Title	Evidence that common variation in NEDD9 is associated with susceptibility to late-onset Alzheimer's and Parkinson's disease
Authors(s)	Li, Yonghong; Grupe, Andrew; Rowland, Charles; Holmans, Peter; Segurado, Ricardo; Abraham, Richard; Jones, Lesley; Catanese, Joseph; Ross, David; Mayo, Kevin; Martinez, Maribel; Hollingworth, Paul; Goate, Alison; Cairns, Nigel J.; Racette, Brad A.; Perlmutter, Joel S.; O'Donovan, Michael C.; Morris, John C.; Brayne, Carol; Rubinsztein, David C.; Lovestone, Simon; Thal, Leon J.; Owen, Michael J.; Williams, Julie
Publication date	2007-11-26
Publication information	Human Molecular Genetics, 17 (5): 759-767
Publisher	Oxford University Press
Item record/more information	http://hdl.handle.net/10197/4379
Publisher's statement	This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/ licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Publisher's version (DOI)	10.1093/hmg/ddm348

Evidence that common variation in *NEDD9* is associated with susceptibility to late-onset Alzheimer's and Parkinson's disease

Yonghong Li[1], Andrew Grupe[1,*], Charles Rowland[1], Peter Holmans[2], Ricardo Segurado[2], Richard Abraham[2], Lesley Jones[2], Joseph Catanese[1], David Ross[1], Kevin Mayo[3], Maribel Martinez[3], Paul Hollingworth[2], Alison Goate[3,4,5], Nigel J. Cairns[4,6], Brad A. Racette[4], Joel S. Perlmutter[4,7,8], Michael C. O'Donovan[2], John C. Morris[3,6], Carol Brayne[9], David C. Rubinsztein[9], Simon Lovestone[10], Leon J. Thal[11,†], Michael J. Owen[2] and Julie Williams[2]

[1]Celera, 1401 Harbor Bay Parkway, Alameda, CA 94502, USA, [2]Department of Psychological Medicine & Biostatistics and Bioinformatics Unit, Wales School of Medicine, Cardiff University, Cardiff CF14 4XN, UK, [3]Department of Psychiatry, [4]Department of Neurology, [5]Department of Genetics, [6]Department of Pathology & Immunology, [7]Department of Radiology and [8]Department of Anatomy & Neurobiology, Washington University School of Medicine, St Louis, MO 63110, USA, [9]Department of Medical Genetics, University of Cambridge, Cambridge Institute for Medical Research, Addenbrookés Hospital, Cambridge CB2 2XY, UK, [10]MRC Centre for Neurodegeneration Research, Institute of Psychiatry, King's College London, London SE5 8AF, UK and [11]Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA

Late-onset Alzheimer's disease (LOAD) and Parkinson's disease (PD) are the most common neurodegenerative disorders and in both diseases susceptibility is known to be influenced by genes. We set out to identify novel susceptibility genes for LOAD by performing a large scale, multi-tiered association study testing 4692 single nucleotide polymorphism (SNPs). We identified a SNP within a putative transcription factor binding site in the *NEDD9* gene (neural precursor cell expressed, developmentally down-regulated), that shows good evidence of association with disease risk in four out of five LOAD samples [$N = 3521$, $P = 5.38 \times 10^{-6}$, odds ratio (OR) = 1.38 (1.20–1.59)] and in addition, we observed a similar pattern of association in two PD sample sets [$N = 1464$, $P = 0.0145$, OR = 1.31 (1.05–1.62)]. In exploring a potential mechanism for the association, we observed that expression of *NEDD9* and *APOE* show a strong inverse correlation in the hippocampus of Alzheimer's cases. These data implicate *NEDD9* as a novel susceptibility gene for LOAD and possibly PD.

INTRODUCTION

Late-onset Alzheimer's disease (LOAD) is a common neurodegenerative disorder which shows strong evidence for genes contributing to disease susceptibility (1). The apolipoprotein E (*APOE*) *ε4* genotype is so far the only undisputed genetic risk factor (2). Although extensive searches over the past decade have suggested numerous other putative genetic associations, none has unequivocal statistical support. This is not unexpected given the complex nature of the disease. Our

study seeks to identify novel susceptibility genes for LOAD with a structured, tiered design implemented to increase statistical power but minimize genotyping. Other similar studies (3), that use multiple, large sample sets, have substantially bolstered confidence in genetic findings, as was recently demonstrated with the identification of *IFIH1* as a very modest risk factor in type 1 diabetes and variants in *IL12B* and *IL23R* in psoriasis (4,5).

In a previous study using a similar design we carried out a large-scale, genome-wide, gene-centric association study of 17 343 SNPs, located in 11 221 unique genes in multiple case–control series ($N = 3870$) from the UK and the USA published recently (6). To increase the number of interrogated genes and variants, this study tests a non-overlapping additional set of 4692 putative functional SNPs in 3664 unique genes. Eighty percent of these markers are either missense variants or occur within predicted transcription factor binding sites; they predominantly add to the number of missense and untranslated region (UTR) variants while increasing the number of markers in transcription factor (TF) binding sites by 75% (see Supplementary Material, Table S1). They are distributed across all autosomes and cover a wide range of allele frequencies (Supplementary Material, Fig. S1). Here we report the findings from a large-scale association study of these additional SNPs in multiple LOAD ($N = 3,521$) and subsequently, Parkinson's disease (PD: $N = 1464$) case–control sample sets. Our results provide good evidence linking *NEDD9* (neural precursor cell expressed, developmentally down-regulated) with LOAD and possibly PD.

RESULTS

Association of *NEDD9* variants with LOAD

We tested 4692 SNPs for association with disease risk in DNA pools from the UK1 sample set. Three hundred and four SNPs passed the follow-up threshold, 19 of which were not pursued because they had not shown significant evidence for association in the UK1, UK2, WU and/or SD sample sets based on individual genotyping in previous studies. Two hundred and eighty-five markers were investigated in pooled cases and controls from the WU sample set. Of these, 19 SNPs met the criteria for genotyping in individual samples of the UK1 and WU sample sets. Six of these 19 SNPs did not remain significant at a $P = 0.05$ level after individual genotyping. The 13 remaining lead SNPs were then individually genotyped and association tests were performed, in the combined UK2 and SD sample sets (Fig. 1).

