A clinical-anatomical signature of Parkinson's Disease identified with partial least squares and magnetic resonance imaging

Yashar Zeighami, Seyed-Mohammad Fereshtehnejad, Mahsa Dadar, D. Louis Collins, Ronald B. Postuma, Bratislav Mišić*, Alain Dagher*

Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

* These authors contributed equally to this work

Correspondence: alain.dagher@mcgill.ca

3801 University St, Montreal QC, Canada H3A 2B4

Words: 6379; Tables: 2; Figures: 4. No supplemental information.

Abstract

Parkinson's disease (PD) is a neurodegenerative disorder characterized by a complex array of motor and non-motor symptoms, such as autonomic and cognitive impairments. It remains unclear whether neurodegeneration in discrete loci gives rise to discrete symptoms, or whether network-wide atrophy gives rise to the unique behavioural clinical profile associated with PD. Here we apply a data-driven strategy to isolate large-scale, multivariate associations between distributed atrophy patterns and clinical phenotypes in PD. In a sample of N = 229 de novo PD patients, we estimate disease-related atrophy using deformation based morphometry (DBM) of T1w MR images. Using partial least squares (PLS), we identify a network of subcortical and cortical regions whose collective atrophy is associated with a robust clinical phenotype with both motor and non-motor features. Despite the relatively early stage of the disease in the sample, the atrophy pattern encompassed lower brainstem, substantia nigra, basal ganglia and cortical areas, consistent with the Braak staging of the disease. In addition, individual variation in this putative atrophy network predicted longitudinal clinical progression in both motor and non-motor symptoms. Altogether, these results demonstrate a pleiotropic mapping between neurodegeneration and the clinical manifestations of PD, and that this mapping can be detected even in de novo patients.

Keywords: Parkinson's disease, MRI, deformation based morphometry, partial least squares, disease progression

1. Introduction:

Parkinson's disease (PD) is a neurodegenerative disease characterized by progressive and widespread neuronal loss associated with intracellular aggregates of α-synuclein giving rise to the classical Lewy pathology (Poewe et al. 2017; Goedert et al. 2013). PD has been traditionally known as a motor disease with bradykinesia, rigidity, and tremor as the cardinal symptoms, and preferential loss of dopamine neurons of the substantia nigra (SN). The motor symptoms have been the main target for diagnosis and treatment (Kalia and Lang 2015). However, it is now clear that PD is a more complex disorder involving several non-motor manifestations that both precede and follow the initial appearance of motor symptoms. The non-motor aspects of PD involve several clinical domains including autonomic, limbic, olfactory, and cognitive impairment (Chaudhuri, Healy, and Schapira 2006; W. Poewe 2008). A 15-year follow-up study shows cognitive decline and dementia in up to 80% of surviving PD patients (Hely et al. 2005).

Over time, PD diagnostic criteria has been modified toward a multifaceted characterization in response to the insufficiency of the narrow motor definition in PD diagnosis and prognosis (Postuma et al. 2016). The increasing attention to non-motor aspects of the disease has allowed detection of different clinical patterns of neurodegeneration in PD. For example, recent studies have subcategorized PD patients based on the dominance of motor, rapid eye movement sleep behavior disorder (RBD), autonomic, and cognitive deterioration (Fereshtehnejad et al. 2015, 2017). Post-mortem and neuroimaging studies have emphasized the preferential loss of dopamine neurons in the substantia nigra (Halliday and McCann 2010). However, post-mortem studies have also shown that the pathological process is neither initiated nor confined to the

substantia nigra, gradually ascending from the olfactory tracts and medulla to the midbrain and cortical layers (Braak et al. 2003; Goedert et al. 2013)

Neuroimaging studies in PD have evolved in the past 30 years (Politis 2014). Initially, the main focus of early studies was on dopaminergic innervation, using single photon emission computed tomography (SPECT) or positron emission tomography (PET). However, the availability of new higher resolution neuroimaging techniques such as magnetic resonance imaging (MRI), metabolic imaging with ^{18}F-FDG PET, and resting or task state functional MRI have provided the opportunity to investigate the non-dopaminergic aspects of PD (Politis 2014; Tuite and Dagher 2013; Yousaf, Wilson, and Politis 2017). Structural analysis using MRI (including T1-w, T2-w, and diffusion weighted MRI) was initially inconclusive, or only sensitive enough to capture disease related differences in late stages of PD once dementia has set in. More recently, with larger sample sizes and higher resolution imaging, it has been possible to study *de novo* PD patients using MRI (Zeighami et al. 2015; Heim et al. 2017). However, these studies mostly focus on brain related differences between PD and healthy control populations, or on a single aspect of the disease (e.g. dementia). To our knowledge, no studies have attempted to model the relationship between brain atrophy and presence and severity of the entire constellation of motor and non-motor symptoms in PD simultaneously.

Here we use a multivariate approach to relate the motor and non-motor aspects of PD to system-wide atrophy patterns. We use data from 235 newly diagnosed PD patients from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data), an observational, multicenter longitudinal study designed to identify PD progression biomarkers (Marek et al. 2011). We use deformation-based morphometry (DBM), which is based on local nonlinear subject-to-template deformations as a measure of structural brain alterations (Ashburner, Good,

and Friston 2000; Penny et al. 2011; Chung et al. 2001; Aubert-Broche et al. 2013), and partial least squares (PLS) (Wold 1966; McIntosh and Lobaugh 2004; McIntosh and Mišić 2013) to capture the relationship between brain atrophy patterns and disease-related clinical measures. Furthermore, we explore the extent to which brain atrophy patterns can predict disease progression by examining longitudinal changes across different measures of disease severity.

1. **Methods**

2.1 PPMI dataset

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI is a cohort of people with *de-novo* idiopathic PD (Marek et al. 2011). Individuals were eligible for recruitment if they were at least 30 years old, diagnosed with PD within the last 2 years, had at least two signs or symptoms of Parkinsonism (tremor, bradykinesia and rigidity), a baseline Hoehn and Yahr Stage of I or II, and did not require symptomatic treatment within six months of the baseline visit. The PPMI is a multi-center international project and the institutional review boards approved the protocol at all participating sites. Participation was voluntary and all individuals signed the written informed consent prior to inclusion.

We obtained data from the baseline visit 3T high-resolution T1-weighted MRI scans in compliance with the PPMI Data Use Agreement. For clinical data, any participant with > 20% missing values at baseline was excluded. Overall, MRI and clinical data were included for 229 drug-naïve participants with PD (6 subjects failed MRI quality control). For each subject, we also obtained demographic and clinical information as well as cerebrospinal fluid (CSF) and

SPECT biomarker values from the dataset in May 2016 (accession date). General information consisted of age at disease onset, gender, years of education, handedness and disease duration. Clinical and laboratory markers are described below.

