Human Molecular Genetics, 2006, Vol. 15, No. 17 **2560–2568**
doi:10.1093/hmg/ddl178
Advance Access published on July 17, 2006

DAPK1 variants are associated with Alzheimer's disease and allele-specific expression

Yonghong Li[1,*], Andrew Grupe[1], Charles Rowland[1], Petra Nowotny[2], John S.K. Kauwe[2], Scott Smemo[2], Anthony Hinrichs[2], Kristina Tacey[1], Timothy A. Toombs[1], Shirley Kwok[1], Joseph Catanese[1], Thomas J. White[1], Taylor J. Maxwell[2], Paul Hollingworth[3], Richard Abraham[3], David C. Rubinsztein[4], Carol Brayne[5], Fabienne Wavrant-De Vrièze[6], John Hardy[6], Michael O'Donovan[3], Simon Lovestone[7], John C. Morris[2], Leon J. Thal[8], Michael Owen[3], Julie Williams[3] and Alison Goate[2]

[1]Celera Diagnostics, 1401 Harbor Bay Parkway, Alameda, CA 94502, USA, [2]Departments of Psychiatry, Neurology, Genetics and Biology, Washington University, St Louis, MO 63110, USA, [3]Department of Psychological Medicine and Biostatistics and Bioinformatics Unit, Cardiff University, Wales College of Medicine, Cardiff CF14 4XN, UK, [4]Department of Medical Genetics, Cambridge Institute for Medical Research, Cambridge CB2 2XY, UK, [5]Department of Public Health and Primary Care, Institute of Public Health, Cambridge CB2 2SR, UK, [6]National Institute on Aging, Bethesda, MD 20892, USA, [7]Departments of Old Age Psychiatry and Neuroscience, MRC Centre for Neurodegeneration Research, King's College London, Institute of Psychiatry, London SE5 8AF, UK and [8]Department of Neurosciences, University of California, San Diego, CA 92093, USA

Received May 2, 2006; Revised July 7, 2006; Accepted July 11, 2006

Genetic factors play an important role in the etiology of late-onset Alzheimer's disease (LOAD). We tested gene-centric single nucleotide polymorphisms (SNPs) on chromosome 9 and identified two SNPs in the death-associated protein kinase, *DAPK1*, that show significant association with LOAD. SNP rs4878104 was significantly associated with LOAD in our discovery case–control sample set (WU) and replicated in each of two initial validation case–control sample sets ($P < 0.05$, UK1, SD). The risk-allele frequency of this SNP showed a similar direction in three other case–control sample sets. A meta-analysis of the six sample sets combined, totaling 2012 cases and 2336 controls, showed an allelic *P*-value of 0.0016 and an odds ratio (OR) of 0.87 (95%CI: 0.79–0.95). Minor allele homozygotes had a consistently lower risk than major allele homozygotes in the discovery and initial two replication sample sets, which remained significant in the meta-analysis of all six sample sets (OR = 0.7, 95%CI: 0.58–0.85), whereas the risk for heterozygous subjects was not significantly different from that of major allele homozygotes. A second SNP, rs4877365, which is in high linkage disequilibrium with rs4878104 ($r^2 = 0.64$), was also significantly associated with LOAD (meta $P = 0.0017$ in the initial three sample sets). Furthermore, *DAPK1* transcripts show differential allelic gene expression, and both rs4878104 and rs4877365 were significantly associated with *DAPK1* allele-specific expression ($P = 0.015$ to <0.0001). These data suggest that genetic variation in *DAPK1* modulates susceptibility to LOAD.

INTRODUCTION

Late-onset Alzheimer's disease (LOAD) is the most common type of dementia. The prevalence of AD is increasing because of a longer lifespan. In addition to age, genetic variation is thought to play a major role in the etiology of LOAD, and is likely to be modulated through complex gene–gene and gene–environment co-actions and interactions. Variation in the apolipoprotein E (*APOE*) gene contributes to both risk and age at disease onset of LOAD (1,2). Although whole

*To whom correspondence should be addressed. Tel: +1 5107496283; Fax: +1 5107496200; Email: yonghong.li@celeradiagnostics.com

Table 1. Sample set characteristics

Sample set	Sample size (cases–controls)	Country of origin	AAO (cases)	AAE (controls)	ApoE4 allele frequency (%) (cases–controls)	Female (cases–controls)
WU	419/377	USA	76.2 ± 6.8	77.5 ± 7.5	33.7/12.3	265/237
SD	201/378	USA	73.3 ± 5.4	79.0 ± 7.3	38.4/12.4	99/240
UK1	360/396	UK	76.7 ± 6.3	76.6 ± 6.2	36.4/13.1	282/295
UK2	286/348	UK	77.4 ± 6.0	76.7 ± 5.6	35.3/11.8	208/198
UK3	473/607	UK	75.1 ± 6.5	76.1 ± 6.4	37.2/14.6	300/354
Linkage	351/335	USA	73.5 ± 6.4	74.6 ± 8.1	36.8/13.7	259/190

AAO, age at disease onset for cases in average ± SD; AAE, age at exam for controls in average ± SD.

genome linkage scans of LOAD have implicated candidate chromosomal regions (3), other susceptibility genes have not been firmly established.

Chromosome 9 is one of the several prominent chromosomes where one or more linkage peaks have been observed by multiple whole genome scans (4–8). A number of genes on this chromosome, including *UBQLN1* and *ABCA1*, have been reported to be associated with LOAD (9–11), but these findings remain controversial (12–16). To identify genetic variants associated with LOAD on chromosome 9, we scanned single nucleotide polymorphisms (SNPs) across the entire chromosome using DNA samples collected from LOAD patients and similar non-demented individuals. The scan was not based on linkage disequilibrium (LD) of markers across the whole chromosome but rather a targeted, gene-centric approach; SNPs were selected from dbSNP and the Celera human genome SNP database. Putative functional SNPs [non-synonymous, UTR, predicted transcription factor binding site (TFBS) and exon splice silencer SNPs] were selected first, followed by other types of gene-centric SNPs. To reduce the likelihood of identifying spurious associations, we employed up to six independently collected LOAD case–control sample sets of Caucasian descent (Table 1). Genotyping and analysis were performed in several phases including a hypothesis-generating discovery phase in one sample set, followed by a hypothesis-validation phase in two other independent sample sets. This led us to identify two SNPs, rs4878104 and rs4877365, in the death-associated protein kinase 1 (*DAPK1*) that are significantly associated with LOAD. SNP rs4878104 was then evaluated in three other independently collected LOAD case–control sample sets (UK2, UK3, Linkage). We also assessed the potential biological consequences of the LOAD-associated SNPs and found that these genetic variants may directly or indirectly modulate allele-specific expression of *DAPK1*.

