GAB2 as an Alzheimer Disease Susceptibility Gene

Follow-up of Genomewide Association Results

Brit-Maren M. Schjeide, BS; Basavaraj Hooli, MS; Michele Parkinson, BS; Meghan F. Hogan, BS; Jason DiVito, BS; Kristina Mullin, BS; Deborah Blacker, MD, ScD; Rudolph E. Tanzi, PhD; Lars Bertram, MD

Background: Genomewide association (GWA) studies have recently implicated 4 novel Alzheimer disease (AD) susceptibility loci (GAB2, GOLM1, and 2 uncharacterized loci to date on chromosomes 9p and 15q). To our knowledge, these findings have not been independently replicated.

Objective: To assess these GWA findings in 4 large data sets of families affected by AD.

Design: Follow-up of genetic association findings in previous studies.

Setting: Academic research.

Participants: More than 4000 DNA samples from almost 1300 families affected with AD.

Main Outcome Measures: Genetic association analysis testing of 4 GWA signals (rs7101429 [GAB2], rs7019241 [GOLM1], rs10519262 [chromosome 15q], and rs9886784 [chromosome 9p]) using family-based methods.

Results: In the combined analyses, only rs7101429 in GAB2 yielded significant evidence of association with the same allele as in the original GWA study (P=.002). The results are in agreement with recent meta-analyses of this and other GAB2 polymorphisms suggesting approximately a 30% decrease in risk for AD among carriers of the minor alleles. None of the other 3 tested loci showed consistent evidence for association with AD across the investigated data sets.

Conclusions: GAB2 contains genetic variants that may lead to a modest change in the risk for AD. Despite these promising results, more data from independent samples are needed to better evaluate the potential contribution of GAB2 to AD risk in the general population.

Arch Neurol. 2009;66(2):250-254

Although the heritability of late-onset Alzheimer disease (AD) is high,1,2 our knowledge of the underlying putative susceptibility genes remains incomplete. The only unequivocally established late-onset AD gene is APOE (OMIM 107741) (encoding apolipoprotein E), whose ε4 allele increases the risk for AD by 4- to 15-fold in a dose-dependent manner.3,4 APOE and most other genetic association findings in AD originate from candidate gene approaches (ie, studies that investigate certain genes based on a prior hypothesis regarding their potential involvement in pathogenesis).5 An alternative to this approach is afforded by recent advances in large-scale genotyping technologies enabling researchers to perform comprehensive unbiased genomewide association (GWA) analyses.6 To date, 3 groups have reported the results of AD GWA analyses.7-9

The first group used a low-density design testing roughly 17 000 single-nucleotide polymorphisms (SNPs) in or near genetic coding regions in almost 4000 combined cases and controls, including several confirmed by autopsy.7 The only SNPs consistently associated with AD risk across different samples were located within or in proximity to APOE and most likely reflect linkage disequilibrium (LD) with the ε4 allele. Although several additional loci were highlighted as potential AD genes by the authors, none showed the same consistency of effect or level of statistical significance as the ε4-related variants.

The second group tested approximately 500 000 SNPs in roughly 1100 unrelated AD cases (all with neuropathologically confirmed diagnoses) and controls.8 Again, except for a single SNP in strong LD with APOE ε4, no other genomewide significant signals were observed. In a follow-up article,10 the same group reported evidence of an association between variants in GRB2-associated binding protein 2 (Gab2 [gene abbreviation, GAB2]) (OMIM 606203) on chromosome 11q14 and AD risk in the same 1100 neuropathologically confirmed persons and in approximately 360 clinically diagnosed AD cases and controls, but this was noted only in carriers of APOE4. The association was observed with 10 different SNPs in GAB2, all displaying
Table 1. Summary of Findings in High-Density Genomewide Association (GWA) Screens in Alzheimer Disease (AD)

