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

Sandip Ray*, Markus Britschgi*, Charles Herbert, Yoshiko Takeda-Uchimura, Adam Boxer, Kaj Blennow, Leah F. Friedman, Douglas R. Galasko, Marek Jutel, Anna Karydas, Jeffrey A. Kaye, Jerzy Leszek, Bruce L. Miller, Lennart Minthon, Joseph F. Quinn, Gil D. Rabinovici, William H. Robinson, Marwan N. Sabbagh, Yuen T. So, D. Larry Sparks, Massimo Tabaton, Jared Tinklenberg, Jerome A. Yesavage, Robert Tibshirani, and Tony Wyss-Coray

* These authors contributed equally to this work.

Supplementary Information

Methods

Plasma samples. We collected a total of 259 archived plasma samples with ethylene diamine tetra acetate (EDTA) as anticoagulant from academic centers specialized in neurological or neurodegenerative diseases (**Supplementary Table S1** online). Plasma was produced by standard blood processing, then frozen and stored in aliquots at –80°C. Informed consent was obtained from human subjects according to the ethics committee guidelines at the respective academic centers.

Supplementary Table S1. Subjects' characteristics

Clinical diagnosis	Number	Age (mean ± SD)	Sex (% female)	MMSE[a] (mean ± SD)	Center[b]
Alzheimer disease (AD)	85	76.2 ± 7.8[c]	47	16.3 ± 7.9	A,O,D,G,K,W,F
Non-demented controls (NDC)	79	71.4 ± 7.8[c]	40	29.4 ± 1.0	A,O,D,S,W,F
Training set					
AD	*43*	*74.3 ± 8.8*	*44*	*16.2 ± 8.0*	A,O,D,G,K,W,F
NDC	*40*	*72.3 ± 7.8*	*35*	*29.3 ± 0.7*	A,O,D,S,W,F
Test set					
AD	*42*	*78.2 ± 6.4*	*33*	*17.1 ± 7.5*	A,O,D,G,K,W,F
NDC	*39*	*70.7 ± 7.7*	*45*	*29.4 ± 1.1*	A,O,D,S,W,F
Other dementia (OD)	11				
Frontotemporal dementia (FTD)	8	64.5 ± 9.1	33	19.7 ± 9.8	K,F
Corticobasal degeneration (CBD)	3	60.0 ± 5.3	66	17.0 ± 6.6	K
Mild cognitive impairment (MCI)	47	71.3 ± 8.1	55	27.3 ± 1.9	
MCI→AD[d]	22	75.3 ± 5.1	61	27.4 ± 1.8	K,G
MCI→OD[d]	8	75.2 ± 8.2	13	27.0 ± 1.6	K,G
MCI→MCI[d]	17	65.8 ± 8.3	47	27.6 ± 2.0	K,G
Other neurological disease (OND)	21				
Parkinson disease	5	79.8 ± 3.1	0	n.a.	S
ALS	2	47.5 ± 12.0	0	n.a.	S
Multiple sclerosis	2	55 ± 0.0	100	n.a.	W
Peripheral neuropathy	12	69.5 ± 8.2	41	n.a.	S
Rheumatoid arthritis	16	63.9 ± 12.0	6	n.a.	S

[a] Mini-mental state exam [1].
[b] A, Sun Health Research Institute; O, Oregon Health Sciences University; D, UC San Diego; G, University of Genoa; K, Göteborg University; W, University of Wroclaw; F, UC San Francisco; S, Stanford University;
[c] Age difference between AD and NDC is not significant ($P = 0.25$, Student's t-test).
[d] Out of 47 subjects diagnosed with MCI at blood draw ("Time point 0" in **Fig. 2c**), 22 converted to Alzheimer disease within 2-5 years (MCI→AD; average conversion time 29.6 ± 14.6 months), 8 converted to other dementias (MCI→OD; average conversion time 27.8 ± 1.6 months), whereas 17 were still diagnosed MCI 4-6 years later (MCI→MCI).
n.a., not available.

Antibody arrays. We measured the relative concentrations of a total of 120 proteins (**Supplementary Table S2** online) with cytokine antibody arrays (Raybiotech Inc., www.raybiotech.com) according to the manufacturer's instructions[2]. Briefly, for each plasma sample two nitrocellulose membranes, each containing 60 different antibodies in duplicate spots, were blocked, incubated with plasma, washed, and then incubated with a cocktail of biotin-conjugated antibodies specific for the different proteins.

