Risk prediction for sporadic Alzheimer’s disease using genetic risk score in the Han Chinese population

Qianyi Xiao1,*, Zhi-Jun Liu2,3,*, Sha Tao1,*, Yi-Min Sun3, Deke Jiang4, Hong-Lei Li2, Haitao Chen1, Xu Liu4, Brittany Lapin5, Chi-Hsiung Wang5, S. Lilly Zheng4,5, Jianfeng Xu1,4,5,6,* and Zhi-Ying Wu2,*

1 Center for Genomic Translational Medicine and Prevention, School of Public Health, Fudan University, Shanghai, China
2 Department of Neurology and Research Center of Neurology in Second Affiliated Hospital, and the Collaborative Innovation Center for Brain Science, Zhejiang University School of Medicine, Hangzhou, China
3 Department of Neurology and Institute of Neurology, Huashan Hospital, Fudan University, Shanghai, China
4 State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China
5 Program for Personalized Cancer Care, NorthShore University Health System, Evanston, IL, USA
6 Fudan Institute of Urology, Huashan Hospital, Fudan University, Shanghai, China

* These authors have contributed equally to this work

Correspondence to: Zhi-Ying Wu, email: zhiyingwu@zju.edu.cn
Correspondence to: Jianfeng Xu, email: jxu@NorthShore.org

Keywords: Alzheimer’s disease, genetic risk score, risk prediction, single nucleotide polymorphism, association, Gerotarget

Received: September 10, 2015 Accepted: September 22, 2015 Published: November 02, 2015

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

More than 30 independent single-nucleotide polymorphisms (SNPs) have been associated with Alzheimer’s disease (AD) risk by genome-wide association studies (GWAS) in European. We aimed to confirm these SNPs in Chinese Han and investigate the utility of these genetic markers. We randomly divided 459 sporadic AD (SAD) patients and 751 cognitively normal controls into two sets (discovery and testing). Thirty-three SAD risk-associated SNPs were firstly tested in the discovery set. Significant SNPs were used to calculate genetic risk score (GRS) in the testing set. Predictive performance of GRS was evaluated using the area under the receiver operating characteristic curve (AUC). In the discovery set, 6 SNPs were confirmed ($P = 7.87 \times 10^{-11} \sim 0.048$), including rs9349407 in CD2AP, rs11218343 in SORL1, rs17125944 in FERMT2, rs6859 in PVRL2, rs157580 and rs2075650 in TOMM40. The first three SNPs were associated with SAD risk independent of APOE genotypes. GRS based on these three SNPs were significantly associated with SAD risk in the independent testing set ($P = 0.002$). The AUC for discriminating cases from controls was 0.58 for GRS, 0.60 for APOE, and 0.64 for GRS and APOE. Our data demonstrated that GRS based on AD risk-associated SNPs may supplement APOE for better assessing individual risk for AD in Chinese.

INTRODUCTION

Alzheimer’s disease (AD) is a neurodegenerative disorder characterized with progressive deterioration in cognition and behavior. AD is the most common form of dementia in aging population with a worldwide prevalence of 35.6 millions in 2010, and is expected to increase to 115.4 millions by 2050 [1]. In China, the burden of AD increased considerably in recent years due to aging population. The incidence of AD among people aged 60 years or older was 6.25 cases per 1000 person-years in 2010 in China [2]. Because several risk factors have been associated with AD risk [3], the disease may be preventable by reducing these risk factors. It is widely believed that targeted prevention for subjects with higher risk for AD is likely a more effective strategy.
AD is highly heritable and its heritability is estimated up to 76% [4]. Previous work suggested that genetic variants play an important role in the development of the disease. Mutations in APP, PSEN1 and PSEN2 lead to early onset familial AD [5-8]. APOE ε4 allele has been found to be the strongest risk factor for sporadic AD (SAD), the most common form of AD including early- and late-onset SAD [9-12]. However, because about 40-50% of SAD do not carry the APOE ε4 allele [11, 13], additional genetic variants that are related to SAD risk likely exist. Several genome-wide association studies (GWAS) in European descent have identified a number of independent AD risk-associated single nucleotide polymorphisms (SNPs) [14-19]. A recent meta-analysis combined four of these AD GWAS samples identified additional 11 independent AD susceptibility SNPs [20]. To date, 11 of these AD risk-associated SNPs (in 9 genes) have been reported to be significantly associated with AD risk in the Han Chinese population [21-27].