RESULTS

DAPK1 variants are associated with LOAD

In the discovery phase, we genotyped 674 SNPs in 347 genes in one sample set (Fig. 1A and Supplementary Material, Fig. S1); 642 SNPs were genotyped first in the WU sample set and 32 SNPs first in the UK1 sample set. Forty-seven of these SNPs met a nominally significant threshold ($P < 0.05$) for association with LOAD based upon a χ^2 test for allelic association (Fig. 1A). In the validation phase, we genotyped these 47 markers in two other sample sets. Three markers were replicated in the combined validation sample sets (allelic $P < 0.05$, one-sided; Table 2, Supplementary Material, Tables S1 and S2), including rs4878104, which was significantly associated with LOAD in each of the three sample sets (Table 2). A meta-analysis of all three sample sets showed an allelic *P*-value of 0.0006 and an odds ratio (OR) of 0.79 (95%CI: 0.69–0.90), suggesting that the minor allele is protective. However, the genotypic analysis suggests that while homozygotes for the minor allele are at reduced risk (OR = 0.55, CI: 0.4–0.74), heterozygous subjects are not at significantly different risk from subjects homozygous for the major allele (Fig. 1B). The most significant associated marker, rs4878104, is located in intron 2 of the *DAPK1* gene, which maps to 9q22 and is located directly under a previously reported LOAD linkage peak (7,10). In addition, two other markers from the chromosome 9 scan were replicated in the combined validation sample set and remained significant in the meta-analysis of all three sample sets (Supplementary Material, Tables S1 and S2). One marker, rs2018621, in protein-*O*-mannosyltransferase 1 (*POMT1*), was significant in one of the validation sample sets. The allele frequency of the marker is relatively low, with 3.1% in controls and 5.3% in cases. Another marker, rs2274159, in *DFNB31* was not significant in either of the two validation sample sets, individually, but reached significance when the two were combined. Defects in *DFNB31* cause hereditary non-syndromic recessive hearing loss. The *POMT1* and *DFNB31* markers are 44.2 and 27.0 Mbp distal from the *DAPK1* marker, respectively, and share little LD (data not shown).

Because of the stronger association of the *DAPK1* variant with LOAD, and since *DAPK1* is an excellent biological candidate gene (see discussion section) within a previously identified linkage region, we focused our further study on *DAPK1* by genotyping the most significant marker in other sample sets and fine-mapping the *DAPK1* region. We examined rs4878104 in three other LOAD sample sets, including two case–control sample sets from the UK and one US sample set generated by selecting one case per family from our genetic linkage sample and similar controls without dementia collected in the St Louis area. In these sample sets, we did not find significant evidence for association with LOAD (Table 2), however, the directionality of the risk allele was the same as in the initial three

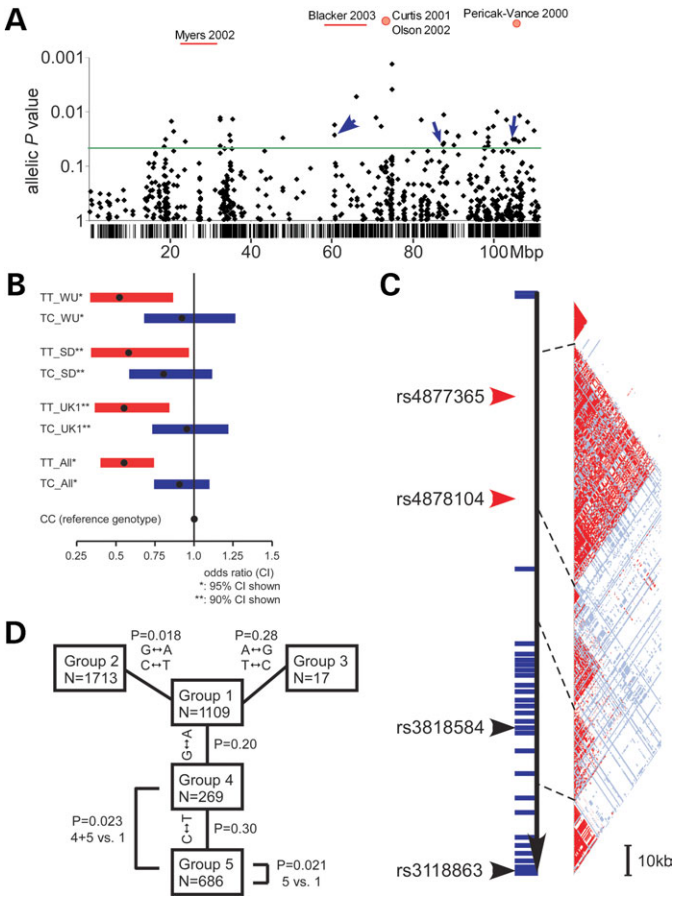

Figure 1. (**A**) Association with LOAD for exploratory markers on chromosome 9. *P*-values of 674 SNPs, their physical positions (Celera assembly R27) and the distribution of annotated genes (bar graph) are shown. A separate line denotes $P = 0.05$. Linkage peak regions are noted along with references. The arrow points to rs4878104 in *DAPK1*, rs2274159 in *DFNB31* and rs2018621 in *POMT1*, respectively (from left to right). (**B**) rs4878104 genotype and LOAD risk. Odds ratios and confidence intervals of minor allele homozygous (TT) and heterozygous (TC) subjects in each individual sample set and the three sample sets combined (All). TT_WU: 0.53 (0.33–0.86); TC_WU: 0.92 (0.68–1.26); TT_SD: 0.57 (0.34–0.96); TC_SD: 0.80 (0.58–1.11); TT_UK1: 0.55 (0.36–0.84); TC_UK1: 0.94 (0.73–1.22); TT_All: 0.55 (0.40–0.74); TC_All: 0.90 (0.74–1.09). (**C**) *DAPK1* gene structure, HapMap LD diagram and positions of the disease-associated and expression markers. (**D**) Haplotype analysis based upon 16 SNPs in a single haplotype block containing the associated SNPs in *DAPK1*. Each rectangle represents a haplotype group. Mutations at rs4877365 are labeled as 'G ↔ A' and 'C ↔ T' for rs4878104. The haplotypes have been grouped to illustrate the key transitions involving rs4877365 and rs4878104. *P*-values were calculated by Fisher's exact test from cases–controls in the adjacent haplotype groups. N = number of chromosomes. The specific haplotypes represented in each group will be provided upon request to the authors.

sample sets. A meta-analysis of rs4878104 with all five replication sample sets remained significant ($P = 0.014$) and the meta-analysis of all six samples showed a *P*-value of 0.0016. rs4878104 did not show a significant interaction with APOE e4 presence or absence in any of the six sample sets, individually or together.

