

University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Health and Biomedical Sciences Faculty
Publications and Presentations

College of Health Professions

2017

Bayesian Cox Proportional Hazards Model in Survival Analysis of HACE1 Gene with Age at Onset of Alzheimer's Disease

Ke-Sheng Wang

East Tennessee State University

Ying Liu

Shaoqing Gong

East Tennessee State University

Chun Xu

The University of Texas Rio Grande Valley

Xin Xie

See next page for additional authors

Follow this and additional works at: https://scholarworks.utrgv.edu/hbs_fac



Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

: Ke-Sheng W, Liu Y, Gong S, Xu C, Xie X, et al. (2017) Bayesian Cox Proportional Hazards Model in Survival Analysis of HACE1 Gene with Age at Onset of Alzheimer's Disease. *Int J Clin Biostat Biom* 3:014. doi.org/10.23937/2469-5831/1510014

This Article is brought to you for free and open access by the College of Health Professions at ScholarWorks @ UTRGV. It has been accepted for inclusion in Health and Biomedical Sciences Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.

Authors

Ke-Sheng Wang, Ying Liu, Shaoqing Gong, Chun Xu, Xin Xie, Liang Wang, and Xingguang Luo



RESEARCH ARTICLE

Bayesian Cox Proportional Hazards Model in Survival Analysis of HACE1 Gene with Age at Onset of Alzheimer's Disease

Ke-Sheng Wang^{1*}, Ying Liu¹, Shaoqing Gong², Chun Xu³, Xin Xie⁴, Liang Wang¹ and Xingguang Luo^{5,6}

¹Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, Johnson City, TN, USA

²School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an, China

³Department of Health and Biomedical Sciences, College of Health Affairs, University of Texas Rio Grande Valley, Brownsville, TX, USA

⁴Department of Economics and Finance, College of Business and Technology, East Tennessee State University, Johnson City, TN, USA

⁵Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

⁶Biological Psychiatry Research Center, Beijing Huilongguan Hospital, Beijing, China

*Corresponding author: Ke-Sheng Wang, Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, PO Box 70259, Lamb Hall, Johnson City, TN 37614-1700, USA; Tel: +1-423-439-4481; Fax: +1 423-439-4606; E-mail: wangk@etsu.edu



Abstract

Alzheimer's Disease (AD), the most common form of dementia, is a chronic neurodegenerative disease. The HECT domain and ankyrin repeat containing E3 ubiquitin protein ligase 1 (HACE1) gene is expressed in human brain and may play a role in the pathogenesis of neurodegenerative disorders. Till now, no previous study has reported the association of the HACE1 gene with the risk and Age at Onset (AAO) of AD; while few studies have checked the proportional hazards assumption in the survival analysis of AAO of AD using Cox proportional hazards model. In this study, we examined the associations of 14 Single Nucleotide Polymorphisms (SNPs) in the HACE1 gene with the risk and the AAO of AD using 791 AD patients and 782 controls. Multiple logistic regression model identified one SNP (rs9499937 with $p = 1.8 \times 10^{-3}$) to be associated with the risk of AD. For survival analysis of AAO, both classic Cox regression model and Bayesian survival analysis using the Cox proportional hazards model were applied to examine the association of each SNP with the AAO. The Hazards Ratio (HR) with its 95% Confidence Interval (CI) was estimated. Survival analysis using the classic Cox regression model showed that 4 SNPs were significantly associated with the AAO (top SNP rs9499937 with HR = 1.33, 95% CI = 1.13-1.57, $p = 5.0 \times 10^{-4}$). Bayesian Cox regression model showed similar but a

slightly stronger associations (top SNP rs9499937 with HR = 1.34, 95% CI = 1.11-1.55) compared with the classic Cox regression model. Using an independent family-based sample, one SNP rs9486018 was associated with the risk of AD ($p = 0.0323$) and the T-T-G haplotype from rs9786015, rs9486018 and rs4079063 showed associations with both the risk and AAO of AD ($p = 2.27 \times 10^{-3}$ and 0.0487, respectively). The findings of this study provide first evidence that several genetic variants in the HACE1 gene were associated with the risk and AAO of AD.

Keywords

Alzheimer's disease, Age at onset, HACE1, Single nucleotide polymorphisms, Survival analysis, Cox model, Bayesian analysis

Introduction

Alzheimer's Disease (AD) is a chronic neurodegenerative disease [1]. It is estimated that, in 2010, about 4.7 million people in the United States (US) aged 65 years or older live with AD, and this number is projected to rise to 13.8 million, a nearly three-fold increase, by 2050 [2]. Weuve et al. [3] estimated the number of adults (aged

≥ 65 years) with AD in each US state and the District of Columbia (DC) and found that the number of older adults (aged ≥ 65 years) with AD in the US ranged from 5.1 (Alaska) to 530 million (California) in 2010; while in 2010, older adults with AD dementia comprised a median of 1.6% of a given state's total population; whereas by 2025, throughout the US, this proportion will keep increasing by nearly one-third on average [3]. The prevalence was estimated to be about 3.12% in 2012 for older adults (≥ 60 years) in the US using the Nationwide Inpatient Sample (NIS) 2002-2012 [4]. A meta-analysis estimated the prevalence for the combined population of African-Americans and Caucasians aged 65-90 in 2013 was 5.7% in the US; the prevalence for African-Americans aged 65-90 years was 8.6%, compared to 5.5% for Caucasians [5]. Globally, 26.6 million people (0.40% of the world population) suffered from AD in 2006 and it was predicted to affect 1 in 85 people by 2050 [6]. A recent meta-analysis showed that the prevalence of AD in Europe was 5.05% for adult (aged ≥ 50 years) (3.31% for males and 7.13% females) [7]. Increasing evidence suggests that autophagy may play a central role in AD [8,9]. The genetic heritability of AD ranges from 49% to 79% based on reviews of twin and family studies [10]. In addition to the risk, the Age at Onset (AAO) of AD has a genetic component with heritability about 42% [11,12].

The HECT domain and ankyrin repeat containing E3 ubiquitin protein ligase 1 (HACE1) gene (also known as KIAA1320) is located at 6q16.3 [13-15]. The HACE1 is expressed in brain, heart, lung, kidney, testis, and ovary [13,15]. Several studies have implicated that HACE1 is a candidate chromosome 6q21 tumor suppressor gene involved in multiple cancers [16-18]. Recently, it has been reported that HACE1 gene may play a role in neurodegeneration [19] and autophagy pathway [20]; while HACE1 mutations are involved in an autosomal recessive neurodevelopmental disorder [21], and glutamine addiction [22]. A Genome-Wide Association Study (GWAS) identified five Single-Nucleotide Polymorphisms (SNPs) (rs17065302, rs11759010, rs6927608, rs4946645 and rs4245525) within the HACE1 gene associated with equol-producing phenotype such as blood pressure, which may implicate HACE1 in intestinal immune responses [23]. Another GWAS identified 5 SNPs (rs4336470, rs9404576, rs4079063, rs24996663, and rs2499667) in HACE1 associated with neuroblastoma susceptibility [24]; while in a replication study, rs4336470 showed moderate association ($p < 0.05$) with risk of neuroblastoma [25]. More recently, it has been reported that the five above SNPs in the HACE1 gene may have a weak combined effect ($p = 0.065$) on neuroblastoma risk in Southern Chinese children [26]. Thus we hypothesized that HACE1 genetic variants may be in association with AD development.

To our best knowledge, no study has focused on the association of the HACE1 gene with the risk and AAO of AD. Even though the Cox proportional hazards model

has been used to detect genetic associations with the AAO of AD [27-29], the proportional hazards assumption may be violated and they might not be carefully checked. Bayesian methods have been widely used recently in genetic association studies and provide alternative ways to traditional statistical methods [30-32]. In this study, we explored the association of HACE1 with the AAO of AD by using a Bayesian proportional hazards model in a population-based sample and then a family-based sample for replication.

