REJUVENATION RESEARCH
Volume 9, Number 1, 2006

Membership in Genetic Groups Predicts Alzheimer Disease

ELIZABETH H. CORDER,[4] RONG HUANG,[1] HEATHER M. CATHCART,[1] IRENE S. LANHAM,[1] GINNY R. PARKER,[1] DANNY CHENG,[1] SUZANNE SMITH,[2] and SHIRLEY E. PODUSLO[1–3,*]

ABSTRACT

The multiple polymorphisms contributing to Alzheimer disease (AD) have been difficult to identify. Three essentially sufficient risk sets were found using a fuzzy latent classification statistical model; that is, grade-of-membership analysis, and genotypes for APOE, APOCI, LDLr, cystatin C, and cathepsin D (180 cases, 120 controls). These were: (a) CST3:GA and CTSD:CT; (b) APOE44 and LDLr8:GG and LDLr13:TT; and (c) APOE34 and LDLr13:TC. Consonance with one of the groups and high aggregate membership carried >800-fold elevated risk for AD. The absence of these combinations defined low risk. APOE3/− with heterozygous promoter and receptor genotypes predicted long life without dementia.

INTRODUCTION

THE SEQUENTIAL BRAIN CHANGES resulting in Alzheimer disease (AD) occur at widely varying rates for individuals. The timing of initial lesions, common by age 55, and the subsequent rate of compromise in brain function are strongly influenced by which pair of apolipoprotein E (APOE) alleles (i.e., genotype) has been inherited.[1–4] Inherited variation in APOE accounts for about one third of cases taken as a single factor at current life expectancies.[5] Regardless, lesions characteristic of AD (i.e., senile plaque and neurofibrillary tangles) are almost universal by age 80. However, not all nominally high-risk APOE4 allele carriers are affected by age 100; some nominally low-risk APOE2 allele carriers are affected nonetheless. Thus, other factors must play a role.

MATERIALS AND METHODS

Many biologically plausible candidate genes are inconsistently associated with AD when investigated singly: sufficient risk sets of genetic variants that result in AD, which might be more easily replicated, have not been identified. The authors defined three multilocus genotypes; that is, pairs of inherited alleles at multiple genomic locations, which had extremely high risk for AD. The candidate loci were APOE, APOE promoter polymorphisms at positions −491 and −427, APOCI alleles (in linkage disequi-

[1]Medical College of Georgia, Institute of Molecular Medicine and Genetics, Augusta, Georgia.
[2]Department of Neurology, Medical College of Georgia, Augusta, Georgia.
[3]Veterans Administration Medical Center, Augusta, Georgia.
[4]Center for Demographic Studies, Duke University, Durham, North Carolina.

librium with APOE alleles), polymorphisms within exons 8 and 13 of the low-density lipoprotein receptor (LDLr) for APOE, as well as exon 1 of cystatin C (CTS3) and exon 2 of cathepsin D (CSTD). Each of the encoded proteins colocalizes to senile plaques or binds with β-amyloid.[5–10] Cystatin C inhibits cathepsin D. There were 180 case subjects (20 < age 60, 46 ages 60 to 69, and 114 ages 70 to 90) and 120 unaffected spouse control subjects. Methods for genotypic determinations in the sample are standard and previously published.[6,11–13]

The high-risk multilocus genotypes were identified by employing a fuzzy latent classification method called grade-of-membership analysis or, more succinctly, GoM[14,15] employing software developed at the Center for Demographic Studies at Duke University. This approach has not been previously used to define multilocus genotypes that result in AD and likely age at onset.

RESULTS

A total of five groups were identified. They were labeled I to V. Three GoM groups were affected (I, II, III) and two groups were not affected (IV, V). The high-risk multilocus genotypes were: I: CST3:GA and CTSD:CT (onset <70 years); II: APOE44, LDLr8:GG, and LDLr13:TT (onset 60 to 75 years); and III: APOE34 and LDLr13:TC (onset <80 years). IV: The absence of these combinations defined low-risk. V: APOE ε23, ε33 or even ε34 predicted long life without dementia when found with heterozygous APOE promoter and LDL receptor genotypes.

