Complete Genomic Screen in Late-Onset Familial Alzheimer Disease

Evidence for a New Locus on Chromosome 12

Margaret A. Pericak-Vance, PhD; Meredyth P. Bass, MS; Larry H. Yamaoka, PhD; Perry C. Gaskell, PA-C; William K. Scott, PhD; Henry A. Terwedow, PhD; Marissa M. Menold, MS; P. Michael Connelly, PhD; Gary W. Small, MD; Jeffery M. Vance, PhD, MD; Ann M. Saunders, PhD; Allen D. Roses, MD; Jonathan L. Haines, PhD

Context.—Four genetic loci have been identified as contributing to Alzheimer disease (AD), including the amyloid precursor protein gene, the presenilin 2 gene, and the apolipoprotein E gene, but do not account for all the genetic risk for AD.

Objective.—To identify additional genetic risk factors for late-onset AD.

Design.—A complete genomic screen was performed (N=280 markers). Critical values for chromosomal regional follow-up were a P value of .05 or less for affected relative pair analysis or sibpair analysis, a parametric lod score of 1.0 or greater, or both. Regional follow-up included analysis of additional markers and a second data set.

Setting.—Clinic populations in the continental United States.

Patients.—From a series of multiplex families affected with late-onset (≥60 years) AD ascertained during the last 14 years (National Institute of Neurological Disorders and Stroke—Alzheimer’s Disease and Related Disorders Association diagnostic criteria) and for which DNA has been obtained, a subset of 16 families (135 total family members, 52 of whom were patients with AD) was used for the genomic screen. A second subset of 38 families (216 total family members, 89 of whom were patients with AD) was used for the follow-up analysis.

Main Outcome Measures.—Linkage analysis results generated using both genetic model-dependent (lod score) and model-independent methods.

Results.—Fifteen chromosomal regions warranted initial follow-up. Follow-up analyses revealed 4 regions of continued interest on chromosomes 4, 6, 12, and 20, with the strongest results observed for chromosome 12. Peak 2-point affected-only lod scores (n=54) were 1.3, 1.6, 2.7, and 2.2 and affected relative pairs P values (n=54) were .04, .03, .14, and .04 for D12S1373, D12S1057, D12S1042, and D12S390, respectively. Sibpair analysis (n=54) resulted in maximum lod scores (MLSs) of 1.5, 2.6, 3.2, and 2.3 for these markers, with a peak multipoint MLS of 3.5.

Conclusions.—A complete genomic screen in families affected with late-onset AD identified 4 regions of interest after follow-up. Chromosome 12 gave the strongest and most consistent results with a peak multipoint MLS of 3.5, suggesting that this region contains a new susceptibility gene for AD. Additional analyses are necessary to identify the chromosome 12 susceptibility gene for AD and to follow up the regions of interest on chromosomes 4, 6, and 20.

ALZHEIMER DISEASE (AD) is a complex genetic disorder and represents the most common form of dementia in the elderly. More than 4 million persons in the United States are affected with AD, and that number continues to increase as the population ages. The cost in ancillary and nursing home care for patients with AD is more than $40 billion annually, while the additional psychological and physical burden to patients and family members is immeasurable.

From the Department of Medicine, Duke University Medical Center, Durham, NC (Drs Pericak-Vance, Yamaoka, Scott, Vance, Saunders, and Roses, Ms Bass and Menold, and Mr Gaskell); Molecular Neurogenetics Unit, Massachusetts General Hospital, Charlestown (Drs Haines and Terwedow); Department of Medical and Molecular Genetics, Indiana University Medical Center, Indianapolis (Dr Connelly); and Department of Psychiatry, University of California, Los Angeles (Dr Small). Dr Haines is now with the Program in Human Genetics, Vanderbilt University, Nashville, Tenn.

