Experimental Gerontology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Experimental Gerontology

journal homepage: www.elsevier.com/locate/expgero

Hidden heterogeneity in Alzheimer's disease: Insights from genetic association studies and other analyses

Anatoliy I. Yashin*, Fang Fang, Mikhail Kovtun, Deqing Wu, Matt Duan, Konstantin Arbeev, Igor Akushevich, Alexander Kulminski, Irina Culminskaya, Ilya Zhbannikov, Arseniy Yashkin, Eric Stallard, Svetlana Ukraintseva*

Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, 2024 W. Main Street, Durham, NC 27705, USA

ARTICLE INFO

Keywords:
Neurodegeneration
Neuroinflammation
Synapse loss
Neuronal death
Pleiotropy
Cancer
Diabetes
Amyloid cascade hypothesis
Response to infection
BBB

ABSTRACT

Despite evident success in clarifying many important features of Alzheimer's disease (AD) the efficient methods of its prevention and treatment are not yet available. The reasons are likely to be the fact that AD is a multifactorial and heterogeneous health disorder with multiple alternative pathways of disease development and progression. The availability of genetic data on individuals participated in longitudinal studies of aging health and longevity, as well as on participants of cross-sectional case-control studies allow for investigating genetic and non-genetic connections with AD and to link the results of these analyses with research findings obtained in clinical, experimental, and molecular biological studies of this health disorder. The objective of this paper is to perform GWAS of AD in several study populations and investigate possible roles of detected genetic factors in developing AD hallmarks and in other health disorders. The data collected in the Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), Health and Retirement Study (HRS) and Late Onset Alzheimer's Disease Family Study (LOADFS) were used in these analyses. The logistic regression and Cox's regression were used as statistical models in GWAS. The results of analyses confirmed strong associations of genetic variants from well-known genes APOE, TOMM40, PVRL2 (NECTIN2), and APOC1 with AD. Possible roles of these genes in pathological mechanisms resulting in development of hallmarks of AD are described. Many genes whose connection with AD was detected in other studies showed nominally significant associations with this health disorder in our study. The evidence on genetic connections between AD and vulnerability to infection, as well as between AD and other health disorders, such as cancer and type 2 diabetes, were investigated. The progress in uncovering hidden heterogeneity in AD would be substantially facilitated if common mechanisms involved in development of AD, its hallmarks, and AD related chronic conditions were investigated in their mutual connection.

1. Introduction

Alzheimer's disease (AD) is a progressive degeneration of the brain, inducing memory decline, learning impairment, language and behavioral disturbances, depressive symptoms, and personality changes, resulting in a marked decline in all mental activities and eventually in death. AD is most prevalent neurological health disorder in developed part of the world today. Recent estimates rank AD as the third cause of death for people older than 75 years. Despite substantial efforts to understand its causes and biological mechanisms the etiology of AD remains largely unknown. There are no medications that can notably influence rate of AD progression when it started. The multifactorial and heterogeneous nature of AD which is manifested in multiple and alternative pathways of its development and progression is likely to be responsible for this situation. This indicates that the improvement in current understanding of AD can be reached by uncovering alternative mechanisms of this health disorder.

The availability of genetic data collected for individuals participated in longitudinal studies of aging, health, and longevity opens a unique opportunity for studying genetic components of hidden heterogeneity using genome wide association studies of AD. Useful insights about the variety of biological mechanisms of AD can also be obtained from findings obtained in clinical, experimental and molecular biological studies of this health disorder. The non-genetic part of heterogeneity in AD can be evaluated from data on aging, health and longevity related traits collected in longitudinal and cross-sectional studies,

* Corresponding authors.
E-mail addresses: aiy@duke.edu (A.I. Yashin), svo@duke.edu (S. Ukraintseva).

http://dx.doi.org/10.1016/j.exger.2017.10.020
Received 31 May 2017; Received in revised form 20 October 2017; Accepted 22 October 2017

Medicare Service Use files and other datasets. An important feature of aging-related health decline observed in longitudinal data is the dependence among chronic conditions detected in epidemiological studies. For example, AD negatively correlates with cancer. Such connections are likely to be caused by common genetic and non-genetic factors as well as by an increase in susceptibility to these disorders with increasing age.

Most of the information and ideas about potential mechanisms connecting the AD with its various biomarkers and risk factors comes from experimental studies of transgenic animal models and from studies of cell cultures, so it is often difficult to directly translate to humans. Still, such studies generated valuable hypotheses that can be further tested using human data, which may substantially improve current understanding of the mechanisms of AD. The availability of rich genetic, behavioral, environmental and other information collected in longitudinal human studies of aging, health and longevity over recent decades opens a unique opportunity for integrating the knowledge from animal and human studies for better understanding of AD. Particularly useful insights about biological pathways involved in AD can be obtained from integration of the results of genome-wide association studies (GWAS) of AD with research findings from clinical, experimental, epidemiological and population studies of AD and the aging-associated physiological and health decline.