These interpretations of each extreme or pure type are based on the group-defining probabilities for the variables disease status, age at onset or at last observation, sex, and genotypes found at the eight loci (Table 1). Group I, having the youngest age at onset and often female, had 100% probability for carrying CST3:GA and CTSD:CT rather than CST3:GG and CTSD:CC found for the other groups. These variants, specifically, were influential in defining the group as indicated by elevated question relevance factor (QRF) scores of 2.2 and 1.3, respectively. The neutral referent value is one. Nominally high-risk APOE ε24 or ε44 genotypes, and low-risk ε23, were consistent with group I. The common denominator was promoter genotypes −491:TT and −427:CT, where the −427 C allele is thought to be permissive, and receptor genotypes LDLr8:GG and LDLr13:CC.

Groups II and III are more familiar, representing APOE44 and APOE34, respectively. Specific LDL receptor genotypes in combination with APOE44 and APOE34 defined high risk. Both groups carried the −491:TT promoter genotype, also thought to be permissive. Group II had onset between ages 60 to 75 and was often female. APOE44 (93%) or APOE24 (7%) was present. Receptor genotypes LDLr8:GG and LDLr13:TT (QRF = 1.8) were group defining. Group III might be regarded as a single-dose version of group II. Onset tended to be later but before age 80. APOE34 (QRF = 1.3) and APOCI:AB (QRF = 1.4) and LDLr13:TC were present.

An absence of these multilocus genotypes defined low risk (i.e., groups IV and V). Group IV was aged 70 to 79 and unaffected. It differed from group III by carrying APOE33 (QRF = 1.4) and APOCI:BB (QRF = 1.5) and, frequently, LDLr13:CC. Group V was over age 80 and unaffected. In addition to lacking the multilocus risk genotypes, there were heterozygous genotypes for the APOE promoter (−491:AT; QRF = 2.4) and LDL receptor (LDLr8:AG and LDLr13:TC). Hence, long life without dementia was consistent with a diversity of APOE genotypes ε23, ε33, and even ε34 when accompanied with certain promoter and receptor variants.

Few study subjects exactly matched any of these extreme pure type groups (i.e., had 100% membership in a group) (I: 0, II: 4; II: 4; IV: 14; V: 2). Membership score ranges from zero (0%, no resemblance) to one (100%, exact match). Scores were assigned to each subject for each group as part of the maximum likelihood estimation procedure.

Logistic models were constructed using GoM membership scores for groups I, II, and III to predict AD status. High membership in any of the groups had more than 800-fold elevated odds of AD. Next, a logistic model was constructed where summed membership scores (I + II + III)

TABLE 1. ALZHEIMER DISEASE RISK GROUPS I TO V*

Attribute	*I*	*II*	*III*	*IV*	*V*	*H*
AD case	100	100	100	0	0	0.68
Age (years)						
<65	31	12	17	0	0	0.90
65−	69	30	0	0	0	
70−	0	57	45	41	0	
75−	0	0	37	59	0	
80+	0	0	0	0	100	
Female						
Yes	96	80	59	51	56	0.07
No	4	20	41	49	44	
APOE						
ε23	40	0	4	0	48	1.17
ε33	0	0	0	100	24	
ε24	42	7	0	0	0	
ε34	0	0	96	0	28	
ε44	19	93	0	0	0	
APOE-491						
AA	0	100	100	100	0	0.90
AT	0	0	0	0	100	
TT	100	0	0	0	0	
APOE-427						
TT	0	100	99	100	74	0.30
TC	100	0	0	0	25	
CC	0	0	1	0	1	
APOCI						
AA	NA	100	0	0	0	0.91
AB	NA	0	100	0	100	
BB	NA	0	0	100	0	
LDLr8						
GG	100	100	100	99	0	0.40
AG	0	0	0	0	100	
AA	0	0	0	1	0	
LDLr13						
TT	0	100	0	0	0	0.82
TC	0	0	100	53	100	
CC	100	0	0	47	0	
CST3						
GG	0	90	84	100	69	0.52
GA	100	0	0	0	0	
AA	0	10	16	0	31	
CTSD						
CC	0	100	100	100	100	0.41
CT	100	0	0	0	0	
Group sizes	41.7	47.9	86.8	83.1	40.5	