Subjects and Methods

Study population

791 patients with AD and 782 controls with complete genotype and phenotype information in a Canadian sample were selected from the Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's disease and the longitudinal follow-up of Genotype-Phenotype Associations in Alzheimer's disease and the Neuroimaging component of Genotype-Phenotype Associations in Alzheimer's disease-Study Accession: phs000219.v1.p1. Covariates include sex and age. The details about these subjects were described in previous studies [27,33]. Genotyping was conducted using the Affymetrix technique. The genotypes of 14 SNPs within the HACE1 gene were available in this data.

A family-based study (1266 AD cases and their relatives, 1070 individuals with the AAO values) were available from the National Institute on Aging - Late Onset Alzheimer's Disease (NIA-LOAD) Family Study: Genome-Wide Association Study for Susceptibility Loci - Study Accession: phs000168.v1.p1. Genotyping by the Center for Inherited Disease Research (CIDR) was performed using the Illumina Infinium II assay protocol. The details about the sample of subjects were described elsewhere [34]. There are 28 SNPs within the HACE1 gene in this family-based sample.

Statistical analysis

Descriptive statistics and genotype quality control: Descriptive statistics were used to characterize participants' sex, age and the AAO of AD stratified by AD case and control status [29,35]. Hardy-Weinberg Equilibrium (HWE) was tested for all SNPs using the controls by HAPLOVIEW software [36]. Then, Minor Allele Frequency (MAF) was determined for each SNP. Pair wise Linkage Disequilibrium (LD) statistics (r^2) among SNPs were assessed using the European sample from the HapMap dataset (<http://hapmap.ncbi.nlm.nih.gov/>) and the founders in the family study.

Multiple logistic and linear regression models in PLINK software: Multiple logistic regression analysis of each SNP with the risk of AD as a binary outcome, adjusted for sex and age, was performed using PLINK [37]; while the asymptotic p-values were obtained and the Odds Ratio (OR) and 95% Confident Interval (CI) were

estimated. The parallel procedure was performed for the multiple linear regression analysis of each SNP with the AAO of AD as a continuous outcome. Bonferroni correction ($\alpha = 0.05/14 = 3.57 \times 10^{-3}$) was used to deal with the multiple comparison issue [38].

Bayesian Cox proportional hazards model in PROC PHREG: The Cox proportional hazards model (1) or Cox regression model [39] is widely used in the analysis of time-to-event data [40-42].

$$h(t|x) = h_0(t) \exp(\beta_1 \text{SNP}_k + \beta_2 \text{Sex} + \beta_3 \text{Age}) \quad (1)$$

where $h(t|x)$ is the hazard at time t for a subject (AAO for this study), $h_0(t)$ is the baseline hazard function. The Hazard Ratio (HR) is defined as the ratio of the predicated hazard function under two different values of a predictor variable. The PHREG procedure in SAS fits the Cox model by maximizing the partial likelihood function. Both the graphical and numerical methods [43] were used to check the proportional hazards assumption in the ASSESS option of PROC PHREG. The ASSESS option plots the cumulative score residuals against time for each independent variable; while the RESAMPLE option computes the p-value of a Kolmogorov-type supremum test based on a sample of 1,000 simulated residual patterns. A significant p-value indicates a poor fit.

The Akaike Information Criterion (AIC) (2) was used as a measure of better fit among candidate models [44,45].

$$\text{AIC} = -2 \ln \{p(x|\hat{\theta})\} + 2k \quad (2)$$

where x is the random variable, $\hat{\theta}$ is the maximum likelihood estimate, and k is the number of parameters. A smaller AIC generally indicates a better fit.

Bayesian statistics is an extension of Bayes theorem, which can be written as (3)

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} \quad (3)$$

where θ is the parameter of interest, Y is the observed evidence, $P(Y)$ is the marginal probability, $P(Y|\theta)$ is the likelihood function, $P(\theta)$ is the prior, and $P(\theta|Y)$ is the posterior probability [32]. Bayesian Cox regression can be requested by using the BAYES statement in the PHREG procedure. Summary statistics (Mean, Standard Deviation, the Highest Posterior Density (HPD) and Credible Intervals, and Correlation Matrix) were com-

puted for each of the parameters. Trace plots, posterior density plots, and autocorrelation function plots were also provided [32]. For Bayesian survival analysis of the AAO of AD, the normal prior was chosen for the coefficients and the Deviance Information Criteria (DIC) was available instead of AIC. DIC is intended as a generalization of AIC [46]. A measure of effective numbers of parameters is defined as p_D in (4), where p_D is the posterior mean deviance minus the deviance measured at the posterior mean of the parameters.

$$p_D = E[-2 \ln \{p(x|\theta)\}] + 2 \ln \{p(x|\hat{\theta})\} \quad (4)$$

Then DIC is defined analogously to AIC as in (5). Models with smaller DIC are better supported by the data.

$$\text{DIC} = -2 \ln \{p(x|\hat{\theta})\} + p_D \quad (5)$$

The PHREG procedure in SAS was used to fit the Cox model. Multiple Cox regression model analysis, adjusted for sex and age, was conducted to examine association of each SNP with the AAO of AD. Descriptive statistics and Cox regression analysis were performed with SAS v.9.4 (SAS Institute, Cary, NC, USA).

Family-based study: A family-based association analysis for AD was performed using PBAT version 3.6.7 [47]. For the risk of AD, the Family-Based Association Test using Generalized Estimating Equations (FBAT-GEE) was used [48]; while for the AAO, FBAT-Wilcoxon statistics were employed [49]. The AAO values for healthy siblings were censored and age at entry into the study was used. Haplotype analysis was conducted in 2 or 3-SNP sliding window.

Results

Descriptive statistics and genotype quality control

The demographic characteristics of the subjects are detailed in Table 1. The mean AAO for cases was 76.4 and 72.3 years, respectively, in the NIA and Canadian samples, respectively. All 14 SNPs had MAF > 5% and were in HWE in the controls ($p > 0.05$); while 1 of 28 SNPs with HWE < 0.10^{-4} in the family sample was removed for further analysis.

Multiple linear and logistic regression analyses using PLINK

We found that one SNP was associated with the risk of AD (rs9499937 with $p = 1.8 \times 10^{-3}$) and four SNPs were

Table 1: Descriptive characteristics of cases and controls.

Variable	Family study (NIA sample)		Case-control study (Canadian sample)	
	Patients	Controls	Patients	Controls
Sample size (n)	1266	1279	791	782
Mean of age at onset (years \pm SD)	76.4 (\pm 6.7)	-	72.3 (\pm 8.5)	-
Median of age at onset (years)	77	-	73	-
Range of age at onset (years)	50-98	-	40-97	-
Mean age at entry (years \pm SD)	-	75.5 (\pm 8.1)	77.6 (\pm 8.6)	73.4 (\pm 7.9)
Median age at entry (years)	-	75	79	79
Range of age at entry (years)	-	42 - 103	43 - 100	48 - 94

associated with the AAO of AD (rs7746856, rs6941988, rs9499937 and rs7770002 with $p = 3.09 \times 10^{-2}$, 3.88×10^{-3} , 7.39×10^{-4} , and 3.14×10^{-2} , respectively) (Table 2). Interestingly, the same SNP rs9499937 showed associations with both the risk and AAO of AD and the results remained significant after Bonferroni correction ($p < 3.57 \times 10^{-3}$).

