

UCLA

UCLA Previously Published Works

Title

Decision tree analysis of genetic risk for clinically heterogeneous Alzheimer's disease

Permalink

<https://escholarship.org/uc/item/0k00g46m>

Journal

BMC Neurology, 15(1)

ISSN

1471-2377

Authors

Yokoyama, Jennifer S

Bonham, Luke W

Sears, Renee L

et al.

Publication Date

2015-12-01

DOI

10.1186/s12883-015-0304-6

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

RESEARCH ARTICLE

Open Access

Decision tree analysis of genetic risk for clinically heterogeneous Alzheimer's disease

Jennifer S Yokoyama^{1*}, Luke W Bonham^{1†}, Renee L Sears^{2†}, Eric Klein², Anna Karydas¹, Joel H Kramer¹, Bruce L Miller¹ and Giovanni Coppola²

Abstract

Background: Heritability of Alzheimer's disease (AD) is estimated at 74% and genetic contributors have been widely sought. The $\epsilon 4$ allele of *apolipoprotein E* (*APOE*) remains the strongest common risk factor for AD, with numerous other common variants contributing only modest risk for disease. Variability in clinical presentation of AD, which is typically amnesic (AmnAD) but can less commonly involve visuospatial, language and/or dysexecutive syndromes (atypical or AtAD), further complicates genetic analyses. Taking a multi-locus approach may increase the ability to identify individuals at highest risk for any AD syndrome. In this study, we sought to develop and investigate the utility of a multi-variant genetic risk assessment on a cohort of phenotypically heterogeneous patients with sporadic AD clinical diagnoses.

Methods: We genotyped 75 variants in our cohort and, using a two-staged study design, we developed a 17-marker AD risk score in a Discovery cohort ($n = 59$ cases, $n = 133$ controls) then assessed its utility in a second Validation cohort ($n = 126$ cases, $n = 150$ controls). We also performed a data-driven decision tree analysis to identify genetic and/or demographic criteria that are most useful for accurately differentiating all AD cases from controls.

Results: We confirmed *APOE* $\epsilon 4$ as a strong risk factor for AD. A 17-marker risk panel predicted AD significantly better than *APOE* genotype alone ($P < 0.00001$) in the Discovery cohort, but not in the Validation cohort. In decision tree analyses, we found that *APOE* best differentiated cases from controls only in AmnAD but not AtAD. In AtAD, *HFE* SNP rs1799945 was the strongest predictor of disease; variation in *HFE* has previously been implicated in AD risk in non- $\epsilon 4$ carriers.

Conclusions: Our study suggests that *APOE* $\epsilon 4$ remains the best predictor of broad AD risk when compared to multiple other genetic factors with modest effects, that phenotypic heterogeneity in broad AD can complicate simple polygenic risk modeling, and supports the association between *HFE* and AD risk in individuals without *APOE* $\epsilon 4$.

Keywords: Alzheimer's disease, Genetics, Decision tree analysis

Background

Alzheimer's disease (AD) is a devastating neurodegenerative disorder that results in memory impairment and can also involve deterioration of language, visuospatial and/or executive functioning abilities. As the world's population ages and the number of individuals with AD grows, it will become increasingly important to identify those at highest risk for AD during the earliest stages of—or prior to—disease.

Genetic predictors of AD hold strong potential for identifying those at risk of developing disease. Indeed, a large clinical study will launch in 2015 to assess the utility of AD therapies given to individuals at highest genetic risk for AD but who are still cognitively healthy [1]. These individuals, who carry the $\epsilon 4$ allele of *apolipoprotein E* (*APOE*), have a 2-10x increased risk for developing AD compared to non-carriers [2,3], but not all $\epsilon 4$ carriers go on to develop disease [3,4]. Despite the vast number of genetic studies of AD, which is estimated to be 74% heritable [5], no other common variants have been identified that confer as high a risk as *APOE* $\epsilon 4$. In rare cases, AD is familial, caused by an autosomal dominant mutation in *APP*, *PSEN1*, or *PSEN2* [6,7]. For sporadic late-onset AD

* Correspondence: Jennifer.Yokoyama@ucsf.edu

†Equal contributors

¹Memory and Aging Center, Department of Neurology, University of California, San Francisco, CA 94158, USA

Full list of author information is available at the end of the article

(LOAD), numerous common variants of very low effect (odds ratio [OR] ~ 1.1-1.3) have been identified through genome-wide association studies (GWAS) and replicated across multiple large [8], and diverse populations [9,10]. More recently, rare variants (<1% allele frequency) of larger effect size have also been identified as risk conferring (*TREM2* 0.3% [11], *PLD3* < 0.5% [12], *MAPT* 0.3% [13]) or protective against (*APP* 0.01% [14] to 0.62% [15]) AD.

In addition to genetic heterogeneity, there is also clinical heterogeneity in AD. The majority of patients present with amnesic syndromes (AmnAD) but approximately 6-14% of AD patients demonstrate atypical clinical syndromes (AtAD) [16]. These include 1) posterior cortical atrophy (PCA), characterized by predominant visuospatial deficits [17]; 2) the logopenic variant of primary progressive aphasia (lvPPA) [18], characterized by loss in phonologic short-term memory; and 3) dysexecutive/behavioral AD [16] characterized by loss of executive function and/or behavioral changes with retention of memory function.

Genetic and phenotypic heterogeneity strongly support the notion that multiple genetic variants of small effect contribute to disease susceptibility. A multi-locus approach may increase the ability to identify individuals at highest risk for any AD syndrome. The multi-locus approach has had modest success in LOAD, with polygenic risk scoring approaches associating better with LOAD diagnoses and age of onset than *APOE* genotype alone [19-21]. However, most studies have focused on clinically homogeneous groups with primary amnesic presentations.

In this study, we investigated two different strategies for polygenic risk assessment of clinically heterogeneous AD. First, we took a traditional approach and developed and assessed the utility of a multi-marker genetic risk score to predict AD. The risk score was based on a Discovery cohort association study that sought to replicate previous AD findings and assess additional candidate variants for their association with disease risk. The risk score was then tested for its predictive ability in a separate Validation cohort. Second, we used a more novel decision tree analysis [22] to identify genetic and demographic risk factors for AD. This data-driven method has been used in diverse clinical contexts [23-26] to predict binary outcomes, but is largely unutilized in the prediction of AD diagnosis. It allowed us to assess step-wise interactions between variables to identify the factors that best predict AD.

Methods

Participants

Individuals 65- to 101-years-old (N = 216 males, N = 232 females) were evaluated at the University of California, San Francisco Memory and Aging Center (UCSF MAC) and had genotype data available for analysis. All participants were unrelated Caucasians (confirmed by multi-dimensional scaling (MDS) plots or self-described for

those without GWAS data available). Non-Caucasians were excluded due to the insufficient number of participants and potential for confounding background genetics. All aspects of the study were approved by the UCSF Institutional Review Board and written informed consent was obtained from all participants and surrogates (as per UCSF Institutional Review Board protocol).

