nature medicine

Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins

Sandip Ray[1,16], Markus Britschgi[2,16], Charles Herbert[1], Yoshiko Takeda-Uchimura[2], Adam Boxer[3], Kaj Blennow[4], Leah F Friedman[5], Douglas R Galasko[6], Marek Jutel[7], Anna Karydas[3], Jeffrey A Kaye[8], Jerzy Leszek[9], Bruce L Miller[3], Lennart Minthon[10], Joseph F Quinn[8], Gil D Rabinovici[3], William H Robinson[11], Marwan N Sabbagh[12], Yuen T So[2], D Larry Sparks[12], Massimo Tabaton[13], Jared Tinklenberg[5], Jerome A Yesavage[5], Robert Tibshirani[14] & Tony Wyss-Coray[2,15]

A molecular test for Alzheimer's disease could lead to better treatment and therapies. We found 18 signaling proteins in blood plasma that can be used to classify blinded samples from Alzheimer's and control subjects with close to 90% accuracy and to identify patients who had mild cognitive impairment that progressed to Alzheimer's disease 2–6 years later. Biological analysis of the 18 proteins points to systemic dysregulation of hematopoiesis, immune responses, apoptosis and neuronal support in presymptomatic Alzheimer's disease.

Alzheimer's disease results in a progressive loss of cognitive function and dementia affecting one in eight people by the time they reach 65 years of age[1]. Diagnosis of Alzheimer's disease is time consuming and requires a combination of psychological testing, imaging and exclusion of other neurological disorders. Patients with presymptomatic Alzheimer's or mild cognitive impairment (MCI) have a greatly increased risk of developing Alzheimer's disease[2]. It is estimated that by the time the typical patient is diagnosed with Alzheimer's, the disease has been progressing for many years, so it is crucial that the disease is detected as early as possible. In light of these facts, a molecular biomarker in blood plasma that could classify Alzheimer's disease and identify those presymptomatic individuals with MCI who will eventually convert to Alzheimer's would be particularly useful. Because the brain controls many body functions via the release of signaling proteins, and because central and peripheral immune and inflammatory mechanisms are increasingly implicated in Alzheimer's[3] and related diseases[4], we hypothesized that the pathological processes leading to Alzheimer's would cause characteristic changes in the concentrations of signaling proteins in the blood, generating a detectable disease-specific molecular phenotype.

We collected a total of 259 archived plasma samples from individuals with presymptomatic to late-stage Alzheimer's disease and from various controls (**Supplementary Table 1** online) and measured the abundance of 120 known signaling proteins (**Supplementary Table 2** online) in these samples with filter-based, arrayed sandwich ELISAs[5] (**Supplementary Fig. 1** online). The Alzheimer's and nondemented control (NDC) samples were divided equally into a training set for predictor discovery and supervised classification and a test set for class prediction of blinded samples (**Supplementary Table 1**). Initial statistical analysis of the training set by significance analysis of microarrays (SAM, **Fig. 1a**)[6] identified 19 proteins with highly significant differences in expression ($q < 3.4\%$) between Alzheimer's and NDC samples (**Supplementary Table 2**). We arranged the training set samples by the similarity in abundance of the 19 markers in the blood with an unsupervised clustering algorithm (**Fig. 1a**), which produced two main clusters that contained mostly Alzheimer's or NDC samples, respectively (**Fig. 1b**). These results show that the plasma concentrations of many secreted signaling proteins differ considerably between subjects with Alzheimer's disease and NDC subjects, and that a protein expression pattern distinct from that in NDC subjects is associated with Alzheimer's disease.