Supremum test for proportional hazards assumption

Figure 1 and Figure 2 display the observed standardized score process with 20 simulated realizations from the null distribution for rs9499937 CC and CT genotypes, respectively. The plots showed that the observed process was atypical compared to the simulated realizations and revealed proportional hazards for the two genotypes compared with TT. The Kolmogorov-type su-

preum test results based on 1,000 simulations for all the covariates were not significant ($p > 0.05$), suggesting the proportional hazards assumption was valid for all the variables in the Canadian sample.

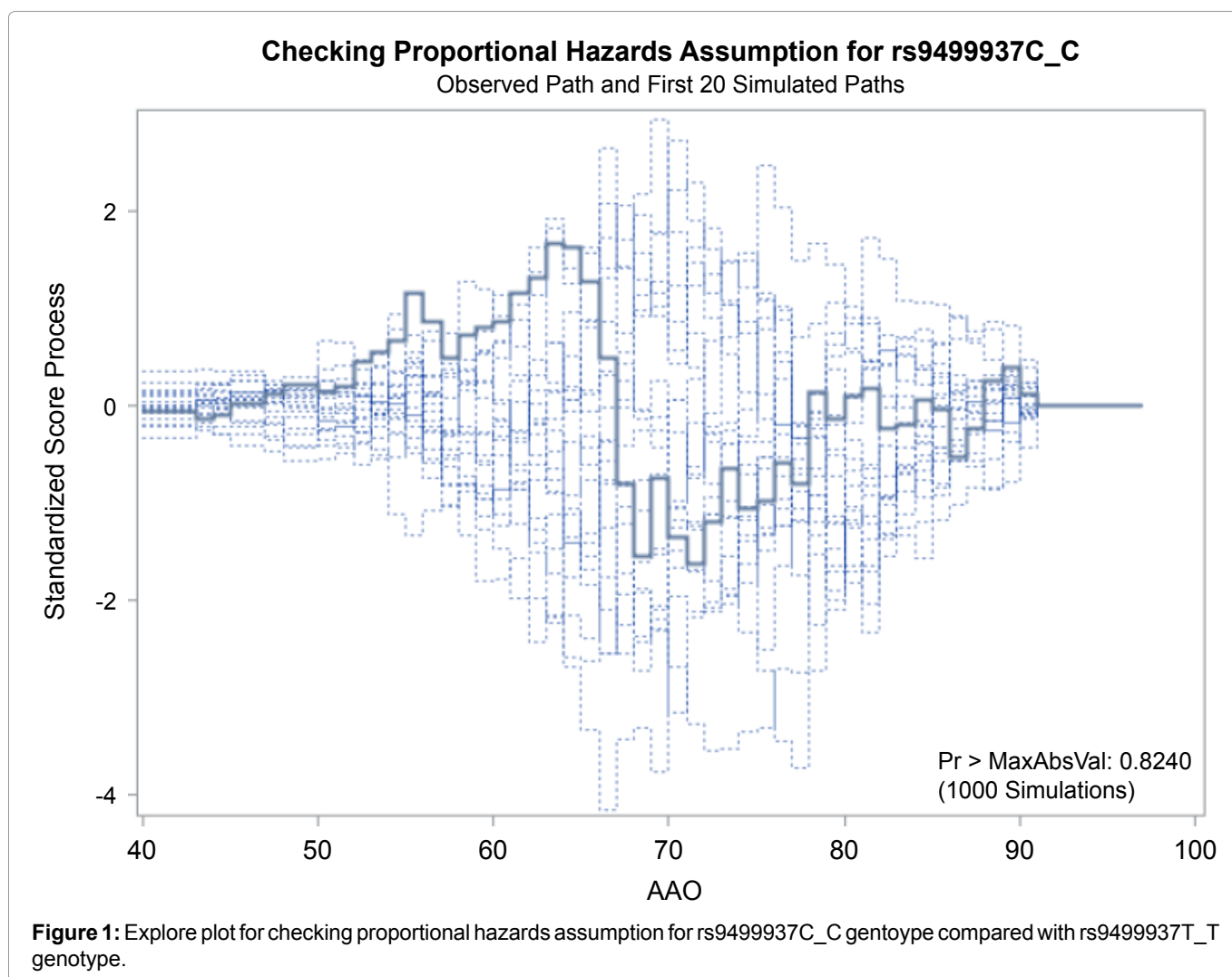
Classic Bayesian Cox proportional hazards model using PROC PHREG

The classic Cox model showed that four SNPs were associated with the AAO of AD (the CT genotype of the top SNP rs9499937 with HR = 1.33, 95% CI = 1.13-1.57, $p = 5.0 \times 10^{-4}$). The HRs based on the Bayesian survival analyses revealed similar but a slightly stronger associations compared with the non-Bayesian analyses results (Table 3). The DIC for the four SNPs were similar to those of AIC. The trace plot, posterior density plot, and autocorrelation function plot based on Bayesian

Table 2: SNPs associated with the risk and/or age at onset of AD ($p < 0.05$).

SNP	Position	Allele ^a	MAF ^b	HWE ^c	OR-AD ^d	p-AD ^e	β -AAO ^f	p-AAO ^g
rs7746856	105253053	A	0.47	0.787	0.92 (0.79 - 1.06)	0.246	-0.367 (-0.70, -0.034)	0.0309
rs6941988	105253349	C	0.47	0.84	0.92 (0.79 - 1.06)	0.239	-0.351 (-0.68, -0.019)	0.00388
rs9499937	105273953	T	0.19	0.649	0.73 (0.60 - 0.89)	1.80×10^{-3}	-0.79 (-1.25, -0.34)	7.39×10^{-4}
rs7770002	105280138	A	0.47	0.96	0.92 (0.80 - 1.07)	0.268	-0.366 (-0.70, -0.033)	0.0314

^aMinor allele; ^bMinor allele frequency; ^cHardy-Weinberg equilibrium test p-value; ^dOdds ratio based on logistic regression; ^ep-value based on logistic regression; ^fRegression coefficient for Age at Onset (AAO) based on multiple linear regression; ^gp-value based on linear regression.



