Results of a high-resolution genome screen of 437 Alzheimer's Disease families

Deborah Blacker[1], Lars Bertram[1], Aleister J. Saunders[1], Thomas J. Moscarillo[1],
Marilyn S. Albert[1], Howard Wiener[2], Rodney T. Perry[2], Julianne S. Collins[2], Lindy E. Harrell[2],
Rodney C.P. Go[2], Amy Mahoney[3], Terri Beaty[3], M. Danielle Fallin[3], Dimitrios Avramopoulos[3],
Gary A. Chase[3], Marshal F. Folstein[3], Melvin G. McInnis[3], Susan S. Bassett[3],
Kimberly J. Doheny[4], Elizabeth W. Pugh[4] and Rudolph E. Tanzi[1],* (The NIMH Genetics Initiative
Alzheimer's Disease Study Group)

[1]Massachusetts General Hospital, Charlestown, MA, USA, [2]University of Alabama at Birmingham, Birmingham, AL,
USA, [3]Johns Hopkins University, Baltimore, MD, USA and [4]Center for Inherited Disease Research, USA

Alzheimer's disease (AD) is a devastating neurodegenerative disorder of late life with complex inheritance. Mutations in three known genes lead to the rare early-onset autosomal dominant form of AD, while a common polymorphism (ε4) in the gene encoding apolipoprotein E (*APOE*) is a risk factor for more typical late-onset (>60 years) AD. A recent study concluded that there are up to four additional genes with an equal or greater contribution to the disease. We performed a 9 cM genome screen of 437 families with AD, the full National Institute of Mental Health (NIMH) sample, which has been carefully ascertained, evaluated and followed by our group over the last decade. Performing standard parametric and non-parametric linkage analyses, we observed a 'highly significant' linkage peak by Lander and Kruglyak criteria on chromosome 19q13, which probably represents *APOE*. Twelve additional locations—on 1q23, 3p26, 4q32, 5p14, 6p21, 6q27, 9q22, 10q24, 11q25, 14q22, 15q26 and 21q22—met criteria for 'suggestive' linkage [i.e. two-point lod score (TLS) ≥1.9 and/ or multipoint lod score (MLS) ≥2.2] in at least one of our analyses. Although some of these will surely prove to be false positives, these linkage signals should provide a valuable framework for future studies aimed at identifying additional susceptibility genes for late-onset AD.

INTRODUCTION

Alzheimer's disease (AD) is a devastating neurodegenerative disorder of late life with complex inheritance. A rare Mendelian subform with early onset exhibits autosomal dominant inheritance and both locus and allelic heterogeneity. This form can be caused by any of over 120 mutations in three known genes (reviewed in 1): the amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*). For the much more common late-onset form of AD, which shows complex inheritance, only one gene has been consistently associated with the disease. A common polymorphism (ε4) in the gene encoding apolipoprotein E (*APOE*) confers increased risk for late-onset AD (2,3), and lowers the age of onset in a dose-dependent fashion (4–6). Although *APOE* plays a major role in the risk for late-onset AD, a variety of genetic and epidemiological studies have shown that genes beyond *APOE* are involved in disease etiology (7,8). A recent study (9) in

which age of onset was examined as a quantitative trait estimated that up to four additional major genes as well as several minor AD genes remain to be identified. Although in theory *age of onset* is distinct from *risk*, for a late-onset disease like AD there is considerable overlap in practice, as has been observed in studies of *APOE*. In any case, efforts to find genes affecting risk and/or age of onset of AD have made limited progress.

Genome screens for AD reported to date (10–12) have selected their samples in a variety of ways, have used both disease risk and age of onset as outcomes, and have employed a wide variety of analytic methods. Many studies use overlapping portions of the NIMH sample used here (10–13), which only adds to the difficulty in understanding the relationship among the various results. Beyond *APOE*, for which nearby markers have shown detectable (but not necessarily statistically significant) linkage on chromosome 19q13, the major linkage findings reported to date have been on chromosomes

12 (11,13–15), 10 (16–18), and 9 (10,13). In addition, there have been over three dozen reports of genetic association, mostly with genes thought to influence pathobiological pathways in AD (1). Here we report our parametric and non-parametric linkage analyses of a 9 cM genome screen of the 437 multiplex families in the full NIMH Genetics Initiative sample.

RESULTS

The results of our affecteds-only parametric two-point analyses in FASTLINK (19,20) and non-parametric multipoint analyses in GENEHUNTER-PLUS (21,22) on the full NIMH sample are summarized in Table 1. Our analyses were performed for the total sample (which includes the 437 families in which no sampled affected subject experienced disease onset before age 50; see Methods), and in strata based on age of onset—a 'late-onset' stratum (families in which all sampled affected subjects onset ≥65), and an 'early/mixed-onset' stratum (all others, i.e. at least one sampled affected subject onsets <65).

Table 1 shows multipoint lod score (MLS; also shown in Figure 1 across the genome) and multipoint likelihood ratio Z-score (Z_{lr}), as well as two-point lod score (TLS) and estimated recombination fraction (θ) for all loci with MLS > 1.0 or 'suggestive linkage' in the two-point analyses alone. There is one region—on 19q13—that meets Lander and Kruglyak criteria for 'highly significant' linkage (23); this is presumed to represent *APOE*, as the best linked marker (*D19S178*) maps within ∼1 Mb upstream of this locus (24). There are 10 additional regions—1q23, 3p26, 4q32, 6p21, 6q27, 9q22, 10q24, 14q22, 15q26 and 21q22—that meet criteria for 'suggestive' linkage in the total sample and/or at least one onset-age stratum in at least one of the analyses (23).

We chose to analyze the data rounding up all rare marker allele frequencies to at least 0.01 (and adjusting the frequencies of the most common alleles to compensate) in order to decrease the likelihood of false-positive linkage due to the effect of rare alleles (25). We also analyzed the data using 'unrounded' allele frequencies estimated from the data, and the results did not change appreciably, although several regions had slightly weaker linkage signals (data not shown). In addition, because of concerns about population heterogeneity, we also analyzed the data excluding the 26 non-Caucasian families; there were no appreciable differences in the results (data not shown).

Last, we also analyzed the data allowing for heterogeneity using the two-point HLOD option in GENEHUNTER. Overall, the results were consistent with the regular two-point analyses. However, after accounting for heterogeneity, two new regions, on 5p14 [at 37 cM, HLOD (late-onset) = 2.6, $\alpha = 0.25$] and 11q25 [at 148 cM, HLOD (total sample) = 2.4, $\alpha = 0.23$], both of which gave modest signals previously, crossed the Lander and Kruglyak threshold, making a total of 12 'suggestive' regions in at least one of our analyses. In addition, one new minor peak appeared on 3p12 [at 114 cM, HLOD (total sample) = 1.6, $\alpha = 0.16$] where in the regular two-point analyses no linkage was observed, and one existing peak on 9q22 [at 101 cM, HLOD (late-onset) = 2.6, $\alpha = 0.29$] had a notably stronger signal.