1.1.Brain Imaging Data Analysis

MRI data consisted of 1×1×1 mm^3 3T T1-weighted scans obtained from the PPMI database. All scans were pre-processed through our MR image processing pipeline, using image de-noising (Coupe et al. 2008), intensity non-uniformity correction (Sled, Zijdenbos, and Evans 1998), and image intensity normalization using histogram matching. The preprocessed images were first linearly (using a 9-parameter rigid registration) and then nonlinearly registered to a standard brain template (MNI ICBM152) (Collins et al. 1994; Collins and Evans 1997). Using the obtained nonlinear transformations, deformation based morphometry (DBM) was performed to calculate local density changes as a measure of tissue expansion or atrophy. For more detail on the processing steps please see (Zeighami et al. 2015). We obtained a single deformation brain map for each subject. The value at each voxel is equal to the determinant of the Jacobian of the transformation matrix obtained from nonlinear registration of participants' T1w MR images and MNI-ICBM152 brain template. The DBM values reflect regional brain deformations and can be used as indirect measures of brain atrophy (Chung et al. 2001; Studholme et al. 2004; Leow et al. 2006; Cardenas et al. 2007)

2.3 Clinical Measures

PD-related motor, cognitive and non-motor clinical manifestations were assessed at baseline and each follow-up visit (Table 1).

Table 1. Demographic and clinical information for individuals with Parkinson's disease from the PPMI used in this study. BP Sys= Systolic Blood Pressure. GDS= Geriatric Depression Scale.

QUIP = Questionnaire for Impulsive-Compulsive Disorders. RBD= REM sleep behaviour disorder. SCOPA= Scales for Outcomes in PD-Autonomic. STAI= State-Trait Anxiety Inventory. UPDRS= Unified Parkinson's Disease Rating Scale. SBR= striatal binding ratio. MoCA= Montreal Cognitive Assessment. HVLT= Hopkins Verbal Learning Test. LNS= letter-number sequencing.

Category	Measure	Value
General Information	Number	229
	Sex (Male / Female / %)	146/ 83
	Age (years)	60.8 ± 9.1
	Education Years	15.5 ± 2.8
	Handedness – Right / Left / Ambidextrous	209/ 15/ 5
	Symptom duration (months)	7 ± 7
Non-motor scores	BP Sys drop	4 ± 11
	Epworth Sleepiness Score	5.9 ± 3.6
	GDS Score	2.3 ± 2.5
	QUIP total	0.3 ± 0.6
	RBD Score	3.5 ± 2.7
	SCOPA AUT Score	9.4 ± 6
	STAI Total Score	64.2 ± 18.3
	UPSIT Score	12.8 ± 17.6
	UPDRS part I	5.5 ± 4
	SBR	1.4 ± 0.4
	UPDRS part II	5.8 ± 4.0
	UPDRS part III	21.9 ± 9
Cognitive scores	MoCA Score	27.4 ± 2.2
	Benton	12 ± 2.8
	HVLT total recall	47.1 ± 11.8
	HVLT delayed recall	47.2 ± 12.1
	HVLT retention Score	50 ± 11.6
	HVLT Recognition	50 ± 13
	LNS	11.4 ± 2.8
	Semantic Fluency Score	50.9 ± 10
	Symbol Digit Score	45.3 ± 8.7
CSF biomarker	Total Tau	44.8 ± 19.1
	pTau	15.4 ± 10.2
	Alpha synuclein	1.8 ± 0.7
	Amyloid beta 42	362 ± 93
	Hoehn and Yahr scale	1.6 ± 0.5

We also included Genetic Risk Score as part of general information in the PLS analysis. This is a single surrogate indicator previously introduced in PPMI that summarizes 30 risk alleles for PD (Nalls et al. 2015). All clinical assessments were repeated in follow-up visits (minimum= 1 year, mean= 2.7 years). In order to evaluate disease progression, we created a putative global composite outcome (GCO) as a single indicator by combining z-scores of the most clinically

relevant motor and non-motor measures of disease severity including Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) part I, II, and III , Schwab and England activities of daily living (SE-ADL) score, and Montreal Cognitive Assessment (MoCA) score (Fereshtehnejad et al. 2017).

2.4 Biomarkers

The striatal binding ratio (SBR), a marker of dopaminergic denervation in caudate and putamen, was obtained by SPECT with the DAT tracer ^{123}I-Ioflupane (Marek et al. 2011) at baseline and follow-up. Cerebrospinal fluid (CSF) biomarkers consisting of amyloid-beta (Aβ1-42), total Tau (T-tau), phosphorylated tau (P-tau181) and α-synuclein were also included in our analysis. Information for all variables is summarized in Table 1.

2.5 Partial least squares analysis

Partial least squares (PLS) is an associative, multivariate method for relating two sets of variables to each other (Wold 1966; Anthony Randal McIntosh and Lobaugh 2004; Abdi and Williams 2010; Anthony R. McIntosh and Mišić 2013). The analysis seeks to find weighted linear combinations of the original variables that maximally covary with each other. Here, the two variable sets were voxel-wise brain atrophy (as measured by DBM) and clinical/demographic measures (Table 1). The respective linear combinations of these variables can be interpreted as atrophy networks and their associated clinical phenotypes.

Singular value decomposition: The imaging and clinical data were organized in two matrices, **X** (DBM) and **Y** (clinical), with participants in the rows of the matrices and variables in the columns (Figure 1). Both matrices were first z-scored by subtracting the mean from each column (variable) and dividing by the standard deviation. The atrophy-clinical covariance matrix was

then computed, representing the covariation of all voxel deformation values and clinical measures across participants. Since the data are z-scored, the atrophy-clinical covariance is effectively a correlation matrix. The resulting matrix was then subjected to singular value decomposition (SVD) (Eckart and Young 1936):

$$\mathbf{X'Y} = \mathbf{U\Delta V'}$$

such that

$$\mathbf{U'U} = \mathbf{V'V} = \mathbf{I}.$$

The decomposition yields a set of mutually orthogonal latent variables (LVs), where **U** and **V** are matrices of left and right singular vectors, and **Δ** is a diagonal matrix of singular values. Each latent variable is a triplet of the ith left singular vector, the ith right singular vector and the ith singular value. The number of latent variables is equal to the rank of the covariance matrix, which is the smaller of its dimensions or the dimension of its constituent matrices. In the present study, the number of clinical measures ($k = 31$) is the smallest dimension, so the rank of the matrix and the total number of latent variables is equal to 31. If there are v voxels, the dimensions of **U**, **V**, and Δ are $v \times k$, $k \times k$, and $k \times k$, respectively.