<table>
<thead>
<tr>
<th>Source</th>
<th>Gene or Locus (SNP), Trait</th>
<th>Model</th>
<th>(P) Value</th>
<th>OR (95% CI)</th>
<th>GWA Data Set Size (AD vs Controls), No. of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coon et al., 2007&lt;sup&gt;b&lt;/sup&gt;</td>
<td>APOC1 (rs4420638), affection&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Additive</td>
<td>1.1 × 10&lt;sup&gt;-20&lt;/sup&gt;</td>
<td>4.01 (NA)</td>
<td>664 vs 422</td>
</tr>
<tr>
<td>Reiman et al., 2007&lt;sup&gt;b&lt;/sup&gt;</td>
<td>GAB2 (rs2373115), affection</td>
<td>Additive</td>
<td>4.6 × 10&lt;sup&gt;-7&lt;/sup&gt;</td>
<td>3.21 (2.04-5.05)</td>
<td>446 vs 290</td>
</tr>
<tr>
<td>Li et al., 2008</td>
<td>GOLM1 (rs7019241), affection&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Additive</td>
<td>2.3 × 10&lt;sup&gt;-14&lt;/sup&gt;</td>
<td>NA</td>
<td>753 vs 736</td>
</tr>
<tr>
<td></td>
<td>GWA_19q21 (rs1019262), age at onset</td>
<td>Recessive</td>
<td>4.5 × 10&lt;sup&gt;-4&lt;/sup&gt;</td>
<td>1.89 (1.46-2.45)</td>
<td>753 vs 736</td>
</tr>
<tr>
<td></td>
<td>GWA_9p24 (rs9886784), affection</td>
<td>Recessive</td>
<td>3.1 × 10&lt;sup&gt;-4&lt;/sup&gt;</td>
<td>3.23 (1.79-5.84)</td>
<td>753 vs 736</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; NA, not available; OR, odds ratio; SNP, single-nucleotide polymorphism.

<sup>a</sup>In the initial GWA analysis as reported in the original publications.

<sup>b</sup>The samples of these 2 publications overlap.

<sup>c</sup>This association is due to linkage disequilibrium with the APOE e4 allele.

Table 2. Demographic Characteristics of Families Affected by Alzheimer Disease

<table>
<thead>
<tr>
<th>Data Set</th>
<th>No. of Families (No. of Subjects)</th>
<th>No. of Affected Individuals (Mean AAO [SD] {Range})</th>
<th>No. of Unaffected Individuals (Mean AAE [SD] {Range})</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIMH</td>
<td>410 (1376)</td>
<td>941 (72.3 [7.7] {50-97})</td>
<td>404 (70.9 [10.7] {31-93})</td>
</tr>
<tr>
<td>NIA</td>
<td>329 (1040)</td>
<td>748 (74.2 [7.1] {52-98})</td>
<td>282 (73.4 [9.6] {36-94})</td>
</tr>
<tr>
<td>NCRAD</td>
<td>331 (1108)</td>
<td>799 (71.3 [7.6] {50-98})</td>
<td>293 (70.6 [8.1] {39-93})</td>
</tr>
<tr>
<td>CAG</td>
<td>215 (483)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>220 (69.3 [9.0] {50-89})</td>
<td>263 (73.3 [8.6] {50-92})</td>
</tr>
</tbody>
</table>

Abbreviations: AAE, age at examination; AAO, age at onset; CAG, Consortium on Alzheimer’s Genetics; NCRAD, National Cell Repository for Alzheimer’s Disease; NIA, National Institute on Aging; NIMH, National Institute of Mental Health.

<sup>a</sup>In the other data sets, affected individuals plus unaffected individuals do not sum to the total number of subjects because of unknown phenotypes.

high degrees of LD, indicating that they likely point to the same underlying signal.

Finally, the third group tested almost 500,000 markers in approximately 1500 clinically diagnosed AD cases and controls from Canada (replicated in approximately 670 AD cases and controls from the United Kingdom). They reported an association between variants in Golgi membrane protein 1 (encoded by GOLM1 [OMIM 606804], also known as GOLPH2) on chromosome 9q22 and 2 uncharacterized loci on chromosomes 9p and 15q.<sup>9</sup> The findings from the 2 high-density GWA screens are summarized in Table 1.

In the present study, we assessed the 4 putative AD loci that emerged from 2 high-density 500,000 GWA screens.

METHODS

SAMPLE

The following 4 data sets tested in this project were initially collected for the study of genetic factors in AD: (1) Consortium on Alzheimer’s Genetics (CAG),<sup>11</sup> (2) National Institute on Aging (NIA [http://www.ncrad.org]), (3) National Institute of Mental Health (NIMH),<sup>12</sup> and (4) National Cell Repository for Alzheimer’s Disease (NCRAD [http://www.ncrad.org]). Characteristics of the samples are given in Table 2. All data sets represented primarily sibships and lacked parental genotypes. Except for the CAG data set, most pedigrees analyzed herein were nuclear families ascertained on the basis of multiple affected individuals. In addition to containing at least 1 affected relative pair, many pedigrees also had DNA samples available from additional affected or unaffected persons (mostly siblings). The diagnosis of definite, probable, or possible AD was made according to National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer’s Disease and Related Disorders Association criteria for affected individuals in all 4 data sets.<sup>13</sup> For all 4 data sets, we included only families of white race/ethnicity and only families in which no affected person had an age at onset younger than 50 years. The NIMH families were collected as part of the NIMH Genetics Initiative. The data set is composed of 1528 persons from 457 families. Only families in which DNA samples were available from at least 2 affected family members were included in these analyses (1376 persons from 410 families). The NIA and NCRAD families were obtained from the NCRAD, and ascertainment and collection details can be found at the NCRAD Web site (http://www.ncrad.org). Our sampling rules yielded 1040 persons from 329 families for the NIA data set and 1108 persons from 331 families for NCRAD data set. The CAG families were recruited from multiple NIA-funded AD research centers under the auspices of the CAG.<sup>11</sup> Probands were included only if they had at least 1 unaffected living sibling willing to participate in the study. This data set was composed of 483 persons from 215 families.