Because of the overlap among our sample and these reported by other groups, we obtained a list of the NIMH subjects used in the report of Kehoe and colleagues from the NIMH Center for Genetic Studies (http://zork.wustl.edu/∼nimh), and repeated all analyses including only those individuals *not* used in the Kehoe report (13) or the follow-up Stage II screen of Myers and colleagues (11), which uses approximately the same sample. The remaining sample has markedly less power: it includes only 238 analyzable families with 808 individuals. Nonetheless, most of the observed peaks were still present, and were sometimes larger (1p36, 1p31, 3p26, 3p14, 6q27, 10p14, 13q32, 14q22, 19q13, and 21q22) in the smaller 'non-Kehoe' sample in at least one of the analyses (Table 2).

DISCUSSION

As expected, there is a 'highly significant' peak (TLS = 3.9, MLS = 7.7) in the early/mixed stratum on chromosome 19q13 at marker *D19S178*, which maps within ∼1 Mb of *APOE* (24). Interestingly, linkage evidence for late-onset families was less pronounced, and peaked 17 cM (∼14 Mb) proximal at marker *D19S433* (TLS = 2.3 at $\theta = 0.2$, MLS = 1.6), which is in agreement with previous reports in this and other samples, suggesting that *APOE*-ε4 acts in part through a decrease of onset age (6,26).

Our findings are consistent with prior findings on chromosome 9, although, as is common in the analysis of complex diseases (27,28), the location information is indistinct within each study, and can vary considerably across studies. Our chromosome 9q22 linkage at ∼101 cM is ∼10 cM distal to a 'suggestive' peak (at ∼91 cM) reported in the Kehoe genome screen in a subset of the NIMH sample (13), and again in the Myers (11) screen (stage II of the Kehoe screen), which used the same subset of the NIMH sample plus additional families from the USA and UK. These 'Kehoe' families appear to be driving the linkage, since there is no peak observed in this region in the 'non-Kehoe' families. In addition, our 9q22 peak and a smaller 9p21 peak at ∼55 cM are flanked by two peaks reported by Pericak-Vance and colleagues (10): a 'significant' *proximal* peak (at ∼43 cM) in the autopsy-confirmed subset of an overlapping sample, and a less pronounced *distal* peak (at ∼151 cM). The Pericak-Vance genome screen pooled portions of the NIMH sample with other samples, and had access to approximately the same families as the Kehoe group. While both chromosome 9 peaks are absent in the non-Kehoe families, Pericak-Vance and colleagues specifically noted that their peaks were observed consistently across the various subsamples (10).

The picture on chromosome 10 is also largely consistent with prior findings, and equally indistinct. The chromosome 10q24 finding at 134 cM was only marginally suggestive in the present analyses (Table 1), but we recently reported significant linkage nearby using a different set of markers in the same sample (18). Our findings were published in tandem with reports from two other groups demonstrating significant linkage ∼30–50 cM proximal (16,17). Very recently, Li and colleagues have reanalyzed their data using age of onset as a quantitative trait (12), and report a peak around 125 cM, i.e. closer to the linkage region that we had identified earlier (between 115 and 127 cM) (18). Interestingly, the multi-point peak in the present analyses (∼92 cM) lies between the peak observed by the latter two studies (16,17) (∼80 cM) and our earlier linkage region, while

Table 1. MLS and Z-scores for the likelihood ratio (Z_{lr}) plus TLS and recombination fractions (θ) for all chromosomal regions with MLS ≥ 1.0 plus those with 'suggestive linkage' (TLS ≥ 1.9) in the two-point analysis alone. 'Total' refers to the full NIMH sample ($n = 437$ families); families are 'late-onset' if all sampled affected individuals had onset ages ≥ 65 years, and 'early/mixed' otherwise. Chromosomal locations are given in Kosambi centi-Morgans according to the map provided by CIDR. Bold indicates regions of at least 'suggestive linkage' by Lander and Kruglyak criteria (23) in either the multipoint or two-point analyses or both

Chromosome	Stratum	Multipoint analyses			Two-point analyses	
		cM	MLS	Z_{lr} (*P*-value)	Marker (cM)	TLS (θ)
1p36	Late	27	1.1	2.3 (0.01)	D1S1597 (30)	1.1 (0.25)
1p31	Early/mixed	121	1.6	2.7 (0.0035)	*D1S551 (125)*	0.4 (0.3)
1q23	**Late**	**178**	**1.1**	**2.2 (0.01)**	***D1S1677* (178)**	**2.1 (0.2)**
3p26	**Early/mixed**	**9**	**2.0**	**3.0 (0.001)**	***D3S2387* (6)**	**2.0 (0.1)**
3p14	Total	88	1.1	2.2 (0.01)	*D3S4542* (90)	1.1 (0.25)
4q32	**Total**	**158**	**1.8**	**2.9 (0.002)**	***D4S1629* (158)**	**1.9 (0.25)**
4q35	Late	204	1.8	2.9 (0.002)	*D4S1652* (209)	0.9 (0.25)
5p14	Late	37	1.4	2.6 (0.005)	*D5S2845* (37)	0.5 (0.3)
5q35	Total	182	1.3	2.4 (0.008)	*D5S211* (182)	1.1 (0.25)
6p21	**Total**	**64**	**1.0**	**2.2 (0.01)**	***D6S1017* (64)**	**1.9 (0.25)**
6q27	**Total**	**181**	**2.2**	**3.2 (0.0007)**	*D6S1027* (187)	1.5 (0.25)
9p21	Total	55	1.3	2.5 (0.006)	*D9S1118* (58)	0.6 (0.3)
9q22	**Late**	**101**	**2.9**	**3.6 (0.0002)**	***D9S283* (95)**	**2.0 (0.2)**
10p14	Early/mixed	19	1.1	2.3 (0.01)	*D10S189* (19)	0.3 (0.35)
10q22	Total	92	1.8	2.9 (0.002)	*D10S1432* (94)	1.1 (0.25)
10q24	**Late**	**135**	**0.9**	**2.0 (0.02)**	***D10S1237* (135)**	**1.9 (0.25)**
11q25	Total	148	2.1	3.1 (0.001)	*D11S968* (148)	1.7 (0.25)
13q32	Late	77	1.1	2.3 (0.01)	*D13S793* (77)	0.5 (0.3)
14q22	**Early/mixed**	**55**	**2.2**	**3.2 (0.0007)**	***D14S587* (55)**	**2.1 (0.15)**
15q26	**Early/mixed**	**123**	**1.3**	**2.5 (0.006)**	***D15S642* (123)**	**2.1 (0.15)**
16q23	Total	99	1.4	2.6 (0.005)	*D16S516* (99)	0.9 (0.3)
19q13	**Early/mixed**	**70**	**7.7**	**5.9 (1.8×10^{-9})**	***D19S178* (68)**	**3.9 (0.1)**
21q22	**Total**	**37**	**0.1**	0.7 (0.24)	D21S1440 **(37)**	**2.5 (0.2)**
Xp21	Late	20	1.0	2.2 (0.01)	*DXS9902* (22)	1.6 (0.25)

