BLOOD-BASED METABOLIC SIGNATURES IN ALZHEIMER'S DISEASE: SUPPLEMENTARY TEXT 3

BASIC RESULTS ON ALL CASES

FRANCISCA A. DE LEEUW*†, CAREL F.W. PEETERS*†,
MAARTJE I. KESTER, AMY C. HARMS, EDUARD A. STRUYS,
THOMAS HANKEMEIER, HERMAN W.T. VAN VLIJMEN,
SVEN J. VAN DER LEE, CORNELIA M. VAN DUIJN, PHILIP SCHELTENS,
AYŞE DEMIRKAN, MARK A. VAN DE WIEL,
WIESJE M. VAN DER FLIER, AND CHARLOTTE E. TEUNISSEN

This supplementary text presents, for purposes of comparison, basic information on the obtained expression and classification signatures when considering data from all patients. That is, patients whose clinical diagnosis was discordant from their CSF-biomarker status were not excluded from the analyzes described below. Processing of the data was analogous to the steps described in Section 2.4 of the main text and *Section 1.2* of *Supplementary Text 2*. Again, metabolites with more than 10% missing observations were removed, leading to the removal of the same 5 metabolites mentioned in *Section 1.2* of *Supplementary Text 2*. Also, again three data samples were removed as their (plasma) quality was deemed unsure and an additional twelve data samples were removed due to instrumental errors in one or more MS platforms. The final metabolic data set for the analyzes below thus contained $n = 285$ data samples (141 AD and 144 SCD) and $p = 230$ metabolic features. Section 1 contains information on the differential expression signature. Section 2 then contains information on the classification signature. Section 3 concludes with some reflections on the findings.

1. Differential Expression Signature

The approach for the evaluation of differential metabolic expression between AD and SCD subjects was described in *Section 2.1.1* of *Supplementary Text 2*. The list of metabolic features that survive multiple testing correction when only sex and age are used as possible confounders can be found in Table S3.1. Table S3.2 then contains the list of metabolic features that survive multiple testing correction when correcting for all clinical variables of interest (see Table 1 of the main text). All compounds in the latter table appear to be underexpressed in the AD group relative to the control group, expect for the Sphingomyelin SM(d18:1/20:1).

* Shared first authorship.
† Corresponding author.

TABLE S3.1. Differentially expressed metabolites when correcting for sex and age only.