[1]Satoris, Inc., 2686 Middlefield Road, Suite E, Redwood City, California 94063, USA. [2]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305-5235, USA. [3]Department of Neurology, Memory and Aging Center, 350 Parnassus Avenue, Suite 706, San Francisco, California 94117, USA. [4]Institute of Clinical Neuroscience, Department of Experimental Neuroscience, Sahlgrenska University Hospital, University of Göteborg, Blå stråket 15, 431 80 Mölndal, Sweden. [5]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305-5550, USA. [6]Department of Neurosciences, University of California, San Diego, 9500 Gilman Drive #9127, La Jolla, California 92093-9127, USA. [7]Department of Internal Medicine and Allergology, Wroclaw Medical University, Traugutta 57, 50-417 Wroclaw, Poland. [8]Layton Aging & Alzheimer's Disease Center, Oregon Health Sciences University, 3181 Southwest Sam Jackson Park Road, CR131, Portland, Oregon 97201-3098, USA. [9]Department of Psychiatry, Wroclaw Medical University, Pasteura 10, 51-622 Wroclaw, Poland. [10]Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Universitetssjukhuset MAS, Ingång 56 plan 7, SE-205 02 Malmö, Sweden. [11]Division of Immunology and Rheumatology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305-5166, USA. [12]Sun Health Research Institute, 10515 West Santa Fe Drive, Sun City, Arizona 85351, USA. [13]Department of Neurosciences, Ophthalmology, and Genetics, University of Genoa, Via A. De Toni 5, 16132 Genoa, Italy. [14]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, 300 Pasteur Drive, California 94305-5405, USA. [15]Geriatric Research, Education and Clinical Center, Veterans Affairs Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, California 94304, USA. [16]These authors contributed equally to this work. Correspondence should be addressed to T.W.-C. (twc@stanford.edu).

Received 2 May; accepted 30 August; published online 14 October 2007; doi:10.1038/nm1653

To find an Alzheimer's-specific signature, we analyzed the above training set with a shrunken centroid algorithm called predictive analysis of microarrays (PAM, **Fig. 1a**)[7]. PAM identified 18 predictors out of the 120 proteins (**Fig. 1c** and **Table 1**) and classified Alzheimer's and NDC samples with 95% positive agreement and 83% negative agreement with the clinical diagnosis, respectively (**Fig. 2a**).

To assess the performance of the 18 predictors in the classification of unknown samples, we carried out a prediction for 'Alzheimer's' or 'non-Alzheimer's' phenotype (a so-called 'two-class' prediction) in a test set containing Alzheimer's and NDC samples as well as samples from individuals with other dementias (**Fig. 1a** and **Supplementary Table 1**). PAM classified them with 90% positive agreement (for the Alzheimer's samples) and 88% negative agreement (for the non-Alzheimer's samples) with the clinical diagnosis (**Fig. 2b**). Eight of nine postmortem-confirmed subjects with Alzheimer's disease were classified correctly and 10 of the 11 other dementia samples received a 'non-Alzheimer's' classification (data not shown). Consistent with these findings, unsupervised clustering based on the 18 markers was able to separate, according to their diagnosis, Alzheimer's and NDC samples that had been combined from the training and test sets