Each singular vector weights the original variables in the multivariate pattern. Thus, the columns of **U** and **V** weight the original voxel deformation values and clinical measures such that they maximally covary. The weighted patterns can be interpreted as a set of maximally covarying atrophy patterns and their corresponding clinical phenotypes. Each such pairing is associated with a singular value from the diagonal matrix, proportional to the covariance between atrophy and behavior captured by the latent variable. Specifically, the effect size associated with each latent variable (proportion of covariance accounted for) can be naturally estimated as the ratio of

the squared singular value to the sum of all squared singular values (Anthony Randal McIntosh and Lobaugh 2004).

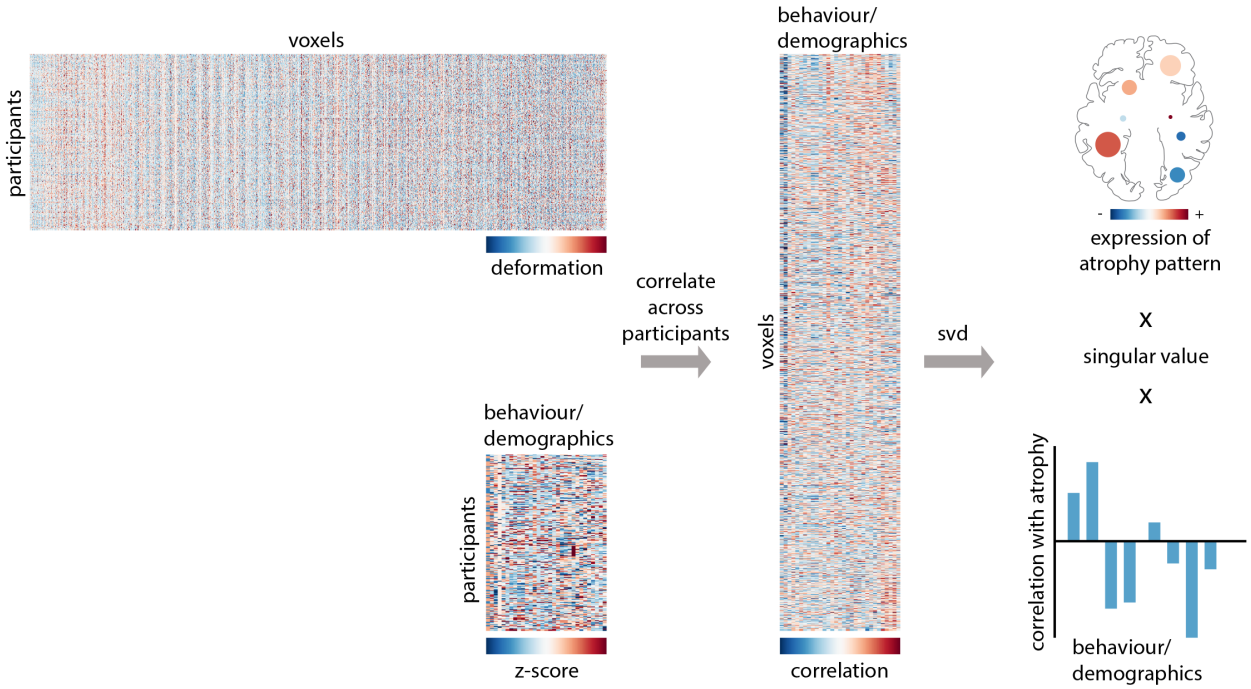

Figure 1. Partial Least Square (PLS) Analysis flowchart.

Significance of multivariate patterns: The statistical significance of each latent variable was assessed by permutation tests. The ordering of observations (i.e. rows) of matrix **X** was randomly permuted, and a set of null atrophy-behaviour matrices were then computed for the permuted brain and non-permuted clinical data matrices. These null correlation matrices were then subjected to SVD as described above, generating a distribution of singular values under the null hypothesis that that there is no relationship between brain deformation and clinical measures. Since singular values are proportional to the magnitude of a latent variable, a non-parametric P value can be estimated for a given latent variable as the probability that a permuted singular value exceeds the original, non-permuted singular value.

Contribution and reliability of individual variables: The contribution of individual variables (voxels or clinical measures) was estimated by bootstrap resampling. Participants (rows of data matrices **X** and **Y**) were randomly sampled with replacement, generating a set of resampled correlation matrices that were then subjected to SVD. This procedure generated a sampling distribution for each individual weight in the singular vectors. A "bootstrap ratio" was calculated for each voxel as the ratio of its singular vector weight and its bootstrap-estimated standard error. Thus, large bootstrap ratios can be used to isolate voxels that make a large contribution to the atrophy pattern (have a large singular vector weight) and are stable across participants (have a small standard error). If the bootstrap distribution is approximately normal, the bootstrap ratio is equivalent to a z-score (Efron and Tibshirani, 1986). Bootstrap ratio maps were thresholded at values corresponding to the 95% confidence interval.

Patient-specific atrophy and clinical scores: To estimate the extent to which individual patients express the atrophy or behavioural patterns derived from the analysis, we calculated patient-specific scores. Namely, we projected the weighted patterns **U** and **V** onto individual-patient data, yielding a scalar atrophy score and clinical score for each patient, analogous to a principal component score or factor score:

$$\text{Atrophy score} = \mathbf{XU}$$

$$\text{Clinical score} = \mathbf{YV}$$

To investigate the predictive utility of the PLS model, we correlated patient-specific atrophy and clinical scores with longitudinal measures of disease progression. These included the Global Composite Outcome (GCO) and SE-ADL scores as measures of general disease severity, MoCA for cognition, MDS-UPDRS III for motor, and MDS-UPDRS I for non-motor aspects of disease.

3. Results

3.1. PLS analysis

The PLS analysis revealed five statistically significant latent variables relating clinical measures in PD and their corresponding brain atrophy patterns (permuted p <0.0001, p<0.005, p<0.05, p<0.05, p< 0.005). These patterns respectively account for 18, 8.5, 8.2, 6, and 4.5% (total of 45%) of the shared covariance between clinical and brain atrophy measures. Based on the variance explained and clinical interpretability of the results, we focus on and discuss only the first latent variable (LV-I) in greater detail for the rest of the report (Figure 2).

