Published in final edited form as:

J Alzheimers Dis. 2019 ; 70(3): 907–915. doi:10.3233/JAD-190168.

Evaluation of the Genetic Variance of Alzheimer's Disease Explained by the Disease-Associated Chromosomal Regions

A Nazarian^{1,*}, AM Kulminski¹

¹Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC, USA

Abstract

Heritability analysis of complex traits/diseases is commonly performed to obtain illustrative information about the potential contribution of the genetic factors to their phenotypic variance. In this study, we investigated the narrow-sense heritability (h^2) of Alzheimer's disease (AD) using genome-wide single-nucleotide polymorphisms (SNPs) data from three independent studies in the linear mixed models framework. Our meta-analyses demonstrated that the estimated h^2 values (and their standard errors) of AD in liability scale were 0.280 (0.091), 0.348 (0.113), and 0.389 (0.126) assuming AD prevalence rates of 10%, 20%, or 30% at ages of 65+, 75+, and 85+ years, respectively. We also found that chromosomal regions containing two or more AD-associated SNPs at $p < 5E-08$ could collectively explain 37% of the additive genetic variance of AD in our samples. AD-associated regions in which at least one SNP had attained $p < 5E-08$ explained 56% of the additive genetic variance of AD. These regions harbored 3% and 11% of SNPs in our analyses. Also, the chromosomal regions containing two or more and one or more AD-associated SNPs at $p < 5E-06$ accounted for 72% and 94% of the additive genetic variance of AD, respectively. These regions harbored 27% and 44% of SNPs in our analyses. Our findings showed that the overall contribution of the additive genetic effects to the AD liability was moderate and age-dependent. Also, they supported the importance of focusing on known AD-associated chromosomal regions to investigate the genetic basis of AD, e.g., through haplotype analysis, analysis of heterogeneity, and functional studies.

Keywords

Aging; complex disorders; dementia; missing heritability; narrow-sense heritability; neurodegenerative disorders; polygenic inheritance

*Correspondence to: Alireza Nazarian, MD, PhD, and Alexander M. Kulminski, PhD, Duke University, Biodemography of Aging Research Unit, Social Science Research Institute, Erwin Mill Building, 2024 W. Main St., Durham, NC 27705, USA. alireza.nazarian@duke.edu. (Alireza Nazarian).

Authors' disclosures available online (<https://www.j-alz.com/manuscript-disclosures/19-0168r1>).

Journal of Alzheimer's Disease. 70(3): 907-915. <https://doi.org/10.3233/JAD-190168>. <https://content.iospress.com/articles/journal-of-alzheimers-disease/jad190168>.

INTRODUCTION

Alzheimer's disease (AD) is the most common neurodegenerative disorder and the most common cause of dementia in people older than 65 years [1]. Late onset AD is believed to occur sporadically with a complex inheritance pattern [2]. As a complex disease of polygenic nature, characteristic for post reproductive period, the phenotypic variance of AD is likely to be attributed to the combined effects of genetic and non-genetic factors, and their interactions. For example, the Apolipoprotein E (*APOE*) $\epsilon 4$ allele, which is known for decades for its remarkably strong association with AD [3,4], is considered as the risk but not causal factor for AD because, for example, even the *APOE* $\epsilon 4$ homozygous individual may remain AD-free until extreme old age [4, 5]. This implies that the effect of the strongest genetic risk factor for AD, the *APOE* $\epsilon 4$ allele, can be modulated by interactions with the environment and/or other genes.

Still, the field intends to characterize the genetic variance of AD by evaluating the narrow-sense heritability (h^2), which is a fraction of phenotypic variance attributed to a 'pure' additive genetic component. The premise of h^2 analyses is to quantify the upper limits of the phenotypic variation of AD that can be collectively caused by the additive genetic effects across genome. Nevertheless, care must be taken when interpreting the results from such analyses due to potential confounding of the additive genetic variance of a trait by the non-additive genetic (i.e., dominance and epistatic), and environmental components [6-9] that, for instance, may arise from the lack of controlled environment condition or appropriate study design [10, 11].

There are various methods for estimating h^2 such as twin-based analyses or linear mixed models (LMMs). In general, LMMs-based methodology is preferred as it benefits from a more relaxed assumption regarding the relatedness of individual in sample, and provides a framework to model non-additive genetic components [12], environmental factors, and gene-environment interactions (e.g., by including a gene-by-environment (GxE) relationship matrix [13] or covariates such as age, sex, or epigenetic modifications into the model [9]) that in turn may attenuate the bias in estimated additive genetic component. The h^2 of AD was estimated to be 58%–74% in twin studies [14, 15], and 24%–53% [16, 17] by LMMs-based methods. In addition, the *APOE* locus coded by rs429358 and rs7412 single-nucleotide polymorphisms (SNPs) was estimated to explain 13.42% of the phenotypic variance of AD. Also, SNPs within ± 50 kb of the *APOE* locus and 27 other well-known AD-associated genes were estimated to collectively account for 31.54% of the phenotypic variance of AD [17]. The differences between h^2 estimates from twin studies and SNPs-based models inform that there might be more genes and genetic variants which can confer risk of AD.

