Influence of Genetic Variation on Plasma Protein Levels in Older Adults Using a Multi-Analyte Panel

Sungeun Kim1,2, Shanker Swaminathan1,3, Mark Inlow1,4, Shannon L. Risacher1, Kwangsik Nho1, Li Shen1,2, Tatiana M. Foroud1,2,3, Ronald C. Petersen5, Paul S. Aisen6, Holly Soares7, Jon B. Toledo8, Leslie M. Shaw9, John Q. Trojanowski8, Michael W. Weiner9,10, Brenna C. McDonald1, Martin R. Farlow1,11, Bernardino Ghetti12, Andrew J. Saykin1,2,3,4, for the Alzheimer’s Disease Neuroimaging Initiative (ADNI)

1 Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, 2 Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, 3 Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, 4 Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, Indiana, United States of America, 5 Department of Neurology, Mayo Clinic, Rochester, Minnesota, United States of America, 6 Department of Neurology, University of California San Diego, San Diego, California, United States of America, 7 Bristol Myers Squibb Co, Wallingford, Connecticut, United States of America, 8 Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, United States of America, 9 Departments of Radiology, Medicine and Psychiatry, University of California, San Francisco, San Francisco, California, United States of America, 10 Department of Veterans Affairs Medical Center, San Francisco, California, United States of America, 11 Department of Neurology, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, 12 Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana, United States of America

Abstract

Proteins, widely studied as potential biomarkers, play important roles in numerous physiological functions and diseases. Genetic variation may modulate corresponding protein levels and point to the role of these variants in disease pathophysiology. Effects of individual single nucleotide polymorphisms (SNPs) within a gene were analyzed for corresponding plasma protein levels using genome-wide association study (GWAS) genotype data and proteomic panel data with 132 quality-controlled analytes from 521 Caucasian participants in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort. Linear regression analysis detected 112 significant (Bonferroni threshold \( p = 2.44 \times 10^{-5} \)) associations between 27 analytes and 112 SNPs. 107 out of these 112 associations were tested in the Indiana Memory and Aging Study (IMAS) cohort for replication and 50 associations were replicated at uncorrected \( p < 0.05 \) in the same direction of effect as those in the ADNI. We identified multiple novel associations including the association of rs7517126 with plasma complement factor H-related protein 1 (CFHR1) level at \( p = 1.46 \times 10^{-10} \), accounting for 40 percent of total variation of the protein level. We serendipitously found the association of rs6677604 with the same protein at \( p = 9.29 \times 10^{-11} \). Although these two SNPs were not in the strong linkage disequilibrium, 61 percent of total variation of CFHR1 was accounted for by rs6677604 without additional variation by rs7517126 when both SNPs were tested together. 78 other SNP-protein associations in the ADNI sample exceeded genome-wide significance \( (5 \times 10^{-8}) \). Our results confirmed previously identified gene-protein associations for interleukin-6 receptor, chemokine CC-4, angiotensin-converting enzyme, and angiotensinogen, although the direction of effect was reversed in some cases. This study is among the first analyses of gene-protein product relationships integrating multiplex-panel proteomics and targeted genes extracted from a GWAS array. With intensive searches taking place for proteomic biomarkers for many diseases, the role of genetic variation takes on new importance and should be considered in interpretation of proteomic results.

Copyright: © 2013 Kim et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; Bioclinica, Inc.; Biogen Idec Inc.; Bristol Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IMCO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Mesro Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG10129 and K01 AG030514, the Dana Foundation, NIA R01AG19771, P30AG010333, U01AG032984, and the Foundation for the NIH for data analysis. Samples from the NCRAD, which receives government support under a cooperative agreement U24AG021886 awarded by the NIA were used in the ADNI. This study was also supported by an investigator-initiated research grant from the National Institutes of Health (R01 AG19771) to A.J.S., National Institute on Aging (P30AG10133-1851 Core Supplement to Drs. B. Ghetti and AJS), National Institute on Aging via Indiana Alzheimer Disease Center pilot grant (P30AG010133) to S.K, National Library of Medicine (R01 LM011360) and National Science Foundation (IIS-1117335) to L.S., National Institutes of Health Clinical and Translational Sciences Institute Pre-doctoral Training Fellowship (Training Grant TL1 RR025759) to S.L.R, National Library of Medicine (K99 LM011384) to K.N. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Citation: Kim S, Swaminathan S, Inlow M, Risacher SL, Nho K, et al. (2013) Influence of Genetic Variation on Plasma Protein Levels in Older Adults Using a Multi-Analyte Panel. PLOS ONE 8(7): e70269. doi:10.1371/journal.pone.0070269

Editor: John B.R Perry, Peninsula College of Medicine and Dentistry, United Kingdom

Received January 22, 2013; Accepted June 17, 2013; Published July 23, 2013
Introduction

Proteins play critical roles in numerous physiological functions and altered protein levels have been associated with disease [1,2,3,4,5,6,7,8]. Protein analytes are increasingly being employed as disease or treatment biomarkers with recent technological advances enabling simultaneous measurement of multiple proteins. However, progress in biomarker discovery and confirmation is likely to be limited without a better understanding of the genetic basis of protein analyte levels which can be analyzed as continuous phenotypes or quantitative traits (QTs) because variations in genes, which contain the information to encode proteins, may affect the production of proteins leading to altered levels and potentially to disease. Therefore, in genetics research for biomarker discovery and confirmation, an important goal is to robustly identify important functional variants in the genome regardless the difference in assessment method of protein levels. While the emphasis in the search for functional variants is often on the transcriptome or expression, protein analytic measurements can provide another level of assessment of association between genes and their corresponding protein products. Especially when proteins are known to play important roles in disease or treatment, influence of genetic variations associated with the encoded proteins should be considered. In this case, identified functional variants may be used to stratify protein analytes in their interpretation as diagnostic, prognostic, or therapeutic response biomarkers for disease or treatment.

Several studies [9,10,11] have investigated the association of single nucleotide polymorphisms (SNPs) with protein levels in humans. The first study [9] used two dimensional difference gel electrophoresis (2D DIGE) technology [12] to measure 544 proteins in 24 human lymphoblastoid cell lines and identified protein expression quantitative trait loci (pQTLs). The second study [10] performed GWAS analysis on the levels of 813 plasma proteins from 96 healthy older individuals, using an aptamer-based proteomic technology [13]. The third study [11] investigated the role of SNPs on levels of 42 serum and plasma proteins measured from 1200 fasting individuals using Enzyme-Linked Immunosorbant Assay (ELISA)-based method (R&D systems, HSTA00C). All three of these studies investigated protein QTLs (pQTLs) either from a large number of subjects with a small number of proteins or from a small number of subjects with a large number of proteins. Although some findings in these studies were replicated in a separate report, none of these studies evaluated the genetic effects on two completely independent cohorts as discovery and replication cohorts.

