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

Running Title: Elements as Parkinson's disease biomarkers

Fabian Maass[a], Bernhard Michalke[c], Andreas Leha[d], Matthias Boerger[a], Inga Zerr[a,f], Jan-Christoph Koch[a,b], Lars Tönges[e], Mathias Bähr[a,b] and Paul Lingor[a,b*]

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

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

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

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

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

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

Keywords

Parkinson's disease; Biomarker; Cerebrospinal fluid; Selenium; Iron

***Corresponding author:**

Prof. Dr. Paul Lingor

Klinik für Neurologie, Universitätsmedizin Göttingen

Robert-Koch-Straße 40, DE 37075 Göttingen

Phone: +49 551 39 66356

Email: paul.lingor@med.uni-goettingen.de

Abbreviations

AUC = area under the curve

AUROC = area under the receiver operating characteristic curve

FWER = family-wise error rate

ICP-OES = inductively coupled plasma optical emission spectrometry

ICP-sf-MS = inductively coupled plasma-sector field mass spectrometry

LOQ = limit of quantification

ML = maximum likelihood

PD = Parkinson's disease

PDNMS = Parkinson's disease non-motor symptoms questionnaire

UPDRS = Unified Parkinson's Disease Rating Scale

Abstract

The diagnosis of Parkinson's disease (PD) still lacks objective diagnostic markers independent of clinical criteria. CSF samples from 36 PD and 42 age-matched control patients were subjected to inductively coupled plasma-sector field mass spectrometry and a total of 28 different elements were quantified. Different machine learning algorithms were applied to the dataset to identify a discriminating set of elements yielding a novel biomarker signature. Using 19 stably-detected elements, the extreme gradient tree boosting model showed the best performance in the discrimination of PD and control patients with high specificity and sensitivity (78.6% and 83.3%, respectively), re-classifying the training data to 100%. The 10 times 10-fold cross validation yielded a good AUROC of 0.83. Arsenic, magnesium and selenium all showed significantly higher mean CSF levels in the PD group compared to the control group ($p = 0.01$, $p = 0.04$ and $p = 0.03$). Reducing the number of elements to a discriminating minimum, we identified an elemental cluster (Se, Fe, As, Ni, Mg, Sr), which most importantly contributed to the sample discrimination. Selenium was identified as the element with the highest impact within this cluster directly followed by iron. After prospective validation, this elemental fingerprint in the CSF could have the potential to be used as independent biomarker for the diagnosis of PD. Next to their value as a biomarker, this data also argues for a prominent role of these highly discriminating six elements in the pathogenesis of PD.

Introduction

The diagnosis of idiopathic Parkinson's disease (PD) is still based on purely clinical criteria, which were recently refined by the Movement Disorder Society (Postuma et al., 2015). A current meta-analysis on the accuracy of the clinical diagnosis of PD shows that still only eight of ten patients are diagnosed correctly (Rizzo et al., 2016). Most studies evaluating the value of alpha-synuclein in the CSF as a potential biomarker show that its mean levels are lower in PD compared to controls, however, individual alpha-synuclein levels lack discrimination power and specificity (Mollenhauer et al., 2011). As an alternative approach other features contributing to the pathogenesis of PD could be used as potential biomarkers. The levels of several bioelements are shown to be altered in PD patients' brains. For example, an increase in iron content and reduction in copper was demonstrated in the Parkinsonian substantia nigra (Davies et al., 2014; Dexter et al., 1989). While the idea of metal-induced oxidative stress in neurodegeneration is known for decades, current studies also link alpha-synuclein pathology and bioelement dysregulation, since alpha-synuclein aggregation can also be triggered by iron, copper, zinc, manganese and arsenic (Cholanians et al., 2016; Uversky et al., 2001). Alzheimer's disease as another example also shows bioelemental dysregulations, which might be associated with pathological protein aggregation and which are reflected by changes in CSF levels of elements like aluminum (Hozumi et al., 2011; Virk and Eslick, 2015). However, results on bioelements in the CSF of PD patients are often not reproducible potentially due to variable detection methods, insufficiently characterized patients and insufficiently controlled blood contamination. Whereas the levels of more common elements, as iron, copper and manganese, were analyzed by several groups yielding inconclusive results, there is insufficient data on less abundant elements (Jiménez-Jiménez et al., 2014). In fact, due to the plethora of mechanisms involved in the pathogenesis of PD, it is unlikely that one single element will emerge as a diagnostic biomarker. We therefore studied the concentrations of up to 28 elements in the CSF of PD patients and age-matched controls. Our data yields an elemental fingerprint, which allows differentiating PD from controls with the potential to become a novel biomarker signature for PD.

Material and methods

Participants

36 PD patients were consecutively selected from the CSF Biobank of the Department of Neurology, University of Goettingen, Germany. Only samples with a sufficiently high amount of CSF required for elemental analysis were included. PD patients (based on UK Brain Bank criteria, compliant with MDS criteria) were arbitrarily recruited from the available patient pool of the out- and in-patient clinics.

Additional clinical parameters increasing diagnostic accuracy were available (32/36 MRI, 25/36 DatScan, 20/36 substantia nigra ultrasound). The majority of the patients belong to a cohort, which has regular scheduled follow-up assessments in the Parkinson's disease outpatient clinics (follow up 13-40 months). 33 patients were under anti-Parkinsonian medication including levodopa, dopamine agonists, amantadine, MAO and COMT inhibitors while 3 patients were drug naïve.

All PD patients underwent a thorough clinical examination and history taking, assessment of motor and non-motor symptoms (MDS-UPDRS, PD-NMS) and routine blood work. Disease duration was defined as the time span since the awareness of the first motor-symptoms. The levodopa equivalent dose was calculated according to Tomlinson et al., 2010.

