AWARD NUMBER: **W81XWH-16-1-0588**

TITLE: **Genetic Variation Underlying Traumatic Brain Injury (TBI) and Late-Onset Alzheimer's Disease (LOAD)**

PRINCIPAL INVESTIGATOR: Badri N. Vardarajan and Sandra Barral

CONTRACTING ORGANIZATION: Columbia University

REPORT DATE: Oct 2019

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
Oct 2019	Annual	09/15/2018 - 09/14/2019

4. TITLE AND SUBTITLE

Genetic Variation Underlying Traumatic Brain Injury (TBI) and Late-Onset Alzheimer's Disease (LOAD)

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0588

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Badri N. Vardarajan and Sandra Barral

E-Mail: bnv2103@columbia.edu, smb2174@cumc.columbia.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Columbia University Medical Center, 630 W 168th St, New York, NY 10032

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We previously computed episodic memory trajectories in 13,979 ethnically diverse elderly (ages 72 to 85 years) with two to 15 years of follow-up, and with known dementia status, age, education and APOE genotypes at baseline. Adjusted trajectories of episodic memory performance over time were estimated using Latent Class Mixed Models. In this reporting cycle, we aimed to identify common and rare genetic variation that may be responsible for uncharacterized genetic risk underlying individual differences in longitudinal performance on memory. We included the UK Biobank dataset to the analysis cohorts, which increased our sample size to 24,769 individuals. The Latent Class Mixed Model was used to derive episodic memory trajectories within each of the study cohorts. The slope of residualized episodic memory scores was used as outcome in an *APOE* stratified (non-*APOE* stratified, *APOE*_ε4 and *APOE*_nonε4 subsamples) genome=wide gene-based analyses. Association results from each study cohort within each subsample were meta-analyzed. Consistent with previous studies, the majority of the study participants maintain their memory performance over time (ranging from 51% to 98%). The strongest association signal for the episodic memory trajectories ($p=4.8 \times 10^{-8}$) was achieved at chromosomal region 6p22 among in the *APOE*_nonε4 subsample. Meta-analysis results in the non-*APOE* stratified and *APOE*_ε4 subsamples also yielded several chromosomal regions associated with memory performance over time, although significance level was diminished ($p\sim 10^{-6}$). Identifying genes associated with progression of episodic memory performance over time in older adulthood provides the possibility of identifying at-risk population subgroups that can benefit from possible interventions to enhance cognitive function.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified		19b. TELEPHONE NUMBER *(include area code)*

1. INTRODUCTION

There is a significant deficit in the literature investigating the possible association between traumatic brain injury (TBI) and increased susceptibility to develop Late Onset Alzheimer's Disease (LOAD) later in life. We hypothesize that traumatic brain injury interferes with aging process by accelerating individual's memory decline and possibly accelerating LOAD like neuro-degeneration. In addition, genetic risk factors including non-coding and highly penetrant coding variants may interact with exposure to TBI to modify risk of LOAD. We will first characterize trajectories of episodic memory change in subjects stratified by LOAD and TBI status using the longitudinal data from multiple cohorts including: The National Institute of Aging Late-Onset Alzheimer's Disease (NIA-LOAD) Family Study, The National Alzheimer's Coordinating Center and Alzheimer's Disease Genetics Consortium (NACC-ADGC), The Religious Orders Study and Rush Memory and Aging Project Alzheimer's Disease Center cohorts (ROSMAP), The Washington Heights-Inwood Columbia Aging Project (WHICAP) and The Chicago Health and Aging Project (CHAP). The Latent Class Mixed Model (LCMM) was used to estimate adjusted trajectories of episodic memory performance over time. Longitudinal studies of cognitive function have consistently clustered subjects into two episodic memory trajectories (EMTs): Stable and Decliner. We will stratified the study's participants into four different groups based on their LOAD and TBI clinical status: Group 1 corresponding to subjects with TBI and LOAD, Group 2 corresponding to subjects with TBI but not LOAD, Group 3 corresponding to subjects with LOAD but not TBI and Group 4 corresponding to subjects without TBI and without LOAD. We hypothesize that episodic memory trajectories signatures will be different across the different groups. We anticipate that study participants within Group 4 (non-demented subjects without TBI) will show a sustained memory performance over time, while study participants within Group 1 (subjects with both TBI and LOAD) will exhibit the most prominent decline in episodic memory. We will test whether the interaction between genetic variants and TBI may modify the risk of LOAD using genome-wide SNP data. We will validate the nominally significant genes ($p<0.001$) in the extreme-phenotype sub-sample including Groups 1 and Group 4. Genes achieving statistical significance in their interaction effects with TBI , will be further follow-up using available whole exome sequencing data in WHICAP and NIA-LOAD cohorts to investigate whether rare genetic variants may underlie these signals.