In this study, we used quality-controlled (QC-ed) genome-wide genotype array data and baseline plasma proteomic data by multiplex immunoassay on the Myriad Rules Based Medicine (RBM) Human DiscoveryMAP panel v1.0 using the Luminex100 platform, different from ones used in the previous studies [9,10,11], from 521 non-Hispanic Caucasian participants in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort for the discovery phase and from 59 non-Hispanic Caucasian participants in the Indiana Memory and Aging Study (IMAS) cohort for the replication phase. We mainly investigated the effect of individual SNPs (eQTL) within a gene on the corresponding plasma protein level, analyzing 140 gene-protein association pairs. In addition, we examined the percent of total variation in plasma protein levels explained by each SNP ($R^2_{\text{SNP}}$) while accounting for the effect of relevant covariates. This study identified novel associations and replicated some existing findings. Approximately half of the current findings from the ADNI cohort were replicated in the IMAS cohort. The current study also demonstrated that individual SNPs showed remarkable variations in their effects ($R^2_{\text{SNP}}$).

Materials and Methods

Ethics Statement

This study was approved by institutional review boards of all participating institutions and written informed consent was obtained from all participants or authorized representatives.

Alzheimer's Disease Neuroimaging Initiative (ADNI)

Data used in this study were obtained from the ADNI database (http://adni.loni.usc.edu/). ADNI was launched in 2004 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations, as a multi-year public-private partnership. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California–San Francisco. ADNI is a multisite longitudinal study, including more than 800 participants, aged 55 to 90, recruited from over 50 sites across the United States and Canada. The participants include approximately 200 cognitively normal older individuals (normal control; NC) to be followed for 3 years, 400 patients diagnosed with mild cognitive impairment (MCI) to be followed for 3 years, and 200 patients diagnosed with early AD to be followed for 2 years at 6- or 12-month intervals. Longitudinal imaging [14,15], performance on neuropsychological and clinical assessments [16] and biological samples [5,17] were collected at baseline and at follow-up visits for all or a subset of participants. APOE ε2/ε3/ε4 genotype and genome-wide genotyping data [18] are available on the full ADNI sample and longitudinal proteomic data [5] was obtained for 566 selected participants. Further information about ADNI can be found at http://www.adni-info.org.

Indiana Memory and Aging Study (IMAS)

IMAS is an ongoing longitudinal study, including euthymic older adults with significant cognitive complaints (CC) including memory concerns in the context of cognitive test performance that is within the normal range, patients with early and late MCI (EMCI and LMCI) or mild AD, and age-matched cognitively normal controls (NC) without significant cognitive complaints. Details regarding participant selection criteria and characterization have been described previously [19,20]. Neuropsychological
and clinical assessments, structural and functional MRI, and blood samples were collected for all participants. APOE ε2/ε3/ε4 genotype and GWAS data were available on the full IMAS sample. Amyloid PET and longitudinal imaging at follow-up were included and their gender and identity-by-descent were checked to identify genotyping or coding error and to avoid the potential confounding effect due to gender ambiguity or consanguinity such as sibling pairs. In addition, to restrict the present analysis to non-Hispanic Caucasians, 988 founders with known ancestry information from HapMap [24] phase 3 (HapMap3) release 2 were used as reference data in the population stratification step and merged with ADNI samples. In short, ADNI and HapMap3 samples were merged and the multidimensional scaling analysis was performed using PLINK with identity-by-state (IBS) pairwise distance matrix of the merged data. This analysis grouped ADNI and HapMap3 samples in the principle component analysis (PCA) space, allowing us to identify which ADNI samples were grouped with which HapMap3 samples with known ancestry. ADNI participants who were grouped with HapMap3 samples with CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection) or TSI (Toscani in Italia) ancestry and had self-reported race/ethnicity as “non-Hispanic/white” were selected as non-Hispanic Caucasian participants. When ADNI samples were separately grouped from any HapMap3 participants, their gender and identity-by-descent were checked to identify genotyping or coding error and to avoid the potential confounding effect due to gender ambiguity or consanguinity such as sibling pairs.

Participants and Overall Quality Control Procedure

To reduce the potential bias of population stratification, analyses were restricted to non-Hispanic Caucasian participants from the ADNI (n = 521) (Table 1) and IMAS (n = 59) (Table 2) cohorts. Samples in other racial/ethnic groups were not included in the study because the number of samples in other racial/ethnic groups was relatively small (less than 10%) for genetic analysis in the ADNI and IMAS cohorts. Included participants had GWAS and plasma proteomic data that passed all quality control (QC) procedures (Figure 1) which were similar for the ADNI and IMAS cohorts. Tables 1 and 2 present demographic information for these samples. Data collection and multi-staged QC steps for genotype and proteomic data, each performed separately, are described below and Figure 1 shows the overall flow of this multi-staged QC procedure.

Genotyping and Quality Control

The ADNI protocol for collecting genomic DNA samples from all 818 ADNI participants has been previously described [18,21]. Genotyping using the Illumina Human610-Quad BeadChip (Illumina, Inc., San Diego, CA), which contains over 600,000 SNP markers, was performed according to the manufacturer’s protocols (Infinium HD Assay; Super Protocol Guide; rev. A, May 2008). A GWAS data set reprocessed in GenomeStudio v2009.1 (Illumina) was downloaded and used for subsequent analyses including all QC procedures.

Table 1. Sample Characteristics in the ADNI data*.

| Characteristics       | All  | NC   | MCI  | EMCI | LMCI | AD   |
|-----------------------|-----|-----|-----|-----|-----|-----|
| Number of samples     | 521 | 53  | 360 | 108 | 63  | 45  |
| Gender (M/F)          | 325/196 | 28/25 | 234/126 | 63/45 |
| Baseline Age (years; mean±SD) | 75.0±7.36 | 75.6±5.73 | 74.9±7.38 | 75.0±8.01 |
| Education (years; mean±SD) | 15.6±2.98 | 15.8±2.62 | 15.7±3.01 | 15.2±3.05 |
| Handedness (R/L)      | 475/46 | 49/4 | 328/32 | 98/10 |
| APOE ε4 status (ε4–/ε+4) | 247/274 | 49/4 | 163/197 | 35/73 |

ADNI: Alzheimer’s Disease Neuroimaging Initiative, NC: normal controls, MCI: mild cognitive impairment, EMCI: early mild cognitive impairment, LMCI: late mild cognitive impairment, AD: Alzheimer’s disease.