In addition, 42 age-matched controls without signs of neurodegenerative, neuroinflammatory or acute ischemic central nervous diseases were included, in most cases having a lumbar puncture for exclusion diagnosis (e.g. headache, dizziness, functional disorders). No specific method of patient randomization was employed. CSF samples were collected in the scope of a local monocentric research project. Except for local ethics approval, no specific trial pre-registration was performed. Sample sizes were chosen in similarity to previously published studies (Jiménez-Jiménez et al., 2014). Due to the lack of pre-existing learning curves for our machine-learning algorithms, there was no definite sample size predetermination available, which would be required to realize more precise sample size considerations for classification models (Mukherjee et al., 2003). A permission of the local ethics committee has been obtained prior to the initiation of the study (Ethics committee of the University Medicine Göttingen, No. 13/11/12). Written consent was provided by all patients or care givers. The study conforms with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

CSF-Procedures

All patients underwent a lumbar puncture, where 10 ml of CSF were collected in polypropylene tubes and treated in a standard procedure. Routine testing included WBC and RBC count, protein and lactate levels. CSF was immediately centrifuged (2000 g; 20 min; 4°C) within 30 minutes of collection and frozen at -80°C until further analysis. Patients were non-fasting. Patients with severe hepatic or renal failure, and those taking mineral supplements or chelating agents were excluded. Samples with a RBC > 100 / µl in the routine testing sample were excluded.

Sample preparation and analysis by ICP-OES and ICP-sf-MS

Frozen aliquots were transported overnight on dry ice to the Research Unit Analytical BioGeoChemistry of the Helmholtz Zentrum, Muenchen for further investigation. The samples were thawed slowly at 4°C before being diluted 1:4 with Milli-Q water. The diluted samples were directly used for element measurements. An ICP-OES „Spectro Ciros Vision“ system (SPECTRO Analytical Instruments GmbH & Co. KG, Kleve, Germany) was used for element determination. An ELEMENT 2, Thermo-Electron (Bremen, Germany) ICP-sf-MS instrument was employed for determination of elements which were below the LOD from ICP-OES. The determination method had been validated previously by regular laboratory intercomparison studies. Certified single element standards were used after every ten measurements (see supplementary methods). Additionally, reference materials were analyzed together with the sample batch. The investigators who performed the elemental analysis (ICP-OES / ICP-sf-MS) were blinded during the entire analytic process and data processing by working with pseudonymized samples, which did not permit to draw conclusions about the group affiliation.

Sample analysis by ICP-OES

Sample introduction was carried out using a peristaltic pump connected to a Meinhard nebulizer with a cyclonic spray chamber. Measured spectral element lines (nm): Ba: 455.404, Ca: 183.801, Fe: 259.941, Li: 670.770, Mg: 279.079, Na: 589.592, P: 177.495, S: 180.731, Si: 251.612, Sr: 407.771, Zn: 213.856. The RF power was set to 1400 W, the plasma gas was 13 L Ar /min, whereas the nebulizer gas was approximately 0.6 L Ar/min after daily optimization.

Sample analysis by ICP-sf-MS

^{103}Rh was administered to each sample at a concentration of 1 µg/L as internal standard. Sample introduction was carried out using a peristaltic pump connected to a Seaspray nebulizer with a cyclonic spray chamber. The RF power was set to 1300 W, the plasma gas was 15 L Ar /min, whereas the nebulizer gas was approximately 0.9 L Ar/min after daily optimization. Measured element isotopes: ^{27}Al, ^{75}As, ^{114}Cd, ^{140}Ce ^{59}Co, ^{52}Cr, ^{63}Cu, ^{202}Hg, ^{127}I, ^{55}Mn, ^{98}Mo, ^{60}Ni, ^{208}Pb, ^{77}Se, ^{120}Sn, ^{48}Ti, ^{204}Tl, ^{51}V.

Statistical analysis

Only elements with more than 60 measurements within the collective group of 78 patients were considered. Measurements were log transformed and values below the limit of quantification were imputed using model-based robust Expectation-Maximization and data imputed this way was used for all analyses if not stated otherwise. The concentrations of each element were compared using regression by maximum likelihood (ML) estimation for left-censored data are given, which specifically handles left-censored data (see Supplementary Methods). As reference values randomly imputed data were compared using t-tests. Resulting p-values were adjusted using Holm's procedure to control the family-wise error rate (FWER) at a 0.05 level.

A heat map was constructed using correlation as measurement of similarity. Different machine learning methods were applied to the data (see Supplementary Methods). The performance was assessed via 10 times 10-fold cross validation. The resulting estimates for the area under the ROC curve were visualized. The consensus specificity and sensitivity across the 10 repeats was calculated according to the Youden Index. Random forest and extreme gradient tree boosting models have been trained on all available test data. A shared scatter plot was constructed showing the interaction of Se and Fe.

The AUC results from the 10 times repeated 10-fold cross validation are used as a feature selection criterion to determine the optimal set of elements for a good classifier. A step-wise procedure removing the least informative element one after the other was applied. To choose the final feature set, starting with the last remaining element, a bigger feature set was

considered if the AUC increased or if the AUC decreased less than 0.01 for only one bigger feature set. The feature importance – measured as information gain – for each element of the optimal set was quantified via 10 times 10-fold cross validation of a classifier trained on those elements only. The performance of the feature selection proceeded classification was assessed by yet another 10 times repeated 10-fold cross validation that included the feature selection. Demographical characteristics were compared between control patients and patients with PD using t-test and chi-square-test. Correlations between the concentration levels and clinical data (disease duration, age, levodopa equivalent dose, MDS-UPDRS part III and sum score, MDS-PDNMS) were tested using Pearson product moment after adjusting for multiple testing according to Bonferroni. Using a logistic regression model, the grouping (PD vs. control patients) was modeled based on the concentration levels of the elements in the final feature set as predictors. Patients were assigned elemental scores using the resulting model (Supplementary Figure S1).