Supplementary Table S2. Proteins measured with cytokine antibody array

Protein Name	Entrez GeneID	Official gene name provided by HUGO Gene Nomenclature Committee (HGNC)	SAM19 AD/NDC [a]	PAM18 Predictors [c]
Adiponectin	9370	adiponectin, C1Q and collagen domain containing		
AGRP	181	agouti related protein homolog (mouse)		
Amphiregulin	374	amphiregulin (schwannoma-derived growth factor)		
ANG-2	285	angiopoietin 2	x	x
Angiogenin	283	angiogenin, ribonuclease, RNase A family, 5		
AXL	558	AXL receptor tyrosine kinase		
basic FGF	2247	fibroblast growth factor 2 (basic)		
BDNF	627	brain-derived neurotrophic factor		
BMP-4	652	bone morphogenetic protein 4		
BMP-6	654	bone morphogenetic protein 6		
BTC	685	betacellulin		
CCL1/I-309	6346	chemokine (C-C motif) ligand 1		
CCL2/MCP-1	6347	chemokine (C-C motif) ligand 2		
CCL3/MIP-1α	6348	chemokine (C-C motif) ligand 3		
CCL4/MIP-1β	6351	chemokine (C-C motif) ligand 4		
CCL5/RANTES	6352	chemokine (C-C motif) ligand 5	x	x
CCL7/MCP-3	6354	chemokine (C-C motif) ligand 7	x	x
CCL8/MCP-2	6355	chemokine (C-C motif) ligand 8		
CCL11/Eotaxin	6356	chemokine (C-C motif) ligand 11		
CCL13/MCP-4	6357	chemokine (C-C motif) ligand 13		
CCL15/MIP-1δ	6359	chemokine (C-C motif) ligand 15	x	x
CCL16/HCC-4	6360	chemokine (C-C motif) ligand 16		
CCL17/TARC	6361	chemokine (C-C motif) ligand 17		
CCL18/PARC	6362	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)	x	x
CCL19/MIP-3β	6363	chemokine (C-C motif) ligand 19		
CCL20/MIP-3α	6364	chemokine (C-C motif) ligand 20		
CCL22/MDC	6367	chemokine (C-C motif) ligand 22	x	
CCL23/CKb8-1	6368	chemokine (C-C motif) ligand 23		
CCL24/ Eotaxin-2	6369	chemokine (C-C motif) ligand 24		
CCL25/TECK	6370	chemokine (C-C motif) ligand 25		
CCL26/ Eotaxin-3	10344	chemokine (C-C motif) ligand 26		
CCL27/CTACK	10850	chemokine (C-C motif) ligand 27		
CNTF	1270	ciliary neurotrophic factor		
CX3CL1/ Fractalkine	6376	chemokine (C-X3-C motif) ligand 1		
CXCL1,2,3/ GRO-α,β,γ	2919 2920 2921	chemokine (C-X-C motif) ligand 1 chemokine (C-X-C motif) ligand 2 chemokine (C-X-C motif) ligand 3		
CXCL1/GRO-α	2919	chemokine (C-X-C motif) ligand 1		
CXCL5/ENA-78	6374	chemokine (C-X-C motif) ligand 5		
CXCL6/GCP-2	6372	chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)		
CXCL7/NAP-2	5473	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)		
CXCL8/IL-8	3576	interleukin 8	x	x
CXCL9/MIG	4283	chemokine (C-X-C motif) ligand 9		
CXCL11/I-TAC	6373	chemokine (C-X-C motif) ligand 11		
CXCL12/SDF-1	6387	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)		
CXCL13/BLC	10563	chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)		
EGF	1950	epidermal growth factor (beta-urogastrone)	x	x
EGFR	1956	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)		
Fas	355	Fas (TNF receptor superfamily, member 6)		
FGF-4	2249	fibroblast growth factor 4 (heparin secretory transforming protein 1)		
FGF-6	2251	fibroblast growth factor 6		
FGF-7	2252	fibroblast growth factor 7 (keratinocyte growth factor)		
FGF-9	2254	fibroblast growth factor 9 (glia-activating factor)		
Flt-3L	2323	fms-related tyrosine kinase 3 ligand		
G-CSF	1440	colony stimulating factor 3 (granulocyte)	x	x
GDNF	2668	glial cell derived neurotrophic factor	x	x

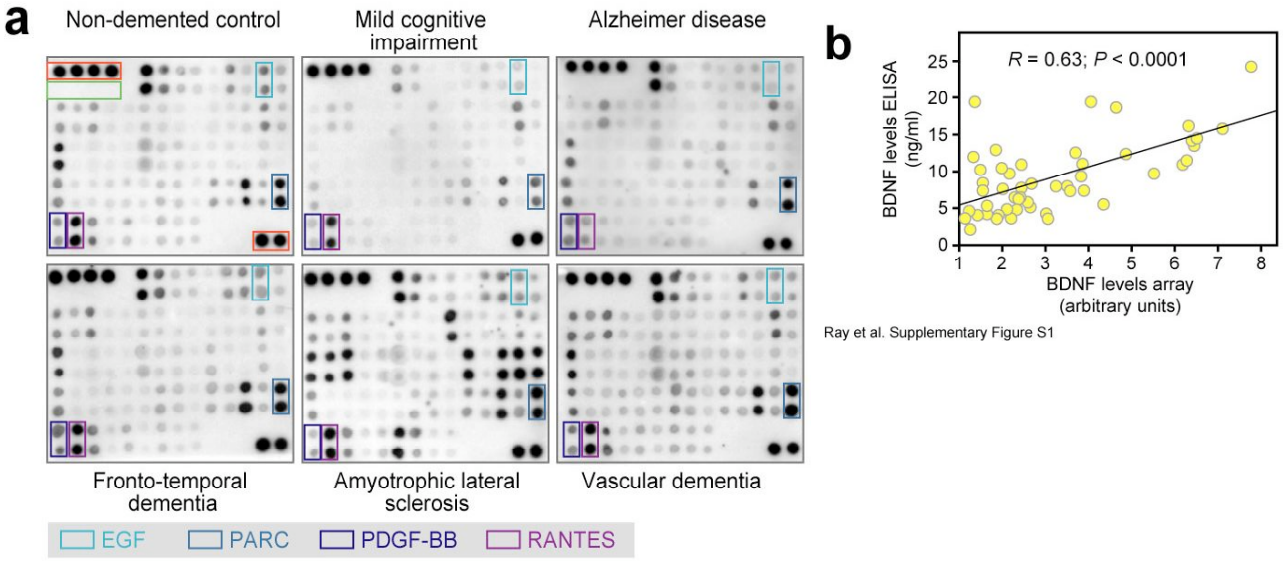
Protein Name	Entrez GeneID	Official gene name provided by HUGO Gene Nomenclature Committee (HGNC)	SAM19 AD/NDC [a]	PAM18 Predictors [c]
GITR	8784	tumor necrosis factor receptor superfamily, member 18		
GITR-L	8995	tumor necrosis factor (ligand) superfamily, member 18		
GM-CSF	1437	colony stimulating factor 2 (granulocyte-macrophage)		
HGF	3082	hepatocyte growth factor (hepapoietin A; scatter factor)		
ICAM-1	3383	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	x	x
ICAM-3	3385	intercellular adhesion molecule 3		
IFN-γ	3458	interferon, gamma		
IGF1-R	3480	insulin-like growth factor 1 receptor		
IGFBP-1	3484	insulin-like growth factor binding protein 1		
IGFBP-2	3485	insulin-like growth factor binding protein 2		
IGFBP-3	3486	insulin-like growth factor binding protein 3		
IGFBP-4	3487	insulin-like growth factor binding protein 4		
IGFBP-6	3489	insulin-like growth factor binding protein 6	x	x
IGF-I	3479	insulin-like growth factor 1 (somatomedin C)		
IL-1 R-like 1	9173	interleukin 1 receptor-like 1		
IL-1 sRI	3554	interleukin 1 receptor, type I		
IL-1ra	3557	interleukin 1 receptor antagonist		
IL-1α	3552	interleukin 1, alpha	x	x
IL-1β	3553	interleukin 1, beta		
IL-2	3558	interleukin 2		
IL-2 sRa	3559	interleukin 2 receptor, alpha		
IL-3	3562	interleukin 3 (colony-stimulating factor, multiple)	x	x
IL-4	3565	interleukin 4		
IL-5	3567	interleukin 5 (colony-stimulating factor, eosinophil)		
IL-6	3569	interleukin 6 (interferon, beta 2)		
IL-6 sR	3570	interleukin 6 receptor		
IL-7	3574	interleukin 7		
IL-10	3586	interleukin 10		
IL-11	3589	interleukin 11	x	x
IL-12p40	3593	interleukin 12B (natural killer cell stimulatory factor 2, p40)		
IL-12p70	3592	interleukin 12A (natural killer cell stimulatory factor 1, p35)		
IL-13	3596	interleukin 13		
IL-15	3600	interleukin 15		
IL-16	3603	interleukin 16 (lymphocyte chemoattractant factor)		
IL-17	3605	interleukin 17A		
Leptin	3952	leptin		
LIGHT	8740	tumor necrosis factor (ligand) superfamily, member 14		
M-CSF	1435	colony stimulating factor 1 (macrophage)	x	x
MIF	4282	macrophage migration inhibitory factor (glycosylation-inhibiting factor)		
MSP α-chain	4485	macrophage stimulating 1 (hepatocyte growth factor-like) alpha-chain		
NGF-β	4803	nerve growth factor, beta polypeptide		
NT-3	4908	neurotrophin 3		
NT-4/5	4909	neurotrophin 5 (neurotrophin 4/5)		
Oncostatin M	5008	oncostatin M		
Osteoprotegerin	4982	tumor necrosis factor receptor superfamily, member 11b		
PDGF-BB	5155	platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)	x	x
PLGF	5228	placental growth factor, vascular endothelial growth factor-related protein		
SCF	4254	KIT ligand		
Sgp130	3572	interleukin 6 signal transducer (gp130, oncostatin M receptor)		
TGF-b1	7040	transforming growth factor, beta 1		
TGF-β3	7043	transforming growth factor, beta 3		
TIMP-1	7076	TIMP metallopeptidase inhibitor 1		
TIMP-2	7077	TIMP metallopeptidase inhibitor 2		
TNFR-1	7132	tumor necrosis factor receptor superfamily, member 1A		
TNFR-2	8764	tumor necrosis factor receptor superfamily, member 14		
TNF-α	7124	tumor necrosis factor (TNF superfamily, member 2)	x	x
TNF-β	4049	lymphotoxin alpha (TNF superfamily, member 1)		