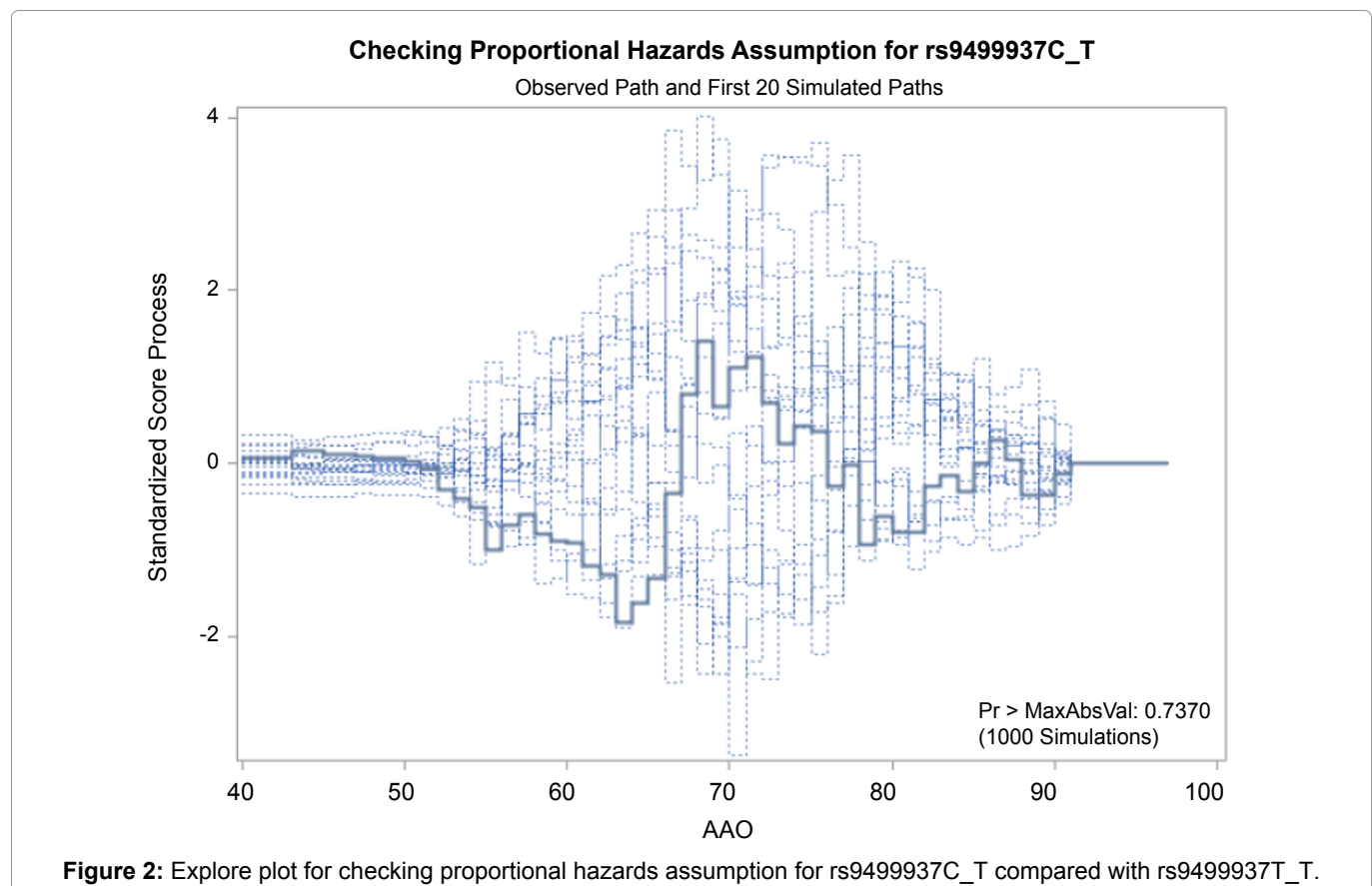


Table 3: SNPs associated with the age at onset of AD using PROC PHREG ($p < 0.05$).

SNP	GT ^a	HR ^b	p ^c	AIC ^d	HR ^e	HPD ^f	DIC ^g
rs7746856	C-C			7622.19			7622.21
	A-A	1.28 (1.04 - 1.57)	0.0208		1.28	1.02 - 1.56	
	A-G	1.02 (0.87 - 1.20)	0.82		1.02	0.86 - 1.19	
rs6941988	T-T			7622.24			7622.32
	C-C	1.27 (1.03 - 1.56)	0.026		1.27	1.02 - 1.54	
	C-T	1.01 (0.85 - 1.18)	0.959		1.01	0.85 - 1.18	
rs9499937	C-C			7616.16			7616.17
	C-T	1.33 (1.13 - 1.57)	5.0×10^{-4}		1.34	1.11 - 1.55	
	T-T	1.35 (0.80 - 2.27)	0.266		1.35	0.68 - 2.06	
rs7770002	A-G			7622.37			7622.34
	A-C	1.27 (1.03 - 1.57)	0.0239		1.28	1.02 - 1.55	
	C-C	1.01 (0.86 - 1.20)	0.864		1.02	0.85 - 1.19	

^aTested genotype comparing with the reference; ^bHazards Ratio (HR) for the tested genotype based on classic Cox regression analysis using PROC PHREG; ^cp-value for the tested genotype based on classic Cox regression analysis; ^dAkaike Information Criterion (AIC) value based on classic Cox regression analysis; ^eHR for the tested genotype based on Bayesian Cox regression analysis; ^fLower and upper 95% Highest Posterior Density (HPD) of HR based on Bayesian Cox regression analysis; ^gDeviance Information Criteria (DIC) value based on Bayesian Cox regression analysis.

analysis (Figure 3) indicated that the Markov chain had stabilized with good mixing for rs9499937. The posterior density plot, which estimates the posterior marginal distributions for the four regression coefficients, showed a smooth and unimodal shape for the posterior marginal distribution (Figure 4).

Family-based association analysis

We observed one SNP associated with the risk of AD (rs9486018 with $p = 0.0323$) by using FBAT-GEE analysis in the family-based study. The T-A haplotype from rs6937432 and rs6940552 revealed mostly significant associations with the risk ($p = 9.7 \times 10^{-4}$). The T-T haplo-

type from rs9786015 and rs9486018 and the T-G haplotype from rs9486018 and rs4079063 showed significant association with the risk ($p = 1.56 \times 10^{-3}$ and 4.46×10^{-3} , respectively) (Table 4). Using the FBAT-Wilcoxon test, the C-C haplotype from rs6937026 and rs6946640 revealed mostly significant associations with the AAO ($p = 0.0223$). The C-A-G and C-C-A haplotype from rs6437026, rs4946640 and rs6910034 showed associations with the AAO ($p = 0.0274$ and 0.0225 , respectively). The T-T-G haplotype from rs9786015, rs9486018 and rs4079063 showed associations with both the risk ($p = 2.27 \times 10^{-3}$) and the AAO ($p = 0.0487$).