2. KEYWORDS

Episodic memory trajectories (EMTs), longitudinal evaluations, late onset Alzheimer's Disease, Traumatic Brain Injury (TBI).

3. ACCOMPLISHMENTS

- **What were the major goals of the project?**

The overall goals of the project were i) to identify trajectories of episodic memory performance using a longitudinal data from subjects with available clinical diagnosis of LOAD and TBI, and ii) to identify genes that interact with TBI and as consequence modify the risk of developing LOAD. Please see below our approved SOW for the project detailing the specific tasks for period of the grant proposal.

Tasks	YEAR 1				YEAR 2				YEAR 3			
	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Qtr 1	Qtr 2	Qtr 3	Qtr 4
SA1. To characterize trajectories of memory change in subjects stratified by AD and TBI status using the longitudinal data in 4,715 samples from the National Institute of Aging Late-Onset Alzheimer's Disease (NIA-LOAD) and Alzheimer's Disease Genetic Consortium (ADGC) cohorts.												
Task 1. Prepare and harmonize longitudinal phenotype data on memory performance from NIA-LOAD and ADGC cohorts												
a) Compute demographically adjusted z-scores of Logical Memory IA and IIA (sex,aged and education adjusted)												
Task 2. Prepare and harmonize the NIA-LOAD and ADGC GWAS datasets												
a) Quality control of genome-wide genotype data in both ADGC and NIA-LOAD cohorts (missingness, HWE etc)												
b) Quality control of sample level data in the two cohorts (relationship and sex checks, missingness etc)												
c) Joint imputation of the two datasets												
d) Joint Population substructure analyses of the two datsets												
Task 3: Implement GMM nested models to compute trajectories of memory decline in the joint NIA-LOAD and ADGC GWAS datasets												
SA2a. To identify genome-wide TBI-interacting genes that predict risk of AD using the non-stratified sample from the NIA-LOAD and ADGC cohorts.												
Task 1. Implement SREBIA on the NIA-LOAD and ADGC GWAS datasets												
Task 2. Implement special cases on the SREBIA algorithm (case-online and enhanced version)												
Task 3. Explore novel set based GXE gene-based methods to be applied on genome-wide data												
SA2b. To assess nominally significant TBI-interacting genes identified in SA2a in the stratified sample from the NIA-LOAD and ADGC cohorts.												

Task 1: To select the genes that are nominally significant at P<10e-03 for TBI genotype interaction in predicting LOAD risk	
Task 2: Select the most homogenous sample of rapid decliners and stable pleateaus (Group 1 and 4) as determined in Aim 1	
Task 3: Test TBI genotype interaction comparing rapid decliners and plateaus in nominally significant genes (SA2, Task1)	
SA2c.To validate the TBI-interacting genes identified in SA2b using independent cohorts with genome-wide SNP data available.	
Task 1. Prepare and harmonize longitudinal phenotype data on memory performance from RADC, CHAP and WHICAP	
Task 2. Prepare and harmonize the GWAS data from RADC, CHAP and WHICAP	
a) Quality control of genome-wide genotype data in the cohorts (missingness, HWE etc)	
b) Quality control of sample level data in the cohorts (relationship and sex checks, missingness etc)	
c) Joint imputation of the datasets	
d) Joint Population substructure analyses of the datsets	
Task 3. Validate the TBI interacting genes identified in SA2a and SA2b	
Task 4. Report results and evaluate potential manuscripts/conference presentations etc	
SA3. To investigate whether rare coding variants in the loci identified in SA2 interact with TBI in predicting risk of AD	
Task 1. Prepare WES from WHICAP and NIA-LOAD	
a) Alignment, variant calling and subsequent to determine high quality variant calls	
b) Harmonize variant data in the two datasets depending on capture kits, depth of coverage etc	
Task 2. Apply set based TBI genotype interaction analyses to determine genes with rare coding variants associated with LOAD in genes that were significant in SA2c	