Clinical assessment

All participants underwent a multi-step screening process with an in-person visit at the MAC that included a neurologic exam, cognitive assessment [27], and medical history. Each participant's study partner was also interviewed regarding functional abilities. A multidisciplinary team composed of a neurologist, neuropsychologist, and nurse then reviewed all potential participants. Participants included in this study had a study partner (*i.e.*, spouse, close friend). The multidisciplinary team established clinical diagnoses for cases according to consensus criteria for AD [16]. Atypical or concomitant diagnoses were established for lvPPA [16,18], PCA syndrome [16,17], primary executive AD [16], vascular disease [28], or dementia with Lewy bodies (DLB) [29] according to consensus criteria. Individuals with primarily amnesic AD presentations were considered "AmnAD" and those with less common clinical syndromes (lvPPA, PCA, primary executive) or comorbidities (vascular disease, DLB) were considered as "AtAD". All control subjects underwent a similar multi-step screening process, including study partner interview and a consensus team of clinicians then reviewed all potential participants. Controls included in this study had Mini-Mental State Exam (MMSE) [30] scores ≥ 26 or a Clinical Dementia Rating Scale (CDR) [31] of 0, no participant or informant report of cognitive decline in the prior year, and no evidence from their screening visit suggesting a neurodegenerative disorder (per team neurologist's clinical judgment). Individuals harboring a known disease mutation were excluded from the study.

Genotypes

Genomic DNA was extracted from peripheral blood using standard protocols (Gentra PureGene Blood Kit, QIAGEN, Inc. – USA, Valencia, CA). Genotyping was performed using one of three platforms: TaqMan, Sequenom, or via array genotyping. The method used for each variant is provided in the Supplement (Additional file 1). TaqMan Allelic Discrimination Assay was used for *APOE* genotyping (rs429358 and rs7412) and others as noted, and was conducted on an ABI 7900HT Fast Real-Time PCR system (Applied Biosystems, Foster City, CA) according to manufacturer's instructions. Sequenom iPLEX Technology (Sequenom, San Diego, CA) was also used for genotyping a subset of variants as per manufacturer's instructions. The SpectroAquire and MassARRAY Typer

Software packages (Sequenom, San Diego, CA) were used for interpretation and Typer analyzer (v3.4.0.18) was used to review and analyze data. Only genotypes with “Conservative” or “Moderate” quality calls were included in analysis. A subset of genotypes was also obtained from the Illumina Omni1-Quad array genotyping platform (Illumina Inc., San Diego, CA), processed using manufacturer’s instructions.

A total of 75 variants were genotyped in all subjects and analyzed for association with AD risk. These variants are a culmination of different, on-going studies to evaluate the effect of genes involved in neurodegenerative disease, neurodevelopment, social function, behavior, neuropsychiatry, and language on diseases like AD and frontotemporal dementia (FTD). These included polymorphisms previously associated with: 1) risk for AD or other neurodegenerative disease; 2) neuropsychiatric phenotypes implicated in dementia risk (e.g., depression [32–34], dyslexia [27]; 3) cognitive protection [35]. A full list of variants, associated phenotypes, and accompanying references is provided in Additional file 1. Inclusion criteria for analyzed markers were: >80% non-missing genotypes, ≥ 0.01 minor allele frequency (MAF), and Hardy-Weinberg equilibrium (HWE) $P > 0.001$. The average call rate was 98% for all variants.

Analysis

Association study

The study cohort was divided into two groups, a first stage “Discovery” cohort for development of the AD risk score and a second stage “Validation” cohort with which to test the risk scoring method developed in the Discovery cohort. We first conducted association analysis of all markers meeting inclusion criteria in the Discovery cohort. Analyses were performed in PLINK as a logistic regression under an additive model [36].

Risk scoring

For scoring, we ranked all findings by p-value and then removed SNPs that were in linkage disequilibrium (LD, $r^2 > 0.8$) in our dataset; the single most strongly associated SNP of a set of linked markers was retained. Using the unlinked markers we created raw scoring files for each top finding, iteratively adding the next most significant finding to each scoring set (i.e., 1st marker in first set, 1st and 2nd markers in second set, etc.). Reference alleles were established in the scoring files such that all effects were in the same direction of conferring risk (e.g., a SNP with an empirical OR 0.1 for the reference minor allele would be switched such that the major allele was the reference allele for scoring). Using this paradigm, we created scoring sets for the top findings that were not in LD.

We implemented the ‘SNP scoring’ algorithm in PLINK to first assess the predictive ability of each score set (A–Z) in the Discovery dataset for evaluative purposes. We compared the risk scores for each set against the true

phenotypes using receiver operating characteristic (ROC) curves and used the resulting area under the curve (AUC) values to determine the optimal score set, with higher AUC values representing better sensitivity and specificity. The optimal score set was determined as follows. First, score sets were evaluated in two ways: 1) by simple consecutive comparisons of AUC values to identify the set at which AUC is largest, and 2) by statistical comparisons of a given set’s ROC curve AUC (AUC_i) versus the previous set’s ROC curve AUC (AUC_{i-1}) and versus the *APOE*-only score’s AUC (AUC_A). We then iteratively evaluated sets to determine the maximum AUC, stopping when two consecutive sets each resulted in decreases of AUC as compared to the previous set (i.e., $AUC_i > AUC_{i+1}$ & $AUC_i > AUC_{i+2}$). After determining this optimal set, we used the same scoring file to create risk scores for the Validation cohort and assessed the AUC of the resulting ROC curve to determine the generalization of our risk scoring method in an independent dataset. All ROC analyses were performed in Stata10/MP (StataCorp LP, College Station, TX).

Decision tree analysis

To explore and evaluate the diagnostic potential of the genetic variants available with ROC curves, we used the ROC4 software platform (ROC4.22.exe; <http://www.stanford.edu/~yesavage/ROC.html>). The software utilizes a user-set weight of sensitivity and specificity (kappa) to choose the predictive variable and value that best divides the sample. The sample is then divided on the value of the variable, which is most predictive based on this sensitivity and specificity. Following this, the program performs the same analysis amongst the subgroups created by the previous step. The process continues until a stopping rule is enforced. The output after stopping rules come into place is a “decision tree” which shows the variables and interactions between them in predicting the outcome of interest. We chose a kappa weight of 0.5 in order to balance efficiency (sensitivity and specificity were equally weighted). There were three stopping rules: when subgroup totals were less than 10, when a significance value corresponding to a multiple-testing-corrected X^2 test greater than $P = 0.01$ was reached, or when a three way interaction was reached. We performed three ROC analyses: one combined analysis of controls and all types of AD patients, one for the controls and AmnAD, and one for controls and AtAD. The ‘gold standard’ binary score was case/control outcome for any AD clinical diagnosis. Additional predictors included sex (0/1 for male/female), age (in years), and all genetic variants passing quality control (0/1/2 for dose of minor frequency allele).

Results

In total, $N = 185$ AD cases and $N = 283$ cognitively normal controls were included in the analysis. Demographics for

Table 1 Sample demographics

Discovery	AmnAD	AtAD	Control
N	42	17	133
Age (mean ± SD)	76.2 ± 7.9	73.3 ± 6.5	73.9 ± 6.2
% Female	42.9%	41.2%	57.1%
Education (years, mean ± SD)	16.3 ± 2.5	16.5 ± 2.9	17.3 ± 2.1
% APOE4 carrier	64.3%	41.2%	20.3%
Validation	AmnAD	AtAD	Control
N	84	22	150
Age (mean ± SD)	80.7 ± 8.4	80.9 ± 9.3	76.8 ± 7.4
% Female	53.6%	27.3%	53.3%
Education (years, mean ± SD)	15.8 ± 3.1	16.1 ± 4.8	17.4 ± 2.1
% APOE4 carrier	49.4%	36.4%	24.0%
Global	AmnAD	AtAD	Control
N	126	39	283
Age (mean ± SD)	79.2 ± 8.5	77.6 ± 9.0	75.4 ± 7.0
% Female	50.0%	33.3%	55.1%
Education (years, mean ± SD)	16.0 ± 2.9	16.7 ± 3.1	17.4 ± 2.1
% APOE4 carrier	55.2%	38.5%	22.3%

Demographic summary of amnesic Alzheimer's disease (AmnAD) cases, atypical AD (AtAD) cases and controls.

each group are shown in Table 1. A total of 192 (59 cases, 133 controls) individuals were in the first stage Discovery cohort and 276 (126 cases, 150 controls) were in the second stage Validation cohort. Of the Discovery cohort, 21.9% were AmnAD and 8.9% were AtAD (17 Total, 7 lvPPA, 3 PCA, 3 primarily executive AD, 2 AD with concomitant vascular disease, 2 AD with concomitant DLB; Figure 1). In the Validation cohort, 30.4% were AmnAD, and 8.0% were AtAD (22 Total, 7 lvPPA, 1 PCA, 13 AD with vascular disease, 1 AD with DLB).

