ARIC Manuscript Proposal #3327r

PC Reviewed: 9/10/19 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: A proteomic analysis of incident dementia: The ARIC Study

b. Abbreviated Title (Length 26 characters):

2. Writing Group:

Writing group members: Keenan Walker; Jingsha Chen; Myriam Fornage; Aozhou Wu; Adrienne Tin; Thomas Mosley; David Knopman; Christie Ballantyne; Eric Boerwinkle; Rebecca Gottesman; Josef Coresh (last author): others welcome

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _____ **[please confirm with your initials electronically or in writing]**

First author: **Keenan Walker, PhD**
Address: Johns Hopkins Hospital, Department of Neurology
Phipps 446
600 North Wolfe Street
Baltimore, MD 21287

Phone: 626-840-6216 Fax: 410-955-0672
E-mail: kwalke26@jhmi.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: **Josef Coresh**
Address: Johns Hopkins Bloomberg School of Public Health and Welch Center for Prevention, Epidemiology, and Clinical Research
2024 E. Monument St.,
Suite 2-600
Baltimore, MD 21287

Phone: 410-995-0495 Fax: 410-955-0476
E-mail: coresh@jhu.edu

3. Timeline: 6-9 months; manuscript submission in summer of fall of 2019

4. Rationale:

There is an urgent need for reliable non-invasive biomarkers that can be used to identify individuals with preclinical Alzheimer's disease and related dementias. While a number of previous studies have examined the degree to which blood protein levels are abnormal among persons living with dementia, the vast majority of these studies have used a targeted approach to examine only a small number of proteins. Although a number of biologically relevant proteins have been associated with Alzheimer's disease, many of these associations have been inconsistent, and as such, no blood-based biomarker has progressed beyond the discovery phase.[1] Given the complexity of Alzheimer's disease pathogenesis and the considerable size of the human proteome, there is likely a wide range of to-be-discovered protein changes that occur during the multi-decade prodromal phase of Alzheimer's disease. Although several case-control studies have evaluated abnormalities in blood protein levels among individuals with Alzheimer's disease, these studies have had little or no follow-up, relatively small sample sizes, and have assayed only a small portion of the human proteome.[2,3,12,13,4–11] Proteomic studies of Alzheimer's disease to date have been almost exclusively cross-sectional and therefore have been unable to determine which proteins are abnormally expressed in non-demented individuals who eventually progress to dementia.

The recent development of high-throughput technology for the characterization of the human proteome has enabled the simultaneous assessment of approximately 5,200 proteins in an accurate and reliable manner using a small amount of blood.[14,15] Although gene expression studies have been informative in identifying transcriptional changes that co-occur with dementia, the limited relationship between mRNA-protein pairs, with regard to expression levels, underscores the need to examine protein levels directly.[16–18] By conducting a large-scale prospective analysis of altered plasma protein levels in non-demented individuals who progress to dementia, we will take a data-driven approach to identify blood-based protein biomarkers that can be used for early disease detection and to better understand the molecular pathways that may be altered in the prodromal phase of dementia. While the former objective may lead to the development of a minimally-invasive method for dementia risk stratification for clinical trials, the latter objective would shed light on the complex systemic changes that occur outside the central nervous system in the years before dementia onset and thereby highlight novel pathways for therapeutic intervention.

The current study will use recently developed SOMAscan Multiplexed Proteomic technology to examine the relationship between the level of ~5,200 plasma proteins and dementia risk within the Atherosclerosis Risks in Communities (ARIC) Cohort, a large, biracial community-based sample of older adults. Candidate proteins identified in the discovery cohort will then be validated within a separate subset of the ARIC cohort. Candidate proteins will also be validated genetically by examining protein-genotype associations and using techniques such as Mendelian randomization to determine whether there is evidence for causal associations. We will conduct a series of secondary analyses to determine whether there are sex-, race-, age-, and *APOE* ε4-specific associations between plasma protein levels and risk of incident dementia. In the event that a number of candidate proteins are identified, we will (1) examine the predictive value of specific proteins and empirically-derived protein risk scores, (2) apply systems-level analyses to determine whether specific biologically informed protein pathways are overrepresented, and (3) determine the degree to which dysregulation of identified protein networks may be independently associated with dementia risk. We will also use machine

