology of Alzheimer's disease, and the development of novel animal models and therapies for this disease[3]. Recent studies using next-generation sequencing have also identified a protective variant in *APP*[4], and a low-frequency variant in *TREM2* associated with Alzheimer's disease risk[5–8] with odds ratio close to that of one *APOE4* allele. These studies have led to the identification of functional variants with large effects on Alzheimer's disease pathogenesis, in contrast to the loci identified through GWAS[1,2]. Low-frequency coding variants not detected by GWAS may be a source of functional variants with a large effect on LOAD risk[5–8]; however, the identification of such variants remains challenging because most study designs require WES in very large data sets. One potential solution is to perform WES or whole-genome-sequencing in a highly selected population at increased risk for disease followed by a combination of genotyping and deep re-sequencing of the variant or gene of interest in large numbers of cases and controls.

We reported previously that families with a clinical history of LOAD in four or more individuals are enriched for genetic risk variants in known Alzheimer's disease and frontotemporal dementia (FTD) genes, but some of these families do not carry pathogenic mutations in the known Alzheimer's disease or FTD genes[9,10], suggesting that additional genes may contribute to LOAD risk. We ranked 868 LOAD families from the National Institute on Aging (NIA)-LOAD study based on number of affected individuals, number of generations affected, the number of

[1]Department of Psychiatry, Washington University, 425 South Euclid Avenue, St. Louis, Missouri 63110, USA. [2]Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University 425 South Euclid Avenue, St. Louis, Missouri 63110, USA. [3]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK. [4]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Building 35 Room 1A1014, 35 Lincoln Drive, Bethesda, Maryland 20892, USA. [5]Institute of Psychiatry, King's College London, 16 De Crespigny Park, London SE5 8AF, UK. [6]Neuroimaging Genetics, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, Queensland 4006, Australia. [7]Istituto di Genetica delle Popolazioni – CNR, Trav. La Crucca, 3 - Reg. Baldinca - 07100 Li Punti, Sassari, Italy. [8]Department of Biology, Brigham Young University, Provo, Utah 84602, USA. [9]Department of Mathematics and Statistics, Utah State University, Logan, Utah 84322, USA. [10]Center for Epidemiologic Studies, Utah State University, Logan, Utah 84322, USA. [11]Department of Psychology, Utah State University, Logan, Utah 84322, USA. [12]Department of Family Consumer and Human Development, Utah State University, Logan, Utah 84322, USA. [13]Department of Nutrition, Dietetics, and Food Sciences, Utah State University, Logan, Utah 84322, USA. [14]Department of Human Genetics, University of Pittsburgh, 130 Desoto Street, Pittsburgh, Pennsylvania 15261, USA. [15]Alzheimer's Disease Research Center, University of Pittsburgh, 130 Desoto Street, Pittsburgh, Pennsylvania 15261, USA. [16]Department of Neurology, University of Pittsburgh, 130 Desoto Street, Pittsburgh, Pennsylvania 15261, USA. [17]Department of Psychiatry, University of Pittsburgh, 130 Desoto Street, Pittsburgh, Pennsylvania 15261, USA. [18]Human Genetics, School of Molecular Medical Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK. [19]Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra, Avenida Pío XII, 55. 31008 Pamplona, Navarra, Spain. [20]Department of Neurology, Clínica Universidad de Navarra, School of Medicine, University of Navarra Avenida Pío XII, 36. 31008 Pamplona, Spain. [21]CIBERNED, Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, Spain. [22]University of Washington, 325 Ninth Avenue, Seattle, Washington 98104-2499, USA. [23]The Scripps Research Institute, La Jolla, California 3344 North Torrey Pines Court, La Jolla, California 92037, USA. [24]Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, 60 Leonard Avenue, Toronto, Ontario M5T 2S8, Canada. [25]Cambridge Institute for Medical Research, and the Department of Clinical Neurosciences, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK. [26]Pathology and Immunology, Washington University, 425 South Euclid Avenue, St. Louis, Missouri 63110, USA. [27]Department of Neurology, Washington University, 425 South Euclid Avenue, St. Louis, Missouri 63110, USA. [28]Knight ADRC, Washington University, 425 South Euclid Avenue, St. Louis, Missouri 63110, USA. [29]Department of Genetics, Washington University, 425 South Euclid Avenue, St. Louis, Missouri 63110, USA.
*These authors contributed equally to this work.
†A list of authors and affiliations appears at the end of the paper.

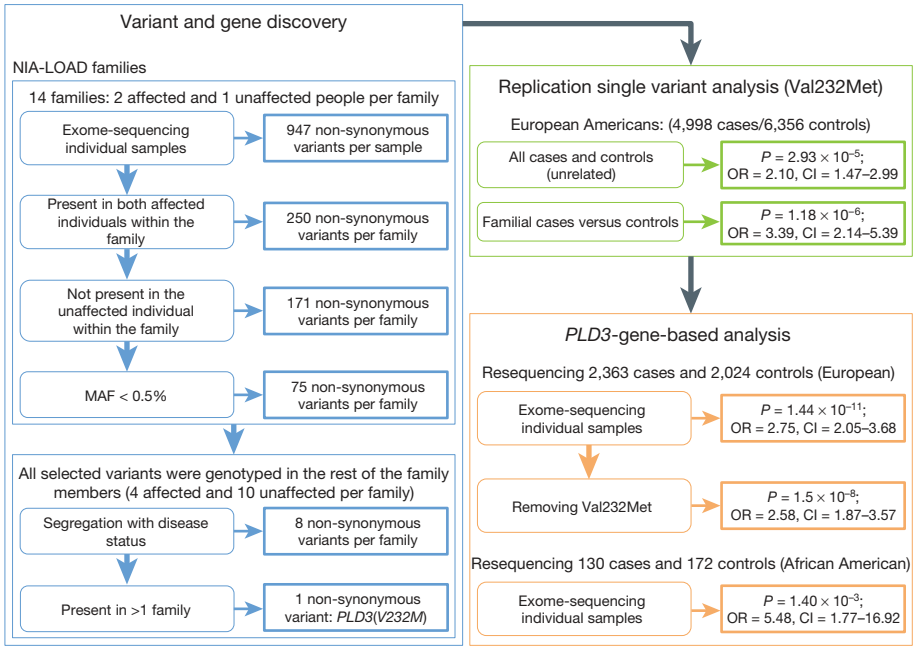


Figure 1 | Summary of the main genetic findings. The diagram shows the steps used to filter the variants identified by exome-sequencing, which led to the identification of the *PLD3*(*V232M*) variant. The diagram also shows the subsequent genetic analyses in large case–control data sets that validated the association of the Val232Met variant and *PLD3* with risk for Alzheimer's disease. CI, confidence interval; OR, odds ratio.

affected and unaffected individuals with DNA available, the number of individuals with a definite or probable diagnosis of Alzheimer's disease, early age at onset (AAO) and *APOE* genotype (discarding families in which *APOE4* segregates with disease status), and 14 were selected to perform WES. In the 14 selected families, there were at least four affected individuals per family, with DNA available for at least three of these individuals. We sequenced at least two affected individuals per family, prioritizing distantly related affected individuals with the earliest AAO. We also sequenced one unaffected individual in nine families and two unaffected individuals in one family. In total, we performed WES on 29 affected individuals and 11 unaffected individuals from 14 families of European American ancestry (Supplementary Table 1 and Supplementary Fig. 2).

All variants shared by affected individuals but absent in unaffected individuals within a family, with a minor allele frequency (MAF) lower than 0.5% in the Exome Variant Server (EVS; http://evs.gs.washington.edu/EVS/) were selected and genotyped in the remaining family members to determine segregation with disease (Supplementary Information). We next examined whether individual variants or variants in the same gene segregated with disease in more than one family. A single variant, rs145999145 (Val232Met, *PLD3,* chromosome 19q13.2), segregated with disease in two independent families (Fig. 1 and Supplementary Fig. 1). We then sought to determine whether this variant was associated with increased risk for sporadic Alzheimer's disease in seven independent data sets (4,998 Alzheimer's disease cases and 6,356 controls of European descent from the Knight Alzheimer's Disease Research Centre (ADRC), NIA-LOAD, NIA-UK data set, Cache-County study, the Universities of Toronto, Nottingham and Pittsburgh, the National Institutes of Mental Health (NIMH) Alzheimer's disease series, and the Wellderly study[7,11–14]; Extended Data Table 1). *PLD3*(*V232M*) was associated with both Alzheimer's disease risk ($P = 2.93 \times 10^{-05}$, odds ratio = 2.10, 95% CI = 1.47–2.99; Table 1) and AAO ($P = 3 \times 10^{-3}$; Extended Data Fig. 1). The frequency of *PLD3*(*V232M*) was higher in Alzheimer's disease cases compared to controls in each age–gender–ethnicity matched data set, with a similar estimated odds ratio for each data set (Extended Data Table 1 and Extended Data Fig. 2), suggesting that the association is unlikely to be a false positive due to population stratification. This was confirmed when population principal components derived from GWAS data were included (Supplementary Information, and Supplementary Figs 2 and 3). The association of the Val232Met variant with Alzheimer's disease risk was also independent of *APOE* genotype (Supplementary Information, Supplementary Table 3 and Supplementary Fig. 4).

LOAD risk variants, such as *APOE4*, are most common in Alzheimer's disease cases with a family history of disease and least common in elderly controls without disease[8,9]. We examined the frequency of Val232Met in three groups of elderly individuals without dementia stratified by age (>65 years, >70 years and >80 years; Table 1) and compared them with

Table 1 | Association between *PLD3*(*V232M*) and Alzheimer's disease risk in individuals of European descent.