Metabolite	Compound class	p-value	Adjusted p-value
2-Aminoadipic acid	Amines	6.085903e-08	1.399758e-05
Methyldopa	Amines	2.583401e-07	2.970911e-05
Valine	Amines	6.498148e-07	4.981913e-05
Tyrosine	Amines	1.697832e-06	9.762531e-05
Lysine	Amines	1.101781e-05	5.068193e-04
S-3-Hydroxyisobutyric acid	Organic acids	2.749051e-05	1.053803e-03
TG(54:6)	Lipids: Triglycerides	5.209223e-05	1.539997e-03
TG(48:0)	Lipids: Triglycerides	5.356511e-05	1.539997e-03
8-iso-PGF2a (15-F2t-IsoP)	Oxidative stress: Isoprostane	7.698703e-05	1.967446e-03
TG(50:4)	Lipids: Triglycerides	8.835841e-05	2.032243e-03
Methylmalonic acid	Organic acids	1.000548e-04	2.092055e-03
TG(52:4)	Lipids: Triglycerides	1.316578e-04	2.523440e-03
TG(48:2)	Lipids: Triglycerides	1.728144e-04	2.798978e-03
TG(50:3)	Lipids: Triglycerides	1.761190e-04	2.798978e-03
Leucine	Amines	1.825420e-04	2.798978e-03
TG(56:8)	Lipids: Triglycerides	2.126884e-04	2.886597e-03
TG(51:3)	Lipids: Triglycerides	2.317948e-04	2.886597e-03
TG(50:1)	Lipids: Triglycerides	2.324632e-04	2.886597e-03
TG(48:3)	Lipids: Triglycerides	2.384581e-04	2.886597e-03
TG(52:5)	Lipids: Triglycerides	2.827883e-04	3.182738e-03
TG(50:2)	Lipids: Triglycerides	2.905978e-04	3.182738e-03
TG(46:2)	Lipids: Triglycerides	3.557556e-04	3.610453e-03
TG(48:1)	Lipids: Triglycerides	3.610453e-04	3.610453e-03
TG(50:0)	Lipids: Triglycerides	4.318757e-04	4.138809e-03
TG(52:3)	Lipids: Triglycerides	5.484129e-04	5.045398e-03
TG(56:7)	Lipids: Triglycerides	8.909861e-04	7.881800e-03
Isoleucine	Amines	9.404737e-04	8.011442e-03
PGD2	Oxidative stress: Prostaglandins	9.789025e-04	8.040984e-03
LPC(18:1)	lipids: Lysophosphatidylcholine	1.210008e-03	9.506953e-03
TG(52:1)	Lipids: Triglycerides	1.240037e-03	9.506953e-03
SM(d18:1/20:1)	Lipids: Sphingomyelins	1.592411e-03	1.176462e-02
1-Methylhistidine	Amines	1.636817e-03	1.176462e-02
5-iPF2a VI	Oxidative stress: Isoprostane	2.014948e-03	1.404357e-02
2-hydroxybutyric acid	Organic acids	2.242011e-03	1.506750e-02
TG(54:5)	Lipids: Triglycerides	2.292880e-03	1.506750e-02
SM(d18:1/23:0)	Lipids: Sphingomyelins	2.583397e-03	1.607458e-02
TG(51:1)	Lipids: Triglycerides	2.639813e-03	1.607458e-02
TG(58:10)	Lipids: Triglycerides	2.655801e-03	1.607458e-02
TG(51:2)	Lipids: Triglycerides	2.892979e-03	1.706116e-02
TG(46:1)	Lipids: Triglycerides	3.683727e-03	2.118143e-02
LPA C14:0	Oxidative stress: Lyso-phosphatidic acid	4.021010e-03	2.255689e-02
3-Hydroxyisovaleric acid	Organic acids	4.778686e-03	2.616899e-02
TG(58:9)	Lipids: Triglycerides	5.031803e-03	2.691430e-02
Histidine	Amines	6.498831e-03	3.397116e-02
DG(36:3)	Lipids: Diacylglycerol	7.216737e-03	3.688554e-02
Phenylalanine	Amines	7.730540e-03	3.812873e-02
TG(52:2)	Lipids: Triglycerides	7.791523e-03	3.812873e-02
SM(d18:1/24:2)	Lipids: Sphingomyelins	8.365093e-03	4.008274e-02
PC(O-44:5)	Lipids: Plasmalogen Phosphatidylcholine	8.717054e-03	4.091679e-02
8,12-iPF2a IV	Oxidative stress: Isoprostane	9.768359e-03	4.493445e-02
TG(46:0)	Lipids: Triglycerides	1.042619e-02	4.702009e-02
Methionine	Amines	1.133966e-02	4.952164e-02
LPA C20:1	Oxidative stress: Lyso-phosphatidic acid	1.141151e-02	4.952164e-02

2. Classification Signature

The approach for the construction of classification signatures was described in in *Section 2.2.1* of *Supplementary Text 2*. Model performances can be found in Figure S3.1. The prediction model carrying the clinical variables only resulted in an AUC of .695. The model that used the Lasso for selection amongst the metabolites sorts a somewhat better classification performance, yielding an AUC of .746. The model that adds a (Lasso-based) selection of metabolites to the clinical variables then improves predictive performance along the full false positive rate range, sorting a AUC of .796. Table S3.3 contains the metabolites selected in the selection-amongst-metabolites-only situation. Table S3.4 then contains the metabolites selected in the selection-amongst-metabolites-whilst-clinical-variables-present situation. We see, for the compounds that also occur in the differential expression signature, that

TABLE S3.2. Differentially expressed metabolites when correcting for all clinical variables.