We examined the LD structure of the region containing rs4878104 in HapMap (HapMap public release #16c.1; http://www.hapmap.org). The marker, rs4878104, is within a 58.5 kb region of high LD from rs913778 to rs888333, which encompasses part of intron 2 of the *DAPK1* gene and a recently predicted gene *LOC643284*, which is homologous to 40S ribosomal protein S29 (Fig. 1C). To fine map this region, we identified 14 tagging SNPs with $r^2 < 0.8$. One of the tagging SNPs was rs4878104 itself and another appeared to have a SNP nearby that prevented us from developing an assay. We therefore genotyped the remaining 12 tagging SNPs in the WU sample set (Supplementary Material, Table S3). One tagging marker, rs4877365, showing a r^2 of 0.68 (in cases + controls) with rs4878104, was also significantly associated with LOAD ($P = 0.020$) in the WU sample set and replicated in the combined initial two validation sample sets. This marker was significant in the SD sample set and trended to significance in the UK1 sample set (Table 2). Similar to rs4878104, it showed a stronger effect under a recessive model. The meta-analysis of all the three sample sets showed a significant association for this marker ($P = 0.0017$, allelic). Two other SNPs in this LD block showed $r^2 > 0.5$ with rs4878104; one of these SNPs, rs7036598, showed significant association in the exploratory sample set, but was not replicated in either of the initial replication sample sets.

We performed a permutation-based haplotype analysis using the program PHASE. Six SNPs were identified using Tag 'n' Tell (http://snp.cgb.ki.se/tagntell/) (rs913778, rs871495, rs7036598, rs3128519, rs3128521, rs4878104), which tag the haplotypes present at a frequency >1% in our combined samples. This analysis failed to detect significant association with disease ($P = 0.35$). Placing haplotypes into their evolutionary context may increase biological information (17). To take advantage of this increase in information, we used tree scanning, a method, which incorporates the evolutionary history of the haplotypes, to analyze our data. Tree scanning in this region failed to detect significant association with LOAD. Inspection of the network suggests that at least 13 of the 16 SNPs used in this analysis (including two SNPs which were significant in the single SNP analyses, rs4878104 and rs4877365) are subject to homoplasy (mutations at the same site are observed in independent parts of the network i.e. the mutation has occurred multiple times in evolutionary history).

A causal SNP that has experienced homoplasy should be significant when tested alone (all haplotypes grouped by state of the SNP). It should also show evidence for effects at each individual branch marked by the SNP if the comparison is made in the immediate area around the branch, provided the sample size is sufficient. In contrast, if the SNP is in LD with a second causal SNP, then the individual branches will only exhibit association if the second causal allele is also present on the haplotype branch. This hypothesis was tested with an alternative approach described in the methods. Markers rs4877365 and rs4878104 were significant in the single SNP analyses; both are inferred to have experienced homoplasy and occur at the same three regions of the tree (Fig. 1D). Using these regions, the tree can be collapsed into five haplotype groups. Each of the three groups with the GC haplotype (groups 2, 3, 5) were individually tested against group 1, which has the AT haplotype and is internal to the others. Two of the three comparisons to group 1 were nominally

Table 2. Allelic association of the *DAPK1* SNPs rs4878104 and rs4877365 with LOAD

Marker	Sample	Case[a]					Control[a]					Allelic P	Allelic OR	Power
		11[e]	12[e]	22[e]	SUM	MAF	11[e]	12[e]	22[e]	SUM	MAF			
rs4878104	WU[b]	36	181	170	387	32.7	55	159	138	352	38.2	0.027	0.79 (0.63: 0.97)	
	SD[c]	17	84	79	180	32.8	46	161	121	328	38.6	0.033	0.78 (0.62: 0.97)	0.54
	UK1[c]	30	169	148	347	33.0	55	181	149	385	37.8	0.028	0.81 (0.68: 0.97)	0.70
	Linkage[c]	52	161	138	351	37.7	55	152	128	335	39.1	0.30	0.94 (0.79: 1.13)	0.68
	UK2[c]	39	123	124	286	35.1	53	159	136	348	38.1	0.14	0.88 (0.73: 1.07)	0.65
	UK3[c]	51	216	194	461	34.5	81	255	252	588	35.5	0.32	0.96 (0.82: 1.12)	0.83
	SD + UK1 + Linkage + UK2 + UK3[c]	189	753	683	1625	34.8	290	908	786	1984	37.5	0.014	0.89 (0.82: 0.96)	0.99
	WU + SD + UK1 + Linkage + UK2 + UK3[d]	225	934	853	2012	34.4	345	1067	924	2336	37.6	0.0016	0.87 (0.79: 0.95)	
rs4877365	WU[b]	24	148	203	375	26.1	41	141	170	352	31.7	0.020	0.76 (0.61: 0.96)	
	SD[c]	11	80	89	180	28.3	35	151	144	330	33.5	0.045	0.79 (0.62: 0.99)	0.59
	UK1[c]	18	144	186	348	25.9	30	165	191	386	29.2	0.080	0.85 (0.70: 1.03)	0.76
	SD + UK1[c]	29	224	275	528	26.7	65	316	335	716	31.2	0.015	0.82 (0.71: 0.95)	0.91
	WU + SD + UK1[d]	53	372	478	903	26.5	106	457	505	1068	31.3	0.0017	0.80 (0.69: 0.92)	

[a]Counts of genotype 11, 12 and 22 and minor allele frequency (MAF).
[b]Exploratory sample set, two-sided *P*-value and OR (95%CI) are shown.
[c]Replication sample set, one-sided *P*-value and OR (90%CI) are shown.
[d]Two-sided *P*-value, Cochran Mantel and Haenzsel test, using sample set as the stratifying variable.
[e]11, 12 and 22 denote TT, TC and CC, respectively, for rs4878104 and 11, 12 and 22 denote AA, AG and GG, respectively, for rs4877365.

significant (group 2 and group 5 or groups 4 and 5; Fig. 1D) with a highly significant Fisher's combined probabilities test ($P = 0.0054$ for groups 2, 3 and 5; $P = 0.0058$ for groups 2, 3 and 4 + 5). The non-significant *P*-value of group 3 compared with group 1 is probably because of the small sample size ($N = 17$). Group 4 only has a change at rs4877365, and was not significantly different from group 5 or group 1, but exhibited an intermediate ratio of cases to controls.