In this study, we investigated the SNPs-based h^2 of AD in three independent cohorts using the genomic best linear unbiased prediction (G-BLUP) methodology [18-20] which integrates the realized genomic relationship matrices, containing the observed proportion of identical-by-descent (IBD) loci that each pair of individual share across their genome [19, 21], into the LMMs framework. In particular, we evaluated the proportions of the additive genetic variance of AD that could be explained by chromosomal regions containing

previously reported AD-associated SNPs at genome-wide (i.e., $p < 5E-08$) and/or suggestive (i.e., $p < 5E-06$) levels of significances.

METHODS

Study participants

Our genetic analyses were performed using the genotype and phenotype information available for 2,544 AD cases and 11,739 controls from three independent studies: 1) the original and offspring cohorts from Framingham Heart Study (FHS) [22, 23], 2) Health and Retirement Study (HRS) [24], and 3) National Institute on Aging's Late-Onset Alzheimer's Disease Family Study (NIA-LOADFS) [25]. All three studies were conducted under the institutional review boards (IRBs) guidelines, and can be accessed through the dbGaP repository (<https://www.ncbi.nlm.nih.gov/gap>) and the University of Michigan restricted access webpage (<http://hrsonline.isr.umich.edu/index.php?p=data>) by the qualified researchers upon approval by the local IRB.

The LOADFS and FHS studies include families and singletons whereas the HRS cohort is a population-based study. Most AD cases were diagnosed based on the clinical findings and routine neurologic examinations (e.g., the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria [26]) without the aid of histopathologic findings (i.e., brain biopsy) or biomarkers. The AD cases and healthy controls were directly identified by the LOADFS and FHS researchers. The International Classification of Disease codes, Ninth revision (ICD-9) were used to determine cases and controls in the HRS cohort based on the Medicare claims available for the study subjects (i.e., ICD-9: 331.0 code). All three studies predominantly enrolled individuals of Caucasian ancestry who were analyzed here as well. We excluded the third generation cohort of FHS in order to make the age of participants comparable among the three datasets. Therefore, based on the inclusion criteria, 3716, 4409, and 6158 subjects from LOADFS, FHS, and HRS, respectively, were included in our analyses. Basic demographic information about the study participants is presented in Table 1. As seen in this Table, the female-to-male ratios in all three studies were larger than one, and the AD cases were on average around 7–17 years older than controls. Also, while LOADFS was primarily initiated to study the genetic basis of AD, the FHS and HRS studies were primarily intended to investigate the cardiovascular diseases and age-related health/well-being/economic issues, respectively. As a result, most AD cases in our study were contributed by LOADFS (i.e., 72.72%, 16.23%, and 11.05% of AD cases, respectively).

Heritability estimates

Our analyses were performed on around 1.5–2 million genotyped and imputed SNPs located on autosomal chromosomes after removing low-quality SNPs (i.e., minor-allele frequencies < 0.01 , missing rates $> 5\%$, p Hardy-Weinberg $< 1E-06$, and squared correlation coefficient (r^2) between the imputed and expected true genotypes < 0.7 for imputed SNPs) and subjects (i.e., calling rates $< 95\%$) [27]. Also, in the case of the LOADFS and FHS datasets that had family-based design, SNPs and subjects/families with Mendel error rates $> 2\%$ were removed. Information about the numbers of SNPs used in our analyses is shown in Table 2.

In each cohort, the G-BLUP methodology [18-20] was used to evaluate the SNPs-based h^2 of AD by fitting a LMM in which the top 5 PCs, birth year, and sex of subjects were considered as fixed-effects covariates, and the additive genetic effects of individuals were considered as a random-effects covariate. The additive genetic effects were modeled by including a normalized marker-based additive relationship matrix [20, 28], generated over the SNPs that passed quality control criteria. The elements of this matrix represented the realized (i.e., observed) proportions of IBD alleles for individuals pairs across the SNPs loci under consideration [19, 21]. The general structure of the fitted LMM was as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of phenotypic values (i.e., case/control status), $\mathbf{X}$ is the incidence matrix for fixed-effects covariates, $\boldsymbol{\beta}$ is the vector of fixed-effects coefficients, $\mathbf{Z}$ is the incidence matrix for additive genetic effects, $\mathbf{a} \sim \text{MVN}(0, \sigma_a^2 \mathbf{A}_G)$ and $\mathbf{e} \sim \text{MVN}(0, \mathbf{I}\sigma_e^2)$ are the vectors of random additive effects and residual errors, respectively. $\mathbf{A}_G$ is the marker-based additive relationship matrix and σ_a^2 is the additive genetic variance. $\mathbf{I}$ is an identity matrix and σ_e^2 is the error variance.

To address the confounding effects of shared environment in the case of the LOADFS and FHS datasets that had family-based designs, another relationship matrix was included in the LMM to capture the GxE interactions for related subjects. The elements of a GxE matrix were the same as those in the additive relationship matrix for family members and zero otherwise [13].