In this paper we present the results of GWAS of AD using data from three longitudinal (CHS, FHS, HRS) and one case-control (LOADFS) human studies. Then we discuss how genes detected in our analyses are involved in mechanisms linking this health disorder with its hallmarks. We emphasize possible involvement of these genes in common biological processes related to health disorders other than AD, such as infectious diseases, cancer, and type 2 diabetes (T2D) - to better understand how exactly the detected genes may contribute to the development of AD, and whether the impact of these genes on AD is a part of their broader pleiotropic influence on organism's vulnerability and resistance to stresses.

2. Data and methods

We used data from the Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), Health and Retirement Study (HRS), and Late Onset Alzheimer Disease Family Study (LOADFS) (Lee et al., 2008), to identify genetic variants associated with AD in genome-wide association study (GWAS). Tables 1.1 and 1.2 provide brief description of these datasets.

More information about the FHS, CHS, HRS, and LOADFS data are given in (Dawber, 1980; Mahmood et al., 2014; D'Agostino et al., 1989; Tucker-Seeley et al., 2011; Lee et al., 2008), respectively. See also dbGaP (https://www.ncbi.nlm.nih.gov/gap).

In the logistic regression model, 'cases' corresponded to study participants with AD, and 'controls' corresponded to study subjects without AD. The Cox regression model was applied to the data where the information on the age at disease onset was available. To take family links in LOADFS into account in the logistic regression model, we used the GLIMMIX program in SAS. The year of birth, gender and race (when available) were used as observed covariates in the analyses.

Table 1.2
The sample sizes in three datasets (by race) before and after quality control (QC) procedure.

By race								
Dataset	Before QC				After QC			
	#White	#Black	#Others		#White	#Black	#Others	
CHS	4434	807	29		4209	781	28	
HRS	7968	1299	374		7960	1297	374	
LOADFS	3894	251	271	145*	3894	251	271	145*

* LOADFS has 145 individuals with missing race values.

To control for possible population stratification we calculated 20 principal components and used them as observed covariate (Price et al., 2006). In the HRS data, the genomic control was used to control for possible population stratification, and to avoid the inflation of association test statistics for both the logistic and Cox regressions (Price et al., 2006).

FHS and CHS CARe data used in this analyses were genotyped on the Illumina IBC chip including ~49 K SNPs in ~2000 candidate genes (for major complex diseases). HRS data were genotyped on the Illumina platform with ~2.5mln SNPs, and LOADFS data – on the Illumina platform with ~600 K SNPs. The quality control (QC) has been performed before running GWAS analyses. Individuals with > 5% missing SNPs we excluded. SNPs are kept only if the genotyping rate is higher than 95% and minor allele frequency (MAF) was higher than 1%. In addition, SNPs failed the Hardy-Weinberg test (p-value < 10^{-7}) were also excluded.

3. Results

3.1. Genetics of AD: results from GWAS of FHS, CHS, HRS, and LOADFS data

Fig. 1 shows the QQ plot and Manhattan plot of the results of GWAS of AD using logistic regression applied to CHS, FHS, HRS, and LOADFS data.

One can see from this figure that in all four analyses the genome wide significant SNPs are located on chromosome 19. Fig. 2 shows the QQ plots and Manhattan plots of the results of GWAS of AD using Cox regression for CHS, FHS and HRS datasets. One can see from Figs. 1 and 2 that in all analyses, SNPs located on chromosome 19 show highly significant associations with AD. Note that CHS and FHS have overall smaller numbers of SNPs with highly significant associations. This is because the genotyping platforms in these datasets contain smaller number of SNPs than those for HRS and LOADFS data (see Data and methods). Tables 2.1 and 2.2 summarize results of these analyses for logistic and Cox regression statistical models, respectively. These tables show that the SNPs on chromosome 19 that demonstrated highly significant associations with AD in more than one dataset (these SNPs are

Table 1.1
The sample sizes (by gender) and the number of SNPs in four datasets before and after quality control (QC) procedure.

By gender								
Dataset	Before QC				After QC			
	#Samples	#Males	#Females	#SNPs	#Samples	#Males	#Females	#SNPs
CHS	5270	2252	3018	49,094	5018	2137	2881	33,328
FHS	3788	1651	2317	49,094	3651	1592	2059	35,259
HRS	9641	4120	5521	2,315,518	9631	4118	5513	1,329,158
LOADFS	4561	1666	2895	590,247	4561	1666	2895	551,330