Group		Count (carriers/non-carriers)	Frequency (%)	Odds ratio (95% CI)	*P* value
Control group	All controls	50/6,306	0.79	NA	NA
	>65 years, no dementia	9/1,690	0.52	NA	NA
	>70 years, no dementia	5/1,248	0.39	NA	NA
	>80 years, no dementia	1/375	0.26	NA	NA
Cases group	All Alzheimer's disease cases	82/4,916	1.64	*2.10 (1.47–2.99)	2.93×10^{-5}
				†3.13 (1.57–6.24)	3.54×10^{-4}
				‡4.16 (1.68–10.29)	2.34×10^{-4}
	Index cases (families)	29/1,077	2.62	*3.39 (2.14–5.39)	1.18×10^{-6}
				†5.05 (2.38–10.41)	5.14×10^{-6}
				‡6.72 (2.59–17.52)	5.23×10^{-6}
	Sporadic Alzheimer's disease cases	53/3,839	1.36	*1.74 (1.18–2.57)	5.70×10^{-3}
				†2.59 (1.27–5.26)	5.20×10^{-3}
				‡3.44 (1.37–8.63)	3.20×10^{-3}

The table shows the counts for minor allele carriers and non-carriers. *P* values were calculated using Fisher's exact test. Only individuals of European descent were included in this analysis. The carrier frequency for the Val232Met variant in the Exome Variant Server (EVS) is 0.99%. *Odds ratio and *P* value in comparison with all controls. †Odds ratio and *P* value in comparison with individuals aged over 65 years who do not have dementia. ‡Odds ratio and *P* value in comparison with individuals aged over 70 years who do not have dementia. NA, not applicable.

sporadic versus familial Alzheimer's disease cases. As predicted for an Alzheimer's disease risk allele, Val232Met showed age-dependent differences in frequency among controls with the lowest frequency in the Wellderly data set, a series composed of healthy individuals without dementia, who were older than 80 years (carrier frequency 0.27%). Similarly, no Val232Met carriers were found among the 303 individuals without dementia who had normal cerebrospinal fluid Aβ42 and tau profiles, suggesting that the calculated odds ratio for the Val232Met variant when compared to all controls may be an underestimation (Supplementary Information and Supplementary Table 4). As proposed, the frequency of Val232Met was higher in familial cases than in sporadic cases (2.62% in familial versus 1.36% in sporadic cases).

Several risk variants have been observed in *APP*, *PSEN1* and *PSEN2* and *APOE,* supporting the role of these genes in Alzheimer's disease risk[3,4]. To identify additional risk variants in *PLD3*, we sequenced the *PLD3* coding region in 2,363 cases and 2,024 controls of European descent (Extended Data Tables 2 and 3). Fourteen variants were observed more frequently in cases than in controls, including nine variants that were unique to cases (Fig. 2a and Supplementary Information). The gene-based burden analysis resulted in a genome-wide significant association of carriers of *PLD3* coding variants among Alzheimer's disease cases (7.99%) compared to controls (3.06%; $P = 1.44 \times 10^{-11}$; odds ratio = 2.75, 95% CI = 2.05–3.68). When the Val232Met variant was excluded, the association remained highly significant, still passing genome-wide multiple-test correction ($P = 1.58 \times 10^{-8}$; odds ratio = 2.58, 95% CI = 1.87–3.57; Extended Data Table 3), indicating that there are additional variants in *PLD3* that increase risk for Alzheimer's disease independent of Val232Met. There were two additional highly conserved variants (Supplementary Fig. 5), that were nominally associated with LOAD risk: Met6Arg ($P = 0.02$; odds ratio = 7.73, 95% CI = 1.09–61), and Ala442Ala ($P = 3.78 \times 10^{-7}$; odds ratio = 2.12, 95% CI = 1.58–2.83). The Ala442Ala variant showed an association with LOAD risk in four independent series (Extended Data Table 4). This variant was included in the gene-based analysis because our bioinformatic and functional analyses indicate that this variant affects splicing and gene expression (see below).

If the association of *PLD3* with Alzheimer's disease risk is real, it is possible that rare coding variants in *PLD3* in other populations will also increase risk for Alzheimer's disease. We therefore sequenced *PLD3* in 302 African American Alzheimer's disease cases and controls. Both the Val232Met and the Ala442Ala variants were found in Alzheimer's disease cases but not controls, and the Ala442Ala variant showed a significant association with Alzheimer's disease risk ($P = 0.03$). There was also a significant association with LOAD risk at the gene level ($P = 1.4 \times 10^{-3}$; odds ratio = 5.48, 95% CI = 1.77–16.92; Fig. 1, Extended Data Table 5 and Supplementary Information). This consistent evidence of association with Alzheimer's disease risk, at the single-nucleotide polymorphism (SNP) and gene level in two different populations strongly supports *PLD3* as an Alzheimer's disease risk gene.

To begin to understand the link between *PLD3* and Alzheimer's disease, we analysed *PLD3* expression in Alzheimer's disease case and control brains. In human brain tissue from cognitively normal individuals, *PLD3* showed high levels of expression in the frontal, temporal and occipital cortices and hippocampus (Supplementary Fig. 6). Using data from gene expression in laser-captured neurons from Alzheimer's disease cases and controls, *PLD3* gene expression was significantly lower in Alzheimer's disease cases compared to controls ($P = 8.10 \times 10^{-10}$; Fig. 2b). This result was replicated in three additional independent data sets (Supplementary Information and Extended Data Fig. 3). Bioinformatic analyses predicted that the Ala442Ala variant affects alternative splicing (Supplementary Fig. 7 and Supplementary Information). We found that Ala442Ala is associated with lower levels of total *PLD3* messenger RNA (Fig. 2D) and lower levels of transcripts containing exon 11 (Fig. 2c and Supplementary Fig. 8), supporting the functional effect of this variant.

PLD3 is a non-classical, poorly characterized member of the PLD superfamily of phospholipases. PLD1 and PLD2 have been previously implicated in APP trafficking and Alzheimer's disease[15–17]. To determine

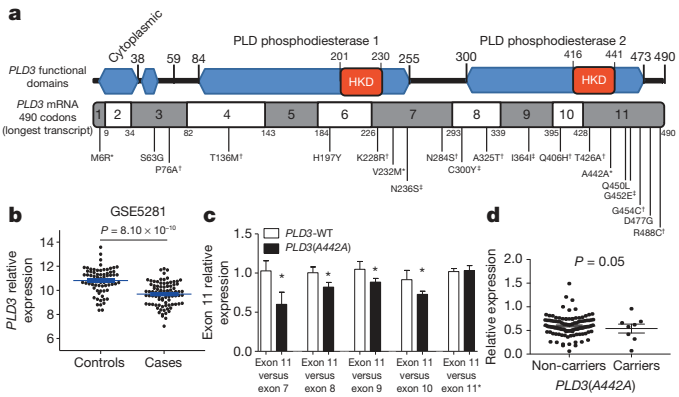


Figure 2 | Most of the *PLD3* coding variants are located in exon 11, and the Ala442Ala variant affects splicing. **a**, Schematic representation of *PLD3* and the relative position of the *PLD3* variants. PLD3 has two PLD phosphodiesterase domains, which contain an HKD signature motif (H-X-K-X(4)-D-X(6)-G-T-X-N, where X represents any amino acid residue). The scheme also shows the exon composition of the longest *PLD3* mRNA and the position of the variants found in this study. *Variants significantly associated with Alzheimer's disease risk. †Variants found only in Alzheimer's disease cases. ‡Variants that are more frequent in Alzheimer's disease cases than in controls. **b**, *PLD3* neuronal gene expression is significantly lower in Alzheimer's disease cases compared to controls. We used the Gene Expression Omnibus data set GSE5281 (ref. 26), in which neurons were laser-captured to analyse whether *PLD3* mRNA expression levels are different between Alzheimer's disease cases and cognitively normal elderly individuals. **c**, **d**, The *PLD3*(*A442A*) variant is associated with lower total *PLD3* mRNA expression and lower levels of exon11 containing transcripts. Primers specific to exons 7, to 11 (two pairs of primers) were designed with PrimerExpress (**c**). cDNA from 8 *PLD3*(*A442A*) carriers and 10 age-, gender-, *APOE*-, clinical dementia rating (CDR)- and post-mortem interval (PMI)-matched individuals were extracted from parietal lobe. Relative expression of exon 11 compared to the other exons was calculated by the ΔCt (changes in cycle threshold) method. Exon-11-containing transcripts were 20% lower in Ala442Ala carriers ($P < 0.05$) in comparison to exon-7–10-containing transcripts. Graphs represent the mean ± s.e.m. Real-time PCR was used to quantify total *PLD3* mRNA and standardized using *GADPH* mRNA as a reference (**d**). *P* value in **d** is for the gene-expression levels of major allele carriers versus minor allele carriers after correcting for dementia severity.

whether PLD3 also affects APP processing, wild-type human PLD3 was overexpressed in mouse neuroblastoma (N2A) cells that stably express wild-type human APP695 (*APP695-WT*; cells termed N2A-695). In this system extracellular Aβ42 and Aβ40 were decreased by 48% and 58%, respectively, compared to the empty vector ($P < 0.0001$; Fig. 3a). Conversely, knockdown of endogenous *PLD3* expression by short hairpin RNA (shRNA) in N2A-695 cells resulted in higher levels of extracellular Aβ42 and Aβ40 than in cells transfected with scrambled shRNA (Fig. 3b). To determine whether the observed effects on APP processing were unique to PLD3 or common among the phospholipase D protein family, we co-expressed APP695-WT with PLD1, PLD2 and PLD3 in human embryonic kidney (HEK293T) cells. Overexpression of PLD3, but not empty vector, PLD1 or PLD2, resulted in a substantial decrease in full-length APP levels (Fig. 3c). Extracellular Aβ42 and Aβ40 levels were significantly reduced in cells overexpressing PLD1, PLD2 and PLD3 compared to control (Fig. 3c). Interestingly, overexpression of catalytically inactive PLD1 and PLD2 variants (*PLD1*(*K898R*) and *PLD2*(*K758R*)) restored extracellular Aβ42 and Aβ40 levels to control values, demonstrating that this is in part a phospholipase-activity-dependent effect (Fig. 3c). Overexpression of a PLD3 dominant-negative variant (*PLD3*(*K418R*)) that inhibits myotube formation[18] failed to restore full-length APP and Aβ42 and Aβ40 to normal levels (Fig. 3c). Furthermore, PLD3 can be co-immunoprecipitated with APP in cultured cells (Extended Data Fig. 4). Together, these studies demonstrate that PLD3 has a role in APP processing that is functionally distinct from PLD1 and PLD2. These findings are consistent with the human