Reprints: Margaret A. Pericak-Vance, PhD, Section of Medical Genetics, Duke University Medical Center, Box 3445, Durham, NC 27710 (e-mail: mvp@locus.mc.duke.edu).

Downloaded from www.jama.com at Indiana University School of Medicine on May 28, 2010
until a second independent data set is tested, and is multianalytical because our approach combines the use of genetic model-dependent (ie, assumption of dominant or recessive inheritance) and model-independent (ie, no assumptions regarding the genetic model) methods of linkage analysis.

METHODS

Family Data

During the past 14 years, we have ascertained and sampled multiplex white families affected with late-onset AD (≥1 patient with AD per family) (present sample, N = 220 families) with a total of 450 affected individuals (affecteds) sampled for genetic studies. All sampled individuals diagnosed as having probable AD were examined by a neurologist or associated personnel of the Joseph and Kathleen Bryan Alzheimer’s Disease Research Center (ADRC) at Duke University, Durham, NC, the Massachusetts General Hospital (MGH) ADRC, Boston, the University of California, Los Angeles Neuropsychiatric Institute, or the Indiana Alzheimer’s Disease Center’s National Cell Repository, Indianapolis. The clinical diagnosis of AD was made using the National Institute of Neurological Disorders and Stroke–Alzheimer’s Disease and Related Disorders Association diagnostic criteria. A subset of 54 of these families (n = 16 for the genomic screen; n = 98 for the follow-up) was chosen for inclusion in a genome-wide screen and follow-up analysis to look for additional major genetic effects in late-onset AD. The 54 families were chosen because of their large size, the substantial number of family members with a clinical and/or pathological diagnosis of AD (ie, extended family history of AD, not just affected sibpairs [AD patients who are siblings]), and the number of AD patients with DNA samples available. DNA samples of the 54 families were negative for mutations in the APP gene and the common PS1 mutations, and showed no evidence of linkage to either of the PS1 and PS2 loci. Age at onset of AD was defined as previously described. The mean (SD) age at onset for the patients with AD in the 54 families was 72.7 (6.5) years.

The families were divided into 2 subsets, 1 for the initial genomic screen and 1 specifically for the follow-up analysis. The first set contained 16 families (genomic screen data set) and was selected to maximize efficiency for the genomic screen by allowing us to perform genotyping testing on as few individuals as possible while still preserving the majority of the power to detect linkage in the families analyzed. Thus, these 16 families were selected for the genomic screen because they were the largest families available at the time of initiation of the genomic screen. Examination of the families after selection for family size revealed that 6 of the original 16 genomic screen families had at least 1 patient with AD who did not have an APOE ε4 allele, suggesting that APOE ε4 may not play a major role in the risk for AD in these families. The 16 families included 195 family members with DNA samples and 52 AD patients with DNA samples. The 16 families had an average of 5.3 (range, 3-9) patients with AD per family and an average of 3.3 (range, 2-5) AD patients with DNA samples per family. The 38 families in the follow-up data set were chosen for the follow-up of interesting regions that were identified in the genomic screen. The 38 families in the follow-up data set included 216 family members with DNA samples and 89 AD patients with DNA samples. The combined data set included 351 family members with DNA samples. These families were selected at the time of follow-up from the then available families based on the same criteria used to select the first 16 families analyzed in the genomic screen. The number of patients and affected relative pairs (ARPs), including cousin pairs and sibpairs (aunt/uncle/nephew or niece, uncle/nephew or niece), in addition to sibpairs within the families is shown in Table 1. The families contained an average of 3.5 (range, 2-10) patients with AD per family and an average 2.3 (range, 2-6) AD patients with DNA samples per family. In addition to the 234 known affected sibpairs (114 with DNA), there were 71 affected avuncular pairs (17 with DNA), 22 affected cousin pairs (13 with DNA), and 62 parent-child pairs (4 with DNA) in the combined screening and follow-up data set. These data (Table 1) indicate that the families were often multigenerational in structure and not limited to affected sibpairs. All pairwise combinations of relative pairs were used in the ARP counts shown in Table 1.