Results

Demographics

Patients in the PD group were 67 ± 11.0 years old, whereas control patients were 65.5 ± 13.1 years old (not significantly different $p > 0.05$). Patients of different disease durations were included with a mean disease duration of 5.0 ± 5.5 years. There was no significant difference in sex distribution between the two groups ($p > 0.05$) (Table 1).

	Control	**PD**
Number of patients	42	36
Age, years	65,5 (13.1) 38-93	67 (11.0) 45-87
Male / female (% female)	24/18 (42.9)	24/12 (50)
Disease duration, years	n.a.	5.0 (5.5) 1-25
MDS-UPDRS total	n.a.	35 (29.8) 15-144
MDS-UPDRS III	n.a.	19 (13.0) 9-53
MDS-PDNMS	n.a.	9 (5.3) 0-24

Table 1: Demographical and clinical characteristics of the study population. Data on age, disease duration and clinical scores is presented as median, (standard deviation), range. PD = Parkinson's disease.

Detection of the elemental profile in the CSF

A total of 28 elements were quantified in the CSF samples (Table 2). Ba, Cd, Ce, Hg, Li, Sn, Tl, and V had measurements over the limit of quantification (LOQ) for less than 60 patients and have been eliminated from the analysis. In addition, iodine was excluded due to extremely high spread of measured values, which most likely was attributed to external factors, such as unreported dietary supplementation, application of radiographic contrast agents or antiseptic agents.

CSF levels of arsenic, magnesium and selenium are significantly higher in Parkinson's disease

The elements arsenic (control 585 ± 630 ng/l; PD 884 ± 634 ng/l; p = 0.01), magnesium (control 21 ± 1.2 mg/l; PD 22 ± 1.3 mg/l; p = 0.04) and selenium (control 5.9 ± 6.6 µg/l; PD 9.4 ± 7.6 µg/l; p = 0.03) all show significantly higher mean CSF levels in the PD group compared with the control group after multiple adjustments (Table2, Figure 1a).

Element	Mean ± sd		Median (min; max)		p value			
	Control	PD	Control	PD	raw	adjusted	Value censored MLE	Value censored MLE adjusted
Al [ng/l]	2.4 ± 0.34	2.4 ± 0.33	2.4 (1.9; 3.4)	2.4 (1.8; 3.4)	0.69	1.00	0.68	1.00
As [ng/l]	2.5 ± 0.56	2.8 ± 0.32	2.4 (1.5; 3.3)	2.9 (2; 3.5)	< 0.01	0.01	< 0.01	0.01
Ba [µg/l]	3.1 ± 0	3.1 ± 0.011	3.1 (3.1; 3.1)	3.1 (3.1; 3.2)	0.90		0.29	
Ca [mg/l]	7.6 ± 0.028	7.6 ± 0.019	7.6 (7.5; 7.7)	7.6 (7.5; 7.6)	0.95	1.00	0.95	1.00

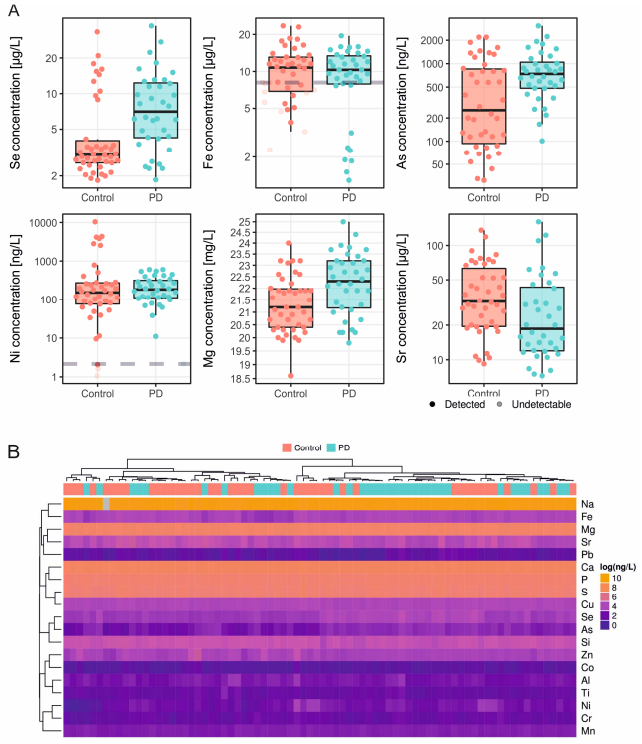
Cd [ng/l]	0.92 ± 0.94	0.3 ± 0.86	0.63 (-1.4; 3.1)	0.2 (-2.4; 2.4)	< 0.01		< 0.01	
Ce [ng/l]	0.48 ± 0.18	0.42 ± 0.02	0.42 (0.42; 1.3)	0.42 (0.42; 0.54)	0.64		0.03	
Co [ng/l]	1 ± 0.38	0.97 ± 0.33	0.93 (0.46; 2.1)	1 (0.46; 1.6)	0.53	1.00	0.56	1.00
Cr [ng/l]	1.7 ± 0.29	1.7 ± 0.32	1.7 (1.1; 2.4)	1.7 (1; 2.4)	0.75	1.00	0.74	1.00
Cu [µg/l]	4.1 ± 0.098	4.1 ± 0.1	4.1 (3.9; 4.3)	4.1 (3.9; 4.3)	0.17	1.00	0.16	1.00
Fe [µg/l]	4.1 ± 0.17	3.9 ± 0.33	4.1 (3.6; 4.4)	4 (3.1; 4.3)	0.84	1.00	0.30	1.00
Hg [ng/l]	0.92 ± 0.63	0.95 ± 0.68	0.7 (0.7; 3)	0.7 (0.7; 3.1)	0.68		0.81	
Li [µg/l]	3.1 ± 0	3.2 ± 0.51	3.1 (3.1; 3.1)	3.1 (3.1; 6.2)	0.02		0.03	0.02
Mg [mg/l]	7.3 ± 0.024	7.3 ± 0.026	7.3 (7.3; 7.4)	7.3 (7.3; 7.4)	< 0.01	0.05	< 0.01	0.04
Mn [ng/l]	2.6 ± 0.11	2.6 ± 0.11	2.6 (2.4; 2.8)	2.6 (2.5; 2.9)	0.79	1.00	0.79	1.00
Na [g/l]	9.6 ± 0.054	9.6 ± 0.029	9.6 (9.4; 9.7)	9.6 (9.5; 9.7)	0.12	1.00	0.12	1.00
Ni [ng/l]	2.2 ± 0.83	2.2 ± 0.48	2.2 (0.31; 4)	2.3 (0.42; 2.8)	0.97	1.00	0.89	1.00
P [mg/l]	7.2 ± 0.046	7.2 ± 0.05	7.1 (7.1; 7.3)	7.2 (7; 7.3)	0.54	1.00	0.53	1.00
Pb [ng/l]	1.3 ± 0.46	1.3 ± 0.52	1.3 (0.46; 2.3)	1.3 (0.46; 2.3)	0.74	1.00	0.69	1.00
S [mg/l]	7 ± 0.11	7.1 ± 0.095	7 (6.9; 7.4)	7.1 (6.9; 7.3)	0.25	1.00	0.25	1.00
Se [µg/l]	3.6 ± 0.34	3.9 ± 0.33	3.5 (3.3; 4.5)	3.8 (3.3; 4.6)	< 0.01	0.04	< 0.01	0.03