Model-based risk groups I to V were identified by grade-of-membership analysis. Each group is defined by the displayed outcome probabilities for the variables. *H* indicates the information content for the variable (Shannon, Bell Laboratories): APOE genotype was most informative ($H = 1.17$), whereas sex was the least informative ($H = 0.07$); zero denotes no information. Influential responses having QRF score >1.2 are shown in bold. Group size is the sum of membership in the group over all the subjects.

predicted AD status. High aggregate membership had more than 1000-fold elevated odds of AD. More directly, none of the 50 subjects having low (<20%) aggregate membership was affected; all 102 subjects having high (>80%) aggregate membership were affected.

DISCUSSION

Using GoM, no genetic model is specified. Maximum likelihood is used to estimate the model parameters; that is, the probabilities that define each group and the membership scores

of individuals in each group. This is an advantage as there is no *a priori* model to pose when many genetic loci are jointly considered. Clearly, the approach allows the joint consideration of multiple loci to identify interpretable sufficient risk sets providing indications as to which APOE4 carriers are at relatively low risk for AD, those having heterozygous LDL receptors, and which APOE2 carriers are at high risk, those with very permissive promoters and homozygous LDLr13:CC receptors. These findings can be tested in other settings.

Membership score in the sufficient sets was highly predictive of risk for individuals. In the future this property may be exploited to estimate inherited risk for individuals regardless of present age and disease status once large prospective population-based studies covering broad age ranges, and including information on additional candidate gene polymorphisms, have been conducted and closely quantify risk. This approach remains a research tool to identify risk sets until predictive information would be useful in applying preventive interventions, not yet in existence.

Note that the study sample was small (180 AD cases, 120 control subjects). Nonetheless potentially useful information emerged. This strongly suggests that joint consideration of multiple candidate loci using GoM has high statistical power compared to current approaches in which allele frequencies for one locus are compared for case and control subjects. This point of view has a statistical basis: GoM can use all of the the Jth order moments where J is the number of variables, rather than being limited to the first and second-order data moments.[15]

Second, GoM uses an L_1 (Minkowski metric of $p = 1.0$; deviations are not squared) distance or stress measure rather than the L_2 criterion (deviation are squared). Studies in physics, electrical engineering, and communication theory methodologies based on an L_1 metric (like GoM) have been shown to have one sixth the distortion in detecting complex signals (e.g., strange attractors) as methods based on an L_2 metric. This result is based on extensive simulations where complex nonlinear patterns were being detected from time series information.[16] Hence, as demonstrated in this study, results can be obtained from relatively small samples. Alternative latent class approaches based on an L_1 metric not allowing the fuzzy mixed membership of individual subjects in several genetic groups would likely require more subjects and identify a larger number of groups.

Clearly, multiple comparisons that decimate power, the bane of genetic epidemiology, are avoided. The approach has proved useful in translating clinical data into interpretable patterns in many contexts.[17–20] Importantly, sampling variation using GoM would be primarily expressed in the prevalence distributions of the membership scores: The relevant combinations of genotypes are rooted in biology and may prove to be robustly replicated.

CONCLUSION

These findings demonstrate sufficient genetic risk sets for AD that may in the future be useful in predicting risk for individuals.