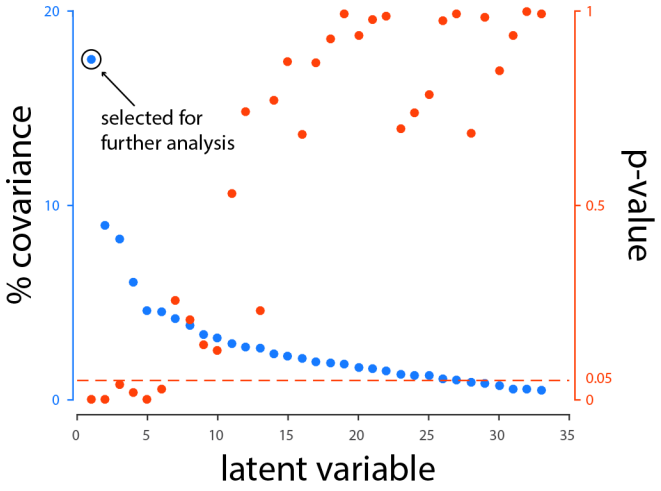

Figure 2. Covariance explained and permutation p-values for all latent variables in PLS analysis. LV-I is selected for analysis based on the variance explained and clinical interpretability of the results. PLS = Partial Least Squares.

3.2. Clinical features and biomarkers patterns

The biomarkers and clinical features (Figure 3) contributing to LV-I are composed of: higher PD-related severity (motor and non-motor) as measured by UPDRS scores, lower striatal dopamine innervation measured by DAT-scan, lower cognitive performance (mainly memory-related), lower amyloid beta level in CSF, and more severe anxiety, depression, and sleep

disorder. We also found the previously reported effects of age (worse with age) and gender (males worse). More specifically, age was the strongest contributor to LV-I (R = 0.69, 95% CI [0.59,0.74]) followed by motor signs measured by UPDRS-III (R = 0.35, 95% CI [0.35,0.52]) and autonomic disturbances (SCOPA-AUT) (R = 0.27, 95% CI [0.25,0.45]). Male gender (R = 0.23, 95% CI [0.07,0.34]) and symptom duration (R = 0.19, 95% CI [0.10,0.32]) were other significant contributors to LV-I. Impaired visuospatial (Benton Line Orientation) (R = -0.25, 95% CI [-0.40,-0.18]) and executive function (Letter-Number Sequencing) (R = -0.25, 95% CI [-0.38,-0.16]) were the strongest cognitive features of LV-I, followed by the global cognitive status measured by MoCA (R = -0.21, 95% CI [-0.35,-0.12]), impaired speed/attention domain (Symbol-Digit Matching) (R = -0.19, 95% CI [-0.35,-0.11]) and memory deficit (HVLT) (R = -0.13, 95% CI [-0.28,-0.05]). CSF concentration of amyloid-beta (R = -0.17, 95% CI [-0.31,-0.03]) and severity of dopaminergic denervation (SBR) (R = -0.15, 95% CI [-0.32,-0.06]) were the only biomarkers that significantly contributed to LV-I, while genetic risk score and α-synuclein failed to reach significance levels.

To ensure the disease specificity of the findings, the PLS analysis was repeated after removing the effect of aging from the atrophy maps, by regressing out age effects calculated based on the healthy subjects in the same dataset (N=117). This analysis was performed similar to previous studies with confounding age effects in diseased populations (Scahill et al. 2003; Franke et al. 2010; Dukart et al. 2011; Moradi et al. 2015). The results remained significant after controlling for normative aging (12% of covariance explained, p-value < 0.0001). Overall, the directionality and significant contributors of the LV-I pattern remained stable after regressing out the effect of age.

3.3. An emerging atrophy network in de novo PD patients

The corresponding brain pattern for the clinical and demographic measures in LV-I involved discrete cortical regions located in multiple parts of the frontal lobes, fusiform gyrus, cingulate gyrus and insular cortex, and subcortical regions including thalamus and basal ganglia (putamen, caudate, and nucleus accumbens), hippocampus and amygdala, brainstem (substantia nigra, red nucleus, subthalamic nucleus, pons, and areas of medulla that overlap with the dorsal motor nucleus of the vagus and nucleus of the solitary tract), and cerebellum. (Table 2, Figure 3.a.)

Figure 3c shows an example of how the putative atrophy network and the associated clinical phenotype relate to each other. For each weighted pattern we estimated patient-specific scores by projecting the patterns onto individual patients' data (see *Methods*). The resulting scalar values (termed atrophy scores and clinical scores), reflect the extent to which an individual patient expresses each pattern. By definition, the two scores are correlated ($r = 0.7$), i.e. patients with greater atrophy in the network in Fig. 3a, also tend to conform more closely to the clinical phenotype in Fig. 3b. Patients who score highly on both therefore have more severe pathology, and we illustrate this by coloring the points (individual patients) by their UPDRS III scores. Individuals with more pronounced atrophy and clinical variable severity also tend to score highly on UPDRS III, a measure of motor symptoms.

Table 2. Peak coordinates in MNI-ICBM152 space for brain PLS scores using bootstrap ratios. PLS = Partial Least Squares. B-ratio= Bootstrap ratio. Structures are ordered within regions by z coordinate.

Region	B-ratios	Structure	MNI coordinates
Brainstem	4.9	Medulla	-6,-42,-58
	3.6	Pons	-3,-30,-47
	4.7/4.5	Substantia Nigra	9,-14,-13/-7,-14,-13
	5.1/4.4	Subthalamic nucleus	8,16,-10/-8,-16,-10
Cerebellum	5.6/4.8	Cerebellum	32,-64,-34
	4.8	Cerebellum	-26,-64,-32
Subcortical	4.2/4	Hippocampus	23,-9,-25/-22,-9,-26
	4.1/4.2	Amygdala	20,-4,-23/-25,-4,-24
	5.6/5.1	Nucleus Accumbens	10,12,-6/-8,12,-10
	5.7/6.0	Globus Pallidus Internal Segment	22,-6,-4/-22,-8,-4
	5.4/4.7	Putamen	24,12,-6/-24,12,-4
	8.6/7.7	Ventrolateral/Ventroposterior Thalamus	12,-26,-2/-10,-24,-2
	7.3/4.7	Caudate	10,14,-2/-10,14,3
Cortical	3.2	Fusiform gyrus	22,8,-48
	3.7	Medial temporopolar region	-22,10,-44
	3.4	Medial/Inferior frontal gyrus	50,8,-38
	3.2/3.3	Anterior/medial Orbital Gyrus	23,49,-14/-23,43,-18
	6.4/5.8	Periaqueductal Gray	6,-32,-12/-4,-32,-12
	5.3/2.8	Fusiform gyrus	26,-66,-8/-24,-64,-9
	3.7	Fusiform gyrus	-22,-66,-4
	5.0	Inferior Frontal gyrus	-30,32,6
	3.4	Medial frontal gyrus	44,52,12
	3.6	Lateral occipital cortex	-24,-78,20
	3.7	Parietal Operculum	70,-30,22
	2.6	Cingulate gyrus	11,22,31
	2.8	Middle Frontal gyrus	24,33,32
	3	Superior Frontal gyrus	26,-4,68
	2.7	Cingulate gyrus	-8,-26,76
	4.8	Cerebellum	-26,-64,-32

