<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Evidence that common variation in NEDD9 is associated with susceptibility to late-onset Alzheimer's and Parkinson's disease</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors(s)</strong></td>
<td>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</td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>2007-11-26</td>
</tr>
<tr>
<td><strong>Publication information</strong></td>
<td>Human Molecular Genetics, 17 (5): 759-767</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>Oxford University Press</td>
</tr>
<tr>
<td><strong>Item record/more information</strong></td>
<td><a href="http://hdl.handle.net/10197/4379">http://hdl.handle.net/10197/4379</a></td>
</tr>
<tr>
<td><strong>Publisher's statement</strong></td>
<td>This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<a href="http://creativecommons.org/licenses/by-nc/2.0/uk/">http://creativecommons.org/licenses/by-nc/2.0/uk/</a>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.</td>
</tr>
<tr>
<td><strong>Publisher's version (DOI)</strong></td>
<td>10.1093/hmg/ddm348</td>
</tr>
</tbody>
</table>
Evidence that common variation in \textit{NEDD9} is associated with susceptibility to late-onset Alzheimer’s and Parkinson’s disease

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

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

Received September 21, 2007; Revised November 21, 2007; Accepted November 26, 2007

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 \textit{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 [\textit{N} = 3521, \textit{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 [\textit{N} = 1464, \textit{P} = 0.0145, OR = 1.31 (1.05–1.62)]. In exploring a potential mechanism for the association, we observed that expression of \textit{NEDD9} and \textit{APOE} show a strong inverse correlation in the hippocampus of Alzheimer’s cases. These data implicate \textit{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 (\textit{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

\textsuperscript{*}To whom correspondence should be addressed. Tel: +1 5107496222; Fax: +1 5107496286; Email: andrew.grupe@celera.com
\textsuperscript{†}This paper is dedicated to the memory of Leon Thal for his many contributions to Alzheimer’s research and patient therapy.

© 2007 The Author(s)

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.
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 IFI1H 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 sense or 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 = 14644) 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 samples. 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 × 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 e4 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 e4 (rs429358) and APOE e2/3 (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 e2,3,4. The rs439401 result validates 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, 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_{allelic} = 2.94 × 10^{-6}, OR = 1.23 95% CI = [1.11,1.35]; P_{recessive} = 5.38 × 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_{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_{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
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 multicolinearity, 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

<table>
<thead>
<tr>
<th>Disease</th>
<th>Sample set</th>
<th>Sample number</th>
<th>Allelic test</th>
<th>Genotypic test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk Allele</td>
<td>Genotypic test</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Freq.</td>
<td>Control</td>
</tr>
<tr>
<td>AD</td>
<td>All AD</td>
<td>1652</td>
<td>1869</td>
<td>0.623</td>
</tr>
<tr>
<td></td>
<td>UK1</td>
<td>342</td>
<td>346</td>
<td>0.610</td>
</tr>
<tr>
<td></td>
<td>WU</td>
<td>396</td>
<td>367</td>
<td>0.636</td>
</tr>
<tr>
<td></td>
<td>UK2</td>
<td>278</td>
<td>311</td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>235</td>
<td>322</td>
<td>0.621</td>
</tr>
<tr>
<td></td>
<td>UK3</td>
<td>401</td>
<td>523</td>
<td>0.626</td>
</tr>
<tr>
<td>PD</td>
<td>All PD</td>
<td>668</td>
<td>796</td>
<td>0.633</td>
</tr>
<tr>
<td></td>
<td>Celera</td>
<td>310</td>
<td>310</td>
<td>0.640</td>
</tr>
<tr>
<td></td>
<td>WU</td>
<td>358</td>
<td>486</td>
<td>0.627</td>
</tr>
</tbody>
</table>

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 marker 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 (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).

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).
UCLH1 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 ($P_{\text{corrected}} = 5.38 \times 10^{-6}$). 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 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 (± SD) 60.56 ± 10.48 years] and were sampled at the age of 41–90 years [mean (± SD) 66.96 ± 10.4 years] (one sample without age data). Controls were sampled at the age of 41–105 years [mean (± 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 \( P < 0.05 \) 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 \( \chi^2 \) 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 \( \chi^2 \) test. Logistic regression models were used to estimate odds ratios and a Wald test was used to assess statistical significance of replicated SNPs. A multiplicative 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 \( APOE \varepsilon4 \) and \( APOE \varepsilon2/3 \) assumed multiplicative allelic risk for each SNP. Tests for heterogeneity of allele or genotype frequencies across sample sets were performed using Pearson’s \( \chi^2 \) test. Gene \( \times \) 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 \( \gamma \)-secreterase and \( \beta \)-amyloid degrading pathways: \( A\)PH\( A \) [MIM 607629], \( A\)PH\( B \) [MIM 607630], \( P\)EN\( 2 \) [MIM 607632], NCS\( T\)N [MIM 605254], PSEN\( 1 \) [MIM 104311], PSEN\( 2 \) [MIM 600759], APP [MIM 104760], BACE\( 1 \) [MIM 604252], BACE\( 2 \) [MIM 605668], ECE\( 1 \) [MIM 600423], ECE\( 2 \) [MIM 610145], NEP [MIM 120520], IDE [MIM 146680] and \( A\)PO\( E \). 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 100°C 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.

ACKNOWLEDGEMENTS

We thank the families/individuals for their invaluable participation in this study, John Sninsky, Thomas J. White and Samuel Broder for stimulating discussions and our colleagues at Celera for expert technical contributions. We acknowledge Pamela Moore and Dragana Turic for providing DNA samples from the MRC UK Genetic Resource for LOAD, Mary Coats and Elizabeth Grant for coordinating the Washington University ADRC material and Mary Sundsmo for coordinating the University of California, San Diego, material. We thank the contributors and the organizers, particularly Dr Katrina Gwinn Hardy and Jeanne Beck, of the NINDS Human Genetics Resources for making the clinical samples (Celera PD) available to the PD research community.

Conflict of Interest statement. Y.L., A.G., C.R., D.R. and J.C. are employed by Celera. A.G. received research funding from and was a consultant to Celera. M.J.O. and J.W. received research funding from Celera.

FUNDING

Funding for this work was partly provided by the Medical Research Council, UK (J.W., M.J.O., M.C.O’D., S.L., D.C.R. and C.B.), the American Parkinson Disease Association (J.S.P. and B.A.R.), the Greater St Louis Chapter of the APDA and the Alzheimer’s Research Trust (J.W., M.J.O., M.C.O’D. and S.L.), the National Institute of Health [Alzheimer’s disease research center grants P50 AG05681 (J.C.M.), P50 AG05131 (L.T.); RO1 AG16208 (A.G.) and PO1 AG03991 (J.C.M.)]. D.C.R. is a Wellcome Trust Senior Fellow in Clinical Science.

REFERENCES

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