Protein Name	Entrez GeneID	Official gene name provided by HUGO Gene Nomenclature Committee (HGNC)	SAM19 AD/NDC [a]	PAM18 Predictors [c]
TPO	7066	thrombopoietin (megakaryocyte growth and development factor)		
TRAIL R3	8794	tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain		
TRAIL R4	8793	tumor necrosis factor receptor superfamily, member 10d, decoy with truncated death domain	x	x
TYRO3	7301	TYRO3 protein tyrosine kinase		
uPAR	5329	plasminogen activator, urokinase receptor		
VEGF-B	7423	vascular endothelial growth factor B		
VEGF-D	2277	c-fos induced growth factor (vascular endothelial growth factor D)		
XCL1/ Lymphotactin	6375	chemokine (C motif) ligand 1		

[a] 19 proteins identified in training set comparing AD vs NDC with *SAM* analysis (see node map in **Fig. 1b**)
[b] 18 proteins identified in training set with *PAM* predictor discovery algorithm (see also **Table 1**)

Membranes were developed with streptavidin-conjugated peroxidase and ECL chemiluminescense reagent and exposed to autoradiographic film (BioMax Lite, Kodak) (**Supplementary Fig. S1a** online). Repeat measurements of the same samples produced Pearson correlation coefficients R^2 > 0.95 for the 120 protein measurements (n = 4 samples; data not shown) indicating high array reproducibility. Additionally, array measurements of individual factors by ELISA confirmed the validity of the assay as shown for brain-derived neurotrophic factor (BDNF; **Supplementary Fig. S1b** online and data not shown). BDNF levels were quantified in 50 plasma samples by sandwich enzyme immunoassay using a commercially available kit (Quantikine, R&D Systems) and filter array results were compared with ELISA measurements by simple regression analysis in StatView version 5.0.1 (SAS Institute Inc.).