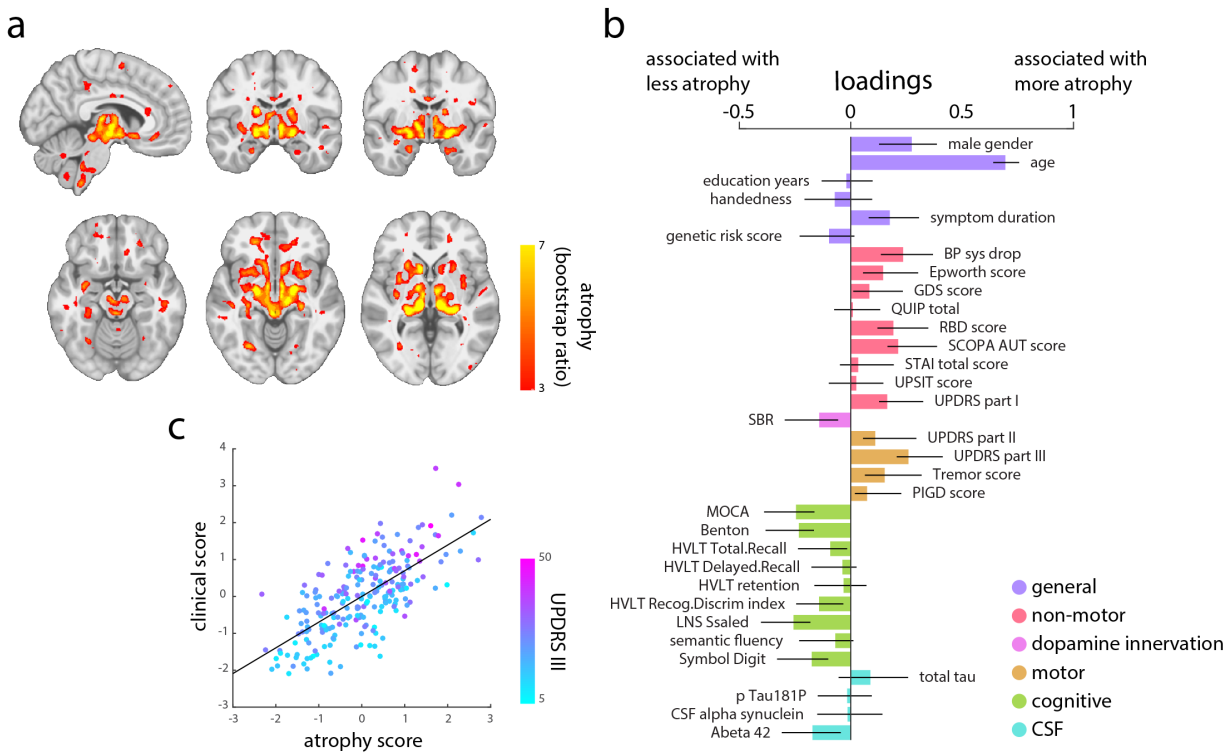

Figure 3. First latent variable (LV-I) obtained from PLS analysis. a) Brain pattern bootstrap ratios in MNI-ICBM152 space (x= -6, y = -12, y= -8, z= -14, z= -6, z= 0) b) Clinical scores pattern (the effect size estimates are derived from SVD analysis and the Confidence Intervals (CI) are calculated by bootstrapping, hence the CI are not necessarily symmetrical), c) individual subjects' Brain versus Clinical PLS score. PLS= Partial Least Squares. SVD= Singular Value Decomposition.

3.4. Atrophy pattern predicts longitudinal disease progression

Baseline LV-I score was significantly related to longitudinal changes in several measures after an average of 2.7 years. Participants with more severe atrophy in the LV-I brain pattern at baseline had significantly greater deterioration in the GCO ($r = 0.22$, $p < 0.001$). Longitudinal decline in activities of daily living, measured by the SE-ADL score (which was not included in the PLS analysis) was significantly greater in those with greater expression of the Brain LV-I pattern ($r = -0.20$, $p=0.003$). We also assessed the correlation between LV-I score at baseline and changes in single clinical measures in different categories. Higher expression of the Brain

LV-I pattern was significantly correlated with decline in cognition demonstrated by decrease in MoCA score (r = -0.28, p<0.0001). However, the association between baseline LV-I expression and changes in motor signs (UPDRS-III) (r = 0.13, p = 0.052) or non-motor symptoms (UPDRS-I) (r = 0.12, p=0.08) marginally failed to reach significance.