Two SNPs met the criteria for replication in the SD and UK2 samples ($P < 0.05$, one-sided, based on the OR direction in the UK1 and WU sample sets combined). The more significant of these two markers was rs439401 [$P = 2.40 \times 10^{-11}$ (UK1+WU), 1.83×10^{-6} (UK2+SD)] near *APOE* on chromosome 19. This marker is in modest linkage disequilibrium (LD) with the *APOE* ε4 allele ($r^2 = 0.15$), and a regression analysis in the combined sample sets (UK1+WU+SD+UK2) controlling for the effects of the *APOE* ε4 (rs429358) and *APOE* ε 2/3 (rs7412) SNPs, was not significant (OR = 0.90 [95%CI: 0.78, 1.03], $P = 0.13$), suggesting that the association of rs439401 can be explained by *APOE* ε2,3,4. The rs439401 result vali-

dates the DNA pooling strategy and confirms that the tested functional SNPs can show association either directly or through LD with the causal variant.

The second most significant marker was rs760678 in *NEDD9* on chromosome 6 [$P = 0.0015$ (UK1+WU), $P = 0.0051$ [UK2+SD]] (Table 1 and Supplementary Material, Table S2). Comparison of the allelic and the genotypic models of rs760678 indicated a recessive mode of inheritance for the major allele in each of the four sample sets, and so we included a 1 degree of freedom test for association under a recessive model in further analyses. We next tested the marker in a fifth LOAD case–control sample set (UK3). The frequency of the risk allele (C) was consistent across all five sample sets as demonstrated by a test for heterogeneity within cases ($P = 0.86$) and within controls ($P = 0.12$), but the UK3 control genotype frequency is significantly different from other sample sets ($P = 0.033$, heterogeneity test of CC genotype within controls of the five sample sets). The combined replication samples showed significant association (UK2, SD, UK3; $P = 0.0060$) and a trend towards association when adjusted for 13 replication tests ($P = 0.078$).

In a combined analysis of the five sample sets, rs760678 showed highly significant association with risk of LOAD ($P_{\text{allelic}} = 2.94 \times 10^{-5}$, OR = 1.23 95% CI = [1.11,1.35]; $P_{\text{recessive}} = 5.38 \times 10^{-6}$, OR = 1.38, 95% CI = [1.20,1.59]; Table1) and was significant after a conservative Bonferroni correction for the 4692 SNPs tested here was applied (recessive genotypic $P_{\text{correction}} = 0.025$). A trend towards significance remains, even after correcting jointly for the 17 343 SNPs tested in our previous association study (6) and the SNPs tested here (22 035 tests, $P_{\text{correction}} = 0.12$). The relatively common risk allele homozygote (0.396 in cases versus 0.323 in controls) has a modest effect size, as expected from the common-disease–common-variant hypothesis. No significant effect on age at disease onset or significant interaction with *APOE4* was observed for the SNP. We note that the present sample size gives optimal power to detect a locus with an odds ratio of ~1.8, and is therefore underpowered to provide absolute evidence for association for loci with effect sizes in the observed range (OR of 1.2–1.3). We therefore await confirmation of the present NEDD9 finding in other samples, which may need to comprise a total of 10 000 cases and controls, to achieve adequate statistical power.

To determine whether the significant association of rs760678 was due to LD with other markers, we examined the LD profile in the HapMap data set (www.hapmap.org) on each side of rs760678 over a 500 kb region. Only seven SNPs, over 20.7 kb, share $r^2 > 0.1$ with rs760678 (Fig. 2). We tested these seven markers in the UK1 sample set and found that three were significantly associated with LOAD at the allelic level (Supplementary Material, Table S3). All three were significant under a recessive genotypic model, and their r^2 with rs760678 among the UK1 controls was 0.309, 0.433 and 0.932. We next tested the three markers in the UK2 sample set; only rs6940151 was significant under the allelic model ($P = 0.038$) and showed a trend for significance assuming a recessive model ($P = 0.10$). Stepwise logistic regression models including the three markers and rs760678 were performed in the combined UK1 and UK2 sample sets to assess whether the effect of rs760678 could
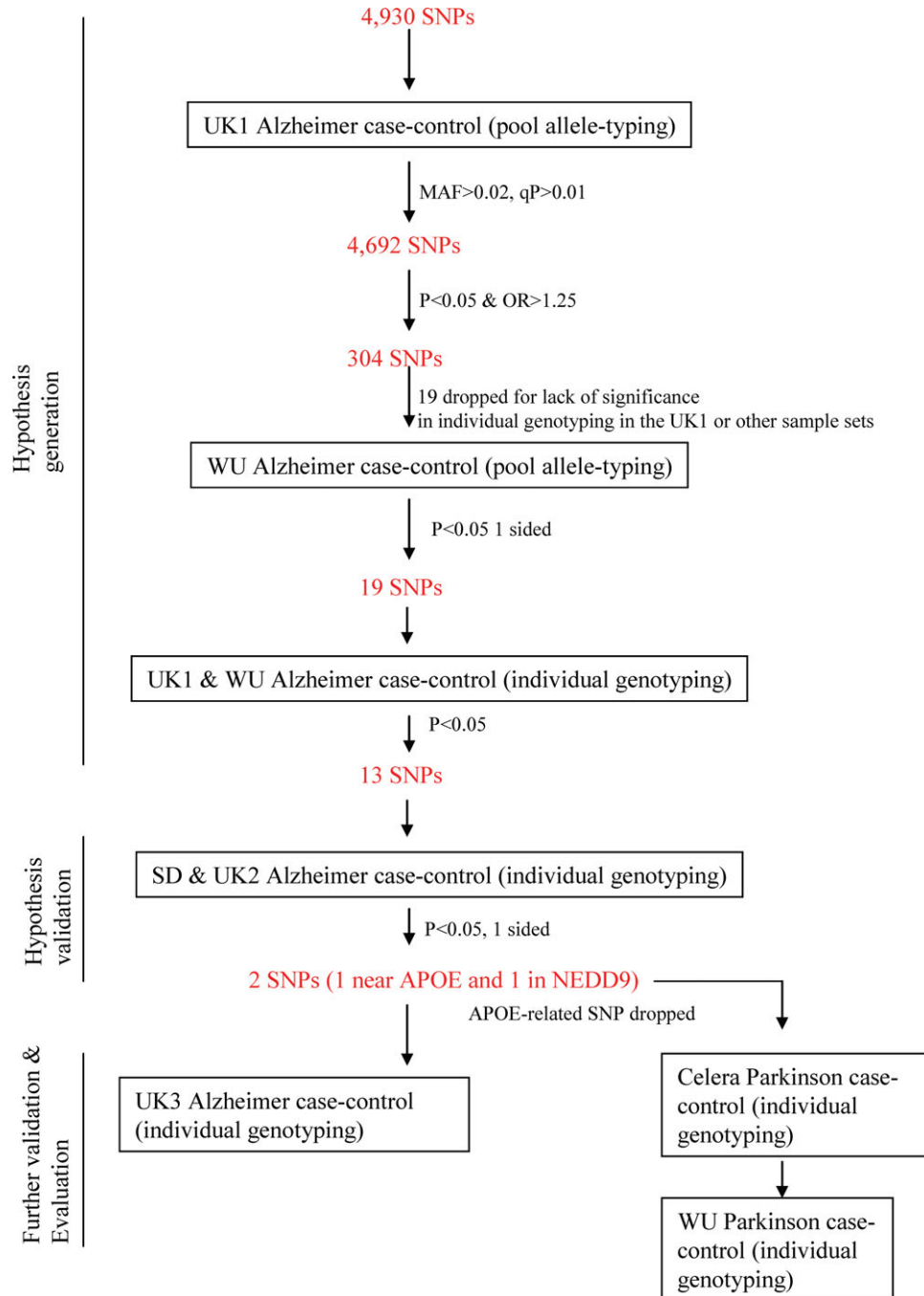
Figure 1. Schematic of the experimental design and follow up strategy. Disease-associated SNPs were identified and validated in three stages. In the hypothesis generation stage, lead markers were identified from sequential allele typing DNA pools of the UK1 sample set and DNA pools of the WU sample set. Significantly replicated markers were re-tested by individual genotyping. Lead markers were generated from the combined analysis of the UK1 and WU sample sets. In the hypothesis validation stage, markers were genotyped in the SD and UK2 sample sets and analyzed in the combined SD and UK2 sample sets. Replicated markers were defined as those reaching $P < 0.05$ (one-sided). One marker was further tested in the UK3 Alzheimer's disease case–control sample set and two late-onset Parkinson's disease case–control sample sets. This overall strategy is more stringent than that recommended by Skol *et al.* (39) and may thus have missed other potential lead candidates, which will be addressed in future studies. MAF, minor allele frequency; qP, quality *P*-value.