Task 3. Conduct genotyping in samples without WES in both cohorts and re-assses TBI-genotype interaction in WES and genotype data												
Task 4. Explore other available WES datasets for replication												
Task 5. Report results and evaluate potential manuscripts/conference presentations etc												
Task 6. Explore possible collaborations with groups with functional expertise (RNA-seq, cell-based methods etc) to functionally characterize the genes and variants identified in Aims 1 through 3												

100% completed
In progress
Planned

What was accomplished under these goals?

- *Major activities, Specific objectives and significant results*

The major accomplishments under the described goals are as follows:
1. Addition of the UK Biobank dataset to augment the initially proposed sample size to derive episodic memory trajectories in over 24,000 individuals.
2. Implementation and optimization of latent class mix model method (LCMM) in a large and ethnically diverse sample.
3. Estimation of trajectories of episodic memory within TBI and LOAD stratified groups.
4. Genome-wide single variant and gene-based analysis to identify genetic variants associated with episodic memory trajectories.
5. Genome-wide single variant and gene-based approaches to TBI*SNP interaction as predictor of LOAD risk.

1. Augmentation of initially proposed sample size with the UK Biobank dataset. In our original proposal, we aimed to characterize trajectories of episodic memory over time change in a longitudinal sample of 4,878 participants from NIA-LOAD) and ADGC cohorts. To improve the reliability of the parameter estimates derived from the LCMM method in Aim1, last year, we included CHAP and ROSMAP as additional cohorts to increase the sample size to over 13000 individuals. Longitudinal assessments of memory and evaluation of dementia in UK Biobank dataset facilitated integration with other datasets available in this study. We applied and secured access to the UK Biobank data[1]. The UKB project is a prospective cohort study with deep genetic and phenotypic data collected on approximately 500,000 individuals from across the United Kingdom. To define cognitive impairment in the UKB cohort, we have used a 1.5-SD cut-off above demographically corrected episodic memory scores (adjusted for age, education, and sex). UKB study participants were classified as non-cognitive impaired (NCI) normal if their standardized adjusted memory scores on the pairs matching episodic test fell greater than 1.5 SD above the mean. As previously described in the UKB[2], scores on the pairs matching test (number of errors before recalling the positions of six matching pairs in a grid of 12 cards) reflect short-term visual episodic memory.
We derived episodic memory trajectories stratified by LOAD and TBI status in over 24000 individuals. Gene-environment interactions using set-based methods are ongoing.