The restricted maximum likelihood (REML) method [29] was applied to estimate the variance-covariance components of the fitted LMMs which were then used for estimating the additive genetic effects for individuals (i.e., best linear unbiased predictions or BLUPs) and the h^2 values (i.e., the portion of phenotypic variance explained by the additive genetic variance, $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$) [9]. To address the loss of power due to ascertainment issue in case-control studies [30], the estimated SNPs-based h^2 values were adjusted and transformed to the underlying liability scale by assuming population prevalence rates of 10%, 20%, or 30% for AD. Three prevalence rates were assumed for these adjustments because the prevalence of AD increases with age from 11% at age 65 to 32% after 85 years [1]. The LMMs of interest were fitted using GCTA package [31].

The estimates of h^2 of AD from the three cohorts under consideration were then combined by an inverse-variance meta-analysis [32] to obtain an h^2 meta-estimate (i.e., h_{meta}^2):

$$h_{meta}^2 = \sum_i \frac{h_i^2}{\text{var}_i} \left(\sum_i \frac{1}{\text{var}_i} \right)^{-1}$$

where h_i^2 and var_i are the estimates of h^2 and its variance in each cohort.

Once the additive genetic variance of AD was estimated using all SNPs located on the autosomal chromosomes, we further investigated the fraction of the genetic variance that could be explained by the SNPs located within ± 1 Mb of the SNPs that have been previously associated with AD at the genome-wide (i.e., $p < 5E-08$) or suggestive (i.e., $p < 5E-06$) significance levels. The list of previously AD-associated SNPs was obtained from the NHGRI-EBI GWAS catalog (release December 2018) [33]. This resulted in 1,032 AD-associated SNPs at $p < 5E-06$. Of these, 253 SNPs were associated with AD at $p < 5E-08$. We considered the ± 1 Mb up/downstream regions of any of these SNPs as an AD-associated chromosomal region, and investigated the extent to which such regions may contribute to the additive genetic variance of AD. We contrasted four alternative scenarios in which chromosomal regions that contained: 1) at least two AD-associated SNPs at $p < 5E-08$ (i.e., R8T regions), 2) at least one AD-associated SNPs at $p < 5E-08$ (i.e., R8 regions), 3) at least two AD-associated SNPs at $p < 5E-06$ (i.e., R6T regions), and 4) at least one AD-associated SNPs at $p < 5E-06$ (i.e., R6 regions) were analyzed. For these analyses, two additive relationship matrices, one generated using SNPs in the chromosomal regions under consideration and the other using the remaining SNPs, were included in the models. The fixed-effects covariates and shared environment effects were also modeled as explained above.

RESULTS

The SNPs-based h^2 values and their standard errors (SE) estimated on the observed scale of AD in the LOADFS, FHS, and HRS cohorts were 0.270 (0.099), 0.081 (0.068), and 0.054 (0.062), respectively, when all SNPs on autosomal chromosomes were included in the genetic analyses. The estimated h^2_{meta} (SE) was 0.103 (0.042) when results from the three datasets were combined by an inverse-variance meta-analysis. Also, the proportions of phenotypic variance explained by the GxE relationship matrices (called the shared environment variance) in the LOADFS and FHS datasets were 0.065 (0.106) and 0.097 (0.080), respectively; and was 0.085 (0.064) when the LOADFS and FHS results were combined by a meta-analysis.

Table 3 summarizes the results of h^2 analyses once the estimated variance components were transformed to the liability scale assuming population prevalence rates of AD were 10%, 20%, or 30% at different ages (i.e., 65+, 75+, and 85+ years, respectively [1]). The adjusted h^2_{meta} values (and their SEs) were 0.280 (0.091), 0.348 (0.113), and 0.389 (0.126) at these three prevalence rates, respectively. Also, the respective adjusted ratios of the shared environment variances to phenotypic variance of AD resulting from the meta-analysis of LOADFS and FHS were 0.107 (0.101), 0.133 (0.126), and 0.149 (0.141).

The genetic analyses were then performed to examine the fraction of the genetic variance of AD that could be attributed to SNPs within the ± 1 Mb up/downstream regions of previously discovered AD-associated SNPs. As seen in Table 3, the chromosomal regions that contained two or more AD-associated SNPs at the genome-wide level of significance (i.e., $p < 5E-08$) accounted for 36.87% of the additive genetic variance of AD (i.e., R8T regions). The regions with one or more AD-associated SNPs at $p < 5E-08$ collectively explained

56.22% of the additive genetic variance in AD (i.e., R8 regions). We also found that the regions that contained at least two AD-associated SNPs at suggestive level of significance (i.e., $p < 5E-06$) accounted for 72.43% of the additive genetic variance of AD (i.e., R6T regions). Finally, the chromosomal regions with at least one AD-associated SNPs at $p < 5E-06$ explained 93.96% of the additive genetic variance of AD (i.e., R6 regions). These four alternative chromosomal regions harbored around 3%, 11%, 27%, and 44% of SNPs in our analyses, respectively (Table 2).