Si [µg/l]	4.8 ± 0.2	4.7 ± 0.15	4.8 (4.1; 5.1)	4.7 (4.3; 5)	0.08	1.00	0.08	1.00
Sn [ng/l]	0.81 ± 0.42	1.1 ± 0.45	0.65 (0.43; 1.7)	1.2 (0.43; 2)	< 0.01		< 0.01	< 0.01
Sr [µg/l]	4.5 ± 0.3	4.4 ± 0.36	4.5 (4; 5.1)	4.3 (3.9; 5.2)	0.03	0.43	0.02	0.36
Ti [ng/l]	1.7 ± 0.24	1.7 ± 0.32	1.6 (1.3; 2.4)	1.6 (1.2; 2.9)	0.78	1.00	0.77	1.00
Tl [ng/l]	0.48 ± 0.058	0.48 ± 0.06	0.46 (0.46;0.73)	0.46 (0.46; 0.78)	0.19		0.47	
V [ng/l]	0.54 ± 0.19	0.54 ± 0.15	0.44 (0.43; 1.1)	0.48 (0.39; 1)	0.69		0.46	
Zn [µg/l]	3.9 ± 0.36	3.9 ± 0.16	3.8 (3.5; 5.2)	3.9 (3.7; 4.6)	0.50	1.00	0.78	1.00

Table 2: Single element comparison: Comparison of CSF concentrations based on regression by maximum likelihood estimation for left-censored data (column ‘Value censored MLE’). P values are adjusted using Holm's procedure (column ‘Value censored MLE adjusted’). Raw and adjusted p values resulting from t-tests on randomly imputed data are given for reference (columns ‘raw’ and ‘adjusted’). PD = Parkinson’s disease.

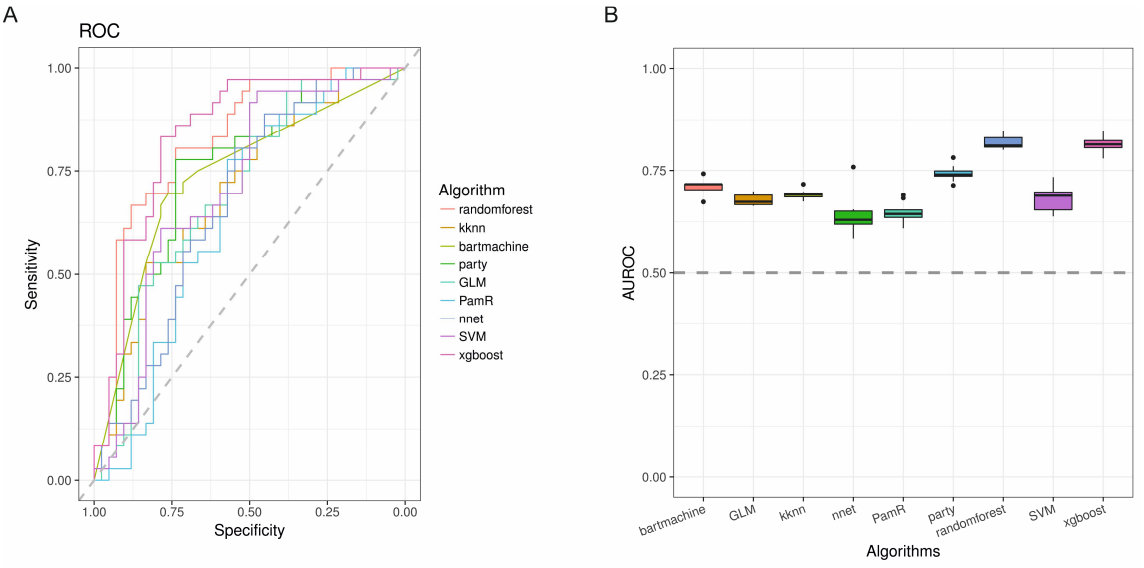
Hierarchical clustering with correlation as measurement of similarity

There is evidence for cluster formation, which partially discriminates PD patients and controls but also shows pronounced overlaps (Figure 1b). Arsenic and selenium, which present with significantly higher CSF levels in the single comparison (Figure 1a) show highly similar profiles and thus are represented directly next to each other. Magnesium, which also presents with significantly higher CSF levels, is directly encompassed by iron and strontium, two elements, which like magnesium itself also contribute to the final discrimination subset of elements described below.