SNPs associated with LOAD are also associated with *DAPK1* allele-specific expression

Because of the intronic nature of the significant SNPs, we sought to determine whether the disease-associated variants have a direct effect (*cis*) on *DAPK1* transcript levels. This information can be obtained directly by measuring the relative expression level of the allele 1 specific transcript and the allele 2 specific transcript. Because this method measures both allele-specific transcripts in the same sample, the measurement is much less impacted by biological variation that occurs when comparing expression levels between different samples. In addition, this method does not require normalization to a house-keeping gene because the transcript level in each sample should be identical for allele 1 and allele 2, unless there is a *cis*-controlling factor that is present on only one allele/haplotype and changes its relative expression level (e.g. a mutation in a repressor element on allele 1/haplotype 1 leads to an expression ratio that is higher for allele 1 than allele 2). Therefore, allele-specific expression analysis is much more sensitive to small relative changes in expression levels, whereas across sample comparisons include much more noise (biological variation between samples and technical noise from normalizing to a house-keeping gene) and can only detect larger expression differences. We thus tested *DAPK1* for allelic expression differences and evaluated whether rs4878104 or rs4877365 is associated with the allelic expression differences. The disease-associated markers could not be used directly for this analysis because they are located within a *DAPK1* intron; therefore, we used a two-step approach. First, we genotyped two high-frequency SNPs that map to the *DAPK1* transcript and identified 69 Caucasians, who were heterozygous for at least one of these expression markers (46 for rs3118863 and 48 for rs3818584; minor allele frequency: 0.44 for rs3118863 and 0.47 for rs3818584). The differences in the two allele-specific transcripts were as large as ~2-fold.

To determine whether the two disease-associated SNPs, rs4878104 and rs4877365, are associated with the observed *DAPK1* allele-specific expression, we next genotyped these two SNPs in all 69 heterozygous carriers of rs3118863 and/or rs3818584. The two expression SNPs are 72 and 129 kb away from rs4878104, the closer of the two disease-associated markers, and do not reside in the high LD region shared by the two disease-associated markers (Fig. 1C and Supplementary Material, Table S4). If either rs4878104 or rs4877365 is the sole causative element, it would be expected that only heterozygous carriers of the disease-associated variant show an allele-specific expression difference, whereas homozygous carriers would not (Fig. 2A). As shown in Figure 2B, both homozygous and heterozygous carriers of rs4878104 and rs4877365 showed allele-specific differences in expression. However, the allele-specific expression ratio was significantly higher in heterozygous carriers of either rs4878104 or rs4877365 when compared with homozygous carriers ($P < 0.05$, Fig. 2B), whereas the genomic DNA control showed no such genotype-dependent difference (data not shown). These results indicate that the genotype status of both LOAD-associated SNPs is significantly associated with *DAPK1* allele-specific expression.

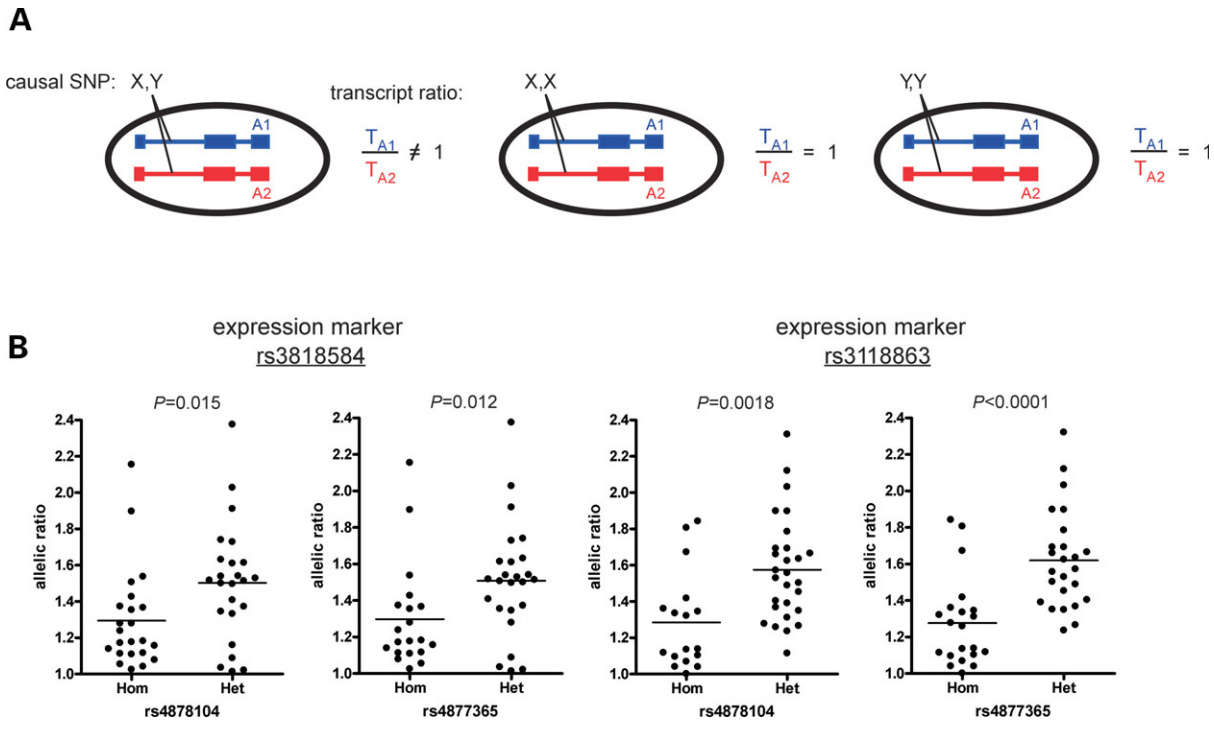

Figure 2. *DAPK1* allele-specific gene expression stratified by the LOAD-associated markers rs4878104 or rs4877365. (**A**) Allele-specific transcript ratios that are different from one require heterozygosity of the causal variant (X,Y). Markers that show association with allele-specific expression, when stratified by genotype, are expected to be in strong LD with the causal variant or represent the causal variant itself. Markers that are not in LD with the causal variant are unlikely to show association with allele-specific expression. A1, A2: allele-specific expression marker. Note that the alleles (X/Y) of the causal SNP can occur on either allele-specific transcript, if the causal SNP is not in perfect LD with the expression marker (A1/A2). (**B**) The ratio of allele-specific gene expression is shown for cDNA. Two markers, rs3818584 and rs3118863, were analyzed to measure allelic gene expression (Hom: homozygotes; Het, heterozygotes). A Mann–Whitney test was performed to assess the association. The relative expression was calculated as $2^{\Delta C_t}$ (ΔC_t was determined by subtracting the smaller C_t value of one allele PCR reaction from the larger C_t value of the other allele PCR reaction).