DISCUSSION

In this study, we investigated the fraction of phenotypic variance of AD which might be explained by its additive genetic variance. The whole-genome SNPs-based h^2 values of AD were estimated in three independent datasets which were then combined by an inverse-variance meta-analysis. We found that the meta-analysis h^2 estimates in liability scale were 28%, 34.8%, and 38.9% assuming population prevalence of AD to be 10%, 20%, or 30%, approximately corresponding to the AD prevalence rates at ages 65+, 75+, and 85+ years [1]. The increase in prevalence rates of AD with age evinced the age-dependent liability thresholds [34] at which AD develops. In theory, the age-dependent liability can be mediated by alterations in genetic, environmental, and/or GXE effects over the life course. The increase in the estimated h^2 of AD with the disease prevalence suggested that the overall additive genetic contribution to the AD liability can be different across lifespan. This is, in fact, in agreement with previous reports suggesting the effects of individual genetic factors associated with complex traits were age-dependent, i.e., their effects may appear at certain ages [35, 36] or even be opposite in different age periods [37].

The estimated h^2 values in our analyses (i.e., $h^2 = 28\%–39\%$) were different from predicted values in previous twin studies (i.e., $h^2 = 58\%–74\%$) [14, 15], explaining 38%–67% of the predicted values. The difference in the proportions of predicted and explained h^2 is known as the missing heritability problem. Literature discusses several potential causes of the missing heritability including the inflation of the estimates of the additive genetic variance by other factors such as non-additive genetic effects, epigenetics modifications, and/or environmental factors [6-9] in the twin studies. The three datasets analyzed here have different designs from those used by twin studies; i.e., the HRS cohort is a population-based study gathering data mostly from independent subjects, and LOADFS and FHS are family-based studies providing data for mixtures of singletons and two/three generations of mostly small-size families. Therefore, different designs may partially account for the discrepancy in the estimated h^2 values between our and previous studies. In fact, the degree of relatedness of individuals in data may have direct impacts on the estimates of genetic parameters. For instance, having data from twins, large number of full-sibs, or multi-generational families may result in higher estimates of h^2 by LMMs-based methods because such datasets provide more inter-connections in the elements of the genomic relationship matrix. Therefore, the additive effect for each individual would be determined based on several highly correlated response values, and the variance components would be estimated more precisely. In such cases information from both family structure and linkage disequilibrium (LD) among markers is exploited by the LMMs framework for estimating the genetic parameters of interest [38]. On the other hand, in cohorts with independent individuals or more distant

relatives the parameter are mainly estimated using LD structure of data as the kinship matrix provides less information regarding additive genetic relationship among individuals [38, 39]. However, it should be noted that, as the degree of relatedness increases among individuals, the requirement of controlling the environmental confounders for obtaining accurate estimate of h^2 becomes more important due to the higher possibility of confounding the estimates of additive genetic component with shared environmental effects.

Also, the missing heritability might partially be accounted for by the density of markers used to estimate the SNP-based h^2 of AD. It has been suggested that many common variants with infinitesimal effects and several rare variants of moderate to large effect sizes may causally contribute to the genetic architecture of complex diseases such as AD [40, 41]. Therefore, it is expected that using denser genotype data may result in capturing a larger fraction of the genetic variance due to LD among the discovered/undiscovered AD causal variants and SNPs included in heritability analysis. For instance, our results were consistent with those from two previous studies that estimated the SNP-based h^2 of AD to be 24% [16] and 33.12% [42] using around half-million and two million SNPs, respectively. However, Ridge et al. (2016) demonstrated that 53.24% of the phenotypic variance of AD can be explained by genotype information from more than 8.7 million SNPs [17] which was more than 4 times the SNPs were used in our analyses.

Most AD cases in our study were contributed by LOADFS in which the case-to-control ratio was nearly one. The AD cases constituted 9.4% and 4.6% of the analyzed subjects in the FHS and HRS studies, respectively. Therefore, the genetic analyses of these two datasets may suffer from insufficient statistical power, which was reflected in the larger standard errors of the estimated h^2 values compared to those obtained from LOADFS. This in turn may slightly bias the meta-analyses results and partially explain the smaller h^2 estimates compared to twin studies.

It has been previously reported that the SNPs mapped to 28 well known AD-associated genes, several of which replicated in independent studies, could account for 59.25% of the genetic variance of AD [17]. However, due to the heterogeneity underlying the genetic architecture of AD it is important to extend such analyses to other chromosomal regions whose association signals were not universally replicated or were detected only at the suggestive level of significance (i.e., $p < 5E-06$). Therefore, we investigated the fraction of the additive genetic variance of AD in our samples that can be attributed to these regions. Our analyses demonstrated that SNPs within the regions with at least two (i.e., R8T regions) and at least one (i.e., R8 regions) AD-associated SNPs at $p < 5E-08$, reported in prior studies, could collectively explain 36.87% and 56.22% of the additive genetic variance of AD, respectively. The R8T and R8 regions harbored 3% and 11% of the SNPs in our analyses. These findings corroborated the results from the aforementioned Ridge et al. study [17]. Interestingly, when AD-associated regions with SNPs at the suggestive level of significance were analyzed, we found that only small fractions of the additive genetic variance of AD remained unexplained as SNPs within the R6T and R6 regions (i.e., 27% and 44% of SNPs in our analyses) accounted for 72.43% and 93.96% of the genetic variance of AD. These findings suggested that the AD-associated loci that did not pass the genome-wide significance threshold of $5E-08$ in previous studies may not be necessarily false-

positive findings, instead, more rigorous studies with larger sample sizes or less heterogeneous samples may help discovering stronger association signals for them. Also, our findings suggested that some additional variants could potentially exist in the up/downstream regions of the already discovered AD-associated markers at $p < 5E-06$ to contribute to the genetic basis of AD. For instance, these can be some SNPs with small effect sizes which require very large samples to be detected, or complex haplotypes affecting AD susceptibility [40, 41, 43].