Confirmation of AD risk variants and establishment of a 17-marker risk assessment

We first performed an association study in the Discovery cohort as a small-scale replication study of previously identified risk variants for AD in our clinically heterogeneous cohort. We then used this analysis to establish a ranked order by which we could iteratively add variants into a polygenic score to evaluate their utility for risk assessment. In our analysis, only the well-established *APOE* $\epsilon 4$ allele ($P = 1.36 \times 10^{-6}$), with an estimated OR = 4.28, met strict significance after Bonferroni correction for multiple testing (Table 2). Seven other variants had nominal p-values of $P < 0.05$. The second strongest association was with the rs1799945 SNP in *HFE* ($P = 1.64 \times 10^{-3}$,

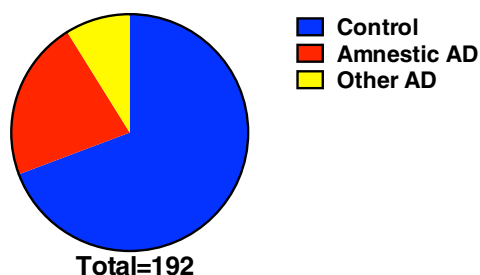
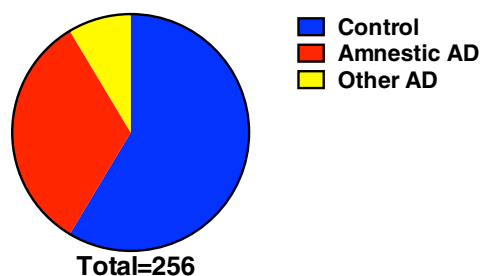
Discovery Cohort by Diagnosis**Replication Cohort by Diagnosis**

Figure 1 Diagnosis breakdown. Each cohort's composition by diagnosis is shown with controls in blue, amnesic AD in red and other (atypical) AD in yellow.

Table 2 Association results

Gene	SNP	OR	STAT	P	MAF
<i>APOE</i>	rs429358/rs7412	4.28	4.83	1.36E-06	0.20
<i>HFE</i>	rs1799945	2.83	3.15	1.64E-03	0.15
<i>PICALM</i>	rs3851179	0.47	-3.04	2.37E-03	0.42
<i>CR1</i>	rs6701713	0.42	-2.65	0.01	0.19
<i>SLC6A4</i>	rs2020942	1.81	2.63	0.01	0.40
<i>TPH1</i>	rs1799913	0.64	-2.02	0.04	0.44
<i>KIAA0319</i>	rs4504469	0.60	-2.02	0.04	0.35
<i>CDC42BPA</i>	rs1320490	1.63	1.93	0.05	0.20
<i>TMEM175</i>	rs6599389	0.41	-1.81	0.07	0.08
<i>SORL1</i>	rs2070045	1.63	1.74	0.08	0.22
<i>CNTNAP2</i>	rs17236239	0.66	-1.61	0.11	0.16
<i>ATP2C2</i>	rs8053211	1.45	1.56	0.12	0.43
<i>CD2AP</i>	rs9349407	1.48	1.55	0.12	0.29
<i>TPD52</i>	rs7814569	1.79	1.53	0.13	0.09
<i>COMT</i>	rs4680	0.72	-1.41	0.16	0.50
<i>C9ORF72</i>	rs3849942	1.42	1.38	0.17	0.24
<i>CPE</i>	rs11186856	0.33	-1.37	0.17	0.33
<i>SORL1</i>	rs12285364	0.42	-1.34	0.18	0.04
<i>RIT2</i>	rs4130047	0.70	-1.33	0.18	0.31
<i>MOBP</i>	rs1768208	0.72	-1.3	0.19	0.29

Top 20 association results in the Discovery cohort. Only *APOE* was significant after Bonferroni correction for multiple testing. OR – odds ratio; STAT – test statistic; MAF – minor allele frequency.

OR = 2.83). Variation in the hemochromatosis gene has previously been associated with AD in numerous large meta-analyses [37–39]. Two established risk factors for AD identified by GWAS were nominally associated in our study but with an opposite direction of association, rs3851179 in *PICALM* ($P = 2.37 \times 10^{-3}$, OR = 1.87) [40,41] and rs6701713 in *CRI* ($P = 0.01$, OR = 0.42) [40,42]. More novel AD risk candidates implicated by our study included rs2020942 ($P = 0.01$, OR = 1.81), a SNP tagging the variable number tandem repeat in the serotonin transporter gene, *SLC6A4*, most often associated with depression [43,44]; rs1799913 ($P = 0.04$, OR = 0.64) in *TPH1*, an established depression risk factor [45] that was recently associated with depression in AD [34]; rs4504469 ($P = 0.04$, OR = 0.60) in *KIAA0319*, which was associated with dyslexia [46]; and rs1320490 ($P = 0.05$, OR = 1.63) in *CDC42BPA*, previously associated with reading ability [47].

By iteratively adding genetic variants, we found that a risk score panel comprising 17 variants (“Q”) was the best predictor of AD status (Table 3; Figure 2). When evaluated alone, *APOE* genotype had modest predictive value for differentiating AD cases from controls. The 17-marker risk score had a significantly better AUC and was better at predicting AD risk than *APOE* alone ($P < 0.00001$; Figure 3).

Genetic risk score does not predict AD better than *APOE* in a separate cohort

When evaluated in the Validation cohort, the “Q” risk scoring method did not perform better than *APOE* alone (Table 4; Figure 3). The 17-marker gene score resulted in 65% maximal correct classification of individuals, with a limited sensitivity (54%) and specificity (73%; Figure 4). Removing excess AmnAD patients from the Validation group to better match the proportion of AtAD individuals in the Discovery cohort did not improve the performance of the multi-marker risk score (Additional file 2).

Decision tree analysis identifies genetic heterogeneity in amnestic versus atypical AD

We postulated that the clinical heterogeneity between the Discovery and Validation cohorts might be contributing to the failure of the 17-variant risk score to differentiate AD cases from controls better than *APOE* genotype alone. Under an alternative model, the genetic risk for AmnAD is different from that for AtAD. In order to identify genetic and/or demographic criteria that are most useful for accurately differentiating all AD cases from controls and to test whether AmnAD and AtAD share disease predictors or are distinct in their risk profiles, we performed data-driven decision tree analyses. We performed three analyses, one in all AD cases ($N = 165$) versus controls ($N = 283$), one with only