Although these AD risk-associated SNPs have a modest effect size individually (odds ratio [OR] of each individual risk allele is typically < 1.3), it is hypothesized that these SNPs may confer a stronger cumulative effect to AD. A genetic risk score (GRS) that captures the cumulative effect of SNPs has been comprehensively studied to stratify individual risk in several complex diseases [28-31]. In this study, we aimed to first determine which AD risk-associated SNPs reported in European descent are associated with SAD risk in Han Chinese in a discovery set, and then to calculate GRS using these implicated SNPs and assess its discriminative performance in a testing set.

**RESULTS**

**Key demographic and clinical information of study subjects**

After quality control analyses, 1210 subjects retained in the study, including 459 SAD patients and 751 control subjects. Key demographic and clinical information for these subjects is presented in Table 1. Because the study was frequency matched for sex, no statistically significant difference in proportion of gender was found ($P > 0.05$). However, due to the frequency match for age (within 5-years), the mean age at examination was slightly, but statistically significantly younger in cases (71.2 years) than controls (72.7 years), $P = 0.004$. The age at onset in cases ranged from 45 to 88 years, with a mean of 68.5 years. About 35% SAD cases were early age onset (< 65 years). The mean MMSE score was significantly lower in cases (14.7) than in controls (25.1), $P < 0.001$. Similarly, APOE ε4 carrier rate was significantly higher in cases (41.4%) than in controls (20.0%), $P < 0.001$.

These subjects were randomly assigned to the discovery set (232 cases and 373 controls) and testing set (227 cases and 378 controls). As shown in Table S1, there was no significant difference between the case subjects from the discovery and testing sets ($P > 0.05$), except that mean age at onset was slightly older in the discovery set (71.2 years) than in the testing set (67.6 years), $P = 0.048$.

Among 33 SNPs selected in the study, 3 SNPs were excluded due to genotyping failure (rs10838725 and minor allele frequency (MAF) < 0.01 (rs7274581 and rs12989701). None of the SNPs significantly deviated

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>SAD (n=459)</th>
<th>Controls (n=751)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at examination, mean ± SD, yr</td>
<td>71.2±9.6</td>
<td>72.7 ± 5.9</td>
<td>0.004</td>
</tr>
<tr>
<td>Age at onset, mean ± SD, yr</td>
<td>68.5±9.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at onset &lt;65, n (%)</td>
<td>161(35.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at onset ≥65, n (%)</td>
<td>298(64.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex*, n (%)</td>
<td></td>
<td></td>
<td>0.407</td>
</tr>
<tr>
<td>Male</td>
<td>228(49.7)</td>
<td>354(47.1)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>231(50.3)</td>
<td>397(52.9)</td>
<td></td>
</tr>
<tr>
<td>MMSE score, mean±SD</td>
<td>14.7±6.6</td>
<td>25.1±3.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE, n (%)</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>APOE ε4 carriers</td>
<td>190 (41.4)</td>
<td>150 (20.0)</td>
<td></td>
</tr>
<tr>
<td>Non-APOE ε4 carriers</td>
<td>269 (58.6)</td>
<td>601 (80.0)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: SAD, Sporadic Alzheimer’s disease; MMSE, Mini-mental state examination; SD, Standard deviation.

* Frequency matched
Table 2: Association of sporadic Alzheimer’s disease risk with candidate SNPs reported in European descent among Chinese subjects in the discovery set.