DNA Analysis

Genomic DNA was extracted from whole blood using methods described previously. Marker genotyping was performed by means of fluorescence imaging (Molecular Dynamics SF Fluorimag, Duke University) or silver staining (MGI). Paternity was confirmed by examining a series of linked polymorphic markers (haplotypes) over several chromosomal regions. Microsatellite marker loci for the genomic screen were chosen for analysis based on a combination of high heterozygosity, approximate 10- to 15-centimorgan (cM) spacing, and ease of use (N = 280 markers). X chromosome markers were not examined as there is no evidence for X chromosome involvement in AD.

Statistical Analysis

The families were analyzed for linkage with a multianalytical approach that used genetic model-dependent (lod score) and model-independent (ARP [SimIRBD]) and sibpair [SIBPAL, ASPEX] linkage methods. Analysis of ARPs was performed in addition to the sibpair studies because many of the families contained ARPs in addition to affected sibpairs (Table 1). Given the suboptimal structure of most AD pedigrees (ie, many deceased family members without DNA information because of the late age of AD onset), our goal was to capture as much linkage information in the families as possible.

Genetic model-dependent methods make assumptions about the mode of inheritance (dominant or recessive) penetrance (ie, a fully penetrant trait vs a nonpenetrant or age-dependent penetrant trait). The use of genetic model-dependent methods is a powerful approach to gene mapping when the necessary assumptions are met. However, the power (the ability to detect linkage) is decreased markedly if the incorrect assumptions are made. Because in genetically complex diseases such as late-onset AD a disease gene model cannot be assumed with certainty, model-dependent methods must be interpreted with caution. Thus, genetic model-independent methods such as sibpair and ARP analyses, which examine allele sharing among ARPs, offer an attractive and complementary approach to linkage analysis.

The VITESSE program package was used in the lod score analysis. An autosomal dominant model was chosen because of the extensive evidence for vertical transmission in the families in our study (Table 1). For the lod score approach in the genomic screen, both age-dependent and low-penetrance (affecteds-only) models were used for analysis with the marker loci. The age curve used in the age-dependent analysis for assigning probabilities of carrying the AD gene to at-risk family members (children of pa-

Table 1.—Data Set Configuration: Number of Affected Relative Pairs*  

<table>
<thead>
<tr>
<th>Affected Relative Pairs</th>
<th>Genomic Screen Data Set (n=16 Families)</th>
<th>Follow-up Data Set (n=38 Families)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients per family, mean</td>
<td>5.3 (3.3)</td>
<td>3.5 (2.3)</td>
</tr>
<tr>
<td>Sibpairs</td>
<td>100 (46)</td>
<td>134 (68)</td>
</tr>
<tr>
<td>Avuncular pairs</td>
<td>40 (7)</td>
<td>31 (10)</td>
</tr>
<tr>
<td>Cousin pairs</td>
<td>15 (9)</td>
<td>7 (4)</td>
</tr>
<tr>
<td>Parent-child pairs</td>
<td>32 (2)</td>
<td>30 (2)</td>
</tr>
</tbody>
</table>

*All possible affected relative pair combinations scored.
†Over all families in set.
The disease allele frequency assumed for the AD locus was 0.001. Marker allele frequencies were estimated using a series of 50 to 100 unrelated white control subjects. The SimIBD program was used for the ARF analyses, the SIBPAL program in the SAGE program package was used in the sibpair analyses in the genomic screen, and the ASPEX program package, which performs exclusion mapping and multipoint maximum lod score (MLS) analysis, was used for the sibpair analysis in the follow-up studies. The order of the markers used in the multipoint analysis was determined from published sources (Collaborative Human Linkage Center, available at http://www.chlc.org) and by additional analysis in our laboratory on a subset of 20 of the largest families from the Centre d’Etude Polymorphisme Humain (M.A.P.-V. and J.L.H., unpublished data, 1997).