the two-point peak here (134 cM) lies beyond our earlier peak and close to the region identified in the age at onset analyses of Li and colleagues (12). Similar to the findings on chromosome 9, almost no linkage evidence was observed in the subset of 'non-Kehoe' families alone (Table 2). In order to gain a better understanding of the broad signal on chromosome 10, additional analyses of the chromosome 10 data were performed in GENEFINDER (29), which estimates the location and a 95% confidence interval for an unobserved trait locus using multipoint marker information to approximate identity by descent (IBD) sharing in affected sib pairs, yielded strong signals with overlapping confidence intervals across a very broad region from 62 to 131 cM in the late-onset stratum [with Z-scores ranging from 2.5 ($P = 0.006$) to 3.1 ($P = 0.001$)], while the early/mixed stratum had a weaker but somewhat narrower signal from 13 to 37 cM [Z-score 1.6 ($P = 0.054$)]. Thus, it is still unclear whether there are one or two underlying genes located on the long arm of chromosome 10.

Many of the remaining linkage peaks lie near interesting candidate genes for AD. Our 'suggestive' linkage peak in late-onset families on chromosome 1q23 [which falls in a region where Kehoe and colleagues reported an MLS of 2.7 (13), but we see virtually no signal in the non-Kehoe sample (see Table 2)] is immediately adjacent to the gene encoding nicastrin (*NCSTN*), which binds presenilin and is required for γ-secretase activity and Aβ generation (30). The peak on 4q32 maps to a general region that also gave weak evidence for linkage in both genome scans by Pericak-Vance and colleagues (10,14). The gene encoding α-synuclein (*SNCA*, at 4q21), which has been suggested as an AD candidate gene (31), is more than 70 Mb proximal, and thus is unlikely to account for