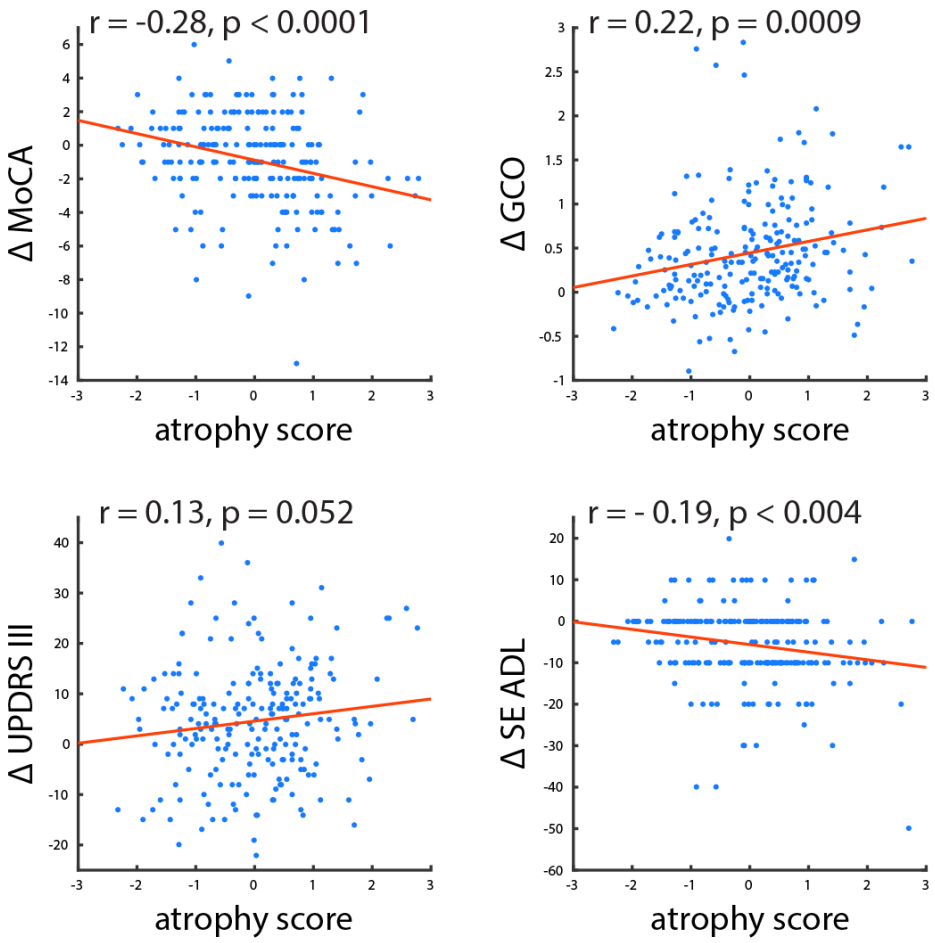

Figure 4. Baseline atrophy is associated with longitudinal clinical progression. Individual patients' atrophy score (expression of the atrophy network from the PLS model) is correlated with longitudinal changes in clinical measures of disease severity. PLS= Partial Least Squares. MoCA= Montreal Cognitive Assessment. GCO= Global Composite Outcome. UPDRS= Unified Parkinson's Disease Rating Scale. SE ADL= Schwab and England ADL score (overall activities of daily living).

4. Discussion

The present study links multiple domains of clinical and biomarker features of PD to the underlying brain atrophy pattern using a single integrated analysis in a recently diagnosed population. In this de novo cohort, in addition to higher age, a wide range of motor and non-motor features were linked to brain atrophy. We hope the PLS approach used here provides a means to investigate the complex combination of motor and non-motor features of PD in relation to patterns of brain atrophy.

Our findings suggest that a broadly distributed spatial pattern of brain atrophy is present in the early stages of PD, which covaries with motor, early cognitive and other non-motor manifestations. This is somewhat at odds with the previous literature, where de novo PD is seldom associated with detectable brain atrophy. The participants of this cohort were all drug-naïve within less than one year of diagnosis. A possible explanation for the greater ability of the multivariate approach to detect atrophy is that the course of PD may be stereotyped and the disease relatively widespread by the time early motor symptoms appear (Braak et al. 2003). Using all the voxels in the brain in a single analysis may confer greater sensitivity to deformation in a disease with a consistent spatial distribution.

PD studies using brain imaging to date have almost always focussed on differences between PD and healthy controls, or on a particular symptom manifestation (such as dementia) to study brain alterations. As a multivariate approach, PLS enables us to investigate brain alterations in PD subjects without a need for a control group and to consider multiple clinical aspects of the disease simultaneously.

We used our standard image analysis pipeline to calculate DBM as a measure of brain alterations. This pipeline (Aubert-Broche et al. 2013) has been previously used for multiple

multi-center and multi-scanner studies and it has been shown to produce robust results by removing site-specific biases (Zeighami et al. 2015; Boucetta et al. 2016; Sanford et al. 2017). Also, in an earlier study, we provided evidence that DBM was a more sensitive measure of atrophy than VBM, especially for subcortical areas (Zeighami et al. 2015).

While the presence of atrophy early in the course of the disease is rarely reported, the direction of associations between atrophy and different clinical features and biomarkers is consistent with the literature. As one notable example, older age of onset and male gender were associated with more atrophy in the PD-related pattern that was later demonstrated to correlate with faster progression. This is in line with previous findings on poorer prognosis of PD in older male patients (Post et al., 2011). Also, key non-motor features such as RBD, somnolence, autonomic disturbance and mood disorders also contributed to the latent variable, consistent with the prognostic importance of these key manifestations shown in other PD cohorts (Fereshtehnejad et al. 2015). Cognitive deficit, even though mild in severity, was also a significant correlate of brain atrophy. Although definite cognitive impairment was an exclusion criterion in PPMI, mild cognitive impairment still significantly correlated with the pattern of atrophy. Up to one fifth of the early PD populations meet the criteria for mild cognitive impairment, which is a strong predictor of earlier onset of dementia and poor prognosis (Pedersen et al. 2013, 2017). It is noteworthy that visuospatial and executive functioning more prominently contributed to the pattern of atrophy than the other cognitive domains. This is consistent with other studies of cognitive impairment in PD compared to Alzheimer's disease (Watson and Leverenz 2010; Wu et al. 2012).

Using PLS, we obtained a disease related atrophy map that included brainstem (more specifically, medulla in the area of the dorsal motor nucleus of the vagus, red nucleus and

substantia nigra), basal ganglia (including putamen, caudate, pallidum and subthalamic nucleus), cortical regions, as well as cerebellar regions. These findings are consistent with the earlier stages of Braak's description of disease spread (Braak et al. 2003), as well as our previously published PD atrophy network map, based on this dataset (Zeighami et al. 2015). It is notable that atrophy was also identified in frontal regions, belonging to Braak Stage V (Braak et al. 2003), and not usually thought to be affected at the time of diagnosis.

One of the main strengths of the proposed approach is the ability to detect brain-clinical manifestations of the disease at an early stage. We further show that the PLS scores relate to disease progression in the follow up visits. These results provide an opportunity to develop a simple comprehensive measure per subject which can be used as a prognostic biomarker of the disease. This approach could also have value in assessing prodromal disease populations, identified through genetic testing or the presence of RBD. We suggest that it could also be applicable to other neurodegenerative or neurodevelopmental diseases.

The findings from this study should be considered in light of some limitations. Using PLS provides the opportunity to comprehensively investigate brain-clinical relations. However, we lose specificity as to how each particular clinical manifestation potentially relates to a specific brain region, rather than the atrophy pattern as a whole. Such individual relationships need to be addressed in future studies using independent PD cohorts. While we investigated the relationship between baseline findings and longitudinal clinical changes, future studies also need to investigate longitudinal brain alterations in PD and how they relate to disease progression.

In this study, we have taken advantage of PLS as a multivariate approach to investigate the collective relationship between brain alterations reflected in DBM measures and various aspects

of the disease reflected in clinical measurements. We used data consisting of people with early diagnosed, drug naïve PD who were followed for an average of 2.7 years from PPMI, a global multi-center study. Clinically speaking, 2.7 years is a relatively short-term follow-up in the course of PD. Yet, the atrophy pattern was significantly associated with the longitudinal rate of decline in several clinical measures. In other words, high-scoring participants with more atrophic patterns at baseline experienced faster progression on the global single indicator of all symptom categories as well as the cognitive measure. Taken together, this study provides a new framework for studying neurodegenerative diseases with multi-faceted clinical measures and the interactions between brain alterations and disease manifestations. In addition, the single collective score summarizing the disease burden for each individual subject can be used as a potential biomarker for both diagnostic and prognostic purposes.