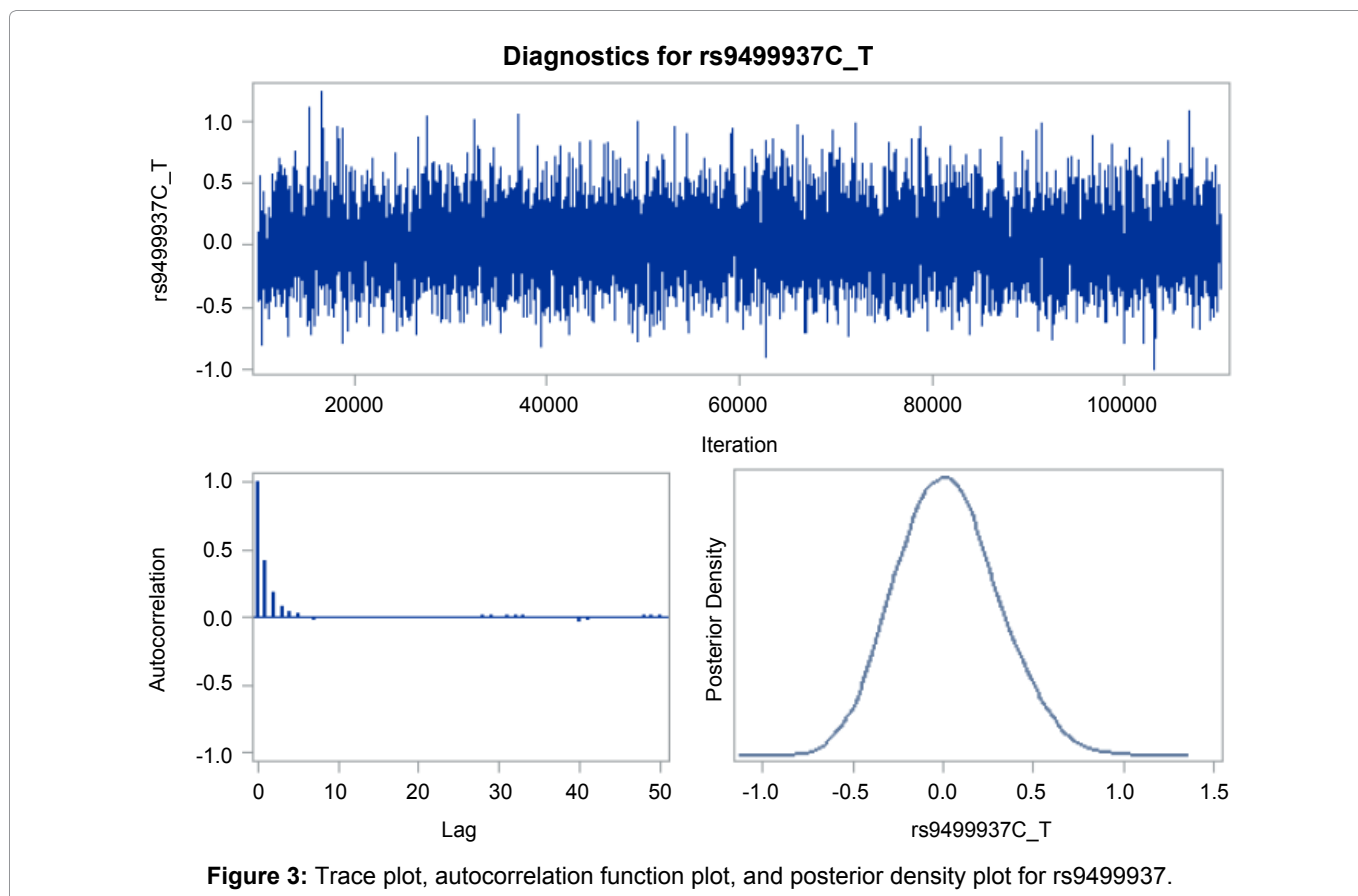


Table 4: Haplotype analysis of the risk and age at onset of AD in the family sample.

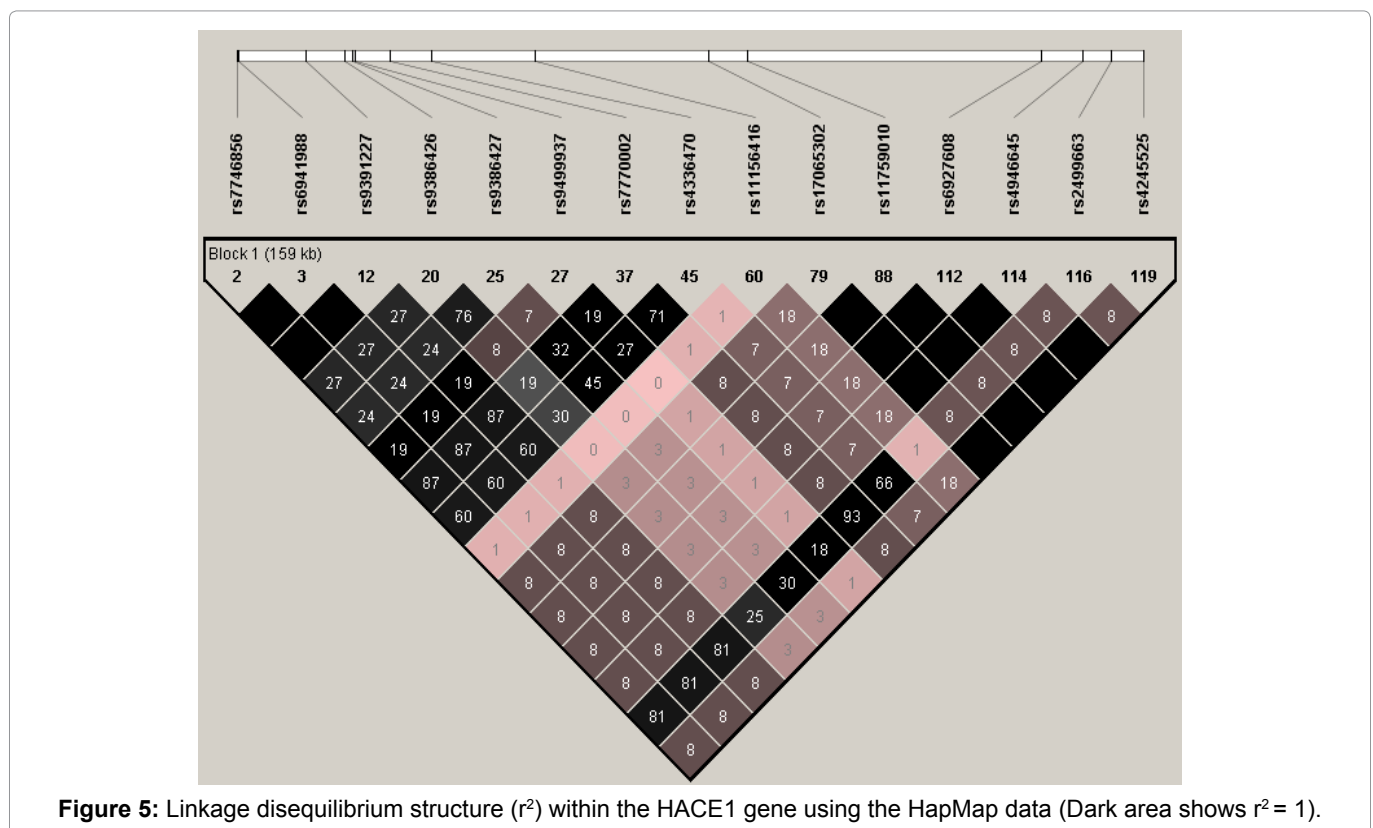
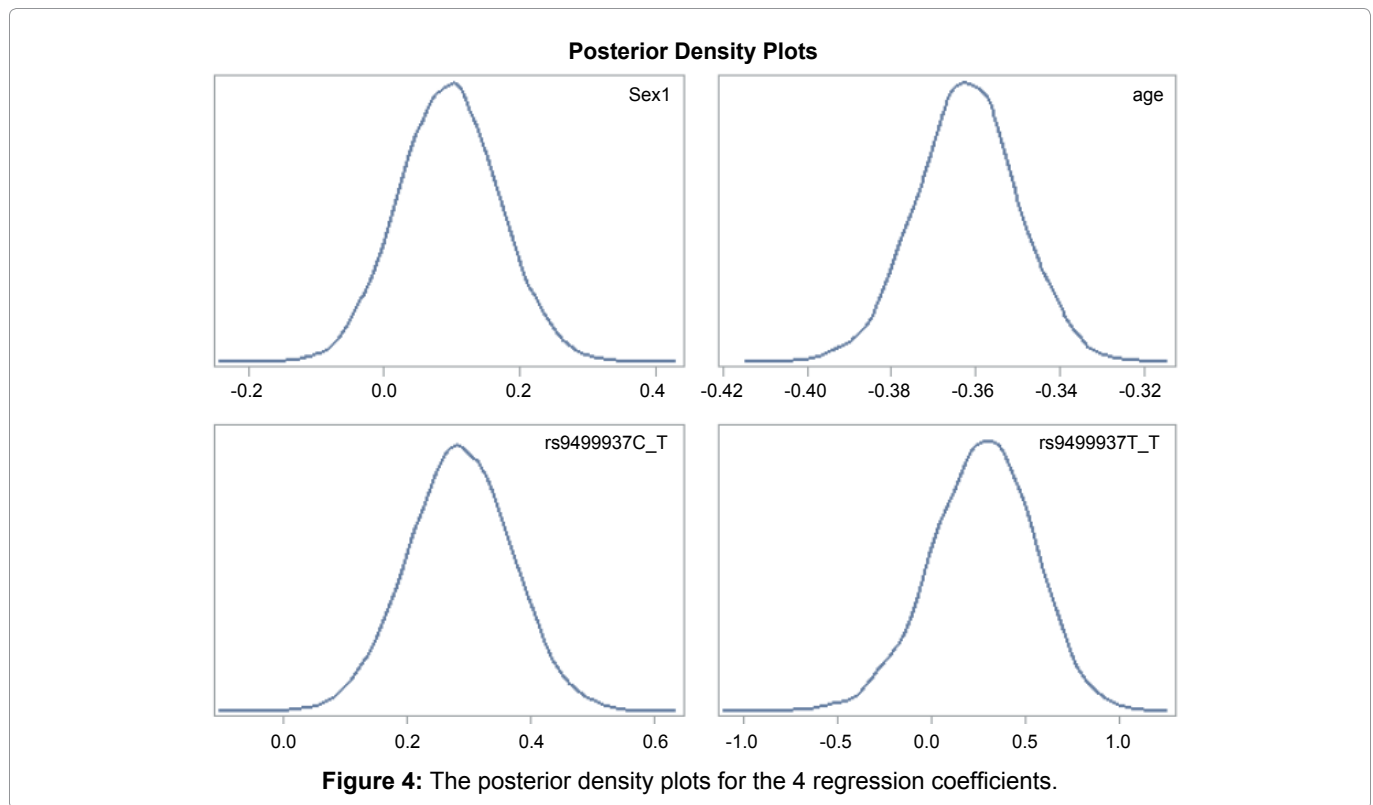
	SNPs		Haplotype ^a	Frequency ^b	Fam# ^c	p-value ^d
Risk of AD						
rs9499927	rs9404573		C-C	0.01	37	0.00144
rs9404573	rs9499934		C-T	0.02	41	0.00355
rs6937432	rs6940552		T-A	0.01	37	0.00097
rs1378720	rs13198196		C-T	0.02	60	0.0147
rs9486015	rs9486018		T-T	0.06	101	0.00156
rs9486018	rs4079063		T-G	0.07	120	0.00446
rs9486015	rs9486018	rs4079063	T-T-G	0.04	54	0.00227
Age at onset of AD						
rs6937026	rs4946640		C-C	0.23	166	0.0223
rs6937026	rs4946640	rs6910034	C-A-G	0.51	156	0.0274
			C-C-A	0.16	100	0.0225
rs9486015	rs9486018	rs4079063	T-T-A	0.01	22	0.0719
			T-T-G	0.04	46	0.0487