this signal. The suggestive two-point signal on 6p21, a region that has also only been weakly linked or associated in previous studies (12–14,32) lies within 15 Mb of the gene encoding the tumor necrosis factor-α (*TNFA*), for which a significant haplotype association has been reported in the NIMH dataset (33). The suggestive multipoint signal on 6q27 (see Table 1) maps within ~8 Mb of the plasminogen gene (*PLG*), which is a protease implicated in cleavage of Aβ (34). This region also shows a fairly strong signal in the 'non-Kehoe' sample (Table 2). Moreover, a similar region has been implicated in a recent re-analysis of the Kehoe-subset of the NIMH dataset (35). The linked interval on chromosome 14q22 in the early/mixed onset stratum could reflect linkage to the early-onset AD gene *PSEN1*, which maps ~20 Mb distal of the best linked marker (*D14S587*; TLS = 2.1 at $\theta = 0.15$) and only ~3 Mb distal of the next best linked marker (*D14S588*; TLS = 1.9 at $\theta = 0.15$). This finding is consistent with recent reports that many early-onset AD patients carry mutations in this gene (36) and that a promoter polymorphism 48 bp upstream of *PSEN1* increases risk for early onset AD (37). To our knowledge, there are no prior reports of linkage or association near chromosome 15q26, and it should be noted that this 'suggestive' peak is driven almost entirely by a single marker (*D15S642*) at the q-terminal and thus is less likely to be confirmed. Finally, it is of note that the suggestive two-point peak on chromosome 21q22 lies ~12 Mb from the early-onset AD gene *APP*, but only 5 Mb from *BACE2* [which has been reported to cleave APP near the α-secretase site within the Aβ-domain and is the homolog of the β-secretase (*BACE*)] (38). Interestingly, there is a linkage signal on chromosome 11q25 (HLOD = 2.4, just over the Lander and Kruglyak threshold for suggestive linkage when
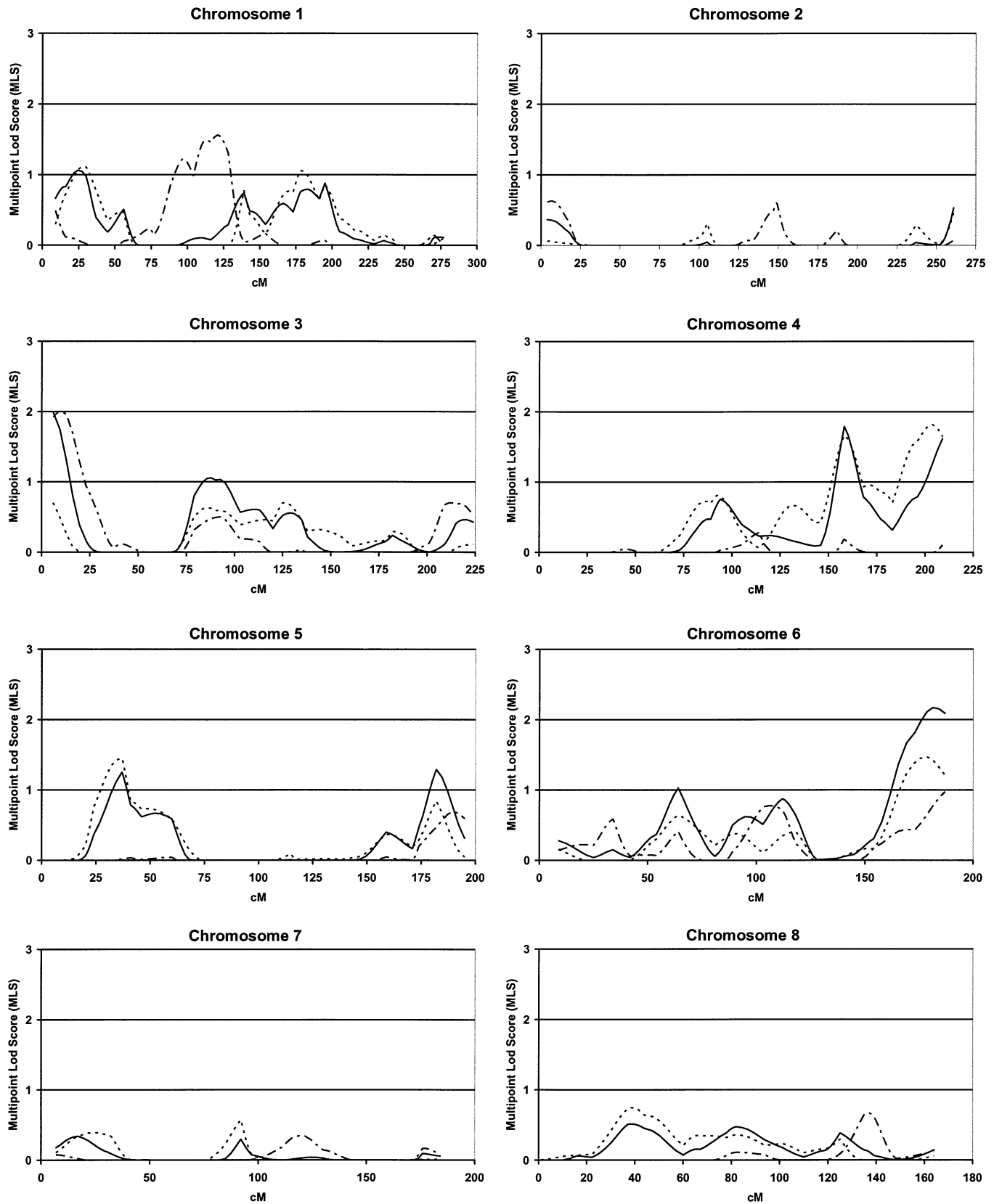
Figure 1. Non-parametric multipoint lod score curves for each chromosome in the genome screen. Families were considered 'late-onset' if all sampled affected individuals had onset ages ≥65 years, and 'early/mixed-onset' otherwise. Marker locations are given in Kosambi centi-Morgans according to the map supplied by CIDR. Note that the scale used for chromosome 19 is different from that for the remainder of the genome.
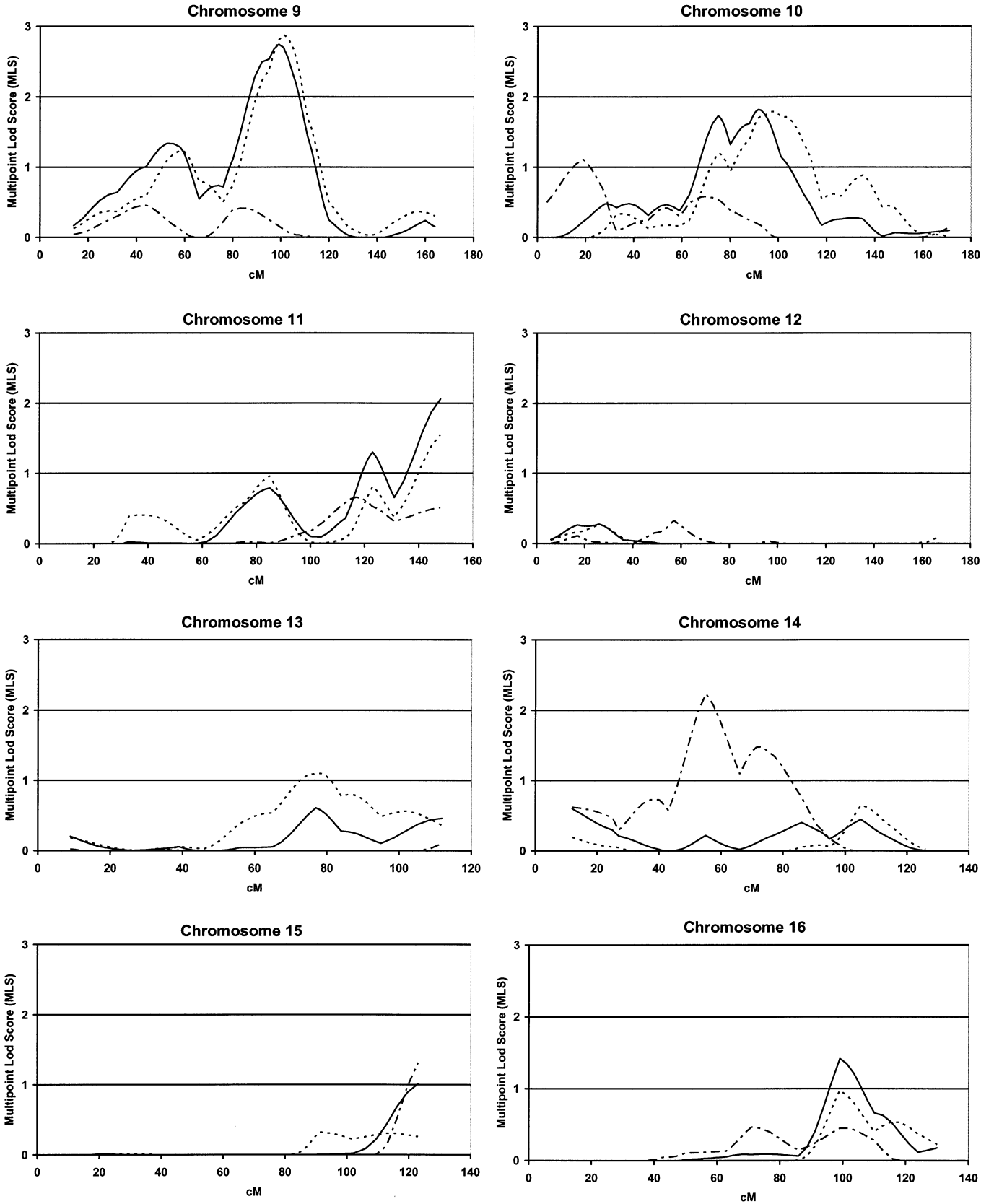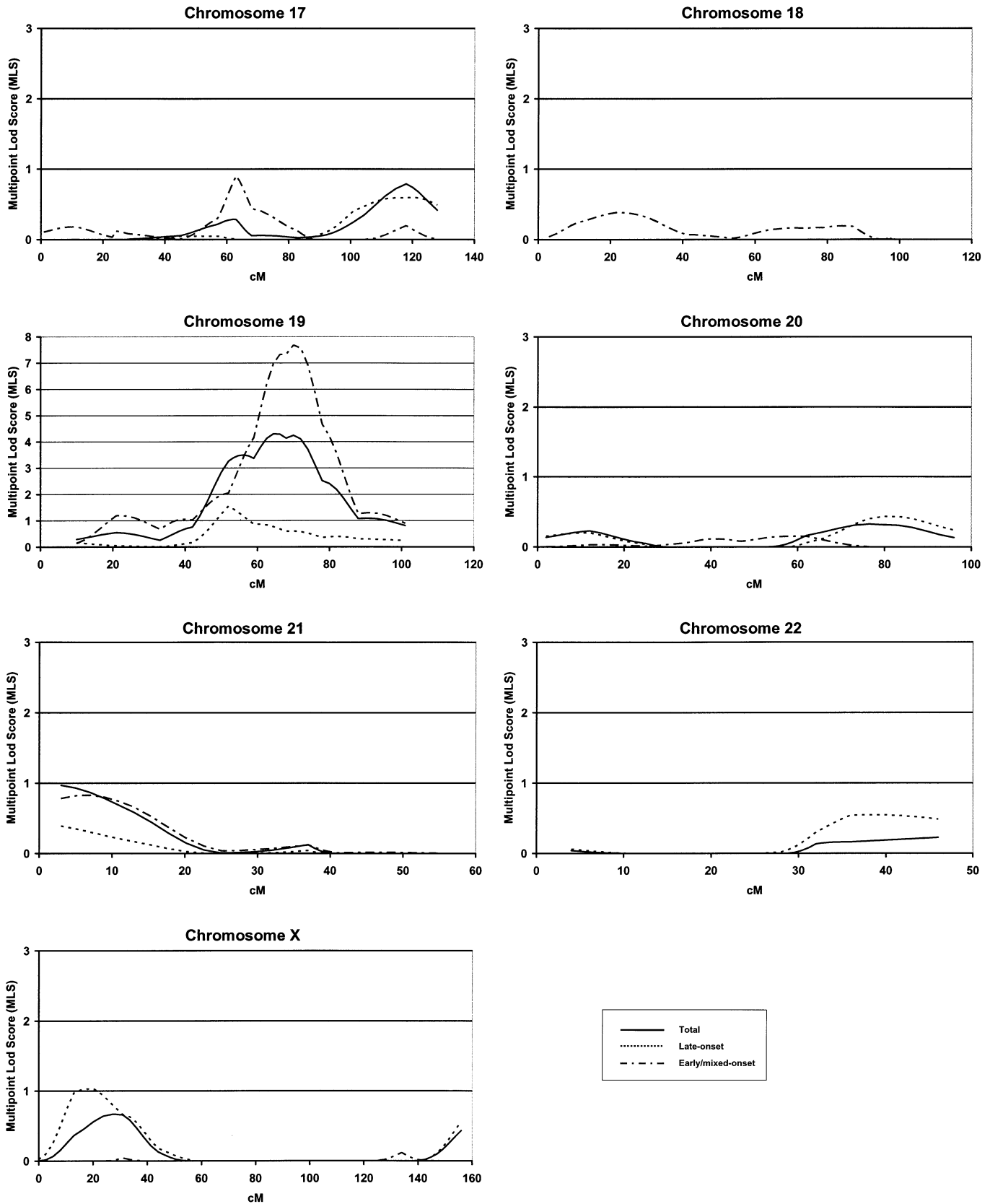
Figure 1 *continued.*

