Supplementary Material

Proteome Network Analysis Identifies Potential Biomarkers for Brain Aging

PROTEIN VALUES AND NETWORK ANALYSIS

As described in the manuscript, the protein measurements were conducted in two phases in FHS. Median intra-assay coefficients of variation (CVs) were 5.2% in Phase 1 and 3.6% in Phase 2; median intrassay CVs were 11.8% in phase 1 and 7.5% in phase 2, and the median interclass correlation in the FHS Offspring Cohort samples was >95% [1]. The 1,913 samples from which the CV values were calculated include a small number of participants (<3%) who were excluded from our analysis because they did not consent to genotyping.

Constructing the protein networks

Protein values were log-transformed and standardized within each phase, then pooled and normalized to means of 0 and standard deviations of 1 using Blom rank-based inverse normal transformations to reduce the potential for outsize effects of outliers on network construction and regression coefficients [2]. Due to significant between-plate differences in means as assessed by Welch's ANOVA, we regressed the values on plate ID using linear regression models to adjust for these batch effects, and residuals from these models were used in the protein network analysis.

To build a protein network, we calculated Pearson correlations between pairs of proteins in each of the two samples from the two phases of protein assays. Correlation matrices from Phases 1 and 2 were combined by calculating weighted averages of correlation coefficients from the two phases. We then used with default parameters to construct networks based on the combined correlation matrix. WGCNA uses hierarchical clustering to group proteins into modules based on the correlation of their expression [3]. We chose an unsigned approach to recognize biologically plausible ways by which an increase in one protein may correlate with an increase or a decrease in another, as in positive or negative regulation within signaling pathways. To choose the beta parameter for the model, we used the *pickSoftThreshold* function in the *WGCNA* R package to identify the lowest beta value for which the scale-free topology fit R^2 exceeded 0.8, which for this analysis was $\beta=2$. This analysis led to the identification of four modules (Supplementary

Figure 1A). We used the STRING (11.5) database with default parameters and hypergeometric tests with a significance threshold of $p<0.05$ to identify overrepresented Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, Gene Ontology Process, Function, and Component annotations, Reactome pathways and Wikipathways among proteins in modules of interest [4].

Eigenproteins

For each module of proteins identified in the network analysis, we calculated a numerical summary measure called an "eigenprotein," which has a single value for each participant. Similar to an eigengene [5], the eigenprotein is a weighted sum of the concentrations of proteins in a module for a given person. The weights used to calculate these eigenproteins were the coefficients of the first principal component of the proteins in each module (Supplementary Figure 1B).

Multiple testing considerations

For the analyses between modules and outcomes, we employed a Bonferroni correction to the resulting p-values (i.e., multiplying by four up to a value of 1.0), to control the family-wise error rate (FWER) for testing of the four modules with a given outcome. For the individual protein analyses, we chose to instead control the false discovery rate (FDR) rather than the FWER, and chose a cutoff value of 0.10, so as to avoid being overly conservative in our corrections and substantially decrease our statistical power for these analyses [6].

THE CHS COHORT

The CHS is a population-based, longitudinal cohort of 5,888 men and women aged 65 years or older at enrollment [7]. In year 2 of the study (1989-1990), 5,201 participants were recruited from four communities: Forsyth County, North Carolina; Washington County, Maryland; Sacramento County, California; and Pittsburgh, Pennsylvania. In year 5, an additional cohort of 687 predominantly African American participants was added. Examinations were performed at baseline and repeated annually until 1999 and included measures of subclinical disease and risk factors for cardiovascular outcomes. Institutional review boards at the University of Washington and at each study sites approved the study. All CHS participants provided written informed consent.

Proteomic analysis

Blood samples were drawn from all participants at their baseline examination and during follow-up clinic visits. Between enrollment and 1998-99, participants were seen in the clinic annually, and contacted by phone at 6-month intervals to collect information about hospitalizations and potential cardiovascular events. Participants with previously unthawed plasma blood samples from the 1992-3 clinic visit were included for proteomics analysis, which was conducted using the SomaScan 5k platform.