*521 non-Hispanic Caucasian participants. doi:10.1371/journal.pone.0070269.t001

Table 2. Sample Characteristics in the IMAS data*.

| Characteristics       | All  | NC   | CC   | EMCI | LMCI | AD   |
|-----------------------|-----|-----|-----|-----|-----|-----|
| Number of Samples     | 59  | 16  | 10  | 9   | 6   | 6   |
| Gender (M/F)          | 22/37 | 5/11 | 9/9 | 3/7 | 4/5 | 1/5 |
| Age (years; mean±SD)  | 72.58±7.07 | 71.79±5.30 | 71.56±7.86 | 72.72±8.09 | 73.49±7.28 | 76.13±7.77 |
| Education (years; mean±SD) | 16.92±2.31 | 17.38±1.71 | 17.44±1.89 | 16.80±2.49 | 14.89±2.47 | 17.33±3.27 |
| APOE ε4 status (ε4–/ε+4) | 39/20 | 10/6 | 12/6 | 9/1 | 6/3 | 2/4 |

IMAS: Indiana Memory and Aging Study, NC: normal controls, CC: cognitive complaints, EMCI: early mild cognitive impairment, LMCI: late mild cognitive impairment, AD: Alzheimer’s disease.

*59 non-Hispanic Caucasian participants. doi:10.1371/journal.pone.0070269.t002
self-reported ethnicity and racial information were used in order to be identified as non-Hispanic Caucasian participants. This population stratification analysis identified 749 ADNI participants as non-Hispanic Caucasians and this information was also used for sample selection for the QC procedures of the plasma proteomic data (Figure 1).

The IMAS employed a highly similar protocol to ADNI for collecting blood samples. Genotyping was performed on 85 genomic DNAs using the Illumina HumanOmniExpress Bead-Chip (Illumina, Inc., San Diego, CA), which contains over 700,000 SNP markers, according to the manufacturer’s protocols (Infinium HD Assay; Super Protocol Guide; Rev. A, May 2008). APOE e2/e3/e4 genotyping was separately performed. The two APOE SNPs were added to the Illumina genotype data prior to assessment of data quality. All genotype data, including two APOE SNPs, underwent the same standard QC assessment (Figure 1) using PLINK v1.07 [22] as the ADNI genotype data. After the final QC step of genotype data, the population stratification, 73 samples were selected as non-Hispanic Caucasians and this information was also used for sample selection for the QC procedures of the plasma proteomic data (Figure 1).

Figure 1. Quality control procedure for genetic and proteomic data.

doI:10.1371/journal.pone.0070269.g001

Plasma Measurement and Quality Control
Plasma samples were collected from blood for all ADNI participants in the morning before breakfast and after an overnight fast at each visit following the ADNI protocol (for further details see http://www.adni-info.org/Scientists/Pdfs/adniproceduresmanual12.pdf). Briefly, blood samples at each visit were collected into two 10 mL EDTA vacutainer® tubes and centrifuged at room temperature, within one hour of collection, at 3000 rpm (1500 g). The plasma fluid was transferred into a labeled 13 mL polypropylene transfer tube, capped, placed upright in dry ice, and shipped to the ADNI Biomarker Core Laboratory at University of Pennsylvania.

A large set of 0.5 mL EDTA plasma samples for a subset of ADNI participants was selected and shipped to Myriad Rules Based Medicine, Inc. (RBM, Austin, TX). Sample selection criteria were explained in the ADNI Biomarker Core Plasma Proteomics Data Primer (http://adni.loni.ucla.edu/wp-content/uploads/2010/11/BC_Plasma_Proteomics_Data_Primer.pdf). A set of 190 protein levels from plasma for each selected individual was measured by multiplex immunoassay on the Human DiscoveryMAP panel v1.0 using the Luminex100 platform by RBM. Additional technical details are available as a white paper from the RBM (http://www.rulesbasedmedicine.com). All QC procedures (Figure 1) by the RBM and ADNI Biomarker Core were previously described in [2,8] and in the ADNI Biomarker Core Plasma Proteomics Data Primer and statistical analysis plan.

In brief, the first QC procedures by the RBM and ADNI
Biomarker Core included data transformation if needed, outlier selection (outside ±five standard deviation from mean) and replacement, and imputation of missing/non-measurable values. After these QC steps, 146 out of the 190 proteins in the RBM Human DiscoveryMAP panel for 356 ADNI participants at the baseline visit passed the QC measures and were used in the subsequent QC analyses.

In the second QC step specific to the present study, non-Hispanic Caucasian participants identified from the population stratification analysis of genotype data were selected. RBM (ADNI Biomarker Core Plasma Proteomics Data Primer) reported serum samples as a potential type of specimens for a few datasets (n = 5) and these five datasets were excluded. 521 sample data from plasma at the baseline visit were chosen after these steps. Next, in order to reduce any effect of extreme outlying analyte levels, defined as larger or smaller than four standard deviations from the mean level of each analyte, these extreme values (maximum n = 4 per analyte) were identified and removed from further statistical analyses.

For the IMAS cohort, the ADNI protocol for collecting plasma samples was adopted and all collected plasma samples were stored at the Specimen Storage Facility (SSF) biorepository at Indiana University. A set of 0.5 mL EDTA plasma samples from 68 IMAS participants was selected and shipped to RBM. A set of 185 protein levels (5 assays were discontinued) from plasma for each selected individual was measured by multiplex immunassay on the RBM Human DiscoveryMAP panel v1.0 that was used for the ADNI samples. The collected proteomic analyte data underwent the initial QC steps, similar to the ADNI Biomarker Core QC steps including outlier detection and replacement, and imputation of missing/non-measurable (“low”) values. Data were not transformed given the limited number of samples in the replication set which did not permit robust distribution analysis. Analytes with less than 10% of missing or non-measurable (“low”) values were imputed (maximum number of imputed values per subject = 3) as follows: missing and non-measurable (“low”) values were imputed to be the mean of the non-missing values and one half of the lowest non-missing value for that analyte, respectively. Finally for each analyte, outliers outside five standard deviations from the mean were assigned the value of the nearest non-outlier point (25 analytes had 1 outlier per analyte). In the second QC step, the QC-ed genotype data were used for selecting non-Hispanic Caucasian participants and one sample was discarded from the analysis due to undetermined diagnosis, resulting in 59 samples with QC-ed proteomic and genotypic data as the replication set. However, further outlier removal outside ±four standard deviation from the mean level of each analyte was not performed due to the limited size of samples.