ACKNOWLEDGMENTS

The authors thank the Texas and Georgia families for their active participation in the DNA Bank. DNA also was obtained from the National Cell Repository for Alzheimer's Disease. This research was supported by MCG startup funds, MCG Dean's summer research fellowship, the summer STAR program, VA VISN 7 service line career development award, and a cooperative agreement grant U24AG21886 from the National Institute of Aging for the National Cell Repository.

REFERENCES

1. Strittmatter WJ, Saunders AM, Schmechel D, et al. 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 1993;90:1977–1981.
2. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. Neurology 1993;43:1467–1472.
3. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of

Alzheimer's disease in late onset families. Science 1993;261:921–923.
4. Ohm TG, Kirca M, Bohl J, et al. Apolipoprotein E polymorphism influences not only cerebral senile plaque load but also Alzheimer-type neurofibrillary tangle formation. Neuroscience 1995;66:583–587.
5. Corder EH, Saunders AM, Risch NJ, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 1994;7:180–184.
6. Poduslo SE, Neal M, Herring K, et al. The apolipoprotein CI A allele as a risk factor for Alzheimer's disease. Neurochem Res 1998:23:361–367.
7. Retz W, Thome J, Durany N, et al. Potential genetic markers of sporadic Alzheimer's dementia. Psychiatr Genet 2001;11:115–122.
8. Crawford FC, Freeman MJ, Schinka JA, et al. A polymorphism in the cystatin C gene is a novel risk factor for late-onset Alzheimer's disease. Neurology 2000;55:763–768.
9. Papassotiropoulos A, Bagli M, Kurz A, et al. A genetic variation of cathepsin D is a major risk factor for Alzheimer's disease. Ann Neurol 2000;47:399–403.
10. Artiga MJ, Bullido MJ, Sastre I, et al. Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. FEBS Lett 1998;421:104–108.
11. Cheng D, Huang R, Lanham IS, et al. Functional interaction between APOE4 and LDL receptor isoforms in Alzheimer's disease. J Med Genet 2005;42:129–131.
12. Cathcart HM, Huang R, Lanham IS. Cystatin C as a risk factor for Alzheimer disease. Neurology 2005;64: 755–757.
13. Parker GR, Cathcart HM, Huang R, et al. Apolipoprotein gene E4 allele promoter polymorphisms as risk factors for Alzheimer's disease. Psychiatr Genet 2005;15:271–275.
14. Woodbury MA, Clive J. Clinical pure types as a fuzzy partition. J Cybernet 1974;4:111–121.
15. Woodbury MA, Clive J, Garson A. Mathematical typology: a grade of membership technique for obtaining disease definition. Computers Biomed Res 1978; 11:277–298.
16. Fraser AM. Information and entropy in strange attractors. IEEE Transact. Informat Theory 1989;35: 245–262.
17. Manton KG, Corder LS, Stallard E. Estimates of change in chronic disability and institutional incidence and prevalence rates in the U.S. elderly population from the 1982, 1984, and 1989 National Long Term Care Surveys. J. Gerontol 1993;48:S153–S166.
18. Corder EH, Woodbury MA, Madsen DK, et al. Density profiles of Alzheimer disease regional brain pathology for the Huddinge brain bank: pattern recognition emulates and expands upon Braak staging. Exp Gerontol 2000;35:851–864.
19. Corder EH, Woodbury MA, Field LL. Grade-of-membership sibpair linkage analysis maps the IDDM11 locus to chromosome 14q24.3-q31: a pattern recognition approach to linkage. Ann Hum Genet 2001;65: 387–394.
20. Hallmayer JF, Kalaydjieva L, Badcock J, et al. Genetic evidence for a distinct subtype of schizophrenia characterized by pervasive cognitive deficit. Am J Hum Genet 2005;77:468–476.

Address reprint requests to:
Shirley E. Poduslo, Ph.D.
Medical College of Georgia
Institute of Molecular Medicine and Genetics
1120 15th Street
Augusta, GA 30912

E-mail: spoduslo@mail.mcg.edu