DISCUSSION

Our chromosome 9 scan resulted in the identification of two SNPs, rs4878104 and rs4877365, in *DAPK1* that show significant association with LOAD. Unlike *APOE* that shows very strong association with LOAD even in small studies, rs4878104 was only marginally associated with LOAD (allelic P meta = 0.0016). However, allele frequencies in cases and controls were consistent in five out of six sample sets, with all six showing the same directionality of the risk allele. Besides allele frequency, homozygosity was more consistently and strongly associated with LOAD than heterozygosity in the initial three sample sets and the meta-analysis of all six sample sets. Several factors may have contributed to the non-significant result in three of the six sample sets. First, the original finding is likely to be an overestimate of the actual risk and thus our calculated power to replicate the initial result may be inflated. Based on the effect size of the five replication sample meta-analysis, we estimate that 1950 cases and 1950 controls are required to achieve 80% power and a one-sided significance of $P = 0.05$. Second, the tested marker may not reflect the causal variant thereby reducing the power of the study. Third, genetic heterogeneity in LOAD may contribute to the non-significant association in some samples. Fourth, family-based association analyses have little power when the effect size of the tested locus is much smaller than at least one other untested locus (18). Fifth, our haplotype analyses suggest that these SNPs are subject to homoplasy, which can reduce the power for tests of single SNP association. Overall, the disease association for a complex disease is generally evaluated through a meta-analysis of all tested sample sets, which also provides greater power than the analysis of multiple smaller sample sets (19,20). In this regard our result derived from six independent sample sets, totaling over 4000 samples, suggests that rs4878104 may be a genuine but weak risk factor for LOAD.

Another SNP, rs4877365, was significantly associated with LOAD in two of the initial three sample sets and trended towards significance in the third sample set. Similar to rs4878104, this marker also showed a stronger effect under a recessive model (data not shown). The results of our haplotype analyses suggest that one or both of these polymorphisms may be causal variants given that they are independently significant on multiple haplotype backgrounds and are more significant when these historically unrelated haplotypes are grouped together in the individual SNP tests. However, because of the lack of power for the contrasts between groups 3 and 1, we cannot rule out the possibility that these variants are in strong LD with unknown causal variants. Such causal variants may affect the transcription of DAPK1 and/or the expression/function of LOC643284, a recently predicted gene similar to 40S ribosomal protein S29 within this region of high LD.

DAPK1 is a Ca^{2+}/calmodulin-dependent serine/threonine kinase that plays a pro-apoptotic role in the programmed cell death cascade, including neuronal apoptosis. *DAPK1* is predominantly expressed in the brain and lung: in embryonic rat brain *DAPK1* mRNA is present at high levels in the cerebral cortex, cerebellar Purkinje cells and hippocampus, but is largely restricted to the hippocampus in adult rat brain (21). Western blot analysis detected DAPK1 protein in rat cortex, hippocampus and olfactory bulb, but not in cerebellum, hindbrain or mesencephalon (22). This expression pattern is particularly relevant to AD, because the hippocampus and cortex are the most severely affected brain regions, whereas regions of low expression, such as cerebellum and basal ganglia are less affected in AD (23). Increased *DAPK1* kinase activity or expression has been observed in neuronal cell death (24), and neurons lacking *DAPK1* are less susceptible to apoptotic insults in cell culture and knockout animal models (25,26). *DAPK1* is a transcriptional target of p53 (27), and is expected to play an important role in p53-mediated apoptosis; p53 expression is increased in the brains of AD patients (28) and the intracellular domain of amyloid precursor protein directly regulates p53 expression (29). In addition, a recent study has reported that DAPK1 kinase activity-deficient mice are more efficient learners and have better spatial memory than wild-type mice (30). *DAPK1* is therefore an excellent positional and biological candidate gene for AD.

Our functional study revealed that *DAPK1* expression shows allelic imbalance, a phenomenon (31) that is seen for genetic risk factors in other complex diseases such as calpain 10 and type 2 diabetes (32), and may explain the genetic association of *DAPK1* with LOAD. In addition, our study shows that the genotypes of the two intronic LOAD-associated SNPs are significantly associated with *DAPK1* allele-specific expression, although they are unlikely to be the sole *cis*-acting variants. Because of the intronic nature of the disease-associated SNPs, we cannot determine which allele is associated with increased expression. Nevertheless, this association suggests that these SNPs may interact with other unidentified polymorphic *cis*-acting regulatory factors to influence the level of *DAPK1* transcripts. We cannot exclude the possibility that they are in high LD with other polymorphic *cis*-acting elements governing *DAPK1* transcription. Regardless of the molecular mechanism of this regulation, considering that *DAPK1* allele-specific expression predicts variation in DAPK1 protein/activity and thus neuronal apoptotic potential, allele-specific expression of *DAPK1* variants provide a plausible explanation linking the genetic association with LOAD to a disease-relevant functional outcome. This hypothesis needs to be tested by correlating *DAPK1* brain expression with genotype. However, expression of *DAPK1* can be induced during neuronal apoptosis (21), which could potentially mask the effect of allele-specific expression of *DAPK1* when samples from patients with different degrees of apoptosis are assayed. Therefore, such an experiment should be ideally performed with tissues from normal brains.

An increasing body of evidence suggests that apoptosis may play a role in AD etiology. For example, changes in presenilin expression or activity have been associated with apoptotic phenotypes in cell-based models (33) and animal knock-down experiments (33,34). Genetic evidence for sporadic AD, such as the disease associations with *DAPK1*, *GAPD* (35) and *LOC439999* (36) variants, also point to apoptosis as a disease-relevant process.

In summary, these results, together with the existing data on the biological functions of *DAPK1* (21,22,24,30,37,38), strongly implicate *DAPK1* as one of the genetic factors affecting susceptibility to LOAD. The evidence includes (i) common SNPs in *DAPK1* are significantly associated with LOAD, (ii) *DAPK1* is in a previously reported linkage peak, (iii) DAPK1 is a pro-apoptotic mediator in the programmed cell death pathway, (iv) *DAPK1* is highly expressed in brain, particularly adult hippocampus and cortex, and abolition of DAPK1 kinase activity enhances learning and memory, and (v) *DAPK1* transcripts show allelic expression differences and the SNPs associated with LOAD risk may directly or indirectly modulate this allele-specific gene expression. *DAPK1* genotype/activity may influence risk for AD by influencing the cell number in the hippocampus and/or by influencing the response to environmental stimuli such as amyloid beta. As a small molecule inhibitor of *DAPK1* has been described for potential use in the treatment of ischemia-induced acute brain injury (37) and for attenuating neuronal damage in a chronic infusion model of amyloid beta toxicity (38), this compound or other classes of DAPK1 inhibitors may be candidates for clinical development and testing in AD.