learning based data-driven approaches, such as feedforward multilayer neural networks, to identify candidate proteins and develop dementia prediction models. These deep learning networks are ideal for modeling complex associations among a large number of variables for the purposes of outcome prediction.[19,20]

5. Main Hypothesis/Study Questions:

H1. Multiple blood-based proteins (measured at Visit 3 and Visit 5) will be abnormally elevated or reduced among non-demented participants (cognitively normal and with mild cognitive impairment) who subsequently progress to dementia.

- **i.** We will identify a number of novel proteins and proteomic features that are associated with incident dementia after correction for multiple comparisons.
- **ii.** A subset of Visit 5 proteins associated incident dementia will also be associated with dementia when measured at Visit 3, i.e., there will be a partial overlap of candidate proteins assayed during midlife (Visit 3) and late-life (Visit 5).
- **iii.** Identified proteins will be used to develop a protein-driven risk score, which will be validated within the ARIC Cohort (i.e., the validation sample) and within external cohorts when data for external validation becomes available.
- **iv.** We will test several proteins that have been previously associated with Alzheimer's disease in multiple cohorts. This will test the utility of a number of Alzheimer's disease biomarkers (e.g., Alpha-1-antitrypsin, Alpha-2-macroglobulin, Apolipoprotein E, Complement C3, Complement factor H, Serum amyloid p-component) as measured in the SomaScan (recognizing predictive utility may vary across platforms which recognize different protein features).
- **v.** Candidate proteins will be validated using genetic information from ARIC participants. Specifically, we will identify *cis*- and *trans*-genotype-protein associations (protein quantitative trait loci [pQTLs]), which will then be examined for overlap with Alzheimer's disease-associated risk variants. Mendelian randomization will also be used to examine whether differentially expressed proteins may be causally related to dementia risk.

H2. Among participants who meet criteria for mild cognitive impairment at Visit 5, we will identify a number of novel blood-based proteins associated with progression from mild cognitive impairment to dementia over approximately five years (Visit 5 to Visit 6). We will also consider examining proteins associated with conversion from normal cognition to mild cognitive impairment over five years to facilitate an understanding of which proteins may be differentially associated with early- versus late-stage cognitive decline.

- **i.** The set of proteins associated with progression from mild cognitive impairment to dementia will be partially overlapping with the set of proteins associated with incident dementia in the full sample (H1). A unique set of candidate proteins will also be identified.
- **ii.** The set of proteins associated with progression from normal cognition to mild cognitive impairment will be partially overlapping with the set of proteins associated with incident dementia in the full sample (H1). A unique set of candidate proteins will also be identified.

H3. In stratified analyses we will identify multiple blood-based proteins at Visit 3 and Visit 5 that are differentially associated with incident dementia according to sex, race, age (dichotomized at the median), and *APOE* ε4 status. This will shed light on interactions and differences (modest quantitative as well as qualitative) in risk relationships and the biology in these subgroups.

H4. Protein pathway enrichment and protein co-expression network analyses using candidate proteins at Visit 3 and Visit 5 will implicate complement and coagulation, cytokine signaling, and synaptic transmission protein networks (among others) in progression to dementia.

H5. We will build a prediction model using machine learning methods. Compared to the prediction model using pre-selected proteins and other predictors, the model using machine learning method will achieve better performance and generalizability. Findings will be validated using 10-fold cross-validation or 2/3rd of the ARIC cohort with validation in the other 1/3rd, and using bootstrapping techniques. Subsequent external validation will be done in collaborating cohorts such as AGES or consortia which have approached ARIC.