The critical values used to consider a region as interesting for follow-up analysis were a nominal P value of 0.05 or less for ARF and sibpair analyses and a 2-point or multipoint lod score of 1.0 or greater for the age-dependent or the affected-only lod score analysis. A multipoint lod score is a lod score that is based on analyzing multiple mapped markers simultaneously for linkage. Using these criteria, a region was marked as "interesting" if a marker exceeded the critical values on 2 of 3 tests or if any one critical value was met for 2 or more adjacent markers in a 30-cM region. Interesting regions then were grouped into 3 different levels to efficiently prioritize laboratory follow-up. Level 1 regions reached critical values on all 3 linkage tests (lod score [either model, ARF, sibpair]). Level 2 regions reached critical values on 2 tests. Level 3 regions reached critical values for 1 test for at least 2 more markers in a 30-cM region or had a multipoint lod score (either MLS or VITESSE) of 1.0 or greater.

Follow-up analyses of the interesting regions included genotyping at least 2 additional flanking markers and filling in all missing data, thereby capturing more than 95% of the potential genotypic information at each locus. Additional analyses were performed in both the genomic screening (n = 16) and follow-up (n = 98) data sets. These analyses included parametric 2-point lod score analysis using an affecteds-only model and assuming a 5% error rate in diagnosis of patients with AD (misdiagnosis parameter, ARF [SimIBD], and sibpair analysis [MLS]). A region remained interesting if there was evidence for linkage in the follow-up data set and if the overall evidence for linkage was stronger or equivalent to the initial finding.

In addition, the entire data set (n = 54) was stratified into 3 families tiers based on the APOE genotypes of diagnosed affected individuals. Tier 1 families were those with at least 1 patient without an APOE*4 allele. Tier 2 families were those in which genotyped patients all possessed at least 1 APOE*4 allele. Tier 3 families were those in which all AD affecteds were APOE*4 homozygotes. The stratification paradigm was developed before the availability of linkage results and was not applied to the data until after the completion of the follow-up studies inclusive of the elimination of laboratory errors by duplicate typing and haplotype analysis.

RESULTS

We genotyped 280 marker loci that covered the entire autosomal genome including the APOE gene and surrounding region on chromosome 19. Figure 1 shows the distribution of the 2-point parametric lod scores generated from the genomic screen using either model (affecteds-only or age-dependent). Only 30 markers (10.7%) generated lod scores of 1.0 or greater. Figure 2 shows the SimIBD P values for the genomic screen; 29 markers (10.4%) had P values of 0.05 or less. Figure 3 shows the distribution of the T statistic for the sibpair sharing results. Only 19 markers (6.8%) had P values of 0.05 or less (a value that corresponds to T = 1.645). Extreme departures from nominal significance levels were not observed; however, this is not unexpected especially if only 1 or a few AD loci are detected. Although some markers reached critical values in common for each linkage test (thus defining regions for follow-up), most markers reached critical value on only a single test. Fifteen regions on 12 chromosomes (1, 8, 10, 14, 15, and 20) met our cri-
Table 2.—Follow-up Analysis (n=38 Families)

<table>
<thead>
<tr>
<th>Region</th>
<th>Interesting Markers</th>
<th>Peak 2-Point Lod Score†</th>
<th>ARP (SimlBD)†</th>
<th>Sibpair (SIBPAL)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 4</td>
<td>D4S1629</td>
<td>&gt;1.0</td>
<td>.03</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>D4S2388</td>
<td>&lt;1.0</td>
<td>.13</td>
<td>.02</td>
</tr>
<tr>
<td>Chromosome 6</td>
<td>D6S1004</td>
<td>&gt;1.5</td>
<td>&lt;.001</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>D6S1019</td>
<td>&gt;1.0</td>
<td>&lt;.001</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>D6S391</td>
<td>&lt;1.0</td>
<td>.01</td>
<td>.14</td>
</tr>
<tr>
<td>Chromosome 12</td>
<td>D12S373</td>
<td>&gt;1.5</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>D12S1057</td>
<td>&gt;1.0</td>
<td>.03</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>D12S1042</td>
<td>&gt;2.0</td>
<td>.14</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>D12S390</td>
<td>&gt;2.1</td>
<td>.04</td>
<td>.02</td>
</tr>
<tr>
<td>Chromosome 20</td>
<td>D20S94</td>
<td>&lt;1.0</td>
<td>.50</td>
<td>.03</td>
</tr>
</tbody>
</table>