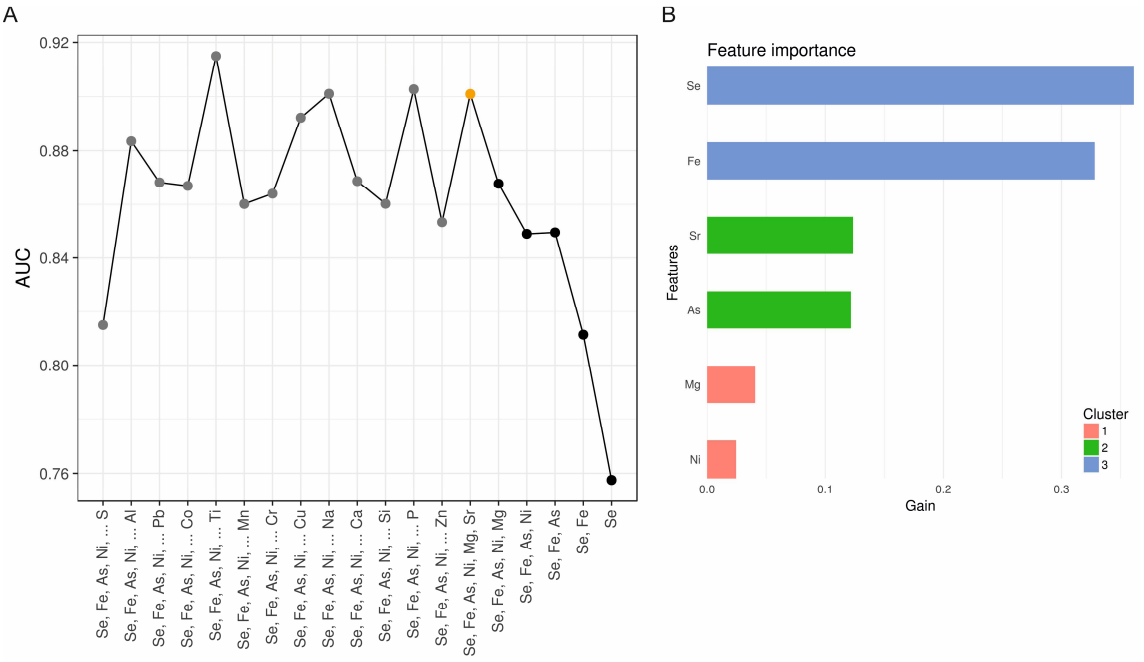
Random forest and extreme gradient tree boosting differentiate best Parkinson's disease from controls

The random forest and tree boosting algorithms showed the best overall performance in the discrimination of PD and control patients and both have a consensus AUROC over 80% (Figure 2a and Figure 2b). The random forest algorithm showed an average AUROC of 82.7%. Sensitivity and specificity at the Youden index are estimated as 66.7% and 88.1%. The tree-boosting algorithm showed the best average AUROC of 83.9%. At the Youden index the achieved sensitivity is 83.3% and the achieved specificity is 78.6%. If trained on the available data, both models can perfectly re-classify the training data.



Selenium, iron, arsenic, nickel, magnesium and strontium are sufficient for classification

We were seeking to reduce the number of elements to a plausible minimum, which still results in a sufficient classification of our samples. This procedure proposed that only six features - Se, Fe, As, Ni, Mg, and Sr - are sufficient to achieve good classification performance (cross validated AUROC: 0.90, Figure 3a).

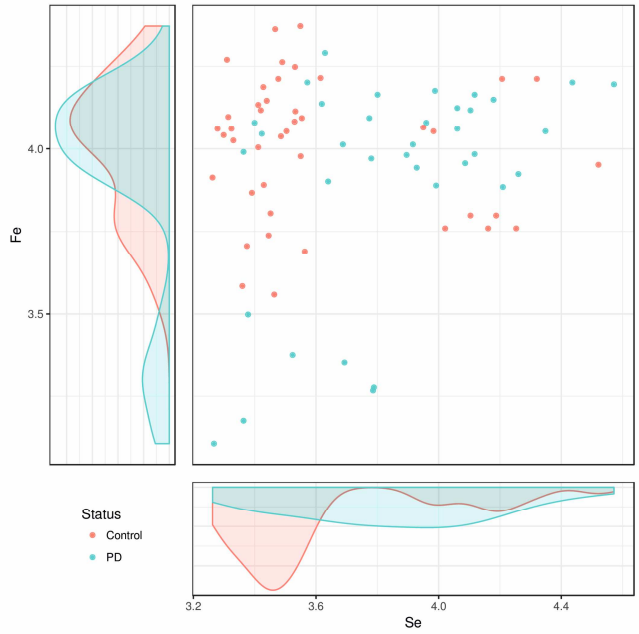


Selenium and iron contribute most to sample discrimination

Based on the six remaining elements from the feature selection analysis, the feature importance analysis presents selenium with the highest impact in this model, directly followed by iron (Figure 3b).

Iron and selenium show a strong interaction within the decision tree

Unlike selenium, iron levels did not differ significantly regarding the single element comparison (Figure 1a). However, iron appears to be a highly informative feature within the discrimination models. While the main effect is clearly contributed by selenium, a second split induced by iron levels can be visualized (Figure 4).



Clinical parameters do not significantly correlate to elemental levels

There was no significant correlation between the 28 single elements neither for the motor or non-motor symptoms (MDS-UPDRS total score and part III, MDS-PDNMS) nor for the disease duration and H&Y stages ($p > 0.05$). The combined regression model also showed no correlations (Supplementary Figure S1). There was also no correlation between the final six elements and the parameter age and levodopa equivalent dose (Supplementary Figure S2).