In summary, we found that the common SNPs could meaningfully contribute to the genetic architecture of AD, explaining up to 39% of its phenotypic variance in liability scale in our samples and between 38%–67% of its predicted h^2 in twin studies. As with any random variable, the h^2 estimates may demonstrate variations among samples because AD is a genetically heterogeneous complex traits and, therefore, its genetic architecture is not universal. Therefore, between-study differences in h^2 estimates of AD may, in part, reflect specifics of the investigated populations. In addition, the differences in the study designs (e.g., sample sizes, marker density, degree of relatedness of subjects, etc.) can contribute to the differences in h^2 estimates. Our analyses demonstrated that the additive genetic contributions to the AD liability did not remain constant across the lifespan; instead, it increased with the increase in the AD prevalence at older ages. Of note, we found that ± 1 Mb flanking regions of AD-associated SNPs could account for major fractions of the additive genetic variance of AD in our samples (i.e., up to 56% for AD-associated regions at $p < 5E-08$ and up to 94% for AD-associated regions at $p < 5E-06$). The fractions of the additive genetic variance of AD explained by AD-associated regions at genome-wide and suggestive significance levels were different by 38%. This difference featured the importance of additive contributions to the genetic basis of AD by discovered/undiscovered variants in the regions that had not attained $p < 5E-08$ in conducted GWAS. These findings may have implications for the future studies of AD as they supported the importance of focusing on known AD-associated chromosomal regions through more rigorous methods such as, haplotype analysis, analysis of heterogeneity, deep sequencing, and functional studies, in order to investigate the genetic architecture of AD.

Despite rigor, we acknowledge some limitations of this study (and similar studies). In this study, we focused on the analysis of additive genetic variance of AD, disregarding the dominance and epistatic variance components. This was not meant to undermine the potential non-additive contributions to the genetic variance of AD. However, it should be noted that the analysis of non-additive genetic variance components of a trait using LMMs framework requires gathering data with specific design (e.g., a large number of large full-sibs families) which may not be feasible in human studies [12]. Also, the cases and controls in the selected studies (i.e., LOADFS, FHS, and HRS) were mostly identified based on clinical criteria without the aid of histopathologic findings (i.e., brain biopsy). It has been suggested that histopathologic findings from brain biopsies may increase the accuracy of AD diagnosis. Therefore, future analysis of large samples of histopathologically diagnosed AD cases and healthy controls may help to obtain more accurate estimates of AD heritability [44].

ACKNOWLEDGMENTS

This research was supported by Grants from the National Institute on Aging (P01AG043352 and R01AG047310). The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors declare no competing interests.

This manuscript was prepared using limited access datasets that are available through dbGaP repository (<https://www.ncbi.nlm.nih.gov/gap>) for qualified researchers (accession numbers: phs000168.v2.p2 (LOADFS), phs000007.v28.p10 (FHS), and phs000428.v2.p2 (HRS)) and through the University of Michigan. Phenotypic HRS data are available publicly and through restricted access from <http://hrsonline.isr.umich.edu/index.php?p=data>. The authors thank Dr. Arseniy P. Yashkin for help preparing the HRS phenotypes.

Funding support for the Late Onset Alzheimer's Disease Family Study (LOADFS) was provided through the Division of Neuroscience, NIA. The LOADFS 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. This manuscript was not prepared in collaboration with LOADFS investigators and does not necessarily reflect the opinions or views of LOADFS.

The Framingham Heart Study (FHS) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195 and HHSN268201500001I). This manuscript was not prepared in collaboration with investigators of the FHS and does not necessarily reflect the opinions or views of the FHS, Boston University, or NHLBI. Funding for SHARe Affymetrix genotyping was provided by NHLBI Contract N02-HL-64278. SHARe Illumina genotyping was provided under an agreement between Illumina and Boston University. Funding for CARE genotyping was provided by NHLBI Contract N01-HC-65226. Funding support for the Framingham Dementia dataset was provided by NIH/NIA grant R01 AG08122. Funding support for the Framingham Inflammatory Markers was provided by NIH grants R01 HL064753, R01 HL076784 and R01 AG028321. Funding support for the Framingham Adiponectin dataset was provided by NIH/NHLBI grant R01-DK-080739. Funding support for the Framingham Interleukin-6 dataset was provided by NIH grants R01 HL064753, R01 HL076784 and R01 AG028321.

The Health and Retirement Study (HRS) genetic data is sponsored by the Genetics Resource with HRS April 21, 2010, version G Page 5 of 7 National Institute on Aging (grant numbers U01AG009740, RC2AG036495, and RC4AG039029) and was conducted by the University of Michigan. This manuscript was not prepared in collaboration with HRS investigators and does not necessarily reflect the opinions or views of HRS.