Table 3 Score set evaluation statistics

Score	AUC \pm SE	P-val vs. A	P-val vs. Prev.	Δ AUC
A	0.69 \pm 0.04	N/A	N/A	N/A
B	0.75 \pm 0.04	0.01	0.01	0.056
C	0.79 \pm 0.03	0.0018	0.02	0.038
D	0.81 \pm 0.03	0.00010	0.18	0.021
E	0.83 \pm 0.03	<0.00001	0.13	0.023
F	0.84 \pm 0.03	<0.00001	0.64	0.005
G	0.85 \pm 0.03	<0.00001	0.36	0.011
H	0.85 \pm 0.03	<0.00001	0.76	0.003
I	0.86 \pm 0.03	<0.00001	0.35	0.008
J	0.86 \pm 0.03	<0.00001	0.42	0.006
K	0.87 \pm 0.03	<0.00001	0.27	0.010
L	0.86 \pm 0.03	<0.00001	0.11	−0.011
M	0.87 \pm 0.03	<0.00001	0.35	0.005
N	0.86 \pm 0.03	<0.00001	0.55	−0.004
O	0.87 \pm 0.03	<0.00001	0.57	0.004
P	0.87 \pm 0.03	<0.00001	0.76	−0.001
Q	0.88 \pm 0.03	<0.00001	0.02	0.007
R	0.87 \pm 0.03	<0.00001	0.90	−0.001
S	0.87 \pm 0.03	<0.00001	0.82	−0.001
T	0.88 \pm 0.03	<0.00001	0.67	0.002
U	0.87 \pm 0.03	<0.00001	0.88	−0.001
V	0.88 \pm 0.03	<0.00001	0.30	0.005
W	0.88 \pm 0.03	<0.00001	0.27	0.005
X	0.89 \pm 0.03	<0.00001	0.89	0.001
Y	0.89 \pm 0.03	<0.00001	0.90	0.001
Z	0.89 \pm 0.02	<0.00001	0.25	0.007

Each score set was evaluated for AUC of the ROC curve to assess predictive ability. Score Q (in bold) was determined the best performing scoring set given the following two sets resulted in consecutively lower AUC values. This resulted in a final score set, Q that had a statistically significant better AUC than just *APOE* (set A) alone, $P < 0.00001$.

AmnAD ($N = 126$) versus controls, and one with AtAD ($N = 39$) versus controls.

In the analysis with all AD cases, carrying an *APOE* $\epsilon 4$ allele was the first differentiator of cases from controls (Figure 5). Amongst individuals carrying the $\epsilon 4$ risk allele, the next risk predictor was being ≥ 77 years old. Of these eldest individuals, the next differentiator was carrying one or more of the minor allele for rs4343 in *ACE*, an AD-risk gene [48,49]. The fourth differentiator of this subgroup was being homozygous for the major allele of rs8053211 in *ATP2C2*, a gene associated with dyslexia and other language traits [50,51], as carriers of one or two copies of the minor allele had a higher risk for diagnosis of AD. Using these predictors, the model had a predictive value positive (PVP) of 0.87, meaning that it correctly predicted a positive AD diagnosis 87% of the time. The sensitivity at this cut point was 0.71 and the

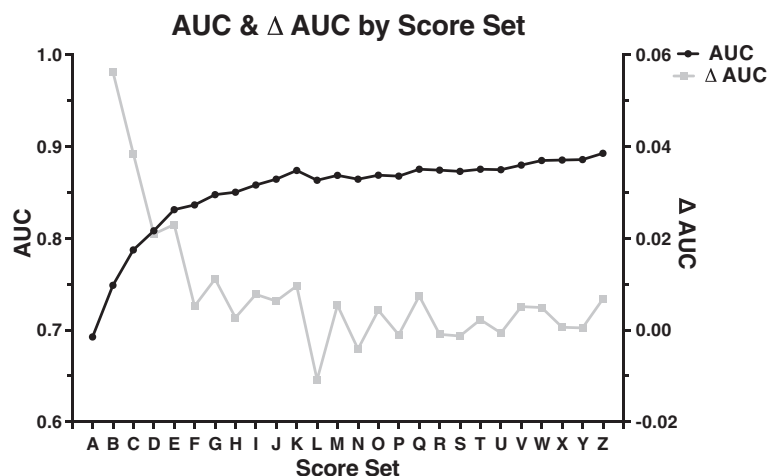


Figure 2 Area under the curve values for each scoring set. Area under the curve (AUC) values are provided for each scoring method (in black) as well as the difference between the current and previous scoring method (Δ AUC, in gray). Score sets were iteratively assessed until two consecutive resulting AUC values were lower than the preceding AUC value. The resulting scoring method [“Q”] had an AUC of 0.88.

specificity was 0.64 (Additional file 3). On the other side of the tree, in individuals carrying no $\epsilon 4$ alleles, the next differentiator of controls from cases was being <83 years old. Of these individuals, not carrying any of the *HFE* SNP, rs1799945, AD risk alleles was more predictive of control status. Finally, carrying two minor alleles of the *DCDC2* SNP rs1091047 (a dyslexia gene [52]) was most predictive of control status. In this final group, the model had a predictive value negative (PVN) of 0.92, meaning it correctly predicted a diagnosis of control 92% of the time. The sensitivity and specificity at this cut point were 0.64 and 0.73, respectively (Additional file 3).

In the analysis of AmnAD cases versus controls, carrying an *APOE* $\epsilon 4$ allele was also the best differentiator of cases from controls (Figure 6). Similar to the all-AD analysis, in individuals carrying the $\epsilon 4$ risk allele, the next

risk predictor was being ≥ 77 years old. Of these eldest individuals, the third differentiator was carrying one or more of the minor allele for rs4343 in *ACE*. In these individuals at this cut point, the PVP was 0.76. The sensitivity at this cut point was 0.83 and the specificity was 0.48. On the other side of the tree, in individuals carrying no $\epsilon 4$ alleles, the next differentiator of controls from cases was being between 66–87 years old. In these older individuals, there was another age differentiation whereby being 66–77 years old predicted control status. In this final group, the PVN was 0.92. The sensitivity and specificity at this cut point were 0.64 and 0.67, respectively.

The analysis of AtAD cases versus controls provided striking contrast to the previous analyses. In this cohort, carrying one or more minor alleles of the *HFE* SNP

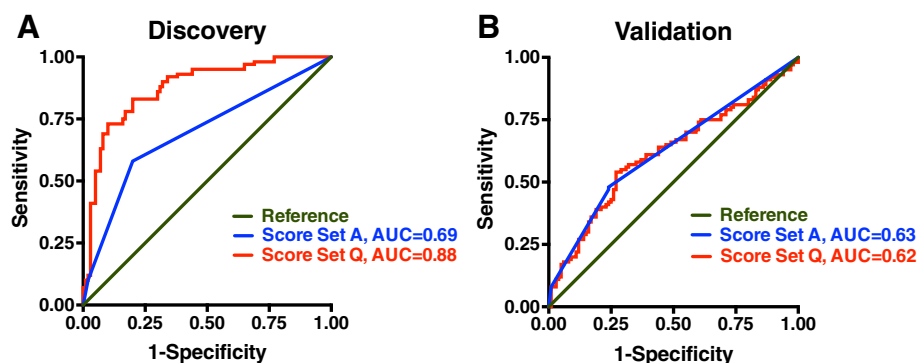


Figure 3 Receiver operating characteristic curves for Discovery and Validation cohorts. Receiver operating characteristic (ROC) curves are shown for scoring with *APOE* only (in blue) and the 17-marker risk score (“Q”, in red). AUC values of each curve are provided. (A) Discovery cohort shows a significant improvement in sensitivity/specificity as compared to *APOE* only, $P < 0.00001$. (B) Risk scoring with the 17-marker risk score in the Validation cohort does not increase prediction beyond *APOE* genotype alone.