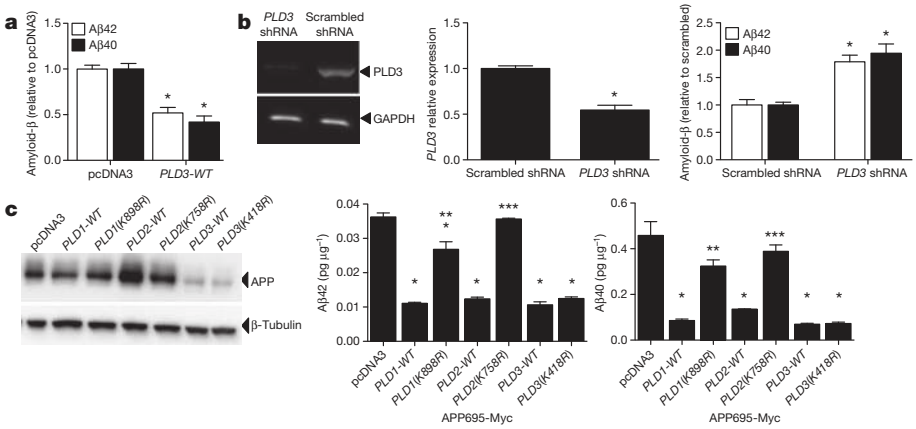


Figure 3 | PLD3 affects APP processing. **a**, **b**, Overexpression and knockdown of PLD3 produce opposing effects on extracellular amyloid-β levels. N2A cells stably expressing human APP695-WT were transiently transfected with vectors containing no insert (*pcDNA3*), human *PLD3-WT*, scrambled shRNA (Origene), or mouse *PLD3* shRNA (Origene) for 48 h. Cell media were analysed with Aβ40 and Aβ42 ELISAs and corrected for total intracellular protein. Amyloid-β levels were then expressed relative to pcDNA3. Graphs represent the mean ± s.e.m. Overexpression of human PLD3 produces significantly less extracellular Aβ42 and Aβ40 (**a**). $*P < 0.0001$. Knockdown of endogenous PLD3 cells produces significantly more extracellular Aβ42 and Aβ40 (**b**). $*P < 0.002$. **c**, Members of the PLD protein family have different effects on APP processing. HEK293T cells were transiently transfected with vectors containing human *APP-WT* and an empty vector (*pcDNA3*), *PLD1, PLD2* or *PLD3-WT*, or *PLD1*, *PLD2*, *PLD3* carrying a dominant-negative mutation. Left panel, PLD3 affects full-length APP levels. Cell lysates were extracted in non-ionic detergent, analysed by SDS–PAGE and immunoblot with antibodies to the Myc-tag on APP (9E10) or β-tubulin. Middle (Aβ42) and right (Aβ40) panels, cell media were analysed with Aβ40 and Aβ42 ELISAs and corrected for total intracellular protein. Graphs represent the mean ± s.e.m. $*P < 0.01$, different from pcDNA3; $**P = 0.002$, different from PLD1-WT; $***P < 0.0001$, different from PLD2-WT. Images are representative of at least three replicate experiments.

genetic and brain expression data presented above; lower PLD3 expression and function is correlated with higher APP and amyloid-β levels and with more extensive Alzheimer's-disease-specific pathology (Supplementary Table 4).

Here we provide extensive genetic evidence that *PLD3* is an Alzheimer's disease risk gene: genome-wide significant evidence that rare variants in *PLD3* increase risk for Alzheimer's disease in multiple data sets and two populations. In addition, our functional studies confirm that PLD3 affects APP processing, in a manner that is consistent with increased risk for Alzheimer's disease[3,19]. This work also provides a second example of a novel gene containing rare variants that influence risk for Alzheimer's disease[5,7,8]. Although these variants have low population attributable fraction (proportion of cases in the population attributable to *PLD3* variants) and diagnostic utility owing to their rarity, they provide important and novel insights into Alzheimer's disease pathogenesis. Our success in identifying multiple families carrying the Val232Met variant and the enrichment of this variant in LOAD families compared to sporadic Alzheimer's disease cases demonstrates the power of using a highly selected sample of multiplex LOAD families for variant discovery. The studies on *TREM2* (refs 5–8), and this report, suggest that next-generation sequencing projects will identify additional low-frequency and rare variants associated with Alzheimer's disease.

METHODS SUMMARY

Participants. Samples were obtained from seven independent data sets totalling 4,998 Alzheimer's disease cases and 6,356 controls of European descent from the Knight ADRC, NIA-LOAD, NIA-UK data set, Cache-County study, the Universities of Toronto, Nottingham and Pittsburgh, the NIMH-AD series, and the Wellderly study[7,11–14].

Exome sequencing. Enrichment of coding exons and flanking intronic regions was carried out using a solution hybrid selection method with the SureSelect human all exon 50-Mb kit (Agilent Technologies) as previously described[20].

SNP genotyping. SNPs were genotyped using the Illumina Golden Gate, Sequenom, KASPar[21,22] and/or Taqman.

***PLD3* sequencing.** *PLD3* was sequenced using a pooled-DNA sequencing design as described previously[9,23,24]. All rare missense or splice site variants were then validated by Sequenom and KASPar genotyping.

Gene-expression and alternative splicing analyses. Total RNA was extracted using the RNeasy mini kit (Qiagen). Complementary DNA was prepared from the total RNA, using the High-Capacity cDNA Archive kit (ABI). Gene-expression levels were analysed by real-time polymerase chain reaction (PCR), using an ABI-7900 real-time PCR system.

Statistical analyses. All of the single SNP analyses were performed using a Fisher's exact test. Allelic association with risk for Alzheimer's disease was tested using 'proc logistic' in SAS, including *APOE* genotype, age, principal component (PC) factors, from population stratification analyses and study as covariates when available. Gene-based analyses were performed using the optimal SNP-set Kernel Association Test (SKAT-O)[25].

Cell-based studies. To assess the effects of PLD3 expression on APP cleavage, vectors containing *PLD3-WT* or *PLD3* shRNA were transiently transfected in mouse N2A cells stably expressing human *APP695-WT*. Aβ40 and Aβ42 were measured in conditioned media by enzyme-linked immunosorbant assay (ELISA) (Invitrogen). *PLD3* silencing was confirmed by quantitative PCR (qPCR). To assess the effects of PLD proteins on APP cleavage, HEK293T cells were transiently transfected with vectors containing *PLD1*, *PLD2* and *PLD3-WT* or dominant-negative mutations. Aβ40 and Aβ42 were measured in conditioned media by ELISA. Full-length APP levels were measured by immunoblot analysis of cell lysates.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 August; accepted 31 October 2013.
Published online 11 December 2013.

1. Bertram, L., McQueen, M., Mullin, K., Blacker, D. & Tanzi, R. The AlzGene Database. Alzheimer Research Forum. http://www.alzgene.org (26 January 2013).
2. Lambert, J. C. *et al.* Extended meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genet.* http://dx.doi.org/10.1038/ng.2802 (27 October 2013).
3. Goate, A. & Hardy, J. Twenty years of Alzheimer's disease-causing mutations. *J. Neurochem.* **120** (Suppl. 1), 3–8 (2012).
4. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488,** 96–99 (2012).
5. Benitez, B. A. *et al. TREM2* is associated with the risk of Alzheimer's disease in Spanish population. *Neurobiol. Aging* **34,** e1715–e31717 (2013).
6. Benitez, B. A. & Cruchaga, C. *TREM2* and neurodegenerative disease. *N. Engl. J. Med.* **369,** 1567–1568 (2013).
7. Guerreiro, R. *et al. TREM2* variants in Alzheimer's disease. *N. Engl. J. Med.* **368,** 117–127 (2013).
8. Jonsson, T. *et al.* Variant of *TREM2* associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* (2013).
9. Cruchaga, C. *et al.* Rare variants in *APP, PSEN1* and *PSEN2* increase risk for AD in late-onset Alzheimer's disease families. *PLoS ONE* **7,** e31039 (2012); correction http://dx.doi.org/10.1371/annotation/c92e16da-7733-421d-b063-1db19488daa6 (2012).