**Annotation of RBM Proteomics**

In order to investigate the genetic influence on each of the 146 protein levels at the baseline visit, we annotated these 146 analytes by identifying their protein-coding genes by mapping the UniProtKB/Swiss-Prot Accession Numbers of the analytes to the Entrez Gene IDs/HUGO Gene Symbols (Table S1). Then, this list of Gene IDs/Symbols was compared to the list of QC-ed ADNI SNPs. In order to map the QC-ed SNPs to the corresponding genes, we used the Illumina annotation information as an initial mapping step. The annotation information was further tuned using SNP Annotation Tool (http://snp-nexus.org/) [25,26] based on NCBI36/hg18 and SNP and CNV Annotation Database (http://www.sctdb.org/newinterface/about.html). All selected SNPs were inside genes or intergenic within 500 kb margin from gene boundary. If SNPs were intergenic between two genes investigated in the study, SNPs were mapped to the closer gene. Ten proteins were excluded because there were no SNPs within corresponding genes in the QC-ed Illumina Human610-Quad genotype data and four proteins which did not have a UniProtKB/Swiss-Prot Accession Number were excluded. Mapping of proteins to genes was not exactly one-to-one in some cases; seven analytes—Fibrinogen, Follicle-Stimulating Hormone (FSH), Luteinizing Hormone (LH), Thyroid-Stimulating Hormone (TSH), Creatine Kinase-MB (CK-MB), Amphiregulin (AR), Ferritin (FRTN) had more than one UniProtKB/Swiss-Prot Accession Numbers and were mapped to multiple genes (Table S1). Cortisol was included, despite not matching to a UniProtKB/Swiss-Prot Accession Number, because some studies have shown an association with AD, memory performance or cognitive performance [27,28,29,30]. Although the main focus of the present analysis was not the effect of proteomic and genetic data on AD pathology, CRH (corticotropin releasing hormone) and Adrenocorticotropic hormone (ACTH), were selected as the corresponding genes to cortisol. After the completion of initial analyses, RBM updated the Swiss-Prot Accession number for Tumor Necrosis Factor receptor 2 (TNFR2) from Q92956 to P20333, changing its protein-coding gene from TNFRSF14 to TNFRSF1B. In addition, the protein name, “complement factor H” has been changed to “complement factor H-related protein 1” together with the Swiss-Prot Accession number from P08603 to Q03591, subsequently changing the protein-coding gene from CFH to C4H2. Regarding “complement factor H-related protein 1”, RBM confirmed that the updated annotation should be used for all analyses. Therefore, we repeated the analysis with new sets of SNPs from TNFRSF1B and CFH1 for TNFR2 and CompFactH, respectively. This slightly increased the number of tests performed. Table S1 contains both of the initial and updated annotation information for TNFR2 and CompFactH.

**Summary of Sample Data and Association**

After all QC procedures for genotype and proteomic data, 132 QC-ed analytes (listed under “Tested” column in Table S1) and 1992 QC-ed SNPs belonging to 137 genes for 521 ADNI participants at the baseline visit remained. Tables 1 and 2 present demographic information for the tested sample of 521 ADNI participants and 59 IMAS participants, respectively. In Table S1, a column, “Tested”, indicates which associations were investigated in the ADNI cohort.

**Statistical Analyses**

This study investigated the genetic influence on each plasma protein level at the single SNP level within the protein-coding gene. We tested the additive genetic model for each association if the minimum sample size criterion (≥10 samples) within each genotype group was satisfied in the ADNI data. When the minimum sample size criterion was not met, the dominant genetic model instead of the additive genetic model was tested. Potential covariates (baseline age, gender, education and handedness) were included in the model if they were significantly associated with the plasma protein level (uncorrected p<0.05) using a linear regression analysis in the ADNI data (Table S2).

Because the aim of the present study was to investigate the genetic influences on plasma protein levels, not specific to MCI or AD, baseline diagnosis and Apolipoprotein E (APOE) e4 allele carrier status, the largest known genetic risk factor for sporadic AD, were included in the statistical model as shown below:
Model. Analyte = constant + SNP + significant covariates + APOE ε4 status (ε4−/ε4+) + baseline diagnosis (NC/MCI/AD) for the ADNI cohort (diagnosis NC/CC/EMCI/LMCI/AD) at the time of plasma collection for the IMAS cohort or error. One exception was plasma ApoE analyte. Because SNPs within APOE gene could be highly correlated with APOE ε4 status, resulting in unstable statistical results, APOE ε4 status was not included in the model for ApoE level. The statistical model was fitted for each association with additive or dominant genetic model depending on satisfaction of minimum sample size criterion, mentioned above. Analyses were performed using PLINK v1.07. The linear regression function in MATLAB R2009b (The MathWorks, Inc., Natick, MA) was used to test associations of SNPs on the X chromosome in order to separately analyze males and females. For the 132 analytes and 1992 SNPs, a total of 2046 association tests were performed by PLINK or MATLAB in the analyses (see Correction for multiple testing section below) from the ADNI data and identified significant associations from the ADNI sample were investigated using the IMAS data for replication. In the replication analysis, significance of potential covariates (age, gender, and education, but not handedness because all subjects were right-handed) was evaluated with the IMAS samples and the minimum sample size criterion was 10% (6 or more) samples due to the limited size of the replication data set. If this minimum sample size criterion was not satisfied in the IMAS cohort, a dominant genetic model was tested instead of the genetic model, tested in the ADNI cohort.

All analytes for the ADNI sample used in this study were examined for normality of distribution within each diagnostic group by the ADNI Biomarker Core and a large set of the analytes were log-transformed ("LOGTRANS in ADNI" in Table S1). However, these initial procedures did not remove the bi-modal nature or skewness of some analytes over all 521 samples. Although one assumption of linear regression, performed in this study, was the normality of error distribution, the error distribution could change from association to association, depending on the dependent variables (analytes) and its main predictors (SNPs). In order to make it feasible to quantitatively assess the error distributions for 2046 associations, we computed the skewness and kurtosis of analytes and visually assessed the distribution. Therefore, the distribution of analytes over all samples was examined and associations were selected for further scrutiny on the basis of (1) the absolute value of skewness > 2, (2) the absolute value of kurtosis > 2, or (3) the subjective assessment of bi-modal distribution from histogram and normal Quantile-Quantile plot. Then, Bootstrap analyses [31] (1000 iterations) were conducted to determine if an analyte with non-normality, e.g., bimodality, resulted in non-normality (Kolmogorov-Smirnov test \( p < 0.05 \)) of the sampling distribution of the regression coefficients and, thus, potentially biased p-values. Also, non-parametric analysis of variance (Kruskal-Wallis test [32]) implemented in MATLAB R2009b was performed for these analytes, pre-adjusted for all covariates used in the parametric analyses by using the regression weights. Finally, the p-values from Kruskal-Wallis test were compared to p-values from the linear regression to determine concordance.

For each of the significant associations in the analyses, the percent of total variation explained by each SNP (\( R^2_{\text{SNP}} \)) from the linear regression model was calculated over all participants while accounting for the effect of other relevant covariates using hierarchical multiple regression as follows:

\[
R^2_{\text{SNP}} = \text{adjusted } R^2 \text{ model with SNP and covariates} - \text{adjusted } R^2 \text{ model with covariates.}
\]

Correction for Multiple Testing

In this study, there were 2046 association tests in the ADNI sample between a set of 132 analytes and a set of 1992 SNPs. Therefore, all associations with uncorrected \( p < 2.44 \times 10^{-5} = 0.05 \div 2046 \) tests (Bonferroni threshold) were considered significant for the ADNI data.