be explained by the other three SNPs. Regardless of the type of stepwise procedure (forward or backward) or assumed inheritance model (allelic or recessive), rs760678 remained the only SNP chosen in the final models (data not shown). The stepwise models could suffer from issues of multicollinearity, particularly between rs760678 and rs6940151.

However, we did not pursue further testing of rs6940151 because the marker is in tight LD with rs760678 and the association with LOAD is slightly weaker. Furthermore, while rs760678 and rs6940151 are both intronic SNPs, rs760678 maps to a putative GATA1 transcription factor binding site but rs6940151 does not.

Table 1. Association of rs760678 with Alzheimer's disease and Parkinson's disease

Disease	Sample set	Allelic test Sample number		Risk Allele Freq.				Genotypic test Risk Genotype Freq.				
		Case	Control	Case	Control	P^*	OR [95% CI]	Case	Control	P_{2df}^{**}	$P_{recessive}^*$	OR [95%CI]
AD	All AD	1652	1869	0.623	0.574	1.47×10^{-5}	1.23 [1.12, 1.35]	0.396	0.323	2.86×10^{-5}	2.69×10^{-6}	1.38 [1.20, 1.59]
	UK1	342	346	0.610	0.545	0.018	1.29 [1.05, 1.59]	0.401	0.289	0.0079	0.0021	1.64 [1.20, 2.26]
	WU	396	367	0.636	0.583	0.035	1.25 [1.02, 1.53]	0.424	0.327	0.018	0.0057	1.52 [1.13, 2.04]
	UK2	278	311	0.615	0.550	0.022	1.32 [1.04, 1.67]	0.367	0.299	0.07	0.08	1.36 [0.96, 1.92]
	SD	235	322	0.621	0.575	0.098	1.24 [0.96, 1.61]	0.379	0.292	0.092	0.032	1.48 [1.04, 2.11]
	UK3	401	523	0.626	0.601	0.29	1.11 [0.92, 1.33]	0.397	0.375	0.51	0.5	1.10 [0.84, 1.43]
PD	All PD	668	796	0.633	0.579	0.0017	1.26 [1.08, 1.46]	0.395	0.333	0.013	0.0073	1.31 [1.05, 1.62]
	Celera	310	310	0.640	0.584	0.039	1.28 [1.01, 1.62]	0.410	0.326	0.09	0.031	1.44 [1.03, 1.99]
	WU	358	486	0.627	0.576	0.035	1.24 [1.02, 1.51]	0.383	0.337	0.071	0.18	1.22 [0.92, 1.62]