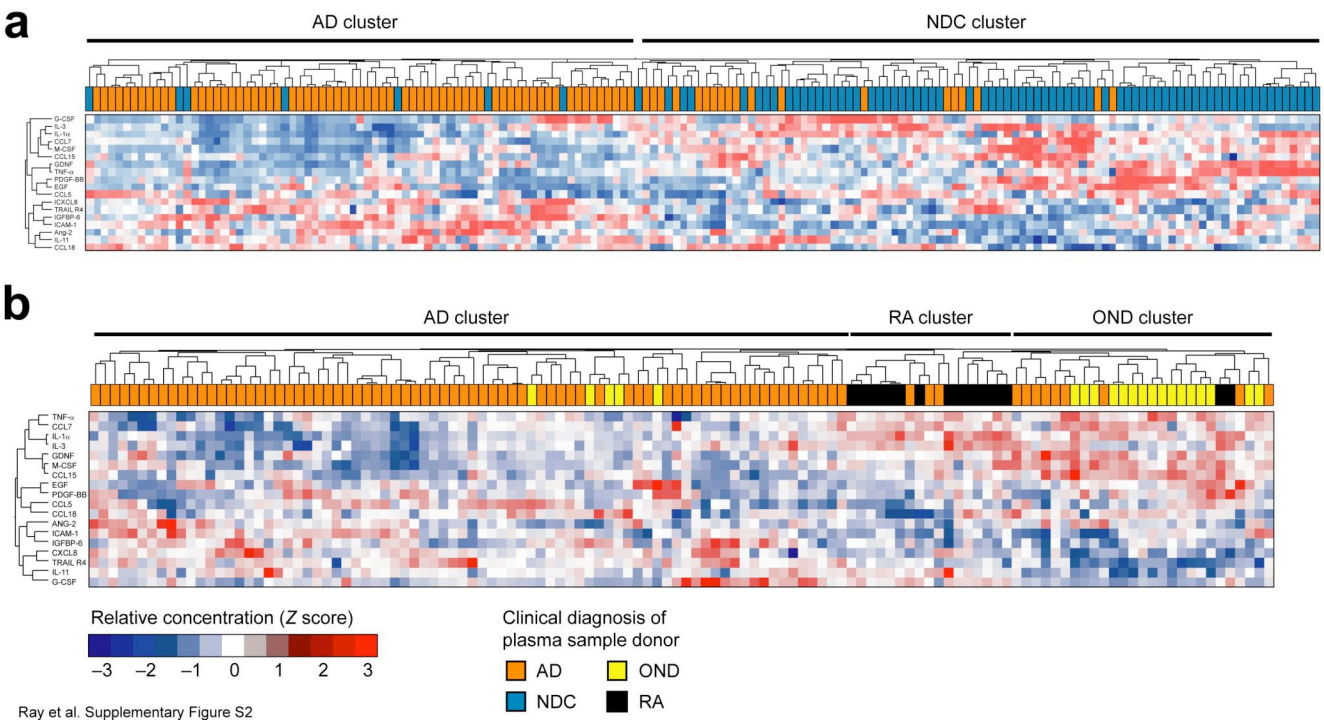
Data extraction and transformation. Autoradiographic films were scanned and digitized spots were quantified with the *Imagene 6.0* data extraction software (BioDiscovery Inc.). Local background intensities were subtracted from each spot, and the average of the duplicate spots for each protein was normalized to the average of six positive controls on each membrane. For statistical analysis expression data from the two filters per sample were normalized to the median expression of all 120 proteins followed by Z score transformation (data file is available online).



Supplementary Fig. S1: Array filter membranes and validation of array with ELISA. (a) Examples of autoradiographs exposed to filters from plasma sample donors with the indicated diagnoses. Human plasma was incubated with an array membrane that detects 60 proteins. Each protein is measured in duplicates. Arrays were developed and exposed to autoradiographic film. Red boxes, positive controls (upper left and lower right corner, high intense spots); green box, negative controls (upper left corner, no spots). Colored boxes indicate the location of the detection of three proteins. Note the differences in expression patterns among the various conditions. **(b)** Regression analysis between filter array and sandwich ELISA for brain-derived neurotrophic factor (BDNF) (n = 50).

Cytokine antibody array data analysis. *Significance analysis of microarray (SAM)*. Group differences in the training set samples were analyzed with the *SAM 3.00* algorithm ([3]; http://www-stat.stanford.edu/~tibs/SAM/index.htm). *SAM* assigns a d-score to each gene or protein on the basis of a multi-comparison analysis of expression changes and indicates significance by q-value.

Unsupervised Clustering. A two-way unsupervised clustering algorithm with a top down repeated bisection approach was used in the clustering package *CLUTO 2.1.1* ([4]; http://glaros.dtc.umn.edu/gkhome/cluto/cluto/download) to group proteins on the basis of similarity in pattern of expression over all samples (**Fig. 1b** and **Supplementary Fig. 2** online).