REFERENCES

- [1]. Alzheimer's Association (2016) 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* 12, 459–509. [PubMed: 27570871]
- [2]. Raghavan N, Tosto G (2017) Genetics of Alzheimer's disease: the importance of polygenic and epistatic components. *Curr Neurol Neurosci Rep* 17, 78. [PubMed: 28825204]
- [3]. Pericak-Vance MA, Bebout JL, Gaskell PC, Yamaoka LH, Hung WY, Alberts MJ, Walker AP, Bartlett RJ, Haynes CA, Welsh KA (1991) Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* 48, 1034–1050. [PubMed: 2035524]
- [4]. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921–923. [PubMed: 8346443]
- [5]. Freudenberg-Hua Y, Freudenberg J, Vacic V, Abhyankar A, Emde A-K, Ben-Avraham D, Barzilai N, Oswald D, Christen E, Koppel J, Greenwald B, Darnell RB, Germer S, Atzmon G, Davies P (2014) Disease variants in genomes of 44 centenarians. *Mol Genet Genomic Med* 2, 438–450. [PubMed: 25333069]
- [6]. Stranger BE, Stahl EA, Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187, 367–383. [PubMed: 21115973]
- [7]. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 109, 1193–1198. [PubMed: 22223662]
- [8]. Vinkhuyzen AAE, Pedersen NL, Yang J, Lee SH, Magnusson PKE, Iacono WG, McGue M, Madden PAF, Heath AC, Luciano M, Payton A, Horan M, Ollier W, Pendleton N, Deary IJ,

Montgomery GW, Martin NG, Visscher PM, Wray NR (2012) Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Transl Psychiatry* 2, e102. [PubMed: 22832902]

- [9]. Tenesa A, Haley CS (2013) The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet* 14, 139. [PubMed: 23329114]
- [10]. Lewontin RC (1974) Annotation: the analysis of variance and the analysis of causes. *Am J Hum Genet* 26, 400–411. [PubMed: 4827368]
- [11]. Rose SPR (2006) Commentary: heritability estimates—long past their sell-by date. *Int J Epidemiol* 35, 525–527. [PubMed: 16645027]
- [12]. Nazarian A, Gezan SA (2016) Integrating nonadditive genomic relationship matrices into the study of genetic architecture of complex traits. *J Hered* 107, 153–162. [PubMed: 26712858]
- [13]. Sul JH, Bilow M, Yang W-Y, Kostem E, Furlotte N, He D, Eskin E (2016) Accounting for population structure in gene-by-environment interactions in genome-wide association studies using mixed models. *PLOS Genet* 12, e1005849. [PubMed: 26943367]
- [14]. Gatz M, Pedersen NL, Berg S, Johansson B, Johansson K, Mortimer JA, Posner SF, Viitanen M, Winblad B, Ahlbom A (1997) Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci* 52, M117–125. [PubMed: 9060980]
- [15]. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL (2006) Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 63, 168–174. [PubMed: 16461860]
- [16]. Lee SH, Harold D, Nyholt DR, ANZGene Consortium, International Endogene Consortium, Genetic and Environmental Risk for Alzheimer's disease Consortium, Goddard ME, Zondervan KT, Williams J, Montgomery GW, Wray NR, Visscher PM (2013) Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet* 22, 832–841. [PubMed: 23193196]
- [17]. Ridge PG, Hoyt KB, Boehme K, Mukherjee S, Crane PK, Haines JL, Mayeux R, Farrer LA, Pericak-Vance MA, Schellenberg GD, Kauwe JSK, Alzheimer's Disease Genetics Consortium (ADGC) (2016) Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol Aging* 41, 200.e13–200.e20.
- [18]. VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91, 4414–4423. [PubMed: 18946147]
- [19]. Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res* 91, 47–60.
- [20]. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42, 565–569. [PubMed: 20562875]
- [21]. Guo SW (1996) Variation in genetic identity among relatives. *Hum Hered* 46, 61–70. [PubMed: 8666414]
- [22]. Dawber TR, Meadors GF, Moore FE (1951) Epidemiological approaches to heart disease: the Framingham study. *Am J Public Health Nations Health* 41, 279–286. [PubMed: 14819398]
- [23]. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP (1975) The Framingham offspring study: design and preliminary data. *Prev Med* 4, 518–525. [PubMed: 1208363]
- [24]. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR (2014) Cohort profile: the health and retirement study (HRS). *Int J Epidemiol* 43, 576–585. [PubMed: 24671021]
- [25]. 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. [PubMed: 19001172]
- [26]. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (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. [PubMed: 6610841]
- [27]. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559–575. [PubMed: 17701901]