Table 4 Risk scoring results for the Discovery and Validation cohorts

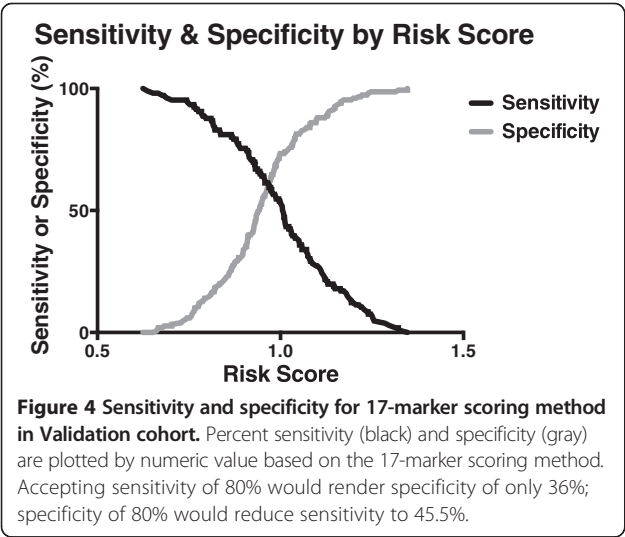
	Score	AUC ± SE	P-val vs. A	N
Discovery	A	0.69 ± 0.04	N/A	192
	Q	0.88 ± 0.03	<0.00001	
Validation	A	0.63 ± 0.03	N/A	256
	Q	0.62 ± 0.04	0.7345	

Evaluation metrics for the *APOE*-only risk score ("A") and the 17-marker risk score ("Q"). The 17-marker risk score predicted AD significantly better than *APOE* alone in the Discovery ($P < 0.00001$) but not the Validation cohort. Area under the curve (AUC) of the Receiver Operating Characteristic curves for A and Q and p-values of their comparisons are provided.

(rs1799945) was the first differentiator (Figure 7). In those with *HFE* risk alleles, the next differentiator was carrying ≥ 1 allele of the *GRN* variant, rs5848, which has been associated with risk for AD [53], hippocampal sclerosis [54,55], FTD [56], and bipolar disorder [57]. In the final at-risk group, the PVP was 0.47, with sensitivity and specificity of 0.62 and 0.74, respectively. On the other side of the tree, the next differentiator predicting control status was being homozygous for the minor allele of *GSK3B* SNP rs13312998, which has also been associated with AD and FTD [58]. At this cut point, the PVN was 0.93. The sensitivity and specificity were 0.43 and 0.87, respectively.

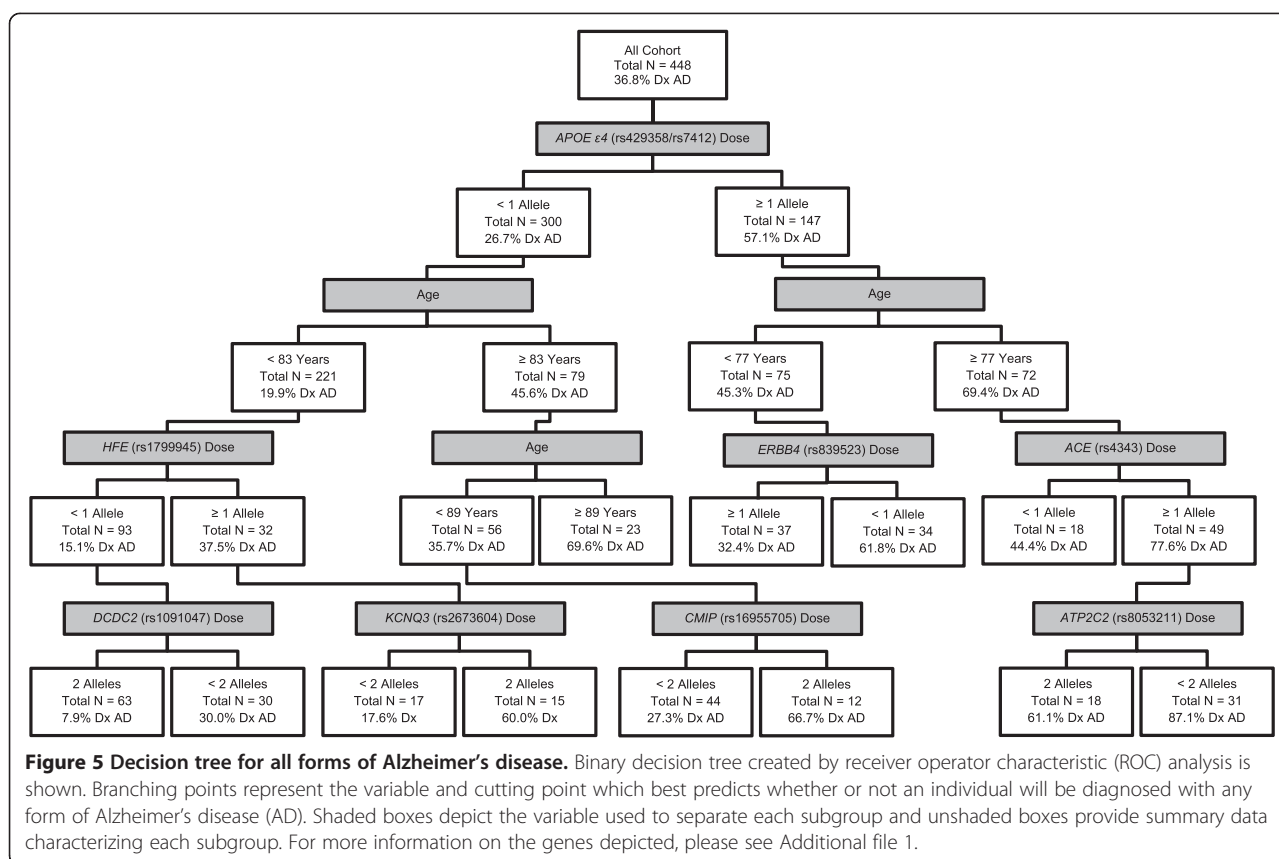
Discussion

In our association study, we found continuing support for *APOE*, *HFE*, *PICALM*, *CR1*, *SLC6A4*, *CDC42BP*, *TPH1*, and *KIAA0319* as genetic risk factors for AD. Using information from 17 variants combined into a genetic risk score allowed us to predict clinically heterogeneous AD cases significantly better than *APOE* genotype alone, supporting the role of these variants as predictors of AD risk in this primary Discovery group.



However, when we attempted to apply this polygenic risk assessment to an independent cohort of clinically heterogeneous AD patients for validation, the utility of analyzing 17 variants was not significantly better than analyzing *APOE* alone. Taken together, this suggests two things. First, it suggests that *APOE* $\epsilon 4$ remains the best predictor of AD risk, likely due to its strong effect, when compared to multiple other risk factors with very modest risk effects. Second, it suggests that phenotypic variability in AD complicates simple genetic risk modeling, particularly when co-morbidities are suspected.

The fact that *APOE* $\epsilon 4$ is the most predictive variant for amnesic AD but does not appear to be associated with risk for atypical AD syndromes such as PCA and lvPPA [59] likely contributes to the decreased specificity of the genetic risk assessment; namely, carrying an $\epsilon 4$ allele is associated with being affected in amnesic AD but is also associated with *not* being affected by PCA or lvPPA. Thus, *APOE* $\epsilon 4$ in the simple context of amnesic AD is quite adept at predicting who will be a case versus control, but is much less specific in the broader context of all AD syndromes, inclusive of atypical presentations and co-morbidities. Indeed, in our entire cohort of Discovery + Validation samples, *APOE* $\epsilon 4$ was significantly enriched in AmnAD but not AtAD cases when compared to controls (AmnAD vs Control $P = 3.08 \times 10^{-7}$; AtAD vs Control $P = 0.1$). A similar discrepancy due to clinical heterogeneity may also underlie our association of variants in *PICALM* and *CR1* in the opposite direction of historical findings. An alternate methodology to identify genetic and demographic factors that predict case/control status in AmnAD and AtAD separately was able to improve differentiation. Utilizing a decision tree methodology, we found that *APOE* best differentiated cases from controls only in AmnAD but not AtAD. In contrast, *HFE* genotype was the best differentiating factor between AtAD cases and controls; the same variant was also the first genetic risk factor for broad AD in individuals without *APOE* $\epsilon 4$. These findings are consistent with prior research implicating *HFE* in AD risk in individuals without *APOE* $\epsilon 4$ [60]. These results also suggest that atypical presentations could represent a distinct genetic class of AD, although the present study was not designed to specifically address this question. A recent study suggests that AtAD is more heritable than AmnAD [61], supporting the theory that there are additional genetic risk factors for AtAD that remain to be elucidated. In the future, GWAS of larger, more diverse cohorts of individuals with specific atypical phenotypes (e.g., PCA) could identify novel genetic risk factors specific to these AD syndromes. Phenotypic specificity in studies of amnesic AD may also provide additional statistical power to identify risk factors of small effect size.