Discussion

Although several studies on the levels of single elements in the CSF have been published, the data is inconclusive due to variable detection methods, inadequately defined patients and insufficiently controlled blood contamination impacting elemental composition (Jiménez-Jiménez et al., 2014; Mariani et al., 2013). Multiple factors influence the levels of single elements, i.e. dietary habits, drug intake, competing diseases, geographical factors, which makes it highly unlikely that one single element will function as disease marker. To overcome

this problem focusing on elemental patterns rather than single elements may increase the robustness of results. After quantification of a total of 28 elements, eight were excluded due to low detection levels and one due to diet- and medication-dependent variability (iodine). Using a total of 19 stably detected elements we were able to create a unique elemental fingerprint, which was able to differentiate PD patients from age-matched controls with high specificity and sensitivity (78.6% and 83.3%, respectively), perfectly re-classifying the training data. Even though our analysis was based purely on the chemical properties of the CSF and did not take into account additional clinical features or supportive diagnostics the 10 times 10-fold cross validation yielded an AUROC of 0.83. In contrast, CSF alpha-synuclein shows a low specificity of ~25% with a corresponding AUROC of 0.69, even when taking age as the biggest risk factor for the development of PD into account (Mollenhauer et al., 2011). Since the analysis of a large number of elements is not economical and might prevent clinical translation, we aimed to reduce the number of required elements to an optimal subset. This resulted in the identification of a cluster of six single elements (Se, Fe, As, Ni, Mg, Sr), which most importantly contributed to the sample discrimination.

Within this most determining cluster, selenium was identified as the element with the highest impact for the overall score directly followed by iron. The knowledge about selenium in neurodegenerative disorders has especially emerged in the last few years. As essential bioelement, dietary selenium converted into selenide (Se^{2-}) serves as a donor for the incorporation into selenoproteins like the glutathione peroxidases family (GPx) or Selenoprotein P (Sepp1), which are expressed in the CNS (Cardoso et al., 2015). Both proteins have anti-oxidative activity protecting against ROS-induced cell stress, which is known to contribute to dopaminergic neuron damage in PD (Bellinger et al., 2012; Kaur and Andersen, 2004). In Parkinson's disease brains, GPx1-positive microglia are co-localized with Lewy bodies (Power and Blumbergs, 2009). GPx4 is also elevated in Parkinson's disease brains and is regulated by DJ-1, a protein deglycase involved in genetic forms of PD (PARK7) (Blackinton et al., 2009). Sepp1, a selenium transport protein and antioxidant, also co-localizes with the core of the Lewy bodies (Bellinger et al., 2012). This could contribute to increased levels of selenium in the CSF, which have been observed in patients with PD (Aguilar et al., 1998; Qureshi et al., 2006). Thus, although selenoproteins appear to be important in PD pathogenesis, the reason for elevated selenium levels is not fully understood. Both, etiologically causative as well as compensatory mechanisms could be involved.

As the second most important discriminator in our algorithm iron is the most abundant bioelement in humans and has been linked to disease mechanisms in PD for almost one

century (Lhermitte et al., 1924). Iron catalyzes the formation of reactive oxygen species (ROS) through Fenton reactions contributing to neuronal protein, lipid and DNA damage (Kaur and Andersen, 2004). Alterations in blood iron levels seem to modulate the risk of developing PD as was shown in a Mendelian randomization study (Pichler et al., 2013). Despite the importance of iron in the pathogenesis of PD, CSF iron levels alone are not sufficient as diagnostic biomarkers. A meta-analysis showed no variation between PD patients and healthy controls (Mariani et al., 2013). This is also confirmed by our data showing no significant difference in iron levels when considered as single factor. Interestingly, our analysis suggests that there is a relevant interaction of iron and selenium, which is underlined by the high discriminating impact of both elements in the machine-learning algorithm. This underlies the strength of this elemental biomarker signature in contrast to the consideration of single elements. Arsenic so far has not been studied in the CSF of PD patients, but a recent study demonstrates that this metalloid induces oligomerization of alpha-synuclein (Cholanians et al., 2016). As a chemical compound in pesticides it might contribute to the increased risk of the development of PD after occupational exposure (Elbaz et al., 2009). There is less evidence linking nickel and neurodegeneration. The metal is discussed to contribute to the development of extrapyramidal signs in technical dentists by occupational exposition (Fabrizio et al., 2007). In addition, mutations in ATP13A2, which encodes a lysosomale ATPase protecting against manganese and nickel toxicity, cause a genetic form of PD (PARK9) (Covy et al., 2012). Again, our single element analysis did not reveal significant changes in nickel CSF levels, which is coherent with previous results (Alimonti et al., 2007). However, nickel significantly contributes to the discrimination algorithm as part of the highly discriminating element cluster.

Magnesium was one of the three elements, which showed significantly increased levels in the single element analysis in the CSF of PD patients compared to controls. It has also been shown to accelerate alpha-synuclein aggregation and lower levels have been found in the cortex, white matter, basal ganglia and brain stem of PD patients (Lowe et al., 2004; Yasui et al., 1992). Data for magnesium in the CSF are contradictory but suggest elevated levels: whereas a recent study showed increased Mg levels in PD, others detected only a trend for an increase and yet two other studies showed no significant changes (Alimonti et al., 2007; Forte et al., 2004; Hozumi et al., 2011; Sanyal et al., 2016). Interestingly, there is also an increase in CSF magnesium levels in patients with Lewy body dementia (Boström et al., 2009).

Our analysis also reveals strontium as non-essential element to contribute to the discrimination algorithm. Strontium is mostly present in osseous structures, i.e. bones and teeth, and its levels were previously shown to be unchanged in the CSF of PD patients compared to controls, matching our results (Alimonti et al., 2007; Peltz-Császma et al., 2005). Its function in the CNS is largely unknown, but due to its size, strontium can replace synaptic calcium and trigger neurotransmitter release, thus contributing to synaptic transmission and plasticity (Xu-Friedman and Regehr, 1999). Interestingly a recent epidemiological study reports a significant association between the PD mortality rates and the soil concentrations of strontium but also selenium and magnesium in the USA, highlighting three of the six elements of the final subset (Sun, 2017).