GENOTYPING

All genetic variants were genotyped on individually optimized single-base extension reactions detected by fluorescent polarization as previously described<sup>11</sup> (genotyping details are available from the corresponding author). In GAB2, we chose a proxy (rs7101429) in lieu of the strongest GWA SNP (rs2373115) because this SNP showed better clustering in our preliminary genotyping. Both SNPs are in almost complete LD in persons of white race/ethnicity (\(r^2=0.92\) and \(D’=1.0\) based on the current genotype release from the International HapMap Project Web site [http://www.hapmap.org]) and can be used interchangeably for the purpose of association testing. Hardy-Weinberg equilibrium was determined separately for all markers in all 4 data...
Table 3. Follow-up of High-Density Genomewide Association Findings (Except for APOE-Related Findings) Among Informative Families

<table>
<thead>
<tr>
<th>Gene or Locus (SNP), Trait</th>
<th>Model</th>
<th>NIMH</th>
<th>NIA</th>
<th>NCRAD</th>
<th>CAG</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P Value</td>
<td>No. of Families</td>
<td>P Value</td>
<td>No. of Families</td>
<td>P Value</td>
<td>No. of Families</td>
</tr>
<tr>
<td>GAB2 (rs7101429), affection</td>
<td>All</td>
<td>Additive</td>
<td>.005</td>
<td>144</td>
<td>.26</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>APOE ε4 positive</td>
<td>Additive</td>
<td>.008</td>
<td>111</td>
<td>.19</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>APOE ε4 negative</td>
<td>Additive</td>
<td>.17</td>
<td>33</td>
<td>.74</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>GOLM1 (rs7019241), affection</td>
<td>Dominant</td>
<td>.09</td>
<td>86</td>
<td>.96</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>GWA_15q21 (rs10519282), age at onset</td>
<td>Recessive</td>
<td>.60</td>
<td>175</td>
<td>.29</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>GWA_9p24 (rs9886784)</td>
<td>Recessive</td>
<td>.72</td>
<td>156</td>
<td>.29</td>
<td>131</td>
</tr>
</tbody>
</table>

Abbreviations: CAG, Consortium on Alzheimer’s Genetics; NCRAD, National Cell Repository for Alzheimer’s Disease (http://www.ncrad.org); NIA, National Institute on Aging (http://www.nicr.org); NIMH, National Institute of Mental Health; SNP, single-nucleotide polymorphism.

Table 4. Effect Size Estimates for the Family-Based Follow-up Findings (Except for APOE-Related Findings)

<table>
<thead>
<tr>
<th>Gene or Locus (SNP), Trait</th>
<th>Model</th>
<th>NIMH</th>
<th>NIA</th>
<th>NCRAD</th>
<th>CAG</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAB2 (rs7101429)</td>
<td>All</td>
<td>Additive</td>
<td>0.58 (0.41-0.83)</td>
<td>1.12 (0.75-1.69)</td>
<td>0.70 (0.44-1.11)</td>
<td>0.80 (0.48-1.34)</td>
</tr>
<tr>
<td></td>
<td>APOE ε4 positive</td>
<td>Additive</td>
<td>0.63 (0.41-0.97)</td>
<td>1.06 (0.68-1.66)</td>
<td>0.70 (0.42-1.14)</td>
<td>0.74 (0.39-1.41)</td>
</tr>
<tr>
<td></td>
<td>APOE ε4 negative</td>
<td>Additive</td>
<td>0.86 (0.33-2.24)</td>
<td>1.72 (0.63-4.70)</td>
<td>0.74 (0.22-2.42)</td>
<td>0.92 (0.39-2.20)</td>
</tr>
<tr>
<td></td>
<td>GOLM1 (rs7019241)</td>
<td>Dominant</td>
<td>0.32 (0.05-3.88)</td>
<td>1.04 (0.24-4.55)</td>
<td>1.33 (0.28-6.26)</td>
<td>1.00 (0.14-7.10)</td>
</tr>
<tr>
<td></td>
<td>GWA_15q21 (rs10519282)</td>
<td>Recessive</td>
<td>1.06 (0.73-1.54)</td>
<td>0.98 (0.67-1.44)</td>
<td>0.83 (0.53-1.30)</td>
<td>0.88 (0.53-1.47)</td>
</tr>
<tr>
<td></td>
<td>GWA_9p24 (rs9886784)</td>
<td>Recessive</td>
<td>0.81 (0.55-1.22)</td>
<td>0.84 (0.55-1.28)</td>
<td>0.85 (0.55-1.32)</td>
<td>1.75 (1.02-3.04)</td>
</tr>
</tbody>
</table>