H6. We will also consider comparing the predictive utility of several the protein-driven risk scores (identified as part of H1) to that of several previously defined genetic risk scores[21] for dementia using the ARIC Cohort.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Inclusion criteria: We will include all participants who (1) have SOMAscan protein measurements available from blood collected at Visit 3 or Visit 5. .
Exclusion Criteria: We will exclude non-white and non-African-American participants and non-white participants in Washington Co. and Minnesota, participants missing the education level variable, and participants missing information needed to classify cognitive status (i.e., normal/MCI/dementia classification) after Visit 3.

Proteomic measurement (exposure variables): Using plasma collected at Visit 3 (1993-95) Visit 5 (2011-13), proteins were measured using a Slow Off-rate Modified Aptamer (SOMAmer)-based capture array (SomaLogic, Inc, Boulder, Colorado). Using chemically modified nucleotides, this process transforms protein signals to a nucleotide signal quantifiable using relative florescence on microarrays. Previous work indicates a median intra- and inter-run coefficient of variation of approximately 5% and intra-class correlation coefficients of ~0.9.[11,12,22,23]

Primary outcome variables:
Incident Dementia after Visit 5: The analysis of Visit 5 proteins will relate Visit 5 protein levels to incident dementia occurring between Visit 5 and Visit 6 (2015-2017). This analysis will include participants who were classified as either cognitively normal or MCI at Visit 5. Dementia will be defined using both the information from the full Visit 6 examination with expert committee diagnosis and information captured in annual follow-up (AFU) interviews using the Six Item Screener (SIS) and the Ascertain Dementia 8-item Informant Questionnaire

(AD8). Date of dementia onset will be captured using the SIS and AD8, and dementia diagnosis will be confirmed at Visit 6 for those who attend Visit 6. Participants who attended Visit 5, but not Visit 6, and have SIS and AD8 information available from the AFU will also be included. For participants who did not attend Visit 6, the SIS, AD8, hospital discharge codes, and death certificates will be used to define dementia diagnosis and date of onset.

Incident Dementia after Visit 3: For the purposes of validation of Visit 5 candidate proteins, we will relate Visit 3 protein levels to incident dementia occurring between Visit 5 and Visit 6. For exploratory purposes, we will also relate Visit 3 protein levels to incident dementia occurring between Visit 3 and Visit 6 (this may be presented in a separate paper). In addition to the dementia cases ascertained after Visit 5 (using methods described above), this analysis will include dementia occurring before Visit 5. Dementia occurring before Visit 5 was ascertained at three levels. Consistent with the dementia classification after Visit 5, the current analysis will use dementia classified in person (level 1), using telephone interview (level 2), and using ICD-9 hospital discharge codes and death certificates (level 3).

Secondary outcomes:
Prevalent Dementia (Visit 5): To determine whether proteins associated with incident dementia overlap with the set of proteins associated with current dementia status (vs. non-dementia status), we will also examine the cross-sectional association between Visit 5 protein levels and prevalent dementia as part of a secondary analysis.

Cognitive Decline: We will examine the association of Visit 3 candidate protein levels (proteins implicated in the primary analysis) with ~20-year cognitive decline between Visits 4 and 6. Additionally, we will examine the association of candidate Visit 5 protein levels with ~5-year cognitive decline between Visits 5 and 6.

Amyloid Status: Using data from participants enrolled in the ARIC-PET study, will examine the association of Visit 3 and Visit 5 candidate proteins levels with cortical amyloid, as defined using florbetapir PET. Cortical amyloid status will be examined as a dichotomous variable (standardized uptake value ratio >1.2) and a continuous variable.