10. Harms, M. *et al. C9orf72* hexanucleotide repeat expansions in clinical alzheimer disease. *JAMA Neurol.* **70,** 736–741 (2013).
11. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's Disease. *Neuron* (2013).
12. Wijsman, E. M. *et al.* Genome-wide association of familial late-onset Alzheimer's disease replicates *BIN1* and *CLU* and nominates *CUGBP2* in interaction with *APOE*. *PLoS Genet.* **7,** e1001308 (2011).
13. Breitner, J. C. *et al.* APOE-ε4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. *Neurology* **53,** 321–331 (1999).
14. Kamboh, M. I. *et al.* Genome-wide association study of Alzheimer's disease. *Transl. Psychiatry* **2,** e117 (2012).
15. Cai, D. *et al.* Phospholipase D1 corrects impaired βAPP trafficking and neurite outgrowth in familial Alzheimer's disease-linked presenilin-1 mutant neurons. *Proc. Natl Acad. Sci. USA* **103,** 1936–1940 (2006).
16. Cai, D. *et al.* Presenilin-1 uses phospholipase D1 as a negative regulator of β-amyloid formation. *Proc. Natl Acad. Sci. USA* **103,** 1941–1946 (2006).
17. Oliveira, T. G. *et al.* Phospholipase d2 ablation ameliorates Alzheimer's disease-linked synaptic dysfunction and cognitive deficits. *J. Neurosci.* **30,** 16419–16428 (2010).
18. Osisami, M., Ali, W. & Frohman, M. A. A role for phospholipase D3 in myotube formation. *PLoS ONE* **7,** e33341 (2012).
19. Hardy, J. & Allsop, D. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends Pharmacol. Sci.* **12,** 383–388 (1991).
20. Benitez, B. A. *et al.* Exome-sequencing confirms DNAJC5 mutations as cause of adult neuronal ceroid-lipofuscinosis. *PLoS ONE* **6,** e26741 (2011).
21. Cruchaga, C. *et al.* Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. *Arch. Neurol.* **68,** 581–586 (2011).
22. Cruchaga, C. *et al.* Association and expression analyses with single-nucleotide polymorphisms in *TOMM40* in Alzheimer disease. *Arch. Neurol.* **68,** 1013–1019 (2011).
23. Jin, S. C. *et al.* Pooled-DNA sequencing identifies novel causative variants in *PSEN1*, *GRN* and *MAPT* in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. *Alzheimer's Res. Ther.* **4,** 34 (2012).
24. Benitez, B. A. *et al.* The *PSEN1*, p.E318G Variant Increases the Risk of Alzheimer's Disease in APOE-ε4 Carriers. *PLoS Genet.* **9,** e1003685 (2013).
25. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89,** 82–93 (2011).
26. Liang, W. S. *et al.* Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc. Natl Acad. Sci. USA* **105,** 4441–4446 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Frohman for providing us with PLD1- and PLD2-WT constructs as well as constructs for the inactive mutations in these genes. This work was supported by grants from the National Institutes of Health (P30-NS069329, R01-AG044546 and R01-AG035083), the Alzheimer Association (NIRG-11-200110) and Barnes Jewish Foundation. This research was conducted while C.C. was a recipient of a New Investigator Award in Alzheimer's Disease from the American Federation for Aging Research. C.C. is a recipient of a BrightFocus Foundation Alzheimer's Disease Research Grant (A2013359S). Sequencing of some of the families included in this study was supported by Genentech and Pfizer. The recruitment and clinical characterization of research participants at Washington University were supported by NIH P50 AG05681, P01 AG03991 and P01 AG026276. This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services, project ZO1 AG000950-11. Samples from the National Cell Repository for Alzheimer's Disease (NCRAD) and NIA-LOAD, which receives government support under a cooperative agreement (U24 AG21886; U24: 5U24AG026395 and 1R01AG041797), were used in this study. We thank our contributors, including the Alzheimer's Disease Centers, that collected samples used in this study, as well as participants and their families, whose help and participation made this work possible. The Cache County Study is supported by National Institutes of Health, RO1-AG11380, RO1-AG18712 and RO1-AG21136. Genotyping and analysis conducted at Brigham Young University was funded by grants from the National Institutes of Health R01-AG042611 and the Alzheimer's Association (MNIRG-11-205368) to J.S.K.K. The sequencing at University of Washington was supported by NIH R01039700 (Z.B.). The sequencing for the NIA-UK samples was supported by the Alzheimer's Research UK (ARUK), by an anonymous donor, by the NINDS (ZO1 AG000950-10), by the Wellcome Trust/MRC Joint Call in Neurodegeneration award (WT089698) to the UK Parkinson's Disease Consortium (UKPDC), by the Big Lottery (to K.M.) and by a fellowship from ARUK to R.G

Some samples and pathological diagnoses were provided by the MRC London Neurodegenerative Diseases Brain Bank and the Manchester Brain Bank from Brains for Dementia Research, jointly funded from ARUK and AS via ABBUK Ltd. This work was also supported by the NIHR Queen Square Dementia BRU and BRC NIHR grant mechanisms. The sample recruitment and genetic studies at University of Pittsburgh are funded by NIH grants AG041718, AG030653, AG005133, AG07562 and AG023652. The Toronto sample studies are funded by Canadian Institutes of Health Research, Wellcome Trust, Medical Research Council, National Institute of Health, National Institute of Health Research, Ontario Research Fund and Alzheimer Society of Ontario (to P.S.G.-H.). The Nottingham Laboratory (K.M.) is funded by ARUK and Big Lottery. ARUK is supported by the UK Medical Research Council through the MRC Sudden Death Brain Bank (C.S.) and by a Project Grant (G0901254) and Training Fellowship (G0802462 to M.R.). P.P. receives funds from the Department of Health of the Government of Navarra, Spain (13085 and 3/2008) and from the UTE project FIMA, Spain. J.T.T. receives funds from the NIA (R01AG21136).

Author Contributions All the authors read and approved the manuscript. C.C. conceived and designed the experiments, supervised research, wrote the manuscript, performed the family and sample selection for exome-sequencing, and analysed the data. C.M.K., S.H., J.C. and A.T.J. performed all the cell-based analysis, and the PLD3 total gene-expression experiments. S.C.J. performed PLD3 pool-sequencing experiments. B.A.B. performed the genotyping of Val232Met and Ala442Ala in the Knight-ADRC and NIA-LOAD data sets, and analysed public gene-expression databases and carried out bioinformatic analysis of the effect of some variants on splicing. O.H., S.B. and Y.C. performed statistical and bioinformatic analyses. J.N. and D.L. recruited and assessed the NIA-LOAD families with the *PLD3* variants. J.B. T.S, D.C. and B.C. performed Sequenom genotyping. R.G., C.S., J.B., M.K.L., J.P., J.R.G., A.S., J.H. P.F., P.G.R., C.D.C., J.T.T., M.C.N., R.G.M., C.S., M.L., J.S.K.K., F.Y.D., M.N.B., X.W., O.L.L., M.G., M.I.K., C.M., J.T., J.L., A.B., I.B., K.B., K.M, O.L.B., P.P., Z.B., E.S., E.T., E.R. and P.S.G.-H., provided genotype data for the NIA-UK and NIMH datasets, Cache-County dataset, University of Pittsburgh dataset, University of Nottingham dataset, NIA-LOAD, the Wellderly dataset and the Toronto dataset. M.R. and D.G.H. performed the co-regulation pathway analysis. N.C. performed the neuropathological examination of the PLD3 Val232Met carriers. J.C.M. supervised recruitment and clinical assessment of the Knight-ADRC subjects, and A.M.G. supervised the functional and genetic experiments and critically reviewed all data and data analysis.

Author Information The authors declare competing financial interests: details are available in the online version of the paper. Exome-sequencing data is available on NIAGADs (https://www.niagads.org, accession number NG00033). Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.C. (ccruchaga@wustl.edu).

The UK Brain Expression Consortium members

John Hardy[3], Mina Ryten[3], Daniah Trabzuni[3], Michael E. Weale[30], Adaikalavan Ramasamy[30] & Colin Smith[31]

[30]Department of Medical and Molecular Genetics, King's College London, 16 De Crespigny Park, London SE5 8AF UK. [31]MRC Sudden Death Brain Bank Project, University of Edinburgh, South Bridge, Edinburgh EH8 9YL UK.

The ARUK Consortium members

Peter Passmore[32], David Craig[32], Janet Johnston[32], Bernadette McGuinness[32], Stephen Todd[32], Reinhard Heun[33], Heike Kölsch[34], Patrick G. Kehoe[35], Nigel M. Hooper[36], Emma R.L.C. Vardy[37], David M. Mann[38], Stuart Pickering-Brown[38], Kristelle Brown[18], Noor Kalsheker[18], James Lowe[18], Kevin Morgan[18], A. David Smith[39], Gordon Wilcock[39], Donald Warden[39] & Clive Holmes[40]

[32]Queen's University Belfast, University Road, Belfast BT7 1NN, UK. [33]Royal Derby Hospital, Uttoxeter Road, Derby, DE22 3NE, UK. [34]University of Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Germany. [35]University of Bristol, Tyndall Avenue, Bristol, City of Bristol BS8 1TH, UK. [36]University of Leeds, Woodhouse Lane, Leeds, West Yorkshire LS2 9JT, UK. [37]University of Newcastle, Newcastle upon Tyne, Tyne and Wear NE1 7RU, UK. [38]University of Manchester, Oxford Road, Manchester, Greater Manchester M13 9PL, UK. [39]University of Oxford (OPTIMA), Wellington Square, Oxford OX1 2JD, UK. [40]University of Southampton, University Road, Southampton SO17 1BJ, UK.

METHODS

Participants and study design. The Institutional Review Board (IRB) at Washington University School of Medicine approved the study. Written informed consent was obtained from participants and their family members by the Clinical Core of the Knight ADRC. The approval number for the Knight ADRC Genetics Core is 93-0006.

Knight-ADRC samples. The Knight-ADRC sample included 1,114 late-onset Alzheimer's disease (LOAD) cases and 913 cognitively normal controls (377 older than 70 years), of European descent, and 302 African American Alzheimer's disease cases and controls, matched for age, gender and ethnicity. These individuals were evaluated by Clinical Core personnel of the Knight ADRC at Washington University. Cases received a clinical diagnosis of Alzheimer's disease dementia in accordance with standard criteria, dementia severity was determined using the Clinical Dementia Rating (CDR)[27].

Cerebrospinal fluid (CSF) levels data set: A subset ($n = 528$) of the Knight-ADRC samples had total tau protein and Aβ42 levels measured in the CSF by ELISA. Of these, 528, 303 did not have dementia (CDR = 0) and were elderly (over 65 years of age), with high CSF Aβ42 levels (>500 pg ml^{-1}). A description of the CSF data set used in this study can be found in another paper[11]. CSF collection and Aβ42, tau and phosphorylated tau181 measurements were performed as described previously[28].

NIA-LOAD. Participants from the National Institute of Ageing Late Onset Alzheimer Disease (NIA-LOAD) Family Study included a single individual with dementia from each of 868 families with at least three Alzheimer's disease-affected individuals, and 881 unrelated control individuals who were elderly and did not have dementia (545 individuals were older than 70 years of age). All Alzheimer's disease cases were diagnosed with dementia of the Alzheimer's type (DAT) using criteria equivalent to the National Institute of Neurological and Communication Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) for probable Alzheimer's disease[29]. NIA-LOAD families were ascertained based on the following criteria: probands (the affected individual through whom the family was recruited into the study) were required to have a diagnosis of definite or probable LOAD (onset after 60 years of age) and a sibling with definite, probable or possible LOAD with a similar age at onset. A third biologically related family member (first, second or third degree) was also required, regardless of affection status. This individual had to be ≥60 years of age if unaffected, or ≥50 years of age if diagnosed with LOAD or mild cognitive impairment[12]. Within each pedigree, we selected a single individual for the case–control series by identifying the youngest affected family member with the most definitive diagnosis (that is, individuals with autopsy confirmation were chosen over those with clinical diagnosis only). Unrelated controls without dementia who were used for the NIA-LOAD case–control series had no family history of Alzheimer's disease and were matched to the cases as previously described[12]. Only individuals of European descent based on the principal component (PC) factors from population stratification analyses were included. Written informed consent was obtained from all participants, and the study was approved by local IRB committees.