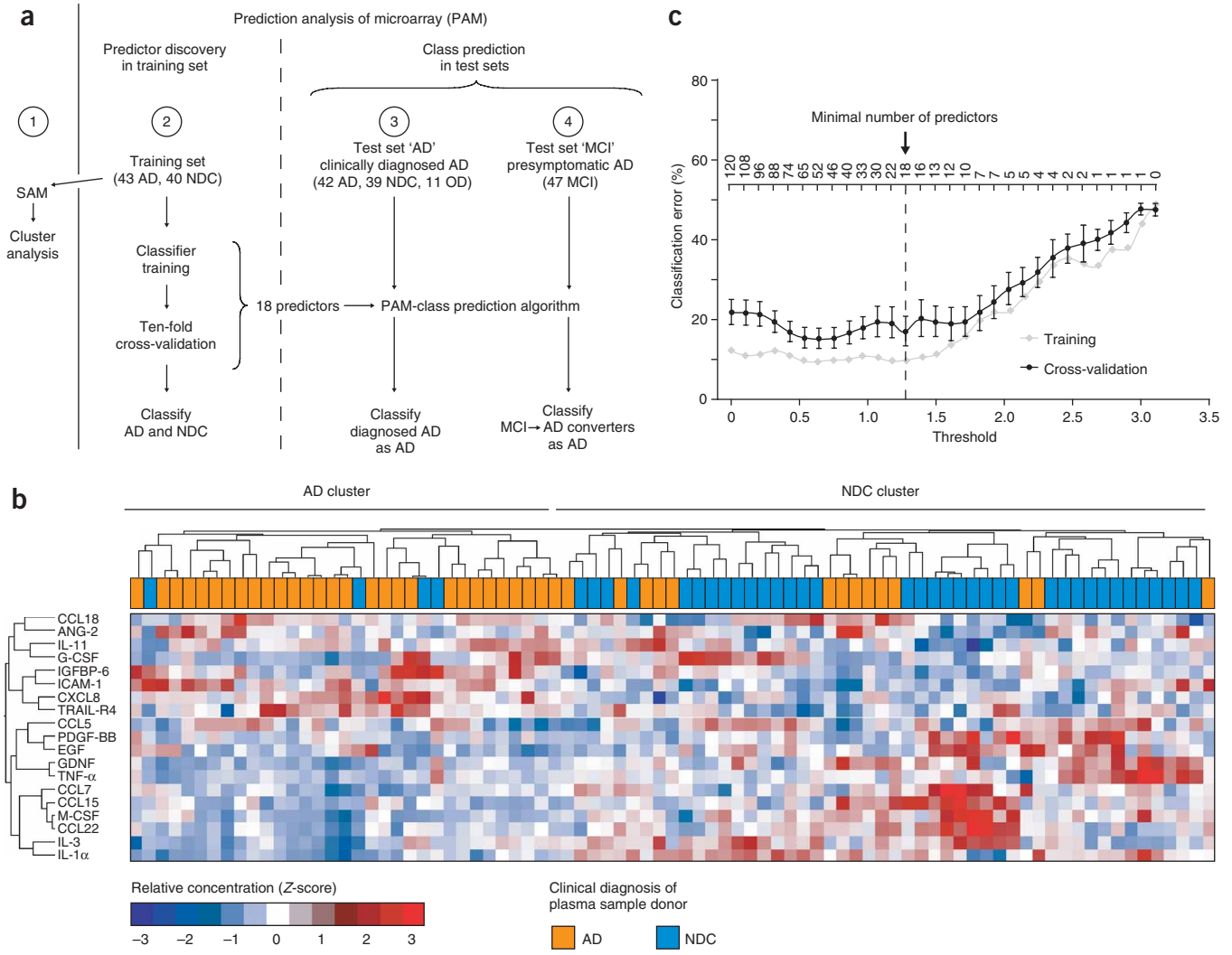

Figure 1 Study outline, clustering of training set, and predictor discovery. (**a**) Informed consent was obtained from all human subjects according to the ethics committee guidelines at the respective academic centers. A total of 223 plasma samples were separated into a training set and two test sets as indicated (**Supplementary Table 1**). Changes in relative signaling protein concentrations were initially analyzed with SAM, and this was followed by cluster analysis (1). To discover predictors for classification, we analyzed the training set by PAM (2), then used the PAM predictors to classify the samples in the independent test set 'AD' (3). Class prediction of presymptomatic Alzheimer's disease (AD) was performed on samples from individuals who were diagnosed with MCI at the date of blood draw (test set 'MCI', 4). None of the samples from the test sets was used for any part of the predictor discovery process. (**b**) We analyzed normalized array measurements of 120 plasma signaling proteins in the training set with SAM to discover differences in protein abundance between samples of subjects with Alzheimer's disease and those of NDC subjects. The nineteen proteins that obtained a significant *d*-score ($q \leq 3.4\%$) are presented in a 'heat map' generated with an unsupervised cluster algorithm. Samples are arranged in columns, proteins in rows. Red shades, increased expression in Alzheimer's disease samples as compared to NDC samples; blue shades, reduced expression; white, median expression. Samples are clustered into AD and NDC categories as indicated by the first-order branches of the dendrogram (two black bars at the top). (**c**) Predictor discovery by PAM was performed with normalized array measurements of 120 signaling proteins in the training set. In training (gray line) and internal cross-validation (black line), decreasing the centroid threshold (lower *x*-axis) resulted in an increase in the number of markers (inserted upper *x*-axis) that were used for classification and calculation of the classification error (*y*-axis). This led to the discovery of a minimal set of 18 predictors with lowest possible classification error.