Our analysis did not reveal a significant correlation neither between single elemental levels nor between an elemental score using the final set of elements and clinical parameters (MDS-UPDRS part III and sum score, MDS-PDNMS, age, disease duration, Hoehn and Yahr stages). There was also no significant correlation between levodopa equivalent doses and the six final elements after adjusting for multiple testing. Since our PD cohort includes patients with variable disease durations, our trial might not be sufficiently powered to establish significant correlations to clinical parameters.

In conclusion, our study demonstrates that an elemental fingerprint in the CSF can be used as novel and independent biomarker for the diagnosis of PD. Our data also argues for a prominent role of six highly discriminating elements in the pathogenesis of PD.

Acknowledgement

The authors grateful appreciate the participation of our patients in this study. We want to thank Woori Koh and Anna Fischbach for their contributions to the patient recruitment. We also thank our Parkinson's and study nurses Gudrun Leyerer and Elisabeth Koch for their excellent assistance.

Author disclosures

None of the authors have anything to disclose regarding this manuscript.

Funding

DFG Research Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB) (Paul Lingor and Mathias Bähr, Grant: CNMPB III.I.9); Else-Kröner-Fresenius-Stiftung (Paul Lingor)

Figure legends

Figure 1 (a) Distribution of the elements selenium, iron, arsenic, nickel, magnesium and strontium. The elements As, Mg, Se are significantly different when compared individually between controls and Parkinson's disease patients on a 0.05 level after multiple adjustment. The element levels are presented on a log10 scale. **(b) Hierarchical clustering with correlation as measurement of similarity.** Patients and elements are hierarchically clustered so that patients with similar element level profiles as well as elements with similar patterns across patients are located close to each other. The one grey coded rectangle represents a missing value for sodium (Na). PD = Parkinson's disease (n = 36), Control (n = 42).

Figure 2 (a) Performance of different machine-learning algorithms, ROC curves. Nine different machine-learning methods were applied to the data. The performance was assessed via 10-times 10-fold cross validation. **(b) Performance of different machine learning algorithms, AUROC.** The area under the ROC-curve (AUROC) was estimated in each 10-fold cross validation.

Figure 3 (a) Feature selection for determination of a minimal number of elements to achieve a good classification performance. The AUC results from the 10 times repeated 10-fold cross validations were used as a feature selection criterion to determine the optimal set of elements for a good classifier. A step-wise procedure removing the least informative (measured by AUC) element one after the other was applied. AUC estimates (y axis) for the backward selected list of feature sets (x axis) are shown. Highlighted in black is the feature selection path starting from the remaining single element including back elements until the AUC decreases again. The orange dot shows the final selection. **(b) Feature importance showing the impact of each element of the final set within the decision tree boosting.** The feature importance – measured as information gain – for each of these six elements is quantified via 10 times 10-fold cross validation of a decision tree boosting classifier trained on these 6 elements only.

Figure 4 Interaction of iron and selenium. The abundances of iron and selenium (in log (ng/L)) are plotted against each other. Marginal densities for both elements are plotted at each side of the plot. While the main effect is clearly in Se, a second split by Fe seems favorable when just assessing the scatter plot visually. PD = Parkinson's disease (n = 36), Control (n = 42).

References

Aguilar M. V., Jiménez-Jiménez F. J., Molina J. A., Meseguer I., Mateos-Vega C. J., González-Muñoz M. J., Bustos F. De, et al. (1998) Cerebrospinal fluid selenium and chromium levels in patients with Parkinson's disease. *J. Neural Transm.* **105**, 1245–1251.

Alimonti A., Bocca B., Pino A., Ruggieri F., Forte G., Sancesario G. (2007) Elemental profile of cerebrospinal fluid in patients with Parkinson's disease. *J. Trace Elem. Med. Biol.* **21**, 234–241.

Bellinger F. P., Raman A. V., Rueli R. H., Bellinger M. T., Dewing A. S., Seale L. A., Andres M. A., et al. (2012) Changes in selenoprotein P in substantia nigra and putamen in Parkinson's disease. *J. Parkinsons. Dis.* **2**, 115–126.

Blackinton J., Kumaran R., Brug M. P. van der, Ahmad R., Olson L., Galter D., Lees A., Bandopadhyay R., Cookson M. R. (2009) Post-transcriptional regulation of mRNA associated with DJ-1 in sporadic Parkinson disease. *Neurosci. Lett.* **452**, 8–11.

Boström F., Hansson O., Gerhardsson L., Lundh T., Minthon L., Stomrud E., Zetterberg H., Londos E. (2009) CSF Mg and Ca as diagnostic markers for dementia with Lewy bodies. *Neurobiol. Aging* **30**, 1265–1271.

Cardoso B. R., Roberts B. R., Bush A. I., Hare D. J. (2015) Selenium, selenoproteins and neurodegenerative diseases. *Metallomics* **7**, 1213–1228.

Cholanians A. B., Phan A. V., Ditzel E. J., Camenisch T. D., Lau S. S., Monks T. J. (2016) Arsenic induces accumulation of α-synuclein: Implications for synucleinopathies and

neurodegeneration. *Toxicol. Sci.* **153**, 271–281.

Covy J. P., Waxman E. A., Giasson B. I. (2012) Characterization of cellular protective effects of ATP13A2/PARK9 expression and alterations resulting from pathogenic mutants. *J. Neurosci. Res.* **90**, 2306–2316.

Davies K. M., Bohic S., Carmona A., Ortega R., Cottam V., Hare D. J., Finberg J. P. M., et al. (2014) Copper pathology in vulnerable brain regions in Parkinson's disease. *Neurobiol. Aging* **35**, 858–866.