Abbreviations: CAG, Consortium on Alzheimer’s Genetics; CI, confidence interval; NCRAD, National Cell Repository for Alzheimer’s Disease (http://www.ncrad.org); NIA, National Institute on Aging (http://www.nicr.org); NIMH, National Institute of Mental Health; OR, odds ratio; SNP, single-nucleotide polymorphism.

STATISTICAL ANALYSIS

To test for association, we used statistical software (PBAT, version 3.6; http://www.biostat.harvard.edu/~clange/default.htm) using the same traits (affection status or age at onset) and genetic models as in the original GWA analyses (using an additive model as a substitute for the allelic tests). For GAB2, we also tested for association in families with at least 1 affected person carrying an APOE ε4 allele, similar to the stratification in the original article. All analyses were restricted to families of self-reported white race/ethnicity. To combine statistical evidence across the association analyses from each independent data set, we used the combined probability test by Fisher. Because the hypothesis of our study was to test for association in the same direction as that observed in the original articles, all P values are 1-tailed. P values were inverted (1 - P) when undertransmission to affected individuals was observed with the opposite allele compared with the original GWA finding. Odds ratios (ORs) were calculated by fitting a conditional logistic regression model to each data set, where family defines the stratum. Summary ORs were calculated for all tested variants using fixed-effects models, as no evidence of significant heterogeneity in the sample-specific ORs was observed across data sets. Results of power calculations (performed in PBAT) suggested that we had good to excellent (≥85%) power in the combined sample to detect the genetic effect sizes (Table 1) estimated in the original GWA studies at α = .05 and a disease prevalence of 10% using the allele frequencies of the respective control populations.

RESULTS

Overall, genotyping efficiency for the 4 tested variants was 98.4%. The mean error rate (based on approximately 10% of the sample run in duplicate) was below 0.5%. None of the markers deviated significantly from Hardy-Weinberg equilibrium. The only variant to show nominally significant association when results were combined across all 4 data sets was rs7101429 in GAB2 (Table 3). Although the association was more pronounced in APOE ε4–positive families vs APOE ε4–negative families (similar to the original article), the association was most significant in the unstratified families possibly because of increased power. The same allele was undertransmitted across all 4 data sets, although this association reached only nominal significance in 2 data sets (NIMH and NCRAD). The undertransmission of the mi-
nor allele is indicative of a “protective” effect (approximate summary OR across all 4 data sets, 0.76 [95% confidence interval, 0.62-0.94]) (Table 4). This is in line with the study by Reiman et al,\textsuperscript{10} who found an increased risk for AD associated with the opposite allele. However, the effect size estimated herein is about 1 order of magnitude smaller than that in the study by Reiman et al, possibly indicative of the “winner’s curse” phenomenon.\textsuperscript{17} None of the other 3 tested SNPs (ie, those in GOLM1 and the 2 loci on chromosomes 9p and 15q) showed overall evidence of association across all data sets. Only the marker in GOLM1 (rs7019241) showed modest evidence for association in the NIMH data set (P=0.09), whereas GWA\textsubscript{9p24} (rs9886784) showed nominal association in the CAG data set (P=0.03) (Table 3).