^aHaplotype inferred from 2 or 3 SNPs; ^bHaplotype frequency; ^cFAM# refers to the number of informative families using an additive model; ^dp-value for the haplotype based on FBAT-GEE analysis for the risk or based on FBAT-Wilcoxon analysis for the age at onset.

The linkage disequilibrium structure of the HACE1 gene

Using the HapMap data, we identified one haplotype block including the four SNPs associated AD and/or AAO in the case-control study. Figure 5 shows the LD structure based on LD statistics (r^2). Based on the rough rule of thumb, values of $r^2 > 1/3$ might indicate sufficiently strong LD that can be used for a fine mapping [50]. The neuroblastoma associated rs4336470 [24,25] had moderate or strong LD with three AAO associated SNPs (rs7746856, rs6941988 and rs7770002 with $r^2 = 0.6, 0.6$ and 0.71 , respectively) and weak LD with the risk and the AAO - asso-

ciated SNP rs9499937 ($r^2 = 0.27$). Furthermore, there was strong LD observed between rs2499663 and three AAO associated SNPs (rs7746856, rs6941988 and rs7770002 with $r^2 = 0.81, 0.81$ and 0.93 , respectively). Using the founders in the family study, LD structure based on (r^2) was constructed for all 28 SNPs (Figure 6). The two neuroblastoma associated SNPs (rs4336470 and rs9404576) [24,25] had moderate LD with SNPs (such as rs6910034, rs9404573, rs9499934, rs696937432, rs9486015, and rs4079063) which built AD and AAO associated haplotype; while another neuroblastoma associated SNPs rs2499663 [24,25] had moderate or strong LD with SNPs involved in AD and AAO associated haplotype.



Discussion

In the present study we explored the association of 14 HACE1 SNPs with the risk and AAO of AD using a case-control study and identified one SNP associated with AD and four SNPs with the AAO of AD using PLINK software. Interestingly, the same SNP rs9499937 showed associations with both the risk and AAO of AD. Bayesian Cox regressions revealed similar but a slight-

ly stronger associations with the AAO of AD. Using an independent family-based sample, one SNP rs9486018 was associated with the risk of AD while haplotype analyses further revealed the associations with the risk and AAO of AD. The findings of this study provide the first evidence that several genetic variants in the HACE1 gene influenced the risk and the AAO of AD.

A previous study suggested that rs9391227 in the

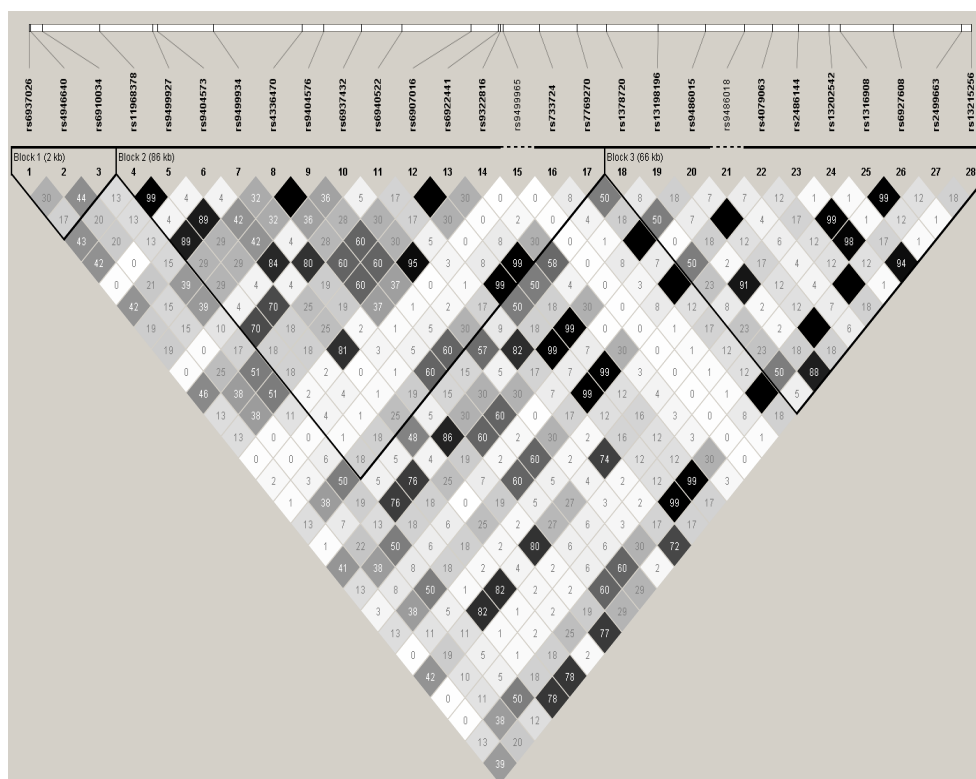


Figure 6: Linkage disequilibrium structure (r^2) within the HACE1 gene of 28 SNPs in the family sample (Dark area shows $r^2 = 1$).

HACE1 gene was associated with celiac disease involved in the immune system and antigen presentation [51]; while another study revealed that five SNPs (rs17065302, rs11759010, rs6927608, rs4946645 and rs4245525) within HACE1 gene associated with equol-producing phenotype such as blood pressure [23] which may implicate HACE1 in immune responses. However, rs9391227 [51] was not available in both the case-control and family-based samples; while equol-producing phenotype associated five SNPs [23] were not associated with the risk or the AAO of AD in the case-control study and rs6927608 was not associated with the risk and AAO of AD in both samples. However, rs9391227 had strong LD with three AAO associated SNPs (rs7746856, rs6941988 and rs7770002 with $r^2 = 1.0$, 1.0 and 0.87 , respectively) in the case-control study (Figure 5); whereas rs6927608 had weak LD with other SNPs (Figure 5 and Figure 6). Previous studies have shown that AD is a chronic neurodegenerative disease while autophagy, immune and inflammatory processes are involved in the pathogenesis of AD [1,8,9,52-56].