Metabolite	Compound class	p-value	Adjusted p-value
Tyrosine	Amines	1.293884e-05	0.001941711
2-Aminoadipic acid	Amines	2.118905e-05	0.001941711
8-iso-PGF2a (15-F2t-IsoP)	Oxidative stress: Isoprostane	2.647672e-05	0.001941711
Methyldopa	Amines	3.376889e-05	0.001941711
TG(54:6)	Lipids: Triglycerides	7.936780e-05	0.003650919
TG(56:8)	Lipids: Triglycerides	1.064855e-04	0.003864998
Valine	Amines	1.176304e-04	0.003864998
TG(50:4)	Lipids: Triglycerides	1.415809e-04	0.004070451
S-3-Hydroxyisobutyric acid	Organic acids	2.261572e-04	0.005413816
TG(52:4)	Lipids: Triglycerides	2.353833e-04	0.005413816
TG(51:3)	Lipids: Triglycerides	2.779465e-04	0.005697531
PGD2	Oxidative stress: Prostaglandins	2.972625e-04	0.005697531
Lysine	Amines	3.235788e-04	0.005724855
TG(52:5)	Lipids: Triglycerides	3.734782e-04	0.005825832
TG(56:7)	Lipids: Triglycerides	3.799456e-04	0.005825832
TG(48:3)	Lipids: Triglycerides	4.059878e-04	0.005836075
TG(50:3)	Lipids: Triglycerides	6.422663e-04	0.008689485
TG(46:2)	Lipids: Triglycerides	8.831594e-04	0.011284814
TG(52:3)	Lipids: Triglycerides	9.803097e-04	0.011866907
TG(58:10)	Lipids: Triglycerides	1.166642e-03	0.013416381
SM(d18:1/23:0)	Lipids: Sphingomyelins	1.259544e-03	0.013795004
TG(48:2)	Lipids: Triglycerides	1.442328e-03	0.014514992
TG(58:9)	Lipids: Triglycerides	1.451499e-03	0.014514992
TG(48:0)	Lipids: Triglycerides	1.516374e-03	0.014531917
5-iPF2a VI	Oxidative stress: Isoprostane	1.584192e-03	0.014574562
SM(d18:1/20:1)	Lipids: Sphingomyelins	1.899763e-03	0.016805592
Methylmalonic acid	Organic acids	2.206887e-03	0.018156831
3-Hydroxyisovaleric acid	Organic acids	2.210397e-03	0.018156831
Ornithine	Amines	3.068127e-03	0.024137545
TG(50:2)	Lipids: Triglycerides	3.148375e-03	0.024137545
Leucine	Amines	3.330067e-03	0.024706952
TG(54:5)	Lipids: Triglycerides	3.448227e-03	0.024784134
DG(36:3)	Lipids: Diacylglycerol	3.614235e-03	0.025190125
TG(48:1)	Lipids: Triglycerides	3.740068e-03	0.025300461
TG(50:0)	Lipids: Triglycerides	3.923739e-03	0.025784569
Phenylalanine	Amines	4.059826e-03	0.025937777
TG(50:1)	Lipids: Triglycerides	4.237719e-03	0.026342577
O-Acetylserine	Amines	4.959665e-03	0.030019024
8,12-iPF2a IV	Oxidative stress: Isoprostane	5.360289e-03	0.031324259
TG(51:2)	Lipids: Triglycerides	5.447697e-03	0.031324259
SM(d18:1/25:0)	Lipids: Sphingomyelins	7.146770e-03	0.039619151
NO2-aLA (C18:3)	Oxidative stress: Nitro-Fatty acid	7.234801e-03	0.039619151
LPA C18	Oxidative stress: Lyso-phosphatidic acid	7.416156e-03	0.039667810
LPA C14:0	Oxidative stress: Lyso-phosphatidic acid	8.866032e-03	0.046345167

the signs of their effects concur with the pattern of AD-associated under- and overexpression.

3. SOME REFLECTIONS

Of the 285 subjects included in the analyzes above a total of 37 had a CSF-biomarker status discordant with their clinical diagnosis. That is, these subjects were either clinically diagnosed with AD while their CSF-markers were normal (t-tau/Aβ_{42} $\leq$ 0.52) or clinically diagnosed as normal while their CSF-markers indicated AD (t-tau/Aβ_{42} $>$ 0.52). The disease status of these subjects is thus unsure as it is unclear which diagnosis (clinical or biomarker-based) should take precedence. Hence, the main analyzes revolved around those cases in which the clinical and biomarker-based diagnoses were concordant. Below we reflect on the findings above in relation to the main analyzes.