AD, Alzheimer's disease; PD, Parkinson's disease.
*P-values derived by logistic regression, 2-sided.
**P-values derived by logistic regression, 2 degree of freedom test for the joint significance of 2 indicator variables denoting additive and dominance effects.
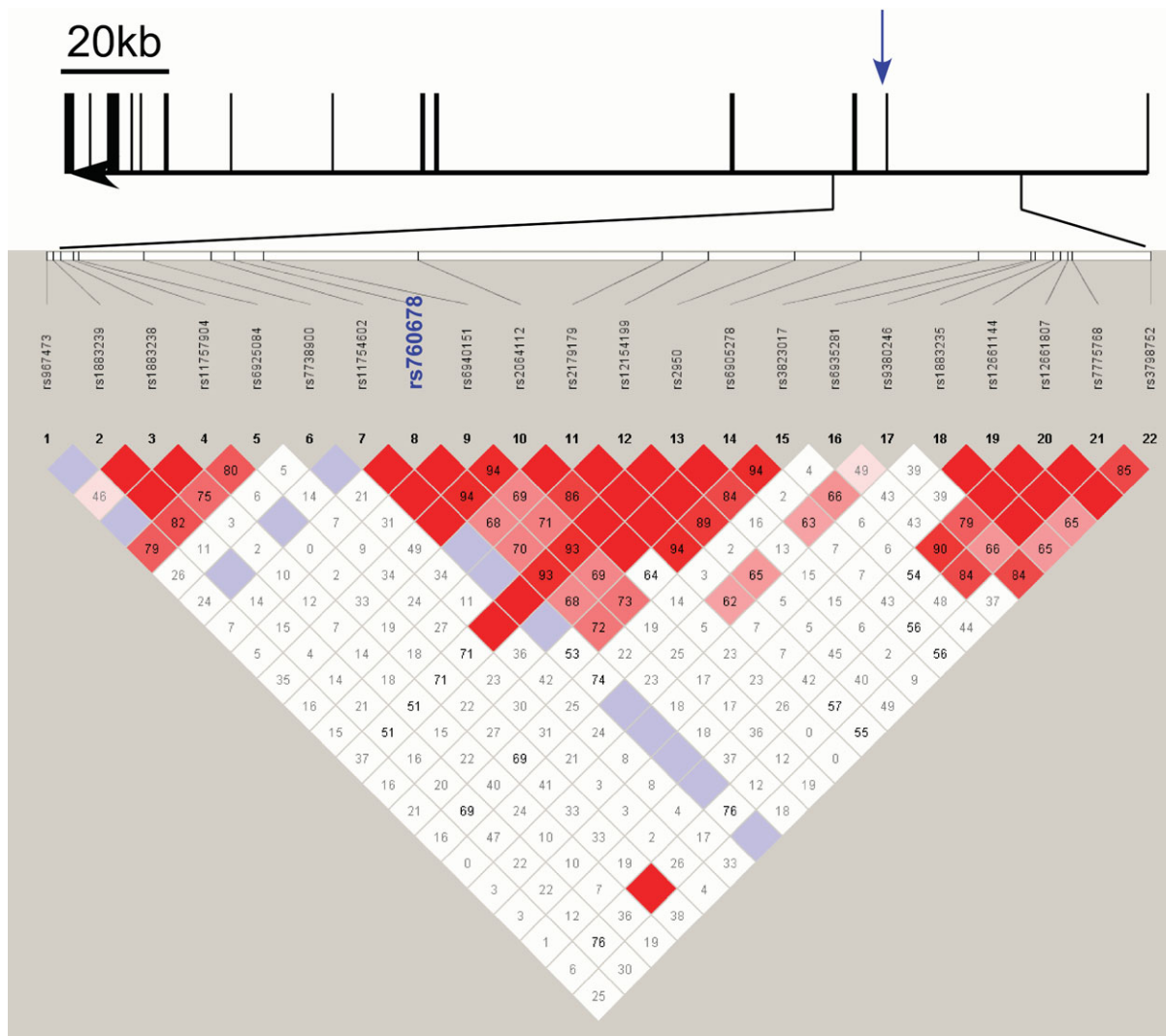

Figure 2. *NEDD9* gene structure, location of the disease-associated SNP and LD diagram. *NEDD9* exons (vertical bars) are based on the Ensemble database annotation (www.ensembl.org), which is consistent with the Celera and Unigene annotation but different from the NCBI annotation (Build 36.2). The Ensemble and Celera annotations are consistent with experimentally confirmed rat NEDD9 transcripts (13). The LD plot is from www.hapmap.org, with the pairwise linkage disequilibrium (D') given below the SNPs. The risk maker rs760678, highlighted in blue, is in intron 2.

To examine whether coding variants in *NEDD9* affect LOAD risk, we sequenced the *NEDD9* exons in 40 LOAD patients and identified one missense SNP rs34044517 (Tyr223Cys, refNP_006394.1). We genotyped this marker in the UK1 and UK2 sample sets, but did not observe a significant association (minor allele frequency of 0.018 in cases versus 0.026 in controls). Four other *NEDD9* missense SNPs are listed in the dbSNP database, but none was detected after sequencing 40 LOAD individuals or genotyping the UK1 and UK2 sample sets. Thus they likely represent rare variants.

Association of NEDD9 variation with PD

A previous study has suggested that common genetic factors influence age at onset for Alzheimer's disease and PD (7). In addition, some Alzheimer's disease and PD patients share similar clinical and pathological characteristics. Specifically, some PD patients develop dementia (8) and some patients with Alzheimer's disease also have Lewy bodies (9). Furthermore cellular and molecular studies have revealed pathways shared by both diseases (10). To test whether variation in NEDD9 might also influence risk for late onset PD, we examined *NEDD9*/rs760678 in a late-onset PD case–control series (Celera) (11,12) and obtained strikingly similar results to those found in LOAD; rs760678 showed a significant association with PD under the allelic and recessive models, and the risk allele and risk genotype frequencies in cases and controls were comparable with those observed in the LOAD sample sets (Table 1 and Supplementary Material, Table S2). To further validate the association in PD, we then tested the marker in a second, independently collected PD case–control series (WU). A replicated significant association of rs760678 with PD was observed at the allelic level; however the recessive effect was not significant but the risk genotype frequencies were in the same direction as in the other PD sample set (Table 1). Meta-analysis of the two PD case–control sample sets showed that the marker was significant at both the allelic and recessive level and the ORs were comparable with the LOAD study.

