Multiple Genetic Interactions Predict Risk for Alzheimer's Disease

Elizabeth H. Corder[4], Rong Huang[1], Heather M. Cathcart[1], Irene S. Lanham[1], Ginny R. Parker[1], Danny Cheng[1], Shirley E. Poduslo[1,2,3],*

[1]Institute of Molecular Medicine and Genetics

[2]Department of Neurology

Medical College of Georgia, Augusta, GA 30912

[3]VA Medical Center. Augusta, GA

[4]Center for Demographic Studies, Duke University, Durham, NC 27708

*To whom correspondence should be addressed.

S.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.

Abstract: The multiple genotypes contributing to Alzheimer's disease have been difficult to define. We used a fuzzy latent classification statistical model, i.e. grade-of-membership analysis, with information on age, sex, disease status, and multilocus genotypes (APOE, APOE promoter, APOCI, LDLr, cystatinC, cathepsinD) in 180 patients and 120 controls to identify five risk sets. Onset age varied for three high-risk sets: 1) CST3:GA and CTSD:CT (<70 years) with permissive APOE promoters; 2)APOE44 and LDLr8:GG and LDLr13:TT (<75 years); 3)APOE34 and LDLr13:TC (<80 years). The absence of these combinations defined low-risk. APOE ε23, ε33 or ε34 with receptor and promoter heterozygosity predicted long-life without dementia.

One Sentence Summary:

Three sufficient sets of risk genotypes, i.e. recipes, for Alzheimer's disease are identified as well as ingredients for a long life without dementia.

APOE polymorphism is the established genetic risk factor for Alzheimer disease (AD). It accounts for about a third of cases taken as a single factor. While there are many biologically plausible candidate genes, sufficient sets of variants that result in AD have not been defined.

We employed information on eight candidate loci to define multilocus genotypes that predict AD. These loci were APOE, promoter polymorphisms at positions –491 & -427, adjacent APOCI, polymorphisms located within exons 8 and 13 low-density lipoprotein receptor (LDLr) for APOE, as well as CTS3 and CSTD which colocalize to senile plaques with β-amyloid[6-14]. Methods for genotypic determinations are presented in the Supplementary Information.

The multilocus genotypes were identified by employing a fuzzy latent classification called grade-of-membership analysis or, more succinctly, GoM[1-2]. Five groups labeled I to V were identified (Table 1). Each group carried probabilities for the genotypes, disease status, age, and sex (Table 2). There were 180 AD patients and 120 unaffected control subjects. Each subject was assigned a membership score in each group indicating resemblance to the group. This approach has not been previously used to define multilocus genotypes that result in

AD and likely age at onset.

Three sets of multilocus genotypes predicted AD. Group I (onset < age 70): Heterozygous CST3:GA and CTSD:CT were found rather than CST3:GG and CTSD:CC. These variants had elevated question relevance factor (QRF) scores of 2.2 and 1.3, respectively, compared to the neutral referent value of 1, and, hence, were influential in identifying group I. Permissive promoter variants were found: i.e. the -491:TT genotype and the -427:CT genotype, regardless of APOE genotype either ε23, ε24 or ε44. LDLr8:GG together with LDLr13:CC was found. Group II (onset between ages 60 to 75): APOE44 (93%) or APOE24 (7%) was found together with LDLr8:GG and LDLr13:TT (QRF=1.8). Group III (onset < age 80): APOE34 (QRF=1.3) was found together with APOCI:AB (QRF=1.4) following the pattern of linkage disequilibrium and LDLr13:TC.

The multilocus genotypes that predicted AD were not found for unaffected groups IV and V. Group IV (aged 70 to 79) conspicuously carried APOE33 (QRF=1.4) together with APOCI:BB (QRF=1.5) following the pattern of linkage disequilibrium and, frequently, LDLr13:CC. Group V (aged 80+) carried heterozygous genotypes for the promoter (-491:AT; QRF=2.4) and receptor (LDLr8:AG & LDLr13:TC). Thus long life without dementia was

consistent with a diversity of APOE genotypes ε23, ε33 and even ε34 when accompanied with promoter and receptor heterozygosity.

The three multilocus genotypes that predicted AD imply that 1) permissive APOE promoters together with CST3 and CSTD variants that influence amyloid peptide secretion and degradation is sufficient to predict AD by age 70, 2) APOE44 interacts with LDLr8GG and LDLr13TT to predict AD by age 75, especially for women[15] and 3) APOE34 interacts with LDLr8GG and LDLr13TC, to predict onset by age 80. The absence of these risk factors predicted good cognition at ages 70 to 79. A long life without AD required heterozygosity for promoters and receptors, and was consistent with ε23, ε33 and even ε34.

No genetic model is specified. Maximum likelihood is used to estimate the model parameters, i.e. the probabilities that define each group and the membership scores of individuals in each group. Using GoM there is 6-fold higher signal detection compared to standard latent class approaches, as it operates in L1, i.e. Minkowski, space rather than L2 space minimizing sums of squared deviations. Hence, results can be obtained from relatively small samples. GoM models do not force individuals into groups. Hence, a parsimonious number of groups is defined. Multiple comparisons are avoided. This approach has been used in numerous medical and genetic applications[3-5].