In the IMAS data, due to a limited number of samples (n = 59) and the relatively small number of tests (only significant associations in the ADNI data were tested for replication), no multiple correction methods were applied and any associations with uncorrected \( p < 0.05 \) were considered significant and replicated.

Results

Discovery Sample (ADNI)

Analyses investigated the effect of individual markers in each gene on corresponding plasma protein levels. Table S3 lists 112 associations between 27 analytes and 112 SNPs in 28 genes at the pre-determined significance level (Bonferroni corrected \( p < 0.05 \), equivalent to uncorrected \( p < 2.44 \times 10^{-5} = 0.05 \div 2046 \) tests) and all the SNPs had at least 11 samples in each genotype group in the ADNI data. (In Table S1, a column, “Identified”, indicates which associations were identified as significant in the ADNI cohort.) Figure 2 summarizes these 112 associations with tested genetic model, linkage disequilibrium (LD) among SNPs within each gene is shown along the x-axis of the heatmap and -log_{10}(uncorrected \( p \)) is visualized using the color scale, shown to the left of the heatmap. Figure S1 shows zoomed association results between the 27 identified analytes and SNPs within the 28 corresponding genes (two gene-protein associations between complement factor H-related protein 1 and SNPs in CFH and CFHRI/ CFHR1 genes is shown within one panel, CompFactH) using LocusZoom [http://csg.sph.umich.edu/locuszoom/] [33]. Just for visualization purpose in Figure S1, the most significant analyte-SNP pair for each panel was used in order to select the genetic model. All identified SNPs were located in or within a 50 kb margin of each gene. The association of TG (Thyroxine-Binding Globulin) levels with SNPs in SERPIN4 (serpin peptidase inhibitor, clade A (alpha-1 antiprotease, antitypsin), member 7) gene was significant only for males. The association of rs7517126 with plasma complement factor H-related protein 1 (CFHR1) level at \( p = 1.46 \times 10^{-6} \), accounts for 40 percent of total variation of the protein level. We serendipitously found the association of rs6677604 with the same protein at \( p < 9.29 \times 10^{-12} \). Although these two SNPs were not in the strong LD (highlighted blocks in Figure 2) based on the default algorithm [34] implemented in HaploView [35], 61 percent of total variation of CFHR1 was accounted for by rs6677604 without additional variation by rs7517126 when these two SNPs were tested together in multiple linear regression model. 78 other associations including novel and confirmed associations had uncorrected \( p < 5 \times 10^{-6} \) (generally accepted genome-wide significance level in many GWAS studies [36]). Figure S2 shows the similar pattern of genetic effect within each diagnostic group to the overall pattern for all subjects. This observation in the ADNI sample at least partially supports that the relationship between SNPs and analytes is independent of AD or MCI diagnosis.

749 associations in the ADNI sample were identified as showing non-normality as described in the Materials and Methods section. Thus, the normality of regression coefficients of identified SNPs was examined and all tests, described in the Materials and Methods section, did not identify any significant non-normal cases, reducing the likelihood of significantly biased results due to non-normal distribution of analytes. In addition, the Kruskal-Wallis
test [32] found five significant associations that were potentially biased (four associations of FAS and one association of HCC-4 (Chemokine CC-4)) from linear regression analyses. However, according to the normality test of regression coefficient of SNPs from bootstrapping, these five associations from linear regression were less likely to be biased cases due to the non-normal distribution of analytes.

Table 3 shows the percent of total variation explained by the most significantly associated SNP ($R^2_{SNP}$) within each gene while accounting for the effect of other relevant covariates over all ADNI participants. These SNPs accounted for 3 to 61 percent of the total variation. The associations in Table 3 were sorted in the descending order of $R^2_{SNP}$. Figure S2 presents scatter plots of the top 12 associations (only rs6677604 for CompFactH) from Table 3 over all ADNI samples and within each diagnostic group. Table S3 shows the complete results of 112 significant associations in the ADNI data.

Replication Sample (IMAS)

In the IMAS cohort, one analyte (GST-alpha), associated with two SNPs in the ADNI cohort, didn’t pass proteomic data QC and two SNPs among the 112 SNPs from the ADNI cohort, associated with AAT (Alpha-1-Antitrypsin), and TBG levels (one per each analyte) didn’t pass QC of genotype/imputed data. Also, only one male participant had a minor allele of rs1804495 in $SERPINA7$ gene. The association of rs1804495 with TBG for males could not be tested. Therefore, a total of 107 associations were tested in the IMAS cohort. For each association test, if the SNP didn’t satisfy the minimum sample criterion (>5 samples), then a dominant model was tested instead of the genetic model, tested in the ADNI cohort. This analysis replicated 50 associations (Table S4) with uncorrected $p<0.05$ ($p$-values were not corrected for multiple testing due to limited detection power with the modest sample size of 39). The direction of genetic effect based on beta coefficients and allele coding (major/minor) for these 50 associations was the same as the ADNI results. In Table S1, the last column, “Replicated” indicates which associations were replicated in the IMAS cohort. The most significant association in the ADNI cohort between rs6677604 and complement factor H-related protein 1 was replicated in the IMAS cohort with a dominant genetic model different from the genetic model (additive) in the ADNI cohort due to the minimum sample criterion. For easy comparison between discovery and replication results, Table 4 summarizes the replicated results among all results in Table 3 in the same order as they are listed in Table 3. In case the association from Table 3 was not replicated in the IMAS sample, the most significant association for each analyte in the IMAS sample is listed.

Discussion

In this study, we were able to identify or confirm the strong influence of genetic variation on circulating plasma protein levels in an older adult population. In some cases this relationship was extraordinarily strong accounting for as much as 61 percent of the variance with $p<9.29\times10^{-112}$ with 79 other associations exceeding conventional GWAS correction for multiple testing ($p<3.5\times10^{-15}$). The biological relevance of the top thirteen gene-protein...
Table 3. List of significant (Bonferroni corrected p<0.05, equivalent to uncorrected p<2.44×10−5) associations in the ADNI cohort.