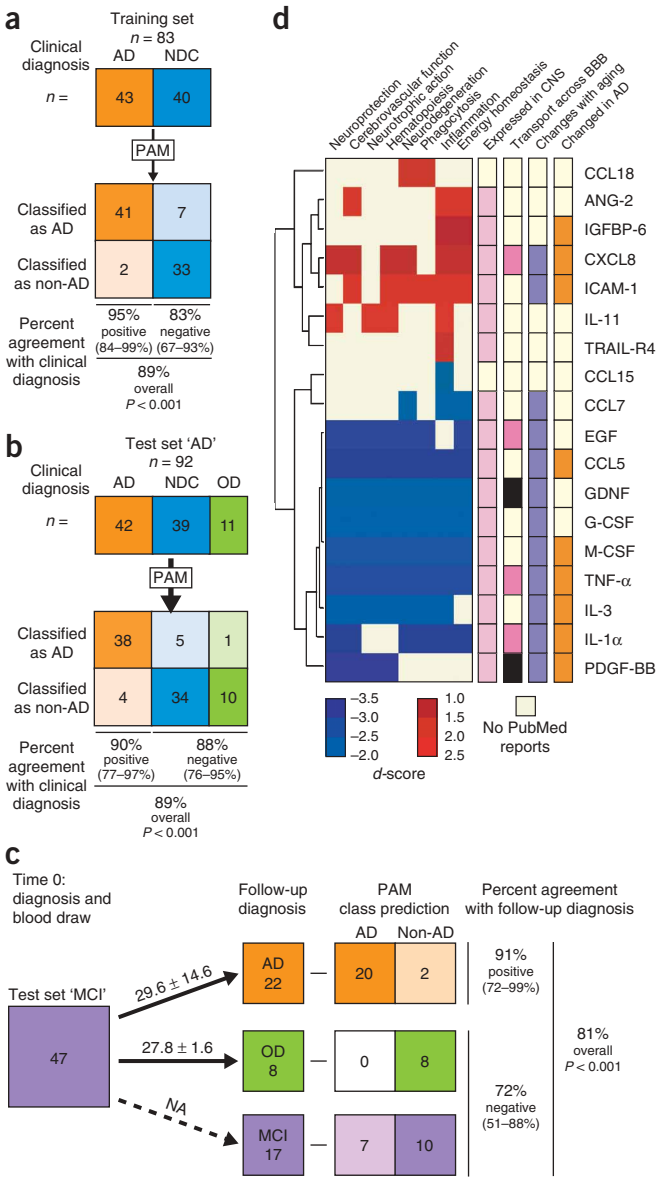

Figure 2 Classification and prediction of clinical Alzheimer's diagnosis in subjects with Alzheimer's disease or MCI and functional analysis of the 18 predictive plasma signaling proteins. (**a**–**c**) The 18 predictors identified with PAM were used for Alzheimer's (AD) and non-Alzheimer's class prediction in the training set (**a**), the blinded test set 'AD' (**b**) and the test set 'MCI' (**c**). Results are shown in modified 2 × 2 contingency tables that were used to calculate the percentage of classifications that agreed with clinical diagnosis. Values in parentheses are 95% confidence intervals and *P*-values were calculated with Fisher's exact test. To calculate negative agreement in the blinded test set (**b**), NDCs and other dementias (ODs) were combined into one group. After an initial diagnosis at blood draw (time 0) 47 subjects with MCI (**c**), who were followed longitudinally, converted to Alzheimer's disease, developed other dementias or remained MCI (follow-up diagnosis; arrow indicates average conversion time in months ± s.d.; NA, not applicable). To calculate the percentage of predictions that agreed with the follow-up diagnosis, we combined NDC and OD classification data into one group, that of patients who did not convert to Alzheimer's disease. (**d**) Result of the PubMed query for additional functional annotations and biological processes of the 18 signaling proteins. Node map lists entries in PubMed reporting that the specific factor modulates the indicated biological function or is regulated by it. Colors indicate *d*-scores as calculated by SAM, representing greater (shades of red) or less (shades of blue) expression in subjects with Alzheimer's disease than in NDC subjects. Production of most of the 18 signaling proteins in the CNS has been reported (light pink), and some have been found to be transported across the blood-brain barrier (BBB; pink) or not (black) in rodents. Additionally, expression levels of several predictors are changed in aging (purple) or Alzheimer's disease (orange). For reports on expression changes in Alzheimer's, see **Supplementary Table 3**.

(**Supplementary Fig. 2a** online). Similarly, unsupervised clustering based on the 18 predictive signaling proteins led to a good separation of all Alzheimer's samples from the plasma samples of individuals with other neurological diseases or with rheumatoid arthritis (**Supplementary Fig. 2b** and **Supplementary Table 1**).