The larger sample size implies that we have more power in finding differentially expressed metabolites. Hence, Tables S3.1 and S3.2 list more metabolites than their corresponding tables in *Supplementary Text 2* (*S2.3* and *S2.2*) and the Main text (Table 3). As the increase in power comes from subjects whose disease status is

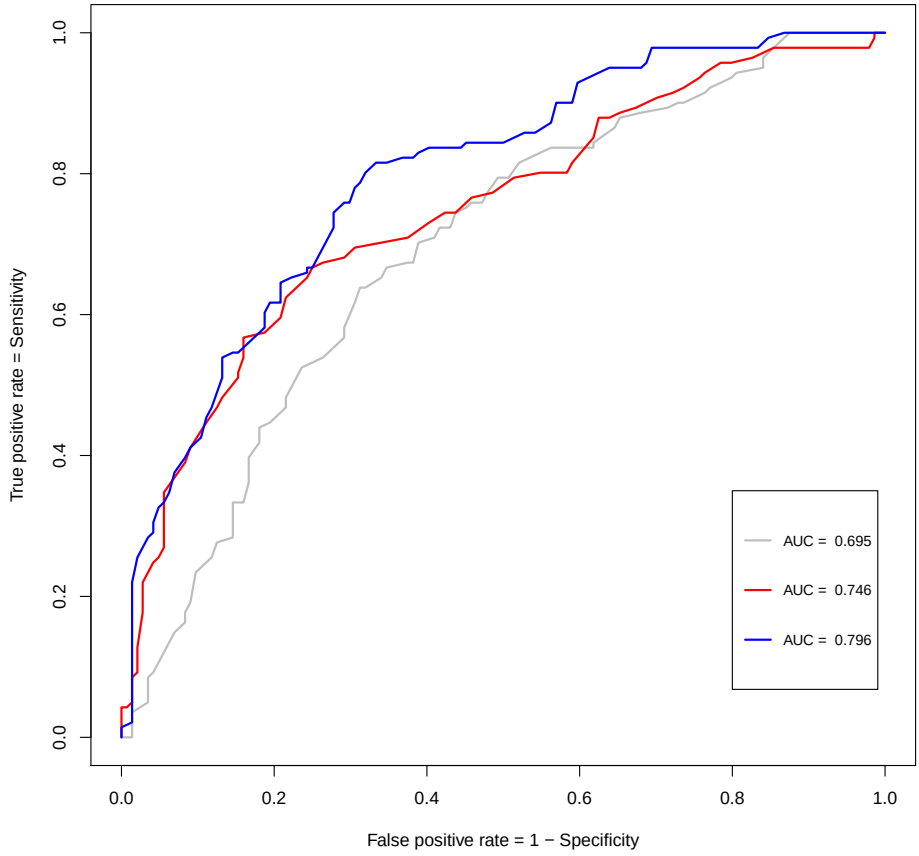

FIGURE S3.1. ROC curves for the classification models. The grey line represents the ROC curve for the unpenalized logistic regression model that entertains the clinical characteristics only. The red line represents the ROC curve for the logistic model in which the Lasso performed variable selection amongst the metabolites (and that does not consider the clinical characteristics). The blue line represents the ROC curve of the logistic model in which the clinical characteristics are present while the Lasso may select amongst the metabolites.

somewhat unsure, the results in Tables S3.1 and S3.2 are somewhat tentative. These results are, however, concordant with the results from the main analyzes, in the sense that they overlap, i.e.: The metabolites listed in Table *S2.3* of *Supplementary Text 2* (see also column 3 of Table 3 of the Main text) are a proper subset (except for Proline, PC(O-34:1), LPC(20:4), SM(18:1/16:0), and Ornithine) of the metabolites listed in Table S3.1; and the metabolites listed in Table *S2.2* of *Supplementary Text 2* (see also column 4 of Table 3 of the Main text) are a proper subset of the metabolites listed in Table S3.2.

TABLE S3.3. Selected metabolites and parameter estimates when considering only metabolites as potential predictors.