| Analyte | Gene | Chr | SNP | Allele (Major/Minor) | Location* | Function_Class | GM** | R²_{SNP} | beta | p-value |
|---------|------|-----|-----|----------------------|-----------|----------------|-------|-----------|------|---------|
| CompFactH | CFH   | 1   | rs6677604 | G/A | Intron | ADD | 0.610 | 1567.000 | 9.29E-112 |
| CompFactH | CFHR1 | 1   | rs7517126 | A/G | Intergenic | ADD | 0.397 | 1487.458 | 1.46E-60 |
| IL-6r  | ILSR  | 1   | rs4129267 | C/T | Intron | ADD | 0.386 | 1.56E-58 |
| ApoA-IV | APOA4 | 11  | rs1263167 | A/G | Intergenic | ADD | 0.223 | 0.145 | 6.84E-31 |
| Fetuin-A | AHSG  | 3   | rs2070633 | C/T | Intron | ADD | 0.219 | 0.061 | 4.50E-31 |
| ACE    | ACE   | 17  | rs4343 | G/A | Intergenic | ADD | 0.6156031 | 5.20E-28 |
| ApoE   | APOE  | 19  | rs429358 | T/C | Intergenic | ADD | 0.196 | 0.125 | 7.27E-30 |
| THP    | UMOD  | 16  | rs4293393 | T/C | 5’ near gene | ADD | 0.192 | 0.153 | 2.48E-28 |
| PARC   | CCL18 | 17  | rs854462 | T/C | Intergenic | DOM | 0.148 | 0.146 | 9.15E-24 |
| HCC-4  | CCL16 | 17  | rs11080369 | A/C | Intron | DOM | 0.138 | 0.218 | 8.03E-19 |
| TBGaa  | SERPNA7 | X | rs1804495 | G/T | Intergenic | DOM | 0.137 | 0.112 | 1.24E-12 |
| Angiotensinogen | AGT | 1 | rs4762 | C/T | Intergenic | DOM | 0.676 | 4.74E-18 |
| IL-16  | IL6   | 15  | rs4778636 | G/A | Intron | DOM | 0.136 | 0.067 | 5.08E-19 |
| MIPF-1 | CCL23 | 17  | rs854656 | A/C | Intergenic | ADD | 0.219 | 0.069 | 2.61E-13 |
| MCP-2  | CCL8  | 17  | rs12602195 | A/G | Intergenic | ADD | 0.223 | 6.84E-31 |
| CD40   | CD40  | 20  | rs1569723 | A/C | Intergenic | ADD | 0.223 | 0.112 | 1.24E-12 |
| ApoH   | APOH  | 17  | rs1788481 | C/T | Intergenic | DOM | 0.059 | 0.133 | 1.18E-08 |
| GST-alpha | GSTA1 | 6   | rs4715326 | T/C | Intergenic | ADD | 0.061 | 4.50E-31 |
| Nr-CAM | NRCAM | 7   | rs10487849 | T/C | Intergenic | ADD | 0.059 | 0.055 | 1.36E-08 |
| FactorVII | F7   | 13  | rs488703 | C/T | Intron | DOM | 0.054 | 0.098 | 2.00E-08 |
| LPA    | LPA   | 6   | rs13202636 | T/C | Intergenic | ADD | 0.048 | 0.223 | 2.27E-07 |
| FAS    | FAS   | 10  | rs1571011 | A/C | Intergenic | ADD | 0.045 | 0.039 | 3.20E-07 |
| ICAM-1 | ICAM1 | 19  | rs1799969 | G/A | Intergenic | DOM | 0.045 | 0.079 | 1.03E-06 |
| AAT    | SERPNA1 | 14 | rs7151526 | C/A | Intergenic | DOM | 0.043 | 0.063 | 7.44E-07 |
| EGFR   | EGFR  | 7   | rs13244925 | C/A | Intron | ADD | 0.041 | 0.245 | 9.12E-07 |
| CDSL   | CDSL  | 1   | rs2765501 | G/A | Intergenic | ADD | 0.040 | 0.047 | 1.52E-06 |
| AFP    | AFP   | 4   | rs10020432 | G/A | 3’ nearGene | ADD | 0.032 | 0.073 | 2.26E-05 |
| Cystatin-C | CST3 | 20  | rs2424590 | T/C | Intergenic | ADD | 0.032 | 0.039 | 4.75E-06 |

The most significant association of each analyte is listed in the descending order of R²_{SNP}, which is the fraction of variation accounted for by a SNP while adjusting for the effect of covariates. R²_{SNP} is defined as follows: R²_{SNP} = adjusted R² of model with SNP and covariates − adjusted R² of model with covariates.

*SNP: Single Nucleotide Polymorphism; ADNI: Alzheimer’s Disease Neuroimaging Initiative; Chr: Chromosome.

**Gene and SNP information are based on the Genome Build 37.3.

***Statistical model did not include APOE ε4 status due to possible collinearity.

▲▲Association was significant only in males.

doi:10.1371/journal.pone.0070269.t003

Further associations based on R²_{SNP} in the ADNI cohort was examined further. Each SNP accounted for 14 to 61 percent of total variation. Among these top 13 gene-protein associations, 9 gene-protein associations were replicated in the IMAS cohort. Two associations, Tamm-Horsfall glycoprotein (THP) and Angiotensinogen, were not replicated in the IMAS cohort. One association, Thromboxine-binding globulin (TBG) with the gene located on the X chromosome, could not be assessed as only a single male participant had the minor allele of the SNP. Association between ApoE level and ApoE gene was not replicated warranting further investigation.

Among the top 13 associations, a SNP in the CFHR1 (complement factor H-related 1) gene (rs7517126) showed the very strong influence (R²_{SNP}) on the plasma level of complement factor H-related protein 1, a complement regulatory protein and a member of complement factor H family. In this study, another SNP in the CFH (complement factor H) gene (rs6677604) showed the larger influence (R²_{SNP}) than rs7517126 although these two SNPs were not in strong LD (Figure 2 and Figure S1). Similar results of these two SNPs on the expression level of CFHR1 gene were observed in the previous study [37]. It was not clearly explained why rs6677604 has the larger influence on the plasma level of complement factor H-related protein 1 than rs7517126, warranting further investigation. Variations in the CFH and CFHR1 genes have been studied for disease susceptibilities, including age-related macular degeneration [38,39], dense deposit disease [40], atypical hemolytic-uremic syndrome [41,42], and systemic lupus erythematosus (SLE) [43]. Plasma complement factor H has been identified as a potential diagnostic biomarker for AD [44]. Interestingly, the SNP with the strongest relationship in
LD (pairwise r² = 0.75) influenced plasma level of Interleukin-16. Two SNPs (rs4778636, rs11857713) in strong

tant for a variety of CD4

serum and plasma [10,11].

for AD susceptibility [51], but

been inconsistent [54,55]. The

function of ApoE is one of the most
tensively studied genes, especially for AD susceptibility [51], but

p < 0.05 in an additional analysis. The

interaction effect between rs429358 and baseline diagnosis on.

functioning as an inhibitor of agonist induced platelet aggregation

hence been recently reported [10].