Analytic Plan

Plasma proteins associated with incident dementia. Participants will be randomly assigned to either a discovery cohort or a validation cohort with a roughly 66%:33% split, respectively. Using the discovery cohort, we will use Cox proportional hazard models to examine the extent to which each protein (after log transformation) is individually associated with incident dementia. Analyses will be adjusted for potentially confounding variables, including age at sample acquisition, sex, education, and *APOE* ε4 status. We may additionally adjust models for cardiovascular risk factors (i.e., BMI, hypertension, and diabetes) and other physiological variables known to affect protein levels and neurocognitive outcomes (eGFR). We will repeat the above analyses after stratifying participants by sex, race (black/white), age at sample acquisition (median split), and *APOE* ε4 allele number (0/≥1). Bonferroni correction will be applied to resulting *p-values* (e.g., .05/5,000 protein=1.00×10^{-5}) to identify candidate proteins. Using only

the set of candidate proteins that surpass this *p*-value threshold, we will repeat this series of analyses using the validation cohort.

Genetic validation. Candidate proteins will be validated using ARIC genetic data. Specifically, we will conduct genome-wide association studies (GWAS) for each candidate protein to identify *cis*- and *trans*-genotype-protein associations (protein quantitative trait loci [pQTLs]). Then, we will examine whether pQTLs overlap with Alzheimer's disease-associated risk variants. Mendelian randomization will also be used to examine protein-associated genetic variants in relation to incident dementia in order to estimate the potential causal relationship between candidate proteins and incident dementia.

Predictive performance of protein-driven risk score. To determine the potential for clinical application of individual proteins for predicting incident dementia among non-demented persons and persons with MCI, we will select multiple predictive proteins based on one or more of the following criteria: 1) top 10 proteins based on Cox regression coefficients, 2) top 5 proteins based on Cox regression coefficients, 3) top 3 proteins based on Cox regression coefficients, and 4) all predictive proteins based on a univariate AUC >0.70 and a correlation coefficient <0.40 between each pair of proteins, indicating high predictive accuracy and relative independence. Using these criteria, we will construct a protein-driven risk score based on a linear combination of selected proteins, with coefficients for each protein in the predictive risk score weighted according to Cox regression coefficients. We will use Cox proportional hazard models to determine whether risk scores act as independent predictors of dementia after adjusting for demographic variables, *APOE* ε4 status, and cardiovascular risk factors. Next, we will use time-dependent AUC to evaluate the predictive utility of each protein-based risk score alone and in combination with demographic characteristics and *APOE* ε4 status. We will also examine a stepwise Cox regression model constructed with p-enter=0.001 and p-exit=0.001.

Build prediction model using machine-learning. In addition to selecting predictors using traditional step-wise based approaches, we will also explore prediction models using machine learning methods. These include: 1) protein selection using penalized regression; 2) prediction model using random forest; 3) prediction model using feedforward multilayer neural network. Penalized regressions, like lasso regression and elastic net, have the advantage that they will conduct variable selection and build up the prediction model simultaneously. We will explore a Cox prediction model incorporating lasso and elastic net. Random forest is a tree-based ensemble method which generally performs well with high dimensional data.[24] It can provide variable importance scores for each predictor, which could help us identify important proteins. Neural networks are good at modeling complex associations in a prediction task. Usually, a neural network with one or two hidden layers and dense connectivity is capable of modeling any arbitrary associations between predictors and outcome in a high dimensional space. We will experiment with several networks with different numbers of layers and nodes. For all the three methods, hyper-parameter tuning will be conducted using cross-validation. We will compare the performance of the models developed using the traditional approach and using machine learning approaches by AUC or equivalent metrics.