In an effort to rule out the possibility of misdiagnosis, particularly in the AtAD group, we performed a *post hoc* chart review of patients for which pathological data was available ($N = 25$ AmnAD and $N = 8$ AtAD). All of these individuals had AD pathology cited as a primary ($N = 24$ AmnAD, $N = 5$ AtAD) or major contributing factor ($N = 1$ AmnAD, $N = 3$ AtAD) that correlated with each patient's clinical presentation (Additional file 4). Although not exhaustive, this data suggests that AD pathology was correctly recognized as a major contributor to patients' clinical syndrome in our patient cohort, and that the differential genetic risk profile of AtAD potentially influences its pathological heterogeneity when compared to AmnAD.

This study benefits from a two-staged discovery-validation study design, inclusion of a broad spectrum of clinical patients representing the phenotypic heterogeneity of AD, well-characterized cognitively normal controls, and inclusion of many of the most replicated genetic loci implicated in AD as well as several, more novel gene candidates. The main limitations of this study include the limited sample size, lack of pathological confirmation in all study participants, and the relatively young age of the controls. In addition, Caucasian individuals were the sole participants in our study, which potentially limits the scope of our findings. Co-morbid depression was not assessed in this analysis and may be

a contributing factor to the associations with the depression associated variants. This hypothesis requires direct testing in a separate study.

We implemented a decision tree analysis to identify genetic and demographic criteria most useful for accurately differentiating AD cases from controls. With an iterative, non-parametric approach, we used recursive partitioning to identify individuals according to a binary outcome of interest [22]. This method benefits from limiting the use of restrictive assumptions like linearity, additivity, and homoscedasticity, which are required by most linear models [23]. This approach has been used in a variety of clinical settings to identify variables of interest in predicting binary outcomes such as identification of AD patients who will have rapid cognitive decline [24], presence of tuberculosis after multiple conflicting tests [25], and ability to succeed in diabetes self-management programs [26]. Decision trees are amenable for use in a clinical setting, where an individual's risk for the outcome of interest—in this case, AD—can be estimated based on multiple predictive variables that follow a logical progression. Testing whether the factors identified in our decision tree analyses have predictive value in a larger, independent cohort will be critical for elucidating whether this risk assessment has clinical utility, particularly with the inclusion of

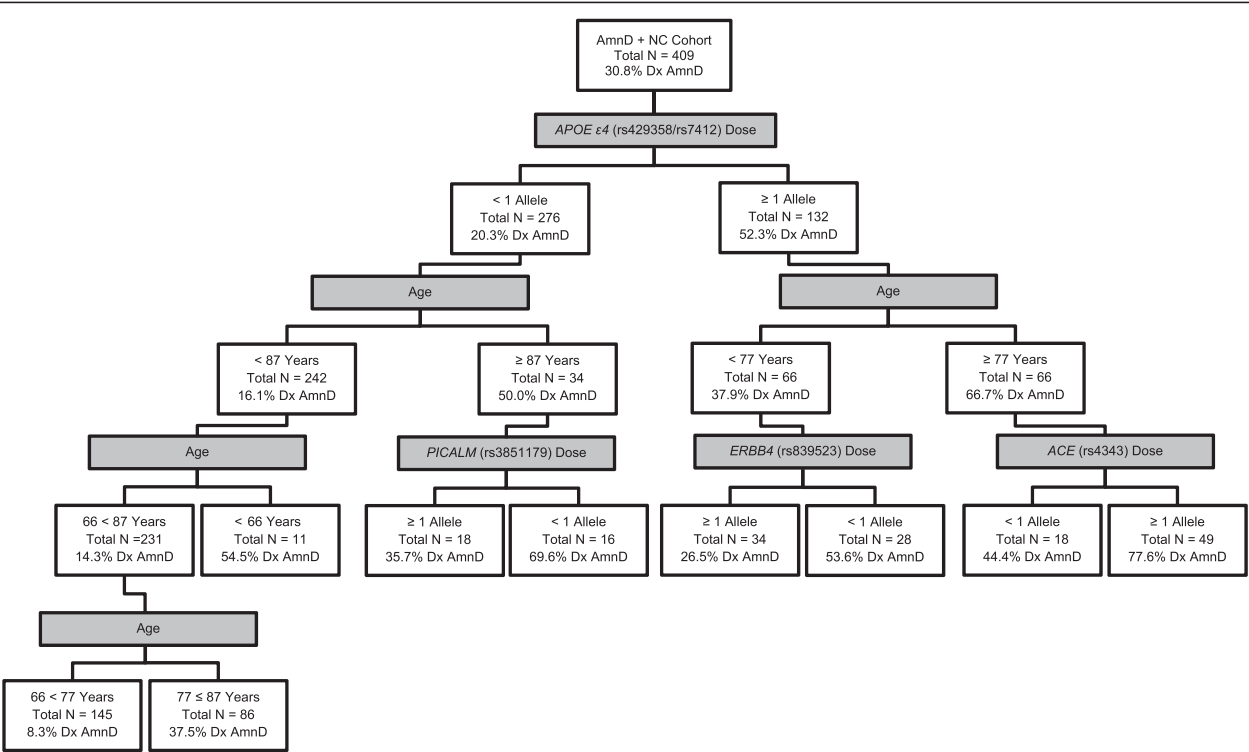


Figure 6 Decision tree for amnestic Alzheimer's disease. Binary decision tree created by receiver operator characteristic (ROC) analysis is shown. Branching points represent the variable and cutting point which best predicts whether or not an individual will be diagnosed with an amnestic form of Alzheimer's disease (AmnD). Shaded boxes depict the variable used to separate each subgroup and unshaded boxes provide summary data characterizing each subgroup. For more information on the genes depicted, please see Additional file 1. NC - normal control.