Metabolite	Compound class	$\hat{\beta}$
LPC(18:1)	lipids: Lysophosphatidylcholine	0.3775846437
Methyldopa	Amines	−0.3280999265
PGD2	Oxidative stress: Prostaglandins	−0.3136491690
NO2-aLA (C18:3)	Oxidative stress: Nitro-Fatty acid	−0.2042163853
O-Acetylserine	Amines	−0.1811852503
SM(d18:1/23:0)	Lipids: Sphingomyelins	−0.1711821392
Tyrosine	Amines	−0.1588698635
SM(d18:1/20:1)	Lipids: Sphingomyelins	0.1375561855
5-iPF2a VI	Oxidative stress: Isoprostane	−0.1296136239
Glyceric acid	Organic acids	−0.1271578713
PC(O-34:3)	Lipids: Plasmalogen Phosphatidylcholine	−0.1154877989
TG(54:6)	Lipids: Triglycerides	−0.1099736486
8,12-iPF2a IV	Oxidative stress: Isoprostane	−0.1091228644
Methylmalonic acid	Organic acids	−0.1021936840
TG(48:2)	Lipids: Triglycerides	−0.0880997011
8-iso-PGF2a (15-F2t-IsoP)	Oxidative stress: Isoprostane	−0.0859696149
LPA C16	Oxidative stress: Lyso-phosphatidic acid	−0.0712895571
2,3-dinor-8-iso-PGF2a	Oxidative stress: Isoprostane	−0.0660850151
TG(O-50:0)	Lipids: Triglycerides	0.0631492242
TG(48:0)	Lipids: Triglycerides	−0.0594940757
PC(O-38:6)	Lipids: Plasmalogen Phosphatidylcholine	−0.0573684699
LPA C14:0	Oxidative stress: Lyso-phosphatidic acid	−0.0462058983
Serine	Amines	0.0427658066
Putrescine	Amines	−0.0402777537
LPA C22:4	Oxidative stress: Lyso-phosphatidic acid	0.0336067651
Lysine	Amines	−0.0308966941
3-Methoxytyramine	Amines	−0.0171013575
cLPA C18:1	Oxidative stress: Cyclic-lyso-phosphatidic acid	−0.0127374834
PE(O-38:5)	Lipids: Plasmalogen Phosphatidylethanolamine	−0.0079640017
2-Aminoadipic acid	Amines	−0.0069680882
Carnosine	Amines	0.0022778247
LPC(20:4)	Lipids: Lysophosphatidylcholine	0.0006342159

TABLE S3.4. Selected metabolites and parameter estimates when considering metabolites as potential predictors on top of the clinical variables.

Metabolite	Compound class	$\hat{\beta}$
PGD2	Oxidative stress: Prostaglandins	-0.48204926
SM(d18:1/20:1)	Lipids: Sphingomyelins	0.33329421
O-Acetylserine	Amines	-0.28384663
Methyldopa	Amines	-0.27782490
8-iso-PGF2a (15-F2t-IsoP)	Oxidative stress: Isoprostane	-0.25085235
NO2-aLA (C18:3)	Oxidative stress: Nitro-Fatty acid	-0.24269712
Methylmalonic acid	Organic acids	-0.22969766
Tyrosine	Amines	-0.17320260
TG(52:4)	Lipids: Triglycerides	-0.15033254
Gamma-glutamylalanine	Amines	0.13963417
Putrescine	Amines	-0.13769196
8,12-iPF2a IV	Oxidative stress: Isoprostane	-0.13656275
TG(51:3)	Lipids: Triglycerides	-0.13258986
SM(d18:1/23:0)	Lipids: Sphingomyelins	-0.11398902
Citrulline	Amines	-0.10005829
Glyceric acid	Organic acids	-0.09690281
LPC(18:1)	lipids: Lysophosphatidylcholine	0.09252004
PC(O-34:3)	Lipids: Plasmalogen Phosphatidylcholine	-0.09149649
3-Hydroxyisovaleric acid	Organic acids	-0.08574230
5-iPF2a VI	Oxidative stress: Isoprostane	-0.08026972
LPA C14:0	Oxidative stress: Lyso-phosphatidic acid	-0.07726029
2,3-dinor-8-iso-PGF2a	Oxidative stress: Isoprostane	-0.07467916
TG(58:10)	Lipids: Triglycerides	-0.06409276
Cysteine	Amines	0.04297397
Carnosine	Amines	0.03408938
PC(O-34:2)	Lipids: Plasmalogen Phosphatidylcholine	-0.03066565
SM(d18:1/25:0)	Lipids: Sphingomyelins	-0.01019553

The main model of interest – selecting metabolites on top of clinical predictors – has a concordant performance on the full and CSF-confirmed data: in both instances the model sorts an AUC of approximately .79. Moreover, most of the

(top) compounds in Table *S2.5* of *Supplementary Text 2* appear as selected (top) compounds in Table S3.4. The insecurity regarding the status of subjects with discordant diagnoses is, however, reflected somewhat in the classification signatures. For example, if we would take, for the main model of interest, the logistic cut-off at the optimal cut-off in terms of accuracy (.42, as determined by 10-fold CV), then we see that those clinically diagnosed with AD while having a normal CSF-status (14) tend to be predominantly classified as AD cases while those clinically diagnosed as normal while having an AD CSF-status (23) seem to be randomly classified as either AD or control cases.