*Affected individuals only, autosomal dominant model.
†Program package used for analysis.
‡Multipoint maximum lod score (D20S173, D2S0171, and D20S94) of 1.0 or greater.

Table 3.—Summary Results for Chromosome 12* 

<table>
<thead>
<tr>
<th>Marker</th>
<th>2-Point Maximum Lod Score‡</th>
<th>APR (SimlBD) P Value‡</th>
<th>Maximum Lod Score§ (ASPEX)</th>
<th>2-Point Maximum Lod Score§</th>
<th>ARP (SimlBD) P Value§</th>
<th>Maximum Lod Score§ (ASPEX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D12S373</td>
<td>1.3 (0.3)</td>
<td>.04 (.71)</td>
<td>1.5</td>
<td>0.9</td>
<td>.13</td>
<td>1.0</td>
</tr>
<tr>
<td>D12S1057</td>
<td>1.6 (1.2)</td>
<td>.03 (.12)</td>
<td>2.6</td>
<td>2.9</td>
<td>.04</td>
<td>2.4</td>
</tr>
<tr>
<td>D12S1042</td>
<td>2.7 (1.2)</td>
<td>1.4 (.15)</td>
<td>3.2</td>
<td>2.2</td>
<td>.06</td>
<td>3.7</td>
</tr>
<tr>
<td>D12S390</td>
<td>2.2 (1.2)</td>
<td>.04 (.10)</td>
<td>2.3</td>
<td>2.9</td>
<td>.16</td>
<td>3.3</td>
</tr>
</tbody>
</table>

*ARP indicates affected relative pair; AD, Alzheimer disease; and SimlBD and ASPEX, program packages used for analysis.
†Results for families in the genomic screen data set (n=16) are given in parentheses.
‡Affected individuals only, autosomal dominant model, with 5% misdiagnosis.
§Peak multipoint maximum lod score is 3.5 for genomic screen and follow-up analyses (n=54 families) and 3.9 for families with at least 1 genotyped AD patient without an APOE*4 allele.

The various methods we used (lod score, ARP, and sibpair) do not examine the data in a fully independent fashion. Instead, these methods extract overlapping subsets of the complete information from the available data set, and the different methods can each provide additional information. For example, the MLS sibpair approach examines only among affected sibpairs whereas the SimlBD approach examines all ARPs, including aunts, uncles, and cousins. Smalley et al have shown that when the genetic model is unknown, a mixture of relative pairs can provide the most powerful approach to detecting linkage. Thus, each method has advantages and disadvantages, and a balanced approach may better decipher the unknown etiologic factors of the disease in question.