The use of biomarkers to predict development of Alzheimer's disease among MCI individuals would yield substantial therapeutic and health-economic benefits. We therefore analyzed plasma samples from two previously published cohorts of MCI patients who were followed longitudinally and who converted to Alzheimer's, developed other dementias or remained MCI (**Supplementary Table 1**)[8,9]. The plasma samples were collected at the initial diagnosis of MCI (time 0), and patients obtained a final follow-up diagnosis for this study after 2–6 years. After application of the 18 predictors to the MCI test set (**Fig. 1a**), PAM classified 20 of 22 MCI patients who developed Alzheimer's disease 2–5 years later as 'Alzheimer's' (91% positive agreement with the clinical diagnosis, **Fig. 2c**). All eight MCI patients who later developed other dementias were correctly classified as 'non-Alzheimer's' (**Fig. 2c**). Of the 17 MCI patients who were still diagnosed as MCI 4–6 years after blood draw, 7 were classified as 'Alzheimer's' and 10 were classified as 'non-Alzheimer's' (**Fig. 2c**). Our data indicate that a highly specific plasma biomarker phenotype can characterize Alzheimer's disease years before a clinical diagnosis can be made.

To understand the potential biological relevance of the 18 signaling proteins that characterize Alzheimer's disease, we used several functional annotation tools and also searched PubMed manually. The computational gene network prediction tool Ingenuity Pathway Analysis (Ingenuity Systems) identified two independent regulatory networks connecting the 18 signaling proteins (**Supplementary Fig. 3** online). One network centered on tumor necrosis factor (TNF)-α and monocyte-colony stimulating factor (M-CSF), whereas the other centered on epidermal growth factor (EGF). Consistent with these findings, gene ontology (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/) and BioCarta (http://www.biocarta.com/) pathway analyses indicated involvement of the 18 markers in immune response, hematopoiesis and apoptosis (**Supplementary Fig. 4** online). The overall effect of up- or downregulation of the observed signaling proteins in the Kyoto Encyclopedia of Genes and Genomes and BioCarta pathways predicts a negative impact on the majority of the pathways (**Supplementary Fig. 4**). A decrease in the abundance of factors linked to hematopoiesis would be particularly noteworthy in light of recent data suggesting that hematopoietic cells can enter the brain in Alzheimer's disease or in Alzheimer's mouse models at increased frequencies and modulate the disease[3,10,11]. Dysfunction of apoptotic pathways has also been linked to Alzheimer's disease[12].

To further investigate the biological relevance of the 18 predictors for Alzheimer's disease, we queried PubMed (**Fig. 2d**). This analysis pointed again to an overall reduction in the abundance of factors associated with hematopoiesis and inflammation during Alzheimer's disease, as well as to deficits in neuroprotection, neurotrophic activity,

Table 1 Eighteen plasma signaling proteins that predict clinical Alzheimer's diagnosis

Predictors	*d*-score	*q*-value (%)
ANG-2	2.1	≤0.05
CCL5	–2.9	≤0.05
CCL7	–1.7	≤0.05
CCL15	–1.6	≤0.05
CCL18	1.9	3.1
CXCL8	1.7	3.1
EGF	–2.7	≤0.05
G-CSF	–1.9	≤0.05
GDNF	–1.8	≤0.05
ICAM-1	2.2	≤0.05
IGFBP-6	1.5	3.1
IL-1α	–2.9	≤0.05
IL-3	–2.0	≤0.05
IL-11	2.1	≤0.05
M-CSF	–2.4	≤0.05
PDGF-BB	–3.4	≤0.05
TNF-α	–2.6	≤0.05
TRAIL-R4	1.8	3.1

In the training set, predictor discovery by PAM identified 18 predictors from the normalized array measurements of 120 signaling proteins. SAM was used to calculate *d*-scores indicating the relative positive (increased) and negative (decreased) changes in concentration of these proteins in plasma of subjects with Alzheimer's disease in comparison to NDC subjects. SAM calculates a minimal false discovery rate (*q*-value) for significance. ANG-2, angiopoietin-2. CCL, chemokine that contains a C-C motif; CXCL, chemokine that contains a C-X-C motif; G-CSF, granulocyte-colony stimulating factor; GDNF, glial-derived neurotrophic factor; ICAM-1, intercellular adhesion molecule-1; IGFBP-1, insulin-like growth factor–binding protein-6; IL, interleukin; PDGF-BB, platelet-derived growth factor BB; TRAIL-R4, TNF-related apoptosis-inducing ligand receptor-4.

phagocytosis and energy homeostasis. Whereas previously reported changes in plasma and cerebrospinal fluid during Alzheimer disease matched many of our findings, there was less overlap between measurements of plasma and brain parenchyma (**Supplementary Table 3** online). Notably, however, an extensive hippocampal gene array analysis reported that most of the pathways identified in this study are abnormal in Alzheimer's disease as well[13].