Insecurity regarding the true status of those with a discordant CSF-biomarker status implies that the class-specific (AD or SCD) samples are heterogeneous. Heterogeneity can lead to the dilution of partial correlations and, hence, may hamper network extraction [S3.1]. Results on the regulatory signature are, for reasons of brevity, not included.

References

[S3.1] van Wieringen, W.N., & van der Vaart, A.W. (2015). Transcriptomic heterogeneity in cancer as a consequence of dysregulation of the gene-gene interaction network. Bulletin of Mathematical Biology, 77: 1768–1786.

(Francisca A. de Leeuw) Alzheimer Center and Dept. of Neurology, Amsterdam Neuroscience, VU University Medical Center Amsterdam, Amsterdam, The Netherlands; and Dept. of Clinical Chemistry, VU University Medical Center Amsterdam, Amsterdam, The Netherlands
E-mail address: f.deleeuw@vumc.nl

(Carel F.W. Peeters) Dept. of Epidemiology & Biostatistics, Amsterdam Public Health research institute, VU University medical center Amsterdam, Amsterdam, The Netherlands
E-mail address: cf.peeters@vumc.nl

(Maartje I. Kester) Alzheimer Center and Dept. of Neurology, Amsterdam Neuroscience, VU University Medical Center Amsterdam, Amsterdam, The Netherlands
E-mail address: m.kester@vumc.nl

(Amy C. Harms) Division of Analytical Biosciences, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands
E-mail address: a.c.harms@lacdr.leidenuniv.nl

(Eduard A. Struys) Dept. of Clinical Chemistry, VU University medical center Amsterdam, Amsterdam, The Netherlands
E-mail address: E.Struys@vumc.nl

(Thomas Hankemeier) Division of Analytical Biosciences, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands
E-mail address: hankemeier@lacdr.leidenuniv.nl

(Herman W.T. van Vlijmen) Discovery Sciences, Janssen Research and Development, Beerse, Belgium; and Division of Medicinal Chemistry, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands
E-mail address: hvvlijme@its.jnj.com

(Sven J. van der Lee) Genetic Epidemiology Unit, Dept. of Epidemiology, Erasmus MC, Rotterdam, The Netherlands; and Alzheimer Center, VU University Medical Center Amsterdam, Amsterdam, The Netherlands
E-mail address: s.j.vanderlee@vumc.nl

(Cornelia M. van Duijn) Genetic Epidemiology Unit, Dept. of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
E-mail address: c.vanduijn@erasmusmc.nl

(Philip Scheltens) Alzheimer Center and Dept. of Neurology, Amsterdam Neuroscience, VU University Medical Center Amsterdam, Amsterdam, The Netherlands
E-mail address: p.scheltens@vumc.nl

(Ayşe Demirkan) Genetic Epidemiology Unit, Dept. of Epidemiology, Erasmus MC, Rotterdam, The Netherlands; and Dept. of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
E-mail address: a.demirkan@erasmusmc.nl

(Mark A. van de Wiel) Dept. of Epidemiology & Biostatistics, Amsterdam Public Health research institute, VU University medical center Amsterdam, Amsterdam, The Netherlands; and Dept. of Mathematics, VU University Amsterdam, Amsterdam, The Netherlands
E-mail address: mark.vdwiel@vumc.nl

(Wiesje M. van der Flier) Alzheimer Center and Dept. of Neurology, Amsterdam Neuroscience, VU University Medical Center Amsterdam, Amsterdam, The Netherlands; and Dept. of Epidemiology & Biostatistics, Amsterdam Public Health research institute, VU University medical center Amsterdam, Amsterdam, The Netherlands
E-mail address: WM.vdFlier@vumc.nl

(Charlotte E. Teunissen) Neurochemistry Laboratory and Biobank, Dept. of Clinical Chemistry, Amsterdam Neuroscience, VU University medical center Amsterdam, Amsterdam, The Netherlands
E-mail address: c.teunissen@vumc.nl