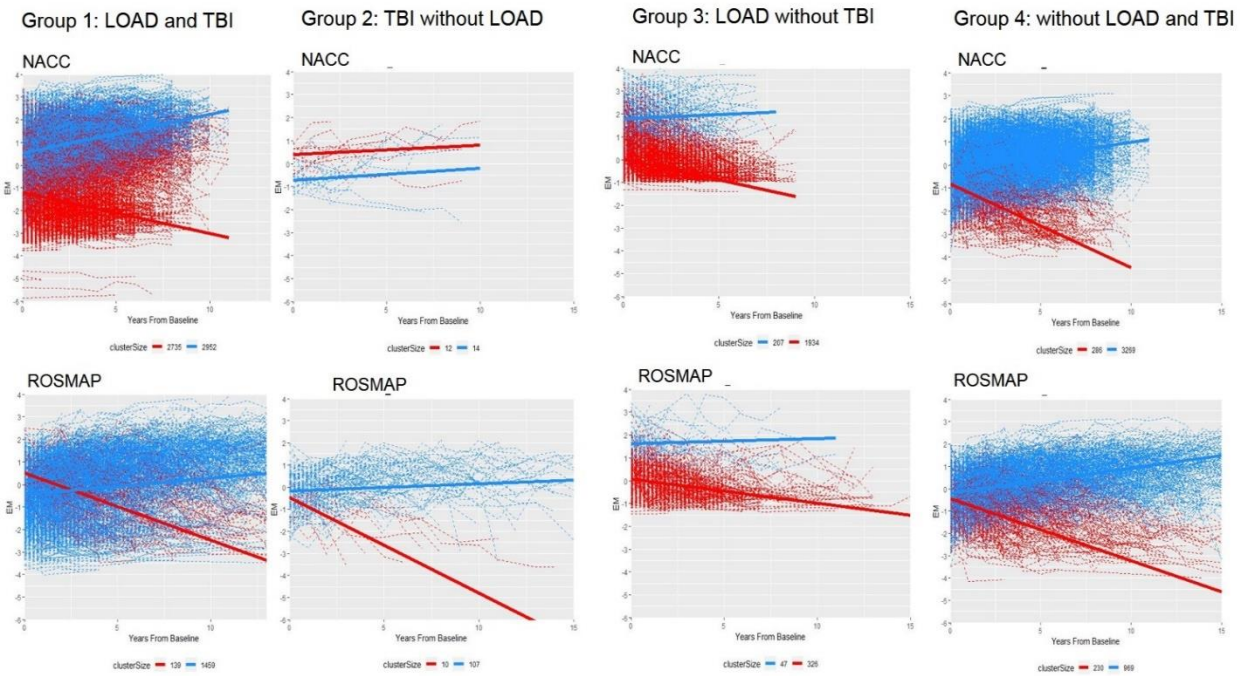
2. Implementation and optimization of LCMM method. Growth curve models represent a powerful analytical framework to model individual differences in cognitive change over time, as well as the variability of patterns of cognitive change between individuals. We used latent curve models to derive growth curve models. The latent model approach allows the possibility of incorporating variables with high degree of inter-individual variability such as the number of follow-up visits. This flexibility becomes especially relevant when participants across study cohorts were enrolled at different ages and/or were followed with different time intervals.

The initial optimization of the LCMM algorithm was conducted using the Washington Heights Columbia Aging Project (WHICAP) as training dataset. The optimized algorithm was subsequently expanded to additional study cohorts: NIA-LOAD, NACC_ADGC, ROSMAP and CHAP. We were able to gather a final sample size of 13,037 ethnically diverse elderly (ages 72 to 85 years) for whom relevant data was available: episodic memory scores from

follow-up clinical visits, dementia status, age, sex, education, and *APOE* genotypes. We identified two major clusters: $EMT_{Stables}$, consisting of individuals whose memory function remains stable or improved over time, and $EMT_{Decliners}$, consisting of individuals whose memory function declined. Consistent with previous studies, the majority of study participants maintain their memory performance over time. Compared to those with stable trajectory, individuals characterized as Decliners exhibited a significant decline of episodic memory, were more likely to have non-white ethnic background, fewer years of education, a higher frequency of ε4 allele at *APOE* gene and five times more likely to develop dementia. We published findings of this analysis in PLoS One[3].

3. Estimation of trajectories of episodic memory within TBI and LOAD stratified groups.
Using LCCM method in samples from NACC_ADGC and ROSMAP cohorts, we estimated episodic memory trajectories (EMTs) over time within each of the four previously defined groups: Group 1 corresponding to subjects with TBI and LOAD, Group 2 corresponding to subjects with TBI but not LOAD, Group 3 corresponding to subjects with LOAD but not TBI and Group 4 corresponding to subjects without TBI and without LOAD.

Figure 1. EMTs in NACC_ADGC and ROSMAP cohorts stratified by LOAD and TBI status.