Wellderly Study. The Scripps Translational Science Institute's Wellderly study has recruited more than 1,000 healthy elderly participants. Inclusion criteria specify informed consent, age >80 years, blood or saliva donation, compliance with protocol-specified procedures, and no or mild ageing-related medical conditions. Exclusion criteria includes self-reported cancer (excluding basal and squamous cell skin cancer), coronary artery disease or myocardial infarction, stroke or transient ischaemic attack, deep vein thrombosis or pulmonary embolism, chronic renal failure or haemodialysis, Alzheimer's or Parkinson's disease, diabetes, aortic or cerebral aneurysm, or the use of oral chemotherapeutic agents, anti-platelet agents (excluding aspirin), cholinesterase inhibitors for Alzheimer's disease, or insulin. All genotyped individuals were of European descent.

Cache-County study. The Cache-County Study was initiated in 1994 to investigate the association of *APOE* genotype and environmental exposures on cognitive function and dementia. A cohort comprised of 5,092 Cache County, Utah, residents (representing 90% of all individuals in the county who were aged 65 or older) has been followed continually for over 15 years, completing four triennial waves of data collection including clinical assessments[13]. Genotypes were obtained for 255 demented individuals and 2,471 elderly cognitively normal individuals[13]. All individuals genotyped were of European descent.

UK-NIA data set. A description of the UK-NIA data set can be found in another paper[7]. In brief, this data set includes WES from 143 Alzheimer's disease cases and 183 elderly control individuals without dementia. All subjects were of European descent.

University of Pittsburgh data set. The *PLD3*(*V232M*) variant was genotyped in 2,211 subjects including 1,253 Alzheimer's disease cases (62.6% females) and 958 elderly control individuals without dementia(64.3% females). A complete description of the data set can be found in another paper[14]. All individuals were of European descent.

Toronto data set. The Toronto data set was composed of 269 unrelated Alzheimer's disease cases (53% females) and 250 unrelated controls without dementia (56% females) of European descent. The mean (s.d.) age at onset of Alzheimer's disease was 73 (±8) years, and the mean age (s.d.) at last examination of the controls was 73 (±10) years. The study was approved by the IRBs of the University of Toronto.

Exome sequencing. Enrichment of coding exons and flanking intronic regions was performed using a solution hybrid selection method with the SureSelect human all exon 50Mb kit (Agilent Technologies) following the manufacturer's standard protocol. This step was performed by the Genome Technology Access Center at Washington University. The captured DNA was sequenced by paired-end reads on the HiSeq 2000 sequencer (Illumina). Raw sequence reads were aligned to the reference genome hg19 using Novoalign (Novocraft Technologies). Base and SNP calling was performed by SNP Samtools. SNP annotation was carried out using version 5.07 of SeattleSeq Annotation server (see URL)[20].

On average, 95% of the exome had greater than eightfold coverage. SNP calls were made using SAM tools[30]. SNPs identified with a quality score lower than 20 and a depth of coverage lower than 5 were removed. More than 2,500 novel variants in the coding region were found per individual. We identified all variants shared by the affected individuals in a family. Variants not present in 1,000 genome project or the Exome Variant Server (EVS: http://evs.gs.washington.edu/EVS/) or with a frequency lower than 0.5% in the EVS were selected. On average, 80 coding variants were selected for each family. The selected variants were then genotyped in the remaining sampled family members. We validated more than 98% of the selected variants, confirming the high specificity of our exome-sequencing method and analysis. On average, we genotyped a total of 13 family members (7 cases and 6 controls) per family.

SNP genotyping. SNPs were genotyped using the Illumina Golden Gate, Sequenom, Kaspar and/or Taqman genotyping technologies. Only SNPs with a genotyping call rate higher than 98% and in Hardy–Weinberg equilibrium were used in the analyses. The principle of the MassARRAY system is PCR-based, with different size products analysed by SEQUENOM MALDI-TOF mass spectrometry[21,31]. The KBioscience Competitive Allele-Specific PCR (KASP) system is FRET-based endpoint-genotyping technology, v4.0 SNP (KBioscience)[21,31]. Genotype call rates were greater than 98%.

***PLD3* sequencing.** *PLD3* was sequenced in 2,363 cases and 2,027 controls of European origin, and 130 cases and 172 controls of African American descent using a pooled-DNA sequencing design as described previously[9,23,32]. In brief, equimolar amounts of individual DNA samples were pooled together following quantification using the Quant-iT PicoGreen reagent. Pools contained 100 ng of DNA per individual, from 94 individuals. The coding exons and flanking regions (a minimum of 50 bp each side) were individually PCR amplified using specific primers and Pfu Ultra high-fidelity polymerase (Stratagene). An average of 20 diploid genomes (approximately 0.14 ng DNA) per individual were used as input. PCR products were cleaned using QIAquick PCR purification kits, quantified using Quant-iT PicoGreen reagent and ligated in equimolar amounts using T4 Ligase and T4 Polynucleotide Kinase. After ligation, concatenated PCR products were randomly sheared by sonication and prepared for sequencing on an Illumina HighSeq2000 according to the manufacturer's specifications. pCMV6-XL5 amplicon (1,908 base pairs) was included in the reaction as a negative control. As positive controls, ten different constructs (*p53* gene) with synthetically engineered mutations at a relative frequency of one mutated copy per 188 normal copies was amplified and pooled with the PCR products.

Paired-end reads (101 bp) were aligned to the human genome reference assembly build 36.1 (hg19) using SPLINTER[32]. SPLINTER uses the positive control to estimate sensitivity and specificity for variant calling. The wild type: mutant ratio in the positive control is similar to the relative frequency expected for a single mutation in one pool (1 chromosome mutated in 94 samples = 1 in 188 chromosomes). SPLINTER uses the negative control (first 900 bp) to model the errors across the 101-bp Illumina reads and to create an error model from each sequencing run. Based on the error model SPLINTER calculates a *P* value for the probability that a predicted variant is a true positive. A *P* value at which all mutants in the positive controls were identified was defined as the cut-off value for the best sensitivity and specificity. All mutants included as part of the amplified positive control vector were found upon achieving >30-fold coverage at mutated sites (sensitivity = 100%) and only ~80 sites in the 1,908-bp negative control vector were predicted to be polymorphic (specificity = 95%). The variants with a *P* value below this cut-off value were considered for follow-up genotyping confirmation. All rare missense or splice-site variants were then validated by Sequenom and KASPar genotyping in each individual included in the pools. To avoid any batch or plate effects, cases and controls were included in each genotyping plate and all genotyping was performed in a single experiment. Finally, to confirm all of the

heterozygous calls, we created a custom DNA plate including all of the heterozygotes (cases and controls) for all of the variants, and then genotyped them again by Sequenom, creating a new Sequenom set.

Gene-expression and alternative splicing analyses. Total RNA was extracted using the RNeasy mini kit (Qiagen) following the manufacturer's protocol from 82 Alzheimer's disease cases and 39 individuals without dementia. Extracted RNA was treated with DNase1 to remove any potential DNA contamination. cDNAs were prepared from the total RNA, using the High-Capacity cDNA Archive kit (ABI). Gene-expression levels were analysed by real-time PCR, using an ABI-7900 real-time PCR system. The *PLD3*(*A442A*) variant was genotyped in DNA extracted from parietal lobe of 82 Alzheimer's disease cases and 39 individuals without dementia by KASPar as explained below. A total of eight carriers for the Ala442Ala variant were identified.

Total *PLD3* expression: gene expression was analysed by real-time PCR, using an ABI-7500 real-time PCR system. TaqMan assays were used to quantify *PLD3* mRNA levels. Primers and TaqMan probe for the reference gene, *GAPDH,* were designed over exon–exon boundaries, using Primer Express software, v3 (ABI) (sequences available on request). *Cyclophilin A* (ABI: 4326316E) was also used as a reference gene. Each real-time PCR run included within-plate triplicates and each experiment was performed at least twice for each sample.

Alternative splicing: we selected eight Ala442Ala carriers as well as eight CDR-, age-, *APOE*- and PMI-matched individuals to analyse the expression level of exon 11 containing transcripts, the exon in which the Ala442Ala variant is located. Real-time PCR assays were used to quantify *PLD3* exon 7 (forward primer, 5′-GCAGC TCCATCCCATCAACT-3′; reverse, 5′-CTTGGTTGTAGCGGGTGTCA-3′), exon 8 (forward primer, 5′-CTCAACGTGGTGGACAATGC-3′; reverse, 5′-AGTGG GCAGGTAGTTCATGACA-3′), 9 (forward primer, 5′-ACGAGCGTGGCGTCA AG-3′; reverse, 5′-CATGGATGGCTCCGAGTGT-3′), 10 (forward primer, 5′-G GTCCCCGCGGATGA-3′; reverse, 5′-GGTTGACACGGGCATATGG-3′) and 11 (first pair of primers: forward primer, 5′-CCAGCTGGAGGCCATTTTC-3′; reverse, 5′-TGTCAAGGTCATGGCTGTAAGG-3′; second pair forward primer, 5′-GCTGCTGGTGACGCAGAAT-3′; reverse, 5′-AGTCCCAGTCCCTCAGGA AAA-3′). Two pairs of primers were designed for exon 11 as an internal control. SYBR-green primers were designed using Primer Express software, v3 (ABI). Each real-time PCR run included within-plate duplicates and each experiment was performed at least twice for each sample. Real-time data were analysed using the comparative Ct method. Only samples with a standard error of $<0.15\%$ were analysed. The Ct values for exon 11 were normalized with the Ct value for the exons 7–10. The relative exon 11 levels for the Ala442Ala carriers versus the non-carriers were compared using a *t*-test.