Inverse relationship between NEDD9 and APOE

NEDD9 was originally identified as a neural precursor cell expressed, developmentally down-regulated gene in the mouse brain (13). We first confirmed NEDD9 expression by immunohistochemistry in brains of LOAD patients and normal controls (Fig. 3A). We next explored a potential mechanism for the observed genetic association. Gene expression correlations, using Affymetrix microarray data, were calculated between probes in *NEDD9* and probes from genes that are relevant to Alzheimer's disease: *APH1A*, *APH1B*, *PEN2*, *NCSTN*, *PSEN1*, *PSEN2*, *APP*, *BACE1*, *BACE2*, *ECE1*, *ECE2*, *NEP*, *IDE* and *APOE*. We analyzed four data sets: expression data from caudate nucleus (33 controls) (14), cortex and cerebellum [79 Huntington's disease (HD) cases and controls] (14) and hippocampus of 22 Alzheimer's disease cases of varying severity reported by Blalock *et al.* (15). Full results are shown in Supplementary Material, Table S4. The correlation between *NEDD9* and *APOE* was
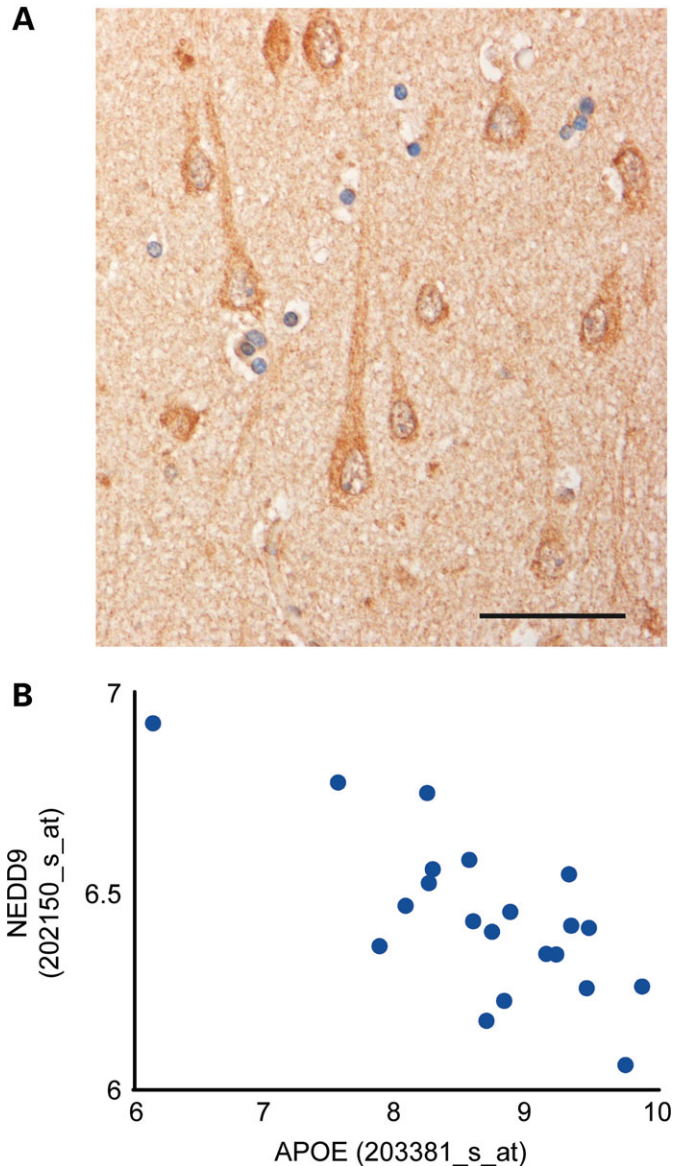

Figure 3. NEDD9 immunohistochemistry. (**A**) Normal adult brain showing diffuse cytoplasmic immunoreactivity in pyramidal neurons of the middle frontal lobe layer III. The characteristic intraneuronal inclusions of Alzheimer's disease (neurofibrillary tangles) and PD (Lewy bodies) were not immunoreactive with the NEDD9 antibody (data not shown). Bar = 50 μm. (**B**) Scatterplot of log expression levels for probes 202150_s_at (*NEDD9*) and 203381_s_at (*APOE*). Blalock Alzheimer's disease cases are denoted by circles. Expression levels were obtained using the RMA method (40), implemented in Bioconductor. (http://www.bioconductor.org).

the strongest we observed across all genes examined ($r = -0.76$) and was specific to the Blalock LOAD cases (Fig. 3B). Other Alzheimer's disease pathway genes showed consistent significant correlations with *NEDD9* across all four data sets, most notably *PSEN1*, *BACE1*, *BACE2* and *ECE1*. Using the same data set, we also analyzed the relationship between NEDD9 and several genes known to be involved in PD. Overall, the correlation was not as striking as the APOE result; however, a probe in PARK7/DJ1 gives a pointwise *P*-value < 0.01 in all four sample sets and several probes in

UCHL1 and SNCA are significant ($P < 0.05$) in cerebellum and caudate nucleus (Supplementary Material, Table S5). These results provide independent evidence linking *NEDD9* to biological pathways connected to Alzheimer's and possibly Parkinson's pathology.

DISCUSSION

The genetic evidence presented above shows a good level of consistency between sample sets and provides strong statistical support that *NEDD9* modulates LOAD risk. Association of rs760678 with LOAD is highly significant in all LOAD case–control sample sets combined, unadjusted for multiple testing ($P_{recessive} = 5.38 \times 10^{-6}$). Association of rs760678 with PD was significant in each of the two case–control sample sets tested, and rs760678 was the only marker followed up in the PD sample sets. Overall the association in individual LOAD and PD sample sets was relatively consistent, however, the effect size was moderate and thus these observations need to be confirmed in independent data sets of adequate statistical power.

As a member of the Cas family of adhesion docking proteins, *NEDD9* is involved in the formation of neurite-like membrane extensions and neurite outgrowth (13,16). A differential effect on neurite outgrowth has been described for APOE ε4 and APOE ε3 (17). Whether the disease associated marker has a similar differential effect requires further experimentation. However, rs760678 is located in a region containing clusters of TATA- and GATA- binding motifs within the *NEDD9* gene (current dbSNP database annotation of rs760678 to an intergenic region is erroneous; see legend to Fig. 2) (Supplementary Material, Fig. S2). GATA-binding proteins play an important role in regulating neuron-specific gene expression, and rs760678 is within the region that surrounds the core consensus sequence to which transcription factor binding extends (18). Differential expression of *NEDD9* transcripts may indirectly affect the number of neuronal cells/synapses in the brain such that risk genotype carriers have a significantly smaller reservoir of cells/synapses to withstand the gradual loss of neurons caused by other factors. Alternatively, the presumed differential effect may be critical in influencing neuronal regeneration under stressful conditions. For example, expression of *NEDD9* is up-regulated in human stem cells exposed to hypoxia (19) and after transient global ischemia in a rat model (13).