Dexter D. T., Wells F. R., Lee A. J., Agid F., Agid Y., Jenner P., Marsden C. D. (1989) Increased Nigral Iron Content and Alterations in Other Metal Ions Occurring in Brain in Parkinson's Disease. *J. Neurochem.* **52**, 1830–1836.

Elbaz A., Clavel J., Rathouz P. J., Moisan F., Galanaud J. P., Delemotte B., Alpérovitch A., Tzourio C. (2009) Professional exposure to pesticides and Parkinson disease. *Ann. Neurol.* **66**, 494–504.

Fabrizio E., Vanacore N., Valente M., Rubino A., Meco G. (2007) High prevalence of extrapyramidal signs and symptoms in a group of Italian dental technicians. *BMC Neurol.* **7**, 24.

Forte G., Bocca B., Senofonte O., Petrucci F., Brusa L., Stanzione P., Zannino S., Violante N., Alimonti A., Sancesario G. (2004) Trace and major elements in whole blood, serum, cerebrospinal fluid and urine of patients with Parkinson's disease. *J. Neural Transm.* **111**, 1031–1040.

Hozumi I., Hasegawa T., Honda A., Ozawa K., Hayashi Y., Hashimoto K., Yamada M., et al. (2011) Patterns of levels of biological metals in CSF differ among neurodegenerative diseases. *J. Neurol. Sci.* **303**, 95–99.

Jiménez-Jiménez F. J., Alonso-Navarro H., García-Martín E., Agúndez J. A. G. (2014) Cerebrospinal fluid biochemical studies in patients with Parkinson's disease: toward a potential search for biomarkers for this disease. *Front. Cell. Neurosci.* **8**, 369.

Kaur D., Andersen J. (2004) Does cellular iron dysregulation play a causative role in Parkinson's disease? *Ageing Res. Rev.* **3**, 327–343.

Lhermitte J., Kraus W. M., McAlpine D. (1924) The Occurence of abnormal deposits of iron in the brain in parkinsonism with special reference to its localisation. *J. Neurol. Psychopathol.* **V**, 195–208.

Lowe R., Pountney D. L., Jensen P. H., Gai W. P., Voelcker N. H. (2004) Calcium(II) selectively induces alpha-synuclein annular oligomers via interaction with the C-terminal domain. *Protein Sci.* **13**, 3245–52.

Mariani S., Ventriglia M., Simonelli I., Donno S., Bucossi S., Vernieri F., Melgari J. M., Pasqualetti P., Rossini P. M., Squitti R. (2013) Fe and Cu do not differ in Parkinson's disease: A replication study plus meta-analysis. *Neurobiol. Aging* **34**, 632–633.

Mukherjee S., Tamayo P., Rogers S., Rifkin R., Engle A., Campbell C., Golub T. R., Mesirov J. P. (2003) Estimating Dataset Size Requirements for Classifying DNA Microarray Data. *J. Comput. Biol.* **10**, 119–142.

Mollenhauer B., Locascio J. J., Schulz-Schaeffer W., Sixel-Döring F., Trenkwalder C., Schlossmacher M. G. (2011) alpha-Synuclein and tau concentrations in cerebrospinal fluid of patients presenting with parkinsonism: A cohort study. *Lancet Neurol.* **10**, 230–240.

Peltz-Császma I., Andrási E., Lásztity A., Kösel S. (2005) Determination of strontium and its relation to other alkaline earth elements in human brain samples. *Microchem. J.* **79**, 375–381.

Pichler I., Greco M F. Del, Gögele M., Lill C. M., Bertram L., Do C. B., Eriksson N., et al. (2013) Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. *PLoS Med.* **10**, e1001462.

Postuma R. B., Berg D., Stern M., Poewe W., Olanow C. W., Oertel W., Obeso J., et al.

(2015) MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591–1601.

Power J. H. T., Blumbergs P. C. (2009) Cellular glutathione peroxidase in human brain: cellular distribution, and its potential role in the degradation of Lewy bodies in Parkinson's disease and dementia with Lewy bodies. *Acta Neuropathol* **117**, 63–73.

Qureshi G. A., Qureshi A. A., Memon S. A., Parvez S. H. (2006) Impact of selenium, iron, copper and zinc in on/off Parkinson's patients on L-dopa therapy. *J. Neural Transm. Suppl.* 71, 229–236.

Rizzo G., Copetti M., Arcuti S., Martino D., Fontana A., Logroscino G. (2016) Accuracy of clinical diagnosis of Parkinson disease. *Neurology* **86**, 566–576.

Sanyal J., Ahmed S. S. S. J., Ng H. K. T., Naiya T., Ghosh E., Banerjee T. K., Lakshmi J., Guha G., Rao V. R. (2016) Metallomic Biomarkers in Cerebrospinal fluid and Serum in patients with Parkinson's disease in Indian population. *Sci. Rep.* **6**, 35097.

Sun H. (2017) Association of soil selenium, strontium, and magnesium concentrations with Parkinson's disease mortality rates in the USA. *Environ. Geochem. Health*, 1–9.

Tomlinson C. L., Stowe R., Patel S., Rick C., Gray R., Clarke C. E. (2010) Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov. Disord.* **25**, 2649–2653.

Uversky V. N., Li J., Fink A. L. (2001) Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein: A possible molecular link between parkinson's disease and heavy metal exposure. *J. Biol. Chem.* **276**, 44284–44296.

Virk S. A., Eslick G. D. (2015) Aluminum Levels in Brain, Serum, and Cerebrospinal Fluid are Higher in Alzheimer's Disease Cases than in Controls: A Series of Meta-Analyses. *J. Alzheimer's Dis.* **47**, 629–638.

Xu-Friedman M. A., Regehr W. G. (1999) Presynaptic Strontium Dynamics and Synaptic Transmission. *Biophys. J.* **76**, 2029–2042.

Yasui M., Kihira T., Ota K. (1992) Calcium, magnesium and aluminum concentrations in Parkinson's disease. *Neurotoxicology* **13**, 593–600.