The ability to identify additional genetic factors for AD and then examine their interaction with both APOE and potential environmental factors is critical to fully dissecting the etiology of this complex disease. Our studies have shown that, at most, APOE accounts for 50% of the total genetic effect in AD, with a substantial amount of genetic variation still unexplained. Linkage analysis has been a powerful approach for unraveling the first 50% of this effect and remains the method of choice for unraveling the remaining genetic effect.
and sibpair P values ≤ 0.05) are not used to establish linkage but to ensure a low probability of type II error (concluding that there is no linkage when there is in fact linkage). These criteria resulted in identifying 15 regions of interest for follow-up. Follow-up analysis in the second data set resulted in 4 regions of primary interest. One region, near the centromere on chromosome 12, generated by far the strongest results. The lod score, ARP, and sibpair analysis of both the genomic screen and the follow-up data set strongly support a major genetic effect in this region. The MLS obtained from sibpair analysis may be interpreted similarly to a parametric lod score. Therefore, the peak multipoint MLS obtained from the chromosome 12 data (MLS of 3.5) exceeds the classic criterion lod score (≥ 3.0) for declaring linkage. The effect of this locus appears to be concentrated in tier 1 families, i.e., those families with several patients without an APOE4 allele. Although these results are not based on formal statistical testing, they suggest that the effect of the chromosome 12 locus is greatest in families in which the APOE4 allele plays a minimal role in AD.

The purpose of our study was to identify additional major susceptibility loci for AD. For this purpose, we chose as our genomic screening the largest multiplex families affected with late-onset AD. These families also tended to have several members with AD who did not have the APOE4/4 genotype. Therefore, these families had little influence of or risk for AD from the APOE4 locus. In addition, the family structure selected for the analysis extended beyond just sibpair data. These factors must be considered when identifying additional follow-up data sets.

This region on chromosome 12 contains many known genes, any of which now becomes a candidate locus for AD. One interesting and highly plausible candidate is the low-density lipoprotein receptor-related protein (LRP1) locus (LRP1 is the major receptor for APOE in neurons) for which evidence for genetic association has been previously reported. However, LRP1 maps about 10 cM distal to our peak region (data not shown). We have examined LRP1 in detail and found no evidence of significant linkage or association in our familial or sporadic AD data sets.

In summary, we have completed a genomic screen in late-onset AD to identify additional AD susceptibility loci. The data were analyzed with a multiallelic approach using several methods of analysis. Based on our criteria, we identified 15 regions of the genome that were targeted for our initial follow-up analysis. Follow-up analysis of these original 15 regions resulted in 4 regions on chromosomes 4, 6, 12, and 20 that remained interesting, with the strongest results lying near the centromere of chromosome 12. Additional analyses of this region are in progress and include genotype typing of new markers in the region, analysis of a comparable data set, association studies, and identification of potential candidate genes for analysis. Further follow-up of the regions on chromosomes 4, 6, and 20 is also ongoing. Finding these additional AD loci will foster new studies of the etiology of AD, including the examination of interactions among genes as well as the interactions between genetic and potential environmental causes. These studies should help to unravel the mystery of this common, genetic disease of older adults, and eventually should lead to effective therapies.

This work was supported by grants from the National Institute of Neurological Disorders and Stroke (Drs Pericak-Vance, Haines, and Vance), National Institute on Aging (Drs. Conneally, Small, and Saunders), the Alzheimer’s Disease and Related Disorders Association (Drs. Saunders Haines, and Pericak-Vance), and from multiple contributions from families and other donors to the Joseph and Kathleen Bryan Alzheimer’s Disease Research Center (ADRC) research fund, Durham, NC. The authors acknowledge the support of the personnel and the Hoffman-La Roche, the Massachusetts General Hospital (MGH) ADRC, the center and staff of the Section of Medical Genetics at Duke University, the Molecular Neurogenetics Unit at MGH, the Center for Human Genetics at Duke University Medical Center, and Valerie Roberts for her careful preparation of the manuscript. The authors also thank Neil Risch, PhD, for his many helpful discussions with respect to the linkage analysis, and the family members and patients with AD who participated in this study. Some of the results of this report were obtained by using the program package SAGE, which is supported by a US Public Health Service Research Grant from the National Center for Research Resources.

References

27. SAGE Statistical Analysis for Genetic Epidemiology [computer program]. Cleveland, Ohio: Dept of Biostatistics, Case Western Reserve University; 1994.

Downloaded from www.jama.com at Indiana University School of Medicine on May 28, 2010.