Follow-up analyses will consider estimating EMTs in additional cohorts, CHAP and NIA-LOAD. Secondary analysis will consider linear mixed models to evaluate the impact of socio-economic (sex, education and ethnicity) , genetic factor (*APOE* genotype) and TBI in the EMTs. The linear mixed models will use the slope of the residualized episodic memory scores (EM_{res}) as a continuous outcome and the socio-economic, *APOE* gene and TBI as independent variables.

4. Genome-wide single variant and gene-based analysis to identify genetic variants associated with episodic memory trajectories.

Understanding genetic modulation of episodic memory will yield insight into the biological mechanisms underlying normal cognitive aging and dementia. Genetic studies based on longitudinal measures of episodic memory are scarce, predominantly focused on candidate genes, and have interrogated common genetic variation. We aimed to identify common and rare genetic variation that may be responsible for uncharacterized genetic risk underlying individual differences in longitudinal performance on memory.

We used the sample of 24,769 elderly individuals was ascertained from eight independent cohorts (ADGC_NACC, NIA-LOAD, ROSMAP, WHICAP, LLFS, CHAP, ADNI and UK Biobank). Data on study participants included socio-demographic variables, longitudinal scores on episodic memory (2 to 15 follow-up visits), clinical diagnosis of cognitive impairment and genomewide imputed genotyped data using the haplotype reference consortium for accurate imputation of low-frequency genetic variants. The Latent Class Mixed Model was used to derive episodic memory trajectories within each of the study cohorts. The slope of residualized episodic memory scores was used as outcome in an *APOE* stratified (non-*APOE* stratified, *APOE*_ε4 and *APOE*_nonε4 subsamples) genomewide gene-based analyses. Association results from each study cohort within each subsample were meta-analyzed.

Consistent with previous studies, the majority of the study participants maintain their memory performance over time (ranging from 51% to 98%). The strongest association signal for the episodic memory trajectories ($p=4.8 \times 10^{-8}$) was achieved at chromosomal region 6p22 among in the *APOE*_nonε4 subsample. Meta-analysis results in the non-*APOE* stratified and *APOE*_ε4 subsamples also yielded several chromosomal regions associated with memory performance over time, although significance level was diminished ($p\sim 10^{-6}$).

To our knowledge, this is the largest gene-based GWAS meta-analysis of high quality imputed common, rare and ultra-rare variants influence on episodic memory performance over time.

Identifying genes associated with progression of episodic memory performance over time in older adulthood provides the possibility of identifying at-risk population subgroups that can benefit from possible interventions to enhance cognitive function.

5. Genome-wide gene-based approaches to TBI*SNP interaction as predictor of LOAD risk.

We conducted preliminary analysis using a genome-wide single variant approach, i.e., individual SNPs in a genome-wide screen are tested for their interaction with TBI and the SNP*TBI interaction term is investigated as predictor of LOAD risk. Analyses were performed using PLINK software 1.07. In a sample of 5,713 participants from NACC_ADGC cohort, we found that the interaction of TBI with SNP variant rs1015290455 significantly increased the risk of LOAD, OR=5.2, $p=3 \times 10^{-6}$. The variant is an intronic SNP within Astrotactin-1 gene (*ASTN1*) located in chromosome 1. Astrotactin gene codifies for a neuronal adhesion molecule required for glial-guided migration of neuroblasts in cortical regions of developing brain, including among others, hippocampus. Replication in a sample of 950 participants from ROSMAP cohort, showed that interaction of TBI with SNP rs1015290455 was not associated with LOAD risk (p=0.417). Our preliminary results suggested that different SNPs in the same gene might be modulating risk of LOAD, further supporting the rationale for using gene-based approaches in the context of gene-environment interaction (i.e., multiple SNPs will be aggregated within a gene).

