Elemental fingerprint as a cerebrospinal fluid biomarker for the diagnosis of Parkinson's disease

- Supplementary material -

Fabian Maass[1], Bernhard Michalke[3], Andreas Leha[4], Matthias Börger[1], Inga Zerr[1,6], Jan-Christoph Koch[1,2], Lars Tönges[5], Mathias Bähr[1,2] and Paul Lingor[1,2,*]

[1] Department of Neurology, University Medical Center Goettingen, Goettingen D-37075, Germany

[2] Cluster of Excellence Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB), Goettingen D-37075, Germany

[3] Research Unit Analytical BioGeoChemistry, German Research Center for Environmental Health, Helmholtz Zentrum Muenchen, Neuherberg D-85764, Germany

[4] Department of Medical Statistics, University Medical Center Goettingen, Goettingen D-37075, Germany

[5] Department of Neurology, Ruhr-University Bochum, St. Josef-Hospital, Bochum D-44791, Germany

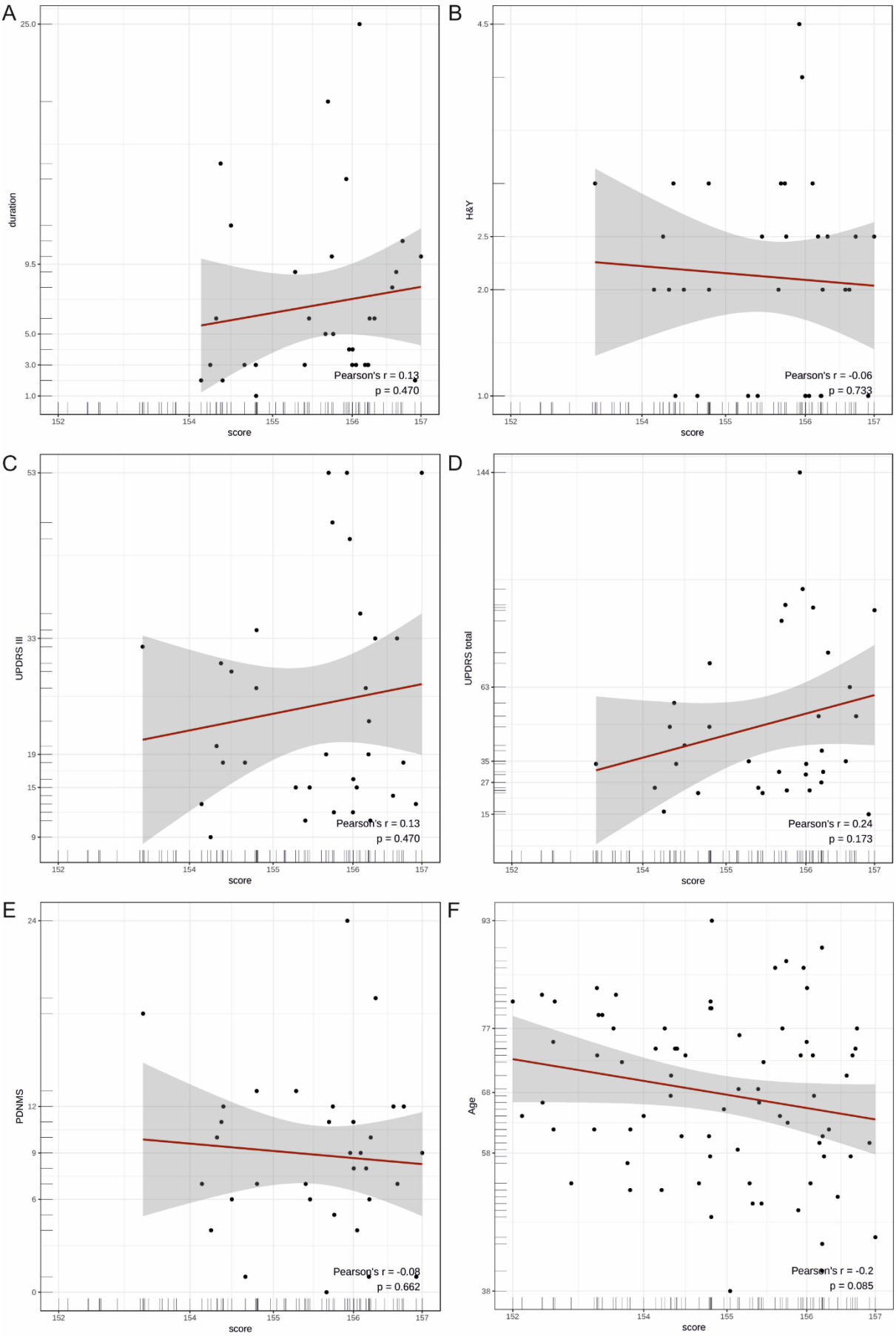
[6] DZNE, German Center for Neurodegenerative Diseases Goettingen, Goettingen D-37075, Germany