Dementia adjudication

A committee of neurologists and psychiatrists used Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria to adjudicate dementia prior to 1998-1999 using neuropsychological and neurological examinations, medical records, physician questionnaires, and informant interviews [8-10].

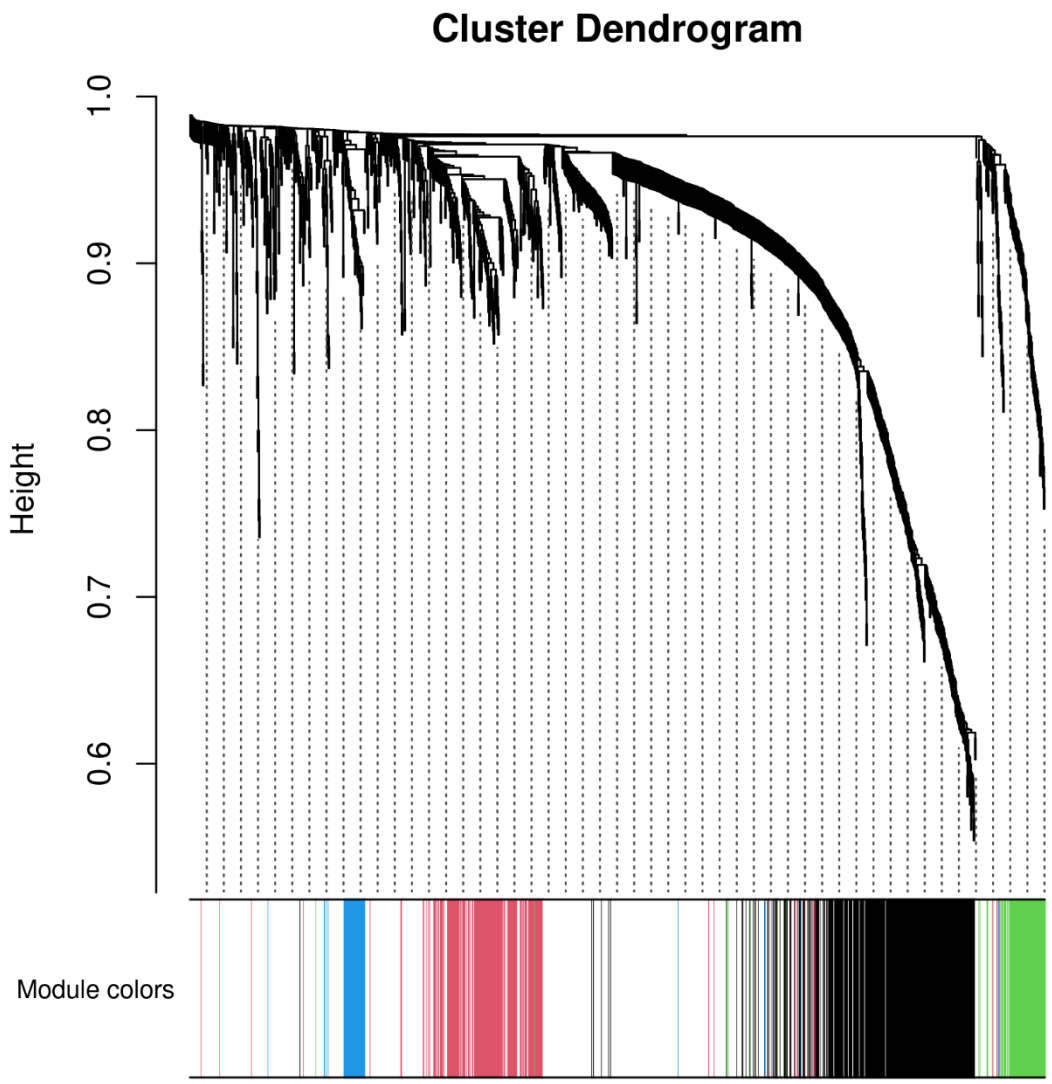
Cranial MRI

Cranial MRI scans were performed in CHS, as described previously [11, 12]. As described previously, neuroradiologists reviewed the MRI scans with a standardized protocol. Volumetric assessments were performed in the higher resolution images of the follow-up scans using FreeSurfer software. Briefly, T1-weighted images were reformatted to 1x1x1 mm^3 voxels, and a surface-based reconstruction of brain was generated, which included the hippocampus and WMH on T1-weighted sequences, which strongly correlates with WMH on FLAIR sequences.