***PLD3* gene expression in public databases.** We also used the GEO data sets GSE15222 (ref. 33) and GSE5281 (ref. 26) to analyse the association of *PLD3* gene expression and case-control status. In the GSE15222 data set, there are genotype and expression data from 486 late onset Alzheimer's Disease cases and 279 neuropathologically normal individuals without dementia. In the GSE5281 data set, samples were laser-captured from cortical regions of 16 normal elderly humans (10 males and 4 females) and from 33 Alzheimer's disease cases (15 males and 18 females). Mean age of cases and controls was 80 years. All samples were run on the Affymetrix U133 Plus 2.0 array. RNA data were re-normalized to an average expression of 8 units on a $\log_2$ scale. As potential covariates we analysed the brain region, gender and age for each sample. Stepwise discriminant analysis was used to identify the potential covariates to be included in the analysis of covariance (ANCOVA). For this data set we also extracted the gene-expression levels for *APP* (probe 211277_x_at), *PSEN1* (1559206_at) and *PSEN2* (203460_s_at) to examine the correlation between *PLD3* and *APP, PSEN1* and *PSEN2* using the Pearson correlation method.

Human brain samples and analysis of the Affymetrix Human Exon 1.0 ST array. Quantification and analysis of *PLD3* gene expression in brains was performed as previously described[34]. In brief, the human data used here were provided by the UK Human Brain Expression Consortium[34] and consisted of 101 control post-mortem brains. All samples originated from individuals with no significant neurological history or neuropathological abnormality and were collected by the MRC Edinburgh Brain Bank[35], ensuring a consistent dissection protocol and sample handling procedure. A summary of the available demographic details of these samples including a thorough analysis of their effects on array quality is provided in another paper[36]. All samples were accompanied by fully informed consent for retrieval and were authorized for ethically approved scientific investigation (Research Ethics Committee number 10/H0716/3). Total RNA was isolated from human post-mortem brain tissues using the miRNeasy 96-well kit (Qiagen). The quality of total RNA was evaluated by the 2100 Bioanalyzer (Agilent) and RNA 6000 Nano Kit (Agilent) before processing with the Ambion WT Expression Kit and Affymetrix GeneChip Whole Transcript Sense Target Labelling Assay and hybridization to the Affymetrix Exon 1.0 ST. All arrays were pre-processed using Robust Multi-array Average using Partek Genomics Suite v6.6 (Partek). The resulting expression data were corrected for individual effects (within which are nested post-mortem interval, brain pH, sex, age at death and cause of death) and experimental batch effects (date of hybridization). Transcript-level expression was calculated for 26,993 genes using Winsorized means (Winsorizing the data below 10% and above 90%).

RNA-pathway analysis. To evaluate the biological and functional relevance of co-expressed genes within the *PLD3*-containing modules, we used Weighted Gene Co-expression Network Analysis (WGCNA) and DAVID v6.7 (http://david.abcc.ncifcrf.gov/), the database for annotation, visualization and integrated discovery[37]. We restricted WGCNA to 15,409 transcripts that passed the Detection Above Background (DABG) criteria ($P < 0.001$ in at least 50% of samples in at least one brain region), had a coefficient of variation $>5\%$ and expression values exceeding 5 in all samples in at least one brain region. We followed a step-by-step network construction and module detection. In short, for each brain region, the Pearson correlations between all genes across all relevant samples were derived. We then calculated a signed-weighted co-expression adjacency matrix, allowing us to consider only positive correlations. A power 12, the default soft-threshold parameter for constructing a signed weighted network[38], was used in all brain regions, after checking that this threshold recapitulated scale-free topology[39]. Topological overlap, a more biologically meaningful measure of node interconnectedness (similarity)[9,23] than correlation, was subsequently calculated and genes were hierarchically clustered using 1 − topological overlap as the distance measure. Finally, modules were determined by using a dynamic tree-cutting algorithm. WGCNA led to the identification of several co-expression modules, ranging in number and size between the ten brain regions. We examined the overrepresentation (that is, enrichment) of the three Gene Ontology (GO) categories (biological processes, cellular components and molecular function) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways for each list of co-expressed genes with *PLD3* for each tissue by comparing numbers of significant genes annotated with this biological category with chance.

Statistical analyses. All of the single SNP analyses were performed using a Fisher's exact test, with no covariates included. Allelic association with risk for Alzheimer's disease was tested using 'proc logistic' in SAS including *APOE* genotype, age, PCs and study as covariates when available. Odds ratios with 95% confidence intervals and relative risks were calculated for the alternative allele compared to the most common allele using SAS. Association with age at onset (AAO) was carried out using the Kaplan–Meier method and tested for significant differences, using a proportional hazards model (proc PHREG, SAS) including gender and study as covariates. Controls without dementia were included in the analyses as censored data. The inclusion of these samples did not change the association. Gene-based analyses were performed using the optimal SNP-set (Sequence) Kernel Association Test (SKAT-O)[25].

Population attributable risk. We calculated the Population attributable risk (PAR) using the relative risk obtained in the study and the MAF from the EVS database (http://evs.gs.washington.edu/EVS/) and in the Cache-County data set, which is a population-based data set, using the equation:

$$\mathrm{PAR} = \frac{P_e(\mathrm{RR}_e - 1)}{(1 + P_e(\mathrm{RR}_e - 1))}$$

where P_e is the carrier frequency in the population and RR_e is the relative risk for the different variants.

Neuropathology studies. All study procedures were approved by Washington University's Human Research Protection Office. At autopsy, brain tissue was obtained from participants according to the protocol of the Knight-ADRC. Alzheimer's disease neuropathologic change was assessed according to the criteria of the National Institute on Ageing-Alzheimer's Association (NIA-AA)[40]. Dementia with Lewy bodies was assessed using the criteria given in another paper[41].

Cell-based studies. The following plasmids were used in this study: *pCMV6-XL5* human *PLD3-WT* (Origene), pCS2-Myc human *APP695-WT*[42], pCGN-*PLD-WT*[43] and Lys758Arg[44], pCGN-*PLD2-WT*[45] and Lys898Arg[44], pGFL-GFP[46], pGFP-V-RS-*PLD3*-shRNA-GI548821 (Origene) and pGFP-V-RS-Scr-shRNA-TR30013 (Origene). A dominant-negative mutation (Lys418Arg)[18] was introduced into the pCMV6-XL5 human *PLD3-WT* vector by site-directed mutagenesis using the QuikChangeII Site-Directed Mutagenesis kit (Agilent). All constructs were verified by Sanger sequencing.

Cell-culture assays. Human embryonic kidney (HEK293T) cells were cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine and penicillin/streptomycin (solution containing penicillin and streptomycin). HEK293T cells were grown in 6-well lysine-coated plates. Mouse neuroblastoma (N2A) cells stably expressing human APP695 wild type were cultured in DMEM and Opti-MEM (50:50) supplemented with 5% FBS, 1% L -glutamine, penicillin/streptomycin and 500 μg ml^{-1} G418. After reaching

confluency, cells were transiently transfected with Lipofectamine 2000 (Invitrogen). Culture media were replaced after 24 h, and cells were incubated for another 24 h. Conditioned media were collected, treated with protease inhibitor cocktail and centrifuged at 3000*g* at 4 °C for 10 min to remove cell debris. Cell pellets were extracted on ice in lysis buffer (50 mM Tris, pH 7.6, 2 mM EDTA, 150 mM NaCl, 1% NP40, 0.5% Triton X-100, protease inhibitor cocktail) and centrifuged at 14,000*g*. Protein concentration was measured by the bicinchoninic acid (BCA) method as described by the manufacturer (Pierce-Thermo).

Real-time PCR and quantitative PCR. To confirm effective knockdown of endogenous mouse *PLD3* in mouse N2A-695 cells, RNA was extracted from cell lysates with an RNeasy kit (Qiagen) according to the manufacture's protocol. Extracted RNA (10 μg) was converted to cDNA by PCR using a High-Capacity cDNA Reverse Transcriptase kit (ABI). Gene expression was analysed by quantitative PCR (qPCR) using an ABI-7900 Real-Time PCR system (ABI). Taqman real-time PCR assays were used to quantify expression for mouse *PLD3* (Mm01171272_m1; ABI) and *GAPDH* (Hs02758991_g1; ABI). Samples were run in triplicate. To avoid amplication interference, expression asays were run in separate wells from the housekeeping gene *GAPDH*. Real-time data were analysed by the comparative C_T method. Average C_T values for each sample were normalized to the average C_T values for the housekeeping gene *GAPDH*. The resulting value was corrected for assay efficiency. Samples with a standard error of 20% or less were analysed.

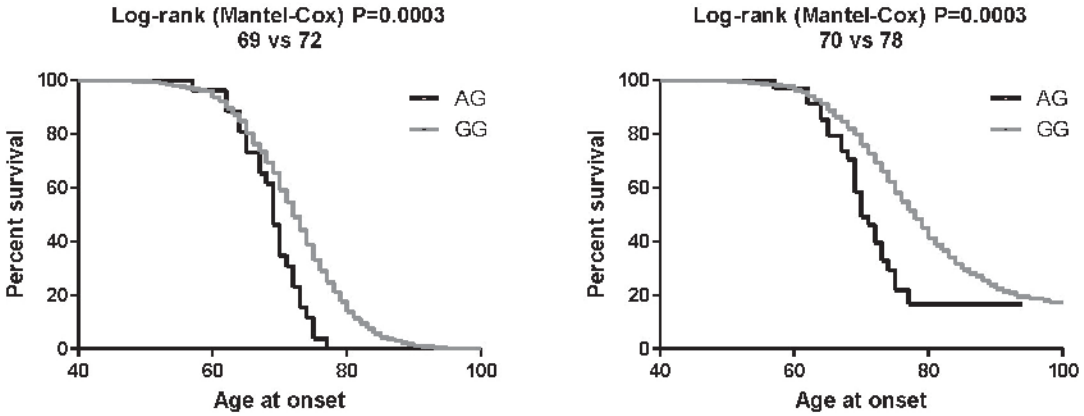
Immunoblot analysis. Standard SDS–PAGE was performed in 4–20% Criterion Tris-HCl gels (Bio-Rad). Samples were boiled for 5 min in Laemmli sample buffer before electrophoresis[47]. Immunoblots were probed with antibodies: PLD3 (Sigma), 9E10 (Sigma) and β-tubulin (Sigma).

Enzyme-linked immunosorbent assay. The levels of Aβ40 and Aβ42 were measured in cell culture media by sandwich ELISA as described by the manufacturer (Invitrogen). ELISA values were obtained (measured in pg ml^{-1}) and corrected for total intracellular protein (measured in μg ml^{-1}) based on BCA measurements.