The observed dysregulation of the signaling pathways represented by the 18 signaling proteins in blood plasma may point to changes in the periphery, the central nervous system or both that are relatively specific to Alzheimer's disease and occur early in the disease process. In support of peripheral manifestations of neurodegeneration, differential gene expression patterns in blood cells can predict early Parkinson's disease[14] and possibly Alzheimer's disease, as well[15]. Other studies have reported differences in the distribution of leukocyte subsets in blood or differential cytokine secretion from blood cells in individuals with MCI or Alzheimer's disease[3,11].

By focusing on signaling proteins or intercellular communication factors rather than on the entire plasma proteome, we were able to identify an Alzheimer's biomarker phenotype that can potentially be used for the diagnosis of early Alzheimer's disease. On the basis of this approach, we propose that the 'cellular communicome', which encompasses those proteins within an organism that carry information from one cell to another, may be an attractive target for unbiased screens in disease research. Similar signatures to the one described here for Alzheimer's disease may exist for other CNS diseases, and may hold potential clues for both treatment and diagnosis.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

We are grateful to the individuals who participated in this study. We also thank W. Hueber for critical reading and helpful comments on the manuscript, and numerous unnamed staff of our institutions for their efforts in subject recruitment, diagnostic assessment and blood processing. This study was supported by the John Douglas French Alzheimer's Foundation (T.W.-C.), the Alzheimer's Association (T.W.-C.), the US National Institute on Aging (T.W.-C., AG20603; J.A.K., J.F.Q., AG08017; D.R.G., AG05131) and Satoris, Inc. We also acknowledge the support of the Veterans Administration Mental Illness Research, Education and Clinical Center and the various Alzheimer's Centers sponsored by the US National Institute on Aging.

AUTHOR CONTRIBUTIONS

Experiments were coordinated by S.R., M.B. and T.W.-C. Filter array experiment was done by S.R. with the help of C.H. Cytokine antibody array experiments, cluster analysis and class prediction were done by S.R. with scientific advice from R.T. Computational analysis of functional annotations was done by M.B. and T.W.-C. Blood processing, sample preparation, ELISA, subject data collection and administration was done by M.B. with the help of Y.T.-U. Recruitment of patients and control individuals, disease assessment and blood processing in center-coordinated studies was directed or done by A.B., K.B., L.F.F. D.R.G., M.J., A.K., J.A.K, J.L., B.L.M., L.M., J.F.Q., G.D.R., W.H.R., M.N.S., Y.T.S., D.L.S., M.T., J.T. and J.A.Y. The project was conceived by S.R. and T.W.-C. and scientifically directed by T.W.-C., and the paper written by M.B., T.W.-C. and S.R.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemedicine/.

Published online at http://www.nature.com/naturemedicine
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Blennow, K., de Leon, M.J. & Zetterberg, H. *Lancet* **368**, 387–403 (2006).
2. Storandt, M., Grant, E.A., Miller, J.P. & Morris, J.C. *Neurology* **59**, 1034–1041 (2002).
3. Wyss-Coray, T. *Nat. Med.* **12**, 1005–1015 (2006).
4. Steinman, L. *Nat. Immunol.* **5**, 575–581 (2004).
5. Huang, R.P. *Methods Mol. Biol.* **264**, 215–231 (2004).
6. Tusher, V.G., Tibshirani, R. & Chu, G. *Proc. Natl. Acad. Sci. USA* **98**, 5116–5121 (2001).
7. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. *Proc. Natl. Acad. Sci. USA* **99**, 6567–6572 (2002).
8. Hansson, O. *et al. Lancet Neurol.* **5**, 228–234 (2006).
9. Assini, A. *et al. Neurology* **63**, 828–831 (2004).
10. Simard, A.R., Soulet, D., Gowing, G., Julien, J.P. & Rivest, S. *Neuron* **49**, 489–502 (2006).
11. Britschgi, M. & Wyss-Coray, T. *Int. Rev. Neurobiol.* **82**, 205–233 (2007).
12. LeBlanc, A.C. *Curr. Alzheimer's Res.* **2**, 389–402 (2005).
13. Blalock, E.M. *et al. Proc. Natl. Acad. Sci. USA* **101**, 2173–2178 (2004).
14. Scherzer, C.R. *et al. Proc. Natl. Acad. Sci. USA* **104**, 955–960 (2007).
15. Maes, O.C. *et al. Neurobiol. Aging*, published online 18 September 2006 (doi:10.1016/j.neurobiolaging.2006.08.004).