MATERIALS AND METHODS

Study samples

Six LOAD case–control sample sets of Caucasian descent, collected with informed consent/assent from the participating individuals and approvals from the participating institutions, were used in this collaborative study (Table 1). The three sample sets that were used for discovery and initial replication are the WU sample set, obtained through the Washington University Alzheimer's Disease Research Center (ADRC) patient registry, the SD sample set obtained from the ADRC of the University of California, San Diego and the UK1 sample set obtained from the Medical Research Council Late Onset AD Genetic Resource that included samples from Cardiff University, Wales School of Medicine, King's College London and Cambridge University. These samples have been used in our recent studies (12,16,35,36,39–41). Further validation samples, the UK2 and UK3 samples, have not been previously described. The same ascertainment and diagnostic instruments and criteria were used in the collection of all UK samples.

An additional LOAD sample set ('linkage sample set') was generated by selecting one case per family from our genetic linkage sample of affected sib pairs and matching them to a corresponding number of Caucasian, non-demented controls collected in St Louis (these controls are independent of the controls used in the exploratory sample above) (Table 1).

SNP genotyping

Genotyping of SNPs was performed by allele-specific real-time PCR for individual samples using primers designed and

validated in-house (42). Previous analyses showed that the accuracy of our genotyping is better than 99%, as determined by internal comparisons of differentially designed assays for the same marker and comparisons for the same marker across different groups (35). Linkage cases and controls were genotyped using Sequenom™ technology by Washington University.

Statistical analysis

Genotyping: Hardy–Weinberg equilibrium was evaluated using an exact test as described by Weir ('Genetic Data Analysis II', Sinauer Associates, Sunderland MA, 1996, 2nd edition). Tests for allelic association of SNPs with disease status were carried out using the χ^2 test or a Fisher's exact test where cell sizes were <5. Markers with minor allele frequency of $\geq 2\%$ in either cases or controls were analyzed.

Meta analyses: When combining data from different sample sets, association was assessed using the method of Cochran Mantel and Haenzsel using sample set as the stratifying variable (43).

Haplotype analyses: We estimated haplotypes of the high LD region containing rs4878104 and rs4877365 using 16 SNPs (four exploratory and 12 fine mapping markers). For the haplotype analyses, all SNPs were genotyped in each of the three sample sets. Haplotypes were estimated for each population separately. Our previous analyses of these populations provided no evidence of stratification within or between these samples (35). To maximize our power to estimate the haplotype network and detect association, we combined the samples for subsequent analyses. Haplotypes were estimated using the software PHASE (44,45). A set of 95% plausible haplotype trees was estimated using statistical parsimony in the program TCS (46,47). A permutation-based haplotype analysis using PHASE was performed. Because of computational constraints, we used Tag 'n' Tell (http://snp.cgb.ki.se/tagntell/) to identify a subset of SNPs, which tag the haplotypes present at a frequency $>1\%$ in our combined samples. Association with LOAD was tested by tree scanning using unique haplotypes observed at least five times in the dataset (17). Tree scanning uses the phylogenetic network to define tests based on each branch of the tree. Each branch represents a pooling of haplotypes: haplotypes on one side of the branch are pooled together and define one allele, whereas the haplotypes on the other side are pooled to define a separate allele. This results in a biallelic locus that can be tested for association with the phenotype. Tree scanning incorporates a permutation-based multiple test correction and was modified to deal with case–control data (48).

In cases where a SNP is causal and subject to homoplasy, tree scanning may fail to detect association. A causal SNP that is subject to homoplasy should be significant when tested alone (all haplotypes grouped by state of the SNP). It should also show evidence for effects at each individual branch marked by the SNP if the comparison is conditioned upon other mutations at that SNP, provided the sample size is sufficient. These tests are performed by comparing the number of cases and controls in each haplotype group (defined by mutations at the SNP of interest) using Fisher's exact test on a 2×2 contingency table.

DAPK1 allele-specific expression study

DNA and total RNA were extracted from peripheral blood mononuclear cells of normal blood donors. A cDNA library was prepared from total RNA using the High Capacity cDNA Archive Kit (ABI). For *DAPK1*, two high-frequency exonic SNPs, rs3118863 or rs3818584, were used to measure allele-specific gene expression. High-frequency markers were selected to increase the number of heterozygotes. Allele-specific expression assays were carried out on cDNA samples with the same real-time PCR procedure as described for genotyping. The same primers were used for genotyping and allele-specific expression assays.

For the examination of allele-specific expression, 92 individual samples of genomic DNA were first genotyped for two expression markers, rs3118863 or rs3818584. Sixty-nine donors of Caucasian descent, heterozygous for rs3118863 or rs3818584, were then examined for allele-specific gene expression. cDNA was arrayed in quadruplicates or duplicates onto 384-well plates, together with appropriate PCR controls, and were run on an ABI-7900 real-time PCR system under standard conditions. Genomic DNA was also arrayed onto the same plate as a control. The relative expression of both alleles for each expression marker was determined by subtracting the smaller C_t value of one allele PCR reaction from the larger C_t value of another allele PCR reaction (dC_t). The ratio of two allele-specific transcripts was calculated as 2^{dC_t} (i.e. a one-cycle difference in our real-time PCR-based assay results in a 2-fold relative difference). For the statistical analysis, dC_t values were obtained as an average of two to four reactions for each sample and data point ($SE = 0.11$ for each assay, averaged across all samples). As a control, the dC_t values were also obtained with heterozygous genomic DNA, which theoretically should equal zero [actual: -0.05 (± 0.13) for rs3118863 and -0.02 (± 0.13) for rs3818584; average dC_t ($\pm$SD)].

For testing the relationship between disease-associated variants and allele-specific expression, the individuals who are heterozygous of the expression markers were then genotyped for the two *DAPK1* disease-associated markers. The allele-specific expression level measured by the expression marker was then stratified by the genotype status of the disease-associated markers (homozygote or heterozygote). The Mann–Whitney test was used to assess whether heterozygosity and homozygosity of the LOAD-associated SNPs are significantly associated with the allele-specific expression ratio. As a control, the same test was performed for the experimentally determined genotype ratio in heterozygous genomic DNA ($P = 0.45$–0.95, data not shown). Statistical significance was calculated separately for each combination of expression marker versus disease-associated marker.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We thank the families/individuals for their invaluable participation in this study, John Sninsky and Sam Broder for

stimulating discussions, and our colleagues at Celera Diagnostics for providing expert technical support. We acknowledge Mary Coats and Elizabeth Grant for coordinating the Washington University material, Mary Sundsmo for coordinating the University of California, San Diego, case material, Pamela Moore and Dragana Turic for providing clinical/DNA samples from the MRC UK Genetic Resource for LOAD. Funding for this work was partly provided by the National Institute of Health [Alzheimer's Disease Research Center Grants P50 AG05681 (J.C.M.), P50 AG05131 (L.T.); RO1 AG16208 (A. Goate) and PO1 AG03991 (J.C.M.)], the Medical Research Council, UK (J.W., M.O., M.O'D. and S.L.), and the Alzheimer's Research Trust (J.W., M.O., M.O'D. and S.L.). P.N. was partly supported by Missouri's Alzheimer's Disease and related Disorders Program. J.S.K.K. is a Ford Foundation Predoctoral Fellow and was supported by NIH training grant T32 HG00045. T.J.M. was supported by MICORTEX and NIH grant GM065509. J.H. was supported by the NIH intramural program and also by the VERUM Foundation (DIADEM project). D.C.R. is a Wellcome Trust Senior Research Fellow in Clinical Science.