In the original project, we proposed the use of the algorithm implemented in SBERIA software to conduct set-based gene-environment interactions. However, we encountered substantial lack of information to successfully implement the program. As an alternative methodology, we have evaluated the performance of two methods to conduct a gene-based gene-environment interaction test-a) gene–environment set association test (GESAT)[4] and rareGE[5]. GESAT and rareGE are both publically available through R. GESAT tests for SNP-set by environment interactions using a variance component test, and estimates the main SNP effects under the null hypothesis using ridge regression. rareGE extends existing gene-environment interaction tests for multiple genetic variants under certain conditions. We tested both methods on genome-wide data, and observed that the type-1error is well controlled

We tested G*E interaction in three datasets- ROSMAP, NACC and NIALOAD and summarized the effects of genes using an inverse variance weighted meta-analysis approach implemented in METAL. We identified several genes with gene-TBI interaction p-values at a genome-wide corrected threshold of $p<10e^{-06}$ (Table 1). Interesting genes included *AGAP1 (p= 1.3E-12)* and *MARCH1 (3.2e-09)*, both involved in endosomal-lysosomal trafficking, and *AUTS2* (p=7.6e-10) which has been implicated in neurological disorders, autism and intellectual disability. We are validating the results in the remaining cohorts including the UK Biobank.

	ROSMAP n=1229	**NACC n=7543**	**NIALOAD n=480**	**METAL n=9252**
Gene	**P**	**P**	**P**	**P meta**
AGAP1	7.70E-28	0.007	0.003	1.30E-12
AUTS2	1.10E-24	0.084	1.80E-04	7.60E-10
AGBL4	3.80E-05	1.90E-05	0.015	3.20E-09
1-Mar	2.10E-12	0.025	2.20E-04	5.80E-08
SLC9A9	1.20E-05	0.006	5.50E-08	1.00E-07
USH2A	1.80E-14	0.218	5.50E-08	2.80E-07
BNC2	1.40E-06	0.014	0.006	4.20E-06
UTRN	6.30E-06	0.011	0.03	9.10E-06
CREB5	1.80E-09	0.142	6.30E-05	9.40E-06

Table 1: META-ANALYSIS results of the TBI*gene interaction

- **What opportunities for training and professional development has the project provided?**

The project has allowed us to create an educational setting in which we have mentored Mr. Yizhe Gao, a Master degree in Applied Mathematics by Columbia University (Class of 2017). Mr. Gao has been responsible for: i) implementing and developing computational methods to analyze phenotypic and omics data; ii) developing and/or applying computational methods and machine learning techniques for genomic analyses; iii) performing statistical genetic analysis of the study's cohorts. Drs. Barral and Vardarajan were jointly supervised Mr. Gao on a daily basis. His supervision included weekly presentations of the project's progress in the laboratory meetings of

the G.H. Sergievsky Center, Department of Neurology, where the work was presented to the other members of the Statistical genetics group.

- **How were the results disseminated to communities of interest?**

Manuscript describing the optimization of the LCMM method to estimate episodic memory trajectories along with its implementation for the analysis of a sample of more than 13,000 is published in the PLoS One journal. We have submitted results from the genome-wide association study of episodic memory for presentation in the AAIC 2020 conference in Amsterdam.

- **What do you plan to do during the next reporting period to accomplish the goals?**

Specific tasks planned for the next reporting period are:
1. Validate the findings from the G*E analysis using GESAT and rareGE in UK Biobank and other available cohorts.
2. Incorporate whole and whole genome sequencing data in WHICAP, Estudio Familiar de la Influencia Genetica Alzheimer (EFIGA) to identify rare variants underlying the G*E findings from the GWAS data in Aim 2.

4. **IMPACT**

- **What was the impact on the development of the principal discipline(s) of the project?**
  - Characterization of EMTs in a sample of 24,000 subjects from different ethnic backgrounds with and without Late Onset Alzheimer's Disease and no Traumatic Brain Injury that can be use as reference for future analyses.
  - Identify of genetic risk factors underlying episodic memory trajectories in over 24,000 individuals.
- **What was the impact on other disciplines?**

Nothing to Report

- **What was the impact on technology transfer?**

Nothing to Report

- **What was the impact on society beyond science and technology?**

Nothing to Report

5. **CHANGES/PROBLEMS**

- **Changes in approach and reasons for change**

The algorithm to be implemented SREBIA was no longer supported and the implementation is missing in the literature. We identified and successfully implemented two independent methods-GESAT (https://rdrr.io/github/lin-lab/iSKAT-GESAT/man/GESAT.html) and rareGE (https://www.ncbi.nlm.nih.gov/pubmed/25060534) implemented in R.