In this regard, it seems interesting that several LOAD candidate genes we have recently proposed, such as *GAPD* (20) and *DAPK1* (21), play a role in neuronal apoptosis, neurodegeneration and cancer (see, for example, refs 22,23). A connection between the risk factors for neurodegeneration and for cancer is not only plausible but has been demonstrated for presenilin (24,25). Furthermore, a recent prospective longitudinal study showed a significant inverse correlation between the risk of developing cancer and Alzheimer's disease (26). In this context it should be noted that *NEDD9* has recently been identified as a melanoma metastasis gene and is one of the lung metastasis signature genes (27,28). The *NEDD9* result and others may warrant the aggregation of robustly associated LOAD risk markers to develop a more compelling predictor of disease risk than *APOE* alone.

In conclusion, these data support a role for variation at a putative transcription factor binding site or at a correlated locus within the *NEDD9* gene, in increasing the risk of developing Alzheimer's disease and possibly PD disease and we encourage further studies in independent sample sets.

MATERIALS AND METHODS

Clinical samples

All samples used in this study were collected or acquired with informed consent/assent from the participating individuals and approvals from the participating institutions. For LOAD, we used a total of five case–control sample sets collected by Cardiff University, King's College London and University of Cambridge (UK 1, 2 & 3: MRC Genetic Resource for LOAD), Washington University, St Louis (WU) and the University of California, San Diego (SD). Cases had a minimum age at disease onset (AAO) of 60 years and a diagnosis of probable or definite Alzheimer's disease (NINCDS-ADRDA), while controls were ascertained at the age of 65 years or older and screened for evidence of dementia (MMSE ≥ 28, Clinical Dementia Rating = 0 or full neurological exam). All individuals were of Caucasian descent. All sample sets show the expected *APOE ε4* distribution in cases and controls. Detailed information can be found in our recent studies (6,29,30).

For PD, one case–control series (Celera) was constructed from cases and matched population/convenience controls that are available through the NINDS Human Genetics Resources at the Coriell Institute. Cases had a minimum AAO of 50 years and met UK Brain Bank criteria for idiopathic PD (31). Controls were neurologically normal. All individuals were of Caucasian descent. Detailed information can be found in our recent studies (11,12). A second case–control series, consisting of 358 cases (39% females) and 486 controls (62% females) was obtained from the Washington University at St Louis (WU). Cases met diagnostic criteria of the British Brain Bank for idiopathic PD as implemented by Racette *et al*. (32). Cases had an AAO ranging from 40 to 85 years [mean ($\pm$ SD) 60.56 ± 10.48 years] and were sampled at the age of 41–90 years [mean ($\pm$ SD) 66.96 ± 10.4 years] (one sample without age data). Controls were sampled at the age of 41–105 years [mean ($\pm$ SD) 72.36 ± 11.49 years] (36 samples without age data) and were neurologically normal.

Generation of DNA pools and frequency estimation

In brief, from each of the UK1 and WU samples (UK1: 380 cases and 396 controls; WU: 376 cases and 344 controls) we constructed 12 DNA pools. Overlapping pools for each sample set were constructed to contain (i) all male cases, (ii) all female cases, (iii) all cases with AAO < 75 years, (iv) all cases with AAO ≥ 75 years, (v) all cases who carry one or more copies of the APOE 4 allele, (vi) all cases who do not carry an APOE 4 allele. The six corresponding pools of control subjects were also constructed with the exception that control subjects were stratified by age at examination rather than AAO. Pools were constructed taking account of phenotypic characteristics to allow future analyses of specific

phenotype–genotype relationships, although none are included in this initial analysis. The allele frequencies of each of the 12 pools were determined from the average of two SYBR green real-time PCR reactions as described by Germer *et al.* (33). Allele numbers (rounded to the nearest integer) for each pool were estimated by multiplying the chromosome count in the pool by the estimated allele frequency. Further details can be found in Grupe *et al.* (6).

SNP genotyping and study design

To identify markers that show the strongest and most consistent association, we used a tiered, multi-stage association design, involving five LOAD case–control sample sets (Fig. 1), mirroring the sample prioritization in our previous study (6). Specifically, in the hypothesis generation phase, we determined allele frequencies of 4930 SNPs in DNA pools from the UK1 sample set. Markers of low frequency (minor allele frequency of < 2% in both cases and controls) and inconsistent allele frequencies among the triplicate pools of identical composition were excluded. Surviving SNPs were tested for association with disease risk. SNPs with an allelic $P < 0.05$ and odds ratio (OR) >1.25 or <0.8 were then assayed in pooled cases and controls in the WU sample set. Association testing revealing an allelic $P < 0.05$ (one-sided, based on the same OR direction as in the UK1 pooled samples) were genotyped in individual samples of the UK1 and WU sample sets to eliminate false positives derived from fluctuations in frequency estimates from DNA pools. To increase the power of our analysis, we tested the association of these markers with disease risk in the combined sample sets. SNPs identified with an allelic $P < 0.05$ were considered as lead candidates for a potential role in LOAD risk. To validate the lead markers, we then individually genotyped each SNP in two independent case–control sample sets (the UK2 and SD sample sets), and performed an association test in a combined sample set. The detection of a one-sided P-value < 0.05 in the UK2+SD sample (OR in the same direction as in the UK1+WU sample) was used as a criterion for genotyping in the UK3 sample.

Genotyping of SNPs was performed by allele-specific real-time PCR using primers designed and validated in-house (33). Genotyping rs760678 in 80 HapMap CEPH samples gave results that are 100% concordant with the HapMap genotypes. Genotyping rs760678 in the UK3 and WU Parkinson's case–control samples was done on the Sequenom platform.