Apolipoprotein E (ApoE) protein plays a role in lipid

metabolism, combining with lipids to form lipoproteins. Also,

ApoE is a major component of very-low-density lipoproteins which

remove excess cholesterol from the blood and are known to be

bound to high density lipoproteins (HDLs), forming HDL-E,

functioning as an inhibitor of agonist induced platelet aggregation

[50]. The

APOE

e

2/

e

3/

e

4 genotypes. Thus, this SNP not only

associated with plasma ApoE in the ADNI cohort is one of two key

SNPs determining

association of ApoE with AD, but the findings are

inconsistent [56,57]. The

significant effect of rs1263167 on the

plasma level of ApoA-IV was replicated in the IMAS cohort, but

has not yet been reported in other studies. One study found the

serum level of ApoA-IV to be up-regulated in AD patients [57]

this study (rs667604) has been previously associated with SLE [43].

Interleukin-6 receptor (IL-6r), rs4129267 in the IL6R (interleukin 6 receptor) gene had the strongest relationship. The

minor allele of the SNP up-regulated the plasma level of IL-6r in

the present cohorts. Previous studies reported this association in

patients with systemic sclerosis [46], and

recently reported [10].

CompFactH

Apolipoprotein A-IV (ApoA-IV) is another apolipoprotein in

plasma that is involved in lipid metabolism. Previous studies have

reported an association of ApoA-IV with AD, but the findings are

inconsistent [56,57]. The

significant effect of rs1263167 on the

plasma level of ApoA-IV was replicated in the IMAS cohort, but

has not yet been reported in other studies. One study found the

serum level of ApoA-IV to be up-regulated in AD patients [57]
and another study observed the association of ApoA-IV deficiency with increased Aβ deposition [56].

Human renin-angiotensin system (RAS) plays a role in the regulation of blood pressure, and angiotensinogen and angiotensin-converting enzyme (ACE) are a part of the RAS. Several studies showed the association of ACE (angiotensin I converting enzyme (peptidyl-dipeptidase A) 1) variants with AD [58,59] as well as type 2 diabetic nephropathy [60], and cerebral amyloid angiopathy-related lobar intracerebral hemorrhage recurrence [61]. In our study, rs4343 showed the strongest effect (R^2 angiopathy-related lobar intracerebral hemorrhage recurrence well as type 2 diabetic nephropathy [60], and cerebral amyloid glycoprotein), synthesized in liver and secreted into the blood stream. Plasma Fetuin-A level has been associated with cardio-vascular disease [66] and AHSG variants have been previously associated with AD [67]. The previous [66] studies identified associations of the same SNPs (rs4917, rs2070633) with plasma Fetuin-A level and in the same direction of effect as was observed in the present study.

Fetuin-A is a serum protein, encoded by AHSG (alpha-2-HS-glycoprotein), synthesized in liver and secreted into the blood stream. Plasma Fetuin-A level has been associated with cardiovascular disease [66] and AHSG variants have been previously associated with AD [67]. The present [66] studies identified associations of the same SNPs (rs4917, rs2070633) with plasma Fetuin-A level and in the same direction of effect as was observed in the present study.

Tamm-Horsfall glycoprotein (THP) is abundant in urine, and in humans it is encoded by the UMOD (uromodulin) gene, which is associated with chronic kidney disease [68,69] and blood pressure [70]. We identified four SNPs (rs11647727, rs4506906, rs4293393, rs13333226) associated with the plasma THP level in the ADNI cohort although they were not replicated in the IMAS cohort. Among these SNPs, rs13333226 has been previously associated with diastolic blood pressure [70]. The strongest SNP effect in our study (rs4293393) has also been associated with urinary THP concentrations in the same direction of the observed effect [69].

Thyroxine-binding globulin (TBG) is a protein that is involved in the transport of thyroxine and triiodothyronine in human serum [71]. Previous studies [72,73] investigated the role of polymorphisms within SERPINA1 (serpin peptidase inhibitor, clade A (alpha-1 antitrypsin), member 7) gene in relation to inherited TBG defects. We found rs1804495 to be associated with plasma TBG level but only in males in the ADNI sample. This polymorphism is in codon 303 replacing TTG (leucine) with TTT (phenylalanine) and the role of this variant has not been previously reported on TBG defects or plasma level of TBG.

The present study has some limitations that may be informative for future studies. First, both cohorts consisted of older adults including a large portion with MCI, AD or cognitive complaints. Although age, diagnosis and APOE genotype were included as covariates, we could not definitively determine the extent to which age and AD risk may have influenced the observed associations. Studies of gene-protein associations in younger and cognitively healthy samples are needed to clarify the generalizability of the present results. Second, although this study included relevant covariates, other factors than those measured and selected for analysis may have influenced the associations we studied. Further investigation of other factors influencing protein levels beyond genetic variation and the current covariates may be important [5]. Third, non-normal distributions may have influenced association statistics. However, this is relatively unlikely because our analyses did not indicate any significant evidence of statistical bias. Fourth, the IMAS replication sample was of modest size, resulting in limited detection power compared to the ADNI cohort. Additional studies with larger sample sizes are needed for confirmation of the observed relationships. Fifth, the genotyping microarray we used shows considerable variation in SNP coverage for the genes of interest, as illustrated in Figure S1. Therefore, some potential influence of genetic variants on protein analyte levels may have been missed due to undersampling of targeted genomic regions. Imputation of SNP data using HapMap or 1000 reference panel can increase the coverage and will be used in the future study. Finally, there might be technical issues with RBM between the discovery and the replication data which were assayed at different times with different antibodies and conditions used in different RBM runs. The technical issues related to assay time/batch differences could have played roles in those that were not replicated and this is also an issue for future validation of candidate analytes. Considerable amount of work to resolve these and other technical issues inherent to the RBM and follow up assays will be required to evaluate the current findings and turn them into research or clinical grade diagnostic assays in the future.

Despite these limitations, the current study identified 112 SNP-protein associations in the ADNI cohort and many (n = 80) of these associations were highly significant relative to generally accepted significance thresholds (<5×10^-6). Approximately half of the 112 SNP-protein associations identified in the ADNI cohort were replicated in the IMAS cohort. However, some findings in the ADNI cohort which were not replicated in the IMAS cohort were previously reported in other studies and therefore continue to warrant additional investigation.

In conclusion, this study investigated the role of genetic variation, specifically cis-effects, on corresponding protein levels. The strong influence of many genes on commonly measured plasma analytes should be considered. This is particularly critical when proteins are known to play an important role in a disease or treatment. In this case, the evaluation of proteins as diagnostic, prognostic or therapeutic response biomarkers may need to be stratified for genetic background. Future studies should examine diagnostic classification after stratification. Our findings should be replicated in additional independent cohorts with larger samples. It is anticipated that future studies will investigate other genetic mechanisms such as trans-effects, haplotypes, copy number variation and epistasis, each of which may influence plasma protein levels. Finally, mRNA sequencing and transcriptome analyses of expression and alternative splicing should provide a more complete picture of functional genetic variations, influencing plasma-gene products.