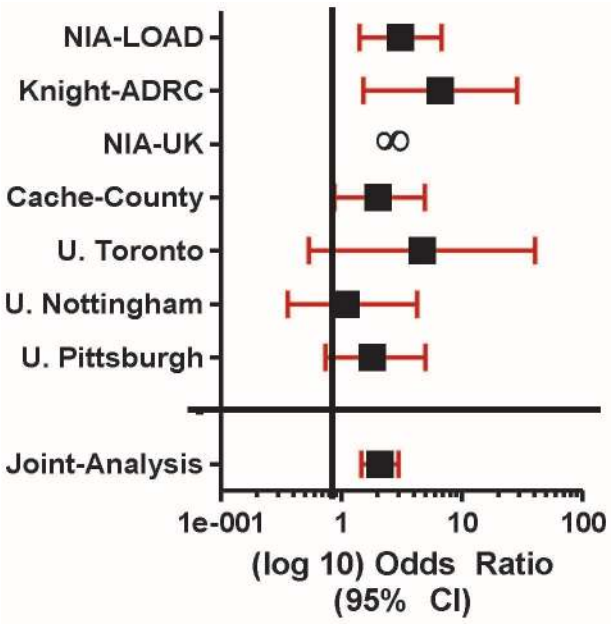
Immunoprecipitation. Cell lysates were incubated with Protein G beads (Thermo Scientific) to remove proteins from the solution that are prone to non-specifically bind to the beads (pre-cleared). Pre-cleared supernatants were incubated overnight at 4 °C with the antibodies indicated. Supernatant–antibody complexes were then incubated with Protein G beads at room temperature for 2 h. After washing, proteins were dissociated from the Protein G beads by incubating the beads in Laemmli sample buffer[47] supplemented with 5% β-mercaptoethanol at 95 °C for 10 min.

Bioinformatics analysis. SIFT (http://sift.jcvi.org/www/SIFT_BLink_submit.html) and Polyphen (http://genetics.bwh.harvard.edu/pph2/) algorithms were used to predict the functional effect of the identified variants. To determine the effect of the Ala442Ala variant on splicing we used the ESEfinder (http://rulai.cshl.edu/tools/ESE). Multiple sequence alignment was performed by ClustalW2, and the PLD3 orthologues were downloaded from Ensembl (http://www.ensembl.org/).

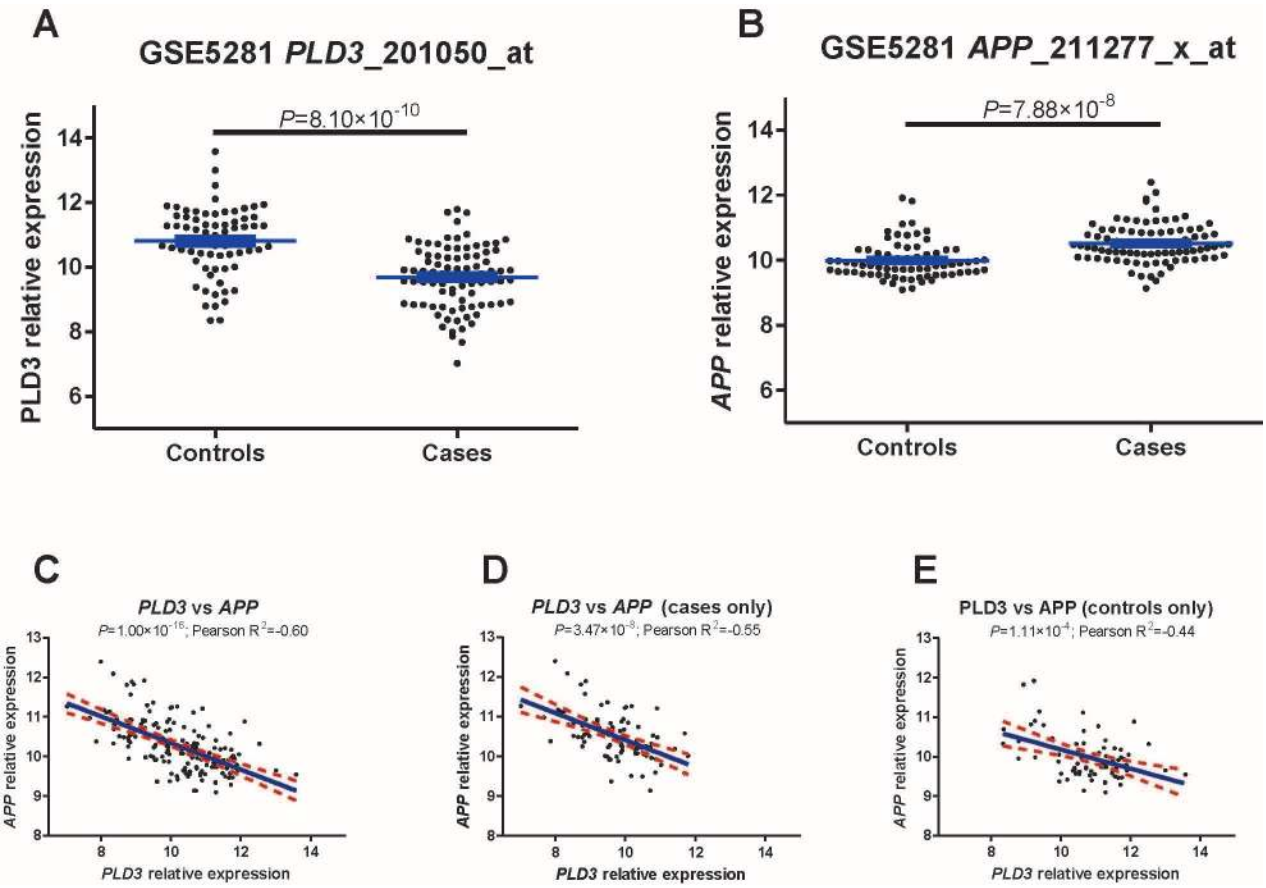
27. Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43,** 2412–2414 (1993).
28. Fagan, A. M. *et al.* Inverse relation between *in vivo* amyloid imaging load and cerebrospinal fluid $A\beta_{42}$ in humans. *Ann. Neurol.* **59,** 512–519 (2006).
29. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34,** 939–944 (1984).
30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
31. Cruchaga, C. *et al.* SNPs associated with cerebrospinal fluid phospho-tau levels influence rate of decline in Alzheimer's disease. *PLoS Genet.* **6,** e1001101 (2010).
32. Vallania, F. L. *et al.* High-throughput discovery of rare insertions and deletions in large cohorts. *Genome Res.* **20,** 1711–1718 (2010).
33. Myers, A. J. *et al.* A survey of genetic human cortical gene expression. *Nature Genet.* **39,** 1494–1499 (2007).
34. Forabosco, P., Ramasamy, A., Hardy, J. & Ryten, M. Insights into TREM2 biology by network analysis of human gene expression data. *Neurobiol. Aging* **34,** 2699–2714 (2013).
35. Millar, T. *et al.* Tissue and organ donation for research in forensic pathology: the MRC Sudden Death Brain and Tissue Bank. *J. Pathol.* **213,** 369–375 (2007).
36. Trabzuni, D. *et al.* Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J. Neurochem.* **119,** 275–282 (2011).
37. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).
38. Mason, M. J., Fan, G., Plath, K., Zhou, Q. & Horvath, S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* **10,** 327 (2009).
39. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statist. Appl. Gen. Mol. Biol.* http:// dx.doi. org/10.2202/1544-6115.1128 (12 August 2005).
40. Montine, T. J. *et al.* National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol.* **123,** 1–11 (2012).
41. McKeith, I. G. *et al.* Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* **65,** 1863–1872 (2005).
42. Schroeter, E. H. *et al.* A presenilin dimer at the core of the γ-secretase enzyme: insights from parallel analysis of Notch 1 and APP proteolysis. *Proc. Natl Acad. Sci. USA* **100,** 13075–13080 (2003).
43. Hammond, S. M. *et al.* Characterization of two alternately spliced forms of phospholipase D1. *J. Biol. Chem.* **272,** 3860–3868 (1997).
44. Sung, T. C. *et al.* Mutagenesis of phospholipase D defines a superfamily including a *trans*-Golgi viral protein required for poxvirus pathogenicity. *EMBO J.* **16,** 4519–4530 (1997).
45. Colley, W. C. *et al.* Phospholipase D2, a distinct phospholipase D isoform with novel regulatory properties that provokes cytoskeletal reorganization. *Curr. Biol.* **7,** 191–201 (1997).
46. Kauwe, J. S. K. *et al.* Fine mapping of genetic variants in BIN1, CLU, CR1 and PICALM for association with cerebrospinal fluid biomarkers for Alzheimer's disease *PLoS One* **6,** e15918 (2011).
47. Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, U. K. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* **252,** 1102–1106 (1977).



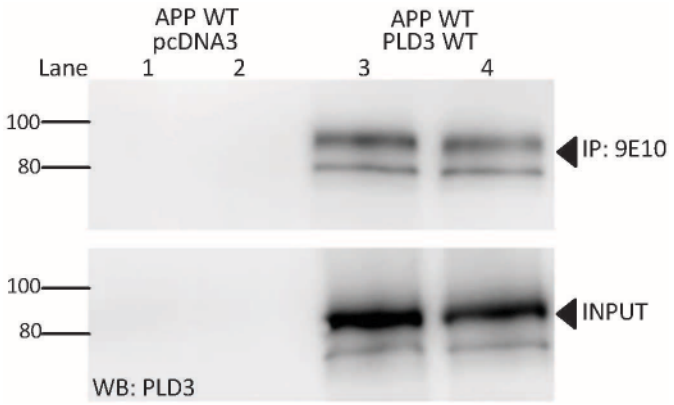
Extended Data Figure 1 | *PLD3*(*V232M*) is associated with age at onset for Alzheimer's disease. **a**, **b**, Age at onset was analysed for association with the *PLD3*(*V232M*) variant in 2,220 cases and 1,841 controls from the Knight-ADRC and NIA-LOAD data sets, by the Kaplan–Meier method. Data were tested for significant differences using the log-rank test. Case-only analysis (**a**); the carriers of the minor allele (AG) have an AAO 3 years lower than the non-carriers (69 versus 73; $P = 3 \times 10^{-3}$). Controls were included as censored data (**b**). The carriers of the minor allele have an AAO 8 years lower than the non-carriers (70 versus 78; $P = 3 \times 10^{-3}$). GG, homozygous for the GG genotype for the *PLD3*(*V232M*) variant.