Figure 1 *continued*.

Table 2. Results for the 'non-Kehoe' families, those not included in the genome screen of Kehoe *et al.* (13). Multipoint non-parametric lod scores (MLS) and Z-scores for the likelihood ratio (Z_{lr}) plus two-point parametric lod scores (TLS) and recombination fractions (θ) for all chromosomal regions appearing in Table 1 [i.e. those with MLS $\geq$ 1.0 plus those with 'suggestive linkage' (TLS $\geq$ 1.9) in the two-point analysis alone in the full NIMH sample]. 'Total non-Kehoe' refers to the entire non-Kehoe sample ($n = 238$ families); families are 'late-onset' if all sampled affected individuals had onset ages $\geq$65 years, and 'early/mixed' otherwise. Chromosomal locations are given in Kosambi centi-Morgans according to the map provided by CIDR. Bold indicates regions of at least 'suggestive linkage' by Lander and Kruglyak criteria (23) in either the multipoint or two-point analyses or both. The shaded regions are those that appear in bold in Table 1 [i.e. those with at least 'suggestive' linkage by Lander and Kruglyak criteria (23) in either the multipoint or two point analyses or both in the full NIMH sample]

Chromosome	Stratum	Multipoint analyses			Two-point analyses	
		cM	MLS	Z_{lr} (*P*-value)	Marker (cM)	TLS (θ)
1p36	Late	27	1.3	2.4 (0.008)	*D1S1597* (30)	1.6 (0.15)
1p31	Early/mixed	121	1.7	2.8 (0.003)	*D1S551* (125)	0.4 (0.2)
1q23	Late	178	0.8	1.9 (0.03)	*D1S1677* (178)	1.1 (0.15)
3p26	**Early/mixed**	**9**	**2.9**	**3.7 (0.0001)**	***D3S2387* (6)**	**2.0 (0.1)**
3p14	**Total**	88	1.5	2.7 (0.004)	***D3S4542* (90)**	**2.1 (0.15)**
4q32	**Total**	158	1.1	2.3 (0.01)	***D4S1629* (158)**	**1.9 (0.2)**
4q35	Late	204	1.1	2.2 (0.01)	*D4S1652* (209)	0.6 (0.2)
5p14	Late	37	0.2	0.9 (0.18)	*D5S2845* (37)	0 (0.3)
5q35	Total	182	0.7	1.8 (0.04)	*D5S211* (182)	0.3 (0.25)
6p21	Total	64	0.2	0.9 (0.18)	*D6S1017* (64)	1.2 (0.2)
6q27	**Total**	**181**	**2.7**	**3.5 (0.0002)**	*D6S1027* (187)	1.8 (0.2)
9p21	Total	55	0.7	2.1 (0.02)	*D9S1118* (58)	0.2 (0.3)
9q22	Late	101	0.2	1.0 (0.16)	*D9S283* (95)	0.8 (0.2)
10p14	Early/mixed	19	1.2	2.3 (0.01)	*D10S189* (19)	0.9 (0.15)
10q22	Total	92	0.2	1.0 (0.16)	*D10S1432* (94)	0.1 (0.35)
10q24	Late	135	0.03	0.4 (0.34)	D10S1237 (135)	0.4 (0.25)
11q25	Total	148	1.1	2.3 (0.01)	*D11S968* (148)	0.7 (0.25)
13q32	Late	77	1.4	2.5 (0.006)	*D13S793* (77)	0.5 (0.3)
14q22	**Early/mixed**	**55**	**2.9**	**3.6 (0.0002)**	***D14S587* (55)**	**2.7 (0.15)**
15q26	**Early/mixed**	123	1.2	2.4 (0.008)	***D15S642* (123)**	**2.1 (0.15)**
16q23	Total	99	0.6	1.6 (0.05)	*D16S516* (99)	0 (0.5)
19q13	**Early/mixed**	**70**	**7.6**	**5.9 (1.8×10^{-9})**	***D19S178* (68)**	**4.5 (0.05)**
21q22	Total	37	0.1	0.5 (0.3)	D21S1440 (37)	1.6 (0.2)
Xp21	Late	20	0.2	1.0 (0.16)	*DXS9902* (22)	1.0 (0.2)

accounting for heterogeneity, and MLS = 2.1 in the regular multipoint analyses) that maps within ∼20 Mb of *BACE* itself.