Supplementary figure legends

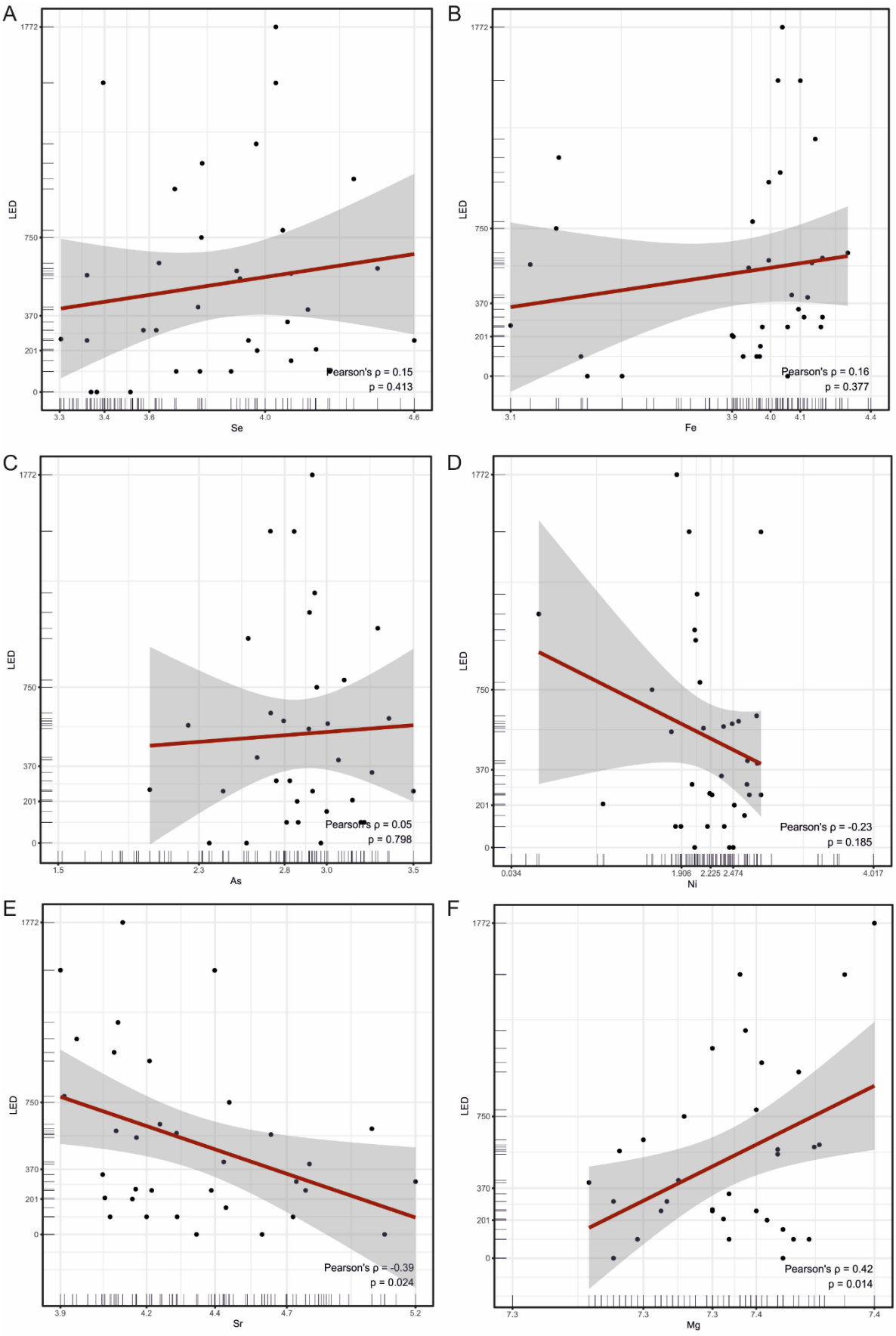
Supplementary Figure S1: Summary of the clinical parameters. The grouping (Control vs Parkinson's disease patients) was modeled using a logistic regression model based on the concentration levels of the elements in the final feature set as predictors. Patients were assigned elemental scores using the resulting model. The model shows no correlation between element concentration scores and clinical parameters (MDS-UPDRS total and MDS-UPDRS part III, PD-NMS, H&Y stage, disease duration and age).