Conflict of Interest statement. Some of the authors (Y.L., A. Grupe, C.R., K.T., T.A.T., S.K., J.C., T.J.W.) are employed by Celera Diagnostics. A. Goate received research funding from and was a consultant to Celera Diagnostics. M.O. and J.W. received research funding from Celera Diagnostics.

REFERENCES

1. Meyer, M.R., Tschanz, J.T., Norton, M.C., Welsh-Bohmer, K.A., Steffens, D.C., Wyse, B.W. and Breitner, J.C. (1998) APOE genotype predicts when–not whether–one is predisposed to develop Alzheimer disease. *Nat. Genet.*, **19**, 321–322.
2. Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S. and Roses, A.D. (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl Acad. Sci. USA*, **90**, 1977–1981.
3. Pastor, P. and Goate, A.M. (2004) Molecular genetics of Alzheimer's disease. *Curr. Psychiat. Rep.*, **6**, 125–133.
4. Pericak-Vance, M.A., Grubber, J., Bailey, L.R., Hedges, D., West, S., Santoro, L., Kemmerer, B., Hall, J.L., Saunders, A.M., Roses, A.D. *et al*. (2000) Identification of novel genes in late-onset Alzheimer's disease. *Exp. Gerontol.*, **35**, 1343–1352.
5. Curtis, D., North, B.V. and Sham, P.C. (2001) A novel method of two-locus linkage analysis applied to a genome scan for late onset Alzheimer's disease. *Ann. Hum. Genet.*, **65**, 473–481.
6. Myers, A., Wavrant De-Vrieze, F., Holmans, P., Hamshere, M., Crook, R., Compton, D., Marshall, H., Meyer, D., Shears, S., Booth, J. *et al*. (2002) Full genome screen for Alzheimer disease: stage II analysis. *Am. J. Med. Genet.*, **114**, 235–244.
7. Blacker, D., Bertram, L., Saunders, A.J., Moscarillo, T.J., Albert, M.S., Wiener, H., Perry, R.T., Collins, J.S., Harrell, L.E., Go, R.C. *et al*. (2003) Results of a high-resolution genome screen of 437 Alzheimer's disease families. *Hum. Mol. Genet.*, **12**, 23–32.
8. Olson, J.M., Goddard, K.A. and Dudek, D.M. (2002) A second locus for very-late-onset Alzheimer disease: a genome scan reveals linkage to 20p and epistasis between 20p and the amyloid precursor protein region. *Am. J. Hum. Genet.*, **71**, 154–161.
9. Wollmer, M.A., Streffer, J.R., Lutjohann, D., Tsolaki, M., Iakovidou, V., Hegi, T., Pasch, T., Jung, H.H., Bergmann, K., Nitsch, R.M. *et al*. (2003) ABCA1 modulates CSF cholesterol levels and influences the age at onset of Alzheimer's disease. *Neurobiol. Aging*, **24**, 421–426.
10. Bertram, L., Hiltunen, M., Parkinson, M., Ingelsson, M., Lange, C., Ramasamy, K., Mullin, K., Menon, R., Sampson, A.J., Hsiao, M.Y. *et al*. (2005) Family-based association between Alzheimer's disease and variants in UBQLN1. *N. Engl. J. Med.*, **352**, 884–894.
11. Katzov, H., Chalmers, K., Palmgren, J., Andreasen, N., Johansson, B., Cairns, N.J., Gatz, M., Wilcock, G.K., Love, S., Pedersen, N.L. *et al*. (2004) Genetic variants of ABCA1 modify Alzheimer disease risk and quantitative traits related to beta-amyloid metabolism. *Hum. Mutat.*, **23**, 358–367.
12. Li, Y., Tacey, K., Doil, L., van Luchene, R., Garcia, V., Rowland, C., Schrodi, S., Leong, D., Lau, K., Catanese, J. *et al*. (2004) Association of ABCA1 with late-onset Alzheimer's disease is not observed in a case–control study. *Neurosci. Lett.*, **366**, 268–271.
13. Brouwers, N., Sleegers, K., Engelborghs, S., Bogaerts, V., van Duijn, C.M., Paul De Deyn, P., Van Broeckhoven, C. and Dermaut, B. (2006) The UBQLN1 polymorphism, UBQ-8i, at 9q22 is not associated with Alzheimer's disease with onset before 70 years. *Neurosci. Lett.*, **392**, 72–74.
14. Kamboh, M.I., Minster, R.L., Feingold, E. and Dekosky, S.T. (2005) Genetic association of ubiquilin with Alzheimer's disease and related quantitative measures. *Mol. Psychiat.*, **11**, 273–279.
15. Slifer, M.A., Martin, E.R., Haines, J.L. and Pericak-Vance, M.A. (2005) The ubiquilin 1 gene and Alzheimer's disease. *N. Engl. J. Med.*, **352**, 2752–2753.
16. Smemo, S., Nowotny, P., Hinrichs, A.L., Kauwe, J.S., Cherny, S., Erickson, K., Myers, A.J., Kaleem, M., Marlowe, L., Gibson, A.M. *et al*. (2006) Ubiquilin 1 polymorphisms are not associated with late-onset Alzheimer's disease. *Ann. Neurol.*, **59**, 21–26.
17. Templeton, A.R., Maxwell, T., Posada, D., Stengard, J.H., Boerwinkle, E. and Sing, C.F. (2005) Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. *Genetics*, **169**, 441–453.
18. Li, M., Boehnke, M. and Abecasis, G.R. (2006) Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *Am. J. Hum. Genet.*, **78**, 778–792.
19. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. and Hirschhorn, J.N. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.*, **33**, 177–182.
20. Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, **38**, 209–203.
21. Yamamoto, M., Takahashi, H., Nakamura, T., Hioki, T., Nagayama, S., Ooashi, N., Sun, X., Ishii, T., Kudo, Y., Nakajima-Iijima, S. *et al*. (1999) Developmental changes in distribution of death-associated protein kinase mRNAs. *J. Neurosci. Res.*, **58**, 674–683.
22. Tian, J.H., Das, S. and Sheng, Z.H. (2003) Ca2+-dependent phosphorylation of syntaxin-1A by the death-associated protein (DAP) kinase regulates its interaction with Munc18. *J. Biol. Chem.*, **278**, 26265–26274.
23. Budson, A.E. and Price, B.H. (2005) Memory dysfunction. *N. Engl. J. Med.*, **352**, 692–699.
24. Yamamoto, M., Hioki, T., Ishii, T., Nakajima-Iijima, S. and Uchino, S. (2002) DAP kinase activity is critical for C(2)-ceramide-induced apoptosis in PC12 cells. *Eur. J. Biochem.*, **269**, 139–147.
25. Pelled, D., Raveh, T., Riebeling, C., Fridkin, M., Berissi, H., Futerman, A.H. and Kimchi, A. (2002) Death-associated protein (DAP) kinase plays a central role in ceramide-induced apoptosis in cultured hippocampal neurons. *J. Biol. Chem.*, **277**, 1957–1961.
26. Schori, H., Yoles, E., Wheeler, L.A., Raveh, T., Kimchi, A. and Schwartz, M. (2002) Immune-related mechanisms participating in resistance and susceptibility to glutamate toxicity. *Eur. J. Neurosci.*, **16**, 557–564.
27. Martoriati, A., Doumont, G., Alcalay, M., Bellefroid, E., Pelicci, P.G. and Marine, J.C. (2005) dapk1, encoding an activator of a p19ARF-p53-mediated apoptotic checkpoint, is a transcription target of p53. *Oncogene*, **24**, 1461–1466.
28. de la Monte, S.M., Sohn, Y.K. and Wands, J.R. (1997) Correlates of p53- and Fas (CD95)-mediated apoptosis in Alzheimer's disease. *J. Neurol. Sci.*, **152**, 73–83.
29. Costa, C.A., Sunyach, C., Pardossi-Piquard, R., Sevalle, J., Vincent, B., Boyer, N., Kawarai, T., Girardot, N., St George-Hyslop, P. and Checler, F. (2006) Presenilin-dependent gamma-secretase-mediated control of p53-associated cell death in Alzheimer's disease. *J. Neurosci.*, **26**, 6377–6385.