- [28]. Nazarian A, Gezan SA (2016) GenoMatrix: a software package for pedigree-based and genomic prediction analyses on complex traits. *J Hered* 107, 372–379. [PubMed: 27025440]
- [29]. Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- [30]. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 88, 294–305. [PubMed: 21376301]
- [31]. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88, 76–82. [PubMed: 21167468]
- [32]. Mägi R, Morris AP (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 11, 288. [PubMed: 20509871]
- [33]. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, Burdett T, Hindorf L, Flicek P, Cunningham F, Parkinson H (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 45, D896–D901. [PubMed: 27899670]
- [34]. Yashin AI, Iachine IA, Christensen K, Holm NV, Vaupel JW (1998) The genetic component of discrete disability traits: an analysis using liability models with age-dependent thresholds. *Behav Genet* 28, 207–214. [PubMed: 9670596]
- [35]. Kulminski AM, Arbeeve KG, Culminkaya I, Arbeeve L, Ukraintseva SV, Stallard E, Christensen K, Schupf N, Province MA, Yashin AI (2014) Age, gender, and cancer but not neurodegenerative and cardiovascular diseases strongly modulate systemic effect of the Apolipoprotein E4 allele on lifespan. *PLoS Genet* 10, e1004141. [PubMed: 24497847]
- [36]. Liu L, Caselli RJ (2018) Age stratification corrects bias in estimated hazard of APOE genotype for Alzheimer's disease. *Alzheimers Dement (N Y)* 4, 602–608. [PubMed: 30450407]
- [37]. Kulminski AM, Culminkaya I, Arbeeve KG, Ukraintseva SV, Stallard E, Arbeeve L, Yashin AI (2013) The role of lipid-related genes, aging-related processes, and environment in healthspan. *Aging Cell* 12, 237–246. [PubMed: 23320904]
- [38]. Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol* 42, 5. [PubMed: 20170500]
- [39]. Zaitlen N, Kraft P (2012) Heritability in the genome-wide association era. *Hum Genet* 131, 1655–1664. [PubMed: 22821350]
- [40]. Plomin R, Haworth CMA, Davis OSP (2009) Common disorders are quantitative traits. *Nat Rev Genet* 10, 872–878. [PubMed: 19859063]
- [41]. Gibson G (2012) Rare and common variants: twenty arguments. *Nat Rev Genet* 13, 135–145. [PubMed: 22251874]
- [42]. Ridge PG, Mukherjee S, Crane PK, Kauwe JSK, Consortium ADG (2013) Alzheimer's disease: analyzing the missing heritability. *PLoS One* 8, e79771. [PubMed: 24244562]
- [43]. Kulminski AM, Huang J, Wang J, He L, Loika Y, Culminkaya I (2018) Apolipoprotein E region molecular signatures of Alzheimer's disease. *Aging Cell* 17, e12779. [PubMed: 29797398]
- [44]. Beach TG, Monsell SE, Phillips LE, Kukull W (2012) Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J Neuropathol Exp Neurol* 71, 266–273. [PubMed: 22437338]

Table 1

Basic demographic information about datasets included in the genetic analyses

Study	N_{Case}	N_{Control}	N_{Total}	Female%	Age_{Case} (SD)	Age_{Control} (SD)
FHS	413	3996	4409	54.77	79.85 (8.49)	62.77 (11.65)
HRS	281	5877	6158	57.31	80.44 (6.71)	73.69 (7.85)
LOADFS	1850	1866	3716	62.43	85.93 (8.39)	71.19(11.53)

FHS, Framingham Heart Study; HRS, Health and Retirement Study; LOADFS, National Institute on Aging's Late-Onset Alzheimer's Disease Family Study; Female%, the percentage of females in the study; Age (SD), the average age and its standard deviation.

Table 2

Numbers (and percentages) of SNPs included in the genetic analyses

SNPs set	FHS	HRS	LOADFS
Total	1548012 (100)	1959419 (100)	1812582 (100)
R6	669747 (43.26)	853777 (43.57)	790340 (43.60)
R6T	410449 (26.51)	524569 (26.77)	485562 (26.79)
R8	166488 (10.75)	211842 (10.81)	195684 (10.80)
R8T	46854 (3.03)	61925 (3.16)	56838 (3.14)

SNP, single-nucleotide polymorphism; FHS, Framingham Heart Study; HRS, the University of Michigan Health and Retirement Study; LOADFS, National Institute on Aging's Late-Onset Alzheimer's Disease Family Study; R6, Regions containing at least one AD-associated SNPs at $p < 5E-06$; R6T, Regions containing at least two AD-associated SNPs at $p < 5E-06$; R8, Regions containing at least one AD-associated SNPs at $p < 5E-08$; R8T, Regions containing at least two AD-associated SNPs at $p < 5E-08$.

Table 3

Estimated SNPs-based narrow-sense heritability (h^2) values (and their standard errors) of Alzheimer's disease (AD) in liability scale assuming three population prevalence rates