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>Position</th>
<th>Genes</th>
<th>Reported risk allele</th>
<th>Allele frequency</th>
<th>Association testb</th>
<th>Association testc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs6656401</td>
<td>207,692,049</td>
<td>CR1</td>
<td>A</td>
<td>0.028</td>
<td>0.021</td>
<td>1.11(0.51-2.39)</td>
</tr>
<tr>
<td>1</td>
<td>rs3818361</td>
<td>207,784,968</td>
<td>CR1</td>
<td>A</td>
<td>0.354</td>
<td>0.378</td>
<td>0.93(0.73-1.19)</td>
</tr>
<tr>
<td>2</td>
<td>rs7561528</td>
<td>127,889,637</td>
<td>BIN1</td>
<td>A</td>
<td>0.123</td>
<td>0.151</td>
<td>0.82(0.58-1.16)</td>
</tr>
<tr>
<td>2</td>
<td>rs744373</td>
<td>127,894,615</td>
<td>BIN1</td>
<td>G</td>
<td>0.366</td>
<td>0.376</td>
<td>0.97(0.76-1.24)</td>
</tr>
<tr>
<td>2</td>
<td>rs35349669</td>
<td>234,068,476</td>
<td>INPP5D</td>
<td>T</td>
<td>0.013</td>
<td>0.011</td>
<td>1.18(0.39-3.52)</td>
</tr>
<tr>
<td>5</td>
<td>rs190982</td>
<td>88,223,420</td>
<td>MEF2C</td>
<td>A</td>
<td>0.864</td>
<td>0.849</td>
<td>1.18(0.84-1.68)</td>
</tr>
<tr>
<td>6</td>
<td>rs9271192</td>
<td>32,578,530</td>
<td>HLA-DRB5-HLA-DRB1</td>
<td>C</td>
<td>0.144</td>
<td>0.112</td>
<td>1.30(0.91-1.85)</td>
</tr>
<tr>
<td>6</td>
<td>rs9349407</td>
<td>47,453,378</td>
<td>CD2AP</td>
<td>C</td>
<td>0.202</td>
<td>0.118</td>
<td>1.95(1.39-2.74)</td>
</tr>
<tr>
<td>6</td>
<td>rs11754661</td>
<td>151,207,078</td>
<td>MTHFD1L</td>
<td>A</td>
<td>0.033</td>
<td>0.027</td>
<td>1.19(0.58-2.41)</td>
</tr>
<tr>
<td>7</td>
<td>rs2718058</td>
<td>37,841,534</td>
<td>NME8</td>
<td>A</td>
<td>0.754</td>
<td>0.797</td>
<td>0.79(0.60-1.05)</td>
</tr>
<tr>
<td>7</td>
<td>rs1476679</td>
<td>100,004,446</td>
<td>ZCWPW1</td>
<td>T</td>
<td>0.708</td>
<td>0.705</td>
<td>1.07(0.83-1.39)</td>
</tr>
<tr>
<td>7</td>
<td>rs11767557</td>
<td>143,109,139</td>
<td>EPHA1</td>
<td>T</td>
<td>0.835</td>
<td>0.873</td>
<td>0.74(0.53-1.03)</td>
</tr>
<tr>
<td>7</td>
<td>rs11771145</td>
<td>143,110,762</td>
<td>EPHA1</td>
<td>G</td>
<td>0.507</td>
<td>0.463</td>
<td>1.12(0.88-1.42)</td>
</tr>
<tr>
<td>8</td>
<td>rs28834970</td>
<td>27,195,121</td>
<td>PTK2B</td>
<td>C</td>
<td>0.276</td>
<td>0.302</td>
<td>0.88(0.67-1.14)</td>
</tr>
<tr>
<td>8</td>
<td>rs11136000</td>