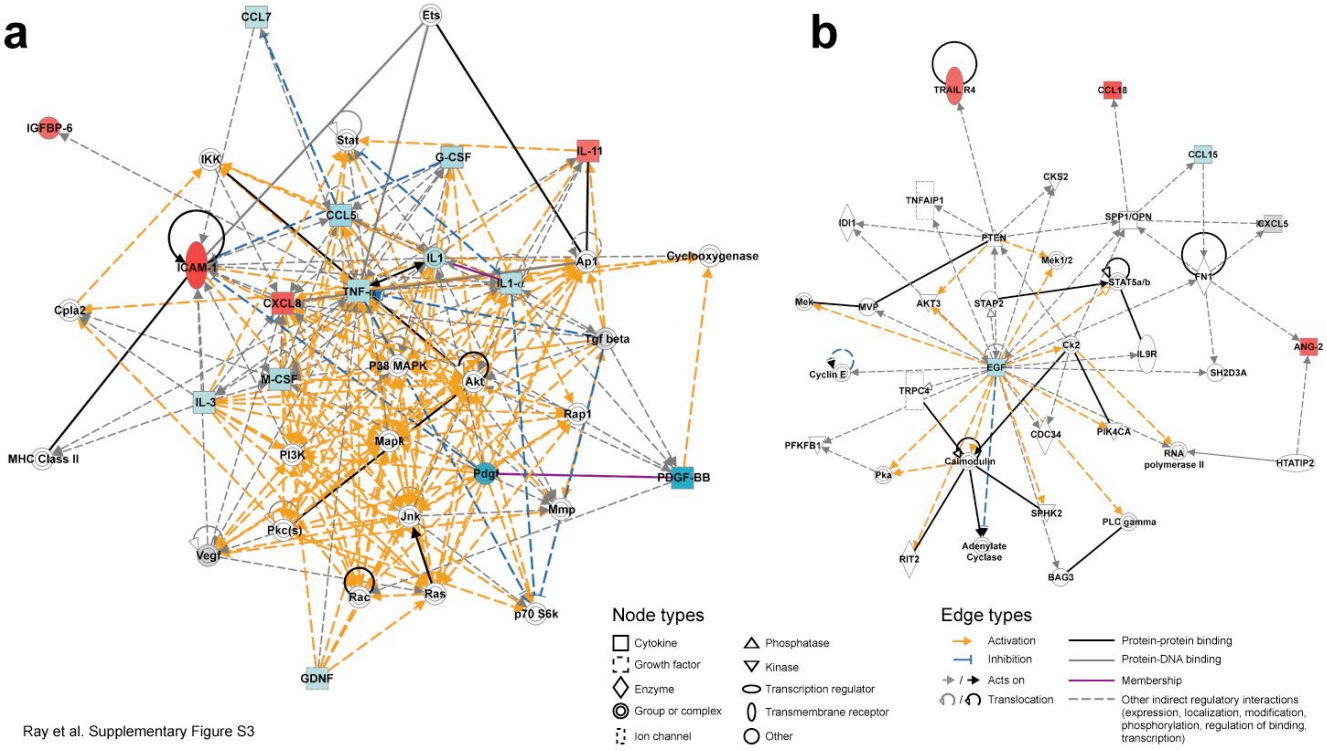


Supplementary Fig. S2: Distinct pattern of signaling protein expression in Alzheimer disease compared with non-demented controls, other neurological disorders, and rheumatoid arthritis. (a) Normalized and *Z* scored array measurements of 18 differentially expressed signaling proteins in plasma from subjects with Alzheimer disease (AD; orange; *n* = 85) and non-demented controls (NDC; blue, *n* = 79) are shown in a node map after unsupervised clustering. Samples are clustered into AD and NDC, respectively, with high accuracy indicated by the first order branches of the dendrogram (two black bars at the top). **(b)** Normalized and *Z* scored array measurements of the 18 predictors in plasma samples from subjects with Alzheimer disease (AD; orange; *n* = 85), other neurological diseases (OND; yellow; *n* = 21), or rheumatoid arthritis (RA; black; *n* = 16) are shown in a node map after unsupervised clustering. Samples are mainly clustered into AD, RA and OND, respectively, (three black bars at the top) indicating distinct disease-specific molecular patterns that allow for differential clustering. This is in line with recent observations obtained in RA with other peripheral immune markers measured on a different platform [5]. **(a and b)** Samples are arranged in columns and proteins in rows. Increased expression is shown in shades of red, reduced expression in shades of blue, median expression is shown in white.

Predictive analysis of microarray (PAM). A semi-supervised prediction analysis was performed using the statistical package *PAM 2.3.1* with the statistical tool *R* ([6]; http://www-stat.stanford.edu/~tibs/PAM/index.html). *PAM* executes a sample classification training routine from expression data via the nearest shrunken centroid procedure to find markers that discriminate best between two or more classes. Then it applies an internal cross-validation by 10-times randomly selecting 90% of the training samples in a class-balanced way to predict each time the class labels on the remaining 10% of samples (10-fold cross-validation; **Fig. 1a**). This assesses and minimizes classification errors and avoids overfitting. (**Fig. 1c**). The obtained minimal number of predictors is then used for a heterogeneity analysis to perform two-class prediction in a test set between a diseased group and a control group. *PAM* demonstrated highest accuracy when compared with other algorithms on several public datasets (http://www-stat.stanford.edu/~tibs/PAM/comparison.html).

Functional profiling of the 18 predictors. *Analysis of regulatory interaction network.* Regulatory network analysis in the mammalian system was done with Ingenuity Pathways Analysis (IPA, Ingenuity® Systems, www.ingenuity.com [7]; **Supplementary Fig. S3** online). A data set containing gene identifiers (Entrez gene ID) and corresponding expression values was uploaded into in the application. Each gene identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledge Base (IPKB). The 18 predictors were flagged and these genes, called *Focus Genes*, were overlaid onto a preliminary global molecular network developed from information contained in the IPKB. Networks of these

Focus Genes were then algorithmically generated based on their connectivity. A network is a graphical representation of the molecular relationships between genes and gene products. Genes or gene products are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one reference from the literature, from a textbook, or from canonical information stored in the IPKB. Human, mouse, and rat orthologs of a gene are stored as separate objects in the IPKB, but are represented as a single node in the network. The intensity of the node color indicates the degree of up- (red) or down- (blue) regulation in our filter array experiment. Nodes are displayed using various shapes that represent the functional class of the gene product. Edges are displayed as solid (direct interaction) or dashed (indirect interactions) lines with and without arrowheads. Various colors that describe the nature of the relationship between the nodes were added manually in Adobe Illustrator (Adobe Systems Inc.). The functional analysis of a network identified the biological functions that were most significant to the genes in the network (Fisher's exact test; data not shown).



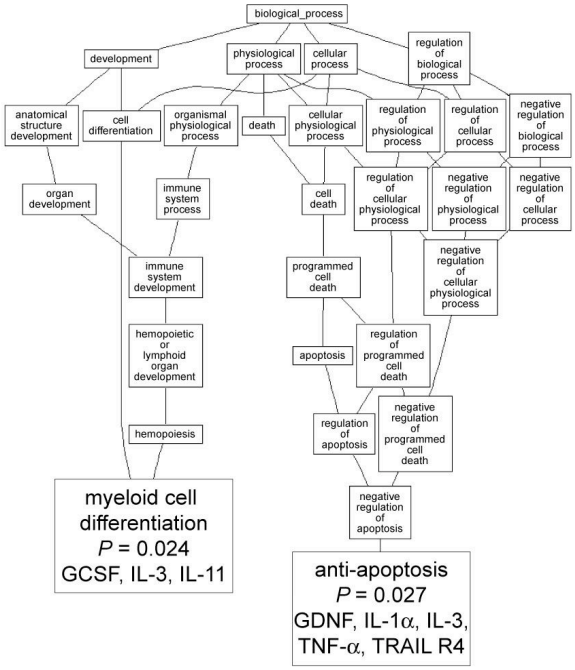
Supplementary Fig. S3: Regulatory gene and protein interaction networks defined by the 18 predictors. Computational molecular interaction network prediction based on genes and proteins associated with the significant pathways in the Ingenuity Pathways Knowledge Base (IPKB). Two independent networks were built from 13 **(a)** and from 5 signaling proteins **(b)** out of the set of 18 predictors, respectively, using *Ingenuity Pathways Analysis* (*IPA*). Predictors are highlighted in red (up-regulation) or blue (down-regulation), and node properties are indicated by shape. Interactions between the different nodes are given as solid (direct interaction) and dashed (indirect interaction) lines (edges) with various colors for the different interaction types. **(a)** This network received a high score by *IPA* and is mostly centered around the signaling proteins TNF-α and M-CSF. Associated functions are cell-to-cell signaling and interaction, cellular growth and proliferation (connective tissue, hematological, immune, and lymphoid system), immune response, and cell death. Of note, the multifunctional cytokine TNF-α, which has neuroprotective as well as neurotoxic effects [8], has previously been reported to be reduced in Alzheimer brains and serum in some studies, although others could not confirm this (**Supplementary Table S3** online) [9]. **(b)** This network received a low score by *IPA* and is mostly centered around EGF. Associated functions are gene expression, cancer, and cellular movement. In both networks most interactions are based on indirect activation of groups or complexes of kinases. Decreased concentrations of many of the predictive markers connected in the two networks may lead to an overall reduced activation of these kinases

Gene ontology analysis. For human gene ontology (www.geneontology.org) term profiling of the 18 predictors the online gene set analysis toolkit *WebGestalt* (http://bioinfo.vanderbilt.edu/webgestalt; [10]) was set to level 4 and $P \leq 0.05$ to stratify the search for gene enrichment in comparison to our own reference list of 120 proteins measured by filter array. Significant gene overrepresentations found by a hypergeometric statistical test were illustrated as an enriched Directed Acyclic Graph (DAG) (**Supplementary Fig. 4a** online).