Acknowledgement:

This research was supported by the Canadian Institutes for Health Research, Natural Sciences and Engineering Research Council of Canada, Michael J Fox Foundation, Weston Brain Institute, and the Alzheimer's Association. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including AbbVie, Avid, Biogen, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi Genzyme, Servier, Teva, and UCB.

References:

Abdi, Hervé, and Lynne J. Williams. 2010. "Principal Component Analysis." *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4): 433–59. doi:10.1002/wics.101.

Ashburner, J., C. Good, and K. J. Friston. 2000. "Tensor Based Morphometry." *NeuroImage* 11 (5 PART I). http://discovery.ucl.ac.uk/1500695/.

Aubert-Broche, B., V. S. Fonov, D. García-Lorenzo, A. Mouiha, N. Guizard, P. Coupé, S. F. Eskildsen, and D. L. Collins. 2013. "A New Method for Structural Volume Analysis of Longitudinal Brain MRI Data and Its Application in Studying the Growth Trajectories of Anatomical Brain Structures in Childhood." *NeuroImage* 82 (November): 393–402. doi:10.1016/j.neuroimage.2013.05.065.

Boucetta, Soufiane, Ali Salimi, Mahsa Dadar, Barbara E. Jones, D. Louis Collins, and Thien Thanh Dang-Vu. 2016. "Structural Brain Alterations Associated with Rapid Eye Movement Sleep Behavior Disorder in Parkinson's Disease." *Scientific Reports* 6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4887790/.

Braak, Heiko, Kelly Del Tredici, Udo Rüb, Rob A. I. de Vos, Ernst N. H. Jansen Steur, and Eva Braak. 2003. "Staging of Brain Pathology Related to Sporadic Parkinson's Disease." *Neurobiology of Aging* 24 (2): 197–211.

Cardenas, Valerie A., Adam L. Boxer, Linda L. Chao, Maria L. Gorno-Tempini, Bruce L. Miller, Michael W. Weiner, and Colin Studholme. 2007. "Deformation-Based Morphometry Reveals Brain Atrophy in Frontotemporal Dementia." *Archives of Neurology* 64 (6): 873–77. doi:10.1001/archneur.64.6.873.

Chaudhuri, K Ray, Daniel G Healy, and Anthony HV Schapira. 2006. "Non-Motor Symptoms of Parkinson's Disease: Diagnosis and Management." *The Lancet Neurology* 5 (3): 235–45. doi:10.1016/S1474-4422(06)70373-8.

Chung, M. K., K. J. Worsley, T. Paus, C. Cherif, D. L. Collins, J. N. Giedd, J. L. Rapoport, and A. C. Evans. 2001. "A Unified Statistical Approach to Deformation-Based Morphometry." *NeuroImage* 14 (3): 595–606. doi:10.1006/nimg.2001.0862.

Collins, D. L., and A. C. Evans. 1997. "Animal: Validation and Applications of Nonlinear Registration-Based Segmentation." *International Journal of Pattern Recognition and Artificial Intelligence* 11 (08): 1271–94. doi:10.1142/S0218001497000597.

Collins, D. L., P. Neelin, T. M. Peters, and A. C. Evans. 1994. "Automatic 3D Intersubject Registration of MR Volumetric Data in Standardized Talairach Space." *Journal of Computer Assisted Tomography* 18 (2): 192–205.

Coupe, P., P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. 2008. "An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images." *IEEE Transactions on Medical Imaging* 27 (4): 425–41. doi:10.1109/TMI.2007.906087.

Dukart, Juergen, Matthias L. Schroeter, Karsten Mueller, and The Alzheimer's Disease Neuroimaging Initiative. 2011. "Age Correction in Dementia – Matching to a Healthy Brain." *PLOS ONE* 6 (7): e22193. doi:10.1371/journal.pone.0022193.

Eckart, Carl, and Gale Young. 1936. "The Approximation of One Matrix by Another of Lower Rank." *Psychometrika* 1 (3): 211–218.

Fereshtehnejad, Seyed-Mohammad, Silvia Rios Romenets, Julius BM Anang, Véronique Latreille, Jean-François Gagnon, and Ronald B. Postuma. 2015. "New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison with Other Phenotypes." *JAMA Neurology* 72 (8): 863–873.

Fereshtehnejad, Seyed-Mohammad, Yashar Zeighami, Alain Dagher, and Ronald B. Postuma. 2017. "Clinical Criteria for Subtyping Parkinson's Disease: Biomarkers and Longitudinal Progression." *Brain*. https://academic.oup.com/brain/article/doi/10.1093/brain/awx118/3855005/Clinical-criteria-for-subtyping-Parkinson-s.

Franke, Katja, Gabriel Ziegler, Stefan Klöppel, and Christian Gaser. 2010. "Estimating the Age of Healthy Subjects from T1-Weighted MRI Scans Using Kernel Methods: Exploring the Influence of Various Parameters." *NeuroImage* 50 (3): 883–92. doi:10.1016/j.neuroimage.2010.01.005.

Goedert, Michel, Maria Grazia Spillantini, Kelly Del Tredici, and Heiko Braak. 2013. "100 Years of Lewy Pathology." *Nature Reviews Neurology* 9 (1): 13–24. doi:10.1038/nrneurol.2012.242.

Halliday, Glenda Margaret, and Heather McCann. 2010. "The Progression of Pathology in Parkinson's Disease." *Annals of the New York Academy of Sciences* 1184 (1): 188–95. doi:10.1111/j.1749-6632.2009.05118.x.

Heim, Beatrice, Florian Krismer, Roberto De Marzi, and Klaus Seppi. 2017. "Magnetic Resonance Imaging for the Diagnosis of Parkinson's Disease." *Journal of Neural Transmission (Vienna, Austria: 1996)*, April. doi:10.1007/s00702-017-1717-8.

Hely, Mariese A., John G. L. Morris, Wayne G. J. Reid, and Robert Trafficante. 2005. "Sydney Multicenter Study of Parkinson's Disease: Non-L-Dopa-Responsive Problems Dominate at 15 Years." *Movement Disorders: Official Journal of the Movement Disorder Society* 20 (2): 190–99. doi:10.1002/mds.20324.

Kalia, Lorraine V., and Anthony E. Lang. 2015. "Parkinson's Disease." *The Lancet* 386 (9996): 896–912. doi:10.1016/S0140-6736(14)61393-3.