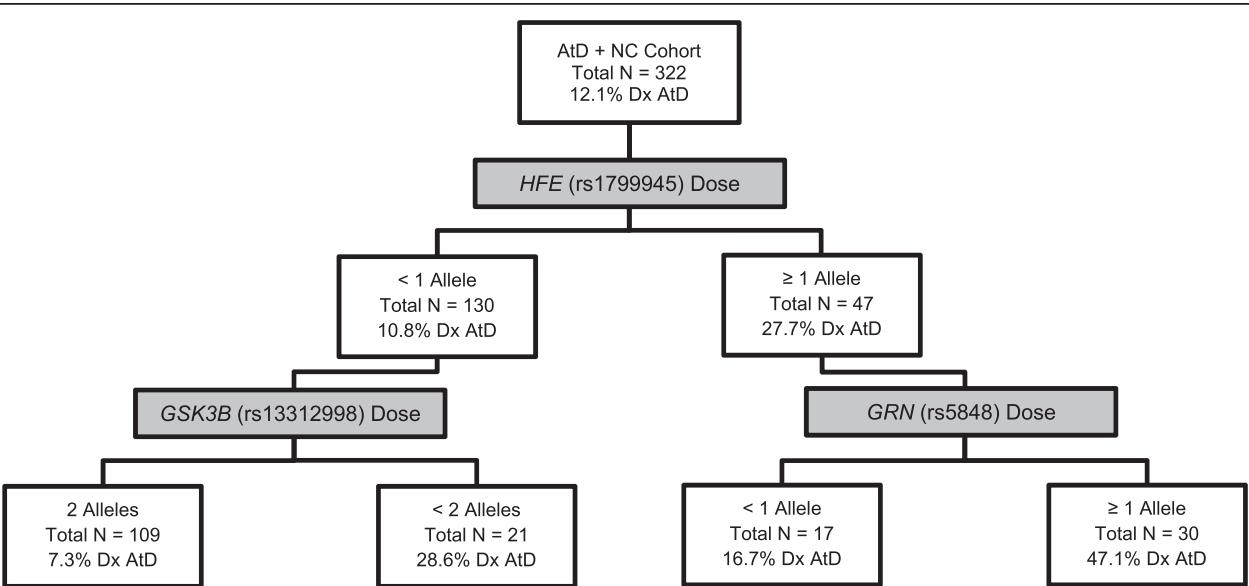


Figure 7 Decision tree for atypical Alzheimer's disease. Binary decision tree created by receiver operator characteristic (ROC) analysis is shown. Branching points represent the variable and cutting point which best predicts whether or not an individual will be diagnosed with an atypical form of Alzheimer's disease (AtD). Shaded boxes depict the variable used to separate each subgroup and unshaded boxes provide summary data characterizing each subgroup. For more information on the genes depicted, please see Additional file 1. NC - normal control.

pathologically confirmed cases and exclusion of amyloid-positive 'controls.'

Conclusions

We found that *APOE* genotype is the best predictor of risk compared to a polygenic risk score when assessing groups of clinically heterogeneous AD patients versus healthy older controls. In decision tree analysis, we found that AmnAD and AtAD have differential genetic risk factors, which may account for the inaccuracy of the traditional polygenic scoring method. Identifying individuals at highest genetic risk for AD could potentially allow for earlier diagnosis and intervention, allowing the opportunity to intervene with pathological processes and/or provide support prior to clinical onset of symptoms. These risk assessments will benefit from future work to characterize genetic risk factors of clinically homogeneous subtypes of AD in large, diverse populations.

Additional files

Additional file 1: Variants and Genotyping Methods.

Additional file 2: Results for Three Different Versions of the Validation Cohort.

Additional file 3: Decision Tree Analysis Results.

Additional file 4: Pathological Diagnoses for Selected Participants.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JSY participated in the design and coordination of the study, conducted statistical analyses and drafted the manuscript. LWB performed the decision tree analysis and drafted the manuscript. RLS performed the risk scoring and carried out genotyping. EK participated in sample processing and carried out genotyping. AK participated in data analysis and sample processing. JHK participated in sample coordination and interpretation of results. BLM participated in sample coordination and interpretation of results. GC conceived of the study, participated in its coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

J.S.Y. was funded by the Larry L. Hillblom Foundation (2012-A-015-FEL) and a diversity supplement from the NIA-NIH (P50-AG023501-08S1, PI: Miller, BL). Additional support was provided by NIH grants P50-AG023501 (BLM) and RC1 AG035610 and R01 AG26938 (G.C.), the Larry L. Hillblom Foundation (BLM), and the John Douglas French Alzheimer's Foundation (G.C.). We acknowledge the support of the NINDS Informatics Center for Neurogenetics and Neurogenomics (P30 NS062691). Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank Dr. Jerome Yesavage and Art Noda for technical advice on the decision tree analysis. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible.

Author details

¹Memory and Aging Center, Department of Neurology, University of California, San Francisco, CA 94158, USA. ²Semel Institute for Neuroscience and Human Behavior, Departments of Neurology and Psychiatry, The David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA 90095, USA.

Received: 11 February 2015 Accepted: 12 March 2015

Published online: 28 March 2015

References

- Banner Alzheimer's Institute Announces Partnership with Novartis in New Study of Alzheimer's Prevention Treatments. Phoenix: Banner Alzheimer's Institute; 2014. p. 1–3.
- Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA*. 1997;278:1349–56.
- Devanand DP, Pelton GH, Zamora D, Liu X, Tabert MH, Goodkind M, et al. Predictive utility of apolipoprotein E genotype for Alzheimer disease in outpatients with mild cognitive impairment. *Arch Neurol*. 2005;62:975–80.
- Petersen RC. Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. *JAMA*. 1995;273:1274–8.
- Gatz M, Pedersen N. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol Med Sci*. 1997;52:117–25.
- Campion D, Dumanchin C, Hannequin D, Dubois B, Belliard S, Puel M, et al. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *Am J Hum Genet*. 1999;65:664–70.
- Avramopoulos D. Genetics of Alzheimer's disease: recent advances. *Genome Med*. 2009;1:34.
- European Alzheimer's Disease I, Genetic, Environmental Risk in Alzheimer's D, Alzheimer's Disease Genetic C, Cohorts for H, Aging Research in Genomic E, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45:1452–8.
- Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang L, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ε4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA*. 2013;309:1483–92.
- Tang MX, Maestre G, Tsai WY, Liu XH, Feng L, Chung WY, et al. Relative risk of Alzheimer disease and age-at-onset distributions, based on APOE genotypes among elderly African Americans, Caucasians, and Hispanics in New York City. *Am J Hum Genet*. 1996;58:574–84.
- Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med*. 2013;368:117–27.
- Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature*. 2014;505:550–4.
- Coppola G, Chinnathambi S, Lee JJ, Dombroski BA, Baker MC, Soto-Ortolaza AI, et al. Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. *Hum Mol Genet*. 2012;21:3500–12.
- Wang L-S, Naj AC, Graham RR, Crane PK, Kunkle BW, Cruchaga C, et al. Rarity of the Alzheimer Disease-Protective APP A673T Variant in the United States. *JAMA Neurol*. 2014;19104:209–16.
- Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature*. 2012;488:96–9.
- Dubois B, Feldman H, Jacova C. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol*. 2014;13:614–29.
- Crutch S, Schott J, Rabinovici G. Shining a light on posterior cortical atrophy. *Alzheimers Dement*. 2013;9:463–5.
- Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. *Neurology*. 2011;76:1006–14.
- Sabuncu MR, Buckner RL, Smoller JW, Lee PH, Fischl B, Sperling RA. The association between a polygenic Alzheimer score and cortical thickness in clinically normal subjects. *Cereb Cortex*. 2012;22:2653–61.
- Marden JR, Walter S, Tchetgen Tchetgen EJ, Kawachi I, Glymour MM. Validation of a polygenic risk score for dementia in black and white individuals. *Brain Behav*. 2014;4:687–97.
- Naj AC, Jun G, Reitz C, Kunkle BW, Perry W, Park YS, et al. Effects of Multiple Genetic Loci on Age at Onset in Late-Onset Alzheimer Disease A Genome-Wide Association Study. *JAMA Neurol*. 2014;71:1394–404.
- Kraemer HC. Evaluating Medical Tests: Objective and Quantitative Guidelines. Thousand Oaks: SAGE Publications, Inc; 1992.
- Ong J, Kuo T, Manber R. Who is at risk for dropout from group cognitive-behavior therapy for insomnia? *J Psychosom Res*. 2008;64:419–25.