Extended Data Figure 2 | Forest plot for each case–control series for the Val232Met variant.



Extended Data Figure 3 | *PLD3* and *APP* mRNA expression are inversely correlated. *PLD3* (probe 201050_at) and *APP* (probe 211277_x_at) expression levels were extracted from the GSE5281 data set. *PLD3* mRNA levels are significantly lower in Alzheimer's disease cases compared to controls ($P = 8.10 \times 10^{-10}$), but *APP* is higher in Alzheimer's disease cases ($P = 7.88 \times 10^{-8}$). *PLD3* mRNA levels are inversely correlated with *APP* mRNA expression levels ($P = 1.00 \times 10^{-16}$). The correlation is stronger in Alzheimer's disease cases (Person correlation coefficient = −0.55), than in controls (Person correlation coefficient = −0.44), but in both scenarios the correlation is highly significant.



Extended Data Figure 4 | PLD3 interacts with APP. HEK293T cells were transiently transfected with vectors containing *APP-WT* and an empty vector (pcDNA3) or *PLD3-WT* for 48 h. Cell lysates were extracted in non-ionic detergent, pre-cleared with Protein A beads and immunoprecipitated with an antibody to the Myc-tag on APP (9E10). Immunoblots were probed with an antibody specific to human PLD3. PLD1 and PLD2 reportedly do not inmunoprecipitate with APP[15,16].

Extended Data Table 1 | Association of the *PLD3*(*V232M*) variant in seven independent case–control data sets

Dataset	Cases	Carrier Freq %	Control	Carrier Freq %	OR (95%CI)	p-value
NIA-LOAD	29/1,077	2.62	8/920	0.86	3.09 (1.41-6.81)	4.00×10^{-03}
Knight-ADRC	16/1,098	1.44	2/911	0.22	6.63 (1.52-28.9)	3.40×10^{-03}
NIA-UK	1/142	0.70	0/183	0.00	∞ (NA)	0.438
Cache-County	6/249	2.35	29/2,442	1.17	2.03 (0.83-4.93)	0.131
U. Toronto	5/260	1.89	1/245	0.41	4.71 (0.54-40.7)	0.212
U. Nottingham	6/519	1.14	3/271	1.09	1.05 (0.26-4.25)	1.000
U. Pittsburgh	15/1,253	1.18	6/958	0.62	1.82 (0.74-5.00)*	0.191
NIMH	4/318	1.24	-	-	N/A	
Wellderly	-	-	1/376	0.27	N/A	
Total	**82/4,916**	**1.64**	**50/6,306**	**0.79**	**2.10 (1.47-2.99)**	2.93×10^{-05}

The table shows the counts for carriers and non-carriers. *P* values were calculated by Fisher's exact-test. *For the University of Pittsburgh data set, age, gender, *APOE* genotype and principal component factors for population stratification were available. Association of the Val232Met variant with Alzheimer's disease risk was performed by logistic regression including age, sex, *APOE* genotype and the first four principal component factors as covariates. N/A, not applicable.

Extended Data Table 2 | Sequence variants found in *PLD3* in the NIA-LOAD, Knight-ADRC and NIA-UK data sets

Chr. position	AA		NIA LOAD	Knight ADRC	NIA-UK	total	MAF %	p-value	OR (95% CI)	EVS MAF%	SIFT	Polyphen
40872407	M6R	CA	0	8	1	9	0.19	0.02	7.73	NP	tolerated	deleterious
		CO	0	1	0	1	0.02		(1.09-61)			
40872764	S63G	CA	3	1	0	4	0.08	0.74	0.68	0.16	tolerated	neutral
		CO	5	0	0	5	0.12		(0.18-2.55)			
40872803	P76A	CA	3	1	0	4	0.08	0.12	NA	0.03	tolerated	benign
		CO	0	0	0	0	0.00					
40873764	T136M	CA	0	1	0	1	0.02	0.54	NA	NP	tolerated	deleterious
		CO	0	0	0	0	0.00					
40876055	H197Y	CA	0	1	0	1	0.02	0.49	0.85	NP	damaging	benign
		CO	0	1	0	1	0.02		(0.05-13-7)			
40877584	K228R	CA	1	1	1	3	0.06	0.25	NA	NP	damaging	deleterious
		CO	0	0	0	0	0.00					
40877595	V232M	CA	29	16	1	46	0.99	$1.05x10^{-05}$	3.99	0.48	damaging	deleterious
		CO	8	2	0	10	0.25		(2.01-7.94)			
40877608	N236S	CA	0	2	0	2	0.04	0.40	1.71	0.01	damaging	deleterious
		CO	0	1	0	1	0.02		(0.15-18.91)			
40877752	N284S	CA	0	1	0	1	0.02	0.54	NA	NP	tolerated	deleterious
		CO	0	0	0	0	0.00					
40880407	C300Y	CA	2	3	0	5	0.10	0.46	2.14	0.09	tolerated	deleterious
		CO	1	0	1	2	0.04		(0.41-11.06)			
40880481	A325T	CA	0	1	0	1	0.02	0.54	NA	NP	damaging	deleterious
		CO	0	0	0	0	0.00					
40883725	Q406H	CA	1	0	0	1	0.02	0.54	NA	NP	tolerated	neutral
		CO	0	0	0	0	0.00					
40883783	T426A	CA	1	0	0	1	0.02	0.54	NA	NP	tolerated	neutral
		CO	0	0	0	0	0.00					
40883911	G435V	CA	0	0	0	0	0.00	0.46	NA	0.02	damaging	deleterious
		CO	1	0	0	1	0.02					
40883933	A442A	CA	48	35	12	95	2.09	$1.08x10^{-05}$	2.31	1.59	-	-
		CO	17	12	7	36	0.90		(1.56- 3.41)			
40883956	Q450L	CA	0	0	0	0	0.00	0.46	NA	NP	tolerated	neutral
		CO	0	0	1	1	0.02					
40883962	G452E	CA	4	6	0	10	0.21	0.16	2.86	0.09	tolerated	deleterious
		CO	0	2	1	3	0.07		(0.78-10.4)			
40883967	G454C	CA	0	1	0	1	0.02	0.54	NA	NP	damaging	deleterious
		CO	0	0	0	0	0.00					
40884037	D477G	CA	0	1	0	1	0.02	0.49	0.42	0.02	damaging	deleterious
		CO	0	1	0	1	0.02		(0.04- 4.72)			
40884069	R488C	CA	0	3	0	3	0.06	0.25	NA	0.02	damaging	deleterious
		CO	0	0	0	0	0.00					
total		CA	1106	1114	143	2363						
total		CO	928	913	183	2024						

The coding region of *PLD3* was sequenced in 2,363 Alzheimer's disease cases and 2,024 controls (see Methods) from the Knight-ADRC, NIA-LOAD and the NIA-UK data sets. The table shows the coding variants identified as well as the number of carriers in each data set. The minor allele frequency (MAF) in cases and in controls, the *P* value and the odds ratio (OR) for the association with case-control status is shown. The MAF of the identified variants in the Exome Variant Server (EVS) is shown. We also used SIFT and Polyphen to predict the impact of the non-synonymous changes on protein function. AA, amino acid; CA, cases; CO, controls; NA, not applicable; NP, not present.

Extended Data Table 3 | Gene-based analysis including all coding variants or only variants predicted to be deleterious

	Benign + deleterious		Only deleterious	
	p-value	OR (CI)	p-value	OR (CI)
All variants	1.44×10^{-11}	2.75 (2.05-3.68)	2.52×10^{-12}	2.86 (2.10-3.88)
Excluding V232M	1.58×10^{-8}	2.58 (1.87-3.57)	2.95×10^{-8}	2.54 (1.81-3.57)
Excluding A442 and V232M	1.61×10^{-3}	2.86 (1.62-5.06)	5.88×10^{-5}	3.20 (1.59-6.45)

Gene-based analyses were performed using SKAT-O. Variants that were predicted to be benign by both SIFT and Polyphen were removed for the second analysis.

Extended Data Table 4 | Association analysis for *PLD3*(*A442A*) in four data sets of individuals of European descent

	CA	CO	p-value	OR (95% CI)
NIA-LOAD	48/1058	17/911	1.40×10^{-03}	2.43 (1.38-4.25)
Knight-ADRC	35/1079	12/901	7.10×10^{-05}	2.43 (1.25-4.71)
NIA-UK	12/131	7/176	9.76×10^{-02}	2.30 (0.88- 6.0)
Cache-County	9/246	50/2421	1.15×10^{-01}	1.77 (0.86-3.65)
Total	104/2514	86/4409	3.78×10^{-07}	2.12 (1.58-2.83)

The table shows the counts for carriers and non-carriers. *P* values were calculated using the Fisher's exact test. CA, cases; Co, controls.

Extended Data Table 5 | *PLD3* is associated with risk for Alzheimer's disease in African Americans

	Cases (n=130)		Controls (n=172)			
Variant	carriers	Carrier Freq %	carriers	Carrier Freq %	p-value	OR (95% CI)
G63S	1	0.77%	0	0.00%	0.43	NA
K228R	1	0.77%	0	0.00%	0.43	NA
V232M	3	2.31%	0	0.00%	**0.07**	NA
I364I	6	4.62%	4	2.33%	0.33	2.02 (0.56-7.29)
A442A	4	3.08%	0	0.00%	**0.03**	NA
Total	15	11.54%	4	2.33%	1.4×10^{-03}	5.48 (1.77-16.92)

PLD3 was sequenced in a total of 302 African Americans. The table shows the counts for single SNPs and the gene-based analysis for *PLD3* in 130 African American cases and 172 controls. *P* values were calculated using the Fisher's exact test. NA, not applicable.