Statistical methods

Exact tests were used to compare genotype frequencies with those expected under Hardy–Weinberg equilibrium as described by Weir ('Genetic Data Analysis II', Sinauer Associates, Sunderland MA 1996, 2nd Edition). A χ^2 test P-value <0.01 was used to exclude assays with inconsistent allele frequencies across triplicates of identical DNA sample pools. Initial testing of SNPs for association with Alzheimer's disease in pooled and individually genotyped sample sets was performed using Pearson's χ^2 test. Logistic regression models were used to estimate odds ratios and a Wald test was used to assess statistical significance of replicated SNPs. A multiplica-

tive risk model was assumed to assess allelic odds ratios. Joint analyses of more than one sample were performed using fixed-effect Mantel–Haenszel methods, and the associated P-values were calculated using a standard test of the Mantel–Haenszel combined odds ratio = 1. For the genotypic tests a logistic regression model was applied; for analyses containing subjects from more than one study, sample set was entered as a covariate. The logistic regression models assessing the effect of rs439401 while correcting for the effects of *APOE* ε4 and *APOE* ε 2/3 assumed multiplicative allelic risk for each SNP. Tests for heterogeneity of allele or genotype frequencies across sample sets were performed using Pearson's χ^2 test. Gene × gene interaction was assessed by testing for significance of the cross-product term of the two SNPs in a logistic regression model. Stepwise regression models were performed using a P-value criterion of 0.05 or less for including and retaining variables in the model. Differences in median age of onset (cases only) by genotype subgroups were evaluated using the Wilcoxon rank sum test.

Expression analysis

We have access to microarray expression data (Affymetrix HG133 A and B chips) on 79 individuals (HD [MIM 143100]) cases and age-matched controls). Correlations were calculated between four probes mapping to *NEDD9* and a total of 32 probes mapping to the following genes involved in the γ-secretase and β-amyloid degrading pathways: *APH1A* [MIM 607629], *APH1B* [MIM 607630], *PEN2* [MIM 607632], *NCSTN* [MIM 605254], *PSEN1* [MIM 104311], *PSEN2* [MIM 600759], *APP* [MIM 104760], *BACE1* [MIM 604252], *BACE2* [MIM 605668], *ECE1* [MIM 600423], *ECE2* [MIM 610145], *NEP* [MIM 120520], *IDE* [MIM 146680] and *APOE*. Three brain regions were analyzed: cerebellum (70 individuals), cortex (67 individuals) and caudate nucleus (controls only; 33 individuals). Differences in sample size between the regions reflect first, that HD cases were not analyzed for the caudate nucleus due to severe and widespread chances of expression associated with HD (14) and secondly, the failure of some chips to meet quality control standards.

For more details of the samples and procedures used to generate the expression data (including quality control) refer to previous works (14,34). Expression correlations were also investigated in a sample of 22 Alzheimer's disease hippocampi from cases of varying severity (15). For these, data were only available on the A chip.

Pearson correlations between the *NEDD9* probes and those from other genes are shown in Supplementary Material, Table S4, together with their P-values. 'Gene-wise' P-values, correcting for multiple probe–probe comparisons within genes, were obtained by randomly permuting the expression values for all probes within a gene simultaneously among individuals (performing a separate permutation for each gene), repeating the correlation analysis, and comparing the resulting correlations to those in the actual data. A similar procedure was used to obtain 'experiment-wise' P-values, correcting not only for multiple probe comparisons within genes, but also for multiple gene–gene comparisons.

NEDD9 immunohistochemistry

After death, the consent of the next-of-kin was obtained for brain removal, following approved Local Ethics Committee procedures in accordance with state law. Brain tissues from clinically and neuropathologically characterized cases of AD, PD and normal aged controls were obtained from the Washington University Alzheimer's Disease Research Center (WUADRC). Cases with Alzheimer's disease or PD displayed the characteristic pathological features of these diseases, as previously described (35–38).

Brain tissue was preserved in buffered 10% formal saline for up to 3 weeks. Antigen retrieval was performed by microwaving tissue sections in a solution of 0.1 M citrate buffer, pH 6.0 at 100ºC for 10 min. Immunohistochemistry was undertaken on 5 µm thick sections prepared from formalin-fixed, paraffin wax-embedded tissue blocks using anti-NEDD9 antibodies (1:5000 mouse monoclonal, Abnova Corporation, Taipei City, Taiwan) and the avidin–biotin complex detection system (Vector Laboratories, Burlingham, CA) and the chromogen 3,3'-diaminobenzidine (DAB) and sections were counterstained with hematoxylin.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