<td>27,464,519</td>
<td>CLU</td>
<td>C</td>
<td>0.832</td>
<td>0.806</td>
<td>1.18(0.87-1.60)</td>
</tr>
<tr>
<td>8</td>
<td>rs569214</td>
<td>27,487,790</td>
<td>CLU</td>
<td>G</td>
<td>0.489</td>
<td>0.505</td>
<td>0.97(0.76-1.24)</td>
</tr>
<tr>
<td>8</td>
<td>rs983392</td>
<td>59,923,508</td>
<td>MS4A6A</td>
<td>A</td>
<td>0.959</td>
<td>0.972</td>
<td>0.77(0.41-1.43)</td>
</tr>
<tr>
<td>8</td>
<td>rs610932</td>
<td>59,939,307</td>
<td>MS4A6A</td>
<td>G</td>
<td>0.667</td>
<td>0.646</td>
<td>1.11(0.87-1.43)</td>
</tr>
<tr>
<td>8</td>
<td>rs4938933</td>
<td>60,034,429</td>
<td>MS4A4A</td>
<td>T</td>
<td>0.754</td>
<td>0.726</td>
<td>1.15(0.88-1.51)</td>
</tr>
<tr>
<td>8</td>
<td>rs2373115</td>
<td>78,091,150</td>
<td>GAB2</td>
<td>C</td>
<td>0.580</td>
<td>0.594</td>
<td>0.95(0.75-1.20)</td>
</tr>
<tr>
<td>8</td>
<td>rs17817600</td>
<td>85,677,471</td>
<td>PICALM</td>
<td>G</td>
<td>0.037</td>
<td>0.043</td>
<td>0.85(0.45-1.59)</td>
</tr>
<tr>
<td>8</td>
<td>rs3851179</td>
<td>85,868,640</td>
<td>PICALM</td>
<td>C</td>
<td>0.635</td>
<td>0.598</td>
<td>1.17(0.92-1.49)</td>
</tr>
<tr>
<td>8</td>
<td>rs11218343</td>
<td>121,435,587</td>
<td>SORL1</td>
<td>T</td>
<td>0.750</td>
<td>0.669</td>
<td>1.60(1.22-2.11)</td>
</tr>
<tr>
<td>14</td>
<td>rs17125944</td>
<td>53,400,629</td>
<td>FERMT2</td>
<td>C</td>
<td>0.263</td>
<td>0.218</td>
<td>1.32(1.00-1.75)</td>
</tr>
<tr>
<td>14</td>
<td>rs10498633</td>
<td>92,926,952</td>
<td>SLC24A4</td>
<td>G</td>
<td>0.884</td>
<td>0.884</td>
<td>1.04(0.72-1.50)</td>
</tr>
<tr>
<td>18</td>
<td>rs3764650</td>
<td>1,046,520</td>
<td>ABCA7</td>
<td>G</td>
<td>0.321</td>
<td>0.276</td>
<td>1.25(0.96-1.61)</td>
</tr>
<tr>
<td>19</td>
<td>rs6859</td>
<td>45,382,034</td>
<td>PVRL2</td>
<td>A</td>
<td>0.403</td>
<td>0.307</td>
<td>1.51(1.18-1.94)</td>
</tr>
<tr>
<td>19</td>
<td>rs157580</td>
<td>45,395,266</td>
<td>TOMM40</td>
<td>A</td>
<td>0.531</td>
<td>0.473</td>
<td>1.29(1.02-1.65)</td>
</tr>
<tr>
<td>19</td>
<td>rs2075650</td>
<td>45,395,619</td>
<td>TOMM40</td>
<td>G</td>
<td>0.236</td>
<td>0.084</td>
<td>3.19(2.25-4.52)</td>
</tr>
<tr>
<td>19</td>
<td>rs3865444</td>
<td>51,727,962</td>
<td>CD33</td>
<td>C</td>
<td>0.838</td>
<td>0.840</td>
<td>0.97(0.70-1.36)</td>
</tr>
</tbody>
</table>