These findings in and of themselves better predict risk for AD for individuals compared to approaches that employ one or few loci. They imply the existence of relevant multilocus genotypes that can be identified using fuzzy latent class analysis in moderately sized samples.

References:

1. Woodbury, M.A. & Manton, K.G. A new procedure for analysis of medical classification. *Methods Inf. Med.* **21**, 210-220 (1982).

2. The GoM software was developed at the Center for Demographic Studies at Duke University.

3. Corder, E.H., Woodbury, M.A., Bogdanovic, N., & Winblad, B. Density profiles of Alzheimer's disease regional brain pathology demonstrate early and late predelection sites and predict clinical status: GoM emulates and expands upon Braak staging. *Exp. Gerontology* 35, 851-864 (2000).

4. 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* 65, 387-94 (2001).

5. Manton,K.G., Woodbury, M.A., & Tolley,H.D. *Statistical applications using fuzzy sets*. John Wiley & Sons, New York,

(1994).

6. Poduslo, S.E., Neal, M., Herring, K., & Shelly, J. The apolipoprotein CI A allele as a risk factor for Alzheimer's disease. *Neurochem. Res.* **23,**361-367 (1998).

7. Retz, W. *et al.* Potential genetic markers of sporadic Alzheimer's dementia. *Psychiatric Genetics* **11,**115-122 (2001).

8. Crawford, F.C. *et al.* A polymorphism in the cystatin C gene is a novel risk factor for late-onset Alzheimer's disease. Neurology **55,** 763-768 (2000).

9. Papassotiropoulos, A. *et al.* A genetic variation of cathepsin D is a risk factor for Alzheimer's disease. *Ann. Neurol.* **47,**399-403 (2000).

10. Artiga, M.J. *et al.* Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. *FEBS Lett.* **421,**104-108 (1998).

11. Dodel, R.C. *et al.* A polymorphism in the cystatin C promoter region is not associated with an increased risk of AD. *Neurology* **58,**664 (2002).

12. Maruyama, H. *et al.* Lack of an association between cystatin C gene polymorphisms in Japanese patients with Alzheimer's disease. *Neurology* **57,**337-339 (2001).

13. McIlroy, S.P., Dynan, K.B., McGleenon, B.M., Lawson, J.T., & Passmore, A.P. Cathepsin D gene exon 2 polymorphism and sporadic Alzheimer's disease. *Neurosci. Lett.* **273**,140-141 (1999).

14. Bertram, L. *et al.* No evidence for genetic association or linkage of the cathepsin D (CTSD) exon 2 polymorphism and Alzheimer disease. *Ann. Neurol.* **49**,114-116 (2001).

15. Corder, E.H. *et al.* The biphasic relationship between regional brain senile plaque and neurofibrillary tangle distributions: Modification by age, sex, and APOE polymorphism. *Annals N.Y. Acad. Sci.* **1019**,1-5 (2004).

Acknowledgments: We thank the Texas and Georgia families for their active participation in the DNA Bank. DNA was also 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 by a cooperative agreement grant U24AG21886 from the National Institute of Aging for the National Cell Repository.

Competing Interests Statement: There are none.

Table 1. The five GoM groups.

Label	Description
I:	CST3:GA, CTSD:CT, permissive APOE promoters (onset <70 years)
II:	APOE44, LDLr8:GG, LDL13r:TT (onset < 75 years)
III:	APOE34, LDLr8:GG, LDL13r:TC (onset < 80 years)
IV:	Absence of these combinations: unaffected at age 70 to 79.
V:	Promoter and receptor heterozygosity: unaffected at ages 80+ for APOE ε23, ε33 or ε34.

Table 2. Frequencies for GoM groups I to V.*

Attribute		**I**	**II**	**III**	**IV**	**V**	**H**
AD case		100	100	100	0	0	0.68
Age (years)	< 65	31	17	12	0	0	0.90
	65-	69	30	0	0	0	
	70-	0	57	45	41	0	
	75-	0	0	36	59	0	
	80+	0	0	0	0	100	
Female		96	80	59	51	56	0.07
APOE	e23	40	0	4	0	48	1.17
	ε33	0	0	0	**100**	24	
	e24	47	7	0	0	0	
	e34	0	0	**96**	0	28	
	e44	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	n/a	100	0	0	0	0.91
	AB	n/a	0	**100**	0	100	
	BB	n/a	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	

* Each GoM group is defined by the displayed probabilities of response for disease status, age category, sex, genotypes for APOE and the other candidate loci. Influential responses in determining the groups are shown in bold, i.e. QRF score > 1.2. H indicates the information content for the variable (Shannon, Bell Laboratories): Each variable contributed to determining the GoM groups; APOE genotype was most informative (H=1.17) while sex was the least informative (H=0.07); zero denotes no information, i.e. the same frequencies were found for each group.