SNPs set	FHS		HRS		LOADFS		Meta-analysis					
	h^2_{G1}	h^2_{G2}	GXE	h^2_{G1}	h^2_{G2}	h^2_{G1}	h^2_{G2}	GXE	h^2_{G1}	h^2_{G2}	GXE	Proportion
AD Prevalence = 0.1												
Total	0.251 (0.212)	NA	0.301 (0.248)	0.326 (0.377)	NA	0.284 (0.104)	NA	0.069 (0.111)	0.280 (0.091)	NA	0.107 (0.101)	100
R6	0.158 (0.161)	0.095 (0.174)	0.300 (0.248)	0.366 (0.267)	0.000 (0.281)	0.294 (0.079)	0.002 (0.082)	0.074 (0.111)	0.274 (0.069)	0.018 (0.072)	0.111 (0.101)	93.96
R6T	0.212 (0.132)	0.043 (0.192)	0.301 (0.248)	0.309 (0.212)	0.036 (0.322)	0.198 (0.063)	0.091 (0.091)	0.073 (0.111)	0.207 (0.055)	0.079 (0.080)	0.111 (0.101)	72.43
R8	0.000 (0.082)	0.267 (0.208)	0.291 (0.248)	0.196 (0.139)	0.157 (0.354)	0.204 (0.044)	0.091 (0.098)	0.077 (0.110)	0.161 (0.038)	0.125 (0.086)	0.112 (0.101)	56.22
R8T	0.000 (0.045)	0.301 (0.212)	0.273 (0.248)	0.061 (0.074)	0.276 (0.370)	0.161 (0.029)	0.157 (0.099)	0.032 (0.108)	0.110 (0.023)	0.188 (0.087)	0.070 (0.099)	36.87
AD Prevalence = 0.2												
Total	0.311 (0.264)	NA	0.374 (0.308)	0.405 (0.468)	NA	0.353 (0.129)	NA	0.085 (0.138)	0.348 (0.113)	NA	0.133 (0.126)	100
R6	0.196 (0.199)	0.118 (0.216)	0.372 (0.308)	0.455 (0.332)	0.000 (0.350)	0.365 (0.099)	0.002 (0.102)	0.092 (0.137)	0.340 (0.085)	0.022 (0.089)	0.138 (0.126)	93.96
R6T	0.263 (0.164)	0.053 (0.238)	0.374 (0.308)	0.384 (0.263)	0.044 (0.399)	0.245 (0.078)	0.113 (0.114)	0.091 (0.138)	0.258 (0.068)	0.098 (0.099)	0.138 (0.126)	72.43
R8	0.000 (0.102)	0.332 (0.258)	0.361 (0.308)	0.243 (0.172)	0.194 (0.439)	0.253 (0.055)	0.113 (0.122)	0.095 (0.137)	0.200 (0.047)	0.155 (0.107)	0.139 (0.125)	56.22
R8T	0.000 (0.056)	0.374 (0.264)	0.339 (0.308)	0.076 (0.092)	0.343 (0.460)	0.200 (0.036)	0.195 (0.123)	0.039 (0.134)	0.136 (0.029)	0.233 (0.109)	0.087 (0.123)	36.87
AD Prevalence = 0.3												
Total	0.348 (0.295)	NA	0.417 (0.344)	0.452 (0.523)	NA	0.394 (0.144)	NA	0.095 (0.154)	0.389 (0.126)	NA	0.149 (0.141)	100
R6	0.219 (0.223)	0.131 (0.241)	0.416 (0.344)	0.508 (0.371)	0.000 (0.390)	0.408 (0.110)	0.003 (0.114)	0.102 (0.153)	0.380 (0.095)	0.024 (0.099)	0.154 (0.140)	93.96
R6T	0.293 (0.183)	0.059 (0.266)	0.418 (0.344)	0.429 (0.293)	0.049 (0.446)	0.274 (0.087)	0.126 (0.127)	0.102 (0.154)	0.288 (0.076)	0.110 (0.111)	0.154 (0.141)	72.43
R8	0.000 (0.114)	0.371 (0.288)	0.403 (0.344)	0.271 (0.192)	0.217 (0.491)	0.283 (0.062)	0.126 (0.137)	0.107 (0.153)	0.223 (0.052)	0.174 (0.120)	0.156 (0.140)	56.22

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

SNPs set	FHS		HRS		LOADFS			Meta-analysis			
	h^2_{G1}	h^2_{G2}	GXE	h^2_{G1}	h^2_{G2}	GXE	h^2_{G1}	h^2_{G2}	GXE	Proportion	
R8T	0.000 (0.062)	0.417 (0.295)	0.378 (0.344)	0.085 (0.103)	0.383 (0.513)	0.223 (0.040)	0.217 (0.138)	0.044 (0.150)	0.261 (0.121)	0.097 (0.137)	36.87

SNP, single-nucleotide polymorphism; FHS, Framingham Heart Study; HRS, the University of Michigan Health and Retirement Study; LOADFS, National Institute on Aging's Late-Onset Alzheimer's Disease Family Study; R6, Regions containing at least one AD-associated SNPs at $p < 5E-06$; R6T, Regions containing at least two AD-associated SNPs at $p < 5E-06$; R8, Regions containing at least one AD-associated SNPs at $p < 5E-08$; R8T, Regions containing at least two AD-associated SNPs at $p < 5E-08$; h^2_{G1} , the portion of phenotypic variance explained by the additive genetic variance corresponding to the G1 set of SNPs (i.e., either all autosomal SNPs or SNPs within the four aforementioned AD-associated regions); h^2_{G2} , the portion of phenotypic variance explained by the additive genetic variance corresponding to the G2 set of SNPs (i.e., SNPs that were not in G1 set); Proportion, the fraction of the additive genetic variance explained by the G1 set of SNPs (i.e., h^2_{G1} / h^2 , where $h^2 = h^2_{G1} + h^2_{G2}$); GXE , the portion of phenotypic variance explained by the gene-by-environment relationship matrix in the case of the LOADFS and FHS datasets that had family-based designs.