30. Yukawa, K., Tanaka, T., Bai, T., Li, L., Tsubota, Y., Owada-Makabe, K., Maeda, M., Hoshino, K., Akira, S. and Iso, H. (2006) Deletion of the kinase domain from death-associated protein kinase enhances spatial memory in mice. *Int. J. Mol. Med.*, **17**, 869–873.
31. Pastinen, T. and Hudson, T.J. (2004) *Cis*-acting regulatory variation in the human genome. *Science*, **306**, 647–650.
32. Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E. *et al*. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.*, **26**, 163–175.
33. Ye, Y. and Fortini, M.E. (1999) Apoptotic activities of wild-type and Alzheimer's disease-related mutant presenilins in Drosophila melanogaster. *J. Cell. Biol.*, **146**, 1351–1364.
34. Campbell, W.A., Yang, H., Zetterberg, H., Baulac, S., Sears, J.A., Liu, T., Wong, S.T., Zhong, T.P. and Xia, W. (2006) Zebrafish lacking Alzheimer presenilin enhancer 2 (Pen-2) demonstrate excessive p53-dependent apoptosis and neuronal loss. *J. Neurochem.*, **96**, 1423–1440.
35. Li, Y., Nowotny, P., Holmans, P., Smemo, S., Kauwe, J.S., Hinrichs, A.L., Tacey, K., Doil, L., van Luchene, R., Garcia, V. *et al*. (2004) Association of late-onset Alzheimer's disease with genetic variation in multiple members of the GAPD gene family. *Proc. Natl Acad. Sci. USA*, **101**, 15688–15693.
36. Grupe, A., Li, Y., Rowland, C., Nowotny, P., Hinrichs, A.L., Smemo, S., Kauwe, J.S., Maxwell, T.J., Cherny, S., Doil, L. *et al*. (2006) A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease. *Am. J. Hum. Genet.*, **78**, 78–88.
37. Velentza, A.V., Wainwright, M.S., Zasadzki, M., Mirzoeva, S., Schumacher, A.M., Haiech, J., Focia, P.J., Egli, M. and Watterson, D.M. (2003) An aminopyridazine-based inhibitor of a pro-apoptotic protein kinase attenuates hypoxia-ischemia induced acute brain injury. *Bioorg. Med. Chem. Lett.*, **13**, 3465–3470.
38. Craft, J.M., Watterson, D.M., Frautschy, S.A. and Van Eldik, L.J. (2004) Aminopyridazines inhibit beta-amyloid-induced glial activation and neuronal damage *in vivo*. *Neurobiol. Aging*, **25**, 1283–1292.
39. Busby, V., Goossens, S., Nowotny, P., Hamilton, G., Smemo, S., Harold, D., Turic, D., Jehu, L., Myers, A., Womick, M. *et al*. (2004) Alpha-T-catenin is expressed in human brain and interacts with the Wnt signaling pathway but is not responsible for linkage to chromosome 10 in Alzheimer's disease. *Neuromol. Med.*, **5**, 133–146.
40. Li, Y., Hollingworth, P., Moore, P., Foy, C., Archer, N., Powell, J., Nowotny, P., Holmans, P., O'Donovan, M., Tacey, K. *et al*. (2005) Genetic association of the APP binding protein 2 gene (APBB2) with late onset Alzheimer disease. *Hum. Mutat.*, **25**, 270–277.
41. Li, Y., Rowland, C., Tacey, K., Catanese, J., Sninsky, J., Hardy, J., Powell, J., Lovestone, S., Morris, J.C., Thal, L. *et al*. (2005) The BDNF Val66Met polymorphism is not associated with late onset Alzheimer's disease in three case–control samples. *Mol. Psychiat.*, **10**, 809–810.
42. Germer, S., Holland, M.J. and Higuchi, R. (2000) High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res.*, **10**, 258–266.
43. Agresti, A. (1990) *Categorical Data Analysis*. John Wiley & Sons.
44. Stephens, M. and Donnelly, P. (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, **73**, 1162–1169.
45. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.
46. Clement, M., Posada, D. and Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Mol. Ecol.*, **9**, 1657–1659.
47. Templeton, A.R., Weiss, K.M., Nickerson, D.A., Boerwinkle, E. and Sing, C.F. (2000) Cladistic structure within the human Lipoprotein lipase gene and its implications for phenotypic association studies. *Genetics*, **156**, 1259–1275.
48. Nowotny, P., Hinrichs, A.L., Smemo, S., Kauwe, J.S., Maxwell, T., Holmans, P., Hamshere, M., Turic, D., Jehu, L., Hollingworth, P. *et al*. (2005) Association studies between risk for late-onset Alzheimer's disease and variants in insulin degrading enzyme. *Am. J. Med. Genet. B: Neuropsychiatr. Genet.*, **136**, 62–68.