Protein pathway enrichment. In order to examine the potential role of identified proteins in the dementia disease process, we will use Ingenuity Pathway Analysis (IPA) and DAVID

bioinformatics resources (the database for annotation, visualization and integrated discovery) to extract biological features/meaning associated with protein lists by identifying over-represented functional annotations assigned to each of the candidate proteins based on biological function.[25] Specifically, we will perform IPA canonical pathway, upstream regulator, and mechanistic pathway analyses. Additionally, we will perform KEGG enrichment analysis based on KEGG PATHWAY, which is a collection of manually drawn pathway maps, which integrates molecular-level information from large scale datasets to represent the current knowledge of molecular interaction and reaction networks. Additionally, we may use the weighted protein co-expression network analysis to construct a co-expression network to define distinct modules of protein co-expression enriched for gene ontologies associated with discrete biological processes.[26] Canonical pathways and co-expression modules will be quantified and examined in relation to dementia risk using Cox proportional hazard models, as described above.

Relative importance of pathways predicting AD. To determine the relative importance (or predictive value) of identified protein pathways as potential determinants of dementia risk, a pathway score will be defined by deriving the linear combination of the coefficients for proteins in each pathway as defined using the techniques outlined in the section above. We will examine time-dependent area under the receiver operating characteristic (ROC) curve for each protein pathway score associated with incident dementia.

Limitations. As in all omic analyses, the validity of measurement of different proteins varies across the the large number of proteins. The discovery effort is focused on identifying and validating new proteomic associations. Lack of sufficiently strong signal is not evidence of no association with a given protein since lack of signal may be due to limited power for a host of reasons including the need to be conservative when adjusting for multiple comparisons. We also recognize that visit 5 and visit 3 data are not exactly comparable since the age ranges are different but we hypothesize that strong proteomic signals will replicate for both later onset and earlier onset dementia.

7.a. Will the data be used for non-CVD analysis in this manuscript? __X__ Yes ____ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? __X__ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __X__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __X__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and

previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/aric/mantrack/maintain/search/dtSearch.html

___X___ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

MP# 3051. The association of middle and late-life blood pressure with conversion to MCI and dementia: The ARIC Study

MP# 3058. The association of late-life glycemia status with 3-year late-life cognitive decline and incident MCI/dementia: The ARIC Study

MP# 3903. Multi-omic data integration using systems approaches for mechanistic understanding of disease in the Atherosclerosis Risk in Communities (ARIC) Study

MP#3113. Identification of novel genetic variants associated with Alzheimer's disease in the Alzheimer's Disease Sequencing Project (ADSP)

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __X__ Yes ____ No

11.b. If yes, is the proposal

__X_ A. primarily the result of an ancillary study (list number* 2017.27_)
"Proteomic longitudinal ARIC study: SOMAscan of multiple visits"

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at https://www2.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.
Understood

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms.