Abbreviation: SAD, sporadic Alzheimer’s disease; Chr, chromosome; SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval.

a Risk allele reported in European population.

b Association test was adjusted for sex, age (age at onset for SAD patients and age at examination for control subjects).

c Association test was adjusted for sex, age (age at onset for SAD patients and age at examination for control subjects) and APOE ε4 status (0 or 1).
from Hardy-Weinberg equilibrium (HWE) among control subjects at \( P < 0.001 \). In the following analysis, 30 SNPs were analyzed.

### Association of SAD risk with candidate SNPs in the discovery set

In the discovery set, 6 of the 30 SNPs were significantly associated with SAD risk in Chinese (\( P < 0.05 \)) after adjustment of sex and age (age at onset for SAD patients and age at examination for control subjects) (Table 2). These 6 SNPs were rs9349407 at 6p12 in CD2AP (\( P = 1.23 \times 10^{-4} \)), rs11218343 at 11q24 in SORL1 (\( P = 8.32 \times 10^{-4} \)), rs17125944 at 14q22 in FERMT2 (\( P = 0.048 \)), rs6859 at 19q13 in PVRL2 (\( P = 0.001 \)), rs157580 at 19q13 in TOMM40 (\( P = 0.036 \)) and rs2075650 at 19q13 in TOMM40 (\( P = 7.87 \times 10^{-11} \)). The direction of association was consistent with that of European descent for all 6 SNPs (Table 2). Four of these SNPs (rs9349407, rs11218343, rs17125944, and rs2075650) remained significant after adjusting for \( APOE \) genotype, \( P < 0.05 \). Meanwhile, we analyzed LD between SNPs at 19 chromosome and \( APOE \) genotype (treating \( \varepsilon3 \) and \( \varepsilon2 \) as the same allele and \( \varepsilon4 \) as another allele), and found the last SNP (rs2075650) was in strong LD with \( APOE \) genotype, \( r^2 = 0.48 \) (Table S2). Associations between SNPs and SAD risk were also tested in the testing set (Table S3).

### GRS calculation and its discriminative performance analysis in the testing set

To assess the cumulative effect of multiple SAD risk-associated SNPs in predicting SAD risk, we calculated GRS for subjects in the independent testing set based on all 6 implicated SNPs in the discovery set (Table 3). The median GRS was significantly higher in SAD cases than in controls, \( P < 0.001 \). AUC of GRS in discriminating SAD cases from controls was 0.63, higher than that of \( APOE \) (0.60). Considering that the effect of GRS may be confounded by \( APOE \), we calculated a modified GRS by removing three SNPs that are related to \( APOE \): two SNPs that were no longer significantly associated with SAD risk after adjusting for \( APOE \) genotypes (rs6859 and rs157580) and one SNP that was in strong LD with \( APOE \) genotypes (rs2075650). In the independent testing set, the modified GRS was significantly higher in SAD cases than in controls, \( P = 0.002 \) (Table 3). The AUC of the modified GRS was 0.58. When this modified GRS was combined with \( APOE \) genotypes, the AUC was 0.64, significantly higher than that of \( APOE \) alone (0.60), \( P = 0.003 \) (Figure 1).

### Association analysis of GRS and SAD risk

When the modified GRS was analyzed in subjects stratified by the \( APOE \varepsilon4 \) status in the testing set, similar trends were observed in both \( APOE \varepsilon4 \) carriers and non-carriers, although the association and discriminative performance of modified GRS was slightly stronger in non-carriers (Table 3).

### DISCUSSION

The primary purpose of this study was to assess performance of multiple risk-associated SNPs for predicting SAD risk in the Han Chinese population. To achieve this goal, we firstly identified SNPs that were associated with SAD risk among Chinese in a discovery set. We then assessed the cumulative effect of these implicated SNPs, as measured by GRS, on association of SAD risk and ability to discriminate SAD patients from non-dementia controls in the testing set. Furthermore, to assess whether the predictive performance of multiple SAD risk-associated SNPs are independent of \( APOE \) genotypes, we calculated a modified GRS that based on three SAD risk-associated SNPs that are independent of \( APOE \) genotypes. With this rigorously strategy, we demonstrated that modified GRS was able to discriminate...
SAD patients from controls, with an AUC of 0.58. When combined the modified GRS with APOE, the AUC increased to 0.64, significantly higher than APOE alone (0.60), \( P = 0.003 \).

To our knowledge, this is the first report assessing cumulative effect of multiple AD risk-associated SNPs on association and discrimination of SAD. It is well recognized that effect of individual SNPs on AD risk is modest. However, it is hypothesized that cumulatively they have a stronger effect. As demonstrated in this study, the AUC of modified GRS based on three implicated SAD risk-associated SNPs (0.58) was similar to that of well-established APOE (0.60). This result offers empirical evidence to support this cumulative effect hypothesis and provides basis for additional larger and more comprehensive studies to further test the hypothesis in Chinese and European descent. With more established AD risk-associated SNPs (such as in European descent), it is expected that cumulative effect will be stronger.