Supplementary Figure S2: Correlation of LED to the six final elements. No significant correlations remain after adjustment for multiple testing. LED = levodopa equivalent dose.

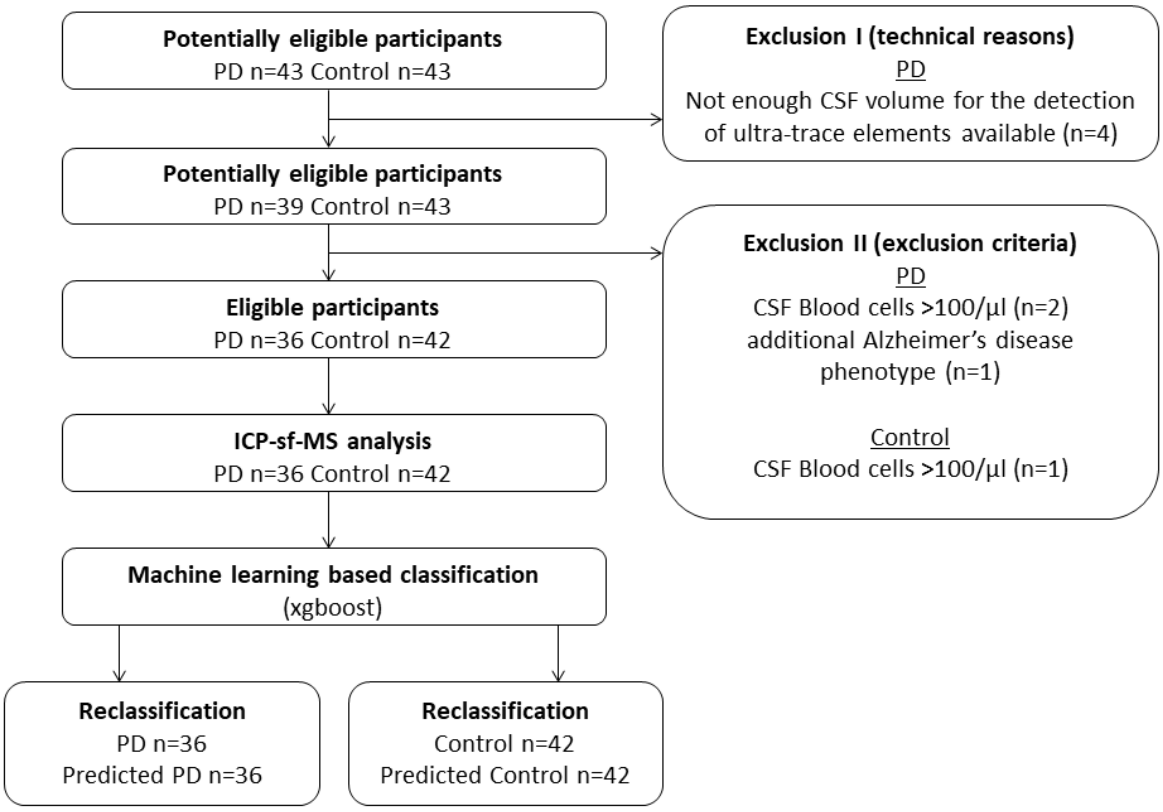
Supplementary Figure S3: Diagram showing flow of patients through the study. PD = Parkinson's disease.



Supplementary Figure S1



Supplementary Figure S2



Supplementary Figure S3

Elements by ICP-OES	measured spectral line , nm	LOQ calculated for undiluted, native CSF, µg/L
Al	167.078	4
B	249.773	8
Ba	455.404	11
Ca	183.801	48
Cu	324.754	8
Fe	259.941	8
K	766.491	48
Li	670.77	2
Mg	279.079	48
Mn	257.611	2
Na	589.592	48
P	177.495	80
S	180.731	80
Sr	407.771	2
Ti	334.941	8
Zn	213.856	4

Elements by ICP-sf-MS	isotope measured	res-mode	LOQ calculated for undiluted, native CSF, ng/L
As	75	high	4
Cd	114	low	2
Co	59	medium	8
Cr	52	high	8
Hg	202	low	4
Mn	55	medium	2
Mo	98	medium	4
Ni	60	medium	4
Pb	208	low	2
Sn	120	low	2
Se	77	high	40
Ti	47	high	4
V	51	medium	4

Supplementary Table S1: Technical information about the Limit of quantification (LOQ) for any measured element using ICP-OES and ICP-sf-MS.