Previous epidemiology studies have suggested comorbidity of AD with certain cancers [57-59]; whereas there may be an inverse link between cancer and AD [59-61]. Recently, a meta-analysis of nine studies supports an association between AD and decreased cancer risk [62]. However, the mechanism still remains unclear [60]. Several SNPs (such as rs4336470, rs9404576, rs4079063, rs2499663, and rs2499667) in the HACE1 have been found to be associated with neuroblastoma susceptibility [24-26]. In the present study, the T-G haplotype inferred from the rs9486018 and rs4079063

was found to be associated with the risk of AD; while the T-T-A haplotype inferred from the rs9486015, rs9486018 and rs4079063 was found to be associated with the AAO of AD in the family-based study (Table 4). Furthermore, four neuroblastoma associated SNPs (rs4336470, rs9404576, rs4079063 and rs2499663) [24-26] had moderate to strong LDs with SNPs which built AD and AAO associated haplotypes (Figure 6). In addition, recently, it has been reported that HACE1 may play a role in neurodevelopment and addiction [19,21,22]. Taken together, the above findings may suggest that HACE1 gene may be involved in the pathogenesis of AD, cancers and blood pressure; however, the mechanism warrants further studies.

Several strengths of this study are worthy of noting. The present study provides the first evidence of several genetic variants within the HACE1 gene associated with the risk and the AAO of AD using a case-control sample and a family-based sample for replication. Furthermore, we checked the proportional hazards assumption using both the graphical and numerical methods for the Cox proportional hazards model and found that the proportional hazards assumption was valid for the AAO of AD data in the Canadian sample. In addition, we conducted Bayesian survival analysis of genetic variants with the AAO of AD. Bayesian method may provide an alternative approach to assessing and verifying associations that alleviates the limitations of p-values at the cost of some additional modeling and it has recently made great inroads in genetic association studies [30]. Like other research studies, some limitations also exist in this study. First, due to different genotyping platforms, only

two SNPs (rs6927608 and rs24999663) overlap in two samples; therefore, our replication results in the family-based sample are gene-based rather than SNP-based. Second, only one SNP rs9499937 was associated with the risk and AAO of AD in the case-control study; while one haplotype T-T-G was associated with the risk and AAO of AD in the family-based study; which revealed heterogeneity between risk and AAO of AD. In addition, our current findings might be subject to type I error and need to be replicated in future studies.

Conclusion

To the best of our knowledge, this is the first candidate gene study which investigated the associations of HACE1 SNPs with the risk of and the AAO of AD. The findings may serve as a resource for replication in other populations for future investigations on target genetic variation and AD. Future functional studies of this gene may help better characterize the genetic architecture of the risk of and AAO of AD.

Acknowledgments

We acknowledge the NIH GWAS Data Repository, the Contributing Investigator(s) who contributed the phenotype data and DNA samples from his/her original study and the primary funding organization that supported the contributing study "Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's disease and longitudinal follow-up of Genotype-Phenotype Associations in Alzheimer's disease and Neuroimaging component of Genotype-Phenotype Associations in Alzheimer's disease" and "National Institute on Aging - Late Onset Alzheimer's Disease Family Study: Genome-Wide Association Study for Susceptibility Loci". The genotypic and associated phenotypic data used in the study, "Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GenADA)" were provided by the GlaxoSmithKline, R&D Limited. The datasets used for analyses described in this manuscript were obtained from dbGaP at <https://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number phs000219.v1.p1. Funding support for the "Genetic Consortium for Late Onset Alzheimer's Disease" was provided through the Division of Neuroscience, NIA. The Genetic Consortium for Late Onset Alzheimer's Disease includes a genome-wide association study funded as part of the Division of Neuroscience, NIA. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by Genetic Consortium for Late Onset Alzheimer's Disease. The datasets used for analyses described in this manuscript were obtained from dbGaP at

https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000168.v1.p1.

Role of the Funding Sources

No funding source is given for the present paper.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- Bird TD (2008) Genetic aspects of Alzheimer disease. *Genet Med* 10: 231-239.
- Hebert LE, Weuve J, Scherr PA, Evans DL (2013) Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* 80: 1778-1783.
- Weuve J, Hebert LE, Scherr PA, Evans DA (2015) Prevalence of Alzheimer disease in US states. *Epidemiology* 26: e4-e6.
- Beydoun MA, Beydoun HA, Gamaldo AA, Rostant OS, Dore GA, et al. (2015) Nationwide Inpatient Prevalence, Predictors, and Outcomes of Alzheimer's Disease among Older Adults in the United States, 2002-2012. *J Alzheimers Dis* 48: 361-375.
- Steenland K, Goldstein FC, Levey A, Wharton W (2016) A Meta-Analysis of Alzheimer's Disease Incidence and Prevalence Comparing African-Americans and Caucasians. *J Alzheimers Dis* 50: 71-76.
- Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* 3: 186-191.
- Niu H, Álvarez-Álvarez I, Guillén-Grima F, Aguinaga-On-toso I (2017) Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. *Neurologia* 32: 523-532.
- Chesser AS, Pritchard SM, Johnson GV (2013) Tau clearance mechanisms and their possible role in the pathogenesis of Alzheimer disease. *Front Neurol* 4: 122.
- Shin JY, Park HJ, Kim HN, Oh SH, Bae JS, et al. (2014) Mesenchymal stem cells enhance autophagy and increase β -amyloid clearance in Alzheimer disease models. *Autophagy* 10: 32-44.
- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, et al. (2006) Role of Genes and Environments for Explaining Alzheimer Disease. *Arch Gen Psychiatry* 63: 168-174.
- Daw EW, Payami H, Nemens EJ, Nochlin D, Bird TD, et al. (2000) The number of trait loci in late-onset Alzheimer disease. *Am J Hum Genet* 66: 196-204.
- Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, et al. (2002) Age at onset in two common neurodegenerative diseases is genetically controlled. *Am J Hum Genet* 70: 985-993.
- Nagase T, Kikuno R, Ishikawa K, Hirose M, Ohara O (2000) Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro. *DNA Res* 7: 65-73.
- Fernandez CV, Lestou VS, Wildish J, Lee CL, Sorensen PH (2001) Detection of a novel t(6;15)(q21;q21) in a pediatric Wilms tumor. *Cancer Genet Cytogenet* 129: 165-167.
- Anglesio MS, Evdokimova V, Melnyk N, Zhang L, Fernandez CV, et al. (2004) Differential expression of a novel ankyrin containing E3 ubiquitin-protein ligase, Hace1, in sporadic Wilms' tumor versus normal kidney. *Hum Mol Genet* 13: 2061-2074.
- Zhang L, Anglesio MS, O'Sullivan M, Zhang F, Yang G, et al. (2007) The E3 ligase HACE1 is a critical chromosome 6q21 tumor suppressor involved in multiple cancers. *Nat Med* 13: 1060-1069.