The findings that SAD risk-associated SNPs are associated with SAD risk in both APOE ε4 allele carriers and non-carriers and that they add value to APOE in discriminating SAD cases from controls are important. On one hand, these results suggest SAD risk-associated SNPs play similar roles in the etiology of SAD among APOE ε4 allele carriers and non-carriers. On the other hand, it is practically important in assessment of SAD risk. The APOE ε4 allele has been found to be the strongest risk factor for SAD [9-12]. A meta-analysis of clinical and autopsy-based studies demonstrated that individuals with ε4 allele have increased AD risk compared with ε3/ε3 genotype in Caucasian population (OR was 2.6, 3.2, and 14.9 for individuals with ε2/ε4, ε3/ε4, and ε4/ε4, respectively) [32]. However, the AUC of APOE alone in discriminating SAD remains moderate (for example, AUC was 0.62 and 0.60 in the discovery set and testing set of our study, respectively). Furthermore, it is recognized that 40–50% of SAD patients do not carry APOE ε4 allele [11, 13]. This number was even higher in our study where 58.6% SAD patients did not carry APOE ε4 allele (Table 1). Therefore, identifying tool for better risk assessment of AD risk, especially among subjects without APOE ε4 allele is necessary. Better risk assessment may identify subjects at higher risk for SAD for targeted prevention. Subjects with a higher risk for SAD may be more motivated to take action to prevent AD through reducing life-style risk factors associated with AD.

Family history is another well-established risk factor for AD. The relative risk of AD for those with at least one first degree relative with dementia was estimated at 3.2-3.8 [33-35]. Unfortunately, we could not assess the effect of family history on AD risk in this study because we focused on SAD in this study, and none of the AD patients had a positive family history by our inclusion criteria. However, the fact that we demonstrated GRS based multiple inherited risk-associated SNPs is a predictor of SAD risk among patients without a known family history indirectly suggests family history alone is not sufficient to capture inherited risk for AD. In contrast to GRS that is based on individual’s own risk-associated SNPs, family history is an indirect measurement of familial risk (inherited and shared household environment) through their relatives, and is therefore influenced by the number, age, and competing mortality of their relatives. This

Figure 1: Receiver operating characteristic curves for genetic models among Chinese subjects in the testing set. a Modified genetic risk score (GRS) based on 3 SNPs (rs9349407, rs11218343 and rs17125944). b APOE ε4 status (0 or 1). c Combination of modified GRS and APOE.
limitation is more prominent for late age onset diseases such as AD. Thus, lack of a known family at the time of examination may not necessarily indicate that individuals are at lower risk for AD. Studies are needed to assess the combined performance of family history, GRS, and APOE in assessing AD risk.

We randomly divided our study subjects into two sets of equal sample size: discovery and testing. The major advantage of this approach is that we can identify SAD risk-associated SNPs in Han Chinese and obtain their OR in the discovery set and then objectively assess the performance of these SNPs in the independent testing set. However, this approach reduced the statistical power to detect association of SNPs with SAD risk. Among the 30 SNPs tested in the discovery set, 19 SNPs had the same direction of association as in the studies of European descent, although only 6 SNPs reached the statistical significance of $P < 0.05$. Larger sample size may be needed to confirm additional AD risk-associated SNPs in Han Chinese.