REFERENCES

1. Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A. and Pedersen, N.L. (2006) Role of genes and environments for explaining Alzheimer disease. *Arch Gen. Psychiatry*, **63**, 168–174.
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. Cardon, L.R. and Bell, J.I. (2001) Association study designs for complex diseases. *Nat. Rev. Genet.*, **2**, 91–99.
4. Smyth, D.J., Cooper, J.D., Bailey, R., Field, S., Burren, O., Smink, L.J., Guja, C., Ionescu-Tirgoviste, C., Widmer, B., Dunger, D.B. *et al.* (2006) A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat. Genet.*, **38**, 617–619.
5. Cargill, M., Schrodi, S.J., Chang, M., Garcia, V.E., Brandon, R., Callis, K.P., Matsunami, N., Ardlie, K.G., Civello, D., Catanese, J.J. *et al.* (2007) A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am. J. Hum. Genet.*, **80**, 273–390.
6. Grupe, A., Abraham, R., Li, Y., Rowland, C., Hollingworth, P., Morgan, A., Jehu, L., Segurado, R., Stone, D., Schadt, E. *et al.* (2007) Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum. Mol. Genet.*, **16**, 865–873.
7. Li, Y.J., Scott, W.K., Hedges, D.J., Zhang, F., Gaskell, P.C., Nance, M.A., Watts, R.L., Hubble, J.P., Koller, W.C., Pahwa, R. *et al.* (2002) Age at onset in two common neurodegenerative diseases is genetically controlled. *Am. J. Hum. Genet.*, **70**, 985–993.
8. Boller, F., Mizutani, T., Roessmann, U. and Gambetti, P. (1980) Parkinson disease, dementia, and Alzheimer disease: clinicopathological correlations. *Ann. Neurol.*, **7**, 329–335.
9. Rosenblum, W.I. and Ghatak, N.R. (1979) Lewy bodies in the presence of Alzheimer's disease. *Arch. Neurol.*, **36**, 170–171.
10. Selkoe, D.J. (2004) Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases. *Nat. Cell Biol.*, **6**, 1054–1061.
11. Li, Y., Schrodi, S., Rowland, C., Tacey, K., Catanese, J. and Grupe, A. (2006) Genetic evidence for ubiquitin-specific proteases USP24 and USP40 as candidate genes for late-onset Parkinson disease. *Hum. Mutat.*, **27**, 1017–1023.
12. Li, Y., Rowland, C., Schrodi, S., Laird, W., Tacey, K., Ross, D., Leong, D., Catanese, J., Sninsky, J. and Grupe, A. (2006) A case-control association study of the 12 single nucleotide polymorphisms implicated in Parkinson's disease by a recent genome scan. *Am. J. Hum. Genet.*, **78**, 1090–1092.
13. Sasaki, T., Iwata, S., Okano, H.J., Urasaki, Y., Hamada, J., Tanaka, H., Dang, N.H., Okano, H. and Morimoto, C. (2005) Nedd9 protein, a Cas-L homologue, is upregulated after transient global ischemia in rats: possible involvement of Nedd9 in the differentiation of neurons after ischemia. *Stroke*, **36**, 2457–2462.
14. Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstag, T., Hughes, G., Elliston, L.A., Hartog, C., Goldstein, D.R., Thu, D. *et al.* (2006) Regional and cellular gene expression changes in human Huntington's disease brain. *Hum. Mol. Genet.*, **15**, 965–977.
15. Blalock, E.M., Geddes, J.W., Chen, K.C., Porter, N.M., Markesbery, W.R. and Landfield, P.W. (2004) Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl Acad. Sci. USA*, **101**, 2173–2178.
16. Bargon, S.D., Gunning, P.W. and O'Neill, G.M. (2005) The Cas family docking protein, HEF1, promotes the formation of neurite-like membrane extensions. *Biochim. Biophys. Acta*, **1746**, 143–154.
17. Nathan, B.P., Bellosta, S., Sanan, D.A., Weisgraber, K.H., Mahley, R.W. and Pitas, R.E. (1994) Differential effects of apolipoproteins E3 and E4 on neuronal growth in vitro. *Science*, **264**, 850–852.
18. Lawson, M.A., Whyte, D.B. and Mellon, P.L. (1996) GATA factors are essential for activity of the neuron-specific enhancer of the gonadotropin-releasing hormone gene. *Mol. Cell Biol.*, **16**, 3596–3605.
19. Martin-Rendon, E., Hale, S.J., Ryan, D., Baban, D., Forde, S.P., Roubelakis, M., Sweeney, D., Moukayed, M., Harris, A.L., Davies, K. *et al.* (2006) Transcriptional profiling of human cord blood CD133+ and cultured bone

marrow mesenchymal stem cells in response to hypoxia. *Stem Cells*, **25**, 1003–1012.

20. 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.

21. Li, Y., Grupe, A., Rowland, C., Nowotny, P., Kauwe, J.S., Smemo, S., Hinrichs, A.L., Tacey, K., Toombs, T.A., Kwok, S. *et al.* (2006) DAPK1 variants are associated with Alzheimer's disease and allele-specific expression. *Hum. Mol. Genet.*, **15**, 2560–2568.

22. 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.

23. Hara, M.R., Agrawal, N., Kim, S.F., Cascio, M.B., Fujimuro, M., Ozeki, Y., Takahashi, M., Cheah, J.H., Tankou, S.K., Hester, L.D. *et al.* (2005) S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. *Nat. Cell Biol.*, **7**, 665–674.

24. Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D.M., Oshima, J., Pettingell, W.H., Yu, C.E., Jondro, P.D., Schmidt, S.D., Wang, K. *et al.* (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science*, **269**, 973–977.

25. Roperch, J.P., Alvaro, V., Prieur, S., Tuynder, M., Nemani, M., Lethrosne, F., Piouffre, L., Gendron, M.C., Israeli, D., Dausset, J. *et al.* (1998) Inhibition of presenilin 1 expression is promoted by p53 and p21WAF-1 and results in apoptosis and tumor suppression. *Nat. Med.*, **4**, 835–838.

26. Roe, C.M., Behrens, M.I., Xiong, C., Miller, J.P. and Morris, J.C. (2005) Alzheimer disease and cancer. *Neurology*, **64**, 895–898.

27. Kim, M., Gans, J.D., Nogueira, C., Wang, A., Paik, J.H., Feng, B., Brennan, C., Hahn, W.C., Cordon-Cardo, C., Wagner, S.N. *et al.* (2006) Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell*, **125**, 1269–1281.

28. Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L. and Massague, J. (2005) Genes that mediate breast cancer metastasis to lung. *Nature*, **436**, 518–524.

29. 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.

30. 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.

31. Hughes, A.J., Daniel, S.E., Kilford, L. and Lees, A.J. (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry*, **55**, 181–184.

32. Racette, B.A., Rundle, M., Parsian, A. and Perlmutter, J.S. (1999) Evaluation of a screening questionnaire for genetic studies of Parkinson's disease. *Am. J. Med. Genet.*, **88**, 539–543.

33. 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.

34. Jones, L., Goldstein, D.R., Hughes, G., Strand, A.D., Collin, F., Dunnett, S.B., Kooperberg, C., Aragaki, A., Olson, J.M., Augood, S.J. *et al.* (2006) Assessment of the relationship between pre-chip and post-chip quality measures for Affymetrix GeneChip expression data. *BMC Bioinformatics*, **7**, 211.

35. The National Institute on Aging and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease (1997) Consensus recommendations for the postmortem diagnosis of Alzheimer's disease The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. *Neurobiol Aging*, **18**, S1–S2.

36. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. and Del Tredici, K. (2006) Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol. (Berl)*, **112**, 389–404.

37. Braak, H., Ghebremedhin, E., Rub, U., Bratzke, H. and Del Tredici, K. (2004) Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res.*, **318**, 121–134.

38. Mirra, S.S., Heyman, A., McKeel, D., Sumi, S.M., Crain, B.J., Brownlee, L.M., Vogel, F.S., Hughes, J.P., van Belle, G. and Berg, L. (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*, **41**, 479–486.

39. 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.

40. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. and Speed, T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.