Of note, no evidence of linkage was observed in the analyses here on chromosome 12 (maximum MLS = 0.2 at 57 cM for the early/mixed stratum; maximum TLS = 0.8 for marker *C12S916* at 57 cM), where our group has reported a highly significant association finding with nearby alpha-2-macroglobulin (*A2M*) (39), and follow-up analyses in the full NIMH sample used here remain strong (Saunders *et al.*, submitted). Although this association finding has been controversial, and the great majority of follow-up studies have been negative (for example, see 40,41), several other groups have reported modest linkage signals in this general region (13–15) or association with A2M itself (for a recent review see 47).

We report here the results of a high-resolution genome screen of a large, uniformly ascertained and evaluated sample of AD families. Although our results are only partially concordant with prior published reports, and the signals generated are not amenable to rapid gene discovery, this result is consistent with efforts to map other complex diseases. Our analyses reveal the expected strong signal for *APOE* on chromosome 19, and a number of more modest linkage peaks that deserve follow-up with linkage analysis of additional microsatellite markers as well as association testing of polymorphisms in positional candidate genes. If these analyses continue to provide support for the presence of AD susceptibility genes in these regions, systematic linkage disequilibrium mapping may be warranted. While some of the more marginal peaks in our study probably represent

false-positive findings, others, particularly those with relatively stronger signals and/or prior reports (e.g. on chromosomes 9q and 10q), are more likely to harbor genuine AD susceptibility genes. The substantial effort required to identify and characterize as many of these genes as possible will greatly facilitate the development of strategies for treatment, early intervention, and prevention of this devastating disease.

MATERIALS AND METHODS

Sample

Subjects were collected as part of the NIMH Genetics Initiative following a standardized protocol developed and tested by the three sites (42). Each site collected a wealth of information from family members, medical records and direct examination of the subjects in order to reach a diagnosis of AD by NINCDS/ADRDA criteria (26,43). Over the nearly 10 years of follow-up reflected in this dataset, autopsies were obtained on 326 affected subjects (48% of the deceased), and a clinical diagnosis of AD has been confirmed at autopsy by standard research criteria (44,45) in 94% of the cases. The NIMH sample constitutes a national archive for AD genetic research; all DNA samples are stored in a centralized cell repository at Rutgers University, and data are stored in a centralized data management center at Washington University in St Louis.

Outside investigators (10–13,16,35,46) have used subsets of the first ~60% of the samples collected.

In order to select a sample more closely resembling more typical AD than had been studied previously, ascertainment criteria for the NIMH sample required only an affected relative pair, with no additional requirements regarding age of onset or family structure. The original sample included a total of 1527 subjects from 457 families. Six families (14 individuals) with only one sampled affected individual available were excluded. In addition, because onset prior to age 50 is strongly associated with *PS1* (47), we excluded 14 additional families (with 74 individuals) in which any sampled individual had an onset age under 50 years.

The total sample for the present analyses is comprised of 1439 individuals from 437 families, including 994 affected individuals (mean age of onset 72.4 ± 7.7 years, range 50–97 years), 411 unaffected, and 34 with phenotype unknown. Families were classified as 'late-onset' (320 families) when all sampled affecteds had onset ages ≥ 65 years, and 'early/mixed' (117 families) otherwise. The family structure of the sample varies widely: 357 families are nuclear, 36 are extended families including one or more affected sibpairs, and 58 are extended families without an affected sibpair.

Genotyping techniques

Genotyping was performed at the Center for Inherited Disease Research (CIDR) using a modification of the CHLC version 9 marker set (381 markers, average spacing 9 cM, average heterozygosity 0.76, maximum gap 19 cM). Allele calls were reviewed by two technicians, blind to family structure and phenotype. Binning of all allele sizes for each marker was performed using the FASTCLUS procedure in SAS. Calls falling outside of bins and Mendelian inconsistencies, identified using GAS, were reviewed for obvious laboratory errors. Four blind duplicates, four positive and two negative control samples were run with every 86 study samples. The error rate for the genome scan, based on 28 023 paired genotypes from these blind duplicates, was 0.22%. The overall missing data rate was 3.6%. The rate of remaining Mendelian inconsistencies was 0.08%. See also www.cidr.jhmi.edu for details on the marker set and genotyping methods.

Statistical techniques

We initially calculated allele frequencies with maximum likelihood estimates using all available data with the program USERM13 (48) based on the program MENDEL (49), but for a fraction of the loci with greater than 20 alleles (six in total) we used a compression algorithm that we developed to reduce the number of alleles at a locus by combining the lowest frequency alleles into a single allele, and renumbering the remaining alleles to maintain consecutive numbering. Except for one marker (*D2S2739*) with 47 alleles, we were able to maintain the heterozygosity of the sample while compressing the most polymorphic loci to fewer than 20 alleles. When comparing linkage results using compressed versus non-compressed alleles, the differences in calculated lod scores never exceeded 0.05 in absolute value (data not shown). In order to protect our results against false-positive findings due to rare alleles (25),

we rounded up rare allele frequencies to 0.01 while adjusting the most common alleles proportionally to compensate, and used these for our primary analyses.