17. Kùçük C, Hu X, Iqbal J, Gaulard P, Klinkebiel D, et al. (2013) HACE1 is a tumor suppressor gene candidate in natural killer cell neoplasms. *Am J Pathol* 182: 49-55.
18. Gao ZF, Wu YN, Bai ZT, Zhang L, Zhou Q, et al. (2016) Tumor-suppressive role of HACE1 in hepatocellular carcinoma and its clinical significance. *Oncol Rep* 36: 3427-3435.
19. Rotblat B, Southwell AL, Ehrnhoefer DE, Skotte NH, Metzler M, et al. (2014) HACE1 reduces oxidative stress and mutant Huntingtin toxicity by promoting the NRF2 response. *Proc Natl Acad Sci U S A* 111: 3032-3037.
20. Zhang L, Chen X, Sharma P, Moon M, Sheftel AD, et al. (2014) HACE1-dependent protein degradation provides cardiac protection in response to haemodynamic stress. *Nat Commun* 5: 3430.
21. Hollstein R, Parry DA, Nalbach L, Logan CV, Strom TM, et al. (2015) HACE1 deficiency causes an autosomal recessive neurodevelopmental syndrome. *J Med Genet* 52: 797-803.
22. Cetinbas N, Daugaard M, Mullen AR, Hajee S, Rotblat B, et al. (2015) DeBerardinis RJ, Sorensen PH. Loss of the tumor suppressor Hace1 leads to ROS-dependent glutamine addiction. *Oncogene* 34: 4005-4010.
23. Hong KW, Ko KP, Ahn Y, Kim CS, Park SJ, et al. (2012) Epidemiological profiles between equol producers and nonproducers: a genomewide association study of the equol-producing phenotype. *Genes Nutr* 7: 567-574.
24. Diskin SJ, Capasso M, Schnepf RW, Cole KA, Attiyeh EF, et al. (2012) Common variation at 6q16 within HACE1 and LIN28B influence susceptibility to neuroblastoma. *Nat Genet* 44: 1126-1130.
25. Capasso M, Diskin SJ, Totaro F, Longo L, De Mariano M, et al. (2013) Replication of GWAS-identified neuroblastoma risk loci strengthens the role of BARD1 and affirms the cumulative effect of genetic variations on disease susceptibility. *Carcinogenesis* 34: 605-611.
26. Zhang Z, Zhang R, Zhu J, Wang F, Yang T, et al. (2017) Common variations within HACE1 gene and neuroblastoma susceptibility in a Southern Chinese population. *Onco Targets Ther* 10: 703-709.
27. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, et al. (2008) Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 65: 45-53.
28. Janicki SC, Park N, Cheng R, Clark LN, Lee JH, et al. (2014) Estrogen receptor α variants affect age at onset of Alzheimer's disease in a multiethnic female cohort. *Dement Geriatr Cogn Disord* 38: 200-213.
29. Wang KS, Liu X, Wang L, Briones D, Xu C (2014) Genetic variants in the SORL1 gene are associated with age at onset of Alzheimer disease: A survival analysis. *International Journal of Medical Genetics* 2014: 689832.
30. Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association studies. *Nat Rev Genet* 10: 681-690.
31. Sullivan SG, Greenland S (2013) Bayesian regression in SAS software. *Int J Epidemiol* 42: 308-317.
32. Stokes M, Chen F, Gunes F (2014) An introduction to Bayesian analysis with SAS/STAT® software. Proceedings of the SAS Global Forum 2014 Conference, SAS Institute Inc, Cary, NC.
33. Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, et al. (2009) Anatomically-distinct genetic associations of APOE epsilon4 allele load with regional cortical atrophy in Alzheimer's disease. *Neuroimage* 44: 724-728.
34. Lee JH, Cheng R, Graff-Radford N, Foroud T, Mayeux R (2008) Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci. *Arch Neurol* 65: 1518-1526.
35. Wang KS, Liu X, Xie C, Liu Y, Xu C (2016) Non-parametric survival analysis of EPG5 gene with age at onset of Alzheimer's disease. *J Mol Neurosci* 60: 436-444.
36. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-265.
37. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.
38. Dunn OJ (1961) Multiple Comparisons Among Means. *Journal of the American Statistical Association* 56: 52-64.
39. Cox DR (1972) Regression models and life-tables. *Journal of the Royal Statistical Society* 34: 187-220.
40. Cantor AB (2007) SAS Survival Analysis Techniques for Medical Research. (3rd edn), Cary, SAS Institute INC.
41. George B, Seals S, Aban I (2014) Survival analysis and regression models. *J Nucl Cardiol* 21: 686-694.
42. Kasza J, Wraith D, Lamb K, Wolfe R (2014) Survival analysis of time-to-event data in respiratory health research studies. *Respirology* 19: 483-492.
43. Lin D, Wei LJ, Ying Z (1993) Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika* 80: 557-572.
44. Akaike H (1979) A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting. *Biometrika* 66: 237-242.
45. Akaike H (1981) Likelihood of a Model and Information Criteria. *Journal of Econometrics* 16: 3-14.
46. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian Measures of Model Complexity and Fit (with Discussion). *J R Statist Soc B* 64: 583-639.
47. Van Steen K, Lange C (2005) PBAT: A comprehensive software package for genome-wide association analysis of complex family-based studies. *Hum Genomics* 2: 67-69.
48. Lange C, Silverman EK, Xu X, Weiss ST, Laird NM (2003) A multivariate family-based association test using generalized estimating equations: FBAT-GEE. *Biostatistics* 4: 195-206.
49. Lange C, Blacker D, Laird NM (2004) Family-based association tests for survival and times-to-onset analysis. *Stat Med* 23: 179-189.
50. Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3: 299-309.
51. Einarsdottir E, Bevova MR, Zhernakova A, Monsuur A, Koskinen LL, et al. (2011) Multiple independent variants in 6q21-22 associated with susceptibility to celiac disease in the Dutch, Finnish and Hungarian populations. *Eur J Hum Genet* 19: 682-686.
52. Holmes C, Cunningham C, Zotova E, Woolford J, Dean C, et al. (2009) Systemic inflammation and disease progression in Alzheimer disease. *Neurology* 73: 768-774.
53. Takeda S, Sato N, Morishita R (2014) Systemic inflammation, blood-brain barrier vulnerability and cognitive/non-cognitive symptoms in Alzheimer disease: relevance to pathogenesis and therapy. *Front Aging Neurosci* 6: 171.

54. Blach-Olszewska Z, Zaczynska E, Gustaw-Rothenberg K, Avila-Rodrigues M, Barreto GE, et al. (2015) The Innate Immunity in Alzheimer Disease- Relevance to Pathogenesis and Therapy. *Curr Pharm Des* 21: 3582-3588.
55. Heppner FL, Ransohoff RM, Becher B (2015) Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci* 16: 358-372.
56. López González I, Garcia-Esparcia P, Llorens F, Ferrer I (2016) Genetic and Transcriptomic Profiles of Inflammation in Neurodegenerative Diseases: Alzheimer, Parkinson, Creutzfeldt-Jakob and Tauopathies. *Int J Mol Sci* 17: 206.
57. Burke WJ, McLaughlin JR, Chung HD, Gillespie KN, Grossberg GT, et al. (1994) Occurrence of cancer in Alzheimer and elderly control patients: an epidemiologic necropsy study. *Alzheimer Dis Assoc Disord* 8: 22-28.
58. Roe CM, Behrens MI, Xiong C, Miller JP, Morris JC (2005) Alzheimer disease and cancer. *Neurology* 64: 895-898.
59. Roe CM, Fitzpatrick AL, Xiong C, Sieh W, Kuller L, et al. (2010) Cancer linked to Alzheimer disease but not vascular dementia. *Neurology* 74: 106-112.
60. Thinnis FP (2012) Why cancer survivors have a lower risk of Alzheimer disease. *Mol Genet Metab* 107: 630-631.
61. Musicco M, Adorni F, Di Santo S, Prinelli F, Pettenati C, et al. (2013) Inverse occurrence of cancer and Alzheimer disease: A population-based incidence study. *Neurology* 81: 322-328.
62. Shi HB, Tang B, Liu YW, Wang XF, Chen GJ (2015) Alzheimer disease and cancer risk: a meta-analysis. *J Cancer Res Clin Oncol* 141: 485-494.