Actual or anticipated problems or delays and actions or plans to resolve them

NA

- **Changes that had a significant impact on expenditures**

Hiring of the post-doctoral research scientist is being more difficult than anticipated and it explains why there is large carryover into the second year. In an attempt to broader our options to find a suitable candidate, we have contacted the Division of Neurology Clinical Outcomes Research and

Population Science (NeuroCORPS) at Columbia University Medical Center. NeuroCORPS receive requests from a number of data science/computer science students at Columbia University looking for analytic opportunities. Therefore, we interviewed master students who have strong background in statistics or/and bioinformatics data analysis. We hired a Masters student for conducting analyses. We also hired a Staff Associate who is actively contributing to tasks in Aims 2 and 3 of the proposed project.

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

Nothing to Report

- **Significant changes in use or care of human subjects**

Nothing to Report

- **Significant changes in use or care of vertebrate animals**

Nothing to Report

- **Significant changes in use of biohazards and/or select agents**

Nothing to Report

6. PRODUCTS

- **Publications, conference papers, and presentations**

Lee S, Zhou X, Gao Y, Vardarajan B, Reyes-Dumeyer D, Rajan KB, Wilson RS, Evans DA, Besser LM, Kukull WA, Bennett DA, Brickman AM, Schupf N, Mayeux R, Barral S. *Episodic memory performance in a multi-ethnic longitudinal study of 13,037 elderly*. PLoS One. 2018 Nov 21;13(11):e0206803. doi:10.1371/journal.pone.0206803. eCollection 2018. PubMed PMID: 30462667; PubMed Central PMCID: PMC6248922.

- **Journals publications**

Nothing to Report

- **Books or other non-periodical, one-time publications**

Nothing to Report

- **Other publications, conference papers, and presentations**

Nothing to Report

- **Website(s) or other Internet site(s)**

Nothing to Report

- **Technologies or techniques**

Nothing to Report

- **Inventions, patent applications, and/or licenses**

Nothing to Report

- **Other Products**

Nothing to Report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?**
  - *Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate "no*

change."

***Example**:*

Name:	Badri Vardarajan
Project Role:	Principal Investigator
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2
Contribution to Project:	Dr. Vardarajan has contributed to the preparation and harmonization of the data from the different study cohorts, and reviewed/edited the manuscript
Funding Support:	N/A

Name:	Sandra Barral
Project Role:	Principal Investigator
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2
Contribution to Project:	Dr. Barral performed the statistical analyses, supervised research assistant analyses, interpreted the data, and wrote the manuscript.
Funding Support:	N/A

Name:	Yizhe Gao
Project Role:	Research Associate
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	6 months
Contribution to Project:	Mr. Gao integrated data from the study cohorts and performed the latent class mix models analyses, as well as implementation and optimization of additional statistical software needed for analyses purposes.
Funding Support:	N/A

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Dr. Vardarajan contributes efforts to additional projects: R21AG061722, U24AG056270-02S1, U01AG058654, U19AG063893, U01AG062943, R56AG063908 in addition to active other support previously reported.