24. O'Hara R, Thompson JM, Kraemer HC, Fenn C, Taylor JL, Ross L, et al. Which Alzheimer Patients Are at Risk for Rapid Cognitive Decline? *J Geriatr Psychiatry Neurol.* 2002;15:233–8.
25. Thanassi W, Noda A, Hernandez B, Newell J, Terpeluk P, Marder D, et al. Delineating a retesting zone using receiver operating characteristic analysis on serial quantiFERON tuberculosis test results in US healthcare workers. *Pulm Med.* 2012;2012:1–7.
26. Glasgow RE, Strycker LA, King DK, Toobert DJ. Understanding who benefits at each step in an internet-based diabetes self-management program: application of a recursive partitioning approach. *Med Decis Making.* 2014;34:180–91.
27. Miller ZA, Mandelli ML, Rankin KP, Henry ML, Babiak MC, Frazier DT, et al. Handedness and language learning disability differentially distribute in progressive aphasia variants. *Brain.* 2013;136:3461–73.
28. Román G, Tatemichi T, Erkinjuntti T. Vascular dementia Diagnostic criteria for research studies: Report of the NINDS-AIREN International Workshop*. *Neurology.* 1993;43:250–60.
29. McKeith I, Dickson D, Lowe J, Emre M. Diagnosis and management of dementia with Lewy bodies third report of the DLB consortium. *Neurology.* 2005;65:1863–72.
30. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12:189–98.
31. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology.* 1993;43:241–4.
32. Gennatas ED, Cholfin JA, Zhou J, Crawford RK, Sasaki DA, Karydas A, et al. COMT Val158Met genotype influences neurodegeneration within dopamine-innervated brain structures. *Neurology.* 2012;78:1663–9.
33. Zuccato C, Cattaneo E. Brain-derived neurotrophic factor in neurodegenerative diseases. *Nat Rev Neurol.* 2009;5:311–22.
34. Arlt S, Demiralay C, Tharun B, Geisel O, Storm N, Eichenlaub M, et al. Genetic risk factors for depression in Alzheimer's disease patients. *Curr Alzheimer Res.* 2013;10:72–81.
35. Dubal DB, Yokoyama JS, Zhu L, Broestl L, Worden K, Wang D, et al. Life Extension Factor Klotho Enhances Cognition. *Cell Rep.* 2014;7:1065–76.
36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
37. Lin M, Zhao L, Fan J, Lian X-G, Ye J-X, Wu L, et al. Association between HFE polymorphisms and susceptibility to Alzheimer's disease: a meta-analysis of 22 studies including 4,365 cases and 8,652 controls. *Mol Biol Rep.* 2012;39:3089–95.
38. Moalem S, Percy M, Andrews D, Kruck T, Wong S, Dalton A, et al. Are Hereditary Hemochromatosis Mutations Involved in Alzheimer Disease? *Am J Med Genet.* 2000;93:58–66.
39. Sampietro M, Caputo L, Casatta A, Meregalli M, Pellagatti A, Tagliabue J, et al. The hemochromatosis gene affects the age of onset of sporadic Alzheimer's disease. *Neurobiol Aging.* 2001;22:563–8.
40. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet.* 2009;41:1088–93.
41. Jun G, Naj AC, Beecham GW, Wang L-S, Buross J, Gallins PJ, et al. Meta-analysis confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. *Arch Neurol.* 2010;67:1473–84.
42. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buross J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet.* 2011;43:436–41.
43. Lazary J, Lazary A, Gonda X, Benko A, Molnar E, Juhasz G, et al. New evidence for the association of the serotonin transporter gene (SLC6A4) haplotypes, threatening life events, and depressive phenotype. *Biol Psychiatry.* 2008;64:498–504.
44. Su S, Zhao J, Bremner J. Serotonin transporter gene, depressive symptoms, and interleukin-6. *Circ Cardiovasc Genet.* 2009;2:614–20.
45. Gizatullin R, Zaboli G, Jönsson EG, Asberg M, Leopardi R. Haplotype analysis reveals tryptophan hydroxylase (TPH) 1 gene variants associated with major depression. *Biol Psychiatry.* 2006;59:295–300.
46. Velayos-Baeza A, Toma C, da Roza S, Paracchini S, Monaco AP. Alternative splicing in the dyslexia-associated gene KIAA0319. *Mamm Genome.* 2007;18:627–34.
47. Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100 K SNP microarrays in a sample of 5760 children. *Mol Psychiatry.* 2008;13:729–40.
48. Lehmann DJ, Cortina-Borja M, Warden DR, Smith AD, Slegers K, Prince JA, et al. Large meta-analysis establishes the ACE insertion-deletion polymorphism as a marker of Alzheimer's disease. *Am J Epidemiol.* 2005;162:305–17.
49. Kehoe PG, Katzov H, Feuk L, Bennet AM, Johansson B, Wilman B, et al. Haplotypes extending across ACE are associated with Alzheimer's disease. *Hum Mol Genet.* 2003;12:859–67.
50. Newbury DF, Winchester L, Addis L, Paracchini S, Buckingham LL, Clark A, et al. CMIP and ATP2C2 Modulate Phonological Short-Term Memory in Language Impairment. *Am J Hum Genet.* 2009;85:264–72.
51. Lesch KP, Timmesfeld N, Renner TJ, Halperin R, Röser C, Nguyen TT, et al. Molecular genetics of adult ADHD: Converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm.* 2008;115:1573–85.
52. Scerri TS, Morris AP, Buckingham L-L, Newbury DF, Miller LL, Monaco AP, et al. DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biol Psychiatry.* 2011;70:237–45.
53. Sheng J, Su L, Xu Z, Chen G. Progranulin polymorphism rs5848 is associated with increased risk of Alzheimer's disease. *Gene.* 2014;542:141–5.
54. Pao WC, Dickson DW, Crook JE, Finch NA, Rademakers R, Graff-Radford NR. Hippocampal sclerosis in the elderly: genetic and pathologic findings, some mimicking Alzheimer disease clinically. *Alzheimer Dis Assoc Disord.* 2011;25:364–8.
55. Dickson DW, Baker M, Rademakers R. Common variant in GRN is a genetic risk factor for hippocampal sclerosis in the elderly. *Neurodegener Dis.* 2010;7:170–4.
56. Rademakers R, Eriksen JL, Baker M, Robinson T, Ahmed Z, Lincoln SJ, et al. Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia. *Hum Mol Genet.* 2008;17:3631–42.
57. Galimberti D, Prunas C, Paoli RA, Dell'Osso B, Fenoglio C, Villa C, et al. Progranulin gene variability influences the risk for bipolar I disorder, but not bipolar II disorder. *Bipolar Disord.* 2014;16:769–72.
58. Schaffer BAJ, Bertram L, Miller BL, Mullin K, Weintraub S, Johnson N, et al. Association of GSK3B with Alzheimer disease and frontotemporal dementia. *Arch Neurol.* 2008;65:1368–74.
59. Mesulam M-M. Primary progressive aphasia—a language-based dementia. *N Engl J Med.* 2003;349:1535–42.
60. Percy M, Somerville MJ, Hicks M, Garcia A, Colelli T, Wright E, et al. Risk factors for development of dementia in a unique six-year cohort study. I. An exploratory, pilot study of involvement of the E4 allele of apolipoprotein E, mutations of the hemochromatosis-HFE gene, type 2 diabetes, and stroke. *J Alzheimers Dis.* 2014;38:907–22.
61. Po K, Leslie FVC, Gracia N, Bartley L, Kwok JBJ, Halliday GM, et al. Heritability in frontotemporal dementia: more missing pieces? *J Neurol.* 2014;261:2170–7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