http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.
Understood

References

1. Shi L, Baird AL, Westwood S, et al. A Decade of Blood Biomarkers for Alzheimer's Disease Research: An Evolving Field, Improving Study Designs, and the Challenge of Replication. *J Alzheimer's Dis*. 2018;62(3):1181-1198. doi:10.3233/JAD-170531.
2. Soares HD, Potter WZ, Pickering E, et al. Plasma Biomarkers Associated With the Apolipoprotein E Genotype and Alzheimer Disease. *Arch Neurol*. 2012;69(10):1310. doi:10.1001/archneurol.2012.1070.
3. Doecke JD, Laws SM, Faux NG, et al. Blood-Based Protein Biomarkers for Diagnosis of Alzheimer Disease. *Arch Neurol*. 2012;69(10):1318. doi:10.1001/archneurol.2012.1282.
4. Hye A, Lynham S, Thambisetty M, et al. Proteome-based plasma biomarkers for Alzheimer's disease. *Brain*. 2006;129(11):3042-3050. doi:10.1093/brain/awl279.
5. Cutler P, Akuffo EL, Bodnar WM, et al. Proteomic identification and early validation of complement 1 inhibitor and pigment epithelium-derived factor: Two novel biomarkers of Alzheimer's disease in human plasma. *Proteomics Clin Appl*. 2008;2(4):467-477. doi:10.1002/prca.200780101.
6. Choi J, Malakowsky CA, Talent JM, Conrad CC, Gracy RW. Identification of oxidized plasma proteins in Alzheimer's disease. *Biochem Biophys Res Commun*. 2002;293(5):1566-1570. doi:10.1016/S0006-291X(02)00420-5.
7. Henkel AW, Müller K, Lewczuk P, et al. Multidimensional plasma protein separation technique for identification of potential Alzheimer's disease plasma biomarkers: a pilot study. *J Neural Transm*. 2012;119(7):779-788. doi:10.1007/s00702-012-0781-3.
8. Ray S, Britschgi M, Herbert C, et al. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat Med*. 2007;13(11):1359-1362. doi:10.1038/nm1653.
9. Ijsselstijn L, Dekker LJM, Stingl C, et al. Serum levels of pregnancy zone protein are elevated in presymptomatic alzheimer's disease. *J Proteome Res*. 2011;10(11):4902-4910. doi:10.1021/pr200270z.
10. O'Bryant SE, Xiao G, Barber R, et al. A serum protein-based algorithm for the detection of Alzheimer disease. *Arch Neurol*. 2010;67(9):1077-1081. doi:10.1001/archneurol.2010.215.
11. Sattlecker M, Kiddle SJ, Newhouse S, et al. Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimer's Dement*. 2014;10(6):724-734. doi:10.1016/j.jalz.2013.09.016.
12. Kiddle SJ, Sattlecker M, Proitsi P, et al. Candidate blood proteome markers of Alzheimer's disease onset and progression: A systematic review and replication study. *J Alzheimer's Dis*. 2014;38(3):515-531. doi:10.3233/JAD-130380.
13. Sattlecker M, Khondoker M, Proitsi P, et al. Longitudinal Protein Changes in Blood Plasma Associated with the Rate of Cognitive Decline in Alzheimer's Disease. *J Alzheimer's Dis*. 2016;49(4):1105-1114. doi:10.3233/JAD-140669.
14. Ganz P, Heidecker B, Hveem K, et al. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA - J Am Med Assoc*. 2016;315(23):2532-2541. doi:10.1001/jama.2016.5951.
15. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. Gelain F, ed. *PLoS One*. 2010;5(12):e15004. doi:10.1371/journal.pone.0015004.
16. Zhang Y, Chen K, Sloan SA, et al. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *J Neurosci*. 2014;34(36):11929-11947. doi:10.1523/JNEUROSCI.1860-14.2014.
17. Sharma K, Schmitt S, Bergner CG, et al. Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci*. 2015;18(12):1819-1831. doi:10.1038/nn.4160.
18. Seyfried NT, Dammer EB, Swarup V, et al. A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Syst*. 2017;4(1):60-72.e4. doi:10.1016/j.cels.2016.11.006.
19. Tu J V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol*. 1996;49(11):1225-1231. doi:10.1016/S0895-4356(96)00002-9.
20. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Networks*. 1989;2(5):359-366. doi:10.1016/0893-6080(89)90020-8.

21. Stocker H, Möllers T, Perna L, Brenner H. The genetic risk of Alzheimer's disease beyond APOE ε4: systematic review of Alzheimer's genetic risk scores. *Transl Psychiatry*. 2018;8(1):166. doi:10.1038/s41398-018-0221-8.
22. Ganz P, Heidecker B, Hveem K, et al. Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. *JAMA*. 2016;315(23):2532-2541. doi:10.1001/jama.2016.5951.
23. Gold L, Ayers D, Bertino J, et al. Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. Gelain F, ed. *PLoS One*. 2010;5(12):e15004. doi:10.1371/journal.pone.0015004.
24. learning LB-M, 2001 undefined. Random forests. *Springer*. https://link.springer.com/article/10.1023/A:1010933404324. Accessed December 26, 2018.
25. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. doi:10.1038/nprot.2008.211.
26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. doi:10.1186/1471-2105-9-559.