Dr. Barral contributes efforts to U19AG063893 in addition to active other support previously reported

- **What other organizations were involved as partners?**

Nothing to Report

8. **SPECIAL REPORTING REQUIREMENTS**

- **COLLABORATIVE AWARDS**

Nothing to Report

- **QUAD CHARTS**

Nothing to Report

9. **APPENDICES**

Nothing to Report

10. ***THE APPENDICES.***

List of abbreviations

Alzheimer's Disease (AD)
Alzheimer Disease Research Centers (ADCs)
Apolipoprotein E (APOE)
The Alzheimer's Disease Genetic Consortium (ADGC)
Binary Sequence Alignment/Map (BAM)
Bootstrap likelihood ratio test (BLRT)
Base pairs (bp)
Base Quality Score Recalibration (BQSR)
Burrows-Wheeler Aligner (BWA)
The Children's Hospital of Philadelphia (CHOP)
Combined Annotation Dependent Depletion (CADD)
The Center for Inherited Disease Research (CIDR)
Copy Number Variation (CNVs)
Database of Single nucleotide polymorphism (dbSNP)
Deoxyribonucleic acid (DNA)
The Department of Defense and Veterans Brain Injury Center (DVBIC)
Genome Analysis Toolkit (GATK)
Gigabites (Gb)
Glasgow Coma Scale (GCS)
Gene environment interaction (GXE)
Genomic Evolutionary Rate Profiling (GERP)
Growth mixture modeling (GMM)

Genome-Wide Association Studies (GWAS)
Human Gene Mutation Database (HGMD)
Hardy-Weinberg Equilibrium (HWE)
Identity-by-descent (IBD)
Latent Class Growth Analysis (LCGA)
Linkage disequilibrium (LD)
Late Onset Alzheimer's Disease (LOAD)
Minimum Allele Frequency (MAF)
Rush Memory and Aging Project (MAP)
Microtubule-Associated Protein Tau (MAPT)
The Minority Aging Research Study (MARS)
Megabases (Mb)
Maximum Likelihood Algorithms (MLE)
The National Alzheimer's Coordinating Center (NACC)
The National Heart, Lung, Blood Institute Grand Opportunity Exome Sequencing Project (NHLBI GO ESP)
National Insitutes of Aging (NIA)
The National Institute of Aging Late-Onset Alzheimer's Disease Family Study (NIA-LOAD)
The National Institute of Neurological and Communicative Diseases and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA)
Nanograms (ng)
Online Mendelian Inheritance in Man (OMIM)
Polymorphism Phenotyping (POLYPHEN)
Principal Components (PCs)
Polymerase Chain Reaction (PCR)
Quality Control (QC)
The Rush Alzheimer's Disease Center Cohorts (RADC)
The Religious Orders Study (ROS)
Sorting Intolerant From Tolerant (SIFT)
Single nucleotide polymorphism (SNP)
Traumatic Brain Injury (TBI)
Uniform Data Set (UDS)
Variant Call Format (VCF)
The Washington Heights Inwood Columbia Aging Project (WHICAP)
Whole Exome Sequence (WES)
Whole Genome Sequence (WGS)

MARKING OF PROPRIETARY INFORMATION: Data that was developed partially or exclusively at private expense shall be marked as "Proprietary Data" and Distribution Statement B included on the cover page of the report. Federal government approval is required before including Distribution Statement B. The recipient/PI shall coordinate with the COR/GOR to obtain approval. **REPORTS NOT PROPERLY MARKED FOR LIMITATION WILL BE DISTRIBUTED AS APPROVED FOR PUBLIC RELEASE.** It is the responsibility of the Principal Investigator to advise the COR/GOR when restricted limitation assigned to a document can be downgraded to "Approved for Public Release." **DO NOT USE THE WORD**

"CONFIDENTIAL" WHEN MARKING DOCUMENTS. DO NOT USE WATERMARKS WHEN MARKING DOCUMENTS.

References

1. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.
2. Cornelis MC, Wang Y, Holland T, Agarwal P, Weintraub S, Morris MC. Age and cognitive decline in the UK Biobank. *PloS one.* 2019;14(3):e0213948.
3. Lee S, Zhou X, Gao Y, et al. Episodic memory performance in a multi-ethnic longitudinal study of 13,037 elderly. *PloS one.* 2018;13(11):e0206803.
4. Lin X, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics.* 2013;14(4):667-681.
5. Chen H, Meigs JB, Dupuis J. Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered.* 2014;78(2):81-90.