Several case-control studies on association of AD risk-associated SNPs reported in GWAS of European descent with AD risk in Chinese population were published in the last several years. However, few SNPs were consistently implicated among these studies. For example, Chen et al. [21] evaluated 7 SNPs (rs3818361 and rs6656401 in CR1; rs11136000, rs2279590, and rs9331888 in CLU; rs3851179 and rs541458 in PACALM) among 462 AD patients and 350 control subjects from southern Chinese population. Of the 7 SNPs, rs6656401 ($P = 0.035$) and rs3818361 ($P = 0.029$) in CR1, and rs11136000 in CLU ($P = 0.038$) were confirmed; rs3851179 in PACALM showed significant association with LOAD only in APOE ε4 non-carriers ($P = 0.028$). Tan et al. [25] assessed a total of 10 SNPs among 612 sporadic late-onset AD (LOAD) patients and 612 control subjects from northern Han Chinese, including 2 in BINI (rs7561528 and rs744373), 2 in ABCA7 (rs3752246 and rs3764650), 3 in the MS4A gene cluster (rs4938933, rs610932, and rs670139), and 1 each in CD2AP (rs9349407), CD33 (rs3865444), and EPHA1 (rs11767557). Based on a multivariate analysis, rs610932 in MS4A6A ($P = 0.019$) and rs3865444 in CD33 ($P = 0.017$) were confirmed; rs7561528 in BINI was confirmed only in APOE ε4 carriers ($P = 0.039$). Ma et al. [26] conducted a replication study of rs11754661 and rs2073067 in MTHFD1L among 582 LOAD subjects and 607 healthy controls from northern Han Chinese. The rs11754661 was confirmed ($P = 0.016$). Liu et al. [23] evaluated and confirmed the association of rs3764650 in ABCA7 with SAD in 350 SAD and 283 non-demented elderly controls from Han Chinese, $P = 0.004$. Ma et al. [27] investigated the association of three SNP (rs157580, rs2075650 and rs11556505 in TOMM40) with LOAD among 787 LOAD patients and 791 healthy subjects. The rs157580 ($P < 0.001$) and rs2075650 ($P = 0.001$) were confirmed. Among these confirmed SNPs (rs3818361, rs6656401, rs7561528, rs11754661, rs11136000, rs610932, rs3851179, rs3764650, rs3865444, rs157580 and rs2075650), 2 SNPs (rs157580 and rs2075650 in TOMM40) were also found to be significantly associated with SAD in our study. Except for 2 SNPs (rs3818361 and rs7561528), the rest SNPs have the same direction of association with previous findings in Chinese. Multiple factors may contribute to these different confirmed SNPs, including small sample size, different criteria for AD patients and controls, and different genetic background between northern and southern Han Chinese.

The three SNPs, independent of APOE genotypes and used in GRS calculating to predict SAD risk in the study, are rs9349407 in CD2AP (intron 1), rs11218343 in SORL1 (intron 21), and rs17125944 in FERMT2 (intron 14). The functions of these 3 genes have been reported to be relevant to the development of AD. Both CD2AP and FERMT2 have been implicated in cell adhesion [36, 37]. In Drosophila model of AD, CD2AP and FERMT2 were identified as modifier of Tau neurotoxicity, which related to neurofibrillary tangle pathology in AD [38]. SORL1 encodes a neuronal sorting protein that binds APP protein and directs it towards the endosome-recycling pathways [39] and variants in SORL1 were significantly associated with cerebrospinal Aβ42 levels [40], which reflect the metabolic process in brain and was used to aid the diagnosis of AD at an early stage of disease.

In summary, results from this well-designed but underpowered study provided preliminary evidence that multiple AD risk-associated SNPs can be used to supplement APOE to better define individual’s risk for AD. Larger studies are justified to formally test the hypothesis and assess its predictive performance.

**MATERIALS AND METHODS**

**Study population**

Subjects included in this study (515 SAD patients and 770 cognitively normal controls) were Chinese Han, and were recruited during 2008-2013. SAD patients, comprising early- and late-onset SAD (age at onset ranged from 45 to 88 years), were recruited from Huashan Hospital in Shanghai and were diagnosed as probable AD according to DSM-IV-R and NINCDS-ADRDA criteria [41, 42]. All SAD patients reported no family history of AD. The cognitively normal controls were recruited from communities in Shanghai and were carefully evaluated based on mini-mental state examination (MMSE) and years of education. They were frequency matched for SAD cases by gender and age. Two senior neurologists reviewed all data and confirmed the diagnosis. APOE genotype status, measured by method described by Donohoe et al. [43], was available in cases and controls. This study was
approved by the ethics committee of Huashan Hospital and written informed consents were completed for all study subjects.