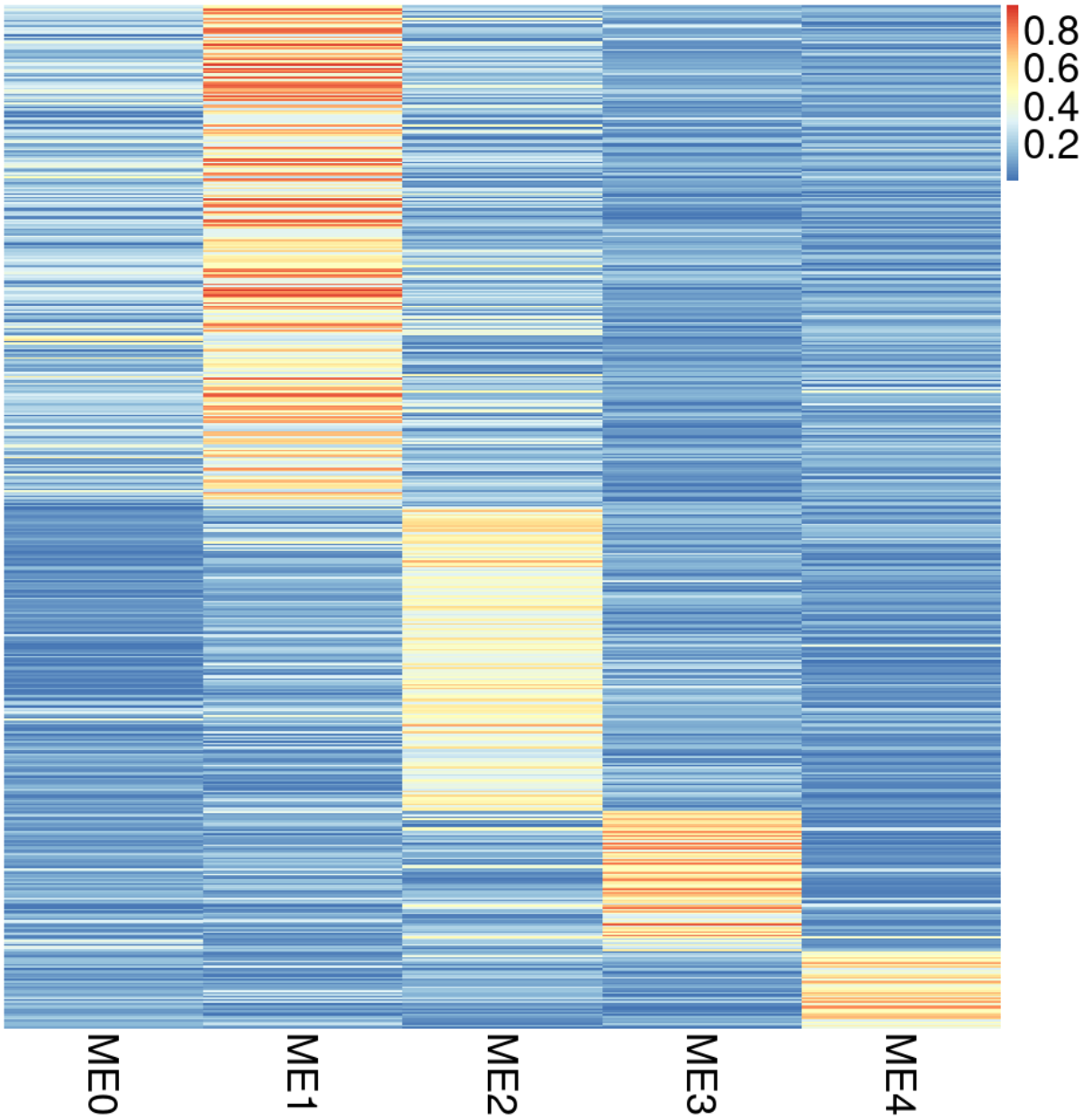
Proteome network analysis replication methods in CHS cohort