Biological pathways analysis. To obtain an overview of human metabolic and regulatory pathways affected by Alzheimer disease, *WebGestalt* was used to query for enriched pathways in the open source databases *Kyoto Encyclopedia of Genes and Genomes* (*KEGG*, www.genome.ad.jp/kegg/; [11]) and

BioCarta (www.biocarta.com/genes/index.asp). From our filter array 103 markers including 15 predictors are present in the *KEGG* pathways and 56 markers including 12 predictors in the *BioCarta* pathways, respectively. Angiopoietin (ANG)-2 and glial cell derived neurotrophic factor (GDNF) are not listed in *KEGG* or *BioCarta* at all. We identified overrepresented pathways by stratifying for at least three significantly ($P \leq 0.05$) enriched markers in each pathway.

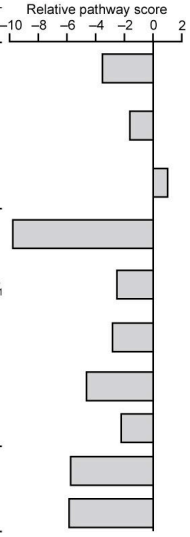
a



Supplementary Fig. S4: Functional profiling of the 18 predictors. **(a)** Analysis of significant enrichment of the 18 predictors in human gene ontology categories. A Direct Acyclic Graph (DAG) generated in *WebGestalt* illustrates the significant enrichment for seven out of 18 predictors in two human gene ontology sub-categories. Significance ($P \leq 0.05$) was calculated by the hypergeometric test. **(b)** Clustering of significant enrichment in biological pathways using *DAVID* selected nine out of 18 proteins with at least three proteins per pathway. The column "*n* =" indicates the number of genes from our 120-protein reference list that are present (P) in a certain pathway, for which a change of expression can be expected (E; given as P x 18 / 120) in this pathway, or which were observed (O) to be changed. Significance ($p \leq 0.05$) of expected versus obtained proteins was calculated by hypergeometric test in *WebGestalt*. *DAVID 2006* clustered similar pathways into the same functional group and used EASE scores and geometric means of EASE scores (EASE geom) for ranking the pathways within a functional group and the functional groups, respectively. To illustrate the impact of the protein expression levels on a particular pathway, a relative pathway score was calculated as the sum of the *d*-scores obtained in *SAM* for the individual markers **(Table 1)**. BC; *BioCarta*, KEGG; *Kyoto Encyclopedia of Genes and Genome*.

b

Functional group	Pathway database	Pathway name	*n* = P / E / O	EASE score	*P* value	Observed
Group 1: Immune response EASE geom = 0.083	BC	h_granulocytesPathway: adhesion and diapedesis of granulocytes	6 / 0.9 / 5	0.033	0.00023	TNF-α, G-CSF, IL-1α, CXCL8, ICAM-1
	BC	h_LairPathway: cells and molecules involved in local acute inflammatory response	5 / 0.8 / 4	0.112	0.00163	TNF-α, IL-1α, CXCL8, ICAM-1
	BC	h_lympathway: adhesion and diapedesis of lymphocytes	3 / 0.5 / 3	0.154	0.00294	IL-1α, CXCL8, ICAM-1
Group 2: Hematopoiesis EASE geom = 0.181	KEGG	HSA04640: hematopoietic cell lineage	18 / 2.7 / 6	0.075	0.02812	TNF-α, M-CSF, G-CSF, IL-1α, IL-3, IL-11
	BC	h_stemPathway: regulation of hematopoiesis by cytokines	11 / 1.7 / 5	0.147	0.01098	M-CSF, G-CSF, CXCL8, IL-3, IL-1
	BC	h_inflamPathway: cytokines and inflammatory response	20 / 3.0 / 7	0.167	0.01185	TNF-α, M-CSF, G-CSF, IL-1α, CXCL8, IL-3, IL-11
	BC	h_erythPathway: erythrocyte differentiation pathway	10 / 1.5 / 4	0.272	0.04120	G-CSF, IL-1α, IL-3, IL-11
	BC	h_il17Pathway: IL17 signaling pathway	6 / 0.9 / 3	0.388	0.04197	G-CSF, CXCL8, IL-3
Group 3: Apoptosis EASE geom = 0.340	KEGG	HSA04210: apoptosis	10 / 1.5 / 4	0.159	0.04120	TNF-α, TRAIL R4, IL-1α, IL-3
	BC	h_cytokinePathway: cytokine network	16 / 2.4 / 4	0.725	0.19336	TNF-α, IL-1α, CXCL8, IL-3



Ray et al. Supplementary Figure S4

We verified the *WebGestalt* findings in the *Database for Annotation, Visualization, and Integrated Discovery* (*DAVID) 2006*, an online graph theory evidence-based method to agglomerate heterogeneous and widely distributed public databases (http://david.abcc.ncifcrf.gov/home.jsp; [12]). *DAVID* allows to search, rank and cluster functional gene or protein similarities within a set of genes or proteins of interest in order to unravel new biological processes associated with their cellular functions and pathways. The online module *Expression Analysis Systematic Explorer* (*EASE*), which is incorporated in *DAVID,* searched for overrepresented markers within the present proteins of interest in comparison to our protein reference list. *EASE* generated a gene representation score (*EASE* score; *P value* of the Jackknife Fisher exact test), which *DAVID* used to cluster the enriched proteins and their corresponding pathways at lowest stringency for functional similarity. Because *DAVID* clustered the biological pathways by functional

similarity into groups without providing a group name, we identified the three obtained groups based on the listed functions in a group (**Supplementary Fig. 4b** online). To better examine the overall effect of up- or downregulation of the enriched predictors on the individual pathways, we calculated a relative pathway score by adding up the respective *SAM*-derived (**Table 1**) positive and negative *d*-scores of the individual markers in each pathway (**Supplementary Fig. 4b** online).