After genotyping, subjects with a missing rate of > 20% were removed from the study (56 SAD patients and 19 control subjects). The retained subjects (459 SAD patients and 751 controls) were randomly divided into discovery set (232 cases and 373 controls) and testing set (227 cases and 378 controls). Association of AD risk-associated SNPs reported in European descent with SAD risk was firstly tested in the discovery set. Significant SNPs were used to calculate GRS in the testing set.

SNP selection

A total of 33 SAD risk-associated SNPs were selected using the following criteria: 1) association with AD risk in European population exceeded the threshold of a genome-wide significance level ($P < 5 \times 10^{-8}$) and published before Jan 2014; 2) if multiple SNPs that are in strong linkage disequilibrium (LD) met the above criterion, defined by pairwise $r^2 > 0.2$ estimated from the HapMap CHB (Han Chinese in Beijing, China) population, the most commonly cited SNP was selected. The information of all SNPs that be chosen was listed in Table S4.

SNP genotyping and Quality control

Genotyping of selected SNPs was performed using the Sequenom MassArray system (iPLEX; Sequenom, Inc. San Diego, CA) at the Centre for Genomic Translational Medicine and Prevention, School of Public Health, Fudan University. Duplicates from two subjects and two water samples (negative control) were included in each 96-well plate for genotyping quality control. All assays were conducted blinded to case-control status. The overall concordance rate was 100% among the duplicated quality control samples. A quality control was conducted to select the samples (mentioned in “Study Population” section) and SNPs for further analysis. SNPs with a missing rate of > 5%, the minor allele frequency (MAF) of < 0.01 in either cases or controls, or with Hardy-Weinberg equilibrium (HWE) test at $P < 0.001$ among controls were excluded.

Statistical analysis

Differences between cases and controls were tested using t-test for quantitative variables and Chi-square test for qualitative variables. Associations between SNPs and SAD risk were tested for each SNP using an additive model adjusted for 1) sex, age (age at onset for SAD patients and age at examination for control subjects) and $APOE \epsilon 4$ status (0 or 1). The allelic odds ratio (OR) and 95% confidence intervals (CI) were estimated using a logistic regression model.

GRS for each subject was calculated based on SAD risk-associated SNPs established in the discovery set using the method described by Pharoah et al. [44]. Briefly, 1) the allelic OR of each SNP was obtained from the discovery set, 2) the genotypic OR of each SNP was estimated from the allelic OR assuming a multiplicative model, 3) the risk relative to the average risk in the population was calculated for each genotype based on genotypic OR and genotype frequency in the HapMap CHB (Han Chinese in Beijing, China) population, and 4) a GRS was obtained by multiplying the risks relative to the population of all SNPs. Therefore, a GRS of 1.0 indicates an average risk in the general population.

Non-parametric analysis (Wilcoxon Rank Sum Test) was used to test association of GRS and SAD risk. The performance of GRS in discriminating SAD cases from controls was evaluated using the area under the receiver operating characteristic curve (AUC). Difference in AUC between two predictive model was tested using the method described by DeLong and colleagues [45].

All statistical analyses were performed using the PLINK software (version 1.07) [46] and SAS software (version 9.2; SAS Institute, Cary, NC). All statistical tests were two-sided.

ACKNOWLEDGMENTS

We sincerely thank all participants who agreed to participate in this study. This work was supported by the grant from the National Natural Science Foundation to Zhi-Ying Wu (81125009, Beijing). This work was also partially supported by the Ellrodt-Schweighauser Family Chair of Cancer Genomic Research of North Shore University HealthSystem to Jianfeng Xu.

Authors’ Contributions

Study conception and design: JFX and ZYW. Acquisition, analysis, or interpretation of data: QYX, ZJL, YMS, HLL, HTC and LX. Statistical analysis: QYX, ST, DKJ, BL and CHW. Manuscript drafting: QYX, ST. Manuscript revision: ZYW, JFX, DKJ and SLZ.

CONFLICTS OF INTEREST

All authors report no disclosures relevant to the manuscript and declare no conflicts of interest.
REFERENCES


46. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81:559-575.