To our knowledge, our study provides the first systematic assessment of genetic association findings that originated from high-density GWA analyses in AD. Although GWA studies represent a leap forward in the analysis of complex diseases such as AD, independent replication (as with earlier methods) remains the first step in distinguishing “real” from false-positive association findings.\textsuperscript{18} Based on the data generated in our study, the GAB2 association represents the most promising of the GWA findings in AD to date. In GWA data by Li et al,\textsuperscript{9} who used the same marker panel as Reiman et al,\textsuperscript{10} all GAB2 variants show nominally significant association on reanalysis of the publicly available genotype data (for more details, see the meta-analyses on the AlzGene Website [http://www.alzgene.org]).\textsuperscript{3} One other study\textsuperscript{19} to date has investigated the potential association between GAB2 and AD risk in case-control samples from France and the United Kingdom. Although the authors reported an overall nonsignificant outcome, in 2 of 3 samples the allele and genotype distributions showed the same pattern as in the 2 GWA case-control studies\textsuperscript{9,10} and in our families with AD. Finally, based on systematic meta-analyses of all published genetic association studies in AD,\textsuperscript{3} GAB2 effect sizes (OR range, 0.66-0.79 for protection [equivalent to 1.27-1.51 for risk for the opposite allele]) are among the strongest and most significant observed in any putative disease gene after APOE in the field of AD. Taken together, these findings suggest that GAB2 ranks among the top contenders for a second genuine AD susceptibility gene.

The lack of confirmation of the 4 GWA findings reported by Li et al\textsuperscript{9} in our sample could be related to several reasons, including low power (eg, because of overestimation of the effect size in the original study), locus or allelic heterogeneity (eg, because the risk effects exerted by some loci or alleles may be present in some populations but not in others), or LD with other as yet unidentified alleles (eg, because LD with the presumed risk alleles may be lower in our sample compared with that of Li et al, reducing our power to detect such effects). Alternatively, the lack of replication of the original findings—which is a common occurrence in genetic studies of complex diseases regardless of study design—could also be due to type 1 error (ie, false-positive results in the original study). More independent follow-up analyses using case-control and family-based designs are needed to more accurately assess these possibilities.

GAB2 is a member of a family of evolutionarily highly conserved scaffolding and adapter proteins that are involved in multiple signaling pathways and particularly in the transduction of cytokine and growth receptor signaling.\textsuperscript{20,21} GAB2 is ubiquitously expressed but is found at high levels in white blood cells, prefrontal cortex, and hypothalamus. Functionally, preliminary evidence reported in the original GWA article\textsuperscript{10} suggested that changes in GAB2 expression could potentially affect Gsk3-dependent phosphorylation of tau and the formation of neurofibrillary tangles. Moreover, growth factor receptor–bound protein 2, which binds GAB2, has been reported to bind tau,\textsuperscript{22} amyloid-β precursor protein,\textsuperscript{23} and presenilin 1 and 2.\textsuperscript{23} These interactions have been proposed to regulate signal transduction (eg, the ERK1/2 pathway).\textsuperscript{21} Therefore, GAB2 is a biologically plausible gene for involvement in the pathogenesis of AD.

In conclusion, results of this first systematic follow-up (to our knowledge) of recent high-density GWA analyses in AD suggest that GAB2 on chromosome 11q14, or a locus nearby, may harbor genetic variants that modulate the risk for AD. Additional independent replications and functional genetic analyses are warranted to elucidate the potential biochemical mechanisms and the epidemiologic relevance of this association.

**Accepted for Publication:** September 11, 2008.

**Correspondence:** Lars Bertram, MD, Genetics and Aging Research Unit, Massachusetts General Institute for Neurodegenerative Disease, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Massachusetts General Hospital, 114 16th St, Charlestown, MA 02129 (bertram.lars@gmail.com).

**Author Contributions:** Study concept and design: Blacker, Tanzi, and Bertram. Acquisition of data: Schjeide, Hooli, Parkinson, Hogan, DiVito, Mullin, Blacker, Tanzi, and Bertram. Analysis and interpretation of data: Schjeide, Tanzi, and Bertram. Drafting of the manuscript: Schjeide and Bertram. Critical revision of the manuscript for important intellectual content: Schjeide, Hooli, Parkinson, Hogan, DiVito, Mullin, Blacker, Tanzi, and Bertram. Statistical analysis: Schjeide and Bertram. Obtained funding: Tanzi and Bertram. Administrative, technical, and material support: Schjeide, Mullin, and Bertram. Study supervision: Bertram.

**Financial Disclosure:** Drs Tanzi and Bertram are consultants to and equity holders in TorreyPines Therapeutics, Inc, and Prana Biotechnology, Ltd. Dr Tanzi is a co-founder of TorreyPines Therapeutics, Inc, and Prana Biotechnology, Ltd.

**Funding/Support:** This study was supported by grant 5RO1AG23667 from the National Institute on Aging (Dr Bertram), by grant R37MH60009 from the National Institute of Mental Health (Dr Tanzi), and by grants from the Extendicare Foundation and the Cure Alzheimer’s Fund (Dr Bertram). The AlzGene database project is funded by the Cure Alzheimer’s Fund (principal investigator Dr Bertram).
Additional Contributions: We thank the families for participating in the study.

REFERENCES