Brain- and Alzheimer disease-specific functions and findings. Because *WebGestalt* and *DAVID* do not allow for a specific search for brain- and Alzheimer disease-specific functions of the 18 predictors and to further confirm the findings of the two online tools used above, we performed our own investigation on *PubMed* (www.pubmed.gov) with the following keywords: *neuroprotection, neurotrophic, neurodegeneration, cerebrovasculature, inflammation, phagocytosis, hematopoiesis,* or *energy metabolism* (**Fig. 2d**). If at least one *PubMed* entry was found linking a given factor with the specified function *in vivo*, the corresponding *SAM d*-score was assigned to that keyword. If no entry was found, null was given. To order the markers and their keywords a hierarchical cluster algorithm with a pair wise similarity function was applied (Open Source Clustering Software *Cluster 3.0*, http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/; [13,14]). Cluster results were displayed as a node map using *Java TreeView* ([15]; http://jtreeview.sourceforge.net/; **Fig. 2d**). Similarly, we searched for *PubMed* entries that describe expression of the 18 predictors in the central nervous system, their ability to cross the blood-brain-barrier, or if they have been studied with regard to aging. In addition, we searched for reports on Alzheimer disease or its mouse models that describe changes of expression levels (RNA and protein) or abnormal presence of the 18 predictive markers in plasma, serum, cerebrospinal fluid, or brain parenchyma (**Supplementary Table S3** online).

Supplementary Table S3. Expression changes in Alzheimer disease reported for seven of the 18 predictors

Protein	Our finding	Plasma/serum	Cerebrospinal fluid	Brain parenchyma
ICAM-1	↑	↑ sICAM-1 in serum [16]		↑ immunoreactivity in and around plaques in humans [17,18]; ↔ hippocampal gene expression in incipient and moderate AD [19] ↑ hippocampal gene expression in severe AD [19] ↑ with progression of disease in activated microglia and in plaques of APP mice [20]
CXCL8	↑		↑ in MCI and AD [21]	↑ in microvasculature [22] Reviewed by Xia et al. [23]
IGFBP-6	↑		↑ in AD [24]	
M-CSF	↓		↑ in AD [25]	↑ neuronal immunoreactivity in proximity to Aβ deposits [25] ↑ hippocampal gene expression [19] ↑ immunoreactivity on neuritic structures near Aβ deposits in APP transgenic mice [26]
IL-1α [a]	↓	↔ serum in AD and multi-infarct dementia [27]		↑ hippocampal gene expression [19]
TNF-α	↓	↑ in plasma in centenarians with AD compared to those without [28] ↓ in serum in mild-moderate AD versus severe AD and vascular dementias [29] ↓ in serum in early and late onset AD versus control [30] ↓ in serum in AD and multi-infarct dementia [31]; ↔ AD vs control [32]	↑ in MCI [33] ↑ in AD and vascular dementia [34] ↓ in AD [35] ↔ AD vs control [32]	↓ in AD frontal cortex, superior temporal gyrus , and entorhinal cortex compared with controls [32]
PDGF-BB	↓			↓ number of PDGF-BB immunoreactive pyramidal neurons in AD [36] ↑ immunoreactivity with neurofibrillary tangles in AD [36] ↑ hippocampal gene expression [19]

[a] An extensive search for articles in *PubMed* for evidence on changes of IL-1α in Alzheimer disease revealed only the two studies cited in the table. Instead, most articles (>400) report on Il-1β or IL1.

Changes in expression levels (RNA or protein) or abnormal presence of the listed proteins in plasma/serum, cerebrospinal fluid, or brain parenchyma in Alzheimer disease or mouse models of the disease reported in *PubMed* articles. We did not list any reports of genetic associations between the listed factors and Alzheimer disease. Updates on systematic meta-analyses of genetic association studies in Alzheimer disease can be found on the Alzgene website (www.alzgene.org) [37]. AD, Alzheimer disease; APP, amyloid precursor protein; MCI, mild cognitive impairment; ↑, increased; ↓ decreased; ↔ no change; empty cells, no reports found or proteins were reported to be undetectable.

Diagnostic test statistics. The mean diagnostic accuracy of Alzheimer disease by neurologists across the US yielded in 2001 81% sensitivity and 70% specificity although Alzheimer research centers such as the ones that provided samples for this study had 90-100% sensitivity [38]. Nevertheless, many patients form our study are still alive and we cannot be 100% certain about the diagnosis for each of them. Therefore, and consistent with recommendations by the US Food and Drug Administration (www.fda.gov/cdrh/osb/guidance/1620.html; [39]), we use the terms "*negative, positive, or overall percent agreement with clinical diagnosis*" rather than "*sensitivity*", "*specificity*", or "*accuracy*" . The calculation and

report of this "*estimate of agreement with clinical diagnosis*" in a 2×2 contingency table is identical to the method for the latter terms. Calculations were done in *InStat 3.0* and 95% confidence intervals and two-sided *P value* of the Fisher's exact test are given (GraphPad Software Inc.).