Leow, Alex D., Andrea D. Klunder, Clifford R. Jack, Arthur W. Toga, Anders M. Dale, Matt A. Bernstein, Paula J. Britson, et al. 2006. "Longitudinal Stability of MRI for Mapping Brain Change Using Tensor-Based Morphometry." *NeuroImage* 31 (2): 627–40. doi:10.1016/j.neuroimage.2005.12.013.

Marek, Kenneth, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, et al. 2011. "The Parkinson Progression Marker Initiative (PPMI)." *Progress in Neurobiology*, Biological Markers for Neurodegenerative Diseases, 95 (4): 629–35. doi:10.1016/j.pneurobio.2011.09.005.

McIntosh, Anthony R., and Bratislav Mišić. 2013. "Multivariate Statistical Analyses for Neuroimaging Data." *Annual Review of Psychology* 64 (1): 499–525. doi:10.1146/annurev-psych-113011-143804.

McIntosh, Anthony Randal, and Nancy J. Lobaugh. 2004. "Partial Least Squares Analysis of Neuroimaging Data: Applications and Advances." *NeuroImage* 23 Suppl 1: S250-263. doi:10.1016/j.neuroimage.2004.07.020.

Moradi, Elaheh, Antonietta Pepe, Christian Gaser, Heikki Huttunen, and Jussi Tohka. 2015. "Machine Learning Framework for Early MRI-Based Alzheimer's Conversion Prediction in MCI Subjects." *NeuroImage* 104 (January): 398–412. doi:10.1016/j.neuroimage.2014.10.002.

Nalls, Mike A., Cory Y. McLean, Jacqueline Rick, Shirley Eberly, Samantha J. Hutten, Katrina Gwinn, Margaret Sutherland, et al. 2015. "Diagnosis of Parkinson's Disease on the Basis of Clinical and Genetic Classification: A Population-Based Modelling Study." *The Lancet Neurology* 14 (10): 1002–9. doi:10.1016/S1474-4422(15)00178-7.

Pedersen, Kenn Freddy, Jan Petter Larsen, Ole-Bjorn Tysnes, and Guido Alves. 2013. "Prognosis of Mild Cognitive Impairment in Early Parkinson Disease: The Norwegian ParkWest Study." *JAMA Neurology* 70 (5): 580–586.

Pedersen, Kenn Freddy, Jan Petter Larsen, Ole-Bjørn Tysnes, and Guido Alves. 2017. "Natural Course of Mild Cognitive Impairment in Parkinson Disease: A 5-Year Population-Based Study." *Neurology* 88 (8): 767–74. doi:10.1212/WNL.0000000000003634.

Penny, William D., Karl J. Friston, John T. Ashburner, Stefan J. Kiebel, and Thomas E. Nichols. 2011. *Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images*. Academic Press.

Poewe, W. 2008. "Non-Motor Symptoms in Parkinson's Disease." *European Journal of Neurology* 15 (April): 14–20. doi:10.1111/j.1468-1331.2008.02056.x.

Poewe, Werner, Klaus Seppi, Caroline M. Tanner, Glenda M. Halliday, Patrik Brundin, Jens Volkmann, Anette-Eleonore Schrag, and Anthony E. Lang. 2017. "Parkinson Disease." *Nature Reviews Disease Primers* 3 (March): 17013. doi:10.1038/nrdp.2017.13.

Politis, Marios. 2014. "Neuroimaging in Parkinson Disease: From Research Setting to Clinical Practice." *Nature Reviews. Neurology* 10 (12): 708–22. doi:10.1038/nrneurol.2014.205.

Postuma, Ronald B., Daniela Berg, Charles H. Adler, Bastiaan R. Bloem, Piu Chan, Günther Deuschl, Thomas Gasser, et al. 2016. "The New Definition and Diagnostic Criteria of Parkinson's Disease." *The Lancet. Neurology* 15 (6): 546–48. doi:10.1016/S1474-4422(16)00116-2.

Sanford, Ryan, Ana Lucia Fernandez Cruz, Susan C. Scott, Nancy E. Mayo, Lesley K. Fellows, Beau M. Ances, and D. Louis Collins. 2017. "Regionally Specific Brain Volumetric and Cortical Thickness Changes in HIV-Infected Patients in the HAART Era:" *JAIDS Journal of Acquired Immune Deficiency Syndromes* 74 (5): 563–70. doi:10.1097/QAI.0000000000001294.

Scahill, Rachael I., Chris Frost, Rhian Jenkins, Jennifer L. Whitwell, Martin N. Rossor, and Nick C. Fox. 2003. "A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging." *Archives of Neurology* 60 (7): 989–94. doi:10.1001/archneur.60.7.989.

Sled, John G., Alex P. Zijdenbos, and Alan C. Evans. 1998. "A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data." *Medical Imaging, IEEE Transactions on* 17 (1): 87–97.

Studholme, C., V. Cardenas, R. Blumenfeld, N. Schuff, H. J. Rosen, B. Miller, and M. Weiner. 2004. "Deformation Tensor Morphometry of Semantic Dementia with Quantitative Validation." *NeuroImage* 21 (4): 1387–98. doi:10.1016/j.neuroimage.2003.12.009.

Tuite, Paul, and Alain Dagher. 2013. *Magnetic Resonance Imaging in Movement Disorders: A Guide for Clinicians and Scientists*. Cambridge University Press.

Watson, G. Stennis, and James B. Leverenz. 2010. "Profile of Cognitive Impairment in Parkinson's Disease." *Brain Pathology (Zurich, Switzerland)* 20 (3): 640–45. doi:10.1111/j.1750-3639.2010.00373.x.

Wold, Herman. 1966. "Estimation of Principal Components and Related Models by Iterative Least Squares." In *Multivariate Analysis.*, 391–420. Academic Press.

Wu, Qi, Ling Chen, Yifan Zheng, Caixia Zhang, Linhuan Huang, Wenyuan Guo, Yingying Fang, et al. 2012. "Cognitive Impairment Is Common in Parkinson's Disease without Dementia in the Early and Middle Stages in a Han Chinese Cohort." *Parkinsonism & Related Disorders* 18 (2): 161–65. doi:10.1016/j.parkreldis.2011.09.009.

Yousaf, Tayyabah, Heather Wilson, and Marios Politis. 2017. "Imaging the Nonmotor Symptoms in Parkinson's Disease." *International Review of Neurobiology*. http://www.sciencedirect.com/science/article/pii/S0074774217300454.

Zeighami, Yashar, Miguel Ulla, Yasser Iturria-Medina, Mahsa Dadar, Yu Zhang, Kevin Michel-Herve Larcher, Vladimir Fonov, Alan C. Evans, D. Louis Collins, and Alain Dagher. 2015. "Network Structure of Brain Atrophy in de Novo Parkinson's Disease." *Elife* 4: e08440.