'Affecteds only' parametric two-point analyses were performed in FASTLINK (19,20) using a dominant inheritance model with a disease gene frequency of 0.02. While affecteds only analysis can sometimes lead to inflation of the lod score, especially in the presence of missing data or misspecified allele frequencies (50,51), we felt that it was still preferable for the present analyses given the uncertainties regarding the true phenotype of 'unaffected' individuals and the shape of the age of onset distribution in a familial sample. Penetrance in affected individuals was set to correspond to phenocopy rates of 5% for definite AD ($n = 289$), 10% for probable AD ($n = 640$), and 20% for possible AD ($n = 65$). Multipoint non-parametric analyses were performed in GENEHUNTER-PLUS using the ASM program (exponential model) to calculate MLS and Z_{lr} scores (21,22). Genetic maps and intermarker distances were provided by CIDR (www.cidr.jhmi.edu) and are similar but not identical to those from the Marshfield Center for Medical Genetics (http://research.marshfieldclinic.org/genetics/), and thus use the Kosambi map function. Both parametric and non-parametric analyses were repeated limiting the sample to the 411 white families, and also limiting the sample to the 238 families not included in the analyses of Kehoe *et al.*, and Myers *et al.* (11,13). In addition, we estimated two-point lod-scores under the assumption of heterogeneity across families (HLOD) in GENEHUNTER PLUS using the same inheritance models as in the TLS analyses (see above). GENEFINDER (29) was used to estimate the location and 95% confidence intervals around an underlying AD locus on chromosome 10q. The program utilizes multipoint marker information to approximate IBD sharing in affected sib pairs, using a robust, model-free approach similar to that used in GENEHUNTER. Based on the assumptions that a single disease locus exists in the map region being studied and that the true IBD sharing between affected siblings will be maximized at the trait locus, GENEFINDER employs generalized estimating equations to estimate the map location of the disease locus.

REFERENCES

1. Tanzi, R.E. and Bertram, L. (2001) New frontiers in Alzheimer's disease genetics. *Neuron*, **32**, 181–184.
2. Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S. and Roses, A.D. (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl Acad. Sci. USA*, **90**, 1977–1981.
3. Saunders, A.M., Strittmatter, W.J., Schmechel, D., George-Hyslop, P.H., Pericak-Vance, M.A., Joo, S.H., Rosi, B.L., Gusella, J.F., Crapper-MacLachlan, D.R., Alberts, M.J. *et al.* (1993) Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*, **43**, 1467–1472.
4. Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L. and Pericak-Vance, M.A. (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, **261**, 921–923.
5. Farrer, L.A., Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W.A., Mayeux, R., Myers, R.H., Pericak-Vance, M.A., Risch, N. and van Duijn, C.M. (1997) 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*, **278**, 1349–1356.
6. Meyer, M.R., Tschanz, J.T., Norton, M.C., Welsh-Bohmer, K.A., Steffens, D.C., Wyse, B.W. and Breitner, J.C. (1998) APOE genotype predicts when—not whether—one is predisposed to develop Alzheimer disease. *Nat. Genet.*, **19**, 321–322.
7. Bergem, A.L., Engedal, K. and Kringlen, E. (1997) The role of heredity in late-onset Alzheimer disease and vascular dementia. A twin study. *Arch. Gen. Psychiat.*, **54**, 264–270.
8. Payami, H., Grimslid, H., Oken, B., Camicioli, R., Sexton, G., Dame, A., Howieson, D. and Kaye, J. (1997) A prospective study of cognitive health in the elderly (Oregon Brain Aging Study): effects of family history and apolipoprotein E genotype. *Am. J. Hum. Genet.*, **60**, 948–956.
9. Daw, E.W., Payami, H., Nemens, E.J., Nochlin, D., Bird, T.D., Schellenberg, G.D. and Wijsman, E.M. (2000) The number of trait loci in late-onset Alzheimer disease. *Am. J. Hum. Genet.*, **66**, 196–204.
10. Pericak-Vance, M.A., Grubber, J., Bailey, L.R., Hedges, D., West, S., Santoro, L., Kemmerer, B., Hall, J.L., Saunders, A.M., Roses, A.D. *et al.* (2000) Identification of novel genes in late-onset Alzheimer's disease. *Exp. Gerontol.*, **35**, 1343–1352.
11. Myers, A., Wavrant De-Vrieze, F., Holmans, P., Hamshere, M., Crook, R., Compton, D., Marshall, H., Meyer, D., Shears, S., Booth, J. *et al.* (2002) Full genome screen for Alzheimer disease: Stage II analysis. *Am. J. Med. Genet.*, **114**, 235–244.
12. Li, Y.J., Scott, W.K., Hedges, D.J., Zhang, F., Gaskell, P.C., Nance, M.A., Watts, R.L., Hubble, J.P., Koller, W.C., Pahwa, R. *et al.* (2002) Age at onset in two common neurodegenerative diseases is genetically controlled. *Am. J. Hum. Genet.*, **70**, 985–993.
13. Kehoe, P., Wavrant-De Vrieze, F., Crook, R., Wu, W.S., Holmans, P., Fenton, I., Spurlock, G., Norton, N., Williams, H., Williams, N. *et al.* (1999) A full genome scan for late onset Alzheimer's disease. *Hum. Mol. Genet.*, **8**, 237–245.
14. Pericak-Vance, M.A., Bass, M.P., Yamaoka, L.H., Gaskell, P.C., Scott, W.K., Terwedow, H.A., Menold, M.M., Conneally, P.M., Small, G.W., Vance, J.M. *et al.* (1997) Complete genomic screen in late-onset familial Alzheimer disease. Evidence for a new locus on chromosome 12. *JAMA*, **278**, 1237–1241.
15. Rogaeva, E., Premkumar, S., Song, Y., Sorbi, S., Brindle, N., Paterson, A., Duara, R., Levesque, G., Yu, G., Nishimura, M. *et al.* (1998) Evidence for an Alzheimer disease susceptibility locus on chromosome 12 and for further locus heterogeneity. *JAMA*, **280**, 614–618.
16. Myers, A., Holmans, P., Marshall, H., Kwon, J., Meyer, D., Ramic, D., Shears, S., Booth, J., DeVrieze, F.W., Crook, R. *et al.* (2000) Susceptibility locus for Alzheimer's disease on chromosome 10. *Science*, **290**, 2304–2305.
17. Ertekin-Taner, N., Graff-Radford, N., Younkin, L.H., Eckman, C., Baker, M., Adamson, J., Ronald, J., Blangero, J., Hutton, M. and Younkin, S.G. (2000) Linkage of plasma Abeta42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Science*, **290**, 2303–2304.
18. Bertram, L., Blacker, D., Mullin, K., Keeney, D., Jones, J., Basu, S., Yhu, S., McInnis, M.G., Go, R.C., Vekrellis, K. *et al.* (2000) Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. *Science*, **290**, 2302–2303.
19. Schaffer, A.A., Gupta, S.K., Shriram, K. and Cottingham, R.W. Jr (1994) Avoiding recomputation in linkage analysis. *Hum. Hered.*, **44**, 225–237.
20. Cottingham, R.W. Jr, Idury, R.M. and Schaffer, A.A. (1993) Faster sequential genetic linkage computations. *Am. J. Hum. Genet.*, **53**, 252–263.
21. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. and Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347–1363.
22. Kong, A. and Cox, N.J. (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am. J. Hum. Genet.*, **61**, 1179–1188.
23. Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.*, **11**, 241–247.
24. UCSC Human Genome Project Working Draft at URL: http://genome. ucsc.edu/index.html (April 2002 freeze).
25. Ott, J. (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. *Am. J. Hum. Genet.*, **51**, 283–290.
26. Blacker, D., Haines, J.L., Rodes, L., Terwedow, H., Go, R.C., Harrell, L.E., Perry, R.T., Bassett, S.S., Chase, G., Meyers, D. *et al.* (1997) ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. *Neurology*, **48**, 139–147.
27. Roberts, S.B., MacLean, C.J., Neale, M.C., Eaves, L.J. and Kendler, K.S. (1999) Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am. J. Hum. Genet.*, **65**, 876–884.
28. Hauser, E.R. and Boehnke, M. (1997) Confirmation of linkage results in affected-sib-pair linkage analyses for complex genetic traits. *Am. J. Hum. Genet.*, **61** (suppl.), A278.
29. Liang, K.Y., Chiu, Y.F. and Beaty, T.H. (2001) A robust identity-by-descent procedure using affected sib pairs: multipoint mapping for complex diseases. *Hum. Hered.*, **51**, 64–78.
30. Yu, G., Nishimura, M., Arawaka, S., Levitan, D., Zhang, L., Tandon, A., Song, Y.Q., Rogaeva, E., Chen, F., Kawarai, T. *et al.* (2000) Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. *Nature*, **407**, 48–54.
31. Xia, Y., Rohan de Silva, H.A., Rosi, B.L., Yamaoka, L.H., Rimmler, J.B., Pericak-Vance, M.A., Roses, A.D., Chen, X., Masliah, E., DeTeresa, R. *et al.* (1996) Genetic studies in Alzheimer's disease with an NACP/ alpha-synuclein polymorphism. *Ann. Neurol.*, **40**, 207–215.
32. Hiltunen, M., Mannermaa, A., Thompson, D., Easton, D., Pirskanen, M., Helisalmi, S., Koivisto, A.M., Lehtovirta, M., Ryynanen, M. and Soininen, H. (2001) Genome-wide linkage disequilibrium mapping of late-onset Alzheimer's disease in Finland. *Neurology*, **57**, 1663–1668.
33. Collins, J.S., Perry, R.T., Watson, B. Jr, Harrell, L.E., Acton, R.T., Blacker, D., Albert, M.S., Tanzi, R.E., Bassett, S.S., McInnis, M.G. *et al.* (2000) Association of a haplotype for tumor necrosis factor in siblings with late-onset Alzheimer disease: the NIMH Alzheimer Disease Genetics Initiative. *Am. J. Med. Genet.*, **96**, 823–830.
34. Selkoe, D.J. (2001) Clearing the brain's amyloid cobwebs. *Neuron*, **32**, 177–180.
35. Olson, J.M., Goddard, K.A. and Dudek, D.M. (2002) A second locus for very-late-onset Alzheimer Disease: a genome scan reveals linkage to 20p and epistasis between 20p and the amyloid precursor protein region. *Am. J. Hum. Genet.*, **71**, 154–161.
36. Rogaeva, E.A., Fafel, K.C., Song, Y.Q., Medeiros, H., Sato, C., Liang, Y., Richard, E., Rogaev, E.I., Frommelt, P., Sadovnick, A.D. *et al.* (2001) Screening for PS1 mutations in a referral-based series of AD cases: 21 novel mutations. *Neurology*, **57**, 621–625.
37. van Duijn, C.M., Cruts, M., Theuns, J., Van Gassen, G., Backhovens, H., van den Broeck, M., Wehnert, A., Serneels, S., Hofman, A. and Van Broeckhoven, C. (1999) Genetic association of the presenilin-1 regulatory region with early-onset Alzheimer's disease in a population-based sample. *Eur. J. Hum. Genet.*, **7**, 801–806.