References

1. Folstein, M.F., Folstein, S.E. & McHugh, P.R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198 (1975).
2. Huang, R.P. Cytokine protein arrays. *Methods Mol Biol* **264**, 215-231 (2004).
3. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121 (2001).
4. Zhao, Y. & Karypis, G. Clustering in life sciences. *Methods Mol Biol* **224**, 183-218 (2003).
5. Hueber, W.*, et al.* Proteomic analysis of secreted proteins in early rheumatoid arthritis: Anti-citrulline reactivity is associated with upregulation of proinflammatory cytokines. *Ann Rheum Dis* (2006).
6. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**, 6567-6572 (2002).
7. Calvano, S.E.*, et al.* A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032-1037 (2005).
8. Wyss-Coray, T. Inflammation in Alzheimer disease: driving force, bystander or beneficial response? *Nat Med* **12**, 1005-1015 (2006).
9. Britschgi, M. & Wyss-Coray, T. Systemic and acquired immune responses in Alzheimer's disease. *International review of neurobiology* **82**, 205-233 (2007).
10. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic acids research* **33**, W741-748 (2005).
11. Kanehisa, M.*, et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research* **34**, D354-357 (2006).
12. Dennis, G., Jr.*, et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
13. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-14868 (1998).
14. de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-1454 (2004).
15. Saldanha, A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248 (2004).
16. Rentzos, M.*, et al.* Serum levels of soluble intercellular adhesion molecule-1 and soluble endothelial leukocyte adhesion molecule-1 in Alzheimer's disease. *J Geriatr Psychiatry Neurol* **17**, 225-231 (2004).
17. Rozemuller, J.M., Eikelenboom, P., Pals, S.T. & Stam, F.C. Microglial cells around amyloid plaques in Alzheimer's disease express leucocyte adhesion molecules of the LFA-1 family. *Neurosci Lett* **101**, 288-292 (1989).
18. Verbeek, M.M.*, et al.* Accumulation of intercellular adhesion molecule-1 in senile plaques in brain tissue of patients with Alzheimer's disease. *Am J Pathol* **144**, 104-116 (1994).
19. Blalock, E.M.*, et al.* Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci U S A* **101**, 2173-2178 (2004).
20. Apelt, J., Lessig, J. & Schliebs, R. Beta-amyloid-associated expression of intercellular adhesion molecule-1 in brain cortical tissue of transgenic Tg2576 mice. *Neurosci Lett* **329**, 111-115 (2002).
21. Galimberti, D.*, et al.* Intrathecal chemokine synthesis in mild cognitive impairment and Alzheimer disease. *Arch Neurol* **63**, 538-543 (2006).
22. Grammas, P., Samany, P.G. & Thirumangalakudi, L. Thrombin and inflammatory proteins are elevated in Alzheimer's disease microvessels: implications for disease pathogenesis. *J Alzheimers Dis* **9**, 51-58 (2006).
23. Xia, M.Q. & Hyman, B.T. Chemokines/chemokine receptors in the central nervous system and Alzheimer's disease. *J Neurovirol* **5**, 32-41 (1999).
24. Tham, A.*, et al.* Insulin-like growth factors and insulin-like growth factor binding proteins in cerebrospinal fluid and serum of patients with dementia of the Alzheimer type. *J Neural Transm Park Dis Dement Sect* **5**, 165-176 (1993).
25. Yan, S.D.*, et al.* Amyloid-b peptide-receptor for advanced glycation endproduct interaction elicits neuronal expression of macrophage-colony stimulating factor: a proinflammatory pathway in Alzheimer disease. *Proc. Natl. Acad. Sci. USA* **94**, 5296-5301 (1997).
26. Murphy, G.M., Jr., Zhao, F., Yang, L. & Cordell, B. Expression of macrophage colony-stimulating factor receptor is increased in the AbetaPP(V717F) transgenic mouse model of Alzheimer's disease. *Am J Pathol* **157**, 895-904 (2000).
27. Cacabelos, R., Franco-Maside, A. & Alvarez, X.A. Interleukin-1 in Alzheimer's disease and multi-infarct dementia: neuropsychological correlations. *Methods Find Exp Clin Pharmacol* **13**, 703-708 (1991).
28. Bruunsgaard, H.*, et al.* A high plasma concentration of TNF-alpha is associated with dementia in centenarians. *J Gerontol A Biol Sci Med Sci* **54**, M357-364 (1999).
29. Paganelli, R.*, et al.* Proinflammatory cytokines in sera of elderly patients with dementia: levels in vascular injury are higher than those of mild-moderate Alzheimer's disease patients. *Exp Gerontol* **37**, 257-263 (2002).
30. Alvarez, X.A., Franco, A., Fernandez-Novoa, L. & Cacabelos, R. Blood levels of histamine, IL-1 beta, and TNF-alpha in patients with mild to moderate Alzheimer disease. *Mol Chem Neuropathol* **29**, 237-252 (1996).
31. Cacabelos, R., Alvarez, X.A., Franco-Maside, A., Fernandez-Novoa, L. & Caamano, J. Serum tumor necrosis factor (TNF) in Alzheimer's disease and multi-infarct dementia. *Methods Find Exp Clin Pharmacol* **16**, 29-35 (1994).
32. Lanzrein, A.S.*, et al.* Longitudinal study of inflammatory factors in serum, cerebrospinal fluid, and brain tissue in Alzheimer disease: interleukin-1beta, interleukin-6, interleukin-1 receptor antagonist, tumor necrosis factor-alpha, the soluble tumor necrosis factor receptors I and II, and alpha1-antichymotrypsin. *Alzheimer Dis Assoc Disord* **12**, 215-227 (1998).
33. Tarkowski, E., Andreasen, N., Tarkowski, A. & Blennow, K. Intrathecal inflammation precedes development of Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **74**, 1200-1205 (2003).
34. Tarkowski, E., Blennow, K., Wallin, A. & Tarkowski, A.h. Intracerebral production of tumor necrosis factor-alpha, a local neuroprotective agent, in Alzheimer disease and vascular dementia. *J Clin Immunol* **19**, 223-230 (1999).
35. Richartz, E.*, et al.* Decline of immune responsiveness: a pathogenetic factor in Alzheimer's disease? *J Psychiatr Res* **39**, 535-543 (2005).
36. Masliah, E., Mallory, M., Alford, M., DeTeresa, R. & Saitoh, T. PDGF is associated with neuronal and glial alterations of Alzheimer's disease. *Neurobiol. Aging* **16**, 549–556 (1995).
37. Bertram, L., McQueen, M.B., Mullin, K., Blacker, D. & Tanzi, R.E. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* **39**, 17-23 (2007).
38. Knopman, D.S.*, et al.* Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* **56**, 1143-1153 (2001).
39. Food and Drug Administration. Guidance for Industry and FDA Staff - Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. (eds. Services, U.S.D.o.H.a.H.*, et al.*) (2007).