In pursuit of replicating the protein network analysis in CHS, we used the modules and corresponding protein weights generated from FHS data and calculated the corresponding eigenproteins in the CHS cohort as the weighted sum of the Blom inverse-rank normalized protein concentrations. We fit separate linear regression models to measure the association between M2 and M4 with total brain volume, and we fit separate Cox proportional hazards models to measure the associations between M2 and M4 with incident dementia and incident clinical AD. In all models, we adjusted for the same sets of covariates in the CHS analyses as in the FHS analyses.

Supplementary Figure 1A. Dendrogram of proteins and modules from WGCNA procedure.



Supplementary Figure 1B. Membership (weights) of proteins in each eigenprotein.



Values represent the correlation of each protein with the first principal component of the given module. Only proteins in one of M1-M4 are included.

REFERENCES

[1] Nayor M, Short MI, Rasheed H, Lin H, Jonasson C, Yang Q, Hveem K, Felix JF, Morrison AC, Wild PS, Morley MP, Cappola TP, Benson MD; CHARGE-Heart Failure Working Group; CHARGE-EchoGen Consortium; Ngo D, Sinha S, Keyes MJ, Shen D, Wang TJ, Larson MG, Brumpton BM, Gerszten RE, Omland T, Vasan RS (2020) Aptamer-based proteomic platform identifies novel protein predictors of incident heart failure and echocardiographic traits. *Circ Heart Fail* **13**, e006749.

[2] Blom G (1958) *Statistical estimates and transformed beta-variables*. Almqvist & Wiksell.

[3] Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform* **9**, 559.

[4] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47(D1)**, D607-D613.

[5] Foroushani A, Agrahari R, Docking R, Chang L, Duns G, Hudoba M, Karsan A, Zare H (2017) Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications. *BMC Med Genom* **10**, 16.

[6] Korthauer K, Kimes PK, Duvallet C, Reyes A, Subramanian A, Teng M, Shukla C, Alm EJ, Hicks SC (2019) A practical guide to methods controlling false discoveries in computational biology. *Genome Biol* **20**, 118.

[7] MPH for the Cardiovascular Health Study Research Group (CHS), Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, O'Leary DH, Psaty B, Rautaharju P, Tracy RP, Weiler PG (1991) The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-276.

[8] Fitzpatrick AL, Kuller LH, Ives DG, Lopez OL, Jagust W, Breitner JC, Jones B, Lyketsos C, Dulberg C (2004) Incidence and prevalence of dementia in the Cardiovascular Health Study. *J Am Geriatr Soc* **52**, 195-204.

[9] Kuller LH, Lopez OL, Jagust WJ, Becker JT, DeKosky ST, Lyketsos C, Kawas C, Breitner JC, Fitzpatrick A, Dulberg C (2005) Determinants of vascular dementia in the Cardiovascular Health Cognition Study. *Neurology* **64**, 1548-1552.

[10] Lopez OL, Kuller LH, Fitzpatrick A, Ives D, Becker JT, Beauchamp N (2003) Evaluation of dementia in the Cardiovascular Health Cognition Study. *Neuroepidemiology* **22**, 1-12.

[11] Longstreth WT Jr, Arnold AM, Beauchamp NJ Jr, Manolio TA, Lefkowitz D, Jungreis C, Hirsch CH, O'Leary DH, Furberg CD (2005) Incidence, manifestations, and predictors of worsening white matter on serial cranial magnetic resonance imaging in the elderly: the Cardiovascular Health Study. *Stroke* **36**, 56-61.

[12] Longstreth WT Jr, Dulberg C, Manolio TA, Lewis MR, Beauchamp NJ Jr, O'Leary D, Carr J, Furberg CD (2002) Incidence, manifestations, and predictors of brain infarcts defined by serial cranial magnetic resonance imaging in the elderly: the Cardiovascular Health Study. *Stroke* **33**, 2376-2382.