38. Saunders, A.J., Kim, T.-W. and Tanzi, R. (1999) BACE maps to chromosome 11 and a BACE homolog, BACE2, reside in the obligate Down syndrome region of chromosome 21. *Science*, **286**, 1255a.

39. Blacker, D., Wilcox, M.A., Laird, N.M., Rodes, L., Horvath, S.M., Go, R.C., Perry, R., Watson, B., Bassett, S.S., McInnis, M.G. *et al.* (1998) Alpha-2 macroglobulin is genetically associated with Alzheimer disease. *Nat. Genet.*, **19**, 357–360.

40. Wang, X., Luedecking, E.K., Minster, R.L., Ganguli, M., DeKosky, S.T. and Kamboh, M.I. (2001) Lack of association between alpha2-macroglobulin polymorphisms and Alzheimer's disease. *Hum. Genet.*, **108**, 105–108.

41. Koster, M.N., Dermaut, B., Cruts, M., Houwing-Duistermaat, J.J., Roks, G., Tol, J., Ott, A., Hofman, A., Munteanu, G., Breteler, M.M. *et al.* (2000) The alpha2-macroglobulin gene in AD: a population-based study and meta-analysis. *Neurology*, **55**, 678–684.

42. Blacker, D., Albert, M.S., Bassett, S.S., Go, R.C., Harrell, L.E. and Folstein, M.F. (1994) Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative. *Arch. Neurol.*, **51**, 1198–1204.

43. McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. and Stadlan, E.M. (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, **34**, 939–944.

44. Khachaturian, Z.S. (1985) Diagnosis of Alzheimer's disease. *Arch. Neurol.*, **42**, 1097–1105.

45. Group, N.-R.I.W. (1997) Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. *Neurobiol. Aging*, **18**, S1–2.

46. Olson, J.M., Goddard, K.A. and Dudek, D.M. (2001) The amyloid precursor protein locus and very-late-onset Alzheimer disease. *Am. J. Hum. Genet.*, **69**, 895–899.

47. Bertram, L. and Tanzi, R.E. (2001) Of replications and refutations: the status of Alzheimer's disease genetic research. *Curr. Neurol. Neurosci. Rep.*, **1**, 442–450.

48. Boehnke, M. (1991) Allele frequency estimation from data on relatives. *Am. J. Hum. Genet.*, **48**, 22–25.

49. Lange, K., Weeks, D. and Boehnke, M. (1988) Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genet. Epidemiol.*, **5**, 471–472.

50. Goldin, L.R. and Gershon, E.S. (1993) Linkage of Alzheimer's disease to chromosome 21 and chromosome 19 markers: effect of age of onset assumptions. *Genet. Epidemiol.*, **10**, 449–454.

51. Knapp, M., Seuchter, S.A. and Baur, M.P. (1993) The effect of misspecifying allele frequencies in incompletely typed families. *Genet. Epidemiol.*, **10**, 413–418.