	LED [mg]
1	0
2	0
3	100
4	567
5	257
6	1110
7	750
8	207
9	626
10	1501
11	300
12	250
13	340
14	300
15	550
16	0
17	785
18	100
19	400
20	985
21	1034
22	250
23	152
24	1237
25	588
26	1500
27	250
28	100
29	1772
30	201
31	412
32	575
33	1204
34	100
35	600
36	400

Supplementary Table S2: Table showing levodopa equivalent dose (LED) for the PD patients.

Supplementary methods – ICP-OES and ICP-sf-MS

Reference standards:

Plasma (simulating protein content of CSF) and urine (simulating the high salt content of CSF) were used.

- Plasma ClinCal 9985 (Recipe, München, Germany)

- Urine Clicheck Control 8847-8849 (Recipe, München, Germany)

Supplementary methods - Machine Learning Algorithms

The machine learning algorithms were trained through the blkbox [1] R-package. For some algorithms blkbox in turn relies on the caret [2] R-package, which by default offers a grid search for the important tuning parameters in each model. Here we list the applied machine learning algorithms with the values of their respective most important tuning parameters:

- randomforest [3]
  500 trees, mtry = 4
- kknn [4]
  k = 7, distance = 3, kernel = optimal, scale = T
- bartmachine [5]
  50 trees, num_burn_in = 250, num_iterations_after_burn_in = 1000, beta = 2, k = 2, q = 0.9, nu = 3, mtry = 4
- party [6]
  500 trees, mtry = 4, teststat = quad, testtype = Univ, replace = FALSE
- elastic net regularized logistic regression : glmnet (through caret) [7]
  alpha = <grid search>, lambda = <grid search>
- Nearest Shrunken Centroids: PamR (through caret) [8]
  threshold: <grid search>
- Neural net: nnet (through caret) [9]
  MaxNWts = 1000000, size = <grid search>, decay = <grid search>
- SVM [10]
  linear kernel, gamma = 1/18, scale = TRUE
- xgboost [11]
  type = logistic, max.depth = 4, eta = 1, nround = 10

Supplementary methods - references

[1] Zachary Davies and Boris Guennewig (2016). blkbox: Data Exploration with Multiple Machine Learning Algorithms. R package version 1.0. https://CRAN.R-project.org/package=blkbox

[2] Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang and Can Candan. (2016). caret: Classification and Regression Training. R package version 6.0-70. https://CRAN.R-project.org/package=caret

[3] A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22.

[4] Klaus Schliep and Klaus Hechenbichler (2016). kknn: Weighted k-Nearest Neighbors. R package version 1.3.1. https://CRAN.R-project.org/package=kknn

[5] Adam Kapelner, Justin Bleich (2016). bartMachine: Machine Learning with Bayesian Additive Regression Trees. Journal of Statistical Software, 70(4), 1-40. doi:10.18637/jss.v070.i04

[6] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis and Torsten Hothorn (2007). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. BMC Bioinformatics, 8(25). URL http://www.biomedcentral.com/1471-2105/8/25.

[7] Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL http://www.jstatsoft.org/v33/i01/.

[8] T. Hastie, R. Tibshirani, Balasubramanian Narasimhan and Gil Chu (2014). pamr: Pam: prediction analysis for microarrays. R package version 1.55. https://CRAN.R-project.org/package=pamr

[9] Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

[10] Chang, Chih-Chung and Lin, Chih-Jen:_LIBSVM: a library for Support Vector Machines_http://www.csie.ntu.edu.tw/~cjlin/libsvm

[11] Tianqi Chen, Tong He and Michael Benesty (2016). xgboost: Extreme Gradient Boosting. R package version 0.4-4. https://CRAN.R-project.org/package=xgboost

[12] Helsel, DR. Nondectects and Data Analysis; Statistics for censored environmental data. John Wiley and Sons, USA, NJ. 2005

[13] Martin-Fernandez JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Model-based replacement of rounded zeros in compositional data: classical and robust approaches. Computational Statistics & Data Analysis 2012; 56: 2688-2704.