**Supporting Information**

**Figure S1** Zoomed view of association results. All association results between 28 gene-protein pairs from analyses were shown using LocusZoom (http://cgp.sph.umich.edu/locuszoom/). In each panel, hg18 and HapMap Phase II CEU were used as Genome build and LD population. The panel for CompFactH shows the association of SNPs within CFIH and CEHRII together. Just for visualization purpose, the genetic mode tested for the most significant protein-SNP pair for each panel was selected and “DOM” represents a dominant genetic model. (TIF)

**Figure S2** Scatter plots of top 12 associations in Table 3 for the ADNI cohort. Different colors represent different diagnoses (black - all, red - AD, green - MCI, blue - NC) and
horizontal bars are the average protein levels within each group.

SNPs (R^2_{SNP}) while adjusting for the effect of covariates is shown. R^2_{SNP} is defined as follows: R^2_{SNP} = adjusted R^2 of model with SNP and covariates – adjusted R^2 of model with covariates.

Table S4 List of all replicated (uncorrected p<0.05) associations in the IMAS cohort. The associations are sorted by chromosome and chromosomal location of each SNP. Percent of total variation explained by individual SNPs (R^2_{SNP}) while adjusting for the effect of covariates is shown. R^2_{SNP} is defined as follows: R^2_{SNP} = adjusted R^2 of model with SNP and covariates – adjusted R^2 of model with covariates.

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. We also thank the following people: genotyping for the ADNI sample at the Translational Genomics Research Institute; Drs. Matthew Huettelman and David Craig for the ADNI sample and sample processing, storage and distribution for the ADNI and IMAS samples at the NCRAD; Kelley Faber and Colleen Mitchell, data collection, review and preprocessing for the IMAS sample; Tamiko R. Magee, Yang Wang, Vijay K. Ramanan, and John D. West.

Author Contributions

Conceived and designed the experiments: SK AJS MI. Performed the experiments: SK MI SS SLR KN LS. Analyzed the data: SK SS MI AJS. Contributed reagents/materials/analysis tools: SLR KN TF RCP PSA HS JBT LMS JQ T MWW BCM MRJ BG. Wrote the paper: SK SS MI SLR JQ T AJS.

References

1. Boot RG, Verhoek M, de Fost M, Hollak CE, Maas M, et al. (2004) Marked elevation of the chondrocyte CCL18/PARC in Gaucher disease: a novel surrogate marker for assessing therapeutic intervention. Blood 103: 33-39.
2. Chen-Potchkin AS, Hu WT, Sidebottom A, Weintraub D, Goldmann Gross R, et al. (2011) Plasma epidermal growth factor levels predict cognitive decline in Parkinson disease. Ann Neurol 69: 655-663.
3. Hu WT, Chen-Potchkin A, Arnold SE, Grossman M, Clark CM, et al. (2010) Biomarker discovery for Alzheimer’s disease, frontotemporal lobar degeneration, and Parkinson’s disease. Acta Neuropathol 120: 385-399.
4. Furlan SE, Xiong G, Barber R, Reisch J, Doody R, et al. (2010) A serum protein-based algorism for the detection of Alzheimer disease. Arch Neurol 67: 1077-1081.
5. Toledo JB, Vanderstichele H, Figurski M, Aisen PS, Petersen RC, et al. (2011) Factors affecting Abeta plasma levels and their utility as biomarkers in ADNI. Acta Neuropathol.
6. Hu WT, Holtzman DM, Fagan AM, Shaw LM, Perrin R, et al. (2012) Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease. Neurology 79: 897-905.
7. Johnstone D, Milward EA, Berretta R, Moscati P (2012) Multivariant protein signatures of pre-clinical Alzheimer’s disease in the Alzheimer’s disease neuroimaging initiative (ADNI) plasma proteome dataset. PLoS One 7: e34341.
8. Soares HD, Potter WZ, Pickering E, Kuhn M, Innemann FW, et al. (2012) Plasma Biomarkers Associated With the Apolipoprotein E Genotype and Alzheimer’s Disease. Arch Neurol. 1-8.
9. Garge N, Pan H, Rowland MD, Cargle BJ, Zhang X, et al. (2010) Identification of quantitative trait loci underlying prostate variation in human plasma. Hum Mol Genet.
10. Moller A, Perry JR, Hernandez D, Corsi AM, Stevens K, et al. (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet 4: e100072.
11. Friedman DB, Lillie KS (2000) Optimizing the difference gel electrophoresis (DIGE) technology. Methods Mol Biol 428: 93-124.
12. Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. (2010) Aptamer-based multiplexed protein technology for biomarker discovery. PLoS One 5: e15404.
13. Jack CR, Jr., Bernstein MA, Fox NC, Thompson P, Alexander G, et al. (2008) The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. J Mag Reson Imaging 27: 685-691.
14. Jagust WJ, Bandy D, Chen K, Foster NL, Landau SM, et al. (2010) The Alzheimer’s Disease Neuroimaging Initiative positron emission tomography core. Alzheimers Dement 6: 221-229.
15. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, et al. (2010) Alzheimer’s Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 74: 209-219.
16. Shaw LM, Vanderstichele H, Knopik-Craka M, Clark CM, Aisen PS, et al. (2009) Cerebrospinal fluid biomarker signature in Alzheimer’s disease neuroimaging initiative subjects. Ann Neurol 65: 403-413.
17. Saykin AJ, Shao L, Foroud TM, Potkin SG, Swaminathan S, et al. (2010) Alzheimer’s Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement 6: 265-273.
18. Saykin AJ, Wishart HA, Rabin LA, Santulli RB, Flashman LA, et al. (2006) Older adults with cognitive complaints show brain atrophy similar to that of amnestic MCI. Neurology 67: 834-842.
19. Risacher SL, Wudunn D, Pepin SM, Magee TR, McDonald BC, et al. (2012) Visual contrast sensitivity in Alzheimer’s disease, mild cognitive impairment, and older adults with cognitive complaints. Neurobiology of aging. 265-273.
20. Risacher SL, Wudunn D, Pepin SM, Magee TR, McDonald BC, et al. (2012) Visual contrast sensitivity in Alzheimer’s disease, mild cognitive impairment, and older adults with cognitive complaints. Neurobiology of aging.
21. Potkin SG, Guflanti L, Lakatos A, Turner JA, Krugell F, et al. (2009) Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer’s disease. PLoS One 4: e6501.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.
23. Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, et al. (2010) Whole Alzheimer’s Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement 6: 265-273.
24. (2003) The International HapMap Project. Nature 426: 789–796.
25. Chelala C, Khan A, Lemoine NR (2009) SNFplexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics 25: 653-661.
26. Saykin AJ, Lemoine NR, Chelala C (2012) SNFplexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic acids research 40: W65-70.
27. Huang CW, Lai CC, Chang WN, Lu CH, Wang YL, et al. (2009) Elevated basal cortisol levels predict lower hippocampal volume and cognitive decline in Alzheimer’s disease. J Clin Neurosci 16: 1203-1206.
